DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 5 | Issue 2 Article 20

11-1-2005

Correlation Between the Number of Epileptic and Healthy Children in Family Size that Follows a Size-Biased Modified Power Series Distribution

Ramalingam Shanmugam
Texas State University, rs25@txstate.edu

Anwar Hassan
University of Kashmir

Peer Bilal Ahmad
University of Kashmir

Journal of Modern Applied Statistical Methods
November, 2006, Vol. 5, No. 2, 458-463

1538 – 9472/06/$95.00

Correlation Between the Number of Epileptic and Healthy Children in Family Size that Follows a Size-Biased Modified Power Series Distribution

Ramalingam Shanmugam
Texas State University

Anwar Hassan Peer Bilal Ahmad
University of Kashmir

An expression for the correlation between the random number of epileptic and healthy children in family whose size follows a size-biased Modified Power Series Distribution (SBMPSD) is obtained and illustrated. As special cases, results are extracted for size biased Modified Negative Binomial Distribution (SBGNBD), size biased Modified Poisson Distribution (SBGPD) and size biased Modified Logarithmic Series Distribution (SBGLSD).

Key words: Size-biased MPSD, GNBD, GLSD, GPD correlation

Introduction

Weighted distributions arise when observations are generated from a stochastic process without an equal chance of being selected from a population. When the sampling mechanism selects units with probability proportional to some measure of the unit size, the resulting distribution is called size-biased. Such distributions arise in life several studies (see Blumenthal (1967), Scheaffer (1972), Gupta (1975, 1984) for details).

Kojima and Kellehar (1962) showed that Negative Binomial Distribution (NBD) is appropriate distribution for the discrete type random observation. The random family size could follow a NBD. Gupta (1976) obtained a general expression for the correlation coefficients ‘ρ’ between the random number of boys and girls in family whose size follows a Modified Power Series Distribution (mpsd). Gupta’s (1974) introduced mpsd and explored its properties. The size biased modified power series distribution (SBMPSD) is considered as a distribution for family size. This class includes among others, size biased GNBD, GPD and GLSD. A general expression for the correlation coefficient for a random number of epileptic and healthy children in a family when the family size follows a size biased Modified Power Series Distribution (SBMPSD) is obtained and illustrated. As special cases, results are extracted for size biased Modified Negative Binomial Distribution (SBGNBD), size biased Modified Poisson Distribution (SBGPD) and size biased Modified Logarithmic Series Distribution (SBGLSD).

Ramalingam Shanmugam is Professor in the College of Health Professions. His research interests include multivariate data modeling and interpretations, informatics, optimal decision making with health/medical data support, health and medical applications of statistics. Anwar Hassan is a member of Faculty in the Department of Statistics. His research interests include discrete statistical distributions and genetic applications of statistics. Peer Bilal Ahmad works in the Department of Statistics. Ahmad's research interests include statistical inference, distribution theory among others.

Main Result: General Expression for ρ

Let N be a discrete random variable denoting the family size. Assume that this random variable is governed by a size-biased Modified Power Series Distribution (SBMPSD) whose probability mass function is

$$P[N=n]=\frac{b(n)\left(g(\alpha)\right)^{n}}{\mu f(\alpha)}, \quad n \in T$$

where T is a subset of the set of positive integers, b(n) > 0, b(n) = n a (n), g(α) and f (α) are positive, finite and differentiable functions. It is easy to establish that

$$\mu = \frac{f'(\alpha)\, g(\alpha)}{g'(\alpha)\, f(\alpha)}$$

is the mean of MPSD(see Gupta 1974). The mean and variance of size biased version of the above modified power series distribution are respectively

$$E_\alpha(N) = 1 + \frac{g(\alpha)\, f''(\alpha)}{g'(\alpha)\, f'(\alpha)} - \frac{g(\alpha) g''(\alpha)}{g'^2(\alpha)} \tag{1}$$

and

$$V_\alpha(N) = \frac{g(\alpha)}{g'(\alpha)}\ E'_\alpha(N) \tag{2}$$

Where

$$E'_\alpha(N) = \frac{\partial}{\partial\alpha} E(N).$$

Let X be the random number of epileptic children in a family of size N. Assume that X follows a binomial random variable with parameters N and p. The correlation coefficient,ρ, between X and N-X (the number of healthy children in a family of size N) is known (Rao et al [1973]) to be

$$\rho = \frac{(p[1-p])^{1/2}\,[V(N) - E(N)]}{\begin{bmatrix} pV(N) + \\ [1-p]E(N) \end{bmatrix}^{1/2} \begin{bmatrix} (1-p)V(N) + \\ pE(N) \end{bmatrix}^{1/2}}. \tag{3}$$

