Implementation of Multivariate Exponential Power Distribution in Discrimination and Classification of Psychological Data and Other applications

A. A. Olósundé[1*] and A. T. Sóyínká[2]

[1] *Department of Mathematics, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria.*
[2] *Department of Statistics, Federal University of Agriculture, Alabata, Abeokuta, Ogun State, Nigeria.*

Authors' contributions

This work was carried out in collaboration between both authors. Author AAO designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Author ATS managed the analyses of the study and the literature searches. Both authors read and approved the final manuscript.

Original Research Article

ABSTRACT

Recent advances have shown that some multivariate psychological data are deviating from usual normal assumption either in the tails or kurtosis. Thereby, allowing the call for modelling of such data using more robust elliptically contoured density which includes the normal distribution as a special case. This allowed more flexibility at the kurtosis and tail regions, which is better in handling non-normality in data analysis and also lower the cost of misclassification. The present study employed a robust model for such cases in the context of discrimination and classification of multivariate psychological disorder data using multivariate exponential distribution as an underlining model. Parameters were estimated using the method of maximum likelihood estimation and the discrimination and classification were based on the log likelihood ratio approach. The resulting models relied solidly on the shape parameter, which regulate the tails and the kurtosis, thereby

---

Corresponding author: E-mail: aolosunde@oauife.edu.ng;

allowed flexibility. This method enable us to lower the cost of misclassification. Some other areas of applications were also considered in the paper.

1 INTRODUCTION

The present study is sequel to Olosunde and Soyinka [1], where we introduced multivariate exponential power distribution as an underlining distribution in discrimination and classification of multivariate data via method of likelihood ratio. The paper considered the case when the shape parameter is unity. But in everyday analysis we encountered several cases where the shape parameter is greater than unity. Such cases are well addressed in the present study, which some how generalized the existing results in the literature with some applications and lay much emphasis on psychological data . Now given a psychological multivariate data with heavier or lighter tail than the usual normal distribution, efforts are made to lower the cost of misclassification, when assigning an unknown subject to one of k classes on the basis of multivariate observation $\mathbf{x} = (\mathbf{x_1}, ..., \mathbf{x_p})^\mathbf{T}$, where p is the number of features in each class. For simplicity of notation k are defined to be integers ranging from 1 to K. We assume that there are n_k observations in class k with each pth multivariate class distributed as

$$x_{1,n_1}, ..., x_{K,n_K} \overset{i.i.d.}{\sim} N_p(\mu_k, \Sigma_k), \qquad k = 1, ..., K$$

where μ_k and Σ_k are the corresponding mean vector and covariance matrix of the $p-$dimensional multivariate normal distribution for each of the known k classes. The total number of observations is $n = n_1 + ... + n_K$.

Let π_k denote the prior probability of observing a class k member with $\pi_1 + ... + \pi_K = 1$. Under the normal distribution assumption, we assign a new subject **x** to class k, which minimizes the following discriminant score

$$D_k(x) = (x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k) + \ln |\Sigma_k| - 2 \ln \pi_k, \tag{1}$$

that is we assign **x** to $\widehat{k} = argmin_k D_k(x)$. This is the so-called quadratic discriminant analysis (QDA) since the boundaries that separate the

disjoint regions belonging to each class are quadratic (Aitchison and Silvey [2]; Cox [3]). The first term on the right-hand side of equation (1) is known as the squared Mahalanobis distance between x and μ_k. When the covariance matrices are all the same, i.e., $\Sigma_k = \Sigma$ for all k, the discriminant score can be simplified as

$$d_k(x) = (x - \mu_k)^T \Sigma^{-1}(x - \mu_k) - 2 \ln \pi_k \tag{2}$$

This is referred to as linear discriminant analysis (LDA). LDA assigns a new subject to $\widehat{k} = argmin_k d_k(x)$ which uses linear boundaries. The mean vectors μ_k and covariance matrices Σ_k when not known are estimated by their maximum-likelihood estimates, $\widehat{\mu}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{k,i}$, $\widehat{\Sigma}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} (x_{k,i} - \widehat{\mu}_k)(x_{k,i} - \widehat{\mu}_k)^T$, $\widehat{\Sigma} = \frac{1}{n} \sum_{k=1}^{K} n_k \widehat{\Sigma}_k$. The prior probabilities are usually estimated by the fraction of each class in the pooled training sample, i.e., $\widehat{\pi}_k = n_k/n$. The sample version rule for QDA is $\ell(x) = argmin_k \widehat{D}_k(x)$, where

$$\widehat{D}_k(x) = (x - \widehat{\mu}_k)^T \widehat{\Sigma}_k^{-1}(x - \widehat{\mu}_k) + ln \left| \widehat{\Sigma}_k \right| - 2 \ln \pi_k,$$

Similarly, the sample version rule for LDA is

$$\ell = argmin_k \widehat{d}_k(x),$$

where

$$\widehat{d}_k(x) = (x - \widehat{\mu}_k)^T \widehat{\Sigma}^{-1}(x - \widehat{\mu}_k) - 2ln\pi_k.$$

However, in situation of unclear boundary of separation, minimizing the cost of misclassification is very paramount. Hence, according to Johnson and Wichern [4], define $c[.|k]$ to be the cost of mis-allocating a new subject **x** to class k then $c[.|k] = 1 - Pr\left(\widehat{D}_k(x)|\widehat{\mu_k}, \widehat{\Sigma_k}\right)$ (Anderson, 1972; Hands and Henley [5]). Also for two populations, logistic discrimination was suggested by Cox [3] and Day and Kerridge [6] with the restriction that estimation of the discriminator was based on samples from the mixture of the populations. This method was extended (Anderson, 1972) to more than two populations and to the more usual plan of sampling from each distribution separately,

using Aitchison and Silvey's [2] method of constrained maximum likelihood estimation. Logistic discriminators can be used in a simple linear form and when the likelihood ratios of the populations are linear in the observations, they are optimal irrespective of the actual likelihoods. They are thus optimal for a much wider class of distributions than standard linear discriminators. The method of logistic discrimination can be extended to the case where the likelihood ratios are quadratic in the observations. The classification rules are known to be sensitive to departures from basic model assumptions on tail flexibility; which the normal and logistic distributions do not accommodate due to their fixed tail region.

