REVSTAT – Statistical Journal
Volume 21, Number 4, October 2023, 557–572
https://doi.org/10.57805/revstat.v21i4.435

Exponential-Gaussian Distribution and Associated Time Series Models

Authors: K.U. Nitha ✉
– Department of Statistics, Farook College (Autonomous), Kozhikode, Kerala, India
nithajilesh@gmail.com

S.D. Krishnarani
– Department of Statistics, University of Calicut, Malappuram, Kerala-673635, India
krishnaranisd@gmail.com

Received: July 2020 Revised: February 2022 Accepted: February 2022

Abstract:

- Exponential-Gaussian distribution has already appeared in the literature and it is widely used in many fields. In this paper, we study its application in time series through a model-based approach. An autoregressive process of order one with exponential-Gaussian distribution as marginals is introduced. Structural aspects of the innovation sequence is derived and analytical properties of the process are studied. Estimation of the parameters is done and the application is established through an illustration with real data.

Keywords:

- *exponential distribution; Gaussian distribution; autoregressive process; stationarity; convolution models.*

✉ Corresponding author.

1. INTRODUCTION

Convolution based models are introduced in the literature as a combination of two random variables. It is used in various fields like physics, engineering, biological studies, etc. Radovic *et al.* [22] studied breakdown voltage distribution in neon using the convolution of distributions. For applications of convolution models, one may refer to Burbeck and Luce [6], Rosso *et al.* [23] and Golubev [11]. Fajriyah [10] noted that beta convolutions and beta convoluted normal can be used in microarray experiments due to the presence of some non-biological noises. Plancade *et al.* [20] introduced gamma-normal convolution to model the background correction of the Illumina BeadArrays. Exponential-lognormal convolution is found to be a good fit for microarray data. Several convolution models based on different underlying distributions such as the exponential-gamma, the normal-gamma and the exponential-normal were studied in the past and different estimation methods of the parameters have been discussed and illustrated with real life data sets. Chen *et al.* [7] used exponential-gamma distribution and its highly skewed behavior for the improved estimation of the detection of differently expressed genes.

For the modeling and analysis of symmetrical and tailed peaks in the data, the exponential-Gaussian distribution has been used by several researchers in the past. The exponential-Gaussian distribution discussed by Xie *et al.* [26], is a convolution distribution for the observed gene expression intensities by assuming that the true signal intensities are exponentially distributed and the noise intensity is normally distributed. Ding *et al.* [9] used the exponential-normal convolution model to correct the background of the Illumina platform by using Markov chain Monte Carlo simulation. With the name exponentially modified Gaussian, the convolution of exponential and Gaussian distribution has been found as a good model in modeling chromatographic peaks as seen in Naish and Hartwell [19]. This distribution is used to model residuals in Ament *et al.* [2]. Application of this distribution in flow injection analysis, quantitation of chromatographic peaks etc. is explained in Jeansonne and Foley [14]. A recent work on exponential-Gaussian distribution is also seen in Jehan *et al.* [13].

Time series analysis using autoregressive models having non-normality assumptions had been an interesting area of researchers of all times. See Lawrance [17], Lawrance and Lewis [16], Popovici [21] and Billard [5] for the details of stationary autoregressive models under the assumption that the marginal distribution is exponential, and refer Sim [24] for gamma distributed marginals. The convolution distribution is relatively less explored in time series data analysis. In the regression context, one may refer to Gori and Rioul [12], where they estimated a linear bound in the presence of outliers under the assumption that the noise is exponential-Gaussian distributed. Also, it is of interest to study time series models developed under the assumption that data is exponential-Gaussian distributed. In this paper, we study the first order autoregressive time series models having exponential-Gaussian as marginals. The paper is systematically organized into various sections as follows.

In Section 2, we consider the probability density function (pdf) of the exponential-Gaussian distribution and bring out its analytical properties. The autoregressive process of order 1(AR(1)) with exponential-Gaussian distribution as marginals is introduced and the distribution of the innovation random variable is identified in Section 3. Important properties of the proposed model are derived in Section 4. The parameters involved in the proposed model are estimated using different methods and the performance of the same are verified using a simulation study in Section 5. Section 6 is devoted to the analysis of real data of GDP growth rate using the proposed model.

2. EXPONENTIAL-GAUSSIAN DISTRIBUTION

Let $U \in \mathbb{R}$ and $V \in \mathbb{R}^+$ be two independent and continuous random variables with pdfs $g(\cdot)$ and $h(\cdot)$ respectively. Then the pdf of the random variable $X = U + V$ is

$$f(x) = (g * h)(x) \ = \int_0^{+\infty} g(x-v)h(v)dv. \tag{2.1}$$

With particular choice of U as Gaussian with parameters μ and σ and V as exponential with mean λ in (2.1), the pdf of X is

$$f(x) = \frac{1}{\lambda\sigma\sqrt{2\pi}} \int_0^\infty e^{-\frac{(x-\mu-v)^2}{2\sigma^2}} e^{-\frac{v}{\lambda}} dv.$$

Using the erfc$(\cdot)$ function, Naish and Hartwell [19] expressed the above integral in a more convenient form as

$$f(x) \ = \ \frac{1}{2\lambda} e^{\frac{1}{\lambda}\left(\frac{\sigma^2}{2\lambda}+\mu-x\right)} \operatorname{erfc}\left(\frac{1}{\sqrt{2}\sigma}\left(\frac{\sigma^2}{\lambda}+\mu-x\right)\right), \tag{2.2}$$
$$-\infty < x < \infty, \lambda > 0, \mu \in \mathbb{R}, \sigma > 0,$$

where

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt.$$

We denote the exponential-Gaussian random variable having pdf (2.2) as EG(λ, μ, σ). A striking feature of such a construction is that, the resultant distribution is capable of capturing the skewed behaviour of the data. X being the sum of independent normal and exponential random variables, it is obvious that

$$E(X) \ = \ \mu + \lambda, \tag{2.3}$$

$$\operatorname{Var}(X) \ = \ \sigma^2 + \lambda^2, \tag{2.4}$$

$$\text{Skewness(X)} \ = \ \frac{2\lambda^3}{(\sigma^2+\lambda^2)^{3/2}} \tag{2.5}$$

and

$$\text{Kurtosis(X)} \ = \ 3\frac{(\sigma^4+2\lambda^2\sigma^2+3\lambda^4)}{(\sigma^2+\lambda^2)^2}. \tag{2.6}$$

When $\lambda \to 0$, the exponential-Gaussian becomes a Normal distribution with skewness zero and kurtosis value 3.
The characteristic function of the EG(λ, μ, σ) is given by

$$\phi_X(t) = \frac{e^{i\mu t - \frac{1}{2}t^2\sigma^2}}{1-\lambda i t}. \tag{2.7}$$

The shape of the pdf of EG(λ, μ, σ) random variable for various values of the parameter λ, by taking $\mu = 0$ and $\sigma = 1$, is depicted in Figure 1. The shape of EG(λ, μ, σ) is determined by the value of $k = \frac{\sigma}{\lambda}$. When $k \to 0$ the EG(λ, μ, σ) density function will be very close to the exponential density, and when k is very large, the distribution is close to the Gaussian distribution. The density plots reveal an apparent similarity in shape, but the peakedness increases significantly and becomes heavy tailed as λ increases.

