SOME ASPECTS OF PROBABILITY THEORY - (N. Isyumov / P. King)

Deterministic Processes:

A deterministic process can be described by a mathematical model, from which the exact state of the process can be predicted. Namely, if we conduct an experiment with known inputs we are able to predict the outcome in exact form.

Stochastic or Random Processes:

A stochastic or random process is one which, although describable by a mathematical model, cannot be predicted exactly either due to the uncertainties associated with the inputs or due to the complexity of the physical process. Consequently, if we carry out an experiment we are not certain of the exact outcome but can only associate certain likelihoods to particular states of the process.

Real Processes:

In real life, all processes are random, even though in many cases if the uncertainties are small we are prepared to disregard the variability of a process and treat it as a deterministic one. Whether or not a particular problem can be treated as a deterministic process depends upon the degree of uncertainty and how sensitive the outcome of the process is to this uncertainty.

If the variability or randomness of a process cannot be disregarded all formulations are based on the mathematic theory of probability. Concepts of probability relate to descriptions of the variable in the amplitude domain. The concept of the probability distribution of the magnitude of a particular variable forms a complete description if only the magnitude rather than both the magnitude and its rate of change in time is important. For example, in defining the rupture strength of steel bars, we are only concerned with the variation in strength from bar to bar and thus defining the probability distribution of the rupture strength completely describes the process. On the other hand, if we are concerned with the response of say an oscillator to a random signal we are not only concerned with the probability distribution of the signal but would also like to know its variation in the time domain. In other words, how is the magnitude of the signal at time $t = t$, related to that at $t = t, \quad + \Delta t$ etc. This implies that we are interested in the frequency description of the variation. This is described by the auto-correlation of the processes or the power spectrum of the process.

If we are dealing with several processes simultaneously, we would further want to know their joint probability distributions and their cross-correlations and cross-spectra.

One-Dimensional Probability Distribution:

If "x" is a random continuous variable we can associate a probability to its magnitude being between any two levels; i.e. x_1 and $x_1 + dx$ as follows:

$p(x_1)dx \equiv$ probability that $x_1 \le x \le x_1 + dx$ where $p(x_1)$ is the probability density of the process at $x = x_1$.

Properties of $p(x)$:

$p(x) \ge 0$ for all x;

$\int_{-\infty}^{\infty} p(x)dx = 1$; and

the probability that $x_1 \le x \le x_2$ becomes

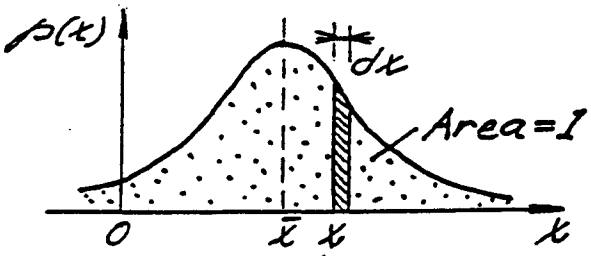
$$P\left[x_1 \le x \le x_2\right] = \int_{x_1}^{x_2} p(x)dx$$

A useful further concept is the cumulative distribution function $F(x_1)$ which represents the probability that $x \le x_1$. Typically

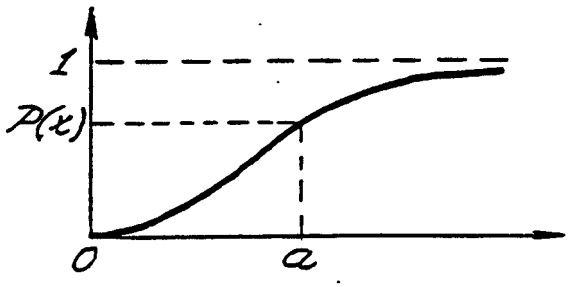
$$F(x_1) = P[x \le x_1] = \int_{-\infty}^{x_1} p(x)dx$$

Typically:

Probability Density



Cumulative Distribution Function



$$Area = F(x_1) = P[x \le x_1] = \int_{-\infty}^{x_1} p(x)dx$$

If we have a variable which only has discrete values; i.e. the age of students in years or numbers on a roulette wheel; the probability is described in terms of the probability mass function or the relative frequency. Typically if x only has discreet values of say $x_1, x_2, x_3 ..... x_N$, the probability that $x = x_1$, i.e. $P\left[x = x_1\right] = r\left(x_1\right)$ where $r\left(x\right)$ is the probability mass function. The following properties of $r\left(x\right)$ are apparent:

