African Journal of Mathematics and Computer Science Research Vol. 2(5), pp. 081-087, June, 2009
Available online at http://www.academicjournals.org/AJMCSR

Full Length Research Paper

Robustness of the maximum likelihood estimation procedure in factor analysis

Nwabueze, Joy Chioma[1]*, Onyeagu, Sidney I.[2] and Ikpegbu Onyedikachi[2]

[1]Department of Statistics, Abia State University, Uturu, Abia State, Nigeria.
[2]Department of Statistics, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.

Accepted 1 June, 2009

Random variables generated from five distributions were used to represent the common and specific factors in factor analysis in order to determine the robustness of the maximum likelihood estimation procedure. Five response variables were chosen for this study each with two factors. The chosen variables were transformed into linear combinations of an underlying set of hypothesized or unobserved components (factors). The result revealed that the estimates of the variance for the first factor were found to be almost the same and closely related to each other in all the distributions considered. The Chi-Square test conducted concluded that maximum likelihood method of estimation is robust in factor analysis.

Key words: Maximum likelihood, factor analysis, robustness, distributions, Random variables, Chi-Square test.

INTRODUCTION

Factor analysis is used as a tool to reduce a large set of variables to a more meaningful smaller set. It is a statistical approach that could be used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying factors. This involves a way of condensing the information contained in a number of original variables into a smaller set of factors with a minimum loss of information.

Factor analysis had a dual development beginning indirectly with the work of Pearson (1927), who used what later became known as principal component to fit regression planes to multivariate data when both dependent and independent variables were subject to error. Spearman (1904, 1913) first used the term factor analysis in the context of psychological testing for general intelligence. It was reported that children's scores were positively correlated with each other which led to a postulation that general mental ability (GMA) underlies and shapes human cognitive performance. The postulate now enjoys broad support in the field of intelligence research, where it is known as the g theory. The robustness of the maximum likelihood estimation procedure in factor analysis for this study involves the generation of sample data using Monte Carlo method for five distributions namely: Normal, Uniform, Exponential, Laplace and Gamma distributions.

Odimegwu (1999) who carried out a research on family planning attitudes and use in Nigeria using factor analysis reported that respondents who associated family planning with health benefits and improved living standard (Factor 1) were more likely to practice contraception than those who did not agree. Factor analysis can thus be used to refer to a class of models that include ordinary principal components, weighted principal components, maximum likelihood factor analysis, certain multidimensional scaling models and others.

The correlation matrixes (Pearson, 1898) of the five distributions used in this study were obtained. The contribution of total sample variance due to the first and second factors considered in this study was also obtained from the five different distributions of our study. The robustness of the maximum likelihood estimation procedure was examined using a chi- square large sample test. A robust statistical test is one that performs well even when its assumptions are violated by the true model from which the data were generated. When a system is robust, it is capable of coping well with variations (sometimes unpredictable) in its operating environment with minimal damage, alteration or loss of functionality. From theoretical point of view, the method of maximum likeli

*Corresponding author. E-mail: teeubueze@yahoo.co.uk.

hood is considered to be more robust (with some exceptions) and yields estimates with good statistical properties. This study seeks to investigate these points using data generated from five distributions. The objectives of this study also include:

1. To generate artificial data independently from each of the five distributions.
2. To calculate random varieties, matrix of constants (factor loadings) and the ith specific factor ε_i which is only associated with the ith response for each of the five distributions.
3. To test the hypothesis of κ common factors using an appropriate $\chi 2$ test for testing the hypothesis.
4. To know whether the method of maximum-likelihood estimation procedure for factor analysis is relatively insensitive to the factors for large number

EXPERIMENTAL DESIGN

Methodology

Generation of Data: The data used for this study were generated from five distributions which included Normal, Uniform, Exponential, Laplace and Gamma distributions. From each of the distributions, five response variables were generated using a sample size of 200. The experiment was replicated fifty times for each of the five distributions in the study.

