

Study on Reliability of KAPS Metric as Degree of Separability for Classification Problems in ATR Systems

Giovanni Marino*

European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen, Maritime affairs, Via Enrico Fermi 2749, I-21027, Ispra (Va), Italy.

* Corresponding author. Tel.: +390332789543; email: giovanni.marino@jrc.ec.europa.eu

Manuscript submitted January 16, 2015; accepted May 26, 2015.

doi: 10.17706/ijcce.2016.5.1.23-30

Abstract: Prediction of separability of potential targets classes is an important issue in ATR systems. Unfortunately despite ATR systems have been adopted since the 1960's, the definition of class-separability is still an open question. Kolmogorov Appropriate Prediction of Separability (KAPS) seems to be a powerful tool to predict the degree of separability of classes. The article is addressed to understand how reliable the metric is in the degree of separability assessment. The methods used for evaluating the performance of metric are order statistics analysis and Monte Carlo simulations. The analysis consisted in understanding how many samples are necessary in order to get a reliable value of KAPS metric. As demonstrated by results, the KAPS value depends on the number of samples, therefore it is suggested to use a number of samples of the order 10^3 at least.

Key words: Automatic target recognition, machine learning, KAPS, class-separability, degree of separability.

1. Introduction

An Automatic Target Recognition (**ATR**) system is a sensor which is usually able to recognize targets or objects based on gathered data. The application of automatic target recognition technology is a critical element of many military and industrial systems.

Currently there are many systems which are able to collect data (e.g. radar sensor, electro-optic sensor, infrared devices) which are commonly used to collect information and detect, recognise and classify potential targets. Despite significant efforts during the last decades, some problems in ATR systems have not been solved yet. Particularly, prediction of separability of potential targets classes is crucial for a good classification of the objects and its definition is still an open question ([1]-[3]). Quickly, two classes are separated if they have no intersection, thereby a necessary condition for class-separability is that the error area (i.e. the overlapping area between two classes) is equal to zero.

In practice the above-mentioned requirement is never satisfied (i.e. PDFs -Probability Density Function- have usually a $(-\infty, +\infty)$ domain), therefore a high degree of separability when a false alarm rate associated with classification problems is considered can be reached by minimizing the Type II error area (i.e. fixed false alarm rate of a misclassified target-class with respect to another one, the highest degree of separability is reached if the maximum misclassification error of the considered class is less or equal to a fixed false alarm rate) as reported in Fig. 1. Therefore, mathematically, the degree of separability of a class with respect to another can be considered as the largest subset of samples which can be classified correctly, and in case of separability it has to be equal to 1.

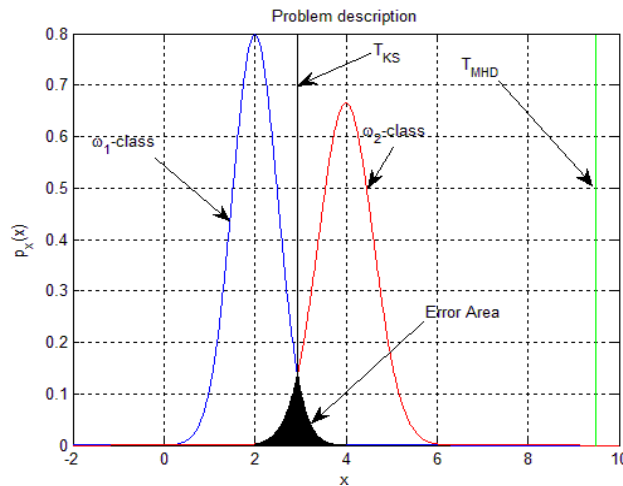


Fig. 1. Problem description: A classification problem consists of labelling samples correctly and the problem can be pictured as in this figure. KAPS metric is based on the idea that the smaller the error area (i.e. the black area) is, the more separated the classes are.

