Bayesian Statistics : Assignment 2

Instructor : Anna Simoni

November 14, 2023

This assignment is part of a project to validate the course. Failure in returning one or more assignments implies failure of the class.

Instructions :

1). Students have to work in **groups of 3 students** (same groups as for the Assignment 1). Only one document per group has to be returned.

2). The due date is **December 5, 2023** (assignment returned after this day will not be considered and the whole project will not be validated).

3). Each group has to upload the document containing the solution to the assignment on `https://app.compilatio.net/v5/document-submission/J23-E55-778` in the pdf format (no other format will be considered).

4). The pdf has to be named in the following way : `FAMILYNAME1_FAMILYNAME2_FAMILYNAME3_Ass_2`.

5). The solution has to be typed (not handwritten).

6). For each question, please provide all the computational details.

7). Question 2 has to be developed with the statistical software that you prefer among Matlab, R or Python.

We consider the same model as in the Assignment 1 : for $t = 1, \ldots, T$,

$$y_t = u_t'\phi + x_t'\beta + \varepsilon_t, \qquad \varepsilon_t \sim^{i.i.d.} \mathcal{N}(0, \sigma^2), \tag{0.1}$$

where x_t is a vector of regressors of dimension k and u_t is a vector of regressors of dimension ℓ. For simplicity, the variance of each regressor is normalized to 1.

The prior for $\theta := (\sigma^2, \phi', \beta') \in \mathbb{R}^{1+\ell+k}$ is specified as in the Assignment 1 : $\pi(\theta) = \pi(\sigma^2) \times \pi(\phi) \times \pi(\beta|\sigma^2, \gamma^2, q)$ with

$$\begin{aligned} \pi(\sigma^2) &= \frac{1}{\sigma^2} \\ \pi(\phi) &= \text{flat} \\ \pi(\beta_i|\sigma^2, \gamma^2, q) &\sim^{i.i.d.} \begin{cases} \mathcal{N}(0, \sigma^2\gamma^2) & \text{with probability } q \\ 0 & \text{with probability } 1-q \end{cases} \qquad \text{for } i = 1, \ldots, k \end{aligned}$$

and γ^2, q are hyperparameters taking respectively nonnegative values and values in $[0, 1]$.

In addition, we specify the prior for the hyperparameters q and γ^2 in the following way. First, we specify the mapping $(\gamma^2, q) \mapsto R^2(\gamma^2, q) := \frac{qk\gamma^2\overline{v}_x}{qk\gamma^2\overline{v}_x+1}$, where $\overline{v}_x := \frac{1}{k}\sum_{j=1}^k \widehat{Var(x_{t,j})}$ is the average sample variance of the predictors. Then, the prior for q and R^2 is

$$\begin{aligned} \pi(q) &\sim Beta(a, b) \\ \pi(R^2) &\sim Beta(A, B). \end{aligned}$$

Introduce a latent variable z_i that takes the value 1 if the corresponding covariate x_i is included in the model (that is, is active) and its coefficient β_i is non-zero, and it takes the value 0 otherwise. The random variable z_i is $Bernoulli(q)$, where q denotes the probability of success. Let $\widetilde{\beta}$ be the vector of the non-zero coefficients (i.e. those associated with $z_i = 1$) and $\widetilde{X}$ be the corresponding regressors. We can recover β from $z := (z_1, \ldots, z_k)'$ and $\widetilde{\beta}$ in the following way : $\beta_i = z_i\widetilde{\beta}_{\bar{z}_i}$, where $\bar{z}_i = \sum_{\ell \leq i} z_\ell$.

Question 1. Show that $\pi(\beta|\sigma^2, R^2, q)$ (which is $\prod_{i=1}^k [q\mathcal{N}(0, \sigma^2\gamma^2) + (1-q)\delta_0(\beta_i)]$, where $\delta_0(\cdot)$ denotes a Dirac in 0) is equal to

$$\prod_{i=1}^k q^{z_i}(1-q)^{1-z_i} \prod_{i=1}^k \left[\mathcal{N}(0, \sigma^2\gamma^2)\right]^{z_i} \left[\delta_0(\beta_i)\right]^{1-z_i}.$$

Question 2. Here you are asked to simulate from the joint posterior distribution. To do so, you have to implement the following Gibbs sampling with blocks (where we denote $z = (z_1, \dots, z_k)'$, $Y = (y_1, \dots, y_T)'$, $U = [u_1, \dots, u_T]'$ (U is $T \times \ell$ matrix) and $X = [x_1, \dots, x_T]'$ (X is $T \times k$ matrix)). The blocks are (R^2, q), ϕ and (β, σ^2) :

