Online Updating for Multivariate Regression Model via Modified Cholesky decomposition

*Joochul Lee**
HeeSeung Kim†

15 December 2018

Abstract

This is the project for the class. We consider the multivariate regression model. Firstly, we consider the model via Modified Cholesky decomposition using Newton-Raphson algorithm. Using the model, we propose a way to update cumulative coefficient estimates.

Keywords: Multivariate Regression Model; Online updated Method; Newton-Raphson algorithm; Modified Cholesky decomposition;

1 Introduction

Online updating is one of the important statistical methods as big data arrives in streams. Standard statistics tool to analyze is challenging for handling big data efficiently. Schifano et al.(2016) developed iterative estimating method without data storage requirements for the linear models and estimating equations. In this project, we will consider the multivariate regression model for the online updating. Pourahmadi (1999) studied the maximum likelihood estimators of a generalised linear model for the covariance matrix. In the researcher's study, mean and covariance of response variables are estimated by using covariates. To estimates parameters corresponding to the mean and covariance, the Newton-Raphson algorithm is used. For the online updating in the multivariate regression model, two estimating equations will be considered; one is for the parameters of the mean, and the other is for the parameters of the covariance. We need to consider two estimating equations at the same time to get cumulative estimators. Two steps should be proposed to handle two equations. For the first step, we fix the parameters of the covariance and then estimated the parameters of the mean. Then, the estimated parameters of the mean are fixed to estimate the parameters of the covariance at the second step.

2 Model Equations

Consider the i.i.d sample $\{\boldsymbol{Y}_i, \boldsymbol{X}_i\}$ for $i = 1, 2, \ldots, n$ where $\mathbf{Y}_i$, $\mathbf{X}_i$ is m, p and dimensional vectors, respectively. The multivariate regression model is

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

where $\boldsymbol{\beta}$ is the p dimensional vector and $\boldsymbol{\varepsilon}_i \sim N(\boldsymbol{\mu}, \Sigma)$

*joochul.lee@uconn.edu; Ph.D. student at Department of Statistics, University of Connecticut.

†hee-seung.kim@uconn.edu; Ph.D. student at Department of Electrical & Computer Engineering, University of Connecticut.

2.1 MLE via Modified Cholesky decomposition

By the modified Cholesky decomposition, Σ^{-1} can be expressed by the following

$$\Sigma^{-1}(\boldsymbol{\phi}, \boldsymbol{\phi}) = C(\boldsymbol{\phi})' D(\boldsymbol{\sigma}) C(\boldsymbol{\phi})$$

. where $C(\boldsymbol{\phi})$ is lower triangular matrix with 1's as diagonal entries and $D(\boldsymbol{\sigma})$ is diagonal matrix. Pourahmadi (2000) drived Likelyhood funcion as follows.

$$\begin{aligned}
-2L(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}) &= n \log |\Sigma| + \sum_{i=1}^{n} (y_i - X_i \beta)' \Sigma^{-1} (y_i - X_i \beta) \\
&= n \sum_{i=1}^{p} log \sigma_t^2 + \sum_{i=1}^{n} (r_i - \mathbf{Z}(i)\gamma)' D^{-1} (r_i - \mathbf{Z}(i)\gamma)
\end{aligned}$$

where $\boldsymbol{Z}(i) = (z(i,1), \ldots, z(i,p))', z(i,t) = \sum_{j=1}^{t-1} r_{ij} z_{tj},\ x_i, z_t, z_{tj}$ are p, q, d dimensinal vectors, respectively.

