THE ANNALS OF "DUNĂREA DE JOS" UNIVERSITY OF GALATI
FASCICLE III, 2013, VOL. 36, NO. 1, ISSN 1221-454X
ELECTROTECHNICS, ELECTRONICS, AUTOMATIC CONTROL, INFORMATICS

SOME PROPERTIES OF ENTROPIES FOR CHANGE DETECTION

Dorel Aiordachioaie

"Dunarea de Jos" University of Galati
Electronics and Telecommunication Department
Domneasca-47, Galati-800008, Romania

Abstract: The segmentation of signals is an important process in change detection applications. Many results reported in the literature use information and entropy, as direct or indirect processed variables. The objective of the paper is to make a short review of the similarities between Shannon and Renyi entropies, and the ways of using them in change detection problem. This paper is on signal description and analysis from information point of view, and it can be viewed as a practical short overview.

Keywords: Change detection, Signal processing, Entropy, Information processing.

1. INTRODUCTION

The general problem of change detection and isolation, as well the definition of the problems, algorithms and solutions, are well presented in a number of solid publication as (Basseville and Benveniste, 1989; Basseville and Nikiforov, 1993; Basseville,1988), of (Gustafsson, 2000) and (Poor, and Hadjiliadis, 2008). The basic field of application is in incipient fault detection and process diagnosis, as presented in (Isermann, 2005; Jardine, *et al*, 2006; Peng and Chu, 2004) or (Venkatsasubramanian, et al, 2003). Automatic segmentation of signals is also an important process in change detection applications. As examples, we might have seismic signals segmentation or images (edge and object new detection), ecosystems, structures of complex processes, etc.

Two basic methods could be considered for change detection: (1) statistical methods, which use estimation of the main moments up to order four (see (Markou and Singh, 2003); (2) time-frequency transforms (Chen and Ling, 2002; Daubechies, 1990) or (Grochenig, 2001). It seems that the second type of method, based on time-frequency analysis, is more powerful, even it is more complex and time consuming. One of the key aspects in the success of these methods is to change the space of measurements, in the sense of transforming. Instead of analyzing the instantaneous values of the raw signal, an auxiliary signal is considered, with behavior more flexible, i.e. more illustrative concerning the moments of changes. One of such transform is based on Renyi entropy.

The objective of the paper is to make a review of the similarities between Shannon and Renyi entropies, and the ways of using them in change detection problem. This paper is on signal description and analysis from information point of view, and it can be viewed as a practical and short overview. The structure of the paper has three main sections. The first one considers the analysis of Shannon and Renyi entropies. The next section, section 3, introduces the basic properties of the considered entropies. Section 4 considers the entropy estimation problem. The last section investigates a short case study related to change detection in seismic signals.

2. SHANNON AND RENYI ENTROPIES

2.1 Discrete case

Discrete case considers a finite set of independent events with associated probabilities under a statistical distribution as

This paper was recommended for publication by Adrian Filipescu

$$S:\begin{pmatrix} s_1 & s_2 & \dots & s_N \\ P(s_1) & P(s_2) & \dots & P(s_N) \end{pmatrix} \quad (1)$$

with

$$0 < P(s_j) \le 1, \quad j = \overline{1,N} \quad (2.a)$$

$$\sum_{j=1}^{N} P(s_j) = \sum_{j=1}^{N} P_j = 1 \quad (2.b)$$

In communication systems, at least, the entropy, as quantitatively measure of information, has been considered mainly by Shannon, which has continued the ideas of Hartley and Nyquist. There is a set of three basic papers, (Hartley, 1928; Nyquist, 2002), and (Shannon, 1948) well knew and accepted as the basic set in information measuring and modelling.