THEORETICAL FRAMEWORK

The factor analytical model (Hills, 1977) is given by

$$
\begin{aligned}
x_1-\mu_1 &= L_{11}F_1+L_{12}F_2+\ldots+L_1mF_m+\varepsilon_1 \\
x_2-\mu_2 &= L_{21}F_1+L_{22}F_2+\ldots+L_2mF_m+\varepsilon_2 \qquad (1)\\
\vdots \quad & \qquad \vdots \qquad\qquad \vdots \qquad\quad \vdots \\
x_p-\mu_p &= Lp_1\,F_1+Lp_2\,F_2+Lp_m\,F_m+\varepsilon_p
\end{aligned}
$$

Which in matrix form is equivalent to

$$x-\mu = \underset{(pxm)}{L}\ \underset{(mx1)}{F}+\underset{(px1)}{\varepsilon} \qquad (2)$$

Where x is the observable random vector with p components, μ is the mean of x, L is the loading of the ith variable on the jth factor. The m unobservable random variables of F were called the common factors. The p additional sources variations of ε called error or sometimes specific factors.

The portion of the variance of the ith variable contributed by the m common factors is called the ith communality. That portion of $\operatorname{var}(x_i) = \sigma_{ii}$ due to the specific factor is called uniqueness or specific variance. Denoting the ith communality by h_i^2, we have that

$$\underset{\operatorname{var}(x_i)}{\sigma_{ii}} = \underset{communality}{L_{i1}^2+L_{i2}^2+\cdots+L_{im}^2}+\underset{Specific\ \ Variance}{\psi_i} \qquad (3)$$

Therefore, $\sigma_{ii} = h_i^2+\psi \quad i = 1, 2, \cdots, p$

Where $h_i^2 = L_u^2+L_{i2}^2+\cdots+L_{im}^2$

The ith communality is the sum of squares of the loadings of the ith variable on the m common factors. For this study, five variables and two factors are chosen, thus p is equal to 5 and m equal to 2 for each of the distributions were chosen and the basic model studied may be represented as

$$\underset{(5x1)}{x-\mu} = \underset{(5x2)}{L}\ \underset{(2x1)}{F}+\underset{(px1)}{\varepsilon} \qquad (4)$$

Maximum likelihood estimation procedure

If X is a continuous random variable with p.d.f $f(x,\theta_1,\theta_2,\cdots,\theta_k)$ where $\theta_1,\theta_2,\cdots,\theta_k$ are κ unknown parameters which need to be estimated. The likelihood function is given by

$$L(x_1,\cdots;x_n|\theta_1,\cdots;\theta_k)=\underset{i=1}{\overset{n}{\pi}} f(x_i,\theta_1,\cdots,\theta_k)\ i=1,\cdots,n \qquad (5)$$

The maximum likelihood estimation procedure chooses the estimators $\hat{\theta}_1, \hat{\theta}_2, \cdots \hat{\theta}_k$ which maximize the log likelihood function. Distributions used in the study

Normal distribution: A random variable is said to follow a normal distribution with mean μ and variance σ^2 if the probability density function is given by

$$F(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty<x<\propto \qquad (6)$$

With likelihood function

$$\underset{i=1}{\overset{n}{\pi}} f(x_i|\mu,\sigma^2) = L(\mu,\sigma^2) =$$

$$\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right\}$$

(7)

And log of the likelihood is

$$L_n L(\mu, \sigma^2) = n\log\frac{1}{\sqrt{2\pi}} - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \qquad (8)$$

Partial differentiation of $\ln L(\mu, \sigma^2)$ with respect to μ when equated to zero yields

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \quad \text{and} \qquad (9)$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

Exponential distribution: An exponential distribution with parameter λ has the probability function given as

$$f(x) = \lambda e^{-\lambda x} \quad x > 0 \qquad (10)$$

With likelihood as

$$L(\lambda) = \underset{i=1}{\overset{n}{\pi}} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} \qquad (11)$$

Taking the log of the likelihood we have

$$\ln L(\lambda) = n\ln\lambda - \lambda\sum_{i=1}^{n} x_i \qquad (12)$$

Partial differentiation of $\ln L(\lambda)$ with respect to λ when equated to zero yields

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{X}} \qquad (13)$$

Uniform Distribution: A continuous random variable has a Uniform distribution if and only if its probability density is given by

$$f(x, \alpha, \beta) = \left\{\frac{1}{\beta-\alpha} \quad \alpha < x < \beta\right. \qquad (14)$$

The likelihood function is given $\left(\frac{1}{\beta-\alpha}\right)^n$. Partial differentiation of $\log$ likelihood $\ln L(\alpha, \beta)$ with respect to α and β when equated to zero yields

