SORT 40 (1) January-June 2016, 153-176

# Transmuted geometric distribution with applications in modelling and regression analysis of count data

Subrata Chakraborty[1,*] and Deepesh Bhati[2]

**Abstract**

A two-parameter transmuted geometric distribution is proposed as a new generalization of the geometric distribution by employing the quadratic transmutation techniques of Shaw and Buckley. The additional parameter plays the role of controlling the tail length. Distributional properties of the proposed distribution are investigated. Maximum likelihood estimation method is discussed along with some data fitting experiments to show its advantages over some existing distributions in literature. The tail flexibility of density of aggregate loss random variable assuming the proposed distribution as primary distribution is outlined and presented along with a illustrative modelling of aggregate claim of a vehicle insurance data. Finally, we present a count regression model based on the proposed distribution and carry out its comparison with some established models.

*MSC:* 62E15

*Keywords:* Aggregate claim, count regression, geometric distribution, transmuted distribution.

## 1. Introduction

A random variable (rv) $X$ follows the geometric distribution with parameter $q$, denoted by $\mathscr{GD}(q)$ (see Johnson et al., 2005), pp. 210, equation (5.8)) if its probability mass function(pmf) is given by

$$P(X=t) = pq^t, \qquad t = 0,1,2,\cdots,0<q<1, p=1-q \tag{1}$$

[*] Corresponding Author: subrata_arya@yahoo.co.in

[1] Department of Statistics, Dibrugarh University, Dibrugarh-786004, Assam, India.

[2] Department of Statistics, Central University of Rajasthan, Ajmer-305817, Rajasthan, India, deepesh.bhati@curaj.ac.in
Received: August 2015
Accepted: April 2016

For the geometric distribution in (1) the cumulative distribution function (cdf) and survival function (sf) are respectively given by

$$F_X(t) = 1 - q^{t+1} \quad \text{and} \quad S_X(t) = P(X \geq t) = q^t.$$

In last few decades, many generalizations of geometric distribution were attempted by researchers by using different methods, for example, see Jain and Consul (1971), Philippou et al. (1983), Tripathi et al. (1987), Makčutek (2008), Gómez (2010), Chakraborty and Gupta (2015), Sastry et al. (2014) and references therein.

The transmutation, in particular the quadratic rank transmutation(QRT) method first introduced by Shaw and Buckley in 2007 has been used by many researchers to generate a large number of new distributions staring with suitable continuous baseline distributions (see Owoloko et al., 2015, Oguntunde and Adejumo, 2015 and Yousof et al., 2015 for details). It is an interesting way of generating a new and more flexible distribution by adding an additional parameter $(\alpha)$ to a baseline distribution. The QRT method produces a new family distribution that can be seen as a mixture of the maximum and minimum order statistics for a sample of size two from the baseline distribution and also as a mixture of the baseline distribution and its exponentiated version with power parameter two. The new family allows a continuum of distributions in the range of the additional parameter $(-1 < \alpha < 1)$. This method is applicable to any type of baseline distribution like symmetric, centred, and defined over $\mathbb{Z}$; provides explicit expression of the cdf, moments for new distribution through those of baseline distribution; and is suitable for simulation through the quantile function of the baseline distribution. Because of the many properties possessed by the method a significant amount of work to develop new flexible continuous distributions by transmutation method has been published in the last few years. The motivation of the present article is to derive a more flexible extension of the geometric distribution by application of the QRT method. The choice of QRT method is not just for its many attractive properties but also due to the fact that so far there is no evidence of any attempt to use transmutation method to generate new discrete distribution.

Accordingly, in this article an attempt is made to derive a new generalization of geometric distribution with two parameters $0 < q < 1$ and $-1 < \alpha < 1$ by using the QRT method of Shaw and Buckley (2007), which is presented in Section 2. Some distributional properties like unimodality, generating function, moments, quantile function are discussed in Section 3. A discussion on the maximum likelihood estimation (MLE) of parameters is presented in Section 4. Finally, in Section 5, applications of the proposed distribution in modelling aggregate claim size data, claim frequency data and in count data regression are presented.

## 2. A new generalization of geometric distribution

Here we first briefly discuss the QRT method and then propose the new transmuted geometric distributon.

### 2.1. Quadratic rank transmutation

The general rank transmutation mapping proposed by Shaw and Buckley (2007) for given pair of cdfs $F_1$ and $F_2$ having same support is defined as $G_{R12}(u) = F_2\left(F_1^{-1}(u)\right)$ and $G_{R21}(u) = F_1\left(F_2^{-1}(u)\right)$ where $F^{-1}(u)$ is the quantile function corresponding to the cdf $F(u)$. Both $G_{R12}(u)$ and $G_{R21}(u)$ map the unit interval in to itself. In particular, the quadratic rank transmutation (QRT) mapping is defined by $G_{R12}(u) = u + \alpha u(1-u)$. This implies

$$F_2\left(F_1^{-1}(u)\right) = u + \alpha u(1-u) = (1+\alpha)u - \alpha u^2 \Rightarrow F_2(x) = (1+\alpha)F_1(x) - \alpha F_1(x)^2$$

A discrete rv $Y$ with cdf $F_Y(.)$ and pmf $P(Y=y)$ is said to be constructed by the QRT method of Shaw and Buckley (2007) by transmuting another discrete rv $X$ with cdf $F_X(.)$ and pmf $P(X=x)$, if

$$\begin{aligned} F_Y(y) &= (1+\alpha)F_X(y) - \alpha F_X(y)^2 \text{and} \\ P(Y=y) &= (1+\alpha-2\alpha F_X(y))P(X=y) + \alpha\,(P(X=y))^2 \end{aligned} \tag{2}$$

The distribution $F_Y$ is then refereed to as the transmuted-$F_X$. In particular, for $\alpha = 0$ it gives the parent distribution function $F_X(y)$, for $\alpha = -1$, $F_X(y)^2$ the distribution of the maximum of two iid rvs with cdf $F_X(x)$, and for $\alpha = 1$, $2F_X(y) - F_X(y)^2$ the distribution of the minimum of two iid rvs with cdf $F_X(x)$.

Mirhossaini and Dolati (2008), expressing the cdf in (2) as $F_Y(y) = F_X(y)(1+ +\alpha\bar{F}_X(y))$ where $\bar{F}_X(y) = 1 - F_X(y)$, viewed it as a univariate counterpart of the Farlie-Gumbel-Morgenstern family (see Drouet-Mari and Kotz (2001)) of bivariate cdf $H_{XY}(x,y)$ generated from two independent univariate cdfs $F_X(x)$ and $F_Y(y)$ by the formula $H_{XY}(x,y) = F_X(x)F_Y(y)\left(1+\alpha\bar{F}_X(x)\bar{F}_Y(y)\right), -1 < \alpha < 1$.

Kozubowski and Podgórski (2016) in a very recent paper have shown that the transmuted-$F_X$ distribution can be seen as the distribution of maxima(or minima) of a random number $N$ of iid rvs with the base distribution $F_X(x)$, where $N$ has a Bernoulli distribution shifted up by one.

More over by rewriting the cdf in (2) as

$$F_Y(y) = \frac{1+\alpha}{2}\left(2F_X(y) - F_X(y)^2\right) + \frac{1-\alpha}{2}\left(F_X(y)\right)^2$$

it can be seen as a convex combination (finite mixture) of the cdfs of the maximum and minimum of two iid rv following $F_X(.)$. This implies $(F_X(y))^2 \leq F_X(y) \leq 2F_X(y) - (F_X(y))^2$ since $(F_X(y))^2 \leq 2F_X(y) - (F_X(y))^2$. Therefore the transmuted-$F_X$ family provides a continuum of distributions over the range of the additional parameter $\alpha \in (-1,1)$.

## *2.2. Transmuted geometric distribution*

Suppose an rv $X$ has $\mathscr{GD}(q)$ in (1). Then the cdf of the transmuted geometric rv $Y$ will be constructed as

$$\begin{aligned} F_Y(y) &= (1+\alpha)\left(1-q^{y+1}\right) - \alpha\left(1-q^{y+1}\right)^2 \\ &= 1-(1-\alpha)q^{y+1} - \alpha q^{2(y+1)}, \quad y = 0,1,2,\cdots; 0 < q < 1, -1 < \alpha < 1. \end{aligned}$$

and the corresponding pmf will then be given by

$$p_y = P(Y=y) = (1-\alpha)q^y(1-q) + \alpha(1-q^2)q^{2y}, \quad y = 0,1,2,\cdots. \tag{3}$$

where $0 < q < 1, -1 < \alpha < 1$. The distribution in (3) will henceforth be referred to as the *transmuted geometric distribution* ($\mathscr{TGD}$) with two parameters $q$ and $\alpha$. In short, $\mathscr{TGD}(q,\alpha)$.

