THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

Date June 4, 2010

WP # 0108MSS-213-2009

A Generalized Poly-logseries Type Distribution

Ram C. Tripathi
Department of Management Science and Statistics
College of Business
University of Texas at san Antonio

UTSA COLLEGE OF BUSINESS KNOWLEDGE FOR A NEW WORLD

ONE UTSA CIRCLE
SAN ANTONIO, TEXAS 78249-0631
210 458-4317 | BUSINESS.UTSA.EDU

Title: A Generalized Poly-logseries Type Distribution

by

Ram C. Tripathi

Department of Management Science and Statistics

College of Business

University of Texas at san Antonio

Abstract

The log-series distribution has been found useful to describe many natural phenomena such as the data on species abundance. These types of data typically have long tails. However, the ordinary log-series distribution does not have sufficiently long tail and is found unsuitable for some data with extra heavy tails. There are many generalized log-series distributions available in the literature. Following the idea of Conway and Maxwell (1962) to generalize the Poisson distribution (COM-Poisson), we start with a COM-negative binomial distribution and use it to generate a COM-log-series distribution. We present derivation of the new model, investigate its various characteristics including the probability generating function, its probability function (pf). We also present recurrence relations for probabilities and moments. We compare graphically the pf's for various members of this class. We also investigate the behavior of the failure rate of this model. Finally, we develop some methods for estimating its parameters and give examples with real data.

Key Words

COM-Poisson, COM-Negative Binomial, Generalized poly-logseries distribution, poly-hypergeometric function, Over- and under-dispersion, failure rate, maximum likelihood estimation,

Acknowledgements: This work is supported by the College of Business Summer Grants 2009, UTSA.

JEL Classification: C16

1 Introduction

The log-series distribution was developed by Fisher to describe the species abundance data such as the distribution of catches of moths in light traps and the distribution of tropical butterflies (Fisher, et. al (1943), Boswell and Patil (1971)). These types of data typically have long tails. This distribution was developed to fit such data and was obtained as the limit of a zero-truncated negative binomial. In practice, however, there are data with even longer tails which can not be described well by the log-series distribution. There have been many generalizations of this model in the literature which are based on some of the extensions of the negative binomial distribution. Jain and Gupta (1973) presented a generalized log series distribution (GLSD) based on the Lagrangian generalized negative binomial distribution and used the method of moments to fit their model to some data. Kempton (1975) developed a GLSD from a negative binomial type distribution which is suitable for describing data with extra long tails. Tripathi and Gupta (1985) presented a GLSD based on the shifted version of the generalized negative binomial distribution of Tripathi and Gurland (1977)(see also Tripathi, Gupta and White(1987)). Recently Kemp (2004)considered a genearalization of log-series model, called the poly-logseries, and investigated its many characteristics.

In this paper, we present an extension of the log-series distribution based on a COM-negative binomial type distribution. We give its probability generating function (pgf), and the probability function (pf). We also provide recurrence relations for its pf as well as its moments and investigate its over- and under-dispersion properties graphically. We provide graphical comparisons for the pf's of some of its members including the ordinary log-series distribution. We also develop some methods for estimating its parameters including the maximum likelihood estimates. The likelihood equations are non-linear functions of the parameters, we obtain the maximum likelihood estimates by using numerical libraries from IMSL (International mathematical and Statistical Libraries). Finally, we include some examples for fitting this model to real data.

2 Model Development

Motivated by the COM-Poisson distribution (Conway and Maxwell (1962), Shmueli et. al. (2005)), and the recent poly-logarithmic distribution of Kemp(2004), we develop a general class of distributions called the COM-logseries type distribution. For this, we define a generalized hypergeometric function, called poly-hypergeometric function (phf) as follows:

$$ {}_pF_q^{\gamma}(a_1, a_2, \cdots, a_p; b_1, b_2, \cdots, b_q; \theta) = \sum_{j=0}^{\infty} \frac{(a_1)_j (a_2)_j \cdots (a_p)_j}{(b_1)_j (b_2)_j \cdots (b_q)_j} \frac{\theta^j}{(j!)^{\gamma}}, $$

where

$$(a)_j = a(a+1)(a+2)\cdots(a+j-1).$$

Now, consider the probability generating function (pgf) of the shifted generalized negative binomial distribution of Tripathi and Gurland (1977) expressed in terms of the phf

$$H(z) = \frac{z^s \ {}_2F_1^{\gamma}(\frac{\alpha}{\beta}+s, 1; \lambda+s; \beta z)}{{}_2F_1^{\gamma}(\frac{\alpha}{\beta}+1, 1; \lambda+1; \beta)}, \ \alpha > 0, \ \lambda > -s, \ 0 < \beta < 1, \tag{1}$$

where s is a non-negative integer. The distribution developed here will be called the phf-generalized logseries distribution (phf-GLSD).

