International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:6, No:12, 2012

Order Statistics-based "Anti-Bayesian" Parametric Classification for Asymmetric Distributions in the Exponential Family

A. Thomas, and B. John Oommen

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

***Keywords*—**Classification using Order Statistics (OS), Exponential family, Moments of OS.

I. Introduction

THE basis for *statistical* pattern classification is that the individual classes are characterized by their *distributions*. These distributions have numerous indicators such as their means, variances etc., and these indices have, traditionally, played a prominent role in achieving pattern classification, and in designing the corresponding training and testing algorithms. It is also well known that a distribution has many other characterizing indicators, for example, those related to its Order Statistics (OS). The interesting point about these indicators is that some of them are quite unrelated to the traditional moments themselves, and in spite of this, have not been used in achieving PR. The main question that we shall consider 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 distribution. 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.

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

Contributions of this Paper: The novel contributions of this paper are:

- We propose an "anti-Bayesian" paradigm for the classification of patterns within the parametric mode of computation, where the distance computations are not with regard to the "mean" but with regard to some samples "distant" from the mean. These points, which are sometimes as few as *two*, are the moments of OS of the distributions;
- We demonstrate that the proposed approach can attain

A. Thomas is a Ph.D. Candidate at School of Computer Science, Carleton University, Ottawa, ON K1S 5B6 Canada (e-mail: athomas1@cmail.carleton.ca).

B. John Oommen is a *Chancellor's Professor* at School of Computer Science, Carleton University, Ottawa, ON K1S 5B6 Canada. The author is a *Fellow* of *IEEE* and *IAPR*, and also an *Adjunct Professor* with the University of Agder in Grimstad, Norway. (e-mail: oommen@scs.carleton.ca; Phone: +1-613-520-2600 x 4358).

the optimal bound for symmetric distributions and near-optimal bound for non-symmetric distributions;
- To justify these claims, we submit a formal analysis and the results of various experiments which have been performed for a few distributions within the exponential family (for which even the closed form expressions of the distributions are not available), and the results are clearly conclusive.

Our results for classification using the OS are both pioneering and novel.

II. Relevant Background Areas Regarding Order Statistics

Let $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$ 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 $\mathbf{x}_{1,n}$, $\mathbf{x}_{2,n}$, ..., $\mathbf{x}_{n,n}$ be the corresponding Order Statistics (OS). The r^{th} OS, $\mathbf{x}_{r,n}$, of the set is the r^{th} smallest value among the given random variables. 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 $r = 1, 2, ..., n$. The reasoning for the above expression is straightforward. If the r^{th} OS appears at a location given by $\mathbf{y} = \mathbf{x}_{r,n}$, it implies that the $r-1$ smaller elements of the set are drawn independently from a Binomial distribution with a probability $\Phi(y)$, and the other $n-r$ samples are drawn using the probability $1-\Phi(y)$. The factorial terms result from the fact that the $(r-1)$ elements can be independently chosen from the set of n elements.

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

$$\frac{n!}{(r-1)!(n-r)!} \int_{-\infty}^{+\infty} y^k \Phi(y)^{k-1}(x)(1-\Phi(y))^{n-r}\varphi(y)dy,$$

provided that both sides of the equality exist [3], [4].

The fundamental theorem concerning the OS that we invoke is found in many papers [4]–[6]. The theorem can be summarized as follows.

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 [5]:

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

where $B(a,b)$ denotes the $Beta$ function, and $B(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$ since its parameters are integers.

The above fundamental result can also be used for characterization purposes as explained in [5], [7]. 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:

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

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

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:

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

and

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

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.

