Middle-East Journal of Scientific Research 11 (9): 1317-1322, 2012
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2012.11.09.1537

Variance Estimation of Linear Regression Coefficients using Markov Chain Monte Carlo Simulation

[1]*K. Fathi Vajargah*, [2]*F. Mehrdoust and* [3]*F. Kamalzadeh*

[1,3]Islamic Azad University, North Branch, Tehran, Iran
[2]Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Guilan, Rasht, Iran

Abstract: The main goal in this paper is to find the inverse of matrix X^tX which is appear in the variance equation of simple linear regression coefficient using Monte Carlo algorithm.

Keywords: Simple linear regression . Monte carlo simulation . Markov chain . Variance estimation

INTRODUCTION

Regression analysis is a statistical technique for investigating and modeling the relationship between variables. Applications of regression are numerous and occur in almost every field, including engineering, the physical sciences, economics and management [5]. In this paper, we employ the Monte Carlo algorithm [2, 3] for finding the variance of simple linear regression coefficient.

Here, we consider the simple linear regression as follows

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1,2,\ldots,n \tag{1}$$

where the intercept β_0 and the slope β_1 are unknown parameters and ε_i are error components. The errors are assumed to have mean zero and unknown variance σ^2. The equation (1) can be rewritten as matrix form

$$Y = X\beta + \varepsilon$$

where

$$Y = [y_1,\ldots,y_n]^T, X = [X_1,\ldots,X_n]^T, \beta = [\beta_0,\beta_1]^T$$

and $\varepsilon = [\varepsilon_1,\ldots,\varepsilon_n]^T$.

The variance of $\beta = [\beta_0,\beta_1]^T$ is found as

$$V(\beta) = V\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} = \begin{bmatrix} V(\beta_o) & \operatorname{cov}(\beta_o,\beta_1)\\ \operatorname{cov}(\beta_o,\beta_1) & V(\beta_1)\end{bmatrix}$$

where β_0 and β_1 obtain using the least square method [2]

$$\beta_0 = \bar{y} - \beta_1\bar{x},$$
$$\beta_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$$

Consequently we have

$$V(\beta) = \begin{bmatrix} \dfrac{\delta^2\sum x_i^2}{n\sum(x_i-\bar{x})^2} & \dfrac{-\bar{x}\delta^2}{\sum(x_i-\bar{x})^2}\\ \dfrac{-\bar{x}\delta^2}{\sum(x_i-\bar{x})^2} & \dfrac{\delta^2}{\sum(x_i-\bar{x})^2}\end{bmatrix}$$
$$= \delta^2\frac{1}{n\sum(x_i-\bar{x})^2}\begin{bmatrix}\sum x_i^2 & -\sum x_i\\ -\sum x_i & n\end{bmatrix} = \delta^2\left(X^TX\right)^{-1}$$

We employ the Monte Carlo (MC) algorithm for calculating $(X^TX)^{-1}$.

MARKOV CHAIN MONTE CARLO ALGORITHM FOR OBTAINING THE INVERSE OF A MATRIX

In the process of calculating the inverse of matrix B in equation BX=f, by Monte Carlo method, firstly, from the Markov chain of $\{x_n\}$ n = 0 with a status of S={1,….,n} (the number of members of the chain space status must be equal to the dimension of the matrix) a stochastic path would be assumed [1-3]

$$i_0 \to i_1 \to i_2 \to \cdots \to i_k$$

We will assume a matrix of probability of transformation in this Markov chain as P= $\{p_{ij}\}$ and the first distribution would be p_{i0}

Corresponding Author: K. Fathi Vajargah, Islamic Azad University, North Branch, Tehran, Iran.

$$T_k(b^{-1}_{mr}) = \sum_{j|i_j=r} W_j$$

where

$$W_j = w_{j-1}\frac{a_{i_{j-1}i_j}}{p_{i_{j-1}i_j}} = \frac{a_{ii_1}\ldots\ldots a_{i_{j-1}i_j}}{p_{ii_1}\ldots\ldots p_{i_{j-1}i_j}}$$

