A NEW INFORMATION THEORETIC MEASURE OF PDF SYMMETRY

Weifeng Liu, P. P. Pokharel, Jose C. Principe

CNEL, University of Florida

ABSTRACT

In this paper, a new quantity called symmetric information potential (SIP) is proposed to measure the reflection symmetry and to estimate the location parameter of probability density functions. SIP is defined as an inner product in the probability density function space and has a close relation to information theoretic learning. A simple nonparametric estimator directly from data exists. Experiments demonstrate that this concept can be very useful dealing with impulsive data distributions, in particular, α -stable distributions.

Index Terms— Symmetric distributions, information theoretic learning, robust statistics

1. INTRODUCTION

A fundamental task in statistical analysis is to characterize the location and variability of a data set. Further characterization of the data includes skewness and kurtosis. In this paper, we mainly address two of these important issues: location and skewness.

The estimation of a location parameter is to find a typical or central value that best describes the data. For univariate data, mean, median, and mode are three common methods [9]. However, if the data is Cauchy distributed, the mean becomes useless, because collecting more data does not provide a more accurate estimate [9].

Another important characteristic of a data distribution is skewness, which is a measure of symmetry. A distribution is symmetric if it looks the same to the left and right of the center point. For the data set $\{x_i\}_{i=1}^N$, the skewness is defined as

$$sk = \sum_{i=1}^{N} (x_i - \overline{x})^3 / (N - 1)s^3$$
(1)

where \overline{x} and s are the mean and the standard deviation of the data. By definition, the skewness measures symmetry with respect to the mean. If the mean estimate is meaningless as in the Cauchy distribution, the skewness is also meaningless although the Cauchy distribution is obviously symmetric. In this paper, we try to solve this dilemma by defining a new concept of center and a new symmetry measure of data distributions based on information theoretic learning [1].

The organization of the paper is as follows. In section 2, a brief review is given about information theoretic learning. Then the definition and properties of symmetric information potential (SIP) are presented, based on which the Euclidean symmetry measure, Cauchy-Schwartz symmetry measure and the reflection point are defined in section 3. Possible applications are discussed in section 4. Finally, section 5 summarizes the main conclusions.

2. INFORMATION THEORETIC LEARNING

Given i.i.d. samples $\{x_i\}_{i=1}^N$ drawn from $f_X(x)$, the Parzen estimator [2] of the PDF is

$$\hat{f}_{X}(x) = \frac{1}{N} \sum_{i=1}^{N} G_{\sigma}(x - x_{i})$$
 (2)

where $G_{\sigma}(x-x_i)$ is the Gaussian kernel and σ is the kernel size.

$$G_{\sigma}(x-x_i) = \exp(-(x-x_i)^2/2\sigma^2)/\sqrt{2\pi}\sigma \qquad (3)$$

Renyi's quadratic entropy of a random variable X with PDF $f_X(x)$ is defined by

$$H_{2}(X) = -\log \int_{-\infty}^{+\infty} f_{X}^{2}(x) dx$$
 (4)

The argument of the log function in (4) is called *information potential* (IP) since the PDF estimated with Parzen kernels can be thought to define an information potential field over the space of the samples [3]. A non-parametric estimator of the IP (and thus of quadratic Renyi's entropy) directly from samples is obtained through (2)

$$I\hat{P}(X) = \frac{1}{N^2} \sum_{j=1}^{N} \sum_{i=1}^{N} G_{\sqrt{2}\sigma}(x_j - x_i)$$
(5)

3. SYMMETRIC INFORMATION POTENTIAL

3.1. Definition

<u>Definition</u>: Suppose the PDF of a random variable X is $f_X(x)$. Symmetric Information Potential (SIP) of X is defined as

$$SIP(X) = \int_{-\infty}^{+\infty} f_X(x) f_X(-x) dx$$
 (6)

With i.i.d. samples $\{x_i\}_{i=1}^N$ drawn from *X*, a nonparametric estimator is obtained as

This work was partially supported by NSF grant ECS-0601271.

$$SI\hat{P}(X) = \frac{1}{N^2} \sum_{j=1}^{N} \sum_{i=1}^{N} G_{\sqrt{2}\sigma}(x_j + x_i)$$
(7)

Comparing (7) with (5), we see that SIP is very similar to IP and furthermore when the distribution is symmetric, i.e. $f_X(x) = f_X(-x)$, SIP reduces to IP by definition.

