Estimating of Eigenvalue with Monte Carlo Method and its Application in the Principal Components (PCA)

Kianoush Fathi Vajargah*, Fatemeh Kamalzadeh

Department of Statistics Islamic Azad University, North branch Tehran, Iran

k_fathi@iau-tnb.ac.ir , *fa_kamalzadeh@gmail.com*

Abstract

One of discussions in multivariable analysis is defining the factor and main vectors by calculating eigenvalue. In this paper we deal with an unbiased estimator of eigenvector and as a result we define eigenvalues. The purpose was introducing a new statistical method that is different from other numerical methods, which it defines the eigenvalue matrix. On the other hand, the efficiency of this method is up when the mass and dimension of matrix are high. Therefore, this is a low cast and efficient method in calculation. This paper covers some background of data compression and how Markov chain Monte Carlo (MCMC) and principal component analysis (PCA) has been and can be used for calculating eigenvalue.

1. Introduction

In this paper we deal with calculating the eigenvalues and eigenvectors and determining the principal components with Monte Carlo method. On the other hand, the Monte Carlo method is illustrated on the

base of repetition systems, so in this paper the answer of liner equations system x=Ax+f with ||A||<1 is determined with Monte Carlo Markov Chain (MCMC). At first, we define an unbiased estimator for defining x factor and analysis in some different algorithms. Suppose that there is a linear equations system Bx=f which B is non-singular matrix $n \times n$ and the vector $f^t = (f_1, f_2, ..., f_n)$ is known. Now, the vector x is estimated by Monte Carlo (MC). At first, we define A=I-B which I is sameness matrix from degree. Therefore, the system Bx=f change to $Bx = f \Leftrightarrow (I - A)x = f \Leftrightarrow x = Ax + f$

and we have $\|A\| = \max\limits_{1 \le i \le n} \sum\limits_{j=1}^{n} |a_{ij}|$

Theorem: The estimator $T_k(h) = \dfrac{h_{i_0}}{p_{i_0}} \sum\limits_{m=0}^{k} w_m f_{i_m}$ is an unbiased estimator for $< h, x^{(k+1)} >$.

2. Principal component analysis (PCA)

Analyze of principal component is depends on variance-covariance structure with helping some liner combinations. The main purposes are 1.data mass reduction 2.infrence of data.

The principal components are the special liner combinations of p random variable $x_1, x_2, ..., x_p$ from algebraic view. From geometric view, these liner combinations show the new coordinate system which is gained from circulate of primer system with $x_1, x_2, ..., x_p$ as coordinate vectors.

Suppose that Σ is a covariance matrix of random vector of $\acute{x} = [x_1, x_2, ..., x_p]$. Suppose that Σ has couple eigenvalue-eigenvector $(\lambda_1, e_1), (\lambda_2, e_2), ... (\lambda_p, e_p)$ hcihw $\quad \lambda_1 \ge \lambda_2 \ge .... \ge \lambda_p \ge 0$ principal $\qquad\qquad$ htiw i tnenopmoc ni

$$Y_i = e_i' X = e_{1i} X_1 + e_{2i} X_2 + ... + e_{pi} X_\rho \qquad\qquad i = 1, 2, ..., p$$

Now, we have

$$Var(Y_i) = e_i' \sum e_i = \lambda_i \qquad\qquad\qquad i = 1, 2, ..., p$$
$$Cov(Y_i, Y_k) = e_i' \sum e_k = 0 \qquad\qquad i \ne k$$

we know; $\max\limits_{l \ne 0} \dfrac{l' \sum l}{l' l} = \lambda_1$ (it is reachable when l=e₁), nevertheless by normaling the eigenvalue then $e_1' e_1 = 1$

Therefore $\max\limits_{l \neq 0} \dfrac{l' \sum l}{l'l} = \lambda_1 = \dfrac{e_1' \sum e_1}{e_1' e_1} = e_1' \sum e_1 = Var(Y_1)$

Result 2-1

suppose that $\acute{x} = [x_1, x_2, \ldots, x_p]$ has covariance matrix Σ with couple eigenvalue-vector value $(\lambda_1, e_1), (\lambda_2, e_2), \ldots (\lambda_p, e_p)$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ niam era neht tnenopmoc

$$\sigma_{11} + \sigma_{22} + \ldots \sigma_{pp} = \sum_{i=1}^{P} Var(X_i) = \lambda_1 + \lambda_2 + \ldots + \lambda_P = \sum_{i=1}^{P} Var(Y_i)$$

Finally, the rate of the final variance of principal component k^{th} is

$$= \dfrac{\lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \qquad k = 1, 2, \ldots, p$$

For large p the most of the final population variance(e.g. %80-90) could attribution to 3 first characters therefore, these characters could replace by p primer variable without losing so many information.

