LAW RECOGNITION VIA HISTOGRAM-BASED ESTIMATION

G. Coq

Laboratoire de Mathématiques et Applications
Université de Poitiers
TELEPORT 2 - BP 30179, 86962 Chasseneuil, France
Phone : (+33) 5 49 49 68 97 Fax : (+33) 5 49 49 69 01
email : coq@math.univ-poitiers.fr

X. Li, O. Alata, Y. Pousset, C. Olivier

Laboratoire Signal, Images et Communications
Université de Poitiers
TELEPORT 2 - BP 30179, 86962 Chasseneuil, France
Phone : (+33) 5 49 49 65 67 Fax : (+33) 5 49 49 65 70
email : alata,pousset,li,olivier@sic.sp2mi.univ-poitiers.fr

ABSTRACT

In this paper, we study the problem of recognizing an unknown probability density function from one of its sample which is of interest in signal and image processing or telecommunication applications. By opposition with the classical Kolmogorov-Smirnov method based on empirical cumulative functions, we consider histogram estimators of the density itself built from our data. Those histograms are generated via model selection, more specifically via a codelength-based Information Criterion. From the histograms, we may compute a Kullback-Leibler distance to any theoretical law which is used to complete the recognition. We apply this histogram-based method for law recognition in a theoretical setup where the true density is known as well as in a real setup where data come from radio channel propagation experimentation.

Index Terms— HF radio propagation, probabilty, information criteria, histograms, law recognition

1. INTRODUCTION

Law recognition is a problem of great interest in various domains such as image processing, shape recognition or telecommunication applications. The most widely used tool to solve this problem is the Kolmogorov-Smirnov (KS) test that will be more precisely presented in the sequel. This test is based on a data-based estimation of the cumulative function of our unknown density. Here, we choose to estimate the density itself by a histogram. This estimation of the density will allow computation of Kullback-Leibler type distance on which the law recognition will be based.

In order to estimate the unknown density by a histogram, one needs to use non-parametrical model selection. More precisely Information Criteria (IC), also called penalized likelihood criteria, will be used. Birgé [1] and Birgé and Al. [2] address this problem in recent works. The authors suggest to use an IC in order to determine from the data which histogram is the most suitable for the estimation. The justification of the use of such a criterion is based on the minimization of the risk of the resulting estimation. From another point of view, Rissanen develops in [3] the notion of Minimum Description Length (MDL) and in [4], that of stochastic complexity, strongly related to the theory of coding as expressed in [5]. From those notions, one may, as in [6], construct an IC suitable for our present concern of histogram selection. The use of that latter criterion is thus justified via coding arguments rather than risk-minimizing arguments as were former criteria from Birgé and Al. [1, 2].

In [5], we worked in the continuity of Rissanen and Al. [6] by developping a two-steps coding technique of our data from which is derived a codelength-based Information Criterion. Both notions and their connections are described in section 2. This criterion allows to select, from the data, a histogram estimating the unknown density. In section 3, we present the method of law recognition based on the previous histogram and on Kullback-Leibler distance. We also recall the classical Kolmogorov-Smirnov method of law recognition. Part 4 and 5 are dedicated to applications in theoretical and experimental backgrounds.

2. THE HISTOGRAM SELECTION CRITERION

The main setup is as follows : f is an unknown density defined on an interval $I = [a, b]$ of $\mathbb{R}$ and $x^n = x_1, \ldots, x_n$ is a sample from f. Given $P = (I_j)_{j=1,\ldots,m}$ a partition of I into m intervals, one constructs a histogram estimator of f by

$$\widehat{f}_P = \sum_{j=1}^{m} \frac{n_j}{nL_j} \mathbb{1}_{I_j},$$

where $\mathbb{1}_X$ denotes the indicator function of a set X, n_j the number of data x_i falling into I_j and L_j the length of I_j. The main problem of histogram selection is to determine which partition P is to be chosen in order to estimate f by $\widehat{f}_P$.

For the sake of simplicity, we choose $r > 0$ a precision, usually small, and denote by $P_{\max}$ the partition of I consisting of R intervals all with length r. Then we restrict ourselves to the 2^{R-1} partitions which intervals are unions of adjacent intervals of $P_{\max}$. Those are called sub-partitions of $P_{\max}$ and their set is denoted by $\mathcal{SP}$. Note that this restriction allows to handle the case where data live in a discrete space.

Previous work [5] allows to design an IC derived from a data-coding technique answering the partition selection problem. Here, we present the coding technique and the IC resulting in our setting.