For the score function and Fisher information,

$$U_1(\boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{i=1}^{n} \boldsymbol{X}_i' \Sigma^{-1} \boldsymbol{r}_i, \quad \boldsymbol{I}_{11} = \sum_{i=1}^{n} \boldsymbol{X}_i' \Sigma^{-1} \boldsymbol{X}_i$$

$$U_2(\boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{i=1}^{n} \boldsymbol{Z}'(i) D^{-1} (\boldsymbol{r}_i - \boldsymbol{Z}(i)\boldsymbol{\phi}), \quad \boldsymbol{I}_{22} = nW$$

where $W = \sum_{i=1}^{n} \sigma_i^{-2} W_t$, $W_t = \sum_{j=1}^{t-1} \sum_{k=1}^{t-1} \sigma_{jk} z_{tk} z_{tl}'$, and $\boldsymbol{r}_i = \boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} = (r_{1i}, \ldots, r_{mi})'$. $\boldsymbol{Z}(i) = (z(i,1), \ldots, z(i,m))'$ is the $m \times q$ matrix where $z(i,t) = \sum_{j=1}^{t-1} r_{ij} z_{tj}$. Given σ_i^2, $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ are obtained using Fisher scoring algorithm.

2.2 Special Case

We can only use one common of covariates to estimate $\boldsymbol{\beta}, \boldsymbol{\phi}$. Cho (2013) studied this case. They drived likelyhood funcion as following.

$$\begin{aligned}
L(\beta^*) &= \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma(X^i; \phi^*)|^{\frac{1}{2}}} exp(-\frac{1}{2}(Y^i - X_i\beta))^\top \Sigma(\tau(X^i; \phi^*)^{-1}(Y^i -))) \\
&= exp(\sum_{i=1}^{n}(-\frac{1}{2} tr[\Sigma(X^i; \phi^*)^{-1} S(Y^i, X^i; \beta)) - \frac{1}{2} log|\Sigma(X^i; \phi^*)^{-1}(\phi)| + \frac{m}{2} log 2\pi))))
\end{aligned}$$

where $S(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\beta})) = (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}_i)(\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}_i)'$ For the score function and Fisher information,

$$U_1^*(\boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{i=1}^{n} \boldsymbol{V}(\boldsymbol{X})_i' \Sigma^{-1} (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}_i), \quad \boldsymbol{I}_{11}^* = \sum_{i=1}^{n} \boldsymbol{V}(\boldsymbol{X})_i' \Sigma^{-1} \boldsymbol{V}(\boldsymbol{X})_i$$

$$U_2^*(\boldsymbol{\beta}, \boldsymbol{\phi}) = -\sum_{i=1}^{n} \boldsymbol{W}(\boldsymbol{X})_i' [G(\phi, S) - vecs(I_m)], \quad \boldsymbol{I}_{22}^* = \boldsymbol{W}(\boldsymbol{X})_i' [H(\phi, S)] \boldsymbol{W}(\boldsymbol{X})$$

where $\Sigma(X^i; \phi^*)^{-1} = C(\phi) C(\phi)'$,$\boldsymbol{V}(\boldsymbol{X}) = I_m \otimes X_i$, $\boldsymbol{W}(\boldsymbol{X})_J = I \otimes X_i$, $\otimes$ is Kronecker product and J is the number of the parameter corresponding to variance. $G(\phi, S)$ and $H(\phi, S)$ is associated with gradient vector and hessian matrix.$G(\phi, S), H(\phi, S)$ and $vecs(I_m)$ can be checked in Cho (2013). For this project, we only drived the case where there are two response variables and one covariates.

Example

When there are two response variables and one covariates,$C(\phi)$, $G(\phi, S)$ and $H(\phi, S)$ are as following.

$$C(\phi) = \begin{pmatrix} e^{\tau_{11}} & 0 \\ \tau_{21} & e^{\tau_{22}} \end{pmatrix}$$

$$G(\phi, S) = 2 \begin{pmatrix} e^{\tau_{11}}(s_{11}e^{\tau_{11}} + s_{21}e^{\tau_{21}}) \\ s_{21}e^{\tau_{11}} + s_{22}\tau_{21} \\ e^{\tau_{22}}(s_{22}e^{\tau_{22}}) \end{pmatrix}$$

$$H(\phi, S) = \begin{pmatrix} 2s_{11}e^{2\tau_{11}} + s_{21}\tau_{21}e^{\tau_{11}} & s_{21}e^{\tau_{11}} & 0 \\ s_{21}e^{\tau_{11}} & s_{22} & 0 \\ 0 & 0 & 2s_{22}e^{2\tau_{22}} \end{pmatrix}$$

where $\tau_{11} = \phi_{11} + \phi_{12}x$, $\tau_{21} = \phi_{21} + \phi_{22}x$, and $\tau_{22} = \phi_{31} + \phi_{32}x$