3.2 Properties

Let X be a random variable.

<u>Property 1</u>: $SIP(X) \ge 0$.

<u>Property 2</u>: $SIP(X) \le IP(X)$. Equality holds if and only if $f_X(x) = f_X(-x)$.

Proof: By Cauchy-Schwartz inequality

$$SIP(X) = \int f_X(x) f_X(-x) dx$$

$$\leq \left(\int f_X(x)^2 dx \int f_X(-x)^2 dx\right)^{1/2}$$
(8)

$$= \int f_X(x)^2 dx = IP(X)$$

Equality holds when $f_X(x) = af_X(-x)$, where *a* is some real positive constant. Since

$$\int f_X(x)dx = \int f_X(-x)dx = 1$$
(9)

a can only be 1. This completes the proof.

<u>Property 3</u>: Define a new random variable $Y = X_1 + X_2$, where X_1 is independent of X_2 but they have the same PDF $f_X(x)$. SIP(X) equals the probability density of Y = 0. Proof: Since X_1 and X_2 are independent, the PDF of Y is

$$f_{Y}(y) = \int f_{X}(x) f_{X}(y - x) dx .$$
 (10)

Therefore,

$$SIP(X) = \int f_X(x) f_X(-x) dx = f_Y(y=0).$$
(11)

This completes the proof. With only the data available and by using (2) in (10), we have

$$\hat{f}_{Y}(y) = \frac{1}{N^{2}} \sum_{j=1}^{N} \sum_{i=1}^{N} G_{\sqrt{2}\sigma}(x_{j} + x_{i} - y)$$
(12)

This is exactly the Parzen estimator of $f_Y(y)$ if $\{(x_i + x_j)\}_{i,j=1}^N$ are regarded as the realizations drawn from *Y*.

3.3. Reflection Symmetry Measure

It is natural to measure symmetry of a data distribution by comparing its PDF $f_X(x)$ with its mirror image $f_X(-x)$ since this is the fundamental definition of the reflection symmetry. If we treat $f_X(x)$ and $f_X(-x)$ as two points in the PDF space, their Euclidean distance is

$$S_{ED} = \int (f_X(x) - f_X(-x))^2 dx$$

= $2(\int (f_X(x)^2 dx - \int f_X(x) f_X(-x) dx)$ (13)
= $2(IP(X) - SIP(X))$

By the properties of SIP, we see that (1) $0 \le S_{ED} < 2IP(X)$; (2) $S_{ED} = 0$ iff $f_X(x) = f_X(-x)$.

 S_{ED} is called the *Euclidean Symmetry measure*. Another way to define 'distance' in a vector space is to measure the angle between two vectors. Therefore, the Cauchy-Schwartz symmetry measure S_{CS} is

$$S_{CS} = \frac{\int f_X(x) f_X(-x) dx}{\left(\int f_X(x) f_X(x) dx\right)^{1/2} \left(\int f_X(-x) f_X(-x) dx\right)^{1/2}}$$

$$= \frac{\int f_X(x) f_X(-x) dx}{\int f_X(x) f_X(x) dx} = \frac{SIP(X)}{IP(X)}$$
(14)

This is a normalized version of SIP and it denotes the cosine value of the angle between $f_X(x)$ and $f_X(-x)$. S_{CS} has the following properties:

(1)
$$0 < S_{CS} \le 1$$
;
(2) $S_{CS} = 1$ iff $f_X(x) = f_X(-x)$.

 S_{ED} and S_{CS} are almost equivalent. Most importantly, they possess nice non-parametric estimators directly from data whereas other forms of divergence such as Kullback-Leibler are computationally expensive. Practically S_{CS} is more appropriate than S_{ED} since it removes the effect of the 'norm' of the PDF.

