# Egyptian Neonatal Mortality Rate Analysis

# between 1960 till 2016 Using a Suggested

# Nonparametric Lomax Density Estimation

**Samah M. Abo-El-Hadid**

Department of Mathematics, Insurance and Applied Statistics
Faculty of Commerce and Business Administration
Helwan University, Cairo, Egypt

## Abstract

Neonatal mortality rate Measures the social, economic, health care and environmental conditions in which children live. In this paper, the density function of neonatal mortality rate in Egypt is estimated nonparametrically using the kernel density estimation method. Suggested kernel density estimator for nonnegative random variables is introduced using the Lomax density as a kernel function. The asymptotic bias, variance, mean squared error (MSE), integrated mean squared error (IMSE), and the optimal bandwidth of the proposed Lomax estimator are investigated. We also introduce a simulation study to compare the proposed estimator with other estimators.

## 1. Introduction

Neonatal mortality rate is the number of neonates dying before reaching 28 days of age, per 1,000 live births in a given year. In this paper the density function of neonatal mortality rate in Egypt is estimated nonparametrically using the kernel density. The kernel density estimator method was introduced by Rosenblatt (1956). He considered $\hat{f}(x)$ as an estimator of the unknown density $f(x)$:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \qquad , -\infty \leq x \leq \infty \qquad (1)$$

Here $n$ is the sample size; $K(\cdot)$ and $h$ are the kernel function and the bandwidth respectively, where the kernel function $K(\cdot)$ is assumed to be symmetric density function.

Parzen (1962) studied the statistical properties of the symmetric kernel density estimator in equation (1), and proved that $\hat{f}(x)$ is biased and consistent estimator. He also obtained the optimal bandwidth $h_{opt}$ which minimizes the integrated mean squared error of $\hat{f}(x)$. Bagai and Rao (1995) studied the statistical properties of the kernel density estimator in the case of using asymmetric kernel function; also they found the optimal kernel function and the optimal bandwidth $h_{opt}$ which minimizes the integrated mean squared errors (IMSE). Chen (1999) proposed using the density of Beta distribution as the kernel function when $x \in [0,1]$. Chen (2000) suggested using the density of Gamma distribution as the kernel function for density estimation when $x \in [0, \infty)$. Scaillet (2004) used inverse Gaussian and reciprocal inverse Gaussian probability density functions as kernels for densities defined on [0;+1) support. Bouezmarni and Scaillet (2005) studied the consistency of both, the asymmetric kernel density estimator and the smoothed histogram. They proved that they both have good finite sample properties. Bouezmarni et al. (2011) suggested using the gamma kernels for the density and the hazard rate functions for right censored data; they also studied IMSE, the asymptotic normality and the law of iterated logarithm of this estimator. Markovich (2016) introduced new kernel estimator as a combination of the asymmetric gamma and Weibull kernels, also the theoretical asymptotic properties of the proposed density estimator and the optimal bandwidth selection for the estimate as a MISE are derived. Abo-El-Hadid (2019) suggested using the Rayleigh distribution as a kernel function and studied the statistical properties of this estimator and obtained the optimal bandwidth.

The rest of this paper is organised as follows: In Section 2, we outline the framework of the Lomax kernel density estimator. In Section 3 the bandwidth selection problem is discussed. Section 4 provides the results of a simulation study in which the behaviour of the Lomax kernel estimator is compared with the other density estimators. The density of the Egyptian Neonatal mortality rate data is estimated in section 5. Finally, in section 6, a brief conclusion is provided.

