

“Anti-Bayesian” *Parametric* Pattern Classification Using Order Statistics Criteria for Some Members of the Exponential Family*

B. John Oommen[†] and A. Thomas[‡]

Abstract

This paper submits a comprehensive report of the use of Order Statistics (OS) for parametric Pattern Recognition (PR) for various distributions within the exponential family. Although the field of parametric PR has been thoroughly studied for over five decades, the use of the OS of the distributions to achieve this has not been reported. The pioneering work on using OS for classification was presented earlier for the Uniform distribution and for some members of the exponential family, 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. Apart from the results for the Gaussian and double-exponential which are merely cited here, our new results include the Rayleigh, Gamma and certain Beta distributions. The new scheme, referred to as Classification by Moments of Order Statistics (CMOS), has an accuracy that attains the Bayes’ bound for symmetric distributions, and is, otherwise, very close to the optimal Bayes’ bound, as has been shown both theoretically and by rigorous experimental testing. The results here also give a theoretical foundation for the families of Border Identification (BI) algorithms reported in the literature.

Keywords : *Pattern Classification, Prototype Reduction Schemes, Classification by Moments of Order Statistics*

1 Introduction

The theory of *statistical* pattern classification is founded on two sets of distributions, namely the *a priori* distributions of the various classes, and the class conditional distributions for the features. Since the *a priori* probabilities are rather straightforward quantities, in reality, the basis

*The authors are grateful for the partial support provided by NSERC, the Natural Sciences and Engineering Research Council of Canada. A preliminary version of this paper was presented as a Keynote/Plenary talk at CIARP’12, the 2012 Iberoamerican Congress on Pattern Recognition, Buenos Aires, Argentina, in September 2012.

[†]This author dedicates this paper to his friends Kåre and Eva Smith, to whom he owes his life in more ways than one. Thanks, Kåre and Eva. This author can be contacted at: *Chancellor’s Professor; Fellow: IEEE and Fellow: IAPR*, School of Computer Science, Carleton University, Ottawa, Canada : K1S 5B6. This author is also an *Adjunct Professor* with the University of Agder in Grimstad, Norway. E-mail address: oommen@scs.carleton.ca.

[‡]This author can be contacted at: School of Computer Science, Carleton University, Ottawa, Canada. E-mail address: athomas1@scs.carleton.ca.

for *statistical* Pattern Recognition (PR) is that the individual classes are characterized by their own *specific* class conditional *distributions*, and their corresponding characteristics. The question of which characterizing indicators are used in achieving the PR is, truly, a study in itself.

Statisticians understand that all distributions have numerous indicators such as their means, variances etc.. Consequently, researchers in PR have, traditionally, used *these* indicators to play a prominent role in achieving pattern classification, and in designing the corresponding training and testing algorithms. In the field of PR, however, there are some families of indicators that have noticeably been *uninvestigated*. Indeed, it is 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 main question that has excited our interest [22] is whether these indicators/indices possess any potential in PR.

The amazing answer to this question 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 distributions. This must be contrasted with Bayesian classifiers which attain the optimal lower bound, and that often reduces to testing the sample point using the corresponding distances/norms to the *means* or the “central points” of the distributions. To be more specific, within a Bayesian paradigm, if we are allowed to compare the testing sample with only *a single* point in the feature space from each class, the *optimal* Bayesian strategy would be to achieve this based on the (Mahalanobis) distance from the corresponding means. The reader should observe that, in this context, the mean, in one sense, is the most *central* point in the respective distribution. The norms themselves are distribution dependent, and so, for example, one uses the Mahalanobis distance for Gaussian distributions. Earlier, in [23], we showed that we could obtain optimal results by the above-mentioned “anti-Bayesian” paradigm by using the OS. Indeed, our counter-intuitive result is that by working with a *very few* points *distant* from the mean, one can obtain remarkable classification accuracies. The number of points can sometimes be as small as *two*. Further, if these points are determined by the *Order Statistics* of the distributions, the accuracy of our method, referred to as Classification by Moments of Order Statistics (CMOS), attains the optimal Bayes’ bound.

The interesting aspect of the study of CMOS is that it is closely related to other families of non-parametric methods in PR. For decades since the initial formulation of PR as a research field, researchers have attempted to develop efficient classification methods in which the schemes achieve their task based on a *subset* of the training patterns. These are commonly referred to as “Prototype Reduction Schemes” (PRS)[10, 25]. For the sake of our work, a PRS will be considered to be a generic method for reducing the number of training vectors, while simultaneously attempting to guarantee that the classifier built on the reduced design set performs as well, or nearly as well, as the classifier built on the original design set [14]. Thus, instead of considering all the training patterns

for the classification, a subset of the whole set is selected based on certain criteria. The learning (or training) is then performed on this reduced training set, which is also called the “Reference” set. This Reference set not only contains the patterns which are closer to the true discriminant’s boundary, but also the patterns from the other regions of the space that can adequately represent the entire training set. Border Identification (BI) algorithms, which are a subset of PRSs, work with a Reference set which only contains “border” points¹. Recent research [17] has shown that for overseeing the task of achieving the classification, the samples extracted by a BI scheme, and which lie *close to the boundaries* of the discriminant function, have significant information when it concerns the classification ability of the classifier. Although this is quite amazing, *the formal analytical reason for this is yet unproven*. Our research seems to close the conceptual gap, because the CMOS classifiers truly utilize “border” points (which could almost be considered to be outliers) to yield near-optimal accuracy.

We conclude by mentioning that as far as we know, like the results in [23] which were both pioneering and novel, the results submitted here represent the first application of OS in PR for the spectrum of distributions within the exponential family.

2 Relevant Background Areas

2.1 Prototype Reduction Schemes

To quote Bezdek[16], “zillions of PRS techniques” have developed over the years, and it is clearly impossible to survey all of these here. These include the Condensed Nearest Neighbor (CNN) rule [12], the Reduced Nearest Neighbor (RNN) rule [11], the Prototypes for Nearest Neighbor (PNN) classifiers [3], the Selective Nearest Neighbor (SNN) rule [20], the Edited Nearest Neighbor (ENN) rule [5], Vector Quantization (VQ) etc..

While some of the above techniques merely *select* a subset of the existing patterns as prototypes, other techniques *create* new prototypes so as to represent *all* the existing patterns in the best manner. Of the above-listed PRS techniques, the CNN, RNN, SNN and ENN merely *select* prototypes from the existing patterns, while the PNN and VQ create new prototypes that collectively represent the entire training set. Comprehensive surveys of the state-of-the-art in PRSs can be found in [10, 13, 25]. The formal algorithms² are also found in [22].

2.2 Border Identification Algorithms

Border Identification (BI) algorithms form a distinct subset of PRSs. The aim of the BI algorithms is also to obtain a Reference set of points close to the discriminant function that can perform near-optimal classification. Duch [7] and Foody [9] proposed schemes to achieve this. But as the

¹In the interest of brevity, detailed description of PRS and BI algorithms are omitted here. They are found in [23].

²A copy of the PhD proposal can be found at <http://people.scs.carleton.ca/~athomas1/Proposal.pdf>.

patterns of the Reference Set described in [7] and [9] are only the “near” borders, they do not have the potential to represent the entire training set, and thus do not perform well. In order to compete with other classification strategies, it has been shown that we need to also include the set of “far” borders to the Reference set [17]. A detailed description of traditional BI algorithms namely Duch’s approach, Foody’s algorithm and the Border Identification in Two Stages can be found in [23].

The CMOS scheme proposed in [23] and developed here, give a rationale for BI algorithms.

2.3 Order Statistics

Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be a univariate random sample of size n that follows a continuous distribution function Φ , where the probability density function (pdf) is $\varphi(\cdot)$. Let $\mathbf{x}_{1,n}, \mathbf{x}_{2,n}, \dots, \mathbf{x}_{n,n}$ be the corresponding Order Statistics³ (OS). The r^{th} OS, $\mathbf{x}_{r,n}$, of the set is the r^{th} smallest value among the given random variables [1]. The pdf of $\mathbf{y} = \mathbf{x}_{r,n}$ is given by:

$$f_{\mathbf{y}}(y) = \frac{n!}{(r-1)!(n-r)!} \{\Phi(y)\}^{r-1} \{1 - \Phi(y)\}^{n-r} \varphi(y),$$

where $r = 1, 2, \dots, n$. The reasoning for the above expression is straightforward and is omitted here. It is found in [23].

Although the distribution $f_{\mathbf{y}}(y)$ contains all the information resident in \mathbf{y} , the literature characterizes the OS in terms of quantities which are of paramount importance, namely its moments [24], as briefly cited below.

Using the distribution $f_{\mathbf{y}}(y)$, one can see that the k^{th} moment of $\mathbf{x}_{r,n}$ can be formulated as:

$$E[\mathbf{x}_{r,n}^k] = \frac{n!}{(r-1)!(n-r)!} \int_{-\infty}^{+\infty} y^k \Phi(y)^{r-1} (1 - \Phi(y))^{n-r} \varphi(y) dy,$$

provided that both sides of the equality exist [2, 19].

The fundamental theorem concerning the OS that we invoke is found in many papers [18, 19, 24]. The result is merely cited below inasmuch as the details of the proof are irrelevant and outside the scope of this study. The theorem, proven in [18], can be summarized as follows.

Theorem 1. *Let $n \geq r \geq k + 1 \geq 2$ be integers. Then, since Φ is a nondecreasing and right-continuous function from $\mathbb{R} \rightarrow \mathbb{R}$, $\Phi(\mathbf{x}_{r,n})$ is uniform in $[0, 1]$. If we now take the k^{th} moment of $\Phi(\mathbf{x}_{r,n})$, it has the form:*

$$E[\Phi^k(\mathbf{x}_{r,n})] = \frac{B(r+k, n-r+1)}{B(r, n-r+1)} = \frac{n! (r+k-1)!}{(n+k)! (r-1)!} \quad (1)$$

where $B(a, b)$ denotes the Beta function, and $B(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$. □

³We are grateful to the anonymous Referee who gave us the pointer to [1]. This subsection has also been modified based on this reference.

The above fundamental result can also be used for characterization purposes as follows [18]. Let $n \geq r \geq k + 1 \geq 2$ be integers, with Φ being nondecreasing and right-continuous. Let G be *any* nondecreasing and right-continuous function from $\mathbb{R} \rightarrow \mathbb{R}$ on the same support as Φ . The relation

$$E[G^k(\mathbf{x}_{r,n})] = \frac{n! (r + k - 1)!}{(n + k)! (r - 1)!} \quad (2)$$

holds if and only if $\forall x, \Phi(x) = G(x)$. In other words, $\Phi(\cdot)$ is the unique function that satisfies Eq. (2), implying that every distribution is characterized by the moments of its OS.

The implications of the above are the following:

1. If $n = 2$, implying that only *two* samples are drawn from \mathbf{x} , we can deduce from Eq. (1) that:

$$E[\Phi^1(\mathbf{x}_{1,2})] = \frac{1}{3}, \implies E[\mathbf{x}_{1,2}] = \Phi^{-1}\left(\frac{1}{3}\right), \text{ and} \quad (3)$$

$$E[\Phi^1(\mathbf{x}_{2,2})] = \frac{2}{3}, \implies E[\mathbf{x}_{2,2}] = \Phi^{-1}\left(\frac{2}{3}\right). \quad (4)$$

Thus, from a computational perspective, the first moment of the first and second 2-order OS would be the values where the cumulative distribution Φ equal $\frac{1}{3}$ and $\frac{2}{3}$ respectively.