Shannon entropy *H(P)* is defined as the mean of the discrete random variable, called information, $i(P_j)$, obtained after or prior of an probabilistic event by

$$H(P) = \overline{i(s_j)} = \sum_{j=1}^{N} P_j \cdot i(P_j) = -k \cdot \sum_{j=1}^{N} P_j \, log(P_j), \quad k > 0 \quad (3)$$

where k is a positive constant depending on the unit of measure of information; *k*=1 for [bit / symbol]. This is called Shannon entropy (discrete case). The logarithmic measure of entropy is indicated by a set of three reasons (Shannon), being usefulness, intuitiveness and mathematically convenience (suitableness). It is the main expression used in the information and coding theory. Equation (3), for the Shannon entropy, uses linear average. By considering of the general theory of means, for any function *g(x)* with inverse g^{-1}, the mean can be computed as, (Principe, 2010a; 2010b):

$$g^{-1}\left(\sum_{j=1}^{N} P_j \cdot g(x_j)\right) \quad (4)$$

Applying this definition to *H(P)* we get

$$H(P) = g^{-1}\left(\sum_{j=1}^{N} P_j \cdot g(I_j)\right) \quad (5)$$

By imposing independent events, we get two possible solutions for *g(x):*

$$g(x) = c \cdot x \quad (6)$$

$$g(x) = c^{-2(1-\alpha)x} \quad (7)$$

The first one gives Shannon entropy and the second gives

$$H_\alpha(P) = \frac{1}{1-\alpha} \log\left(\sum_{j=1}^{N} P_j^\alpha\right), \quad \alpha > 0, \alpha \ne 1 \quad (8)$$

which means a parametric family of information measures that are called Renyi’s entropies. The Renyi entropies tend to Shannon entropy as $\alpha \to 1$. Fig.1 shows the behaviour of the two information measures. The Renyi’s entropies contain Shannon as a special case, (Bromiley, *et al*, 2004).

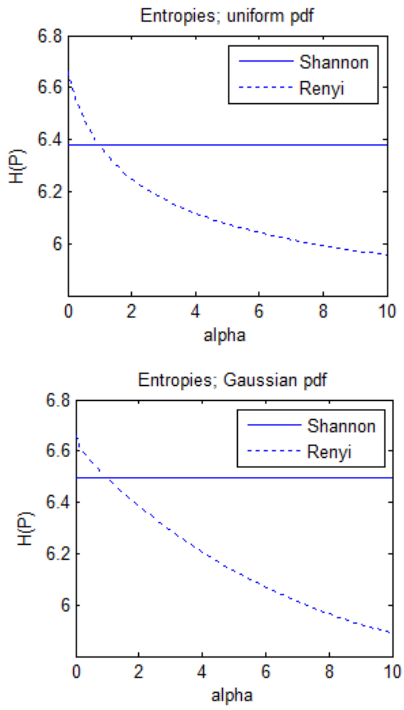

Fig. 1. Evolution of the entropies with alpha

An important feature of Shannon’s entropy is that, for a fixed variance, it is maximized for Gaussian distributions (see properties of the Shannon entropy). By simulation, the Renyi entropy is maximized as well by the Gaussian pdf as it is was presented in Fig. 2, right side. It is useful to know the behaviour around Gaussian distribution, i.e. distributions which are sub-Gaussian or super-Gaussian. The expression of the generalized Gaussian pdf (GGpdf) is, (Domínguez-Molina, *et al*, 2003),:

$$p(x) = B \cdot exp\left(-C \cdot |x|^\beta\right) \quad (9)$$

where $\beta > 2$ refers to the super-Gaussian region and $\beta < 2$ refers to the sub-Gaussian region. Parameters *B* and *C* are function of β that ensure the pdf integrates to 1 and that yield a pdf corresponding to a unit-variance random variable. The values of $\beta = 1$, 2, and infinite, correspond to Laplacian, Gaussian and uniform pdfs, respectively. Fig. 2, right side, shows the GGpdf for various values of parameter β, values of 0.5, 1, and 2.