2.3 Online Updating Method

Schifano (2016) fomulated the Online Updated estimates. Let $\{(\boldsymbol{Y}_{ki}, \boldsymbol{X}_{ki}, \boldsymbol{Z}_{ki}, i = 1, \ldots, n_k\}$ be the k^{th} subset. The score functions for subset k are

$$U_{1,k}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{i=1}^{n_k} \Psi_1(\boldsymbol{X}_{ki}, \boldsymbol{Z}_{ki}, \boldsymbol{\beta}, \boldsymbol{\phi})$$

$$U_{2,k}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{i=1}^{n_k} \Psi_2(\boldsymbol{X}_{ki}, \boldsymbol{Z}_{ki}, \boldsymbol{\beta}, \boldsymbol{\phi})$$

Define $\boldsymbol{A}_{1,k} = \sum_{i=1}^{n_k} \dfrac{\partial \Psi_1(\boldsymbol{X}_{ki}, \boldsymbol{\beta}, \boldsymbol{\phi})}{\partial \boldsymbol{\beta}}$ and $\boldsymbol{A}_{2,k} = \sum_{i=1}^{n_k} \dfrac{\partial \Psi_2(\boldsymbol{X}_{ki}, \boldsymbol{\beta}, \boldsymbol{\phi})}{\partial \boldsymbol{\phi}}$ A Taylor Expansion of $-U_{1,k}(\boldsymbol{\beta}, \boldsymbol{\phi}), U_{2,k}(\boldsymbol{\beta}, \boldsymbol{\phi})$ at $\hat{\boldsymbol{\beta}}_{n_k,k}, \hat{\boldsymbol{\phi}}_{n_k,k}$ respectively is given by

$$-U_{1,k}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \boldsymbol{A}_{1,k}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{n_k,k}) + \boldsymbol{R}_1$$

$$-U_{2,k}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \boldsymbol{A}_{2,k}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}_{n_k,k}) + \boldsymbol{R}_2$$

The cumulative coefficient estimates are

$$\hat{\boldsymbol{\beta}}_K = (\sum_{k=1}^{K} \boldsymbol{A}_{1,k}(\phi_k))^{-1} \sum_{k=1}^{K} \boldsymbol{A}_{1,k}\hat{\boldsymbol{\beta}}_{n_k,k} \tag{1}$$

$$\hat{\boldsymbol{\phi}}_K = (\sum_{k=1}^{K} \boldsymbol{A}_{2,k}(\beta_k))^{-1} \sum_{k=1}^{K} \boldsymbol{A}_{2,k}\hat{\boldsymbol{\phi}}_{n_k,k} \tag{2}$$

For β_k and ϕ_k in $(\boldsymbol{A}_{1,k}(\phi_k), \boldsymbol{A}_{2,k}(\beta_k)))$, we can use $\hat{\boldsymbol{\beta}}_{n_k,k}$ and $\hat{\boldsymbol{\phi}}_{n_k,k}$

Thus, we propose the cumulative coefficient estimators.