2. For any $n > 2$, implying that we are considering the k^{th} -OS from n samples drawn from \mathbf{x} , we can deduce from Eq. (1) that:

$$E[\Phi^1(\mathbf{x}_{k,n})] = \frac{k}{n+1}, \implies E[\mathbf{x}_{k,n}] = \Phi^{-1}\left(\frac{k}{n+1}\right), \text{ and} \quad (5)$$

$$E[\Phi^1(\mathbf{x}_{n-k,n})] = \frac{n-k+1}{n+1}, \implies E[\mathbf{x}_{n-k,n}] = \Phi^{-1}\left(\frac{n-k+1}{n+1}\right). \quad (6)$$

Again, computationally, the first moment of the k^{th} and $n - k^{th}$ n -order OS would be the values where the cumulative distribution Φ equal $\frac{k}{n+1}$ and $\frac{n-k+1}{n+1}$ respectively.

Although the analogous expressions can be derived for the higher order *moments* of these OS, for the rest of this paper we shall merely focus on the *first* moment of these OS, and derive the consequences of using them in classification.

3 Optimal Bayesian Classification using *Two* Order Statistics

3.1 The Generic Classifier

Having characterized the moments of the OS of arbitrary distributions, we shall now consider how they can be used to design a 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)⁴. Let ν_1 and ν_2 be the corresponding *medians* of the distributions. Then, classification based on ν_1 and ν_1 would be the strategy that classifies samples based on a *single* OS. We shall show the fairly straightforward result that for all symmetric distributions, the classification accuracy of this classifier 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. Consequently, we shall show that for a large number of distributions, mostly from the exponential family, the classification based on *these* OS again attains the Bayes' bound.

The results we obtained for some distributions of the exponential family namely Uniform, Doubly Exponential and Gaussian distributions are mentioned here, but the details of the analysis and experiments are omitted in the interest of brevity even though they can be found in [23].

3.2 The Uniform Distribution

The continuous Uniform distribution is characterized by a constant function $U(a, b)$, where a and b are the minimum and the maximum values that the random variable \mathbf{x} can take. If the class conditional densities of ω_1 and ω_2 are uniformly distributed,

$$f_1(x) = \begin{cases} \frac{1}{b_1 - a_1} & \text{if } a_1 \leq x \leq b_1; \\ 0 & \text{if } x < a_1 \text{ or } x > b_1, \text{ and} \end{cases}$$

$$f_2(x) = \begin{cases} \frac{1}{b_2 - a_2} & \text{if } a_2 \leq x \leq b_2; \\ 0 & \text{if } x < a_2 \text{ or } x > b_2. \end{cases}$$

The results obtained for Uniform distribution are as follows:

Theorem 2. *For the 2-class problem in which the two class conditional distributions are Uniform and identical, CMOS, the classification using two OS, attains the optimal Bayes' bound.* \square

The proof of the theorem can be found in [23] and omitted here to avoid repetition. However, the argument is based on the following. By virtue of Eq. (3) and (4), the expected values of the first moment of the 2-order OS for uniformly distributed random variables can be seen to be $E[\mathbf{x}_{1,2}] = \frac{1}{3}$, and $E[\mathbf{x}_{2,2}] = \frac{2}{3}$. Similarly, for the distribution $U(h, 1 + h)$, the expected values are $E[\mathbf{x}_{1,2}] = h + \frac{1}{3}$ and $E[\mathbf{x}_{2,2}] = h + \frac{2}{3}$. If we perform the classification with respect to these CMOS points, we have shown in [23] that the optimal Bayes' bound can be attained.

⁴Throughout this section, we will assume that the *a priori* probabilities are equal. If they are unequal, the above densities must be weighted with the respective *a priori* probabilities.

This method has been rigorously tested for various uniform distributions with 2-OS. In the interest of brevity, a few typical results are given below.

For each of the experiments, we generated 1,000 points for the classes ω_1 and ω_2 characterized by $U(0, 1)$ and $U(h, 1 + h)$ respectively. We then invoked a classification procedure by utilizing the Bayesian and the CMOS strategies. In every case, CMOS was compared with the Bayesian classifier for different values of h , as tabulated in Table 1. The results in Table 1 were obtained by executing each algorithm 50 times using a 10-fold cross-validation scheme.

h	0.95	0.90	0.85	0.80	0.75	0.70
Bayesian	97.58	95.1	92.42	90.23	87.82	85.4
CMOS	97.58	95.1	92.42	90.23	87.82	85.4

Table 1: Classification of Uniformly distributed classes by the CMOS 2-OS method for different values of h .

Theorem 3. *For the 2-class problem in which the two class conditional distributions are Uniform and identical as $U(0,1)$ and $U(h, 1+h)$, 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 $k > \frac{(n+1)(1-h)}{2}$. If $k < \frac{(n+1)(1-h)}{2}$, optimal Bayesian classification can be achieved by using the Dual symmetric pairs of the n -OS, i.e., the k OS for ω_1 and the $n - k$ OS for ω_2 . \square*

The proof of this theorem⁵ is omitted here in the interest of brevity and can be found in [23]. But, the argument can be summarized as follows. By virtue of Eq. (3) and (4), the expected values of the first moment of the k -order OS have the form $E[\mathbf{x}_{k,n}] = \frac{k}{n+1}$. Our claim is based on the classification in which we can choose any of the symmetric pairs of the n -OS, i.e., the $n - k$ OS for ω_1 and the k OS for ω_2 , whose expected values are $\frac{n-k+1}{n+1}$ and $h + \frac{k}{n+1}$ respectively. At any instant, we must enforce the ordering of the OS of the two distributions, and this requires that:

$$\frac{n - k + 1}{n + 1} < h + \frac{k}{n + 1} \implies k > \frac{(n + 1)(1 - h)}{2}. \quad (7)$$

If we perform the classification with regard to these symmetric points, provided the condition $k > \frac{(n+1)(1-h)}{2}$ is enforced, the optimal Bayes' bound can be attained. The fact that the Dual criterion is valid when the condition is not satisfied can also be proven with identical arguments, and the details are again omitted to avoid repetition.

The experimental results obtained for k -OS are depicted in Table 2.

⁵This theorem is slightly more powerful than the one proved in [23] because it considers both the cases when the condition is satisfied and when it is violated.

Trial No.	Order(n)	Moments	OS_1	OS_2	CMOS	CMOS/Dual CMOS
1	Two	$\{\frac{i}{3} 1 \leq i \leq 2\}$	$\frac{2}{3}$	$h + \frac{1}{3}$	90.23	CMOS
2	Three	$\{\frac{i}{4} 1 \leq i \leq 3\}$	$\frac{3}{4}$	$h + \frac{1}{4}$	90.23	CMOS
3	Four	$\{\frac{i}{5} 1 \leq i \leq 4\}$	$\frac{4}{5}$	$h + \frac{1}{5}$	90.23	CMOS
4	Five	$\{\frac{i}{6} 1 \leq i \leq 5\}$	$\frac{4}{6}$	$h + \frac{2}{6}$	90.23	CMOS
5	Six	$\{\frac{i}{7} 1 \leq i \leq 6\}$	$\frac{4}{7}$	$h + \frac{2}{7}$	90.23	CMOS
6	Seven	$\{\frac{i}{8} 1 \leq i \leq 7\}$	$\frac{5}{8}$	$h + \frac{3}{8}$	90.23	CMOS
7	Eight	$\{\frac{i}{9} 1 \leq i \leq 8\}$	$\frac{6}{9}$	$h + \frac{3}{9}$	90.23	CMOS
8	Nine	$\{\frac{i}{10} 1 \leq i \leq 9\}$	$\frac{7}{10}$	$h + \frac{3}{10}$	90.23	CMOS
9	Ten	$\{\frac{i}{11} 1 \leq i \leq 10\}$	$\frac{10}{11}$	$h + \frac{1}{11}$	90.23	Dual CMOS
10	Ten	$\{\frac{i}{11} 1 \leq i \leq 10\}$	$\frac{9}{11}$	$h + \frac{2}{11}$	90.23	CMOS
11	Ten	$\{\frac{i}{11} 1 \leq i \leq 10\}$	$\frac{7}{11}$	$h + \frac{4}{11}$	90.23	CMOS
12	Ten	$\{\frac{i}{11} 1 \leq i \leq 10\}$	$\frac{6}{11}$	$h + \frac{5}{11}$	90.23	CMOS

Table 2: Results of the classification of Uniformly distributed classes obtained by using the symmetric pairs of the OS for different values of n . The value of h was set to be 0.8. Note that in every case, the accuracy attained the Bayes' value whenever the conditions stated in Theorem 3 were satisfied.

3.3 The Laplace (or Doubly-Exponential) Distribution

The *Laplace distribution* is a continuous uni-dimensional pdf named after Pierre-Simon Laplace. It is sometimes called the *doubly exponential distribution*, because it can be perceived as being a combination of two exponential distributions, with an additional location parameter, spliced together back-to-back.

If the class conditional densities of ω_1 and ω_2 are doubly exponentially distributed,

$$f_1(x) = \frac{\lambda_1}{2} e^{-\lambda_1|x-c_1|}, \quad -\infty < x < \infty, \text{ and}$$

$$f_2(x) = \frac{\lambda_2}{2} e^{-\lambda_2|x-c_2|}, \quad -\infty < x < \infty,$$

where c_1 and c_2 are the respective means of the distributions. By elementary integration and straightforward algebraic simplifications, the variances of the distributions can be seen to be $\frac{2}{\lambda_1^2}$ and $\frac{2}{\lambda_2^2}$ respectively.

Theorem 4. *For the 2-class problem in which the two class conditional distributions are Doubly Exponential and identical, CMOS, the classification using two OS, attains the optimal Bayes' bound.*

□

By virtue of Eq. (3) and (4), 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 $\frac{1}{3}$ and $\frac{2}{3}$. 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. By straightforward integrations and simplifications, these points are obtained as:

$$u_1 = c_1 - \frac{1}{\lambda_1} \log\left(\frac{2}{3}\right), \quad (8)$$

$$u_2 = c_2 + \frac{1}{\lambda_2} \log\left(\frac{2}{3}\right). \quad (9)$$

Classification can be performed with regard to these points, and the results are depicted in Table 3. From the experimental results and the theoretical analysis, we conclude that the expected values

c₁	0	0	0	0	0	0	0	0	0
c₂	10	9	8	7	6	5	4	3	2
Bayesian	99.75	99.65	99.25	99.05	98.9	97.85	96.8	94.05	89.9
CMOS	99.75	99.65	99.25	99.05	98.9	97.85	96.8	94.05	89.9

Table 3: Classification for the Doubly Exponential Distribution by the CMOS.

of the first moment of the 2-OS of the Doubly Exponential distribution can always be utilized to yield the exact accuracy as that of the Bayes' bound, even though this is a drastically anti-Bayesian operation.

Theorem 5. *For the 2-class problem in which the two class conditional distributions are Doubly Exponential and identical, the 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 $\log\left(\frac{2k}{n+1}\right) > \frac{c_1 - c_2}{2}$. Again, if the latter condition is violated, optimal Bayesian classification can be achieved by using the Dual symmetric pairs of the n -OS, i.e., the k OS for ω_1 and the $n - k$ OS for ω_2 . □*

As in 2-OS, by virtue of Eq. (3) and (4), the CMOS positions can be obtained by straightforward integrations and simplifications as:

$$u_1 = c_1 - \log\left(\frac{2k}{n+1}\right), \quad (10)$$

$$u_2 = c_2 + \log\left(\frac{2k}{n+1}\right). \quad (11)$$

CMOS can achieve optimal classification with these positions, and the formal proof of this assertion is found in [23]. This has also been rigorously tested with different possibilities of k -OS

and for various values of n , and the test results are given in Table 4.

