Revista Colombiana de Estadística
January 2016, Volume 39, Issue 1, pp. 1 to 15
DOI: http://dx.doi.org/10.15446/rce.v39n1.55131

A Trinomial Difference Distribution

Una distribución de diferencia trinomial

Maha A. Omair[1,a], Abdulhamid Alzaid[1,b], Omalsad Odhah[2,c]

[1]Department of Statistics and Operations Research, College of Sciences, King Saud University, Riyadh, Saudi Arabia

[2]Department of Mathematics, College of Sciences and Humanities, Prince Sattam Bin Abdulaziz University, Hotat Bani Tamim, Saudi Arabia

Abstract

A trinomial difference distribution is defined and its distributional properties are illustrated. This distribution present the binomial difference distribution as a special case. The moment estimators and maximum likelihood estimators of the trinomial difference distribution are compared via simulation study. Two applications are modeled with the trinomial difference distribution and compared with other possible distributions.

***Key words*:** Binomial Distribution, Discrete Distribution, Maximum Likelihood Estimate, Moment Estimate, Poisson Distribution, Trinomial Distribution.

Resumen

Una distribución de diferencia trinomial se define en este artículo así como sus propiedades distribucionales. Esta distribución cuenta con la distribución de diferencia binomial como un caso particular. Los estimadores de momentos y de máxima verosimilitud son comparados vía un estudio de simulación. Dos aplicaciones son modelados con la distribución diferencia trinomial y se comparan con otras distribuciones posibles.

***Palabras clave*:** distribución binomial, distribución discretos, estimación de máxima verosimilitud, momento estimar, distribución de Poisson, distribución binomial.

[a]Assistant Professor. E-mail: maomair@ksu.edu.sa
[b]Professor. E-mail: Alzaid@ksu.edu.sa
[c]Assistant Professor. E-mail: o.odhah@psau.edu.sa

1. Introduction

Recently, discrete distributions defined on the set of integers have attracted the attention of many researchers. The most popular ways to define distributions on $\mathbb{Z}$ are: the differences between two nonnegative discrete random variables and the discrete version of continuous distributions on R. The main distributions on the set $\mathbb{Z}$ are Poisson difference (Skellam 1946), discrete normal and discrete Laplace.

The Poisson difference distribution has many applications in different fields. Karlis & Ntzoufras (2009) modeled the difference of the number of goals in football games with the Poisson difference distribution and the zero inflated Poisson difference distribution. Karlis & Ntzoufras (2006) also used the Poisson difference and the zero inflated Poisson difference distributions to model the difference between the decayed, missing and filled teeth (DMFT) index before and after treatment. Alzaid & Omair (2010) have used the Poisson difference distribution to model the difference in the price of a share every minute as number of ticks and the occupancy in a nursery intensive care unit in a hospital. Alzaid & Omair (2014) defined the integer valued autoregressive model of first order with the Poisson difference marginal distribution. Bakouch, Kachour & Nadarajah (2013) proposed the first reflected version of the Poisson distribution over the set of all integers, referred to as the extended Poisson (E-Po) distribution. They applied the extended Poisson distribution to the change in number of students between two sessions of a course in a specific group from the Bachelor program at IDRAC International Management School (Lyon, France).

Ong, Shimizu & Min Ng (2008) defined the general difference between two discrete random variables from the Panjer family. As a special case, they defined the difference between two negative binomial distributions. Goodness of fit and tests have been considered. In order to illustrate, an application to the difference for the assessment number of palpable lymph nodes deleted by two physicians on 32 participants on a prospective study of men with AIDS or an AIDS-related conditions was provided.

Inusah & Kozubowski (2006) studied discrete skew Laplace distribution (DL). This distribution shares many properties of the skew Laplace distribution. In the symmetric case, this leads to a discrete analogue of the classical Laplace distribution.

Kemp (1997) presented the discrete Normal distribution (DN-distributions) and showed that among all discrete distributions with given expectation and variance and support equal to the set of integers, DN maximizes entropy.

The difference between two independent random variables following the same binomial distribution has been derived by Castro (1952). Kotz, Johnson & Kemp (1992) discussed the distribution of the difference between two independent binomial distributions having the same probability of success and possibly a different number of trials. They also showed that when the probability of success equals the probability of failure the distribution of the difference reduces to the binomial distribution with the general form. The standardized difference of two independent binomial distributions with different parameters tends to unit normal distribution as both numbers of trials tend to infinity.

Definition 1. For any pair of variables (X, Y) that can be written as $X = W_1+W_3$ and $Y = W_2 + W_3$ with $W_1 \sim Bin(n, p)$ independent of $W_2 \sim Bin(n, q)$ and W_3 following any distribution, the probability mass function $(p.m.f)$ of $Z^* = X - Y = W_1 - W_2$ is given by

$$f(z^*) = \sum_{r=\max(0,-z^*)}^{\min(n,n-z^*)} \binom{n}{z^*+r}\binom{n}{r} p^{z^*+r} q^r (1-p)^{n-z^*-r}(1-q)^{n-r},$$

$$z^* = -n, -n+1, \dots, n.$$

Z^* is said to have the binomial difference distribution denoted by $BD(n, p, q)$. The random variable W_3 can be viewed as a background effect or a correlation between X and Y that is totally removed after differencing.

