American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-10, Issue-10, pp: 85-94
www.ajer.org
Research Paper
Open Access

Supervised and Unsupervised Recognition of Laws from Finite Sample

Ahmed Alzahmi
Department of Electrical Engineering, Faculty of Engineering, University of Tabuk, Saudi Arabia.
aalzahmi@ut.edu.sa, ahmed.alzahmi@yahoo.com

ABSTRACT
In this paper, we address the problem of law recognition from samples with size varying from 100 to 10000 or more. The application context covers the modeling of radio-mobile channels in a situation of visibility or non-direct visibility between transmitter and receiver. This problem is crucial for improving digital communications risks. In the digital transmissions community, it is common to use distance by Kolmogorov-Smirnov. More rarely, a kernel method is considered before the com- test parative. We propose to use the information criteria (IC) to approach the probability laws by a histogram and to select the best model of law. Supervised and unsupervised cases along with utilized method have been studied and compared in a realistic situations. The results show the interest of using information criteria based methods (CIs).

KEYWORDS: *recognition of laws, supervised and unsupervised approaches, information criteriation, kernel density estimator, digital communication.*

Date of Submission: 12-10-2021 Date of acceptance: 26-10-2021

I. INTRODUCTION

The problem studied in this article concerns the recognition of statistical distributions in a wireless transmission context. This recognition is crucial for good communication, knowing that it suffers in particular from fainting fast fadings due to the different echoes of the electromagnetic wave in the environment, which disrupt transmission (Goldsmith, 2005). It is possible to predict this electromagnetic phenomenon, either rigorously via deterministic propagation models, or stochastically through adequate statistical distributions (Sarkar et al., 2003; Iskander, Yun, 2002). The first one approach is precise in estimation but very costly in computation time. In contrast, the statistical distributions are relatively fast in computation time. The challenge is therefore to identify with precision the distribution suitable for “fast fading” from field measurements, in the goal thereafter, to improve the reception of the communication. Indeed, a good estimation of these propagation channel fading, for a configuration of a given study, results in an adapted transmission strategy (coding, modulation ...) (Proakis, 1995; Goldsmith, 2005; Rappaport, 1996). In this application context of digital communications, much research has focused on this problematic of recognizing the adequate statistical law to model the variations fast.

We can cite the work of (Barbot et al., 1992; Bultitude et al., 1998; Babich, Lombardi, 2000; Hashemi, Tholl, 1994; Seitadji, Levy, 1994) as well as those of (Fryziej, 2001). However, the latter only consider the test of Kolmogorov-Smirnov which only applies to distribution functions statistics. Only (Zayen, Hayar, 2011) and (Santamaria et al., 2002) propose respectively work based on the AIC information criterion (Akaike's Information Criterion) and on the kernel method. These methods have only been used as simple tools for estimating adequate probability densities signal-to-noise ratio (AIC method) or quite simply rapid variations of the received signal (kernel method). However, these have not been compared statistically to existing selection methods such as those we will discuss afterwards. Thus, no statistical study such as the one we are proposing has not been, until now and to our knowledge, carried out in this context of communication. We will distinguish on the one hand the LOS (Light of Sight) configurations and the NLOS (Non LOS), depending on whether the transmitters and receivers are visible or non-visibility of each other. The most common statistical distribution models utilized in our study context are the laws of Rayleigh, Weibull and by Nakagami (Sarkar et al., 2003).

There are, of course, other hypotheses of laws associated with extremely marginal application contexts that will not be discussed here.

There are two possible strategies for recognition to consider the field data as a significant representation of laws. Consider field data as a significant representation of laws, the most used method in digital communication adopts this strategy by considering the distribution functions directly obtained by field measurements. A mere distance from Kolmogorov-Smirnov then allows the recognition of distributions (Bultitude et al., 1998; Babich, Lombardi, 2000). Another strategy consists in first estimating the density functions of probability optimally. Let us quote for main non-parametric methods: the kernel method and the histogram method. Once a representation of laws chosen, it is again a matter of considering recognition tests. We choose the Kullback-Leibler divergence as the most common and classic of -divergences (Basseville, 1989).

The objective of this article is to show the behavior of the three methods in the context presented above. The kernel method is often applied for estimation radio channel (Santamaria et al., 2002). Among the range of methods available on the principle of the kernel method, we choose to implement the GCE method (Generalized Cross Entropy) that is significantly efficient (Botev, 2007) and has been already considered in this application context. The proposed histogram approximation method exploits information criteria (IC for Information Criterion) (Rissanen et al., 1992; Coq, Alata, Pousset et al., 2009).

On the one hand, we place ourselves in the supervised context, where the laws are known and where we seek to recognize them, and on the other hand. In the supervised context, the laws are known and recognized. On the other hand, in unsupervised context, recognizing the law in real situation of digital communication is more likely based on the records of channel behavior. LOS and NLOS cases will be considered each time, and the performance of these methods analyzed depending on the number of signal measurement samples. Finally, CIs are effective tools in the selection of parametric models, thus overcoming the shortcomings maximum likelihood (MV) (Olivier, Alata, 2009). We will use them in the unsupervised case, as selection tools for different laws, depending on the number parameters defining them.

The use of CI is based on the following consideration: maximum likelihood used to estimate the dimension of a parametric model overestimates the number of free parameters. In this context, precisely on autoregressive models (AR), that CIs were first proposed as the AIC criterion already mentioned (Akaike, 1974). This criterion consists in minimizing the sum the likelihood term (−logMV) and a so-called penalty term. Note that AIC is itself an improvement of the FPE criterion (for Final Prediction Error) of the same author, and is based on a minimization of the Kullback-Leibler information.

