# STA 602 HW 9

Ryan Tang

November 4th 2022

## 1  Exercise 7.1

**(a) Jeffrey Prior on the Joint Parameters**   The Jeffrey prior is $p_J(\theta, \Sigma) \propto |\Sigma|^{-(p+2)/2}$. It is not a proper probability density function because it doesn't integrate to 1 and doesn't ensure $\Sigma$ to be positive definite, and symmetric. The only possible way is when the Inverse-Wishart distribution's $\nu_o > p - 1$, where $p$ is the dimension of $X \in \mathbb{R}^p$ in our observations. But the Jeffrey prior implies $\nu_o = 0$, which doesn't guarantee a proper $\Sigma$. Well, we can use it as long as the sample size exceeds the $p - 1$ threshold.

**(b) Full conditionals**

$$Y_i \overset{iid}{\sim} \mathcal{N}(\theta, \Sigma) \qquad\qquad Y_i \in \mathbb{R}^p$$

$$p(Y|\theta, \Sigma) = (2\pi)^{-pN/2} |\Sigma|^{-N/2} \exp\left[-\frac{1}{2}\sum_i^N (y_i - \theta)^\mathsf{T}\Sigma^{-1}(y_i - \theta)\right]$$

$$p_J(\theta, \Sigma) \propto |\Sigma|^{-(p+2)/2}$$
$$p_J(\theta, \Sigma|Y) \propto p_J(\theta, \Sigma)p(Y|\theta, \Sigma)$$

$$\propto |\Sigma|^{-N/2}|\Sigma|^{-(p+2)/2}\exp\left[-\frac{1}{2}\sum_i^N (y_i - \theta)^\mathsf{T}\Sigma^{-1}(y_i - \theta)\right]$$

$$\propto |\Sigma|^{-(N+p+2)/2}\exp\left[-\frac{1}{2}\sum_i^N (y_i - \theta)^\mathsf{T}\Sigma^{-1}(y_i - \theta)\right]$$

$$p_J(\theta|Y, \Sigma) \propto \exp\left[-\frac{1}{2}(\theta^\mathsf{T}(N\Sigma^{-1})\theta + \theta^\mathsf{T}(N\Sigma^{-1}\bar{y}))\right]$$
$$\sim \mathcal{N}(\theta|(N\Sigma^{-1})^{-1}N\Sigma^{-1}\bar{Y}, N\Sigma^{-1})$$

$$p_J(\Sigma|Y, \theta) \propto |\Sigma|^{-(N+p+2)/2}\exp\left[-\frac{1}{2}\sum_i^N (y_i - \theta)^\mathsf{T}\Sigma^{-1}(y_i - \theta)\right]$$

$$\propto |\Sigma|^{-(N+1+p+1)/2}\exp\left[-\frac{1}{2}tr(S\Sigma^{-1})\right]$$

$$\sim \text{inverse-Wishart}(\Sigma|N + 1, S)$$

$$S = \sum_i^N (y_i - \theta)^\mathsf{T}(y_i - \theta)$$

## 2 Exercise 7.2

**(a) Multivariate Normal MLE & Unit Information Priors**   Here we show that the joint unit information priors consist of two proper conjugate priors for the multivariate Gaussian model. The parameterization of the two priors is just the mean $\bar{y}$ and the unit scatter matrix $\frac{1}{N}S_{\bar{y}} = \frac{1}{N}\sum_i^N (y_i - \bar{y})(y_i - \bar{y})^{\mathsf{T}}$ from MLE.

$$\Lambda = \Sigma^{-1}$$

$$p(X|\theta,\Sigma) = (2\pi)^{-pN/2}|\Lambda|^{N/2}\exp\left[-\frac{1}{2}\sum_i^N (y_i - \theta)^{\mathsf{T}}\Lambda(y_i - \theta)\right]$$

$$\ell(\theta,\Sigma) = \frac{N}{2}\log|\Lambda| - \frac{1}{2}\sum_i^N (y_i - \theta)^{\mathsf{T}}\Lambda(y_i - \theta)$$

$$p_U(\theta,\Lambda) \propto \exp[\frac{1}{N}\ell(\theta,\Lambda)]$$

$$\propto \exp\left[\frac{1}{2}\log|\Lambda| - \frac{1}{2N}\sum_i^N (y_i - \theta)^{\mathsf{T}}\Lambda(y_i - \theta)\right]$$

$$\propto |\Lambda|^{\frac{1}{2}}\exp\left[-\frac{1}{2N}\sum_i^N (y_i - \bar{y} + \bar{y} + \theta)^{\mathsf{T}}\Lambda(y_i - \bar{y} + \bar{y} + \theta)\right]$$