**Particular cases:**

1. For $\alpha = 0$, (3) reduces to $\mathscr{GD}(q)$ in (1).
2. For $\alpha = -1$, (3) reduces to a special case of the exponentiated geometric distribution of Chakraborty and Gupta (2015) with power parameter equal to 2. This is the distribution of the maximum of two iid $\mathscr{GD}(q)$ rvs.
3. For $\alpha = 1$, (3) reduces to $\mathscr{GD}(q^2)$ with pmf $(1-q^2)q^{2y}$, which is the distribution of the minimum of two iid $\mathscr{GD}(q)$ rvs.

**Remark 1** $\mathscr{TGD}(q,\alpha)$ forms a continuous bridge between the distributions of the minimum to maximum in a sample of size two from $\mathscr{GD}(q)$.

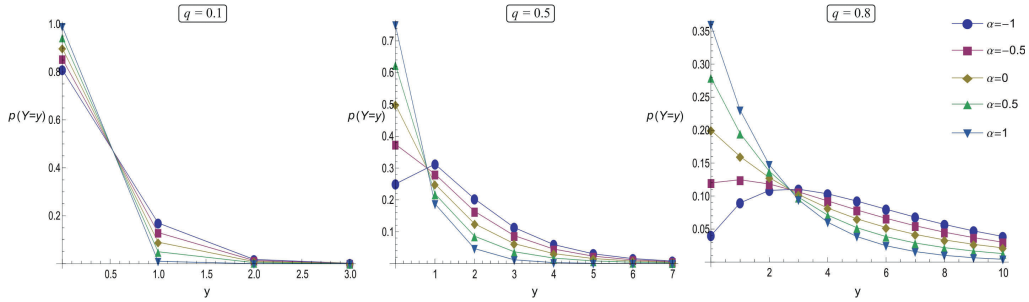


***Figure 1:*** *PMF plot of $\mathscr{TGD}(q,\alpha)$ for different value of parameter.*

## 3. Distributional properties

### 3.1. Shape of the $\mathscr{TGD}(q,\alpha)$

The graphs of the pmf of $\mathscr{TGD}(q,\alpha)$ are plotted for various combinations of the values of the two parameters $q$ and $\alpha$ in Figure 1. When $-1<\alpha<0$, the pmf is unimodal with either zero or non-zero mode, while for $0\leq\alpha<1$, the pmf is always a decreasing function with unique mode at $Y=0$. The above assertions are mathematically established later in Section 3.3. Moreover, the spread of $\mathscr{TGD}(q,\alpha)$ increases with $q$ and decreases with $\alpha$.

Furthermore, $\mathscr{TGD}(q,\alpha)$ has at most a tail as long as $\mathscr{GD}(q)$ can be seen from the pmf plots in the Figure 1 and also from the monotonicity of the ratio of the successive probabilities(see theorem 1). The shortest tail occurs when $\alpha=1$.

### 3.2. Monotonicity

Here we briefly discuss some useful monotonic properties of $\mathscr{TGD}(q,\alpha)$ and its direct consequences.

**Theorem 1** *For $0<\alpha<1$ the $\mathscr{TGD}(q,\alpha)$ distribution with pmf given in (3), the ratio $p_y/p_{y-1}$, $y=1,2,\cdots$, forms a monotone increasing sequence.*

*Proof.* Firstly, we have $p_0\neq 0, p_1\neq 0$ and $0<\alpha<1$. Now

$$\mathscr{Q}(y)=p_y/p_{y-1}=\frac{(1-\alpha)(1-q)q^y+\alpha(1-q^2)q^{2y}}{(1-\alpha)(1-q)q^{y-1}+\alpha(1-q^2)q^{2(y-1)}}$$

$$=q\left(1+\frac{\alpha(1+q)q^y}{(1-\alpha)}\Big/1+\frac{\alpha(1+q)q^{y-1}}{(1-\alpha)}\right)$$

further,

$$\triangle\mathscr{Q}(y)=\mathscr{Q}(y+1)-\mathscr{Q}(y)=\frac{(1-q)^2q^{y+1}(1+q)(1-\alpha)\alpha}{(q^2(1-\alpha)+\alpha(1+q)q^y)\,(q(1-\alpha)+\alpha(1+q)q^y)}$$

Since, for $0<\alpha<1$, $\mathscr{Q}(y)>0$, therefore $p_y/p_{y-1}$ forms a monotone increasing sequence for $0<\alpha<1$. ■

The following results follow as a consequence of Theorem 1. For $0<\alpha<1$, $\mathscr{TGD}(q,\alpha)$

i. is infinitely divisible (see Warde and Katti, 1971).

ii. pmf is a decreasing sequence (see Johnson and Kotz, 2005 p.75), which in turn indicates that, $\mathscr{TGD}$ has a zero vertex (see Warde and Katti, 1971). This fact was also mentioned in Remark 3.

iii. is DFR(decreasing failure rate), which in turn implies IMRL(increasing mean residual life).

iv. an upper bound for the variance of the $\mathscr{TGD}(q,\alpha)$ can be obtained for $0<\alpha<1$ as

$$Var(Y)\geq\frac{p_1}{p_0}=\frac{q(1-\alpha)+\alpha q^2(1+q)}{1-\alpha+\alpha(1+q)}$$

**Corollary 1** *For $-1<\alpha<0$, $\mathscr{TGD}(q,\alpha)$ distribution with pmf given in (3) is log-concave.*

*Proof.* The result follows from that fact that $p_y/p_{y-1}$, $y=1,2,\dots$, forms a monotone decreasing sequence for $-1<\alpha<0$ that is $p_{y+1}/p_y<p_y/p_{y-1}\Rightarrow p_y^2>p_{y-1}p_{y+1}\ \forall y$. ■

The following results follow as a consequence of corollary 1: For $-1<\alpha<0$, $\mathscr{TGD}(q,\alpha)$ distribution is

i. IFR (increasing failure rate), which in turn implies DMRL (decreasing mean residual life).

ii. Strongly unimodal.

iii. At most has a geometric tail.

### 3.3. Mode

**Theorem 2** *$\mathscr{TGD}(q,\alpha)$ is unimodal with a nonzero mode for $-1<\alpha<-(q(2+q))^{-1}$ provided that $q>0.414$.*

*Proof.* A pmf $P(Y=y)$ with support $y=0,1,2,\cdots$, is uni modal if there exists a unique point $M(\neq 0)$, in the support of $Y$ such that $P(Y=y)$ is increasing on $(0,1,\ldots,M)$ and decreasing on $(M,M+1,\ldots)$. $M$ is then the unique mode of $P(Y=y)$. Thus $\mathscr{TGD}(q,\alpha)$ will have a non zero mode if,

$$\begin{aligned}
&P(Y=1)>P(Y=0)\\
&\Rightarrow (1-\alpha)(1-q)q+\alpha q^2(1-q^2)>(1-\alpha)(1-q)+\alpha(1-q^2)\\
&\Rightarrow (1-\alpha)(1-q)^2+\alpha(1-q^2)(1-q^2)<0\\
&\Rightarrow \alpha<-(1-q)^2/\left((1-q^2)^2-(1-q)^2\right)=-1/(q(2+q)
\end{aligned}$$

But the condition $-1<\alpha<-(q(2+q)^{-1}$ makes sense only if $q(2+q)>1$ which implies $q>\sqrt{2}-1\cong 0.414$. ∎

For example, with $q=0.8$ non zero modes occur when $-1<\alpha<-0.4464$ as can be clearly seen in the third plot of the pmfs in the Figure 1.

**Remark 2** For $q<0.414$, the condition of non-zero unimodality leads to $\alpha$ outside its permissible range of $-1<\alpha$

**Remark 3** For $0\leq\alpha\leq 1$, the pmf is decreasing, and the mode occurs at the point 0. This indicates the suitability of the proposed distribution for count data which feature, relatively, a large number of zeros. Moreover the proportion of zeros in $\mathscr{TGD}(q,\alpha)$ is more(less) than that of $\mathscr{GD}(q)$ depending on $\alpha>(<)0$.

