I. Time Series and Stochastic Processes

Purpose of this Module

- Introduce time series analysis as a method for understanding real-world dynamic phenomena
- Define different types of time series
- Explain the need for understanding stochastic processes in time series analysis

Definition: a time series is a variation with time in amplitude and polarity of a measured physical quantity.

Definition: a continuous time series is one that exists at all instants of time during which it occurs.

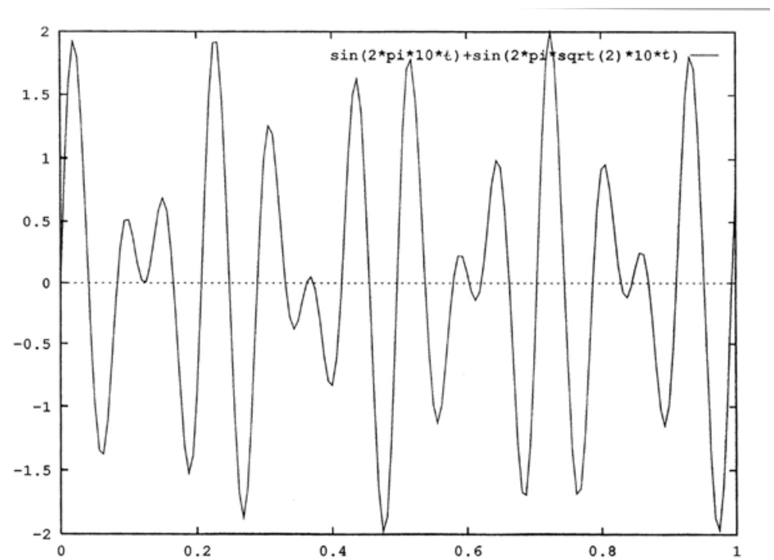
Definition: a discrete time series is one which exists only at discrete instants of time.

Definition: a repetitive time series is one which contains a pattern that recurs for all time over which the time series exists.
Example: the electrocardiogram is a repetitive time series. It repeats with every heartbeat.

Definition: a periodic time series is a repetitive time series in which the repetition occurs at uniformly spaced time intervals.
The electrocardiogram is near-periodic, but no function of biological origin is exactly periodic unless driven by external periodic stimulation.

Definition: a deterministic time series is one which can be expressed explicitly by an analytic expression. It has no random or probabilistic aspects. It. In mathematical terms, it can be described exactly for all time in terms of a Taylor series expansion provided that all its derivatives are known at some arbitrary time. Its past and future are completely specified by the values of these derivatives at that time. If so, then we can always predict its future behavior and state how it behaved in the past.

Example: x(t) = sin(2pi*f*t) + sin(2pi*sqrt(2)*f*t) is a deterministic function of time. Although its 2 components are periodic, this function is aperiodic since it is impossible to find a finite value of t corresponding to a repetition period.



A <u>nondeterministic time series</u> is one which cannot be described by an analytic expression. It has some random aspect that prevents its behavior from being described explicitly.

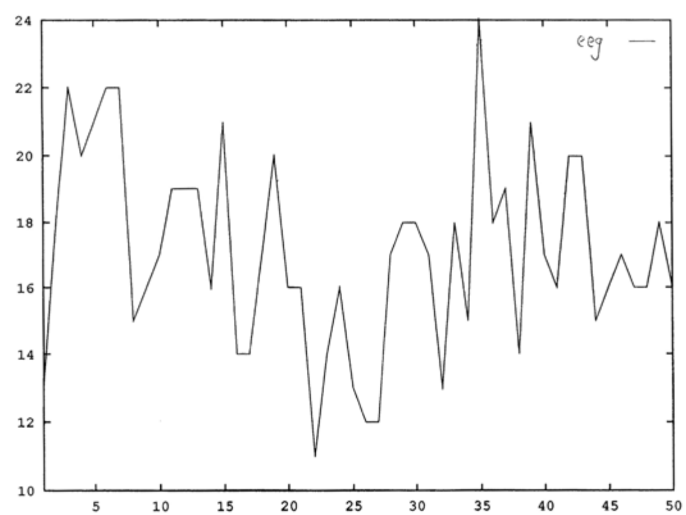
A time series may be nondeterministic because:

(1) all the information necessary to describe it explicitly is not available, although it might be in principle.

or

(2) the nature of the generating process is inherently random.

Since nondeterministic time series have a random aspect, they follow probabilistic rather than deterministic laws. Random data are not defined by explicit mathematical relations, but rather in statistical terms, i.e. by probability distributions and averages of various forms, such as means and variances.



An EEG time series

Definition of the Stochastic Process

Nondeterministic time series may be analyzed by assuming they are the manifestations of stochastic (random) processes.