2.1. Two steps coding

The main idea is to choose a partition $P \in \mathcal{SP}$ and, with help of it, to encode our data x^n. We consider here the natural idea of coding: transforming our data x^n in a sequence of bits that is decodable if encoder and decoder agree. Then, via the principle of Minimum Description Length [7], the partition to be chosen is the one that realizes the best encoding of our datas. That encoding is now described ; it is lossless up to the precision r and presents two steps.

2.1.1. First step : arithmetic coding

In the first step, data x^n are transformed into y^n as follows

$$y_i = \sum_{j=1}^{m} j.\mathbb{1}_{\{x_i \in I_j\}},\ i = 1, \dots, n.$$

In other words, y_i denotes the number of the interval of P in which x_i falls. In order to encode y^n, we use a version of the arithmetic coding technique presented in [5] as Predictive Adaptive Arithmetic Coding of order 0, 0-PAAC for short. In the sequel, $L(y^n|P)$ denotes the length of the binary string resulting from such a coding. As also discussed in the previous reference, the PAAC is strongly related to the work of Rissanen [4] in the sense that $L(y^n|P)$ is asymptotically estimated by the so-called stochastic complexity of y^n:

$$L(y^n|P) \approx -\sum_{j=1}^{n} n_j \log \frac{n_j}{n} + \frac{m-1}{2} \log(n). \qquad (1)$$

As m grows, that quantity tends to grow as well since encoding symbols y_i, that may take m different values, gets harder.

2.1.2. Second step : fixed length coding

For a fixed $i = 1, \dots, n$, the information $y_i = j$ alone does not allow to recover x_i. In order to do this, one needs to precise where x_i is located inside the interval I_j. Up to the precision r, there are L_j/r real numbers in this interval. Consequently, the precision of x_i may be done with an encoding of (ideal) fixed length equal to $\log L_j/r$.

Now, the total number of bits required to precise all the x_i's equals

$$L(x^n|y^n) = \sum_{j=1}^{m} n_j \log \frac{L_j}{r}. \qquad (2)$$

By opposition to (1), this quantity tends to decreases as m increases. Indeed, the larger m, the smaller the intervals of P, the easier it is to precise where each x_i is.

2.1.3. The criterion

Via our two-steps encoding method, the estimated total lossless codelenght of the data x^n with help of the partition P writes as the sum of (1) and (2), that is

$$\text{CRIT}(x^n, P) = -\sum_{j=1}^{n} n_j \log \frac{rn_j}{nL_j} + \frac{m-1}{2} \log(n). \qquad (3)$$

This quantity enters the formalism of Information Criteria (IC), widely used tools in model selection problem for which one may for instance refer to [8, 9, 2].

The MDL principle thus suggests to choose $\widehat{P}$ as

$$\widehat{P} = \text{Argmin}\, \{\text{CRIT}(x^n, P),\ P \in \mathcal{SP}\} \qquad (4)$$

and consider $\widehat{f}_{\widehat{P}}$ as an estimator of the unknown density f. Note that this minimization does not depend on the chosen precision r.

The opposite behaviors of (1) and (2) described earlier reflect the usual fact that this minimization of the IC (3) realizes the best compromise between the complexity of the partition and how well it fits the data.

The resulting histogram is referred to in the sequel as dynamic histogram. This word is inherited from the dynamic programming method introduced by Rissanen [6] that allows to determine $\widehat{P}$ in (4) in a number of operations of the order R^2 instead of having to compute the 2^{R-1} values of the criterion for each $P \in \mathcal{SP}$. We may also restrict ourselves to the class of regular histograms on I. Those are built on partitions P_m that have m intervals all of length $(b-a)/m$ for $m = 1, \dots, M$. In this case, the resulting estimation is referred to as regular histogram. We use the term optimal histograms to refer to either dynamic or regular ones, by opposition to the empirical histogram described in part 3.2.

3. METHODS

Let $\mathcal{F}$ be a family of density functions on I ; they are the laws in competition.

3.1. Optimal histogram method

Once the optimal histogram estimator $\widehat{f}$ is selected via (3) and (4), we may compute the Kullback-Leibler distance from it to any $f \in \mathcal{F}$ as follows :

$$\text{KL}(\widehat{f}, f) = \frac{1}{2} \int_I (\widehat{f} - f) \log \frac{\widehat{f}}{f} d\mu. \qquad (5)$$

Then, the law recognition is done via

$$f_{\text{KL},opt} = \text{Argmin}(\text{KL}(\widehat{f}, f),\ f \in \mathcal{F}). \qquad (6)$$

3.2. Empirical histogram method

From our set of data x^n, it is usual to build an empirical histogram describing the distribution. Classically this histogram $\widehat{f}_{emp}$ is built on the regular partition of I that counts $\lfloor 2\sqrt{n}-1 \rfloor$ intervals where $\lfloor . \rfloor$ denotes the inferior integer part. From it, one may solve the law recognition problem by

$$f_{\mathrm{KL},emp} = \mathrm{Argmin}(\mathrm{KL}(\widehat{f}_{emp}, f),\ f \in \mathcal{F}). \qquad (7)$$

