Continuous belief functions and α-stable distributions

Anthony Fiche, Arnaud Martin, Jean-Christophe Cexus, Ali Khenchaf

$E^3I^2 - EA3876$

ENSIETA

2 rue François Verny, 29806 Brest Cedex 9, France

anthony.fiche,arnaud.martin,jean-christophe.cexus,ali.khenchaf@ensieta.fr

Abstract − *The theory of belief functions has been formalized in continuous domain for pattern recognition. Some applications use assumption of Gaussian models. However, this assumption is reductive. Indeed, some data are not symmetric and present property of heavy tails. It is possible to solve these problems by using a class of distributions called α-stable distributions. Consequently, we present in this paper a way to calculate pignistic probabilities with plausibility functions where the knowledge of the sources of information is represented by symmetric α-stable distributions. To validate our approach, we compare our results in special case of Gaussian distributions with existing methods. To illustrate our work, we generate arbitrary distributions which represents speed of planes and take decisions. A comparison with a Bayesian approach is made to show the interest of the theory of belief functions.*

Keywords: Belief functions, pignistic probabilities, plausibility functions, symmetric α-stable distributions.

1 Introduction

Stable distributions have been developed by Paul Lévy in his study of normalized sums of independant and identically distributed terms [1]. It is a class of distributions which includes special cases like Gaussian or Cauchy distributions. Some applications are listed in the literature like in radar [2], engineering [3], finance [4], (etc).

In general, classic statistical problem is dominated by methods based on Gaussian models. However, in many cases, signals are rarely Gaussian and are often complex. Indeed, they can present properties of skewness and heavy tails. A distribution is said to have heavy tails if the tails decays slower than the tail of the Gaussian distribution. Thefore, property of skewness means that it is impossible to find a mode where curve is symmetric. One of avantages of α-stable distributions compared to Gaussian cases is that allows to represent these properties. Consequently, α-stable distributions can be a class of distributions which is used for example to follow texture parameters for pattern recognition.

A Bayesian approach is often used to solve problem of pattern recognition [5]. However, Bayesian approach requires to know prior probabilities. To avoid to estimate prior probabilities, it is possible to use belief functions.

The theory of belief functions in discret domain has been developed by Dempster [6], and Shafer [7]. It can be seen as an extension of theory of probabilities. One of avantages of belief functions is that allows to consider imprecision and uncertainty of measures.

Recently, some works extend theory of belief functions in continuous domain [8]. The theory of belief functions has been applied to Gaussian cases in [9]. Consequently, we try to develop the theory of belief functions to α-stable distributions.

The structure of the paper is as follows: definitions about α-stable distributions are given in section one. Therefore, belief functions on discret and continuous domain are explained in section two. Finally, in section three, a method is proposed to calculate belief functions where information transmitted by sensors are modelized by α-stable distributions and the interest of belief functions is showed compared a Bayesian approach.

2 α-stable distributions

α-stable distributions allow to represent phenomena of skewness and heavy tails. Consequently, a definition of stable random variable is given, then a way to calculate α-stable distributions is developed.

2.1 Stable random variable

The notion of stability has been introduced by P. Lévy in [1]. A random variable X is said stable if $\forall (a, b) \in (\mathbb{R}^+)^2$, it exists $c \in \mathbb{R}^+$ and $d \in \mathbb{R}$ such as:

$$aX_1 + bX_2 = cX + d \tag{1}$$

with X_1 and X_2 independent copies of X. Equation (1) define the notion of stability but gives no information how to parametrize α-stable distributions. Consequently, α-stable distributions are defined from characteristic functions.

2.2 Characteristic functions

There are in the literature several parameterizations of α-stable distributions, noted $S_\alpha(\beta, \gamma, \delta)$. The most known definition is that used by Samorodnitsky and Taqqu [10]. However, the most employed definition is given by Zolotarev [11]. A random variable is said stable if his characteristic function verifies:

if $\alpha \neq 1$
$$\phi(t) = \exp(it\delta - |\gamma t|^\alpha [1 + i\beta\tan(\frac{\pi\alpha}{2})sign(t)(|t|^{1-\alpha} - 1)])$$
if $\alpha = 1$
$$\phi(t) = \exp(it\delta - |\gamma t|[1 + i\beta\frac{2}{\pi}sign(t)\log|t|])$$

with $\alpha \in]0, 2]$, $\beta \in [-1, 1]$, $\gamma \in \mathbb{R}^{+*}$ and $\delta \in \mathbb{R}$. The four parameters can be interpreted as:

- α is called characteristic exponent
- β is called skewness parameter
- γ represents the scale parameter
- δ designates location parameter

It is possible to obtain probability density function (*pdf*) by applying a Fourier transform to characteristic function:

$$pdf(x) = \int_{-\infty}^{+\infty} \phi(t)\exp(-itx)dt \quad (2)$$

However, it is difficult to represent probability density functions for two reasons. Indeed, the characteristic function is complex and the interval of integration is infinite. Nolan [12] suggests to make variable modifications to have a set of integration finite. This method has been programmed in a Matlab code [13].

Each parameter has an influence on plots. Indeed, if α is small, the distribution presents a more important peak (*cf.* figure 1). When $\beta \to 1$, distributions is shapeless on left and conversely when $\beta \to -1$ (*cf.* figure 2). Moreover, parameter γ allows to dilate and compress distributions (*cf.* figure 3). Distributions are translated on the abscissa when parameter δ varies (*cf.* figure 4).

2.3 Examples of pdf

Despite the lack of closed formulas, it is possible to describe known distributions. Indeed, when $\alpha = 2$ and $\beta = 0$, Gaussian distribution is defined:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x - \delta)^2}{2\sigma^2}\right) \quad (3)$$
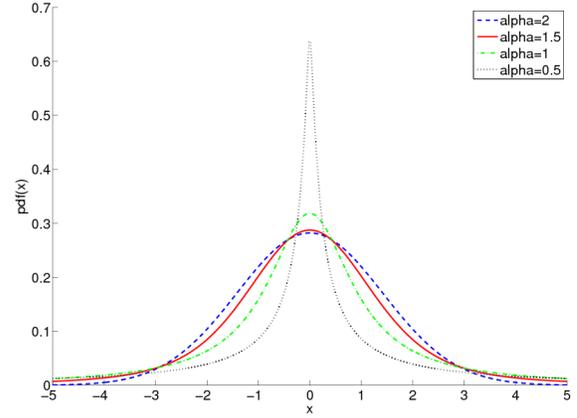

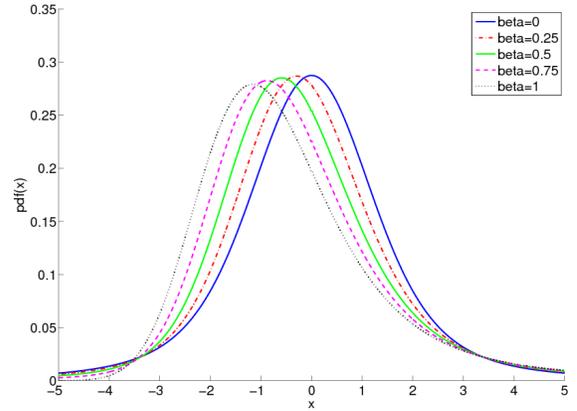Figure 1: Influence of parameter α with $\beta = 0$, $\gamma = 1$ and $\delta = 0$.

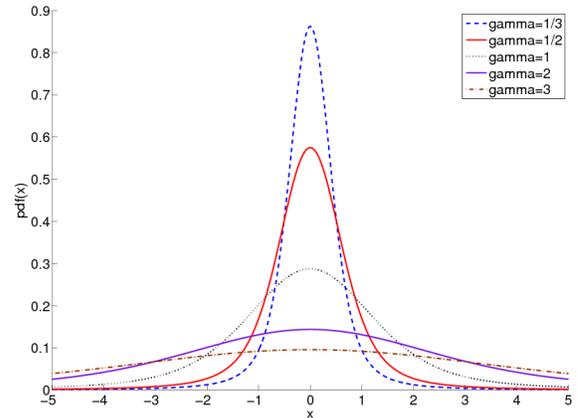Figure 2: Influence of parameter β with $\alpha = 1.5$, $\gamma = 1$ and $\delta = 0$.