Result 2-2

Another common way for the rate of different characters [Cotel 1966] is ' slope point plot' (it means λ versus i is plotted). Most of the time this plot shows that where the large special values sre stopped and where the small special values are started. Thus, if the variance-covariance matrix is Σ then we will use bellow ratio:

The rate of total variance $= \dfrac{\hat{\lambda}j}{\sum\limits_{j=1}^{n} \hat{\lambda}_j}$

Result2-3

The Bartlet test could show the number of including characters. Special structure always $cov(x_i, x_k) = \sqrt{\sigma_{ii} \sigma_{kk} \rho}$ or $corr(x_i, x_k) = \rho$ for each i$\neq$k is the sagittal structure that eigenvalues Σ are not different and previous results do not have application. The test is as follow

$$H_0 : \rho = \rho_{0(p \times p)} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & \rho & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} \qquad H_1 : r^1 r_0$$

The test of H_0 v.s. H_1 could on the base of likelihood ratio statistics. The test H_0 v.s. could stablish on the base of maximum likelihood ratio statistics. In this method the quantities are

$$\overline{r}_k = \frac{1}{p-1} \sum_{i=1}^{P} r_{ik} \qquad\qquad k = 1,2,\ldots, p$$

$$i \neq k$$

$$\overline{r} = \frac{2}{P(P-1)} \sum_{i<k} \sum r_{ik} \qquad\qquad \hat{\gamma} = \frac{(P-1)^2\left[1-(1-\overline{r})^2\right]}{P-(p-2)(1-\overline{r})^2}$$

$\overline{r}_k$ the average of non-diagonal in Colom (or row) k^{th} the R and $\overline{r}$ is the average of all non-diagonal members. The approximately test of large samples in level α is rejected H_0 when

$$T = \frac{(n-1)}{(1-\overline{r})^2}\left[\sum_{i<k} \sum (r_{ik}-\overline{r})^2 - \hat{\gamma}\sum_{k=1}^{P}(\overline{r}_k - \overline{r})^2 \right] > \chi^2_{\frac{(p+1)(p-2)}{2}}(\alpha)$$

2. Monte Carlo Method and calculating eigenvalues

Calculating of eigenvalue of one matrix is one of the important parameters. Suppose that

$A_{n\times n}$ dna xirtam cirtemys a si $f = (f_1, f_2, \ldots, f_n)^t \in R^n$ esoprup ehT .rotcev mouluc si

esoppuS .dohtem noitateper olraC etnoM yb eulavnegie fo gnitaulave si $Af = \lambda_{(A)}f$ dna

$a_{ij} = a_{ji}$, n,...,,1,2=j,i

xirtam cirtemys si A snaem tI

$$\lambda_{\min} = \lambda_n < \lambda_{n-1} \leq \lambda_{n-2} \leq \cdots \leq \lambda_2 < \lambda_1 = \lambda_{\max} \text{ then}$$

$$\lambda_{\max}(A) = \lim_{i\to\infty} \frac{<h, A^i f>}{<h, A^{i-1} f>} \text{ .A xirtam eulavnegie tsegib eht etamitse ew}$$

niahc vokram eht ,tsrif tA $\{X_n\}_{n\geq 0}$ si htap modnar wolleb htiw {n,...,1,2}=s ecaps sutats htiw

redisnoc $k_0 \to k_1 \to k_2 \to \cdots \to k_i$. we define $h = \langle h_1, \ldots, h_n \rangle$ dna

$$x = \left\langle x_\backslash, \ldots, x_n \right\rangle \text{ hcihw}$$

$$< h, x >= \sum_{i=1}^{n} h_i \cdot x_i = h_1 x_1 + h_2 x_2 + \cdots + h_n x_n \text{ then we define sequences}$$

$$P = \{p_{ij}\}_{i,j=\backslash}^{n} \text{ dna } P = \{p_i\}_{i=1}^{n} \text{ hcihw}$$

$$\begin{cases} p_i > 0 \\ p_i = 0 \end{cases} \quad \text{if} \quad \begin{array}{l} h_i \neq 0 \ P_r\{k_0 = \alpha\} = p_\alpha \\ h_i = 0 \ P_r\{k_j = \beta | k_{j-1} = \alpha\} = p_{\alpha\beta} \end{array}$$

sa senifed n,...,1,2=j elbairav modnar eht ,lla retfa

$$W_j = \begin{cases} \dfrac{h_{k_0}}{p_{k_0}} & j = 0 \\[4mm] \dfrac{h_{k_0}}{p_{k_0}} \cdot \dfrac{a_{k_0 k_1} \ldots a_{k_{j-1} k_j}}{p_{k_0 k_1} \cdots p_{k_{j-1} k_j}} & j \geq 1 \end{cases}$$

$$W_j = W_{j-1} \frac{a_{k_{j-1} k_j}}{p_{k_{j-1} k_j}} \quad j = 1, 2, \ldots, i$$

neht hguone egral si i fi snoitpmussa evoba gniredisnoc yB :meroehT

$$\frac{E\{w_i f_{k_i}\}}{E\{w_{i-1} f_{k_{i-1}}\}} \approx \lambda_{\max}(A)$$

Example: Data of bellow table shows the information of 5 social economical variables for 14 zones of Madison Wisconsin:

