INTRODUCTION TO RANDOM SIGNALS

MULTIVARIATE PROBABILITIES AND TIME SERIES ANALYSIS

P.R. WHITE

MULTIVARIATE PDFS

CORRELATION AND INDEPENDENCE

RANDOM TIME HISTORIES

STATIONARITY AND ERGODICITY

CORRELATION FUNCTIONS

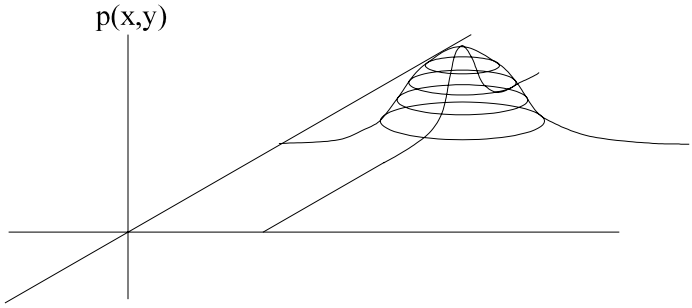
 © ISVR 2007

MULTIVARIATE PROCESSES

So far we have discussed univariate processes, *i.e.* experiments that yield a single value.

A multivariate process is one that generates multiple outputs, *e.g.* rolling several dice, they are defined through their joint pdfs, we shall often consider two variables (the bivariate case).

Analogously to the univariate case the bivariate pdf, $p(x,y)$, can be thought of as the limit of a normalised 2-D histogram.



The basic property of joint distributions is:

$$P\{a \le X < b, c \le Y < d\} = \int_a^b \int_c^d p(x,y)\,dxdy$$

 © ISVR 2007

INDEPENDENCE

Often one is interested in whether two RVs are inter-related or if they are independent, *e.g.* in medicine: "Is a patient's outcome independent of whether a particular drug, or a placebo, administered?"

Consider rolling 2 dice (one red, one blue), if X=score on the red dice, Y=score on the blue dice and Z=sum of the scores on the two dice ($X+Y$), then:

X and Y are independent, since knowing the score on the *blue* dice tells you nothing about the score on the *red* dice.

X and Z are dependent, since knowing the score on the red dice allows you to narrow the range of possible values of Z.

If two processes are statistically independent then their joint pdf, $p(x,y)$, is the product of the individual pdfs, *i.e.*

$$p(x,y) = p(x)\,p(y)$$

Recall Bayes' Rule: $p(y \mid x) = \dfrac{p(x,y)}{p(x)}$

For independent variables: $p(y \mid x) = \dfrac{p(x)\,p(y)}{p(x)} = p(y)$

This states that knowing x does not influence the value of y.

 © ISVR 2007

MARGINALISATION

Often one seeks to estimate the pdf of a single variable, say x, when the joint distribution, say $p(x,y)$, is known.

To achieve this one integrates across the variables of no interest, for example:

$$p(x) = \int p(x,y)\,dy$$

Multi-dimensional integrals can be used to integrate out (marginalise) across many variables.

$$p(x_1) = \int_{x_2} \cdots \int_{x_M} p(x_1, x_2, \cdots x_M)\,dx_2 \cdots dx_M$$

For independent variables (in the bivariate case) we have

$$p(x) = \int p(x,y)\,dy = \int p(x)\,p(y)\,dy = p(x)\int p(y)\,dy = p(x)$$

 © ISVR 2007

COVARIANCE AND CORRELATIONS

The interdependence of two RVs is commonly measured via their correlation and/or their covariance.

The **covariance** is the second central cross moment:

$$E\left[\left(X-\mu_x\right)\left(Y-\mu_y\right)\right]$$

where μ_x and μ_y are the means of X and Y, respectively.

The **correlation** is second uncentred cross moment:

$$E\left[XY\right]$$

In most engineering applications one commonly assumes $\mu_x=\mu_y=0$, so the correlation and covariance are identical.

The engineering convention is to refer to correlation – which assumes the means are zero.

If the data has a non-zero mean then one "pre-processes" the data by estimating the mean and then subtracting it from all of the measurements.

 © ISVR 2007

PROPERTIES OF CORRELATIONS

The correlation (covariance) is bounded in magnitude:

$$E\left[XY\right]^2 \leq \sigma_x^2 \sigma_y^2$$

where σ_x^2 is the variance (second moment) of X and σ_y^2 is the variance of Y.

One can define a normalised measure of correlation, termed the correlation coefficient, r, as

$$r = \frac{E\left[XY\right]}{\sigma_x \sigma_y}$$

If $|r|=1$ then X and Y are "perfectly" related $\{X=aY\}$.
If $r=0$ then X and Y are "perfectly" uncorrelated.

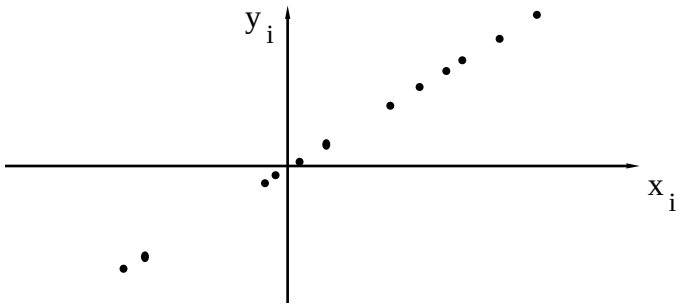
Note that if $Y=X^3$, then X and Y contain the same "information" – knowing X one can find Y (and vice versa), but the correlation between X and Y is normally low since correlation assumes a *linear* relationship between X and Y.

 © ISVR 2007

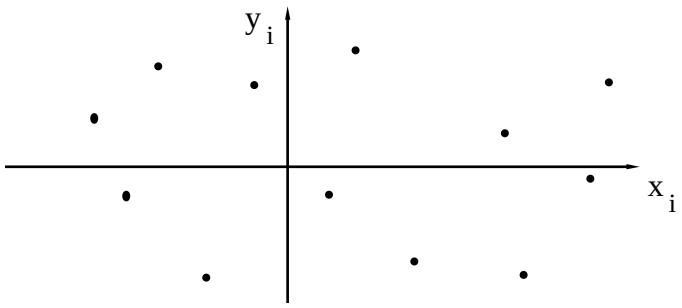
INTERPRETING THE CORRELATION COEFFICIENT

The following are 'scatter diagrams' obtained by plotting realisations of X against Y.