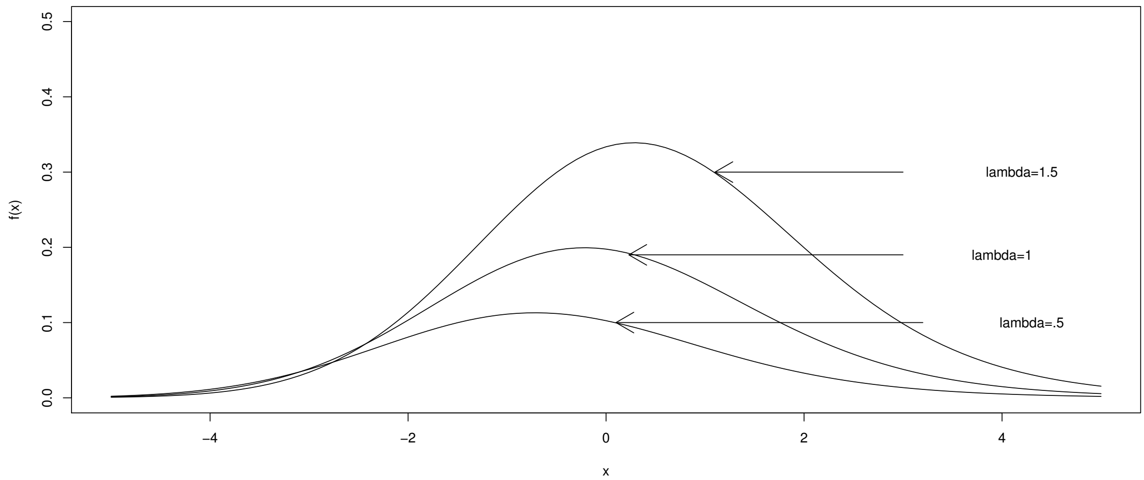

Figure 1: Shape of the density function of exponential-Gaussian for the different values of $\lambda \in \{0.5, 1, 1.5\}$, $\mu = 0$, $\sigma = 1$.

To discuss its application in time series, we propose an AR(1) process with EG(λ, μ, σ) distribution as marginals in the next section.

3. AR(1) MODEL WITH EXPONENTIAL-GAUSSIAN AS MARGINAL

Let $\{X_n\}$ be a first order autoregressive process having the linear structure

$$X_n = aX_{n-1} + \epsilon_n, \ |a| < 1. \tag{3.1}$$

Assume that $\{X_n\}$ is a stationary process with exponential-Gaussian distribution as marginals and $\{\epsilon_n\}$ is a sequence of independent and identically distributed (i.i.d) random variables independent of $\{X_t\}$, where $t < n$.
Since X_n's are stationary, by using the characteristic function of X_n we can write

$$\phi_{\epsilon_n}(t) = \frac{\phi_X(t)}{\phi_X(at)}. \tag{3.2}$$

Since X_n is following EG(λ, μ, σ) distribution, substituting (2.7) we obtain

$$\phi_{\epsilon_n}(t) = e^{it\mu(1-a) - \frac{\sigma^2 t^2 (1-a^2)}{2}} \left[\frac{1 - \lambda iat}{1 - \lambda it} \right]. \tag{3.3}$$

Using the expression of $\phi_{\epsilon_n}(t)$, we can represent the random variable $\{X_n\}$ as

$$X_n = aX_{n-1} + \begin{cases} Z_n, & \text{with probability } a, \\ W_n, & \text{with probability } 1-a, \end{cases} \tag{3.4}$$

where $Z_n \sim N(\mu(1-a), \sigma\sqrt{1-a^2})$ and $W_n \sim \text{EG}(\lambda, \mu(1-a), \sigma\sqrt{(1-a^2)})$. Alternatively (3.3) can be expressed as

$$\phi_{\epsilon_n}(t) = e^{it\mu(1-a)-\frac{\sigma^2 t^2(1-a^2)}{2}}\left[a + (1-a)\frac{1}{1-\lambda it}\right]. \tag{3.5}$$

Further it may be noted that, using the tailed exponential random variable of Littlejohn [18], X_n may be written as

$$X_n = aX_{n-1} + Y_{1n} + Y_{2n}, \tag{3.6}$$

where $Y_{1n} \sim N(\mu(1-a), \sigma\sqrt{1-a^2})$ and

$$Y_{2n} = \begin{cases} 0, & \text{with probability } a, \\ Exp(\lambda) & \text{with probability } 1-a, \end{cases} \tag{3.7}$$

and $Exp(\lambda)$ is the exponential distributed random variable with mean λ.

Now we define the first order exponential-Gaussian autoregressive process (EGAR(1)) as given below.

Definition 3.1. A Markovian sequence $\{X_n\}$ defined according to (3.1), is said to be an exponential-Gaussian autoregressive process of order 1 (EGAR(1)) with EG(λ, μ, σ) distribution as marginals if and only if $\{\epsilon_n\}$ admits the following representation

$$\epsilon_n = \begin{cases} Z_n, & \text{with probability } a, \\ W_n, & \text{with probability } 1-a, \end{cases} \tag{3.8}$$

where $Z_n \sim N(\mu(1-a), \sigma\sqrt{1-a^2})$ and $W_n \sim \text{EG}(\lambda, \mu(1-a), \sigma\sqrt{(1-a^2)})$.

From (3.8), the pdf of ϵ_n can be written as

$$f_{\epsilon_n}(x) = af_{Z_n}(x) + (1-a)f_{W_n}(x),$$

where

$$f_{Z_n}(x) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{1-a^2}} e^{\frac{-1}{2(1-a^2)}\left(\frac{x-\mu(1-a)}{\sigma}\right)^2},$$

$$f_{W_n}(x) = \frac{1}{2\lambda} e^{\frac{1}{\lambda}\left(\frac{\sigma^2(1-a^2)}{2\lambda}+\mu(1-a)-x\right)} \operatorname{erfc}\left(\frac{1}{\sqrt{2(1-a^2)}\sigma}\left(\frac{\sigma^2(1-a^2)}{\lambda} + \mu(1-a) - x\right)\right).$$

In the next section, we shall bring together the important properties of EGAR(1).