No.	Order(n)	Moments	OS_1	OS_2	CMOS	CMOS/Dual CMOS
1	Two	$(\frac{2}{3}, \frac{1}{3})$	$c_1 - \frac{1}{\lambda_1} \log(\frac{2}{3})$	$c_2 + \frac{1}{\lambda_2} \log(\frac{2}{3})$	95.2	CMOS
2	Three	$(\frac{3}{4}, \frac{1}{4})$	$c_1 - \frac{1}{\lambda_1} \log(\frac{1}{2})$	$c_2 + \frac{1}{\lambda_2} \log(\frac{1}{2})$	95.2	CMOS
3	Four	$(\frac{5-i}{5}, \frac{i}{5}), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1} \log(\frac{4}{5})$	$c_2 + \frac{1}{\lambda_2} \log(\frac{4}{5})$	95.2	CMOS
4	Five	$(\frac{6-i}{6}, \frac{i}{6}), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1} \log(\frac{1}{3})$	$c_2 + \frac{1}{\lambda_2} \log(\frac{1}{3})$	95.2	CMOS
5	Six	$(\frac{7-i}{7}, \frac{i}{7}), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1} \log(\frac{4}{7})$	$c_2 + \frac{1}{\lambda_2} \log(\frac{4}{7})$	95.2	CMOS
6	Seven	$(\frac{8-i}{8}, \frac{i}{8}), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1} \log(\frac{1}{4})$	$c_2 + \frac{1}{\lambda_2} \log(\frac{1}{4})$	95.2	CMOS
7	Eight	$(\frac{9-i}{9}, \frac{i}{9}), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1} \log(\frac{2}{9})$	$c_2 + \frac{1}{\lambda_2} \log(\frac{2}{9})$	95.2	Dual CMOS
8	Eight	$(\frac{9-i}{9}, \frac{i}{9}), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1} \log(\frac{4}{9})$	$c_2 + \frac{1}{\lambda_2} \log(\frac{4}{9})$	95.2	CMOS
9	Nine	$(\frac{10-i}{10}, \frac{i}{10}), 1 \leq i \leq \frac{n}{2}$	$c_1 - \frac{1}{\lambda_1} \log(\frac{3}{5})$	$c_2 + \frac{1}{\lambda_2} \log(\frac{3}{5})$	95.2	CMOS

Table 4: Results of the classification obtained by using the symmetric pairs of the OS for different values of n . The value of c_1 and c_2 were set to be 0 and 3. Note that in every case, the accuracy attained the Bayes' value whenever the conditions stated in Theorem 5 were satisfied.

3.4 The Gaussian Distribution

The Normal (or Gaussian) distribution is a continuous probability distribution that is often used as a first approximation to describe real-valued random variables that tend to cluster around a single mean value. It is particularly pertinent due to the so-called Central Limit Theorem. The univariate pdf of the distribution is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Theorem 6. *For the 2-class problem in which the two class conditional distributions are Gaussian and identical, CMOS, the classification using 2-OS, attains the optimal Bayes' bound.* \square

The moments of the OS for the Normal distribution can be determined from the generalized expression:

$$E[\mathbf{x}_{k,n}^r] = \frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{+\infty} x^r \Phi^{k-1}(x) (1 - \Phi(x))^{n-k} \varphi(x) dx,$$

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $\Phi(x) = \int_{-\infty}^x \varphi(t) dt$. From this expression, the expected values of the first moment of the 2-OS can be determined as:

$$E[\mathbf{x}_{1,2}] = \mu - \frac{\sigma}{\sqrt{2\pi}}, \quad \text{and} \quad (12)$$

$$E[\mathbf{x}_{2,2}] = \mu + \frac{\sigma}{\sqrt{2\pi}}, \quad (13)$$

as shown in [2]. With these values, we showed that for identically distributed classes differing only in the means, the CMOS with 2-OS can yield the same Bayesian accuracy. The analysis and proofs can be found in [23]. The CMOS classifier was rigorously tested for a number of experiments with various Gaussian distributions having means μ_1 and μ_2 . In every case, the 2-OS CMOS gave *exactly* the same accuracy as that of the Bayesian classifier. The method was executed 50 times with the 10-fold cross validation scheme. The test results are displayed in Table 5, whence the power of the scheme is clear.

μ_1	0	0	0	0	0	0
μ_2	14	12	10	8	6	4
Bayesian	99.2	96.5	95.1	95	90	85
CMOS	99.2	96.5	95.1	95	90	85

Table 5: Classification of Normally distributed classes by the CMOS 2-OS method for different means.

If we intend to use higher order CMOS pairs, one can easily see that the $\left(\frac{k}{n+1}\right)^{th}$ and the $\left(\frac{n-k+1}{n+1}\right)^{th}$ percentiles of the Normal function are precisely the CMOS points which are to be used in the corresponding classification strategy. Again, if the $\left(\frac{n-k+1}{n+1}\right)^{th}$ percentile of ω_1 is greater than the $\left(\frac{k}{n+1}\right)^{th}$ percentile of ω_2 , the optimal bound is attained by invoking the corresponding *Dual* classifier.

Using these, the CMOS method has been rigorously tested with different possibilities of k -OS and for various values of n , and the test results are given in Table 6.

No.	Order(n)	Moments	CMOS	CMOS/Dual CMOS
1	Two	$\left(\frac{2}{3}, \frac{1}{3}\right)$	91.865	CMOS
2	Four	$\left(\frac{4}{5}, \frac{1}{5}\right)$	91.865	CMOS
3	Six	$\left(\frac{6}{7}, \frac{1}{7}\right)$	91.865	CMOS
4	Eight	$\left(\frac{8}{9}, \frac{1}{9}\right)$	91.865	CMOS
5	Ten	$\left(\frac{10}{11}, \frac{1}{11}\right)$	91.865	Dual CMOS
6	Ten	$\left(\frac{9}{11}, \frac{2}{11}\right)$	91.865	CMOS
7	Twelve	$\left(\frac{12}{13}, \frac{1}{13}\right)$	91.865	Dual CMOS
8	Twelve	$\left(\frac{10}{13}, \frac{3}{13}\right)$	91.865	CMOS

Table 6: Results of the classification obtained by using the symmetric pairs of the k -OS for different values of n .

This concludes our brief report on the study of symmetric distributions within the exponential family, that have already been reported earlier [23].

4 The Rayleigh Distribution

The Rayleigh distribution is a continuous probability distribution which is often observed when the overall magnitude of a vector is related to its directional components. The pdf of the Rayleigh distribution, with parameter $\sigma > 0$ is:

$$\varphi(x, \sigma) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}, \quad x \geq 0,$$

and the cumulative distribution function is:

$$\Phi(x) = 1 - e^{-x^2/2\sigma^2}, \quad x \geq 0.$$

The mean, the variance and the median of the Rayleigh distribution are $\mu(\mathbf{x}) = \sigma\sqrt{\frac{\pi}{2}}$, $Var(\mathbf{x}) = \frac{4-\pi}{2}\sigma^2$ and $Median(x) = \sigma\sqrt{\ln 4}$, respectively.

4.1 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 σ_2 respectively. As in the previous cases, 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. By virtue of Eq. (3) and (4), 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:

$$\begin{aligned} \int_0^{u_1} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} dx &= \frac{2}{3} \implies 1 - e^{-\frac{u_1^2}{2\sigma^2}} = \frac{2}{3} \\ &\implies u_1 = \sigma\sqrt{2 \ln 3}. \end{aligned} \tag{14}$$

Using a similar argument, u_2 can be evaluated as:

$$u_2 = \theta + \sigma\sqrt{2 \ln \frac{3}{2}}. \tag{15}$$

We now derive the result concerning the efficiency of the CMOS when compared to the Bayesian

classifier.

Theorem 7. *For the 2-class problem in which the two class conditional distributions are Rayleigh and identical, the accuracy obtained by CMOS, the classification using two OS, deviates from the optimal Bayes' bound as the solution of the transcendental equality $\ln \frac{x}{x-\theta} = \frac{-\theta^2+2\theta x}{2\sigma^2}$ deviates from $\frac{\theta}{2} + \frac{\sigma}{\sqrt{2}} \left(\sqrt{\ln 3} + \sqrt{\ln \frac{3}{2}} \right)$.*

Proof. Without loss of generality, let the distributions of ω_1 and ω_2 be $R(x, \sigma)$ and $R(x - \theta, \sigma)$, where σ is the identical scale parameter. Then, to get the Bayes' classifier, we argue that:

$$\begin{aligned}
p(x|\omega_1)P(\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(x|\omega_2)P(\omega_2) &\implies \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{x-\theta}{\sigma^2} e^{-\frac{(x-\theta)^2}{2\sigma^2}} \\
&\implies \frac{x}{x-\theta} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} e^{\frac{-(x-\theta)^2}{2\sigma^2} + \frac{x^2}{2\sigma^2}} \\
&\implies \ln \frac{x}{x-\theta} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{-\theta^2 + 2\theta x}{2\sigma^2}. \tag{16}
\end{aligned}$$

The discriminant is then the solution to the transcendental equation:

$$\ln \frac{x}{x-\theta} = \frac{-\theta^2 + 2\theta x}{2\sigma^2}. \tag{17}$$

We now consider the classification with respect to the expected values of the moments of the 2-OS, u_1 and u_2 , where as per Eq. (14) and (15), $u_1 = \sigma\sqrt{2 \ln 3}$ and $u_2 = \theta + \sigma\sqrt{2 \ln \frac{3}{2}}$. The discriminant enforced by the 2-OS classifier satisfies:

$$D(x, u_1) = D(x, u_2). \tag{18}$$

The condition imposed by Eq. (18) leads to the following:

$$\begin{aligned}
D(x, u_1) = D(x, u_2) &\implies D\left(x, \sigma\sqrt{2 \ln 3}\right) = D\left(x, \theta + \sigma\sqrt{2 \ln \frac{3}{2}}\right) \\
&\implies 2x = \theta + \sigma\sqrt{2 \ln 3} + \sigma\sqrt{2 \ln \frac{3}{2}} \\
&\implies x = \frac{\theta}{2} + \frac{\sigma}{\sqrt{2}} \left(\sqrt{\ln 3} + \sqrt{\ln \frac{3}{2}} \right). \tag{19}
\end{aligned}$$

The difference in the errors of the two classifiers is clearly related to differences in the corresponding discriminant functions specified by Eq. (17) and (19). The result follows. \square

Remark:

Another way of comparing the approaches is by obtaining the error difference created by the CMOS classifier when compared to the Bayesian classifier. In Figure 1, the small area marked as

“Error Difference” is the difference between the probability of error formed by the CMOS classifier when compared to the Bayesian counterpart, and we can evaluate this area by using the corresponding definite integrals. As Eq. (17) is transcendental in nature, the only way to find the Bayesian classifier is to resort to a numerical strategy, for example, by using a Taylor series expansion. The area depicting the differences in classification accuracy (in percentage) is reported in Table 7. Since the accuracy of the Taylor’s expansion depends on the point around which the expansion is done, in Table 7, we have also recorded this point, i.e., the one around which the Taylor’s expansion has been invoked for each specific scenario. From this table, we can see that the CMOS classifier is bounded by an error difference of less than 0.15%, which is truly, negligible.

