Journal of Rajasthan Academy of Physical Sciences
ISSN: 0972-6306; URL: http://raops.org.in
Vol.23, No.1&2, January-June, 2024, 82-91

SOME PROBABILITY DISTRIBUTIONS WITH STATISTICAL PROPERTIES AND THEIR APPLICATION TO WAITING AND SURVIVAL TIME DATA

Abhishek Agarwal[1] and **Himanshu Pandey**[2]
[1, 2]Department of Mathematics & Statistics, D.D.U. Gorakhpur University, Gorakhpur, U.P., India

Abstract: Real-world problems are described by probability distributions. In this paper we developed length biased Zeghdoudi distribution, generalized exponential distribution with two parameters and its length biased form with some statistical properties.

The parameters of the proposed distributions were estimated using the maximum likelihood approach and application of the proposed distributions were applied to a real life data set related to waiting time to first conception for females and survival time of Guinea individuals infected with Ebola virus.

Keywords: Generalized Exponential Distribution; Length Biased; Maximum Likelihood Estimation; Waiting Time; Conception; Virus.

2020 AMS subject classifications: 62E10, 62F10, 62P25, 62-08, 62-11.

1. Introduction

A variety of distributions have been applied to lifetime data in recent years, revealing functions of diminishing or growing failure rates as well as significant use in modeling natural life data that are factorized by risk and uncertainty. Exponential families are crucial to probability and statistical research, which are widely applied in a variety of disciplines including the medical, biological, and engineering sciences. Pommeret[10]. Researchers from several disciplines, including (Adewusi et al.[1], Afaq et al.[2], Lee et al.[7], Al-Aqtash et al.[3], Alzaatreh et al.[4] and Artur[5]), have applied probability distribution theory to a variety of fields, including medicine, biology, demography, the environment, and more.

The objective of our present work is to develop generalized exponential distribution (GED), length-biased generalized exponential distribution (LBGED) and length-biased zeghdoudi distribution (LBZD), as well as to describe some relevant statistical properties and parameter estimation to some continuous data sets to test their applicability.

2. Generalized Exponential Distribution (GED)

The probability density function of exponential distribution with parameter λ is given by-

$$P(X=x)=\lambda e^{-\lambda x} \quad ; \quad x>0\,,\lambda>0 \tag{1}$$

Now we define generalized form of exponential distribution and its density function is follows as:

$$P(X=x)=\lambda^n e^{-\lambda^n x} \quad ; \quad x>0\,,\lambda>0, n\in I \tag{2}$$

with two parameters λ as scale parameter and n as shape parameter.

2.1 Statistical Properties of Generalized Exponential Distribution

Moment and Related Measures

$$\text{Mean} = \int_0^\infty x\lambda^n e^{-\lambda^n x} dx = \frac{1}{\lambda^n}$$

$$\mu_r' = \int_0^\infty x^r \lambda^n e^{-\lambda^n x}\, dx = \frac{\lambda^n \Gamma(\mathrm{r}+1)}{(\lambda^n)^{\mathrm{r}+1}}$$

putting r = 2,3,4 we get

$$\mu_2' = \frac{2}{\lambda^{2n}}\,, \mu_3' = \frac{6}{\lambda^{3n}}\,, \mu_4' = \frac{24}{\lambda^{4n}}$$

$$\text{Variance} = \mu_2' - (\mu_1')^2 = \frac{1}{\lambda^{2n}}$$

$$\mu_3 = \mu_3' - 3\mu_2'(\mu_1')^2 + 2(\mu_1')^3 = \frac{2}{\lambda^{3n}}$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 = \frac{9}{\lambda^{4n}}$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 4 \;\Rightarrow \text{Distribution is positively skewed.}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 9 \;\Rightarrow \text{Distribution is leptokurtic.}$$

$$M_x(t) = \int_0^\infty e^{tx}\lambda^n e^{-\lambda^n x}\, dx = \left(1-\frac{t}{\lambda^n}\right)^{-1}$$

$$\phi_x(t) = \int_0^\infty e^{itx}\lambda^n e^{-\lambda^n x}\, dx = \left(1-\frac{it}{\lambda^n}\right)^{-1}$$

$$K_x(t) = \log M_x(t) = -\log\left(1-\frac{t}{\lambda^n}\right)$$

$$F(x) = \int_0^x \lambda^n e^{-\lambda^n x}\, dx = 1 - e^{-\lambda^n x}$$

Survival function $s(x) = 1 - F(x) = e^{-\lambda^n x}$

Hazard function $h(x) = \frac{f(x)}{s(x)} = \lambda^n$

Cumulative hazard Rate $H(t) = \int_0^t h(x)dx = \lambda^n t \; ; \; \begin{matrix} t > 0 \\ n \in I \\ \lambda > 0 \end{matrix}$