Given multivariate data with heavier or lighter tails than normal and logistic distribution, the resulting probability of misclassification will be higher in values and thus we obtain classification rule not suitable or reliable for future classification of any new entrant from the same population density. Since it is well known that, multivariate normal assumption are not realistic in many applications. In fact, Azzalini [7,8] emphasized that multivariate normal does not practically obtainable except only on theoretical assumptions. In the case of logistic regression procedure Effron [9] and Press and Wilson [10] showed that logistics regression must be less efficient than the exponential families of distribution which normal discrimination procedure under model (1) and (2) belong to, at least asymptotically, as n goes to infinity, since the families are based on the full maximum likelihood estimator for the population parameters. Hence, there is need to study a generalized class of the family of this elliptical density that will accommodate for tail flexibility; and one which multivariate normal is a special case. Section two gave a brief introduction into the exponential power distribution and the significance of it in the study, section three presented the theoretical development of linear and quadratic discriminant model for classification. Finally, section four present the application of the model in section three to primary and secondary data.

2   MULTIVARIATE EXPONENTIAL POWER DISTRIBUTION (MEPD)

The family of elliptical density which also belong to exponential family under consideration is the exponential power distribution with the pdf

$$f(y; \mu, \sigma, \beta) = \frac{1}{\sigma \, 2^{1+1/2\beta} \Gamma(1 + \frac{1}{2\beta})} \exp\left\{ -\frac{|y - \mu|^{2\beta}}{2\sigma^{2\beta}} \right\} \tag{3}$$

where $-\infty < y < \infty$, $-\infty < \mu < \infty$, $\beta > 0$ and $\sigma > 0$. (2.1) is called exponential power distribution with shape parameter β which regulates the tail region. The multivariate extension is

$$f(y; \mu, \Sigma, \beta, p) = \frac{p\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}} \sqrt{|\Sigma|} \Gamma\left(1 + \frac{p}{2\beta}\right) 2^{1 + \frac{p}{2\beta}}} \exp\left\{ -\frac{1}{2} \left[(y - \mu)^T \Sigma^{-1} (y - \mu) \right]^{\beta} \right\} \tag{4}$$

where the mean and variance are $E(Y) = \mu$, $var(Y) = \frac{2^{\frac{1}{\beta}} \Gamma\left(\frac{p+2}{2\beta}\right)}{p\Gamma\left(\frac{p}{2\beta}\right)} \Sigma$ and β determines the kurtosis

(Gomez et al. [11]). Thus, the correlation structure can be obtained directly from Σ in the usual way. However, when $\beta = 1$, we have a multivariate normal distribution; when $\beta = 1/2$, it becomes multivariate Laplace (double exponential) distribution; and when $\beta \to \infty$, we have a multivariate uniform distribution. Hence, when $\beta < 1$, the distribution has heavier tails than the multivariate normal distribution and this property can be useful in providing robustness against outliers (Lindsey [12]). Parameters of the exponential power distribution were estimated using the method of maximum likelihood see Saralees (2005). The resulting equations were not in close form, therefore, we developed code in R environment to estimate the parameters of any given data. Similar code were developed for univariate case by Ruggeri and Mineo [13] and Agro [14]. Some areas where normal distribution assumptions had not been reliable, thus necessitated the replacement by exponential power distribution

for such elliptically contoured data are found in poultry feeds data Olosunde [15], also several authors have shown that financial data cannot be justified by normal distribution hypo-thesis, see Eberlein and Keller (1995), In Bayesian network, Main and Navarro [16] proved that exponential power distribution was preferred for some platikurtic conditional distributions where conditional regression functions are constant. The results gave conditions to avoid compatibility problems when distributions with lighter tails than the normal are used in the description of conditional densities to specify joint densities, like in Bayesian networks. Hence, a generalized elliptical densities which includes the normal distribution with better flexible tail, that can represent features of the data as adequately as possible and reduce unrealistic assumptions. Furthermore, the kind of data and applications that follows multivariate exponential power distribution can be seen in (Gomez et al. [11]), Lindsey [12] and Olosunde [15] just to mention few. This present study was motivated by the behavioural pattern of some Secondary school Teachers which has lead to unreported psychological disorder among them. The effects were obvious in the quality of students they produce for higher education learning. The sample collected using General Health Questionnaire (GHQ) informed the idea of separating those with psychological disorder from those who are not. But the sample collected did not followed the usual normal distribution hitherto commonly assumed in the analysis. Therefore, in a parametric case where maximum likelihood estimation method is embraced for parameter's estimation, there is need to substitute the underlining normal distribution for a generalized one which will account for any deviation from normal distribution both in the kurtosis and tails.