The aim of the paper is to define the trinomial difference distribution on Z, which extends the binomial difference distribution. It is shown that this distribution can be used to model accident data while the binomial difference fails to model such data. In the trinomial difference distribution, not all the correlation between the differenced random variables is removed by differencing. This is not the case in the definition of the binomial difference distribution when the two differenced variables are either independent are either independent or if there is some background effect then all the dependency is removed by differencing. The trinomial difference distribution can be overdispersed i.e. the variance is greater than the absolute value of the mean, equidispersed or underdispersed and can also be symmetric, positively skewed or negatively skewed depending on the value of the parameters. The remainder of this paper is organized as follows: a definition of the trinomial difference distribution with some properties is introduced in Section 2. In Section 3, moment and maximum likelihood estimators were derived and compared via simulation. In Section 4, two real applications are modeled with the trinomial difference distribution and other possible distributions. Finally, a conclusion is present in Section.

2. Trinomial Difference Definition and Properties

A random vector (X, Y) has bivariate binomial distribution denoted by $BBin(n, \alpha, \beta, \gamma)$ if its p.m.f. is given by

$$f(x, y) = \sum_{i=0}^{\min(x,y)} \frac{n!}{i!(x-i)!(y-i)!(n-x-y+i)!} \alpha^{x-i}\beta^{y-i}\gamma^i(1-\alpha-\beta-\gamma)^{n-x-y+i},$$

$$x, y = 0, 1, \dots, n,\ x+y \le n. \tag{1}$$

The corresponding characteristic function (c.f.) is

$$Q_{x,y}(t_1,t_2) = [1-\alpha-\beta-\gamma+\alpha e^{it_1}+\beta e^{it_2}+\gamma e^{i(t_1+t_2)}]^n. \tag{2}$$

Two important cases arise from the bivariate binomial distribution

(i) The independence case:

In this case the two random variables are independent such that $X \sim Bin(n,p)$ and $Y \sim Bin(n,q)$. The c.f. of the bivariate binomial distribution is

$$\begin{aligned} Q_{x,y}(t_1,t_2) = [1-p(1-q)-q(1-p)-pq+p(1-q)e^{it_1} \\ +q(1-p)e^{it_2}+pqe^{i(t_1+t_2)}]^n \end{aligned} \tag{3}$$

Hence in the independent case, $pq=\gamma, p(1-q)=\alpha$ and $q(1-p)=\beta$.

(ii) The trinomial case:

In this case the random variables X and Y follow the trinomial distribution. The c.f. is

$$Q_{x,y}(t_1,t_2) = [1-\alpha-\beta+\alpha e^{it_1}+\beta e^{it_2}]^n \ where\ \gamma=0. \tag{4}$$

Let (X,Y) have the bivariate binomial distribution (1) and let Z be the difference random variable i.e. $Z=X-Y$ then from (2) the c.f. of Z is

$$Q_z(t) = [1-\alpha-\beta+\alpha e^{it}+\beta e^{-it}]^n \tag{5}$$

Note that the c.f. does not depend on γ and is identical with the one resulting from differencing X and Y in the trinomial case (4). The c.f. of the binomial difference distribution $Z^*=X-Y$ that arise from the independent case (3) is

$$Q_{Z^*}(t) = [1-p(1-q)-q(1-p)+p(1-q)e^{it}+q(1-p)e^{-it}]^n$$

To distinguish the general definition of difference of dependent binomial variable from the independence special case we give the following definition.

Definition 2. We say that the random variable Z has trinomial difference distribution with parameters $n \in \mathbb{Z},\ 0 \le \alpha,\ \beta \le 1$ and $\alpha+\beta<1$ denoted by $Z \sim TD(n,\alpha,\beta)$ if its c.f. is given by (5). The probability function of the trinomial distribution is given by

$$f(x,y) = \binom{n}{x,y}\alpha^x\beta^y(1-\alpha-\beta)^{n-x-y},\ x,y=0,1,\dots,n;$$

$$x+y \le n \quad \text{where} \quad \binom{n}{x,y} = \frac{n!}{x!y!(n-x-y)!}.$$

The probability function of the trinomial difference is present as:

$$g(z) = \sum_{y=\max(0,-z)}^{\left[\frac{n-z}{2}\right]} \binom{n}{z+y,y}\alpha^{z+y}\beta^y(1-\alpha-\beta)^{n-z-2y},\ z=0,\pm1,\pm2,\dots,\pm n. \tag{6}$$

***Note* 1.** If X and Y follow trinomial distribution and we observed $X+W$ and $Y+W$ with W being any random variable, the definition of trinomial difference is not affected and W can be viewed as background effect.

The above definition implies that the difference between the two independent binomial random variables with the same number of trials can be thought of as a difference between the components of trinomial distribution. The reverse of this conclusion is not true, as can be seen from the following proposition.

