Study of Correlation using Bayes Approach under bivariate Distributions

N.S.Pandharikar* and. M.N.Deshpande**

*Govt.Vidarbha Institute of Science And Humanities ,Amravati.(M.S.) 444604.

**Ex -Director ,Institute of Science, Nagpur (M.S.)

Abstract

Coefficient of correlation plays a major role in statistical analysis. It is used as a measure of extent of association between the two random variables understudy. Present article contains the study of correlation in Bayesian framework for discrete bivariate distributions, where one variable is treated as a parameter. Estimators for linear and non-linear functions of the parameter are derived and correlations between them are studied.

1. Introduction

In statistical methodology Pearson's coefficient is one the most important concept. This concept is first introduced by Sir R. A. Fisher. It is frequently used as a measure of association between the variables in regression analysis, principal component analysis and observational studies. This concept has been considered in Bayesian framework, probably, for the first time by Dasgupta et al. (2000).They employed the concept of correlation in Bayesian framework and studied the correlations between the functions of random variables and parameter θ for univariate distribution having a parameter θ. In Bayesian framework, parameter θ is treated as a random variable having some prior distribution. They studied the correlation between two arbitrary functions of the data and parameter θ . Taking one of the variables as θ and other variable as estimator of θ, correlations between parameter and various estimators are also studied. They also studied correlation between two estimators of the same parameter.

This work for univariate distributions gave us a new insight to study the correlation and motivated us to think for the bivariate distributions. This study is carried out for discrete bivariate distributions. We have considered those discrete bivariate distributions whose marginals are Binomial, Negative Binomial (hence, in particular ,Geometric), Hypergeometric and Poisson distributions. Six types of discrete bivariate distributions are considered for this study. For these distributions, the concept of correlation between parameter and its estimator in Bayesian framework is employed in the following way.

For the bivariate distribution of (X,Y), random variable Y is considered as parameter and it has a prior distribution which involves a parameter θ (in single parameter case) .We assume that θ is known. Taking the loss function as squared error loss Bayes estimator of Y and that of Y^2 are obtained. Then the following coefficients of correlation are computed and studied.

ISSN: 2278 – 7798

1) Coefficient of correlation between Y and Bayes estimator of Y.

2) Coefficient of correlation between unbiased estimator of Y and Bayes estimator of Y.

3) Coefficient of correlation between unbiased estimator of Y^2 and Bayes estimator of Y^2.

2. Six Bivariate Distributions

The study of correlation is carried out for six types of bivariate distributions of the type given below:

I. Bivariate distributions having single parameter and θ and $0<x<\infty$, $0<y<\infty$. Here ranges of variables are independent of each other. We refer to this as bivariate distribution of type I.

II. Bivariate distribution having single parameter θ and range of one variable depends on other variable. We refer to this as bivariate distribution of type II.

III. Bivariate distribution having two parameters and $0<x<\infty$, $0<y<\infty$, that is, range of random variables are independent of each other. We refer to this as bivariate distribution of type III.

IV. Discrete bivariate distribution having two parameters and range of one variable depends on the other variables. We refer to this as bivariate distribution of type IV.

V. Bivariate distribution having three parameters and range of one variable is independent of other variable. We refer to this as bivariate distribution of type V.

VI. Bivariate distribution having three parameters, range of one variable is independent of other variable and variables are sum of the two mutually independent random variables. We refer to this as bivariate distribution of type VI.

These six bivariate distribution along with their probability mass functions and ranges of random variables are given in the following table.