In formula (2) the elements of the numerator are very similar to entries of frequency matrix A and the elements of denominator are the probability of choosing the elements of the numerator, i.e. for instance, the entry $a_{i_{k-1}i_k}$ of matrix A (which would be stated later) will be chosen with a probability of $P_{i_{k-1}i_k}$ [3]. We will show that $T_k\left(b^{-1}_{mr}\right)$ is non-transverse for the vector of answer. In general, we will have

$$BX = f \leftrightarrow x = AX + f$$

$$X^{k+1} = \sum_{m=0}^{k} A^m f$$

$$\lim_{k\to\infty} X^{k+1} = \lim_{k\to\infty}\sum_{m=0}^{k} A^m f = \frac{I}{I-A}\ f = (II-A)^{-1}f\ = B^{-1} f = X$$

$$B^{-1} = \left[b_{ij}\right]^{-1} = I + A + A^2 + \ldots\ldots + A^m + \ldots..$$

The m^{th} element of vector X is

$$X_m = \sum_{r=1}^{n} b^{-1}_{mr} f_r$$

It is enough to assume

$$f^T = e_r = (0,\ldots,0,1,\ldots,0)$$

in this case we will have

$$X_m = b^{-1}_{mr}\ \ m=1,2,\ldots, n$$

Theorem: The inverse of square and nonsingular matrix B_{nxn} by Monte Carlo method will be calculated as follows:

$$b^{-1}_{mr} = \frac{1}{N}\sum_{s=1}^{N}\left[\sum_{j|i_j=r} W_j^{(S)}\right]$$

Remark: Note that that if K will be large enough, for h = (0,…,0,1,0,…,0) and f = (0,…,0,1,0,…,0) we will have

$$\lim_{k\to\infty} E(\lim_{k\to\infty} E\left(T_k\left(b^{-1}_{mr}\right)\right) = E\left(T_\infty\left(b^{-1}_{mr}\right)\right) = \langle h, x\rangle = b^{-1}_{mr}$$

CALCULATION OF THE INVERSE OF MATRIX B

In this section, we will show iterative form in a way that it will be convergent to the main answer. Therefore, we should present particular decompositions for the nonsingular matrix of coefficients $B_{n\times n}$. In this case we can rewrite matrix B as a composition of prevalent diagonal matrix and a balance matrix. By this decomposition, the inverse of matrix B will be obtained from the inverse of the diagonal prevalent matrix. We assume the decomposition D=B+E in a way that D become a prevalent diagonal matrix. D^{-1} will be calculated according to the algorithm presented below. Afterwards we calculate the inverse of matrix B by the below algorithm by D^{-1}. The residual matrix $E = \sum_{i=1}^{n} s_i$ is, in which s_i is a matrix of order 1 in which all of the elements except elements (i,i) are zero. Assume that

$$B^{(n)} = B^{(n-1)} + E,\ \ B^{(0)} = B$$

in this case the inverse $[B^{(i)}]^{-1}$ would be calculated as follows:

$$\left[B^{(i)}\right]^{-1} = \left[B^{(i+1)}\right]^{-1} + \frac{\left(B^{(i+1)}\right)^{-1} s_{i+1}\left(B^{i+1}\right)^{-1}}{1-\mathrm{trace}\left(\left(B^{i+1}\right)s_{i+1}\right)}$$

Algorithm for calculating the prevalent diagonal matrix $D(D^{-1})$

Step 1: Receiving prevalent diagonal matrix D_{nxn} and the values for parameters ε and δ will be received.

Step 2: Put D= B – B_1 which B_{nxn} is a diagonal matrix in a way that i=1,..,n, $b_{ii} = d_{ii}$

Step 3: calculating T= B^{-1} B_1

Step 4: Calculating ||T|| and

$$N = \left(\frac{0.6745}{\varepsilon}.\frac{1}{(1-\|T\|)}\right)^2$$

Step 5: We assume i=1 (where I varies from 1 to N) and j=1 (where j varies from 1 to n).

Step 6: We assume w_0=1, sum(i)=0, point=i and t_k =0.

Step 7: Randomly and steadily we produce the next point number.

Step 8: If the value of T[point] [next point]≠0 we will proceed to the next step, otherwise it refers to other stochastic number to the variable point so that a non zero number will be chosen.

Step 9: We calculate

$$W_j = W_{j-1}\frac{T[\text{point}][\text{nextpoint}]}{p[\text{point}][\text{nextpoint}]}$$

Step 10: It refers to the value of next point to the variable point and assume:

$$\text{Sum (i)} = \text{sum (i)} + w_j$$

Step 11: If $|W_j| < d$ we assume $t_K = t_K+1$ (otherwise we assume j=j+1 and if j<n we go back to step seven and follow the steps again)

Step 12: If $t_k = N$ the fifth step (the loop corresponding to i,j) will be up, otherwise we go back to step seven and follow the trend until comes true $t_k = N$.