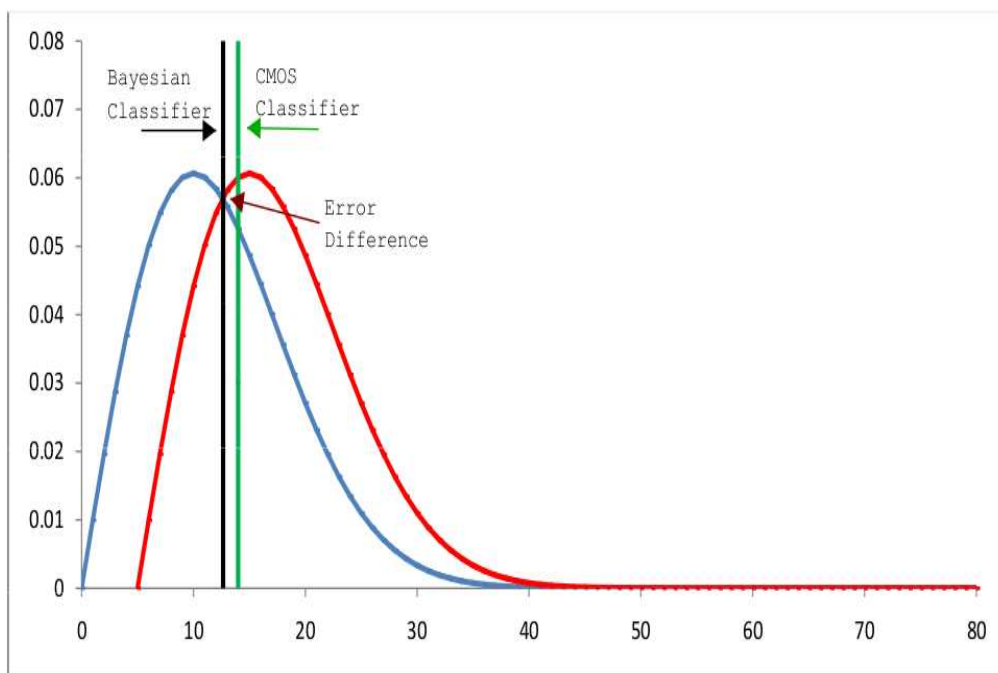


Figure 1: The differences of the error probability quantified by the differences between the areas under the curves of the resulting errors.

θ	1	1.5	2	2.5	3
a	3	4	5	5.3	6.5
Max. Bounded Error(in %)	0.15	0.06	0.05	0.001	0

Table 7: The maximum bounded error by the CMOS classifier when compared to the Bayesian classifier, for different values of θ of the Rayleigh Distribution. In each case, $\sigma = 2$, and the Taylor’s expansion was invoked around the point a .

Theorem 8. For the 2-class problem in which the two class conditional distributions are Rayleigh and identical, CMOS, the accuracy obtained by CMOS, the classification using two OS, deviates from the classifier which discriminates based on the distance from the corresponding **medians** as $\frac{\theta}{2} + \sigma\sqrt{\ln 4}$ deviates from $\frac{\theta}{2} + \frac{\sigma}{\sqrt{2}} \left(\sqrt{\ln 3} + \sqrt{\ln \frac{3}{2}} \right)$.

Proof. As the curve of the Rayleigh distribution is not symmetric, for the present analysis, we shall consider the scenario that the classification is done based on the median, which is the most central point of the distribution, other than the mean. In order to prove the theorem, we shall first show that when the class conditional distributions are Rayleigh and identical, the accuracy of the corresponding near-optimal discriminant obtained by a comparison to the corresponding medians is almost equal to the accuracy of the CMOS. Again, as in the case of Theorem 7, as the equations are transcendental, we can consider the classification based on the medians of the given distributions, namely $\nu_1 = \sigma\sqrt{\ln 4}$ and $\nu_2 = \theta + \sigma\sqrt{\ln 4}$, respectively. The classification will be based on the distances that the testing point has with respect to the respective medians. Thus,

$$\begin{aligned}
D(x, \nu_1) < D(x, \nu_2) &\implies x - \sigma\sqrt{\ln 4} < \theta + \sigma\sqrt{\ln 4} - x \\
&\implies 2x < \theta + 2\sigma\sqrt{\ln 4} \\
&\implies x < \frac{\theta}{2} + \sigma\sqrt{\ln 4}.
\end{aligned} \tag{20}$$

The discriminant function with regard to the medians of the distributions is: $x = \frac{\theta}{2} + \sigma\sqrt{\ln 4}$.

We now consider the classification with respect to the expected values of the moments of the 2-OS, u_1 and u_2 , where as per Eq. (14) and (15), $u_1 = \sigma\sqrt{2 \ln 3}$ and $u_2 = \theta + \sigma\sqrt{2 \ln \frac{3}{2}}$. The discriminant enforced by 2-OS CMOS is:

$$D(x, u_1) = D(x, u_2). \tag{21}$$

This equation simplifies to:

$$\begin{aligned}
D(x, u_1) = D(x, u_2) &\implies D\left(x, \sigma\sqrt{2 \ln 3}\right) = D\left(x, \theta + \sigma\sqrt{2 \ln \frac{3}{2}}\right) \\
&\implies 2x = \theta + \sigma\sqrt{2 \ln 3} + \sigma\sqrt{2 \ln \frac{3}{2}} \\
&\implies x = \frac{\theta}{2} + \frac{\sigma}{\sqrt{2}} \left(\sqrt{\ln 3} + \sqrt{\ln \frac{3}{2}} \right).
\end{aligned} \tag{22}$$

The difference in the errors of the two classifiers is clearly related to differences in the corresponding discriminant functions specified by Eq. (20) and (22). Hence the theorem. \square

Remark: As in Theorem 7, we can show that Eqs. (20) and (22) are almost identical by obtaining

the error difference created by the CMOS classifier when compared to the classifier based on the corresponding medians for different values of θ . The area depicting the differences in classification accuracy (in percentage) is reported⁶ in Table 8. From this table, we can see that the CMOS classifier is bounded by an error difference of less than 0.42%, which is again, negligible.

θ	1	1.5	2	2.5	3	3.5
Maximum Bounded Error(in %)	0.40	0.34	0.20	0.14	0	0

Table 8: Maximum bounded error by the CMOS classifier when compared to the classifier obtained with regard to the median, for different values of θ of the Rayleigh Distribution. In each case, $\sigma = 1$.

Corollary 1. *By virtue of the almost-identical nature of the two expressions for the Rayleigh distribution, the classification using the proximity to the median is almost indistinguishable from that of the Bayesian classifier.*

Proof. This result is an indirect implied consequence of Theorems 7 and 8. □

4.2 Data Generation: Rayleigh Generation

To experimentally verify our results, we made use of a Uniform (0,1) random variable generator to generate data values that follow a Rayleigh distribution. The expression $\mathbf{x} = \sigma\sqrt{-2 \ln (1 - \mathbf{u})}$, where σ is the parameter and \mathbf{u} is a random variate from the Uniform distribution $U(0, 1)$, generates Rayleigh distributed values [6].

4.3 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 9 and justify Theorem 7.

The results presented justify the claims of Theorems 7 and 8.

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

Table 9: A comparison of the accuracy of the Bayesian and the 2-OS CMOS classifier for the Rayleigh Distribution.

⁶Since the expressions are directly solvable, we do not need to resort to a Taylor’s expansion in this case.

4.4 Theoretical Analysis: Rayleigh Distribution - k -OS

We have seen from Theorem 7 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. The formal result pertaining to this is given in Theorem 9.

Theorem 9. *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 $\sqrt{\ln \frac{n+1}{k}} - \sqrt{\ln \frac{n+1}{n+1-k}} < \frac{\theta}{\sigma\sqrt{2}}$. If this condition is violated, the CMOS classifier uses the Dual condition, i.e., the k OS for ω_1 and the $n - k$ OS for ω_2 . In both these cases, the classification obtained by CMOS deviates from the optimal Bayes' bound as the solution of the transcendental equality $\ln \frac{x}{x-\theta} = \frac{-\theta^2+2\theta x}{2\sigma^2}$ deviates from $\frac{\theta}{2} + \frac{\sigma}{\sqrt{2}} \left[\sqrt{\ln \frac{n+1}{k}} + \sqrt{\ln \frac{n+1}{n+1-k}} \right]$.*

Proof. First of all, we invoke the result of Corollary 1 that classification based on the proximity to the median is almost equivalent to the Bayesian classification. We shall now show the result for the k -OS, that the classification is almost identical to the classification based on the medians. Before proceeding further, we have to show that the expected values of the first moment of the k -order OS for the Rayleigh distribution have the form $E[\mathbf{x}_{k,n}] = \sigma\sqrt{2 \ln \frac{n+1}{k}}$. Let u_1 be the point for the percentile $\frac{n+1-k}{n+1}$ (the $(n - k)^{th}$ -OS) of the first distribution, and u_2 be the point for the percentile $\frac{k}{n+1}$ (the k -OS) of the second distribution. Then:

$$\begin{aligned} \int_0^{u_1} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} dx &= \frac{n+1-k}{n+1} \implies 1 - e^{-\frac{u_1^2}{2\sigma^2}} = \frac{n+1-k}{n+1} \\ &\implies u_1 = \sigma\sqrt{2 \ln \frac{n+1}{k}}. \end{aligned} \quad (23)$$

Using a similar argument, u_2 can be evaluated as:

$$u_2 = \theta + \sigma\sqrt{2 \ln \frac{n+1}{n+1-k}}. \quad (24)$$

Our present claim is based on the classification in which we can choose any of the symmetric pairs of the n -OS, i.e., the $n - k$ OS for ω_1 and the k OS for ω_2 , where these quantities are $\sigma\sqrt{2 \ln \frac{n+1}{k}}$ and $\theta + \sigma\sqrt{2 \ln \frac{n+1}{n+1-k}}$ respectively.

It is obvious that an OS value can correctly classify a testing point only when their positions

have the correct ordering, i.e., $u_1 < u_2$. We can resolve this condition by solving this inequality as:

$$\begin{aligned}
& \sigma \sqrt{\left(2 \ln \frac{n+1}{k}\right)} < \theta + \sigma \sqrt{\left(2 \ln \frac{n+1}{n+1-k}\right)} \\
\implies & \sigma \sqrt{2} \left[\sqrt{\ln \frac{n+1}{k}} - \sqrt{\ln \frac{n+1}{n+1-k}} \right] < \theta \\
\implies & \sqrt{\ln \frac{n+1}{k}} - \sqrt{\ln \frac{n+1}{n+1-k}} < \frac{\theta}{\sigma \sqrt{2}}. \tag{25}
\end{aligned}$$

We have seen from Eq. (17) that the Bayesian classifier is the solution to the transcendental equation:

$$\ln \frac{x}{x-\theta} = \frac{-\theta^2 + 2\theta x}{2\sigma^2}. \tag{26}$$

The discriminant enforced by the k -OS CMOS classifier is $D(x, u_1) = D(x, u_2)$ which can further be simplified to:

$$\begin{aligned}
D(x, u_1) = D(x, u_2) & \implies D\left(x, \sigma \sqrt{\left(2 \ln \frac{n+1}{k}\right)}\right) = D\left(x, \theta + \sigma \sqrt{\left(2 \ln \frac{n+1}{n+1-k}\right)}\right) \\
& \implies x - \left(\sigma \sqrt{\left(2 \ln \frac{n+1}{k}\right)}\right) = \left(\theta + \sigma \sqrt{\left(2 \ln \frac{n+1}{n+1-k}\right)}\right) - x \\
& \implies x = \frac{\theta}{2} + \frac{\sigma}{\sqrt{2}} \left[\sqrt{\ln \frac{n+1}{k}} + \sqrt{\ln \frac{n+1}{n+1-k}} \right]. \tag{27}
\end{aligned}$$

The difference in the errors of the two classifiers is clearly related to differences in the corresponding discriminant functions specified by Eq. (26) and (27). The case when the *Dual* condition has to be invoked follows in an analogous manner and is omitted in the interest of brevity. Hence the theorem. \square

Remark: As in the case of the 2-OS, if we examine the error bounded by the CMOS classifier with regard to the classifier which discriminates based on the distance from the corresponding medians for different values of θ , k and n , we can see that the classifiers are almost identical. This is demonstrated by the results tabulated in Table 10.