4. PROPERTIES OF THE EGAR(1) PROCESS

Proposition 4.1. *If X_0 is distributed arbitrarily, the Markovian process (3.1) is again exponential-Gaussian distributed asymptotically.*

Proof: We can rewrite $X_n = aX_{n-1} + \epsilon_n$, as

$$X_n = a^n X_0 + \sum_{k=0}^{n-1} a^k \epsilon_{n-k}.$$

Consequently, the characteristic function is

$$\begin{aligned} \phi_{X_n}(t) &= \phi_{X_0}(a^n t) \prod_{k=0}^{n-1} \phi_\epsilon(a^k t) \\ &= \phi_{X_0}(a^n t) \left[exp\left(i\mu(1-a) \sum_{k=0}^{n-1} a^k t - \frac{1}{2}\sigma^2(1-a^2) \sum_{k=0}^{n-1} a^{2k} t^2 \right) \right] \prod_{k=0}^{n-1} \frac{1 - ia^{k+1} t\lambda}{1 - ia^k t\lambda}. \end{aligned}$$

As $n \to \infty$,

$$\phi_{X_n}(t) \to e^{i\mu t - \frac{\sigma^2 t^2}{2}} \left[\frac{1}{1 - it\lambda} \right], \tag{4.1}$$

implying that X_n is asymptotically EG(λ, μ, σ) distributed. □

Proposition 4.2. *For the EGAR(1) process, the $(k+1)$ step ahead conditional mean is given by*

$$E(X_{n+k} | X_{n-1} = x_{n-1}) = a^{k+1} x_{n-1} + (1 - a^{k+1})(\lambda + \mu). \tag{4.2}$$

Proof: Using (3.1), we have

$$X_{n+k} = a^{k+1} X_{n-1} + a^k \epsilon_n + a^{k-1} \epsilon_{n+1} + \cdots + \epsilon_{n+k}. \tag{4.3}$$

By taking expectation conditionally on $X_{n-1} = x_{n-1}$ on both sides, we obtain the desired result. □

Remark 4.1. When $k \to \infty$,

$$E(X_{n+k} | X_{n-1} = x_{n-1}) \to \lambda + \mu, \tag{4.4}$$

which is the unconditional mean of the process.

Proposition 4.3. *For the EGAR(1) process, the $(k+1)$ step ahead conditional variance is given by*

$$\mathrm{Var}(X_{n+k} | X_{n-1} = x_{n-1}) = (1 - a^{2(k+1)})(\sigma^2 + \lambda^2). \tag{4.5}$$

Remark 4.2. As $k \to \infty$,

$$\text{Var}(X_{n+k}|X_{n-1} = x_{n-1}) \to (\sigma^2 + \lambda^2). \tag{4.6}$$

Proposition 4.4. *EGAR(1) process is not time-reversible.*

Proof: The joint characteristic function of (X_n, X_{n+1}) is

$$\begin{aligned}
\phi_{X_n,X_{n+1}}(t) &= E\big(e^{it_1X_n+it_2X_{n+1}}\big)\\
&= E\Big(e^{(it_1X_n+it_2(aX_n+\epsilon_{n+1})}\Big)\\
&= \phi_{X_n}(t_1+at_2)\phi_{\epsilon_{n+1}}(t_2)\\
&= e^{i\mu(t_1+t_2)-\frac{\sigma^2}{2}(t_1^2+t_2^2+2at_1t_2)}\frac{1-ia\lambda t_2}{(1-i\lambda t_2)(1-i\lambda(t_1+at_2))},
\end{aligned}$$

which is not symmetric in t_1 and t_2. So the process EGAR(1) is not time reversible. □

Remark 4.3. From the model defined in (3.1),

$$E(X_n|X_{n-1} = x) = ax + (1-a)(\lambda+\mu).$$

Therefore, we can see that regression in the forward direction is linear and the conditional variance is constant.

Following the steps in Lawrance [17], the joint moment generating function (m.g.f) of (X_n, X_{n+1}) is

$$M_{X_n,X_{n+1}}(t_1,t_2) = \frac{M_X(t_1+at_2)M_X(t_2)}{M_X(at_2)}. \tag{4.7}$$

Differentiating this with respect to t_1 and setting $t_1 = 0, t_2 = t$,

$$\begin{aligned}
E(e^{tX_{n+1}}E(X_n|X_{n+1})) &= \frac{M'_X(at)M_X(t)}{M_X(at)}\\
&= M'_X(at)M_\epsilon(t)\\
&= e^{t\mu+\frac{\sigma^2t^2}{2}}\left[\frac{\lambda a+(1-\lambda at)(\mu a+a^2\sigma^2t)}{(1-\lambda at)(1-\lambda t)}\right]. \qquad (4.8)
\end{aligned}$$

Also differentiating (4.7) with respect to t_2 and setting $t_2 = 0, t_1 = 0$, we get $E(X_n) = \lambda + \mu$.

Proposition 4.5. *The characteristic function of the partial sums $S_r = X_n + X_{n+1} + \cdots + X_{n+r-1}$ is*

$$\begin{aligned}
\phi_{S_r}(t) \;=\; &\left[exp\left(i\mu\frac{1-a^r}{1-a}t-\frac{\sigma^2}{2}\left(\frac{1-a^r}{1-a}\right)^2t^2\right)\right]\frac{1}{1-\lambda i\left(\frac{1-a^r}{1-a}\right)t}\\
&\cdot\prod_{j=1}^{r-1}\left[exp\left(i\mu(1-a^{r-j})t-\frac{\sigma^2}{2}(1-a^2)\left(\frac{1-a^{r-j}}{1-a}\right)^2t^2\right)\right]\frac{1-a\lambda i\left(\frac{1-a^{r-j}}{1-a}\right)t}{1-\lambda i\left(\frac{1-a^{r-j}}{1-a}\right)t}.
\end{aligned}$$