$$\hat{\boldsymbol{\beta}}_K = (\sum_{k=1}^{K} \boldsymbol{A}_{1,k}(\boldsymbol{\phi}_{n_k,k}))^{-1} \sum_{k=1}^{K} \boldsymbol{A}_{1,k}\hat{\boldsymbol{\beta}}_{n_k,k} \tag{3}$$

$$\hat{\boldsymbol{\phi}}_K = (\sum_{k=1}^{K} \boldsymbol{A}_{2,k}(\hat{\boldsymbol{\beta}}_{n_k,k}))^{-1} \sum_{k=1}^{K} \boldsymbol{A}_{2,k}\hat{\boldsymbol{\phi}}_{n_k,k} \tag{4}$$

3 Simulation Study

For the simulation study, we used two block sample sizes, n_1, n_2 to generate observation $\boldsymbol{Y}_i$ and X_i for $i = 1, ..., 1000$. X_i is generated from normal distribution with mean 1, and variance 1. And $\boldsymbol{Y}_i$ is generated from normal $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ where $\boldsymbol{\mu}_i = (\beta_1 + \beta_2 x_i \; \beta_1 + \beta_2 x_i)'$ and $\boldsymbol{\Sigma}_i^{-1} = CC'$,

$$C = \begin{bmatrix} \phi_{1,1} + \phi_{1,2}x_i & 0 \\ \phi_{2,1} + \phi_{2,2}x_i & \phi_{3,1} + \phi_{3,2}x_i \end{bmatrix}.$$

We set

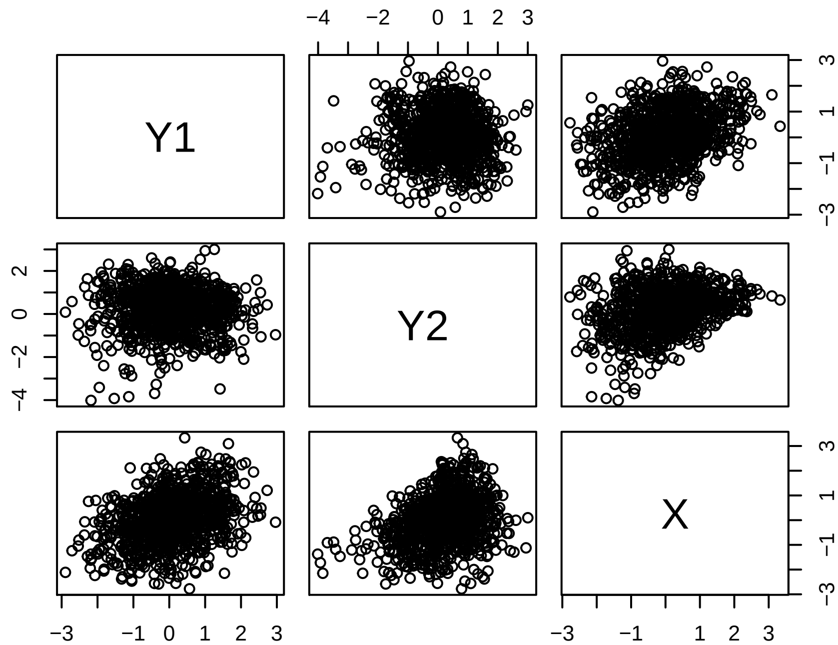
$$(\beta_1, \beta_2, \beta_3, \beta_4) = (0.1, 0.3, 0.2, 0.3)$$

$$(\phi_{1,1}, \phi_{1,2}, \phi_{2,1}, \phi_{2,2}, \phi_{3,1}, \phi_{3,2}) = (0.1, 0.1, 0.2, 0.2, 0.1, 0.2).$$