$0 \le r\left(x\right) \le 1$ for all x

$$\sum_{all\ x} r\left(x\right) = 1$$

$$P\left[a \le x \le b\right] = \sum_{x_i = a}^{x_i = b} r\left(x_i\right)$$

2-Dimensional Probability Distribution:

If we have two random variables, say x and y, we can define a joint probability as follows:

$p_{xy}\left(x_1, y_1\right)dx\,dy \equiv$ Probability that
$x_1 \le x \le x_1 + dx$ and
$y_1 \le y \le y_1 + dy$

where $p_{xy}\left(x,y\right)$ is the joint probability density function of the variables x and y.

If x and y are independent

$$p_{xy}\left(x,y\right) = p_x\left(x\right)p_y\left(y\right)$$

By its definition $p_{xy}\left(x,y\right) \ge 0$ for all x and y and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{xy}\left(x,y\right)dx\,dy = 1$$

For example, a two-dimensional normal probability distribution becomes;

$$p_{xy}\left(x,y\right) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} e^{-\frac{1}{\left(1-\rho_{xy}^2\right)}\{-\}}$$

where $\{-\} = \dfrac{(x-\bar{x})^2}{\sigma_x^2} - 2\rho_{xy}\dfrac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y} + \dfrac{(y-\bar{y})^2}{\sigma_y^2}$

where $\rho_{xy} = (x-\bar{x})(y-\bar{y})/\sigma_x\sigma_y$; $\bar{x}$ and $\bar{y}$ are mean values of x and y; and σ_x^2 and σ_y^2 are variances of x and y.

If x and y are independent

$$p_{xy}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\frac{(x-\bar{x})^2}{\sigma_x^2} + \frac{(y-\bar{y})^2}{\sigma 2_y}\right]}$$

Practical Estimates of the Prob. Distribution:

If we have at our disposal samples of a process, these can be sorted into a histogram. A histogram is a plot that gives the relative frequency of occurrence of the variable x in different intervals of x. Typically, if we have sorted M values of the variable x into a histogram and say N_i of these fall into the interval between x_i and $x_i + \Delta x$ then,

$\dfrac{N_i}{M\,dx}$ = an estimate of the probability that

$$x_i \le x \le x_i + \Delta x$$

This estimate becomes exact as $M \to \infty$. If our variable "x" is continuous and Δx is small, i.e. $\Delta x \to dx$, an estimate of the probability density becomes

$$p(x_i) = \frac{N_i}{M\,dx}$$

Usually Δx is finite and a good estimate of $p(x)$ at the midpoint of the interval is,

$$p\left(x = x_i + \frac{\Delta x}{2}\right) = \frac{N_i}{M\,\Delta x}$$

- Having obtained an estimate of $p(x)$ one can sometimes fit $p(x)$ by a mathematical model using least square fitting or maximum likelihood fitting etc. Having done this one of course is very much interested in the "goodness of fit".

- The theory of moments also provides a useful fitting procedure. Furthermore, from moments of the probability density function estimate, one can make very useful statements about the variable.

Theory of Moments

If we know either $p(x)$ for a continuous variable or $r(x)$ for a discrete variable the following moments can be formed.

Moments About the Origin:

n^{th} moment for a continuous variable x becomes;

$$\mu_n(x) = \int_{-\infty}^{\infty} x^n \; p(x)\, dx$$

n^{th} moment for a discrete variable x becomes;

$$\mu_n(x) = \sum_{all\ x_i} x_i^n \; r(x_i)$$

Central Moments or Moments About the Mean:

n^{th} moment for a continuous variable x becomes

$$\upsilon_n(x) = \int_{-\infty}^{\infty} [x - \mu_1(x)]^n \; p(x)\, dx$$

n^{th} moment for a discreet variable x becomes

$$\upsilon_n(x) = \sum_{all\ x_i} [x_i - \mu_1(x)]^n \; r(x_i)$$

The most important moments are the first moment about the origin, which represents the mean value, and the second moment about the mean which represents the variance.