### 3.4. An alternative derivation of the $\mathscr{TGD}(q,\alpha)$

**Theorem 3** *$\mathscr{TGD}(q,\alpha)$ is the discrete analogue of the skew exponential distribution of Shaw and Buckley (2007).*

*Proof.* The pdf and cdf of the skew exponential distribution derived using the quadratic rank transmutation (Shaw and Buckley, 2007) are respectively given by

$$f_X(x)=(1-\alpha)\beta e^{-\beta x}+2\alpha\beta e^{-2\beta x},\quad x>0,\beta>0,-1<\alpha<1$$

and

$$F_X(x)=(1+\alpha)(1-e^{-\beta x})-\alpha(1-e^{-2\beta x})^2,\quad x>0,\beta>0,-1<\alpha<1.$$

Hence, the pmf of the discrete analogue (see Chakraborty, 2015, for a detail review of various methods of construction of discrete analogues of continuous distributions.) of $X$, $Y = \lfloor X \rfloor$, where $\lfloor X \rfloor$ is the floor function, is given by the formula $P(Y = y) = S_X(y) - S_X(y+1) = F_X(y+1) - F_X(y)$. On simplification, this reduces to the pmf of $\mathscr{TGD}(q = e^{-\beta}, \alpha)$. ■

### 3.5. Generating functions

**Theorem 4** *The probability generating function (PGF) of $\mathscr{TGD}(q,\alpha)$ is given by*

$$G_Y(z) = \frac{(1-q)(1-\alpha q(1-z)-q^2 z)}{(1-qz)(1-q^2 z)}, \qquad |q^2 z| < 1$$

*Proof.* It is known that the pgf $\mathbb{E}(z^X)$ of $X \sim \mathscr{GD}(q)$ is equal to $\frac{1-q}{1-qz}$ (see p. 215, Johnson et al., 2005).

Therefore pgf of $Y \sim \mathscr{TGD}(q,\alpha)$ is given by

$$G_Y(z) = \mathbb{E}(z^Y) = \sum_{y=0}^{\infty} z^y P(Y = y) = \sum_{y=0}^{\infty} z^y \left((1-\alpha)(1-q)q^y + \alpha(1-q^2)q^{2y}\right)$$

$$= \frac{(1-q)(1-\alpha)}{1-qz} + \frac{\alpha(1-q^2)}{1-q^2 z}$$

The result follows on simplification. ■

**Remark 4** The other generating functions like characteristic function, moment generating function and cumulant generating function can be easily derived from the PGF by using the results $\Phi_Y(z) = G_Y(e^{iz})$, $\mathbb{M}_Y(z) = G_Y(e^z)$ and $\mathbb{K}_Y(z) = \log(G_Y(e^z))$ respectively.

### 3.6. Moments and related measures

Here we derive various moments and related measures of $\mathscr{TGD}(q,\alpha)$.

**Theorem 5** *The $r^{th}$ factorial moment of $Y \sim \mathscr{TGD}(q,\alpha)$ is given by*

$$\mathbb{E}\left(Y_{(r)}\right) = (1-\alpha)r!\left(\frac{q}{1-q}\right)^r + \alpha r!\left(\frac{q^2}{1-q^2}\right)^r.$$

*where $Y_{(r)} = Y(Y-1)\cdots(Y-r+1)$*

***Table 1:*** *Expressions for various measures of $\mathscr{TGD}(\alpha,q)$.*

| S.No. | Measures | Expression |
|---|---|---|
| 1 | Mean $\mathbb{E}(Y)$ | $\frac{q(1-\alpha)+q^2}{1-q^2}$ |
| 2 | Variance $\mathbb{V}(Y)$ | $\frac{q\left(1-\alpha^2+q(1-\alpha^2+q(1-\alpha)+2)\right)}{(1-q^2)^2}$ |
| 3 | Index of Dispersion (ID) | $\frac{q\left(1-\alpha^2+q(1-\alpha^2+q(1-\alpha)+2)\right)}{(1-q^2)(q(1-\alpha)+q^2)}$ |
| 4 | $\gamma^{th} quantile\ (y_\gamma)$ | $\left\lfloor \frac{\log\left(\alpha-1+\sqrt{\alpha^2-2\alpha(1-2\gamma)+1}\right)-\log(2\alpha)}{\log q} \right\rfloor -1$ |
| 5 | Median $(y_{0.5})$ | $\left\lfloor \frac{\log\left(\alpha-1+\sqrt{\alpha^2+1}\right)-\log(2\alpha)}{\log q} \right\rfloor -1$ |

*Proof.* It is known that the $r^{th}$ factorial moment $\mathbb{E}(X_{(r)})$ of $X \sim \mathscr{GD}(q)$ is given by

$$\mathbb{E}\left(X_{(r)}\right) = r!\left(\frac{q}{1-q}\right)^r \tag{4}$$

Therefore the $r^{th}$ factorial moment of $Y \sim \mathscr{TGD}(q,\alpha)$ using equation (3) is given by

$$\mathbb{E}\left(Y_{(r)}\right) = (1-\alpha)(1-q)\sum_{y=r}^{\infty} y_{(r)}q^y + \alpha(1-q^2)\sum_{y=r}^{\infty} y_{(r)}q^{2y} \tag{5}$$

The result then follows upon using (4). ∎

*Note 1.* Alternatively, the above theorem can also be proved using the result $\mathbb{E}(Y_{(r)}) = \frac{d^r}{dz^r}G_Y(z)|_{z=1}$.

By using Theorem 5, the descriptive statistics mean, variance, index of dispersion quantile functions as well as median are given in Table 1. However, we do not present the expressions for skewness as well as kurtosis as they are quite gigantic, instead we present 3-D surface plot of these two measures in Figure 2(a) and 2(b). In Figure 2(a), the $q$-$\alpha$ surface cuts the skewness surface at zero indicated in blue, hence $\mathscr{TGD}(\alpha,q)$ possess positive skewness above $q$-$\alpha$ surface and negative skewness below $q$-$\alpha$ surface. Moreover, if we look in Figure 2(b) horizontal $q$-$\alpha$ surface drawn at value 3 which never intersect the kurtosis surface, indicating leptokurtic nature of $\mathscr{TGD}(\alpha,q)$. Further, Figure 2(c) shows that the horizontal $q$-$\alpha$ surface cuts the ID surface at 1 indicating under or

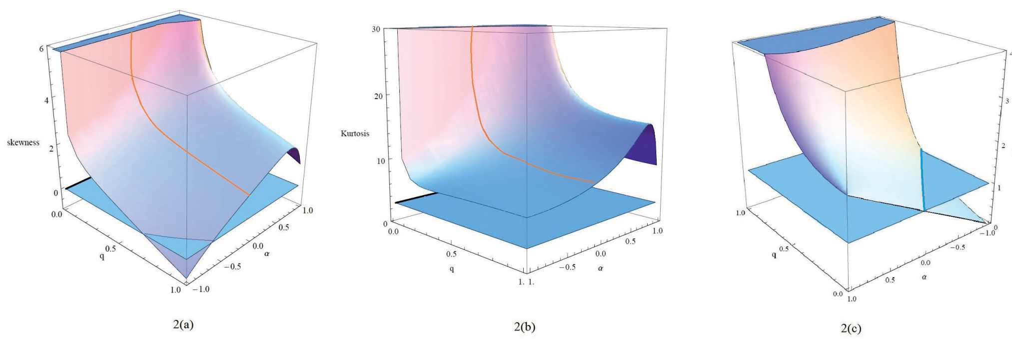


***Figure 2:*** *q-$\alpha$ surface plot of 2(a) Skewness, 2(b) Kurtosis and 2(c) Index of Dispersion of $\mathscr{TGD}(q,\alpha)$.*

over-dispersion for $\alpha \in (-1,0)$ or $(0,1)$ respectively (see Remark 3). Finally skewness and kurtosis of $\mathscr{GD}(q)$ is depicted in red curve on respective surfaces.

**Remark 5** A random number $Y \sim \mathscr{TGD}(q,\alpha)$ can be drawn by first generating a uniform random number $U$ in $(0,1)$ and then using the method of inversion to get a sampled observation $Y$ by using result 4 of Table 1.

## 4. Maximum likelihood estimator

In this section, we focus on the maximum likelihood estimator (MLE), though other estimators can also be derived easily, such as (i) sample proportion of 1's and 0's, (ii) sample quantiles, (iii) method of moments.

