DATA QUALITY

Experimental Data Quality

- Errors and Error Analysis
- Statistical Analysis
- Time and Frequency Domain Analysis
- Smoothing, Windowing, etc…

Errors

Sources

- Instrument Errors – construction, variability, temperature effects (not known to experimentalist).
- Systematic errors (fixed) – are of consistent form that result from conditions or procedures that are correctable (zero shift, calibration factor, etc…).
- Random Errors – accidental errors that occur in all measurements. They are characterized by their inconsistent nature, and their origin can not be determined in the measurement process (corrected by statistical analysis), i.e. fluctuations, noise, etc…

Error Analysis

- Identify possible errors and associate with given data.

i.e. $\sigma = P/A$

$F = F_0 \pm \Delta F$ …$\pm \Delta F$: Δ possible random error

$A = b \cdot h \rightarrow$

$b = b_0 \pm \Delta b$ …$\pm \Delta b$: Δ possible random error

$h = h_0 \pm \Delta h$ …$\pm \Delta h$: Δ possible random error

$\sigma_0 = F_0 / b_0 h_0$ …$\Delta\sigma$ = ??

$\Delta F, \Delta h, \Delta b$ can be determined from accuracy of instruments, deviation in reading or data transfer, truncation in recordings, etc…

Determine errors for complex calculations

i.e. $f = f_0(x_1, x_2, ... x_n)$.

$$df = \frac{\partial f_0}{\partial x_1} dx_1 + \frac{\partial f_0}{\partial x_2} dx_2 + ... \frac{\partial f_0}{\partial x_n} dx_n$$

if $dx_1, dx_2 .... dx_n$ are errors

$\frac{\partial f_0}{\partial x_i}$ are weighting factors.

df is the resulting error.

Determine initial contribution:

$$f = f_0\left(x_1, x_2, \ldots x_n\right)$$

$$df = \frac{\partial f_0}{\partial x_1} dx_1 + \frac{\partial f_0}{\partial x_2} dx_2 + \ldots \frac{\partial f_0}{\partial x_n} dx_n$$

$$df = \varepsilon_F, \qquad dx_i = \varepsilon_i$$

Then:

$$\varepsilon_F = \frac{\partial f}{\partial x_1} \varepsilon_1 + \frac{\partial f}{\partial x_2} \varepsilon_2 + \ldots \frac{\partial f_n}{\partial x_n} \varepsilon_n$$

Example:

$$\sigma = \frac{F_0}{b_0 h_0} \qquad \begin{cases} F = 10K \pm 1kip \\ b_0 = 2'' \pm 0.01in \\ h_0 = 1'' \pm 0.01in \end{cases} \qquad \sigma_0 = \frac{10}{2 \times 1} = 5ksi$$

$$\frac{\partial \sigma}{\partial F_0} = \frac{1}{b_0 h_0} = \frac{\sigma_0}{F_0} = \frac{5}{10K}$$

$$\frac{\partial \sigma}{\partial b_0} = -\frac{F_0}{b_0^2 h_0} = -\frac{\sigma_0}{b_0} = \frac{5}{2''}$$

$$\frac{\partial \sigma}{\partial h_0} = -\frac{1}{b_0 h_0^2} = -\frac{\sigma_0}{h_0} = \frac{5}{1''}$$

Calculation of error:

$$\varepsilon_\sigma = \frac{5}{10} \times 1 \pm \frac{5}{2} \times 0.01 \pm \frac{5}{2} \times 0.01 =$$

$$\varepsilon_\sigma = 0.5 \times 1 \pm 0.025 \pm 0.05 = \pm 0.575$$

$$\varepsilon_\sigma = \frac{1}{10} \pm \frac{0.01}{2} \pm \frac{0.01}{1} = 0.1 \pm 0.005 \pm 0.01 = 0.115$$

Propagation of error:

$$\varepsilon_R^2 = \sum \left(\frac{\partial R}{\partial x_i} \right)^2 \cdot \varepsilon_{xi}^2$$

where R = calculated quantity = f (x_1, x_2, …)
x_i = measured value
ε_{xi} = measurement error

For $R = \left(k; x_1^a, x_2^b, x_3^c, \ldots . x_n^m\right)$

$$\left(\frac{\varepsilon_R}{R}\right)^2 = a^2\left(\frac{\varepsilon_{xi}}{x_i}\right)^2 + b^2\left(\frac{\varepsilon_{xi}}{x_i}\right)^2 + \ldots . m^2\left(\frac{\varepsilon_{xn}}{x_n}\right)^2$$

(obtained from a calculus of variables)

This leads to simply:

$$\left(\frac{\varepsilon_R}{R}\right) = \sqrt{\left(\frac{\varepsilon_{x1}}{x_1}\right)^2 + \left(\frac{\varepsilon_{x1}}{x_1}\right)^2 + \ldots . \left(\frac{\varepsilon_{xn}}{x_n}\right)^2} = SRSS$$

For previous example:

$$\frac{\varepsilon_\sigma}{\sigma_0} = \sqrt{(0.1)^2 + (0.005)^2 + (0.01)^2} = 0.1006$$

$$\varepsilon_\sigma = 0.503$$

$$\sigma = 5 \pm 0.503 kips$$

Statistical analysis: Mathematical expectations:

$$E[x] = \mu = \sum_{i=1}^{N} \frac{x_i}{N}$$

$$E[x-\mu]^2 = \text{variance} = \frac{\sum (x-\mu)^2}{N} = \sigma^2$$ (shows distribution…)

$$E[x]^2 = \sigma^2 + \mu^2$$

$$E[x+y] = E[x] + E[y]$$

$$E[x+y]^2 = E[x]^2 + E[y]^2 + 2E[xy]$$

where 2E[xy] is referred to as ϕ, an independent variable

$E[xy] = E[x] \cdot E[y]$ if independent variables $\Rightarrow$ 0.

