Section 2: Bayesian inference in Gaussian models

2.1 Bayesian inference in a simple Gaussian model

Let's start with a simple, one-dimensional Gaussian example, where

$$y_i|\mu, \sigma^2 \sim \mathrm{N}(\mu, \sigma^2).$$

We will assume that μ and σ are unknown, and will put conjugate priors on them both, so that

$$\begin{aligned} \sigma^2 \sim & \text{Inv-Gamma}(\alpha_0, \beta_0) \\ \mu|\sigma^2 \sim & \text{Normal}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right) \end{aligned}$$

or, equivalently,

$$\begin{aligned} y_i|\mu, \omega \sim & \mathrm{N}(\mu, 1/\omega) \\ \omega \sim & \text{Gamma}(\alpha_0, \beta_0) \\ \mu|\omega \sim & \text{Normal}\left(\mu_0, \frac{1}{\omega\kappa_0}\right) \end{aligned}$$

We refer to this as a normal/inverse gamma prior on μ and σ^2 (or a normal/gamma prior on μ and ω). We will now explore the posterior distributions on μ and $\omega(/\sigma^2)$ – much of this will involve similar results to those obtained in the first set of exercises.

Exercise 2.1 *Derive the conditional posterior distributions $p(\mu, \omega|y_1, \ldots, y_n)$ (or $p(\mu, \sigma^2|y_1, \ldots, y_n)$) and show that it is in the same family as $p(\mu, \omega)$. What are the updated parameters α_n, β_n, μ_n and κ_n?*

Solution:
Likelihood is following:

$$p(y_i|\mu, \omega) = (\frac{\omega}{2\pi})^{\frac{n}{2}} \exp\left(- \frac{\omega}{2} \sum_{i=1}^{n} (y_i - \mu)^2 \right) = (\frac{\omega}{2\pi})^{\frac{n}{2}} \exp\left(- \frac{\omega}{2} [n(\mu - \overline{y})^2 + \sum_{i=1}^{n} (y_i - \overline{y})^2] \right)$$

Therefore, we need to find the prior $p(\mu, \omega)$ Since $p(\mu, \omega) = p(\mu|\omega)p(\omega)$
we have the prior on μ and the prior on ω. Thus, the conjugate prior on μ, ω is following:

$$p(\mu, \omega|\mu_0, k_0, \alpha_o, \beta_0) = N\left(\mu_0, \frac{1}{k_0\omega}\right) * Gamma(\omega|\alpha_0, \beta_0)$$

$$p(\mu|\omega) = \left(\frac{wk_0}{2\pi}\right)^{\frac{1}{2}} exp\left[- \frac{\omega k_0}{2} (\mu - \mu_0)^2 \right]$$

$$p(\omega) = \frac{\beta_0^\alpha}{\Gamma(\alpha)}\omega^\alpha - 1exp(-\beta_0\omega)$$

we obtained,

$$p(\mu,\omega) = \left(\frac{\omega k_0}{2\pi}\right)^{\frac{1}{2}} exp\left[-\frac{\omega k_0}{2}(\mu-\overline{\mu})^2\right]\frac{\beta_0^a}{\Gamma(\alpha_0)}\omega^{\alpha_0-1}exp(-\beta_0\omega) = \left(\frac{k_0}{2\pi}\right)^{\frac{1}{2}}\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\omega^{\alpha_0-\frac{1}{2}}exp\left[-\frac{\omega}{2}(k_0(\mu-\overline{\mu})^2+2\beta_0)\right]$$
$$\propto NG(\mu_0, k_0, \alpha_0, \beta 0)$$

Through, we could estimate the posterior

$$p(\mu,\omega|y_i) \propto p(y_i|\mu,\omega)p(\mu,\omega) \propto \omega^{\frac{n}{2}} exp\left[-\frac{\omega}{2}\sum_{i=1}^{n}(y_i-\mu)^2\right]\omega^{\alpha_0-\frac{1}{2}}exp\left[-\frac{\omega}{2}[k_0(\mu-\overline{\mu})^2+2\beta_0]\right]$$