Proposition 1. *A $TD(n,\alpha,\beta)$ can be written as binomial difference if and only if (iff)*

$$(1-\alpha-\beta)^2-4\alpha\beta \geq 0. \tag{7}$$

***Proof*.** Note that $TD(n,\alpha,\beta)$ can be written as the difference between two independent (n,p) and (n,q) binomials iff

$$1-\alpha-\beta+\alpha e^{it}+\beta e^{-it}=pq+\overline{p}\;\overline{q}+p\;\overline{q}\;e^{it}+q\;\overline{p}\;e^{-it}$$

where, $\overline{p}=1-p$ and $\overline{q}=1-q$. This holds iff $\alpha=p(1-q)$ and $\beta=q(1-p)$. Solving these in equations for p and q in terms of α and β, we get

$$p=\frac{1+\alpha-\beta\pm\sqrt{(1-\alpha-\beta)^2-4\alpha\beta}}{2}$$

and

$$q=\frac{1-\alpha+\beta\pm\sqrt{(1-\alpha-\beta)^2-4\alpha\beta}}{2}.$$

Thus a real p and q can be achieved iff $(1-\alpha-\beta)^2-4\alpha\beta\geq 0$ which is equivalent to (7).

To distinguish the case when (7) holds we calling the corresponding distribution binomial difference distribution $(BD(n,p,q))$ as in definition 1. □

Proposition 2. *Let $X\sim TD(n,\alpha,\beta)$ then $E(X)=n(\alpha-\beta)$, $Var(X)=n[\alpha+\beta-(\alpha-\beta)^2]$, and $\gamma_1=\frac{n(\alpha-\beta)[2(\alpha-\beta)^2-3(\alpha+\beta)+1]}{[n(\alpha+\beta-(\alpha-\beta)^2]^{\frac{3}{2}}}$ where γ_1 is Pearson's coefficient of skewness.*

The trinomial distribution is symmetric when $\alpha=\beta$. This distribution can be positively skewed or negatively skewed. Figure 1 exhibits the probability function of $TD(n,\alpha,\beta)$ for different values of $n,\ \alpha$ and β.

Proposition 3. *Let $X_i, i=1,\ldots,n$ be independent random variables having $TD(n_i,\alpha,\beta)$ for $i=1,\ldots,n$, then $Z=\sum_{i=1}^n X_i\sim TD(\sum_{i=1}^n n_i,\alpha,\beta)$.*

Proposition 4. *If $X\sim TD(n,\alpha,\beta)$ then $Y=-X\sim TD(n,\beta,\alpha)$.*

Proposition 5. *For a fixed positive λ_1,λ_2 and $n>\lambda_1,\lambda_2$, consider a sequence of random variables $X_n\sim TD(n,\lambda_1/n,\lambda_2/n)$, the limiting distribution of X_n has Poisson difference distribution.*

Proof.

$$Q_{x_n}(t) = (1-(\lambda_1/n)-(\lambda_2/n)+(\lambda_1/n)e^{it}+(\lambda_2/n)e^{-it})^n, \ (\lambda_1+\lambda_2)/n \ \leq \ 1.$$

$$\lim_{n\to\infty} Q_{x_n}(t) = \lim_{n\to\infty}(1+\frac{\lambda_1(e^{it}-1)+\lambda_2(e^{-it}-1)}{n})^n = e^{\lambda_1(e^{it}-1)+\lambda_2(e^{-it}-1)}.$$

This is the Characteristic function of $PD(\lambda_1, \lambda_2)$. □

Alzaid & Omair (2012) introduced the extended binomial distribution as for any given two independent Poisson difference distributions the conditional distribution of one given the sum using a special constraint in order to obtain a linear mean in the parameter.

Definition 3. A random variable X in $\mathbb{Z}$ has extended binomial distribution with parameters $0 < p < 1, (q = 1-p, \theta > 0)$ and $z \in \mathbb{Z}$, denoted by $X \sim EB(z, p, \theta)$ if

$$P(X=x) = \frac{p^x q^{z-x} \ {}_0\widetilde{F}_1(;x+1;p^2\theta){}_0\widetilde{F}_1(;z-x+1;q^2\theta)}{{}_0\widetilde{F}_1(;z+1;\theta)}, \ x = \ldots, -1, 0, 1, \ldots$$

where ${}_0\widetilde{F}_1$ is the regularized hypergeometric function defined by

$${}_0\widetilde{F}_1(;b;z) = \sum_{k=0}^{\infty} \frac{z^k}{k!\Gamma(b+k)}.$$

Bakouch et al. (2013) extended the Poisson distribution over the set of all integers as follows.

Definition 4. A random variable X in $\mathbb{Z}$ has extended Poisson distribution with parameters $0 \leq p \leq 1$ and $\lambda > 0$, denoted by $X \sim E - P_0(p, \lambda)$ if its probability mass function is defined as

$$P(X=x) = \begin{cases} e^{-\lambda} & x = 0 \\ pe^{-\lambda} \ \frac{\lambda^x}{x!} & x = 1, 2, \ldots \\ (1-p)e^{-\lambda} \ \frac{\lambda^{|x|}}{|x|!} & x = -1, -2, \ldots \end{cases}$$

***Note* 2.**

1. The $BD(n, p, q)$ is a convolution of two independent strongly unimodal random variables which implies that it is strongly unimodal.

2. When $(1-\alpha-\beta)^2 - 4\alpha\beta < 0$, the trinomial difference distribution is no a longer binomial difference distribution and it may not be unimodal as illustrated in Figure 1. However the Poisson difference distribution is strongly unimodal and the extended binomial distribution has no proof of unimodality. In all the parameters explored till now it appears to be unimodal.