If we have N samples of the process x, i.e. $x_1, x_2 ..... x_N$, estimates of the mean and variance are;

$$\hat{\mu}_1(x) = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \approx \bar{x}$$

$$\upsilon_2(x) = \frac{1}{N}\sum_{i=1}^{N} \left[x_i - \hat{\mu}_1(x)\right]^2 \quad \approx \sigma_x^2$$

- The above two moments permit approximate predictions of the probability of exceeding particular values of x if the process is approximately normal.

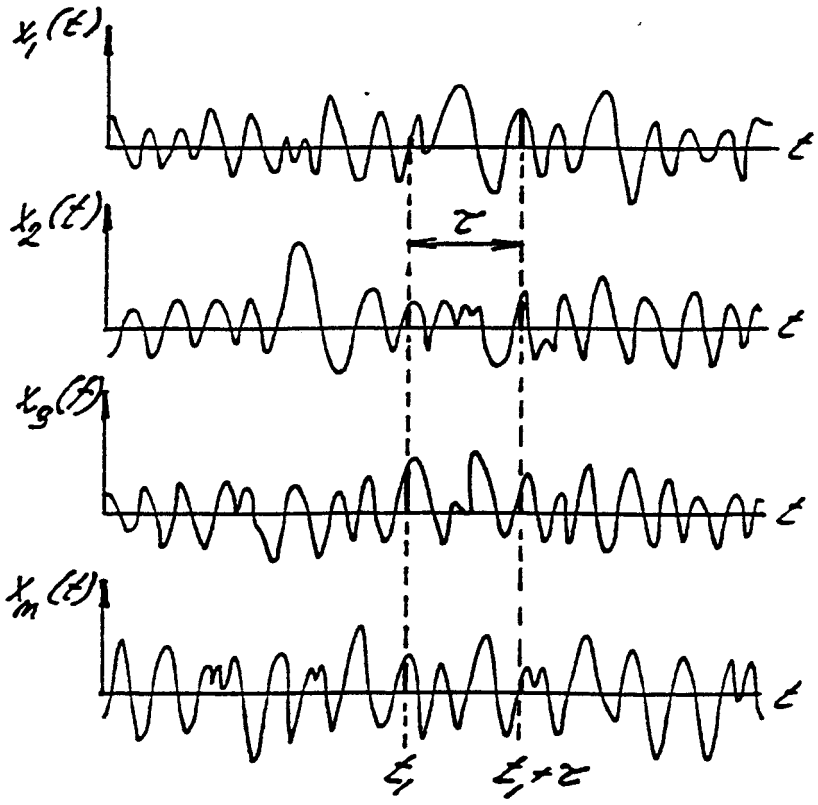
- In any event use can be made of Chebyshev's Inequality to make exact statements of the lower bound of probability;

i.e. $$P\left[\left(\bar{x} - h\sigma_x\right) \le x \le \left(\bar{x} + h\sigma_x\right)\right] \ge \left(1 - \frac{1}{h^2}\right)$$

SOME ASPECTS OF TIME VARYING RANDOM VARIABLES

General Description:

Suppose we have an ensemble of N sample functions of a random variable $x(t)$; i.e.



In general, all statistical properties of the process may vary from sample to sample and as a result, the statistical properties are functions of time. Typically, the ensemble mean at $t = t_1$ becomes

$$\mu_x(t_1) = \lim_{N\to\infty} \sum_{k=1}^{N} \approx k(t_1)$$

Similarly the variance and covariance or correlation functions depend on "t"; i.e.

Variance:

$$\sigma_x^2(t_1) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} [x_k(t_1) - \mu_x(t_1)]^2$$

Covariance or Correlation Function

$$\sigma_x(t_1,t_2) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} [x_k(t_1) - \mu_x(t_1)][x_k(t_2) - \mu_x(t_2)]$$

STATONARY PROCESS

A stationary process is one in which the statistical properties are invariant with time. This obviously permits great simplifications.

Strict Definition of Stationarity: implies all statistical properties are invariant.

Weak Stationarity (usually sufficient for practical application):
Only requires that,

$$\mu_x(t_1) = \mu_x(t_2) = ... = \mu_x \text{; and}$$

if $\rho_x(t_1,t_2) = \rho_x(t_1,t_1+\tau)$ where $t_2 = t_1 + \tau$,

that $\rho_x(t_1,t_1+\tau) = \rho_x(t_2,t_2+\tau) = ... = \rho_x(\tau)$

for all values of τ.