Figure 3: Influence of parameter γ with $\alpha = 1.5$, $\beta = 0$ and $\delta = 0$.
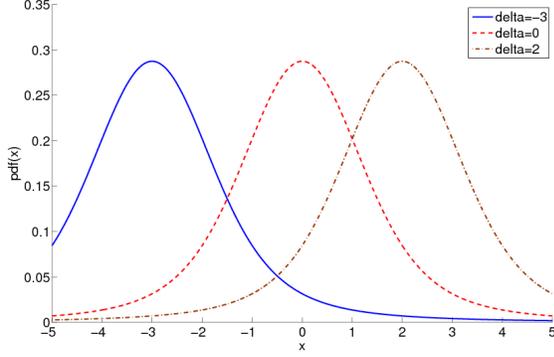
Figure 4: Influence of parameter γ with $\alpha = 1.5$, $\beta = 0$ and $\delta = 0$.

with δ corresponding to the mean and σ^2 to the variance. To recognize α-stable distribution, we have $\sigma^2 = 2\gamma^2$. When $\alpha = 1$ and $\beta = 0$, it is possible to build a Cauchy distribution:

$$f(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x - \delta)^2} \qquad (4)$$

It is usual to define symmetric α-stable distributions, *i.e.* a class of distributions where $\beta = 0$.

3 Theory of belief functions

The theory of probabilities is limited because it is difficult to consider imprecisions of measures and it is necessary to know prior probabilities. However, these problems can be solved by using works on belief functions. In this part, the theory of belief functions is developed in discret and continuous domains.

3.1 Discret belief functions

Firstly, it is necessary to consider a frame of discernement $\Theta = \{C_1, C_2, \ldots, C_n\}$, where C_j correspond to a class. The belief is represented by a basic belief assignment m defined on 2^Θ and in $[0, 1]$. m verifies by construction $\sum_{A \in 2^\Theta} m(A) = 1$. When $m(A) > 0$, A is focal element. With these definitions, it is possible to construct several functions:

- credibility function, noted bel, we can calculate it by the formula:

$$bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B) \qquad (5)$$

- plausibility function, noted pl, We can define it by:

$$pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \qquad (6)$$

We can also calculate pignistic probability used for decision step [14]. It is an operator which can give probability on every $C_j \in \Theta$. It is possible to determinate this probability with:

$$BetP(C_j) = \sum_{A \subseteq \Theta, C_j \in A} \frac{m(A)}{|A|(1 - m(\emptyset))} \qquad (7)$$

$|A|$ represents the cardinality of A. The combinaison between two mass functions can be realized with orthogonal rule of Dempster. Indeed, m_1 and m_2 and $\forall X \in 2^\Theta$:

$$m(X) = \frac{\sum\limits_{Y_1 \cap Y_2 = X} m_1(Y_1)m_2(Y_2)}{1 - \sum\limits_{Y_1 \cap Y_2 = \emptyset} m_1(Y_1)m_2(Y_2)} \qquad (8)$$

3.2 Continuous belief functions

The theory of belief functions on continuous domain has been developed by Smets [8]. Compared to discret domain, mass functions become mass density defined on sets $[a, b]$ of $\mathbb{R}$ with $m([a, b]) = f(a, b)$ where f is probability density of $\{(z, y) \in \mathbb{R}^2 | z \leq y\}$ in $[0, \infty]$. Credibility degree on set $[a, b]$, noted $bel([a, b])$, is defined with:

$$bel([a, b]) = \int_{z=a}^{z=b} \int_{y=z}^{y=b} f(z, y)dydz \qquad (9)$$

It is possible to define plausibility degree with:

$$pl([a, b]) = \int_{z=-\infty}^{z=b} \int_{y=max(a,z)}^{y=+\infty} f(z, y)dydz \qquad (10)$$

Mass densities can be determinated from probability density functions. However, many mass densities can induce the same probability density function. The transformation is said isopignistic. To avoid this problem, the hypothesis of consonant mass densities is made, *ie* focal elements are nested. It is possible to create an index u where focal elements are labeled I_u, verifying $I_u \subseteq I'_u$ with $u' > u$. It is necessary to apply the principle of least commitment to have this property: Authors [15] use this method to calculate plausibility functions where probability densities are unimodal Gaussian distributions. Caron *et al.* [9] generalize this approach in multidimensional case. The plausibility of a point mass x is given by:

$$pl(x) = 1 - F_{n+2}((x - \mu)^T \Sigma^{-1}(x - \mu)) \qquad (11)$$

where μ is the mean, Σ represent matrix of covariance and F_{n+2} is the cumulative density function of a χ^2 distribution with $n + 2$ degrees of freedom, defined by:

$$F_{n+2}(\chi^2) = \int_0^{\chi^2} \frac{u^{\frac{n+2}{2} - 1}}{2^{\frac{n+2}{2}} \Gamma(\frac{n+2}{2})} \exp(-\frac{u}{2})du \qquad (12)$$

