Analysis of the characteristic features of the density functions for gamma, Weibull and log-normal distributions through RBF network pruning with QLP

EDWIRDE LUIZ SILVA AND PAULO LISBOA
Departamento de Matemática e Estatística
Universidade Estadual da Paraíba - UEPB
Rua Juvêncio Arruda, S/N Campus Universitário (Bodocongó).
Cep: 58.109 - 790 Campina Grande - Paraíba
BRASIL
School of Computing and Mathematical Sciences
Liverpool John Moores University
Byrom Street, Liverpool, L3 3AF
ENGLAND
http://www.uepb.edu.br
http:/www.ilmu.ac.uk

Abstract: - The Weibull, Gamma and Log-normal densities are similar in shape for the same coefficient of variation, making it difficult to identify the differences between the three densities. This paper is concerned with separating these three probability densities using feature descriptors, identified by pruning a Radial Basis Function (RBF) network using pivoted QLP decomposition generated for each density function with the same mean and coefficients of variation. The QLP method proves efficient for reducing the network size by pruning hidden nodes, resulting is a parsimonious model which identifies four main features (namely kurtosis, skewness, inter-quartile range and mean). This application tool can be used to identify the correct distribution function from empirical data in cases where traditional statistical tests are inconclusive.

Key-Words: - Radial Basis Function, Pivoted QLP Decomposition, probability density function.

1 Introduction

Radial Basis Function neural networks have been widely applied to pattern classification. This paper applies this model to the discrimination, from data, between three similar probability density functions. The characteristic features resulting from the network decomposition using QLP (a lower diagonal matrix L between orthogonal matrices Q and P [2]) for this particular choice of density functions has a very interesting interpretation. The distributions of interest are the Gamma, Lognormal and Weibull density functions within particular parameters designed to make the three distributions particularly difficult to distinguish. The analysis presented in this paper identifies four key discriminant features (kurtosis, skewness, inter-quartile range and mean).

The rest the paper is organized as follows. Section II presents the three densities probabilities, section III presents the detection of the Numerical rank of the QLP, section IV presents the simulated data to test the proposed approach to discrimination, section V reports on the quantitative performance comparisons for the pruned RBF networks, section VI experiment results. Finally, section VII offers conclusions from this study.

2 Probability density function

2.1 The gamma distribution.

A positive random variable X is said to be gamma (α,λ) distributed when it has the probability density

$$\rho(t)=\frac{\lambda^{\alpha}t^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda t},t\geq 0 \qquad (1)$$

Where $\alpha>0$ is the shape parameter and $\lambda>0$ is the scale parameter. The symbol $\Gamma(\alpha)$ denotes the complete gamma function. The mean (E) and the squared coefficients of variation (C_x^2) of a gamma (α,λ) distributed random variable X are given by

$$E_{gamma}(X)=\frac{\alpha}{\lambda} \text{ and } C_x^2=\frac{1}{\alpha}. \qquad (2)$$

The gamma density is always unimodal; that is, the density has only one maximum at $t=(\alpha-1)/\lambda>0$ and next decreases to zero when $t\to\infty$, whereas for the case $C_x^2\geq 1$ the identified variable has its maximum at $t=0$ and thus decreases from $t=0$ onwards. The failure rate function increases from zero

to λ if $C_x^2 < 1$ and decreases from infinity to zero if $C_x^2 > 1$ [1].

2.2 The Weibull distribution.

A positive random variable X is said to be Weibull distributed when it has the probability density

$$\rho(t) = \alpha\lambda(\lambda t)^{\alpha-1}\exp[-(\lambda t)^{\alpha}], t > 0 \qquad (3)$$

Where $\alpha > 0$ shape parameter and $\lambda > 0$ scale parameter. The mean and squared coefficients of variation of the Weibull random variable X are:

$$E_{weibull}(X) = \frac{1}{\lambda}\Gamma\left(1+\frac{1}{\lambda}\right) \quad \text{and} \quad C_x^2 = \frac{\Gamma(1+2/\alpha)}{[\Gamma(1+1/\alpha)]^2} - 1$$

The Weibull density is always unimodal with a maximum at $t = \lambda^{-1}(1-1/\alpha)^{1/\alpha}$ if $C_x^2 < 1 (\alpha > 1)$, and at $t = 0$ if $C_x^2 \geq 1$ $(\alpha \leq 1)$. The failure rate function increases from 0 to infinity if $C_x^2 < 1$ and decreases from infinity to zero if $C_x^2 > 1$.

