BULLETIN *of the*
Malaysian Mathematical
Sciences Society
http://math.usm.my/bulletin

Bull. Malays. Math. Sci. Soc. (2) **33***(1) (2010), 17–45*

On The Poisson Difference Distribution Inference and Applications

[1]Abdulhamid A. Alzaid and [2]Maha A. Omair
[1,2]Department of Statistics and Operations Research, College of Sciences,
King Saud University, BOX 2455 Riyadh 11451, Kingdom of Saudi Arabia
[1]alzaid@ksu.edu.sa, [2]maomair@ksu.edu.sa

Abstract. The distribution of the difference between two independent Poisson random variables involves the modified Bessel function of the first kind. Using properties of this function, maximum likelihood estimates of the parameters of the Poisson difference were derived. Asymptotic distribution property of the maximum likelihood estimates is discussed. Maximum likelihood estimates were compared with the moment estimates in a Monte Carlo study. Hypothesis testing using likelihood ratio tests was considered. Some new formulas concerning the modified Bessel function of the first kind were provided. Alternative formulas for the probability mass function of the Poisson difference (PD) distribution are introduced. Finally, two new applications for the PD distribution are presented. The first is from the Saudi stock exchange (TASI) and the second is from Dallah hospital.

2000 Mathematics Subject Classification: Primary 60E05; Secondary 62F86, 46N30

Key words and phrases: Poisson difference distribution, Skellam distribution, Bessel function, regularized hypergeometric function, maximum likelihood estimate, likelihood ratio test.

1. Introduction

The distribution of the difference between two independent Poisson random variables was derived by Irwin [5] for the case of equal parameters. Skellam [13] and Prekopa [11] discussed the case of unequal parameters. The distribution of the difference between two correlated Poisson random variables was recently introduced by Karlis and Ntzoufras [8] who proved that it reduces to the Skellam distribution (Poisson difference of two independent Poisson). Strakee and van der Gon [14] presented tables of the cumulative distribution function of the PD distribution to four decimal places for some combinations of values of the two parameters. Their tables also show the differences between the normal approximations (see [3]). Romani [12] showed that all the odd cummulants of the PD distribution ($\mathrm{PD}(\theta_1, \theta_2)$) equal to $\theta_1 - \theta_2$,

Communicated by Rosihan M. Ali, Dato'.
Received: December 4, 2008; *Revised:* July 3, 2009.

and that all the even cummulant equal to $\theta_1 + \theta_2$. He also discussed the properties of the maximum likelihood estimator of $E(X_1 - X_2) = \theta_1 - \theta_2$. Katti [10] studied $E|X_1 - X_2|$. Karlis and Ntzoufras [7] discussed in details properties of the PD distribution and obtained the maximum likelihood estimates via the Expectation Maximization (EM) algorithm. Karlis and Ntzoufras [8] derived Bayesian estimates and used the Bayesian approach for testing the equality of the two parameters of the PD distribution.

The PD distribution has many applications in different fields. Karlis and Ntzoufras [9] applied the PD distribution for modeling the difference of the number of goals in football games. Karlis and Ntzoufras [8] used the PD distribution and the zero inflated PD distribution to model the difference in the decayed, missing and filled teeth (DMFT) index before and after treatment. Hwang *et al.* [4] showed that the Skellam distribution can be used to measure the intensity difference of pixels in cameras. Strackee and van der Gon [14] state, "In a steady state the number of light quanta, emitted or absorbed in a definite time, is distributed according to a Poisson distribution. In view thereof, the physical limit of perceptible contrast in vision can be studied in terms of the difference between two independent variates each following a Poisson distribution". The distribution of differences may also be relevant when a physical effect is estimated as the difference between two counts, one when a "cause" is acting, and the other a "control" to estimate the "background effect". For more applications see Alvarez [2].

The aim of this paper is to obtain some inference results for the parameters of the PD distribution and give application on share and occupancy modeling. Maximum likelihood estimates of θ_1 and θ_2 are obtained by maximizing the likelihood function (or equivalently the log likelihood), using the properties of the modified Bessel function of the first kind. A Monte Carlo study is conducted to compare two estimation methods, the method of moment and the maximum likelihood. Moreover, since regularity conditions hold, asymptotic distribution of the maximum likelihood estimates is obtained. Moreover, hypothesis testing using Likelihood ratio test for equality of the two parameters is introduced and Monte Carlo study is presented with the empirical power being calculated. For simplification alternative formulas of the PD distribution are presented for which Poisson distribution and negative of Poisson distribution can be shown by direct substitution to be special cases of the PD distribution. These formulas are used for estimation and testing. The applications considered in this study are such that only the difference of two variables could be estimated while each one by its own is not easily estimated. Our considered data could take both positive and negative integer values. Hence, PD distribution could be a good candidate for such data. The first is from the Saudi stock exchange (TASI) and the second from Dallah hospital at Riyadh.

The remainder of this paper proceeds as follows: Properties of the PD distribution are revised with some properties of the modified Bessel function of the first kind and new formulas for the Bessel function are derived in Section 2. In Section 3, new representation of the PD distribution is presented. Maximum likelihood estimates are considered in details with their asymptotic properties in Section 4. In Section 5, likelihood ratio tests for equality of means and for testing if one of the parameters has zero value are presented. A simulation study is conducted in Section 6. Finally, two new applications of the PD distribution are illustrated in Section 7.

2. Definition and basic properties

Definition 2.1. *For any pair of variables (X, Y) that can be written as $X = W_1 + W_3$ and $Y = W_2 + W_3$ with $W_1 \sim Poisson\,(\theta_1)$ independent of $W_2 \sim Poisson\,(\theta_2)$ and W_3 following any distribution, the probability mass function of $Z = X - Y$ is given by*

$$P(Z = z) = e^{-\theta_1 - \theta_2} \left(\frac{\theta_1}{\theta_2}\right)^{z/2} I_z\left(2\sqrt{\theta_1\theta_2}\right), \quad z = \cdots, -1, 0, 1, \cdots \tag{2.1}$$

where

$$I_y(x) = \left(\frac{x}{2}\right)^y \sum_{k=0}^{\infty} \frac{\left(x^2/4\right)^k}{k!\,(y+k)!}$$

is the modified Bessel function of the first kind and Z is said to have the PD distribution (Skellam distribution) denoted by $\mathrm{PD}(\theta_1, \theta_2)$*. See* [9].

An interesting property is a type of symmetry given by

$$P\left(Z = z \mid \theta_1, \theta_2\right) = P\left(Z = -z \mid \theta_2, \theta_1\right).$$

The moment generating function is given by

$$M_Z(t) = \exp\left[-\left(\theta_1 + \theta_2\right) + \theta_1 e^t + \theta_2 e^{-t}\right]. \tag{2.2}$$

The expected value is $E(Z) = \theta_1 - \theta_2$, while the variance is $V(Z) = \theta_1 + \theta_2$. The odd cummulants are equal to $\theta_1 - \theta_2$ while the even cummulants are equal to $\theta_1 + \theta_2$. The skewness coefficient is given by $\beta_1 = (\theta_1 - \theta_2)/(\theta_1 + \theta_2)^{3/2}$, that is the distribution is positively skewed when $\theta_1 > \theta_2$, negatively skewed when $\theta_1 < \theta_2$ and symmetric when $\theta_1 = \theta_2$. The kurtosis coefficient $\beta_2 = 3 + 1/(\theta_1 + \theta_2)$. As either θ_1 or θ_2 tends to infinity kurtosis coefficient tends to 3 and for a constant difference $\theta_1 - \theta_2$, skewness coefficient tends to zero implying that the distribution approaches the normal distribution. The PD distribution is strongly unimodal.

If $Y_1 \sim \mathrm{PD}(\theta_1, \theta_2)$ independent of $Y_2 \sim \mathrm{PD}(\theta_3, \theta_4)$ then

(1) $Y_1 + Y_2 \sim \mathrm{PD}(\theta_1 + \theta_3,\ \theta_2 + \theta_4)$

(2) $Y_1 - Y_2 \sim \mathrm{PD}(\theta_1 + \theta_4,\ \theta_2 + \theta_3)$.

More properties of the PD distribution can be found in [8].

The following are some known identities for the modified Bessel function of the first kind (see [1]) :

For any $\theta > 0$ and $y \in Z$,

$$I_y(\theta) = I_{-y}(\theta), \tag{2.3}$$

$$\sum_{y=-\infty}^{\infty} I_y(\theta) = e^{\theta}, \tag{2.4}$$

$$\sum_{y=-\infty}^{\infty} y I_y(\theta) = 0, \tag{2.5}$$

$$I_y(\theta) = \left(\frac{\theta}{2}\right)^y {}_0\tilde{F}_1\left(y+1, \frac{\theta^2}{4}\right), \tag{2.6}$$

where

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

is the regularized hypergeometric function and $\Gamma(x)$ *is the gamma function*

$$\frac{\partial I_y(\theta)}{\partial \theta} = \frac{y}{\theta} I_y(\theta) + I_{y+1}(\theta), \tag{2.7}$$

$$I_y(\theta) = \frac{2(y+1)}{\theta} I_{y+1}(\theta) + I_{y+2}(\theta). \tag{2.8}$$

In the following proposition, other relations for the Bessel function which can be easily driven from (2.1) and (2.2) are presented.

Proposition 2.1. *For any* $\theta > 0$, $\theta_1 > 0$ *and* $\theta_2 > 0$ *then*

$$\sum_{y=-\infty}^{\infty} \left(\frac{\theta_1}{\theta_2}\right)^{y/2} I_y\left(2\sqrt{\theta_1\theta_2}\right) = e^{\theta_1+\theta_2}, \tag{2.9}$$

$$\sum_{y=-\infty}^{\infty} y\left(\frac{\theta_1}{\theta_2}\right)^{y/2} I_y\left(2\sqrt{\theta_1\theta_2}\right) = (\theta_1-\theta_2)\, e^{\theta_1+\theta_2}, \tag{2.10}$$

$$\sum_{y=-\infty}^{\infty} y^2\left(\frac{\theta_1}{\theta_2}\right)^{y/2} I_y\left(2\sqrt{\theta_1\theta_2}\right) = \left(\theta_1+\theta_2+(\theta_1-\theta_2)^2\right) e^{\theta_1+\theta_2}, \tag{2.11}$$

$$\sum_{y=-\infty}^{\infty} y^2 I_y(\theta) = \theta e^{\theta} \text{ for any } \theta > 0, \tag{2.12}$$

$$\sum_{y=-\infty}^{\infty} y^4 I_y(\theta) = \theta e^{\theta}(3\theta+1) \text{ for any } \theta > 0. \tag{2.13}$$

Proof. (2.9) is obtained from the fact that (2.1) is a probability mass function. (2.10) and (2.11) follow from the mean and the variance representations. (2.12) is a special case of (2.11) by setting $\theta_1 = \theta_2 = \theta/2$. (2.13) follows from the fact that the fourth cummulant $K_4 = \mu_4 - 3\mu_2^2$.