When p = q = ½ , the correlation coefficient reduces to

$$\rho = \frac{V(N) - E(N)}{V(N) + E(N)} \tag{4}$$

Similar results for the correlation coefficient in the case of size biased MPSD can be obtained. After algebraic simplifications, it is

$$\rho = \frac{(pq)^{1/2}\left[g(\alpha)g'(\alpha)f'(\alpha)E'_\alpha(N) - \left\{g'^2(\alpha)f'(\alpha) + f''(\alpha)g'(\alpha)g(\alpha) - g''(\alpha)g(\alpha)f'(\alpha)\right\}\right]}{\left[p\ g(\alpha)g'(\alpha)f'(\alpha)E'_\alpha(N) + q\left\{g'^2(\alpha)f'(\alpha) + f''(\alpha)g'(\alpha)g(\alpha) - g''(\alpha)g(\alpha)f'(\alpha)\right\}\right]^{1/2}} \times$$

$$\frac{1}{\left[\begin{array}{l}(1-p)g(\alpha)g'(\alpha)f'(\alpha)E'_\alpha(N) \\ +p\left\{\begin{array}{l}g'^2(\alpha)f'(\alpha) \\ +f''(\alpha)g'(\alpha)g(\alpha) \\ -g''(\alpha)g(\alpha)f'(\alpha)\end{array}\right\}\end{array}\right]^{1/2}} \tag{5}$$

where the prime denotes derivate with respect to α. For p = q = ½, expression (5) reduces to

$$\rho=\frac{\left[g(\alpha)g'(\alpha)f'(\alpha)E'_\alpha(N)-\left\{\begin{array}{l}g'^2(\alpha)f'(\alpha)+ \\ f''(\alpha)g'(\alpha)g(\alpha)- \\ g''(\alpha)g(\alpha)f'(\alpha)\end{array}\right\}\right]}{\left[g(\alpha)g'(\alpha)f'(\alpha)E'_\alpha(N)+\left\{\begin{array}{l}g'^2(\alpha)f'(\alpha)+ \\ f''(\alpha)g'(\alpha)g(\alpha)- \\ g''(\alpha)g(\alpha)f'(\alpha)\end{array}\right\}\right]} \tag{6}$$

Particular Cases

In this section, expressions are obtained for particular cases such as modified negative binomial, Poisson, and log series distributions.

Modified Negative Binomial Distribution:

In this section, a particular case of the modified negative binomial distribution (see Jain and Consul 1971) is obtained. That is,

$$P[N=n]=$$

$$\frac{\delta\Gamma(\delta+\beta n)}{x!\Gamma(\delta+(\beta-1)n+1)}\frac{\left[\alpha(1-\alpha)^{\beta-1}\right]^n}{(1-\alpha)^{-\delta}}$$

Here $f(\alpha)=(1-\alpha)^{-\delta}, g(\alpha)=\alpha(1-\alpha)^{\beta-1}$.
Note that

$$E_\alpha(N)=\frac{\delta\alpha}{(1-\beta\alpha)}+\frac{1-\alpha}{(1-\beta\alpha)^2}$$

and

$$E'_\alpha(N)=\frac{\delta-\delta\alpha\beta+2\beta-1-\alpha\beta}{(1-\alpha\beta)^3}$$

Using these values in (5) and (6), the following is obtained

$$\rho=\frac{(pq)^{1/2}\left[\begin{array}{l}\alpha(1-\alpha)(\delta-\delta\alpha\beta+2\beta-1-\alpha\beta) \\ -(1-\alpha\beta)^2(1+\delta\alpha-\alpha-\delta\alpha^2\beta)\end{array}\right]}{\left[\begin{array}{l}p\alpha(1-\alpha)(\delta-\delta\alpha\beta+2\beta-1-\alpha\beta) \\ +q(1-\alpha\beta)^2(1+\delta\alpha-\alpha-\delta\alpha^2\beta)\end{array}\right]^{1/2}}\times$$

$$\frac{1}{\left[\begin{array}{l}q\alpha(1-\alpha)(\delta-\delta\alpha\beta+2\beta-1-\alpha\beta) \\ +p(1-\alpha\beta)^2(1+\delta\alpha-\alpha-\delta\alpha^2\beta)\end{array}\right]^{1/2}}$$