3 DISCRIMINATION AND CLASSIFICATION UNDER MULTIVARIATE EXPONENTIAL POWER DISTRIBUTION

3.1 Discriminant Function

Proposition 1. The discriminant function between two Multivariate Exponential Power Distribution classes i and j from among $k = 1, 2, ., i, .j, ., K$ classes can be derived as

$$-\frac{1}{2}\ln\left(\frac{|\Sigma_j|}{|\Sigma_i|}\right) - \frac{1}{2}\left[L^\beta(y_j) - M^\beta(y_i)\right] \tag{5}$$

where $L(y_j) = \left|y_j'\Sigma_j^{-1}y_j - 2\mu_j'\Sigma_j^{-1}y_j + \mu_j'\Sigma_j^{-1}\mu_j\right|$ and
$M(y_i) = \left|y_i'\Sigma_i^{-1}y_i - 2\mu_i'\Sigma_i^{-1}y_i + \mu_i'\Sigma_i^{-1}\mu_i\right|$

Proof. Suppose y_i and y_j are multivariate observations from classes i and j respectively with density (2.2), then the ratio $\frac{f_j(y_j)}{f_i(y_i)}$ can be obtained as

$$\frac{|\Sigma_i|^{\frac{1}{2}}}{|\Sigma_j|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\left\{\left|y_j'\Sigma_j^{-1}y_j - 2\mu_j'\Sigma_j^{-1}y_j + \mu_j'\Sigma_j^{-1}\mu_j\right|^\beta - \left|y_i'\Sigma_i^{-1}y_i - 2\mu_i'\Sigma_i^{-1}y_i + \mu_i'\Sigma_i^{-1}\mu_i\right|^\beta\right\}\right). \tag{6}$$

Taking the natural logarithm of (6) $\ln\left[\frac{f_j(y_j)}{f_i(y_i)}\right]$ we have

$$-\frac{1}{2}\ln\left(\frac{|\Sigma_j|}{|\Sigma_i|}\right) - \frac{1}{2}\left[\left|y_j'\Sigma_j^{-1}y_j - 2\mu_j'\Sigma_j^{-1}y_j + \mu_j'\Sigma_j^{-1}\mu_j\right|^\beta - \left|y_i'\Sigma_i^{-1}y_i - 2\mu_i'\Sigma_i^{-1}y_i + \mu_i'\Sigma_i^{-1}\mu_i\right|^\beta\right]. \tag{7}$$

From (3.3), let $L(y_j) = \left|y_j'\Sigma_j^{-1}y_j - 2\mu_j'\Sigma_j^{-1}y_j + \mu_j'\Sigma_j^{-1}\mu_j\right|$ and
$M(y_i) = \left|y_i'\Sigma_i^{-1}y_i - 2\mu_i'\Sigma_i^{-1}y_i + \mu_i'\Sigma_i^{-1}\mu_i\right|$ then the result (5) follows which is the MEPD discriminant function between two classes.

3.2 Classification Rule

However from (7), suppose y_o is an arbitrary observation that could be classified into any of the classes i and j then we define a classification rule aimed at minimizing the misclassification rate of allocating y_o wrongly as follows: Allocate y_o to class j if

$$\ln\left(\frac{f_j(y_o)}{f_i(y_o)}\right) \geqslant \ln H = \ln\left[\left(\frac{c[j|i]}{c[i|j]}\right)\left(\frac{\pi_i}{\pi_j}\right)\right] \tag{8}$$

and allocate y_o to class i if otherwise; where $\left(\frac{c[j|i]}{c[i|j]}\right)$ and $\left(\frac{\pi_i}{\pi_j}\right)$ are the cost and prior probability ratio respectively, that can take different values depending on initial assumptions. Although the most common assumption in literature is when both cost and prior probability ratio are unity; we can evaluate the cost ratio as follows when such assumptions are not necessary.

Proposition 2. If y_o have the pdf (4) then the distribution of the discriminant function in (5) is

$$L^\beta(y_o) - M^\beta(y_o) \approx \frac{\gamma(\frac{p}{2\beta}, y_{oj})}{2\Gamma(\frac{p}{2\beta})} - \frac{\gamma(\frac{p}{2\beta}, y_{oi})}{2\Gamma(\frac{p}{2\beta})}.$$

Proof. From (4), if MEPD exponent for class j is $z_{oj} = \left|(y_o - \mu)'\Sigma^{-1}(y_o - \mu)\right|^\beta$ then

$$Pr(z < z_{oj}) = \int f(z_{oj})dz_{oj} = \int f(y_o)dy_o = \frac{1}{2} + \frac{1}{2\Gamma(\frac{p}{2\beta})}\int_0^{y_{oj}} z_o^{\frac{p}{2\beta}-1}\exp(-z_o)dz_o.$$

Similarly for class i, we have $Pr(z < z_{oi}) = \frac{1}{2} + \frac{1}{2\Gamma(\frac{p}{2\beta})}\int_0^{y_{oi}} z_o^{\frac{p}{2\beta}-1}\exp(-z_o)dz_o$. Since the discriminant function is a difference of two MEPD exponent of $[L^\beta(y_o) - M^\beta(y_o)]$, then the probability over the discriminant function is

$$Pr(z < z_{oj}) - Pr(z < z_{oi}) = \frac{\gamma(\frac{p}{2\beta}, y_{oj})}{2\Gamma(\frac{p}{2\beta})} - \frac{\gamma(\frac{p}{2\beta}, y_{oi})}{2\Gamma(\frac{p}{2\beta})}.$$