2.3 The lognormal distribution

The lognormal has shape parameter $\alpha > 0$ and the scale parameter λ may assume each real value. The mean and the squared coefficient of variation of the lognormal distribution are given by

$$E_{\log n} = \exp\left(\lambda + \frac{1}{2}\alpha^2\right) \text{ and } C_x^2 = \exp(\alpha^2) - 1$$

The lognormal density is also unimodal with a maximum at $t = \exp(\lambda - \alpha^2)$.The failure rate function first increases and next decreases to zero when $t \to \infty$ and thus the failure rate only decreases in the long-life range.

The Weibull and gamma densities are similar in shape, and for $C_x^2 < 1$ the lognormal density takes on shapes similar to the Weibull and gamma densities. For the case $C_x^2 \geq 1$ the Weibull and gamma densities have their maximum value at $t = 0$; most outcomes tend to be small and very large outcomes occur only occasionally. The differences between the gamma, Weibull and lognormal densities become most significant in their tail behavior [1]. Distributions with a relatively large tail are called leptokurtic. A distribution with the same kurtosis as the normal distribution is called mesokurtic.

The densities for large t go down $\exp[-\lambda t]$, $\exp[-(\lambda t)^{\alpha}]$ y $\exp[-\frac{1}{2}[\ln(t)-\lambda]^2/\alpha^2]$. Thus, for given values of the mean and the coefficient of variation, the lognormal density always has the longest tail. The gamma density has the second longest tail only if α This is significant for a mean and coefficient of variation, the density lognormal always has a large tail. The density gamma has the second largest tail only if $\alpha > 1$ that is, only if its coefficient of variation is less than 1. In Fig. 1 we illustrate these facts by drawing the gamma, Weibull and lognormal densities for $C_x^2 = 0.25$, where $E(x)$ is taken to be 1.

There are Weibull densities with parameters of shape $\alpha > 0$ and scale parameter $\lambda > 0$ and also gamma with parameters $\alpha > 0$ and $\lambda > 0$ in the interval 0<x<3. There are 200 data created for each of the densities, and $x_i's$ are uniformly random in the interval (0,3).

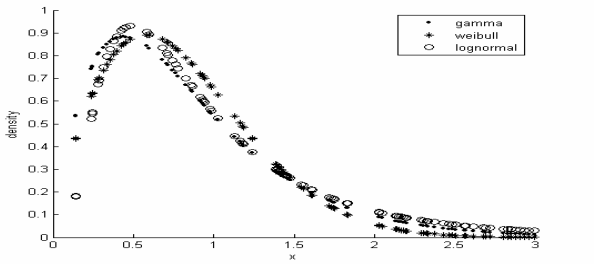

Fig. 1 The gamma, Weibull and lognormal densities

3 Detection of the Numerical rank of the QLP

The algorithm to compute the QLP decomposition can be used as an alternative to SVD and QR [2]. Since the gaps and singular subspace are defined in terms of the singular value decomposition, the natural way to compute them is to compute the singular value decomposition. Unfortunately, the computation of this decomposition is expensive. For this reason, researchers have proposed many alternatives. Of these the pivoted QR decomposition is widely recommended because of its simplicity. The pivoted QR decomposition has certain drawbacks: it gives only fuzzy approximations to the singular values, and fails to provide orthonormal bases for some of the fundamental subspaces [3].

The gap in the L-value of the decomposition provides orthogonal bases of analogues of row, column, and null space provided by A. The decomposition requires no more than twice the work required for a pivoted QR decomposition. The implementation of the

Matlab package permits the QLP: [P,Q,L,pr,pl]=qlp(A) to determine the numerical rank of matrix A.