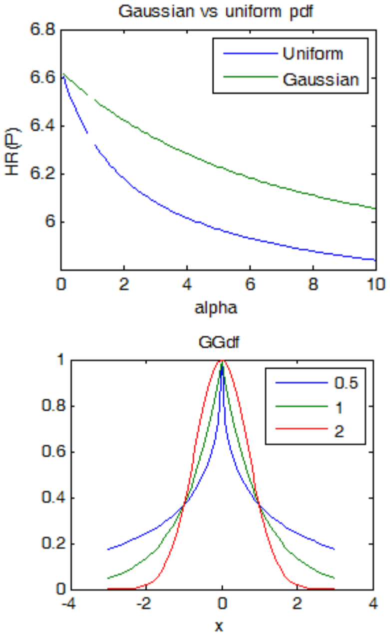

Fig. 2. Evolution of Renyi entropy ($\alpha = 1$) with various pdfs and GGpdfs

The values of Renyi's quadratic entropy and Shannon's Entropy for a uniform random variable are identical (this is true for all α > 0). For the Generalized Gaussian family, Shannon's entropy is maximized for β=2, as expected, and Renyi's entropy is maximized of β equal to 4, (Hild, *et al*, 2006).

2.2. Continuous case

It is often necessary to define and use the entropy of a continuous ensemble, *X*, described by the probability density function (pdf), *p(x)*. The Shannon entropy of *X* is defined by, (Gallager, 1968):

$$H(X) = -\int_{-\infty}^{\infty} p(x) \log p(x) dx \tag{10}$$

and, as example, a conditional entropy is defined by

$$H(X/Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y) \log p(x/y) dxdy \tag{11}$$

Similar expressions could be developed for other common entropies as *H(Y/X)* and *I(X,Y)*. The last one is the average mutual information between *X* and *Y*. These entropies are not necessarily positive, not necessarily finite, not invariant to transformations of the outcomes, and thus not interpretable as average of information, (Gallager, 1968). The absolute values are meaningless. Therefore, they can generally only be used in comparative or differential processes, (Bromiley, *et al*, 2004). In the continuous case, the Renyi's entropy of order α is defined as:

$$H_{R_\alpha}(Y) = \frac{1}{1-\alpha} \log \int f_Y(y)^\alpha dy, \quad \alpha > 0, \alpha \neq 1 \tag{12}$$

If $H_S(Y)$ is Shannon entropy for *Y*, then

$$\lim_{\alpha \to 1} H_{R_\alpha}(Y) = H_S(Y) \tag{13}$$

and

$$H_{R_\beta}(Y) \geq H_S(Y) \geq H_{R_\gamma}(Y), \quad 0 < \beta < 1; \alpha > 1 \tag{14}$$

When α=2, Renyi's entropy $H_{R_2}(Y)$ is also called quadratic entropy.

3. PROPERTIES OF THE ENTROPIES

3.1. Properties of the Shannon entropy

PS1 (*positivity*): The entropy is a positive function

$$H(S) = -\sum_{i=1}^{N} \underbrace{P(s_i)}_{>0} \cdot \underbrace{\log P(s_i)}_{\leq 0} \geq 0 \tag{15}$$

PS2 (*continuity*): The entropy is a continuous function with reference to each discrete variable defined on (0,1], being the sum of some elementary continuing functions (logarithms). The continuity property shows that a small variation in the distribution probability P_j of the events implies small variation in the information measure (entropy). So, if ΔP_j is small then $\Delta H(\Delta P_j)$ is also small.

PS3 (*symmetry*): The entropy is a symmetric function of variables P_j:

$$H(P_1, P_2, ..., P_N) = H(perm(P_1, P_2, ..., P_N)) \tag{16}$$

The change of the order in the list of probabilities does not change the value of the entropy.

PS4 (*additivity*): The entropy of the reunion of some independent events S_j is the sum of the entropies of the considered events.

$$S = \bigcup_{j=1}^{n} S_j, \bigcap_{j=1}^{n} S_j = \Phi \rightarrow H(S) = \sum_{j=1}^{n} H(S_j) \tag{17}$$

PS5 (*superior margin*): The maximum value of the entropy is reached when the events have the same probabilities:

$$H(S) = \log N, \text{ iff } P_i = 1/N, \forall i \tag{18}$$

3.2. Properties of the Renyi's entropy, (Principe, 2010b)

The expression (8) is considered as reference.

