Adv. & Appl. in Stat. **3**(1) (2003), 67-76

A GENERALIZED GEOMETRIC DISTRIBUTION AND ITS APPLICATION

M. E. GHITANY and D. K. AL-MUTAIRI

Department of Statistics and Operations Research, Faculty of Science
Kuwait University, P. O. Box 5969, Safat 13060, Kuwait

Abstract

In this paper we derive a two-parameter discrete distribution, called generalized geometric distribution, which includes the Poisson-Lomax, geometric, and Poisson-Lindley distributions as special cases. Recurrence relations for calculating the probabilities and factorial moments of the distribution are given. For all values of the parameters, it is shown that the distribution is unimodal. Also, it is shown that the distribution has a failure rate (mean residual life) which may be decreasing, constant, or increasing (increasing, constant, or decreasing). Application of the proposed distribution to real data is given and its goodness-of-fit is demonstrated.

1. Introduction

A certain mixed Poisson distribution arises when the Poisson parameter vary according to some probability distribution, called the *mixing distribution*. The mixing distribution may be discrete or continuous. A well known example of mixed Poisson distribution is the negative binomial distribution, where the mixing distribution is the gamma distribution.

Let X given $\Lambda = \lambda$, denoted by $X \mid \lambda$, has a Poisson distribution with

---

2000 Mathematics Subject Classification: 62E15, 62F10.

Key words and phrases: mixed Poisson distribution, factorial moments, unimodality, log-convexity, log-concavity, failure rate, mean residual life.

Received March 10, 2003

probability mass function (p.m.f.)

$$p(x \mid \lambda) = P(X = x \mid \lambda) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, \ldots, \quad \lambda > 0. \tag{1}$$

Now suppose Λ is a continuous random variable with probability density function (p.d.f.),

$$f(\lambda) = \frac{1}{U(1, \alpha + 2, \beta)} (1 + \lambda)^{\alpha} e^{-\beta\lambda}, \quad \lambda > 0, \tag{2}$$

where $-\infty < \alpha < \infty$, $\beta > 0$, $U(\cdot, \cdot, \cdot)$ is the confluent hypergeometric function of the second kind with integral representation:

$$U(a, b, z) = \frac{1}{\Gamma(a)} \int_0^{\infty} e^{-zt} t^{a-1} (1 + t)^{b-a-1} dt, \quad a, z > 0$$

with $\Gamma(\cdot)$ as the gamma function. More details about the hypergeometric function of the second kind can be found in Abramowitz and Stegun [1].

Special cases:

(i) $\alpha = -(\nu + 1)$, $\nu > 0$, $\beta \to 0$: $f(\lambda) = \dfrac{\nu}{(1 + \lambda)^{\nu+1}}$, the Lomax distribution.

(ii) $\alpha = 0$: $f(\lambda) = \beta e^{-\beta\lambda}$, the exponential distribution.

(iii) $\alpha = 1$: $f(\lambda) = \dfrac{\beta^2}{\beta + 1} (1 + \lambda) e^{-\beta\lambda}$, the Lindley [5] distribution.

The unconditional p.m.f. of X, i.e., $p(x) = P(X = x)$, is given by

$$\begin{aligned} p(x) &= \int_0^{\infty} p(x \mid \lambda) f(\lambda) d\lambda \\ &= \frac{1}{U(1, \alpha + 2, \beta) x!} \int_0^{\infty} e^{-(\beta+1)\lambda} \lambda^x (1 + \lambda)^{\alpha} d\lambda \\ &= \frac{U(x + 1, x + \alpha + 2, \beta + 1)}{U(1, \alpha + 2, \beta)}, \quad x = 0, 1, \ldots, \end{aligned} \tag{3}$$

where $-\infty < \alpha < \infty$, $\beta > 0$.

Special cases:

(i) $\alpha = -(\nu + 1), \ \nu > 0, \ \beta \to 0:$ since $U(1, 1 - \nu, 0) = \nu^{-1},$ (3) reduces to

$$p(x) = \nu U(x + 1, x - \nu + 1, 1), \quad x = 0, 1, \ldots, \quad \nu > 0, \tag{4}$$

the Poisson-Lomax distribution.