4.5 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 11. For the distribution under consideration, the Bayesian approach provides an accuracy of 82.5%, 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

θ	$x : 4\text{-OS}, k = 2$	$x : 6\text{-OS}, k = 3$	$x : 8\text{-OS}, k = 4$
1	0.05	0.03	0.01
1.5	0.08	0.04	0.03
2	0.08	0.05	0.02
2.5	0.04	0.02	0.01
3	0	0	0

Table 10: Maximum bounded error (in %) by the CMOS classifier when compared to the classifier with respect to the medians of the distributions, for different values of θ , k and n of the Rayleigh Distribution. In each case, $\sigma = 2$.

9 is not satisfied.

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	$(\frac{5-i}{5}, \frac{i}{5}), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{(2 \ln \frac{5}{1})}$	$\theta + \sigma\sqrt{(2 \ln \frac{5}{4})}$	81.8	CMOS
3	Four	$(\frac{5-i}{5}, \frac{i}{5}), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{(2 \ln \frac{5}{2})}$	$\theta + \sigma\sqrt{(2 \ln \frac{5}{3})}$	82.0	CMOS
4	Six	$(\frac{7-i}{7}, \frac{i}{7}), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{(2 \ln \frac{7}{1})}$	$\theta + \sigma\sqrt{(2 \ln \frac{7}{6})}$	81.6	Dual CMOS
5	Six	$(\frac{7-i}{7}, \frac{i}{7}), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{(2 \ln \frac{7}{2})}$	$\theta + \sigma\sqrt{(2 \ln \frac{7}{5})}$	82.10	CMOS
6	Six	$(\frac{7-i}{7}, \frac{i}{7}), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{(2 \ln \frac{7}{3})}$	$\theta + \sigma\sqrt{(2 \ln \frac{7}{4})}$	82.15	CMOS
7	Eight	$(\frac{9-i}{9}, \frac{i}{9}), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{(2 \ln \frac{9}{1})}$	$\theta + \sigma\sqrt{(2 \ln \frac{9}{8})}$	81.55	Dual CMOS
8	Eight	$(\frac{9-i}{9}, \frac{i}{9}), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{(2 \ln \frac{9}{2})}$	$\theta + \sigma\sqrt{(2 \ln \frac{9}{7})}$	82.05	CMOS
9	Eight	$(\frac{9-i}{9}, \frac{i}{9}), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{(2 \ln \frac{9}{4})}$	$\theta + \sigma\sqrt{(2 \ln \frac{9}{5})}$	82.15	CMOS

Table 11: A comparison of the accuracy of the Bayesian (i.e., 82.5%) and the k -OS CMOS classifier for the Rayleigh Distribution by using the symmetric pairs of the OS for different values of n . The value of σ and θ were set to be 2 and 1.5 respectively. Note that in every case, CMOS attained near-optimal accuracy whenever the conditions stated in Theorem 9 were satisfied.

To clarify the table, consider the cases in which the 6-OS were invoked for the classification. For 6-OS, the possible symmetric OS pairs could be $\langle 1, 6 \rangle$, $\langle 2, 5 \rangle$, and $\langle 3, 4 \rangle$ respectively. Observe that the expected values for the first moment of the k -OS has the form $E[\mathbf{x}_{k,n}] = \sigma\sqrt{(2 \ln \frac{n+1}{k})}$. For the cases where the condition $\sqrt{\ln \frac{n+1}{k}} - \sqrt{\ln \frac{n+1}{n+1-k}} < \frac{\theta}{\sigma\sqrt{2}}$, the accuracy attained is either optimal or near-optimal, as indicated by the results in the table (denoted by Trial Nos. 5 and 6). Now, consider the results presented in the row denoted by Trial No. 7. In this case where the CMOS positions were $\sigma\sqrt{(2 \ln \frac{7}{1})}$ and $\theta + \sigma\sqrt{(2 \ln \frac{7}{6})}$, the inequality of the condition imposed in Theorem 9 simplified

to $1.002339 < 0.88388$, which is not valid. Observe that if $\sqrt{\ln \frac{n+1}{k}} - \sqrt{\ln \frac{n+1}{n+1-k}} > \frac{\theta}{\sigma\sqrt{2}}$, the Dual CMOS should be invoked in which the symmetric pairs are reversed to obtain the near-optimal Bayes' bound.

This concludes our study on the CMOS for the Rayleigh distribution.

5 The Gamma Distribution

The Gamma distribution is a continuous probability distribution with two parameters - a , a shape parameter and b , a scale parameter. Another parametrization which is commonly used in Bayesian statistics has the parameters α , the shape parameter and the inverse scale parameter or the rate parameter, $\beta = \frac{1}{b}$. The pdf of the Gamma distribution with the parameters a and b is:

$$\frac{1}{\Gamma(a) b^a} x^{a-1} e^{-\frac{x}{b}}; a > 0, b > 0, \quad (28)$$

with mean ab and variance ab^2 . The pdf of the Gamma distribution with the parameters a and β is:

$$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}; \alpha > 0, \beta > 0, \quad (29)$$

with mean $\frac{\alpha}{\beta}$ and variance $\frac{\alpha}{\beta^2}$. Unfortunately, the cumulative distribution function does not have a closed form expression [15, 21, 26]⁷.

5.1 Theoretical Analysis: Gamma Distribution - 2-OS

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. As in the previous cases, we consider the scenario when $a_1 = a_2 = a$ and $b_1 = b_2 = b$.

In the interest of simplicity, consider the distributions: $f(x, 2, 1) = x e^{-x}$ 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. By virtue of Eq. (3) and (4), 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

⁷The closed form expression (described in terms of the incomplete Gamma function) is defined and readily available in mathematical software such as *Wolfram*, *Mathematica* and *Matlab*. However, what we communicate here is the concept stated in [15, 21, 26] that the cdf does not have a form that it is described independent of the (complete or incomplete) Gamma function itself.

distribution, and u_2 be the point for the percentile $\frac{1}{3}$ of the second distribution. Then:

$$\begin{aligned} \int_0^{u_1} x e^{-x} dx = \frac{2}{3} &\implies 1 - u_1 e^{-u_1} - e^{-u_1} = \frac{2}{3} \\ &\implies \ln u_1 - 2u_1 = \ln \frac{1}{3}. \end{aligned} \quad (30)$$

By a similar argument, the CMOS point for the $\frac{1}{3}$ percentile of the second distribution leads to the equation:

$$\ln(u_2 - \theta) - 2(u_2 - \theta) = \ln \frac{1}{3} - \ln \theta. \quad (31)$$

We now prove that the classification with regard to the CMOS points are almost identical to the classification based on the median⁸.

Without loss of generality, let the distributions of ω_1 and ω_2 be $G(x, 2, 1)$ and $G(x - \theta, 2, 1)$, where θ is the displacement. As the curve of the Gamma distribution is not symmetric, we shall consider the scenario that the classification is done based on the median, which is the most central point of the distribution, other than the mean. The claim can be stated as in Theorem 10.

Theorem 10. *For the 2-class problem in which the two class conditional distributions are Gamma and identical with $a = 2$ and $b = 1$, the accuracy obtained by CMOS, the classification using two OS, 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 $1.7391 + \frac{\theta}{2}$ and $1.6783 + \frac{\theta}{2}$.*

Proof. The claim of this theorem is that CMOS classification can attain an accuracy which is almost identical to the one obtained with regard to the corresponding medians of the distributions.

As Eqs. (30) and (31) are of a transcendental nature, they cannot be simplified further, and hence it is not possible to obtain a closed form expression for the CMOS positions. The reason for this phenomenon is that the Gamma distribution lacks a closed form expression for its cumulative distribution function. Consequently, the only possible way by which we can proceed further to prove the claim is through a numerical formulation. The 2-OS CMOS positions u_1 and u_2 for $\Gamma(x, 2, 1)$ and $\Gamma(x - \theta, 2, 1)$ can be obtained by making use of the built-in functions available in standard software packages [8] as $u_1 = 2.2893$ and $u_2 = \theta + 1.1888$. Also, we can obtain the values of ν_1 and ν_2 for the same distributions as $\nu_1 = 1.6783$ and $\nu_2 = 1.6783 + \theta$ respectively. Then, the classifier

⁸The Bayes' classifier, in this case when we are only using the 2-OS, is more distant than the CMOS, because of the skewed asymmetric form of the Gamma distribution. However, as we shall see later, other k -OS CMOS classifiers become more near-optimal.

with regard to the medians of the distributions can be obtained as:

$$\begin{aligned}
D(x, \nu_1) < D(x, \nu_2) &\implies D(x, 1.6783) < D(x, 1.6783 + \theta) \\
&\implies x - 1.6783 < 1.6783 + \theta - x \\
&\implies x < 1.6783 + \frac{\theta}{2}.
\end{aligned} \tag{32}$$

The discriminant function with regard to the medians of the distributions is thus:

$$x = 1.6783 + \frac{\theta}{2}. \tag{33}$$

We now consider the classification with respect to the expected values of the moments of the 2-OS. We can see that the discriminant enforced by 2-OS CMOS is $D(x, u_1) = D(x, u_2)$ which can further be simplified to:

$$\begin{aligned}
D(x, u_1) = D(x, u_2) &\implies D(x, 2.2893) = D(x, 1.1888 + \theta) \\
&\implies x - 2.2893 = 1.1888 + \theta - x \\
&\implies x = 1.7391 + \frac{\theta}{2}.
\end{aligned} \tag{34}$$

The difference in the errors of the two classifiers is clearly related to differences in the corresponding discriminant functions specified by Eq. (35) and (34). \square

Remark: As in the case of the Rayleigh distribution, we can show that the resulting classifiers are almost identical by considering the differences of the error probabilities quantified by the differences between the areas under the curves of the resulting errors. These error differences can be calculated by evaluating the corresponding definite integrals. Since closed form expressions for the integrals are not available, this has to be achieved numerically. The maximum area differences created by the CMOS classifier and the classifier based on the medians of the distributions for different values of θ are listed in Table 12. The claim of Theorem 10 is thus justified.

θ	1	1.5	2	2.5	3
Max. Bounded Error(in %)	0.71	0.95	0.90	0.83	0.23

Table 12: Maximum bounded error by the CMOS classifier when compared to the classifier with regard to the medians, for different values of θ of the Gamma Distribution.

5.2 Data Generation: Gamma Generation

There are a number of data generation algorithms reported for the Gamma distribution, all of which make use of the Uniform random variate $U(0, 1)$. In our experiments, data was generated using the built-in function available in MatLab [4], namely $gamrnd(a, b, sz)$, where a is the shape parameter, b is the scale parameter, and sz is the size of the array. To be specific, $gamrnd(2, 1, 10)$ will generate 100 values that follow the Gamma distribution with the shape parameter 2 and the scale parameter 1. For our experiments, we generated 1,000 points for each of the distributions, where the second distribution was displaced by a constant, θ .