Errors evaluation based on probability and expected values.

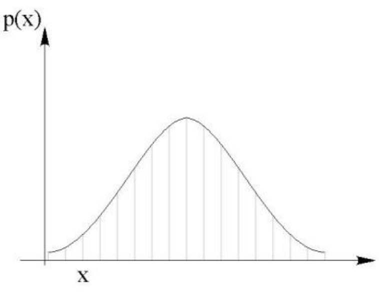
$$E[x] = \int xp(x) = \mu \qquad \Delta \frac{\int xp(x)}{\int_A p(x)} = \mu$$

$$E[x-\mu]^2 = \int (x-\mu)^2 p(x) = \text{var} = \sigma^2$$

$$\text{var}[x+y] = \text{var}[x] + \text{var}[y]$$

$$\sigma^2[x+y] = \sigma^2[x] + \sigma^2[y]$$

$$\sigma^2[x+y] = \sigma^2[x] + \sigma^2[y]$$



Correlation coefficient:

$$f_{xy} = \frac{E\left[(x-\mu_x)(y-\mu_y)\right]}{E\left[(x-\mu_x)^2\right]^{\frac{1}{2}} - E\left[(y-\mu_y)^2\right]^{\frac{1}{2}}} = \frac{E\left[(x-\mu_x)(y-\mu_y)\right]}{\left[\operatorname{var} x \times \operatorname{var} y\right]^{\frac{1}{2}}}$$

This measures how correlated is the data…

i.e. :

For two identical data sets :

$x \equiv y$ then $\sigma_x = \sigma_y$;

$$E(x-\mu_x)(y-\mu_y) = E(x-\mu_x)^2 = \sigma_x^2$$

$f_{xy} = 1$

For uncorrelated data:

$f_{xy} = 0$

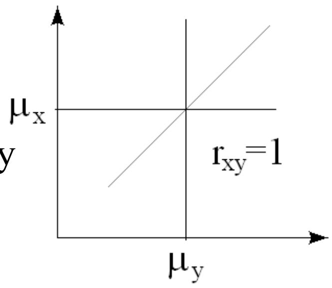
5. Correlation Analysis

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y} \quad :$$ this is the correlation coefficient

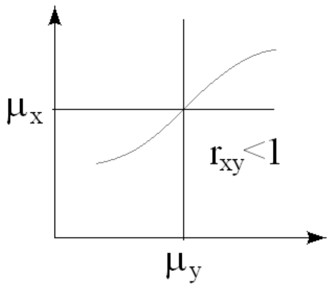
C_{xy} = covariance of x and y =

$$= \int_{-\infty}^{\infty}\int (x-\mu_x)(y-\mu_y)\, p(x \cdot y)\, dx\, dy$$

$$= \frac{1}{N}\sum (x_i - \bar{x})(y_i - \bar{y})$$



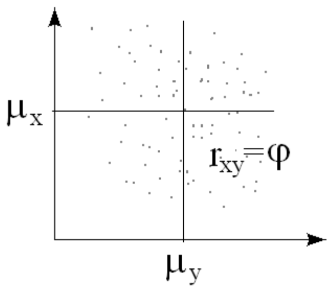
linear correlation



nonlinear correlation

or

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}$$



no correlation

Definitions

Precision $= S_x = \left[\dfrac{\sum (x_i - \mu_x)^2}{n-1}\right]^{\frac{1}{2}}$ ≅ standard deviation $\sigma \times \sqrt{\dfrac{n}{n-1}}$

(small variation to standard deviation)

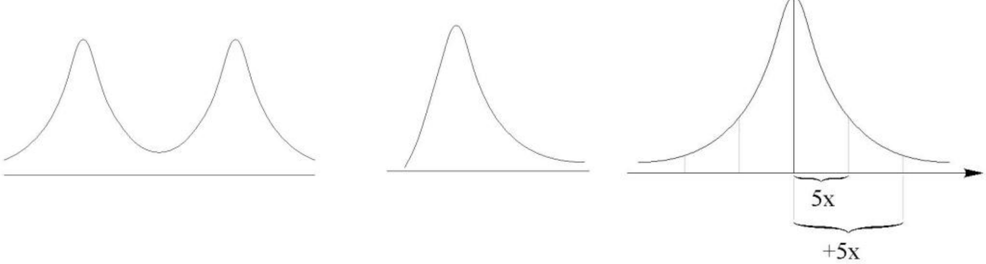
Confidence level ⇒ precision limit = $(t \cdot S_x)$

