On Achieving Near-Optimal "Anti-Bayesian" Order Statistics-Based Classification for Asymmetric Exponential Distributions

Anu Thomas and B. John Oommen*

School of Computer Science, Carleton University, Ottawa, Canada : K1S 5B6

Abstract. This paper considers the use of Order Statistics (OS) in the theory of Pattern Recognition (PR). The pioneering work on using OS for classification was presented in [1] for the Uniform distribution, where it was shown that optimal PR can be achieved in a counter-intuitive manner, diametrically opposed to the Bayesian paradigm, i.e., by comparing the testing sample to a few samples *distant from the mean* - which is distinct from the optimal Bayesian paradigm. In [2], we showed that the results could be extended for a few *symmetric* distributions within the exponential family. In this paper, we attempt to extend these results significantly by considering asymmetric distributions within the exponential family, for some of which even the closed form expressions of the cumulative distribution functions are not available. These distributions include the Rayleigh, Gamma and certain Beta distributions. As in [1] and [2], the new scheme, referred to as Classification by Moments of Order Statistics (CMOS), attains an accuracy very close to the optimal Bayes' bound, as has been shown both theoretically and by rigorous experimental testing.

Keywords: Classification using Order Statistics (OS), Moments of OS.

1 Introduction

Class conditional distributions have numerous indicators such as their means, variances etc., and these indices have, traditionally, played a prominent role in achieving pattern classification, and in designing the corresponding training and testing algorithms. It is also well known that a distribution has many other characterizing indicators, for example, those related to its Order Statistics (OS). The interesting point about these indicators is that some of them are quite unrelated to the traditional moments themselves, and in spite of this, have not been used in achieving PR. The amazing fact, demonstrated in [3] is that OS can be used in PR, and that such classifiers operate in a completely "anti-Bayesian" manner, i.e., by only considering *certain* outliers of the distribution.

* *Chancellor's Professor* ; *Fellow: IEEE* and *Fellow: IAPR*. This author is also an *Adjunct Professor* with the University of Agder in Grimstad, Norway. The work of this author was partially supported by NSERC, the Natural Sciences and Engineering Research Council of Canada.

R. Wilson et al. (Eds.): CAIP 2013, Part I, LNCS 8047, pp. 368–376, 2013.

Earlier, in [1] and [2], we showed that we could obtain optimal results by an "anti-Bayesian" paradigm by using the OS. Interestingly enough, the novel methodology that we propose, referred to as Classification by Moments of Order Statistics (CMOS), is computationally not any more complex than working with the Bayesian paradigm itself. This was done in [1] for the Uniform distribution and in [2] for certain distributions within the exponential family. In this paper, we attempt to extend these results significantly by considering asymmetric distributions within the exponential family, for some of which even the closed form expressions of the cumulative distribution functions are not available. Examples of these distributions are the Rayleigh, Gamma and certain Beta distributions. Again, as in [1] and [2], we show the completely counter-intuitive result that by working with a *very few* (sometimes as small as two) points *distant* from the mean, one can obtain remarkable classification accuracies, and this has been demonstrated both theoretically and by experimental verification.

2 Optimal OS-Based Classification: The Generic Classifier

Let us assume that we are dealing with the 2-class problem with classes ω_1 and ω_2, where their class-conditional densities are $f_1(x)$ and $f_2(x)$ respectively (i.e, their corresponding distributions are $F_1(x)$ and $F_2(x)$ respectively)[1]. Let ν_1 and ν_2 be the corresponding *medians* of the distributions. Then, classification based on ν_1 and ν_2 would be the strategy that classifies samples based on a *single* OS. We can see that for all symmetric distributions, this classification accuracy attains the Bayes' accuracy.

This result is not too astonishing because the median is centrally located close to (if not exactly) on the mean. The result for higher order OS is actually far more intriguing because the higher order OS are not located centrally (close to the means), but rather distant from the means. In [2], we have shown that for a large number of distributions, mostly from the exponential family, the classification based on *these* OS again attains the Bayes' bound. These results are now extended for asymmetric exponential distributions.