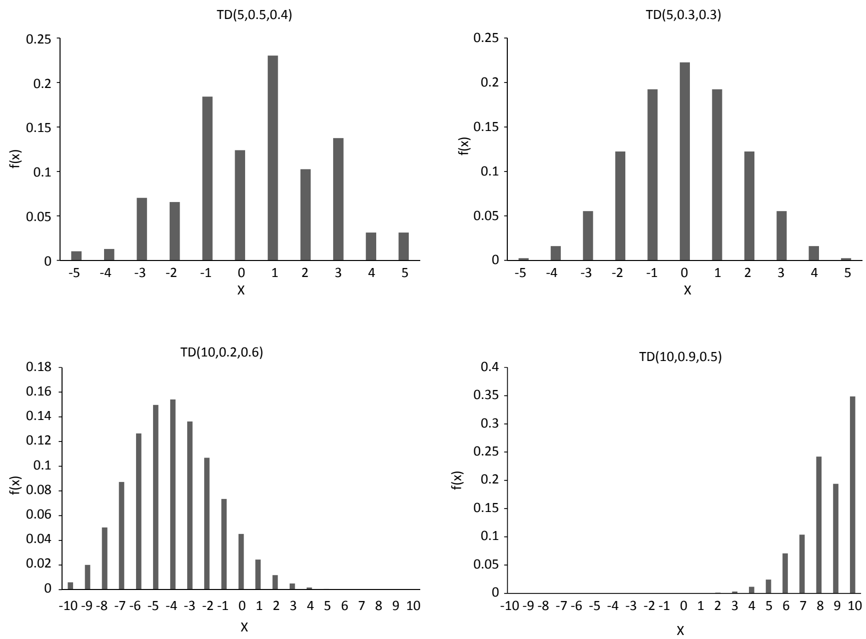

FIGURE 1: The probability mass function of $TD(n, \alpha, \beta)$ for different values of the parameters.

3. Unlike the Poisson difference distribution, which is always overdispersed, the trinomial difference distribution can be overdispersed or equidispersed or underdispersed, depending on the values of α and β as follows:

   If $(\alpha - \beta)^2 = 2\beta$ or 2α then the distribution is equidispersed.

   If$(\alpha - \beta)^2 < \min(2\alpha, 2\beta)$ then the distribution is overdispersed.

   Otherwise the distribution is underdipersed.

4. When $\beta = 0$ the $TD(n, \alpha, \beta) \equiv$ *binomial*(n, α).
5. When $\alpha = 0$ the $TD(n, \alpha, \beta) \equiv$ negation of *binomial*(n, β).

3. Inference

Let $X_1, \dots, X_m$ be independent identically distributed (i.i.d.) random variables (r.v.) from $TD(n, \alpha, \beta)$ the moment estimators of α and β are $\widehat{\alpha}_{mm} = \frac{nS^2 + \bar{X}^2 + n\bar{X}}{2n^2}$ and $\widehat{\beta}_{mm} = \widehat{\alpha} - \frac{\bar{X}}{n}$ where $\bar{X}$ and S^2 are the sample mean and the sample variance, respectively.

The Likelihood function is given by

$$L(\alpha, \beta; \mathbf{x}) = \prod_{i=1}^{m} \sum_{y=\max(0 \text{ instead of}; -x_i)}^{\left[\frac{n-x_i}{2}\right]} \binom{n}{x_i + y, y} \alpha^{x_i+y} \ \beta^y (1 - \alpha - \beta)^{n-x_i-2y}$$

The maximum likelihood estimators $\widehat{\alpha}_{mle}$ and $\widehat{\beta}_{mle}$ are obtained by solving the following two nonlinear equations

$$\frac{\partial \ln L}{\partial \alpha} = 0 \tag{8}$$

and

$$\frac{\partial \ln L}{\partial \beta} = 0. \tag{9}$$

After working the last two equation, it is found that

$$\widehat{\alpha}_{mle} - \widehat{\beta}_{mle} = \frac{\bar{x}}{n}. \tag{10}$$

Hence, to find the maximum likelihood estimators we used (10) and solved only one nonlinear equation.

We ran a simulation study for computing the estimates of the parameters of $TD(n, \alpha, \beta)$ using the method of moments and the maximum likelihood method. We simulated 1000 samples of size m from trinomial difference distribution, and we calculated the bias and used the mean square error as a measure of the estimates performance properties in all the methods of estimation that were considered, as follows:

1. $BIAS(\widehat{\alpha}) = \frac{1}{k}\sum_{j=1}^{k}(\widehat{\alpha}_j - \alpha),\ BIAS(\widehat{\beta}) = \frac{1}{k}\sum_{j=1}^{k}(\widehat{\beta}_j - \beta)$

2. $MSE(\widehat{\alpha}) = \frac{1}{k}\sum_{j=1}^{k}(\widehat{\alpha}_j - \alpha)^2$, $MSE(\widehat{\beta}) = \frac{1}{k}\sum_{j=1}^{k}(\widehat{\beta}_j - \beta)^2$, for $k = 1000$ runs.

We undertook these processes for different values of the parameters n, α and β and for different sample sizes m.

The following are the main findings:

1. The MSE of the maximum likelihood estimates is always smaller than the MSE of the moment estimates, except in few cases when $n = 1$.