Models included in (1) as special cases

1. For $s = 0, \ \lambda = \gamma = 1$, (1) reduces to the Katz family of distributions (Katz(1965)) which includes the Poisson distribution for $\beta \to 0$; negative binomial distribution, and the binomial distribution and their generalizations.
2. For $s = 0, \ \gamma = 1$, it reduces to the hyper Poisson distribution of Crow and Bardwell (19) as $\beta \to 0$.
3. For $s = 0, \ \gamma \neq 1$, the above includes COM versions of Poisson (Conway and Maxwell (1962), Shmuelli et. al. (2005)), hyper Poisson, negative binomial and the binomial distributions depending on various conditions on β and λ.

Now, we utilize the above model to derive the proposed phf-GLSD which is given in the following result:

Result:

The pgf of the proposed phf-GLSD is given by

$$G(z) = \frac{z \ {}_2F_1^{\gamma}(1, 1; \lambda+1; \beta z)}{{}_2F_1^{\gamma}(1, 1 : \lambda+1; \beta)}, \ \alpha > 0, \lambda > -1, \ 0 < \beta < 1. \tag{2}$$

The above result is obtained by expanding the phf involved in (1) and taking the limit as $\alpha \to 0$. The probabilities of the new model are the coefficients of z^j in the expansion of $G(z)$. If X denotes the corresponding random variable, it can be seen that

$$P_1 = P(X=1) = \frac{1}{{}_2F_1^{\gamma}(1, 1; \lambda+1; \beta)},$$

and

$$P_j = P(X=j) = \frac{(j-1)!)^2 \beta^{j-1}}{P_1 (\lambda+1)_{j-1} [(j-1)!]^{\gamma}}, \ j = 2, 3, \cdots.$$

3 Some Characterizations of the Model

In this section we investigate some useful characterizations of this such as the recurrence relations for its probabilities and the moments. We also present graphical characterizations involving its over- and under-dispersion as well its failure rate which exhibits both increasing failure rate (IFR) and decreasing failure rate (DFR) properties.

3.1 Recurrence Relations for Probabilities

The recurrence relation for the probabilities is given by

$$\frac{P_{j+1}}{P_j} = \frac{\beta}{(\lambda + j)(j^{\gamma-2})}.$$

This can be written as

$$(\lambda + j)P_{j+1} = \beta j^{2-\gamma} P_j, \ j = 1, 2, \cdots. \tag{3}$$

This relation can be used to compute the probabilities of the model starting from P_1.

3.2 Recurrence relations for moments

On multiplying both sides of (3) by $(j+1)^r$ and summing over j, we get

$$\sum_{j=1}^{\infty}(\lambda + j)(j+1)^r P_{j+1} = \beta \ \sum_{j=1}^{\infty}(j+1)^r j^{2-\gamma} P_j.$$

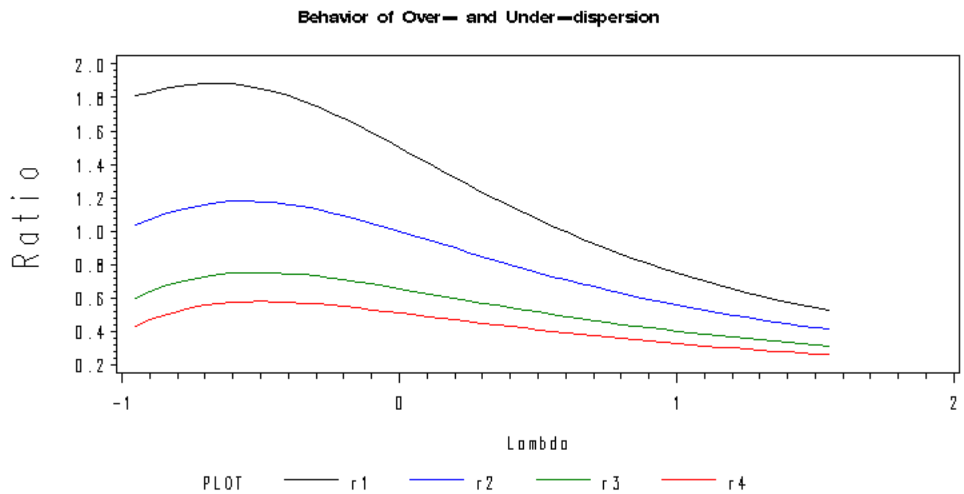
This can be simplified to

$$\mu'_{r+1} + (\lambda - 1)\mu'_r - \lambda P_1 = \beta \ \sum_{i=0}^{r} \binom{r}{i} \mu'_{i+2-\gamma}, \ r = 0, 1, 2, \cdots. \tag{4}$$

Closed form expressions for the mean and the variance for this distribution are not easy to derive.