## 2.    The Lomax Kernel

Let $x_1, \ldots, x_n$ be a random sample from a distribution with an unknown density function $f(x)$. We propose the use of Lomax kernel for density estimation in this paper.  The Lomax kernel function is defined as:

$$K(u) = \frac{\alpha}{\lambda}\left[1 + \frac{u}{\lambda}\right]^{-\alpha-1} \qquad\qquad , \ u \geq 0, \ \ \alpha, \lambda > 0 \qquad\qquad (2)$$

where

$$E(u) = \int_0^\infty uK(u)\,du = \frac{\lambda}{\alpha-1} \qquad\qquad (3)$$

$$E(u^r) = \frac{\lambda^r \lceil(\alpha-r)\lceil(1+r)}{\lceil r} \qquad\qquad (4)$$

Where $\lambda$ is the scale parameter, and $\alpha$ is the shape parameter. Then the Lomax kernel density estimator is as follows:

$$\hat{f}(x) = \frac{1}{nh\lambda}\sum_{i=1}^{n}\left[1 + \frac{\left(\frac{x-x_i}{h}\right)}{\lambda}\right]^{-\alpha-1} \qquad , \frac{x-x_i}{h} \geq 0 \quad (5)$$

It can be shown that the expectation of the kernel density estimator is:

$$E[\hat{f}(x)] = E\left[\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right)\right] = \frac{1}{h}\int K\left(\frac{x-x_i}{h}\right)f(x)dx \qquad (6)$$

Let $u = \frac{x-x_i}{h}$

$$\therefore E[\hat{f}(x)] = \frac{1}{h}\int K(u)f(x-uh)hdu \qquad (7)$$

Using Taylor expansion for f(x − uh) yields that:

$$Bias[\hat{f}(x)] \simeq -hf'(x)\int uK(u)\,du + \frac{h^2}{2}f''(x)\int u^2K(u)\,du \qquad (8)$$

Note that for the Lomax kernel function from equations (3) and (4):

$$\int_0^\infty uK(u)\,du = \frac{\lambda}{\alpha-1}, \quad \int_0^\infty u^2K(u)\,du = \frac{2\lambda^2}{\alpha^2-3\alpha+2} \qquad (9)$$

Then the asymptotic bias of the Lomax kernel density estimator is:

$$Bias[\hat{f}(x)] \simeq \frac{-h\lambda}{\alpha-1}f'(x) + \frac{h^2\lambda^2}{\alpha^2-3\alpha+2}f''(x) \qquad (10)$$

Also, it can be shown that the variance of the kernel density estimator is:

$$Var[\hat{f}(x)] = \frac{1}{nh^2}\left\{E\left[K\left(\frac{x-x_i}{h}\right)\right]^2 - \left[EK\left(\frac{x-x_i}{h}\right)\right]^2\right\} \qquad (11)$$

$$= \frac{1}{nh^2}\left\{\int\left[K\left(\frac{x-x_i}{h}\right)\right]^2 f(x_i)dx_i - \left[\int K\left(\frac{x-x_i}{h}\right)f(x_i)dx_i\right]^2\right\}$$

let $u = \frac{x-x_i}{h}$, then

$$Var[\hat{f}(x)] = \frac{1}{nh^2}\{h\int K^2(u)f(x-uh)du - [h\int K(u)f(x-uh)du]^2\}$$

Again by Taylor expansion: $f(x-uh) = f(x) - uhf'(x) + \frac{u^2h^2}{2!}f''(x) + \cdots$

$$\therefore Var[\hat{f}(x)] \simeq \frac{1}{nh}f(x)\int K^2(u)du \qquad (12)$$

Using the Lomax kernel function:

$$\int_0^\infty K^2(u)\,du = \int_0^\infty \frac{\alpha^2}{\lambda^2}\left[1 + \frac{u}{\lambda}\right]^{-2\alpha-2}du \qquad (13)$$

let $z = 1 + \frac{u}{\lambda}$, then: $\frac{dz}{du} = \frac{1}{\lambda}$ , $du = \lambda\,dz$

$$\therefore \int_0^\infty \frac{\alpha^2}{\lambda^2}\left[1 + \frac{u}{\lambda}\right]^{-2\alpha-2}du = \int_0^\infty \frac{\alpha^2}{\lambda^2}z^{-2\alpha-2}\lambda\,dz \qquad (14)$$

$$= \frac{\alpha^2}{\lambda}\int_0^\infty z^{-2\alpha-2}\,dz = \frac{\alpha^2}{\lambda}\left[\frac{z^{-2\alpha-1}}{-2\alpha-1}\right]_0^\infty \qquad (15)$$