5.3 Experimental Results: Gamma Distribution - 2-OS

The CMOS classifier was rigorously tested for a number of experiments for 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 9.

n	Median	CMOS
4.5	94.825	95.01
4.0	94.25	94.49
3.5	92.74	92.915
3.0	90.765	90.425
2.5	86.51	85.985
2.0	80.145	79.54
1.5	72.64	72.34

Table 13: A comparison of the accuracy with respect to the median and the 2-OS CMOS classifier for the Gamma Distribution.

5.4 Theoretical Analysis: Gamma Distribution - k -OS

We have already seen in Theorem 10 that the 2-OS CMOS can obtain classification accuracy comparable to that obtained by comparing the testing sample with respect to the medians of the distributions. We shall now move on to examine the k -OS CMOS.

For the sake of the argument, let the distributions of ω_1 and ω_2 be $G(x, 2, 1)$ and $G(x - \theta, 2, 1)$, where θ is the displacement. Then, our claim for the k -OS can be stated as in Theorem 11.

Theorem 11. For the 2-class problem in which the two class conditional distributions are Gamma and identical, the classification obtained 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) is arbitrarily close to the classification based on the medians if and only if $u_1 < u_2$. If $u_1 > u_2$, CMOS involves invoking the Dual n -OS pairs, i.e., the k^{th} OS for ω_1 and the $(n - k)^{th}$ OS for ω_2 .

Proof. We shall now extend the result of Theorem 10 for k -OS, so as to determine if the CMOS classification is almost identical to the classification based on the medians. Let u_1 be the point for the percentile $\frac{n+1-k}{n+1}$ (the $(n - k)^{th}$ -OS) of the first distribution, and u_2 be the point for the percentile $\frac{k}{n+1}$ (the k -OS) of the second distribution. Our task is to compare the classification with respect to the CMOS positions and with the classifier obtained with regard to the medians of the distributions. The classifier with regard to the medians of the distributions can be obtained as:

$$\begin{aligned}
D(x, \nu_1) < D(x, \nu_2) &\implies D(x, 1.6783) < D(x, 1.6783 + \theta) \\
&\implies x - 1.6783 < 1.6783 + \theta - x \\
&\implies x < 1.6783 + \frac{\theta}{2}.
\end{aligned} \tag{35}$$

Again, in order to compare the Bayesian, CMOS and median classifiers, as in the 2-OS case, we can provide a numerical comparison of the schemes by evaluating the differences of the error probabilities quantified by the differences between the areas under the curves of the resulting errors. If we proceed in this manner, we are to first obtain the values for the CMOS positions for different k -OS, and these are tabulated in Table 14.

n	Percentile	CMOS	n	Percentile	CMOS
2	$\frac{1}{2}$	1.6783	Corresponds to the Median		
3	$\frac{1}{3}$	1.1888	3	$\frac{2}{3}$	2.2893
5	$\frac{1}{5}$	0.8244	5	$\frac{2}{5}$	1.3764
5	$\frac{3}{5}$	2.0223	5	$\frac{4}{5}$	2.9943
7	$\frac{1}{7}$	0.6624	7	$\frac{2}{7}$	1.0584
7	$\frac{3}{7}$	1.4596	7	$\frac{4}{7}$	1.9183
7	$\frac{5}{7}$	2.5077	7	$\frac{6}{7}$	3.4356
9	$\frac{1}{9}$	0.5669	9	$\frac{2}{9}$	0.8855
9	$\frac{4}{9}$	1.5068	9	$\frac{5}{9}$	1.8627
9	$\frac{7}{9}$	2.8529	9	$\frac{8}{9}$	3.7568

Table 14: CMOS positions for Gamma distribution $\Gamma(2, 1)$ for different percentiles.

With these values on hand, we can now verify the claim that the CMOS classifier and its median-based counterpart are almost identical for different values of k , n and θ by computing the respective errors areas that they yield. For different values of θ , the areas are tabulated in Table 12 for certain specific CMOS pairs (k , $n-k+1$). From Table 12, the reader can observe that the classifiers are almost identical.

θ	2	2.5	3	3.5	4
n	8	8	8	8	6
k	4	4	4	2	1
Max. Bounded Error	0.11	0.08	0.0040	0.01	0.01

Table 15: Maximum bounded error (in %) by the CMOS classifier when compared to the classifier with regard to the medians of the distributions, for different values of θ , k and n of the Gamma Distribution.

We thus conclude that a classification that is arbitrarily close to the one obtained by comparing to the medians 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).

The proof of the case when the *Dual* condition has to be invoked follows in an identical manner and is omitted here. \square

Remark: Similar results are available for the comparison of CMOS and the corresponding Bayesian classifier. To get the Bayes' classifier, we argue that:

$$\begin{aligned}
 p(x|\omega_1)P(\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(x|\omega_2)P(\omega_2) &\implies x e^{-x} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} (x-\theta) e^{-(x-\theta)} \\
 &\implies \frac{x-\theta}{x} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} e^{-\theta} \\
 &\implies x \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{\theta}{1-e^{-\theta}}, \tag{36}
 \end{aligned}$$

whence the differences between the areas under the curves can be evaluated and seen to be almost negligible. The details are omitted here to avoid repetition.

5.5 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 the test results are given in Table 16. Experiments have been performed for different values of θ , and we can see that the CMOS attained a near-optimal Bayes' accuracy. Also, we can see that the Dual CMOS has to be invoked if the condition stated in Theorem 11 is

not satisfied.

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	-	96.04	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	77.875 (D)
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	82.71 (D)	76.435 (D)
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	86.905 (D)	80.59 (D)	75.94 (D)
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

Table 16: A comparison of the k -OS CMOS classifier when compared to the Bayes’ classifier and the classifier with respect to the median and mean for the Gamma Distribution for different values of n . In each column, the value which is near-optimal is rendered bold. The scenarios when we have invoked the *Dual* condition are specified by noting them using the notation “(D)”.

By way of example, consider the case in Trial # 7 when $\theta = 3.0$, where the condition $u_1 < u_2$ was not satisfied. Here, as the condition yielded an invalid inequality, i.e., $3.7568 < 3.5669$, the Dual CMOS has to be invoked by reversing the CMOS values to obtain the near-optimal accuracy. Interestingly enough, if we examine the table, we can see that the Bayes’ accuracy is the highest for all cases except for the scenario when $\theta = 3.0$. This result must, in fact, be considered as an aberration.

This concludes the study of the Gamma distribution with regard to the CMOS classification.

6 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:

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}. \quad (37)$$

The mean and the variance of the distribution are $\frac{\alpha}{\alpha+\beta}$ and $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ respectively.

The Beta distribution can take different shapes based on the values of the shape parameters, and hence any systematic analysis will have to be performed on a case-by-case basis. Some of the cases are given in Table 17 and are plotted in Figure ??.

No	α, β	Distribution
1	$\alpha = 1, \beta = 1$	Uniform Distribution
2	$\alpha < 1, \beta < 1$	U-shaped
3	$\alpha = 1, \beta = 2$	Straight line
4	$\alpha = \beta$	Symmetric about $\frac{1}{2}$
5	$\alpha = \frac{1}{2}, \beta = \frac{1}{2}$	Arcsine Distribution
6	$\alpha > 1, \beta > 1$	Unimodal Distribution
7	$\alpha = 1, \beta > 1$	Strictly Convex
8	$\alpha = 1, 1 < \beta < 2$	Strictly Concave

Table 17: Different forms of the Beta distribution for the various values of its parameters α and β .

For this study, we mainly consider three cases:

- $\alpha = 1, \beta = 1$: Uniform Distribution.
- $\alpha = \beta$: Symmetric about $\frac{1}{2}$. In this case, in this present work, we merely deal with the scenario when $\alpha = \beta = 2$. The results for the more general case (i.e., $\alpha = \beta > 2$) will be briefly alluded to, but is left as an avenue for future work.
- $\alpha > 1, \beta > 1$: Unimodal Distribution.

Earlier, in [23], where we had initially introduced the concept of CMOS-based PR, we had analyzed the 2-OS and k -OS CMOS schemes for the Uniform distribution, and had provided the corresponding theoretical analysis and the experimental results. These were briefly catalogued in Section 3.2, from which we can see 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. To avoid repetition, the analysis for the Beta distribution, $B(1,1)$ (which reduces to the analysis for Uniform $U(0,1)$) is omitted here, closing the first of the above three cases.

We now proceed to consider the Beta distribution in which $\alpha = \beta$.

6.1 Theoretical Analysis: Beta Distribution ($\alpha = \beta$) - 2-OS

Consider two classes ω_1 and ω_2 where the class ω_2 is displaced by a quantity θ , and the values of the shape parameters are identical. As in the previous cases, we consider the scenario when

$\alpha_1 = \alpha_2 = \alpha$, $\beta_1 = \beta_2 = \beta$, and for the sake of simplicity, $\alpha = \beta = 2$. With these settings, the respective distributions become: $f(x, 2, 2) = 6x(1 - x)$ and $f(x - \theta, 2, 2) = 6(x - \theta)(1 - x + \theta)$.

We first derive the moments of the 2-OS, which are the points of interest for the CMOS, for the Beta distribution. By virtue of Eq. (3) and (4), the expected values of the first moments of the two OS can be obtained by determining the points where the cumulative distribution functions attain the values of $\frac{1}{3}$ and $\frac{2}{3}$ respectively. As the distribution takes different forms based on the values of the shape parameters, we have to solve each case separately, which we shall embark on. 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:

$$\int_0^{u_1} 6x(1 - x)dx = \frac{2}{3} \implies -6u_1^3 + 9u_1^2 - 2 = 0. \quad (38)$$

By a similar argument, the CMOS point for the $\frac{1}{3}$ percentile of the second distribution (if we don't take the displacement, θ , into consideration) leads to the equation:

$$-6u_2^3 + 9u_2^2 - 1 = 0. \quad (39)$$

We shall now prove that in this symmetric case, CMOS, indeed, attains the optimal Bayes' bound.

Theorem 12. *For the 2-class problem in which the two class conditional distributions are Beta(α, β) ($\alpha = \beta$) and identical, CMOS, the classification using two OS, attains an accuracy that is exactly identical to the optimal Bayes' bound.*

Proof. Without loss of generality, let the distributions of ω_1 and ω_2 be $B(x, 2, 2)$ and $B(x - \theta, 2, 2)$, where θ is the displacement for the second distribution. Then, to get the Bayes' classifier, we argue that:

$$\begin{aligned} p(x|\omega_1)P(\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(x|\omega_2)P(\omega_2) &\implies 6x(1 - x) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 6(x - \theta)(1 - (x - \theta)) \\ &\implies x \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{\theta + 1}{2}. \end{aligned} \quad (40)$$

We now consider the classification with respect to the expected values of the moments of the 2-OS, u_1 and u_2 . In order to prove our claim, we need to show that

$$x \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{\theta + 1}{2} \implies D(x, u_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} D(x, u_2). \quad (41)$$

If we examine Eqs. (38) and (39), we can see that Eq. (39) can be obtained by substituting

$1 - u_2$ for u_1 in Eq. (38) since:

$$-6(1 - u_2)^3 + 9(1 - u_2)^2 - 2 = 0 \implies -6u_2^3 + 9u_2^2 - 1 = 0. \quad (42)$$

Consequently, it is obvious that $u_2 = \theta + u_1 - 1$, implying that the RHS of the claim given by Eq. (41) leads to the following:

$$\begin{aligned} D(x, u_1) \underset{\omega_2}{\overset{\omega_1}{\leq}} D(x, u_2) &\implies D(x, u_1) \underset{\omega_2}{\overset{\omega_1}{\leq}} D(x, \theta + 1 - u_1) \\ &\implies x - u_1 \underset{\omega_2}{\overset{\omega_1}{\leq}} \theta + 1 - u_1 - x \\ &\implies x \underset{\omega_2}{\overset{\omega_1}{\leq}} \frac{\theta + 1}{2}. \end{aligned} \quad (43)$$

The result follows by observing that Eqs. (40) and (41) are identical comparisons. Hence the theorem. \square

6.2 Data Generation: Beta Generation

As in the case of the Gamma distribution, the data is generated using the built-in function available in MatLab, namely *betarnd*(α, β, r), where α and β are the shape parameters, and where the function returns a square matrix with the dimension r . To be specific, *betarnd*(2,2,10) will generate 100 values that follow the Beta distribution with 2 being the value for the shape parameters. For our experiments, we generated 1,000 points for each of the distributions, where the second distribution was displaced by a constant, θ .