Probability of $x_i < X = \int_{-\Delta}^{x} p(x)dx$ where X is the precision limit

(p(x) ≡ probability distribution)

Probability density or distribution could be :

Binomial Poisson Normal



Statistical Analysis and Probability Issues

Mean : $x_m = \dfrac{1}{n}\sum x_i = E(x_i)$

Deviation : $d_i = x_i - x_m \Rightarrow E(d_i) = \phi$

Variance : $\text{var}\, x = \dfrac{1}{n}\sum d_i^2 = E\left[(x_i - x_m)^2\right]$

Standard dev. : $\sigma = \sqrt{\text{var}\, x} = \left[\dfrac{1}{n}\sum (x_i - x_m)^2\right]^{\frac{1}{2}}$

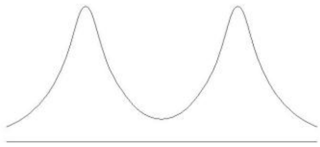
The above have meaning in experiments:

Mean: Computed value – used for engineering purposes

Distribution: How samples are related to a mean:

(1) **Binomial Distribution** (for true/false tests):

$$p(n) = \frac{N!}{(N-n)!n!} p^n (1-p)^{N-n}$$



p – probability of failure, occurrence of any variable x

N – number of tries

n – number of successes (n ≤ N)

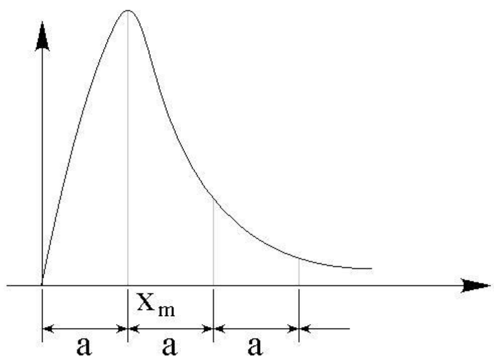
(2) **Poisson Distribution**

$$p_a(n) = \frac{a^n e^{-h}}{n!}$$

n can be a real number -variable

$a = \sigma^2$ (or $\sigma = \sqrt{a}$)

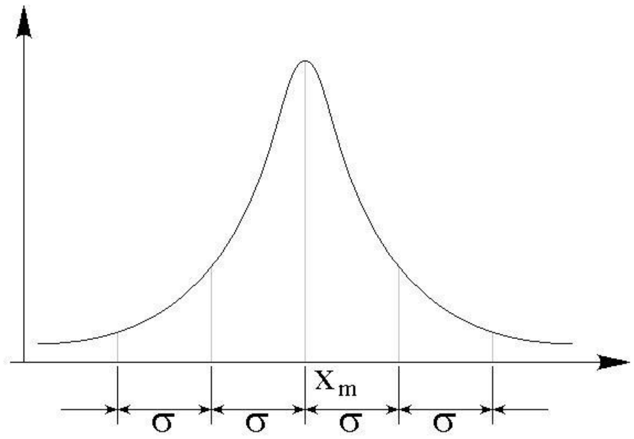
n = number of successful events



(Possible examples are 'distance from a target', 'tensile strength', i.e. positive quantities)

(3) **Gaussian Distribution**

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-x_m)^2}{2\sigma^2}}$$



Cumulative Distribution

$$P(x) = \int_{-\infty}^{x} p(x)dx$$

$$P(\infty) = 1.0$$

$$P(x_m) = \int_{-\infty}^{x_m} p(x)dx = 0.5$$

$$p(x_m) = \frac{1}{\sigma\sqrt{2\pi}}$$ maximum probability of occurrence appears for the mean, which justifies using this value for estimates.

Error distribution around the mean:

$$z = \frac{x_i - x_m}{\sigma}$$

is measured in terms of standard deviation and number of samples for the range (x – x_m) with their cumulative probability (see tables).

Example : Non-Standard Normal Distribution

When $\bar{x} \neq 0$ and $\sigma \neq 1$,

Then $z = \dfrac{x - \bar{x}}{\sigma}$ (non-standard random variable)

Then use previous table to find cumulative

i.e. if $\bar{x} = 50\ ft/\sec$ (mean failure velocity)

$\sigma = 0.2\bar{x} = 10\ ft/\sec$

These two variables are found from experimental testing.

What is the probability of success for a 30 ft/sec strike?

$$z = \frac{30 - 50}{10} = -2$$

From standard normal curve tables

p(z) = 0.0228 chance of failure

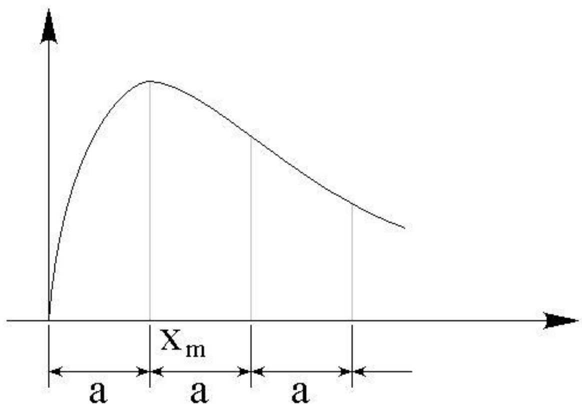
or 1 - 0.0228 = 0.9772 chance of success