Undo substitution $z = 1 + \frac{u}{\lambda}$

$$\int_0^\infty K^2(u)\,du = \frac{\alpha^2}{\lambda}\left[\frac{\left(1+\frac{u}{\lambda}\right)^{-2\alpha-1}}{-2\alpha-1}\right]_0^\infty \qquad (16)$$

$$= \left[\frac{\alpha^2}{\lambda(-2\alpha-1)\left(1+\frac{u}{\lambda}\right)\left(1+\frac{u}{\lambda}\right)^{2\alpha}}\right]_0^\infty \qquad (17)$$

$$= \left[-\frac{\alpha^2}{\lambda(2\alpha+1)(u+\lambda)\left(1+\frac{u}{\lambda}\right)^{2\alpha}}\right]_0^\infty \tag{18}$$

$$\int_0^\infty K^2(u)\,du = \frac{\alpha^2}{2\alpha\lambda+\lambda} \tag{19}$$

Substitute equation (19) into equation (12), then the asymptotic variance of the Lomax kernel density estimator is:

$$Var[\hat{f}(x)] = \frac{\alpha^2\,f(x)}{nh(2\alpha\lambda+\lambda)} \tag{20}$$

Combining (10) and (20), the mean squared errors for $\hat{f}(x)$ is:

$$MSE[\hat{f}(x)] = Var[\hat{f}(x)] + Bias^2[\hat{f}(x)] = \frac{\alpha^2\,f(x)}{nh(2\alpha\lambda+\lambda)} + \frac{h^2\lambda^2}{(\alpha-1)^2}\left(f'(x)\right)^2 +$$
$$o(h^2) \quad (21)$$

Where $o(h^2)$ higher than second order terms of $h$. Also, the asymptotic IMSE for $\hat{f}(x)$ is:

$$IMSE[\hat{f}(x)] \simeq \frac{\alpha^2}{nh(2\alpha\lambda+\lambda)} + \frac{h^2\lambda^2}{(\alpha-1)^2}\int\left(f'(x)\right)^2\,dx \tag{22}$$

## 3.    The optimal bandwidth

The optimal bandwidths which minimize the IMSE for $\hat{f}(x)$ is obtained as follows:

$$\frac{\partial\,IMSE[\hat{f}(x)]}{\partial h} = -\frac{\alpha^2}{nh^2(2\alpha\lambda+\lambda)} + \frac{2h\lambda^2}{(\alpha-1)^2}\int\left(f'(x)\right)^2\,dx = 0 \tag{23}$$

$$\therefore h_{opt} = \left[\frac{2n\lambda^2(2\alpha\lambda+\lambda)}{\alpha^2(\alpha-1)^2}\int\left(f'(x)\right)^2\right]^{-1/3} \tag{24}$$

Now let us replace the unknown term $\int\left(f'(x)\right)^2$ in (24) by the Lomax density as a reference distribution. Let:

$$f(x) = \frac{\alpha}{\lambda}\left[1+\frac{x}{\lambda}\right]^{-\alpha-1} \qquad ,\ x \geq 0,\ \alpha,\lambda > 0 \tag{25}$$

$$\therefore f'(x) = \frac{\alpha(-\alpha-1)}{\lambda}\left[1+\frac{x}{\lambda}\right]^{-\alpha-2}\left(\frac{1}{\lambda}\right) \tag{26}$$

$$\therefore f'(x) = \frac{\alpha(-\alpha-1)}{\lambda^2}\left[1+\frac{x}{\lambda}\right]^{-\alpha-2} \tag{27}$$

and hence

$$\int\left(f'(x)\right)^2\,dx = \int_0^{-\infty}\frac{\alpha^2(-\alpha-1)^2}{\lambda^4}\left[1+\frac{x}{\lambda}\right]^{-2\alpha-4} \tag{28}$$