Mean Residual life $= m(t) = \frac{1}{1 - F(t)} \int_t^\infty \{1 - F(x)\}dx = \frac{1}{\lambda^n}$

Renyi Entropy − An entropy is a measure of uncertainity.

$$e(\eta) = \frac{1}{1-\eta} \log\left[\int_0^\infty f^\eta(x)\, dx\right]$$

$$= \frac{1}{1-\eta} \log \int_0^\infty \left(\lambda^n e^{-\lambda^n} x\right)^\eta dx$$

$$= \frac{1}{1-\eta} \log\left\{\frac{\lambda^{(\eta-1)n}}{\eta}\right\}$$

3. Length Biased Generalized Exponential Distribution (LBGED)

According to Patil and Rao [9], if $f(x;\theta)$ be the probability density function of a random variable X and the unknown parameter θ, the weighted distribution is defined as:

$$f(x;\theta) = \frac{w(x)\, f^*(x;\theta)}{E[w(x)]} \quad ; \quad x \in R, \theta > 0$$

where $w(x)$ is the weight function and $f^*(x;\theta)$ is the base line distribution.

If we take $w(x) = x \; ; \; x > 0$

Then Length biased form of Generalized exponential distribution is given by-

$$f(x) = x\lambda^{2n} e^{-\lambda^n x} \quad ; \; x > 0, \lambda > 0, n \in I \tag{3}$$

where λ and n are scale and shape parameters respectively.

3.1 Statistical Properties of Length Biased Generalized Exponential Distribution

Moment and Related Measures

$$\mu_r' = \int_0^\infty x^{r+1} \lambda^{2n} e^{-\lambda^n x}\, dx = \lambda^{2n} \frac{\Gamma(\mathrm{r}+2)}{(\lambda^n)^{r+2}} \quad ; \; \lambda > 0, n \in I$$

putting r $= 1, 2, 3, 4$ we get –

$$\mu_1' = \text{Mean} = \frac{2}{\lambda^n}\,, \qquad \mu_2' = \frac{6}{\lambda^{2n}}\,, \qquad \mu_3' = \frac{24}{\lambda^{3n}}\,, \qquad \mu_4' = \frac{120}{\lambda^{4n}}$$

$$\text{Variance} = \mu_2' - (\mu_1')^2 = \frac{2}{\lambda^{2n}}$$

$$\mu_3 = \mu_3' - 3\mu_2'(\mu_1') + 2(\mu_1')^3 = \frac{4}{\lambda^{3n}}$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 = \frac{24}{\lambda^{4n}}$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 2 \; and \; \gamma_1 = \sqrt{2} \;\Rightarrow \text{Distribution is positively skewed.}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 6 \; and \; \gamma_2 = 3 \;\Rightarrow \text{Distribution is leptokurtic.}$$

$$M_x(t) = \int_0^\infty e^{tx}\, x\lambda^{2n} e^{-\lambda^n x}\, dx = \left(1 - \frac{t}{\lambda^n}\right)^{-2}$$

$$\phi_x(t) = \int_0^\infty e^{itx}\, x\lambda^{2n} e^{-\lambda^n x}\, dx = \left(1 - \frac{it}{\lambda^n}\right)^{-2}$$

$$K_x(t) = \log M_x(t) = -2\log\left(1 - \frac{t}{\lambda^n}\right)$$

$$F(x) = \int_0^x x\lambda^{2n} e^{-\lambda^n x}\, dx = 1 - e^{-x\lambda^n}(1 + x\lambda^n)$$