So we evaluate the probability functions $c[i|j] = 1 - Pr(z < z_{oj}|\mu_j, \Sigma_j)$ as the probability of misclassifying object y_o in i instead of j and $c[j|i] = Pr(z < z_{oi}|\mu_i, \Sigma_i)$ as the probability of misclassifying object y_o in j instead of i for the null hypothesis $H_o : \mu_i, \Sigma_i$ and alternative hypothesis $H_a : \mu_j, \Sigma_j$. Note that since $0 < y_o < z_{o(i,j)} < \infty$, then its integral will be a regularised incomplete gamma. Hence in summary from (7) and (8), the overall discriminant function for MEPD is

$$-\frac{1}{2}\left[L^\beta(y_j) - M^\beta(y_i)\right] \tag{9}$$

while the classification rule is

$$\ln H + \frac{1}{2}\ln\left(\frac{|\Sigma_j|}{|\Sigma_i|}\right). \tag{10}$$

So, the allocation rule for all arbitrary observation y_o to class j is

$$-\frac{1}{2}\left[L^\beta(y_o) - M^\beta(y_o)\right] \geqslant \ln H + \frac{1}{2}\ln\left(\frac{|\Sigma_j|}{|\Sigma_i|}\right) \tag{11}$$

and to class i is

$$-\frac{1}{2}\left[L^\beta(y_o) - M^\beta(y_o)\right] < \ln H + \frac{1}{2}\ln\left(\frac{|\Sigma_j|}{|\Sigma_i|}\right). \tag{12}$$

Where $L(y_j) = \left|y_j'\Sigma_j^{-1}y_j - 2\mu_j'\Sigma_j^{-1}y_j + \mu_j'\Sigma_j^{-1}\mu_j\right|$ and $M(y_i) = \left|y_i'\Sigma_i^{-1}y_i - 2\mu_i'\Sigma_i^{-1}y_i + \mu_i'\Sigma_i^{-1}\mu_i\right|$ are Mahalanobis distance for multivariate observations y_j and y_i for groups parameter (μ_j, Σ_j) and (μ_i, Σ_i) respectively for j and i classes, at the same shape parameter β. Next we obtain the discriminant function for shape parameter from $\beta = 1$ up to $\beta = 6$.

3.3 Some Polynomial Results of MEPD Discriminant Function

We established some results polynomial discriminant functions for $\beta = 1$ to $\beta = 6$ for multivariate exponential power distribution.

Corollary 1: From (9), when $\beta = 1$ we obtained the known discriminant model
$\left[y' \left(\Sigma_j^{-1} \mu_j - \Sigma_i^{-1} \mu_i \right) - \frac{y' \left(\Sigma_j^{-1} - \Sigma_i^{-1} \right) y'}{2} \right]$ and its classification rule
$\ln H + \frac{1}{2} \ln \left(\frac{|\Sigma_j|}{|\Sigma_i|} \right) + \frac{1}{2} \left(\mu_j \Sigma_j^{-1} \mu_j - \mu_i \Sigma_i^{-1} \mu_i \right)$ for the multivariate normal distribution. So we have allocate y_o to group j if
$\left[y_o' \left(\Sigma_j^{-1} \mu_j - \Sigma_i^{-1} \mu_i \right) - \frac{y_o' \left(\Sigma_j^{-1} - \Sigma_i^{-1} \right) y_o}{2} \right] \geqslant \ln H + \frac{1}{2} \ln \left(\frac{|\Sigma_j|}{|\Sigma_i|} \right) + \frac{1}{2} \left(\mu_j \Sigma_j^{-1} \mu_j - \mu_i \Sigma_i^{-1} \mu_i \right)$, otherwise
allocate y_o to group i. The obtained discriminant function is a difference of two Mahalanobis distance. Noting that the discriminant model is quadratic in nature, it can be reduced to the linear discriminant model if $\Sigma_j^{-1} = \Sigma_i^{-1} = \Sigma^{-1}$, that is homogeneous covariances, resulting to the cancellation of the quadratic term in the discriminant model. See text for further modification (Johnson and Wichern, 2006; Hands and Henley, 1997). Note that for $\beta > 1$ the polynomial becomes tedious to evaluate; however restricting the shape parameter to the highest integers $\beta \in Z^+$ we can obtain the following discriminant functions for EPD of shape parameter $\beta = 2$ up to $\beta = 6$ via its polynomial relations defined for multivariate random variables $y_j = x$ in group j and $y_i = w$ in group i respectively.

Proposition 3. The MEPD discriminant function when $\beta = 2$ can be derived as

$$ - \frac{D_Q(xw)}{2} \left[L(x) + M(w) \right] \tag{13} $$

Proof. From (8), we have $-\frac{1}{2} \left[L^2(x) - M^2(w) \right]$ whose discriminant function is a difference of two squares $-\frac{1}{2} \left[L(x) - M(w) \right] \left[L(x) + M(w) \right]$ where $L(x) = (x - \mu_j)' \Sigma_j^{-1} (x - \mu_j)$, $M(w) = (w - \mu_i)' \Sigma_i^{-1} (w - \mu_i)$ and $D_Q(xw) = L(x) - M(w)$. When $\Sigma_j = \Sigma_i$ then (3.9) becomes
$-\frac{D_L(xw)}{2} \left[L(x) + M(w) \right]$.