Step 13: The average and SUM (i) will be calculated.

ESTIMATION OF THE MONTE CARLO PARAMETERS

Parameters T (the length of Markov chain) and N (the number of sample systems) have essential role in calculation of the inverse of a matrix. These parameters are important since the time and the cost are directly dependant on them.

Estimation of parameter T: Assume that for every i = 1,…,n, the elements of transition probability matrix P, for every non zero matrix A, is defined as below

$$P_{ij} = \frac{|a_{ij}|}{\sum |a_{ij}|}, \quad j = 1,2,\ldots,n$$

Now, we define the Monte Carlo estimator Θ^* as below

$$\Theta^*[g] = \frac{g_{k_\alpha}}{p_{k_\alpha}}\sum_{j=0}^{\infty} W_j \varphi_{k_j}$$

where

$$W_j = W_{j-1}\frac{a_{k_{j-1}k_j}}{p_{k_{j-1}k_j}}, \quad W_0 = 1$$

On the other hand we stop the calculation process of sum of Θ^*, Whenever $|W_i|<\delta$, in which δ is a positive small number, we will have

$$|W_i| = \left|\frac{a_{\alpha_0\alpha_1}a_{\alpha_1\alpha_2}\cdots a_{\alpha_{i-1}\alpha_i}}{p_{\alpha_0\alpha_1}p_{\alpha_1\alpha_2}\cdots p_{\alpha_{i-1}\alpha_i}}\right| \le \left|\frac{a_{\alpha_0\alpha_1}a_{\alpha_1\alpha_2}\cdots a_{\alpha_{i-1}\alpha_i}}{\frac{a_{\alpha_0\alpha_1}}{\|A\|}\frac{a_{\alpha_1\alpha_2}}{\|A\|}\cdots\frac{a_{\alpha_{i-1}\alpha_i}}{\|A\|}}\right|$$

$$= \|A\|^i = \delta$$

We can write

$$i\log\|A\| = \log\delta$$

Therefore we have

$$T = i \le \left[\frac{\log\delta}{\log\|A\|}\right] + 1$$

Estimation of parameter N: Regarding to the convergence condition of Newman series we can write

$$|\theta^*| \le \frac{\|f\|}{1-\|A\|} \le \frac{1}{1-\|A\|}$$

On the other hand the variance of estimation of θ^* is finite. Therefore we will have:

$$\mathrm{Var}(\theta^*) \le E(\theta^{*2}) \le \frac{\|f\|^2}{(1-\|A\|)^2} \le \frac{1}{(1-\|A\|)^2}$$

If the estimation of θ is a Monte Carlo estimation, considering to the theorem of central limit for large quantities of n we will have:

$$\bar{q} \sim N\left(E(q), \frac{\mathrm{var}(q)}{N}\right)$$

Now we define the probability of error for Monte Carlo method as below:

$$P\left\{|\bar{\theta} - E(\theta)| < r_N\right\} \approx \frac{1}{2} \approx P\left\{|\bar{\theta} - E(\theta)| > r_N\right\}$$

In this case, regarding the theorem of central limit, we will have

$$r_N = 0.6745\sqrt{\frac{\mathrm{Var}(\theta)}{N}}$$

Therefore if the accuracy of error be $\varepsilon = |\bar{\theta} - E(\theta)|$, for large number then

$$N \geq \frac{(0.6745)^2 . \mathrm{Var}(\theta)}{\varepsilon^2}$$

On the other hand, considering the two previous relations we can have:

$$N \geq \frac{(0.6745)^2}{\varepsilon^2} \frac{1}{(1-\|A\|)^2}$$

NUMERICAL EXPERIMENTS

Example 1: Calculating the relation between the amount of food (Y) eaten by a turkey and the average weight (X_1) is desired. Ten turkey have been chosen. Data have been presented in the Table 1.