If $Y \sim \mathrm{PD}\left(\frac{\theta}{2}, \frac{\theta}{2}\right)$, then the fourth cummulant is θ and $\mu_2 = \theta$. Hence

$$E\left(Y^4\right) = \theta + 3\theta^2$$

and

$$\sum_{y=-\infty}^{\infty} y^4 I_y(\theta) = \theta e^{\theta}(3\theta+1)$$

for any $\theta > 0$. ∎

3. New representation of the Poisson difference distribution

The regularized hypergeometric function ${}_0\tilde{F}_1$ is linked with the modified Bessel function of the first kind through the identity given by equation (2.6). It has the property

$$ {}_0\tilde{F}_1\left(;-y+1;\theta\right)=\theta^{y}{}_0\tilde{F}_1\left(;y+1;\theta\right). \tag{3.1}$$

Using (2.6) and (3.1) the PD distribution can be expressed using any of the following equivalent formulas:

Formula I

$$P\left(Y=y\right)=e^{-\theta_1-\theta_2}\left(\frac{\theta_1}{\theta_2}\right)^{y/2}I_y\left(2\sqrt{\theta_1\theta_2}\right),\quad y=\cdots,-1,0,1,\cdots$$

Formula II

$$P\left(Y=y\right)=e^{-\theta_1-\theta_2}\theta_1^{y}\ {}_0\tilde{F}_1\left(y+1,\theta_1\theta_2\right),\quad y=\cdots,-1,0,1,\cdots$$

Formula III

$$P\left(Y=y\right)=e^{-\theta_1-\theta_2}\theta_2^{-y}\ {}_0\tilde{F}_1\left(-y+1,\theta_1\theta_2\right),\quad y=\cdots,-1,0,1,\cdots$$

Formula IV

$$P\left(Y=y\right)=e^{-\theta_1-\theta_2}\left(\theta_1\theta_2\right)^{\max\{0,-y\}}\theta_1^{y}\ {}_0\tilde{F}_1\left(|y|+1,\theta_1\theta_2\right),\quad y=\cdots,-1,0,1,\cdots.$$

The advantages of the new formulas are:

(1) Easier and more direct notation. Following the steps of deriving the PD distribution, it is more logical to use the regularized hypergeometric function instead of the Bessel function as follows.
Let X_1 ~Poisson(θ_1) be independent of X_2 ~Poisson(θ_2) then $Y=X_1-X_2$ ~PD(θ_1, θ_2).

$$\begin{aligned}P\left(Y=y\right)=P\left(X_1-X_2=y\right)&=\sum_{k=0}^{\infty}P\left(X_1-X_2=y\,|X_2=k\right)P(X_2=k)\\&=\sum_{k=\max(-y,0)}^{\infty}P\left(X_1=y+k\right)P\left(X_2=k\right)\\&=e^{-\theta_1-\theta_2}\theta_1^{y}\sum_{k=\max(-y,0)}^{\infty}\frac{\left(\theta_1\theta_2\right)^k}{k!\,(y+k)!}\end{aligned}$$

$$P\left(Y=y\right)=e^{-\theta_1-\theta_2}\theta_1^{y}\sum_{k=0}^{\infty}\frac{\left(\theta_1\theta_2\right)^k}{k!\,(y+k)!},\quad y=\cdots,-1,0,1,\cdots$$

with the convention that any term with negative factorial in the denominator is zero.

$$P\left(Y=y\right)=e^{-\theta_1-\theta_2}\theta_1^{y}{}_0\tilde{F}_1\left(;y+1;\theta_1\theta_2\right),\quad y=\cdots,-1,0,1,\cdots$$

(2) The special case when $\theta_2=0$ can be considered directly using Formula II to get the Poisson difference(θ_1,0)≡Poisson(θ_1). Let Y ~PD(θ_1,θ_2), and assume that $\theta_2=0$ then

$$P(Y=y)=e^{-\theta_1}\theta_1^{y}{}_0\tilde{F}_1\left(;y+1;0\right),\quad y=\cdots,-1,0,1,\cdots$$

$$= \begin{cases} \frac{e^{-\theta_1}\theta_1^y}{y!}, & y = 0, 1, 2, \cdots \\ 0, & \text{otherwise,} \end{cases}$$

since

$$_0\tilde{F}_1\left(;y+1;0\right) = \begin{cases} \frac{1}{y!}, & y = 0, 1, 2, \cdots \\ 0, & \text{otherwise.} \end{cases}$$

This special case is not applicable when using the notation with the modified Bessel function of the first kind since θ_2 appears in the denominator.

(3) The special case when $\theta_1 = 0$ can be considered directly using Formula III to get the Poisson difference $(0, \theta_2) \equiv$'negative' Poisson(θ_2). Let $Y \sim$PD(θ_1,θ_2), and assume that $\theta_1 = 0$ then

$$P(Y = y) = e^{-\theta_2}\theta_2^{-y}{}_0\tilde{F}_1\left(;-y+1;0\right), \quad y = \cdots, -1, 0, 1, \cdots$$
$$= \begin{cases} \frac{e^{-\theta_2}\theta_2^{-y}}{(-y)!}, & y = 0, -1, -2, \cdots \\ 0, & \text{otherwise,} \end{cases}$$

since

$$_0\tilde{F}_1\left(;-y+1;0\right) = \begin{cases} \frac{1}{(-y)!}, & y = 0, -1, -2, \cdots \\ 0, & \text{otherwise.} \end{cases}$$

This special case is not applicable using the notation with the modified Bessel function of the first kind since θ_1 appears in the numerator and a direct substitution will yield zero.

(4) A more general formula for the probability mass function of the PD distribution for which Formula II and III are special cases is as follows:
Formula IV

$$P\left(Y = y\right) = e^{-\theta_1-\theta_2}\left(\theta_1\theta_2\right)^{\max\{0,-y\}}\theta_1^y{}_0\tilde{F}_1\left(;|y|+1;\theta_1\theta_2\right),$$

for $y = \cdots, -1, 0, 1, \cdots$.

4. Estimation

The PD distribution had been introduced more than 70 years ago. Till now only moment estimates are used in the literature and recently maximum likelihood estimates via EM algorithm were obtained by Karlis and Ntzoufras [7] avoiding to maximize the likelihood directly. Karlis and Ntzoufras [8] derived also Bayesian estimates and used the Bayesian approach for testing the equality of the two parameters of the PD distribution.

In this section, we focus on the estimation of the parameters θ_1 and θ_2 of the PD distribution. The maximum likelihood estimates are presented and are compared with the moment estimates via a Monte Carlo study. Asymptotic properties of the maximum likelihood estimates are exploited and confidence interval for each parameter is obtained for the first time. Likelihood ratio test for testing the equality of the two parameters is introduced.

4.1. The method of moments

Let $Z_1, Z_2, \cdots, Z_n$ be i.i.d. PD (θ_1, θ_2), then

$$\hat{\theta}_{1MM} = \frac{1}{2}\left(S^2 + \bar{Z}\right) \tag{4.1}$$

and

$$\hat{\theta}_{2MM} = \frac{1}{2}\left(S^2 - \bar{Z}\right), \tag{4.2}$$

where $\bar{Z}$ is the sample mean and S^2 is the sample variance. The moment estimators are unbiased estimators. The moment estimates do not exist if $S^2 - \left|\bar{Z}\right| < 0$ since in this case we would obtain negative estimates of θ_1 or θ_2 [7]. That is, moment estimates do not exist when the sample variance is less than the absolute value of the sample mean. In simulated samples or real data, cases like this happen usually when one of the parameters is very small compared to the other i.e. $\theta_i/\theta_j \geq 10$ for $i, j = 1, 2$. To solve this problem, a modification is done such that the negative estimate is set to zero since zero is the smallest possible value and the other estimate is set to equal the absolute value of the mean.

4.2. Maximum likelihood estimation

Let $Z_1, Z_2, \cdots, Z_n$ be i.i.d. PD (θ_1, θ_2). The likelihood function is given by

$$L = \prod_{i=1}^{n} P\left(Z_i = z_i\right) = \prod_{i=1}^{n}\left[e^{-\theta_1-\theta_2}\left(\frac{\theta_1}{\theta_2}\right)^{z_i/2} I_{z_i}\left(2\sqrt{\theta_1\theta_2}\right)\right].$$

Using the differentiation formula for the modified Bessel function we differentiate the log-likelihood with respect to θ_1 and θ_2 as follows

$$\begin{aligned}
\frac{\partial \ln L}{\partial \theta_1} &= -n + \frac{\sum_{i=1}^{n} z_i}{2\theta_1} + \frac{\theta_2}{\sqrt{\theta_1\theta_2}} \sum_{i=1}^{n} \frac{\frac{z_i}{2\sqrt{\theta_1\theta_2}} I_{z_i}\left(2\sqrt{\theta_1\theta_2}\right) + I_{z_i+1}\left(2\sqrt{\theta_1\theta_2}\right)}{I_{z_i}\left(2\sqrt{\theta_1\theta_2}\right)} \\
&= -n + \frac{\sum_{i=1}^{n} z_i}{\theta_1} + \frac{\theta_2}{\sqrt{\theta_1\theta_2}} \sum_{i=1}^{n} \frac{I_{z_i+1}\left(2\sqrt{\theta_1\theta_2}\right)}{I_{z_i}\left(2\sqrt{\theta_1\theta_2}\right)} \qquad (4.3)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln L}{\partial \theta_2} &= -n - \frac{\sum_{i=1}^{n} z_i}{2\theta_2} + \frac{\theta_1}{\sqrt{\theta_1\theta_2}} \sum_{i=1}^{n} \frac{\frac{z_i}{2\sqrt{\theta_1\theta_2}} I_{z_i}\left(2\sqrt{\theta_1\theta_2}\right) + I_{z_i+1}\left(2\sqrt{\theta_1\theta_2}\right)}{I_{z_i}\left(2\sqrt{\theta_1\theta_2}\right)} \\
&= -n + \frac{\theta_1}{\sqrt{\theta_1\theta_2}} \sum_{i=1}^{n} \frac{I_{z_i+1}\left(2\sqrt{\theta_1\theta_2}\right)}{I_{z_i}\left(2\sqrt{\theta_1\theta_2}\right)}. \qquad (4.4)
\end{aligned}$$