$\hat{\alpha} = x_{(1)}$ = smallest sample observation

$\hat{\beta} = x_{(n)}$ = largest sample observation

Gamma distribution: A random variable x has a Gamma distribution if its probability density is given by

$$f(x) = \left\{\frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad x > 0\right. \qquad (15)$$

Where $\alpha > 0\, and\, \beta > 0$

The log likelihood is given as

$$\ln L(\alpha, \beta) = (\alpha-1)\sum_{i=1}^{n}\ln xi - \frac{\sum_{i=1}^{n} xi}{\beta} \_ n\alpha\ln(\beta) \_ n\ln\Gamma(\alpha) \qquad (16)$$

Partial differentiation of $\ln L(\alpha, \beta)$ with respect to β when equated to zero yields $\hat{\beta} = \frac{1}{n\alpha}\sum_{i=1}^{n} xi$ (17)

There is no closed form solution for α. The function is numerically very well behaved. If a numerical solution is desired, it can be found using iterative method for example Newton's Method.

Laplace Distribution: A random variable has a Laplace (u, b) distribution if its probability function is given by

$$f(x/\mu, b) = \frac{1}{2b}\begin{cases} \exp\left[\frac{-u-x}{b}\right] if\, x < u \\ \exp\left[\frac{-x-u}{b}\right] if\;\; x \geq u \end{cases} \qquad (18)$$

The likelihood function is given as

$$\text{L (u, b) = -n In 2b - } \frac{\sum_{i=1}^{n}|Xi-u|}{b} \qquad (19)$$

Which yields;

Table 1. Correlation Matrix R_1 for the Normal distribution.

	X_1	X_2	X_3	X_4	X_5
X_1	1.00000				
X_2	0.01398	1.00000			
X_3	-0.07063	0.06021	1.00000		
X_4	0.11251	0.02745	-0.00087	1.00000	
X_5	-0.04029	0.07090	-0.06888	-0.11335	1.00000

Table 2. Correlation Matrix R_2 from Exponential Distribution.

	X_1	X_2	X_3	X_4	X_5
X_1	1				
X_2	-0.00056	1			
X_3	-0.05192	0.08722	1		
X_4	0.00900	0.07219	0.03748	1	
X_5	-0.04840	-0.10320	-0.04307	0.07314	1

$$\hat{b}=\frac{1}{n}\sum_{i=1}^{n}\left|xi-\hat{\mu}\right| \qquad (20)$$

Large sample test for the number of common factors: A large sample test was employed to test the adequacy of the two common factors using the test hypothesis.

$$Ho:\ \sum{}_{pxp} = \ \hat{L}_{pxm}\ \hat{L}_{mxp} + \ \hat{\psi}_{pxp}$$

While the alternative hypothesis is H_1 Σ is equal to any other positive definite matrix. (21)

The calculated χ^2 test statistic using Bartlett's correction is

$$\chi^2_{cal} = \left[n-1-\frac{(2p+4m+5)}{6}\right]\ln\frac{\left|\hat{L}\hat{L}+\hat{\psi}\right|}{\left|S_n\right|} \qquad (22)$$

The decision rule is rejected H_0 at α level of significance if

$$\left(n-1-\frac{(2p+4m+5)}{6}\right)\ln\frac{\left|\hat{L}\hat{L}+\hat{\psi}\right|}{S_n} > \chi^2_{\frac{\left[(p-m)^2-p-m\right]}{2},\alpha} \qquad (23)$$

Provided n and n-p are large otherwise do not reject.

RESULTS AND DISCUSSION

Correlation matrix and factor loadings

The analysis of the data generated using was performed using MINITAB statistical software. The correlation matrix and the maximum likelihood factor analysis of the distributions were obtained. The result is displayed on Table 1. This is a diagonal matrix.

From Table 1, it was observed that there was no strong correlation between the variables. Variable 1 is positively correlated with x_2 and x_4 while it is negatively correlated with x_3 and x_5. The variable x_2 is positively correlated with the other variables namely x_1, x_3, x_4, and x_5 variable x_3 is negatively correlated with x_1, x_4, and x_5 while it is positively correlated with x_2 variable x_4 is negatively correlated with x_5. In all these variables the correlation level with each other is not strong whether negative or positive.

Table 2 showed poor correlation between the variables under study, variable 1 is negatively correlated with all the other variables, X_2 is positively correlated with X_3 and X_4 while X_3 is positively correlated with X_4 but negatively correlated with X_5.