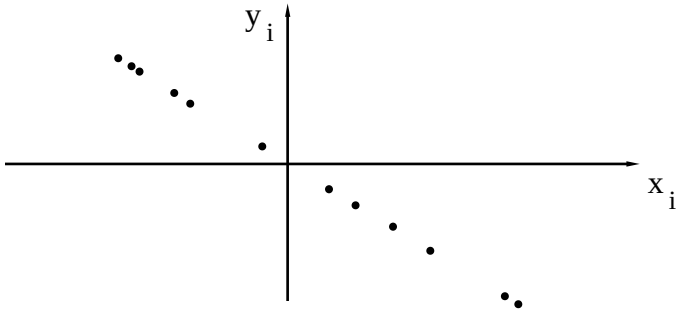
If r=1 then $Y=aX$ ($a > 0$)



If r=0 then X and Y are statistically independent



If r=-1 then $Y=aX$ ($a < 0$)



 © ISVR 2007

MULTIDIMENSIONAL GAUSSIAN DISTRIBUTIONS

The most commonly considered multivariate model is the joint Gaussian distribution.

We shall consider the distribution of the set of variables x_1, x_2, …, x_N which we shall denote $p(x_1, x_2, \cdots, x_N)$.

It is convenient to write the variables in the form of a column vector $\mathbf{x} = [x_1, x_2 . \cdots, x_N]^t$.

The joint Gaussian pdf is then:

$$p(x_1, x_2, \cdots, x_N) = p(\mathbf{x}) = \frac{1}{\sqrt{2\pi|\mathbf{R}|}} e^{-(\mathbf{x}-\boldsymbol{\mu})^t \mathbf{R}^{-1}(\mathbf{x}-\boldsymbol{\mu})/2}$$

where $\mathbf{R}$ is the covariance matrix and $\boldsymbol{\mu}$ is the mean vector.

$$\boldsymbol{\mu} = E[\mathbf{x}] \qquad \mathbf{R} = E\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^t\right]$$

In many applications we shall assume that $\boldsymbol{\mu}=\mathbf{0}$ and the covariance matrix is equivalent to the correlation matrix

$$\mathbf{R} = E\left[\mathbf{x}\mathbf{x}^t\right] = \left\{r_{ij}\right\}_{ij} = \left\{E\left[x_i x_j\right]\right\}_{ij}$$

Gaussian variables have the unique property that if they are uncorrelated then they are necessarily independent.

 © ISVR 2007

UNCORRELATED GAUSSIAN RVS

By definition if variables are uncorrelated (and zero mean) then

$$E\left[x_i x_j \right] = 0 \qquad i \neq j$$

$$E\left[x_i x_j \right] = \sigma_i^2 \qquad i = j$$

In which case the correlation matrix **R** is diagonal

$$\mathbf{R} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix}$$

So the pdf is

$$p(\mathbf{x}) = \frac{1}{\sqrt{2\pi \prod_{i=1}^{N} \sigma_i^2}} e^{-\frac{1}{2}\sum_{i=1}^{N} x_i^2 / \sigma_i^2}$$

Further, if the processes all have the same variance, σ^2, this simplifies further

$$p(\mathbf{x}) = \frac{1}{\sigma^N \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N} x_i^2}$$

 © ISVR 2007

BIVARIATE GAUSSIAN VARIABLES

One can generate any bivariate Gaussian random variable using uncorrelated RVs.

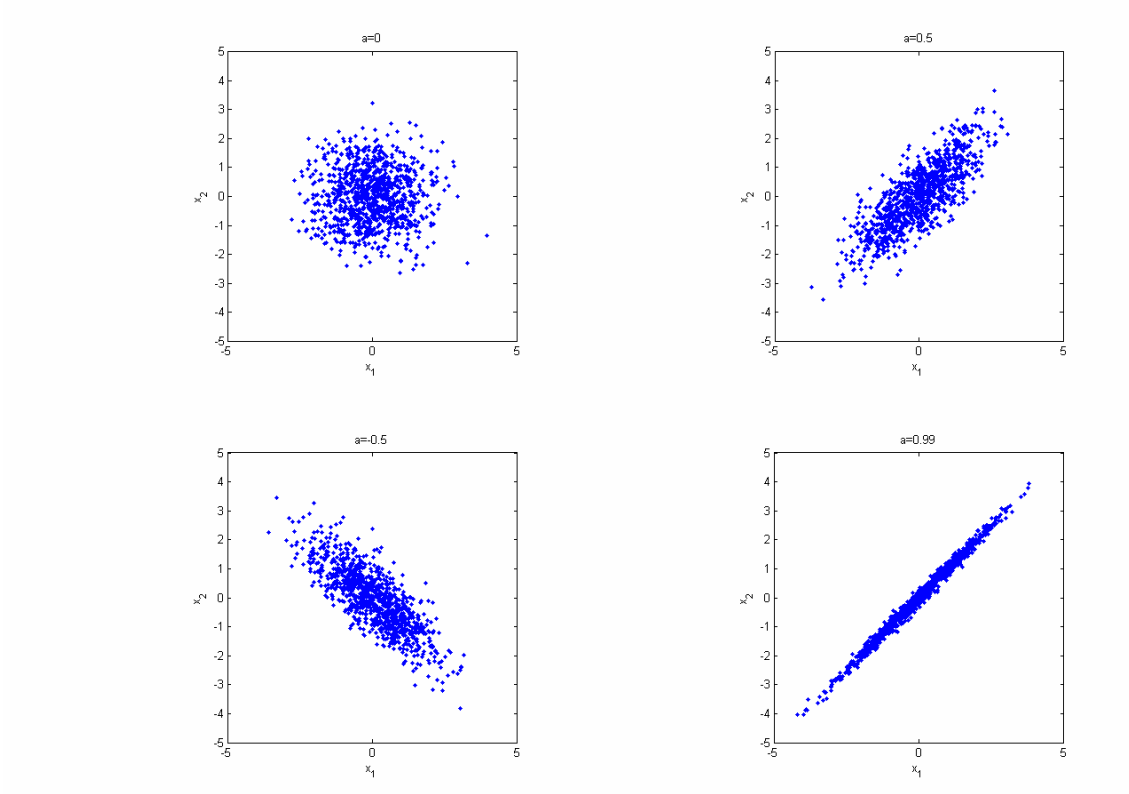
If **u** is a vector of 2 uncorrelated Gaussian RVs, with zero mean and variance 1 then

$$E\left[\mathbf{u}\mathbf{u}^t\right] = \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Leading to

$$\mathbf{R} = E\left[\mathbf{x}\mathbf{x}^t\right] = E\left[A\mathbf{u}\mathbf{u}^t A^t\right] = AE\left[\mathbf{u}\mathbf{u}^t\right]A^t = AA^t$$

Consider the particular case $A = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}$



Note the shape of these plots is always ellipsoidal.