Poisson's Distribution

$$p(x) = \frac{a^x e^{-a}}{x!}$$

where a = σ^2 (usually rate/time or rate/area)

x = number of successful points



i.e. Let x = number of flaws on a surface of a randomly selected boiler .

suppose a = 5 (from observation of many boilers)

What is the probability that a randomly selected boiler has exactly 2 flaws?

$$p(x = 2) = \frac{5^2 e^{-5}}{2!} = 0.084$$

At most 2 flaws? (none, 1, or 2)

$$p(x \le 2) = \sum_{x=0}^{2} \frac{5^x e^{-5}}{x!} = e^{-5}\left(1 + 5 + \frac{25}{2}\right) = 0.125$$

Binomial Distribution

$$P(x) = \frac{N!}{(N-x)!\,x!}\,p^x (1-p)^{N-x}$$

where P(x) = probability of success

N = number of data points

x = number of successful points with (x ≤ N)

p = individual probability of a successful Bernoullian test (pass or fail).

i.e. 6 randomly selected beer drinkers are given a glass of beer S and beer F. The glasses are identical in appearance such that there is no tendency to prefer one beer from the other.

⇒ p = 0.5

What is the probability that 3 will select beer S?

$$P(x = 3, N = 6, p = 0.5) = \frac{6!}{(6-3)!\,3!}\,0.5^3 (1-0.5)^{6-3} = 0.313$$

at least 3 select beer S?

(x = 3, 4, 5, 6)

$$P(x \geq 3, N = 6, p = 0.5) = \sum_{x=3}^{6} \frac{6!}{(6-x)!\,x!}\,0.5^x (1-0.5)^{6-x} = 0.656$$

Binomial Distribution: Example: SBCCI Impact Standard for Glass

Requirements: Sets A and B consist of 3 test specimens

Set ‘A’ corresponds to missile hits near the center of the glass

Set ‘B’ corresponds to missile hits near the corner of the glass

Missiles are 9^{lb} 2×4s

Missile velocity is $50 – 52^{fps}$

The glass is considered successful if two of the three specimens pass in both sets ‘A’ and ‘B’.

p_A and p_B (individual probabilities of success) are unknown.

Evaluate probabilistically!!

Since each specimen test is a pass or fail, it is a Bernoullian test, and we use a Binomial distribution.

$$P(x) = \frac{N!}{(N-x)!\,x!} p^x (1-p)^{N-x}$$

For any set; N = 3, x = 2 or 3 for a pass

$$\Rightarrow P(2) = \frac{3!}{(3-2)!\,2!} p^2 (1-p)^{3-2} = \frac{3!}{2!} p^2 (1-p)^1 = 3p^2(1-p)$$

$$P(3) = \frac{3!}{(3-3)!\,3!} p^3 (1-p)^{3-3} = \frac{3!}{3!} p^3 (1-p)^0 = p^3$$

$$\Rightarrow P(x \geq 2) = 3p^2(1-p) + p^3 = 3p^2 - 2p^3$$

this is the probability of passing any single set!

To pass both sets

$$p(A \text{ and } B) = p(A) \times p(B)$$

$$= (3p_A^2 - 2p_A^3) \cdot (3p_B^2 - 2p_B^3)$$

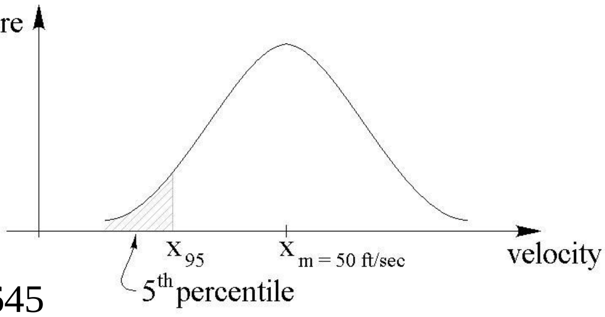
$$= 9\, p_A^2\, p_B^2 - 6(p_A^3\, p_B^2 + p_A^2\, p_B^3) + 4\, p_A^3\, p_B^3$$

$p(A \text{ and } B) \equiv$ probability of passing standard

Gaussian distribution – Example

If the mean failure velocity is 50 ft/sec, what velocity would ensure 95% confidence of passing? Assume $\sigma = 0.2x_m = 10$ ft/sec.

Recall Normal Distribution (Non-Standard)



For fifth percentile,

$$z = -1.645$$

$$\Rightarrow z = \frac{x-50}{\sigma}$$

$$-1.645 = \frac{x-50}{10}$$

$$x = 33.55\,\text{ft/sec}$$

There if we pass the standard, we are 95% certain that the glass will not fail by a missile impact at 33.55 ft/sec!!

Data Processing

(Digital)

Data conditioning

Digital filtering

Windowing – zooming

Commercial software for data processing

Data Conditioning

Data: N samples ‘equally’ spaced at sampling interval Δt.

