ISSN 2250-3153

Important Summaries and Graphs of the Mixture of Two Logarithmic Distributions

Epimaco A. Cabanlit, Jr. [*], Jennifer T. Bulawan[**],

[*] Mathematics Department, Mindanao State University, General Santos City, Philippines
[**] Mathematics Department, Mindanao State University, General Santos City, Philippines

Abstract- The mixture of distributions can serve as a model to some realities where the population consists of heterogeneous components. In this paper we present some of the important summaries of the mixtures of two logarithmic distributions such as the moment generating function, mean and variance. Moreover, various graphs of the mixture with different values of mixing parameters are presented.

Index Terms- Probability Distributions, Logarithmic Distribution, Mixture of Distributions, Moment Generating Functions, Mean and Variance.

I. Introduction

The logarithmic series distribution was introduced by Fisher, Corbett and Williams (1943) to investigate the distribution of butterflies in the Malayan Peninsula. It has been used in the sampling of quadrants for plant species, the distribution of animal species, population growth and in economic applications. Chatfield et al (1966) used the logarithmic series distribution to represent the distribution of numbers of items of a product purchased by a buyer in a specified time period. They point out that the logarithmic series has the advantage of dependency on only one parameter θ [13].

In probability and statistics, the logarithmic distribution also known as the logarithmic series distribution or the log-series distribution is a discrete probability distribution derived from the Maclaurin series expansion,

$$-\ln(1-p) = p + \frac{p^2}{2} + \frac{p^3}{3} + \cdots. \tag{1}$$

From this we obtain the identity

$$\sum_{x=1}^{\infty} \frac{-1}{\ln(1-p)} \frac{p^x}{x} = 1\ . \tag{2}$$

This leads directly to the probability mass function of a Log(p)-distributed random variable:

$$f(x) = \frac{-1}{\ln(1-p)} \frac{p^x}{x} \tag{3}$$

for $x \geq 1$ an integer and where $0 < p < 1$. Because of the identity above, the distribution is properly normalized.

The cumulative distribution function is

$$F(x) = 1 + \frac{B(\,p; x+1, 0)}{\ln(1-x)} \tag{4}$$

where B is the incomplete beta function.

The formula for the mixture of two logarithmic distributions is defined as

$$P(x) = \varphi_1 \frac{1}{-\ln(1-p_1)} \frac{{p_1}^{x}}{x} + \varphi_2 \frac{1}{-\ln(1-p_2)} \frac{{p_2}^{x}}{x}, \tag{5}$$

where $x \in \mathbb{N}$ such that $\varphi_1 + \varphi_2 = 1$.

This paper presents some important properties of the mixture of two logarithmic distributions, such as the moment generating function, mean and variance. Moreover, various graphs of some mixtures are presented.

II. Important Summaries

In this section, we present some important properties such as the moment generating function, the mean and the variance.

Theorem 1. The moment generating function of the mixture of two logarithmic distributions is,

$$M(t) = \varphi_1 \frac{\ln(1-p_1\, e^t)}{\ln(1-p_1)} + \varphi_2 \frac{\ln(1-p_2\, e^t)}{\ln(1-p_2)}$$

Proof:

The moment generating function of a probability distribution is given by,

$$M(t) = \sum\nolimits_x e^{tx}\, P(x)$$

where $P(x)$ is the probability mass function of the mixture of two logarithmic distributions in equation (5). We have