3 The Rayleigh Distribution

The pdf of the Rayleigh distribution, whose applications are found in [4], with parameter $\sigma > 0$ is $\varphi(x, \sigma) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}, x \geq 0$ and the cumulative distribution function is $\Phi(x) = 1 - e^{-x^2/2\sigma^2}, x \geq 0$. The mean, the variance and the median of the Rayleigh distribution are $\sigma\sqrt{\frac{\pi}{2}}$, $\frac{4-\pi}{2}\sigma^2$ and $\sigma\sqrt{ln(4)}$, respectively.

Theoretical Analysis: Rayleigh Distribution - 2-OS. The typical PR problem involving the Rayleigh distribution would consider two classes ω_1 and ω_2 where the class ω_2 is displaced by a quantity θ, and the values of σ are σ_1 and

[1] Throughout this section, we will assume that the *a priori* probabilities are equal.

σ_2 respectively. We consider the scenario when $\sigma_1 = \sigma_2 = \sigma$. Consider the distributions $f(x,\sigma) = \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}$ and $f(x-\theta,\sigma) = \frac{x-\theta}{\sigma^2} e^{\frac{-(x-\theta)^2}{2\sigma^2}}$. In order to do the classification based on CMOS, we shall first derive the moments of the 2-OS for the Rayleigh distribution. The expected values of the first moments of the two OS can be obtained by determining the points where the cumulative distribution function attains the values of $\frac{1}{3}$ and $\frac{2}{3}$ respectively. Let u_1 be the point for the percentile $\frac{2}{3}$ of the first distribution, and u_2 be the point for the percentile $\frac{1}{3}$ of the second distribution. Then, $\int_0^{u_1} \frac{x}{\sigma^2}\ e^{-x^2/2\sigma^2} dx = \frac{2}{3} \implies u_1 = \sigma\sqrt{2\ ln(3)}$ and $u_2 = \theta + \sigma\sqrt{2ln\left(\frac{3}{2}\right)}$.

Theorem 1. *For the 2-class problem in which the two class conditional distributions are Rayleigh and identical, the accuracy obtained by CMOS deviates from the optimal Bayes' bound as the solution of the transcendental equality* $ln\left(\frac{x}{x-\theta}\right) = \frac{-\theta^2+2\theta x}{2\sigma^2}$ *deviates from* $\frac{\theta}{2} + \frac{\sigma}{\sqrt{2}}\left(\sqrt{ln(3)} + \sqrt{ln\left(\frac{3}{2}\right)}\right)$.

Proof. The proof of the theorem can be found in [4]. □

Remark: Another way of comparing the approaches is by obtaining the error difference created by the CMOS classifier when compared to the Bayesian classifier. The details of this can be found in [4].

Theorem 2. *For the 2-class problem in which the two class conditional distributions are Rayleigh and identical, the accuracy obtained by using 2-OS CMOS deviates from the classifier which discriminates based on the distance from the corresponding medians as* $\frac{\theta}{2} + \sigma\sqrt{ln(4)}$ *deviates from* $\frac{\theta}{2} + \frac{\sigma}{\sqrt{2}}\left(\sqrt{ln(3)} + \sqrt{ln\left(\frac{3}{2}\right)}\right)$.

Proof. The proof is omitted here but can be seen in [4]. □

Experimental Results: Rayleigh Distribution - 2-OS. The CMOS classifier was rigorously tested for a number of experiments with various Rayleigh distributions having the identical parameter σ. In every case, the 2-OS CMOS gave almost the same classification as that of the Bayesian classifier. The method was executed 50 times with the 10-fold cross validation scheme. The test results are tabulated in Table 1. The results presented justify the claims of Theorems 1 and 2.