Definition: $\Delta t \rightarrow \infty$ static measurements

Δt finite small dynamic data

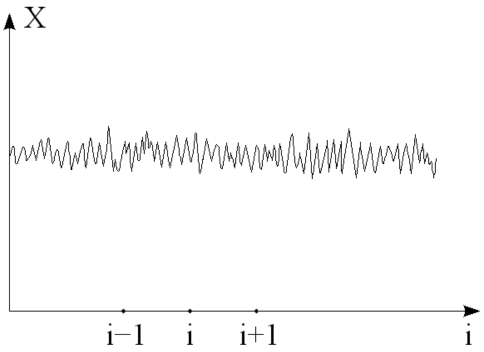
(test: if Δt is significantly smaller or larger than T_0, then T_0 is the largest significant period of the testing system)

*1. **Static measurements***

Sources of inaccuracies and their remedies:

- Noise from instrumentation – smoothing
- Drift – time and temperature effects – trend removal
- DC offset removal – averaging

Noise from instrumentation or other sources



$$\overline{X}_i^* = X_i + n_i$$

Assume that $\overline{n}_i \neq 0$, where $\overline{\ }$ = average

By definition:

$$X_{i,avg}^* = \overline{X}_i^* = \frac{1}{n+1}\sum_{i-\frac{n}{2}}^{i+\frac{n}{2}}\left(X_i + n_i\right) = \overline{X}_i + \overline{n}_i = \overline{X}_i$$

Using averaging of data noise can be eliminated.

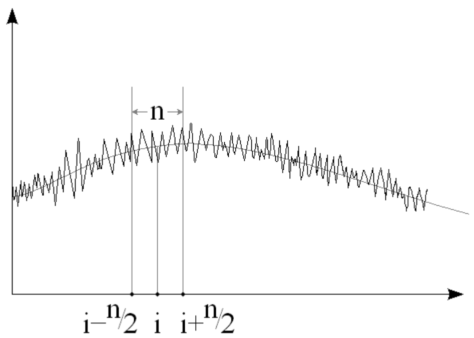
a) Smoothing: $\overline{X}_i^* = \frac{1}{n+1}\sum_{i-\frac{n}{2}}^{i+\frac{n}{2}} X_i^*$

Note that 'noise' is a relative definition. Any fluctuating signal of no importance to the time variation of the main signal qualifies. Smoothing can therefore be used to eliminate AC components from static data.

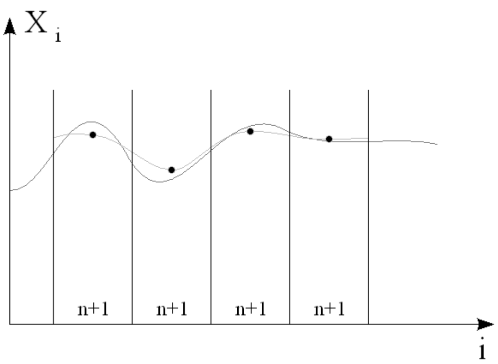
Determine n : Large enough to obtain $\overline{n} = 0$

n : Small enough not to disturb variations of main signal

Use at least three points (n = 2), though seven points (n = 3) is optimal



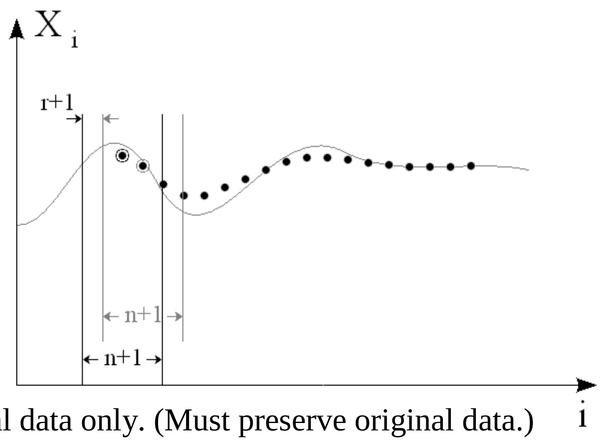
b) Sequential versus overlapping (moving) windows.



Sequential windows

- Smoothing through averaging done in one window (does not preserve original data).
- Loses data points – Data is 'decimated' by division (n + 1)

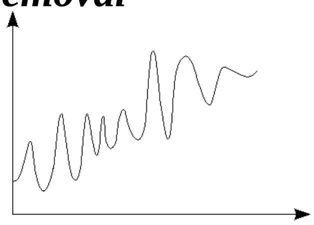
Overlapping (moving) windows



- Smoothing done on original data only. (Must preserve original data.)
- Preserves data (r = 0) or reduces points to $r\Delta t = \Delta t^*$

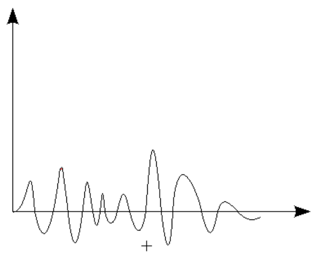
Drift due to time and temperature – trend removal

Drifted signal



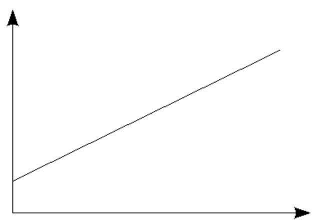
=

Actual



+

Drift



$X_i^* = X_i - \bar{x}_{ni}$; $\bar{x}_{ni} = \sum_{k=0}^{K} b_K (n\Delta t)^K$,

which is a polynomial

Use 'least square fit':

$$b_0 = 2(2N+1)\sum_{n=1}^{n} x_n - 6\sum_{n=1}^{N} n x_n$$

$$b_1 = \frac{\left[12\sum_{n=1}^{N} n x_n - 6(n+1)\sum_{n=1}^{N} x_n\right]}{\Delta t N (N-1)(N+1)}$$

(i) DC offset removal

- This is a particular case of drift, and it affects b_0 only

$$b_0 = \overline{X}$$

This is subtracted from the original (measured) signal.