Again let $z = 1 + \frac{x}{\lambda}$, then $dx = \lambda\,dz$

$$\therefore \int\left(f'(x)\right)^2\,dx = \int_0^{-\infty}\frac{\alpha^2(-\alpha-1)^2}{\lambda^4}z^{-2\alpha-4}\ \lambda\,dz \tag{29}$$

$$= \left[\frac{\alpha^2(-\alpha-1)^2\ z^{-2\alpha-3}}{\lambda^3(-2\alpha-3)}\right]_0^\infty \tag{30}$$

Undo substitution $z = 1 + \frac{x}{\lambda}$

$$\therefore \int\left(f'(x)\right)^2\,dx = \left[\frac{\alpha^2(-\alpha-1)^2}{\lambda^3(-2\alpha-3)\left(1+\frac{x}{\lambda}\right)^{2\alpha+3}}\right]_0^\infty \tag{31}$$

$$= \left[\frac{\alpha^2(\alpha+1)^2}{\lambda^3(2\alpha+3)}\right] \tag{32}$$

Substituting (32) into (24), we get

$$h_{opt} = \left[\frac{2n\lambda^2(2\alpha\lambda+\lambda)}{\alpha^2(\alpha-1)^2} \cdot \frac{\alpha^2(\alpha+1)^2}{\lambda^3(2\alpha+3)}\right]^{-1/3} \tag{33}$$

$$= \left[\frac{2n\lambda^2(2\alpha\lambda+\lambda)(\alpha+1)^2}{\lambda(\alpha-1)^2(2\alpha+3)}\right]^{-1/3} \tag{34}$$

## 4.    Simulation

In this section, the influence of Lomax kernel estimator is examined using a simulation study. The Lomax kernel was given in equation (2). The Lomax kernel is compared with the most widely used kernel functions:

1) The Gaussian kernel which is symmetric about zero:

$$K(u) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{\left(-\frac{1}{2}u^2\right)}, \qquad\qquad 0 < u < \infty$$

2) The Gamma kernel which is asymmetric kernel:

$$K(u) = \frac{\beta^r u^{r-1} e^{-\beta u}}{\Gamma r} \qquad\qquad r, \beta > 0, \quad 0 < u < \infty$$

To evaluate the suggested Lomax estimator, we select the exponential random variable with ($\theta = .5$, $\theta =1$, and $\theta =2$). We generate the exponential i.i.d samples with sample sizes $n \in \{10, 100, 1000, \}$. For each simulated sample, the optimal bandwidths are calculated for each distribution. Then, the actual and the estimated densities are plotted; and the following errors' measures are computed:

$$\text{Mean squared error (MSE)} = \frac{\sum_{i=1}^{n}\left(f(x_i)-\hat{f}(x_i)\right)^2}{n} \tag{35}$$

$$\text{Mean absolute error (MAE)} = \frac{\sum_{i=1}^{n}|f(x_i)-\hat{f}(x_i)|}{n} \tag{36}$$

$$\text{Mean absolute percentage error (MAPE)} = \sum_{i=1}^{n}\left|\frac{f(x_i)-\hat{f}(x_i)}{n \cdot f(x_i)}\right| \tag{37}$$

These above measures are used to compare the fits obtained by different kernels. For all three measures, smaller values indicate a better fitting model. MSE is commonly-used measure of accuracy of fitted values but it is highly affected by outliers than MAE. MAPE expresses accuracy as a percentage of the error

The values of the above goodness of fit measures are given in tables (1,2,3) below:

Table (1): Goodness of fit measure's of the difference between the actual density and the estimated densities with $\theta = .5$

| Sample Size | Measure | Estimated density | | |
|---|---|---|---|---|
| | | Gaussian kernel | Gamma kernel | Lomax |
| $n = 10$ | MSE | 0.115271 | 0.0835603 | 0.0108429 |
| | MAE | 0.286971 | 0.243102 | 0.0792544 |
| | MAPE | 0.735817 | 0.67068 | 0.196878 |
| $n=100$ | MSE | 0.0659428 | 0.0501063 | 0.0166454 |
| | MAE | 0.199204 | 0.166958 | 0.0869625 |
| | MAPE | 1.22588 | 0.78857 | 0.469903 |
| $n=1000$ | MSE | 0.0713222 | 0.0532124 | 0.0152329 |
| | MAE | 0.210702 | 0.174348 | 0.0785587 |
| | MAPE | 1.06445 | 0.753061 | 0.237755 |

Table (2): Goodness of fit measure's of the difference between the actual density and the estimated densities with $\theta = 1$

| Sample Size | Measure | Estimated density | | |
|---|---|---|---|---|
| | | Gaussian kernel | Gamma kernel | Lomax |
| $n = 10$ | MSE | 0.437947 | 0.411517 | 0.254002 |
| | MAE | 0.603266 | 0.584701 | 0.456779 |
| | MAPE | 1.72574 | 1.4069 | 0.945377 |
| $n=100$ | MSE | 0.379856 | 0.332857 | 0.214483 |
| | MAE | 0.543075 | 0.469493 | 0.365412 |
| | MAPE | 1.42119 | 1.37828 | 0.776492 |
| $n=1000$ | MSE | 0.355446 | 0.296394 | 0.182367 |
| | MAE | 0.538791 | 0.447678 | 0.339461 |
| | MAPE | 1.28718 | 1.0154 | 0.681375 |

Table (3): Goodness of fit measure's of the difference between the actual density and the estimated densities with $\theta = 2$

| Sample Size | Measure | Estimated density | | |
|---|---|---|---|---|
| | | Gaussian kernel | Gamma kernel | Lomax |
| $n = 10$ | MSE | 1.32134 | 1.10478 | 1.01706 |
| | MAE | 0.969244 | 0.835095 | 0.797891 |
| | MAPE | 1.53093 | 1.40952 | 1.1792 |
| $n=100$ | MSE | 1.19339 | 0.978609 | 0.909661 |
| | MAE | 0.919006 | 0.787027 | 0.737417 |
| | MAPE | 1.52037 | 1.08411 | 0.98922 |
| $n=1000$ | MSE | 0.948175 | 0.835279 | 0.724036 |
| | MAE | 0.752042 | 0.653885 | 0.643588 |
| | MAPE | 1.21059 | 0.889223 | 0.755419 |

The above tables shows that under the generated exponential i.i.d samples, the estimated densities get closer to the original density function as the parameter $\theta$ decreases; and also the estimated densities get closer to the actual density as the sample size increases, and the suggested Lomax kernel always outperforms the others kernels, while the Gaussian kernel is the worst one.

