Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2015, 7(3):841-845

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Statistical techniques for the analysis of electroencephalography signals from epileptic patients

R. Harikumar and P. Sunil Kumar

Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam, Erode, Tamil Nadu, India

---

ABSTRACT

The aim of this paper is to explore the statistical techniques used for the analysis of Electroencephalography Signals obtained from Epileptic Patients. The Electroencephalogram (EEG) is actually a measure of the cumulative firing of neurons in various parts of the brain. The EEG contains the information with regard to the changes in the electrical potential of the brain which is obtained from a set of recording electrodes. The study of the statistical analysis techniques paves the way for the easy classification of epilepsy risk levels from EEG signals as a future step.

Keywords: EEG, epilepsy, electrodes

INTRODUCTION

The Electroencephalogram (EEG) is actually a measure of the cumulative firing of neurons in various parts of the brain [1]. The EEG contains the information with regard to the changes in the electrical potential of the brain which is obtained from a set of recording electrodes. Generally the EEG data includes both the standard waveforms and shortly occurring electrical patterns. The standard waveforms may accompany with rapid variations in amplitude, frequency and phase. The electrical patterns such as sharp and spike waveforms and spindles may also be present. EEG patterns can be modified using a wide range of variables including hormonal, biochemical, metabolic, circulatory, neuro electric and behavioral factors. Earlier, just by visual inspection the encephalographer was able to distinguish the normal EEG activity from the abnormal EEG activity. The most important activity detected from the EEG is the epilepsy and it is characterized by the excessive activity by a part or all of the central nervous system. By observing the different EEG waveform patterns the different types of epileptic seizures are characterized. In order to quantify the changes occurring based on the EEG signals, the application of computers has made it possible to effectively apply a host of methods for the real time monitoring and detection of epileptic seizures. The EEG is a vital tool used for the diagnosis, monitoring and managing of neurological disorders related to epilepsy.

2. STUDY OF STATISTICAL PARAMETERS USED FOR THE ANALYSIS OF EEG SIGNALS OF EPILEPTIC PATIENTS

The statistical parameters taken here for the study are Mean, Variance, Standard Deviation, Skewness, Kurtosis, Hurst exponent, Entropy, Mutual Information, Expectation Maximization and Modified Expectation Maximization.

2.1 MEAN

The mean or average is the sum of the total collection or set of numbers divided by the total number of numbers in the collection. The set of results usually is from a survey or an experiment. The mean taken here is arithmetic mean and one should not get confused with geometric mean or harmonic mean.

The average value of a distribution is the mean, which is given by

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

2.2 VARIANCE

The Variance generally shows that how far a set or range or series of numbers is spread out. A Variance is always or mostly non-negative. If the data points are close to the mean, then it signifies it has a small variance. Similarly if the data points are spread out around the mean and from each other, then it indicates a very high variance. In general, the probability distribution has several descriptors and variance is just one of them. The variance is also one of the moments of a distribution. Actually, the variance is a parameter that describes both the actual probability distribution and theoretical probability distribution of observed population of numbers or a not fully observed population from which the sample of numbers can be easily drawn. It is an index of dispersion expressed in the same units as the observations from which it is calculated. It is always given by the following expression.

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x-\mu)^2}{N}$$

2.3 STANDARD DEVIATION

Standard Deviation is calculated as the square root of Variance.

$$StdDeviation = \sqrt{Variance}$$

It generally measures the amount of variation from the average. If the data points are very close to the mean, then it shows a low standard deviation and if the data points are spread out over a large range of values then it indicates a high standard deviation. The Standard Deviation of anything, say random variable, standard data set or probability distribution is nothing but the square root of the variance. The greatest advantage of taking Standard Deviation is that it has the property to express the same units as the data.

2.4 SKEWNESS

It is just the measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. Sometimes the skew values can be positive or either negative or even undefined at times. Interpreting the skew value in a qualitative manner is very complicated and it is significant to note that the relationship of mean and median is not determined by skewness. Skewness has a wide range of benefits in many areas. The indication whether deviations from the mean are going to be positive or negative are made clear by understanding the skewness of the dataset.