3.3. Kolmogorov-Smirnov method

The Kolmogorov-Smirnov (KS) method is the tool used classically for law recognition. From the data x^n, we construct the empirical cumulative function denoted by $\widehat{F}$. Let us denote by $\mathcal{F}_c$ the set of cumulative functions of the laws $f \in \mathcal{F}$. The KS distance between $\widehat{F}$ and $F \in \mathcal{F}_c$ is

$$\mathrm{KS}(\widehat{F}, F) = \sup_{t \in I} \left| \widehat{F}(t) - F(t) \right|. \qquad (8)$$

From that distance, the law recognition is done via

$$F_{\mathrm{KS}} = \mathrm{Argmin}(\mathrm{KS}(\widehat{F}, F),\ F \in \mathcal{F}_c). \qquad (9)$$

This method does not require to estimate the density itself.

3.4. The model

It contains three densities defined for $x \geq 0$:

$$\begin{aligned} \text{Rayleigh} \quad &: \quad f_R(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \\ \text{Nakagami} \quad &: \quad f_N(x) = \frac{2\mu^\mu x^{2\mu-1}}{(\mu-1)!\Omega^\mu} \exp\left(-\frac{\mu x^2}{\Omega}\right) \qquad (10) \\ \text{Weibull} \quad &: \quad f_W(x) = \frac{k x^{k-1}}{\lambda^k} \exp\left(-\frac{x^k}{\lambda^k}\right) \end{aligned}$$

This choice of model is usually done for radio channel propagation modelization. All coefficients $\sigma, \mu, \Omega, k, \lambda$ are shape parameters to be described later. Note that choosing $\mu = 1$ and $k = 2$ in Nakagami and Weibull laws make them similar to a Rayleigh law.

4. APPLICATION IN A THEORETICAL SETUP

Our aim in this part is to show that the recognition method defined in (6) is efficient and to compare it with the usual KS method (9) and the empirical histogram method (7).

The shape parameters $\sigma, \mu, \Omega, k, \lambda$ in (10) are all set to obtain a mean of 73 and a standard deviation of 1.2, the resulting values are given in Table 1(a). We generate 30 samples of sizes n ranging from 100 to 3000 of the laws in model (10). On each of those sample, we apply the three recognition methods discussed earlier: optimal histograms (6), empirical histogram (7) and KS distance (9). Since the generating law is known, we may compute successfull recognition rates (RR) of the true family and plot them in figure 1.

We choose to show only RR of the Rayleigh law. Results are similar for other generating distributions. We see that using optimal histograms, especially dynamic one, yields a better RR than using the usual KS method. This is remarkable since, in our setting, optimal histograms never contain more than 50 bins. This means that resuming the data information to about 50 classes allows a better recognition of the underlying law than conserving all the information in the n steps of the empirical cumulative function used in the KS method. Moreover, if one wants to avoid KS method for law recognition, optimal histograms should be used since the empirical one gives poor results.

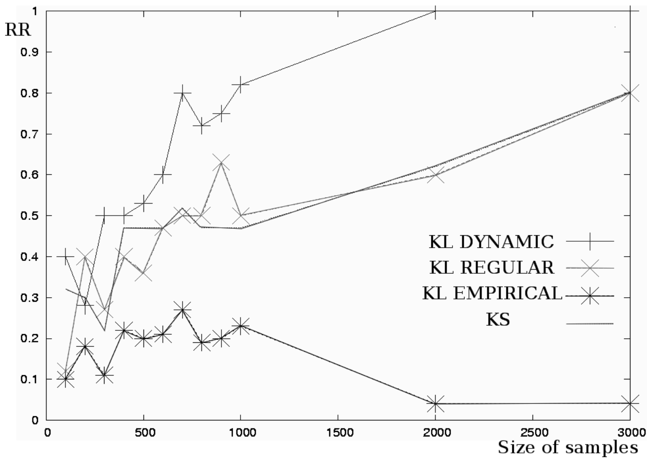

Fig. 1. Rayleigh RR using optimal histograms, empirical histogram and KS methods.

5. APPLICATION IN A REAL SETUP

Laboratory SIC-XLIM developped a software allowing to simulate the fast fading behavior of a radio propagation channel in various environments, see [10]. From this software, we collected $n = 700$ data representing the attenuation (dB) of the signal in both Line Of Sight (LOS) and Non Line Of Sight (NLOS) configurations, see figure 2. We choose to modelize the radio channel by either a Rayleigh, Weibull, or Nakagami distribution from the model (10). In order to determine which of those laws best suits the radio channel, we apply recognition method (6) with optimal histograms.