Linear regression model of Y = $\beta_0+\beta_1 X_1$ will be used for processing the data. Matrix X and matrix Y are as below:

$$Y = \begin{bmatrix} 87.1 \\ 93.1 \\ . \\ . \\ . \\ 94.4 \end{bmatrix}, X = \begin{bmatrix} 1 & 4.6 \\ 1 & 5.1 \\ . & . \\ . & . \\ . & . \\ 1 & 5.1 \end{bmatrix}, \beta = \begin{bmatrix} \beta_{.} \\ \beta_1 \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

We need to estimate the variance of regression coefficient β

$$V(\beta) = \delta^2 (X'X)^{-1}$$

$$(x'x)^{-1} = \begin{bmatrix} \dfrac{\sum x_i^2}{n\sum (x_i - \bar{x})^2} & \dfrac{-\bar{x}}{\sum (x_i - \bar{x})^2} \\ \dfrac{-\bar{x}}{\sum (x_i - \bar{x})^2} & \dfrac{1}{\sum (x_i - \bar{x})^2} \end{bmatrix}$$

$$= \begin{bmatrix} 16.24609375 & -3.2421875 \\ -3.2421875 & 0.65104167 \end{bmatrix}$$

$$V(\beta) = \begin{bmatrix} 16.24609375 & -3.2421875 \\ -3.241875 & 0.65104167 \end{bmatrix} (5.587)$$

$$= \begin{bmatrix} 90.76692578 & -18.11410156 \\ -18.1410152 & 3.63736981 \end{bmatrix}$$

On the other hand we need to solve the inverse of this matrix $(XX')^{-1}$, by Monte Carlo method:

$$(X'X)_{MC}^{-1} = \begin{bmatrix} 13.177 & 11.768 \\ 65.661 & 58.828 \end{bmatrix}$$

$$(X'X)^{-1}_{MATLAB} = \begin{bmatrix} 13.169 & 11.768 \\ 65.681 & 58.820 \end{bmatrix}$$

Table 1: The average of body weight (X) and the amount of food (Y) eaten by 10 turkeys within 8 hours

Observation	Consuming (Y) Food	Weight (X)
1	87.1	4.6
2	93.1	5.1
3	89.8	4.8
4	91.4	4.4
5	99.5	5.9
6	92.1	4.7
7	95.5	5.1
8	99.3	5.2
9	93.4	4.9
10	94.4	5.1

Table 2: Regression Coefficients by Monte Carlo method

	Y	X_1	X_2	X_3	X_3	X_4
1	43	51	30	39	61	92
2	63	64	51	54	63	73
3	71	70	68	69	76	86
4	61	63	45	47	54	84
5	81	78	56	66	71	83
6	43	55	49	44	54	49
7	58	67	42	56	66	68
8	71	75	50	55	70	66
9	72	82	72	67	71	83
10	67	61	45	47	62	80
11	64	53	53	58	58	67
12	67	60	47	39	59	74
13	69	62	37	42	55	63
14	68	83	83	45	59	77
15	77	77	54	72	79	77
16	81	90	50	72	60	54
17	74	85	64	69	79	79
18	65	60	65	75	55	80
19	65	70	46	57	75	85
20	50	58	68	54	64	78

Example 2: Relation between the grade which clerks give to assessor (Y) with study of clerk's complain (X_1),and preventing from distinguish (X_2),and learning some new things (X_3),and giving overtime work (X_4) and extravagance in criticizing of weak works (X_5) are being considered. Therefore we gather 20 observations from these 6 variable from insurance company, which is shown in the following Table 2. A linear regression model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

is becoming on data.The matrix form is Y = $X\beta+\beta$. The form of matrix X and vector Y is

$$Y = \begin{bmatrix} 43 \\ 63 \\ 71 \\ . \\ . \\ . \\ 65 \\ 65 \\ 50 \end{bmatrix}_{20\times 1}, X = \begin{bmatrix} 1 & 51 & 30 & 39 & 61 & 92 \\ 1 & 64 & 51 & 54 & 63 & 73 \\ 1 & 70 & 68 & 69 & 76 & 86 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & 70 & 46 & 57 & 75 & 85 \\ 1 & 58 & 68 & 54 & 64 & 78 \end{bmatrix}$$