Then, several plausibility functions can be combined by using general Bayesian theorem to obtain mass function at x:

$$m(x)(A) = \prod_{C_j \in A} pl_j(x) \prod_{C_j \in A^c} (1 - pl_j(x)) \qquad (13)$$

The next step consists to convert mass functions into pignistic probabilities by using equation (7). This criterion distributes uniformly mass of focal elements on singletons. Finally, the decision is taken by using the maximum of pignistic probability.

4 Belief functions and α-stable distributions

Belief functions have been applied to Gaussian cases in pattern recognition. However, α-stable distribution have better properties to represent noise [16]. Consequently, we propose a way to calculate plausibility functions where probability density functions are modelized by symmetric α-stable distributions. Therefore, a comparison is made between method develop in section 3.2 and our method in Gaussian case. Bayesian approach is compared with belief functions in stable cases. Finally, a problem of classification is simulated to show the interest of α-stable distribution compared Gaussian case.

4.1 Calculation of plausibility

Smets [8] defines plausibility functions in the case of unimodal probability density, noticed $Betf$, of mode μ. For $x > \mu$, plausibility functions can be calculated by:

$$pl(x) = \int_{t=x}^{t=+\infty} (\nu(t) - t) \frac{dBetf(t)}{dt} dt \qquad (14)$$

where $\nu(t)$ verifying $Betf(\nu(t)) = Betf(t)$. In symmetric case, equation (14) can be simplified [8]:

$$pl(x) = 2(x - \mu)Betf(x) + 2 \int_{t=x}^{t=+\infty} Betf(t)dt \qquad (15)$$

From equation (15), we use relation of Chasles to obtain:

$$\int_{t=x}^{t=+\infty} pdf(t)dt = \int_{t=-\infty}^{t=+\infty} pdf(t)dt - \int_{t=-\infty}^{t=x} pdf(t)dt \qquad (16)$$

where pdf designate the probability density functions of α-stable distributions. However, $\int_{t=-\infty}^{t=x} pdf(t)dt$ correspond to cumulative density function, also written cdf, calculated at point x and by definition, $\int_{t=-\infty}^{t=+\infty} pdf(t)dt = 1$. Consequently, equation (15) becomes:

$$pl(x) = 2(x - \mu)pdf(x) + 2(1 - cdf(x)) \qquad (17)$$

It will be useful to compare this result with plausibility functions obtained by Caron *et al.* [9], where probability densities are Gaussian, to validate it.

4.2 Validation of plausibility functions

We take up the application of identification of flying-object published in [15]. Speed can be considered as a feature. These features can be measure by sensors. In general, these measures are imprecise. Distribution of speed can be represented by a Gaussian or α-stable distributions. In our example, three α-stable distributions are chosen arbitrary given by notations defined in section 2.2:

- $S_2(0, 8, 722.5)$

- $S_2(0, 7, 690)$

- $S_2(0, 10, 730)$

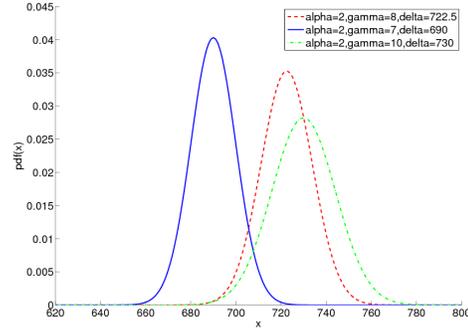These three distributions are plotted in figure 5.

Figure 5: Plot of three Gaussian distributions

As it is explained in section 2.3, these three distributions are Gaussian because $\alpha = 2$ and $\beta = 0$. For each distribution, a plausibility is calculated at every point x between $[620; 800]$ with a step of 1. They are transformed into mass functions by using general Bayesian theorem. Finally, the decision is taken by using maximum of pignistic probability. Plots can be observed at figure 6(a).