Table 1. A comparison of the accuracy of the Bayesian and the 2-OS CMOS classifier for the Rayleigh Distribution

θ	3	2.5	2	1.5	1
Bayesian	99.1	97.35	94.45	87.75	78.80
CMOS	99.1	97.35	94.40	87.70	78.65

Theoretical Analysis: Rayleigh Distribution - k-OS. We have seen from Theorem 1 that for the Rayleigh distribution, the moments of the 2-OS are sufficient for a near-optimal classification. As in the case of the other distributions, we shall now consider the scenario when we utilize other k-OS. Let u_1 be the point for the percentile $\frac{n+1-k}{n+1}$ of the first distribution, and u_2 be the point for the percentile $\frac{k}{n+1}$ of the second distribution. Then, $\int_0^{u_1} \frac{x}{\sigma^2}\, e^{-x^2/2\sigma^2} dx = \frac{n+1-k}{n+1} \implies u_1 = \sigma\sqrt{2\, ln\left(\frac{n+1}{k}\right)}$ and $u_2 = \theta + \sigma\sqrt{2\, ln\left(\frac{n+1}{n+1-k}\right)}$.

Theorem 3. *For the 2-class problem in which the two class conditional distributions are Rayleigh and identical, a near-optimal Bayesian classification can be achieved by using symmetric pairs of the n-OS, i.e., the* $n-k$ *OS for* ω_1 *and the* k *OS for* ω_2 *if and only if* $\sqrt{ln\left(\frac{n+1}{k}\right)} - \sqrt{ln\left(\frac{n+1}{n+1-k}\right)} < \frac{\theta}{\sigma\sqrt{2}}$. *The classification obtained by CMOS deviates from the optimal Bayes' bound as the solution of the transcendental equality* $ln\left(\frac{x}{x-\theta}\right) = \frac{-\theta^2+2\theta x}{2\sigma^2}$ *deviates from* $\frac{\theta}{2} + \frac{\sigma}{\sqrt{2}}\left[\sqrt{ln\left(\frac{n+1}{k}\right)} + \sqrt{ln\left(\frac{n+1}{n+1-k}\right)}\right]$.

Proof. The proof of this theorem is omitted here, but is included in [4]. □

Experimental Results: Rayleigh Distribution - k-OS. The CMOS method has been rigorously tested with different possibilities of the k-OS and for various values of n, and the test results are given in Table 2. The Bayesian approach provides an accuracy of 82.15%, and from the table, it is obvious that some of the considered k-OSs attains the optimal accuracy and the rest of the cases attain near-optimal accuracy. Also, we can see that the Dual CMOS has to be invoked if the condition stated in Theorem 3 is not satisfied.

Table 2. A comparison of the accuracy of the Bayesian(i.e., 82.15%) and the k-OS CMOS classifier for the Rayleigh Distribution by using the symmetric pairs of the OS for different values of n (where $\sigma = 2$ and $\theta = 1.5$)

No.	Order(n)	Moments	OS_1	OS_2	CMOS	CMOS/ Dual CMOS
1	Two	$(\frac{2}{3}, \frac{1}{3})$	$\sigma\sqrt{(2\, ln\,(\frac{3}{1}))}$	$\theta + \sigma\sqrt{(2\, ln\,(\frac{3}{2}))}$	82.05	CMOS
2	Four	$\left(\frac{5-i}{5}, \frac{i}{5}\right), 1 \le i \le \frac{n}{2}$	$\sigma\sqrt{(2\, ln\,(\frac{5}{2}))}$	$\theta + \sigma\sqrt{(2\, ln\,(\frac{5}{3}))}$	82.0	CMOS
3	Six	$\left(\frac{7-i}{7}, \frac{i}{7}\right) 1 \le i \le \frac{n}{2}$	$\sigma\sqrt{(2\, ln\,(\frac{7}{1}))}$	$\theta + \sigma\sqrt{(2\, ln\,(\frac{7}{6}))}$	81.6	Dual CMOS
4	Eight	$\left(\frac{9-i}{9}, \frac{i}{9}\right), 1 \le i \le \frac{n}{2}$	$\sigma\sqrt{(2\, ln\,(\frac{9}{4}))}$	$\theta + \sigma\sqrt{(2\, ln\,(\frac{9}{5}))}$	82.15	CMOS

Details of when the original OS-based criteria and when the Dual criteria are used, are found in [4]. These are omitted here in the interest of space.