3.3 Over- and under-dispersion of the phf-GLSD model

The following graph presents a comparison of the phf-GLSD for four combination of parameters.



The four curves in the above graph plot the ratios r_1, r_2, r_3 and r_4 for $\gamma = .9, 1.0, 1.2, 1.4.$ respectively for $\beta = .5$ for $\lambda = -.95$ to .60 in increments of .05. It can be seen that for $\gamma > 1.0$, the model is under-dispersed regardless of the values of λ. However, for $\gamma \leq 1.0$, the model exhibits both the over- and under-dispersion. If γ is small, the model remains over-dispersed for larger positive values of λ. Thus, smaller values of γ and small negative values of λ both make the model over-dispersed. Larger values of γ and positive values of λ make the model under-dispersed.

3.4 Comparison of Log-series and Com-Log-Series Distributions

In this section we present some comparisons between the pf's of the log-series and the proposed phf-GLSD models. The calculations needed for the graphical comparison are in progress and the comparisons will be presented shortly. A similar graphical comparison of the failure rate of this model will also be presented.

Effect of λ, β and γ on the tail of the distribution

λ	β	γ	Terms	Var	Mean	Var/Mean
1.5	.8	1.5	11	.650	1.462	.445
1.5	.9	2.0	9	.476	1.417	.336
.8	.5	1.5	10	.443	1.353	.327
.8	.5	1.0	15	.944	1.507	.626
.8	.5	.95	17	1.108	1.544	.718
.8	.5	.9	20	1.371	1.594	.860
.8	.5	.85	27	1.908	1.671	1.142
.8	.5	.83	34	2.394	1.721	1.391
.8	.5	.82	48	3.034	1.761	1.723

The above table gives the effect of various parameters on the tail of the distribution. The column titled "Terms" effectively represents the tail of the distribution. The larger the values of "term", the longer is the tail, smaller the value of "Terms", the smaller the tail. From the table, it can be seen that larger values of λ and/or γ both lead to a short-tailed model, whereas, smaller values of λ and/or γ both lead to a long-tailed tailed model.

4 Estimation of Parameters

In this section, we develop methods to estimate the parameters of this model which will help in fitting this model to real data. We consider two types of estimators: Moment type estimators and maximum likelihood estimators.

4.1 Some useful relations in parameter estimation

Since the likelihood equations for this model are non-linear (as seen in the following section), we need initial estimates for the parameters to start the iteration. For this, we use the following three relations to yield initial estimates. These are

$$P_1 = [{}_2F_1^{\gamma}(1, 1; (\lambda + 1; \beta)]^{-1},$$

$$\lambda(1 - P_1) - \beta\mu'_{2-\gamma} = 1 - \mu,$$

and

$$\lambda(\mu - 1) - \beta\mu'_{3-\gamma} = 2\mu - \mu'_2.$$

The last two equations are obtained by taking $r = 0,$ and 1 in (3).

4.2 Maximum likelihood estimation

In this section, we present likelihood equations for the parameters of this model. These equations can be solved to give the mle's for the parameters β, λ and γ. Suppose, that the random sample is available as a grouped data with the frequencies $n_1, n_2, \cdots, n_k$ at counts $1, 2, \cdots, k$; with k being the largest count. Then, the likelihood functions is

$$L = \prod_{j=1}^{k} P_j^{n_j},$$

and the log likelihood is

$$\begin{aligned} ln(L) &= \sum_{j=1}^{k} n_j ln(P_j) \\ &= \sum_{j=1}^{k} n_j \ ln\left[\frac{[(j-1)!]^2 \beta^{j-1}}{P_1(\lambda+1)_{j-1}[(j-1)!]^\gamma}\right] \\ &= 2\sum_{j=1}^{k} n_j \ ln[(j-1)!] + n(1-\bar{x})ln(\beta) - n \ ln(P_1) \\ &\quad -\sum_{j=1}^{k} n_j \left[\sum_{i=1}^{j-1} ln(\lambda+i)\right] - \gamma \sum_{j=1}^{k} n_j \ ln(j-1)!. \end{aligned}$$