Exercise 2.2 *Derive the conditional posterior distribution* $p(\mu|\omega, y_1, \ldots, y_n)$ *and* $p(\omega|y_1, \ldots, y_n)$ *(or if you'd prefer,* $p(\mu|\sigma^2, y_1, \ldots, y_n)$ *and* $p(\sigma^2|y_1, \ldots, y_n)$*). Based on this and the previous exercise, what are reasonable interpretations for the parameters* $\mu_0, \kappa_0, \alpha_0$ *and* β_0*?*

Solution: We can utilize the Normal-gamma posterior that ultimately gives a normal distribution:

$$p(\mu|\omega,y) \propto exp\left[-\frac{\omega}{2}(n+k_0)\left(\mu-\frac{n\overline{y}+k_0\mu_0}{n+k_0}\right)^2 = N\left[\frac{k_0\mu_0+n\overline{y}}{k_0+n}, \omega(n+k_0)\right] = N(\mu_n, \omega_n)\right.$$

Then, marginalize out μ, we could obtain following for μ

$$p(\omega|y) = \int p(\omega,\mu|y)d\mu = Gamma(\alpha_n, \beta_n)$$

Exercise 2.3 *Show that the marginal distribution over* μ *is a centered, scaled t-distribution (note we showed something very similar in the last set of exercises!), i.e.*

$$p(\mu) \propto \left(1+\frac{1}{\nu}\frac{(\mu-m)^2}{s^2}\right)^{-\frac{\nu+1}{2}}$$

What are the location parameter m*, scale parameter* s*, and degree of freedom* ν*?*

$$p(\mu) \propto \int_0^\infty p(\mu,\omega)d\omega$$

when,

$$p(\mu,\omega) = \left(\frac{k_0}{2\pi}\right)^{\frac{1}{2}}\frac{\beta_0^{\alpha 0}}{\Gamma(\alpha_0)}\omega^{\alpha_0-\frac{1}{2}}exp\left[-\frac{\omega}{2}[k_0(\mu-\overline{\mu})^2+2\beta_0]\right]$$
$$p(\mu) \propto \int_0^\infty \omega^{\alpha_0+\frac{1}{2}-1}exp\left[-\frac{\omega}{2}[k_0(\mu-\overline{\mu})^2+2\beta_0]\right]dw$$

So, the Kernel is Gamma$\left(\alpha_0+\frac{1}{2},\beta_0+\frac{k_0(\mu-\mu_0)^2}{2}\right)$. Therefore, we have the normalization constant that distributed by the following:

$$p(\mu)\propto\frac{\Gamma(\alpha_0+\frac{1}{2})}{[\beta_0+\frac{k_0(\mu-\mu_0)^2}{2}]^{a_0+\frac{1}{2}}}\propto\left[\beta_0+\frac{k_0(\mu-\mu_0)^2}{2}\right]^{-a_0-\frac{1}{2}}\propto\left[1+\frac{1}{2a_0}\frac{a_0k_0(\mu-\mu_0)^2}{\beta_0}\right]^{-\frac{(2a_0+1)}{2}}$$

Thus, $m=\mu_0$, $v=2\alpha_0$, $s^2=\frac{b_0}{a_0k_0}$

Exercise 2.4 *The marginal posterior $p(\mu|y_1,\ldots,y_n)$ is also a centered, scaled t-distribution. Find the updated location, scale and degrees of freedom.*

Solution:

$$p(\mu|y_i)\propto\int_0^\infty p(\mu,\omega|y)i)dw$$

$$p(\mu)\propto\frac{\Gamma(a)}{\beta^a}\propto\beta^{-a}=\left[\frac{k_0}{2}(\mu-\mu_0)^2+\beta_0\right]^{-a_0-\frac{1}{2}}=\left[1+\frac{1}{2a_0}\frac{a_0k_0(\mu-\mu_0)^2}{b_0}\right]^{-a_0-\frac{1}{2}}$$