Proof:

$$S_r = X_n + X_{n+1} + \cdots + X_{n+r-1}$$
$$= \sum_{j=0}^{r-1} a^j X_n + \sum_{j=0}^{r-2} a^j \epsilon_{n+1} + \sum_{j=0}^{r-3} a^j \epsilon_{n+2} + \cdots + \epsilon_{n+r-1}$$
$$= X_n\left(\frac{1-a^r}{1-a}\right) + \sum_{j=1}^{r-1} \epsilon_{n+j}\left(\frac{1-a^{r-j}}{1-a}\right),$$

$$\phi_{S_r}(t) = \phi_{X_n}\left(\frac{1-a^r}{1-a}t\right)\prod_{j=1}^{r-1}\phi_\epsilon\left(\frac{1-a^{r-j}}{1-a}t\right)$$
$$= \left[exp\left(i\mu\frac{1-a^r}{1-a}t - \frac{\sigma^2}{2}\left(\frac{1-a^r}{1-a}\right)^2 t^2\right)\right]\frac{1}{1-\lambda i\left(\frac{1-a^r}{1-a}\right)t}$$
$$\prod_{j=1}^{r-1}\left[exp\left(i\mu(1-a^{r-j})t - \frac{\sigma^2}{2}(1-a^2)\left(\frac{1-a^{r-j}}{1-a}\right)^2 t^2\right)\right]\frac{1-a\lambda i\left(\frac{1-a^{r-j}}{1-a}\right)t}{1-\lambda i\left(\frac{1-a^{r-j}}{1-a}\right)t}.$$

On inverting the above expression of the characteristic function of S_r, one may obtain its distribution. □

5. ESTIMATION

In this section we will discuss the estimation of the parameters. The parameters involved in the process are μ, a, σ and λ. Let $(X_1, ..., X_n)$ be the realizations from the EGAR(1) process. Method of moments, conditional least square method, and Gaussian estimation method are discussed in the following sections. A simulation study is also conducted.

5.1. Estimation using the Method of Moments

Using (2.3), (2.4), and (2.5), we can identify the estimates for the parameters μ, σ and λ under the method of moments estimation. The autoregressive parameter a can be estimated by the sample autocorrelation function (ACF), that is $\hat{a} = corr(X_n, X_{n-1})$. Other moment estimates are given by

(5.1) $$\hat{\mu} = m - s\left(\frac{\gamma}{2}\right)^{1/3},$$

(5.2) $$\hat{\sigma^2} = s^2\left[1 - \left(\frac{\gamma}{2}\right)^{2/3}\right]$$

and

(5.3) $$\hat{\lambda} = s\left(\frac{\gamma}{2}\right)^{1/3},$$

where m is the sample mean, s is the sample standard deviation and γ is the skewness.

It may be noted that explicit expression for the mean and variance of the above estimators are not available.

5.2. Conditional Least Square Estimation

The conditional least square estimates of the parameters are obtained by minimizing the conditional sum of squares function

$$D_n(a,\mu,\sigma,\lambda)=\sum_{i=1}^{n}(x_i-E(X_i|X_{i-1}=x_{i-1}))^2. \tag{5.4}$$

From the linearity of the regression of EGAR(1) process, we have

$$E(X_i|X_{i-1}=x)=ax_{i-1}+(1-a)(\lambda+\mu). \tag{5.5}$$

Therefore, (5.4) can be written as

$$D_n(a,\mu,\sigma,\lambda)=\sum_{i=1}^{n}[x_i-ax_{i-1}-(1-a)(\lambda+\mu)]^2. \tag{5.6}$$

Solving the normal equations obtained from (5.6) we obtain estimates of a and μ in terms of $\hat{\lambda}$ as

$$\hat{a} = \frac{n\sum x_i x_{i-1}-\sum x_i\sum x_{i-1}}{n\sum x_{i-1}^2-(\sum x_i)^2}, \tag{5.7}$$

$$\hat{\mu} = \frac{\sum x_i-\hat{a}\sum x_{i-1}}{n(1-\hat{a})}-\hat{\lambda}. \tag{5.8}$$

Estimates of σ and λ can be identified numerically through other methods, like maximizing the conditional likelihood function, and also by making use of (5.7) and (5.8). The conditional likelihood function is given by

$$\begin{aligned}L(x;a,\mu,\sigma,\lambda) &= \left[\prod_{i=1}^{n} f_{X_i|X_{i-1}}(x_i|x_{i-1})\right] f_{X_0}(x_0)\\ &= \left[\prod_{i=1}^{n} f_{X_i|X_{i-1}}(x_i|x_{i-1})\right]\frac{1}{2\lambda}e^{\frac{1}{\lambda}\left(\frac{\sigma^2}{2\lambda}+\mu-x_0\right)}\\ &\quad\cdot\ \operatorname{erfc}\left(\frac{1}{\sqrt{2}\sigma}\left(\frac{\sigma^2}{\lambda}+\mu-x_0\right)\right),\end{aligned}$$

where

$$\begin{aligned}f_{X_i|X_{i-1}}(x_i|x_{i-1}) &= af_{Z_n}(x_i-ax_{i-1})+(1-a)f_{W_n}(x_i-ax_{i-1})\\ &= a\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma\sqrt{1-a^2}}e^{\frac{-1}{2}\frac{(x_i-ax_{i-1}-\mu(1-a))^2}{\sigma^2(1-a^2)}}\\ &\quad+\ (1-a)\frac{1}{2\lambda}e^{\frac{1}{\lambda}\left(\frac{\sigma^2(1-a^2)}{2\lambda}+\mu(1-a)-x_i+ax_{i-1}\right)}\\ &\quad\cdot\ \operatorname{erfc}\left(\frac{1}{\sqrt{2(1-a^2)}\sigma}\left(\frac{\sigma^2(1-a^2)}{\lambda}+\mu(1-a)-x_i+ax_{i-1}\right)\right).\end{aligned}$$

Since EGAR(1) is a stationary process and the moments are finite, using the regularity conditions of Klimko and Nelson [15], it is verified that the conditional least square estimators obtained are consistent and asymptotically normal. That is, $\sqrt{n}\Big[(\hat{a},\hat{\lambda})'-(a,\lambda)'\Big]\to N(0,\Sigma)$ where $N(0,\Sigma)$ is a bivariate normal distribution with mean 0 and dispersion matrix

$$\Sigma=\begin{bmatrix}(1-a^2) & 0\\ 0 & \frac{1+a}{1-a}\lambda^2\end{bmatrix}.$$