III. Optimal OS-based Classification: The Generic Classifier

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

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

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

IV. 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 whose applications are found in [8] and [9].

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

Theoretical Analysis: Rayleigh Distribution - 2-OS

The typical PR problem involving the Rayleigh distribution would consider two classes ω_1 and ω_2 where the class ω_2 is displaced by a quantity θ, and the values of σ are σ_1 and σ_2 respectively. We consider the scenario when $\sigma_1 = \sigma_2 = \sigma$. Consider the distributions: $f(x,\sigma) = \frac{x}{\sigma^2}e^{\frac{-x^2}{2\sigma^2}}$ and $f(x-\theta,\sigma) = \frac{x-\theta}{\sigma^2}e^{\frac{-(x-\theta)^2}{2\sigma^2}}$. In order to do the classification based on CMOS, we shall first derive the moments of the 2-OS for the Rayleigh distribution. By virtue of Eq. (2) and (3), the expected values of the first moments of the two OS can be obtained by determining the points where the cumulative distribution function attains the values of $\frac{1}{3}$ and $\frac{2}{3}$ respectively. Let u_1 be the point for the percentile $\frac{2}{3}$ of the first distribution, and u_2 be the point for the percentile $\frac{1}{3}$ of the second distribution. Then:

$$\int_0^{u_1} \frac{x}{\sigma^2}\, e^{-x^2/2\sigma^2} dx \;=\; \frac{2}{3} \implies u_1 = \sigma\sqrt{2\ ln(3)}, \text{ and} \quad (6)$$

$$u_2 = \theta + \sigma\sqrt{2ln\left(\frac{3}{2}\right)}. \quad (7)$$

We now consider the efficiency of the CMOS.

Theorem 1: 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 $ln\left(\frac{x}{x-\theta}\right) = \frac{-\theta^2+2\theta x}{2\sigma^2}$ deviates from $\frac{\theta}{2} + \frac{\sigma}{\sqrt{2}}\left(\sqrt{ln(3)} + \sqrt{ln\left(\frac{3}{2}\right)}\right)$.

Proof: 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:

$$\begin{aligned}
p(x|\omega_1)P(\omega_1) &\underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(x|\omega_2)P(\omega_2) \\
\implies \frac{x}{\sigma^2}e^{\frac{-x^2}{2\sigma^2}} &\underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{x-\theta}{\sigma^2}e^{\frac{-(x-\theta)^2}{2\sigma^2}} \\
\implies \frac{x}{x-\theta} &\underset{\omega_2}{\overset{\omega_1}{\lessgtr}} e^{\frac{-(x-\theta)^2}{2\sigma^2}+\frac{x^2}{2\sigma^2}} \\
\implies ln\left(\frac{x}{x-\theta}\right) &\underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{-\theta^2+2\theta x}{2\sigma^2}. \quad (8)
\end{aligned}$$

The discriminant is then the solution to the transcendental equation:

$$ln\left(\frac{x}{x-\theta}\right) = \frac{-\theta^2+2\theta x}{2\sigma^2}. \quad (9)$$

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. (15) and (16), $u_1 = \sigma\sqrt{2\ ln(3)}$ and $u_2 = \theta + \sigma\sqrt{2ln\left(\frac{3}{2}\right)}$. The discriminant enforced by the 2-OS classifier satisfies:

$$D(x,u_1) = D(x,u_2). \quad (10)$$

The condition imposed by Eq. (10) leads to the following:

$$\begin{aligned}
D(x,u_1) &= D(x,u_2) \\
\implies D\left(x,\ \sigma\sqrt{2\ ln(3)}\right) &= D\left(x,\ \theta + \sigma\sqrt{2ln\left(\frac{3}{2}\right)}\right) \\
\implies 2x &= \theta + \sigma\sqrt{2ln(3)} + \sigma\sqrt{2ln\left(\frac{3}{2}\right)} \\
\implies x &= \frac{\theta}{2} + \frac{\sigma}{\sqrt{2}}\left(\sqrt{ln(3)} + \sqrt{ln\left(\frac{3}{2}\right)}\right). \quad (11)
\end{aligned}$$

The difference in the errors of the two classifiers is clearly related to differences in the corresponding discriminant functions specified by Eq. (9) and (11). Hence the theorem. ■

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. (9) 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 under the curve (in percentage) is depicted in Table I. 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.

TABLE I
The differences in Error Obtained By The CMOS Classifier When Compared To The Bayesian Classifier, For Different Values Of θ Of The Rayleigh Distribution. In Each Case, $\sigma = 2$.