Unfortunately, this AIC criterion still leads to an over-parameterization of the order of the AR model (Shibata, 1976; Alata, Olivier, 2003). To correct this deficiency of AIC, various authors have proposed other criteria such as the weakly consistent (convergence only in probability) (Hannan, Quinn, 1979), or the famous BIC criterion (Bayesian Information Criterion) (Schwarz, 1978). J. Rissanen introduced in the context of information encoding, the MDL criterion (for Minimum description Length) (Rissanen, 1978; Grünwald, 2005). MDL is quite comparable to the BIC criterion in model selection, insofar as different additive terms distinguish it from BIC (like the Fischer information), but that these terms are often negligible. BIC and MDL are strongly consistent (convergence almost sure) by penalizing the likelihood term more than AIC and. Another criterion that we will use is the criterion (El Matouat, Hallin, 1996; Jouzel et al., 1998), where is a parameter varying from 0 to 1 to have a good asymptotic behavior.

This variation interval has also been refined in (Olivier et al., 1999). This criterion can be seen as a generalization of the work of J. Rissanen on the complexity stochastic (Rissanen, 1989). This criterion is still very consistent and offers a good compromise between BIC/MDL and. From, it is also possible to find the different penalty terms of the other criteria by the following. Finally, other criteria exist but are used in remote configurations ours or very specific, such as when we have few observational data for a large number of parameters of an autoregressive model (Hurvich, Tsai, 1989; Broersen, 2000). Most often, the criteria used are not then not consistent.

II. METHODOLOGIES

The observation $xn = (xi), i = 1, \cdots, n,$ is assumed to be the realization of a process X independent and identically distributed (i.i.d.) for which each variable random admits a density of unknown probability $f(x)$, absolutely continuous by compared to the Lebesgue measure.

2.1. Approximation of probability densities

In this section, we present two density approximation methods probability from samples.

2.1.1. GCE Core Method

It is always possible to define a minimum and maximum value for the studied samples, and we assume here these normalized values between 0 and 1. There is no effect not loss of generality than to consider this hypothesis because a simple transformation invertible allows to pass necessarily bounded values of the process X (we are in an application) at interval [0, 1]. So all the random variables of the i.i.d. can be considered as admitting the realization space Ω= [0, 1]. Using the kernel method, the most direct way to approach *f (x)* from the observation is to consider the empirical distribution defined by $\Delta(x) = \frac{1}{n}\sum_{i=1}^{n} \delta(x - x_i)$which is of course suboptimal. In (Botev, 2007), an estimator by kernel method has been proposed:

$$f(t,x) = p(x)\sum_{k=0}^{+\infty} e^{\frac{\lambda kt}{2}} \varphi_k \, \Phi_k \,(x) \quad (1)$$

With: p (x) is the a priori density of the random variables contained in X;$\{1, \emptyset_1, \emptyset_2 \ldots.\}$ are the normalized eigenfunctions of the problem of Sturm- Liouville on [0, 1]. They form an orthonormal basis of L2 ([0, 1], *p (x) dx*); λ_k is the eigenvalue associated with $\emptyset_k$ $\lambda_0 = 0 > \lambda_1 > \lambda_2 > \cdots..$; $\varphi_k = \mathbb{E}_\Delta[\emptyset_k(x)] = \frac{1}{n}\sum_{i=1}^{n} \emptyset_k(x - x_i)$; *t* is an unknown parameter (the variance of the kernel) and *xi* the mean of each core. Following asymptotic considerations and assuming a nucleus Gaussian, Botev showed that the estimator *ˆf (t, x)* is consistent for the criterion of Mean Integrated Squared Error: MISE (Botev, 2007). In order to obtain *ˆf (t, x),* it suffices to determine *t* by minimizing the criterion BET which gives the following formula:

$$t = (2n\sqrt{\pi} \, \left\|(\frac{f}{p})"\right\|^2)^{-2/5} \quad (2)$$

Since (2) depends on the unknown density *f*, Botev suggests estimating *t* by finding the unique positive solution t ∗ of the following nonlinear equation (the first term of equality is indeed a monotonic function of *t*):

$$\sum_{k=1}^{+\infty} \frac{\mathbb{E}_\Delta[\emptyset_k{}^2]}{-\lambda_k} \; (e^{\frac{\lambda kt}{2}} - 1)^2 + \frac{1}{n}\sqrt{\frac{t}{\pi}} = \frac{1}{n}\int_0^1 F_\Delta(x)[1 - F_\Delta(x)]dx \quad (3)$$

Where *F (x)* is the distribution function associated with (*Δx*). We obtain the following algorithm by initializing *p (x)* to the uniform density on [0, 1]:

1. With the solution t^* from eq. (3), calculate $f_{GCE}(t^*, x)$ using eq. (1);
2. use $f_{GCE}(t^*, x)$ in eq. (2) to obtain a new estimate of t called *t*;
3. The approximation of the ddp is obtained using now t and $f_{GCE}(t^*, x)$ instead of *p (x)* in eq. (1).

2.1.2. Approximation by histogram obtained by information

Criterion as we mentioned in the introduction, most of the information criteria (AIC, BIC) were established for the selection of parametric models such as AR models. The use of criteria was then extended to find the number of classes of a histogram, approximation of a theoretical probability density *f (x)* of a random variable *X*. Suppose that the realization space of *X* is an interval I and that the only other information known is the x^n observation. The goal is to find the best histogram which summarizes *f (x)* in the sense of the criterion: number and optimal sizes of intervals or classes of this histogram. Let (Pk) be a partition in *k* intervals I_j of *I*. The criteria IC are then written:

$$IC(P_{k)} = -\sum_{j=1}^{k} n_j \log\frac{n_j}{nl_j} + kC(n) \quad (4)$$

Where L_j*, j = 1, · · ·, k* is the size of *Ij and nj, j = 1, · · ·, k* is the number of x_i in I_j. Recall that the first term of (4) is the log-likelihood term and that *C (n)* is the penalty. The partition with *k* intervals retained is that which minimizes IC. She is obtained by dynamic programming as in (Rissanen et al., 1992), which allows a reasonable complexity of the algorithm and not very time-consuming calculation.