$$\begin{aligned} M(t) &= \sum_{x=1}^{\infty} e^{tx}\left[\varphi_1\left(\frac{1}{-\ln(1-p_1)}\right)\left(\frac{{p_1}^x}{x}\right) + \varphi_2\left(\frac{1}{-\ln(1-p_2)}\right)\left(\frac{{p_2}^x}{x}\right)\right] \\ &= \varphi_1\left(\frac{1}{-\ln(1-p_1)}\right)\sum_{x=1}^{\infty}\frac{e^{tx}{p_1}^x}{x} + \varphi_2\left(\frac{1}{-\ln(1-p_2)}\right)\sum_{x=0}^{n}\frac{e^{tx}{p_2}^x}{x} \\ &= \varphi_1\left[\frac{1}{-\ln(1-p_1)}\right][-\ln(1-p_1e^t)] + \varphi_2\left[\frac{1}{-\ln(1-p_2)}\right][-\ln(1-p_2e^t)] \\ &= \varphi_1\left[\frac{\ln(1-p_1\, e^t)}{\ln(1-p_1)}\right] + \varphi_2\left[\frac{\ln(1-p_2\, e^t)}{\ln(1-p_2)}\right]. \end{aligned}$$

Theorem 2. The mean of the mixture of two logarithmic distributions is,

$$E(x) = \varphi_1\left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)}\right] + \varphi_2\left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)}\right].$$

Proof: The mean is defined as

$$E(x) = \sum_{x=1}^{\infty} x \bullet P(x)$$

where $P(x)$ is the probability density function of the mixture of two logarithmic distributions in equation (5).

Then,

$$E(x) = \sum_{x=1}^{\infty} x \bullet P(x)$$

$$= \sum_{x=1}^{\infty} x \left[\varphi_1 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)}\right] + \varphi_2 \left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)}\right]\right]$$

$$= \sum_{x=1}^{\infty} \varphi_1 x \left[\frac{1}{-\ln(1-p_1)}\right]\left(\frac{{p_1}^x}{x}\right) + \sum_{x=1}^{\infty} \varphi_2\, x \left[\frac{1}{-\ln(1-p_2)}\right]\left(\frac{{p_2}^x}{x}\right)$$

$$= \frac{\varphi_1}{-\ln(1-p_1)} \sum_{x=1}^{\infty} {p_1}^x + \frac{\varphi_2}{-\ln(1-p_2)} \sum_{x=1}^{\infty} {p_2}^x$$

$$= \left[\frac{\varphi_1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{-\ln(1-p_1)}\right] + \left[\frac{\varphi_2}{-\ln(1-p_2)}\right]\left[\frac{p_2}{-\ln(1-p_2)}\right].$$

Lemma 1. Given the mixture of two logarithmic distributions in (5) and for $x \geq 1, x \in \mathbb{N}$.

$$E(x^2) = \varphi_1 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)^2}\right] + \varphi_2 \left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)^2}\right]$$

Proof:

$$E(x^2) = \sum_{x=1}^{\infty} x^2 \bullet P(x)$$

$$E(x^2) = \sum_{x=1}^{\infty} x^2 \left[\varphi_1 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)^2}\right] + \varphi_2 \left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)^2}\right]\right]$$

$$= \sum_{x=1}^{\infty} \varphi_1 x^2 \left[\frac{1}{-\ln(1-p_1)}\right]\left(\frac{{p_1}^x}{x}\right) + \sum_{x=1}^{\infty} \varphi_2\, x^2 \left[\frac{1}{-\ln(1-p_2)}\right]\left(\frac{{p_2}^x}{x}\right)$$

$$= \varphi_1 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)^2}\right] + \varphi_2 \left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)^2}\right].$$

Theorem 3. The Variance of the Mixture of two Logarithmic Distributions is

$$Var(X) = \varphi_1 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)^2}\right] + \varphi_2 \left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)^2}\right] -$$

$$\left[\varphi_1 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)}\right]\right]^2 - 2\varphi_1\varphi_2 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)}\right]\left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)}\right]$$

$$-\left[\varphi_2 \left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)}\right]\right]^2.$$