Figures (1), (2) and (3), present the actual density with the estimated densities using: Lomax kernel; Gaussian Kernel and Gamma kernel at the different values of parameter $\theta$ and different sample size.
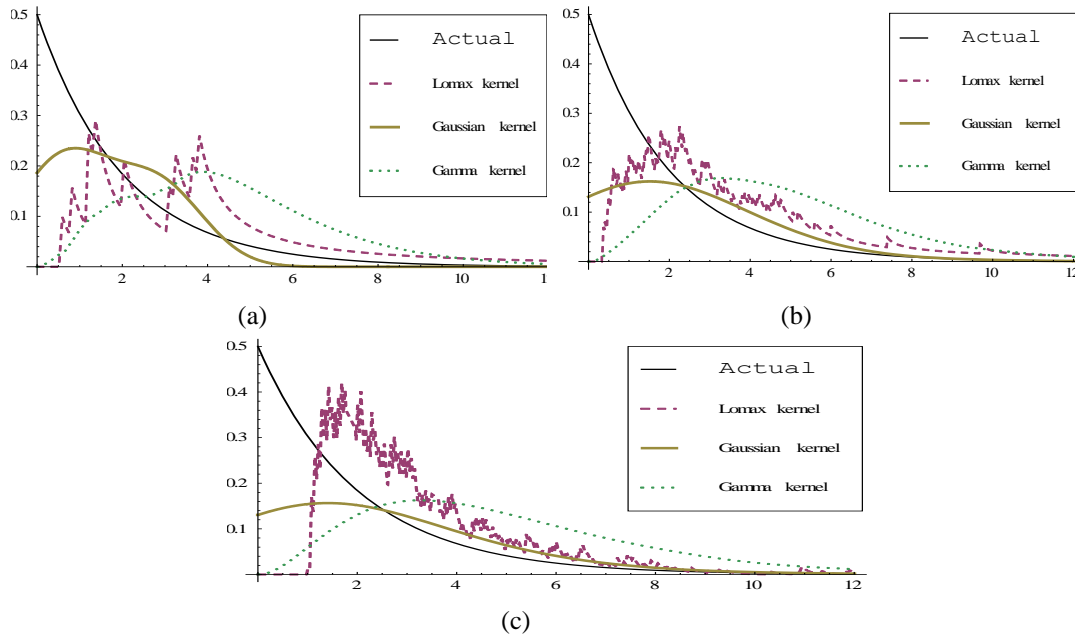


(a)

(b)

(c)