5.3. Gaussian Estimation Method

Whittle [25] introduced this method by taking, Gaussian likelihood function as the baseline distribution for the estimation. Later, Crowder [8] used this method of estimation for the analysis of correlated binomial data. Al-Nachawati *et al.* [1] and Alwasel *et al.* [3] used the same estimation procedure in the context of first order autoregressive process. Although this method has an approximate nature, this gives a good estimation to our model and also the possibility to estimate all the parameters in the model. The conditional maximum likelihood function is given by

$$L=f(x_1)\prod_{t=2}^{n}f(x_t|x_{t-1}). \tag{5.9}$$

Here $f(x_t|x_{t-1})$ and $f(x_1)$ are the conditional and marginal probability function of $X_t|X_{t-1}$ and X_t, respectively. We assume Gaussian pdf for $f(x_1)$ and $f(x_t|x_{t-1})$ with conditional mean and conditional variance as the parameters. Then the log-likelihood function can be written as

$$log(L)=n\;log\frac{1}{\sqrt{2\pi}}-\frac{1}{2}\sum_{t=2}^{n}\left(log(\sigma^2_{x_{t-1}})+\frac{(x_t-m_{x_{t-1}})^2}{\sigma^2_{x_{t-1}}}\right), \tag{5.10}$$

where $m_{x_{t-1}}=E(X_t|X_{t-1})=ax_{t-1}+(1-a)(\lambda+\mu)$ and $\sigma^2_{x_{t-1}}=\mathrm{Var}(X_t|X_{t-1})=(1-a^2)(\lambda^2+\sigma^2)$. So, the Gaussian log-likelihood function corresponding to $EGAR(1)$ process becomes

(5.11)
$$log(L)\;=\;n\;log\left(\frac{1}{\sqrt{2\pi}}\right)-\frac{1}{2}\sum_{t=2}^{n}\left[log\big((1-a^2)(\lambda^2+\sigma^2)\big)+\frac{\left(x_t-ax_{t-1}-(1-a)(\lambda+\mu)\right)^2}{(1-a^2)(\lambda^2+\sigma^2)}\right].$$

The Gaussian estimators are, thus, obtained by maximising the above non linear equation. But explicit expressions as the solution for the parameters a, λ, μ and σ are not available. Therefore, we have used numerical methods for identifying the value for these parameters. We use the `nlminb()` function in R with the Nelder–Mead method for this purpose. Crowder [8] pointed out that under Gaussian method of estimation of the parameter θ, $\sqrt{n}(\hat{\theta}-\theta)$ is asymptotically normally distributed with mean zero and asymptotic variance $[J(\theta)]^{-1}$, where $J(\theta)$ is the conditional expected information matrix. An approximation of the same can be done using the observed conditional information matrix, see Bakouch and Popovic [4].
To check the performance of the estimates, we have conducted a simulation study and the mean square error (MSE) is used for the comparison purpose.

5.4. Simulation Study

For checking the validity of the model, we simulated 100 samples of sizes 100, 500, 1000, 5000 and 10000 for different values of the parameters. The values considered are:

(**1**) $a = 0.1,\ \lambda = 1,\ \mu = 2$ and $\sigma = 1$;

(**2**) $a = 0.2,\ \lambda = 3,\ \mu = 5$ and $\sigma = 4$;

(**3**) $a = 0.5,\ \lambda = 8,\ \mu = 10$ and $\sigma = 8$.

We obtained the estimates of the parameters and corresponding MSE values and the results are presented in Table 1. It can be seen from the table that the Gaussian estimators are very close to the true value of the parameters and it shows better performance when the sample size increases. Further it may be noted that the MSE values decrease as the sample size increases for different parameter values.

Table 1: Estimated values of a, λ, μ and σ corresponding mean squared error (MSE).

True values are: $a = 0.1,\ \lambda = 1,\ \mu = 2,\ \sigma = 1$								
Sample size	$\widehat{a}$	$\widehat{\lambda}$	$\widehat{\mu}$	$\widehat{\sigma}$	MSE($\widehat{a}$)	MSE($\widehat{\lambda}$)	MSE($\widehat{\mu}$)	MSE($\widehat{\sigma}$)
100	0.1371	0.9158	1.8221	1.1812	0.0613	0.1514	0.1034	0.1588
500	0.1158	0.9821	1.9948	1.1609	0.0227	0.0863	0.0949	0.1368
1000	0.1117	0.9969	1.9951	1.2321	0.0184	0.0546	0.0602	0.1211
5000	0.1041	1.0028	1.9934	1.0860	0.0124	0.0299	0.0287	0.0947
10000	0.1030	1.0003	2.0002	1.0636	0.0055	0.0269	0.0272	0.0782

True values are: $a = 0.2,\ \lambda = 3,\ \mu = 5,\ \sigma = 4$								
Sample size	$\widehat{a}$	$\widehat{\lambda}$	$\widehat{\mu}$	$\widehat{\sigma}$	MSE($\widehat{a}$)	MSE($\widehat{\lambda}$)	MSE($\widehat{\mu}$)	MSE($\widehat{\sigma}$)
100	0.1895	3.1918	5.1041	3.7119	0.0629	0.4664	0.2149	0.5053
500	0.1908	2.9903	4.9784	3.7372	0.0369	0.3011	0.1347	0.4146
1000	0.1946	2.9960	4.9863	3.7753	0.0299	0.1769	0.0738	0.3238
5000	0.1980	2.9986	4.9975	3.8110	0.0261	0.1369	0.0573	0.3009
10000	0.2004	3.0001	4.9993	3.8410	0.0095	0.0605	0.0393	0.2363

True values are: $a = 0.5,\ \lambda = 8,\ \mu = 10,\ \sigma = 8$								
Sample size	$\widehat{a}$	$\widehat{\lambda}$	$\widehat{\mu}$	$\widehat{\sigma}$	MSE($\widehat{a}$)	MSE($\widehat{\lambda}$)	MSE($\widehat{\mu}$)	MSE($\widehat{\sigma}$)
100	0.4842	7.8667	10.1034	7.5170	0.0631	0.8189	0.5678	1.0678
500	0.4951	7.9766	9.9780	7.6179	0.0383	0.6861	0.4468	0.9906
1000	0.4973	7.9827	9.9863	7.8915	0.0288	0.4723	0.3221	0.8955
5000	0.4985	7.9930	9.9945	7.8955	0.0120	0.2179	0.1620	0.8915
10000	0.5001	7.9989	10.0005	7.9301	0.0080	0.1339	0.1059	0.8207

Next we shall look into the sample path behaviour of the EGAR(1) process. We simulated 500 observations from the proposed process by taking $a = 0.5,\ \mu = 0,\ \sigma = 1$ and $\lambda = 0.2, 0.4, 0.6, 0.8,$ and the same is plotted in Figure 2. The sample path clearly shows that the simulated data is stationary.