For a sample $(y_1, y_2, \cdots, y_n)$ of size $n$ drawn from $\mathscr{TGD}(q,\alpha)$, the likelihood function is given by $L = \prod_{i=1}^{n} \left((1-\alpha)q^{y_i}(1-q) + \alpha q^{2y_i}(1-q^2)\right)$. Taking logarithms on both sides gives the log-likelihood function as

$$l = \log L = n\log(1-q) + n\bar{y}\log(q) + \sum_{i=1}^{n} \log\left((1-\alpha) + \alpha q^{y_i}(1+q)\right) \tag{6}$$

By differentiating (6) with respect to $q$ and $\alpha$ and equating to 0, the following likelihood equations are obtained.

$$\frac{\partial l}{\partial q} = -\frac{n}{1-q} + \frac{n\bar{y}}{q} + \sum_{i=1}^{n} \frac{\alpha q^{y_i} + \alpha y_i (1+q) q^{y_i - 1}}{1-\alpha+\alpha(1+q)q^{y_i}} = 0$$

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n} \frac{(1+q)q^{y_i} - 1}{1-\alpha+\alpha(1+q)q^{y_i}} = 0$$

Since the likelihood equations have no closed form solution, the MLEs $\hat{q}$ and $\hat{\alpha}$ of the parameters $q$ and $\alpha$ can be obtained by maximizing the log-likelihood function using global numerical maximization techniques. Further, the second order partial derivatives of the log-likelihood function are given by

$$\frac{\partial^2 l}{\partial q^2} = -\frac{n}{(1-q)^2} - \frac{n\bar{y}}{q^2} - \sum_{i=1}^{n}\left(\frac{\alpha(1+q)(y_i-1)y_i q^{y_i-2} + 2\alpha y_i q^{y_i-1}}{1-\alpha+\alpha(1+q)q^{y_i}} - \left(\frac{\alpha(1+q)y_i q^{y_i-1} + \alpha q^{y_i}}{1-\alpha+\alpha(1+q)q^{y_i}}\right)^2\right)$$

$$\frac{\partial^2 l}{\partial q \partial \alpha} = \sum_{i=1}^{n}\left(\frac{(1+q)y_i q^{y_i-1} + q^{y_i}}{1-\alpha+\alpha(1+q)q^{y_i}} - \frac{\left(\alpha(1+q)y_i q^{y_i-1} + \alpha q^{y_i}\right)\left((1+q)q^{y_i}-1\right)}{\left(1-\alpha+\alpha(1+q)q^{y_i}\right)^2}\right)$$

$$\frac{\partial^2 l}{\partial \alpha^2} = -\sum_{i=1}^{n}\left(\frac{\left((1+q)q^{y_i}-1\right)^2}{1-\alpha+\alpha(1+q)q^{y_i}}\right)$$

The approximate Fisher information matrix can then be obtained as

$$\begin{pmatrix} \frac{\partial^2 l}{\partial q^2} & \frac{\partial^2 l}{\partial q \partial \alpha} \\ \frac{\partial^2 l}{\partial q \partial \alpha} & \frac{\partial^2 l}{\partial \alpha^2} \end{pmatrix}_{q=\hat{q},\alpha=\hat{\alpha}} \tag{7}$$

where $\hat{q}$ and $\hat{\alpha}$ are the MLEs of $q$ and $\alpha$ respectively.

## 5. Application and data analysis

### 5.1. An actuarial application

In an actuarial context, non-life insurance companies are often interested in modelling the aggregate claim of a portfolio of policies. Let $Z_j$, $j = 1, 2, \cdots$ be the rv denoting the size or amount of the $j^{th}$ claim and $Y$ be the rv denoting the number of claims. Then the aggregate claim of that portfolio is defined as $S = \sum_{j=1}^{Y} Z_j$. Assuming that the claim amounts $Z_j$ to be identically and independently distributed among themselves as well as with claim frequency $Y$, the pdf of S can be obtained as $g_S(s) = \sum_{j=1}^{\infty} p_j f^{*j}(s)$ where $p_j$ denotes the probability of the $j$th claim (called the primary distribution) and $f^{*j}(s)$ is the $j$-fold convolution of $f(s)$, the pdf of the claim amount (the secondary distribution). For more details one can see Rolski et al. (1999), Antzoulakos and Chadjiconstantinidis (2004), Klugman et al. (2008)) and the references therein.

In the following theorem, we present the distribution of aggregate claim when the primary distribution is $\mathscr{TGD}(q,\alpha)$ and the secondary distribution is exponential with mean $(1/\theta)$.

**Theorem 6** *If $\mathscr{TGD}(q,\alpha)$ distribution is the primary distribution and the exponential distribution with parameter $\theta > 0$ is the secondary distribution, then the pdf of rv $S = \sum_{j=1}^{Y} Z_j$ is given by*

$$g_S(s) = \begin{cases} (1-\alpha)(1-q) + \alpha\left(1-q^2\right) & \text{for} \quad s = 0 \\ (1-q)q\theta\left((1-\alpha)e^{-(1-q)s\theta} + q(1+q)\alpha e^{-\left(1-q^2\right)s\theta}\right) & \text{for} \quad s > 0 \end{cases} \tag{8}$$

*Proof.* Since the claim severity distribution follows an exponential distribution with parameter $\theta > 0$, the $j$-fold convolution of the exponential distribution is a gamma distribution with parameter $j$ and $\theta$, having density function

$$f^{*j}(z) = \frac{\theta^j}{(j-1)!} z^{j-1} e^{-\theta z}, \qquad j = 1, 2, \cdots,$$

Hence, the pdf of the rv $S$ is given by

$$\begin{aligned} g_S(s) = \sum_{j=1}^{\infty} p_j f^{*j}(s) &= \sum_{j=1}^{\infty} \frac{\theta^j}{(j-1)!} s^{j-1} e^{-\theta s}\left((1-\alpha)(1-q)q^j + \alpha\left(1-q^2\right)q^{2j}\right) \\ &= (1-q)q\theta\left((1-\alpha)e^{-(1-q)s\theta} + q(1+q)\alpha e^{-\left(1-q^2\right)s\theta}\right) \end{aligned}$$

where $g_S(s)$ has a jump at $s = 0$ with probability $(1-\alpha)(1-q) + \alpha(1-q^2)$. ■

Henceforth, we denote the distribution of $S$ with $\mathscr{TGD}(q,\alpha)$ as primary and exponential as secondary distribution as $\mathscr{CTG}$-$\mathscr{ED}(q,\alpha,\theta)$. Further, it is also well-known that the mean of the aggregate rv is the product of the respective means of the primary and secondary rvs, hence in our proposed model

$$\mathbb{E}(S) = \frac{q(1-\alpha) + q^2}{1-q^2}\frac{1}{\theta}$$

We now compare the aggregate loss model as defined in (8) with the aggregate loss model obtained by considering the geometric distribution as the primary distribution and exponential as the secondary distribution for claim severity, hence the density of the compound geometric-exponential distribution $\mathscr{CG}$-$\mathscr{ED}$ (see pp.152 of Tse, 2009) is given as

$$g_S(s) = \begin{cases} 1-q_1 & \text{for} \quad s=0 \\ (1-q_1)\,q_1\theta e^{-(1-q_1)s\theta} & \text{for} \quad s>0 \end{cases} \tag{9}$$

with mean $\mathbb{E}(X) = \frac{1-q_1}{q_1}\frac{1}{\theta}$.

It is a well known that in the case of reinsurance, the reinsurance company will be interested in those aggregate claim models that are suitable for modelling extreme value. In the following theorem we show that with the same mean and different parameter values, $\mathscr{CTG}$-$\mathscr{ED}(q,\alpha,\theta)$ captures heavy tail values as compared to $\mathscr{CG}$-$\mathscr{ED}(q_1,\theta)$.