Kolmogorov Appropriate Prediction of Separability (**KAPS**) [1] is a new criterion based on the Kolmogorov-Smirnov test [4] which is able to manage the separability when a constraint about the maximum tolerated Type II error value is introduced, to predict if two classes are separated or not and at what degree. It is based on the idea that the smaller the error area is, the more separated the classes are.

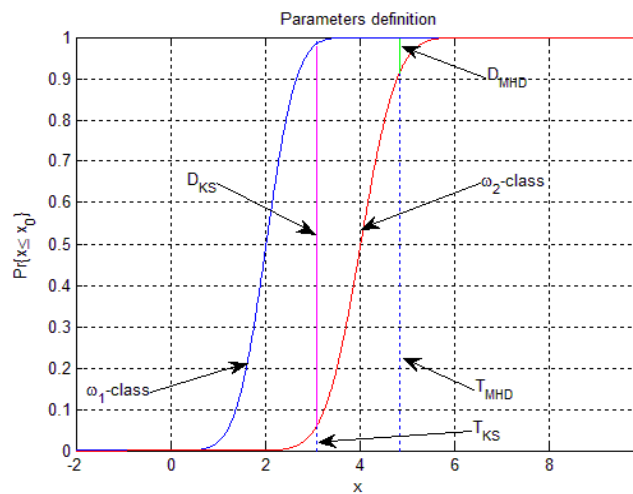


Fig. 2. Graphical representation of the parameters of KAPS metric.

In Fig. 2 a graphical representation of the most important parameters of KAPS metric is given: Mathematically KAPS metric is defined as:

$$KAPS = \frac{D_{KS}}{D_{MHD}} \quad (1)$$

where D_{KS} is the Kolmogorov-Smirnov distance (i.e. the distance between two classes CDFs (Cumulative Density Function) corresponding at T_{KS}), whereas D_{MHD} is the analogous distance corresponding at T_{MHD} . T_{KS} and T_{MHD} are defined as:

$$T_{KS} = \max_{x \in \mathbb{R}} \left\{ \int_{-\infty}^{+\infty} |p_1(x) - p_2(x)| dx \right\} \Rightarrow p_1(x) = p_2(x) \quad (2)$$

$$T_{MHD} = \min_{x \in \mathbb{R}} \left\{ \int_x^{+\infty} p(\xi) d\xi \leq \alpha \right\}, \alpha \ll 1 \quad (3)$$

The aim of the article is to understand how the value of the KAPS metric varies as a function of sample size. The methods used for assessing the performance of metrics are order statistics analysis [5] and Monte Carlo simulations [6].

As for the structure of the article, it is organized as follows: Firstly an introduction of the adopted methodologies is given, then the results and conclusions are reported.

2. Methodologies

2.1. Order Statistics

Order statistics analysis [5] is a fundamental tool in nonparametric statistics. It consists usually of computing the distribution of a sorted (in increasing order) set of the realizations of a random variable. A classic example of order statistics can be described as follows: Drawn n -samples independent and identically distributed $\{x_1, \dots, x_n\}$, from a distribution with PDF $f(x)$ and CDF $F(x)$, find the distribution of the following random variables:

$$U = \max\{x_1, \dots, x_n\} \quad (4)$$

$$V = \min\{x_1, \dots, x_n\} \quad (5)$$

The problem of (4) can be solved easily by defining the set of sample values as follows:

$$A_U = \{U \leq u\} \text{ (the maximum is at most } u) = \{x_1 \leq u\} \cap \{x_2 \leq u\} \cap \dots \cap \{x_n \leq u\} \quad (6)$$