θ	1	1.5	2	2.5	3
Max. Bounded Error(in %)	0.15	0.06	0.05	0.001	0

Theorem 2: For the 2-class problem in which the two class conditional distributions are Rayleigh and identical, CMOS, the accuracy obtained by 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 $\frac{\theta}{2} + \frac{\sigma}{\sqrt{2}}\left(\sqrt{ln(3)} + \sqrt{ln\left(\frac{3}{2}\right)}\right)$.

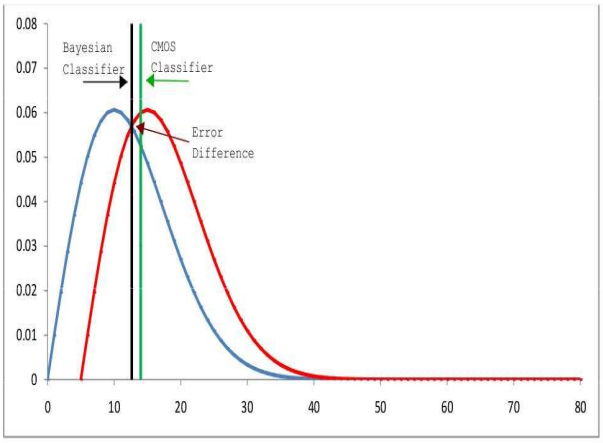

Fig. 1. The differences of the error probabilities.

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 1, as the equations are transcendental, we first 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,

$$\begin{aligned} D(x,\nu_1) &< D(x,\nu_2) \\ \implies & x - \sigma\sqrt{ln(4)} < \theta + \sigma\sqrt{ln(4)} - x \\ \implies & 2x < \theta + 2\sigma\sqrt{ln(4)} \\ \implies & x < \frac{\theta}{2} + \sigma\sqrt{ln(4)}. \end{aligned} \tag{12}$$

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. (15) and (16), $u_1 = \sigma\sqrt{2\ ln(3)}$ and $u_2 = \theta + \sigma\sqrt{2ln\left(\frac{3}{2}\right)}$. The discriminant enforced by 2-OS CMOS is:

$$D(x,u_1) = D(x,u_2). \tag{13}$$

This equation simplifies to:

$$\begin{aligned} D(x,u_1) &= D(x,u_2) \\ \implies & D\left(x,\ \sigma\sqrt{2\ ln(3)}\right) = D\left(x,\ \theta + \sigma\sqrt{2ln\left(\frac{3}{2}\right)}\right) \\ \implies & 2x = \theta + \sigma\sqrt{2ln(3)} + \sigma\sqrt{2ln\left(\frac{3}{2}\right)} \\ \implies & x = \frac{\theta}{2} + \frac{\sigma}{\sqrt{2}}\left(\sqrt{ln(3)} + \sqrt{ln\left(\frac{3}{2}\right)}\right). \end{aligned} \tag{14}$$

The difference in the errors of the two classifiers is clearly related to differences in the corresponding discriminant functions specified by Eq. (12) and (14). Hence the theorem. ■

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: The proof of this corollary is straightforward and omitted here, but can be found in [9]. ■

Data Generation: Rayleigh Generation 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 [10].

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 II. The results presented justify the claims of Theorems 1 and 2.

Theoretical Analysis: Rayleigh Distribution - k-OS We have seen from Theorem 1 that for the Rayleigh distribution, the moments of the 2-OS are sufficient for a near-optimal classification. As in the case of the other distributions, we shall now consider the scenario when we utilize other k-OS. Let u_1 be the point for the percentile $\frac{n+1-k}{n+1}$ of the first distribution, and u_2 be the point for the percentile $\frac{k}{n+1}$ of the second distribution. Then:

$$\int_0^{u_1} \frac{x}{\sigma^2}\ e^{-x^2/2\sigma^2} dx = \frac{n+1-k}{n+1}$$

$$\implies u_1 = \sigma\sqrt{2\ ln\left(\frac{n+1}{k}\right)},\ \text{and} \tag{15}$$

$$u_2 = \theta + \sigma\sqrt{2\ ln\left(\frac{n+1}{n+1-k}\right)}. \tag{16}$$

The k-OS results of CMOS follow.