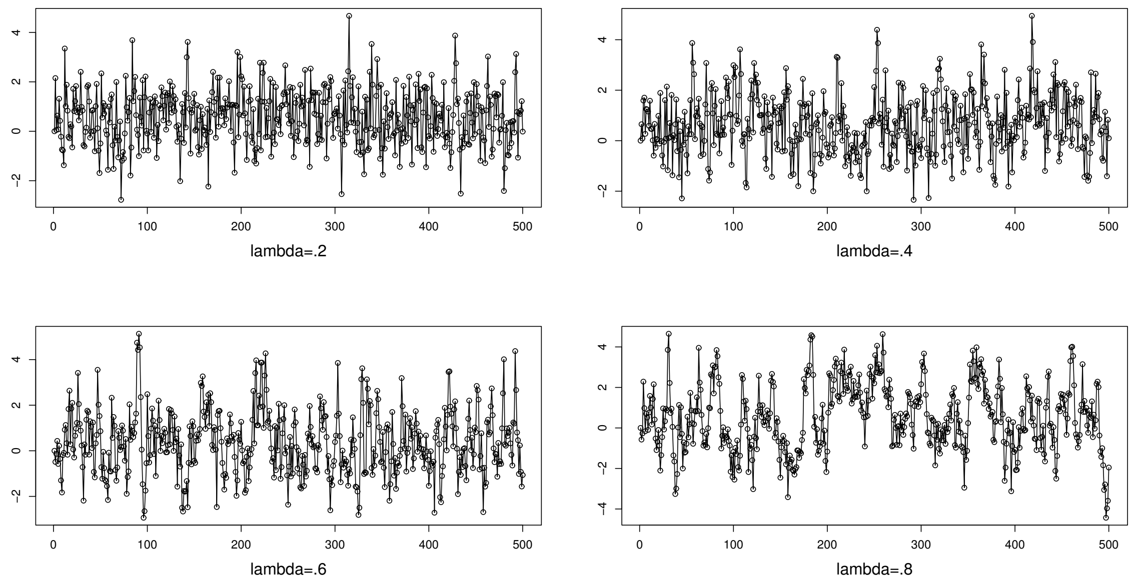

Figure 2: Sample path for the values of lambda=0.2, 0.4, 0.6, 0.8, μ=0 and σ=1.

6. DATA ANALYSIS

For establishing the applicability of the model, we considered the US GDP growth rate data for the period from 1961 to 2018 which is available in `https://data.worldbank.org`. The exponential-Gaussian distribution is found to be a suitable distribution for the data and the fitted density is ploted in Figure 3. We performed the Kolmogorov–Smirnov test of goodness of fit for the data set to check the adequacy of the exponential-Gaussian distribution and obtained the p-value as $0.28 > 0.05$. Significantly high p-value indicates the acceptance of the hypothesis that EG is a good fit to the data.

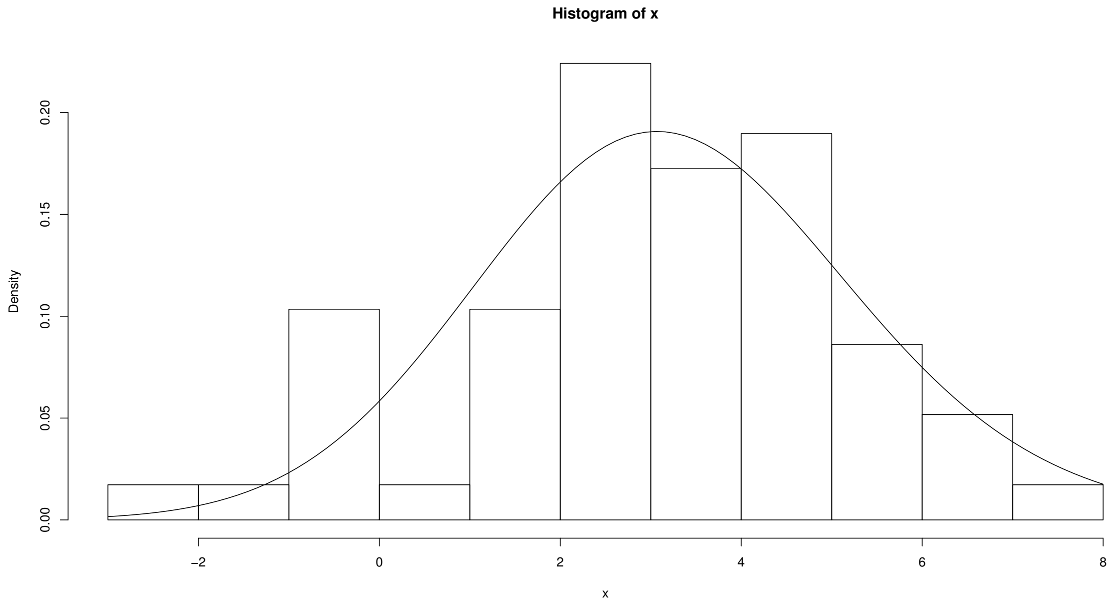

Figure 3: Fitted Exponential-Gaussian distribution for the GDP data.

Hence we tried to model the data using the proposed EGAR(1) model. The time series plot, the plots of the ACF and the partial autocorrelation function (PACF) of the data are presented in Figure 4. We can find that the ACF and PACF is significant only at lag 1.