The penalty *C (n)* differs according to the criteria (Table 1). We choose the criterion whose $\Phi_\beta$penalty is equal to C (n) = n^β log log n, with 0 <β<1 (ElMatouat, Hallin, 1996). We gave in (Olivier et al., 1999) the additional condition $\beta_{\min} = \frac{\log\log n}{\log n} \leq \beta < 1$ which allows to readjust according to the number of observations x^n. Moreover, from the penalty of Φ_β, it is possible to calculate that of the criterion BIC or AIC by assigning the following values to: β: $\beta_{\text{BIC}} = \frac{\log\log N - \log\log\log N}{\log N}$ and $\beta_{\text{AIC}} = \frac{\log 2 - \log\log\log N}{\log N}$. These results are true regardless of the application area CIs (histograms or selection of models).

Table 1. The criteria used and their penalty

Criterion	Penalty *C(n)*
AIC	$C(n) = 2$
BIC	$C(n) = \log n$
Φ_β	$C(n) = n^\beta \log \log n$

2.2. Law recognition procedures

2.2.1. Presentation of the laws used

Let us give ourselves a family of parameterized probability densities. As indicated in the introduction, we consider the family made up of the laws of Rayleigh, Weibull and Nakagami, with probability densities respectively given by:

$$f_R(x) = \frac{x}{\sigma^2}\exp(-\frac{x^2}{2\sigma^2})$$

$$f_W(x) = (\frac{k_W}{\lambda})\left(\frac{x}{\lambda}\right)^{(k_W-1)} \exp[-(\frac{x}{\lambda})^{k_W}]$$

$$f_N(x) = (\frac{2\mu^\mu x^{2\mu-1}}{\Gamma(\mu)\Omega^\mu})\exp[-\frac{\mu}{\Omega}x^2]$$

Where $\sigma, k_w, \lambda, \mu$ and Ω are the shape parameters of these laws. We notice that for $k = 2$ or $\mu = 1$, the laws of Weibull or Nakagami correspond to that of Rayleigh.

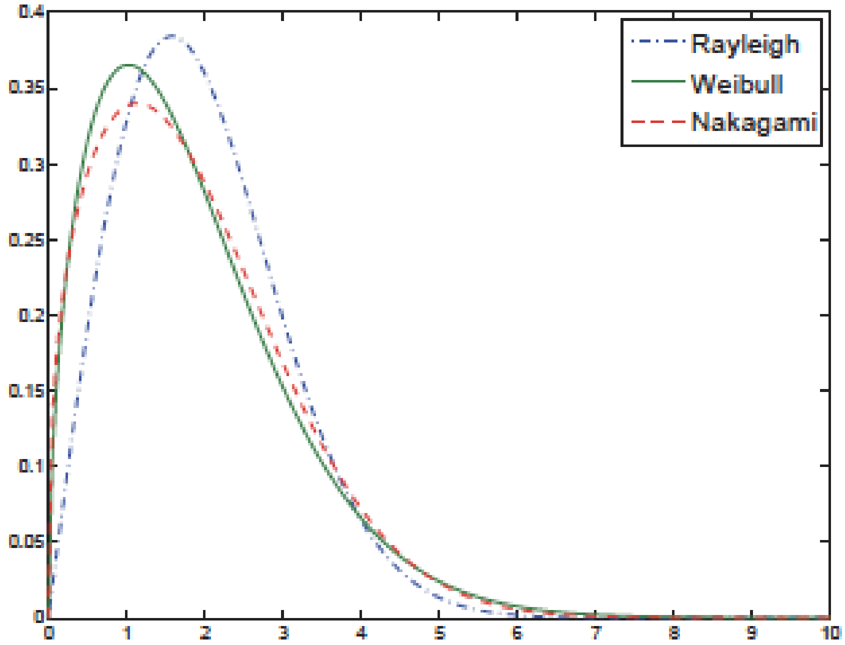
Note that Rayleigh's law is used when the multipaths constituting the propagation channel are approximately of the same amplitude. Also, the laws of Weibull and Nakagami are more general than the previous one and include, in addition to the previous case, the one or one of the paths can be predominant in amplitude compared to the others. We have not chosen to use the Rice distribution, generally considered in the case of a predominant journey, from the remark of (Dholer, 2003) who emphasizes that this model is not appropriate in our application context.

We will be interested in the cases where the transmitters and receivers are in position of LOS visibility or NLOS non-visibility. The NLOS case is the most delicate. Indeed, a simulation of the propagation of radio waves in NLOS configuration shows that the candidate probability densities (see Fig. 1) are closer visually than in the LOS configuration. Note that these two configurations (LOS and NLOS) summarize all the cases encountered where the distributions have densities of likelihood similar or not.