PR1: $H_\alpha(X)$ is nonnegative: $H_\alpha(X) \geq 0$ (19)

PR2: $H_\alpha(X)$ is symmetric: $H_\alpha(0,1) = H_\alpha(1,0)$ (20)

PR3: $H_\alpha(X)$ is concave for $\alpha \leq 1$. For $\alpha > 1$ is not pure convex nor pure concave; It loses concavity for $\alpha > \alpha^* > 1$, where α^* depends on N, as

$$\alpha^* \leq 1 + ln(4)/ln(N-1) \tag{21}$$

PR4: $(\alpha - 1) \cdot H_\alpha(X)$ is a concave function of *X*. Fig. 3 shows the shape of the Renyi entropy for various values of alfa parameter.

PR5: $H_\alpha(X)$ is bounded, continuous and not increasing function of α. (see Fig. 3)

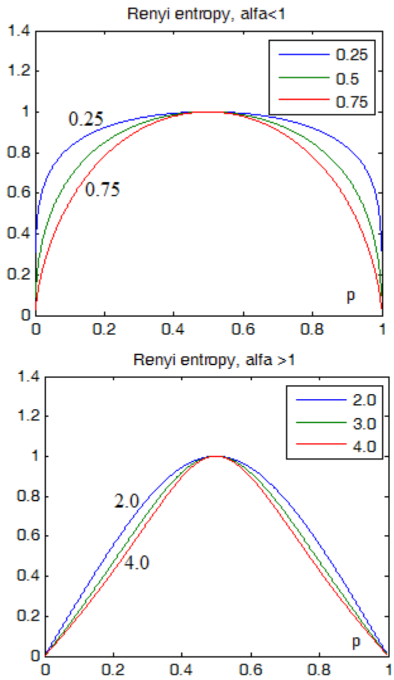

Fig. 3. Renyi entropy vs alfa

PR6: $H_\alpha(X)$ is a monotonic function of the information. Indeed, if we consider two experiments with the number of events *N* and, respectively, *N*+1, the Renyi entropy of the new one is grater than the old one.

$$H_\alpha^{(N+1)}(P) = \frac{1}{1-\alpha} log\left(\sum_{j=1}^{N+1} P_j^\alpha\right) =$$

$$\frac{1}{1-\alpha} log\left(\sum_{j=1}^{N} P_j^\alpha + P_{N+1}^\alpha\right) = \tag{22}$$

$$= H_\alpha^{(N)}(P) + \frac{1}{1-\alpha} log\left(P_{N+1}^\alpha\right) \geq H_\alpha^{(N)}(P)$$

3.3. Properties of Renyi's entropy with reference to Shannon entropy

P1: If $\alpha = 1$ then the Renyi's entropy is equal with the Shannon's entropy: $H_{R1}(P) = H_S(P)$ (23)

This can be observed by simulation, e.g. as it is presented in Fig. 1, or by observing that, (Bromiley, 2004):

$$H_{R1}(P) = \lim_{\alpha \to 1} \frac{1}{1-\alpha} log\left(\sum_{j=1}^{N} P_j^\alpha\right) =$$

$$= \lim_{\alpha \to 1} \frac{\frac{d}{d\alpha}\left[log\left(\sum_{j=1}^{N} P_j^\alpha\right)\right]}{\frac{d}{d\alpha}(1-\alpha)} = -\lim_{\alpha \to 1} \frac{\frac{1}{ln\, a}\sum_{j=1}^{N} P_j^\alpha\, ln(P_j)}{\sum_{j=1}^{N} P_j^\alpha}$$

$$= -\sum_{j=1}^{N} P_j \cdot log(P_j) = H_S(P) \tag{24}$$

where, if $y(x) = log_a\, \varphi(x), u > 0$, then $\frac{dy(x)}{dx} = \frac{\varphi'(x)}{\varphi(x)} \frac{1}{ln\, a} = \frac{1}{ln\, a} \cdot \frac{\varphi'(x)}{\varphi(x)}$, $\frac{da^x}{dx} = a^x \ln a \cdot$