Now we calculate the equation by variance of regression coefficients with Monte Carlo method

$$X'X = \begin{bmatrix} 1 & 1 & 1 & . & . & . & 1 & 1 \\ 51 & 64 & 70 & . & . & . & 70 & 58 \\ 30 & 51 & 68 & . & . & . & 46 & 68 \\ 39 & 54 & 69 & . & . & . & 57 & 54 \\ 61 & 63 & 76 & . & . & . & 75 & 64 \\ 92 & 73 & 86 & . & . & . & 85 & 78 \end{bmatrix} \begin{bmatrix} 1 & 51 & 30 & 39 & 61 & 92 \\ 1 & 64 & 51 & 54 & 63 & 73 \\ 1 & 70 & 68 & 69 & 76 & 86 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & 70 & 46 & 57 & 75 & 85 \\ 1 & 58 & 68 & 54 & 64 & 78 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 25 & 1364 & 1075 & 1127 & 1291 & 1498 \\ 1364 & 95494 & 74639 & 78306 & 89030 & 102103 \\ 1075 & 74639 & 60913 & 61893 & 69915 & 80971 \\ 1127 & 78306 & 61893 & 66151 & 73787 & 84628 \\ 1291 & 89030 & 69915 & 73787 & 84659 & 97374 \\ 1498 & 102103 & 80971 & 84628 & 97374 & 114462 \end{bmatrix}$$

The inverse of matrix is being calculated by Matlab software and MAO method as beow.
It is considering that there is not any notorious error in these answers so we do not need to filter them.

$$(X'X)^{-1}_{MATLAB} = \begin{bmatrix} 0.1920 & -0.0008 & -0.0001 & 0.0000 & -0.0008 & -0.0011 \\ -0.0008 & 0.0008 & -0.0002 & -0.0001 & -0.0006 & 0.0001 \\ -0.0001 & -0.0002 & 0.0005 & -0.0002 & 0.0002 & -0.0002 \\ 0.0000 & -0.0001 & -0.0002 & 0.0007 & -0.0004 & 0.0001 \\ -0.0008 & -0.0006 & 0.0002 & -0.0004 & 0.0016 & -0.0006 \\ -0.0011 & 0.0001 & -0.0002 & 0.0001 & -0.0006 & 0.0005 \end{bmatrix}$$

$$(XX')^{-1}_{MC} = \begin{bmatrix} 0.1922 & -0.0008 & -0.0001 & 0.0000 & -0.0008 & -0.0010 \\ -0.0008 & 0.0008 & -0.0003 & -0.0001 & -0.0006 & 0.0001 \\ -0.0002 & -0.0002 & 0.0004 & -0.0002 & 0.0002 & -0.0002 \\ 0.0000 & -0.0001 & -0.0002 & 0.0007 & -0.0004 & 0.0001 \\ -0.0008 & -0.0006 & 0.0002 & -0.0004 & 0.0014 & -0.0007 \\ -0.0010 & 0.0001 & -0.0002 & 0.0002 & -0.0006 & 0.0006 \end{bmatrix}$$

The estimate of variance of regression coefficient is:

$$V(\beta) = \delta^2 (X'X)^{-1} = \begin{bmatrix} 1359.2 & -5.6577 & -0.7072 & 0.0000 & -5.6577 & 7.072 \\ -5.6577 & 5.6577 & -2.1216 & -0.7072 & -4.2432 & 0.7072 \\ -1.4144 & -1.4144 & 2.8288 & -1.4144 & 1.4144 & -1.4144 \\ 0.0000 & -0.7072 & 1.4144 & 4.9504 & -2.8288 & 0.7072 \\ -5.6577 & -4.2432 & 1.4441 & -2.8288 & 9.9009 & -4.9504 \\ -7.0721 & 0.7072 & -1.4441 & 1.4144 & -4.2432 & 4.2432 \end{bmatrix}$$

CONCLUSION

We have seen that both of Matlab and Monte Carlo method have the same answers and there is not any notorious difference between their answers. However, in large scale problems the Monte Carlo method is efficient.

REFERENCES

1. Alexandrov, V.N., A. Rau-Chaplin, F. Dehne and K. Taft, 1998. Efficient Coarse Grained Monte Carlo Algorithms for Matrix Computations using PVM, LNCS 1497, Springer, pp: 323-330.
2. Dimov, I.T. and V.N. Alexandrov, 1998. A new highly convergent Monte Carlo method for matrix computations. Mathematics and Computers in Simulation, Bulgaria Academy of Science, 47: 165-181.
3. Fathi Vajargah, B. and K. Fathi Vajargah, 2006. Parallel Monte Carlo computation for solving SLAE with minimum communication. Applied Mathematics and Computation, pp: 1-9.
4. Forsythe, S. and Liebler, 1950. Matrix Inversion by a Monte Carlo method. Math. Tables other Aids Compul., 4: 127-129.
5. Montgomery, D.C. and E.A. Peck, 1991. Introduction to Linear Regression Analysis.