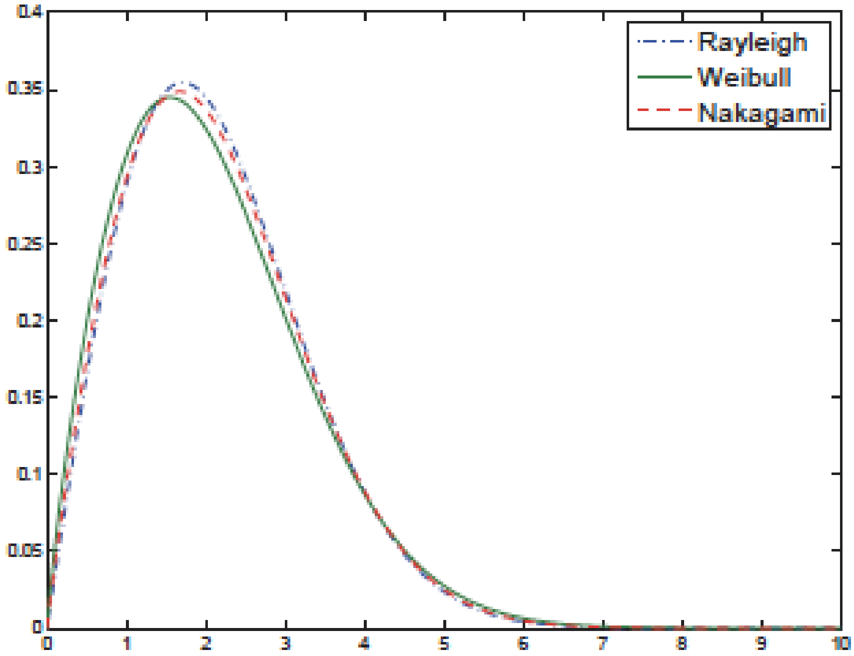
The parameters of the three laws of Rayleigh, Nakagami and Weibull, given in the Table 2, were therefore learned from simulations. We propose, in a first time, from the observation x^n of one of the 3 laws defined by the parameters of the Table 2, to statistically validate the proposed selection methods, according to the two supervised and unsupervised contexts.

Table 2. Parameters of the different laws learned from simulated samples

Law and parameter (s)	The case LOS	The case NLOS
Rayleigh (σ)	1.579	1.711
Weibull (k, λ)	1.529, 2.050	1.845, 2.367
Nakagami (μ, Ω)	0.675, 4.985	0.956, 5.854



(a) LOS



(b) NLOS

Fig. (1) Distribution of the laws of Rayleigh, Weibull and Nakagami, depending on the case LOS (a) and NLOS (b)

2.2.2. Supervised case

In this context, we want to know what the most appropriate density among the three known densities is (see Table 2 for the LOS and NLOS cases). So, there is no parameter estimation from x^n. Therefore, an exploitative approach the notion of maximum likelihood cannot be used because the parameters learned are not a priori those who will maximize the likelihood of the news observation. To compare two processes *X* and *Y*, it is common to use the distance of Kolmogorov-Smirnov (KS) between *FX* and *FY* distribution functions:

$$KS(X,Y) = \sup\{|F_X(x) - F_Y(x)|\} \quad (5)$$

Although easy to calculate and widely used in the application context considered, its low discriminating power is known. Divergences are classic tools for measuring the distance between densities of probability. We choose the Kullback-Leibler (KL) divergence, which is at the origin, for example, of obtaining the AIC criterion:

$$KL(X,Y) = \frac{1}{2}\int (f_X - f_Y)\log\left(\frac{f_X}{f_Y}\right)dx \quad (6)$$

However, while it is almost immediate to calculate *KS* from the observation xn of the process *X*, we need to estimate a probability density to use *KL*. For that, we will take either the kernel method presented in, or the IC-optimized histogram method presented in (2.1.2). In the supervised context, the recognition algorithm therefore operates from the as follows, in the case of using the Kolmogorov-Smirnov distance eq. (5):

1. From an observation of size n, xn, calculate the three distances of Kolmogorov-Smirnov between the empirical distribution function and the three functions distribution of the laws whose parameters are given in Table 2.
2. Assign the law that returns the smallest distance to the observation. If an approximation of the probability density is used (see section 2.1), the Kolmogorov-Smirnov distance is replaced by the Kullback-Leibler divergence eq. (6), which then measures the difference between the probability density estimated by the nucleus, or using an optimal histogram, and the density of one of the three laws whose parameters are provided in Table 2.

2.2.3. Unsupervised case

In this context, we give ourselves just as a priori the set of densities probability with which we want to approximate the density. For observation x^n, the parameters associated with the three densities are estimated in the sense of the maximum of likelihood. As before, we use the distance *KS* between distribution functions as well as the *KL* divergence between probability densities estimated by the *GCE* method or by histogram optimized by *IC*. The recognition algorithm then works as follows, in the case of using the Kolmogorov-Smirnov distance (5):

1. Estimate in the sense of maximum likelihood the parameters of the three starting from an observation of size *n*, x^n;
2. Compute the three Kolmogorov-Smirnov distances between the distribution function empirical and the three distribution functions of the laws whose parameters have been estimated in the previous step.
3. Assign the law that returns the smallest distance to the observation. If an approximation of the probability density is used, the Kolmogorov-Smirnov distance is replaced by the Kullback-Leibler divergence eq. (6), which then measures the difference between the probability density estimated by the nucleus, or using an optimal histogram, and the density of one of the three laws whose parameters were estimated in step 1 of the algorithm.

In addition, to improve the discrimination of densities having a different number of parameters (here, Rayleigh's law has one parameter while the two others have two), we also propose to use the CIs but in the context selection of models. Thus, if a density *fm* with m free parameters is one of these models and if $\hat{f}_m$ denotes this probability density configured with the parameters maximizing the likelihood, the IC criteria are written:

$$IC(f_m) = -2\log\left(\hat{f}_m(x^n)\right) + mC(n) \quad (7)$$

The recognition algorithm then works as follows:
1. Estimate in the sense of maximum likelihood the parameters of the three starting from an observation of size n, x^n;
2. Calculate the three information criteria associated with the three laws whose parameters were estimated in the previous step.
3. Assign the law that returns the lowest information criterion to the observation.