**Theorem 7** *With the same mean, $\mathscr{CTG}$-$\mathscr{ED}(q,\alpha,\theta)$ has thinner (thicker) tail as compared to $\mathscr{CG}$-$\mathscr{ED}(q_1,\theta)$ for $-1<\alpha<0(0<\alpha<1)$.*

*Proof.* Without loss of generality, we consider $\theta=1$. By equating the means of $\mathscr{CTG}$-$\mathscr{ED}$ with $\mathscr{CG}$-$\mathscr{ED}$, we get

$$\frac{q(1-\alpha)+q^2}{1-q^2} = \frac{1-q_1}{q_1} \qquad \text{which gives} \qquad q_1 = \frac{1-q^2}{1+q(1-\alpha)}.$$

We now compare the tail behaviour of two distributions by taking the limiting ratio (LR) of their sf (see pp. 60, Tse, 2009):

$$LR = \lim_{t\to\infty} \frac{\bar{G}_{CTG\text{-}ED}(t)}{\bar{H}_{CG\text{-}ED}(t)}$$

where $\bar{G}_{CTG\text{-}ED}(t) = q\left((1-\alpha)e^{-(1-q)t} + \alpha q e^{-(1-q^2)t}\right)$ and $\bar{H}_{CG\text{-}ED}(t) = \frac{q(q+1-\alpha)}{1+q(1-\alpha)}$ $\exp[-\frac{(1-q^2)t}{1+q(1-\alpha)}]$ are respectively the sf of $\mathscr{CTG}$-$\mathscr{ED}(q,\alpha,\theta)$ and $\mathscr{CG}$-$\mathscr{ED}(q_1,\theta)$.

Substituting these values in LR, we obtain

$$LR = \lim_{t\to\infty} \left((1-\alpha)e^{\frac{\alpha q(1-q)t}{1+q(1-\alpha)}} + \alpha q e^{-\frac{(1-\alpha)q(1-q^2)t}{1+q(1-\alpha)}}\right)$$

Now observe that for $-1<\alpha<0$, $LR = \lim_{t\to\infty} \frac{\bar{G}_{CTG\text{-}ED}(t)}{\bar{H}_{CG\text{-}ED}(t)} = 0$.

$\Rightarrow \mathscr{CTG}$-$\mathscr{ED}$ has thinner tail then $\mathscr{CG}$-$\mathscr{ED}$.

whereas for $0<\alpha<1$, $LR = \lim_{t\to\infty} \frac{\bar{G}_{CTG\text{-}ED}(t)}{\bar{H}_{CG\text{-}ED}(t)} = \infty$.

$\Rightarrow \mathscr{CTG}$-$\mathscr{ED}$ has thicker tail then $\mathscr{CG}$-$\mathscr{ED}$. ∎

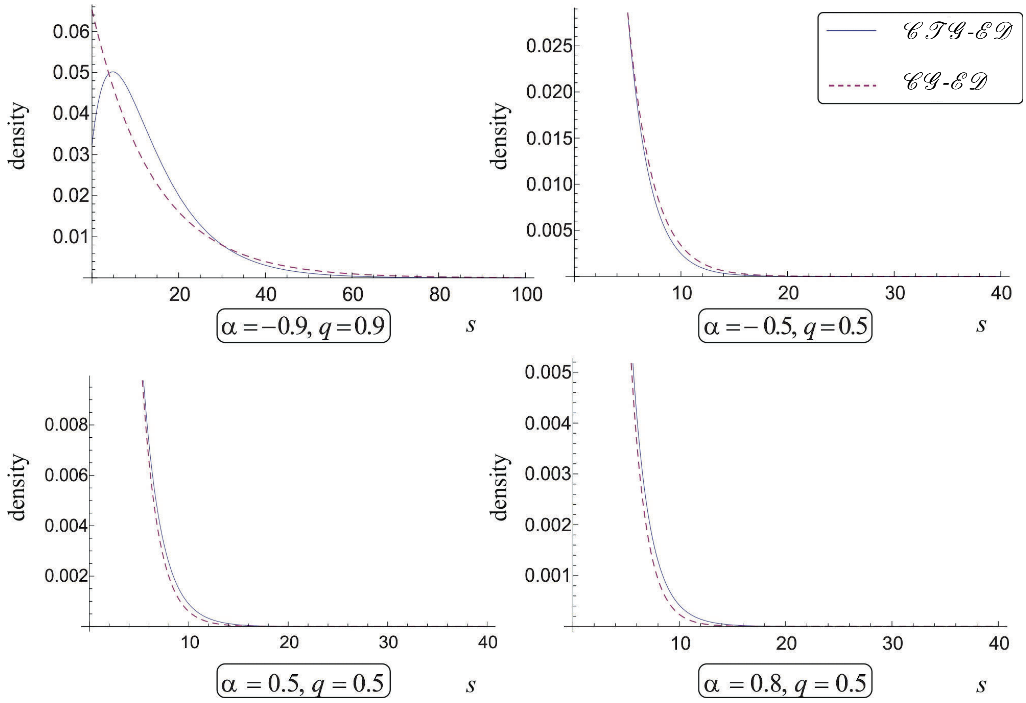


***Figure 3:*** *PDF of aggregate loss rv (compound geometric-exponential distribution in red dashed lines and compound transmuted geometric-exponential distribution in blue lines) for different values of parameter q and* $\alpha$.

Tail behaviour of $\mathscr{CTG}$-$\mathscr{ED}$ and $\mathscr{CG}$-$\mathscr{ED}$ distributions for different parameter values are presented in Figure 3.

### *5.1.1. Illustration: aggregate loss modelling*

To illustrate the applicability and superiority of the proposed aggregate model compared to other existing aggregate models such as Poisson-exponential, negative binomial-exponential and geometric-exponential, in short $X$-exponential models having densities indicated in Table 2, we consider a vehicle insurance data set of one-year vehicle insurance policies taken out in 2004 or 2005. There are 67856 policies of which 4624 (6.8%) had at least one claim. Table 3 gives some in-depth information about the claims frequency ($X$) and total claim($S$) for the data set. Full access to this dataset is available on the webiste of the Faculty of Business and Economics, Macquarie University, Australia – see also Jong and Heller (2008). As the variability in total claim data is very high, we scale these observations by scale factor 0.001, remembering the fact that scaling will not effect the comparison, and apply the maximum likelihood method to estimate the parameters of aggregate model. The log-likelihood function for proposed $\mathscr{CTG}$-$\mathscr{ED}(q,\alpha,\theta)$ model is given as

$$
\begin{aligned}
l =&(n-m)\log(\theta(1-q)q)+m\log\left(\alpha\left(1-q^2\right)+(1-\alpha)(1-q)\right) \\
&+\sum_{s_i>0}\log\left((1-\alpha)e^{-\theta(1-q)s_i}+\alpha q(q+1)e^{-\theta(1-q^2)s_i}\right)
\end{aligned}
$$

where $m$ is the number of policies having no claim, $(n-m)$ is the number of policies having at least one claim and $n$ be the total number of policies. As we can see the log-likelihood equations obtained from the log-likelihood function cannot help in determining the estimates of parameter, hence we make use of numerical techniques to search global maximum of log-likelihood surface. We make use of `FindMaximum` function of Mathematica software package v.10.0. The estimates and other comparative measures such as log-likelihood value(LL), Akaike Information Criteria(AIC) are shown in Table 4. Based on the AIC value it can be claimed that the proposed $\mathscr{CTG}$-$\mathscr{ED}(q,\alpha,\theta)$ model gives the best fit for the vehicle insurance data among all the models considered.

***Table 2:*** *Density of X-exponential models.*

| S.No. | distribution of $X$ | Density of aggregate rv. |
|---|---|---|
| 1 | Poisson | $g_S(s)=\begin{cases} e^{-\lambda} & \text{for} \quad s=0 \\ \sqrt{\frac{\theta\lambda}{s}}e^{-\theta s-\lambda}\mathscr{I}_1\left(2\sqrt{\lambda\theta s}\right) & \text{for} \quad s>0 \end{cases}$ |
| | | where, $\mathscr{I}_1(.)$ is the modified Bessel function of first kind |
| 2 | Negative binomial | $g_S(s)=\begin{cases} (1-q)^r & \text{for} \quad s=0 \\ q\theta r(1-q)^r e^{-\theta s}\,{}_1F_1(r+1;2;\theta qs) & \text{for} \quad s>0 \end{cases}$ |
| | | where ${}_1F_1(.;.;.)$ is the confluent hypergeometric function |
| 3 | Geometric | $g_S(s)=\begin{cases} 1-q & \text{for} \quad s=0 \\ (1-q)\,q\theta e^{-(1-q)s\theta} & \text{for} \quad s>0 \end{cases}$ |

***Table 3:*** *Descriptive statistics of the vehicle insurance dataset.*

| | Number of claims | Total claim amount |
|---|---|---|
| Mean | 0.072 | 137.27 |
| variance | 0.077 | 1115769.69 |
| Index of Dispersion | 1.0734 | 8128.29 |
| min | 0 | 0 |
| max | 4 | 55922.1 |