Again the variables in Uniform distribution are poorly correlated with the highest negative correlation coefficient of 9 percentage occurring between the variables X_3 and X_4.

Poor correlation coefficient was also observed among the variables in the Gamma distribution with the highest negative correlation of 13% between X_2 and X_5 and highest positive correlation coefficient of 15% occurring between X_2 and X_4.

In the Laplace distribution, X_1 variable was negatively correlated with the other variables except X_5 where it had a positive correlation coefficient of 1 percent. X_2 Was positively correlated with X_3 and X_5. The variables X_3 was positively correlated with X_4 and X_5 while X_4 was negatively correlated with X_5.

The correlation matrices displayed in Table 1 - 5 were used in the maximum likelihood factor analysis for all the distributions in our study and the results were displayed Normal distribution, factor 1 is made up by mainly X_3 in Table 6. From the Table 6, it was clear that for the (be-

Table 3. Correlation Matrix R_3 from Uniform Distribution.

	X_1	X_2	X_3	X_4	X_5
X_1	1				
X_2	0.00640	1			
X_3	0.07942	0.02709	1		
X_4	0.06236	0.03939	-0.08653	1	
X_5	0.06787	0.01265	-0.06582	0.06529	1

Table 4. Correlation Matrix R_4 from Gamma Distribution.

	X_1	X_2	X_3	X_4	X_5
X_1	1				
X_2	0.03097	1			
X_3	0.04287	-0.04395	1		
X_4	-0.08892	0.14816	0.02664	1	
X_5	0.10722	-0.12989	-0.03323	-0.03890	1

Table 5. Correlation Matrix R_5 from Laplace Distribution.

	X_1	X_2	X_3	X_4	X_5
X_1	1				
X_2	-0.13866	1			
X_3	-0.01383	0.01842	1		
X_4	-0.19041	-0.02756	0.04767	1	
X_5	0.00609	0.10366	0.05393	-0.16229	1

cause of its high loading) while factor 2, is explained by the variable X_4 because of its high loading of 0.651 in absolute terms. In this research the recommendation of Stevens (1992) that factor loading with absolute values greater than 0.4 (which explained around 16% of variance) is meaningfully and should be accepted while those less than 0.4 should be ignored was adopted. Using the variance for each factor and the factor loadings, the contribution of the total. Sample variance due to each of the factors was obtained Laplace.

For example, for the Normal distribution, contribution of total sample variance due to first factor was

$$= \frac{0.177^2 + 0.060^2 + 1.000^2 + 0.001^2 + 0.069^2}{1.0134 + 0.4859} x \frac{100}{1} = 68\% \qquad (24)$$

Similarly, contribution of total sample variance due to second factor for the normal distribution is

$$= \frac{0.177^2 + 0.026^2 + 0.000^2 + 0.651^2 + 0.175^2}{1.0134 + 0.4859} x \frac{100}{1} = 32\% \qquad (25)$$

The contributions of the total variance due to the first and second factors in the other distributions were obtained in a similar manner.

For the Exponential distribution, factor 1 is explained only by variable X_5 with its high factor loading of 0.996 in absolute terms. None of the variables significantly explain factor 2 because of their low factor loadings. Variable 3 has a high factor loading of 1 for factor 1 in Uniform distribution while none of the variables have a high factor loading for factor 2 in the Uniform distribution.

In Gamma distribution, the results are similar for factor 1 but factor 2 is being explained by the variable X_2 unlike the factor 2 of the Uniform distribution that was not explained by any of the variables.

Finally, for the Laplace distribution, the results are very similar to the results obtained in Gamma distribution in that variable X_2 explains factor 2 with factor loadings of 1.000 and 0.658 respectively in absolute terms.

Contributions of the total sample variance due to each of the factors were calculated for each of the remaining distributions. For factor 1, they are 82, 83, 72 and 68% for exponential, uniform, gamma and Laplace distributions respectively. For factor 2, they are 18, 17, 28 and 32% for Exponential, Uniform, Gamma and Laplace distributions respectively.

The large sample test of was applied to all the distributions and the results is summarized in Table 7.

For all the distributions considered, the Chi-Square calculated is less than the critical value of Chi-Square tabulated at 5% level of significance; we therefore accept the hypothesis of common factor for the five distributions considered.