$$\propto |\Lambda|^{\frac{1}{2}}\exp\left[-\frac{1}{2N}tr[(\sum_i^N (y_i - \bar{y})(y_i - \bar{y})^{\mathsf{T}} + \sum_i^N (\bar{y} - \theta)(\bar{y} - \theta)^{\mathsf{T}} + \sum_i^N (y_i - \bar{y})(\bar{y} - \theta)^{\mathsf{T}})\Lambda]\right]$$

$$\sum_i^N (y_i - \bar{y})(\bar{y} - \theta)^{\mathsf{T}} = 0$$

$$\propto |\Lambda|^{\frac{1}{2}}\exp\left[-\frac{1}{2}tr[(\frac{1}{N}\sum_i^N (y_i - \bar{y})(y_i - \bar{y})^{\mathsf{T}} + \frac{1}{N}\sum_i^N (\bar{y} - \theta)(\bar{y} - \theta)^{\mathsf{T}})\Lambda]\right]$$

$$\propto |\Lambda|^{\frac{1}{2}}\exp\left[-\frac{1}{2}tr[\frac{1}{N}\sum_i^N (y_i - \bar{y})(y_i - \bar{y})^{\mathsf{T}}\Lambda + (\bar{y} - \theta)(\bar{y} - \theta)^{\mathsf{T}}\Lambda]\right]$$

$$\propto |\Lambda|^{\frac{1}{2}}\exp\left[-\frac{1}{2}tr[\frac{1}{N}S_{\bar{y}}\Lambda] - \frac{1}{2}(\bar{y} - \theta)^{\mathsf{T}}\Lambda(\bar{y} - \theta)\right]$$

$$\frac{1}{N}S_{\bar{y}} = \frac{1}{N}\sum_i^N (y_i - \bar{y})(y_i - \bar{y})^{\mathsf{T}}$$

$$\propto |\Lambda|^{\frac{p+1-p-1}{2}}\exp\left[-\frac{1}{2}tr[\frac{1}{N}S_{\bar{y}}\Lambda]\right]|\Lambda|^{\frac{1}{2}}\exp\left[-\frac{1}{2}(\bar{y} - \theta)^{\mathsf{T}}\Lambda(\bar{y} - \theta)\right]$$

$$= Wi(\Lambda|p+1, \frac{1}{N}S_{\bar{y}})\mathcal{N}(\theta|\bar{y}, \Lambda^{-1})$$

$$= p(\theta|\Lambda)p(\Lambda)$$

$$\theta|\Lambda \sim \mathcal{N}(\bar{y}, \Lambda^{-1})$$

$$\Lambda \sim Wi(p+1, \frac{1}{N}S_{\bar{y}})$$

**(b) Unit Information Posterior**  If we use MLE's unit information priors, the posterior with the same data happens to be the same form with the effective sample size increased with $N$.

$$p_U(\theta, \Lambda | Y) \propto p_U(\Lambda) p_U(\theta|\Lambda) p(Y|\theta, \Lambda)$$

$$\propto |\Lambda|^{\frac{p+1-p-1}{2}} \exp\left[-\frac{1}{2} tr[\frac{1}{N} S_{\bar{y}}\Lambda]\right] |\Lambda|^{\frac{1}{2}} \exp\left[-\frac{1}{2}[\theta^\intercal \Lambda \theta - \theta^\intercal \Lambda \bar{y}]\right]$$

$$\cdot |\Lambda|^{N/2} \exp\left[-\frac{1}{2}[tr(S_{\bar{y}}\Lambda) + \theta^\intercal N\Lambda\theta - \theta^\intercal N\Lambda\bar{y}]\right]$$

$$\propto |\Lambda|^{\frac{N+p+1-p-1}{2}} \exp\left[-\frac{1}{2} tr[\frac{N+1}{N} S_{\bar{y}}\Lambda]\right] |\Lambda|^{\frac{1}{2}} \exp\left[-\frac{1}{2}[\theta^\intercal(\Lambda + N\Lambda)\theta - \theta^\intercal(\Lambda + N\Lambda)\bar{y}]\right]$$

$$\propto Wi(N + p + 1, \frac{N+1}{N} S_{\bar{y}}) \mathcal{N}((\Lambda + N\Lambda)^{-1}(\Lambda + N\Lambda)\bar{y}, (\Lambda + N\Lambda)^{-1})$$