***Table 4:*** *Estimated value of parameters of X-exponential models.*

| S.No. | Distribution of $X$ | Estimated parameter | LL | AIC |
|---|---|---|---|---|
| 1 | Poisson | $\hat{\lambda} = 0.12057$, $\hat{\theta} = 0.87832$ | $-25699.3$ | 51402.6 |
| 2 | Negative binomial | $\hat{r} = 0.51168$, $\hat{q} = 0.1291$, $\hat{\theta} = 0.55250$ | $-24740.6$ | 49487.2 |
| 3 | Geometric | $\hat{q} = 0.06814$, $\hat{\theta} = 0.53273$ | $-24745.7$ | 49495.4 |
| 4 | Transmuted Geometric | $\hat{q} = 0.2313$, $\hat{\alpha} = 0.9147$, $\hat{\theta} = 0.5693$ | $-24702.0$ | 49410.0 |

## 5.2. Count data modelling

In this section we demonstrate the utility of $\mathscr{TGD}(q,\alpha)$ in count data modelling considering a real data set on the number of automobile insurance claims per policy in portfolios from Great Britain and Zaire (Willmot, 1987). This data set contain 87% of zeros as well as with variance to mean ratio 1.051 indicating the presence of over-dispersion in the data set. Hence the proposed model is expected to provide adequate fit. Here $\mathscr{TGD}(q,\alpha)$ is compared with the following existing ones.

i. Negative binomial ($\mathscr{NB}$) (Johnson et al., 2005).

ii. Poisson inverse Gaussian (Willmot, 1987) ($\mathscr{P}-\mathscr{IG}$) with pmf defined as

$$P(X=x) = \frac{1}{x!}\sqrt{\frac{2\phi}{\pi}}e^{\phi/\mu}\phi^{-\frac{1}{4}+\frac{x}{2}}\left(2+\frac{\phi}{\mu^2}\right)^{\frac{1-2x}{4}}K_{\frac{1}{2}-x}\left(\sqrt{2\phi+\frac{\phi^2}{\mu^2}}\right)$$

where $x = 0,1,2,\ldots,$ $\phi,\mu > 0$ and $K_a(.)$ is modified Bessel function of the third kind.

iii. New discrete distribution (Gómez et al., 2011) ($\mathscr{ND}$) with pmf

$$P(X=x) = \frac{\log(1-\alpha\theta^x)-\log(1-\alpha\theta^{x+1})}{\log(1-\alpha)}$$

where $x = 0,1,2,\ldots,$ $\alpha < 1, 0 < \theta < 1$, and

iv. Zero distorted generalized geometric (Sastry et al., 2014) ($\mathscr{ZDGGD}$) with pmf

$$P(X=x) = \begin{cases} 1-q^{\alpha+1} & \text{if} \quad x=0 \\ (1-q)q^{\alpha+x+1} & \text{if} \quad x>0 \end{cases}$$

where $0 < q < 1, -1 < \alpha < 1$.

***Table 5:*** *Fit of automobile claim data in Great Britain, 1968 (Willmot, 1987).*

| # claims | Observed | Expected frequency | | | | |
|---|---|---|---|---|---|---|
| | Frequency | $\mathscr{NB}$ | $\mathscr{P}-\mathscr{IG}$ | $\mathscr{ND}$ | $\mathscr{XDGGD}$ | $\mathscr{TGD}$ |
| 0 | 370412 | 370438.99 | 370435 | 370413 | 370412 | 370412 |
| 1 | 46545 | 46451.28 | 46476.4 | 46538.3 | 46555.16 | 46546.7 |
| 2 | 3935 | 4030.50 | 3995.76 | 3942.39 | 3913.70 | 3929.19 |
| 3 | 317 | 297.82 | 307.67 | 318.57 | 329.00 | 323.23 |
| 4 | 28 | 20.09 | 23.12 | 25.64 | 27.76 | 26.53 |
| 5 | 3 | 1.28 | 1.74 | 2.06 | 2.38 | 2.38 |
| Total | 421240 | 421240 | 421240 | 421240 | 421240 | 421240 |
| estimated | | $\hat{p}=0.338$ | $\hat{\phi}=0.338$ | $\hat{\alpha}=-1.349$ | $\hat{q}=0.0845$ | $\hat{q}=0.0821$ |
| parameter | | $\hat{r}=0.131$ | $\hat{\mu}=0.131$ | $\hat{\theta}=0.080$ | $\hat{\alpha}=-0.146$ | $\hat{\alpha}=-0.5121$ |
| $\chi^2$-statistic | | 9.15 | 2.74 | 0.71 | 0.72 | **0.31** |
| df | | 3 | 3 | 3 | 3 | 3 |
| p-value | | 0.03 | 0.43 | 0.87 | 0.87 | **0.96** |
| $l_{\max}$ | | –171136.9 | –171134.4 | –171133.0 | –171134.1 | **–171133.0** |

***Table 6:*** *SE, CI, and CL of estimated parameters for the data sets in Table 5.*

| Models | Parameters | ML Estimate | S.E. | CI | CL |
|---|---|---|---|---|---|
| $\mathscr{NB}$ | $\hat{r}$ | 0.131 | 0.5684 | (–0.983, 1.255) | 0.2228 |
| | $\hat{p}$ | 0.338 | 0.0011 | (0.336, 0.340) | 0.0039 |
| $\mathscr{P}-\mathscr{IG}$ | $\hat{\phi}$ | 0.338 | 0.0188 | (0.3017, 0.3756) | 0.0739 |
| | $\hat{\nu}$ | 0.131 | 0.0005 | (0.1306,0.1328) | 0.0022 |
| $\mathscr{ND}$ | $\hat{\alpha}$ | –1.349 | 0.1120 | (–1.5686, –1.1295) | 0.4390 |
| | $\hat{\theta}$ | 0.080 | 0.0018 | (0.0768, 0.0840) | 0.0071 |
| $\mathscr{XDGGD}$ | $\hat{q}$ | 0.0845 | 0.0011 | (0.0817, 0.0863) | 0.0046 |
| | $\hat{\alpha}$ | –0.146 | 0.0051 | (–0.1160, –0.1359) | 0.0200 |
| $\mathscr{TGD}$ | $\hat{q}$ | 0.0821 | 0.0011 | (0.079, 0.0844) | 0.0046 |
| | $\hat{\alpha}$ | –0.5121 | 0.0236 | (–0.558, –0.465) | 0.0920 |

The data fitting results for the above four distributions in (i) to (iv) presented in Table 5 are taken from the respective papers. From the findings of the data fitting presented in Table 5, to assess the fit of the competing models we first compare the expected frequencies with the observed one for each model, which reveals that the $\mathscr{TGD}(q,\alpha)$ predicts most of the observed counts more closely than the other models. The $\chi^2$ statistics and its $p$-values implies lack of fit for NB and also for PIG. The rest of the models provides good fit, with $\mathscr{TGD}(q,\alpha)$ being the best among the lot with highest with $p$-value of 0.96. Moreover, we also compute standard error (SE), confidence interval (CI) and confidence length (CL) for the parameter estimates. It can be clearly seen from Table 6, that the SE of the estimates of proposed distribution is smaller compared to other distributions. Hence, it is envisaged that the proposed distribution may serve as an alternative model for modelling data with a large proportion of zeros and over-dispersion.