2. For values of the parameter $n \geq 5$, the ML estimates are much better than the moment estimates in terms of MSE.

3. For sample sizes $m \geq 30$ the MSE of the moment estimates approach to the MSE of the ML estimates.

4. The ML estimates are always negatively biased except when $n = 1$. The moment estimates are positively biased.

5. In all cases except $n = 1$, the moment estimates have a smaller absolute bias than the ML estimates.

6. The moment and ML estimates are consistent.

Figures 2-3 show some results.

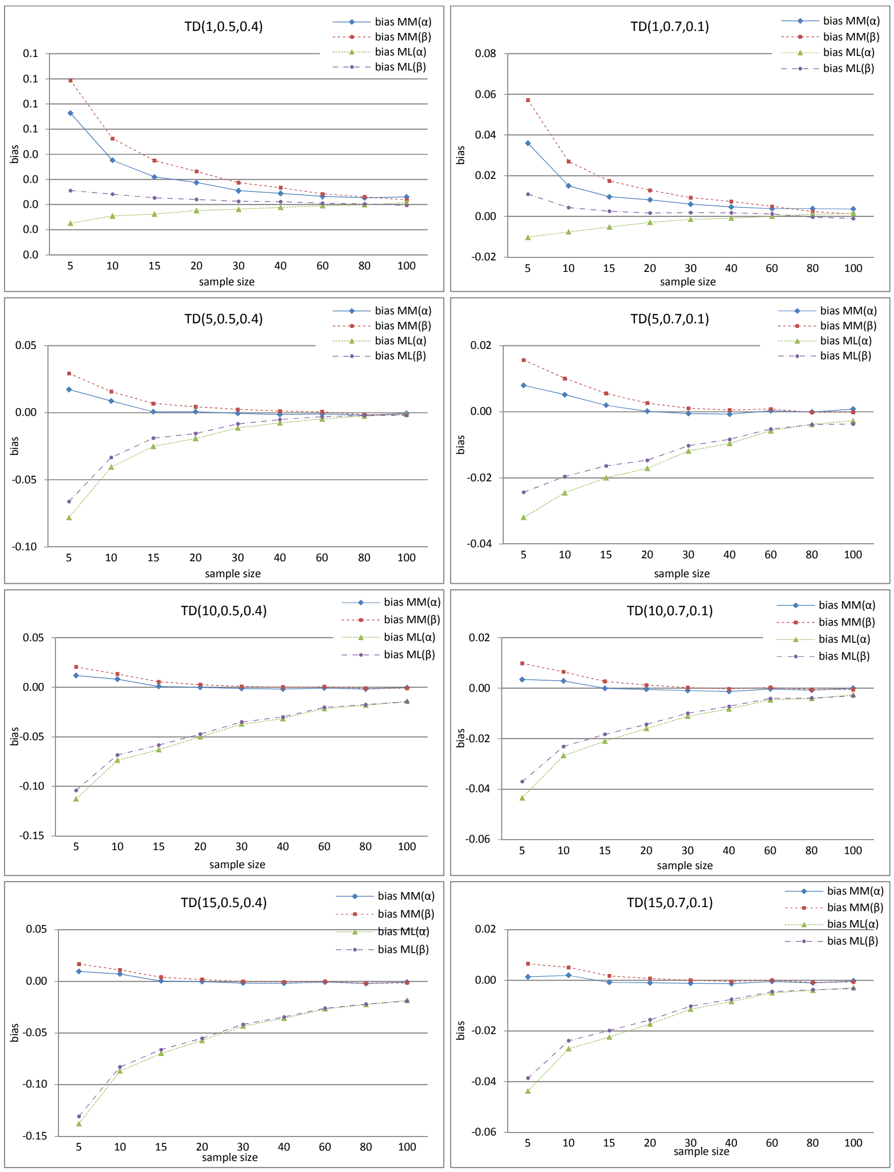

Figure 2: Plot of bias for moment and maximum likelihood estimators versus sample size.