$$\text{Survival function } s(x) = 1 - F(x) = e^{-x\lambda^n}(1 + x\lambda^n)$$

$$\text{Hazard function } h(x) = \frac{f(x)}{s(x)} = \frac{x\lambda^{2n}}{1 + x\lambda^n}$$

$$\text{Cumulative hazard Rate } H(t) = \int_0^t h(x)dx = \lambda^n t - \log(1 + \lambda^n t)\,;\, \lambda > 0\,,\, t > 0$$

$$\text{Mean Residual life} = m(t) = \frac{1}{1 - F(t)}\int_t^\infty \{1 - F(x)\}dx = \frac{\lambda^n t + 2}{\lambda^n(\lambda^n t + 1)}$$

$$\text{Renyi Entropy} - e(\eta) = \frac{1}{1 - \eta}\log\left[\int_0^\infty f^\eta(x)\, dx\right]$$

$$= \frac{1}{1-\eta} \log \int_0^\infty \left(x\lambda^{2n} e^{-\lambda^n} x\right)^\eta dx$$

$$= \frac{1}{1-\eta} \log \left\{\lambda^{(\eta-1)n} \frac{\Gamma(\eta+1)}{\eta^{\eta+1}}\right\}$$

3.2 Maximum Likelihood Estimation

The maximum likelihood estimation function of length-biased generalized exponential distribution can be given as:

$$L(\lambda;\, x_1, x_2, \dots, x_t) = (x_1 . x_2 \dots x_t)(\lambda^{2n})^t\, e^{-\lambda^n} \sum_{i=1}^{t} x_i$$

Taking log on both sides and partially differentiate with respect to $\boldsymbol{\lambda}$, we get:

$$\frac{\partial}{\partial\lambda} \log L = \frac{2nt}{\lambda} - \sum x_i\;\, n\lambda^{n-1} = 0$$

$$\Rightarrow \hat{\lambda} = \left(\frac{2}{\bar{x}}\right)^{\frac{1}{n}}$$

$w$here λ is parameter of the distribution

4. Length Biased Zeghdoudi Distribution (LBZD)

Messaadia and Zeghdoudi [8] has suggested a new distribution called as Zeghdoudi distribution (ZD) which is given as:

$$f(x;\theta) = \begin{cases} \dfrac{\theta^3 x(1+x)e^{-\theta x}}{2+\theta} & ; \quad x, \theta > 0 \\ 0 & ; \quad \text{otherwise} \end{cases} \tag{4}$$

with Mean $= \dfrac{2(\theta+3)}{\theta(\theta+2)}$ and Variance $= \dfrac{2(\theta^2+6\theta+6)}{\theta^2(\theta+2)^2}$

Now the length-biased form of Zeghdoudi distribution can be defined as -

$$f(x;\theta) = \begin{cases} \dfrac{\theta^4}{2(\theta+3)}\, x^2(1+x)e^{-\theta x} & ;\ x > 0, \theta > 0 \\ 0 & ;\ \text{otherwise} \end{cases} \tag{5}$$

with Mean $= \dfrac{3(\theta+4)}{\theta(\theta+3)}$

4.1 Maximum Likelihood Estimation

Let $\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}$ be the random sample drawn from the population the following length-biased Zeghdoudi Distribution having parameter $\boldsymbol{\theta}$. Then the likelihood function can be given as:

$$L = \prod_{i=1}^{n} f(x_i;\theta)$$

$$= \left\{\frac{\boldsymbol{\theta}^4}{\mathbf{2(\boldsymbol{\theta}+3)}}\right\}^{\boldsymbol{n}} \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_i^2(1+x_i)e^{-\theta\sum x_i}}$$

Now taking log on both sides and partially differentiating w.r.t. θ we get –

$$\frac{\partial}{\partial\theta}\log L = \frac{n}{\theta}\left(\frac{6\theta+24}{2\theta+6}\right) - \sum x_i = 0$$

$$\Rightarrow\ \theta^2\bar{x} + 3\theta\bar{x} - 3\theta = 12$$

$$\Rightarrow\ \hat{\theta} = \frac{(3-3\bar{x}) \pm \sqrt{3}\sqrt{3\bar{x}^2+10\bar{x}+3}}{2\bar{x}}$$

$$\because \theta > 0 \text{ we consider } \hat{\theta} = \frac{(3-3\bar{x}) + \sqrt{3}\sqrt{3\bar{x}^2+10\bar{x}+3}}{2\bar{x}}$$

5. Application

The proposed probability models (ii) and (iii) at n=2, (iv) and (v) are fitted using some real data sets collected from different sample surveys related to waiting time for first conception, taken from NFHS IV [6] for Varanasi district of Uttar Pradesh, India. Only those females are considered in our data whose age at marriage is more than or equal to 16 years and marital duration is more than 10 years at the time of survey, and the data for survival times of (94, 91) Guinea individus infected with Ebola virus used by Messadia and Zeghdoudi [8].