For cumulative coefficient estimates, first block sample size n_1 and second block sample size n_2 are 1000 and the repeat is 100.

```r
library(mvtnorm)
DATA_GENERATION = function(n, beta,  phi)
{
   X = rnorm(n, 0, 1)
   Dat = cbind(1, X)
   Y = matrix(0,n,2)
   for(i in 1:n)
   {
      Mu =  Dat %*% beta
      tau = Dat %*% phi[,2]
      diag.phi1 = exp(Dat %*% phi[,1])
      diag.phi2 = exp(Dat %*% phi[,3])
      C = diag(c(diag.phi1[i],diag.phi2[i]))
      C[lower.tri(C)] = tau[i]
      Covari = solve( C %*% t(C))
      Y[i,] = rmvnorm(1,mean = Mu[i,], sigma = Covari)
   }
   return(list( Dat= Dat, Y=Y))
}
set.seed(12345)
beta = matrix(c(0.1, 0.3, 0.2, 0.3),2,2)
phi = matrix(c(0.1, 0.1,0.2, 0.2,0.1,0.2),2,3)
n = 1000
DATA = DATA_GENERATION(n, beta, phi)
```

```
Y1 = DATA$Y[,1]; Y2 = DATA$Y[,2]; X = DATA$Dat[,2]
pairs(~Y1 + Y2 + X)
```



The plot show the relationship between response variables and a predictor from the generated one dataset. When X incease, Y1 and Y2 tend to incease. Also, when X is increased, the variance of Y2 look like decrease at a point x.

The simulation result is in the table 1 and 2. For the simulation, we used the initial values as follows.

$$(\beta_1, \beta_2, \beta_3, \beta_4) = (0.1, 0.1, 0.1, 0.1)$$

$$(\phi_{1,1}, \phi_{1,2}, \phi_{2,1}, \phi_{2,2}, \phi_{3,1}, \phi_{3,2}) = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1).$$

When you look at the tables, the average cumulative coefficient estimates look like very close to the true value. When we focus on the estimates corresponding the covariate, the standard errors of estimates for mean are very similar. The standard errors of estimates for variance are a little different.The standard error for $\phi_{2,1}$ is higher than for $\phi_{1,2}$ and $\phi_{3,2}$.

Table 1: Result for β's

	β_1	β_2	β_3	β_4
Ture value	0.1	0.3	0.2	0.3
Average Cumulative Coefficient Estimates	0.0998	0.2925	0.1979	0.2920
Standard Error	0.0205	0.0451	0.0287	0.04197

4 Summary

In summary, we proposed one method to update estimates in data stream under multivatiate regression model. Schifano (2016) developed cumulative coefficient estimates for the linear models and estimating equations. In this project, we will consider the multivariate regression model for the

Table 2: Result for ϕ's

	$\phi_{1,1}$	$\phi_{1,2}$	$\phi_{2,1}$	$\phi_{2,2}$	$\phi_{3,1}$	$\phi_{3,2}$
Ture value	0.1	0.1	0.2	0.2	0.1	0.2
Average Cumulative Coefficient Estimates	0.10234	0.1005	0.1938	0.1961	0.1040	0.1897
Standard Error	0.0149	0.01482	0.0286	0.0249	0.0159	0.01628

online updating. Pourahmadi (2000) studied the maximum likelihood estimators of a generalised linear model for the covariance matrix. In this project, mean and covariance of response variables are estimated by using a covariate. To estimates parameters corresponding to the mean and covariance, the Newton-Raphson algorithm is used. We considered two estimating equations to update cumulative estimators. Firstly, we fixed the parameters of the covariance and then estimated the parameters of the mean. Then, the estimated parameters of the mean are fixed to estimate the parameters of the covariance. Through the simulation study, we checked the perfomance of the proposed method. As the result, the average of cumulative coefficient estimates are very close to the true values.

5 Code

We submit two R code. *Simulation.R* code is for the result of simulation study and *Project.R* code is used by *Simulation.R*. You can check the result in tables by *Simulation.R*.

References

Cho, K.-D. (2013), *Simultaneously modeling for mean vector and covariance matrix*, Ph.d Thesis, university of Korea, Library.

Pourahmadi, M. (2000), *Maximum likelihood estimation of generalised linear models for multivariate normal covariance matirx*, Biometrika, 87,2,pp,425-435.

Schifano, Elizabeth D., e. a. (2016), "Online Updating of Statistical Inference in the Big Data Setting," *Technometrics*, 393–403.