III. EXPERIMENTAL RESULTS

During our experiments, we studied the influence of the number n of samples on the recognition results by varying n from 100 to 10,000 show the results for the values 100, 500, 1000, 5000 and 10000 samples. In addition to the good recognition rates of each law, we give a parameter confidence statistic, the PPV, for positive predictive value, equal to the probability of having been generated by a law when it is recognized as that law. The result in recognition rate will be even more credible as the VPP is close to 1.

Other statistics have been used in (Coq, Alata, Olivier et al., 2009) to judge the quality of the methods used such as sensitivity and specificity factors, precision as well as the negative VP. In addition, we have shown that in some case, the optimal histogram method was significantly the most efficient in terms of disjunction of the confidence intervals that the method exploiting the distance of Kolmogorov-Smirnov in particular for the laws of Weibull and Nakagami. We refer the reader to the article cited above. We have not provided here of study on confidence intervals but average results from 450 experiments practiced for each law to be identified.

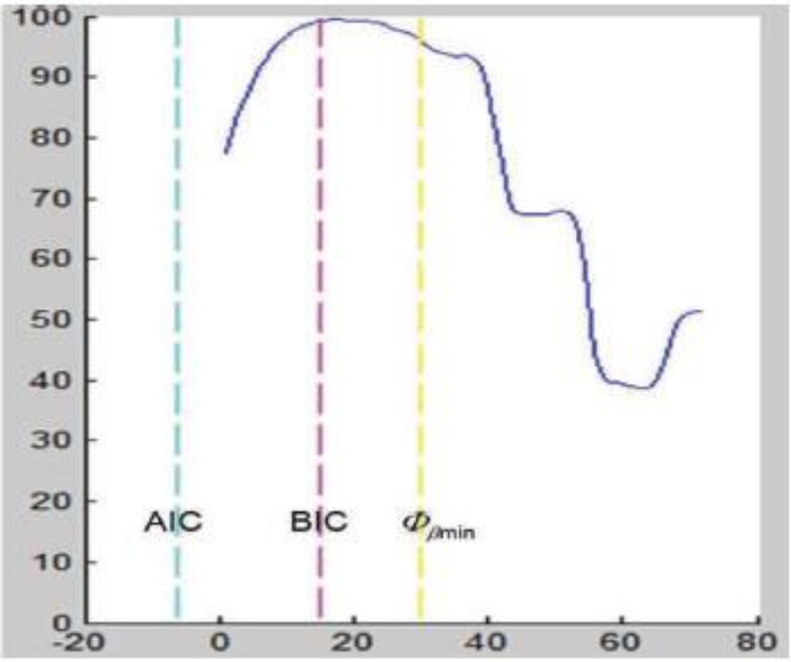
Thus, if a is the number of times a law considered L has been well recognized (rate of good recognition a/450) and if b is the number of times L is recognized then that the samples come from another law, then VPP = a/(a + b). We note, in tables 3, 4 and 5, “Histo” and “Kernel” for the methods using the approximation of probability densities by histogram or by method kernel and the distance KL, “KS” for the one using the distance KS on the function’s distribution. In tables 4 and 5, we add “IC” for the selection method by IC (7). We will denote by *R, W, N,* the three laws of Rayleigh, Weibull and Nakagami, respectively.

3.1. Supervised case

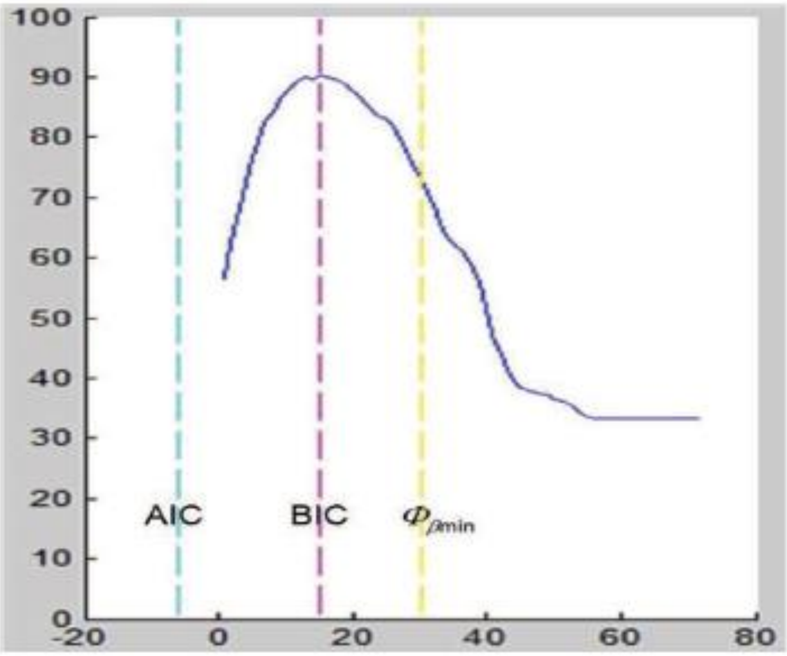
In this section we give average results of good recognition rate on samples for which the law from which they are derived is known (case supervised) to compare the performance of the different methods in LOS and NLOS (Table 3). We give the results for the “Histo” method with the criterion Φ_β giving the best performance and in fact corresponding to the BIC criterion in LOS and NLOS. Figure 2 justifies this choice by Φ_β: it shows the criterion Φ_β and $\beta_{\min}$. for min (yellow), BIC (magenta) and AIC (green). We can observe that BIC gives practically the optimal rate in the two cases presented.