 © ISVR 2007

RANDOM TIME HISTORIES

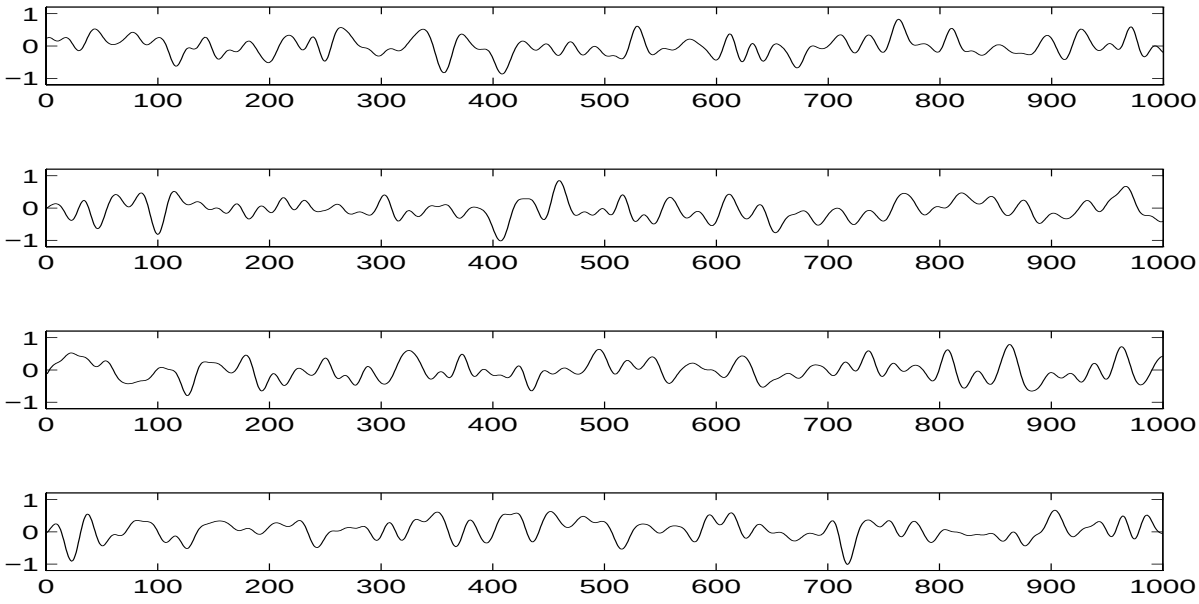
We are now going to be concerned with random variables which are also functions of time, *e.g.* $X(t)$.

Now we can talk about quantities such as:

$$P\left\{X\left(t\right) > a\right\} \text{ or } E\left[X\left(t\right)\right]$$

Each realisation of this random process, $x(t)$, should be different.

Consider a machine which measures thermal noise. The random variable is $X(t)$. Every time the machine is switched on a signal $x(t)$ may be observed, each observation will be different but the statistics of each should be the same.



 © ISVR 2007

STATISTICAL DESCRIPTIONS OF RANDOM TIME HISTORIES

A pdf can be defined for every point in time within a random time series. So one can define:

$p(x,t)$ which is the pdf of $X(t)$.

e.g. The probability of a signal $X(t)$ lying between a and b at a time $t=T$, is given by:

$$\int_a^b p(x,T)\,dx$$

Similarly a joint pdf between two points in the time series can be defined:

$p(x,t_1,t_2)$ which is the joint pdf of $X(t_1)$ and $X(t_2)$

(Joint pdfs between more than 2 points can also be defined)

Moments can be used as partial descriptions of the pdfs and joint pdfs.

 © ISVR 2007

STATIONARITY

A signal is said to be Strict-Sense Stationary (SSS) if its statistics are unaltered by shifts of the time axis.

$$\Rightarrow \quad p(x,t+T)=p(x,t)$$

and $$p(x,t_1+T,t_2+T)=p(x,t_1,t_2) \quad \{\text{and higher orders}\}$$

This implies that the signal has the same statistical structure throughout, *i.e.* it is homogeneous.

In practice it is difficult to verify that a signal is SSS. A lesser condition of Wide-Sense Stationarity (WSS) is normally taken as being sufficient.

A signal is WSS if

$E[\ X(t)\ ]$ = Constant

$E[\ X(t+t_1)X(t+t_2)\ ]$ is only dependent upon t_1-t_2.

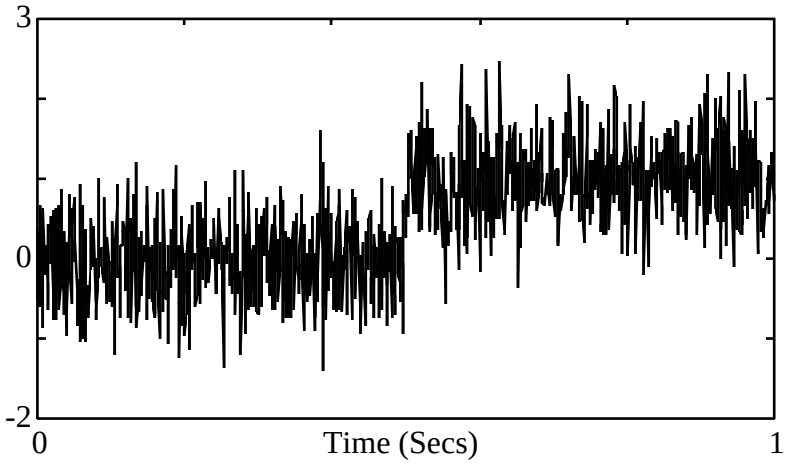
The later condition says that the correlation between two points only depends upon their separation in time.

 © ISVR 2007

SIMPLE EXAMPLES OF NON-STATIONARITIES

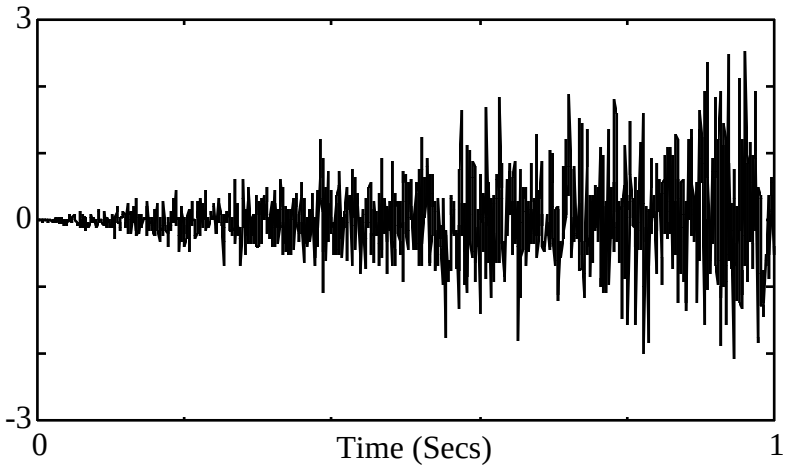
A Non-Stationary Mean

e.g. Measurement of sea levels, as the tide comes in.



Non-Stationary Variance

e.g. Measurements of sea level, when a storm occurs.



 © ISVR 2007

ENSEMBLES

An Ensemble is an infinite collection of statistically identical trials.

e.g. It is an infinite set of different time histories from our thermal noise machine.