4 The Gamma Distribution

The Gamma distribution is a continuous probability distribution with two parameters - a, a shape parameter and b, a scale parameter. The pdf of the Gamma distribution is $\frac{1}{\Gamma(a)\ b^a}\ x^{a-1}e^{\frac{-x}{b}};\ a>0,\ b>0$, with mean ab and variance ab^2 where a and b are the parameters. Unfortunately, the cumulative distribution function does not have a closed form expression [5,6,7].

Theoretical Analysis: Gamma Distribution. The typical PR problem invoking the Gamma distribution would consider two classes ω_1 and ω_2 where the class ω_2 is displaced by a quantity θ, and in the case analogous to the ones we have analyzed, the values of the scale and shape parameters are identical. We consider the scenario when $a_1 = a_2 = a$ and $b_1 = b_2 = b$. Thus, we consider the distributions: $f(x, 2, 1) = xe^{-x}$ and $f(x-\theta, 2, 1) = (x-\theta)e^{-(x-\theta)}$.

We first derive the moments of the 2-OS, which are the points of interest for CMOS, for the Gamma distribution. Let u_1 be the point for the percentile $\frac{2}{3}$ of the first distribution, and u_2 be the point for the percentile $\frac{1}{3}$ of the second distribution. Then, $\int_0^{u_1} x\ e^{-x}dx = \frac{2}{3} \implies ln(u_1)-2u_1 = ln\left(\frac{1}{3}\right)$ and $ln(u_2-\theta)-2(u_2-\theta) = ln\left(\frac{1}{3}\right)-ln(\theta)$. The following results hold for the Gamma distribution.

Theorem 4. *For the 2-class problem in which the two class conditional distributions are Gamma and identical, the accuracy obtained by CMOS deviates from the accuracy attained by the classifier with regard to the distance from the corresponding medians as* $1.7391 + \frac{\theta}{2}$ *deviates from* $1.6783 + \frac{\theta}{2}$.

Proof. The proof of this theorem can be found in [4]. □

Experimental Results: Gamma Distribution - 2-OS. The CMOS classifier was rigorously tested for a number of experiments with various Gamma distributions having the identical shape and scale parameters $a_1 = a_2 = 2$, and $b_1 = b_2 = 1$. In every case, the 2-OS CMOS gave almost the same classification as that of the classifier based on the central moments, namely, the mean and the median. The method was executed 50 times with the 10-fold cross validation scheme. The test results are tabulated in Table 3.

Table 3. A comparison of the accuracy with respect to the median and the 2-OS CMOS classifier for the Gamma Distribution

n	4.5	4.0	3.5	3.0	2.5	2.0	1.5
Median	94.83	94.25	92.74	90.77	86.51	80.15	72.64
CMOS	95.01	94.49	92.92	90.43	85.99	79.54	72.34

Theorem 5. *For the* 2*-class problem in which the two class conditional distributions are Gamma and identical, a near-optimal Bayesian classification can be achieved by using certain symmetric pairs of the n-OS, i.e., the* $(n-k)^{th}$ *OS for* ω_1 *(represented as* u_1*) and the* k^{th} *OS for* ω_2 *(represented as* u_2*) if and only if* $u_1 < u_2$.

Proof. The proof of this theorem is included in [4]. □

Experimental Results: Gamma Distribution - k-OS. The CMOS method has been rigorously tested for numerous symmetric pairs of the k-OS and for various values of n, and a subset of the test results are given in Table 4. Experiments have been performed for different values of θ, and we can see that the CMOS attained near-optimal Bayes' bound. Also, we can see that the Dual CMOS has to be invoked if the condition stated in Theorem 5 is not satisfied.