5.1. LOS configuration

In this experiment, the computed average attenuation and its standard deviation are respectively 73 dB and 1.2 dB. Shape parameters in (10) are set in consequence and shown in Table 1(a). Here, the Weibull distribution suits the best the fast

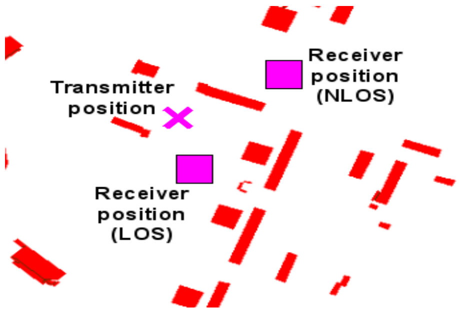

Fig. 2. The environment of the simulation.

fading behavior of the propagation channel in a LOS configuration. This is the usual conclusion as in [11].

5.2. NLOS configuration

Here, average and standard deviation equal respectively 99.8 dB and 3.5 dB. Shape parameters in (10) are set in consequence and shown in Table 1(b). It is not obvious to decide which law suits the best the data but it does not matter since shape parameters $\mu \approx 1$ and $k \approx 2$ for Nakagami and Weibull laws actually correspond to a Rayleigh law. As in [12], the Rayleigh modelization appears as the best in the NLOS configuration.

Law / Histo.	Rayleigh $\sigma = 1.58$	Nakagami $\mu = 0.68$ $\Omega = 4.99$	Weibull $k = 1.53$ $\lambda = 2.05$
Dynamic	0.12	0.05	0.04
Regular	0.12	0.08	0.06

(a)

Law / Histo.	Rayleigh $\sigma = 4.09$	Nakagami $\mu = 0.99$ $\Omega = 33.43$	Weibull $k = 2.01$ $\lambda = 5.79$
Dynamic	0.033	0.034	0.035
Regular	0.114	0.115	0.115

(b)

Table 1. KL distances from optimal histograms to the laws in competition in LOS(a) and NLOS(b) cases.

6. CONCLUSION

In this paper, we developped an information-theoretic criteria (3) allowing to estimate an unknown probability law by a histogram. This histogram, summing our data to a few number of parameters and used along with Kullback-Leibler distance, is shown to allow a rate of succesfull law recognition as good as or even better than the usual Kolmogorov-Smirnov method. It is then applied to a realistic environment where it matches usual results of modelization.

7. REFERENCES

[1] L. Birgé, "Statistical estimation with model selection," *Indag. Math. (N.S.)*, vol. 17, no. 4, pp. 497–537, 2006.

[2] L. Birgé and Y. Rozenholc, "How many bins should be put in a regular histogram," *ESAIM Probab. Stat.*, vol. 10, pp. 24–45 (electronic), 2006.

[3] J. Rissanen, "Modeling by the shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[4] J. Rissanen, "Stochastic complexity and modeling," *Ann. Statist.*, vol. 14, no. 3, pp. 1080–1100, 1986.

[5] G. Coq, C. Olivier, O. Alata, and M. Arnaudon, "Information criteria and arithmetic codings : an illustration on raw image.," in *15th European Signal Processing Conference proceedings*, pp. 634–638.

[6] J. Rissanen, T. P. Speed, and B. Yu, "Density estimation by stochastic complexity.," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 315–323, 1992.

[7] P. D. Grunwald, In Jae Myung, and M. A. Pitt, *Advances in Minimum Description Length: Theory and Applications (Neural Information Processing)*, The MIT Press, 2005.

[8] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pp. 267–281. Akadémiai Kiadó, Budapest, 1973.

[9] P. Hall, "Akaike's information criterion and Kullback-Leibler loss for histogram density estimation," *Probab. Theory Related Fields*, vol. 85, no. 4, pp. 449–467, 1990.

[10] F. Escarieu, Y. Pousset, R. Vauzelle, and L. Aveneau, "Outdoor and indoor channel characterization by a 3d simulation software.," Septembre 2001, PIMRC '2001, San diego, USA.

[11] N. C. Sagias and G. K. Karagiannidis, "Gaussian class multivariate Weibull distributions: Theory and applications in fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3608–3619, Oct 2005.

[12] T.K. Sarkar, Zhong Ji, Kyungjung Kim, A. Medouri, and M. Salazar-Palma, "A survey of various propagation models for mobile communication," *Antennas and Propagation Magazine, IEEE*, vol. 45, no. 3, pp. 51–82, 2003.