6.3 Experimental Results: Beta Distribution ($\alpha = \beta$) - 2-OS

The CMOS has been rigorously tested for various Beta distributions with 2-OS with $\alpha = \beta = 2$. In the interest of brevity, a few typical results are given below. For each of the experiments, we generated 1,000 points for the classes ω_1 and ω_2 characterized by $B(x, 2, 2)$ and $B(x - \theta, 2, 2)$ respectively. We then invoked a classification procedure by utilizing the Bayesian and the CMOS strategies. In every case, CMOS was compared with the Bayesian classifier for different values of θ , as tabulated in Table 18. The results were obtained by executing each algorithm 50 times using a 10-fold cross-validation scheme.

The results given in this table justify the claim of Theorem 12. We conjecture that this claim is true for any $\alpha = \beta = t$, but it is presently considered as an unsolved problem.

6.4 Theoretical Analysis: Beta Distribution ($\alpha = \beta$)- k -OS

We have seen from Theorem 12 that the moments of the 2-OS are sufficient for the classification to attain a Bayes' bound. We shall now examine the scenario where the k -OS CMOS is invoked, and

θ	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
Bayesian	99.845	99.45	98.185	96.94	94.95	92.86	90.31	88.075
CMOS	99.845	99.45	98.185	96.94	94.95	92.86	90.31	88.075

Table 18: A comparison of the accuracy of the Bayesian and the 2-OS CMOS classifier for the Beta distribution $B(2, 2)$ for different values of θ .

thus determine the strength of the proposed method.

Theorem 13. *For the 2-class problem in which the two class conditional distributions are Beta and identical as $B(x, \alpha, \beta)$ and $B(x - \theta, \alpha, \beta)$ where $\alpha = \beta = 2$, optimal Bayesian classification can be achieved by using symmetric pairs of the n -OS, i.e., the $n - k$ OS for ω_1 (represented by u_1) and the k OS for ω_2 (represented by u_2) if and only if $u_1 < u_2$. If $u_1 > u_2$, optimal Bayesian classification can be achieved by using the Dual symmetric pairs of the n -OS, i.e., the k OS for ω_1 and the $n - k$ OS for ω_2 .*

Proof. Our claim is that we can choose any of the symmetric pairs of the n -OS, i.e., the $n - k$ OS for ω_1 and the k OS for ω_2 to obtain an optimal classification. 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:

$$\int_0^{u_1} 6x(1-x)dx = \frac{n+1-k}{n+1} \implies -2u_1^3 + 3u_1^2 - \frac{n+1-k}{n+1} = 0. \quad (44)$$

By a similar argument, if we ignore the displacement θ , the CMOS point for the $\frac{k}{n+1}$ percentile of the second distribution leads to the equation:

$$-2u_2^3 + 3u_2^2 - \frac{k}{n+1} = 0. \quad (45)$$

We have already shown in Eq. (40) that the Bayes' classifier is equivalent to the inequality $x \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{\theta+1}{2}$. Thus, in order to prove our claim, we need to show that the same classification criterion results for any symmetric pairs of the n -OS. Thus, our claim is:

$$x \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{\theta+1}{2} \implies D(x, u_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} D(x, u_2). \quad (46)$$

As in the case of 2-OS, if we substitute $u_1 = 1 - u_2$ in Eq. (44), the equation reduces to $-2u_2^3 + 3u_2^2 - \frac{k}{n+1} = 0$, proving the fact that for the distributions ω_1 and ω_2 , the $\left\langle \frac{n+1-k}{n+1}, \frac{k}{n+1} \right\rangle$ CMOS positions (represented by u_1 and u_2 respectively) have the relation $u_2 = 1 - u_1$. Thus, the

the RHS of the claim given by Eq. (46) simplifies to:

$$\begin{aligned}
D(x, u_1) \underset{\omega_2}{\overset{\omega_1}{\leq}} D(x, u_2) &\implies D(x, u_1) \underset{\omega_2}{\overset{\omega_1}{\leq}} D(x, \theta + 1 - u_1) \\
&\implies x - u_1 \underset{\omega_2}{\overset{\omega_1}{\leq}} \theta + 1 - u_1 - x \\
&\implies x \underset{\omega_2}{\overset{\omega_1}{\leq}} \frac{\theta + 1}{2},
\end{aligned} \tag{47}$$

which is the condition sought for.

The arguments for the cases when the Dual condition has to be invoked follow in an identical manner and are omitted. The theorem follows. \square

6.5 Experimental Results: Beta Distribution ($\alpha = \beta$) - k -OS

The CMOS method has also been tested for the Beta distribution for other k OS when $\alpha = \beta = 2$. In the interest of brevity, we merely cite one example where the distributions for ω_1 and ω_2 were characterized by $\beta(x, 2, 2)$ and $b(x - \theta, 2, 2)$ respectively. For each of the experiments, we generated 1,000 points for each class, and the testing samples were classified based on the selected *symmetric* pairs for values k and $n - k$ respectively. The results are displayed in Table 19.

Trial No.	Order(n)	Moments	OS_1	OS_2	CMOS	CMOS/Dual CMOS
1	Two	$\langle \frac{2}{3}, \frac{1}{3} \rangle$	0.61304	$\theta + 0.38696$	87.3	CMOS
2	Four	$\langle \frac{4}{5}, \frac{1}{5} \rangle$	0.71286	$\theta + 0.28714$	87.3	CMOS
3	Four	$\langle \frac{3}{5}, \frac{2}{5} \rangle$	0.56707	$\theta + 0.43293$	87.3	CMOS
4	Six	$\langle \frac{6}{7}, \frac{1}{7} \rangle$	0.7621	$\theta + 0.23790$	87.3	CMOS
5	Six	$\langle \frac{5}{7}, \frac{2}{7} \rangle$	0.6471	$\theta + 0.3529$	87.3	CMOS
6	Six	$\langle \frac{4}{7}, \frac{3}{7} \rangle$	0.54776	$\theta + 0.45224$	87.3	CMOS
7	Eight	$\langle \frac{8}{9}, \frac{1}{9} \rangle$	0.79269	$\theta + 0.20731$	87.3	Dual CMOS
8	Eight	$\langle \frac{7}{9}, \frac{2}{9} \rangle$	0.69508	$\theta + 0.30492$	87.3	CMOS
9	Eight	$\langle \frac{5}{9}, \frac{4}{9} \rangle$	0.53711	$\theta + 0.46289$	87.3	CMOS

Table 19: A comparison of the accuracy of the Bayesian and the k -OS CMOS classifier for the Beta Distribution by using the symmetric pairs of the OS for different values of n . The value of θ was set to be 0.58. Note that in every case, the accuracy attained the Bayes' value whenever the conditions stated in Theorem 13 were satisfied.

To clarify the table, consider the cases in which the 8-OS were invoked for the classification. For 8-OS, the possible symmetric OS pairs could be $\langle 1, 8 \rangle$, $\langle 2, 7 \rangle$, and $\langle 4, 5 \rangle$ respectively. Wherever the condition $u_1 < u_2$ is satisfied, the CMOS attained the optimal Bayes' bound, as indicated by

the results in the table (denoted by Trial Nos. 8 and 9). Now, consider the results presented in the row denoted by Trial No. 7. In this case where the CMOS positions were 0.79269 and $\theta + 0.20731$, the inequality of the condition imposed in Theorem 13 simplified to $0.79269 < 0.78731$, which is not valid. Observe that if $u_1 > u_2$, the symmetric pairs should be reversed to obtain the optimal Bayes' bound.

This concludes our study on the symmetric Beta distribution.

We now move on to the unimodal Beta distribution characterized by the shape parameters $\alpha > 1$ and $\beta > 1$, $\alpha \neq \beta$.

6.6 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 preciseness⁹, we consider the case when $\alpha = 2$ and $\beta = 5$. Then, the distributions are:

$$f(x, 2, 5) = 30x(1 - x)^4 \quad (48)$$

and

$$f(x - \theta, 2, 5) = 30(x - \theta)(1 - x + \theta)^4. \quad (49)$$

We first derive the moments of the 2-OS, namely u_1 and u_2 where u_1 represents the point for the percentile $\frac{2}{3}$ of the first distribution, and u_2 represents the point for the percentile $\frac{1}{3}$ of the second distribution. Then:

$$\int_0^{u_1} 30x(1 - x)^4 dx = \frac{2}{3} \quad (50)$$

and

$$\int_0^{u_2} 30(x - \theta)(1 - x + \theta)^4 dx = \frac{1}{3}. \quad (51)$$

These positions u_1 and u_2 can be obtained by making use of the built-in functions available in standard software packages [8] as $u_1 = 0.34249$ and $u_2 = \theta + 0.1954$. 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 14. *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, the*

⁹Any 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.

classification using two OS, 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. Without loss of generality, let the distributions of ω_1 and ω_2 be $B(x, 2, 5)$ and $B(x - \theta, \sigma)$, where σ is the identical scale parameter. As already stated, the 2-OS CMOS positions and the medians of the distributions can be obtained using the standard software packages, whence we can determine that $u_1 = 0.34249$, $u_2 = \theta + 0.1954$, $\nu_1 = 0.26445$, and $\nu_2 = \theta + 0.26445$, where ν_1 and ν_2 are the medians of the distributions. The claim of this part is that CMOS classification can attain an accuracy which is almost identical to the one obtained with regard to the corresponding medians of the distributions.

Using the values of the medians of the distributions, the classifier can be obtained as:

$$\begin{aligned} D(x, \nu_1) < D(x, \nu_2) &\implies D(x, 0.26445) < D(x, 0.26445 + \theta) \\ &\implies x - 0.26445 < 0.26445 + \theta - x \\ &\implies x < 0.26445 + \frac{\theta}{2}. \end{aligned} \tag{52}$$

Thus, the discriminant function with respect to the medians of the distributions is:

$$x = 0.26445 + \frac{\theta}{2}. \tag{53}$$

If we consider the classifier with regard to the expected moments of the 2-OS, we can see that the discriminant enforced by 2-OS CMOS is $D(x, u_1) = D(x, u_2)$, which simplifies to:

$$\begin{aligned} D(x, u_1) = D(x, u_2) &\implies D(x, 0.34249) = D(x, 0.1954 + \theta) \\ &\implies x - 0.34249 = 0.1954 + \theta - x \\ &\implies x = 0.2689 + \frac{\theta}{2}. \end{aligned} \tag{54}$$

The difference in the errors of the two classifiers is clearly related to differences in the corresponding discriminant functions specified by Eq. (53) and (54). Hence the theorem. \square

Remark:

1. As in the other asymmetric distributions, we can quantify the differences of the error probabilities by the differences between the areas under the curves of the resulting errors of the considered Beta distributions. The maximum bounded error by the CMOS classifier when compared to the classifier with regard to the medians, for different values of θ , are tabulated in Table 20. From this table, we can conclude that the classifiers obtained with respect to the medians and the CMOS

positions are almost indistinguishable. Clearly, CMOS, the classification using two OS, attains the near-identical bound obtained by comparison to the corresponding medians.

