Homework 3

Instructor: Hedibert Freitas Lopes
Course: STP 598 Advanced Bayesian Statistical Learning (Class # 31199)
Semester: Spring 2022
Due date: 1:30pm, March 21st, 2022.

Posterior inference via Gibbs sampler

Here is the data ($n = 50$ observations)

```
y = c(-2.991,-2.845,-2.640,-1.962,-1.646,-1.627,-1.185,-0.997,-0.917,-0.726,-0.630,
-0.382,-0.335,-0.300,-0.216,-0.178,-0.136,-0.013,0.185,0.255,0.367,0.422,0.440,
0.447,0.454,0.539,0.558,0.568,0.587,0.696,0.716,0.832,0.861,0.869,0.898,1.012,
1.073,1.081,1.112,1.159,1.332,1.342,1.402,1.560,1.728,1.785,2.364,2.491,2.558,3.575)
```

Let us model the data as Student's t with known ν degrees of freedom, but unknown location, μ, and unknown scale, σ^2. More precisely,

$$y_i|\mu, \sigma^2 \sim t_\nu(\mu, \sigma^2),$$

for $i = 1, \ldots, n$. Also, let us assume that μ and σ^2 are, *a priori*, independent,

$$\mu \sim N(m_0, C_0) \quad \text{and} \quad \sigma^2 \sim IG(a_0, b_0),$$

for known hyperparameters (m_0, C_0, a_0, b_0). Assume that data $= \{y_1, \ldots, y_n\}$.

Scale mixture of Gaussians. The Student's t likelihood is not easy to handle, but there is an important probability result that links the Student's t_ν with the Gaussian and the inverse-gamma distributions. More precisely, the Student's t is a scale mixture of Gaussian distributions:

$$\text{If } y_i|\lambda_i, \mu, \sigma^2 \sim N(\mu, \lambda_i\sigma^2) \quad \text{and} \quad \lambda_i \sim IG(\nu/2, \nu/2),$$

then $y_i|\mu, \sigma^2 \sim t_\nu(\mu, \sigma^2)$. This result allows us to reinterpret the model hierarchically:

$$\begin{aligned} y_i|\lambda_i &\sim N(\mu, \lambda_i\sigma^2), \qquad i = 1, \ldots, n \\ \lambda_i &\sim IG(\nu/2, \nu/2), \qquad i = 1, \ldots, n \\ (\mu, \sigma^2) &\sim N(m_0, C_0)IG(a_0, b_0). \end{aligned}$$

The (Monte Carlo) price one pays is in the additional parameters, (μ, σ^2, λ), where $\lambda = (\lambda_1, \ldots, \lambda_n)$.

1. Let $z|\omega \sim N(0, \omega)$ and $\omega|\nu \sim IG(\nu/2, \nu/2)$, for $\nu > 0$.
   Show that $z|\nu \sim t_\nu(0, 1)$.
   Hint: $p(z|\nu) = \int_0^\infty p(z|\omega)p(\omega|\nu)d\omega$.

2. Derive the full conditional of μ, i.e. $p(\mu|\text{data}, \sigma^2, \lambda)$

3. Derive the full conditional of σ^2, i.e. $p(\sigma^2|\text{data}, \mu, \lambda)$

4. Derive the full conditional of λ_i, i.e. $p(\lambda_i|\text{data}, \mu, \sigma^2, \lambda_{-i})$, for $i = 1, \ldots, n$ and λ_{-i} is λ without the λ_i.

5. For $\nu = 5$ and hyperparameters $m_0 = 0$, $C_0 = 1$, $a_0 = 5$ and $b_0 = 1$, perform posterior inference for (μ, σ^2), based on a Gibbs sampler that cycles through the above $n + 2$ full conditionals. Report your code/implementation and your posterior summaries and findings.

6. Let us incorporate ν into the analysis with the following discrete prior

$$Pr(\nu) \propto \nu e^{-0.1\nu} \qquad \nu = 1, 2, \ldots, 100.$$

   Derive its full conditional distribution $Pr(\nu|\text{data}, \mu, \sigma^2)$ (easy, since it is a discrete distribution), and revisit question 5 above.

7. Simulate yourself $n = 50$ from the standard Gaussian. Assuming the previous modeling structure (prior and likelihood), plot the posterior distribution of ν. Does it concentrate on large values of ν, which would indiate approximate Gaussian behavior? Comment your findings.

8. What would change in your Gibbs sampler if μ, the location parameter, is replaced by a linear preditor

$$\beta_0 + \beta_1 x_{i1} + \cdots + x_{iq},$$

   for known and q fixed predictors x_{ij}, for $i = 1, \ldots, n$ and $j = 1, \ldots, q$, and prior

$$\beta = (\beta_0, \beta_1, \ldots, \beta_q) \sim N_q(m_0, V_0),$$

   for known hyperparameters b_0, a q-dimensional vector and B_0, a $q \times q$ covariance matrix?