Department of Statistics and Data Science at SUSTech

MAT7035: Computational Statistics

Tutorial 10: MCMC Methods (II): DA Algorithm and Gibbs Sampler

C. The Data Augmentation (DA) Algorithm

Let Y_{obs} denote the observations and $\boldsymbol{\theta}$ the unknown parameter vector.

(a) Goal: Based on the observed-data likelihood function $L(\boldsymbol{\theta}|Y_{\text{obs}})$ and the prior density $\pi(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$, we

— Find the observed-data posterior density $p(\boldsymbol{\theta}|Y_{\text{obs}})$;

— Generate posterior samples from $p(\boldsymbol{\theta}|Y_{\text{obs}})$.

(b) Method: Introduce a latent variable/vector Z such that both the complete-data posterior distribution

$$p_{(\boldsymbol{\theta}|Y_{\text{obs}},Z)}(\boldsymbol{\theta}|Y_{\text{obs}}, z^{(t+1)})$$

and the conditional predictive distribution

$$f_{(Z|Y_{\text{obs}},\boldsymbol{\theta})}(z|Y_{\text{obs}}, \boldsymbol{\theta}^{(t)})$$

are available.

(c) DA algorithm: Let $(\boldsymbol{\theta}^{(0)}, z^{(0)})$ be initial values:

— I-step: Draw $z^{(t+1)} \sim f_{(Z|Y_{\text{obs}},\boldsymbol{\theta})}(z|Y_{\text{obs}}, \boldsymbol{\theta}^{(t)})$;

— P-step: Draw $\boldsymbol{\theta}^{(t+1)} \sim p_{(\boldsymbol{\theta}|Y_{\text{obs}},Z)}(\boldsymbol{\theta}|Y_{\text{obs}}, z^{(t+1)})$.

D. The Gibbs Sampler

D.1 The formulation of the Gibbs sampling

Let $\mathbf{x} = (X_1, X_2, \dots, X_d)^\top$ be a random vector. Assume that the full conditional density of the i-th component is given by $f_i(x_i|x_1, , \dots, x_{i-1}, x_{i+1}, \dots, x_d) = f_i(x_i|\boldsymbol{x}_{-i})$.

(a) Goal: Simulate $\mathbf{x} = \boldsymbol{x}$ from the joint density $f(x_1, x_2, \dots, x_d) = f(\boldsymbol{x})$.

(b) Gibbs sampling: Set $t = 0$, and choose a starting value $\boldsymbol{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_d^{(0)})^\top$:

— **Draw** $X_1^{(t+1)} = x_1^{(t+1)} \sim f_1(x_1|x_2^{(t)}, \ldots, x_d^{(t)})$;

— **Draw** $X_2^{(t+1)} = x_2^{(t+1)} \sim f_2(x_2|x_1^{(t+1)}, x_3^{(t)}, \ldots, x_d^{(t)})$;

— $\cdots$

— **Draw** $X_d^{(t+1)} = x_d^{(t+1)} \sim f_d(x_d|x_1^{(t+1)}, \ldots, x_{d-1}^{(t+1)})$.

D.2 The two-block Gibbs sampling

(a) Rationale: Grouping (or blocking) highly correlated components together in the Gibbs sampler can greatly improve its efficiency.

(b) Method: We can split $(X_1, X_2, \ldots, X_d)^\top$ into

$$\mathbf{y}_1 = (X_1, X_2, \ldots, X_{d'})^\top \quad \text{and} \quad \mathbf{y}_2 = (X_{d'+1}, \ldots, X_d)^\top.$$

(c) two-block Gibbs sampling: Set $t = 0$, and choose $\boldsymbol{y}_1^{(0)} = (x_1^{(0)}, \ldots, x_{d'}^{(0)})^\top$ and $\boldsymbol{y}_2^{(0)} = (x_{d'+1}^{(0)}, \ldots, x_d^{(0)})^\top$:

— **Draw** $\mathbf{y}_1^{(t+1)} = \boldsymbol{y}_1^{(t+1)} \sim f_{(\mathbf{y}_1|\mathbf{y}_2)}(\boldsymbol{y}_1|\boldsymbol{y}_2^{(t)})$;

— **Draw** $\mathbf{y}_2^{(t+1)} = \boldsymbol{y}_2^{(t+1)} \sim f_{(\mathbf{y}_2|\mathbf{y}_1)}(\boldsymbol{y}_2|\boldsymbol{y}_1^{(t+1)})$.

Example T10.1 (Bivariate normal model). Using Gibbs Sampler to generate a random sample from bivariate normal distribution:

$$(X, Y)^\top \sim N_2\left(\mathbf{0}_2, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Remark: The density function of $(X, Y)^\top$ is given by

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2}{2} - \frac{(y-\rho x)^2}{2(1-\rho^2)}\right\}.$$

Solution: To draw $(X, Y)^\top \sim N_2\left(\mathbf{0}_2, \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)\right)$ using the Gibbs sampler, we first derive the full conditional distributions. The marginal density of $X \sim N(0, 1)$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

then we obtain the conditional density

$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{ -\frac{(y-\rho x)^2}{2(1-\rho^2)} \right\};$$

that is,

$$Y|(X = x) \sim N(\rho x, 1 - \rho^2).$$

By symmetry, we also have

$$X|(Y = y) \sim N(\rho y, 1 - \rho^2).$$

Hence, we can choose a starting vector $(X^{(0)}, Y^{(0)})$ and independently simulate

$$X^{(t+1)} \sim N(\rho Y^{(t)}, 1 - \rho^2), \qquad \text{and} \qquad Y^{(t+1)} \sim N(\rho X^{(t+1)}, 1 - \rho^2).$$

As $t \to +\infty$, the distribution of $(X^{(t)}, Y^{(t)})^\top$ converges to the desired bivariate normal density.

Finally, we finish our Gibbs sampler by examining the sample covariance matrix of the Gibbs sampler distribution. We choose $(X^{(0)}, Y^{(0)})^\top = (0, 0)^\top$ and the R code is as following:

R code:

```
> Gibbs <- function (n, rho)
{   # Function name: Gibbs
    # ------------ Input -------------
    # n   = the number of samples
    # rho = covariance
    # ------------ Onput -------------
    # C   = Sample Covariance Matrix
    #       of X and Y
    # --------------------------------
    mat <- matrix(ncol = 2, nrow = n)
    x <- 0;
    y <- 0;
    mat[1, ] <- c(x, y)
    for (i in 2:n){
        x<-rnorm(1,rho*y,sqrt(1-rho^2))
        y<-rnorm(1,rho*x,sqrt(1-rho^2))
        mat[i, ] <- c(x, y)
    }
    C = cov(mat);
    return (C)
}
```

Calculated result:

Choose $n = 200000$ and $\rho = 0.7$, the above code print the following output, which is the sample covariance matrix:

```
> Gibbs(200000,0.7)
          [,1]      [,2]
[1,] 0.9987326 0.6978068
[2,] 0.6978068 1.0025710
```