Table 4. A comparison of the k-OS CMOS classifier when compared to the Bayes' classifier and the classifier with respect to median and mean for the Gamma Distribution for different values of n. In each column, the value which is near-optimal is rendered bold.

No.	**Classifier**	**Moments**	$\theta = $ **4.5**	**4.0**	**3.5**	**3.0**	**2.5**	**2.0**
1	Bayes	-	97.06	95.085	93.145	90.68	86.93	81.53
2	Mean	-	96.165	94.875	92.52	88.335	83.105	77.035
3	Median	-	90.04	93.57	92.735	90.775	86.275	80.115
4	2-OS	$(\frac{2}{3}, \frac{1}{3})$	95.285	93.865	92.87	90.61	86.085	79.48
5	4-OS	$(\frac{4}{5}, \frac{1}{5})$	95.905	94.605	93.11	89.57	84.68	22.125
6	4-OS	$(\frac{3}{5}, \frac{2}{5})$	95.185	93.675	92.82	**90.855**	86.02	**80.32**
7	6-OS	$(\frac{6}{7}, \frac{1}{7})$	96.405	**95.01**	92.125	88.005	17.29	23.565
8	6-OS	$(\frac{5}{7}, \frac{2}{7})$	95.47	94.11	**93.135**	90.16	85.495	79.55
9	6-OS	$(\frac{4}{7}, \frac{3}{7})$	95.135	93.625	92.78	90.745	**86.135**	80.165
10	8-OS	$(\frac{8}{9}, \frac{1}{9})$	**96.815**	94.895	91.555	13.095	19.41	24.06
11	8-OS	$(\frac{7}{9}, \frac{2}{9})$	95.8	94.445	93.11	89.885	84.81	78.535
12	8-OS	$(\frac{5}{9}, \frac{4}{9})$	95.135	93.625	92.735	90.7	86.085	80.045

5 The Beta Distribution

The Beta distribution is a family of continuous probability distributions defined in $(0, 1)$ parameterized by two shape parameters α and β. The distribution can take different shapes based on the specific values of the parameters. If the parameters are identical, the distribution is symmetric with respect to $\frac{1}{2}$. Further, if $\alpha = \beta = 1$, $B(1, 1)$ becomes $U(0, 1)$. The pdf of the Beta distribution is $f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)}\ x^{\alpha-1}(1-x)^{\beta-1}$. The mean and the variance of the distribution are $\frac{\alpha}{\alpha+\beta}$ and $\frac{\alpha\beta}{(\alpha+\beta)^2\ (\alpha+\beta+1)}$ respectively. We consider the case when $\alpha = \beta > 1$. Earlier, in paper [3], when we first introduced the concept of CMOS-based PR, we had analyzed the 2-OS and k-OS CMOS for the Uniform distribution, and had provided the corresponding theoretical analysis and the experimental results. We had concluded that, for the 2-class problem in which the two class conditional distributions are Uniform and identical, CMOS can, indeed, attain the optimal Bayes' bound. So, in this paper, to avoid repetition, we skip the analysis for the Beta distribution, B(1,1), as this case reduces to the analysis for Uniform U(0,1). Thus, we reckon that the first of these cases (i.e., when $\alpha = 1$ and $\beta = 1$) as being closed. We also discussed the symmetric Beta distribution when the values of the shape parameters α and β are identical in [4]. In this paper, we now move on to the unimodal Beta distribution characterized by the shape parameters $\alpha > 1$ and $\beta > 1$, $\alpha \neq \beta$.