θ	0.30	0.35	0.4	0.45	0.5	0.55
Max. Bounded Error(in %)	0.3	0.25	0.17	0.03	0.16	0.21

Table 20: Maximum bounded error by the CMOS classifier when compared to the classifier with regard to the medians, for different values of θ of the Beta Distribution.

2. A similar comparison can be done by considering the efficiency of CMOS and comparing it with the optimal *Bayesian* classifier. Again, the accuracies of the two are almost indistinguishable, and the details are omitted here to avoid repetition.

6.7 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 21. The results were obtained by executing each algorithm 50 times using a 10-fold cross-validation scheme. The quality of the classifier is obvious.

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

Table 21: 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 .

6.8 Theoretical Analysis: Beta Distribution ($\alpha > 1, \beta > 1$) - k -OS

We have seen in Theorem 12 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 follows.

Theorem 15. *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 $\beta = 5$, 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 u_1) and the*

k^{th} OS for ω_2 (represented as u_2) if and only if $u_1 < u_2$. If this condition is violated, the CMOS classifier uses the Dual condition, i.e., the k OS for ω_1 and the $n - k$ OS for ω_2 .

Proof. In order to prove this claim, we shall now extend the result of Theorem 14 for k -OS, that the classification is almost identical to the classification based on the medians. Let u_1 be the point for the percentile $\frac{n+1-k}{n+1}$ (the $(n - k)^{th}$ -OS) of the first distribution, and u_2 be the point for the percentile $\frac{k}{n+1}$ (the k -OS) of the second distribution. We have to compare the CMOS classifier with the classifier obtained with regard to the medians of the distributions. To achieve this, we need to obtain the values for the CMOS positions for different k -OS, and these are tabulated in Table 22.

n	Percentile	CMOS	n	Percentile	CMOS
2	$\frac{1}{2}$	0.26445	Corresponds to the Median		
3	$\frac{1}{3}$	0.1954	3	$\frac{2}{3}$	0.3425
5	$\frac{1}{5}$	0.1399	5	$\frac{2}{5}$	0.2226
5	$\frac{3}{5}$	0.3095	5	$\frac{4}{5}$	0.4225
7	$\frac{1}{7}$	0.1140	7	$\frac{2}{7}$	0.1760
7	$\frac{3}{7}$	0.2344	7	$\frac{4}{7}$	0.2961
7	$\frac{5}{7}$	0.3684	7	$\frac{6}{7}$	0.4675
9	$\frac{1}{9}$	0.0984	9	$\frac{2}{9}$	0.1495
9	$\frac{4}{9}$	0.2409	9	$\frac{5}{9}$	0.2889
9	$\frac{7}{9}$	0.4072	9	$\frac{8}{9}$	0.4982

Table 22: CMOS positions for Beta distribution $B(2, 5)$ for different percentiles.

With these values on hand, we can verify the claim that the CMOS classifier and its Bayesian counterpart are almost identical for different values of k , n and θ by computing the differences of the error probabilities quantified by the differences between the areas under the curves of the resulting errors of the respective distributions. The computed areas are depicted in Table 23 for certain CMOS pairs ($k, n-k+1$) for different values of θ . From the tabulated values, we can see that the classifiers are almost identical.

The arguments are identical for the case when the Dual condition has to be invoked and are omitted here. The result follows. \square

6.9 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 24. Various experiments were performed

θ	0.35	0.45	0.55	0.65
n	8	4,6,8	8	8
k	4	2,3,4	4	4
Max. Bounded Error	0.02	0	0.03	0.02

Table 23: Maximum bounded error (in %) by the CMOS classifier when compared to the classifier with regard to the medians of the distributions, for different values of θ , k and n of the Beta Distribution.

for different values of θ , and from them, we can see that CMOS attained a near-optimal Bayes' accuracy. Also, we can see that the Dual CMOS has to be invoked if the condition stated in Theorem 15 is not satisfied.

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
9	8-OS	$(\frac{8}{9}, \frac{1}{9})$	84.225 (D)	92.225	96.225	98.35	99.5	99.375
10	8-OS	$(\frac{7}{9}, \frac{2}{9})$	85.675	92.5	96.175	98.15	99.325	99.375
11	8-OS	$(\frac{5}{9}, \frac{4}{9})$	86.575	92.775	95.525	97.975	99.125	99.275

Table 24: 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 when we have invoked the *Dual* condition are specified by noting them using the notation "(D)".

For example, if we examine Table 24, we see that CMOS attained the near-optimal value for certain k -OS when compared to the accuracy obtained with regard to the medians of the distributions. However, if we consider the case in Trial # 9 when $\theta = 0.35$, where the condition $u_1 < u_2 \implies 0.46753 < 0.46401$. In such cases, the Dual CMOS (CMOS values have to be reversed) has to be invoked in order to yield near-optimal accuracy.

This concludes the study of the Beta distribution with regard to the CMOS classification.

7 Conclusions and Future Work

In this paper we have presented a comprehensive study on the use of Order Statistics (OS) for parametric Pattern Recognition (PR) for various distributions within the exponential family. Within the traditional Bayesian paradigm, if we are allowed to compare the testing sample with only *a single* point in the feature space from each class, the *optimal* strategy would be to achieve this based on the (Mahalanobis) distance from the corresponding *central* points, for example, the means. As opposed to this, the paradigm has a different philosophy if we are working with the OS. The pioneering work on using OS for classification (Classification by Moments of Order Statistics (CMOS)) was presented in [23] for the Uniform distribution, by comparing the testing sample to a few samples distant from the mean. In [23], we also showed that the results could be extended for a few *symmetric* distributions within the exponential family. In this paper, we extended these results significantly by considering a spectrum of symmetric and asymmetric distributions within the exponential family, for some of which even the closed form expressions of the cumulative distribution functions are not available. The paper initially cites (without going into any great detail) the results for the Gaussian and double-exponential distributions. Thereafter, we have presented our new results which involve the Rayleigh, Gamma and certain Beta distributions. As in [23], we show that the new scheme has an accuracy that attains the Bayes' bound for symmetric distributions, and is, otherwise, very close to the optimal Bayes' bound. This has been demonstrated by both a theoretical and rigorous experimental testing.

As far as we know, our results for classification using the OS are both pioneering and novel.

With regard to future work, we intend to extend these results for multi-dimensional distributions and to also test CMOS for real-life data sets¹⁰. The unsupervised aspects of OS-based PR are also unsolved. Finally, the relationship between CMOS and the reported Border Identification schemes is also currently being studied.

References

- [1] David H. A. *Order Statistics*. John Wiley & Sons Inc., 1981.
- [2] M. Ahsanullah and V. B. Nevzorov. *Order Statistics: Examples and Exercises*. Nova Science Publishers, Inc, 2005.
- [3] C. L. Chang. Finding Prototypes for Nearest Neighbor Classifiers. In *IEEE Transactions on Computing*, volume 23, pages 1179–1184, 1974.
- [4] Stephen J. Chapman. *Matlab Programming for Engineers*. Thomson Learnnig, 2008.

¹⁰The details of the initial results for such data sets can be provided to interested readers on request. It will also soon be available at <http://people.scs.carleton.ca/~athomas1/AntiBayesRealLife.pdf>.

- [5] P. A. Devijver and J. Kittler. On the Edited Nearest Neighbor Rule. In *Fifth International Conference on Pattern Recognition*, pages 72–80, December 1980.
- [6] Luc Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.
- [7] W. Duch. Similarity based methods: a general framework for Classification, Approximation and Association. *Control and Cybernetics*, 29(4):937–968, 2000.
- [8] John W. Eaton. Gnu octave, 1994.
- [9] G. M. Foody. Issues in Training Set Selection and Refinement for Classification by a Feedforward Neural Network. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pages 409–411, 1998.
- [10] S. Garcia, J. Derrac, J. Ramon Cano, and F. Herrera. Prototype Selection for Nearest Neighbor Classification: Taxonomy and Empirical Study. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 34(3):417–435, 2012.
- [11] G. W. Gates. The Reduced Nearest Neighbor Rule. In *IEEE Transactions on Information Theory*, volume 18, pages 431–433, 1972.
- [12] P. E. Hart. The Condensed Nearest Neighbor Rule. In *IEEE Transactions on Information Theory*, volume 14, pages 515–516, 1968.
- [13] <http://sci2s.ugr.es/pr/>.
- [14] S. Kim and B. J. Oommen. On Using Prototype Reduction Schemes and Classifier Fusion Strategies to Optimize Kernel-Based Nonlinear Subspace Methods. In *IEEE Transactions on Pattern Analysis and machine Intelligence*, volume 27, pages 455–460, 2005.
- [15] P. R. Krishnaiah and M. Haseeb Rizvi. A note on Moments of Gamma Order Statistics. *Technometrics*, 9:315–318, 1967.
- [16] L. I. Kuncheva, J. C. Bezdek, and R. P. W. Duin. Decision Templates for Multiple Classifier Fusion: An Experimental Comparison. *Pattern Recognition - The Journal of the Pattern Recognition Society*, 34:299–314, 2001.
- [17] G. Li, N. Japkowicz, T. J. Stocki, and R. K. Ungar. Full Border Identification for Reduction of Training Sets. In *Proceedings of the Canadian Society for computational studies of intelligence, 21st conference on Advances in artificial intelligence*, pages 203–215, 2008.
- [18] G. D. Lin. Characterizations of Continuous Distributions via Expected values of two functions of Order Statistics. *Sankhya: The Indian Journal of Statistics*, 52:84–90, 1990.

- [19] K. W. Morris and D. Szynal. A goodness-of-fit for the Uniform Distribution based on a Characterization. *Journal of Mathematical Science*, 106:2719–2724, 2001.
- [20] G. L. Ritter, H. B. Woodruff, S. R. Lowry, and T. L. Isenhour. An Algorithm for a Selective Nearest Neighbor Rule. In *IEEE Transactions on Information Theory*, volume 21, pages 665–669, 1975.
- [21] Pandu R. Tadikamalla. An approximation to the moments and the percentiles of Gamma Order Statistics. *Sankhya: The Indian Journal of Statistics*, 39:372–381, 1977.
- [22] A. Thomas. *Pattern Classification using Novel Order Statistics and Border Identification Methods*. PhD thesis, School of Computer Science, Carleton University, 2012. (To be Submitted).
- [23] A. Thomas and B. J. Oommen. The Fundamental Theory of Optimal “Anti-Bayesian” Parametric Pattern Classification Using Order Statistics Criteria. *Pattern Recognition*, 46:376–388, 2013.
- [24] Y. Too and G. D. Lin. Characterizations of Uniform and Exponential Distributions. *Academia Sinica*, 7(5):357–359, 1989.
- [25] I. Triguero, J. Derrac, S. Garcia, and F. Herrera. A Taxonomy and Experimental Study on Prototype Generation for Nearest Neighbor Classification. *IEEE transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, 42:86–100, 2012.
- [26] D. H. Young. Moment relations for Order Statistics of the standardized Gamma distribution and the inverse multinomial distribution. *Biometrika*, 58:637–640, 1971.