TABLE II
A COMPARISON OF THE ACCURACY OF THE BAYESIAN AND THE 2-OS CMOS CLASSIFIER FOR THE RAYLEIGH DISTRIBUTION.

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

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

Proof: The proof of this theorem is omitted here, but is included in [9]. ■

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 III. 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 approach fails if the condition stated in Theorem 3 is not 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 $E[\mathbf{x}_{k,n}] = \sigma\sqrt{(2\ ln\left(\frac{n+1}{k}\right))}$. For the cases where the condition $\sqrt{ln\left(\frac{n+1}{k}\right)} - \sqrt{ln\left(\frac{n+1}{n+1-k}\right)} < \frac{\theta}{\sigma\sqrt{2}}$, 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 $\sigma\sqrt{(2\ ln\left(\frac{7}{1}\right))}$ and $\theta + \sigma\sqrt{(2\ ln\left(\frac{7}{6}\right))}$, the inequality of the condition imposed in Theorem 3 simplified to $1.002339 < 0.88388$, which is not valid. Observe that if $\sqrt{ln\left(\frac{n+1}{k}\right)} - \sqrt{ln\left(\frac{n+1}{n+1-k}\right)} > \frac{\theta}{\sigma\sqrt{2}}$, the symmetric pairs should be reversed to obtain the near-optimal Bayes' bound. This concludes our study on the CMOS for the Rayleigh distribution.

V. THE GAMMA DISTRIBUTION

The Gamma distribution is a continuous probability distribution with two parameters - a, a shape parameter and b, a scale parameter. The pdf of the Gamma distribution is $\frac{1}{\Gamma(a)\ b^a}\ x^{a-1}e^{\frac{-x}{b}};\ a>0,\ b>0,$

with mean ab and variance ab^2 where a and b are the parameters. Unfortunately, the cumulative distribution function does not have a closed form expression [11]–[13].

Theoretical Analysis: Gamma Distribution

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

We first derive the moments of the 2-OS, which are the points of interest for CMOS, for the Gamma distribution. Let u_1 be the point for the percentile $\frac{2}{3}$ of the first distribution, and u_2 be the point for the percentile $\frac{1}{3}$ of the second distribution. Then:

$$\int_0^{u_1} x\ e^{-x}dx\ =\ \frac{2}{3}$$

$$\implies ln(u_1) - 2u_1 = ln\left(\frac{1}{3}\right),\ \text{and} \qquad (17)$$

$$ln(u_2-\theta) - 2(u_2-\theta) = ln\left(\frac{1}{3}\right) - ln(\theta). \qquad (18)$$

The following results hold for the Gamma distribution.

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

Proof: The proof of this theorem can be found in [9]. ■

Data Generation: Gamma Distribution There are a number of data generation algorithms reported for the Gamma distribution, all of which make use of the Uniform random variate $U(0,1)$. The data is generated using the built-in function available in MatLab, 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

TABLE III
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 3 WERE SATISFIED.