Since, $a=a_0+\frac{1}{2}$, $\beta=\frac{k0}{2}(\mu-\mu_0)^2+\beta_0$, under the kernel of gamma distribution.
Therefore, scaled t distribution with $m=\mu_0$, $v=2\alpha_0$, $s^2=\frac{b_0}{a_0k_0}$, which is same with exercise 2-3

Exercise 2.5 *Derive the posterior predictive distribution $p(y_{n+1},\ldots,y_{n+m}|y_1,\ldots,y_m)$.*

Solution:
Posterior predictive is,

$$p(y_{newi}|y_i)=\frac{p(_{ynew_i},yi)}{p(y_i)}$$

The first step to derive the posterior predictive distribution is to find $p(y_i)$, the marginal distribution of y_i.

$$p(\mu,\omega|y_i)=\frac{p(y_i|\mu,\omega)p(\mu,\omega)}{p(y_i)}$$

Likelihood $=p(y_i|\mu,\omega)=\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\prod_i^n N(y_i|\mu,\omega)$, as prior is $p(\mu,\omega)=\frac{1}{z0}NG(\mu_0,k_0,a_0,\beta_0)$. when Z_0 is a normalizing constant. Thus, posterior is $p(\mu,\omega|y_i)=\frac{1}{Z_n}NG(\mu_n,K_n,a_n,\beta_n)$
Thus, the posterior predictive is following:

$$p(y_{newi}|y_i)=\frac{p(y_{newi},y_i)}{p(y_i)}=\frac{Z_{n+m}}{z_0}\left(\frac{1}{2\pi}\right)^{\frac{n+m}{2}}\frac{Z_0}{Z_n}(2\pi)^{\frac{n}{2}}=\frac{Z_{n+m}}{Z_n}\left(\frac{1}{2\pi}\right)^{\frac{m}{2}}$$

$$p(y_{newi}|y_i)=\frac{\Gamma(\alpha_{n+m})}{\Gamma(\alpha_n)}\frac{B_n^{a_n}}{B_{n+m}^{a_{n+m}}}\left(\frac{k_n}{k_nn+m}\right)^{\frac{1}{2}}\left(\frac{1}{2\pi}\right)^{\frac{m}{2}}$$

Thus, we have the followings: $a_{n+1}=a_n+\frac{1}{2},\beta_{n+1}=\beta_n+\frac{1}{2}\frac{k_n(y-\mu_n)^2}{k_n+1},k_{n+1}=k_n+1$

Ultimately, we obtained T-distribution where, $m=\mu_n,v=2a_n,s^2=\frac{\beta_n(k_n+1)}{a_nk_n}$

Exercise 2.6 *Derive the marginal distribution over $y_1,\ldots,y_n$.*

Solution:

$$p(\mu, \omega|y) = \frac{p(y, \mu, \omega}{p(y)}$$

$$p(y) = \frac{p(y, \mu, \omega)}{p(\mu, \omega|y)}$$

2.2 Bayesian inference in a multivariate Gaussian model

Let's now assume that each y_i is a d-dimensional vector, such that

$$y_i \sim \mathrm{N}(\mu, \Sigma)$$

for d-dimensional mean vector μ and $d \times d$ covariance matrix Σ.

We will put an *inverse Wishart* prior on Σ. The inverse Wishart distribution is a distribution over positive-definite matrices parametrized by $\nu_0 > d - 1$ degrees of freedom and positive definite matrix Λ_0^{-1}, with pdf

$$p(\Sigma|\nu_0, \Lambda_0^{-1}) = \frac{|\Lambda|^{\nu_0/2}}{2^{(\nu_0+d)/2}\Gamma_d(\nu_0/2)}|\Sigma|^{-\frac{\nu_0+d+1}{2}}e^{-\frac{1}{2}\mathrm{tr}(Lambda\Sigma^{-1})}$$

where

$\Gamma_d(x) = \pi^{d(d-1)/4}\prod_{i=1}^{d}\Gamma\left(x - \frac{j-1}{2}\right)$.

