Block Gibbs Sampling for SLDS

March 29, 2023

Bayesian linear regression and VAR

A necessary premise for a Bayesian treatment of linear systems is Bayesian linear regression. In (multivariate) linear regression we consider the system

$$\mathbf{y}_t = A\mathbf{x}_t + \boldsymbol{\epsilon}_t$$

where $\mathbf{y}_t \in \mathbb{R}^m$, $\mathbf{x}_t \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $t = 1, \ldots, T$

$$\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, Q)$$

with $Q \in \mathbb{R}^{m\times m}$. Introducing matrix notations $\{\mathbf{y}_t\} \equiv Y$ with $Y \in \mathbb{R}^{m\times T}$ (with Y_{it} is the t-th observation of the i-th component of $\mathbf{y}$) and $\{\mathbf{x}_t\} \equiv X$ with $X \in \mathbb{R}^{n\times T}$ (with X_{it} is the t-th observation of the i-th component of x) and $\{\boldsymbol{\epsilon}_t\} \equiv E$ with $E \in \mathbb{R}^{m\times T}$

$$Y_{it} = \sum_j A_{ij}X_{jt} + E_{it} \qquad \text{or} \qquad Y = AX + E$$

We can assume that $\mathbf{y}_t, \mathbf{x}_t$ have zero mean, oterwise simply

$$\mathbf{y}_t \to \mathbf{y}_t - \frac{1}{T}\sum_t \mathbf{y}_t, \qquad \mathbf{x}_t \to \mathbf{x}_t - \frac{1}{T}\sum_t \mathbf{x}_t$$

Hence, we assume that all rows of matrices Y, X have zero mean. The likelihood is

$$f(X, Y|A, Q) = |Q|^{-T/2} exp\{-\frac{1}{2}Tr[(Y - AX)^T Q^{-1}(Y - AX)]\}$$

The max-likelihood solution for A is

$$\hat{A} = YX^T(XX^T)^{-1}$$

In the Bayesian scenario, we introduce a (conjugate) Matrix-normal inverse-Wishart prior for A, Q:

$$A, Q \sim MNIW(M, \Lambda, \Psi, \nu)$$

i.e.,

$$f(A,Q|M,\Lambda,\Psi,\nu) = |Q|^{-m/2}|\Lambda|^{-n/2} exp\Big(-\frac{1}{2}Tr\big[(A-M)^T Q^{-1}(A-M)\Lambda^{-1}\big]\Big)\times$$
$$\times|\Psi|^{\nu/2}|Q|^{-(m+\nu+1)/2} exp\Big(-\frac{1}{2}Tr\big[\Psi Q^{-1}\big]\Big)$$

It can be easily shown by some algebra that the posterior is also Matrix-normal inverse-Wishart with updated parameters

$$A,Q|Y,X \sim MNIW(M',\Lambda',\Psi',\nu')$$

with

$$M' = \hat{A}XX^T\Lambda' + M\Lambda^{-1}\Lambda'$$

$$\Lambda' = (\Lambda^{-1} + XX^T)^{-1}$$

$$\Psi' = \Psi + M\Lambda^{-1}M^T + YY^T - A'\Lambda'^{-1}A'^T$$
$$\nu' = \nu + T$$

In a vector-autoregressive process, we consider the system

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t$$

where $\mathbf{x}_t \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $t = 2,\dots,T$, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0,Q)$ with $Q \in \mathbb{R}^{m\times m}$. Introducing matrix notations $Y = \{\mathbf{x}_t\}$, $t = 2,\dots,T$ with $Y \in \mathbb{R}^{m\times T-1}$ and $X = \{\mathbf{x}_{t-1}\}$, $t = 2,\dots,T$ with $X \in \mathbb{R}^{n\times T-1}$and $\{\boldsymbol{\epsilon}_t\} \equiv E$ with $E \in \mathbb{R}^{m\times T-1}$

$$Y_{it} = \sum_j A_{ij}X_{j(t-1)} + E_{it} \qquad \text{or} \qquad Y = AX + E$$

We can assume that $\mathbf{y}_t, \mathbf{x}_t$ have zero mean. A, Q can be simply found with (Bayesian) linear regression.