ISSN: 2278 – 7798

Table 2.1: Six types of bivariate distribution

Type of distribution	Probability mass function	Range of X, Y and θ
I	$\binom{x+y}{x}\theta^{y+2}(1-\theta)^{x+y}$ 0	x, y≥ 0 , 0 < θ < 1 Otherwise
II	$\theta^3\binom{x}{y}.(1-\theta^2)^x\left(\frac{\theta^2}{1+\theta}\right)^y$ 0	0≤y ≤ x , 0< θ < 1 Otherwise
III	$\binom{x+y}{y}\theta_1^x\theta_2^y\theta_0$ 0	x, y =0,1,2… θ_0=1-θ_1-θ_2 0<θ i <θ_1+θ_2<1 i=1,2 Otherwise
IV	$\left\{\frac{\theta_2!(1-\theta_1)^{\theta_2}\theta_1^{x+2y}}{x!.y!(\theta_2-x-y)!(1-\theta_1^3)^{\theta_2}}\right.$ 0	x+y ≤θ_2 x,y >0 and 0<θ_1<1 and θ_2 integer Otherwise
V	$\frac{\binom{\theta_1}{x}\binom{\theta_2}{y}\binom{\theta_0}{n-x-y}}{\binom{N}{n}}$ 0	x =0,1,….,min(n, θ_1) y =0,1,….min(n, θ_2) Max(0,n-θ_0)$\leq x+y\leq n$ Otherwise
VI	$e^{-(\theta_1+\theta_2+\theta_{12})}\sum_{i=0}^{\min(x,y)}\frac{\theta_1^{x-i}\ \theta_2^{y-i}\ \theta_{12}^{i}}{(x-i)!\ (y-i)!\ i!}$ 0	$x\geq 0, y\geq 0$ $\theta_1,\theta_2,\theta_{12}>0$ Otherwise

ISSN: 2278 – 7798

3. Correlation in Bayesian framework

Now we study correlation in Bayesian style for the above discrete bivariate distributions where we treat random variable Y as a parameter having a prior distribution which involves parameter θ. Hence the joint probability function of (X,Y) can be expressed as

$$p(x,y) = p(x/y).h(y)$$

where p(x/y) is the probability mass function of X given y and h(y) is prior distribution of Y which involves parameter θ. Here we derive unbiased estimator of Y, Bayes estimator of Y, unbiased estimator of Y^2 and Bayes estimator of Y^2 and then study their correlations. We consider squared error as loss function. We have derived conditional distribution of X given y, p(x /y) and posterior distribution of Y, that is, p(y /X=x). Those are given in the following Table.

Table 3.1: p(x /y) and p(y /X=x)

Type of Distribution	Distribution of X given y p(x /y)	Posterior Distribution of Y p(y /X=x)
I	$NB(y+1.\theta)$	$NB(x+1,1-\theta+\theta^2)$
II	$NB(y+1,\theta)$	$B(x, \theta^2/1+\theta+\theta^2)$
III	$NB(y+1, 1-\theta_1)$	$NB(x+1, 1-\theta_2)$
IV	$B(\theta_2-y, \theta_1/(1+\theta_1))$	$B(\theta_2-x, \theta_1^2/(1+\theta_1^2))$
V	$HG(\theta_0+\theta_1, \theta_1, n-y)$	$HG(\theta_0+\theta_2, \theta_2, n-x)$
VI	$B(y, \theta_{12}/(\theta_2+\theta_{12}))+P(\theta_1)$	$B(x, \theta_{12}/(\theta_1+\theta_{12}))+P(\theta_2)$

Unbiased estimator and Bayes estimator of y and y^2 for all the six distributions can be derived easily. Those are displayed in the table.

ISSN: 2278 – 7798

Table 3.2: Unbiased estimator of Y and Bayes estimator of Y

p(y /X=x)Type of distribution	Unbiased estimator of Y $\hat{Y}$	Bayes estimator of Y $\hat{Y}_B$
I	$\left(\frac{\theta}{1-\theta}\right)X-1$	$(X+1)\left(\frac{\theta(1-\theta)}{1-\theta+\theta^2}\right)$
II	$\frac{\theta^2}{1-\theta^2}X-1$	$\left(\frac{\theta^2}{1+\theta+\theta^2}\right)$X
III	$\frac{(1-\theta_1)X}{\theta_1}-1$	$\frac{\theta_2}{(1-\theta_2)}(X+1)$
IV	$\theta_2-\left(\frac{1+\theta_1}{\theta_1}\right)X$	$(\theta_2-X)\frac{{\theta_1}^2}{(1+{\theta_1}^2)}$
V	$n-X\left(\frac{\theta_1+\theta_0}{\theta_1}\right)$	$(n-X)\frac{\theta_2}{\theta_0+\theta_2}$
VI	$(X-\theta_1)\frac{(\theta_2+\theta_{12})}{\theta_{12}}$	$\theta_2+\frac{\theta_{12}}{\theta_1+\theta_{12}}X$