Thus, we have a simple QLP algorithm as follows

1. **define** matrix A, which consists of the hidden node activations in the RBF network.
2. **calculate** the orthogonal matrices Q and P which reduce the matrix A to lower diagonal form.
3. **identify the** diagonal of lower-triangular matrix L
4. **sort** the diagonal elements by size.

The set of input features used in this analysis are shown in Table 1[4].

Descriptive features	order
Mean	1
Median	2
Q1 – first quartile	3
Q2 – second quartile	4
interquartile range (iqr)	5
sqrt	6
kurtosis	7
skewness	8

4 Design RBF Neural Classifiers

The network consists of n input features x, M hidden units with center C_j and Y output units' y_j. The activation functions φ in the hidden units are the Gaussian functions defined as:

$$\varphi_j(x) = \exp(-\|x - C_j\|^2 / 2\rho_j^2) \qquad (8)$$

The output is calculated by

$$y_K(x) = \sigma\left[\sum_{j=1}^{M} w_{kj}\varphi_j(x) + w_{ko}\right],\ k = 1,2,...,L$$

where σ_k is a linear transfer function and ρ_j is the width of the radial basis function.

4.1 Proposed reduction RBF for the identification of important covariates

The RBF neural network is applied here to distinguish between three density functions in pairs, for example discriminating between the Weibull (α, λ), versus, the lognormal (μ, σ^2).

The first process is the creation of the matrix P and vector T (target matrix) of training for the problem. Each distribution has a matrix of size 200x500 comprising 500 samples of the distribution, each containing 200 observations, where $x_i's$ are uniformly randomly distributed in the (0,3). 8 features are extracted from each sample of 200 points, the features being described in table 1. The sample of 200x1 reduces the data matrix to 8x1 for each distribution, resulting in a data matrix of size 8x1000 for each pair of distributions. The training data matrix uses 8 features calculated from each column of the matrix, resulting in a data matrix of size 8x1000 when the data from the two distributions is concatenated. RBF functions are trained to distinguish between each pair of distributions. Starting with the number of hidden nodes (i.e. radial basis functions) equal to the number of data points, the hidden nodes are reduced using QLP decomposition. The algorithm is based in the reduction of number hidden nodes with QLP.

For each trial, a training set (x_i, p_i) was established, where p is an indicator function representing the density function which generated the sample vector x. An out of sample test set was generated for performance estimation. The data were generated in a similar way to the training data, but with 300 instead of 500 samples.

The QLP algorithm is applied twice: first, to the training data matrix, of size 8x1000, to identify the most important inputs, which are features of the distribution types; then, again, to the matrix of output node weights, of size number of hidden nodes x number of output nodes, this time to identify the most important hidden nodes, i.e. radial basis functions. The key variables identified by the QLP method are those with the highest ranking values of the diagonal matrix L.

The reduction in the number of hidden nodes is carried out by sorting the diagonal elements generated by the QLP decomposition and selecting the corresponding nodes so as to give good classification accuracy with a small number of hidden nodes. The selected RBF networks (case Weibull versus log-normal) and their performance are indicated in Table 3, Show us the 7 neurons which are most significant for distinguish these two densities.

The results of network training are analyzed for various conditions used in the learning process. Thus, for RBF models, the number of hidden neurons is regarded as the result of training (which depends on two key parameters, namely mean squared error goal and spread). Note that of the results reported here, the reported computational effort does not include the centre selection phase and is solely for the training of the RBF with reduction neurons. The algorithm was the same the whole analysis and the stopping criterion for training was that the error function should be less than

$1x10^{-2}$ and a maximum value of 0.5 for the radial function widths. In the Matlab package, the function RANDPERM Random permutation is used to present random form in the network to avoid effects from the order in which the data are presented.

5 Performance estimation

5.1 Kolmogorov_Smirnov

The performs a two-sample Kolmogorov-Smirnov test to compare the distributions of values in the two data distributions D1 and D2 of length n1 and n2, respectively. The test statistic is based on the difference between them. The null hypothesis is that the population has identical distribution functions. Here, each pairs of distributions are Weibull versus lognormal, Weibull versus gamma, lognormal versus, and Normal versus gamma.