### *5.3. Count regression modelling including covariates*

In this section, we present the count regression modelling assuming the discrete response variable $(Y)$ as a function of a set of independent (exogenous) variables. Furthermore, we also consider that the mean $(\theta)$ of response variable is related with the set of exogenous variables by the positive valued function $\theta = \theta(\boldsymbol{x})$. There are several possible choices for the selection of function $\theta(\boldsymbol{x})$ and thus to ensure the non-negativity of the mean of the response variable, we consider the log-link function as $\theta_i(\boldsymbol{x}) = e^{\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta}}$, where $\boldsymbol{x}_i^{\mathsf{T}} = (x_{i1}, x_{i2}, \cdots, x_{ip})$ and $\boldsymbol{\beta}^{\mathsf{T}} = (\beta_1, \beta_2, \cdots, \beta_p)$ be the set of covariates and their coefficients. This selection of log-link function includes both random and fixed effects on the same exponential scale. Further, to estimate the parameters, we use following reparametrization

$$\nu = 1 - \alpha \quad \text{and} \qquad q = \left(-\nu + \sqrt{4\theta + 4\theta^2 + \nu^2}\right)/2(1+\theta)$$

where $\theta_i(\boldsymbol{x}) = e^{\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta}}$. The above re parametrization enable us to bring the regression coefficients $(\boldsymbol{\beta})$ and parameters of the response variable into the log-likelihood functions. The log-likelihood function for a random sample $(y_i, \boldsymbol{x}_i)$ of size $n$ with count $y_i$ and a vector $\boldsymbol{x}_i$ of covariates for $i = 1, 2, \cdots, n$ can be written as

$$l(\nu, \theta | y, \boldsymbol{x}) = \textstyle\sum_{i=1}^{n} \log \left( \nu \left( 1 - \dfrac{-\nu + \sqrt{4\theta_i + 4\theta_i^2 + \nu^2}}{2(1+\theta_i)} \right) \left( \dfrac{-\nu + \sqrt{4\theta_i + 4\theta_i^2 + \nu^2}}{2(1+\theta_i)} \right)^{y_i} \right.$$
$$\left. + (1-\nu) \left( 1 - \left( \dfrac{-\nu + \sqrt{4\theta_i + 4\theta_i^2 + \nu^2}}{2(1+\theta_i)} \right)^2 \right) \left( \dfrac{-\nu + \sqrt{4\theta_i + 4\theta_i^2 + \nu^2}}{2(1+\theta_i)} \right)^{2y_i} \right)$$

The parameters $(\nu, \beta_1, \beta_2, \cdots, \beta_p)$ in the above log-likelihood function can be estimated by maximizing the log-likelihood function for a given data set using the `optim()` function in R (for more details one can browse `https://stat.ethz.ch/R-manual/R-devel/library/stats/html/optim.html`), where the initial values of the parameters were chosen from Poisson regression model.

In the next section we present an application of the proposed count regression model to a real life data set and compare its performance with following popular regression models:

i. Poisson regression model

$$P(Y_i = y_i|\mu_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \qquad y_i = 0,1,2,... \tag{10}$$

where $\mu_i > 0$. The regression model is obtained by putting $\mu_i = e^{\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}}$.

ii. Generalized Poisson model ($\mathscr{GP}$-2): The pmf of a generalized Poisson ($\mathscr{GP}$-2) regression model (Consul and Famoye , 1992, Yang et al., 2009) is given as

$$P(Y_i = y_i|\theta_i,\nu_i) = \frac{\mu_i\left(\mu_i + \phi\mu_i y_i\right)^{y_i-1}}{(1+\phi\mu_i)^{y_i} y_i!} e^{\left(-\frac{\mu_i+\phi\mu_i y_i}{1+\phi\mu_i}\right)}, \quad y_i = 0,1,2,... \tag{11}$$

where $\phi > 0$ is dispersion parameter and $\mu_i = e^{\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}}$ in (11). For more details refer Yang et al. (2009) and finally with

iii. Generalized Negative Binomial ($\mathscr{NB}$-2) (Greene, 2008): The pmf of a generalized negative binomial ($\mathscr{NB}$-2) regression model is given as

$$P(Y_i = y_i|\theta,r_i) = \frac{\Gamma(\theta+y_i)r_i^{\theta}(1-r_i)^{y_i}}{y_i!\Gamma(\theta)} \tag{12}$$

where $y_i = 0,1,2,\ldots$ and $r_i = \theta/(\theta+\lambda_i)$ and $\lambda_i = e^{\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}}$.

***Table 7:*** *Exploratory data description.*

| Variable | Nature of variable | Measurement | Mean | Variance |
|---|---|---|---|---|
| OFP | Response | Number of physician visits | 6.046 | 57.169 |
| HOSP | Explanatory | Number of days of hospital stays | 0.297 | 0.513 |
| POORHLTH | | Self-perceived health status, poor =1, else =0. | 0.13 | 0.113 |
| EXCLHLTH | | Self-perceived health status, excellent =1, else 0 | 0.071 | 0.066 |
| NUMCHRON | | Number of chronic conditions | 1.533 | 1.788 |
| MALE | | Gender; male = 1, else =0 | 0.408 | 0.241 |
| SCHOOL | | Number of year of education | 10.355 | 13.25 |
| PRIVINS | | Private insurance indicator, yes =1, no = 0 | 0.794 | 0.164 |

***Table 8:*** *Maximum likelihood estimates of the parameters of different regression models.*

| Parameter | Poisson | | | $\mathscr{GP}$-2 | | | $\mathscr{NB}$-2 | | | $\mathscr{TGM}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | est. (s.e) | t-statistic | p-value | est. (s.e.) | t-statistic | p-value | est. (s.e.) | t-statistic | p-value | est.(s.e.) | t-statistic | p-value |
| Intercept($\beta_1$) | 0.99(0.04) | 27.51 | $< 0.00001$ | 0.82 (0.09) | 9.12 | 0.00001 | 0.88(0.09) | 10.17 | $< 0.00001$ | 0.205 (0.17) | 1.207 | 0.22758 |
| HOSP($\beta_2$) | 0.19(0.01) | 20.39 | $< 0.00001$ | 0.28 (0.04) | 6.22 | 0.00001 | 0.24(0.03) | 7.14 | $< 0.00001$ | 0.090 (0.04) | 2.0397 | 0.04249 |
| POORHLTH($\beta_3$) | 0.21(0.03) | 8.28 | $< 0.00001$ | 0.39 (0.09) | 4.53 | 0.00001 | 0.31(0.07) | 4.32 | 0.000015 | –1.174 (0.65) | –1.8106 | 0.07035 |
| EXCLHLTH($\beta_4$) | –0.21(0.04) | -4.84 | $< 0.00001$ | –0.17 (0.09) | –1.8 | 0.07201 | –0.18(0.01) | -1.92 | 0.054857 | 1.132 (0.11) | 10.1784 | 0.00001 |
| NUMCHRON($\beta_5$) | 0.16(0.01) | 24.48 | $< 0.00001$ | 0.20 (0.02) | 9.54 | 0.00001 | 0.19(0.02) | 10.04 | $< 0.00001$ | 0.166 (0.03) | 5.7383 | 0.00001 |
| MALE($\beta_6$) | –0.1(0.02) | –5.31 | $< 0.00001$ | –0.13 (0.05) | -2.54 | 0.01116 | –0.12(0.05) | –2.52 | 0.01181 | 0.073 (0.10) | 0.7439 | 0.45702 |
| SCHOOL($\beta_7$) | 0.03(0.001) | 12.19 | $< 0.00001$ | 0.04 (0.01) | 5.29 | 0.00001 | 0.04(0.01) | 5.29 | $< 0.00001$ | 0.069 (0.02) | 3.6983 | 0.00022 |
| PRIVINS($\beta_8$) | 0.15(0.02) | 5.85 | $< 0.00001$ | 0.15 (0.07) | 2.27 | 0.02331 | 0.14(0.06) | 2.28 | 0.02271 | 0.471 (0.11) | 4.2646 | 0.00001 |
| Dispersion parameter | — | — | | 0.28 (0.01) | 30.08 | 0.00001 | 1.16 (0.04) | 25.03 | $< 0.00001$ | 0.066(0.034) | 1.9064 | 0.05674 |
| $V$ | | 10.6736 | | | –1.188642 | | | –2.28383 | | | — | |
| $l_{\max}$ | | –8813.74 | | | –5614.72 | | | –5607.2 | | | **–4973.1** | |
| AIC $(-2l_{\max}+2k)$ | | 17643 | | | 11247.44 | | | 11232.39 | | | **9964.202** | |

### 5.3.1. A numerical illustration of count regression

We examine the US National Medical Expenditure Survey 1987/88 (NMES) data obtained from Journal of Applied Econometrics 1997 Data Archive at `http://qed.econ.queensu.ca/jae/1997-v12.3/ deb-trivedi/`, which were originally employed by Deb and Trivedi (1997) in their analysis of various measures of health-care utilization. For illustration purpose we consider the first 2000 observations for fitting the regression model. The exploratory data description of the response variable as well as the set of explanatory variables is given in Table 7, from where it can be seen that the mean and variance of the number of physician visit (OFP) variable indicates presence of the over-dispersion as well as existence of large number of zeros. Hence it seems appropriate to apply our model for the present data set with the number of physician visits (OFP) as the response variable and remaining seven as explanatory variables.