The same scheme is followed for approach of Caron *et al.* [9]. We observe in figure 6(b) results of pignistic probabilities. A comparison between two functions can be made by estimating a coefficient of correlation given by the formula: $c = \sum_{n=0}^{N-1} x_n y_n^*$. Coefficients of correlation are superior than 0.99 in the three distributions. The difference is due to numerical approximations during calculations of α-stable distributions. Consequently, we have shown that approach with α-stable distributions, in the particular case of three Gaussian distributions, is equivalent to current approach of Caron. To conclude, a way to calculate belief functions by supposing probability densities are α-stable distributions has been developed.
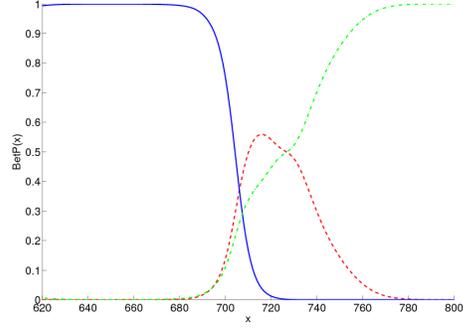
(a) pignistic probability of three α-stable distributions
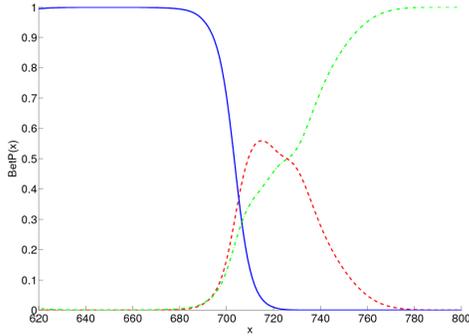
(b) pignistic probability by using Caron method

Figure 6: Plots of pignistic probability

4.3 Comparison with Bayesian approach

In this section, we take now three distributions that are not Gaussian. The three symmetric α-stable distributions are chosen with parameters:

- $S_{1.5}(0, 8, 722.5)$

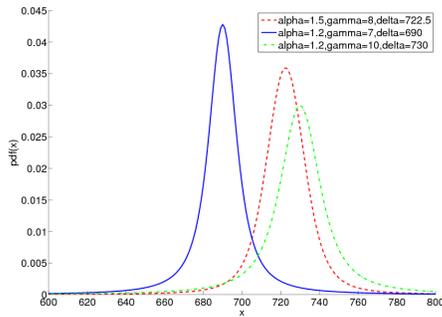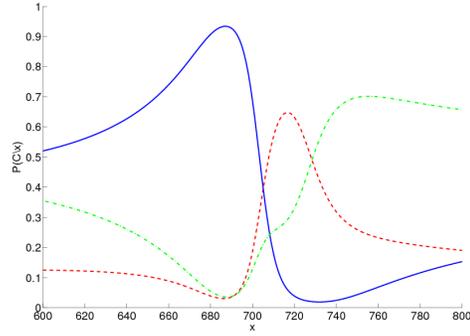- $S_{1.2}(0, 7, 690)$

- $S_{1.2}(0, 10, 730)$

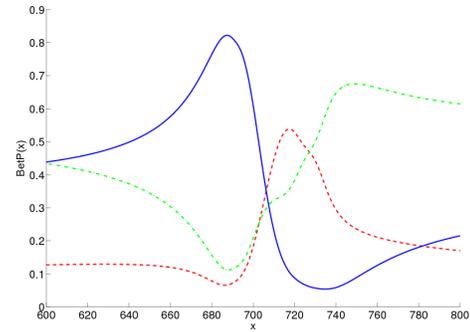Figure 7: Plots of three α-stable distributions.

These distributions can modelize speed feature of different planes. These graphics are represented at figure 7.

It is necessary to explain Bayesian approach in several steps. Firstly, prior probability must be calculated and can be modelized by:

$$p(x/C_j) = \int_{-\infty}^{+\infty} \phi(t) \exp(-itx)dt \qquad (18)$$

(a) Class probabilities

(b) Pignistic class probabilities

Figure 8: Plots of Bayesian analysis and belief functions analysis.