3.4. Reflection Point

In the previous definitions, we set the reflection point at the origin by default. However, it would be appealing in some cases to have the symmetry as an internal property of the data distribution and independent of the external coordinate system. As defined in (1), the skewness measures symmetry with respect to the data mean, and is therefore shiftinvariant. By analogy, the center of the data can be first estimated with a subsequent shift to the origin. However, this means our definition of symmetry depends on the particular choice of which point is the center and in practice depends on what kind of method is used to estimate it. Intuitively, the concepts of reflection point and reflection symmetry are co-dependent. Further, it is logically sound to define the reflection point, with respect to which the maximal symmetry is achieved. The reflection point R_p can be used to represent the 'center' of the data distribution.

Assume we shift the data by t. The S_{CS} becomes

$$S_{CS}(t) = \frac{\int f_X(x-t) f_X(-x+t) dx}{\left(\int f_X(x-t)^2 dx\right)^{1/2} \left(\int f_X(-x+t)^2 dx\right)^{1/2}}$$

$$= \frac{\int f_X(x) f_X(2t-x) dx}{\int f_X(x) f_X(x) dx}$$
(15)

The denominator, i.e. the IP, is shift-invariant and the numerator turns out to be $f_y(y = 2t)$. As a result, the reflection point is defined as

$$Rp = \arg\max_{t} \int f_{X}(x) f_{X}(2t - x) dx$$

= $\arg\max_{t} f_{Y}(y = 2t)$ (16)

Consequently, finding the reflection point of $f_X(x)$ is equivalent to finding the main mode of $f_Y(y)$. Since Y is the sum of two independent random variables, Y would be 'closer' to the normal distribution than X by the central limit theorem. Thus it is advantageous to deal with $f_Y(y)$ instead of $f_X(x)$ except for the increasing computation. Finding the main mode of $f_Y(y)$ can be accomplished by the mean shift algorithm or by gradient methods [4].

4. APPLICATIONS

4.1 Measuring Symmetry of data distributions

In the first example, we compare our method with skewness in assessing symmetry of data drawn from Cauchy, Laplacian and exponential distributions (Table I).

TABLE I Distributions Used in the First Example									
Distribution	PDF								
Cauchy	$f(x) = 1/\pi(1+x^2)$								
Laplacian	$f(x) = \exp(- x)/2$								
Exponential	$f(x) = \exp(-x), x \ge 0$								
2-Cauchy Mixture	$f(x) = \frac{1}{2} \frac{1}{\pi (1 + (x + 2)^2)} + \frac{1}{2} \frac{1}{\pi (1 + (x - 2)^2)}$								

For each distribution, 1000 points are drawn to estimate the mean, skewness, Rp and S_{CS} . 100 Monte Carlo realizations are run for each distribution. After that, we calculate the average and the standard deviation of these estimates. All distributions and their parameters used in the simulation are listed in Table I. The kernel size is chosen by the Silverman's rule [5]. The results are summarized in Table II. S_{CS} has much smaller estimation variance than the skewness. In the case of symmetric, heavy-tailed distributions like Cauchy and Laplacian, Rp also provides a much more accurate estimate of the center than the mean. In most cases, the skewness has such a large variance that its estimate is uninformative. The reason why S_{CS} is very suitable to describe impulsive distributions is beyond the scope of this paper (see [8]).

The α -stable distributions are a family of heavy-tailed distributions widely used in financial analysis [6]. The α -stable distribution requires four parameters for complete description: an index of stability $\alpha \in (0,2]$, a skewness parameter $\beta \in [-1,1]$, a scale parameter $\sigma > 0$ and a location parameter μ , denoted as $S_{\alpha}(\sigma,\beta,\mu)$. When $\alpha = 2$, the Gaussian distribution results. The Cauchy distribution is a special case when $\alpha = 1$ and $\beta = 0$. When $\alpha < 2$, the

variance is infinite and the tails are asymptotically equivalent to a Pareto Law. The estimation of stable law parameters is in general severely hampered by the lack of known closed-form density functions. As far as we know, simple descriptors for α -stable distributions are still lacking due to the flat tails and asymmetry. In this example, we show that the newly defined S_{CS} can be used to characterize the skewness of these distributions. In this example, we fix three parameters $\alpha = 0.8$, $\sigma = 1$, $\mu = 0$ and vary β from 0 to 0.9. For each β , 2000 data are generated to estimate the skewness and S_{CS}. 100 Monte Carlo simulations are run so that the average and the standard deviation are calculated for both estimates (Fig. 1). The skewness is estimated by two ways: one is with entire data and the other with 5% data trimmed off. As we see, a smooth monotonic curve is obtained between S_{CS} and β whereas the skewness is almost uninformative.