Proposition 4. The MEPD discriminant function when $\beta = 3$ can be derived as

$$ - \frac{D_Q(xw)}{2} \left[D_Q^2(xw) + 3\rho_{ij}^2 \right] \tag{14} $$

Proof. From (8) when $\beta = 3$ we have $-\frac{1}{2} \left[L^3(x) - M^3(w) \right]$ which by polynomial expansion can be written as

$-\frac{1}{2} \left[L(x) - M(w) \right] \left[(L(x) - M(w))^2 + 3L(x)M(w) \right]$. Note that $L(x)M(w) = (x - \mu_j)' \Sigma_j^{-1} (x - \mu_j)(w - \mu_i)' \Sigma_i^{-1} (w - \mu_i) = \frac{(x - \mu_j)'(x - \mu_j)(w - \mu_i)'(w - \mu_i)}{\Sigma_i \Sigma_j}$ which is our known correlation matrix

$\rho_{ij}^2 = \frac{\left[(x - \mu_j)(w - \mu_i) \right]' \left[(x - \mu_j)(w - \mu_i) \right]}{\Sigma_i \Sigma_j}$. When $\Sigma_j = \Sigma_i$ then (3.9) becomes $-\frac{D_L(xw)}{2} \left[D_L^2(xw) + 3\rho_{ij}^2 \right]$

Proposition 5. The MEPD discriminant function when $\beta = 4$ can be derived as

$$ - \frac{\left[L(x) + M(w) \right] D_Q(xw)}{2} \left[D_Q^2(xw) + 2\rho_{ij}^2 \right] \tag{15} $$

Proof. From (8) when $\beta = 4$, we have $-\frac{1}{2} \left[L^4(x) - M^4(w) \right]$ which by polynomial expansion can be written as $-\frac{1}{2} [L^2(x)]^2 - [M^2(w)]^2$ which becomes a difference of two squares and can be expressed fully as
$-\frac{1}{2} \left[L(x) + M(w) \right] \left[L(x) - M(w) \right] \left([L(x) - M(w)]^2 + 2L(x)M(w) \right)$. Having defined the difference

and product of $L(x)$ and $M(w)$ as $D_Q(xw)$ and ρ_{ij}^2 respectively, then replacing appropriately we have (3.11). When $\Sigma_j = \Sigma_i$ then (14) becomes

$$- \frac{[L(x)+M(w)]D_L(xw)}{2}\left[D_L^2(xw) + 2\rho_{ij}^2\right].$$

Proposition 6. The MEPD discriminant function when $\beta = 5$ can be derived as

$$-\tfrac{1}{2}\left([L(x) + M(w)]^2 D_Q(xw)[D_Q^2(xw) + 2\rho_{ij}^2] - D_Q(xw)\rho_{ij}^2[D_Q^2(xw) + 3\rho_{ij}^2]\right) \tag{16}$$

Proof. From (8) when $\beta = 5$ we have $-\tfrac{1}{2}\left[L^5(x) - M^5(w)\right]$ which by substituting earlier results obtained when $\beta = 3$ and $\beta = 4$ to the expansion $-\tfrac{1}{2}\left(\left[L^4(x) - M^4(w)\right][L(x) + M(w)] - (L(x)M(w)\left[L^3(x) - M^3(w)\right])\right)$ gives (3.12). When $\Sigma_j = \Sigma_i$ then (15) becomes $-\tfrac{1}{2}\left(\left[[L(x) + M(w)]^2 D_L(xw)\left(D_L^2(xw) + 2\rho_{ij}^2\right)\right] - \left[D_L(xw)\rho_{ij}^2\left(D_L^2(xw) + 3\rho_{ij}^2\right)\right]\right).$

Proposition 7. When $\beta = 6$, the MEPD discriminant function is

$$-\frac{1}{2}\left[L(x) + M(w)\right]D_Q(xw)\left(D_Q^2(xw) + 3\rho_{ij}^2\right)\left(D_Q^2(xw) + \rho_{ij}^2\right) \tag{17}$$

Proof. From (8) when $\beta = 6$ we have $-\tfrac{1}{2}\left[L^6(x) - M^6(w)\right]$
$= [L(x) + M(w)][L(x) - M(w)]\left([L(x) - M(w)]^2 + L(x)M(w)\right)\left([L(x) - M(w)]^2 + 3L(x)M(w)\right).$
Substituting the earlier defined expression, we have (3.13). Note that when $\Sigma_j = \Sigma_i$, (16) becomes $-\tfrac{1}{2}\left[L(x) + M(w)\right]D_L(xw)\left(D_L^2(xw) + 3\rho_{ij}^2\right)\left(D_L^2(xw) + \rho_{ij}^2\right).$

Remarks: Similar result can be obtained for β values beyond $\beta = 6$ as the highest positive integer value of β increases. For instance

$\left[L^7(x) - M^7(w)\right] = \left(\left[L^6(x) - M^6(w)\right][L(x) + M(w)]\right) - \left((L(x)M(w)\left[L^5(x) - M^5(w)\right]\right).$
Likewise $\left[L^8(x) - M^8(w)\right] =$
$\left[L^4(x) - M^4(w)\right]\left[\left(\left[L^3(x) - M^3(w)\right][L(x) - M(w)]\right) + \left(\left[L^2(x) - M^2(w)\right][L(x)M(w)]\right)\right].$

Note that using similar approach, we can obtain the discriminant model for $\beta \in Z^+$ as β increases beyond $\beta = 8$. Next we obtain the solution to the discrimination of some practical data with the aim of obtaining the separation boundary (behavioural pattern) with the least misclassification error.

4 APPLICATIONS

In this section we apply the obtained discriminant function and the classification rule from (7)-(11), we proceed to test the performance of our obtained results on some data some are primary data and secondary data already in the literature. The purpose of various applications were to demonstrate that replacement of normal distribution with more flexible generalized form which is exponential power distribution is appropriate in data analysis. It is noteworthy that the exponential power distribution recourse to normal when the shape parameter $\beta = 1$. Therefore, normal is embedded in exponential distribution and also they belong to the same exponential families of distributions. In the Tables, $\overline{x}$ and σ are the estimated sample mean and standard deviations respectively. Also, S and U are the skewness and kurtosis respectively for each variables considered in the analysis.