Definition: a stochastic (random) process is a statistical phenomenon consisting of a collection of random variables ordered in time.

The stochastic process evolves in time according to probabilistic laws.

Features of the Stochastic Process

1. The stochastic process is a model for the analysis of time series.

2. The stochastic process is considered to generate the infinite collection (called the ensemble) of all possible time series that might have been observed. Every member of the ensemble is a possible realization of the stochastic process.

3. The ensemble of a stochastic process is a statistical population. An observed time series is considered to be one realization of a stochastic process. A set of observed time series is considered to be a sample of the population.

Stationarity

Definition: a stationary stochastic process is one whose ensemble statistics are the same for any value of time.

A time series is said to be stationary if there is no systematic change in mean (no trend), if there is no systematic change in variance, and if it contains no strictly periodic variations.

The assumption of stationarity is often important for statistical analysis.

Strict and Weak Stationarity

Definition: A stochastic process is strictly stationary if all statistical measures on it are stationary, i.e. do not depend on t.

The distribution of a strictly stationary stochastic process is the same at time t as at any other time $t + \tau$.

Definition: a stochastic process, denoted X(t), is weakly stationary (or second-order stationary) if both of the following conditions are satisfied:

1. $E[X(t)] = \mu$
2. $cov[X(t), X(t + \tau)] = Cxx(\tau)$

   where $cov(X,Y)$ is defined as $E[(X - \mu_X) (Y - \mu_Y)]$

The first condition states that the expected value of X(t) is equal to the ensemble mean regardless of t. This condition means that the mean of the process does not depend on time, i.e. it is stationary.

The second condition states that the autocovariance of X(t) also does not depend on time, only on time-difference (τ).
(When $\tau = 0$, the autocovariance reduces to the variance. Therefore, this condition also means that the variance is stationary.)

Generally, many of the useful properties of stationary stochastic processes depend only on the structure of the process as specified by the first and second moments (i.e. the first and second conditions are satisfied).

Therefore, the weaker definition of stationarity is commonly taken to be sufficient.

If the process is Gaussian, we will see that higher-order moments are all zero. Therefore, second-order stationarity implies strict stationarity for a Gaussian process.