Table-1. Data of survival time (in months) of 91 Guinea individuals infected with Ebola Virus

Survival time (in months)	Observed Frequency	Expected Frequency			
		GED	**LBGED**	**ZD**	**LBZD**
0-2	35	44.50	35.34	34.56	27.51
2-4	32	22.73	32.80	31.49	42.83
4-6	16	11.61	14.71	16.20	16.19
6-8	6 }8	12.15	8.15	7.98	4.46
8-10	2 }				
Total	**91**	**91**	**91**	**91**	**91**
Estimates of parameter		**0.5794**	**0.8195**	**0.896**	**1.244**
Chi square value		**8.86**	**0.13**	**0.19**	**7.57**
d.f		**2**	**2**	**2**	**2**
p-value		**0.0119**	**0.9370**	**0.9093**	**0.0227**

Table-2. Data of survival time (in months) of 94 Guinea individuals infected with Ebola Virus

Survival time (in months)	Observed Frequency	Expected Frequency			
		GED	LBGED	ZD	LBZD
0-2	32	43.97	33.80	33.25	25.19
2-4	35	23.40	33.64	32.36	43.75
4-6	17	12.45	16.34	17.79	18.90
6-8	7 } 10	14.17	10.22	9.51	6.16
8-10	3 }				
Total	**94**	**94**	**94**	**94**	**94**
Estimates of parameter		**0.5616**	**0.7943**	**0.852**	**1.173**
Chi square value		**11.88**	**0.17**	**0.40**	**6.17**
d.f		**2**	**2**	**2**	**2**
p-value		**0.0026**	**0.9185**	**0.8187**	**0.0457**

Table-3. Goodness of fit on waiting time to first conception for females (NFHS-IV data).

Waiting time to first conception (in months)	Observed no. of Females	Expected no. of Females			
		GED	LBGED	ZD	LBZD
0-3	12	53.40	12.20	4.19	1.09
3-15	165	157.10	150.01	131.49	113.04
15-27	125	94.55	135.87	161.65	185.16
27-39	72	56.90	78.77	91.36	100.60
39-51	40	34.24	39.21	38.59	35.25
51-63	17	20.61	18.06	14.01	9.815
63-75	10	12.40	7.93	4.63	2.37
75-87	4 }				
87-99	1 } 7	18.8	5.95	2.07	0.675
99-111	2 }				
Total	**448**	**448**	**448**	**448**	**448**
Estimates of parameter		**0.2057**	**0.2908**	**0.1244**	**0.1670**
Chi square value		**55.74**	**3.75**	**41.49**	**97.69**
d.f		**6**	**6**	**4**	**3**
p-value		**$3.2854\ e^{-10}$**	**0.7104**	**$2.1277\ e^{-08}$**	**0.00001**