4.1 Psychological Disorder Data, (Courtesy of Federal Neuropsychiatric Hospital, Abeokuta Medical Records)

The main data that motivated this study were actually samples collected for Federal Neuropsychiatric Hospital, Abeokuta, Ogun State, Nigeria. The data which has to do with the behavioural pattern of some high school teachers, which in turn is having an effect on their productivity. Therefore, the data used in this study was a survey of secondary school

teachers in Ogun state, Nigeria. The survey was to detect cases of unreported psychological disorder tendencies among 393 teachers across secondary schools in Ogun state. All the 393 teachers were screened to find out their general health status using General Health Questionnaire (GHQ). The study proceeded with teachers with GHQ positive while those with negative GHQ status were dropped. For each of the teacher some socio-demographic variables were obtained which includes: Age of teachers, No of children, Length of service, income and Job satisfaction. In respect of the study at hand, out of the 393 teachers captured in the survey, 317 were GHQ negative while the remaining have positive GHQ status. A training set of about 297 teachers from the GHQ negative group and 49 teachers from the GHQ positive group was used to develop the discriminant function meant for onward classification. The remaining teachers with sizes 20 and 27 for GHQ negative and GHQ positive status were used as validating set to determine the accuracy of the discriminant function. The summary of the data is given in Table 1 below. Also a Q-Q probability plots for the sample data to show the deviation from normal distribution are shown in Fig. 1 below, in the figure some data exhibited tails either shorter or longer than the normal distribution, this has necessitated for the intervention and substitution of a generalized case of normal distribution called exponential power which also belong to family of elliptical density. This distribution will give a good account of the deviations in the kurtosis and tails of the distribution. We used the procedure of discrimination described in section 3.0, which depends on the estimation on the shape parameter β could discriminate between teachers with GHQ negative and GHQ positive status based on Age of teachers, No of children, Length of service, income and Job satisfaction. The discriminant function with the least error of misclassification is the main focus of the study. The obtained discriminant model have an apparent error rate of of 51.06% when $\beta = 1$, 48.94% apparent error rate when $\beta = 5$ and when $\beta = 8$ it recorded an apparent error of 36.17%.

4.2 Bumpus Data, (Hands [17])

The Bumpus data of 1898, published by Hands [17] was the another example to be classified. The data is divided into two groups of female sparrows, the first group were those female sparrows that survived severe weather condition, while the second group were those that died as a result of the weather condition. Some features about each sparrow was measured namely: total length of the sparrow; alar, beak, humerus and sternum length were also measured. The summary and the results are given in Table 2, now for each of the two groups, five different variables were measured from each bird. The discriminant model on the training data set was able to correctly classify 69.39% of the data set; given an apparent error rate of 30.61%. In this case the shape parameter $\beta = 1$ gave the best classification for the data. Therefore, exponential power recourse to normal in this sense within the exponential families.

4.3 Margolese Data, (Hands [17])

Also applying our result to the data that have been used in several studies which is the data on discrimination of urinary androsterone and etiocholanolone in healthy heterosexual and homosexual males in mg/24 hours from Margolese, 1970; The discriminant model revealed the same apparent error rate between the normal and the exponential distribution regardless of the shape parameter value used. This suggested that the data obtained by Margolese in 1970 actually followed a normal distribution.

4.4 Job Satisfaction and Income Data, (Soyinka et al. [18])

Next, we investigate the discriminant model between secondary school teachers who are married and those that are yet to marry, in terms of their Job satisfaction and Income (Soyinka et al., 2017). The EPD model predicted accurately 75% of the data with an apparent error rate of 25%; compared to normal distribution with an apparent error rate of 45%. In this case, the EPD

with shape parameter $\beta = 8$ performed better than the normal $\beta = 1$.

4.5 Artisan Data, (Courtesy of Ministry of Youth, Sport and Development, Ogun State)

For further illustration, we then considered developing a discriminant model for the discrimination of artisans data, who are working with government from those who are self employed. The secondary data was obtained from the records of artisans who visited the Ministry of Youth and Sport Development, Ogun state, Nigeria to obtain government approval to practice as artisan within the state. However, due to the fee charged for the test, majority of the artisans usually come for the test, after they have been practising but cannot obtain contract from government or for the purpose of promotion to a higher cadre for those working with government. The trade test of each of the artisans (those working with government and those that are self employed) were used along with their individual monthly saving. The obtained discriminant model depending on the the estimated values for β, the shape parameter, we have an apparent error rate of 6.67% when $\beta = 1$, 5% apparent error rate when $\beta = 6$ and when $\beta = 8$ it recorded an apparent error of 8.33%. The exponential power discrimination when $\beta = 6$ produced the least misclassification error.