When we talk of expectations, say $E[\ X(t_1)\ ]$, the averaging occurs across such an ensemble, and so is referred to as 'ensemble averaging'.

e.g. To estimate $E[\ X(10)\ ]$ using an ensemble average:
One has to turn the machine on wait 10 secs and take a reading, then turn the machine off and repeat, to get the readings $x_i(10)$, where $x_i(10)$ is the reading on the ith trial after 10 secs. These readings are then averaged.

$$E[\ X(10)\ ] = \underset{N \to \infty}{Lim} \frac{1}{N} \sum_{i=1}^{N} x_i(10)$$

In practice one rarely has more than one recording of an event. The problem is then how can we perform averaging if there is no ensemble?

 © ISVR 2007

ERGODICITY

Definition:

> *If the stochastic process is such that an ensemble average maybe replaced by a time average then the process is said to be ergodic.*

All ergodic processes are stationary, but not all stationary processes are ergodic.

For an ergodic process, when only one measurement, $x(t)$, is available one can use expressions like:

$$E\left[X\left(t\right)\right] = \underset{T \to \infty}{Lim} \frac{1}{T} \int_0^T x\left(t\right) dt$$

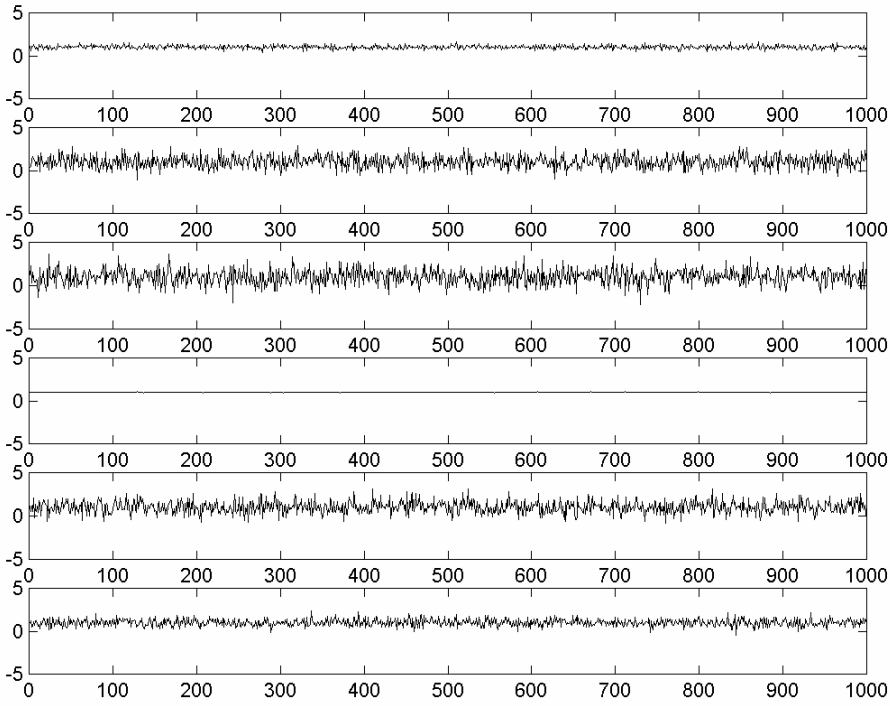
If a process is ergodic then a single time history is sufficient to estimate the statistics of a signal.

In practice one nearly always assumes that (wide-sense) stationarity implies ergodicity.

 © ISVR 2007

EXAMPLE OF A STATIONARY, NON-ERGODIC, PROCESS

Consider a machine that when switched on produces a random output whose mean is always 1 by whose variance is chosen randomly, say between 0 and 1.



This process is stationary (the statistics at each point in time are the same).

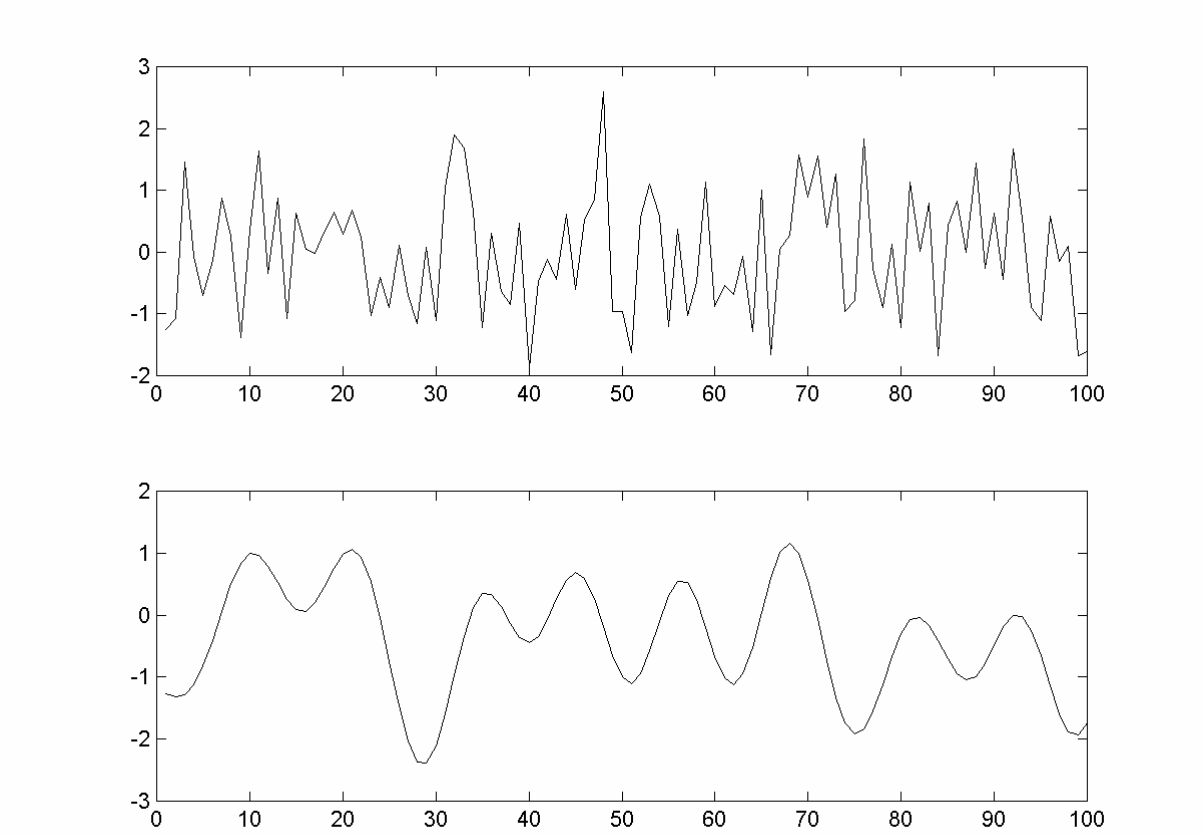
Given one time series you can estimate the true mean, since in each trace it is one, so this process is ergodic for the mean.