4.2 Signal detection in symmetrically distributed noises In this example, we utilize the property of R_p in a signal detection problem. In a digital communication system, let S and R be respectively the transmitted signal and the received signal corrupted by the additive noise N.

R

$$=S+N \tag{17}$$

Suppose $S = s_i$ which is either 0 or 1 with equal probably. The noise can be any symmetric distribution with reflection point at the origin. $S_{1,2}(1,0,0)$ and 2-Cauchy mixture noises (in Table I) are used in this simulation. The signal-noise-ratio (SNR) is simply controlled by scaling the noise. When an interval of signal is observed, T samples are obtained, based on which a decision is made whether the transmission is 0 or 1. The mean square error criterion is commonly used while M-estimation is more effective in impulsive noises [7]. Assume the sampled received signal is $\{r_i\}_{i=1}^T$. For the *Rp* detection method, the following criterion is used

Distributions	data mean		data skewness		<i>Rp</i> of data		S_{cs} of data	
	average	std	average	std	average	std	average	std
Cauchy	-2.7038	7.3509	-1.1888	24.037	-0.00073	0.04510	0.99951	0.00026
Laplacian	0.0194	0.0534	0.00556	0.19704	-0.00448	0.02650	0.99634	0.00245
Exponential	0.9969	0.0270	2.112	0.20936	0.48942	0.04184	0.84084	0.01616

TABLE II ESTIMATION RESULTS OF CAUCHY DISTRIBUTION

Decide 1 if
$$\sum_{\tau=1}^{T} \sum_{t=1}^{T} G(r_t + r_\tau - 2) > \sum_{\tau=1}^{T} \sum_{t=1}^{T} G(r_t + r_\tau);$$
 (18)

Decide 0 otherwise.

We set T=20 and transmit 10^4 bits in each simulation to calculate the bit error rate (BER). Kernel size is set according to the Silverman's rule.



Fig.2. Detection performance with α -stable noise



Fig.3. Detection performance with 2-Cauchy mixture noise

The noise power and trimmed MSE is estimated from 5% trimmed data. As we see in Fig.3, the mode detection

fails in the case of 2-Cauchy mixture because there is a dip in the noise PDF at the center.

5. CONCLUSIONS

In this paper, a new quantity Symmetric Information Potential (SIP) is proposed. Its mathematical meaning, probabilistic interpretation and relation to the information potential are presented. Based on this understanding, descriptors S_{ED} , S_{CS} and Rp are defined to quantify the reflection symmetry and to estimate the location parameter of data distributions. All these methods are non-parametric and robust so they are very useful to characterize impulsive data distributions. Examples demonstrate the newly proposed methods outperform the conventional mean and skewness in estimating the location parameter and quantifying symmetry in Laplacian, a-stable and other mixture models. Future work includes the bias and variance analysis of the SIP, detailed parameter estimation methods for α -stable distributions and its possible application in supervised learning.

6. REFERENCES

[1] Deniz Erdogmus, Jose Principe, "From linear adaptive filtering to nonlinear information processing," *IEEE Signal Processing Magazine*, Nov. 2006.

[2] E. Parzen, "On the estimation of a probability density function and the mode", *Ann. Math. Stat.* 33, p1065, 1962.

[3] Deniz Erdogmus, Jose C. Principe, "Generalized information potential criterion for adaptive system training," *Trans. on Neural Networks*, Vol. 13, No. 5, pp. 1035-1044, Sept. 2002.

[4] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790--799, Aug. 1995.

[5] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

[6] Nolan, J. P., "Numerical calculation of stable densities and distribution functions," *Communications in Statistics-Stochastic Models*, 13: 759-774, 1997.

[7] Weifeng Liu, P. P. Pokharel, J. C. Principe, "Correntropy: A Localized Similarity Measure," in Proc. of *Intl. Joint Conf. on Neural Networks*, 2006.

[8] Weifeng Liu, P. P. Pokharel, J. C. Principe, "Error Entropy, Correntropy and M-Estimation," in Proc. *IEEE Int. Workshop on Machine Learning for Signal Processing*, 2006.

[9] *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/, 2006.