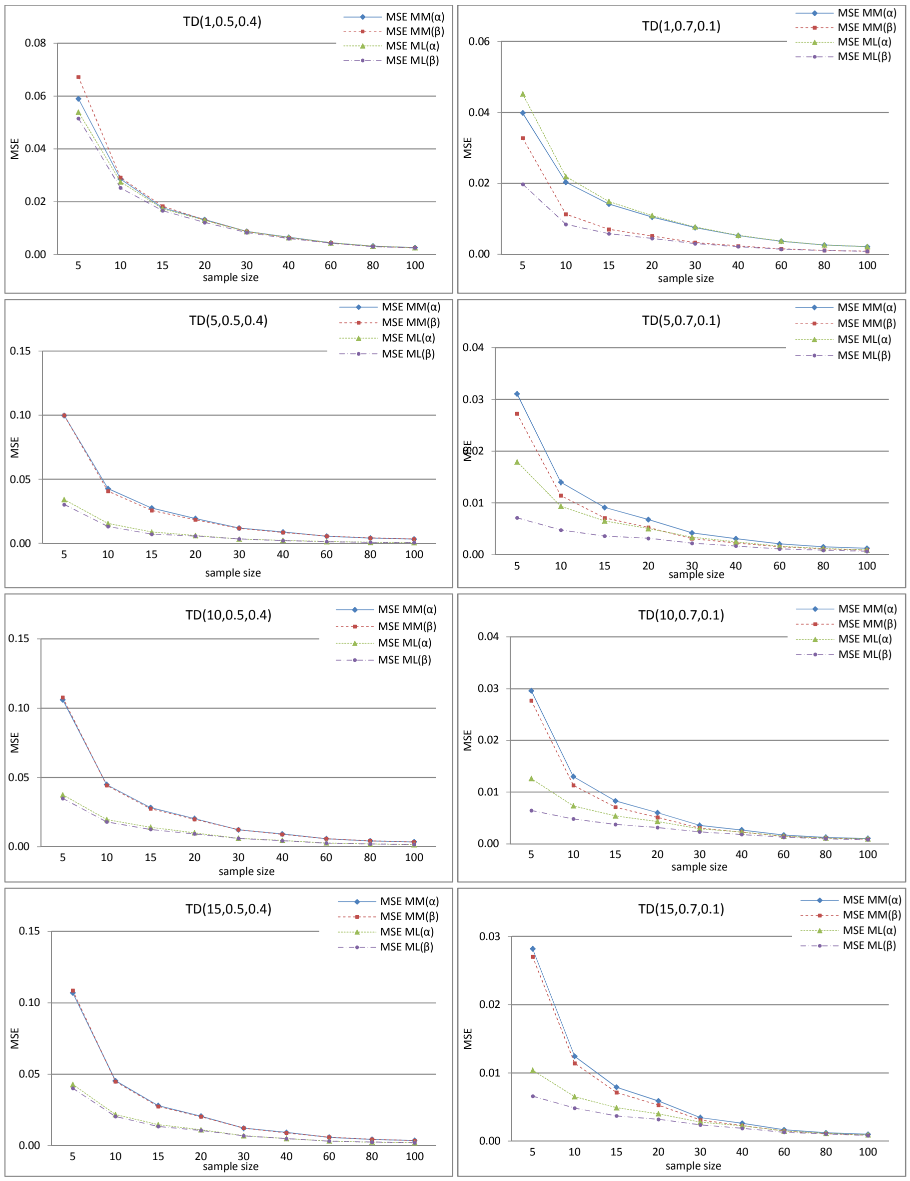

Figure 3: Plot of MSE for moment and maximum likelihood estimators versus sample size.

4. Applications

4.1. Motor Cycle Accidents

In order to examine the so-called underreporting of figures, since 1996 Statistics Denmark has conducted a study in which data on persons treated by casualty wards included. The results are published in the MOERKE table (`http://www.statbank.dk/statbank5a/default.asp?w=1920`). The data used below are the number of monthly motor cycle accidents involving persons under the influence of alcohol in Denmark from 1997 to 2008. The statistics only include injuries reported by the police. The number of accidents was found to be correlated and are affected by seasonality. Applying the runs test for motor cycle accidents, the $p-value = 0.000$, which implies that the original data is not independent. Taking a lag 12 difference of the data and applying the runs test, $p-value = 0.598$, which implies that the lag 12 difference is independent. This means that the dependence in the monthly data is due to seasonal additive effect and has been removed by seasonal differencing. The lag 12 difference correspond to the monthly difference in number of traffic accidents for a corresponding month between two consecutive years.

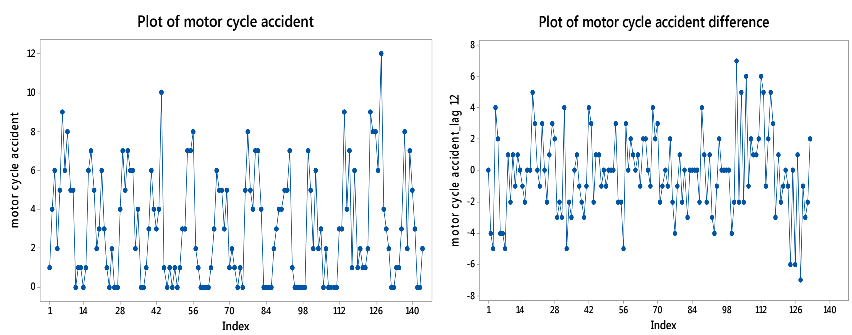

Figure 4: Time series plot of motor cycle accidents and differenced data.

Some descriptive statistics for the lag 12 difference of motor cycle data are illustrated in Table 1.

Table 1: Descriptive statistics of lag 12 difference of motor cycle data.

Variable	Size	Mean	Variance	Minimum	Maximum
Motor cycle	132	-0.152	6.9998	-7	7

The mean change $= -0.152$ indicates that the average monthly difference in the number of traffic accidents for any corresponding month between two consecutive years decreases slightly. Sine the data ranges from -7 to 7, the trinomial difference distribution with n greater than or equal to 7 is a candidate to model this data. The trinomial difference distribution is fitted assuming different values of n as well as the Poisson difference distribution, the extended Poisson and the extended binomial distribution assuming different values of the parameter z.