P2: If the events have the same probabilities, i.e. P_j=1/*N*, *j*=1,2,…,*N* , then the two entropies are equal:

$$H_{R\alpha}(P) = \frac{1}{1-\alpha} log\left(\sum_{j=1}^{N} P_j^\alpha\right) =$$

$$= \frac{1}{1-\alpha} log\left(\sum_{j=1}^{N} \frac{1}{N^\alpha}\right) = \frac{1}{1-\alpha} log \frac{1}{N^\alpha} \cdot N = \tag{25}$$

$$\frac{1}{1-\alpha} log \frac{1}{N^{\alpha-1}} = log \sqrt[1-\alpha]{N^{1-\alpha}} = log(N), \quad \forall \alpha > 0$$

$$H_S(P) = -\sum_{j=1}^{N} P_j\, log\, P_j = -\sum_{j=1}^{N} \frac{1}{N} log \frac{1}{N} = \tag{26}$$

$$= -\frac{1}{N} \cdot N \cdot log \frac{1}{N} = -log \frac{1}{N} = log(N)$$

P3. A bound for Shannon's entropy is, (Nyquist, 2002):

$$H_2(X) \leq H_S(X) \leq ln(N) + 1/N - exp(-H_2(X)) \tag{27}$$

4. ENTROPY ESTIMATION

There is a motivation to use Renyi's entropy, which is related to the computation. As relation (1) and (3) show, the computation of the entropies needs the availability of the exact or estimated pdf. (The probabilities' set in discrete case and pdf in the continuous case). The real examples provide information in terms of data samples $\mathbf{a} \in R^{mx1}$, as observations of a random vector $Y_m \in R^{mx1}$. If we consider sets with N observation vectors, $\mathbf{a}_i$, the problem of computation of entropies become thus mandatory. The pdf is estimated using a Parzen window (Parzen, 1962) with Gaussian kernel

$$\hat{p}(\boldsymbol{y}/\boldsymbol{a}_i) = \frac{1}{N}\sum_{i=1}^{N} G(\boldsymbol{y}-\boldsymbol{a}_i, \sigma^2 \cdot \boldsymbol{I}) \quad (28)$$

where $G(.,.)$ is the Gaussian function

$$G(\boldsymbol{y}-\boldsymbol{a}_i, \sigma^2 \cdot \boldsymbol{I}) = \frac{1}{\sqrt{\left(2\pi\sigma^2\right)^m}} exp\left(-(\boldsymbol{y}-\boldsymbol{a}_i)^T \cdot (\boldsymbol{y}-\boldsymbol{a}_i)/2\sigma^2\right) \quad (29)$$

σ^2 is the variance, and $\mathbf{I} \in R^{mxm}$ is the identity matrix. The Renyi entropy estimator is, (Erdogmus, *et al*, 2002) for $\alpha > 0, \alpha \neq 1$:

$$\hat{H}_\alpha(Y_m, \sigma) = \frac{1}{1-\alpha} log \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{N}\sum_{k=1}^{N}\left(G(y_m(n)-y_m(k), 2\sigma^2\right)\right)^{\alpha-1} \quad (30)$$

It is observable that Eq. 8 is not valid for $\alpha = 1$. By using the idea of (Hild, *et al*, 2006), which starts by writing the Shannon entropy in terms of an expectation:

$$H_1(Y_m) = E\{i(Y_m)\} = -E\{log\, P(Y_m)\} \quad (31)$$

the expectation is replaced by the sample mean and using Parzen window estimation of pdf, the following estimator for Shannon entropy is obtained, (Hild, *et al*, 2006):

$$\hat{H}_1(Y_m, \sigma) = -\frac{1}{N}\sum_{n=1}^{N} log \frac{1}{N}\sum_{k=1}^{N}\left(G(y_m(n)-y_m(k), \sigma^2\right) \quad (32)$$

When Shannon's entropy definition is used along with this pdf estimation, the estimation becomes very complex. The second order Renyi entropy could be estimated more efficiently (*O(N)* instead of $O(N^2)$) by, (Hild, *et al*, 2006):

$$\hat{H}_2(Y_m, \sigma) = -log \frac{1}{N}\sum_{k=1}^{N}\left(G(y_m(k)-y_m(k-q), 2\sigma^2\right) \quad (33)$$

where the difference in time between the outputs, q, is user defined (the recommended value is q=1). Good approximation can be guaranteed if either one of the following two conditions is met, (Hild, *et al*, 2006): (1) Multiple entropy estimates are averaged, where the (time) indices of the data are uniformly randomized for each estimate; (2) The data is i.i.d. (independent and identically distributed) and N is sufficiently large, e.g. N>1000 requires only a single estimate. The entropy estimators require the selection of the kernel size, σ. This should be small small (relative to the standard deviation of the data). Values between 0.1 and 2 for unit-variance signals, are good choices, (Hild, *et al*, 2006).

