Chapter 5

Certain Properties of Generalized Poisson-Lindley Distribution

5.1 Introduction

The aim of this chapter is to investigate certain properties including recurrence relations of probabilities, factorial moments and cumulants etc. of the generalized Poisson- Lindley (GPL) distribution which includes the Poisson –Lindley distribution of Sankaran (1970) as a special case mentioned in chapter 2, which is derived from the Poisson distribution (λ) by mixing with Lindley distribution due to Lindley (1958) having probability density function:

$$g(\lambda;\theta) = \frac{\theta^2}{1+\theta}(1+\lambda)e^{-\theta\lambda}\ , \quad \lambda > 0,\ \theta > 0 \qquad (5.1.1)$$

Ghitany et al. (2008) showed that in many ways model (5.1.1) is a better model for some application than one based on the exponential distribution.

Because of having only one parameter, the Poisson- Lindley distribution does not provide enough flexibility for analyzing different types of life data. To increase the flexibility for modeling purpose it will be useful to consider further alternative of this distribution. To increase the flexibility Mohmoudi and Zakerzadeh (2010) obtained an extended version of the Poisson- Lindley distribution known as Generalized Poisson-Lindley distribution and estimated its parameter using the moments and maximum likelihood method.

5.2 Probability mass function

The generalized Poisson-Lindley (GPL) distribution or two-parameter Poisson-Lindley distribution derived from the Poisson distribution with probability mass function

$$g(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \quad , \; x = 0, 1, 2, \dots , \;\; \lambda > 0 \tag{5.2.1}$$

when its parameter λ has a generalized Lindley distribution of Zakerzadeh and Dolati (2009) with probability density function

$$h(x; \alpha, \theta) = \frac{\theta^{1+\alpha}}{1+\theta} \frac{x^{\alpha-1}}{\Gamma(\alpha+1)} (\alpha + x) e^{-\theta x} , \;\; x > 0, \theta > 0, \alpha > 0 \tag{5.2.2}$$

Then the resultant compound Poisson distribution is the generalized Poisson-Lindley (GPL) distribution obtained by Mahmoudi and Zakerzadeh (2010) with pmf

$$P_x(\alpha, \theta) = \frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha+1)} \frac{\theta^{\alpha+1}}{(\theta+1)^{x+\alpha+1}} \left(\alpha + \frac{x+\alpha}{\theta+1}\right) , \quad x = 0, 1, 2, \; \dots \tag{5.2.3}$$

where $\theta > 0$ and $\alpha > 0$ be the two parameters of the GPL distribution and denote it by GPL(α, θ) .

As $\alpha \to 1$ in equation (5.2.3), it reduces to probability mass function of Poisson-Lindley distribution of Sankaran (1970). That is, Poisson-Lindley distribution is a particular case of the GPL distribution at $\alpha = 1$.

Proposition 5.1 If X is a r.v with GPL(α,θ) distribution. Then the corresponding probability generating function (pgf) is given by

$$G(t)=\left(\frac{\theta}{\theta-t+1}\right)^{1+\alpha}\left(\frac{\theta-t+2}{1+\theta}\right) \tag{5.2.4}$$

[cf.Mahmoudi and Zakerzadeh (2010)]

and the factorial moment generating function(fmgf) is written as

$$G(t+1)=\left(\frac{\theta}{\theta-t}\right)^{1+\alpha}\left(\frac{\theta-t+1}{1+\theta}\right) \tag{5.2.5}$$

Proposition 5.2 Let $X_1,X_2,\dots,X_n$ denote independent random variables from GPL (α_i,θ); for $i=1,2,\dots,n$. Then the generalized form of probability generating function (pgf) of $X=\sum_{i=1}^{n}X_i$ is given by

$$G_X(t)=\left(\frac{\theta}{\theta-t+1}\right)^{n+\alpha_1+\alpha_2+\cdots+\alpha_n}\left(\frac{\theta-t+2}{1+\theta}\right)^{n}$$

$$=\left(\frac{\theta}{\theta-t+1}\right)^{n(1+\alpha)}\left(\frac{\theta-t+2}{1+\theta}\right)^{n}, \text{when } (\alpha_1=\alpha_2=\cdots=\alpha_n=\alpha) \tag{5.2.6}$$

For $n=1$, the pgf (5.2.6) is same as that of the GPL distribution. For $n=1$ and $\alpha=1$, the equation (5.2.6) reduces to the pgf of one-parameter Poisson-Lindley distribution of Sankaran (1970).

5.3 Statistical Properties

This section is devoted to studying certain statistical properties of the GPL distribution, specifically recurrence relations for probabilities, factorial moments, cumulants, index of dispersion, cumulative distribution function and reliability function.