Figure 5 demonstrates the fitted distributions. The maximum likelihood estimates and the values of the Akaike information criterion AIC are illustrated in Table 2. The trinomial difference distribution with $n = 8$ has the smallest AIC. Note that in this case

$$(1 - \widehat{\alpha} - \widehat{\beta})^2 - 4\widehat{\alpha}\widehat{\beta} = -0.726 < 0$$

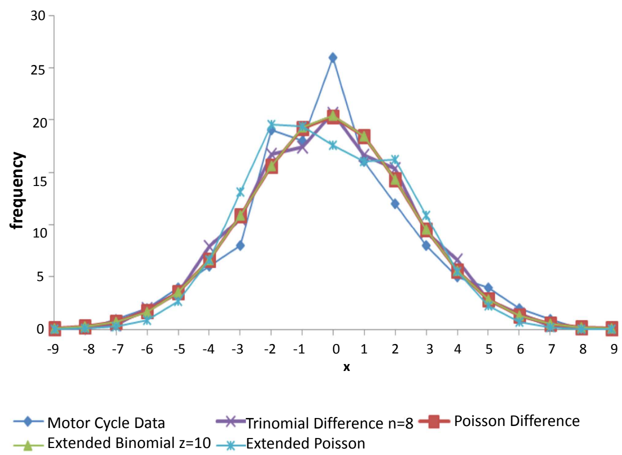

FIGURE 5: Motor cycle data and fitted distributions.

TABLE 2: Maximum likelihood estimates (standard errors) and AIC of the fitted distributions for motor cycle data.

Distribution	MLE	AIC
Trinomial Difference $n = 7$	$\widehat{\alpha} = 0.404\ (0.038)$, $\widehat{\beta} = 0.425(0.038)$	638.904
Trinomial Difference $n = 8$	$\widehat{\alpha} = 0.422(0.037)$, $\widehat{\beta} = 0.441(0.037)$	634.254
Trinomial Difference $n = 9$	$\widehat{\alpha} = 0.376(0.043)$, $\widehat{\beta} = 0.393(0.042)$	634.676
Trinomial Difference $n = 10$	$\widehat{\alpha} = 0.341(0.041)$, $\widehat{\beta} = 0.356(0.041)$	634.596
Extended Poisson	$\widehat{p} = 0.453(0.046)$, $\widehat{\lambda} = 2.015(0.124)$	642.796
Poisson Difference	$\widehat{\theta}_1 = 3.391(0.455)$, $\widehat{\theta}_2 = 3.543(0.457)$	634.332
Extended Binomial $z = -1$	$\widehat{p} = 0.145(0.231)$, $\widehat{\theta} = 794.384(2086.4)$	634.37
Extended Binomial $z = -2$	$\widehat{p} = 0.074(0.115)$, $\widehat{\theta} = 2579.58(7341.4)$	634.35
Extended Binomial $z = -4$	$\widehat{p} = 0.037(0.057)$, $\widehat{\theta} = 9289.86(273485)$	634.34
Extended Binomial $z = -10$	$\widehat{p} = 0.015(0.023)$, $\widehat{\theta} = 54538(163603)$	634.334
Extended Binomial $z = -20$	$\widehat{p} = 0.008(0.011)$, $\widehat{\theta} = 213676(644870)$	634.334

Hence, the data can be modeled by the TD distribution but not by the special case of binomial difference distribution. The Pearson Chi-square test is performed to test the fitting of the trinomial difference with $n = 8$. The test statistic $= 5.28 < \chi^2_{10,0.95} = 18.307$ indicates that the trinomial difference fits the data well.

4.2. Students Number at IDRAC International Management School

Bakouch et al. (2013) fitted the change in number of students between two sessions in a specific (test) group from the Bachelor program (first year) at IDRAC International Management school (Lyon, France) who had 60 sessions in marketing from the period, 1/9/2012 to 1/4/2013 with the extended Poisson distribution and other discrete distributions on $\mathbb{Z}$. In order to have more of understanding about the performance of the trinomial difference distribution we compared the fitting of this data using trinomial, extended Poisson, extended binomial and Poisson difference distributions. More information about the data can be found in Bakouch et al. (2013). Table 3 displays the descriptive statistics for the differenced data. Since the data ranges from -5 to 7 the trinomial difference distribution with $n \geq 7$ is a candidate to model this data. The trinomial difference distribution is fitted assuming different values of n as well as the Poisson difference distribution, the extended Poisson and the extended binomial distribution assuming different values of the parameter z. Figure 6 demonstrates the fitted distribution. The maximum likelihood estimates and the AIC values are illustrated in Table 4. The trinomial difference distribution with $n = 9$ has the smallest AIC. Note that in this case

$$(1 - \widehat{\alpha} - \widehat{\beta})^2 - 4\widehat{\alpha}\widehat{\beta} = -0.8036 < 0$$

Hence, the data can be modeled by the TD distribution but not by the special case of binomial difference distribution. The Pearson Chi-square test is performed to test the fitting of the trinomial difference with $n = 9$. The test statistic$= 11.87 < \chi^2_{7,0.95} = 14.067$ indicates that the trinomial difference fits the data well.