For p = q = ½, the following is obtained

$$\rho=\frac{\left[\begin{array}{l}\alpha(1-\alpha)(\delta-\delta\alpha\beta+2\beta-1-\alpha\beta) \\ -(1-\alpha\beta)^2(1+\delta\alpha-\alpha-\delta\alpha^2\beta)\end{array}\right]}{\left[\begin{array}{l}\alpha(1-\alpha)(\delta-\delta\alpha\beta+2\beta-1-\alpha\beta) \\ +(1-\alpha\beta)^2(1+\delta\alpha-\alpha-\delta\alpha^2\beta)\end{array}\right]}$$

Modified Poisson Distribution

The pdf. of GPD given by Consul and Jain (1971) and Shoukri and Consul (1989) is slightly altered to represent in terms of α and β. In this section, a particular case of modified Poisson distribution (see Shoukri & Consul, 1989) is obtained. That is,

$$P[N=n]=\frac{(1+\beta n)^{n-1}\alpha^n e^{-\alpha(1+\beta n)}}{n!};$$
$$\alpha>0, |\alpha\beta|<1,\ n=0,1,2,.......$$

Here $f(\alpha)=e^\alpha,\ g(\alpha)=\alpha e^{-\alpha\beta}$

$$E_\alpha(N)=\frac{\alpha}{(1-\alpha\beta)}+\frac{1}{(1-\alpha\beta)^2}$$

$$V_\alpha(N)=\frac{2\alpha\beta}{(1-\alpha\beta)^4}+\frac{\alpha}{(1-\alpha\beta)^3}$$

Using these values in (5) and (6), the following is obtained

$$\rho=\frac{(pq)^{1/2}\left[\alpha(2\beta+1-\alpha\beta)-(1-\alpha\beta)^2\left\{1+\alpha-\alpha^2\beta\right\}\right]}{\left[p\alpha(2\beta+1-\alpha\beta)+q(1-\alpha\beta)^2\left\{1+\alpha-\alpha^2\beta\right\}\right]^{1/2}}\times\frac{1}{\left[q\alpha(2\beta+1-\alpha\beta)+p(1-\alpha\beta)^2\left\{1+\alpha-\alpha^2\beta\right\}\right]^{1/2}}$$

For p = q = ½, the following is obtained

$$\rho=\frac{\left[\alpha(2\beta+1-\alpha\beta)-(1-\alpha\beta)^2\left\{1+\alpha-\alpha^2\beta\right\}\right]}{\left[\alpha(2\beta+1-\alpha\beta)+(1-\alpha\beta)^2\left\{1+\alpha-\alpha^2\beta\right\}\right]}$$

Modified Logarithmic Series Distribution.

In this section, a particular case of modified logarithmic series distribution (see Jain & Gupta, 1973) is obtained. That is,

$$P[N=n]=\frac{\Gamma(n\beta)}{\Gamma(n+1)\Gamma([\beta-1]n+1)}\frac{\alpha^n(1-\alpha)^{(\beta-1)n}}{[-\log(1-\alpha)]}$$

Here

$$f(\alpha)=-\log(1-\alpha)=\theta^{-1}(say),\ g(\alpha)=\alpha(1\text{-}\alpha)^{\beta\text{-}1}$$

$$E_\alpha(N)=\frac{1-\alpha}{(1-\alpha\beta)^2}$$

$$V_\alpha(N)=\frac{\alpha(1-\alpha)(2\beta-\alpha\beta-1)}{(1-\alpha\beta)^4}$$

Using these values in (5) and (6), the following is obtained

$$\rho=\frac{(pq)^{1/2}\begin{bmatrix}\alpha(1-\alpha)(2\beta-\alpha\beta-1)\\-(1-\alpha\beta)^2(1-\alpha)\end{bmatrix}}{\begin{bmatrix}p\alpha(1-\alpha)(2\beta-\alpha\beta-1)\\+q(1-\alpha\beta)^2(1-\alpha)\end{bmatrix}^{1/2}\begin{bmatrix}q\alpha(1-\alpha)(2\beta-\alpha\beta-1)\\+p(1-\alpha\beta)^2(1-\alpha)\end{bmatrix}^{1/2}}$$

$$\rho=\frac{(pq)^{1/2}\begin{bmatrix}\alpha(2\beta-\alpha\beta-1)\\-(1-\alpha\beta)^2\end{bmatrix}}{\begin{bmatrix}p\alpha(2\beta-\alpha\beta-1)\\+q(1-\alpha\beta)^2\end{bmatrix}^{1/2}\begin{bmatrix}q\alpha(2\beta-\alpha\beta-1)\\+p(1-\alpha\beta)^2\end{bmatrix}^{1/2}}$$

For p = q = ½, the following is obtained

$$\rho=\frac{\left[\alpha(2\beta-\alpha\beta-1)-(1-\alpha\beta)^2\right]}{\left[\alpha(2\beta-\alpha\beta-1)+(1-\alpha\beta)^2\right]}$$

Illustration.