The maximum likelihood estimators $\hat{\theta}_{1MLE}$ and $\hat{\theta}_{2MLE}$ are obtained by setting (4.3) and (4.4) to zero and solving the two nonlinear equations

$$0 = -n + \frac{\sum_{i=1}^{n} z_i}{\hat{\theta}_{1MLE}} + \frac{\hat{\theta}_{2MLE}}{\sqrt{\hat{\theta}_{1MLE}\hat{\theta}_{2MLE}}} \sum_{i=1}^{n} \frac{I_{z_i+1}\left(2\sqrt{\hat{\theta}_{1MLE}\hat{\theta}_{2MLE}}\right)}{I_{z_i}\left(2\sqrt{\hat{\theta}_{1MLE}\hat{\theta}_{2MLE}}\right)} \tag{4.5}$$

and

$$0 = -n + \frac{\hat{\theta}_{1MLE}}{\sqrt{\hat{\theta}_{1MLE}\hat{\theta}_{2MLE}}} \sum_{i=1}^{n} \frac{I_{z_i+1}\left(2\sqrt{\hat{\theta}_{1MLE}\hat{\theta}_{2MLE}}\right)}{I_{z_i}\left(2\sqrt{\hat{\theta}_{1MLE}\hat{\theta}_{2MLE}}\right)}. \tag{4.6}$$

Note that multiplying equation (4.5) by $\hat{\theta}_{1MLE}$ and equation (4.6) by $\hat{\theta}_{2MLE}$ and subtracting them we get

$$-n\hat{\theta}_{1MLE} - n\hat{\theta}_{2MLE} + \sum_{i=1}^{n} z_i = 0,$$

$$\hat{\theta}_{1MLE} = \hat{\theta}_{2MLE} + \overline{z}. \tag{4.7}$$

Now, substituting equation (4.7) into equation (4.6) we obtain

$$0 = -n + \frac{\left(\hat{\theta}_{2MLE} + \overline{z}\right)}{\sqrt{\left(\hat{\theta}_{2MLE} + \overline{z}\right)\hat{\theta}_{2MLE}}} \sum_{i=1}^{n} \frac{I_{z_i+1}\left(2\sqrt{\left(\hat{\theta}_{2MLE} + \overline{z}\right)\hat{\theta}_{2MLE}}\right)}{I_{z_i}\left(2\sqrt{\left(\hat{\theta}_{2MLE} + \overline{z}\right)\hat{\theta}_{2MLE}}\right)}. \tag{4.8}$$

Hence, we can find $\hat{\theta}_{2MLE}$ by solving the nonlinear equation (4.8) and then find $\hat{\theta}_{1MLE}$ using equation (4.7).

Using the identity, $(\partial {}_0\tilde{F}_1(;x+1;\theta))/\partial\theta = {}_0\tilde{F}_1(;x+2;\theta)$, maximum likelihood estimates could also be obtained using Formulas II and III.

Using Formula II, one can find $\hat{\theta}_{2MLE}$ by solving the nonlinear equation

$$0 = -n + \left(\hat{\theta}_{2MLE} + \overline{X}\right) \sum_{i=1}^{n} \frac{{}_0\tilde{F}_1\left(;x_i+2;\left(\hat{\theta}_{2MLE} + \overline{X}\right)\hat{\theta}_{2MLE}\right)}{{}_0\tilde{F}_1\left(;x_i+1;\left(\hat{\theta}_{2MLE} + \overline{X}\right)\hat{\theta}_{2MLE}\right)} \tag{4.9}$$

and

$$\hat{\theta}_{1MLE} = \hat{\theta}_{2MLE} + \overline{X}. \tag{4.10}$$

Remark 4.1. All three Formulas (I, II and III) gave identical maximum likelihood estimates when the relative difference between θ_1 and θ_2 was not large (less than 10) when solving the nonlinear equation. But when $\theta_1 > 10\theta_2$, the nonlinear equations using Formulas I or III were not as fast to converge as using Formula II and were more willing to obtain negative estimate for θ_2 than Formula II. On the other side, for $\theta_2 > 10\theta_1$, the nonlinear equations using Formulas I or II were not as fast to converge as using Formula III and were more willing to obtain negative estimates for θ_1 than Formula III. Hence, for maximum likelihood estimation, when the relative difference between θ_1 and θ_2 is not large any formula can be used. If θ_1 is much larger than θ_2, Formula II gives better estimate. If θ_2 is much larger than θ_1, Formula III gives better estimate. This is also an advantage of using the new representation. It is possible (but very rare) that the maximum likelihood estimates result as negative values when the relative difference between the two estimates is very large a modification as stated in the method of moments is considered.

4.3. Asymptotic properties of the maximum likelihood estimates

Tests and confidence intervals can be based on the fact that the maximum likelihood estimator $\hat{\Theta} = (\hat{\theta}_{1MLE}, \hat{\theta}_{2MLE})$ is asymptotically normally distributed $N_2\left(\Theta, I^{-1}(\Theta)\right)$ or more accurately that $\sqrt{n}(\hat{\Theta} - \Theta)$ is asymptotically $N_2(0, nI^{-1}(\Theta))$, where $I(\Theta)$ is the Fisher information matrix with entries

$$I_{i,j}\left(\Theta\right) = E\left(\frac{-\partial^2 \log L\left(\Theta\right)}{\partial\theta_i \partial\theta_j}\right), \quad i,j = 1,2. \tag{4.11}$$

Under mild regularity conditions, n^{-1} times the observed information matrix $I(\hat{\Theta})$ is a consistent estimator of $I(\Theta)/n$.

The observed information matrix using Formula II is given by

$$I\left(\hat{\Theta}\right) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$$

where

$$\begin{aligned} I_{11} &= -\frac{\partial^2 \log L}{\partial\theta_1^2}\bigg|_{\theta_1=\hat{\theta}_1} \\ &= \frac{\sum\limits_{i=1}^{n} z_i}{\hat{\theta}_{1MLE}^2} - \hat{\theta}_{2MLE}^2 \sum_{i=1}^{n}\left[\frac{{}_0\tilde{F}_1\left(z_i+1, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right){}_0\tilde{F}_1\left(z_i+3, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)}{{}_0\tilde{F}_1\left(z_i+1, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)^2}\right. \\ &\quad \left. - \frac{{}_0\tilde{F}_1\left(z_i+2, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)^2}{{}_0\tilde{F}_1\left(z_i+1, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)^2}\right], \end{aligned}$$

$$\begin{aligned} I_{22} &= -\frac{\partial^2 \log L}{\partial\theta_2^2}\bigg|_{\theta_2=\hat{\theta}_2} \\ &= -\hat{\theta}_{1MLE}^2 \sum_{i=1}^{n}\left[\frac{{}_0\tilde{F}_1\left(z_i+1, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right){}_0\tilde{F}_1\left(z_i+3, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)}{{}_0\tilde{F}_1\left(z_i+1, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)^2}\right. \\ &\quad \left. - \frac{{}_0\tilde{F}_1\left(z_i+2, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)^2}{{}_0\tilde{F}_1\left(z_i+1, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)^2}\right] \end{aligned}$$

and

$$\begin{aligned} I_{21} = I_{12} &= -\frac{\partial^2 \log L}{\partial\theta_1 \partial\theta_2}\bigg|_{\theta_1=\hat{\theta}_1, \theta_2=\hat{\theta}_2} \\ &= -\sum_{i=1}^{n} \frac{{}_0\tilde{F}_1\left(z_i+2, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)}{{}_0\tilde{F}_1\left(z_i+1, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)} \end{aligned}$$

$$- \hat{\theta}_{1MLE}\hat{\theta}_{2MLE} \sum_{i=1}^{n} \left[\frac{{}_0\tilde{F}_1\left(z_i+1, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right) {}_0\tilde{F}_1\left(z_i+3, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)}{{}_0\tilde{F}_1\left(z_i+1, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)^2} \right.$$
$$\left. - \frac{{}_0\tilde{F}_1\left(z_i+2, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)^2}{{}_0\tilde{F}_1\left(z_i+1, \hat{\theta}_{1MLE}\hat{\theta}_{2MLE}\right)^2} \right].$$

The 95% confidence intervals for θ_1 and θ_2 are obtained by

$$\hat{\theta}_1 \pm 1.96\sqrt{\frac{I_{22}}{I_{11}I_{22} - I_{12}^2}} \quad \text{and} \quad \hat{\theta}_2 \pm 1.96\sqrt{\frac{I_{11}}{I_{11}I_{22} - I_{12}^2}}. \tag{4.12}$$

5. Testing

The likelihood ratio test is a statistical test for making a decision between two hypotheses based on the value of this ratio.

5.1. Likelihood ratio test for equality of the parameters

Let $x_1, x_2, \cdots, x_n$ be the outcome of a random sample of size n with respect to the variable X. We consider the likelihood ratio test (LRT) for the null hypothesis:
H_0 : The data is drawn from PD (θ, θ) against the alternative,
H_1 : The data is drawn from PD (θ_1, θ_2).

The LRT statistic is written as

$$\lambda_n = \frac{f\left(x_1, x_2, \cdots, x_n; \hat{\theta}\right)}{f\left(x_1, x_2, \cdots, x_n; \hat{\theta}_1, \hat{\theta}_2\right)}, \tag{5.1}$$

where $f(x_1, x_2, \cdots, x_n; \hat{\theta})$ denotes the likelihood function of the sample under the null hypothesis calculated at maximum likelihood estimate of θ and $f(x_1, x_2, \cdots, x_n; \hat{\theta}_1, \hat{\theta}_2)$ denotes the likelihood function of the sample under the alternative hypothesis calculated at maximum likelihood estimates of θ_1 and θ_2.