(ii) $\alpha = 0:$ since $U(x + 1, x + 2, z) = z^{-(x+1)},$ (3) reduces to

$$p(x) = \frac{\beta}{\beta + 1}\left(\frac{1}{\beta + 1}\right)^{x}, \quad x = 0, 1, \ldots, \quad \beta > 0, \tag{5}$$

the geometric distribution.

(iii) $\alpha = 1:$ since $U(x + 1, x + 3, z) = z^{-(x+2)}(z + 1 + x),$ (3) reduces to

$$p(x) = \frac{\beta^2(x + \beta + 2)}{(\beta + 1)^{x+3}}, \quad x = 0, 1, \ldots, \quad \beta > 0, \tag{6}$$

the Poisson-Lindley distribution (Sankaran [6]).

In the following we use the notation $GGD(\alpha, \beta)$ to denote the generalized geometric distribution with p.m.f. (3).

For the proposed generalized geometric distribution, we provide recurrence relations for probabilities and factorial moments, Section 2. Also, we show that the proposed distribution is unimodal, Section 3, and its failure rate (mean residual life) may be decreasing, constant, or increasing (increasing, constant, or decreasing), Section 4. Finally, fitting the proposed distribution to real data on the number of daily thunderstorms is given and its goodness-of-fit is demonstrated, Section 5.

2. Recurrence Relations

The following theorem provides a useful three-term recurrence relation for calculating the probabilities $p(x)$ of the $GGD(\alpha, \beta)$.

Theorem 1. *For all* $-\infty < \alpha < \infty, \ \beta > 0,$ *the* $GGD(\alpha, \beta)$ *satisfies the recurrence relation*:

$$p(x+1) = \frac{1}{(\beta+1)(x+1)}[(x+\alpha-\beta)p(x) + p(x-1)], \quad x = 1, 2, \ldots$$

where

$$p(0) = \frac{U(1, \alpha+2, \beta+1)}{U(1, \alpha+2, \beta)}, \quad p(1) = \frac{U(2, \alpha+3, \beta+1)}{U(1, \alpha+2, \beta)}.$$

Proof. Using formula 13.4.27 of Abramowitz and Stegun [1, p.507]:

$$zaU(a+1, b+1, z) + (1-b+z)U(a, b, z) - U(a-1, b-1, z) = 0 \qquad (7)$$

with $a = x+1$, $b = x+2+\alpha$ and $z = \beta+1$, we obtain

$$(\beta+1)(x+1)U(x+2, x+3+\alpha, \beta+1)$$

$$+ (\beta-\alpha-x)U(x+1, x+2+\alpha, \beta+1) - U(x, x+1+\alpha, \beta+1) = 0. \qquad (8)$$

Using (3) in (8), we obtain, for all $x = 1, 2, \ldots,$

$$(\beta+1)(x+1)p(x+1) + (\beta-\alpha-x)p(x) - p(x-1) = 0$$

proving the theorem.

To show how Theorem 1 works, consider the case: $\alpha = -3$, $\beta = 1$. Since, in this case, $p(0) = 0.746737$ and $p(1) = 0.183402$, we have

$$p(2) = \frac{1}{4}[-3(0.183402) + 0.746737] = 0.049133,$$

$$p(3) = \frac{1}{6}[-2(0.049133) + 0.183402] = 0.014189, \quad \text{etc.}$$

Next, we provide a recurrence relation for the rth factorial moment of the $GGD(\alpha, \beta)$, i.e., $\mu'_{[r]} = E\left[\prod_{i=1}^{r}(X-i+1)\right]$, $r = 1, 2, \ldots$.

Theorem 2. *For all* $-\infty < \alpha < \infty$, $\beta > 0$, *the* $GGD(\alpha, \beta)$ *satisfies the recurrence relation*:

$$\mu'_{[r+1]} = \frac{1}{\beta}\{(r+\alpha-\beta+1)\mu'_{[r]} + r\mu'_{[r-1]}\}, \quad r = 1, 2, \ldots$$

where

$$\mu'_{[0]} = 1, \qquad \mu'_{[1]} = \frac{U(2,\ \alpha + 3,\ \beta)}{U(1,\ \alpha + 2,\ \beta)}.$$