Therefore we obtain:

$$F(U) = \Pr\{U \leq u\} = \prod_{i=1}^n \Pr\{x_i \leq u\} = [F(u)]^n \quad (7)$$

$$f(U) = \frac{d}{dx} [F(u)]^n = n[F(u)]^{n-1} f_x(U) \quad (8)$$

Problem defined in (5) can be solved in a similar way. Our problem however is different. We are indeed interested in computing the statistics of the j -th smallest value of the samples. Suppose that our x_i , $i = 1, \dots, n$, is drawn from a continuous distribution with CDF $F_x(x)$ and PDF $f_x(x)$. The j -th order statistics is hence equal to:

$$F_{x(j)}(x) = \Pr\{x(j) \leq \tilde{x}\} = \sum_{k=j}^n \binom{n}{k} [F_x(x)]^k [1 - F_x(x)]^{n-k} \quad (9)$$

$$f_{x(j)}(x) = \frac{n!}{(j-1)!(n-j)!} f_x(x) [F_x(x)]^{j-1} [1 - F_x(x)]^{n-j} \quad (10)$$

i.e. (9) gives us the probability that a sample is placed at position j when a set of n -samples is ordered. Equation (10) is obtained as derivative of (9).

2.2. Monte Carlo Simulations

From a mathematical point view [6], consider a (possibly multidimensional) random variable X having probability mass function or probability density function $f_x(x)$ which is greater than zero on a set of values

$\{X\}$. Then the expected value of a function g of X is:

$$E\{g(X)\} = \sum_{x \in \{X\}} g(x) f_X(x) \quad (11)$$

Now, if we were to take an n -sample of X 's, (x_1, \dots, x_n) , and we computed the mean of $g(x)$ over the sample, then we would have the Monte Carlo estimate of the (11):

$$\tilde{g}_n(x) = \frac{1}{n} \sum_{i=1}^n g(x_i) \quad (12)$$

We could, alternatively, speak of the random variable:

$$\tilde{g}_n(x) = \frac{1}{n} \sum_{i=1}^n g(X) \quad (13)$$

which we call the Monte Carlo estimator of (11).

If $E\{g(X)\}$ exists then the weak law of large numbers tells us that for any arbitrarily small $\varepsilon > 0$:

$$\lim_{n \rightarrow +\infty} \Pr\{|\tilde{g}_n(x) - E\{g(X)\}| \geq \varepsilon\} = 0 \quad (14)$$

The limit (14) tells us that as n gets large, then there is a small probability that $\tilde{g}_n(x)$ deviates much from $E\{g(X)\}$. For our purposes, the strong law of large numbers says the same thing (the important part being that as long as n is large enough), $\tilde{g}_n(x)$ arising from a Monte Carlo experiment shall be close to $E\{g(X)\}$, as desired. Moreover $\tilde{g}_n(x)$ is unbiased for $E\{g(X)\}$:

$$E\{g(X)\} = E\left\{\frac{1}{n} \sum_{i=1}^n g(x_i)\right\} = \frac{1}{n} \sum_{i=1}^n E\{g(x_i)\} = E\{g(X)\} \quad (15)$$

Therefore a Monte Carlo simulation becomes useful when one realizes that very many quantities of interest may be cast as expectations.

2.3. Experiment Description

Determining the performance of the KAPS metric as a function of sample size rises two kinds of problems: first, the computation of order statistics (9) and (10) is quite hard in many cases, whereas Monte Carlo requires a high number of samples, which is very time consuming. In order to solve both problems the following criteria have been adopted: As for assessing values of the T_{KS} and D_{KS} as function of samples, Monte Carlo analysis has been adopted, whereas for T_{MHD} and D_{MHD} an analytical study has been performed.

The considered experiment consists in computing the value of KAPS parameters when a class of the sample is drawn from an exponential distribution (with scale parameter $\lambda = 1$) whereas the second one is drawn from a Gaussian distribution with mean value $E\{g(X) = 5\}$ and the standard deviation equal to $\sigma_g = 1$.