5.3.1 Recurrence relation for probabilities

The probability generating function (pgf.), $G(t)$ of the GPL distribution derived from the model (5.2.3) can be given as

$$G(t) = \left(\frac{\theta}{\theta-t+1}\right)^{1+\alpha}\left(\frac{\theta-t+2}{1+\theta}\right) \tag{5.3.1}$$

This result for probability generating function of the GPL distribution is obtained by Mahmoudi and Zakerzadeh (2010).

Now, differentiating (5.3.1) w.r.to 't' and equating the coefficients of t^r from both sides, we get the recurrence relation for probabilities of the GPL distribution as

$$P_{r+1} = \frac{1}{(1+\theta)(\theta+2)(r+1)}\left[\{1+\alpha(2+\theta)+(3+2\theta)r\}P_r - (\alpha+r-1)P_{r-1}\right],\ r = 1, 2, \dots, \tag{5.3.2}$$

where $P_0 = \left(\frac{\theta}{1+\theta}\right)^{1+\alpha}\left(\frac{2+\theta}{1+\theta}\right)$

$$P_1 = \frac{1+\alpha(2+\theta)}{(1+\theta)(\theta+2)}P_0$$

Putting $r = 1, 2, 3, \dots,$ in equation (5.3.2), the higher order probabilities may be computed easily. However, it is easy to write another recurrence relation for the probabilities as

$$P_r = \frac{r+\alpha-1}{r(1+\theta)}\ \frac{\alpha(\theta+2)+r}{\alpha(\theta+2)+r-1}P_{r-1},\ r = 1, 2, 3, \dots \tag{5.3.3}$$

where $P_0 = \left(\frac{\theta}{1+\theta}\right)^{1+\alpha}\left(\frac{2+\theta}{1+\theta}\right)$

5.3.2 Recurrence relation for factorial moments

The moments of the GPL distribution may be obtain in terms of its factorial moments as the computation of factorial moment is very simple. The recurrence relation for factorial moments of the GPL distribution obtained from fmgf (5.2.5) is given by

$$\mu'_{(r+1)} = \frac{1}{\theta(1+\theta)}\left[\{1+\theta r+(1+\theta)(\alpha+r)\}\mu'_{(r)} - \{r(1+\alpha)-1\}\mu'_{(r-1)}\right],\ r \geq 1 \quad (5.3.4)$$

where, $\mu'_{(1)} = \frac{(1+\alpha+\alpha\theta)}{\theta(1+\theta)}$

The higher order factorial moments obtained from expression (5.3.4) are given below as

$$\mu'_{(2)} = \frac{(1+\alpha)(2+\alpha+\alpha\theta)}{\theta^2(1+\theta)} \quad (5.3.5)$$

$$\mu'_{(3)} = \frac{(1+\alpha)(2+\alpha)(3+\alpha+\alpha\theta)}{\theta^3(1+\theta)} \quad (5.3.6)$$

$$\mu'_{(4)} = \frac{(1+\alpha)(2+\alpha)(3+\alpha)(4+\alpha+\alpha\theta)}{\theta^4(1+\theta)} \quad (5.3.7)$$

The general expression for the r^{th} order factorial moment of the GPL distribution may also be given as

$$\mu'_{(r)} = \frac{(\alpha+r-1)!(r+\alpha+\alpha\theta)}{\alpha!\theta^r(1+\theta)}. \quad (5.3.8)$$

After obtaining the first four factorial moments, by substituting $r = 1,2,3,4$ in the relation (5.3.8) and then using the relationship between factorial moments and moment about origin, the first four moments about origin of the GPL distribution are obtained very easily. It can be easily verified that at $\alpha = 1$, the moments of the GPL distribution reduces to the respective moments of the Poisson-Lindley distribution.

5.3.3 Recurrence Relation for Cumulants

The moment generating function of the GPL distribution corresponding to the pgf (5.3.1) is of the following form

$$M(t) = \left(\frac{\theta}{1+\theta-e^t}\right)^{1+\alpha}\left(\frac{2+\theta-e^t}{1+\theta}\right) \tag{5.3.9}$$

The cumulant generating function (cgf) of the GPL distribution is obtained by taking log on both sides of (5.3.9) as

$$K(t) = \log\left\{\left(\frac{\theta}{1+\theta-e^t}\right)^{1+\alpha}\left(\frac{2+\theta-e^t}{1+\theta}\right)\right\}. \tag{5.3.10}$$

From (5.3.9), we get the cumulant recurrence relation as

$$K_{r+1} = \frac{1}{\theta(1+\theta)}\left[\{1+\alpha(2+\theta)\} - \alpha 2^r + \sum_{j=1}^{r}(3+2\theta-2^j)\binom{r}{j}K_{r+1-j}\right], \quad r = 1, 2, \ldots, \tag{5.3.11}$$