Proof. By conditioning on the mixing random variable Λ, we obtain

$$\mu'_{[r]} = E\left\{E\left[\prod_{i=1}^{r}(X - i + 1) \mid \Lambda\right]\right\}$$

$$= E\{\Lambda^r\}$$

$$= \frac{1}{U(1,\ \alpha + 2;\ \beta)}\int_0^\infty e^{-\beta\lambda}\lambda^r(1 + \lambda)^\alpha d\lambda$$

$$= r!\frac{U(r + 1,\ r + \alpha + 2,\ \beta)}{U(1,\ \alpha + 2,\ \beta)}, \quad r = 1, \ldots. \tag{9}$$

Using identity (7) with $a = r + 1$, $b = r + \alpha + 2$ and $z = \beta$, we obtain

$$\beta(r + 1)U(x + 2,\ x + 3 + \alpha,\ \beta)$$

$$+ (\beta - \alpha - r)U(r + 1,\ r + 2 + \alpha,\ \beta) - U(r,\ r + 1 + \alpha,\ \beta) = 0. \tag{10}$$

Finally, using (9) in (10), we obtain, for all $r = 1, 2, \ldots$,

$$\beta(r + 1)\frac{1}{(r + 1)!}\mu'_{[r+1]} + (\beta - \alpha - r - 1)\frac{1}{r!}\mu'_{[r]} - \frac{1}{(r - 1)!}\mu'_{[r-1]} = 0$$

proving the theorem.

To show how Theorem 2 works, consider the case: $\alpha = -3$, $\beta = 1$. Since, in this case, $\mu'_{[0]} = 1$ and $\mu'_{[1]} = 0.353750$, we have

$$\mu'_{[2]} = -2(0.353750) + 1 = 0.292500$$

$$\mu'_{[3]} = -(0.29250) + 2(0.353750) = 0.415000, \quad \text{etc.}$$

Remark. The mean and variance of the $GGD(\alpha, \beta)$, respectively, are given by

$$\mu = \mu'_{[1]} = \frac{U(2,\ 3 + \alpha,\ \beta)}{U(1,\ 2 + \alpha,\ \beta)},$$

$$\sigma^2 = \mu'_{[2]} + \mu'_{[1]} - (\mu'_{[1]})^2 = \frac{1}{\beta} + \frac{\alpha + 2}{\beta}\mu - \mu^2.$$

3. Unimodality

A discrete (continuous) distribution with p.m.f. (p.d.f.) $\eta(x)$ is said to be *unimodal* if there is a value x_0 such that $\eta(x)$ is steadily increasing for $x \le x_0$ and steadily decreasing for $x \ge x_0$. The value x_0 is called a *mode* and may not be unique. The class of discrete (continuous) unimodal distributions includes those distributions for which $\eta(x)$ is *monotone*.

Theorem 3. *For all* $-\infty < \alpha < \infty$, $\beta > 0$, *the* $GGD(\alpha, \beta)$ *is unimodal.*

Proof. Holgate [4] showed that any mixed Poisson distribution is unimodal if the mixing distribution is positive continuous and unimodal.

Note that, for all $-\infty < \alpha < \infty$, $\beta > 0$, $f(\lambda)$ is positive continuous function with

$$f(0) = \frac{1}{U(1, \alpha + 2, \beta)}, \quad f(\infty) = 0.$$

The derivative of $f(\lambda)$ with respect to λ can be written as

$$f'(\lambda) = -\left(\beta - \frac{\alpha}{1 + \lambda}\right) f(\lambda).$$

Hence, if $-\infty < \alpha \le \beta$, then $f(\lambda)$ is decreasing. Also, if $\alpha > \beta > 0$, then $f(\lambda)$ is unimodal at the point $\lambda_0 = (\alpha - \beta)/\beta$. This completes the proof.

Corollary 1. *The Poisson-Lomax, geometric and Poisson-Lindley distributions are unimodal.*

4. Failure Rate and Mean Residual Life

In reliability studies, the failure rate, $r(x)$, and mean residual life, $m(x)$, functions play important roles in distinguishing between failure distributions. For a discrete distribution with p.m.f. $p(x)$, $x = 0, 1, \ldots$, these functions, respectively, are defined as:

$$r(x) = \frac{p(x)}{R(x)},$$

$$m(x) = E(X - x \mid X \geq x) = \frac{1}{R(x)} \sum_{i \geq x} (i - x) p(i),$$

where $R(x) = P(X \geq x) = \sum_{i \geq x} p(i)$ is called the *reliability function* of X.