No.	Order(n)	Moments	OS_1	OS_2	CMOS	Pass/Fail
1	Two	$\left(\frac{2}{3}, \frac{1}{3}\right)$	$\sigma\sqrt{\left(2\ ln\left(\frac{3}{1}\right)\right)}$	$\theta+\sigma\sqrt{\left(2\ ln\left(\frac{3}{2}\right)\right)}$	82.05	Passed
2	Four	$\left(\frac{5-i}{5}, \frac{i}{5}\right), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{5}{1}\right)\right)}$	$\theta+\sigma\sqrt{\left(2\ ln\left(\frac{5}{4}\right)\right)}$	81.8	Passed
3	Four	$\left(\frac{5-i}{5}, \frac{i}{5}\right), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{5}{2}\right)\right)}$	$\theta+\sigma\sqrt{\left(2\ ln\left(\frac{5}{3}\right)\right)}$	82.0	Passed
4	Six	$\left(\frac{7-i}{7}, \frac{i}{7}\right) 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{7}{1}\right)\right)}$	$\theta+\sigma\sqrt{\left(2\ ln\left(\frac{7}{6}\right)\right)}$	18.4	Failed
5	Six	$\left(\frac{7-i}{7}, \frac{i}{7}\right) 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{7}{2}\right)\right)}$	$\theta+\sigma\sqrt{\left(2\ ln\left(\frac{7}{5}\right)\right)}$	82.10	Passed
6	Six	$\left(\frac{7-i}{7}, \frac{i}{7}\right) 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{7}{3}\right)\right)}$	$\theta+\sigma\sqrt{\left(2\ ln\left(\frac{7}{4}\right)\right)}$	82.15	Passed
7	Eight	$\left(\frac{9-i}{9}, \frac{i}{9}\right), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{9}{1}\right)\right)}$	$\theta+\sigma\sqrt{\left(2\ ln\left(\frac{9}{8}\right)\right)}$	18.45	Failed
8	Eight	$\left(\frac{9-i}{9}, \frac{i}{9}\right), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{9}{2}\right)\right)}$	$\theta+\sigma\sqrt{\left(2\ ln\left(\frac{9}{7}\right)\right)}$	82.05	Passed
9	Eight	$\left(\frac{9-i}{9}, \frac{i}{9}\right), 1 \leq i \leq \frac{n}{2}$	$\sigma\sqrt{\left(2\ ln\left(\frac{9}{4}\right)\right)}$	$\theta+\sigma\sqrt{\left(2\ ln\left(\frac{9}{5}\right)\right)}$	82.15	Passed

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.

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

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

Proof: The proof of this theorem is included in [9]. ■

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

Interestingly enough, if we examine the table, we can see that the Bayes' accuracy is the highest value except for the case where $\theta = 3.0$, although this result must, in fact, be be considered as an aberration. This concludes the study of the Gamma distribution with regard to the CMOS classification.

VI. THE BETA DISTRIBUTION

The Beta distribution is a family of continuous probability distributions defined in $(0, 1)$ parameterized by two shape parameters α and β. The distribution can take different shapes based on the specific values of the parameters. If the parameters are identical, the distribution is symmetric with respect to $\frac{1}{2}$. Further, if $\alpha = \beta = 1$, $B(1, 1)$ becomes $U(0, 1)$. The pdf of the Beta distribution is $f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)}\ x^{\alpha-1}(1-x)^{\beta-1}$. The mean and the variance of the distribution are $\frac{\alpha}{\alpha+\beta}$ and $\frac{\alpha\beta}{(\alpha+\beta)^2\ (\alpha+\beta+1)}$ respectively. We consider the case when $\alpha = \beta > 1$.

For this study, we mainly consider three cases:

- $\alpha = 1,\ \beta = 1$: Uniform Distribution.
- $\alpha = \beta$: Symmetric about $\frac{1}{2}$.
- $\alpha > 1,\ \beta > 1$: Unimodal Distribution.

Earlier, in paper [8], when we first introduced the concept of CMOS-based PR, we had analyzed the 2-OS and k-OS CMOS for the Uniform distribution, and had provided the corresponding theoretical analysis and the experimental results. We had concluded that, for the 2-class problem in which the two class conditional distributions are Uniform and identical, CMOS can, indeed, attain the optimal Bayes' bound. So, in this paper, to avoid repetition, we skip the analysis for the Beta distribution, B(1,1), as this case reduces to the analysis for Uniform U(0,1). Thus, we reckon that the first of these cases (i.e., when $\alpha = 1$ and $\beta = 1$) as being closed.

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

Theoretical Analysis: Beta Distribution ($\alpha = \beta$) 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. We consider the scenario when $\alpha_1 = \alpha_2 = \alpha$, $\beta_1 = \beta_2 = \beta$, and for the sake of simplicity, $\alpha = \beta = 2$. Then, the distributions are: $f(x, 2, 2) = 6x(1-x)$ and $f(x-\theta, 2, 2) = 6(x-\theta)(1-x+\theta)$.