Table 3.3: Unbiased estimator of Y^2 and Bayes estimator of Y^2

Type of distribution	Unbiased estimator of Y^2 $\hat{Y}$	Bayes estimator of Y^2 $\hat{Y}_B$
I	$\frac{\theta^2}{(1-\theta)^2}X^2+\frac{(2\theta^2-3\theta)}{(1-\theta)^2}X+1$	$\left[\frac{\theta^2(1-\theta)^2}{(1-\theta+\theta^2)^2}\right]X^2+\frac{\theta(1-\theta)(2\theta-2\theta^2+1)}{(1-\theta+\theta^2)^2}+\frac{\theta(1-\theta)(\theta-\theta^2+1)}{(1-\theta+\theta^2)^2}$
II	$X(X-1)\left(\frac{\theta^2}{1+\theta+\theta^2}\right)^2+X\frac{\theta^2}{1+\theta+\theta^2}$	$\frac{X\theta^2}{1+\theta+\theta^2}\left[\frac{(X-1)\theta^2}{(1+\theta+\theta^2)}+1\right]$

ISSN: 2278 – 7798

III	$\frac{(1-\theta_1)^2}{\theta_1^{\ 2}}X(X+1)-\frac{(2+\theta_1)(1-\theta_1)}{\theta_1^{\ 2}}+1$	$\theta_2^{\ 2}\frac{(X+1)(X+2)}{(1-\theta_2)^2}+\frac{\theta_2(X+1)}{(1-\theta_2)}$
IV	$\frac{(1+\theta_1)^2}{\theta_1^{\ 2}}X(X-1)-\frac{(1-2\theta_2)(1+\theta_1)}{\theta_1}X+\theta_2^{\ 2}$	$\frac{\theta_1^{\ 2}}{1+\theta_1^{\ 2}}(\theta_2-X)\left[\frac{\theta_1^{\ 2}(\theta_2-X-1)}{(1+\theta_1^{\ 2})}+1\right]$
V	$\frac{(\theta_1+\theta_0)}{\theta_1(\theta_1-1)}(\theta_1+\theta_0-1)X^2-\frac{(\theta_0+\theta_1)[2n(\theta_1-1)+\theta_0]}{\theta_1(\theta_1-1)}X+n^2$	$\frac{(n-X)\theta_2}{(\theta_0+\theta_2)}\left[\frac{(n-X-1)(\theta_2-1)}{(\theta_0+\theta_2-1)}+1\right]$
VI	$[X(X-1)-\theta_1^2]\left(\frac{\theta_{12}}{\theta_2+\theta_{12}}\right)^2+\left(\frac{\theta_2+\theta_{12}}{\theta_{12}}\right)(X-\theta_1)$ $\left(\frac{\theta_2+\theta_{12}}{\theta_{12}}\right)\left[\left(\frac{\theta_2+\theta_{12}}{\theta_{12}}\right)X^2-\frac{\theta_2 X}{\theta_{12}}-\theta_1\left\{\frac{\theta_1(\theta_2+\theta_{12})}{\theta_{12}}+1\right\}\right]$	$\left(\frac{\theta_{12}}{\theta_2+\theta_{12}}\right)^2X^2+\frac{\theta_{12}}{(\theta_1+\theta_{12})^2}X+\theta_2(\theta_2+1)$

The coefficients of correlation between parameter Y, its unbiased estimator $\hat{Y}$ and its Bayes estimator $\hat{Y}_B$ are derived for all the distribution discussed earlier.Those are listed in the following table.