Proof: The variance is defined as

$$Var(X) = E(x^2) - [E(x)]^2.$$

By Lemma 1and Theorem 2, we have

$$Var(X) = \varphi_1 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)^2}\right] + \varphi_2 \left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)^2}\right] -$$

$$\left[\varphi_1 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)}\right] + \varphi_2 \left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)}\right]\right]^2$$

$$= \varphi_1 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)^2}\right] + \varphi_2 \left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)^2}\right] -$$

$$\left[\varphi_1 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)}\right]\right]^2 - 2\varphi_1\varphi_2 \left[\frac{1}{-\ln(1-p_1)}\right]\left[\frac{p_1}{(1-p_1)}\right]\left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)}\right]$$

$$-\left[\varphi_2 \left[\frac{1}{-\ln(1-p_2)}\right]\left[\frac{p_2}{(1-p_2)}\right]\right]^2.$$

III. GRAPHS OF THE MIXTURE

The following are some of the graphs of the mixture of two logarithmic distributions with different values of their parameters.

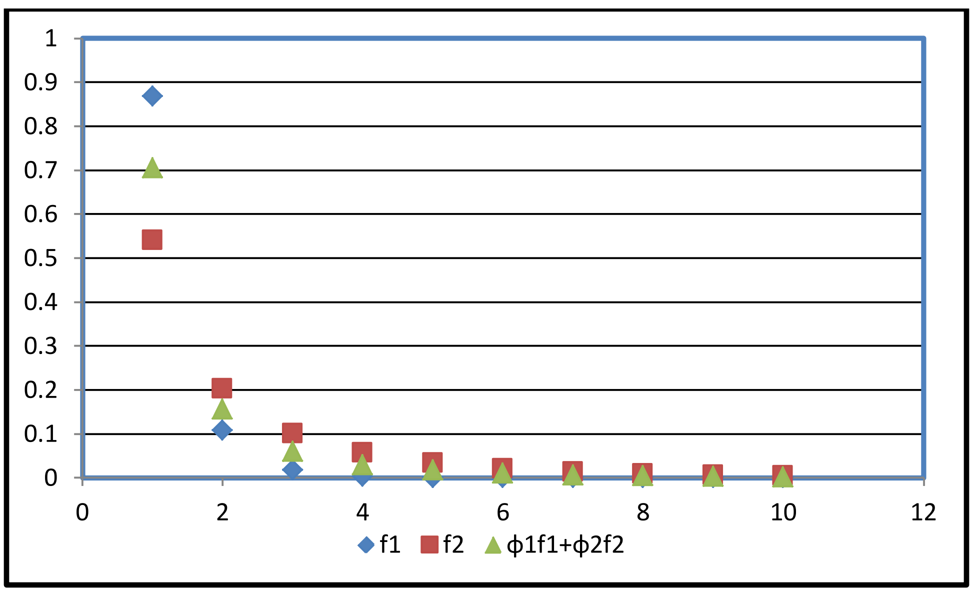

Figure 1.The Two Logarithmic Distributions and its Mixture ($\varphi_1 = 0.5, \varphi_2 = 0.5$)

Figure 1 represents the graph of probability distributions of f_1, f_2 and the mixture $\varphi_1 f_1 + \varphi_2 f_2$ where f_1 is the graph of the logarithmic distributions where $p_1 = 0.25$, f_2 is the graph of logarithmic distribution where $p_2 = 0.75$. $\varphi_1 f_1 + \varphi_2 f_2$ is the graph of the mixture of two logarithmic distributions. The mixture has the average of the graphs of f_1 and f_2 since $\varphi_1 = \varphi_2 = 0.5$.