When the portion of the frequency polygon to the right of the mean is the mirror image of the portion to the left, the distribution is said to be symmetrical. If more observations lie on one side of the mean than on the other side, the distribution is skewed. The skewness is characterized by the mean of the distribution shifting from the median. The Pearson's coefficient of skewness is given as follows

$$S = \frac{\sum_{i=1}^{N} \frac{(x_i-\mu)^3}{N}}{\left(\sum_{i=1}^{N} \frac{(x_i-\mu)^2}{N}\right)^{3/2}}$$

2.5 KURTOSIS

Kurtosis denotes mainly the peakedness of probability distribution function of real valued random variable. It also denotes the shape of the distribution function and their different ways of quantifying it. Interpretation of kurtosis can be done in several ways and it mainly includes peakedness, tail weight and lack of shoulders. Higher Kurtosis means that the variance is more and it results in infrequent extreme deviations. The most commonly used Kurtosis includes the Pearson's Kurtosis and the excess Kurtosis which generally provides a comparison of the shape of a given distribution to that of the normal distribution. Other types of Kurtosis are also available such as platykurtic distributions where it deals with negative excess kurtosis and leptokurtic distributions where it deals with positive excess kurtosis respectively.

The coefficient of kurtosis is calculated using the formula.

$$K = \frac{\sum_{i=1}^{N} \frac{(x_i - \mu)^4}{N}}{\left(\sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N} \right)^2} - 3$$

(Note: where μ is not known, the sample mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ is used.)

Ideally for a normal distribution, the skewness and kurtosis are zero. Hence, a nonzero value indicates that the distribution is not purely normal Gaussian distribution.

2.6 HURST EXPONENT

As Hurst exponent is a measure of self-similarity, predictability and the degree of long-range dependence in a time-series [2]. It is also a measure of the smoothness of a fractal time-series based on asymptotic behavior of the rescaled range of the process. Hurst's generalized equation of time series, Hurst exponent H is defined as the follows

$$H = \frac{\log(R / S)}{\log(T)}$$

where T is the duration of the sample of data and R/S is the corresponding value of rescaled range. R is the difference between the maximum and minimum deviation from the mean while S represents the standard deviation. Hurst exponent is estimated by plotting (R/S) versus T in log-log axes. The slope of the regression line approximates the Hurst exponent.

2.7 ENTROPY

It is nothing but a measure of uncertainity of a particular random variable. The entropy H(X) for a particular discrete random variable X is defined as follows

$$H(X) = -\sum_{x \in N} p(x) \log p(x)$$

The entropy of X is also expressed as the expected value of $\log \frac{1}{p(X)}$, where X is drawn according to probability mass function p(x).

Thus, $H(X) = E_p \log \frac{1}{p(X)}$

Usually the above definition is related to the definition of entropy in thermodynamics.

2.8 MUTUAL INFORMATION

Mutual Information is nothing but the measure of the amount of information that one random variable contains about another random variable [3]. By having the knowledge about the other random variable, the reduction in the uncertainity of one random variable can be easily done using mutual information. Two random variables X and Y are considered which has a joint probability mass function(x,y) and marginal probability mass function p(x) and p(y) .

The relative entropy between the joint distribution and the product distribution (i.e.) p(x) and p(y) is given as the mutual information I(X; Y) as is expressed mathematically as follows:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

2.9 EXPECTATION MAXIMIZATION

The Expectation Maximization (EM) is often defined as a statistical technique for maximizing complex likelihoods and handling incomplete data problem [5]. EM algorithm generally consists of two steps namely:-*Expectation Step (E Step):* Say for instance consider data x_1 which has an estimate of the parameter and the observed data, the expected value is initially computed easily. For a given measurement y_1 and based on the current estimate of the parameter, the expected value of x_1 is computed as given below:

$$x_1^{[k+1]} = E[x_1 \mid y_1, p^k]$$

This implies, $x_1^{[k+1]} = y_1 \dfrac{1/4}{\frac{1}{4} + \frac{p^{[k]}}{2}}$

Maximization Step (M Step): From the expectation step, we use the data which was actually measured to determine the Maximum Likelihood estimate of the parameter. The set of unit vectors is considered to be as X. Considering x_i X, the likelihood of X is:

P (X | μ, K, μ, k) = P ($x_{1.......}x_n$| μ, k x_i …,x | μ, K x_i, …x|μ,K)

$$= \prod_{i=1}^{n} f(x_i \mid \mu, k) \prod_{i=1}^{n} c_d(k) e^{k\mu^T x_i}$$

The log likelihood of the above equation can be written as:

L (X|μ, k) = ln P (X|μ, k) = n ln c_d (k) + k μ^Tr

where r = $\sum_i x_i$

2.10 MODIFIED EXPECTATION MAXIMIZATION (MEM) ALGORITHM

In this paper, a Maximum Likelihood (ML) approach which uses a modified Expectation Maximization (EM) algorithm for pattern optimization is used [4]. Similar to the conventional EM algorithm, this algorithm alternated between the estimation of the complete log – likelihood function (E – Step) and the maximization of this estimate over values of the unknown parameters (M - Step). Because of the difficulties in the evaluation of the ML function, modifications are made to the EM algorithm.

1. The initial values of the maximum likelihood parameters like mean, covariance and mixing weights are found out.
2. Each x_i to its nearest cluster centre c_k by Euclidean Distance (d) is assigned.
3. In maximization step, use Maximization $Q(\theta, \theta')$ .The likelihood function is written as:-

$$Q(\theta^{j+1}, \theta^{j}) = \max Q(\theta^{j}, \theta), \theta^{j+1} = \arg\max Q(\theta, \theta^{j})$$

$$d(p,q) = d(p,q) = \sqrt{\sum_{i+1}^{n} (q_i - p_i)^2}$$

4. The iterations are repeated and the loop should not be stopped until $\|\theta^{i+1} - \theta^{i}\|$ becomes small enough.

The algorithm ends when the difference between the log likelihood for the previous iteration and current iteration fulfills the tolerance level. The method of maximum likelihood corresponds for many well – known statistical estimation methods. For instance , one may be very interested to learn about the heights of adult female giraffes in a particular zoo, but it might be unable because of time and permission constraints, to measure the height of each and every single giraffe in that population. If the heights are assumed to be normally Gaussian distributed with some unknown mean and variance, then the mean and variance can be estimated mostly with Maximization Likelihood Equalization (MLE) by just knowing the heights of some of the samples of the overall population.

CONCLUSION

The aim of this paper is to explore the statistical techniques used for the analysis of Electroencephalography Signals obtained from Epileptic Patients. The above mentioned statistical techniques are very much useful for the analysis of dimensionally reduced EEG data signals and also for various other statistical analyses. Thus the study of the statistical techniques for the analysis of electroencephalography signals from epileptic patients holds a great significance in the field of bio-signal processing.

REFERENCES

[1] H Adeli; Z Zhou; N Dadmehr, *J. Neurosci. Methods*, **2003**,123(1,) 69-87.
[2] N Arunkumar, VS Balaji, Subhashree Ramesh, Sharmila Natarajan, Vellanki Ratna Likhita & Sivakama Sundari, 'Automatic Detection of Epileptic Seizures Using Independent Component Analysis Algorithm', *IEEE–International Conference on Advances in Engineering, Science and Management* (ICAESM), **2012**, 542-544
[3] RA Choudrey; SJ Roberts; Bayesian; 'ICA with Hidden markov model sources', *International Conference on Independent Component Analysis*, **2003**, Japan, 809-814.
[4] KH Finley, JB Dynes, *Brain*, **1942,** vol.65, 256-265
[5] PE McSharry, *Medical*, **2001**, vol. 40, 12-26.