Table 8 presents the maximum likelihood estimates of the parameters of the models Poisson($\mathscr{P}$), negative binomial($\mathscr{NB}$-2), generalized Poisson ($\mathscr{GP}$-2), and transmuted geometric ($\mathscr{TGM}$), their standard errors, $t$-statistics and $p-$values.

For comparison between the different fitted models, we have used the value of the maximum of the log-likelihood function ($l_{\max}$) and the Akaike information criterion (AIC). The model with the lowest AIC value is considered to be the best. It can be observed that the estimates of all parameters except the parameters of POORHLTH, MALE and dispersion parameter are found significant at 5% level of significance. Unlike the other models considered here the number of physician visit has not been influenced by the gender profile and poor health status of the patient. Most of the estimated parameters values under the $\mathscr{TGM}$ model differs in values obtained under other competitive models. The estimate of dispersion parameter for $\mathscr{TGM}$ found significant at 5% level of significance as opposed to $\mathscr{GP}$-2 and $\mathscr{NB}$-2 models which gives an indication of capturing dispersion of data. Moreover, with respect to the values of $l_{\max}$ and consequently $AIC$, our proposed model turns out to be the best. Hence, we conclude that proposed $\mathscr{TGM}$ regression model gives satisfactory fit and can be considered suitable for count data regression analysis.

Since the models under consideration namely $\mathscr{P}$, $\mathscr{NB}$-2, $\mathscr{GP}$-2, are not nested within $\mathscr{TGM}$, it may of interest to employ the Vuong test (see Vuong (1989)) for non-nested models to discriminate among these models. The Vuong statistic is given by

$$V = \frac{1}{\zeta\sqrt{n}}\left(l_{\mathscr{TGM}}(\hat{\Theta}_1) - l_g(\hat{\Theta}_2)\right) \tag{13}$$

where

$$\zeta^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{f_{\mathscr{TGM}}(\hat{\Theta}_1|y_i,x_i)}{g(\hat{\Theta}_2|y_i,x_i)}\right)\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{f_{\mathscr{TGM}}(\hat{\Theta}_1|y_i,x_i)}{g(\hat{\Theta}_2|y_i,x_i)}\right)\right)^2$$

where $f_{\mathscr{TGM}}$ and $g$ represent $\mathscr{TGM}$ and the other competing model respectively.

As statistic $V$ ia asymptotically standard normal, the rejection of test in favour of $\mathscr{TGM}$ occurs if $V > 1.96$, at the 5% level of significance . From our findings in Table 8, it is seen that the proposed $\mathscr{TGM}$ regression model is preferred over Poisson (since $V > 1.96$), but do not distinguish between $\mathscr{GP}$-2 model (since $-1.96 < V < 1.96$). However the test rejects the $\mathscr{TGM}$ model when compared with $\mathscr{NB}$-2 (since $V < -1.96$).

## 6. Concluding remarks

In this paper the transmutation technique is used to offer a new flexible generalization of the geometric distribution as a viable alternative to some existing models. Different distributional properties of the distribution are found to be simple and attractive. The theoretical result regarding possibility of applying this new distribution to model aggregate claim in the actuarial context is presented and its suitability for modelling large aggregate claims is established and complimented with a real life data set. Illustrative data fitting with the proposed model for a popular data set from automobile insurance sector having over-dispersion turned out to be very useful. Finally, a count regression model based on the proposed distribution provided best fit in terms of the AIC value when compared with some existing models for analysing a data set from the health sector. Based on these findings, it is envisaged that the transmuted geometric distribution with two parameters can be very useful in modelling and analysis of count data of different types. Further, this idea of applying transmutation to discrete distribution may be applied to construct new generalizations of other distributions.

## Acknowledgments

The authors gratefully acknowledge the suggestions of the editor-in-chief and three anonymous referees on earlier versions of the manuscript which resulted in much improved presentation.

## References

Antzoulakos, D. and Chadjiconstantinidis, S. (2004). On mixed and compound mixed Poisson distributions. *Scandinavian Actuarial Journal*, 3, 161–188.

Chakraborty, S. and Gupta, R. D. (2015). Exponentiated geometric distribution: another generalization of geometric distribution. *Communication in Statistics-Theory and Methods*, 44, 1143–1157.

Chakraborty, S.(2015). Generation of discrete analogues of continuous distributions-a survey of methods and constructions. *Journal of Statistical Distributions and Applications*, 2, 6.

Consul, P. C. and Famoye, F. (1992). Generalized Poisson regression model. *Communication in Statistics-Theory and Methods*, 2, 89–109.

Deb, P. and Trivedi, P. K. (1997). Demand for medical care by the Elderly: a finite mixture approach. *Journal of Applied Econometrics*, 12, 313–336.

Drouet-Mari, D, Kotz, S. (2001). *Correlation and Dependence*. Imperial College press, London.

Gómez-Déniz, E. (2010). Another generalization of the geometric distribution. *Test*, 19, 399–415.

Gómez-Déniz, E., Sarabia, J. M. and Odeja, E. C. (2011). A new discrete distribution with actuarial applications. *Insurance: Mathematics and Economics*, 48, 406–412.

Greene, W. (2008). Functional forms for the negative binomial model for count data. *Economics Letter*, 99, 585–590.

Jain, G. C. and Consul, P. C. (1971). A generalized negative binomial distribution. *SIAM Journal of Applied Mathematics*, 21, 501–513.

Johnson, N. L. Kemp, A. W. and Kotz, S. (2005). *Univariate Discrete Distributions*. 2nd ed. Wiley, New York.

Jong, D. P. and Heller, G. Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press, Cambridge.

Klugman, S. A., Panjer, H. H. and Willmot, G. E. (2008). *Loss Models: From Data to Decisions*. 3rd ed. John Wiley and Sons, New York, 3, 161–188.

Kozubowski, T.J. and Podgórski, K. (2016). Transmuted distributions and random extrema. *Statistics and Probability Letters*, available online 13 April 2016, in press.

Makčutek, J. (2008). A generalization of the geometric distribution and its application in quantitative linguistics. *Romanian Rep Phys*, 60, 501–509.

Mirhossaini, S. M. and Dolati, A. (2008). On a new generalization of the exponential distribution. *Journal of Mathematical Extension*, 3, 27–42.

Oguntunde, P. E. and Adejumo, A. O. (2015). The transmuted inverse exponential distribution. *International Journal of Advanced Statistics and Probability*, 3, 1–7.

Owoloko, E. A., Oguntunde, P. E. and Adejumo, A. O. (2015). Performance rating of the transmuted exponential distribution: an analytical approach. *Scandinavian Actuarial Journal*, 4, 1–15.

Philippou, A. N., Georghiou, C. and Philippou, G. N. (1983). A generalized geometric distribution and some of its properties. *Statistics and Probability Letters*, 1, 171–175.

Rolski, T., Schmidli, H., Schmidt, V. and Teugel, J. (1999). *Stochastic Processes for Insurance and Finance*. John Wiley and Sons, New York.

Sastry, D. V. S., Bhati, D., Rattihalli, R. N. and Gómez-Déniz, E.(2004). On zero distorted generalized geometric distribution. *Communication in Statistics-Theory and Methods*, accepted.

Shaw, W. and Buckley, I. (2007). The alchemy of probability distributions: beyond Gram- Charlier expansions and a skew-kurtotic-normal distribution from a rank transmutation map. *Research Report*. Available in arXiv:0901.0434v1 [q-fin.ST].

Tripathi, R. C., Gupta, R. C. and White, T. J.(1987). Some generalizations of the geometric distribution. *Sankhyā*, Series B, 49, 218–223.

Tse, Yiu-Kuen (2009). *Non-life Actuarial Models Theory, Methods and Evaluation*. Cambridge University Press, UK.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307–333.

Warde, W. D. and Katti, S. K.(1971). Infinite divisibility of discrete distributions II. *Annual of Mathematical Statistics*, 42, 1088–1090.

Willmot, G. E. (1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial Journal*, 3-4, 113–127.

Yang, Z., Hardin, J. W. and Addy, C. L. (2009). A score test for overdispersion in Poisson regression based on the generalized Poisson-2 model. *Journal of Statistical Planning and Inference*, 139, 1514–1521.

Yousof, H. M., Afify, A. Z., Alizadeh, M., Butt, N. S., Hamedani, G. G. and Ali, M. M. (2015). The Transmuted exponentiated generalized-G family of distributions. *Pakistan Journal of Statistics and Operations Research*, DOI: 10.18187/pjsor.v11i4.1164.