Note that the correlation function becomes independent of the origin of time. Also note that a weakly stationary process if Gaussian is also stationary in the strict sense.

Ergodic Process

If the statistical properties of a single sample record averaged with time are identical to the ensemble statistics the process is called ergodic.

Namely, if for the k^{th} sample record we compute;

$$x_k = \lim_{T\to\infty} \frac{1}{T} \int_0^T x_k \, dt \text{ ; and}$$

$$R_{x_k}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T [x_k(t) - \bar{x}_k][x_k(t+\tau) - \bar{x}_k] dt$$

The process is ergodic if

$$\bar{x}_k = \bar{x} = \mu_x \text{; and}$$

$$R_{x_k}(\tau) = R_x(\tau) = \rho_x(\tau)$$

This implies that

Temporal Properties = Ensemble Properties

Ergodic processes in view of their tractability are very attractive and usually every attempt is made to make use of the ergodicity principle. If the process is ergodic all information about the statistical properties can be obtained from a single record of the variable.

In reality of course we usually only have estimates of the statistical properties as T is finite rather than ∞.

Auto-Correlation Functions

The auto-correlation function for an ergodic process $x(t)$, as already defined, becomes;

$$R_x(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T [x(t) - \bar{x}][x(t+\tau) - \bar{x}]\, dt$$

usually we only have an estimate of $R_x(\tau)$, i.e. $\hat{R}_x(\tau)$ if T is finite.

The auto-correlation function $R_x(\tau)$ represents the average correlation between the variable x at time t and the same variable at time $t+\tau$. Here τ is a separation in time, sometimes called the lag time. Since the auto-correlation function is a measure of the correlation between the signal at different point in time, it provides information on the frequency content of the signal.

Frequently it is convenient to express the auto-correlation function in non-dimensional form, namely

$$R_x'(\tau) = \frac{T \lim_{T\to\infty} \frac{1}{T} \int_0^T [x(t) - \bar{x}][x(t+\tau) - \bar{x}]\, dt}{\lim_{T\to\infty} \frac{1}{T} \int_0^T [x(t) - \bar{x}]^2\, dt}$$

$$= R_x(\tau) / \sigma_x^2$$

Properties of Auto-Correlation Functions

i) $R_x(\tau)$ is an even or symmetric function; namely

$$R_x(\tau) = R_x(-\tau)$$

ii) $R_x(\tau) \le \sigma_x^2$ for all τ

iii) If $x(t) = y_1(t) \pm y_2(t) \pm ..... R_{y_n}(\tau)$ where

y_1, y_2 ..... y_n are independent, then

$$R_x(\tau) = R_{y_1}(\tau) + R_{y_2}(\tau) + ... R_{y_n}(\tau)$$

Measurements of Auto-Correlation

i) Analogue methods

ii) Digital method: if we sample a process $x(t)$ at a sampling rate of $1/\Delta\tau$ apart in time, then an estimate of R_x becomes;

$$R_x(k\Delta\tau) = \frac{1}{\sigma_x^2 (N-k)} \sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})$$

where $k = 0, 1, 2, 3 ..... M$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x; \sigma_x^2 = \frac{1}{N} i = 1 \sum^{N} (x_i - x)^2 \text{; and}$$

$M \equiv$ maximum lag number

$\tau_{\max} = M \Delta\tau \equiv$ maximum lag time.

Power Spectra:

The auto-correlation function, although useful in providing information on the frequency content, is of greater direct use if transformed to the power spectrum form.

There exists a Fourier transform pair;

$$G_x(\omega) = \frac{\sigma_x^2}{2\pi} \int_{-\infty}^{\infty} R_x(\tau)\, e^{-i\omega\tau}\, d\tau$$

$$R_x(\tau) = \frac{1}{\sigma_x^2} \int_{-\infty}^{\infty} G_x(\omega)\, e^{i\omega\tau}\, d\omega$$

where "ω" is the circular frequency and $G_x(\omega)$ is the two-sided power spectral density of the process $x(t)$.

$G_x(\omega)$ can be expressed in terms of a one-sided power spectral density $\phi_x(n)$. Since $R_x(\tau)$ is a real even function the above general Fourier transform pair can be replaced by a real cosine transform pair as follows:

$$\phi_x(n) = 4\,\sigma_x^2 \int_o^{\infty} R(\tau) \cos 2\pi n\tau\, d\tau$$

$$R_x(\tau) = \frac{1}{\sigma_x^2} \int_o^{\infty} \phi(n) \cos 2\pi n\tau\, d\tau$$

where n is the frequency in cps.