Theoretical Analysis: Beta Distribution ($\alpha > 1, \beta > 1$) - 2-OS. Consider the two classes ω_1 and ω_2 where the class ω_2 is displaced by a quantity θ. In this section, we consider the case when the shape parameters take the values $\alpha > 1$ and $\beta > 1$, and for the interest of preciseness[2], we consider the case when $\alpha = 2$ and $\beta = 5$. Then, the distributions are $f(x, 2, 5) = 30x(1-x)^4$ and $f(x-\theta, 2, 5) = 30(x-\theta)(1-x+\theta)^4$.

We first derive the moments of the 2-OS, namely o_1 and o_2 where o_1 represents the point for the percentile $\frac{2}{3}$ of the first distribution, and o_2 represents the point for the percentile $\frac{1}{3}$ of the second distribution. Then, $\int_0^{o_1} 30x(1-x)^4 dx = \frac{2}{3}$ and $\int_0^{o_2} 30(x-\theta)(1-x+\theta)^4 dx = \frac{1}{3}$.

These positions o_1 and o_2 can be obtained by making use of the built-in functions available in standard software packages as $o_1 = 0.34249$ and $o_2 = \theta + 0.1954$. Thus, our aim is to show that the classification based on these points can attain near optimal accuracies when compared to the accuracy obtained by the classifier with regard to the medians, the most central points of the distributions.

Theorem 6. *For the 2-class problem in which the two class conditional distributions are Beta(α, β) ($\alpha > 1,\ \beta > 1$) and identical with $\alpha = 2$ and $\beta = 5$, the accuracy obtained by CMOS deviates from the accuracy attained by the classifier with regard to the distance from the corresponding medians as the areas under the error curves deviate from the positions $0.26445 + \frac{\theta}{2}$ and $0.2689 + \frac{\theta}{2}$.*

Proof. The proof of this theorem is omitted here, but can be found in [4]. □

Experimental Results: Beta Distribution ($\alpha > 1, \beta > 1$) - 2-OS. The CMOS has been rigorously tested for various Beta distributions with 2-OS. For each of the experiments, we generated 1,000 points for the classes ω_1 and ω_2 characterized by $B(x, 2, 5)$ and $B(x-\theta, 2, 5)$ respectively. We then performed the classification based on the CMOS strategy and with regard to the medians of the distributions. In every case, CMOS was compared with the accuracy obtained with respect to the medians for different values of θ, as tabulated in Table 5. The results were obtained by executing each algorithm 50 times using a 10-fold cross-validation scheme. The quality of the classifier is obvious.

Table 5. A comparison of the accuracy of the 2-OS CMOS classifier with the classification with respect to the medians for the Beta Distribution for different values of θ

θ	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8
Median	89.625	92.9	94.3	95.525	97.3	97.975	98.375	99.05	99.15
CMOS	89.475	92.775	94.525	95.75	97.3	98.05	98.375	99.2	99.225

[2] Any analysis will clearly have to involve specific values for α and β. The analyses for other values of α and β will follow the same arguments and are not included here.

Theoretical Analysis: Beta Distribution ($\alpha > 1, \beta > 1$) - k-OS. We have seen in Theorem 6 that the 2-OS CMOS can attain a near-optimal classification when compared to the classification obtained with regard to the medians of the distributions. We shall now prove that the k-OS CMOS can also attain almost indistinguishable bounds for some symmetric pairs of the n-OS. The formal theorem, proven in [4], follows.

Theorem 7. *For the 2-class problem in which the two class conditional distributions are Beta(α, β) ($\alpha > 1, \beta > 1$) and identical with $\alpha = 2$, $\beta = 5$, a near-optimal classification can be achieved by using certain symmetric pairs of the n-OS, i.e., the $(n-k)^{th}$ OS for ω_1 (represented as o_1) and the k^{th} OS for ω_2 (represented as o_2) if and only if $o_1 < o_2$. If $o_1 > o_2$, the CMOS classifier uses the* Dual *condition, i.e., the k OS for ω_1 and the $n - k$ OS for ω_2.*

□

Experimental Results: Beta Distribution ($\alpha > 1, \beta > 1$) - k-OS. The CMOS method has been rigorously tested for certain symmetric pairs of the k-OS and for various values of n, and the test results are given in Table 6.From the table, we can see that CMOS attained a near-optimal Bayes' accuracy when $o_1 < o_2$. Also, we can see that the Dual CMOS has to be invoked if $o_1 > o_2$.

