

Lab assignment 8 [Hoff 7.4. p.239]

Marriage data: The file `agehw.dat` contains data on the ages of 100 married couples sampled from the U.S. population.

(<http://www.stat.washington.edu/hoff/Book/Data/hwdata/>)

- a) Before you look at the data, use your own knowledge to formulate a semiconjugate prior distribution for $\boldsymbol{\theta} = (\theta_h, \theta_w)^T$ and Σ , where θ_h , θ_w are mean husband and wife ages, and Σ is the covariance matrix.
- b) Generate a *prior predictive dataset* of size $n = 100$, by sampling $(\boldsymbol{\theta}, \Sigma)$ from your prior distribution and then simulating $\mathbf{Y}_1, \dots, \mathbf{Y}_n \sim \text{i.i.d. multivariate normal}(\boldsymbol{\theta}, \Sigma)$. Generate several such datasets, make bivariate scatterplots for each dataset, and make sure they roughly represent your prior beliefs about what such a dataset would actually look like. If your prior predictive datasets do not conform to your beliefs, go back to part a) and formulate a new prior. Report the prior that you eventually decide upon, and provide scatterplots for at least three prior predictive datasets.
- c) Using your prior distribution and the 100 values in the dataset, obtain an MCMC approximation to $p(\boldsymbol{\theta}, \Sigma | \mathbf{y}_1, \dots, \mathbf{y}_{100})$. Plot the joint posterior distribution of θ_h and θ_w , and also the marginal posterior density of the correlation between \mathbf{Y}_h and \mathbf{Y}_w , the ages of a husband and wife. Obtain 95% posterior confidence intervals for θ_h , θ_w and the correlation coefficient.
- d) Obtain 95% posterior confidence intervals for θ_h , θ_w and the correlation coefficient using a diffuse prior with $\mu_0 = \mathbf{0}$, $\Lambda_0 = 10^5 \times \mathbf{I}$, $\mathbf{S}_0 = 1000 \times \mathbf{I}$ and $\nu_0 = 3$. Compare the confidence intervals to those obtained in c).
- e) Estimate $P[\theta_h > \theta_w | \mathbf{y}_1, \dots, \mathbf{y}_{100}]$ using the prior in d).