The experiment has been performed for an exponential distribution with scale parameter $\lambda = 1$, but similar consideration are still valid for other kind of exponential distributions. Moreover, as proved in [1], the exponential distribution is an upper bound for many clutter models, therefore the experiments are a good example for classification of potential targets in a radar scenario. In order to compute how the T_{MHD} varies when the sample size changes, we derived from (10) the order statistics PDF for i.i.d. samples drawn from an exponential distribution:

$$\frac{n!}{(j-1)!(n-j)!} (1 - e^{-x})^{j-1} (e^{-x})^{n-j+1} \tag{16}$$

Since usually for $x = T_{MHD}$ (in this case $T_{MHD} = 13.8155$, so that the Type error II is $\leq 10^{-6}$, [1]) $e^{-x} \ll 1$, the (16) can be rewritten as:

$$\frac{n-j+1}{n} \left[\frac{(n+1)!}{(j-1)!(n-j+1)!} x^{j-1} (1-x)^{n-j+1} \right] \tag{17}$$

By considering the Taylor approximation stopped at first order of e^{-x} . The term between brackets in (17) is a $Beta(j, n - j + 2)$ distribution whose mean value and variance are respectively:

$$E_x = \frac{n-j+1}{n} \frac{j}{n+2} \tag{18}$$

$$Var_x \leq \frac{n-j+1}{n} \frac{j(n-j+2)}{(n+2)^2(n+3)} \tag{19}$$

The inequality of (19) is valid because $\frac{n-j+1}{n} \leq 1$ in (17).

Because of the linear approximation, we are interested in the median value of j (i.e. we are interested in computing the order statistics at the approximation point which is the median value of the approximated Beta distribution) which gives the highest outcomes, as expected, therefore we can assess for $j = n/2$ the value of the (19).

As for T_{KS} value, the (17) is usually no more valid because of the order of magnitude of the function in the variable range, therefore only a numerical Monte Carlo simulation can be performed (number of trials equal to 10^3).

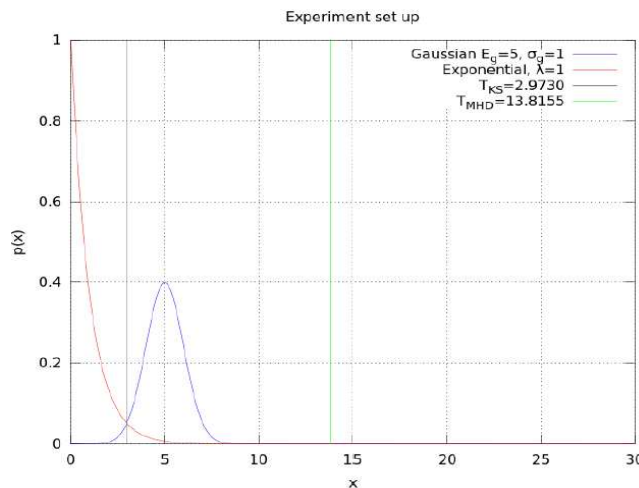


Fig. 3. Experimental set up. The exponential and the Gaussian distribution are depicted. T_{KS} is represented by the vertical black line, as defined in [1]. The green line represents the T_{MHD} threshold and it is the value of variable x where $p_1(x)$, i.e. the exponential distribution produces a false alarm rate, i.e. the Type error II 10^{-6} .

As described in [1], the value of T_{KS} is equal to the value of the random variable x where $p_1(x) = p_2(x)$, $p_i(x), i = 1, 2$ are the PDFs of the samples. For our experiments T_{KS} has been set to 2.9730. The experimental set up is reported in Fig. 3.

The algorithm used for the simulations can be summarized as follows:

- 1) m samples are drawn from an exponential distribution, $m = 10, 100, 1000$.
- 2) The closest sample to the value 2.9730 is selected.
- 3) The error between the selected sample and the value 2.9730, as value difference, is calculated.
- 4) Go to point 1 until the number of iterations is 1000.
- 5) Compute the mean and standard deviation values of the error.