Table 1. Summary statistics

Variables	Group 1	Group 2
	GHQ Negative ($n = 297$)	GHQ Positive ($n = 49$)
Age	$\overline{x} = 40.98, \sigma_x = 40.53, S = -0.2541, U = -0.5250$	$\overline{x} = 38.04, \sigma_x = 13.46, S = -0.036, U = -1.34$
No of children	$\overline{x} = 2.687, \sigma_x = 1.2028, Sk = 0.2649, U = -0.1518$	$\overline{x} = 2.68, \sigma_x = 0.6433, S = 0.1273, U = -0.8397$
Length of Service	$\overline{x} = 12.87, \sigma_x = 54.73, S = 0.4044, U = -0.9087$	$\overline{x} = 7.6, \sigma_x = 28.17, S = 0.6788, U = -1.0157$
Salary	$\overline{x} = 54.57, \sigma_x = 650.66, S = 1.014, U = 0.6776$	$\overline{x} = 40.74, \sigma_x = 266.75, S = 1.1326, U = 0.8453$
Job Satis	$\overline{x} = 68.24, \sigma_x = 125.15, S = -0.5601, U = 0.2359$	$\overline{x} = 65.52, \sigma_x = 168.01, S = -0.325, U = -0.4239$

Table 2. Summary statistics

Variables	Group 1	Group 2
	Survived ($n = 8$)	Dead ($n = 8$)
Total length	$\overline{x} = 155.5, \sigma_x = 1.6903, S = 0.0776, U = -1.5682$	$\overline{x} = 160.625, \sigma_x = 3.021, S = -0.2027, U = -1.445$
Alar length	$\overline{x} = 238.81, \sigma_x = 2.532, S = -0.0628, U = -1.2094$	$\overline{x} = 245.75, \sigma_x = 4.2003, S = 0.1531, U = -1.6698$
Beak length	$\overline{x} = 31.21, \sigma_x = 0.7864, S = 0.6973, U = -0.5466$	$\overline{x} = 32.26, \sigma_x = 0.715, S = 0.3891, U = -1.6479$
Humerus length	$\overline{x} = 18.388, \sigma_x = 0.2588, S = -0.7892, U = -1.0601$	$\overline{x} = 19.062, \sigma_x = 0.518, S = -0.1144, U = -1.3745$
Sternum length	$\overline{x} = 20.475, \sigma_x = 0.787, S = 0.737, U = -0.8797$	$\overline{x} = 21.9, \sigma_x = 0.7783, S = -0.2816, U = -1.5509$

Table 3. Summary statistics from various real-life illustrations

Variables	Group 1	Group 2
	androsterone ($n = 11$)	etiocholanolone ($n = 15$)
heterosexual	$\overline{x} = 3.518, \sigma_x = 0.7209, S = -0.1717, U = -1.4233$	$\overline{x} = 2.5, \sigma_x = 0.9227, S = 0.4949, U = -1.1483$
homosexual	$\overline{x} = 2.1, \sigma_x = 0.6899, Sk = 0.578, K = -1.1118$	$\overline{x} = 3.24, \sigma_x = 1.244, S = 0.0077, U = -1.1614$

Table 4. Summary statistics

Variables	Group 1	Group 2
	Never married ($n = 10$)	Married ($n = 10$)
Job satisfaction	$\overline{x} = 46.22, \sigma_x = 29.87, S = 1.133, U = -0.545$	$\overline{x} = 54.61, \sigma_x = 12.7, S = 0.795, U = -0.0925$
Income	$\overline{x} = 67.59, \sigma_x = 3.36, S = -0.1581, U = -1.5497$	$\overline{x} = 67.84, \sigma_x = 2.398, S = -0.8197, U = -0.3529$

Table 5. Summary statistics

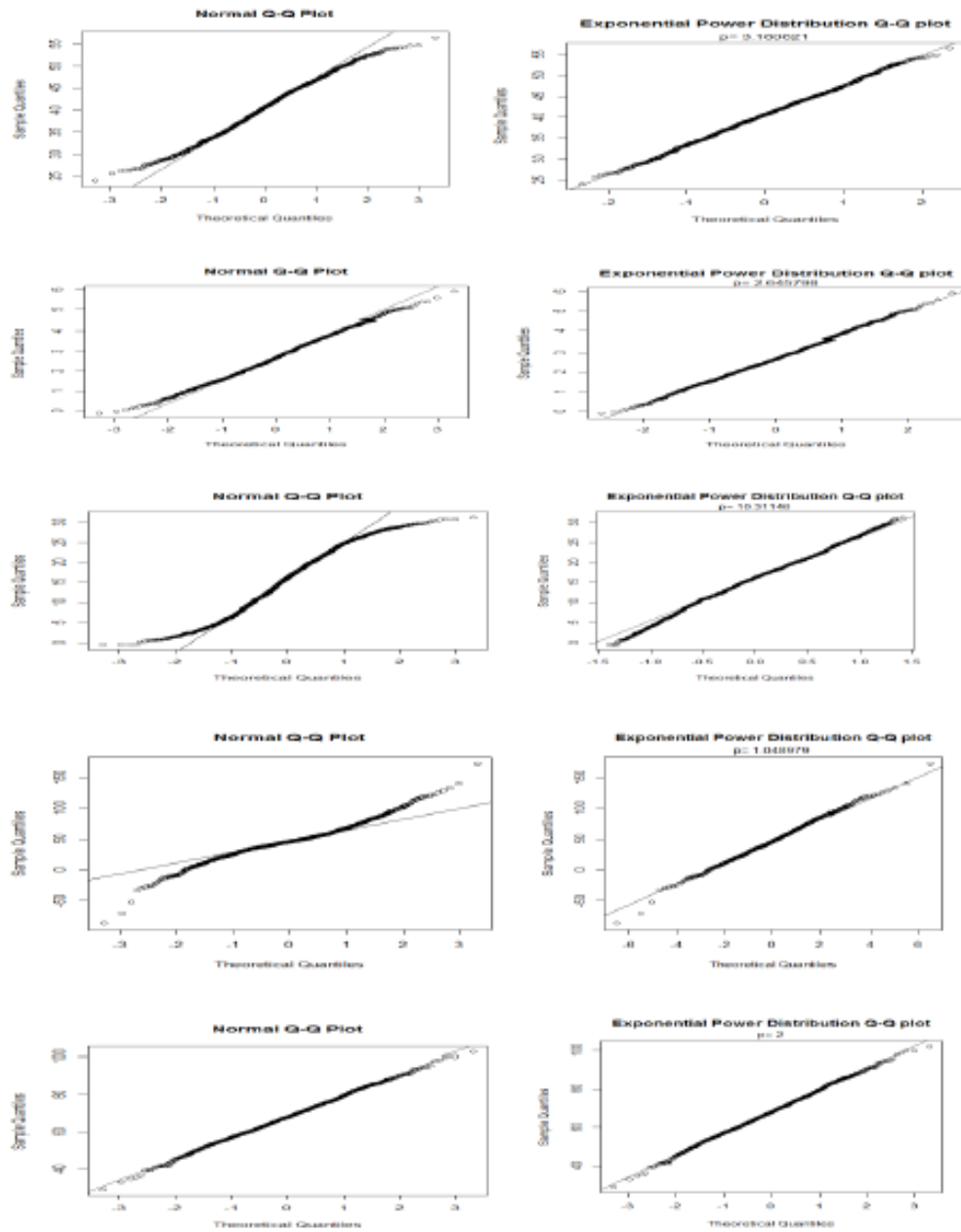| Variables	Group 1	Group 2
	Govt. Artisans ($n = 15$)	Self-employed Artisans ($n = 15$)
Test score	$\overline{x} = 46.31, \sigma_x = 1.2997, S = -0.3637, U = -1.7564$	$\overline{x} = 54.97, \sigma_x = 1.1742, S = -0.4323, U = -0.8975$
Saving	$\overline{x} = 67.53, \sigma_x = 0.8962, Sk = 0.3983, K = -1.0991$	$\overline{x} = 67.98, \sigma_x = 0.7828, S = 0.3573, U = -0.8545$