However given one trace you would not compute the overall variance of the process, so it is not ergodic for the variance.

 © ISVR 2007

IN WHAT WAYS CAN TWO TIME SERIES DIFFER?

Consider the following two random time histories:



The two signals look different (and indeed sound different if played over a loudspeaker).

Both processes have the same mean (0) and variance (1).

What is the different then?
In the lower trace the signal "changes less" from sample to sample, knowing the value at time *n* allows one to make confident predictions about the value at time *n*+1. Whereas in the upper trace the signal jumps about more erratically.

 © ISVR 2007

CHARACTERISING RANDOM PROCESSES

To quantify the "erraticness" of a random process one can define the so-called correlation function.

The previous example illustrated that for some random signals the value of the process at time n+1 is heavily dependent on the value of the process at time n, whilst in some processes $x(n)$ and $x(n+1)$ are relatively independent.

Let us revert to considering a continuous time signal. One seeks to quantify how independent $x(t-\tau)$ and $x(t)$ are.

From earlier discussions we know that correlations can be used to quantify the degree of dependence (or otherwise), so one can define:

$$r(t,\tau) = E\left[x(t-\tau)x(t)\right]$$

which is referred to as the (auto-)correlation function.

 © ISVR 2007

CORRELATION FUNCTIONS

The correlation function, $r(t,\tau)$, is defined as the correlation measured between two points, $x(t-\tau)$ and $x(t)$.

For a stationary process then by definition the statistics of the signal do not depend on t and so the correlation function can be written as:

$$r(\tau) = E\left[x(t-\tau)x(t)\right]$$

This means that for a stationary process the correlation function depends on the single variable, τ.

Since the correlation function defines the correlation between any two points separated by τ seconds, we also have:

$$r(-\tau) = E\left[x(t+\tau)x(t)\right] = r(\tau)$$

So that the correlation function is necessarily symmetric.

 © ISVR 2007

SUMMARY OF THE PROPERTIES OF CORRELATION FUNCTIONS OF STATIONARY PROCESSES

(i) $r(\tau)$ is a function only of the lag.

(ii) The correlation function is symmetric:

$$r(-\tau) = r(\tau)$$

(iii) The correlation function attains its maximum value at zero

$$r(0) \geq |r(\tau)| \qquad \text{for all } \tau$$

(iv) Evaluating the correlation function at zero lag gives the variance of the process.

$$r(0) = E[\, X(t)^2 \,] = \sigma^2$$

 © ISVR 2007

EXAMPLE

Consider a sine wave:

$$x(t) = A\sin(2\pi f_0 t + \phi)$$

where ϕ is the initial phase chosen randomly between [0,2π].

The correlation function for this signal is:

$$r(\tau) = \frac{A^2}{2}\cos(2\pi f_0 \tau)$$

Proof using ensemble averages:

$$E\left[x(t)x(t-\tau)\right] = A^2 E\left[\sin(2\pi f_0 t + \phi)\sin\left(2\pi f_0 (t-\tau) + \phi\right)\right]$$

$$= \frac{A^2}{2}E\left[\cos(2\pi f_0 \tau)\right] + \frac{A^2}{2}E\left[\cos\left(2\pi f_0 (2t-\tau) + 2\phi\right)\right]$$

The averaging occurs over the phase angle ϕ so this becomes:

$$\frac{A^2}{2}\cos(2\pi f_0 \tau)$$

Proof using time averages:

$$\frac{1}{2T}\int_{-T}^{T} x(t)x(t-\tau)\,dt = \frac{1}{2T}\int_{-T}^{T} A^2\sin(2\pi f_0 t + \phi)\sin\left(2\pi f_0 (t-\tau) + \phi\right)dt$$

$$= \frac{A^2}{2}\frac{1}{2T}\int_{-T}^{T}\cos(2\pi f_0 \tau)\,dt + \frac{A^2}{2}\frac{1}{2T}\int_{-T}^{T}\cos\left(2\pi f_0 (2t-\tau) + 2\phi\right)dt$$

The second term goes to zero as T increases and the first integral is cos(2$\pi f_0\tau$) for all T and the same result is obtained.

Since time and ensemble averages give the same answer we can conclude that is signal is ergodic w.r.t. $r(\tau)$.

 © ISVR 2007

ESTIMATION OF THE CORRELATION FUNCTION

When computing the correlation function from a time series one almost always is forced to assume ergodicity and use a time average.

Assume that a record of $x(t)$ is available in the interval $0 \le t \le T$.

The correlation function can be estimated using, for example:

$$\hat{r}(\tau) = \frac{1}{T}\int_0^T x(t)\,x(t-\tau)\,dt$$

where the $\hat{r}(\tau)$ is an estimate of the correlation function. (The 'hat' notation emphasises that $\hat{r}(\tau)$ is only an estimate of $r(\tau)$ and the two will, in general, not be equal.)

 © ISVR 2007

RELATIONSHIP WITH CONVOLUTION

The time averaged correlation function is very closely related to convolution.

Recall the convolution operation is defined as:

$$h(t) * x(t) = \int h(t-\tau)\, x(\tau)\, d\tau$$

If one sets $h(t)=x(t)$, and renaming the time axis τ then:

$$x(\tau) * x(\tau) = \int x(\tau-u)\, x(u)\, du$$

Compare this with the time averaged correlation function (the dummy variable in the integral has been changed to u).

$$\hat{r}(\tau) = \frac{1}{T}\int_0^T x(u)\, x(u-\tau)\, du$$

This allows one to write:

$$\hat{r}(\tau) = \frac{1}{T}\, x(\tau) * x(-\tau)$$

So the time averaged correlation function is a scaled version the convolution of the signal with itself reversed in time.

 © ISVR 2007

FREQUENCY DOMAIN REPRESENTATION

Recall that the Fourier transform of a convolution can be readily expressed through the frequency domain:

$$\mathbb{F}\left\{h(t) * x(t)\right\} = H(f)\,X(f)$$

where $H(f)$ and $X(f)$ are the Fourier transforms of $h(t)$ and $x(t)$ respectively and $\mathbb{F}\{\ \}$ denotes the Fourier operator.

Also it is also well known that:

$$\mathbb{F}\left\{x(-t)\right\} = X^{*}(f)$$

where * denotes complex conjugation.

Hence:

$$\hat{r}(\tau) = \frac{1}{T}\mathbb{F}^{-1}\left\{X(f)\,X^{*}(f)\right\} = \frac{1}{T}\mathbb{F}^{-1}\left\{\left|X(f)\right|^{2}\right\}$$

That is to say that the time integrated correlation function can be computed by inverse Fourier transforming the squared magnitude of the Fourier transform of $x(t)$ and scaling by $1/T$.

 © ISVR 2007

DIGITAL IMPLEMENTATION IN THE FREQUENCY DOMAIN

The previous scheme offers one approach to implementing the estimation of correlation functions for digital signals.