Under H_0 the likelihood function is given by

$$f\left(x_1, x_2, \cdots, x_n; \theta\right) = e^{-2n\theta}\theta^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} {}_0\tilde{F}_1\left(x_i+1, \theta^2\right). \tag{5.2}$$

The log-likelihood is given by

$$\ln f\left(x_1, x_2, \cdots, x_n; \theta\right) = -2n\theta + \left(\sum_{i=1}^{n} x_i\right)\ln\theta + \sum_{i=1}^{n} \ln {}_0\tilde{F}_1\left(x_i+1, \theta^2\right). \tag{5.3}$$

The maximum likelihood estimate $\hat{\theta}$ of θ is obtained by solving the nonlinear equation

$$\frac{\partial \ln f\left(x_1, x_2, \cdots, x_n; \theta\right)}{\partial\theta} = -2n + \frac{\sum_{i=1}^{n} x_i}{\theta} + 2\theta\sum_{i=1}^{n} \frac{{}_0\tilde{F}_1\left(x_i+2, \theta^2\right)}{{}_0\tilde{F}_1\left(x_i+1, \theta^2\right)} = 0. \tag{5.4}$$

And hence,

$$f\left(x_1, x_2, \cdots, x_n; \hat{\theta}\right) = e^{-2n\hat{\theta}} \hat{\theta}^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} {}_0\tilde{F}_1\left(x_i + 1, \hat{\theta}^2\right). \tag{5.5}$$

Under H_1,

$$f\left(x_1, x_2, \cdots, x_n; \hat{\theta}_1, \hat{\theta}_2\right) = e^{-n\hat{\theta}_1 - n\hat{\theta}_2} \hat{\theta}_1^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} {}_0\tilde{F}_1\left(x_i + 1, \hat{\theta}_1\hat{\theta}_2\right), \tag{5.6}$$

$$\begin{aligned} & -2\ln\lambda_n \\ & = -2\left(\begin{array}{c} -2n\hat{\theta} + \left(\sum_{i=1}^{n} x_i\right)\ln\hat{\theta} + \sum_{i=1}^{n} \ln {}_0\tilde{F}_1\left(x_i + 1, \hat{\theta}^2\right) \\ -\left(-n\hat{\theta}_1 - n\hat{\theta}_2 + \left(\sum_{i=1}^{n} x_i\right)\ln\hat{\theta}_1 + \sum_{i=1}^{n} \ln {}_0\tilde{F}_1\left(x_i + 1, \hat{\theta}_1\hat{\theta}_2\right)\right) \end{array}\right). \end{aligned} \tag{5.7}$$

Under regularity condition for large values of n, $-2\ln\lambda_n$ has chi-square distribution with one degree of freedom. We reject H_0 if $-2\ln\lambda_n > \chi^2_{1-\alpha,1}$.

5.2. Likelihood ratio test for $\theta_2 = 0$

If the observed data were all nonnegative integer values even though they are differences, it is interesting to test if Poisson distribution can fits the data as well as the Poisson difference or not.

Let $x_1, x_2, \cdots, x_n$ be the outcomes of a random sample of size n with respect to the variable X where all these outcomes are nonnegative integer values. We consider the LRT for the null hypothesis:
H_0 : The data is drawn from $Poisson\,(\theta_1)$ (i.e. $\theta_2 = 0$) against the alternative,
H_1 : The data is drawn from PD (θ_1, θ_2).

The LRT statistic is written as

$$\lambda_n = \frac{f\left(x_1, x_2, \cdots, x_n; \hat{\theta}_{01}\right)}{f\left(x_1, x_2, \cdots, x_n; \hat{\theta}_1, \hat{\theta}_2\right)}, \tag{5.8}$$

where $f(x_1, x_2, \cdots, x_n; \hat{\theta}_{01})$ denotes the likelihood function of the sample under the null hypothesis calculated at maximum likelihood estimate of θ_1 and $f(x_1, x_2, \cdots, x_n; \hat{\theta}_1, \hat{\theta}_2)$ denotes the likelihood function of the sample under the alternative hypothesis calculated at maximum likelihood estimates of θ_1 and θ_2.

Under H_0 the likelihood function is given by

$$f\left(x_1, x_2, \cdots, x_n; \hat{\theta}_{01}\right) = e^{-n\bar{x}} \bar{x}^{\sum_{i=1}^{n} x_i} / \prod_{i=1}^{n} x_i! \tag{5.9}$$

Under H_1,

$$f\left(x_1, x_2, \cdots, x_n; \hat{\theta}_1, \hat{\theta}_2\right) = e^{-n\hat{\theta}_1 - n\hat{\theta}_2} \hat{\theta}_1^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} {}_0\tilde{F}_1\left(x_i + 1, \hat{\theta}_1\hat{\theta}_2\right). \tag{5.10}$$

Therefore,

$$-2\ln\lambda_n = -2\left[-n\bar{X} + \left(\sum_{i=1}^{n} x_i\right)\ln\bar{X} - \sum_{i=1}^{n} \ln x_i!\right.$$

$$-\left(-n\hat{\theta}_1 - n\hat{\theta}_2 + \left(\sum_{i=1}^{n} x_i\right)\ln\hat{\theta}_1 + \sum_{i=1}^{n}\ln{}_0\tilde{F}_1\left(x_i+1,\hat{\theta}_1\hat{\theta}_2\right)\right)\Bigg]. \tag{5.11}$$

Under regularity condition for large values of n, $-2\ln\lambda_n$ has chi-square distribution with one degree of freedom. We reject H_0 if $-2\ln\lambda_n > \chi^2_{1-\alpha,1}$.

6. Simulation study

The main objective of this section is to discuss some simulation results for computing the estimates of the parameters of PD (θ_1, θ_2) using the method of moments and the maximum likelihood method.

To generate one observation, Z, from PD (θ_1, θ_2) we generated one observation, X, from the Poisson distribution with parameter θ_1 and an independent observation, Y, from the Poisson distribution with parameter θ_2 and computed $Z = X - Y$.

In this simulation study we used 1000 samples of size $n = 10$, 20, 30, 50, 100, 150 and 200 and different values of θ_1 and θ_2.

We calculated the bias and used the relative mean square error (RMSE) as measures of the performance of the estimates in all the considered methods of estimation, where

$$\mathrm{BIAS}\left(\hat{\theta}_i\right) = \frac{1}{r}\sum_{j=1}^{r}\left(\hat{\theta}_{ji} - \theta_i\right), \tag{6.1}$$

$$\mathrm{RMSE}\left(\hat{\theta}_i\right) = \frac{1}{\theta_i}\left[\frac{1}{r}\sum_{j=1}^{r}\left(\hat{\theta}_{ji} - \theta_i\right)^2\right]^{1/2} \tag{6.2}$$

for $i = 1, 2$ and $r = 1000$.

Tables 1–3 and Figures 1–8 illustrate some of the results.

Table 1. Estimation result when $\theta_1 = 0.1$ and $\theta_2 = 0.1, 0.5, 1, 5, 10, 20, 40, 100$

θ_1	θ_2	n	BIAS θ_1MLE	BIAS θ_2MLE	RMSE θ_1MLE	RMSE θ_2MLE	BIAS θ_1MM	BIAS θ_2MM	RMSE θ_1MM	RMSE θ_2MM
0.1	0.1	10	-0.0132	-0.0054	1.2930	1.0322	-0.0002	0.0076	1.1212	1.1572
0.1	0.1	30	0.0013	-0.0018	0.6077	0.5903	0.0019	-0.0011	0.6379	0.6200
0.1	0.1	50	-0.0010	0.0004	0.4777	0.4769	-0.0007	0.0007	0.4990	0.5052
0.1	0.5	10	0.0046	0.0000	1.3411	0.4941	0.0085	0.0039	1.8019	0.5563
0.1	0.5	30	-0.0054	-0.0025	0.7032	0.2721	-0.0081	-0.0052	0.9085	0.2986
0.1	0.5	50	-0.0041	0.0003	0.5905	0.2083	-0.0041	0.0003	0.7721	0.2264
0.1	1	10	-0.0002	0.0006	1.7204	0.3463	0.0718	0.0726	2.4962	0.4272
0.1	1	30	-0.0004	-0.0061	0.9780	0.1993	0.0196	0.0139	1.2906	0.2223
0.1	1	50	-0.0039	-0.0018	0.7238	0.1494	0.0102	0.0124	1.0872	0.1768
0.1	5	10	-0.0865	-0.0641	1.3821	0.1442	0.3465	0.3689	8.9107	0.2359
0.1	5	30	-0.0785	-0.0782	1.2607	0.0835	0.2210	0.2214	5.5154	0.1390
0.1	5	50	-0.0646	-0.0498	1.1949	0.0684	0.1444	0.1592	3.9078	0.1062
0.1	10	10	-0.0971	-0.0599	0.9916	0.1013	0.7334	0.7706	17.4268	0.2055
0.1	10	30	-0.0942	-0.0942	0.9819	0.0588	0.4983	0.4983	10.9426	0.1244
0.1	10	50	-0.0919	-0.0726	0.9752	0.0468	0.3464	0.3657	7.7686	0.0935
0.1	20	10	-0.0984	-0.1232	0.9937	0.0730	1.8747	1.8499	37.9693	0.2089
0.1	20	30	-0.0964	-0.0974	0.9895	0.0407	0.9964	0.9954	20.5929	0.1123
0.1	20	50	-0.0954	-0.0866	0.9856	0.0319	0.7507	0.7595	15.1794	0.0808
0.1	40	10	-0.0992	-0.1304	0.9968	0.0514	3.8079	3.7767	76.1753	0.2009
0.1	40	30	-0.0978	-0.0974	0.9931	0.0286	2.0356	2.0360	41.0485	0.1077
0.1	40	50	-0.0982	-0.0855	0.9929	0.0224	1.5613	1.5741	30.3452	0.0779
0.1	100	10	-0.0997	-0.1514	0.9990	0.0324	9.5382	9.4865	187.956	0.1927
0.1	100	30	-0.0996	-0.1001	0.9985	0.0180	5.1811	5.1806	102.720	0.1049
0.1	100	50	-0.0993	-0.0773	0.9976	0.0141	4.0074	4.0294	76.3022	0.0769

Table 2. Estimation result when $\theta_1 = 3$ and $\theta_2 = 0.1, 0.5, 1, 5, 10, 20, 40, 100$