Following (on the abscissa) which varies from 0 to 0.7 in 75 values. The results with AIC, BIC and min are shown in green, magenta, and yellow respectively. For n> 500 and the LOS case, the “Histo” method gives superior results to the other two methods, with a PPV almost always close to one. For little one’s samples (n = 100), no method is relevant, but in this borderline case, it may be preferable to take the classic “KS” method, the “Histo” method being based on asymptotic criteria, hence its very bad behavior. n = 500, the methods are equivalent in LOS, with the Weibull law with difficulty recognizable.

In NLOS, for n <1000, the rates are generally insufficient and therefore unusable for n up to 1000 samples. The laws are indeed too close (figure 1) to be identified when the number of observations is insufficient. Let’s remember that the “Kernel” and “Histo” methods are asymptotic methods. Nevertheless, in the case of small samples, there are other information criteria such as (Hurvich, Tsai, 1989; Broersen, 2000), but to our knowledge not justifiable in the case histograms. To summarize, in view of the average results, it seems that the histogram method overall gives the best results, the best deviations being obtained for the NLOS case with n 1000. However, the results presented do not allow to prove statistically speaking a superiority of this method over the others.



(a) LOS



(b) NLOS

Fig. 2. Evolution of the recognition rate (ordinate) for LOS cases (a) and NLOS (b).

Table 3. Comparison of supervised methods. The first number gives the good recognition rate (in %), the second in brackets gives the PPV.

		LOS			NLOS		
		KS	**Kernel**	**Histo**	**KS**	**Kernel**	**Histo**
100	**R**	81.5(0.80)	35.3(0.98)	51.3(0.65)	46.6(0.43)	11.1(0.62)	40.4(0.43)
	W	59.3(0.54)	44.7(0.67)	52.9(0.46)	64.6(0.43)	95.3(0.40)	67.1(0.39)
	N	45.6(0.51)	82.0(0.41)	45.3(0.43)	16.6(0.38)	15.3(0.40)	8.9(0.37)
	Avg.	62.2	54.0	49.8	42.7	40.6	38.1
500	**R**	99.3(0.99)	96.7(1)	98.2(0.99)	54.7(0.55)	20.2(0.85)	40.9(0.66)
	W	67.8(0.67)	52.2(0.9)	55.8(0.84)	71.1(0.57)	96.7(0.52)	83.8(0.56)
	N	65.3(0.67)	94.4(0.65)	89.5(0.67)	27.8(0.37)	36.2(0.4)	36.2(0.41)
	Avg.	77.4	81.1	81.2	51.1	51	53.6
1000	**R**	100(1)	100(1)	100(1)	59.6(0.61)	22.4(0.89)	56.4(0.66)
	W	77.6(0.75)	59.6(0.99)	76.2(0.92)	79.8(0.7)	97.3(0.63)	83.8(0.73)
	N	74.4(0.77)	99.1(0.71)	93.3(0.8)	38.4(0.44)	54.4(0.45)	50.2(0.5)
	Avg.	84	86.2	89.8	59.2	58.1	63.5
5000	**R**	100(1)	100(1)	100(1)	78.7(0.77)	45.1(0.97)	81.3(0.91)
	W	96.0(0.95)	91.3(1)	98.7(1)	95.8(0.98)	100(0.95)	98.9(0.98)
	N	95.3(0.96)	100(0.92)	100(0.99)	74.4(0.74)	93.1(0.63)	90.2(0.82)
	Avg.	97.1	97.1	99.5	83	79.4	90.1
10000	**R**	100(1)	100(1)	100(1)	85.1(0.85)	62.7(1)	95.5(0.95)
	W	99.1(0.99)	98.5(1)	100(1)	99.1(1)	100(1)	99.5(1)
	N	98.7(0.99)	100(0.98)	100(1)	84.4(0.84)	99.5(0.73)	94.9(0.95)
	Avg.	99.2	99.5	100	89.5	87.4	96.7

3.2. Unsupervised case

In this case, the samples are taken from the three laws, but, as the parameters are not known, they are estimated in the sense of maximum likelihood from the xn observation. We can introduce here the model selection method called "IC" eq. (7) because it can be used in this context, which was not consistent in the context supervised. The two IC-based methods, ie “Histo” and “IC”, are still given with the criterion Φ_β producing the optimal rates, or with Φ_β close to the BIC criterion.

Table 4. Comparison of unsupervised methods for the LOS case. The first number gives the rate of good recognition (in %), the second in brackets gives the VPP.

		LOS			
		KS	**Kernel**	**Histo**	**IC**
100	**R**	21.5(0.88)	43.8(0.99)	44.0(0.99)	92.7(0.88)
	W	77.1(0.43)	39.8(0.54)	40.4(0.53)	62.9(0.64)
	N	47.3(0.49)	86.2(0.47)	81.8(0.46)	61.3(0.63)
	Avg.	48.7	40.6	55.4	72.3
500	**R**	26.0(1)	44.7(1)	34.7(0.99)	100(1)
	W	86.2(0.5)	43.1(0.64)	54.4(0.58)	82.9(0.84)
	N	64.2(0.64)	97.3(0.52)	90.0(0.52)	84.0(0.83)
	Avg.	58.8	61.7	59.7	88.9
1000	**R**	23.6(1)	47.8(1)	38.4(1)	100(1)
	W	92.0(0.55)	53.3(0.72)	62.4(0.68)	90.2(0.92)
	N	78.7(0.71)	99.3(0.56)	97.6(0.58)	91.8(0.9)
	Avg.	64.74	66.81	66.14	94
5000	**R**	22.0(1)	48.7(1)	41.3(1)	100(1)
	W	99.1(0.65)	87.1(0.83)	94.9(0.79)	100(0.99)
	N	96.2(0.77)	100(0.68)	100(0.72)	99.6(1)
	Avg.	72.4	78.6	78.7	99.8
10000	**R**	21.6(1)	49.6(1)	48.9(1)	100(1)
	W	100(0.66)	97.3(0.85)	98.9(0.83)	100(1)
	N	99.1(0.78)	100(0.73)	100(0.76)	100(1)
	Avg.	73.6	82.3	82.6	100

Table 5. Comparison of unsupervised methods for the NLOS case. The first number gives the rate of good recognition (in %), the second between parentheses gives the VPP.