1) Fourier transform $x(n)$ using an FFT to give $X(k)$
2) Form the squared magnitude of $X(k)$
3) Inverse FFT $\left|X\left(k\right)\right|^2$

Some care has to be exercised when implementing such a scheme to avoid *wrap-around* effects:-

If two sequences $x(n)$ and $h(n)$ of lengths L_1 and L_2 are convolved to form $y(n)$ then the length of $y(n)$ is L_1+L_2-1.

Hence one expects $\hat{r}\left(\tau\right)$ to be $2N-1$ samples long if $x(n)$ is N samples in duration.

This points to the flaw in the above scheme, namely it only produces an output with N samples.

The solution is simple, at step 1) you need to ensure that the data is zero-padded to a length of at least $2N-1$.

 © ISVR 2007

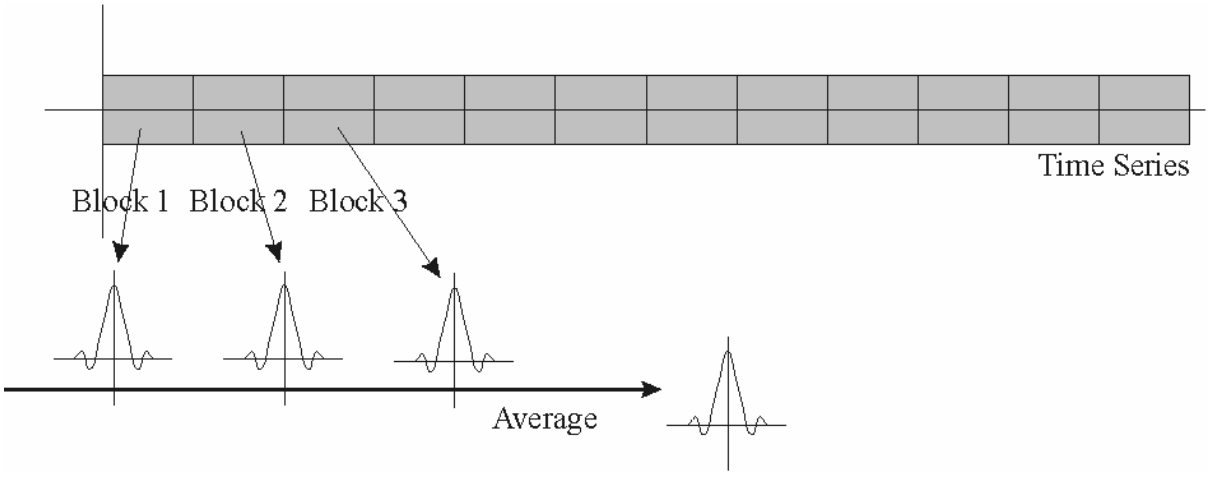
SEGMENT AVERAGING

If one seeks to compute the correlation function of a very long time series the even using FFTs can be computationally demanding.

The direct approach using the whole data set computes the correlation function for a large number of lags that are often then discarded.

One way to reduce the computational load, at the expense of some loss of accuracy, is to use segment averaging.

- The data is divided into small sub-blocks.
- Estimate the correlation function in each block, usually using FFTs.
- Finally average all the estimates.



 © ISVR 2007

CHOICE OF CORRELATION FUNCTION ESTIMATOR

The time averaged correlation function has been defined as:

$$\hat{r}(\tau) = \frac{1}{T}\int_0^T x(t)x(t-\tau)dt$$

We shall restrict our attention to the case where $\tau \geq 0$, but since $\hat{r}(-\tau) = \hat{r}(\tau)$ this is sufficient to completely define the correlation function.

Recall that for $t<0$ the signal is unknown, so that the integral in the above is more correctly written:

$$\int_\tau^T x(t)x(t-\tau)dt$$

which is an average over T-τ seconds and so one might consider using the estimator:

$$\tilde{r}(\tau) = \frac{1}{T-\tau}\int_\tau^T x(t)x(t-\tau)dt = \frac{T}{T-\tau}\hat{r}(\tau)$$

Note this is simply a re-weighting of the original correlation function.

To allow one to choose which estimator to use we now consider their statistical properties.

 © ISVR 2007

STATISTICAL PROPERTIES OF AN ESTIMATOR

Consider $\hat{\theta}$ that is an estimator of a quantity θ. Three measures of accuracy are commonly considered:

1) The bias:

$$b = \theta - E\left[\hat{\theta}\right]$$

2) The variance:

$$Var\left\{\hat{\theta}\right\} = E\left[\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)^2\right] = E\left[\hat{\theta}^2\right] - E\left[\hat{\theta}\right]^2$$

3) The mean squared error:

$$MSE\left\{\hat{\theta}\right\} = E\left[\left(\theta - \hat{\theta}\right)^2\right] = Var\left\{\hat{\theta}\right\} + b^2$$

For a good estimator all 3 quantities should be as small as possible.

The mean squared error is a combination of bias and variance and is the "gold standard", but in practice difficult to compute, since θ is not known.

 © ISVR 2007

PROPERTIES OF CORRELATION ESTIMATORS I: BIAS

1) The average value of the estimator $\hat{r}(\tau)$ is given by:

$$E\left[\hat{r}(\tau)\right] = \frac{1}{T}\int_{\tau}^{T} E\left[x(t)x(t-\tau)\right]dt = \frac{1}{T}\int_{\tau}^{T} r(\tau)dt = \frac{r(\tau)}{T}\int_{\tau}^{T} dt$$

$$= \frac{T-\tau}{T}r(\tau)$$

The bias, *b*, of the estimator $\hat{r}(\tau)$ is thus:

$$b = r(\tau) - E\left[\hat{r}(\tau)\right] = \frac{\tau}{T}r(\tau)$$

Hence the estimator $\hat{r}(\tau)$ is *biased*, *i.e.* on average we get the wrong answer.

2) The average value of the estimator $\tilde{r}(\tau)$ is given by:

$$E\left[\tilde{r}(\tau)\right] = \frac{1}{T-\tau}\int_{\tau}^{T} E\left[x(t)x(t-\tau)\right]dt = \frac{1}{T-\tau}\int_{\tau}^{T} r(\tau)dt$$

$$= \frac{r(\tau)}{T-\tau}\int_{\tau}^{T} dt = r(\tau)$$

Hence the estimator $\tilde{r}(\tau)$ is *unbiased*, *i.e.* on average we get the correct answer.

 © ISVR 2007

PROPERTIES OF CORRELATION ESTIMATORS II: VARIANCE

The variance of the two estimators can be expressed in terms of the following quantity:

$$\gamma(\tau) = \int_{u=-T+\tau}^{T-\tau} \left(T - \tau - |u|\right)\left(r(u)^2 + r(u+\tau)r(u-\tau)\right)du$$

For large values of T this can be approximated as:

$$\gamma(\tau) = T\int_{u=-\infty}^{\infty} r(u)^2 + r(u+\tau)r(u-\tau)\,du$$

This integral has the property that $\gamma(\tau)\rightarrow 0$ as $\tau\rightarrow T$.