Exercise 2.7 *Show that in the univariate case, the inverse Wishart distribution reduces to the inverse gamma distribution.*

Exercise 2.8 *Let $\Sigma \sim$ Inv-Wishart(ν_0, Λ_0^{-1}) and $\mu|\Sigma \sim N(\mu_0, \Sigma/\kappa_0)$, so that*

$$p(\mu, \Sigma) \propto |\Sigma|^{-\frac{\nu_0+d+2}{2}}e^{-\frac{1}{2}tr(\Lambda_0\Sigma^{-1})-\frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)}$$

and let

$$y_i \sim N(\mu, \Sigma)$$

Show that $p(\mu, \Sigma|y_1, \ldots, y_n)$ is also normal-inverse Wishart distributed, and give the form of the updated parameters μ_n, κ_n, ν_n and Λ_n. It will be helpful to note that

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu) =& \sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{k=1}^{d}(x_{ij}-\mu_j)(\Sigma^{-1})_{jk}(x_{ik}-\mu_k)\\
=& \sum_{j=1}^{d}\sum_{k=1}^{d}(\Sigma^{-1})_{ab}\sum_{i=1}^{n}(x_{ij}-\mu_j)(x_{ik}-\mu_k)\\
=& tr\left(\Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T\right)
\end{aligned}
$$

Based on this, give interpretations for the prior parameters.

2.3 A Gaussian linear model

Lets now add in covariates, so that

$$\mathbf{y}|\beta, X \sim \text{Normal}(X\beta, (\omega\Lambda)^{-1})$$

where $\mathbf{y}$ is a vector of n responses; X is a $n \times d$ matrix of covariates; and Λ is a known positive definite matrix.

Let's assume $\beta \sim \text{Normal}(\mu, (\omega K)^{-1})$ and $\omega \sim \text{Gamma}(a, b)$, where K is assumed fixed.

Exercise 2.9 *Derive the conditional posterior* $p(\beta|\omega, y_1, \dots, y_n)$

Exercise 2.10 *Derive the marginal posterior* $p(\omega|y_1, \dots, y_n)$

Exercise 2.11 *Derive the marginal posterior,* $p(\beta|y_1, \dots, y_n)$

Exercise 2.12 *Download the dataset dental.csv from Github. This dataset measures a dental distance (specifically, the distance between the center of the pituitary to the pterygomaxillary fissure) in 27 children. Add a column of ones to correspond to the intercept. Fit the above Bayesian model to the dataset, using* $\Lambda = I$ *and* $K = I$*, and picking vague priors for the hyperparameters, and plot the resulting fit. How does it compare to the frequentist LS and ridge regression results?*

2.4 A hierarchical Gaussian linear model

The dental dataset has heavier tailed residuals than we would expect under a Gaussian model. We've seen previously that we can model a scaled t-distribution using a scale mixture of Gaussians; let's put that into effect here. Concretely, let

$$\begin{aligned}
\mathbf{y}|\beta, \omega, \Lambda \sim& \text{N}(X\beta, (\omega\Lambda)^{-1})\\
\Lambda =& \text{diag}(\lambda_1, \dots, \lambda_n)\\
\lambda_i \stackrel{iid}{\sim}& \text{Gamma}(\tau, \tau)\\
\beta|\omega \sim& \text{N}(\mu, (\omega K)^{-1})\\
\omega \sim& \text{Gamma}(a, b)
\end{aligned}$$

Exercise 2.13 *What is the conditional posterior,* $p(\lambda_i|\mathbf{y}, \beta, \omega)$*?*

Exercise 2.14 *Write a Gibbs sampler that alternates between sampling from the conditional posteriors of* λ_i*,* β *and* ω*, and run it for a couple of thousand samplers to fit the model to the dental dataset.*

Exercise 2.15 *Compare the two fits. Does the new fit capture everything we would like? What assumptions is it making? In particular, look at the fit for just male and just female subjects. Suggest ways in which we could modify the model, and for at least one of the suggestions, write an updated Gibbs sampler and run it on your model.*