TABLE IV
A COMPARISON OF THE ACCURACY WITH RESPECT TO THE MEDIAN AND THE 2-OS CMOS CLASSIFIER FOR THE GAMMA DISTRIBUTION.

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

TABLE V
A COMPARISON OF THE k-OS CMOS CLASSIFIER WHEN COMPARED TO THE BAYES' CLASSIFIER AND THE CLASSIFIER WITH RESPECT TO MEDIAN AND MEAN FOR THE GAMMA DISTRIBUTION FOR DIFFERENT VALUES OF n. IN EACH COLUMN, THE VALUE WHICH IS NEAR-OPTIMAL IS RENDERED BOLD.

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

We first derive the moments of the 2-OS, which are the points of interest for CMOS, for the Beta distribution. By virtue of Eq. (2) and (3), 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. As the distribution takes different forms based on the values of the shape parameters, we have to solve each case separately, and so we can obtain numerical values for the CMOS positions. 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:

$$\begin{aligned} \int_0^{u_1} 6x(1-x)dx &= \frac{2}{3} \\ \implies -6u_1^3 + 9u_1^2 - 2 &= 0. \end{aligned} \tag{19}$$

Similarly, if we don't take the displacement, θ, into consideration, the form for u_2 leads to the equation:

$$-6u_2^3 + 9u_2^2 - 1 = 0. \tag{20}$$

We shall now prove that the CMOS, indeed, attains the Bayes' bound.

Theorem 6: 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:

$$\begin{aligned} p(x|\omega_1)P(\omega_1) &\underset{\omega_2}{\overset{\omega_1}{\gtrless}} p(x|\omega_2)P(\omega_2) \\ \implies \quad 6x(1-x) &\underset{\omega_2}{\overset{\omega_1}{\gtrless}} 6(x-\theta)(1-(x-\theta)) \\ \implies \quad x &\underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{\theta+1}{2}. \end{aligned} \tag{21}$$

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

$$x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{\theta+1}{2} \implies D(x, u_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(x, u_2). \tag{22}$$

If we examine the Eqs. (19) and (20), we can see that Eq. (20) can be obtained by substituting $1 - u_2$ for u_1 in Eq. (19) as:

$$-6(1-u_2)^3 + 9(1-u_2)^2 - 2 = 0 \implies -6u_2^3 + 9u_2^2 - 1 = 0. \tag{23}$$

From this, it is obvious that $u_2 = \theta + u_1 - 1$. Consequently,

the RHS of the claim given by Eq. (22) leads to the following:

$$\begin{aligned} & D(x,u_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(x,u_2) \\ \Longrightarrow \quad & D(x,u_1) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} D(x,\theta+1-u_1) \\ \Longrightarrow \quad & x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \frac{\theta+1}{2}. \end{aligned} \tag{24}$$

The result follows by observing that Eqs. (21) and (22) are identical comparisons. Hence the theorem. ■

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 VI. 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 6.

We have seen from Theorem 6 that the moments of the 2-OS are sufficient for the classification to attain a Bayes' bound. Now we shall examine the scenario where the k-OS CMOS is invoked, and thus determine the strength of the proposed method.

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:

$$\begin{aligned} \int_0^{u_1} 6x(1-x)dx &= \frac{n+1-k}{n+1} \\ \Longrightarrow -2u_1^3+3u_1^2-\frac{n+1-k}{n+1} &= 0. \end{aligned} \tag{25}$$

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:

$$-2u_2^3+3u_2^2-\frac{k}{n+1}=0. \tag{26}$$

We shall now prove that the CMOS attains the Bayes' bound.

Theorem 7: 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 $u_1 < u_2$.

Proof: The proof of this theorem is included in [9] and omitted here in the interest of space. ■

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. A subset of the results are found in Table VII.

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 7 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 the study on the symmetric Beta distribution.

VII. Conclusions

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

References

[1] A. Thomas and B. J. Oommen, "Optimal "Anti-Bayesian" Parametric Pattern Classification Using Order Statistics Criteria," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, vol. 7441 of *Lecture Notes in Computer Science*, pp. 1–13, Springer Berlin / Heidelberg, 2012. This was a Plenary/Keynote Talk at the Conference.