This is obtained by differentiating (5.3.10) w. r. t. 't' and then equating the coefficients of $\frac{t^r}{r!}$ on both sides. First three cumulants of the distribution are respectively given by

$$K_1 = \frac{1+\alpha(1+\theta)}{\theta(1+\theta)} \quad \text{(Mean)} \tag{5.3.12}$$

$$K_2 = \frac{\alpha(1+\theta)^3+\theta^2+3\theta+1}{\theta^2(1+\theta)^2} \quad \text{(Variance)} \tag{5.3.13}$$

$$K_3 = \frac{\alpha(1+\theta)^4(\theta+2)+(\theta^3+6\theta^2+4\theta+1)(\theta+2)}{\theta^3(1+\theta)^3} \tag{5.3.14}$$

From the cumulants the moments can be easily obtained. Hence, the mean and variance of the GPL distribution are respectively

$$\mu = \frac{1+\alpha(1+\theta)}{\theta(1+\theta)} \tag{5.3.15}$$

and

$$\sigma^2 = \frac{\alpha(1+\theta)^3+\theta^2+3\theta+1}{\theta^2(1+\theta)^2} \tag{5.3.16}$$

5.3.4 Index of dispersion

The index of dispersion for the GPL distribution is obtained as

$$i_d = \frac{\sigma^2}{\mu} = \frac{\alpha(1+\theta)^3+\theta^2+3\theta+1}{\theta^2(1+\theta)^2}\frac{\theta(1+\theta)}{\alpha(1+\theta)+1}$$

$$= 1 + \frac{\alpha(1+\theta)^2+2\theta+1}{\alpha\theta(1+\theta)^2+\theta(1+\theta)};\ \theta > 0, \alpha > 0 \tag{5.3.17}$$

The above expression shows that, the two-parameter GPL distribution (α, θ) is over-dispersed $(\sigma^2 > \mu)$ for all values of α and θ, equi-dispersed $(\sigma^2 = \mu)$ for large amount of θ . [cf. Mahmoudi and Zakerzadeh (2010)]

5.3.5 Cumulative distribution function and Reliability function

The cumulative distribution function of the GPL distribution may be written as

$$F(x) = P(X \leq x) = \frac{\theta^{1+\alpha}}{\Gamma(1+\alpha)(1+\theta)^{1+\alpha}}\sum_{j=0}^{x}\frac{\Gamma(j+\alpha)}{j!(1+\theta)^j}\left(\alpha + \frac{j+\alpha}{1+\theta}\right);\ \ x = 0,1,2,\ldots \tag{5.3.18}$$

and the reliability function is given a

$$R(x) = 1 - F(x) = \frac{\Gamma(1+\alpha)(1+\theta)^{1+\alpha} - \theta^{1+\alpha}\sum_{j=0}^{x}\frac{\Gamma(j+\alpha)}{j!(1+\theta)^{j}}\left(\alpha+\frac{j+\alpha}{1+\theta}\right)}{\Gamma(1+\alpha)(1+\theta)^{1+\alpha}}\;;\; x = 0,1,2,.... \qquad (5.3.19)$$

where $\theta > 0$ and α is positive integer value.

5.4 Zero-modified version of GPL distribution

A zero-modified distribution alters the probability of occurrence of zeros. The major motivation force behind the development of zero-modified distribution is that, many distributions obtained in the course of experimental investigations often have an excess frequency of the observed event at zero point. This is already discussed in chapter 4.

A combination of the original distribution with probability mass function (pmf) $P_x\,; x = 0,1,2,...$ together with the degenerate distribution with all probabilities concentrated at the origin gives a zero-modified GPL distribution with pmf

$$P[X = 0] = w + (1 - w)P_0 \qquad (5.4.1)$$

$$P[X = x] = (1 - w)P_x \;;\; x \geq 1 \qquad (5.4.2)$$

where w is a parameter assuming arbitrary value in the interval $0 < w < 1$. It is also possible to take $w < 0$, provided $w + (1 - w)P_0 \geq 0$ [cf. Johnson et al. (2005)]. P_0 be the zero order probability of the GPL distribution which is discussed in section 5.3.1 and P_x be the probability mass function of the model.

The pgf of the zero-modified GPL distribution may be written as

$$G(t) = w + (1 - w)\left(\frac{\theta}{\theta - t + 1}\right)^{1+\alpha}\left(\frac{\theta - t + 2}{1+\theta}\right), \qquad (5.4.3)$$

where $0 < w < 1\,;\, \theta > 0,\, \alpha > 0$ be the parameters of the distribution.