The likelihood equations are given by

$$\frac{\partial lnL}{\partial \gamma} = -\frac{n}{P_1}\frac{\partial P_1}{\partial \gamma} - \sum_{j=1}^{k} n_j \ ln(j-1)!,$$

$$\frac{\partial lnL}{\partial \beta} = -\frac{n}{P_1}\frac{\partial P_1}{\partial \beta} + \frac{n}{\beta}(\bar{x}-1),$$

and

$$\frac{\partial lnL}{\partial \lambda} = -\frac{n}{P_1}\frac{\partial P_1}{\partial \lambda} - \sum_{j=1}^{k} n_j \sum_{i=1}^{j-1} \frac{1}{\lambda+i}.$$

The partial derivatives of P_1 with respect to $\beta,\ \lambda,$ and γ are given in the Appendix. In order to solve the non-linear likelihood equations given above, we need the hessian, the second order partial derivatives of $ln(L)$. These also appear in the Appendix.

Appendix

Partial derivatives of P_1

We need the first and second order partial derivatives of P_1 which are given below.

$$P_i = [{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)]^{-1} = \left[\sum_{j=1}^{k} \frac{(j!)^2}{(\lambda+1)_j} \frac{\beta^j}{(j!)^{\gamma}}\right]^{-1}.$$

$$\frac{\partial P_1}{\partial \beta} = -\left[\sum_{j=1}^{k} \frac{(j!)^2}{(\lambda+1)_j} \frac{\beta^j}{(j!)^{\gamma}}\right]^{-2} \left[\sum_{j=1}^{k} \frac{(j!)^2}{(\lambda+1)_j} \frac{j\ \beta^{j-1}}{(j!)^{\gamma}}\right],$$

$$\frac{\partial P_1}{\partial \lambda} = \left[\sum_{j=1}^{k} \frac{(j!)^2}{(\lambda+1)_j} \frac{\beta^j}{(j!)^{\gamma}}\right]^{-2} \left[\sum_{j=1}^{k} \frac{(j!)^2}{(\lambda+1)_j} \frac{j\ \beta^{j-1}}{(j!)^{\gamma}} \sum_{k=1}^{j} \frac{1}{\lambda+k}\right],$$

and

$$\frac{\partial P_1}{\partial \gamma} = \left[\sum_{j=1}^{k} \frac{(j!)^2}{(\lambda+1)_j} \frac{\beta^j}{(j!)^{\gamma}}\right]^{-2} \left[\sum_{j=1}^{k} \frac{(j!)^2}{(\lambda+1)_j} \frac{j\ \beta^{j-1}}{(j!)^{\gamma}}\ ln(j!)\right].$$

The second order partial derivatives of $ln(L)$ are as follows;

$$\frac{\partial^2 ln(L)}{\partial \beta^2} = -\frac{n(\bar{x}-1)}{\beta^2} + \frac{n}{P_1^2}\left(\frac{\partial P_1}{\partial \beta}\right)^2 - \frac{n}{P_1}\left(\frac{\partial^2 P_1}{\partial \beta^2}\right),$$

$$\frac{\partial^2 ln(L)}{\partial \gamma \partial \beta} = \frac{n}{P_1^2}\left(\frac{\partial P_1}{\partial \beta}\right)\left(\frac{\partial P_1}{\partial \gamma}\right) - \frac{n}{P_1}\left(\frac{\partial^2 P_1}{\partial \gamma \partial \beta}\right),$$

$$\frac{\partial^2 ln(L)}{\partial \lambda \partial \beta} = \frac{n}{P_1^2}\left(\frac{\partial P_1}{\partial \beta}\right)\left(\frac{\partial P_1}{\partial \lambda}\right) - \frac{n}{P_1}\left(\frac{\partial^2 P_1}{\partial \lambda \partial \beta}\right),$$

$$\frac{\partial^2 ln(L)}{\partial \lambda^2} = \frac{n}{P_1^2}\left(\frac{\partial P_1}{\partial \lambda}\right)^2 - \frac{n}{P_1}\left(\frac{\partial^2 P_1}{\partial \lambda^2}\right) + \sum_{j=1}^{k} n_j \sum_{i=1}^{j-1}(\lambda+i)^{-2},$$

$$\frac{\partial^2 ln(L)}{\partial \gamma \partial \lambda} = \frac{n}{P_1^2}\left(\frac{\partial P_1}{\partial \gamma}\right)\left(\frac{\partial P_1}{\partial \lambda}\right) - \frac{n}{P_1}\left(\frac{\partial^2 P_1}{\partial \gamma \partial \lambda}\right).$$

$$\frac{\partial^2 ln(L)}{\partial \gamma^2} = \frac{n}{P_1^2}\left(\frac{\partial P_1}{\partial \gamma}\right)^2 - \frac{n}{P_1}\left(\frac{\partial^2 P_1}{\partial \gamma^2}\right).$$