3. Results

As depicted in Fig. 4, the mean value behaviour of (18) as function of j and for $n = 10$ (similar behaviour for $n = 100$ and $n = 1000$, as reported in Table 1) number of samples presents a local maximum at the median value, as expected. In Fig. 5 however is reported how the variance (19) varies as a function of j . As for the Monte Carlo simulations, the value of T_{KS} has been set as described in [1] (i.e. 2.9730) and the results are reported in Table 2.

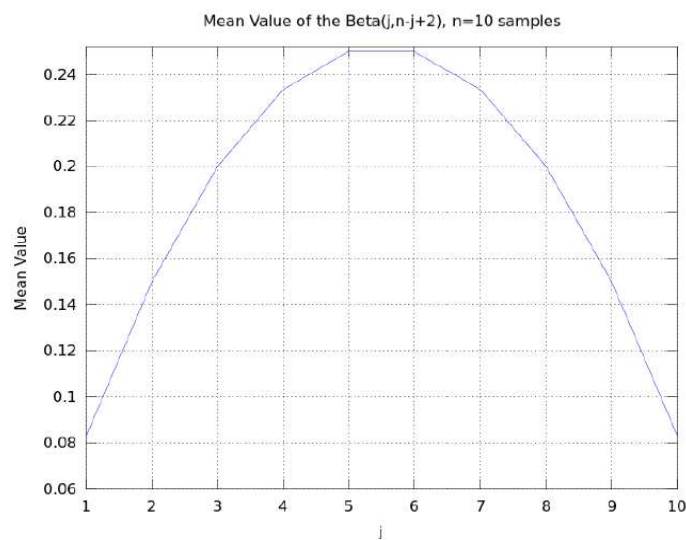


Fig. 4. The mean value behaviour of (18) as a function of j and for $n = 10$. The mean presents a peak in $j = 5$ as expected.

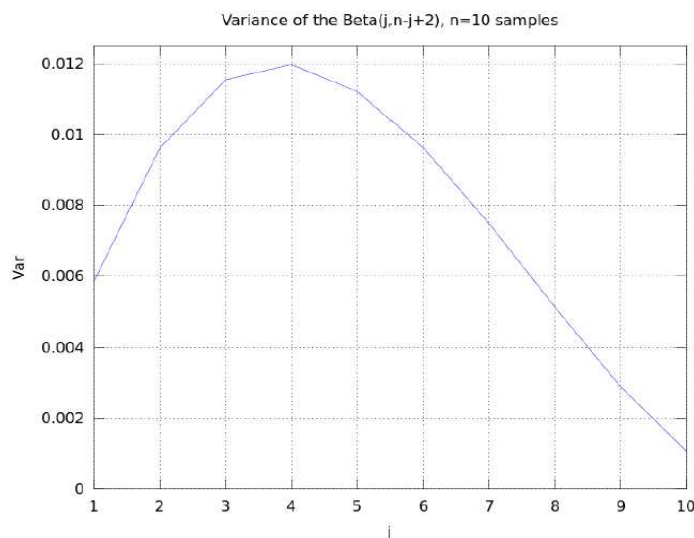


Fig. 5. Variance as defined in (19) for $n = 10$. The value selected corresponding to $j = 5$ is equal to 0.011218.

Table 1: Mean and Variance Values of the Peak of $Beta(j, n - j + 2)$

Sample size	E_x	Var_x
$n = 10$	0.25000	0.011218
$n = 100$	0.25000	0.0012374
$n = 1000$	0.25000	$1.2487e - 04$

Table 2: Mean and Standard Deviation Values of Monte Carlo simulations

Sample size	E_x	σ_x
$n = 10$	0.48537	0.66497
$n = 100$	0.023552	0.12820
$n = 1000$	$-7.1466e - 5$	0.013363

4. Discussion

As mentioned in the previous sections the mean value of (18) presents a maximum in $j = 5$ (also for $j = 50$ and $j = 500$, because of how the mean value was defined) as expected. As known indeed in probability theory, the expected value refers, intuitively, to the value of a random variable one would “expect” to find if one could repeat the random variable process an infinite number of times and take the average of the values obtained, therefore in a range centred in T_{MHD} where the Taylor’s approximation is valid, the maximum expected value corresponds to the sample j of T_{MHD} .