θ_1	θ_2	n	BIAS θ_1MLE	BIAS θ_2MLE	RMSE θ_1MLE	RMSE θ_2MLE	BIAS θ_1MM	BIAS θ_2MM	RMSE θ_1MM	RMSE θ_2MM
3	0.1	10	-0.0400	-0.0363	0.1968	2.4765	0.2439	0.2476	0.2856	6.2152
3	0.1	30	-0.0352	-0.0322	0.1164	1.6012	0.1039	0.1069	0.1503	3.1133
3	0.1	50	-0.0259	-0.0302	0.0918	1.3618	0.0719	0.0676	0.1152	2.3882
3	0.5	10	-0.1100	-0.1131	0.2514	1.1856	0.0088	0.0057	0.3295	1.7551
3	0.5	30	-0.0570	-0.0482	0.1558	0.7619	-0.0064	0.0024	0.1791	0.9222
3	0.5	50	-0.0404	-0.0471	0.1254	0.6125	-0.0124	-0.0191	0.1412	0.7116
3	1	10	-0.1593	-0.1720	0.3146	0.8777	0.0037	-0.0090	0.3568	1.0020
3	1	30	-0.0661	-0.0544	0.1866	0.5056	-0.0101	0.0016	0.1977	0.5373
3	1	50	-0.0436	-0.0501	0.1461	0.3867	-0.0138	-0.0203	0.1525	0.4104
3	5	10	-0.3663	-0.3912	0.6018	0.3752	0.0061	-0.0188	0.6450	0.4043
3	5	30	-0.1499	-0.1201	0.3542	0.2171	-0.0219	0.0080	0.3559	0.2209
3	5	50	-0.0834	-0.0901	0.2618	0.1620	-0.0137	-0.0204	0.2679	0.1655
3	10	10	-0.5394	-0.5760	0.8937	0.2858	0.0221	-0.0145	1.0264	0.3236
3	10	30	-0.2851	-0.2418	0.5563	0.1716	-0.0466	-0.0033	0.5682	0.1763
3	10	50	-0.1359	-0.1421	0.4144	0.1300	-0.0005	-0.0067	0.4256	0.1331
3	20	10	-0.6859	-0.7406	1.0528	0.1714	0.8143	0.7596	1.5387	0.2439
3	20	30	-0.3527	-0.2938	0.8026	0.1237	0.0872	0.1460	0.9013	0.1400
3	20	50	-0.1722	-0.1821	0.6868	0.1070	0.0792	0.0692	0.7179	0.1115
3	40	10	-2.0415	-2.1197	0.9312	0.0866	2.4659	2.3877	2.7106	0.2109
3	40	30	-1.2048	-1.1220	0.8843	0.0717	0.7948	0.8776	1.4775	0.1140
3	40	50	-0.9959	-1.0106	0.8395	0.0669	0.5640	0.5493	1.2099	0.0933
3	100	10	-2.5942	-2.7140	0.9712	0.0439	7.9947	7.8749	6.4012	0.1954
3	100	30	-2.3632	-2.2271	0.9413	0.0331	3.4991	3.6352	3.2574	0.0992
3	100	50	-2.2448	-2.2589	0.9291	0.0311	2.8107	2.7965	2.6622	0.0810

Table 3. Estimation result when $\theta_1 = 20$ and $\theta_2 = 0.1, 0.5, 1, 5, 10, 20, 40, 100$

θ_1	θ_2	n	BIAS θ_1MLE	BIAS θ_2MLE	RMSE θ_1MLE	RMSE θ_2MLE	BIAS θ_1MM	BIAS θ_2MM	RMSE θ_1MM	RMSE θ_2MM
20	0.1	10	-0.1042	-0.0982	0.0710	0.9941	1.8702	1.8762	0.2116	39.4132
20	0.1	30	-0.0957	-0.0965	0.0427	0.9909	0.9647	0.9640	0.1050	19.3010
20	0.1	50	-0.0830	-0.0957	0.0320	0.9824	0.7267	0.7140	0.0811	14.8676
20	0.5	10	-0.4656	-0.4664	0.0749	0.9784	1.6849	1.6841	0.2135	8.0079
20	0.5	30	-0.4115	-0.4064	0.0494	0.9462	0.7716	0.7767	0.1069	3.9424
20	0.5	50	-0.3680	-0.3831	0.0394	0.9215	0.5373	0.5222	0.0841	3.0937
20	1	10	-0.7847	-0.7951	0.0847	0.9564	1.4707	1.4603	0.2171	4.0909
20	1	30	-0.6256	-0.6177	0.0627	0.9020	0.5748	0.5828	0.1127	2.0982
20	1	50	-0.5054	-0.5204	0.0539	0.8781	0.3618	0.3468	0.0894	1.6696
20	5	10	-0.8775	-0.9001	0.2545	0.9957	0.0209	-0.0017	0.3039	1.2054
20	5	30	-0.4062	-0.3801	0.1517	0.5914	-0.0152	0.0109	0.1598	0.6277
20	5	50	-0.3174	-0.3326	0.1253	0.4896	-0.0684	-0.0836	0.1273	0.4976
20	10	10	-2.2817	-2.3030	0.3474	0.6982	-0.4740	-0.4953	0.3404	0.6837
20	10	30	-0.5415	-0.4884	0.2070	0.4139	0.0875	0.1405	0.2000	0.3995
20	10	50	-0.2659	-0.2673	0.1495	0.2989	0.0295	0.0281	0.1499	0.2992
20	20	10	-2.4851	-2.5179	0.4270	0.4317	-0.6085	-0.6413	0.4537	0.4584
20	20	30	-0.4914	-0.4246	0.2564	0.2570	0.1574	0.2242	0.2639	0.2649
20	20	50	-0.3720	-0.3744	0.1966	0.1978	0.0098	0.0074	0.1996	0.2008
20	40	10	-3.3294	-3.4121	0.7015	0.3503	0.4396	0.3569	0.7121	0.3549
20	40	30	-1.1249	-1.1077	0.4036	0.2034	0.1001	0.1173	0.3873	0.1956
20	40	50	-0.5934	-0.6097	0.3193	0.1605	0.1552	0.1390	0.3087	0.1553
20	100	10	-2.4862	-2.5904	1.0489	0.2105	0.7053	0.6011	1.4005	0.2800
20	100	30	-1.1913	-1.1674	0.7057	0.1421	0.3341	0.3580	0.7704	0.1550
20	100	50	-0.8106	-0.8495	0.5775	0.1161	0.3308	0.2919	0.6079	0.1221

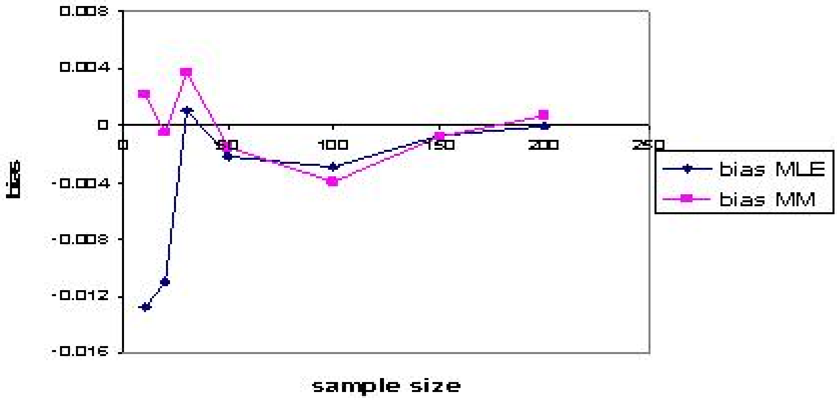

Figure 1. Bias of θ_1 using MM and ML versus sample size when $\theta_1 = 0.3,\ \theta_2 = 0.3$

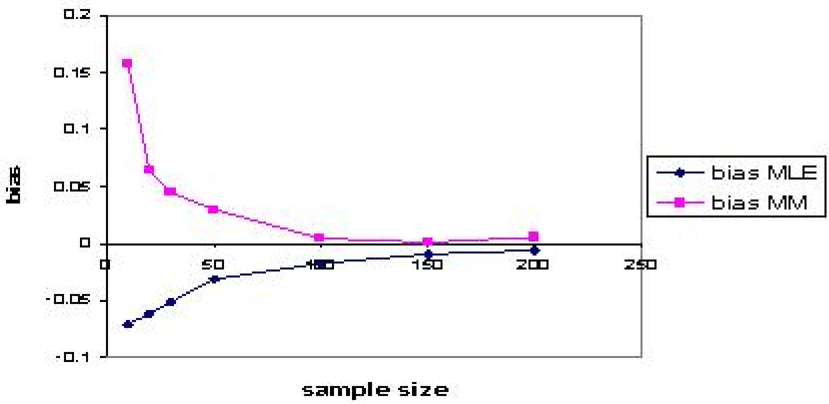

Figure 2. Bias of θ_1 using MM and ML versus sample size when $\theta_1 = 0.3,\ \theta_2 = 3$

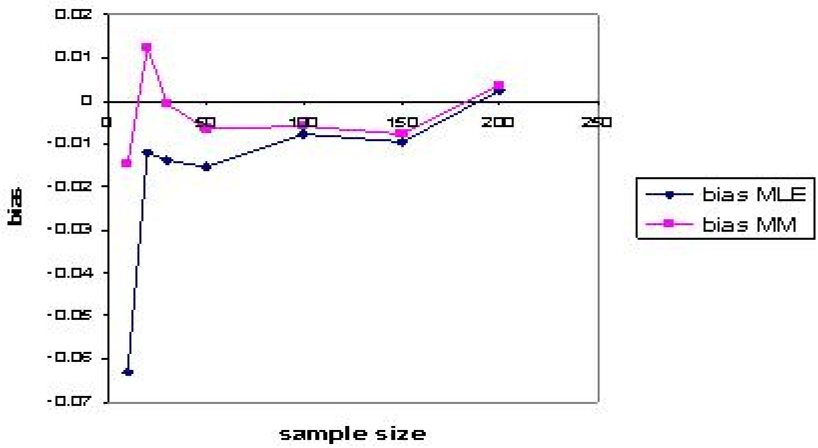

Figure 3. Bias of θ_1 using MM and ML versus sample size when $\theta_1 = 1, \ \theta_2 = 1$