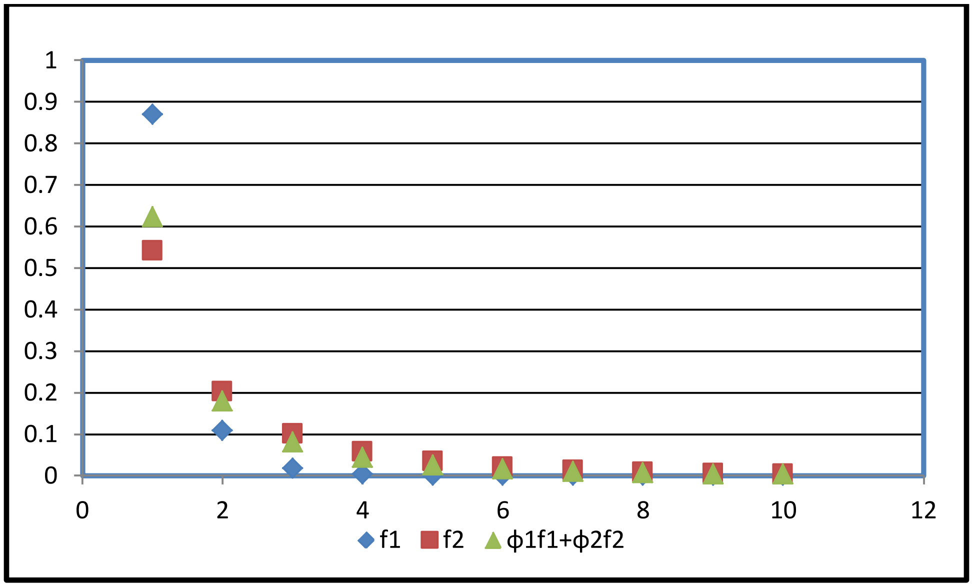

Figure 2. The Two Logarithmic Distributions and its Mixture ($\varphi_1 = 0.25, \varphi_2 = 0.75$)

Figure 2 represents the graph of probability distributions of f_1, f_2 and the mixture $\varphi_1 f_1 + \varphi_2 f_2$ where f_1 is the graph of the logarithmic distributions where $p_1 = 0.25$. f_2 is the graph of logarithmic distribution where $p_2 = 0.75$. $\varphi_1 f_1 + \varphi_2 f_2$ is the graph of the mixture of two logarithmic distributions. The mixture has a graph that tends to f_2 since $\varphi_2 > \varphi_1$, $\varphi_2 = 0.75$ while $\varphi_1 = 0.25$.

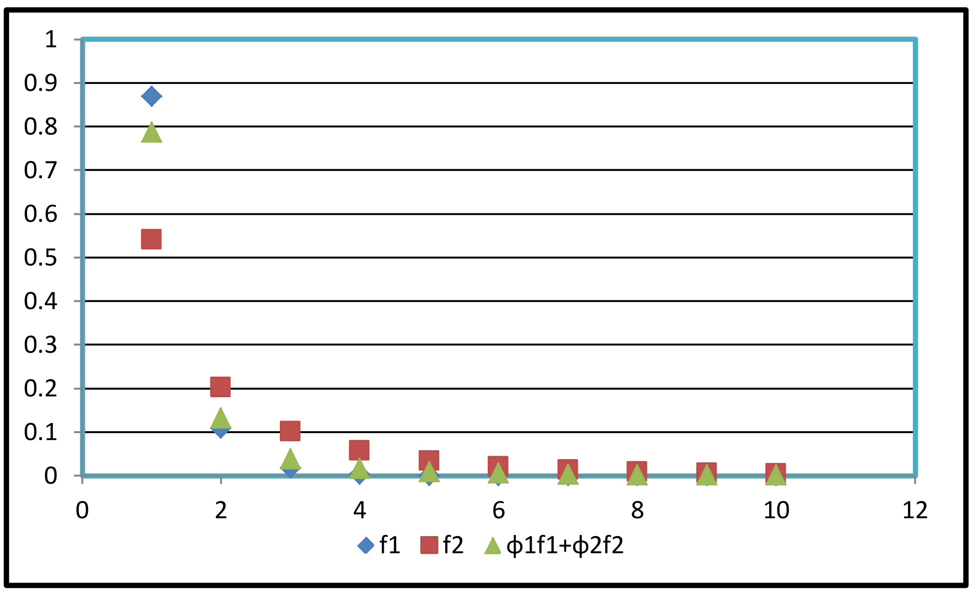

Figure 3. The Two Logarithmic Distributions and its Mixture ($\varphi_1 = 0.75, \varphi_2 = 0.25$)