The second partial derivatives of P_1 needed in the above expressions are given below.

$$\begin{aligned}\frac{\partial^2 P_1}{\partial \beta^2} &= 2\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-3}\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{j\beta^{j-1}}{(j!)^{\gamma}}\right]^2 \\ &\quad -\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-2}\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{j(j-1)\beta^{j-2}}{(j!)^{\gamma}}\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 P_1}{\partial \lambda \partial \beta} &= -2\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-3}\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{j\beta^{j-1}}{(j!)^{\gamma}}\right]\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{j\beta^{j-1}}{(j!)^{\gamma}}\sum_{k=1}^{j}\frac{1}{\lambda+k}\right] \\ &\quad +\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-2}\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{j\beta^{j-1}}{(j!)^{\gamma}}\sum_{k=1}^{j}\frac{1}{\lambda+k}\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 P_1}{\partial \gamma \partial \beta} &= 2\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-3}\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{j\beta^{j-1}}{(j!)^{\gamma}}\right]\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{j\beta^{j-1}}{(j!)^{\gamma}}ln(j!)\right] \\ &\quad +\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-2}\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{j\beta^{j-1}}{(j!)^{\gamma}}ln(j!)\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 P_1}{\partial \lambda^2} &= -2\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-3}\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{\beta^{j}}{(j!)^{\gamma}}\sum_{k=1}^{j}\frac{1}{\lambda+k}\right]^2 \\ &\quad -\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-2}\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{\beta^{j}}{(j!)^{\gamma}}\left(\sum_{k=1}^{j}\frac{1}{\lambda+k}\right)^2+\sum_{k=1}^{j}\frac{1}{(\lambda+k)^2}\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 P_1}{\partial \gamma \partial \lambda} &= 2\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-3}\left[\sum_{j=0}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{\beta^{j}}{(j!)^{\gamma}}ln(j!)\right]\left[\sum_{j=0}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{\beta^{j}}{(j!)^{\gamma}}\sum_{k=1}^{j}\frac{1}{\lambda+k}\right] \\ &\quad +\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-2}\left[\sum_{j=0}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{\beta^{j}}{(j!)^{\gamma}}ln(j!)\sum_{k=1}^{j}\frac{1}{\lambda+k}\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 P_1}{\partial \gamma^2} &= 2\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-3}\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{\beta^{j}}{(j!)^{\gamma}}[ln(j!)]^2\right] \\ &\quad -\left[{}_2F_1^{\gamma}(1,1;\lambda+1;\beta)\right]^{-2}\left[\sum_{j=1}^{k}\frac{(j!)^2}{(\lambda+1)_j}\frac{\beta^{j}}{(j!)^{\gamma}}(ln(j!))^2\right].\end{aligned}$$

References

1. Boswell, M. T. and Patil, G. P. (1971). Chance mechanisms generating the logarithmic series distributions used in the analysis of number of species and individuals. In International Symposium on Statistical Ecology, Eds. G. P. Patil, E. C. Pielou and D. E. Waters, Penn. State Univ. Press, pp. 99-130.

2. Conway, R. W. and Maxwell, W. L. (1962). A queuing model with state dependent service rates. Journal of Industrial Engineering, 12, pp. 132-136.

3. Erdelyi, A. et al (1953). Higher transcendental functions. McGraw Hill, New York.

4. Fisher, R. A., Corbet, A. S., and Wiliams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. J. Animal Ecology, 12, pp. 42-58.

5. Jain, G. C. and Gupta, R. P. (1973). A logarithmic series type distribution. Trab. Estadistica, 24, pp. 99-105.

6. Kemp, A. W. (2004). Pologarithmic Distributions. Encyclopedia of Statistical Sciences.

7. Kempton, R. A. (1975). A generalized form of Fisher's logarithmic series. Biometrika, 62, pp. 28-38.

8. Stacy, E. W. (1962). A generalization of the gamma distribution. Annals of Mathematical Statistics, 33, 1187-1192.

9. Tripathi, R. C. and Gurland, J. (1977). A general family of discrete distributions with hypergeometric probabilities. Journal of the Royal Statistical Society, b39, pp. 349-356.

10. Tripathi, R. C. and Gupta, R. C. (1985). A generalization of the log-series distribution. Communications in Statistics, A14, pp. 1779-1799.

11. Tripathi, R. C., Gupta, R. C. and White, T. J. (1987). Some generalizations of the geometric distribution. Sankhya, B49, pp. 218-223.