5.2 Apparent error rate - APER

We find that the overall performance of the RBF reductions classification approach is fairly good. A good classification procedure should result in few misclassifications. The apparent error rate (Aper) is the fraction of observations in the test set that are misclassified by RBF reductions. $Aper = \sum_{i=1}^{k} n_i / \sum_{i=1}^{k} N_i$, where N_i is the number of observations into probabilities i, $i = 1,2,3$ three cases particulars, and n_i is the number of items misclassified in another probabilities other than i.

5.3 Correct classification rate – CCR.

The validity of the RBF reduction classification procedure can also be measured by calculating the rate of correct classification from the following equation: $CCR = 1 - Aper$. This test represents the item in the test set that are correctly classified. This measure does not depend on the form parent of the distribution; also it can be calculated for any classification procedure.

6 Experimental Results

6.1 Weibull versus log-normal

The QLP shown in Fig. 3 has 4 significant inputs, that are: pr = {**7,1,8,3,5**,4,2,6}.

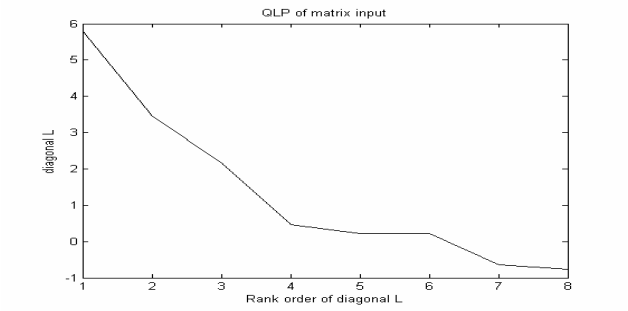

Fig. 4 Decomposition QLP between Weibull versus log-normal.

A preliminary calculation of the mean values of each feature, for each distribution, is shows in Table 2. In the Fig.4 shows us that the kurtosis (7), mean (1), skewness (8), Q1(3) and iqr (5) were the principal features between these two densities. In table 2, note that the kurtosis is highly significant value; this is justified because the Weibull presents a small kurtosis in comparison to the log-normal (Fig.1).

Regarding the training in newrb (Matlab package) the number of neurons was 710, the sum of square error was 0.11 and the number of iterations was 700. Fig. 5 shows us that the QLP decomposition ranks the hidden nodes by order and it is empirically established that when 15 neurons is the threshold for 70% accuracy. Table 7 shows us that of 300 values of Weibull and 300 values of lognormal, the RBF reduction has detected a total of 289 value for the Weibull and 290 for the lognormal. In this case the reduced RBF (15 neurons) has better identification between these two densities highlighted by the characteristic statistics. The apparent error rates were 0.96 for Weibull and log-normal. The order of rows in Table 2 is the order of covariates in Table 1.

Table 2 Descriptive characteristics of the probability density functions for each distribution and their logarithms

Descriptive features	Lognormal	Weibull	Gamma
Mean	1.07	0.99	0.99
Median	0.98	0.96	0.88
Q1	0.22	0.08	0.12
Q2	0.30	0.09	0.18
iqr	0.67	0.73	0.67
sqrt	0.49	0.48	0.52
kurtosis	3.49	2.58	3.70
skewness	0.82	0.36	0.87

Table 3 Neurons that remain after pruning in the case Weibull versus log-normal. The table shows the values of the neurons that remain, from the initial 362 radial basis functions.