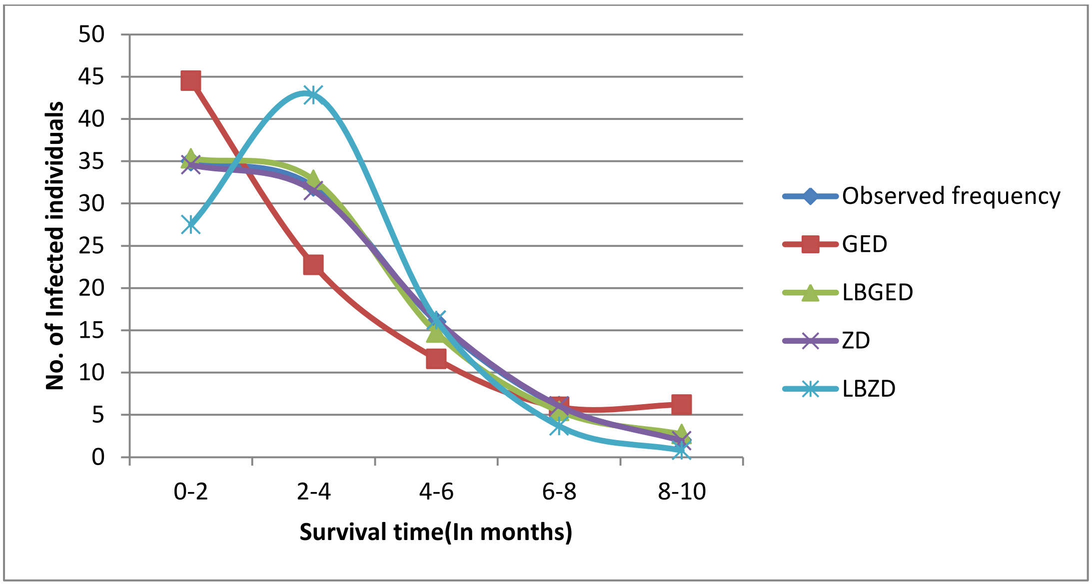

Figure 1: Graphical presentation showing observed and expected number of (91) Guinea individuals infected with Ebola Virus.

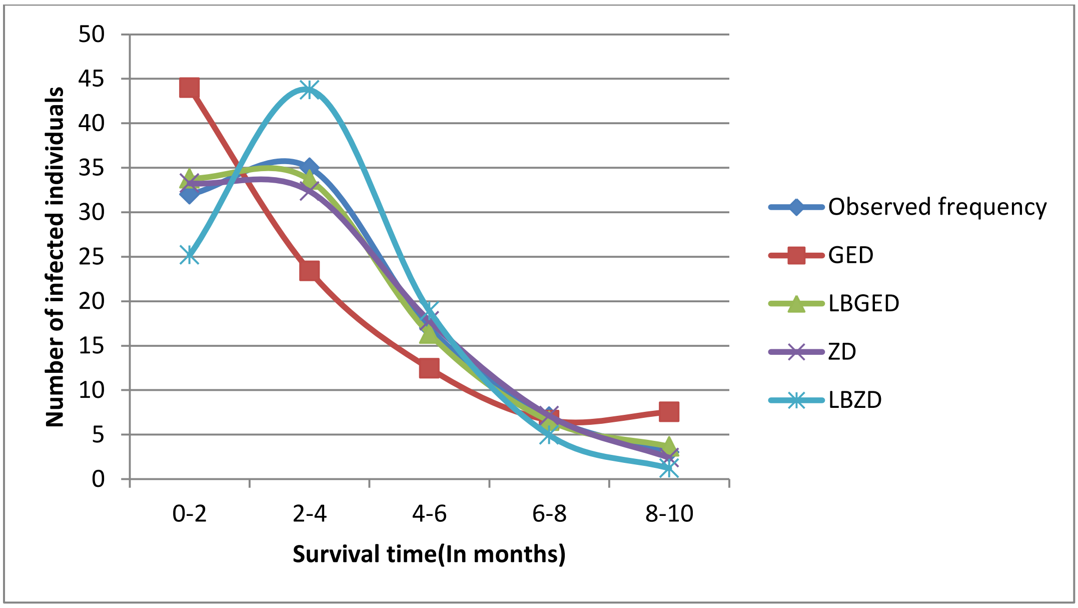

Figure 2: Graphical presentation showing observed and expected number of (94) Guinea individuals infected with Ebola Virus.