Remark 5.1 Note that as $w \to 0, \alpha \to 1$ in (5.4.3), then it reduces to the probability generating function of the Poisson-Lindley distribution of Sankaran (1970) mentioned in chapter 2.

5.5 Parameter Estimation

In this section, we discuss about the estimation of the parameters of GPL distribution by using composite method i.e. θ can be estimated by using Newton-Raphson method whereas α is estimated by the method of moment.

In case of the GPL distribution, we have

$$\mu = \frac{1+\alpha(1+\theta)}{\theta(1+\theta)} \tag{5. 5.1}$$

$$\Rightarrow \alpha = \frac{\theta(1+\theta)\mu-1}{1+\theta} \tag{5.5.2}$$

and

$$\sigma^2 = \frac{\alpha(1+\theta)^3+\theta^2+3\theta+1}{\theta^2(1+\theta)^2} \tag{5.5.3}$$

$$\Rightarrow \alpha = \frac{\sigma^2\theta^2(1+\theta)^2-\theta^2-3\theta-1}{(1+\theta)^3} \tag{5.5.4}$$

After some suitable simplification of equations (5.5.2) and (5.5.4), a functional equation for θ in terms of μ and σ^2 may be obtained as

$$f(\theta) = A\theta^3 + B\theta^2 + C\theta + D \tag{5.5.5}$$

where

$$A = (\sigma^2 - \mu)$$
$$B = (2\sigma^2 - 3\mu)$$
$$C = (\sigma^2 - 3\mu)$$
$$D = -(\mu + 1)$$

Now replacing the population mean and variance by the respective sample mean and variance an estimate of A, B, C and D can be obtained and using it in (5.5.5), an estimate $\hat{\theta}$ of θ can be obtained by using the Newton-Raphson iteration method.

Again, substituting the value of $\hat{\theta}$ in (5.5.2) and replacing the population mean by the sample mean $\bar{x}$, an estimate of $\hat{\alpha}$ is obtained as

$$\hat{\alpha} = \frac{\theta(1+\theta)\bar{x}-1}{1+\theta} \tag{5.5.6}$$

5.6 Goodness of fit

For the fitting of the GPL distribution to illustrate its applications, two sets of reported data are considered.

In **Table 5.1**, we have considered the problem of mistakes in copying groups of random digits [data from Kemp and Kemp (1965)] and the problem of number of accidents to 647 women working on high explosive shells in 5 weeks [data from Greenwood and Yule (1920)] is considered in **Table 5.2** for which single parameter Poisson-Lindley distribution was fitted by Sankaran (1970). To estimate the single parameter of Poisson-Lindley distribution [Sankaran(1970)] method of moment is used, but in case of the GPL distribution parameters are estimated by using a composite method. i.e., the parameter θ is estimated by Newton-Raphson method, whereas the parameter α is estimated by the method of moments.

The **Table 5.1** and **Table 5.2** give the comparison of observed and expected frequencies of the fitted distribution and χ^2-statistic which has been used for test its goodness of fit. Observing the values of χ^2 and comparison of the observed frequencies with the expected frequencies of fitted GPL distribution in **Table 5.1** and **5.2**, it is clearly seen that the GPL distribution describe the data very well than the other distributions.

Table 5.1 Comparison of observed frequencies of mistakes in copying groups of random digits with expected frequencies of fitted GPL distribution. [data from Kemp and Kemp (1965)]

No. of errors per group	Observed frequencies	PLD Sankaran(1970) $\hat{\theta}$=1.743	Fitted GPLD $\hat{\theta}$=1.227 $\hat{\alpha}$=0.513
0	35	33.1	35.3
1	11	15.3	13.0
2	8	6.8	7.1
3	4	2.9	2.9
4	2	1.2	1.4
Total	60	59.3	59.7
χ^2		2.20	1.56
$d.f$		1	1
$p-value$		0.14	0.21

Note: PLD: Poisson-Lindley distribution.

GPLD: Generalized Poisson-Lindley distribution.

Table 5.2 Comparison of observed frequencies for accidents to 647 women working on high explosive shells with expected frequencies of fitted GPL distribution. [data from greenwood and Yule (1920)]

No. of accidents	Observed frequencies	PLD Sankaran(1970) $\hat{\theta}$=2.729	Fitted GPLD $\hat{\theta}$=2.433 $\hat{\alpha}$=0.841
0	447	439.5	443.3
1	132	142.8	137.7
2	42	45.0	44.7
3	21	13.9	14.5
4	3	4.2	4.7
≥5	2	1.3	1.5
Total	647	646.7	646.4
χ^2		4.82	1.79
$d.f$		2	2
$p-value$		0.089	0.409

Note: PLD: Poisson-Lindley distribution.

GPLD: Generalized Poisson-Lindley distribution.

****