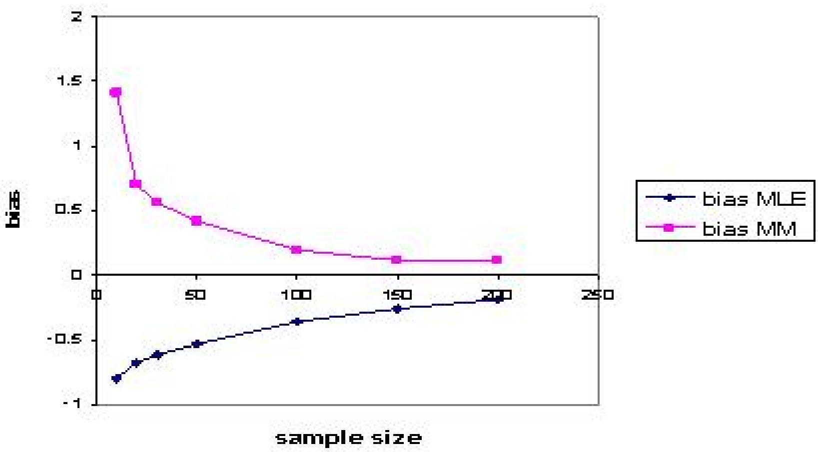

Figure 4. Bias of θ_1 using MM and ML versus sample size when $\theta_1 = 1, \ \theta_2 = 20$

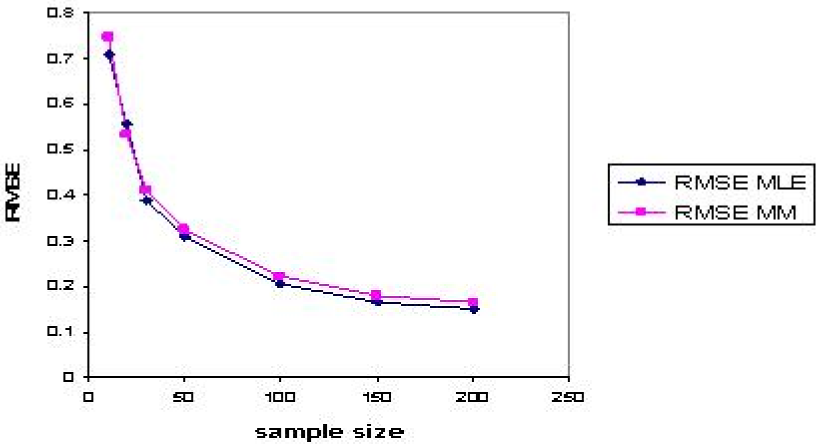

Figure 5. RMSE of θ_1 using MM and ML versus sample size when $\theta_1 = 0.3,\ \theta_2 = 0.3$

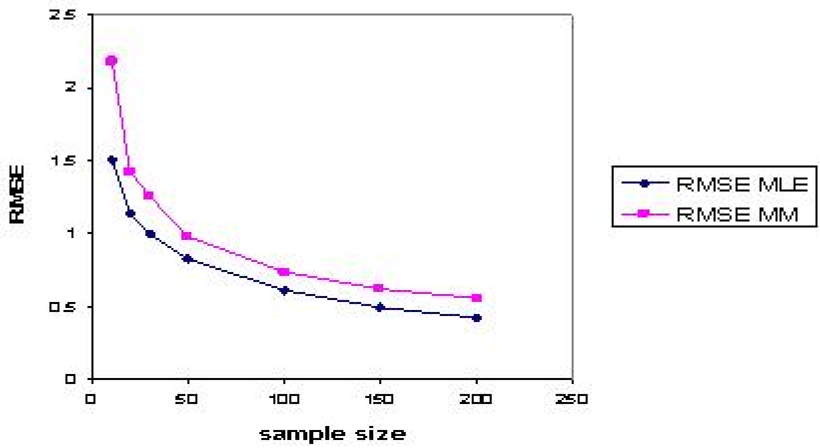

Figure 6. RMSE of θ_1 using MM and ML versus sample size when $\theta_1 = 0.3,\ \theta_2 = 3$

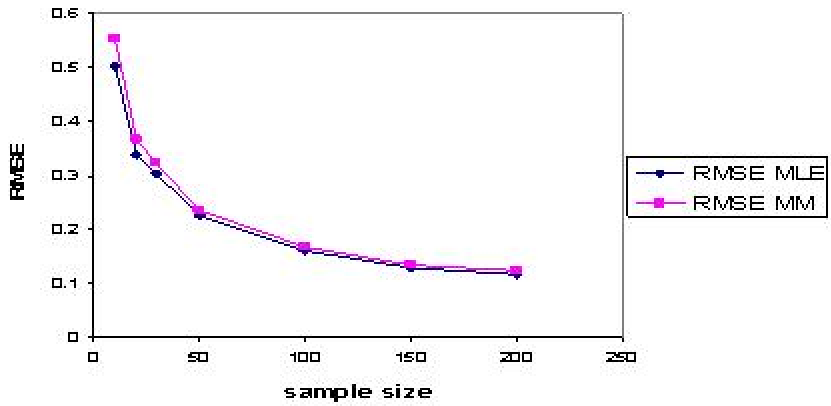

Figure 7. RMSE of θ_1 using MM and ML versus sample size when $\theta_1 = 1,\ \theta_2 = 1$

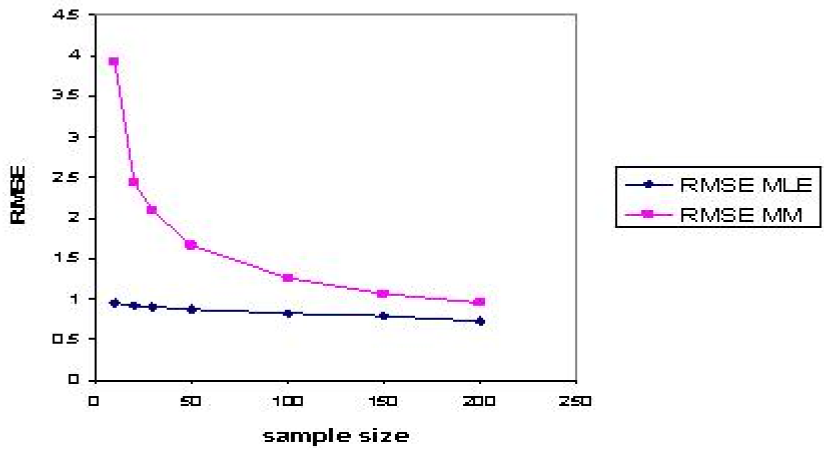

Figure 8. RMSE of θ_1 using MM and ML versus sample size when $\theta_1 = 1,\ \theta_2 = 20$

In order to investigate the power of LRT for equality of the two parameters, the empirical power of the test was examined. The empirical power of the test is defined as the proportion of times the null hypothesis was rejected when the data actually were generated under the alternative hypothesis using 1000 replications.

For each of a sample size $n = 30$, 50 and 100, the power of the test is computed under various choices for the parameters of the alternative distribution. We obtain the power at $\theta_1 = 0.1,\ 0.3,\ 0.5,\ 1,\ 3,\ 5,\ 10,\ 20$ and 50 and $\theta_2 = c\theta_1$ for $c = 0.1,\ 0.2,\ 0.3,\ 0.5,\ 1,\ 1.1,\ 1.2,\ 1.3$ and 1.5. Note that $c = 1$ corresponds to the null hypothesis and the calculated values are the empirical type one error of the test.

Table 4 shows the power of the test when the significance level is 5%, while Table 5 shows the power of the test when the significance level is 1%.

Table 4. Power of the LRT of equal parameters when the significance level is 5%

θ_1	n	c=0.1	c=0.2	c=0.3	c=0.5	c=1	c=1.1	c=1.2	c=1.3	c=1.5
	30	0.498	0.373	0.314	0.179	0.064	0.053	0.037	0.039	0.049
0.1	50	0.645	0.463	0.324	0.152	0.033	0.05	0.046	0.063	0.1
	100	0.868	0.702	0.535	0.262	0.047	0.046	0.063	0.1	0.171
	30	0.83	0.653	0.467	0.233	0.036	0.064	0.076	0.095	0.147
0.3	50	0.96	0.853	0.689	0.339	0.052	0.056	0.079	0.125	0.239
	100	1	0.994	0.949	0.645	0.041	0.043	0.092	0.206	0.42
	30	0.958	0.853	0.688	0.337	0.048	0.058	0.088	0.127	0.247
0.5	50	1	0.983	0.892	0.542	0.053	0.066	0.099	0.165	0.347
	100	1	1	0.998	0.86	0.039	0.067	0.136	0.307	0.635
	30	0.999	0.992	0.931	0.612	0.064	0.076	0.131	0.21	0.411
1	50	1	1	0.996	0.827	0.046	0.081	0.156	0.283	0.598
	100	1	1	1	0.991	0.045	0.112	0.282	0.531	0.904
	30	1	1	0.999	0.965	0.056	0.115	0.261	0.474	0.837
3	50	1	1	1	0.999	0.055	0.136	0.372	0.674	0.98
	100	1	1	1	1	0.041	0.219	0.63	0.939	1
	30	1	1	1	1	0.06	0.13	0.368	0.664	0.972
5	50	1	1	1	1	0.061	0.189	0.536	0.885	1
	100	1	1	1	1	0.045	0.334	0.835	0.993	1
	30	1	1	1	1	0.064	0.215	0.639	0.911	1
10	50	1	1	1	1	0.057	0.308	0.844	0.991	1
	100	1	1	1	1	0.043	0.563	0.994	1	1
	30	1	1	1	1	0.063	0.373	0.914	0.999	1
20	50	1	1	1	1	0.06	0.58	0.987	1	1
	100	1	1	1	1	0.042	0.855	1	1	1
	30	1	1	1	1	0.057	0.769	0.999	1	1
50	50	1	1	1	1	0.061	0.923	1	1	1
	100	1	1	1	1	0.041	1	1	1	1