In general, the failure rate and mean residual life functions cannot be expressed in simple closed form and hence direct investigation of their monotonicity properties is extremely difficult.

In the following, we use the notation DFR, CFR, IFR (DMRL, CMRL, IMRL) to denote a decreasing, constant, increasing failure rate (a decreasing, constant, increasing mean residual life).

Theorem 4. *For all* $\beta > 0$, *the* $GGD(\alpha, \beta)$ *is DFR*, *CFR*, *IFR* (*IMRL*, *CMRL*, *DMRL*) *if* $\alpha < 0$, $\alpha = 0$, $\alpha > 0$, *respectively*.

Proof. When $\alpha = 0$, it is clear that the geometric distribution is both CFR and CMRL. Hence, for $\alpha < 0$ $(\alpha > 0)$, it suffices to show that the mixing p.d.f (2) is log-convex (log-concave), see, e.g., Steutel [7]. This is, indeed, the case, since

$$\frac{d^2}{d\lambda^2} \log f(\lambda) = \frac{\alpha}{(1 + \lambda)^2}$$

is < 0 (> 0) if $\alpha < 0$ $(\alpha > 0)$. The statement regarding the MRL is a direct implication of monotone FR.

Corollary 2. *The Poisson-Lomax*, *geometric and Poisson-Lindley distributions are*, *respectively*, *DFR*, *CFR*, *IFR* (*IMRL*, *CMRL*, *DMRL*).

5. Applications

We consider two published data sets on:

(i) the number of accidents (x_i) to women working on H.E. shells during five weeks, Table 1, see Greenwood and Yule [3],

(ii) the number of thunderstorm events (x_i) per day at Cape Kennedy, Florida during the month of June, Table 2, see Falls et al. [2].

We fit the proposed generalized geometric distribution to the above data sets using the method of maximum likelihood to estimate the parameters α and β.

Tables 1 and 2 give the observed frequencies O_i as well as the expected frequencies E_i for the considered data sets. For both data sets, the expected frequency of each of the last two cells are less than 5. Hence, these two cells are combined with the last cell before them. Based on the reported P-values in both tables, we conclude that the $GGD(\alpha, \beta)$ provides good fit for both data sets.

Table 1. Distribution of the number of accidents to women working on H.E. Shells for five weeks

x_i	O_i	E_i
0	447	443.663
1	132	137.883
2	42	43.992
3	21	14.322
4	3	4.737
≥ 5	2	2.403
Total	647	647
$\hat{\alpha}$		-1.642
$\hat{\beta}$		1.597
χ^2		4.122
d.f.		2
P-value		0.127

Table 2. Distribution of the number of thunderstorm events per day during the month of June at Cape Kennedy, Florida

x_i	O_i	E_i
0	187	183.776
1	77	84.905
2	40	36.576
3	17	15.040
4	6	5.981
5	2	2.319
6	1	1.403
Total	330	330
$\hat{\alpha}$		1.625
$\hat{\beta}$		2.110
χ^2		1.419
d.f.		3
P-value		0.701

References

[1] M. Abramowitz and I. Stegun, Handbook of Mathematical Functions, Dover, New York, 1965.

[2] L. W. Falls, W. O. Williford and M. C. Carter, Probability distributions for thunderstorm activity at Cape Kennedy, Florida, J. Appl. Meteorol. 10 (1971), 97-104.

[3] V. Greenwood and G. U. Yule, An inquiry into the nature of frequency distribution representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or repeated accidents, J. Roy. Statist. Soc. Series A 83 (1920), 255-279.

[4] P. Holgate, The modality of some compound Poisson distributions, Biometrika 57 (1970), 666-667.

[5] D. V. Lindley, Fiducial distributions and Bayes' theorem, J. Roy. Statist. Soc. Series A 20 (1958), 102-107.

[6] M. Sankaran, The discrete Poisson-Lindley distribution, Biometrics 26 (1970), 145-149.

[7] F. W. Steutel, Log-concave and log-convex distributions, in: Encyclopedia of Statistical Sciences, Kotz, S., Johnson, N. L. and Read, C. H. (Editors-in-Chief), Wiley, New York 5 (1985), 116-117.

■