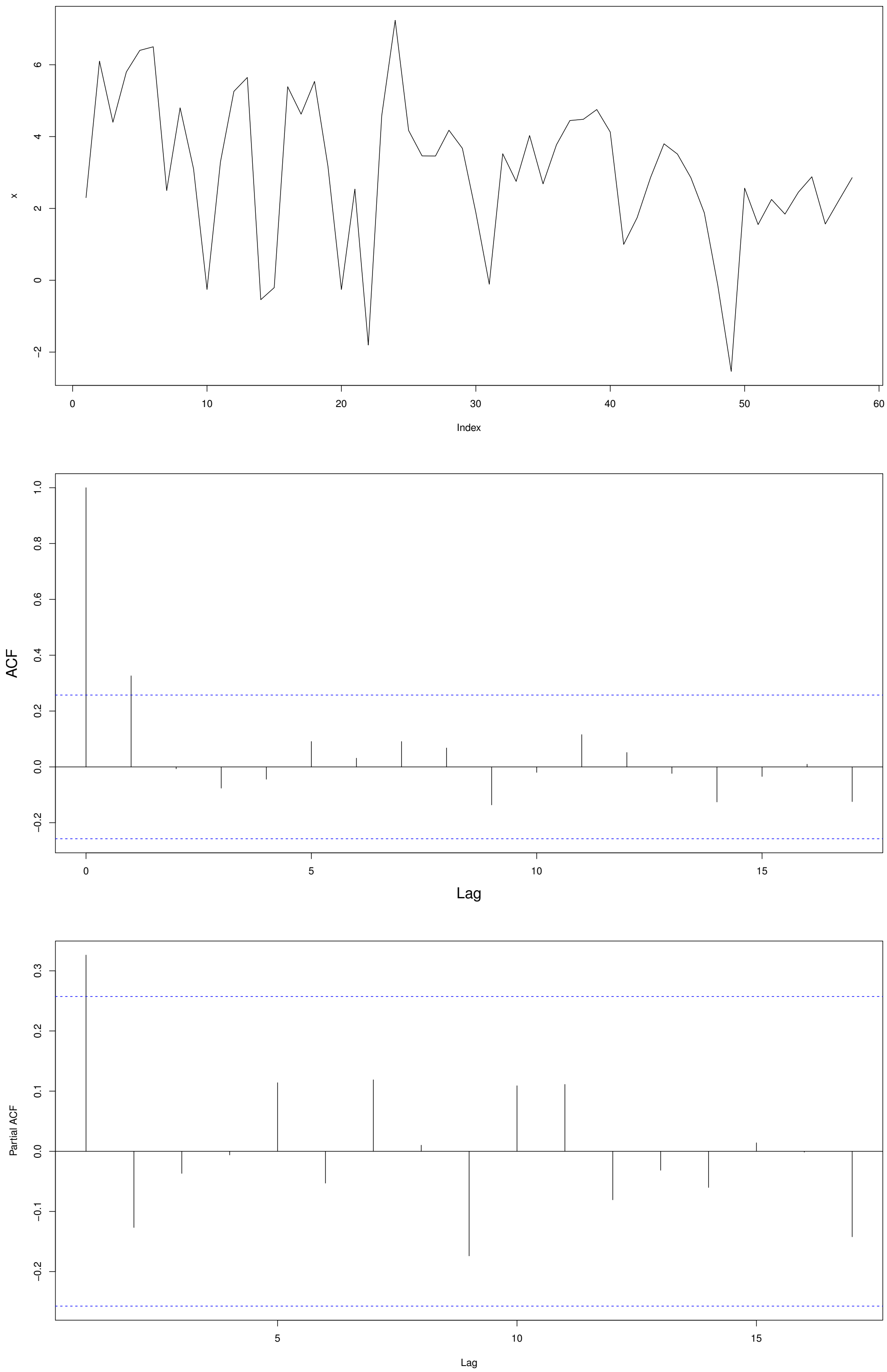

Figure 4: GDP data, ACF, PACF plots.

Therefore, we use AR(1) model for this data. The Gaussian estimation discussed in Section 5.3 is performed and the values of the parameters of EGAR(1) process are obtained as $\widehat{a} = 0.33$, $\widehat{\lambda} = 1.88$, $\widehat{\mu} = 1.18$ and $\widehat{\sigma} = 0.19$. Residual analysis has been carried out and the model adequacy has been checked. The p-value of Ljung–Box test is $0.79 > 0.05$, accepting the null hypothesis that the residuals are white noise. Also, the ACF and PACF of the residual are within the limits as represented in Figure 5. We have calculated the standard errors of the estimates using the Hessian matrix and got the standard errors of a as 0.05, μ as 0.08 and λ as 0.08.

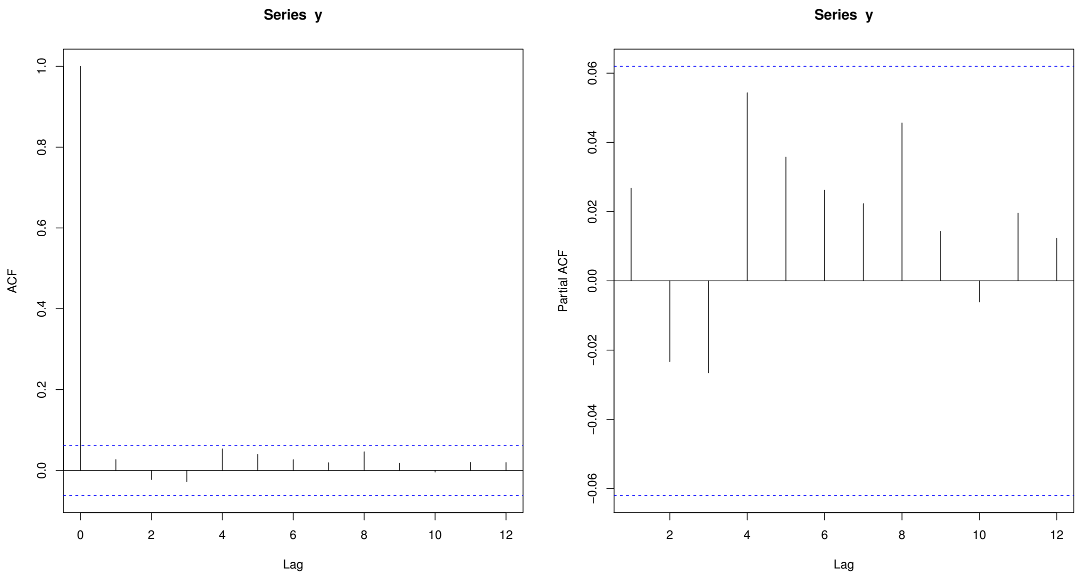

Figure 5: ACF and PACF plots of residuals.

Due to the evaluation of the likelihood function outside the given range of the parameter values, the standard error of σ is not evaluated correctly. We have predicted the GDP values for the next years and plotted them in Figure 6. In particular, note that the predicted value of GDP growth rate in the year 2019 was 2.32, where the actual value is 2.161.

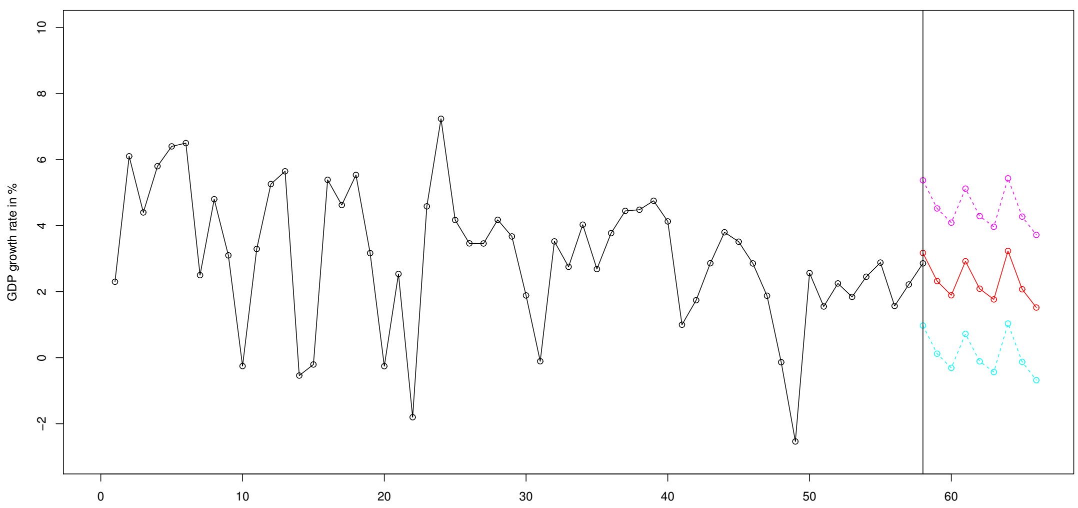

Figure 6: Prediction for the GDP data.