Table 5. Power of the LRT of equal parameters when the significance level is 1%

θ_1	n	c=0.1	c=0.2	c=0.3	c=0.5	c=1	c=1.1	c=1.2	c=1.3	c=1.5
0.1	30	0.091	0.059	0.045	0.031	0.007	0.005	0.001	0.004	0.006
	50	0.38	0.248	0.162	0.061	0.009	0.013	0.016	0.015	0.029
	100	0.711	0.472	0.302	0.119	0.009	0.014	0.018	0.024	0.065
0.3	30	0.605	0.382	0.235	0.082	0.006	0.008	0.011	0.025	0.052
	50	0.877	0.647	0.42	0.153	0.011	0.011	0.018	0.033	0.082
	100	1	0.962	0.807	0.387	0.006	0.009	0.022	0.068	0.215
0.5	30	0.854	0.625	0.425	0.149	0.013	0.012	0.022	0.04	0.082
	50	0.989	0.908	0.707	0.3	0.01	0.015	0.026	0.046	0.152
	100	1	1	0.979	0.658	0.01	0.012	0.038	0.135	0.388
1	30	0.996	0.937	0.797	0.335	0.017	0.026	0.054	0.086	0.212
	50	1	1	0.97	0.625	0.01	0.02	0.054	0.115	0.345
	100	1	1	1	0.943	0.012	0.026	0.099	0.293	0.732
3	30	1	1	0.997	0.885	0.019	0.028	0.123	0.242	0.634
	50	1	1	1	0.984	0.013	0.042	0.17	0.429	0.896
	100	1	1	1	1	0.008	0.078	0.378	0.817	0.998
5	30	1	1	1	0.985	0.02	0.046	0.166	0.41	0.867
	50	1	1	1	1	0.011	0.068	0.304	0.69	0.99
	100	1	1	1	1	0.006	0.133	0.661	0.964	1
10	30	1	1	1	1	0.025	0.083	0.371	0.761	0.998
	50	1	1	1	1	0.014	0.128	0.654	0.966	1
	100	1	1	1	1	0.006	0.333	0.946	1	1
20	30	1	1	1	1	0.023	0.161	0.746	0.983	1
	50	1	1	1	1	0.01	0.328	0.949	1	1
	100	1	1	1	1	0.005	0.685	1	1	1
50	30	1	1	1	1	0.019	0.492	0.993	1	1
	50	1	1	1	1	0.014	0.786	1	1	1
	100	1	1	1	1	0.008	0.993	1	1	1

Discussion of the simulation results:

(1) The maximum likelihood estimates are better than the moment estimates in terms of relative mean square error. Out of the 700 different cases considered in this simulation the RMSE of $\hat{\theta}_{1MLE}$ is less than $\hat{\theta}_{1MM}$ in 669 cases and RMSE of $\hat{\theta}_{2MLE}$ is less than $\hat{\theta}_{2MM}$ in 672 cases.
(2) The RMSE differ substantially between the two methods of estimation when there is a large relative difference between the two parameters. In these cases the maximum likelihood estimates are much better than the moment estimates in terms of relative mean square error as shown by the Figures 1–8.
(3) In terms of bias, the method of moment is better than the maximum likelihood method when the relative difference between the two parameters is small or moderate since the moment estimates are unbiased. The maximum likelihood estimates are much better than the moment estimates in terms of the bias when the relative difference between the two parameters is large and the sample size is small, while the method of moment becomes better for large sample size.
(4) When there is no large relative difference between the two parameters, both methods are good. As well as, for large sample size, both methods can be used even when there is large relative difference.
(5) When both θ_1 and θ_2 are large, moment estimates and ML estimates are very close as expected since the distribution approaches normality.
(6) As expected, the RMSE always decreases as the sample size increases in both methods of estimation. It has been noticed that the RMSE increases with the decreases of the parameter in both methods.
(7) The maximum likelihood estimators are frequently negatively biased and the bias decreases as the sample size increases.
(8) The LRT test has lower performance when it is used to detect components that are very close, in other words the power of the test increase with the relative distance of the components.
(9) Considering the value of c fixed, the power increases as the values of θ_1 and θ_2 increase.
(10) When we increase the sample size, the power improves as expected.
(11) At $c = 1$, the type one error is smaller or around 0.05 in the 5% level of significance table and is smaller or around 0.01 in the 1% level of significance table.

7. Applications

7.1. Application to the Saudi Stock Exchange data

The data has been downloaded from the Saudi Stock Exchange and further filtered. Trading in Saudi Basic Industry (SABIC) and Arabian Shield from the Saudi stock exchange (TASI) were recorded at June 30, 2007, every minute. The Saudi Stock Exchange opens at 11.00 am and closes at 3.30 pm. Missing minutes have been added with a zero price change. The first and final 15 minutes of the trading day, were deleted from the data. The reason for this is that we only focus on studying the price formation during ordinary trading. The minimum amount a price can move is SAR 0.25 in Saudi stock i.e. the tick size is 0.25. The price change is therefore characterized by discrete jumps. The data consists of *the difference in price every minute as number of ticks* = (*close price-open price*) $*$ 4. Note that, our considered data could take both positive and negative integer values.

In Figure 9 and 10, the price change at every minute is illustrated in terms of number of ticks for SABIC and Arabian Shield. Descriptive statistics of the considered data are presented in Table 6.

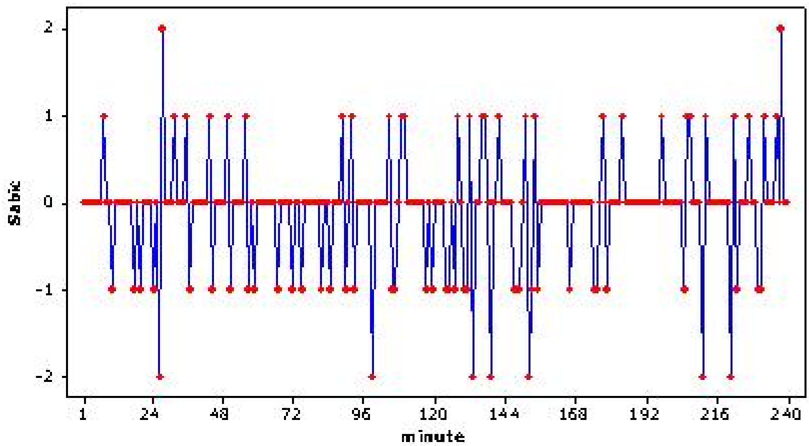

Figure 9. Plot of the price change every minute for SABIC

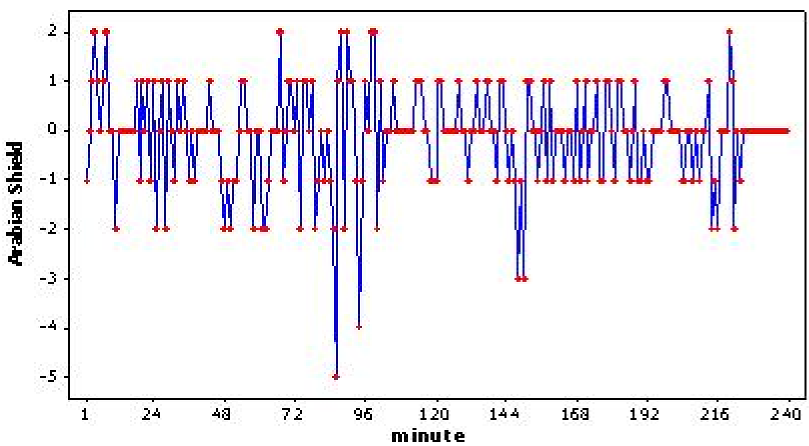

Figure 10. Plot of the price change every minute for Arabian Shield

Table 6. Descriptive statistics for SABIC and Arabian Shield

Variable	Sample size	Mean	Standard Deviation	Minimum	Maximum
SABIC	240	-0.0833	0.6479	-2	2
Arabian Shield	240	-0.1042	1.0276	-5	2

In order to test if our samples are random samples we conduct the runs test on every sample. (Runs tests test whether or not the data order is random. No assumptions are made about population distribution parameters.)

For SABIC the p-value = 0.852 and for Arabian Shield the p-value = 0.123. Since both p-values are greater than 0.05, our samples can be considered as random samples.

The numbers of ticks of price change take values on the integer numbers. An appropriate distribution to fit these samples could be the PD distribution. Maximum likelihood and moment estimates of θ_1 and θ_2 are obtained using methods discussed in the previous section and illustrated in the Table 7.

Table 7. Estimation result for SABIC and Arabian Shield

Stock	$\hat{\theta}_{1MLE}$	$\hat{\theta}_{2MLE}$	$\hat{\theta}_{1MM}$	$\hat{\theta}_{2MM}$
SABIC	0.1681	0.2514	0.1682	0.2516
Arabian Shield	0.451	0.5551	0.4737	0.5779

The Pearson Chi-square test is performed to both samples to test if PD distribution gives good fit to the data. The null hypothesis is that the sample comes from PD distribution and the alternative hypothesis is that sample does not come from PD distribution. For SABIC the p-value = 0.449862, which implies that PD(0.168,0.251) fits the data well. For Arabian Shield the p-value = 0.137931, which implies that PD(0.451,0.5551) fits the data well.

The 95% confidence intervals for θ_1 and θ_2 are calculated for SABIC and Arabian Shield.

SABIC: The 95% confidence intervals for θ_1 is (0.1088, 0.2273) and 95% CI for θ_2 is (0.1818, 0.3211).

Arabian Shield: The 95% confidence intervals for θ_1 is (0.3355, 0.5664) and 95% CI for θ_2 is (0.4327, 0.6776).

In both cases, SABIC and Arabian Shield the two confidence intervals overlapped indicating that θ_1 and θ_2 could be equal. The likelihood ratio test for equality of means is conducted and the statistic for SABIC = 3.979 and for Arabian Shield = 2.582. The tabulated value to compare with is $\chi^2_{1,0.95} = 3.84146$. For SABIC, we reject the hypothesis that θ_1 and θ_2 are equal. While for Arabian Shield we fail to reject the hypothesis that θ_1 and θ_2 are equal confirming the overlapping in the confidence intervals. The maximum likelihood estimate of θ is $\hat{\theta}_{MLE} = 0.507185$. Hence PD(0.507,0.507) gives a good fit for these data.

Bar charts of the relative frequency, PD distribution estimated using the method of moments and the maximum likelihood method for the two stocks are plotted in the

Figures 11 and 12. In Figure 12, PD distribution with equal parameters estimated using maximum likelihood is also plotted. Clearly, from the graphs that PD fits well both data and no significance difference is found from the two methods of estimation for SABIC while this is not the case for Arabian Shield.