Fig. 1. Q-Q plots for the psychological disorder data, left is the Q-Q fit with normal distribution while right is the fit with exponential power with different β estimated from each data

5   CONCLUSION

The discrimation model of multivariate exponential power distribution generalizes for the discrimation model of normal and double exponential distribution. The study also established a relationship to the known fisher discriminant function. The various examples used in the study revealed that the more the deviation of the data from normal distribution ($\beta > 1$), the more effective the discriminant model of the EPD. This was however confirmed for small samples > 30. However for sample size above thirty, the exponential and the normal were both effective. The r environment code for multivariate exponential power discrimination model to investigate future data using the model obtained in this study is available. The code can be extended to nth variables.

ACKNOWLEDGEMENT

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

[1] Olosunde AA, Soyinka AT. Discrimination and Classification for exponential power distribution and applications. Electronic Journal of Applied Statistics and Applications. 2020;13(2):284-292.

[2] Aitchison J, Silvey SD. Maximum likelihood estimation of parameters subject to restraints. Ann. Math. Stat. 1958;29:813-829.

[3] Cox DR. Some procedures associated with the logistic qualitative response curve. Research Ropers in Statistics: Festschrift for J. Neyman, (F.N. David, Ed.), Wiley, London. 1966;55-71.

[4] Johnson RA, Wichern DW. Applied multivariate statistical analysis. Englewood Cliffs, NJ: Prentice-Hall, Inc.; 2006.

[5] Hands DJ, Henley WE. Statistical classification methods in consumer credit scoring: A review. J.R. Statistist. Sos. 1997;160(3):523-541.

[6] Day NE, Kerridge DF. A general maximum likelihood discriminant. Biometrics. 1967;23:313-323.

[7] Azzalini A. The multivariate skew normal distribution. BIometrika. 1996;83:715-726.

[8] Azzalini A. References on the skew normal distribution and the related ones; 2003. Available:http://azzalini.stat.unipd.it/SN/list-publ.pdf

[9] Effron B. The efficiency of logistics regression compare to normal discriminant analysis. Journal of American Statistical Association. 1975;70(352):892-898.

[10] Press SJ, Wilson S. Choosing between logistic regression and discriminant analysis. Journal of American Statistical Association. 1978;73(364):699-705.

[11] Gomez E, Gomez-Villegas MA, Marin JM. A multivariate generalization of the power exponential family of distributions. Communications in Statistics, Theory and Methods. 1998;27:589-600.

[12] Lindsey JK. Multivariate Elliptically Contoured Distributions for Repeated Measurements. Biometrics. 1999;55:1277-1280.

[13] Mineo AM, Ruggieri M. A software tool for exponential power distribution: The normalp package. Journal of Statistical Software. 2005;12(4): Available:http://www.jstatsoft.org/

[14] Agro G. Maximum likelihood estimation for the exponential power distribution. Communications in statistics (Simulations and Computations). 1995;24(2):523-536.

[15] Olosunde AA. On exponential power distribution and poultry feeds data: A case study. Journal Iran Statistical Society. 2013;12(2):253-270.

[16] Main P, Navarro H. Conditional specification with exponential power distributions. Communications in Statistics - Theory and Methods. 2010;39(12):2231-2240.

[17] Hands DJ. Discrimination and classification. Wiley; Chichester. 1981;218:76.

[18] Soyinka Taiwo, Olosunde Akin, Oyetola Shola, Akinhanmi Akin. Modelling job satisfaction and age using a composite function of gamma and power transformation function. International Journal of Statistics and Applications. 2017;7(2):107-112.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/67946