		NLOS			
		KS	**Kernel**	**Histo**	**IC**
100	**R**	24.4(0.40)	50.0(0.49)	49.5(0.49)	74.6(0.36)
	W	56.4(0.36)	35.6(0.44)	34.0(0.46)	24.0(0.50)
	N	24.7(0.30)	36.9(0.31)	41.3(0.32)	13.3(0.30)
	Avg.	35.2	40.8	41.6	37.3
500	**R**	27.3(0.56)	50.7(0.75)	45.3(0.74)	83.6(0.49)
	W	64.9(0.4)	34.4(0.45)	40.4(0.44)	48.7(0.68)
	N	32.0(0.35)	57.8(0.37)	54.7(0.37)	16.7(0.29)
	Avg.	41.4	47.6	46.8	49.6
1000	**R**	22.9(0.57)	43.1(0.75)	45.6(0.75)	84.7(0.56)
	W	67.6(0.4)	33.1(0.43)	36.2(0.44)	61.3(0.71)
	N	33.3(0.38)	62.7(0.38)	59.8(0.38)	20.4(0.33)
	Avg.	41.2	46.3	47.2	55.5
5000	**R**	25.6(0.86)	48.9(0.97)	39.3(0.97)	90.9(0.81)
	W	79.1(0.45)	34.0(0.54)	56.9(0.54)	82.2(0.72)
	N	47.6(0.51)	87.3(0.54)	78.2(0.54)	49.1(0.68)
	Avg.	50.7	56.7	58.1	74.1
10000	**R**	24.2(0.95)	45.8(1)	45.1(0.99)	95.8(0.93)
	W	87.8(0.49)	38.7(0.59)	60.4(0.62)	91.1(0.74)
	N	56.7(0.59)	94.4(0.50)	87.5(0.56)	64.4(0.86)
	Avg.	56.2	59.6	64.4	83.7

We give in tables 4 and 5, the rates of good recognition and the PPV, according to the number n of samples. We see that the model selection method "IC" gives the best results in all cases, followed at a distance by the other three methods in the simplest case LOS (Table 4). As in the supervised case, and for n = 100, 500, 1000, these three methods give comparable but insufficient results, both in terms of recognition rate in terms of credibility (the VPP) with an advantage for the asymptotic method "Histo" for n> 1 000.

For the NLOS case (Table 5), if we exclude the borderline case of very few samples (n = 100) where the results lack credibility with the four methods, the "IC" method is also better in terms of recognition rate, but the VPP are still too low to allow a conclusion on the nature of the law for n up to 1000 samples. For n = 5,000 samples, we can judge that only the results by the "IC" method are credible in terms of PPV, which is confirmed also for 10,000 samples.

IV. CONCLUSION

We have shown in this article the interest, in a supervised context like in an unsupervised context, the use of information criteria (IC) as tools that can intervene in the recognition of probability laws, on the one hand for the construction of optimal histograms, and on the other hand, for model selection. For this, we relied on a particular and difficult application context recognition of laws in digital communication, with or without visibility between transmitter and receiver (LOS and NLOS case), where knowledge of the law of propagation is crucial if we want to ensure a guaranteed quality of service at the reception. We were able to observe the good performance of the two IC-based methods compared to a high-performance kernel method, called GCE, and the usual method used in digital propagation. The results presented are not limited to communications digital and are likely to be of interest to people working in numerous application contexts in signal and image as soon as it is necessary to call on the description of the distribution of sample values.

These results globally show the interest of the use of IC for the law recognition in the supervised case, by an optimal histogram, and in the unsupervised case, by minimizing an IC. These results may interest persons working in numerous application contexts in signal and in image domains as soon as it is necessary to use an approximation of the distribution of the samples. The problem of the selection of a statistical distribution studied in this paper is linked to the wireless transmission context. In such a context, the electromagnetic wave undergoes multiple interactions like reflection, diffraction, and their combinations on obstacles such as buildings, vehicles, etc. Thus, we speak about the "multi-path phenomenon" between transmitter and receiver antennae. It induces that the amplitude of the received signal, resulting from the sum and from the complexity of the paths, has undergone small scale fading in the spatio-temporal domain.