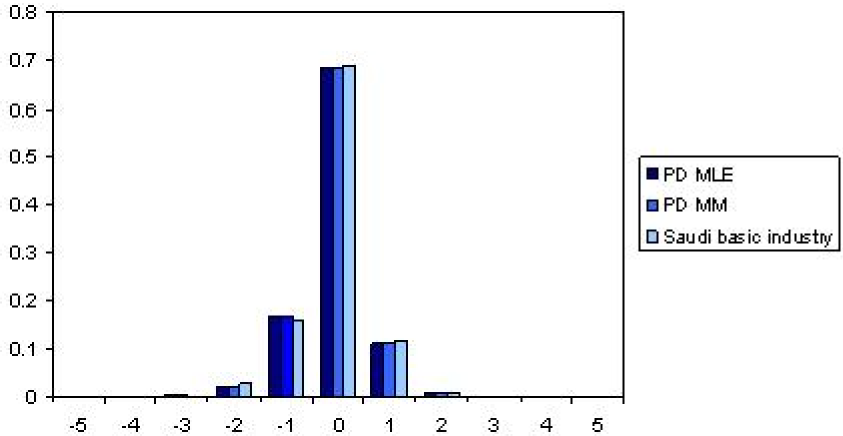

Figure 11. SABIC and fitted distributions

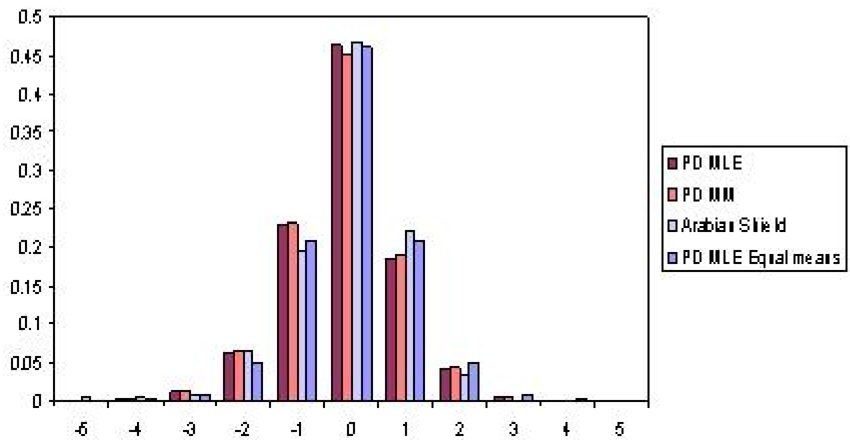

Figure 12. Arabian Shield and fitted distributions

7.2. Application to nursery intensive care unit data

The numbers of occupied beds of the NICU (nursery intensive care unit) in Dallah hospital at Riyadh, Saudi Arabia from December 12, 2005 to March 22, 2006 are time dependent but after taking difference of every two consecutive days the resulting data is a random sample of size 100. This data represents the change in number of beds during 24 hours in NICU.

Descriptive statistics of the original and the differenced data are as in Table 8.

Table 8. Number of occupied beds and the difference

Variable	Sample size	Mean	Standard Deviation	Minimum	Maximum
No. of occupied beds	100	9.861	2.250	5	17
Difference	101	-0.0300	1.654	-4	4

Figures 13 and 14 illustrate the plots of the number of occupied beds and the difference of every two consecutive days, respectively. The runs test was applied to both the original and the differenced data. The result of the test was that the number of occupied beds is not a random sample (p-value=0), while after differentiating the resulting data is a random sample (p-value=0.979).

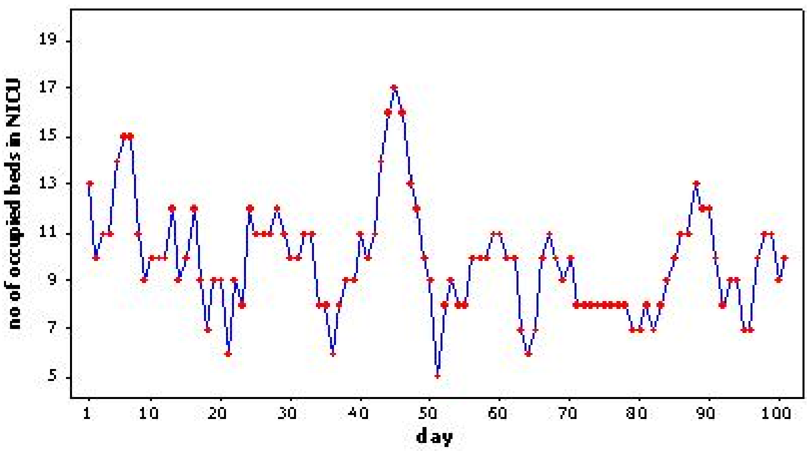

Figure 13. Number of occupied beds in NICU

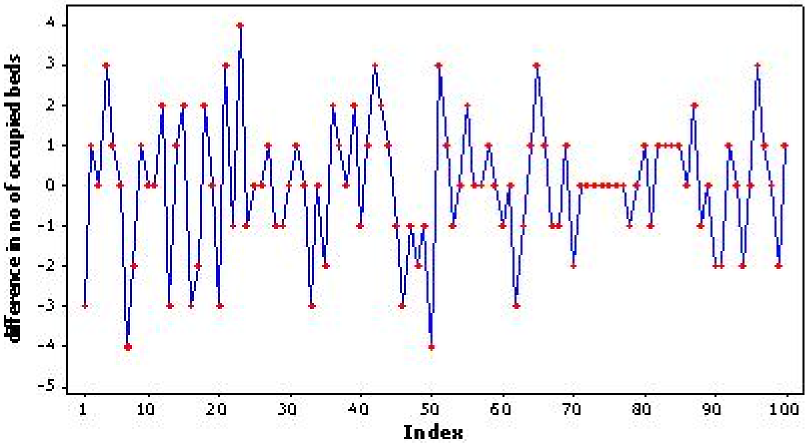

Figure 14. Difference in number of occupied beds in NICU

The changes in number of occupied beds during 24 hours in NICU take values on the integer numbers. A good candidate to fit this sample could be the PD distribution. The estimation result are summarized in Table 9. Figure 15 represents the fitted PD distributions using both methods with the relative frequency of the data.

Table 9. Estimation result for NICU

$\hat{\theta}_{1MLE}$	$\hat{\theta}_{2MLE}$	$\hat{\theta}_{1MM}$	$\hat{\theta}_{2MM}$
1.34236	1.37236	1.35323	1.38323

The Pearson Chi-square test is performed to test if PD distribution gives good fit to the data. The p-value = 0.407497, which implies that PD(1.342, 1.372) fits the data well.

The 95% CI for θ_1 is (0.9089, 1.7758) and the 95% CI for θ_2 is (0.93761, 1.8071).

The likelihood ratio test for equality of the parameters is conducted. The statistic $= 0.0331479 < \chi^2_{1,0.95} = 3.84146$ which implies that a Poisson difference with equal parameters fits the data well.

The maximum likelihood estimate of θ is given by $\hat{\theta} = 1.3578$.

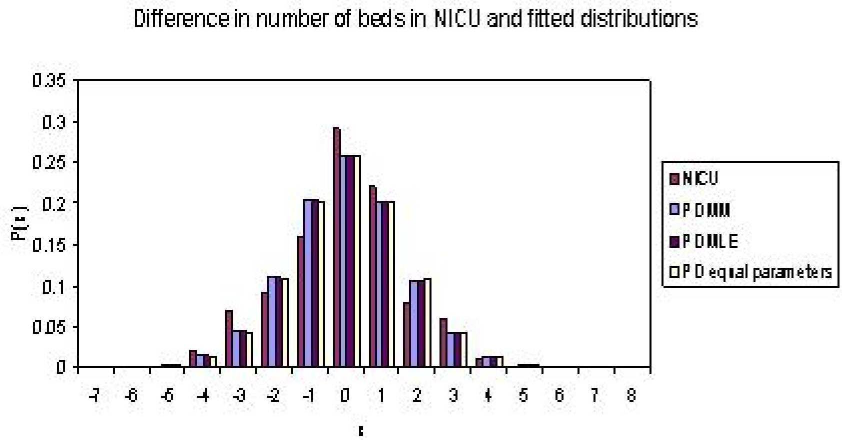

Figure 15. Relative frequency of occupied beds of NICU and fitted distributions

Interpretation:

If a is positive, then $P(X = a)$ represent the probability that the number of occupied beds in NICU increase by a during 24 hours, and if a is negative it is the probability that the number of occupied beds in NICU decrease by a during 24 hours, while a zero value of a gives the probability that the number of occupied beds remain the same during 24 hours.

References

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, 1992.

[2] J. L. Álvarez, Poisson-based detection limit and signal confidence intervals for few total counts, *Health Phys.* **93** (2007), 120–126.

[3] M. Fisz, The limiting distribution of the difference of two Poisson random variables, *Zastosowania Mat.* **1** (1953), 41–45.

[4] Y. Hwang, J. Kim and I. Kweon, Sensor noise modeling using the Skellam distribution: Application to the color edge detection, *IEEE Conference on Computer Vision and Patern Recognition* (2007), pp. 1–8.

[5] J. O. Irwin, The frequency distribution of the difference between two independent variates following the same Poisson distribution, *Journal of the Royal Statistical Society Series A* **100** (1937), pp. 415.

[6] N. L. Johnson, S. Kotz and A. W. Kemp, *Univariate Discrete Distributions*, Second edition, Wiley, New York, 1992.

[7] D. Karlis and I. Ntzoufras, Distributions based on Poisson differences with applications in sports, *Technical Report 101*, Department of Statistics, Athens University of Economics, 2000.

[8] D. Karlis and I. Ntzoufras, Bayesian analysis of the differences of count data, *Stat. Med.* **25** (2006), no. 11, 1885–1905.

[9] D. Karlis and I. Ntzoufras, Bayesian modeling of football outcomes: Using the Skellam's distribution for the goal difference, *IMA J. Management Math.*, (2008).

[10] S. K. Katti, Moments of the absolute difference and absolute deviation of discrete distributions, *Ann. Math. Statist.* **31** (1960), 78–85.

[11] A. Prékopa, On composed Poisson distributions. IV. Remarks on the theory of differential processes, *Acta Math. Acad. Sci. Hungar.* **3** (1952), 317–325 (1953).

[12] J. Romani, Distribution of the algebraic sum of Poisson variables, *Trabajos Estadist.* **7** (1956), 175–181.

[13] J. G. Skellam, The frequency distribution of the difference between two Poisson variates belonging to different populations, *J. Roy. Statist. Soc.* (N.S.) **109** (1946), 296.

[14] J. Strackee and J. J. D. van der Gon, The frequency distribution of the difference between two Poisson variates, *Statistica Neerlindica* **16** (1962), 17–23.