(ii) Removal of slope: subtract from signal $b_1 X$

Removal of random errors through "Smoothing & Filtering.":

(see notes)

Statistical processing

1) Samples and parameters
2) Distributions
3) Hypothesis tests
4) Chi-square goodness-of-fit tests
5) Correlation analysis
6) Linear and non-linear regression

1. Samples and parameters

Mean: $\mu=\bar{x}=E(x)=\sum xp(x)$ where p(x) is a weighting factor. We assume equal weight.

$$=\frac{1}{N}\sum x$$

Variance: $\sigma_x^2=\frac{1}{N}\sum\left(x_i-\bar{x}\right)^2$ or:

$$=\frac{1}{N-1}\sum\left(x_i-\bar{x}\right)^2$$

produced as a result of an unbiased estimator.

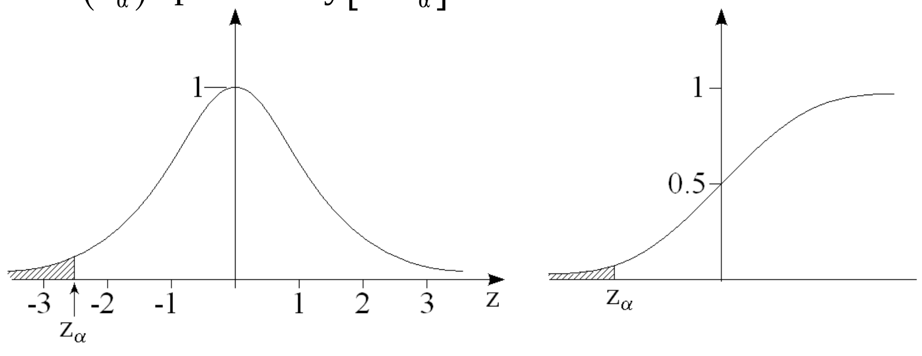
2. Distributions

- **Normal** $z=\frac{\left(x-\mu_x\right)}{\sigma_x}$ normalized (standardized) variable

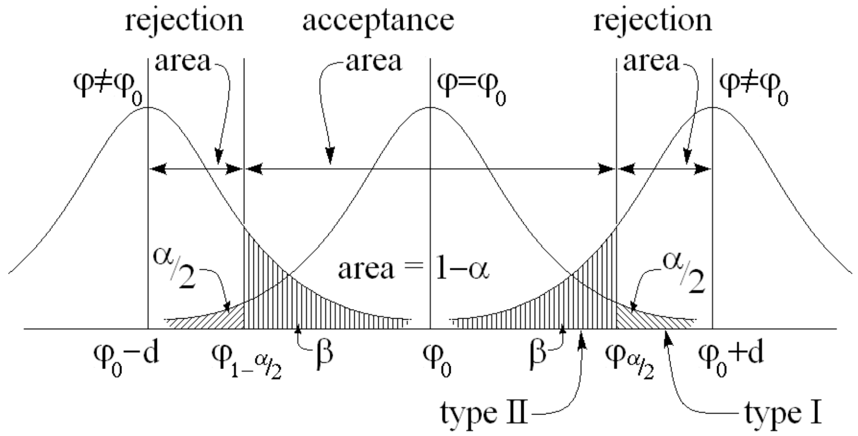
$$p(z)=\left(\sqrt{2\pi}\right)^{-1}\cdot e^{\frac{-z^2}{2}}$$

$$P(z_\alpha)=\text{probability}\left[z\le z_\alpha\right]=1\text{-}\alpha$$

$$1\text{-}P(z_\alpha)=\text{probability}\left[z>z_\alpha\right]=\alpha$$



3. Hypothesis test: measure $\hat{\varphi}$; can we say that population $\varphi=\varphi_0$?



$$P\left(\hat{\varphi}\le\varphi_{1-\frac{\alpha}{2}}\right)=\frac{\alpha}{2}\quad ; \qquad P\left(\hat{\varphi}>\varphi_{\frac{\alpha}{2}}\right)=\frac{\alpha}{2}$$

Type I error: Hypothesis is rejected but is true. (Probability α.)

Type II error: Hypothesis is accepted but is false. (i.e. $\varphi=\varphi_0\text{-}d$ or $\varphi=\varphi_0+d$ (Probability β.)