As reported in Table 1, the mean value of the order statistic is independent of sample size, as a consequence the variance is reduced by increasing it, as confirmed by the values reported in the second column of Table 1. The Monte Carlo simulations however show that the assessment of the T_{KS} value, reported in Table 2, and its standard deviation are improved by increasing the sample size. In conclusion one can state that the KAPS metric is reliable for a sample size of the order of magnitude of 10^3 at least.

Prediction of separability of potential targets classes is crucial for an ATR system performance evaluation, and its definition is still an open question. In the article a definition of a new metric for the assessment of the degree of separability of potential targets classes were given, besides the paper was addressed to understand how reliable the metric is in the degree of separability assessment. It has been proved that the reliability of the metric is a function of the sample size and the metric behaviour can be considered correct when at least a number of samples equal to 10^3 is considered (see Fig. 6).

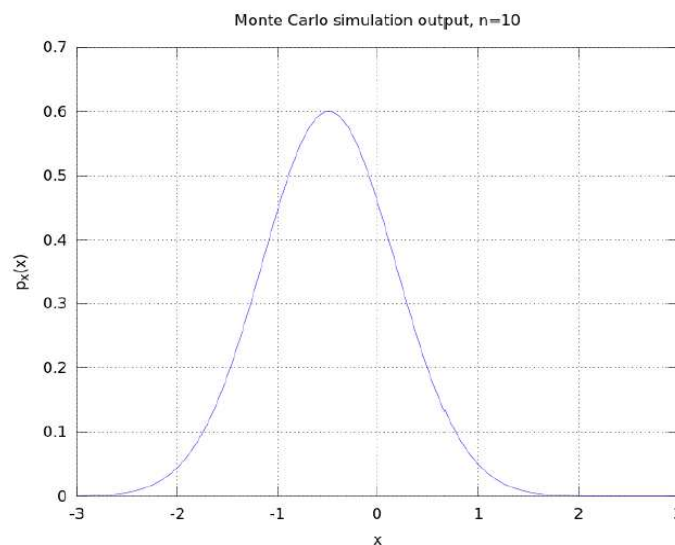


Fig. 6. Monte Carlo simulation of variation of T_{KS} for $n = 10$.

References

- [1] Corder, G. W., & Foreman, D. I. (2009). *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*. John Wiley & Sons.
- [2] Marino, G. (2012). Analysis of performance of automatic target recognition systems. Ph.D. dissertation, DISE, Cranfield Univ., Shrivenham, UK.
- [3] David, H. A., & Nagaraja, H. N. (2003). *Order Statistics* (3rd ed.). John Wiley & Sons.
- [4] Metropolis, N. (1987). The beginning of the Monte Carlo method. *Los Alamos Science*, Special Issue.
- [5] Tarkkola, K. (2003). Feature extraction by non-parametric mutual information maximization. *Journal of Machine Learning Research*, 3, 1415-1438.
- [6] Horne, A. M. (2001). An information theory for prediction of SAR target classification performance. *Proceedings of SPIE: Vol. 4382, Algorithms for Synthetic Aperture Radar Imagery VII*.



Giovanni Marino received his degree in telecommunication engineering from Università La Sapienza di Roma, Rome, Italy, in 2005. In 2006 he joined FGAN, now FHR/Fraunhofer, Wachtberg, Germany and he started his research in imaging radar. In 2012 he received his Ph.D. degree from Cranfield University, Shrivenham, United Kingdom. His research interests are in the areas of wireless sensor networks, self-organizing networks and radar systems with emphasis on mathematical modeling, performance analysis and

classification.