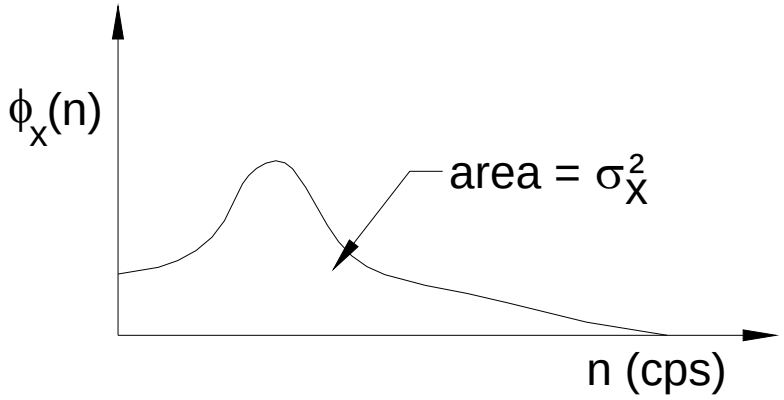
The power spectral density $\phi_x(n)$ is a measure of the proportion of the total energy or the total variance of the process at frequency "n" per unit frequency. Namely

$\phi_x(n_1)\, dn$ = contribution to the variance at frequencies n
where $n_1 \le n \le n_1 + dn$

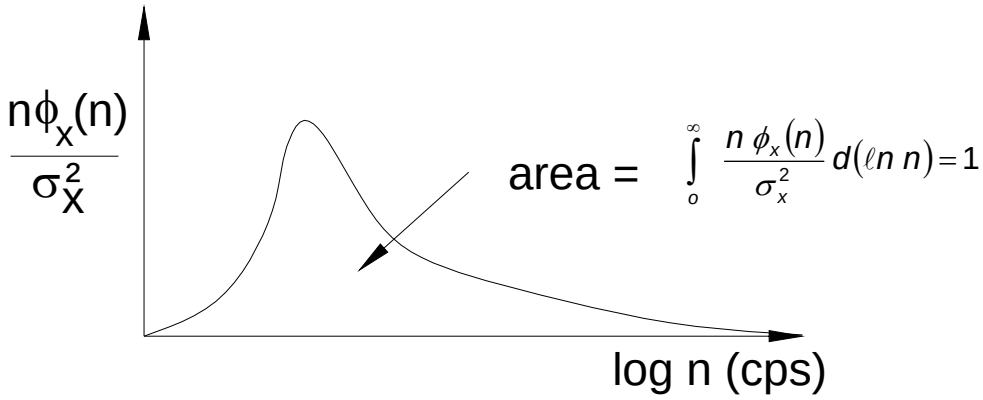
Consequently

$$\int_o^{\infty} \phi_x(n)\, dn = \phi_x^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2\, p(x)\, dx$$

The above equation forms the link between the descriptions of the random variable in the amplitude and frequency domains.



Another useful form of the power spectrum is



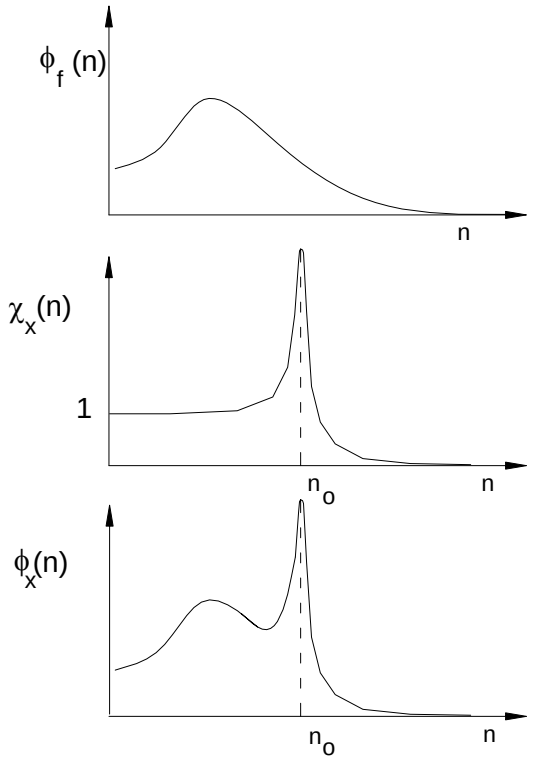
Typical Example:

If a mass, spring, dashpot single degree of freedom oscillator is subjected to a random force $F(t)$ having a power spectral density of $\phi_F(n)$, the power spectrum of the response "x" to this excitation becomes

$$\phi_x(n) = \frac{1}{K^2} \quad |\chi_m(n)|^2 \; \phi_F(n)$$

where $|\chi_m(n)| \equiv$ mechanical admittance or mechanical magnification

$$\equiv \left[\frac{1}{\left(1 - \frac{n^2}{n_o^2}\right)^2 + 4\,\zeta^2 \frac{n^2}{n_o^2}} \right]^{1/2}$$



$$\sigma_x^2 = \int_o^\infty \phi_x(n)\,dn = \frac{1}{K^2} \int_o^\infty |\chi_m(n)|^2 \; \phi_F(n)\,dn$$

where $n_o = \frac{1}{2\pi}\sqrt{\frac{K}{M}} \equiv$ undamped natural frequency

In contrast if the excitation $F(t) = A \sin 2\pi n_1$ (i.e. a sine wave at frequency n_1), the variance of the response would be

$$\sigma_x^2 = \frac{1}{2}\frac{A^2}{K^2}\,|\chi_m(n_1)|^2 = \frac{1}{K^2}\,|\chi_m(n_1)|^2 \; \sigma_F^2$$

Namely all energy of the response is at the frequency of the excitation "n_1".

Measurements of Power Spectra

1. Analogue Methods: In these methods the signal if passed through a narrow band pass filter and the variance of the signal associated with different frequency bands is measured. It is important in such measurements to choose a sufficiently small band width of the filter to permit adequate resolution in the frequency domain.

2. Digital Methods:

   i) measure the auto-correlation function and transform mathematically to obtain the power spectrum.

   ii) transform the data, i.e. $x(t)$, to obtain an amplitude spectral density which when squared becomes a power spectral density.

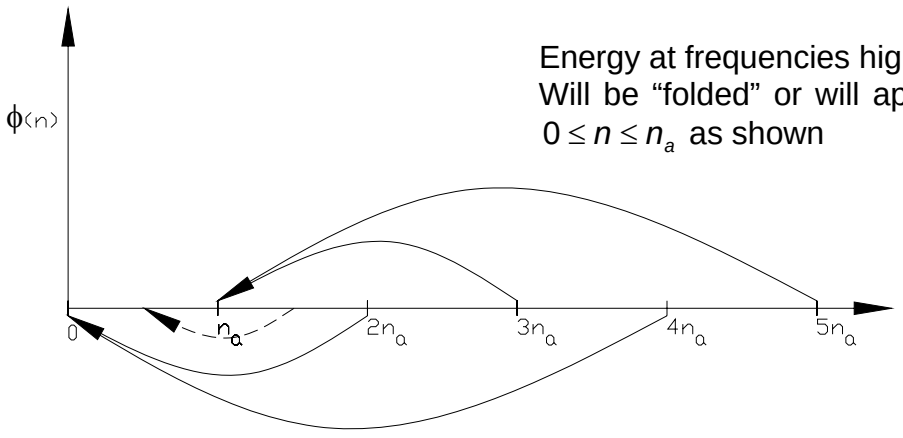
   iii) use of a "FFT" (Fast Fourier Transform) algorithm to do ii) more efficiently.

Calculation of $\phi(n)$ from $R(\tau)$:

If we are planning to compute $\phi(n)$ from $R(\tau)$ there are a number of considerations for such calculation. These are as follows:

i) The choice of $\Delta\tau$, i.e. the spacing in time of the sampled data, has to be chosen sufficiently small to provide information at high frequencies.