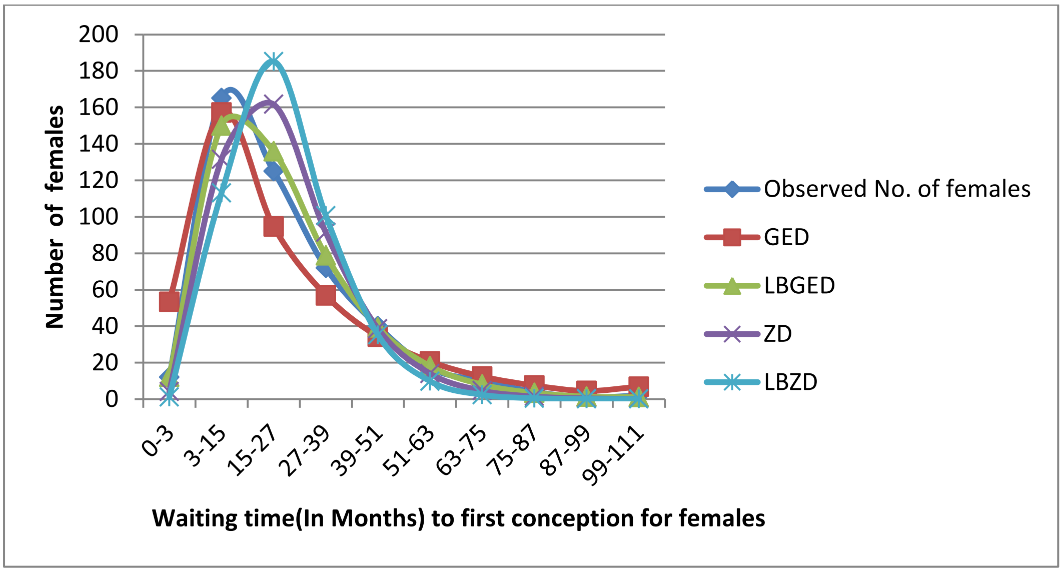

Figure 3: Graphical presentation showing observed and expected number of females having waiting time for first conception (NFHS-IV) data.

6. Conclusions

In the area of probability and statistics, several distributions have been developed, defined and widely used. In this paper we defined, studied and established a new generalized exponential distribution and its length biased form their statistical properties including moments, skewness, kurtosis, hazard rate function, mean residual life function have also been discussed. The ML estimation technique is used to estimate its parameter. Finally the goodness of fit using the chi square statistic has been shown to demonstrate the applicability and comparability of GED, LBGED, ZD, and LBZD for modeling life time data. According to the value of $\chi^2$and p-value from table 1-3 and graphical representation between O_i and E_i, the nature and behavior of proposed LBGED model founds best suitable in modeling life time data.

Acknowledgement: The author is thankful to Referee for valuable comments and suggestions.

References

[1] Adewusi, O.A., Ayeni, T.M. and Ogunwale, O.D. (2019). On The Exponential-Gamma-X Mixed Distribution, International Journal of Research and Innovation in Applied Science, **4**(11), 139-141.

[2] Afaq, Ahmad, S.P. and Ahmad, A. (2017). Characterization and Estimation of Weibull-Rayleigh Distribution with Applications to Life Time Data, Applied Mathematics and Information Sciences, **5**(2), 71-79. http://dx.doi.org/10.18576/amisl/050206

[3] Al-Aqtash, R., Lee, C. and Famoye, F. (2013). A new method for generating families of continuous distributions, Metron **71**(1), 63-79.

[4] Alzaatreh, Famoye, F. and Lee, C. (2013). Weibull-Pareto Distribution and Its Applications, Communications in Statistics-Theory and Methods, **42**(9), 1673-1691.

[5] Artur, J.L. (2014). The beta log-logistic distribution, Brazilian Journal of Probability and Statistics **28**(3), 313-332. http://dx.doi.org/10.1214/12-BJPS209

[6] IIPS. National family health survey (NFHS-4).2015-16. International Institute for Population Sciences (IIPS), Mumbai, India, 2017.

[7] Lee, C., Al-Aqtash and Famoye, F. (2014). Gumbel-Weibull distribution: Properties and applications, Journal of Modern Applied Statistical Methods, **13**(2), 201-225.

[8] Messaadia, H. and Zeghdoudi, H. (2018). Zeghdoudi Distribution and Its Applications, International Journal Computing Science and Mathematics, **9**(1), 58-65.

[9] Patil, G.P. and Rao, C.R. (1977). The weighted distribution: A survey of their application, In Applications of Statistics, Krishnaiah PR (Ed.), North Holland Publishing Company. 383-405.

[10] Pommeret, D. (1997). Bhattacharyya matrices and exponential families, Journal of Multivariate Analysis, **63**(1), 105-118.