7. CONCLUSION

In this paper, we studied exponential-Gaussian distribution as a suitable model for data having symmetry or heavy tailed behaviour. The time series application of the EG distribution has been explored using an AR(1) process. The estimation of the model parameters and the problem of fitting of the model to real time series data and simulated data are perused. Application of the similar convolution model like Lindley-Gaussian, Gamma-Gaussian etc. are themes for future works. It may be interesting to investigate non-linear time series models and stochastic volatility models based on EG distribution.

ACKNOWLEDGMENTS

The authors express their sincere thanks to the editor-in-chief and anonymous reviewers for their constructive comments which improved the quality of the paper. Also we would like to acknowledge DST, India for the financial assistance through the FIST programme.

REFERENCES

[1] Al-Nachawati, H.; Alwasel, I. and Alzaid, A.A. (1997). Estimating the parameters of the generalized Poisson AR(1) process, *Journal of Statistical Computation and Simulation*, **56**, 337–352.

[2] Ament, S.; Gregoire, J. and Gomes, C. (2019). Exponentially-Modified Gaussian mixture model: Applications in Spectroscopy, *ArXiv preprint arXiv:1902.05601 (arxiv.org).*

[3] Alwasel, I.; Alzaid, A. and Al-Nachawati, H. (1998). Estimating the parameters of the binomial autoregressive process of order one, *Applied Mathematics and Computation*, **95**, 193–204.

[4] Bakouch, H.S. and Popovic, B.V. (2016). Lindley first-order autoregressive model with applications, *Communications in Statistics – Theory and Methods*, **45**(17), 1–20.

[5] Billard, L. (2014). Exponential autoregressive models, *Wiley StatsRef: Statistics Reference Online*, John Wiley & Sons Ltd.

[6] Burbeck, S.L. and Luce, R.D. (1982). Evidence from auditory simple reaction times for both change and level detectors, *Perception and Psychophysics*, **32**(2), 117–133.

[7] Chen, M.; Xie, Y. and Story, M. (2011). An exponential-gamma convolution model for background correction of Illumina BeadArray data, *Communications in Statistics – Theory and Methods*, **40**(17), 3055–3069.

[8] Crowder, M. (1985). Gaussian estimation for correlated binomial data, *Journal of Royal Statistical Society*, **47**(2), 229–237.

[9] Ding, L.H.; Xie, Y.; Park, S.; Xiao, G. and Story, M.D. (2008). Enhanced identification and biological validation of differential gene expression via illumina hole genome expression arise through the use of the model based background correction methodology, *Nucleic Acids Research*, **36**, 10:e58.

[10] Fajriyah, R. (2016). A Study of convolution models for background correction of BeadArrays, *Austrian Journal of Statistics*, **45**, 15–33.

[11] Golubev, A. (2010). Exponentially modified Gaussian(EMG) relevance to distributions related to cell proliferation and differentiation, *Journal of Theoretical Biology*, **262**, 257–266.

[12] Gori, J. and Rioul, O. (2019). Regression to a linear lower bound with outliers an exponentially modified Gaussian noise model, *Eusipco 2019 27th European Signal Processing Conference*, Sep. 2019, A Coruna, Spain.

[13] Jehan, A.; Sajid A. and Ismail, S. (2020). Monitoring schedule time using exponentially modified Gaussian distribution, *Quality Technology & Quantitative Management*, **17**(4), 448–469.

[14] Jeansonne, M.S. and Foley, J.P. (1991). Review of the exponentially modified Gaussian (EMG) Function Since 1993, *Journal of Chromatographic Science*, **29**(6), 258–266.

[15] Klimko, L.A. and Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes, *The Annals of Statistics*, **6**(3), 629–642.

[16] Lawrance, A.J. and Lewis, P.A.W. (1981). A new autoregressive time Series model in exponential variables (NEAR(1)), *Advances in Applied Probability*, **13**(04), 826–845.

[17] Lawrance, A.J. (1978). Some autoregressive models for point processes, *Colloquia Mathematica Societatis Janos Bolyai*, **24**, 257–275.

[18] Littlejohn, R.P. (1994). A reversibility relationship for two markovian time series models with stationary exponential tailed distribution, *Journal of Applied Probability*, **31**, 575–581.

[19] Naish, P.J. and Hartwell, S. (1988). Exponentially modified Gaussian functions – a good model for chromatographic peaks in isocratic HPLC, *Chromatographia*, **26**, 285–296.

[20] Plancade, S.Y.; Rozenholc, Y. and Lund, E. (2012). Generalization of the normal-exponential model: exploration of a more accurate parametrisation for the signal distribution on Illumina BeadArrays, *BMC Bioinfomatics*, **78**, 493–506.

[21] Popovici, G. (2010). On three classes of time series involving exponential distribution, *Mathematical Reports*, **12**(1), 45–57.

[22] Radovic, M.K.; Maluckov, C.A.; Karamarkovic, J.P.; Rancev, S.A. and Mitic, S.D. (2014). Convolution based model of breakdown voltage distributions in neon at 1.33 mbar with corona appearance in pre-breakdown regime, *Romanian Reports in Physics*, **66**(2), 472–480.

[23] Rosso, L.; Cathryn, S.B.; James, M.G.; Azeem, S.; Patricia, M.P.; Federico, E.T. and Eric, O.A. (2009). A new model for prediction of drug distribution in tumor and normal tissues: Pharmacokinetics of temozolomide in glioma patients, *American Association for Cancer Research*, 120–127.

[24] Sim, C.H. (1994). Modelling non-normal first-order autoregressive time series, *Journal of Forecasting*, **13**, 369–381.

[25] Whittle, P. (1961). Gaussian estimation in stationary time series, *Bulletin of the International Statistical Institute*, **39**, 105–130.

[26] Xie, Y.; Wang, X. and Story, M. (2009). Statistical methods of background correction for Illumina BeadArray data, *Bioinformatics*, **25**(6), 751–757.