154	104	577	641	310	8	663

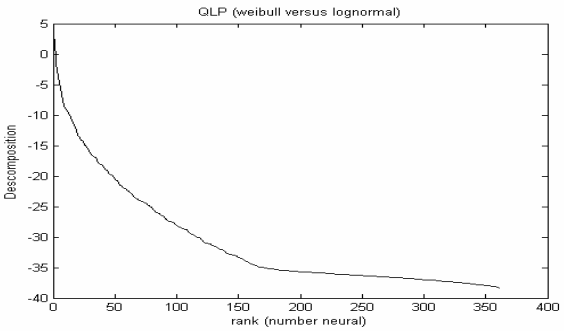

Fig. 5 Decomposition QLP for identification number neurons RBF

6.2 Gamma versus log-normal

In this case, there are 975 iterations, a sum of squared error of 0.34 and 982 neurons. The QLP shown in Fig. 6 has 4 or 5 significant inputs, that are: pr = {7,1,8,3,5}. The same analysis in section 6.1, the principal characteristics were kurtosis (7), mean (1), skewness (8), Q1 (3) and iqr (5). The kurtosis is based on the size of the distribution's tail and skewness (8) presents a significantly high value in comparison the Weibull. This shows that the sensible significant differences between gamma and log-normal are in the tail. Thus justifying that the differences between the gamma, Weibull and lognormal densities become most significant in their tail behavior [1]. This is perfect match of result in the theory.

Table 8 shows us that of 300 values of gamma and 300 values of lognormal, the RBF reduction with 35 neurons has detected a total of 250 value for the gamma and 240 for the log-normal. This case the reduced RBF has good identification between these two densities highlighted by the characteristic statistics with apparent error rate of 0.83 and 0.80 for gamma and log-normal respectively.

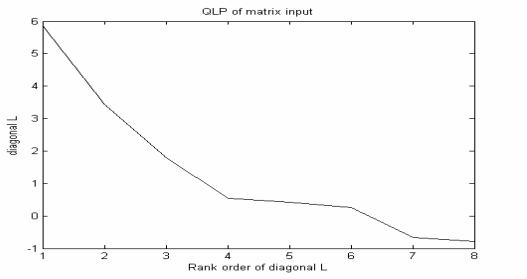

Fig.6 Decomposition QLP of the RBF data matrix of dimension 1000x8.

6.3 Gamma versus Weibull

This case had a large number of neurons, approximately 400 to identify the principal characteristics between two distributions. A QLP decomposition shows that there are 4 important characteristics for accurate discrimination: kurtosis (7), mean(1), skewness (8) and Q1 percentile (0.25)(3). These features were the principal characteristic statistics to differentiate between the gamma versus lognormal and Weibull versus lognormal densities. The RBF had learned these two densities with a sum squared error value of 0.1058, with 976 iteration and total neurons of 993. In the Table 2, the skewness and kurtosis has better differences in comparison the distribution lognormal. The reduced RBF by QLP identified these differences with approximately 400 neurons because the curve of the frequency Weibull is bell-shaped and broadly similar to the gamma distribution; that for the Weibull distribution is less peaked and has longer tails than for the gamma distribution (Fig.1). For both densities there was a not good APER value of 0.76 and 0.60 respectively (Table 9).

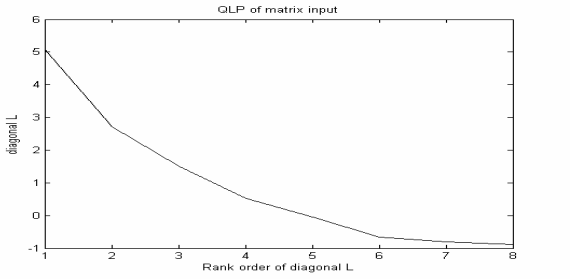

Fig.6 Decomposition QLP of the RBF data matrix of dimension 1000x8.

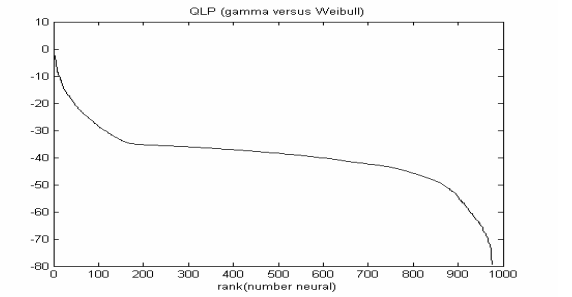

Fig.7 Decomposition QLP for identification number neurons RBF

The final result in this case was worse in comparison others two results. In this case it is necessary for further analysis to reduce the inputs of the network.