5. APPLICATION IN SIGNAL SEGMENTATION

A synthetic signal is considered. It is composed of seven multicomponent epochs (*Ne*=7) with the duration of 6 seconds, excepting the last epoch of 5 seconds. The sampling frequency is *Fs* = 100 Hz. The parameter values of the signals are presented by Eq. 37. The free-noise signal is built by

$$u_i(t) = \sum_{j=1}^{n} A_{ij} \cdot cos\left(2\pi \cdot f_{ij} \cdot t\right), \quad i = 1,2,...,Ne \quad (35)$$

A Gaussian noise signal is added to the previous one:

$$s(t) = u(t) + n(t) \quad (36)$$

with zero mean and variance of 0.1. This signal might be a prototype signal for seismic signal class, where detection of the changes is quite important.

$$\mathbf{A} = \begin{bmatrix} 0.5 & 1.5 & 4.0 & 0.0 \\ 1.5 & 4.0 & 0.0 & 0.0 \\ 1.0 & 4.5 & 0.0 & 0.0 \\ 0.5 & 1.5 & 0.8 & 3.5 \\ 0.5 & 2.5 & 0.0 & 0.0 \\ 0.5 & 2.5 & 0.0 & 0.0 \\ 0.8 & 1.0 & 3.0 & 0.0 \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} 0.5 & 2 & 4 & 0 \\ 1 & 7 & 0 & 0 \\ 0.5 & 3.5 & 0 & 0 \\ 0.5 & 1 & 3 & 8 \\ 3 & 8 & 0 & 0 \\ 1.5 & 4 & 0 & 0 \\ 1.5 & 2.5 & 4 & 0 \end{bmatrix} \quad (37)$$

The resulted signal is presented in Fig. 4, upper side. The vertical lines show the limits of the epoch, i.e. the moments when changes occur. The next two subfigures show the evolution of the Renyi entropies, first and second orders. From the evolution of the entropies, and compared with the raw signal, it is clear that the change moments will be much easier detected, with ordinary methods based on statistical signal processing, e.g. mean and variance estimations. In the computation of the entropies, a moving (Hanning) weighted window of length 32 was used.

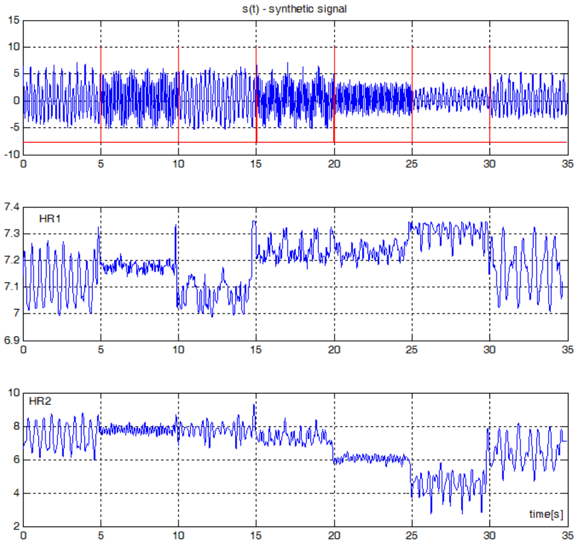

Fig.4. Raw signal, the 1st & 2nd order Renyi entropies

6. CONCLUSION

The Renyi entropies are important measures of information. The measures are scale-dependent when applied to continuous distributions, so their absolute values are meaningless. Therefore, they can generally only be used in comparative or differentiable processes. The Shannon's entropy can be viewed as one member of Renyi's entropy family. Renyi's entropy is more general and includes the Shannon's entropy as a special case. Estimation of Renyi's quadratic entropy form a finite data set is much easier than in Shannon's entropy case. The information content and the complexity of a probability density function can be measured by the entropy function. One of the key reasons to do this study was to prepare the next work, which is related to the use of Renyi entropy to measure the complexity and the information content of non-stationary multicomponent signals based on time-frequency transforms.