Table 1: The census data

emoh fo egarevA etar ($10.000)	tnemyolpmE yhtlaeh eitilicafs (tnecrep)	latoT tnemyolpme	fo naideM ga loohcse	noitaluppop latoT (sdnasuoht)	enoz
2.91	2.27	2.265	14.2	5.935	1
2.62	0.75	0.597	13.1	1.523	2
1.72	.11	1.237	12.7	2.599	3
3.02	0.81	1.649	15.2	4.009	4
2.22	2.50	2.312	14.7	4 .687	5
2.36	4.51	3.648	15.6	8.04	6
1.97	1.03	1.244	13.3	2.766	7
1.85	2.39	2.618	17.0	6.538	8
2.01	5.52	3.147	12.9	6.451	9
1.82	2.18	1.606	12.2	3.314	10
1.80	2.83	2.119	13.0	3.777	11
4.25	0.84	0 .798	13.8	1.530	12
2.64	1.75	1.336	13.6	2.768	13
3.17	1.91	2.763	14.9	6.585	14

The Bellow statistics are gained by these data:

$$\overline{X}' = \begin{bmatrix} 4.32 & 14.01 & 1.95 & 2.17 & 2.45 \end{bmatrix}$$

$$S = \begin{bmatrix} 4.308 & 1.683 & 1.803 & 2.155 & -0.253 \\ 1.683 & 1.768 & 0.588 & 0.177 & 0.176 \\ 1.803 & 0.588 & 0.801 & 1.065 & -0.158 \\ -2.155 & 0.177 & 1.065 & 1.970 & -0.357 \\ -0.253 & 0.176 & -0.158 & -0.357 & 0.504 \end{bmatrix}$$

The follow table is:

Table 2: The principal component coefficients (numbers in parentheses are the correlation coefficients)

$\hat{e}_5$	$\hat{e}_4$	$\hat{e}_3$	$\hat{e}_2\left(r_{\hat{y}_2, x_k}\right)$	$\hat{e}_1\left(r_{\hat{y}_1, x_k}\right)$	elbairav
-0.302	0.542	0.004	-0 .071(-0.04)	0.781(0.99)	noitalupop latoT
-0 .010	-.0.545	-0.162	-0.264 (-0.76)	03.6 (0.61)	ega loohcs fo naideM
0 .937	0.05	0 .015	0.083(0.12)	0.334(0.98)	Total employments
-0 .173	-0 .636	0.220	0 .579(0 .55)	0.426(0.80)	yhtlaeH tnemyolpmE seitilicaf
-0 .024	-0 .051	0 .962	-0.262(-0.42)	.054(- 0..20) -	eulav emoh fo naideM
0.014	0 .230	0 .390	1 .786	6 .639	ecnairav $\left(\hat{\lambda}_i\right)$
100	99.9	97.4	93.2	74.1	ecnairav latoT tnecrep evitalumuc

The first principal component expresses %74.4 variance of total sample. The first couple of principal component expresses %93.3 of total variance. As a result, good changes are obvious with 2 principal component and one reduction of 14 observation on the 5 variables on principal component is much more better.

Numerical comparison of statistical methods and numerical methods versus eigenvalue:

Table 3: Comparison of the eigenvalues calculated by the Monte Carlo method and Matlab software

λ_{mtlb}	$\hat{\lambda}_{mcmc}$	$\left|\lambda_{mtlb} - \hat{\lambda}_{mcmc}\right|$
6.639	6.8791	0.2401
1.786	1.8077	0.0217
0.390	0.3881	0.0019
0.230	0.2311	0.0011
0.014	0.0123	0.0017

4. Conclusion

The result of statistical methods and numerical methods in this table shows that they have not obvious difference in eigenvalue estimation. Estimating of eigenvalue even by not large enough matrix is not enough efficient. On the other hand, calculating $\det(\lambda I - A)$=0 has large error and take large mass of memory and long time calculation. However, the statistical methods in high dimensions are much more efficient than numerical methods.

5. References

[1] Alexandrov V.N., Rau-Chaplin A., Dehne F., and Taft K., Efficient Coarse Grained Monte Carlo Algorithms for Matrix Computations using PVM, LNCS 1497, pp.323-330, Springer, August 1998.

[2] Dimov I.T., Alexandrov V.N., A new highly convergent Monte Carlo method for matrix computations, Mathematics and Computers in Simulation 47 (1998) 165- 81, Bulgaria Academy of science.

[3] Fathi Vajargah B and Fathi Vajargah K., Parallel Monte Carlo computation for solving SLAE with minimum communication, Applied Mathematics and Computation (2006), 1-9.

[4] Orsythe, S. and Liebler, Matrix Inversion by a Monte Carlo method .Math. Tables other Aids Compul., 1950, 4:127-129.

[5] Montgomery D.C. and Peck E. A., Introduction to linear regression analysis, 1991.

[6] Fathi Vajargah B, Heidary- Harzavily A.,Random Numbers and Monte Carlo Approximation In Fuzzy Riemann Integral,2012,4:93-101.

[7] Mehrdoust F., Monte Carlo Simulation for Numerical Integration Based on Antithetic Variance Reduction and Haltons Sequences,2012,4:48-52.