[2] A. Thomas and B. J. Oommen, "Optimal "Anti-Bayesian" Parametric Pattern Classification for the Exponential Family Using Order Statistics Criteria," in *Image Analysis and Recognition*, vol. 7324 of *Lecture Notes in Computer Science*, pp. 11–18, Springer Berlin / Heidelberg, 2012.

[3] M. Ahsanullah and V. B. Nevzorov, *Order Statistics: Examples and Exercises*. Nova Science Publishers, Inc, 2005.

[4] K. W. Morris and D. Szynal, "A goodness-of-fit for the Uniform Distribution based on a Characterization," *Journal of Mathematical Science*, vol. 106, pp. 2719–2724, 2001.

[5] G. D. Lin, "Characterizations of Continuous Distributions via Expected values of two functions of Order Statistics," *Sankhya: The Indian Journal of Statistics*, vol. 52, pp. 84–90, 1990.

[6] Y. Too and G. D. Lin, "Characterizations of Uniform and Exponential Distributions," *Academia Sinica*, vol. 7, no. 5, pp. 357–359, 1989.

[7] A. Thomas, *Pattern Classification using Novel Order Statistics and Border Identification Methods*. PhD thesis, School of Computer Science, Carleton University, 2012. (To be Submitted).

[8] A. Thomas and B. J. Oommen, "The Fundamental Theory of Optimal "Anti-Bayesian" Parametric Pattern Classification Using Order Statistics Criteria," *Pattern Recognition*, vol. 46, pp. 376–388, 2013.

[9] A. Thomas and B. J. Oommen, "Optimal Order Statistics-based "Anti-Bayesian" Parametric Pattern Classification for the exponential family," 2012. (To be submitted).

TABLE VI
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.

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 VII
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 7 WERE SATISFIED.

Trial No.	**Order(n)**	**Moments**	OS_1	OS_2	**CMOS**	**Pass/Fail**
1	Two	$\left\langle \frac{2}{3}, \frac{1}{3} \right\rangle$	0.61304	$\theta + 0.38696$	87.3	Passed
2	Four	$\left\langle \frac{4}{5}, \frac{1}{5} \right\rangle$	0.71286	$\theta + 0.28714$	87.3	Passed
3	Four	$\left\langle \frac{3}{5}, \frac{2}{5} \right\rangle$	0.56707	$\theta + 0.43293$	87.3	Passed
4	Six	$\left\langle \frac{6}{7}, \frac{1}{7} \right\rangle$	0.7621	$\theta + 0.23790$	87.3	Passed
5	Six	$\left\langle \frac{5}{7}, \frac{2}{7} \right\rangle$	0.6471	$\theta + 0.3529$	87.3	Passed
6	Six	$\left\langle \frac{4}{7}, \frac{3}{7} \right\rangle$	0.54776	$\theta + 0.45224$	87.3	Passed
7	Eight	$\left\langle \frac{8}{9}, \frac{1}{9} \right\rangle$	0.79269	$\theta + 0.20731$	12.7	Failed
8	Eight	$\left\langle \frac{7}{9}, \frac{2}{9} \right\rangle$	0.69508	$\theta + 0.30492$	87.3	Passed
9	Eight	$\left\langle \frac{5}{9}, \frac{4}{9} \right\rangle$	0.53711	$\theta + 0.46289$	87.3	Passed

[10] L. Devroye, *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.
[11] P. R. Krishnaih and M. H. Rizvi, "A note on Moments of Gamma Order Statistics," *Technometrics*, vol. 9, pp. 315–318, 1967.
[12] P. R. Tadikamalla, "An approximation to the moments and the percentiles of Gamma Order Statistics," *Sankhya: The Indian Journal of Statistics*, vol. 39, pp. 372–381, 1977.
[13] D. H. Young, "Moment relations for Order Statistics of the standardized Gamma distribution and the inverse multinomial distribution," *Biometrika*, vol. 58, pp. 637–640, 1971.