(I). Draw from the conditional posterior of (R^2, q) given by :

$$\pi(R^2, q|Y, U, X, \theta, z) \propto \left[e^{-\frac{1}{2\sigma^2}\frac{k\bar{v}_x q(1-R^2)}{R^2}\beta' diag(z)\beta} \right] \\ \times q^{s(z)+\frac{s(z)}{2}+a-1}(1-q)^{k-s(z)+b-1}(R^2)^{A-1-s(z)/2}(1-R^2)^{s(z)/2+B-1},$$

where $s(z) := \sum_{i=1}^k z_i$. To draw from this distributions you have to sample from a discrete approximation of this distribution. More specifically, let us discretize the support of R^2 and q by interlacing two grids defined over the unit interval as follows : the grid for q is $[0 : .001 : .1\ .11 : .01 : .9\ .901 : .001 : 1]$ and for R^2 is $[0 : .001 : .1\ .11 : .01 : .9\ .901 : .001 : 1]$ (each grid has increments of 0.01, and finer increments of 0.001 near the boundaries).

(II). Draw from the conditional posterior of ϕ given by :

$$\pi(\phi|Y, U, X, z, \beta, R^2, q, \sigma^2) = \pi(\phi|Y, U, X, \beta, \sigma^2) \\ = \mathcal{N}\left((U'U)^{-1}U'(Y - X\beta), \sigma^2(U'U)^{-1}\right).$$

(III). To sample from the conditional posterior of z observe that

$$\pi(z|Y, U, X, \phi, R^2, q) \propto q^{s(z)}(1-q)^{k-s(z)} \left(\frac{1}{\gamma^2}\right)^{\frac{s(z)}{2}} |\widetilde{W}|^{-1/2} \left[\frac{\widetilde{Y}'\widetilde{Y} - \widehat{\widetilde{\beta}}'\widetilde{W}\widehat{\widetilde{\beta}}}{2}\right]^{-T/2} \Gamma\left(\frac{T}{2}\right),$$

where $\widehat{\widetilde{\beta}} := \widetilde{W}^{-1}\widetilde{X}'\widetilde{Y}$, $\widetilde{W} := (\widetilde{X}'\widetilde{X} + I_{s(z)}/\gamma^2)$, and $\widetilde{Y} := Y - U\phi$. Therefore, to sample from the posterior of $z|Y, U, X, \phi, R^2, q$ we can draw iteratively from the distribution of $z_i|Y, U, X, \phi, R^2, q, z_{-i}$, $i = 1, \dots, k$, using a Gibbs sampling.

(IV). Draw from the conditional posterior of σ^2 given by :

$$\pi(\sigma^2|Y,U,X,\phi,R^2,q,z) = I\Gamma\left(\frac{T}{2}, \frac{\widetilde{Y}'\widetilde{Y} - \widehat{\widetilde{\beta}}'(\widetilde{X}'\widetilde{X} + I_{s(z)}/\gamma^2)\widehat{\widetilde{\beta}}}{2}\right).$$

(V.) Let $\widetilde{\beta}$ be the subvector of β containing only the non-zero coefficients (notice that we know the identity of the non-zero coefficients thanks to the vector z). Draw from the conditional posterior of $\widetilde{\beta}$ given by :

$$\pi(\widetilde{\beta}|Y,U,X,\phi,R^2,q,\sigma^2,z)$$
$$= \mathcal{N}\left(\left(I_{s(z)}/\gamma^2 + \widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'(Y-U\phi), \sigma^2\left(I_{s(z)}/\gamma^2 + \widetilde{X}'\widetilde{X}\right)^{-1}\right).$$

where $\widetilde{X}$ are the regressors corresponding to $\widetilde{\beta}$ and the other β_i's are equal to 0.

You will repeat these steps for 110 000 iterations and you will eliminate the first 10000 iterations (burn-in period).

The data are generated according to model (0.1) with $\ell = 0$, $k = 100$, and $T = 200$. The predictors x_t are drawn from a zero-mean Normal distribution with a Toeplitz correlation matrix with $corr(x_{t,i}, x_{t,j}) = \rho^{|i-j|}$, where we set $\rho = 0.75$.

We set $k - s$ elements of β (chosen at random) at zero and draw the remaining s from independent $\mathcal{N}(0,1)$ distributions. The parameter s is set equal to 5, 10 and 100.

The error term ε_t is $i.i.d.\mathcal{N}(0,\sigma^2)$ with $\sigma^2 = (1/R_y - 1)\sum_{t=1}^{T}(\beta' x_t)^2/T$ and where $R_y = 2\%, 25\%, 50\%$ denotes the ratio between explained and total variance.

The values for the hyperparameters of the priors are : $a = b = A = B = 1$.

You have to :

1. Simulate 100 data sets, standardize the data.
2. For each value of $s \in \{5, 10, 100\}$ and of $R_y \in \{2\%, 25\%, 50\%\}$:
   (a) For each data set, compute the posterior median of q and plot an histogram (or kernel) approximation of the distribution of the posterior median of q across the 100 data sets.
   (b) For one data set, plot the marginal posterior distribution of q (by using the histogram).

(c) Comment and compare the results for the different values of s and R_y by recalling that q is the posterior of the probability of inclusion.