6.4 Comparing classical statistics

Suppose that we intend to verify these differences through a classical test multivariate of normality. Kolmogorov-Smirnov normality contrast is used if three functions provide the same distribution. A two-sample

Kolmogorov-Smirnov test is performed to compare the distributions of values in the two data distributions with of length 200, respectively. The alternative hypothesis is that they have different continuous distributions. The result in table 5 shows that the alternative hypothesis can be rejected when utilizing the criterion p-value. The results obtained show the distributions cannot be separated at the 5% level. Classic statistics would use the Kolmogorov-Smirnov normality contrast for analysis if that statistic province were of the same distribution.

Table 5 Resulting Kolmogorov-Smirnov classification of objects into 3 distributions.

Distribution.	p_value
Weibull x gamma	0.3124
Weibull x log-normal	0.1040
Gamma x log-normal	0.3766

The correct classification rate (CCR) is represented with RBF reduction. The small value represents a better classification.

Table 6 Resulting CCR between three density functions in pairs.

Classification	Weib	Gam	Logn
Weib x Gam	0.15	0.61	-
Gam xLogn	-	0.17	0.20
Logn x Weib	0.04	-	0.04

The RBF reduction in the Weibull and Lognormal case has better identification. The apparent error rate (APER) is represented between brackets in Table 6.

7 Conclusion

In this paper, a simple idea of using RBF reduction to identify characteristics among three densities has been developed. The resulting reduced networks show us that the kurtosis, mean and skewness and inter-quartile range were the principal characteristics that accurately separated the three densities, where the identification by classic statistics was very difficult, and the Kolmogorov-Smirnov normality contrast is unable to distinguish between the three distributions. The RBF reduction can be effectively used for identification purposes.

The final results for two densities are shown in Table 7, 8 and 9. It can be seen from The Weibull and log normal show us that 7 neurons with 4 inputs the RBF reduction identified a value of CCR=0.20 (Weibull) and CCR=0.20(log- normal) this is a better identification. The gamma versus log-normal had second better identification with a value of CCR of 0.17 and a value of 0.20 respectively. The differentiation between the gamma and Weibull distributions was less successful.

Table 7 Resulting of RBF reduced of relationship for Webull (Wei) and log-normal(ln)

Case	Number inputs	Number neural	300 Wei (approx)	300 ln (approx)
Network original	8	710	295(.98)	296(.98)
Reduction inputs and neural	4	50	292(.97)	293(.97)
	4	35	291(.97)	294(.98)
	4	20	289(.96)	292(.97)
	4	15	289(.96)	290(.96)
	4	**7**	**267(.87)**	**268(.89)**

Table 8 Resulting of RBF reduced of relationship for gamma (gam) and log-normal.

Case	Number inputs	Number neural	300 gam (approx)	300 ln (approx)
Network original	8	300	274(.91)	250(.81)
Reduction inputs and neural Case	4	300	254 (.85)	247(.82)
	4	100	253(.84)	247(.82)
	4	60	253(.84)	244(.81)
	4	**35**	**250(.83)**	**240(.80)**

Table 9 Resulting of RBF reduced of relationship for gamma and Weibull

Case	Number inputs	Number neural	300 gamma (approx)	300 Weibull (approx)
Network original	8	993	264 (.88)	287(.95)
Reduction inputs and neural Case	**4**	**400**	**230(.76)**	**179(.60)**
	4	300	117(.39)	255(.85)
	4	100	22(.07)	39(.13)

References:

[1] Henk C. Tijms, *A First Course in Stochastic Models*. Vrije Universiteit, Ámsterdam, Wiley, The Netherlands, 2003

[2] Stewart, G.W, *Matrix Algorithms: Basic Decompositions*. SIAM Publications. Philadelphia, U.S.A., 1998.

[3] Sterwart, G.W, On an Inexpensive Triangular Approximation to the Singular Value Decomposition*, A Conference in Honor of G. W. (Pete) Stewart*, 1997. pp.1-16.

[4] López P.C, *Matlab y sus Aplicaciones en las Ciencias y la Ingeniería.* Prentice Hall. Universidad Complutense de Madrid, Madrid, España, 2002