7. REFERENCES

Basseville, M., Benveniste, A. (Eds) (1985). *Detection of Abrupt Changes in Signals and Dynamical Systems*. Lecture Notes in Control and Information Sciences, LNCIS-77, Springer, Berlin, DE, Russian Translation, Mir, Moscow.

Basseville, M., Nikiforov, I.V. (1993). *Detection of Abrupt Changes - Theory and Application*. Prentice-Hall, Inc.

Basseville, M., (1988). Detecting Changes in Signals and Systems – A Survey, *Automatica*, **24** (3), pp. 309-326.

Bromiley, P.A., Thacker, N.A., and Bouhova-Thaker E. (2004). Shannon Entropy, Renyi Entropy, and Information, *Statistics and Information Series* (004), www.tina-vision.net (accsesed Oct. 2013)

Chen, V.C., Ling, H. (2002). *Time-Frequency Transforms for Radar Imaging and Signal Analysis*, Artech House.

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis, *Inf. Theory, IEEE Trans. on*, **36** (5), pp. 961-1005.

Domínguez-Molina, J. A., González-Farías, G., Rodríguez-Dagnino, R.M., (2003). *A practical procedure to estimate the shape parameter in the generalized Gaussian distribution*. http://www.cimat.mx/reportes/enlinea/I-01-18_eng.pdf.

Erdogmus, D., et al., (2002). Blind source separation using Renyi's alpha-marginal entropies, *Neurocomputing*, **49**(1), pp. 25–38.

Gallager, R.G. (1968). *Information theory and reliable communication*, Wiley.

Grochenig, K.. (2001). *Foundations of time-frequency analysis*, Birkhauser.

Gustafsson F., (2000). *Adaptive filtering and change detection*, John Wiley & Sons, Ltd.

Hartley, R.V.L. (1928). Transmission of information. *Bell Systems Tech. Journal*, **7**(3), pp. 535–563.

Hild II, K.E. , Erdogmus, D., Principe, J. C. (2006). An analysis of entropy estimators for blind source separation, *SP*., **86**(1), pp.182-194.

Isermann, R. (2005), Model-based fault-detection and diagnosis–status and applications, *Annual Reviews in Control*, **29**, pp. 71-75.

Jardine, A.K.S., Lin, D., Banjevic, D. (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical systems and Signal processing*, **20**, pp. 1483-1510.

Markou, M., Singh, S. (2003). Novelty detection: a review-part 1: statistical approaches, *Signal Processing*, **83**, pp. 2481-2497.

Nyquist, H. (2002). Certain topics in telegraph transmission theory, *Trans. AIEE*, **47**, pp. 617–644, Apr. 1928 Reprint in: Proc. IEEE, **90** (2).

Shannon, C.E. (1948). A math. theory of comm., *Bell Syst. Tech. Journal*, **27**, 379–423 and 623–656.

Parzen, E. (1962). On estimation of a prob. function &mode, *Ann. Math. Statist*.,**33**(3),pp.1065–1076.

Peng, Z.K., Chu, F.L. (2004). Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, *Mec. syst. and SP*, **18**, 199-221.

Poor, H. V. and Hadjiliadis, O. (2008). *Quickest Detection*, Cambridge Univ. Press.

Principe, J.C. (2010a). *The Renyi entropy*, Course EEL 6814, University of Florida, USA.

Principe, J.C. (2010b). *Information Theoretic Learning. Renyi's Entropy and Kernel Perspectives*, Springer.

Venkatsasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S.N. (2003). A review of process fault detection and diagnosis. Part I: Quantitative model-based methods, *Computers and Chemical Engineering*, **27**, pp. 293-311.