1) The variance of $\hat{r}(\tau)$ is given by:

$$Var\left\{\hat{r}(\tau)\right\} = \frac{1}{T^2}\gamma(\tau)$$

2) The variance of $\tilde{r}(\tau)$ is given by:

$$Var\left\{\tilde{r}(\tau)\right\} = \frac{1}{(T-\tau)^2}\gamma(\tau)$$

As $\tau\rightarrow T$ the variance of $\tilde{r}(\tau)$ increases, because whilst $\gamma(\tau)$ decreases for large τ, the factor $(T\text{-}\tau)^{-2}\rightarrow\infty$ in manner that ensures divergence of the variance.

 © ISVR 2007

SUMMARY STATISTICAL PROPERTIES

The estimator $\tilde{r}(\tau)$ is:

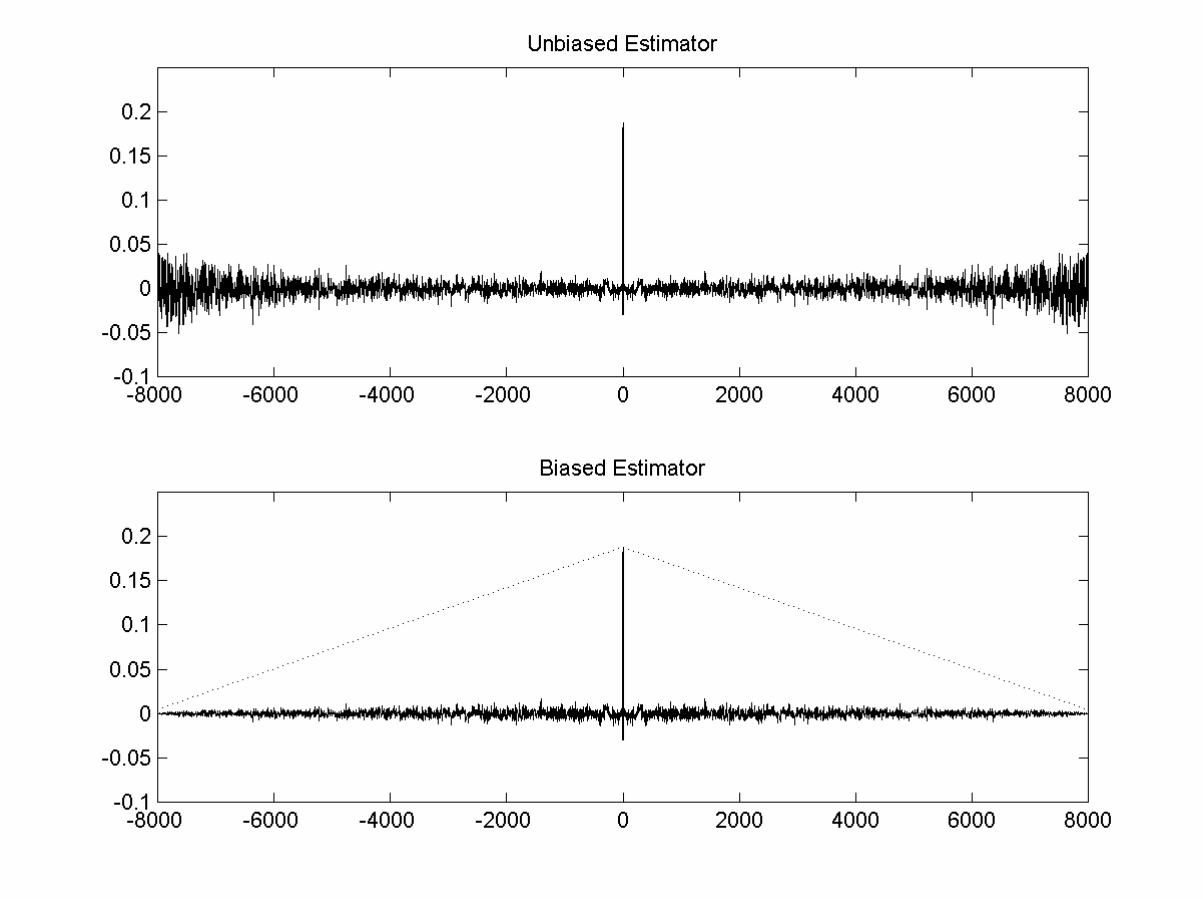
1) Is unbiased.

3) Has a large variance for τ close to T.

The estimator $\hat{r}(\tau)$ is:

1) Is biased.

2) Has a variance that is small for values of τ close to T.



 © ISVR 2007

MEAN SQUARED ERROR

The MSE is a good overall measure of the performance of an estimator and is formed as the sum of the variance with the squared bias.

We seek now to determine which of two estimators is best in terms of MSE.

The general shape of any correlation function is such that it is maximal at $\tau=0$ and decays as $|\tau|$ approaches T.

The bias of $\tilde{r}(\tau)$ is $r(\tau)|\tau|/T$. The factor $|\tau|/T$ becomes large as $|\tau|$ approaches T, but this is as $r(\tau)$ decays.

These two factors act so that the overall bias tends not to become excessive.

The effect is that the biased estimator, $\tilde{r}(\tau)$, always (?) has a lower MSE than the unbiased estimator, $\hat{r}(\tau)$.

Hence the biased estimator is usually used in preference to the unbiased estimator.

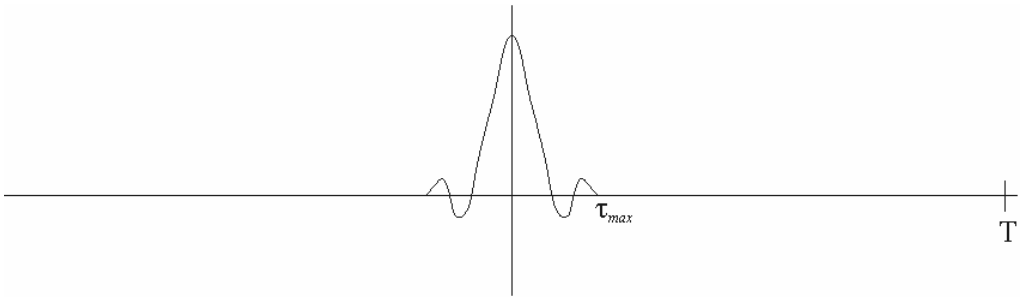
 © ISVR 2007

COMMENTS ON ACHIEVING 'GOOD' CORRELATION FUNCTION ESTIMATES

Let us consider a correlation function that decays so that for lags greater than τ_{max} the correlation function is negligible.

To estimate this correlation function one needs to have an observation length much greater than τ_{max}.

In which case the only important values of the correlation function will appear close to the centre of your estimate.