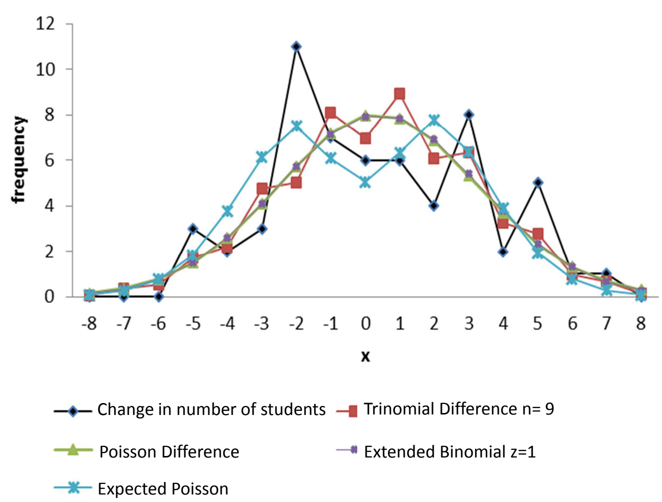

FIGURE 6: Change in number of students data and fitted distributions.

Table 3: Descriptive statistics of the change in number of students.

Variable	Size	Mean	Variance	Minimum	Maximum
difference	59	0.3898	8.828	-5	7

Table 4: Maximum likelihood estimates (standard errors) and AIC of the fitted distributions for change in number of students.

Distribution	MLE	AIC
Trinomial Difference $n = 7$	$\widehat{\alpha} = 0.475(0.053)$, $\widehat{\beta} = 0.419(0.053)$	299.913
Trinomial Difference $n = 8$	$\widehat{\alpha} = 0.44(0.059)$, $\widehat{\beta} = 0.391(0.058)$	301.376
Trinomial Difference $n = 9$	$\widehat{\alpha} = 0.473(0.056)$, $\widehat{\beta} = 0.43(0.056)$	296.96
Trinomial Difference $n = 10$	$\widehat{\alpha} = 0.423(0.060)$, $\widehat{\beta} = 0.384(0.060)$	298.612
Extended Poisson	$\widehat{p} = 0.509(0.068)$, $\widehat{\lambda} = 2.458(0.204)$	299.134
Poisson Difference	$\widehat{\theta}_1 = 4.617(2.251)$, $\widehat{\theta}_2 = 4.228(2.231)$	299.305
Extended Binomial $z = 1$	$\widehat{p} = 0.401(0.352)$, $\widehat{\theta} = 339.3(230.8)$	299.045
Extended Binomial $z = 2$	$\widehat{p} = 0.207(0.188)$, $\widehat{\theta} = 727.8(1013.3)$	299.127
Extended Binomial $z = 3$	$\widehat{p} = 0.137(0.128)$, $\widehat{\theta} = 1406(2262.3)$	299.177

5. Conclusions

Recently, discrete distributions on $\mathbb{Z}$ have attracted the attention of many researchers. In this paper, we defined the difference of two random variables following trinomial distributions. This distribution can be overdispersed or underdispersed, unlike the Poisson difference distribution that is always overdispersed. It is also symmetric or positively skewed or negatively skewed.

Through two real applications, the paper shows that the trinomial difference distribution is compatible with the Poisson difference distribution, the extended Poisson distribution and the extended binomial distribution. It also indicates that the binomial difference distribution failed to fit these data.

Acknowledgment

The authors extend their appreciation to the Dean of Scientific Research at King Saud University for funding the work through the research group project RGB-VPP-053. The authors are thankful to the referees for the several suggestions that helped improve the clarity of this paper.

[Received: June 2014 — Accepted: February 2015]

References

Alzaid, A. A. & Omair, M. A. (2010), 'On the poisson difference distribution inference and applications', *Bulletin of the Malaysian Mathematical Sciences Society* **8**(33), 17–45.

Alzaid, A. A. & Omair, M. A. (2012), 'An extended binomial distribution with applications', *Communications in Statistics-Theory and Methods* **41**(19), 3511–3527.

Alzaid, A. A. & Omair, M. A. (2014), 'Poisson difference integer valued autoregressive model of order one', *Bulletin of the Malaysian Mathematical Sciences Society* **2**(37), 465–485.

Bakouch, H., Kachour, M. & Nadarajah, S. (2013), An extended Poisson distribution. Working paper or preprint.
*https://hal.archives-ouvertes.fr/hal-00959426

Castro, G. (1952), 'Note on differences of bernoulli and poisson variables', *Portugaliae mathematica* **11**(4), 173–175.

Inusah, S. & Kozubowski, T. J. (2006), 'A discrete analogue of the laplace distribution', *Journal of statistical planning and inference* **136**(3), 1090–1102.

Karlis, D. & Ntzoufras, I. (2006), 'Bayesian analysis of the differences of count data', *Statistics in medicine* **25**(11), 1885–1905.

Karlis, D. & Ntzoufras, I. (2009), 'Bayesian modelling of football outcomes: using the skellam's distribution for the goal difference', *IMA Journal of Management Mathematics* **20**(2), 133–145.

Kemp, A. W. (1997), 'Characterizations of a discrete normal distribution', *Journal of Statistical Planning and Inference* **63**(2), 223–229.

Kotz, S., Johnson, N. & Kemp, A. (1992), *Univariate Discrete Distributions*, 2 edn, John Wiley & Sons, Inc., New York.

Ong, S., Shimizu, K. & Min Ng, C. (2008), 'A class of discrete distributions arising from difference of two random variables', *Computational Statistics & Data Analysis* **52**(3), 1490–1499.

Skellam, J. G. (1946), 'The frequency distribution of the difference between two poisson variates belonging to different populations', *Journal of the Royal Statistical Society. Series A* **109**(3), 296.