Table 3.4: Coefficients of correlation

Type of distribution	coefficient of correlation		
	$\rho(Y,\hat{Y})$	$\rho(Y,\hat{Y}_B)$	$\rho(\hat{Y},\hat{Y}_B)$
I	$\frac{(1-\theta)}{\sqrt{1-\theta+\theta^2}}$	$\frac{(1-\theta)}{\sqrt{1-\theta+\theta^2}}$	1
II	$\sqrt{\frac{1}{1+\theta+\theta^2}}$	$\sqrt{\frac{1}{1+\theta+\theta^2}}$	1
III	$\sqrt{\frac{\theta_1}{(\theta_0+\theta_2)}\frac{\theta_2}{(\theta_0+\theta_1)}}$	$\sqrt{\frac{\theta_1}{(\theta_0+\theta_2)}\frac{\theta_2}{(\theta_0+\theta_1)}}$	1

ISSN: 2278 – 7798

IV	$-\sqrt{\dfrac{\theta_1^3}{(1+\theta_1^2)(1+\theta_1\ )}}$	$-\sqrt{\dfrac{\theta_1^3}{(1+\theta_1^2)(1+\theta_1\ )}}$	1
V	$-\sqrt{\dfrac{\theta_1\theta_2}{(N-\theta_1)(N-\theta_2)}}$	$-\sqrt{\dfrac{\theta_1\theta_2}{(N-\theta_1)(N-\theta_2)}}$	1
VI	$\dfrac{\theta_{12}}{\sqrt{\theta_1+\theta_{12}}\sqrt{\theta_2+\theta_{12}}}$	$\dfrac{\theta_{12}}{\sqrt{\theta_1+\theta_{12}}\sqrt{\theta_2+\theta_{12}}}$	1

4. Conclusion

From all the derivations and discussions in previous chapters we note that for such type bivariate distributions correlation between unbiased estimator and Bayes estimator of Y is always unity. But correlation between unbiased estimator and Bayes estimator of $Y^{2,}$ which is a non linear function of Y, is not unity.We also conclude that the coefficient of correlation between the parameter and the estimator is positive. But if correlation between the parameter and estimator δ is unity it does not imply that δ(X) is an appropriate estimator of the parameter θ but its linear function [aδ(X)+b] would be an appropriate for some value of a and b.

References

1) Anderson, T. W. (1984): *An introduction to multivariate analysis*, second edition, Wiley eastern private limited.
2) Berger, J. O. (1980): *Statistical decision theory and Bayesian analysis* ,second edition, Springer verlag .
3) Campbell, J. T. (1938):The Poisson correlation function, *Proceedings of the Edinburgh Mathematical Society(series2) ,vol. 4,14-26.*
4) Casella, G. and Berger, R .L., (2002): *Statistical Inference*, second edition, Thomson, Duxbury.
5) Dasgupta, A. , Casella, Dalampady, M., Ganest, C., Rubin, H. and Strawderman, W. E. (2000): Correlation in Bayesian framework*, The Canadian Journal of Statistics,* vol.28.no.4, 675-687.
6) Deshpande, M.N. and Welukar R.M. (2006): A coin tossing experiment and nineteen distributions, *Teaching Statistics, vol.28 , no.2,54-55.*

ISSN: 2278 – 7798

7) Johnson, N.L. and Kotz, S. (1969): *Discrete Distributions,* Awiley interscience publications.

8) Johnson,N.L., Kotz,S. and Balakrishnan,N. (1997):*Discrete Multivariate distributions*, Wiley series in Probability and Statistics .

9) Kocherlakota, S.and Kocherlakota,K. . (1992): *Bivariate discrete distributions*, Marcel Dekkar; Inc.

10) Phatak, A.G. and Sreehari, M. (1981):Some Characterizations of a bivariate geometric distribution, *Journal of Indian Statistical Association,vol.19,141-146.*

11) Sreehari, M. (2005): Characterizations via conditional distributions, *Journal of Indian Statistical Association , vol.43, 77-93.*

12) Sreehari,M.(2003):Bivariate Geometric Distributions revisited, *Sci. Lett. Vol.27, 13-26.*

ISSN: 2278 – 7798