In this region the bias will be small and the variance will not have begun to increase, since you are well away from *T*.

In such circumstances the two correlation function estimators will achieve similar performance.

A small value of τ_{max} implies a large bandwidth, *B*, and so good performance will occur when the product *BT* is large.

 © ISVR 2007

TRADING VARIANCE AGAINST BIAS

Let us consider a scheme when one seeks to truncate the unbiased correlation function estimate at some value τ_1.

This truncation is a form of windowing, much as the biased estimator windows the unbiased estimator using a triangular window.

The new estimate, $\breve{r}(\tau)$, will be:

$$\begin{aligned} \breve{r}(\tau) &= \tilde{r}(\tau) & |\tau| \le \tau_1 \\ &= 0 & \text{Elsewhere} \end{aligned}$$

The choice of τ_1 is critical:

- If it is too small then important parts of the correlation function will be deleted leading to bias errors.
- If it is too large the unnecessary variance will be introduced.

Any attempt to window the correlation function will encounter a similar trade-off between bias and variance.

 © ISVR 2007

EXAMPLE SIGNAL: A MARKOV PROCESS

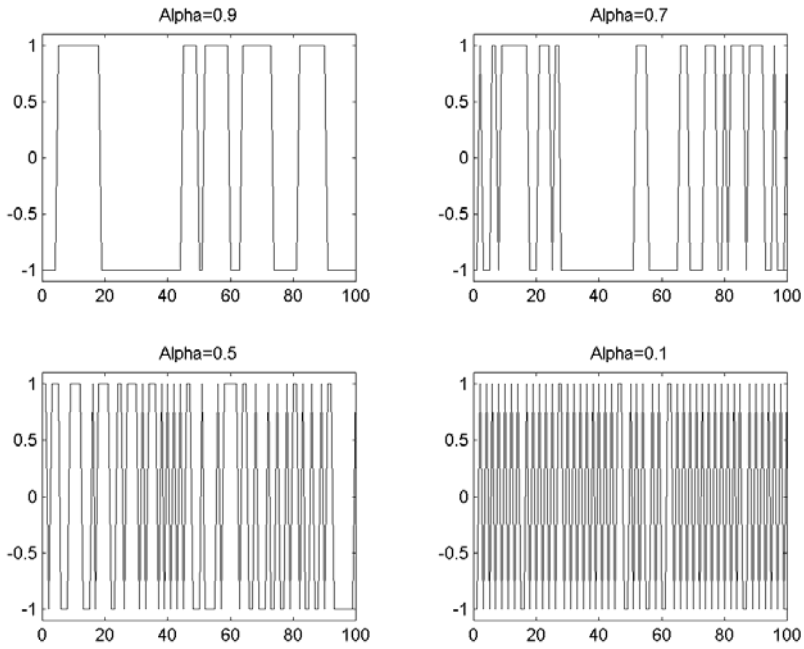
Consider a digital signal which only takes on the values of ±1 and is constructed as follows:

At time *n* the signal is either at +1 or -1, to create the signal at *n*+1 randomly choose between:

a) Keeping the signal the same, with fixed probability α.

b) Changing the sign of the signal, with probability 1-α.

This generates a sequence of +1 and -1's. If α is close to 1 then these changes are infrequent and become more frequent as α reduces.



 © ISVR 2007

THE CORRELATION FUNCTION FOR A MARKOV SIGNAL (I)

We seek to compute the digital correlation function:

$$r(m) = E[x(n)x(n-m)]$$

Since $x(n)$ can only be ± 1 then $x(n)x(n\text{-}m)$ can also only be ± 1.

To compute $E[x(n)x(n\text{-}m)]$ one needs to calculate:

$$(+1)\times\Pr\{\, x(n)x(n\text{-}m)=+1\} + (-1)\times\Pr\{\, x(n)x(n\text{-}m)=-1\}$$
$$= \Pr\{\, x(n)x(n\text{-}m)=+1\} - \Pr\{\, x(n)x(n\text{-}m)=-1\}$$

Note that $\Pr\{\, x(n)x(n\text{-}m)=+1\}$ is equivalent to the probability that $x(n)$ has the same value as $x(n\text{-}m)$ and $\Pr\{\, x(n)x(n\text{-}m)=-1\}$ is equivalent to the probability that $x(n)$ and $x(n\text{-}m)$ have different values.

How does one compute these probabilities?

Let us denote the probability of $x(n)=1$ as being $p_{1,n}$ and the probability of $x(n)=-1$ as being $p_{\text{-}1,n}$ and defined the vector $\underline{p}_n$:

$$\underline{p}_n = \begin{bmatrix} p_{1,n} \\ p_{-1,n} \end{bmatrix}$$

 © ISVR 2007

THE CORRELATION FUNCTION FOR A MARKOV SIGNAL (II)

These probabilities obey the following recursion:

$$p_{1,n+1} = \alpha p_{1,n} + (1-\alpha) p_{-1,n}$$
$$p_{-1,n+1} = \alpha p_{-1,n} + (1-\alpha) p_{1,n}$$

which may be expressed in matrix form:

$$\underline{p}_{n+1} = \begin{bmatrix} \alpha & (1-\alpha) \\ (1-\alpha) & \alpha \end{bmatrix} \underline{p}_n = A \underline{p}_n$$

Thus:

$$\underline{p}_n = A^n \underline{p}_0$$

One can show that:

$$A^n = \frac{1}{2} \begin{bmatrix} 1+(2\alpha-1)^n & 1-(2\alpha-1)^n \\ 1-(2\alpha-1)^n & 1+(2\alpha-1)^n \end{bmatrix}$$

The probability of $x(n)$ being in the same state as $x(0)$ is given by the diagonal term of this matrix and the probability of being in different states is given by the off-diagonal term.

Hence:

$$E\left[x(n-m)x(n)\right] = \frac{1}{2}\left(1+(2\alpha-1)^m\right) - \frac{1}{2}\left(1-(2\alpha-1)^m\right) = (2\alpha-1)^m$$

$$r(m) = (2\alpha-1)^m$$

 © ISVR 2007

COMMENTS ON MARKOV EXAMPLE

For α close to 1 then $(2\alpha-1)$ is also close to unity and so the correlation function decays slowly.

This reflects the fact that the sequence $x(n)$ stays constant for relatively long periods, so is quite predictable.

For α close to 0.5 then $(2\alpha-1)$ is close to zero and so the correlation function decays rapidly.

This reflects the fact that the sequence $x(n)$ changes rapidly and so is quite unpredictable.

For α close to 0 then $(2\alpha-1)$ is close to -1 and so the correlation function changes very rapidly and rarely has two values at the same level.

For the case $\alpha=0.5$, the correlation function is zero everywhere except $m=0$, and $x(n)$ is a *white noise* sequence.

This reflects the fact that knowing the value at $x(n)$ does not allow one to say anything about the value of $x(n+1)$.

 © ISVR 2007