Figure 3 represents the graph of probability distributions of f_1, f_2 and the mixture $\varphi_1 f_1 + \varphi_2 f_2$ where f_1 is the graph of the logarithmic distributions where $p_1 = 0.25$. f_2 is the graph of logarithmic distribution where $p_2 = 0.75$. $\varphi_1 f_1 + \varphi_2 f_2$ is the graph of the mixture of two logarithmic distributions. The mixture has a graph that tends to f_1 since $\varphi_1 > \varphi_2$, $\varphi_1 = 0.75$ while $\varphi_2 = 0.25$.

IV. CONCLUDING REMARKS

Based on the graphs, this paper shows that if $\varphi_1 = \varphi_2$, the graph of the mixture of two logarithmic distributions is the average of f_1 and f_2. If $\varphi_2 > \varphi_1$,the graph of the mixture of two logarithmic distributions tends to f_2. And if $\varphi_1 > \varphi_2$, the graph of the mixture of two logarithmic distributions tends to f_1. The graph of the mixture of two logarithmic distributions tends to the graph of the logarithmic distribution that has a greater value of weight φ.

We recommend the following for further studies: 1. On the mixture of more than two logarithmic distributions, and 2. On the estimation of parameters of the mixture of two logarithmic distributions.

REFERENCES

[1] Birkhäuser Boston, "*Generalized Logarithmic Series Distribution*", pp 223-239, 2006, 10.1007/0-8176-4477-6_11, 978-0-8176-4365-2.

[2] Farewell, V.T., Sprott D.A. (1988). "*The use of mixture model in the analysis of count data*". Biometrics, 11, 1191-1194.

[3] Feller, W. (1968). "*An Introduction to Probability Theory and Its Applications*", Vol. 1, 3rd Edition

[4] Fisher, R.A.; Corbet, A.S.; Williams, C.B. (1943). *"The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population"*. Journal of Animal Ecology : 42–58. doi:10.2307/1411. JSTOR

[5] Gupta, R.C. (1974). "*Power Series Distributions and some of its Applications*". Sankhya, Ser. B, 35, 288-298.

[6] *Johnson, Norman Lloyd; Kemp, Adrienne W; Kotz, Samuel (2005). "Chapter 7: Logarithmic and Lagrangian distributions". Univariate discrete distributions (3 ed.). John Wiley & Sons.*

[7] Khatri (1959). "*On certain properties of power series distributions*". Biometrika, 46, 486-490.

[8] Murat, M., Szynal, D. (2003). "*Moments of certain deformed probability distributions*". Commun. Statist. Theory Meth.32(2). 291-313.

[9] Noak, A. (1950). "*A class of random variable with discrete distribution*". Annals of the Institute of Statistical Mathematics. 21(1), 127-132

[10] Patil (1962). "*On certain properties of the generalized power series distribution*". Annals of the Institute of Statistical Mathematics. Vol 14(1), 179-182.

[11] Walck, Christian "*Hand-book on Statistical Distributions for experimentalists*" Particle Physics Group Fysikum University of Stockholm

[12] Weisstein, Eric W., "*Log-Series Distribution"*, MathWorld.

[13] Wilson, Jeffrey R, "*Logarithmic Series Distribution and its use in Analyzing Discerete Data*" Arizona State University Tempe, Arizona 85287

AUTHORS

First Author –Dr. Epimaco A. Cabanlit, Jr., Professor VI, Mathematics Department, Mindanao State University, General Santos City, Philippines. E-mail: maco_727@yahoo.com

Second Author – Jennifer T. Bulawan , BS Mathematics Graduate, Mathematics Department, Mindanao State University, General Santos City, Philippines.

Correspondence Author – Dr. Epimaco A. Cabanlit, Jr., Professor VI, Mathematics Department, Mindanao State University, General Santos City, Philippines. E-mail: maco_727@yahoo.com