REFERENCES

[1]. Akaike H. (1974). A new look at the statistical model identification. IEEE Trans. on Automatic Control, vol. 19, no 6, p. 716-723.
[2]. Alata O., Olivier C. (2003). Choice of a 2d causal ar texture model using information criteria. Pattern Recognition Letters, vol. 24, no 9-10, p. 1191-1201.
[3]. Babich F., Lombardi G. (2000). Statistical analysis and characterization of the indoor propagation channel. IEEE Trans. on Communications, vol. 48, no 3, p. 455-464.
[4]. Barbot J. P., Levy A. J., Bic J. C. (1992, June). Estimation of fast fading distribution functions. In proc. of the ursi commission f symposium. Ravenscar, (U.K.).
[5]. Basseville M. (1989). Distance measures for signal processing and pattern recognition. Signal Processing, vol. 18, no 4, p. 349-369.
[6]. Botev Z. I. (2007). A novel nonparametric density estimator. Technical Report. University of Queensland, Brisbane (Australia).
[7]. Broersen P.M. T. (2000, December). Finite Sample Criteria for Autoregressive Order Selection. IEEE Trans. on Signal Processing, vol. 48, no 12, p. 3550-3558.
[8]. Bultitude R. J. C., Hahn R. F., Davies R. J. (1998). Propagation considerations for the design of an indoor broad-band communications system at ehf. IEEE Trans. on Vehicular Technology, vol. 47, no 1, p. 235-245.
[9]. Coq G. (2008). Utilisation d'approches probabilistes basées sur les critères entropiques pour la recherche d'information sur support multimédia. Thèse de doctorat, Université de Poitiers (France).
[10]. Coq G., Alata O., Olivier C., Pousset Y., Li X. (2009, September). Reconnaissance de loi via une estimation de densités par histogramme pour lamodélisation de canaux de transmission. In Actes de la conférence gretsi. Dijon (France).
[11]. Coq G., Alata O., Pousset Y., Li X., Olivier C. (2009, April). Law recognition via histogrambased estimation. In in proc. ieee iccasp, p. 3425-3428. Taïpei (Taïwan).
[12]. Coq G., Olivier C., Alata O., Arnaudon M. (2007, Sept.). Information criteria and arithmetic codings: an illustration on raw images. In in proc. 15th eusipco, p. 634-638. Poznan (Poland).
[13]. Dholer M. (2003). Virtual antenna arrays. PhD, King's College London, London (UK).
[14]. El Matouat A., Hallin M. (1996). Order selection, stochastic complexity and Kullback-Leibler information. In Athens conference on applied probability and time series analysis, vol. ii (1995), vol. 115, p. 291–299. New York, Springer.
[15]. Fryziej M. (2001). Caractérisation large bande du canal radio intra-bâtiment à 60 ghz. Thèse de doctorat, CNAM (Paris, France).
[16]. Goldsmith A. (2005). Wireless communications. Cambridge University Press.
[17]. Grünwald P. (2005). A tutorial introduction to the minimum description length principle. In
[18]. Advances in minimum description length: Theory and applications. MIT Press.
[19]. Hannan E., Quinn B. (1979). The determination of the order of an autoregression. Journal of the Royal Statistic Society, vol. 41, no 2, p. 190-195.
[20]. Hashemi H., Tholl D. (1994). Statistical modeling and simulation of the rms delay spread of indoor radio propagation channels. IEEE Transactions on Vehicular Technology, vol. 43, no 1, p. 110-120.
[21]. Hurvich C.M., Tsai C.-L. (1989). Regression and time series model selection in small samples.
[22]. Biometrika, vol. 76, no 2, p. 297-307.
[23]. Iskander M., Yun Z. (2002). Propagation prediction models for wireless communication systems. IEEE Trans. on Microwave Theory and Techniques, vol. 50, no 3, p. 662-673.
[24]. Jouzel F., Olivier C., El Matouat A. (1998, September). Information criteria based edge detection.
[25]. In EUSIPCO-signal processing IX, Rhodes (Greece), vol. 2, p. 997-1000. Elsevier.
[26]. Olivier C., Alata O. (2009, July). The information criteria: examples of applications in image and signal processing. In Optimisation in image and signal processing, desip series, p. 105- 136. Wiley & ISTE Ltd ed.
[27]. Olivier C., Jouzel F., Matouat A. E. (1999, May). Choice of the Number of Component Clusters in Mixture Models by Information Criteria. In Proc. iapr vision interface, p. 74-81.
[28]. Proakis J. (1995). Digital communications, 3th edition. MacGraw-Hill, New-York.
[29]. Rappaport T. (1996). Wireless communications: Principle and practice. Englewood Cliffs, NJ: Prentice-Hall.
[30]. Rissanen J. (1976). Generalized kraft inequality and arithmetic coding. IBM Journal of Research and Development, vol. 20, no 3.
[31]. Rissanen J. (1978). Modeling by the shortest data description. Automatica, vol. 14, p. 465-471.
[32]. Rissanen J. (1989). Stochastic complexity in statistical inquiry (vol. 15). Teaneck, NJ, World Scientific Publishing Co. Inc.
[33]. Rissanen J., Speed T. P., Yu B. (1992). Density estimation by stochastic complexity. IEEE
[34]. Trans. on Information Theory, vol. 38, no 2, p. 315-323. Santamaria I., Erdogmus D., Principe J. (2002). Entropy minimization for supervised digital communications channel equalization. IEEE Trans. on Signal Processing, vol. 50, no 5, p. 1184-1192.
[35]. Sarkar T. K., Zhong J., Kyungjung K., Medouri A., Salazar-Palma M. (2003). A survey of various propagation models for mobile communication. IEEE Trans. on Antennas and Propagation Magazine, vol. 45, no 3, p. 51-82.
[36]. Schwarz G. (1978). Estimating the dimension of a model. Ann. Statist., vol. 6, no 2, p. 461–464.
[37]. Seitadji B., Levy A. J. (1994). A statistical model for the simulation of time-varying multipath mobile radio propagation channel. In In proc. ieee int. conf. on acoustics, speech and signal processing (ICASSP), vol. 6, p. 149–152.
[38]. Shibata R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika, vol. 63, no 1, p. 117–126.
[39]. Zayen B., Hayar A. (2011). A performance study of kullback-leibler distance-based spectrum sensing algorithm. In In proc. 3rd ieee international conference on ultra-modern telecommunications (icumt), p. 1–5.

Ahmed Alzahmi. "Supervised and Unsupervised Recognition of Laws from Finite Sample."*American Journal of Engineering Research (AJER)*, vol. 10(10), 2021, pp. 85-94.