Fig. (1): The actual Exponential density ($\theta = .5$) with its kernel estimation using Lomax kernel; Gaussian kernel; and Gamma kernel with sample sizes (a) n=10, (b) n=100, and (c) n=1000
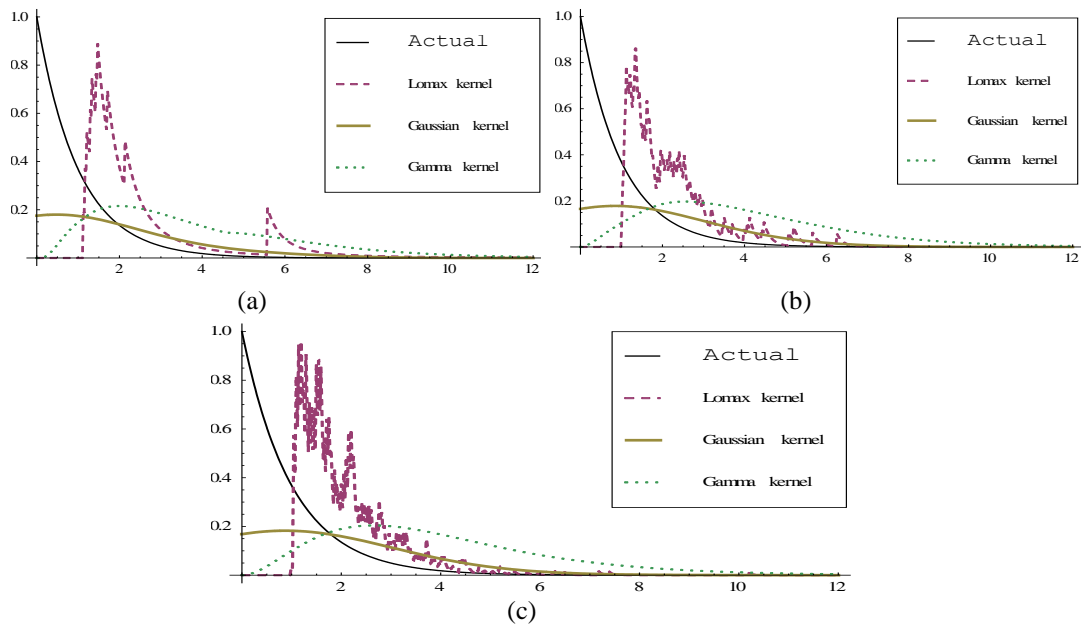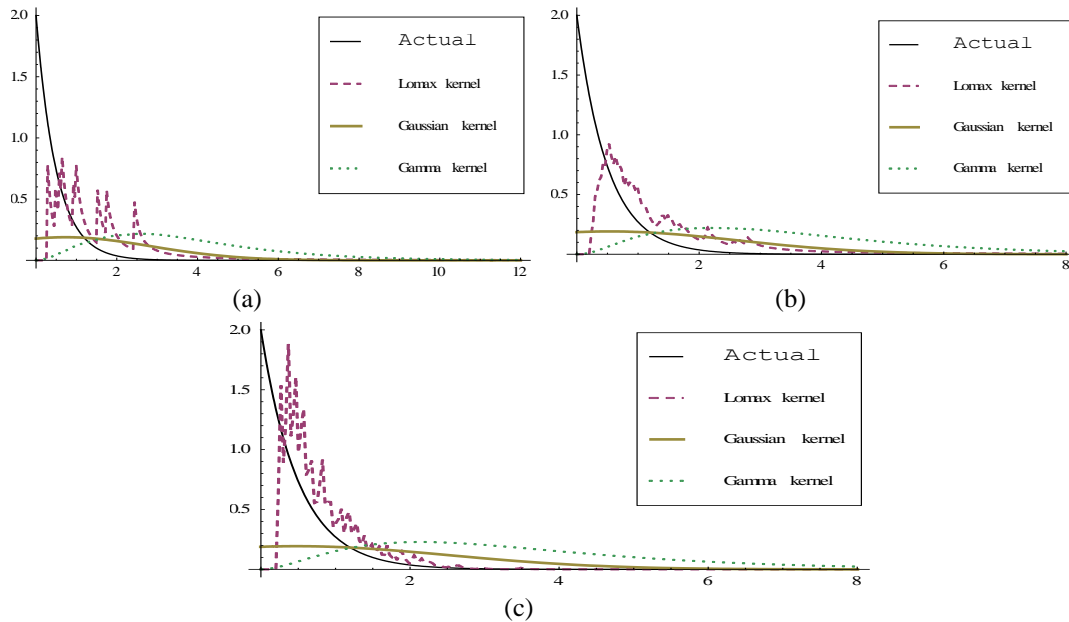


(a)

(b)

(c)

Fig. (2): The actual Exponential density ($\theta = 1$) with its kernel estimation using Lomax kernel; Gaussian kernel; and Gamma kernel with sample sizes (a) n=10, (b) n=100, and (c) n=1000

Fig. (3): The actual Exponential density ($\theta = 2$) with its kernel estimation using Lomax kernel; Gaussian kernel; and Gamma kernel with sample sizes  (a) n=10,  (b) n=100, and (c) n=1000

The above figures show that under the generated exponential i.i.d samples, the three kernel functions gets closer to the original density function as the sample size increases for different values of parameter $\theta$. Also, among the three kernel functions, the suggested Lomax kernel always outperforms the others while the Gaussian kernel is the worst one.

## 5.    Application

In this section, we apply the proposed Lomax kernel estimator to the Egyptian annual neonatal mortality rate data from 1960 to 2016 (www.indexmundi.com/facts/egypt/mortality-rate). Figure (4) shows the estimated distribution of Neonatal mortality rate data using both the parametric Lomax distribution and the nonparametric Lomax distribution.
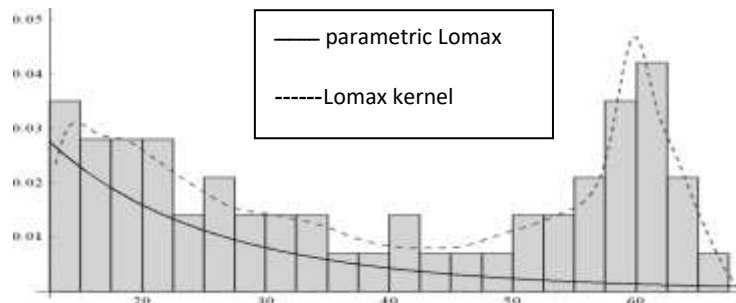