For example if the sampling frequency is $n_s = \frac{1}{\Delta\tau}$ we can only obtain undistorted information on $\phi(n)$ up to the aliasing frequency n_a, where $n_a = \frac{n_s}{2} = \frac{1}{2\Delta\tau}$. Energy at frequencies higher than n_a will erroneously appear at lower frequencies as shown below:



Energy at frequencies higher than
Will be "folded" or will appear in the range $0 \le n \le n_a$ as shown

Solutions: a) Choose $n_s = 2n_a$ sufficiently high to minimize distortion of $\phi(n)$ estimates.

b) Pass the signal through a low pass filter to eliminate the contribution to the power at $n > n_a$.

ii) The choice of the maximum lag $\tau_{\max} = M\Delta\tau$ limits the information at low frequencies and also determines the resolution of the $\phi(n)$ estimate:

practical minimum frequency = $\dfrac{1}{2\tau_{\max}}$

resolution $\equiv$ effective band width = $\dfrac{1}{\tau_{\max}} \approx b_e$.

Consequently, if we have computed $R(\tau)$ at τ values of $k\Delta\tau$ where $k = 0, 1, 2, ... M$, we can obtain estimates of $\phi(n)$ for the following frequencies:

$$n = \underbrace{\frac{1}{2\tau_{\max}}}_{\text{minimum frequency}}, \quad \frac{2}{2\tau_{\max}}, \quad .... \frac{k}{2\tau_{\max}}, \quad .... \underbrace{\frac{M}{2\tau_{\max}} = \frac{1}{2\Delta\tau}}_{\text{maximum frequency}}$$

iii) The total number of data points N and $\tau_{\max}$ also provide information on the statistical reliability of the $\phi(n)$ estimate

- Coefficient of Variation = $\dfrac{1}{\sqrt{b_e\, T}}$, where T = length of record of $\phi(n)$

$\therefore$ Coeff. of Variation = $\dfrac{1}{\sqrt{\dfrac{1}{M\Delta\tau} N\Delta\tau}} = \dfrac{2N}{M}$

- Confidence limits for $\phi(n)$ estimates can be obtained from those of a χ^2 distribution with k degrees of freedom, where

$$k = 2b_e T = \frac{2T}{\tau_{\max}} = \frac{2N}{M}$$

Reasonable estimates of $\phi(n)$. Theoretically we have;

$$\phi(n) = 4\sigma_x^2 \int_o^\infty R(\tau) \cos 2\pi n \tau$$

If we have an $R(\tau)$ function defined for $0 \le \tau \le \tau_{\max}$, an estimate of $\phi(n)$ becomes

$$\hat{\phi}(n) = 4\sigma_x^2 \int_o^{\tau_{\max}} R(\tau) \cos 2\pi n \tau$$

where the integration is carried out numerically. Since $\tau_{\max}$ is finite, the estimate obtained is not very good. In fact, if ϕ_1 is the true spectral density the above estimate will be identical to

$$\hat{\phi}(n) = 4 \int_o^\infty \phi_1(n-\eta) \frac{\sin 2\pi\eta\, \tau_{\max}}{\pi\eta} \quad d\eta$$

where η is a dummy variable of integration and $\frac{\sin 2\pi\eta\, \tau_{\max}}{\pi\eta}$ is referred to as a "spectral window".

An improved "smoothed" estimate of $\phi(n)$ is obtained if $R(\tau)$ is forced to zero at $\tau = \tau_{\max}$. In other words, the measured $R(\tau)$ is multiplied by a "lag window", called $D(\tau)$ and then transformed. Several such windows are used in practice. Typical lag windows are:

Bartlett Window:

$$D_1(\tau) = 1 - \frac{|\tau|}{\tau_{\max}};\ -\tau_{\max} \le \tau \le \tau_{\max} \text{ and = 0 elsewhere}$$

Hanning Window:

$$D_2(\tau) = \frac{1}{2}\left(1 + \cos\frac{\pi\tau}{\tau_{\max}}\right);\ -\tau_{\max} \le \tau \le \tau_{\max} \text{ and = 0 elsewhere}$$

Hamming Window:

$$D_3(\tau) = 0.54 + 0.46\cos\frac{\pi\tau}{\tau_{\max}};\ -\tau_{\max} \le \tau \le \tau_{\max} \text{ and = 0 elsewhere}$$

Using the above lag windows is only a mathematical procedure to improve the $\phi(n)$ estimates.

Using a lag window, we would numerically integrate the following:

$$\hat{\phi}(n) = 4\sigma_x^2 \int_0^{\tau_{\max}} D_i(\tau) R(\tau) \cos 2\pi n \tau d\tau$$