Conclusion

From the analysis, the robustness of the maximum likelihood estimation procedure was established in factor analysis on five different distributions with five variables and two factors. For the Normal distribution, the variance of the first and second factors respectively is 1.0134 and 0.4859 respectively while the percentage contributions of the first and second factors respectively are 68 and 32%. For the Exponential distribution, the variances of the first and second factors are 1.0121 and 0.2264 respectively while the percentage contributions of the first and second factors are 82 and 18% respectively.

For the Uniform distribution, the variance of the first and second factor is 1.0189 and 0.2087 respectively while the percentage contributions of the first and second factors are 83 and 17% respectively. For the Gamma dis-

Table 6. Maximum Likelihood Factor Analysis of the Correlation Matrix for the Distributions.

	Normal			Exponential			Uniform			Gamma			Laplace		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
X_1	0.071	-0.177	0.036	0.049	-0.071	0.007	0.079	-0.270	0.079	0.043	-0.195	0.040	0.190	0.217	0.083
X_2	-0.060	-0.026	0.004	0.104	0.344	0.129	0.027	-0.077	0.007	-0.044	0.415	0.174	0.028	-0.658	0.434
X_3	-1.000	-0.000	1.000	0.043	0.244	0.061	1.000	-0.000	1.000	1.000	-0.000	1.000	-0.048	0.040	0.004
X_4	0.001	-0.651	0.423	-0.073	0.208	0.049	-0.087	-0.262	0.076	0.027	0.312	0.098	-1.000	0.000	1.000
X_5	0.069	0.175	0.035	-0.996	0.000	0.992	-0.066	-0.248	0.066	-0.033	-0.284	0.082	0.162	-0.151	0.049
Variance	1.0134	0.4859	1.4993	1.0121	0.2264	1.2385	1.0189	0.2087	1.2275	1.0056	0.3887	1.3943	1.0656	0.5050	1.5706

1 mean factor 1, 2 means factor 2 and 3 means communalities.

Table 7. Analysis of a large sample test for the number of common factor for the five distributions.

Hypothesis	Test values	Critical value	Decision
H$_0$: $\Sigma = \hat{L}\hat{L}^1 + \hat{\psi}$ **H$_1$: Any other positive matrix**	$\chi_{cal} = \left(\dfrac{n-1-2p+4p+5}{6}\right) X \ln \dfrac{\lvert \hat{L}\,\hat{L}' + \hat{\psi} \rvert}{\lvert R \rvert}$	$\chi^2_{\frac{1}{2}}\left[(5-2)^2 - 5 - 2\right], 0.05$	
Normal	0.1190	3.86	Accept H$_0$ at m = 2
Exponential	0.5262	3.86	Accept H$_0$ at m = 2
Uniform	0.1298	3.86	Accept H$_0$ at m = 2
Gamma	2.3649	3.86	Accept H$_0$ at m = 2
Laplace	0.6819	$\chi^2_{(0.05)} = 3.86$	Accept H$_0$ at m = 2

distribution, the variances of the first and second factors are 1.0056 and 0.3887 respectively while the percentage contributions of the first and second factors are 72 and 28% respectively.

For the Laplace distribution the variance of the first and second factors are 1.0656 and 0.5050 respectively, while the percentage contribution of the first and seconds are 68 and 32 percent respectively. Since the variance of the first factor of the all the distributions considered are all within the range of1.0056 to 1.0656, and the percentage contribution from 68 - 83%, we conclude that the maximum likelihood estimation on factor analysis is robust to all the distributions considered in this study.

REFERENCES

Hills M (1977). Book Review, Appl. Stat 26; 339-340

Odimegwu CO (1999). Family Planning Attitude and use in Nigeria, A Factor Analysis www.guttmacher.org/pubs/journals

Pearson K (1898). On the Problem Error Freq. Constants on the Influence of Random Sel. Variation Correl. Phyl. Trans. R. Soc. (A) 191: 229-292.

Pearson K (1927). Math. Intell. The Sampling Errors in the Sampling Theory of a Generalized Factor, Biometrica, 19: 246-292

Spearman C (1904). Gen. Intell. Objectively Determ. and Measured, AM. J. Psych. 15: 201-293

Spearman C (1913). Correl. of Sums and Differ. Br. J. Pshych. 5: 417-426.

Stevens JP (1992). Appl. Multivar. Stat. for the Soc Sci (2nd Ed). Hillsdale, N.J. Erlbaim.