# 3   Exercise 7.3

**(a) Posterior**  Since we have a known form in semi-conjugate, we can directly sample from it.

$$\theta \sim \mathcal{N}(\mu_o, \Sigma_o = \Lambda^{-1})$$
$$\theta|Y, \Sigma \sim \mathcal{N}(\mu_n = \Sigma_n(\Lambda_o\mu_o + N\Lambda y), \Sigma_n = (\Lambda_o + N\Lambda)^{-1})$$
$$\Sigma \sim IW(\nu_o, S_o)$$
$$\Sigma|Y \sim IW(v_o + n, (S_0 + S_\theta)^{-1}$$
$$S_\theta = \sum_i^N (y_i - \theta)(y_i - \theta)^\intercal$$

**(b) $\theta$ posteriors**  After the Bayesian updates, we can see Blue crabs generally have smaller body widths and rear widths than orange crabs.
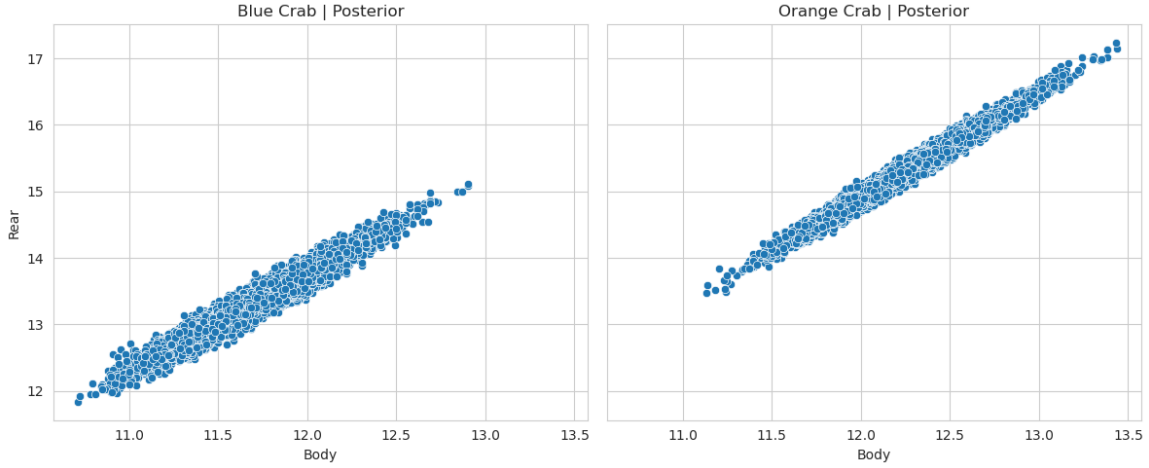


Figure 1: $\theta$ Posteriors Comparison of two Crabs

**(c) Correlation on $\Sigma$ posteriors** Lastly, we plot the posterior correlation between the two crab species. We can see that the Orange crabs' widths are quite consistent between the body width and the rear width. The Blue crabs are relatively less consistent. $P[\rho_{blue} < \rho_{orange}|Y] = 0.9915$.
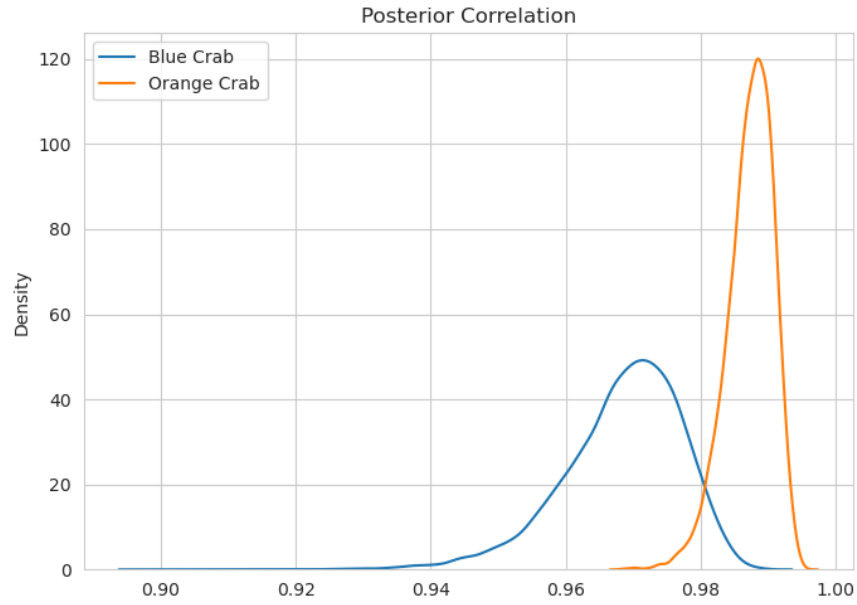


Figure 2: Correlation Posteriors Comparison of two Crabs.