Example: $\mu_x=10$ $\sigma_x^2=4$

Determine N such that $\mu_x = 20$ at 5% level of significance (= α)
with a probability of type II error of 5% (= β)
in detecting a 10% difference from the hypothesized value (10% μ_x = d).

$$\bar{x}=\frac{\sigma_x}{\sqrt{N}}z+\mu_x$$

Equation 1 upper limit = $\frac{\sigma_x}{\sqrt{N}}z_{\frac{\alpha}{2}}+\mu_x$

For acceptance of $\mu_x = 10$

Equation 2 lower limit = $\frac{\sigma_x}{\sqrt{N}}z_{1-\frac{\alpha}{2}}+\mu_x$

Equation 3 lower limit = $\frac{\sigma_x}{\sqrt{N}}z_{1-\beta}+\mu_x+d$

For error detection of d

Equation 4 upper limit = $\frac{\sigma_x}{\sqrt{N}}z_{\beta}+\mu_x-d$

Equation 1 = Equation 3 ; Equation 2= Equation 4

$$z_{\frac{\alpha}{2}}=z_{1-\beta}+\frac{\sqrt{N}}{\sigma_x}d=-z_{\beta}+\frac{\sqrt{N}}{\sigma_x}d$$

$$N=\left[\frac{\sigma_x\left(z_{\alpha/2}+z_{\beta}\right)}{d}\right]^2$$

For the example:

$\sigma_x=\sqrt{4}=2$; given

$Z_{\alpha/2}$=Z(2.5%)= 1.960

Z_{β} =Z(5.0%) =1.645

d = 10% × 20 = 2 given

N =13

upper limit = 21.0872,

lower limit = 18.9128

4.Chi Square Goodness of Fit Test

Chi-square definition: $\chi_n^2 = z_1^2 + z_2^2 + z_3^2 + ... + z_n^2$

z_i are independent random variables with a mean of zero and a variance of 1 (normally distributed).

n is the number of degrees of freedom (the number of independent variables).

$$p\left(\chi_n^2\right) = \left[2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)\right]^{-1}\left(\chi^2\right)^{\left[\left(\frac{n}{2}\right)-1\right]} e^{\frac{-x^2}{2}}$$ where $\Gamma\left(\frac{n}{2}\right)$ is a Γ function

$$E\left(\chi_n^2\right) = n = \mu_{x^2}$$

$$E\left[\left(\chi_n^2 - \mu_{x^2}\right)^2\right] = 2n = \sigma_{x^2}^2$$

for large values of n; $p\left(\chi_n^2\right) \to p(z)$, normally distributed

Distribution of Sample Mean with Known Variance.

$\bar{x} = \frac{1}{N}\sum x_i$, which is normally distributed, with μ_x and σ_x^2

$$\mu_{\bar{x}} = \mu_x$$

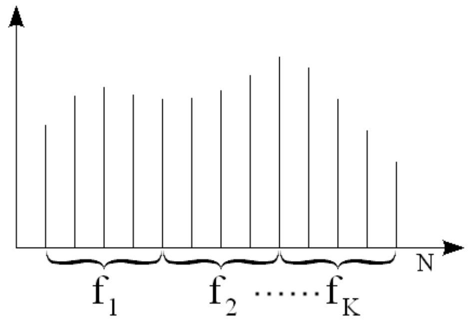
$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{N}$ proven by:

$$z = \frac{\left(\bar{x} - \mu_x\right)\sqrt{N}}{\sigma_x}$$

$$P\left[\bar{x} > \left(\frac{\sigma_x z_\alpha}{\sqrt{N}} + \mu_x\right)\right] = \alpha$$

Uses: Verify equivalence of a probability density function to a theoretical density.

N samples of x with p(x)



Assume F_i is the theoretical frequency of samples in the i^{th} interval.

f_i - F_i is the discrepancy

$$X^2=\sum_{i=1}^{K}\frac{\left(f_i-F_i\right)^2}{F_i}$$

Test for acceptance:

$X^2 \le X^2_{n_i\alpha}$ α is of known (probability of rejection) level of significance

Example: N = 200 observations of temperature, (or strains, or load in a long test (constant load?). (This is given in a table.) Check to see if data is normally distributed.

$$2.43=X^2 \le X^2_{9,0.05}=16.95$$

6. Linear and non-linear regressions

Least squares – multiparameters

(see reference books)

Dynamic measurements:

Transfer Functions and Random Vibrations

Excitation $\rightarrow$ Structure $\rightarrow$ Response

$$m\ddot{x} + kx + c\dot{x} = Fe^{i\omega t} \Rightarrow x = X \cdot e^{i\omega t}$$

$$e^{i\omega t}\left[mX\left(-\omega^2\right) + kX + eX\left(i\omega\right)\right] = Fe^{i\omega t}$$

$$X\left[\omega_0^2 - \omega^2 + 2i\xi\omega\omega_0\right] = \frac{F}{m} \qquad \omega_0^2 = \frac{k}{m}$$

$$\frac{c}{m} = 2\xi\omega_0$$

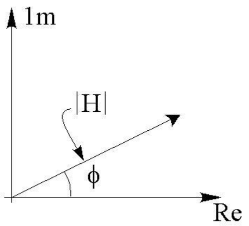
$$F = \frac{1}{m}\left[\omega_0^2 - \omega^2 + 2i\xi\omega\omega_0\right]^{-1} \cdot X$$