Table 6. A comparison of the k-OS CMOS classifier when compared to the classifier with respect to means and medians for the Beta Distribution for different values of n. The scenarios for the *Dual* condition are specified by "(D)".

No.	Classifier	Moments	$\theta = 0.35$	0.45	0.55	0.65	0.75	0.85
1	Mean	-	85.325	92.575	96.55	98.3	99.4	99.475
2	Median	-	86.675	92.775	95.525	97.975	99.05	99.275
3	2-OS	$(\frac{2}{3}, \frac{1}{3})$	86.2	92.575	95.75	98.05	99.2	99.275
4	4-OS	$(\frac{4}{5}, \frac{1}{5})$	85.375	92.525	96.225	98.225	99.325	99.475
5	4-OS	$(\frac{3}{5}, \frac{2}{5})$	86.475	**92.775**	95.6	98.05	99.125	99.275
6	6-OS	$(\frac{6}{7}, \frac{1}{7})$	85.2 (D)	92.425	**96.475**	**98.35**	99.45	99.625
7	6-OS	$(\frac{5}{7}, \frac{2}{7})$	86.125	92.625	96.0	98.075	99.2	99.275
8	6-OS	$(\frac{4}{7}, \frac{3}{7})$	86.55	**92.775**	95.525	97.975	99.125	**99.75**

6 Conclusions

In this paper, we have shown that optimal classification for symmetric distributions and near-optimal bound for asymmetric distributions can be attained by an "anti-Bayesian" approach, i.e., by working with a *very few* (sometimes as small as two) points *distant* from the mean. This scheme, referred to as CMOS, Classification by Moments of Order Statistics, operates by using these points determined by the *Order Statistics* of the distributions. In this paper, we have proven the claim for some distributions within the exponential family, and the theoretical results have been verified by rigorous experimental testing. Our results for classification using the OS are both pioneering and novel.

References

1. Thomas, A., Oommen, B.J.: Optimal “Anti-Bayesian” Parametric Pattern Classification Using Order Statistics Criteria. In: Alvarez, L., Mejail, M., Gomez, L., Jacobo, J. (eds.) CIARP 2012. LNCS, vol. 7441, pp. 1–13. Springer, Heidelberg (2012)
2. Thomas, A., Oommen, B.J.: Optimal “Anti-Bayesian” Parametric Pattern Classification for the Exponential Family Using Order Statistics Criteria. In: Campilho, A., Kamel, M. (eds.) ICIAR 2012, Part I. LNCS, vol. 7324, pp. 11–18. Springer, Heidelberg (2012)
3. Thomas, A., Oommen, B.J.: The Fundamental Theory of Optimal “Anti-Bayesian” Parametric Pattern Classification Using Order Statistics Criteria. Pattern Recognition 46, 376–388 (2013)
4. Oommen, B.J., Thomas, A.: Optimal Order Statistics-based “Anti-Bayesian” Parametric Pattern Classification for the Exponential Family. Pattern Recognition (accepted for publication, 2013)
5. Krishnaih, P.R., Rizvi, M.H.: A Note on Moments of Gamma Order Statistics. Technometrics 9, 315–318 (1967)
6. Tadikamalla, P.R.: An Approximation to the Moments and the Percentiles of Gamma Order Statistics. Sankhya: The Indian Journal of Statistics 39, 372–381 (1977)
7. Young, D.H.: Moment Relations for Order Statistics of the Standardized Gamma Distribution and the Inverse Multinomial Distribution. Biometrika 58, 637–640 (1971)