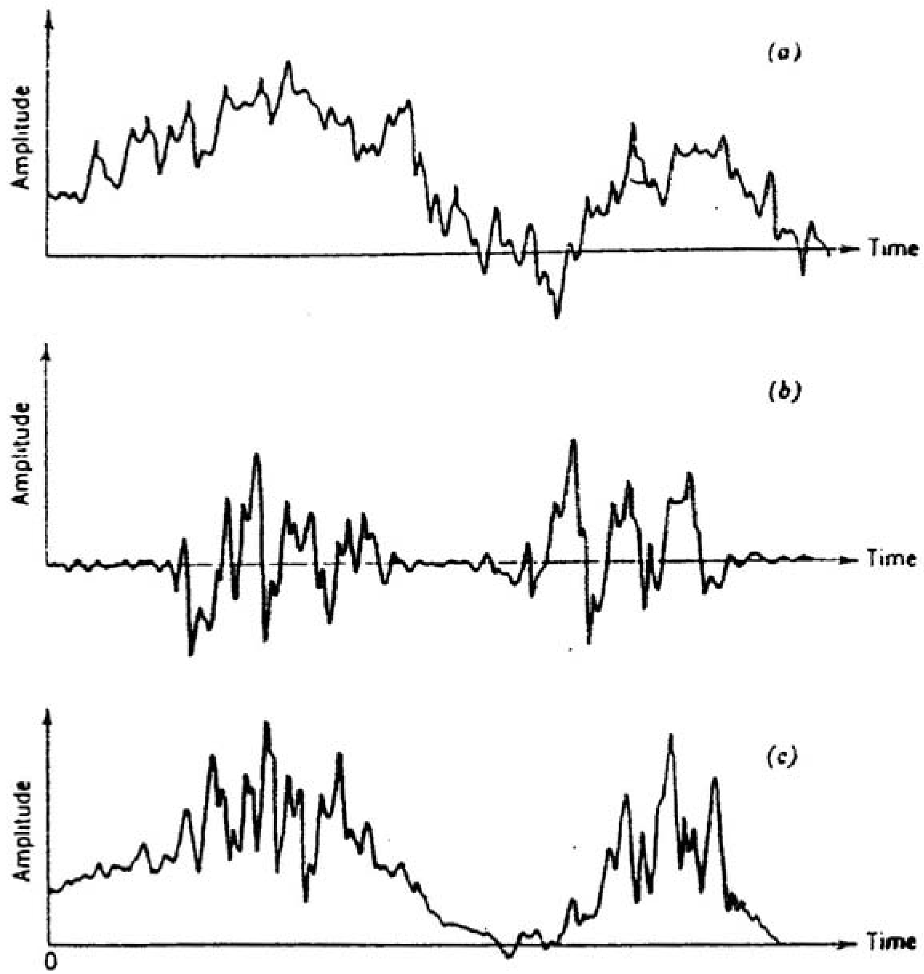


Fig. 5. *Examples of non-stationary data.* **(*a*)** Time-varying mean value. **(*b*)** Time-varying mean square value. (*c*) Time-varying mean and mean square value. The amplitude distribution has negative skewness in (*a*), positive excess (Kurtosis > 3) in (*b*) and negative skewness as well as positive excess in (*c*). (Bendat and Piersol 1966)

Ergodicity

Definition: an ergodic stochastic process is one whose time statistics equal its ensemble statistics.

In experiments carried out in the physical world, one can usually collect time series data for only limited lengths of time. If an ergodic stochastic process is generating the time series, then the statistical behavior of one time series, if observed long enough, will be characteristic of the entire ensemble of realizations.

In other words, we would like to obtain consistent estimates of the properties of a stationary stochastic process from a single finite-length realization. Theorems have been proven, called ergodic theorems, showing that, for most stationary processes likely to be met in practice, the statistics of an observed time series converge to the corresponding population statistics.

Describing a Stationarity Stochastic Processes

1. A stationary stochastic process is described by its mean value and the distribution around the mean.

2. If that distribution is Gaussian and the observed values of the stochastic process are mutually independent, then the mean and variance are sufficient descriptors of the stochastic process.

3. If the distribution is non-Gaussian, then higher-order moments are also needed to describe the process.

4. If the observed values of the stochastic process show interdependence, then the autocovariance is also needed to describe the process.

The Ensemble Mean of a Stationary Stochastic Process

We next consider the statistics of a stationary stochastic process. We will assume that it is ergodic.

We begin with the ensemble mean (a first-order statistic), and then consider higher-order statistics (second-order and above).

The definition of the ensemble mean depends on the expectation operation:

Definition: Let X be a random variable, and let $\Psi(X)$ represent any quantity derived from it. The expectation of Ψ is :

$$E[\Psi(X)] \equiv \int_{-\infty}^{\infty} \Psi(X)\, p(X)\, dX$$

where p(X) is the probability density function of X.

Now consider the ensemble of realizations at a particular time t_0, denoted $\{x(t_0)\}$.

The expected value of $x(t_0)$ is $E[x(t_0)]$.

That is, the expected value is $E[\Psi(x)]$ where $\Psi(x) = x(t_0)$.

This expected value is called the ensemble mean, and is given by:

$$E[x(t_0)] = \mu(t_0) \equiv \int_{-\infty}^{\infty} x(t_0)\, p(x)\, dx$$

where $x(t_0)$ is the value of one realization of the random process X at time t_0, and $p(x)$ is the probability of occurrence of that realization. Note that the integration extends over the entire (infinite) ensemble of realizations.

1. Since the stochastic process is stationary, $\mu(t_0)$ is independent of time. That is:

$$\mu(t_0) = \mu(t_1) = \mu_x$$

2. Since the stochastic process is ergodic:

$$\mu_x = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t)dt$$

This says that the limit of the average over time interval T of one realization, as T approaches infinity, equals the ensemble mean.

Thus the ensemble mean can be determined by the time average based on a single realization.

Higher-Order Statistics

The ensemble mean is a first-order statistic. We now define higher-order statistics of a stochastic process by introducing the concept of moments. (As with the ensemble mean, we assume stationarity and ergodicity here.)

Given a random variable X,

the k^{th} moment of X is the value $E[X^k]$,

and the k^{th} central moment of X is the value $E[(X – E[X])^k]$.

The <u>mean</u> of the random variable is equal to its first moment. (Note that the first central moment is equal to zero):

$$m_1 = \overline{X} = E[X]$$

The <u>variance</u> of the random variable is equal to its second central moment:

$$m_2 = \sigma_x^{\ 2} = E[(X - \overline{X}\ )^2\ ]$$

The third central moment is called skewness:

$$m_3 = E[(X - \overline{X})^3]$$

It is used to derive the skewness factor:

$$\beta_1 = \frac{m_3}{(m_2)^{\frac{3}{2}}}$$

which measures the deviation from symmetry.