Fig. (4): The histogram; parametric and nonparametric Lomax estimator of Neonatal mortality rate data

The above figure indicates the flexibility of Lomax kernel in modelling the Neonatal mortality rate distributions. It shows that the nonparametric Lomax estimator agrees well with the neonatal mortality rate data.

## Conclusion

We consider the nonparametric estimation of the neonatal mortality rate density function using the Lomax Kernel function. The theoretical asymptotic properties of the proposed density estimator are derived. Also a simulation study was introduced. the Gaussian, Gamma, and Lomax kernels get closer to the actual density function as the sample size increases for different values of parameter $\theta$, but the suggested Lomax kernel always outperforms the others while the Gaussian kernel is the worst one.

## References

[1]     Abo-El-Hadid, S., Density Function Estimation Using Rayleigh Kernel, *Journal of Applied Probability & Statistics*, **14** (2019), 1-11.

[2]     Bagai, I. and P. Rao, Kernel Type Density Estimates for Positive Valued Random Variables, *The Indian Journal of Statistics*, **57** (1995), 56- 67.

[3]     Bouezmarni, T and O. Scaillet, Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data, *Econometric Theory*, **21** (2005), 390–412.
https://doi.org/10.1017/s0266466605050218

[4]     Bouezmarni, T., A. El Ghouch and M. Mesfioui, Gamma Kernel Estimators for Density and Hazard Rate of Right Censored Data, *Journal of Probability & Statistics*, **2011** (2011), 1-16. https://doi.org/10.1155/2011/937574

[5]     Chen, S. X., Beta Kernel Estimators for Density Functions, *Computational Statistics and Data Analysis*, **31** (1999), 131-145.
https://doi.org/10.1016/s0167-9473(99)00010-9

[6]     Chen, S. X., Probability Density Function Estimating Using Gamma Kernels, *Annals Institute of Mathematical Statistics*, **52** (2000), 471-480.
https://doi.org/10.1023/a:1004165218295

[7]     Indicators for Monitoring the Millennium Development Goals, The United Nations, New York, 2003.

[8]     Kleiber, Christian; Kotz, Samuel, *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley Series in Probability and Statistics, **470**, John Wiley & Sons, 2003. https://doi.org/10.1002/0471457175

[9]     Lomax, K. S., Business Failures; Another example of the analysis of failure data, *Journal of the American Statistical Association*, **49** (1954), 847–852.
https://doi.org/10.1080/01621459.1954.10501239

[10]    Markovich, L. A., Gamma Kernel Estimation of Multivariate Density and its Derivative on the Nonnegative Semi-axis by Dependent Data, Cornell University Library, 2015, arXiv preprint ArXiv:1410.2507

[11]    Markovich, L. A., Gamma-Weibull kernel estimation of the heavy tailed densities, Cornell University Library, 2016, ArXiv preprint arXiv:1604.06522v1.

[12]    Parzen, E., On Estimation of a Probability Density Function and Mode, *Annals of Mathematical Statistics*, **33** (1962), 1065-1076. https://doi.org/10.1214/aoms/1177704472

[13]    Rosenblatt, M., Remarks on Some Nonparametric Estimates of Density Function, *Annals of Mathematical Statistics*, **27** (1956), 832-837. https://doi.org/10.1214/aoms/1177728190

[14]    Scaillet, O., Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels, *Journal of Nonparametric Statistics*, **16** (2004), 217-226. https://doi.org/10.1080/10485250310001624819