The application of Bayes theorem gives posterior probability:

$$p(C_j/x) = \frac{p(x/C_j)p(C_j)}{\sum\limits_{j=1}^{n} p(x/C_j)p(C_j)} \qquad (19)$$

Finally, the decision is chosen by using the maximum of posterior probability. Same steps are followed as them described in section 4.2 to calculate pignistic probabilities. Plots of theses two approach can be observed at figure 8.

When we analyse class probabilities in figure 8(a), decisions are clearly defined at the end of the tails. On the contrary, pignistic class probabilities in figure 8(b) show that it is difficult to take a decision. Indeed, tails are mixed and it is difficult to choose one or the other class. This remark is interesting for the classification when we have a lack of data for learning database.

To show interest of belief functions, we try to classify

generated data. For example, 300 samples for each α-stable distributions can be simulated by using the same approach given in [17]. The rate of correct classification are calculated. The two approaches are roughly the same because rates are near 70 %. However, the criterion of decision choosen considers only singletons. In the literature, a criteron of decision, introduced in [18] and used in [19] , allows to take decison on union of classes. Consequently, the decision A of 2^{Θ} is obtained by:

$$A = argmax(m_b(X)(x)pl(X)(x)) \qquad (20)$$

where m_b is a basic belief assignment given by:

$$m_b(X) = K_b \lambda_X \left(\frac{1}{|X|^r} \right) \qquad (21)$$

r is a parameter in $[0,1]$. When $r \to 0$, we have a lack of information and more weight is given to union of classes. On the contrary, when $r \to 1$, more weight is allocated to singletons. K_b is a constant of normalization, which respects $\sum_{X \in 2^{\Theta}} m_b^j(X) = 1$. λ_X allows the integration of the lack of knowledge on one of the elements $X \in 2^{\Theta}$. For our application, we choose $r = 0.4$
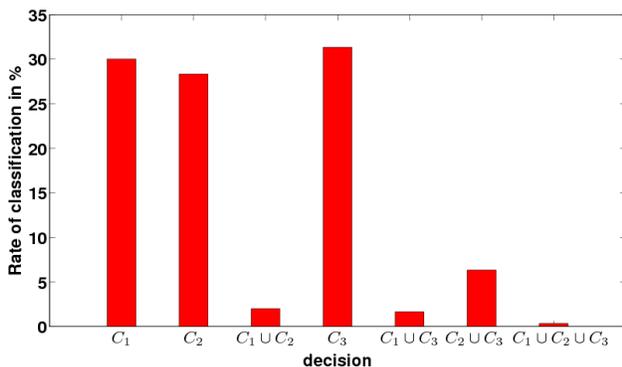

Figure 9: Classification with possible decision on union

to consider ambiguity between the classes. The results are illustrated in figure 9. The proportion given to the subset $C_2 \cup C_3$ is important compared to the others. It is logical due to the pignistic class probabilities. Consequently, belief functions consider imprecisions of measures whereas it is impossible with a Bayesian approach.

4.4 Gaussian vs α-stable model

In this section, we try to show α-stable models are more robust in classification compared to Gaussian models. Firstly, 1000 samples of the three symmetric α-stable distributions, defined in section 4.3, are generated. 1/3 is used for the learning database and 2/3 for the test database. Therefore, in Gaussian case, mean and standard-deviation are estimated. In α-stable case, parameters α, γ and δ are estimated by using [20]. The next step consists to applicate belief functions, developed in last parts, on test database. A final rate of

correct classification is calculated as a mean of 10 successive runs. The obtained results are:

- 58.06 % in Gaussian case

- 67.24 % in α-stable distribution

The model of probability density functions with α-stable distribution gives results roughly better than the hypothesis of Gaussian case. We must take care to the model of estimation.

5 Conclusions

This paper proposes a way to calculate belief functions by using α-stable distributions in symmetric case. We have confirm this approach with a measure a correlation in Gaussian case. We have showed avantages of belief functions compared a Bayesian approach in a perspective of classification. Indeed, this theory does not need to know prior probability. Furthermore, the theory of belief functions decide not necessary an exclusive class: it is possible to consider a decision as being a union of classes.