The fourth central moment is called kurtosis:

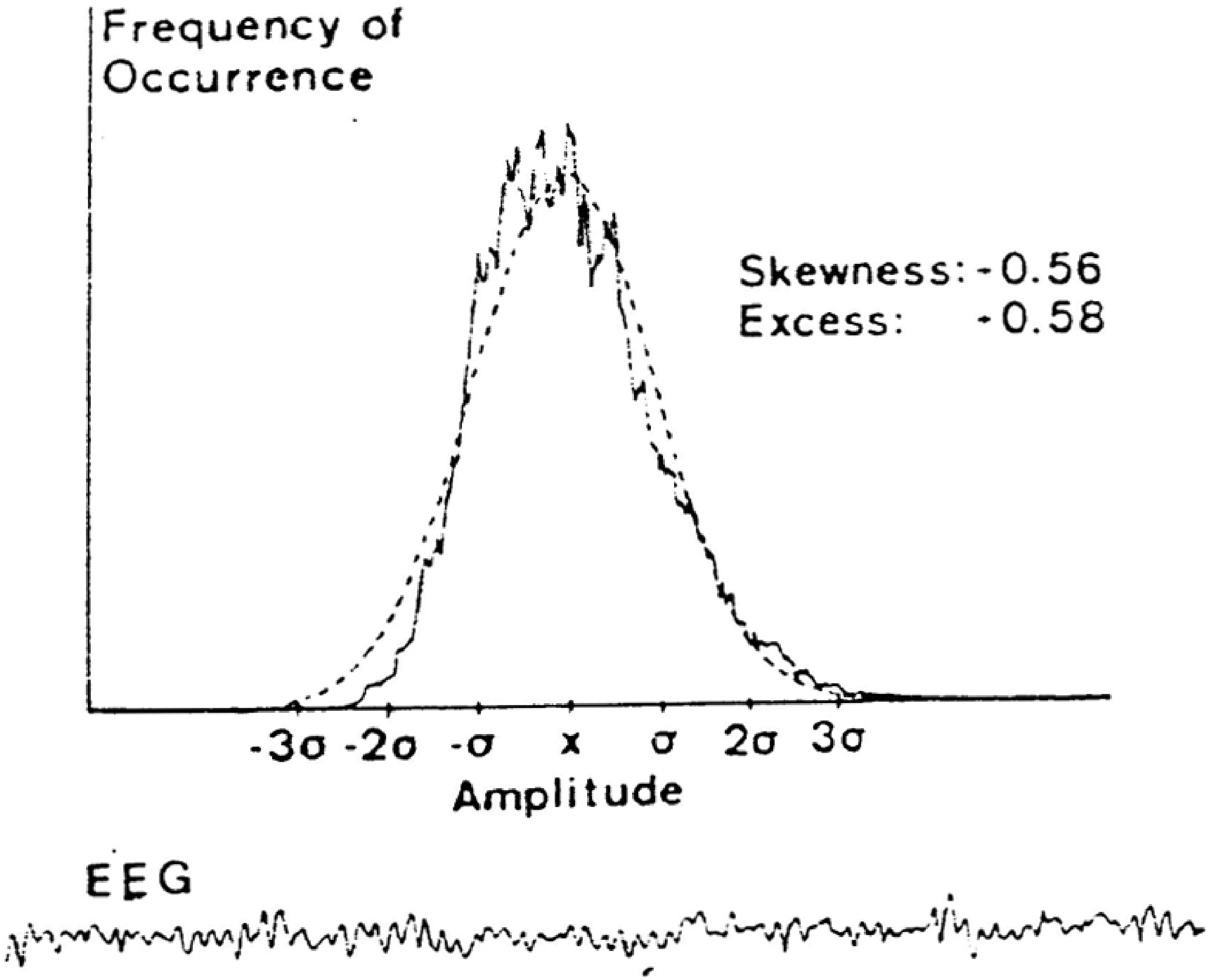
$$m_4 = E[(X - \overline{X})^4]$$

It is used to derive the kurtosis factor:

$$\beta_2 = \frac{m_4}{(m_2)^2}$$

which measures the relative peakedness or flatness of the distribution.

The Gaussian distribution has a kurtosis factor of 3, so the quantity $\beta_2 - 3$ is called excess of kurtosis.


Frequency of Occurrence
Skewness: -0.56
Excess: -0.58
-3σ -2σ -σ x σ 2σ 3σ
Amplitude
EEG

Lessons for Analysis of Experimental Time Series

Real-world (experimental) time series usually are:

1. Non-deterministic (cannot be described analytically)
2. Non-stationary (statistics differ at different times)
3. Non-Gaussian (mean & variance not sufficient statistics)
4. Interdependent (amplitude values at different time points are correlated)

These problems can be dealt with by the following:

1. Use statistical descriptions
2. Separate time series data into “quasi-stationary” segments
3. Consider higher-order moments
4. Consider the autocovariance function