Bayesian treatment of switching linear systems

The switching linear dynamical system model is a generative model for observed time series $\mathbf{x}_t$ where $\mathbf{x}_t \in \mathbb{R}^d$ and $t = 1,\dots,T$. The model assumes a latent categorical variable

$$z_t \in \{1,\dots,K\}$$

The latent variable follows a simple stochastic process, a (time-invariant) Markov chain

$$P(z_{t+1} = k|z_t = l) = \Pi_{lk}$$

where Π is the Markov transition matrix. Given z_t, The dynamics of $\mathbf{x}_t$ follows an (order 1) vector-autoregressive process

$$\mathbf{x}_{t+1} = A_{z_t}\mathbf{x}_t + \boldsymbol{\epsilon}$$

where A_k are $d\times d$ matrices (with $max(eig(A)) < 1$ for linear stability), and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, Q_{z_t})$ where Q_k are positive-definite matrices.

Priors and conditional posteriors

We use conjugate priors. For $\boldsymbol{\pi}_k$ (the k-th column of Π) we assume a Dirichlet prior,

$$\boldsymbol{\pi}_k \sim Dir(\boldsymbol{\alpha}_k)$$

The *conditional posterior* of $\boldsymbol{\pi}_k$ (all other parameters fixed) is

$$\boldsymbol{\pi}_k|all \sim Dir(\boldsymbol{\alpha}_k + \mathbf{n}_k)$$

where the matrix $n_{lk} = \#(z_{t+1} = k|z_t = l)$ has entries counting the number of observed state transitions from k to l, and $\mathbf{n}_k$ is its k-th row.

For A_k, Q_k, we assume matrix-normal inverse-wishart prior

$$A_k, Q_k \sim MNIW(M_k, \Lambda_k, \Psi_k, \nu)$$

Now, let

$$\mathbf{t}_k = \{t > 1 \mid z_t = k\}, \qquad |\mathbf{t}_k| = T_k$$

And consider the matrix $X^{(k)} = \{\mathbf{x}_{t-1}\}$ with $t \in \mathbf{t}_k$, $X \in \mathbb{R}^{n \times T_k}$; and $Y = \{\mathbf{x}_t\}$ with $t \in \mathbf{t}_k$, $Y \in \mathbb{R}^{n \times T_k}$. We assume that the rows of matrices $Y^{(k)}, X^{(k)}$ have zero mean (otherwise, subtract the mean of each row from the matrices). The *conditional posterior* A_k, Q_k (all other parameters fixed) is

$$A_k, Q_k|all \sim MNIW(M_k', \Lambda_k', \Psi_k', \nu_k')$$

with

$$M_k' = \hat{A}_k X^{(k)} X^{(k)T} \Lambda' + M_k \Lambda_k^{-1} \Lambda_k'$$

$$\Lambda_k' = (\Lambda_k^{-1} + X^{(k)} X^{(k)T})^{-1}$$

$$\Psi_k' = \Psi + M_k \Lambda_k^{-1} M_k^T + Y^{(k)} Y^{(k)T} - A_k' (\Lambda')^{-1} A_k'^{\,T}$$

$$\nu_k' = \nu_k + T_k$$

(here, $\hat{A}_k = Y^{(k)} X^{(k)T} (X^{(k)} X^{(k)T})^{-1}$.

The conditional posterior of z_t, given all other parameters, is

$$Prob(z_t = k) = \frac{r_{tk}}{\sum_k r_{tk}}$$

where

$$r_{tk} = \Pi_{Z_{t-1}k} |Q_k|^{-1/2} exp\{-\frac{1}{2}(\mathbf{x}_t - A_k \mathbf{x}_{t-1})^T Q_k^{-1} (\mathbf{x}_t - A_k \mathbf{x}_{t-1})^T]\}$$

Block Gibbs sampling

Block Gibbs sampling proceeds with as follows

Initialization. Randomly sample z_t uniformly. Determine $\mathbf{n}_k$, $X^{(k)}, Y^{(k)}$, $M'_k, \Lambda'_k, \Psi'_k, \nu'_k$. Randomly sample Π and A_k, Q_k from the priors.

Iterations

1. For each k, sample
$$\boldsymbol{\pi}_k \sim Dir(\boldsymbol{\alpha}_k + \mathbf{n}_k)$$

2. For each k, sample
$$A_k, Q_k \sim MNIW(M'_k, \Lambda'_k, \Psi'_k, \nu')$$

3. For each $t = 2, \ldots, T$, sample z_t from
$$Prob(z_t = k) = \frac{r_{tk}}{\sum_k r_{tk}}$$

4. Recompute $\mathbf{n}_k$, $X^{(k)}, Y^{(k)}$, $M'_k, \Lambda'_k, \Psi'_k, \nu'_k$.