In our future works, we try to estimate unknown distributions. However, it is necessary to solve several problems. We must generalize belief functions in non symmetric case. Indeed, it is difficult to define ν as $Betf(\nu(t)) = Betf(t)$. In symmetric case, the mode is known and equal to δ and the function $\nu(t) = \delta - t$. On the contrary, the mode and ν are unknown in non symmetric case. Multimodal distributions can be estimated with a mixture of α- stable distributions as it is realized in [21]. However, it is difficult to define belief functions in this case. However, it is possible to use works published in [22] where consonant belief functions are build form a multimodal distributions. Finally, our final objective is to develop classification from several SONAR images by using α-stable distributions. Results have been already given in Gaussian case [23] and it will be interesting to compare with α-stable distributions.

References

[1] P. Lévy. Théorie des erreurs : La loi de Gauss et les lois exceptionelles, *Bull. Soc. Math. France*, 52: 49–85, 1924.

[2] A. Achim, and P. Tsakalides, and A. Bezerianos. SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modelling, *IEEE Transactions on Geoscience and Remote Sensing*, 41(8): 1773–1784, 2003.

[3] C.L. Nikias and M. Shao. *Signal processing with alpha-stable distributions and applications*, 1995.

[4] J.H. McCulloch. Financial applications of stable distributions, *Handbook of statistics*, 14: 393–425, 1996.

[5] D.P. Williams. Bayesian Data Fusion of Multi-View Synthetic Aperture Sonar Imagery for Seabed Classification, *IEEE Transactions on Image Processing*, 18: 1239–1254, 2009.

[6] A. Dempster. Upper and Lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics*,38: 325-339, 1967.

[7] G. Shafer. *A mathematical theory of evidence*, Princeton University Press, 1976.

[8] Ph. Smets. Belief functions on real numbers, *International journal of approximate reasoning*, 40(3): 181-223, 2005.

[9] F. Caron, and B. Ristic, and E. Duflos, and P. Vanheeghe. Least Committed basic belief density induced by a multivariate Gaussian pdf, *International Conference on Information Fusion (FUSION'2006)*. Florence, Italie, 2006.

[10] M.S. Taqqu, and G. Samorodnisky. Stable nongaussian random processes, *Chapman and Hall*, 1994.

[11] V.M. Zolotarev. One-dimensional stable distributions, *Amer. Math. Soc. Transl. of Math. Monographs, Vol. 65. Amer Mathematical Society, Providence, RI. (Transl. of the original 1983 Russian*, 1986.

[12] J.P. Nolan. Numerical calculation of stable densities and distribution functions, *Communications in Statistics-Stochastic Models*, 13(4): 759–774, 1997.

[13] http://math.bu.edu./people/mveillet/research.html.

[14] Ph. Smets. Constructing the pignistic probability function in a context of uncertainty, *Uncertainty in artificial intelligence*, 8: 29–39, 1990.

[15] B. Ristic, and Ph. Smets, Belief function theory on the continuous space with an application to model based classification, *Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems* : 4–9, 2004.

[16] P.G. Georgiou, and P. Tsakalides, C. Kyriakakis, Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise, *IEEE Transactions on Multimedia*, 1(3): 291–301, 1999.

[17] J.M. Chambers, and C.L. Mallows, and B.W. Stuck. A method for simulating stable random variables, *Journal of the American Statistical Association*, 71(354): 340–344, 1976.

[18] A. Appriou. Approche générique de la gestion de l'incertain dans les processus de fusion multisenseur, *Traitement du signal*, 22(4): 307–319, 2005.

[19] A. Martin, and I. Quidu, Decision support belief functions theory for seabed characterization, *International Conference on Information Fusion, Cologne, Germany*, 2008.

[20] J.H. McCulloch, Simple consistent estimators of the parameters of stable laws, *Journal of the american statistical association*, 75(372): 918–928, 1980.

[21] D. Salas-Gonzalez, and E.E. Kuruoglu, and D.P. Ruiz. Finite mixture of α-stable distributions, *Digital Signal Processing* 19(2): 250–264, 2009.

[22] P.-E. Doré, and A. Martin, and A. Khenchaf, Constructing consonant belief function induced by a multimodal probability, *COGnitive systems with Interactive Sensors (COGIS2009)*, 2009.

[23] A. Fiche, and A. Martin. Bayesian approach and continuous belief functions for classification, *Rencontre francophone sur la Logique Floue et ses Applications (LFA2009)*, 2009.