In this section, the main results of Section 3.2 are illustrated using data in Table 1. Results of Section 3.1 and 3.3 can be similarly made but are omitted.

Note that $\hat{p}=73/136=0.54$. Using Moment Method of estimation for estimate of α and β, the following is obtained

$$\mu_1'=\frac{\alpha}{(1-\alpha\beta)}+\frac{1}{(1-\alpha\beta)^2}$$

$$\mu_2=\frac{2\alpha\beta}{(1-\alpha\beta)^4}+\frac{\alpha}{(1-\alpha\beta)^3}$$

For convenience Let 1-αβ=θ or $\alpha\beta=1-\theta$ the following is obtained

$$\mu_1'=\frac{\alpha\theta+1}{\theta^2}$$

$$\mu_2=\frac{2(1-\theta)+\alpha\theta}{\theta^4}$$

Table 1. Number of Epileptic (x) and Healthy (N-X) children in random sample of family size (N) among 50 families

N=n	X=x	f=# families
1	0	9
2	1	13
3	2	15
4	2	9
5	3	4
#children= $\sum nf = 136$	# Epileptic Children = $\sum xf = 73$	# families= $\sum f = 50$

These give an equation in θ as

$$\mu_2\theta^4 - \mu_1'\theta^2 + 2\theta - 1 = 0$$

Replacing μ_1' and μ_2 by corresponding sample values $\bar{x}$ and S^2 respectively, the following is obtained

$$S^2\theta^4 - \bar{X}\theta^2 + 2\theta - 1 = 0$$

It is a polynomial of degree four and can be solved by using the Newton –Raphson method and so an estimate of θ can be obtained. An estimate of α is then obtained as

$$\hat{\alpha} = \frac{\hat{\theta}^2\bar{x} - 1}{\hat{\theta}}$$

After estimating $\hat{\alpha}$ and $\hat{\theta}$, $\hat{\beta}$ is obtained as

$$\hat{\beta} = \frac{1 - \hat{\theta}}{\hat{\alpha}}$$

After using Revised Table -1, the estimated values of α and β are $\hat{\alpha}$ = 2.0807, $\hat{\beta}$ = –0.0478, and $\rho = -0.310$.

In comparison to the critical chi-squared score $\chi^2_{2df} = 5.991$, the computed chi-squared score for testing the goodness of fit is $\chi^2(calculated) = 0.825$ highly insignificant implying the best fit.

References

Blumenthal, S. (1967). Proportional sampling in life length studies. *Technometrics, 9*, 205-219.

Gupta, R. C. (1975). Some characterizations of discrete distributions. *Communications in Statistics, A5*(1), 45-48.

Gupta, R. C. (1984). Some characterizations of renewal densities with emphasis in reliability. *Mathematische Operations Far Schung Und Statistics, 15*, 571-579.

Gupta, R. C. (1974). Modified power series distributions and some of its applications. *Sankhya, series B, 36*, 288-298.

Gupta, R. C. (1976). Some applications of modified power series distributions in genetics. *Sankhya, series B, 38*, 187-191.

Kojima, K. I. & Kelleher, T. M. (1962). Survival of mutant genes. *The American Naturalist, XCVI*, 329-45.

Jain, G. C. & Consul, P. C. (1971). A modified negative binomial distribution. *SIAM, Journal of Applied Mathematics, 21*, 501-513

Jain, G. C. & Gupta, R. P. (1973). A logarithmic type distribution. *Trabjos Estadiest., 24*, 99-105.

Shoukri, M. M. & Consul, P. C. (1989). Bayesian analysis of modified Poisson distribution. *Communication and Statistical Theoretical Mathematics, 18*(4), 1465-1480.

Scheaffer, R. L. (1972). Size biased sampling. *Technometrics, 14*, 635-644.

Rao, B. R., Mazumdar, S., Waller, J. H., & Li, C. C. (1973). Correlation between the numbers of two types of children in a family, *Biometrics, 29*, 271-279.