Note that: F ≡ amplitude of harmonic forcing function

X ≡ amplitude of response harmonic function

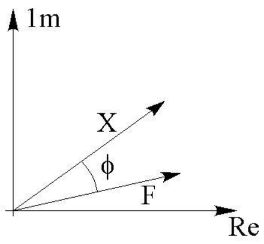
$F = H(i\omega) \cdot X$ complex function

$H(i\omega)$ = complex frequency response function



$$\|H\| = \frac{1}{m\sqrt{\left(\omega - \omega_0\right)^2 + \left(2\xi\omega\omega_0\right)^2}}$$

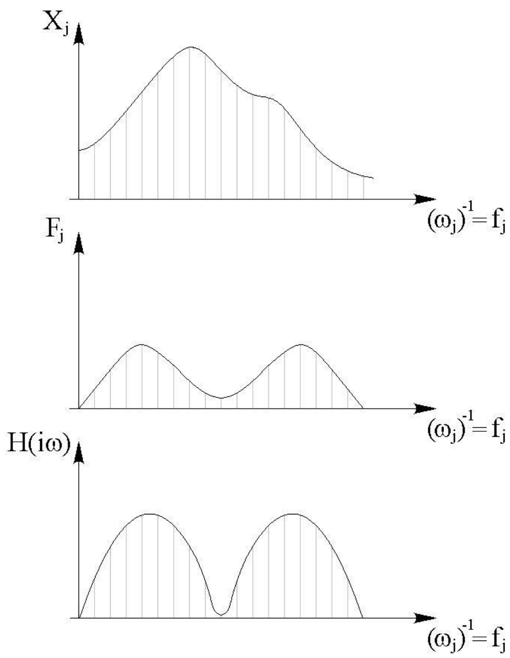
$$\phi = \frac{2\xi\omega\omega_0}{\omega_0^2 - \omega^2}$$

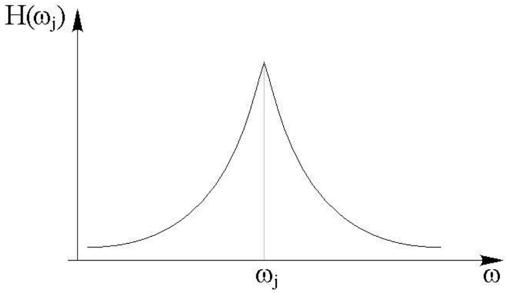


For functions:

$$f = \sum_{j=2}^{\infty} F_j e^{i\omega_j t} \qquad ; \qquad x = \sum_{j=1}^{\infty} X_j e^{i\omega_j t}$$

$$F_j = H(i\omega_j) \cdot X_j$$





$$H(i\omega) = \frac{F_j}{X_j}$$ = (complex numbers)

$$= \frac{F_j X_j^*}{X_j X_j^*} =$$ Transfer function

$S_{HH} = H(i\omega) \cdot H^*(i\omega) =$ real numbers = "power spectrum"

$S_{xy} = X(i\omega) \cdot Y^*(i\omega)$ = cross spectrum

$$\gamma_{xy} = \frac{(S_{xy})^2}{S_{xx} S_{yy}} = \frac{\left[X(i\omega) \cdot Y^*(i\omega) \right]^2}{\|X_{i\omega}\| * \|Y_{i\omega}\|}$$ === coherence function

$$\|X_{i\omega}\| = \left[X(i\omega) \cdot X^*(i\omega) \right] \qquad \|Y_{i\omega}\| = \left[Y(i\omega) \cdot Y^*(i\omega) \right]$$

Modal Analysis / Testing

Modal analysis is the identification of modal characteristics from vibration data.

Modal Characteristics

(1) Modal frequencies

(2) Modal shapes

(3) Phase lags in modal shapes

Testing procedures

(1) Random input – single input, multiple inputs

(2) Harmonic input

The following procedures will consider single random input for the sake of simplicity.

Principles

Identify frequency response functions [H(f)] using frequency domain analysis.

Background information

- Spectral density functions

$$S_{xx} = \int_{-\infty}^{\infty} R_{xx}(T) e^{i2\pi f\tau} d\tau = \lim_{T\to\infty} \frac{1}{T} E\left[X^*(f,T) \cdot X(f,T)\right]$$

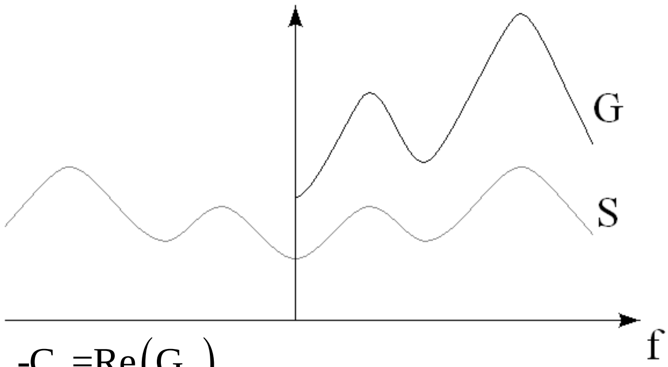
T is the length of the time record in seconds

X, X^* are Fourier Transforms of signal X(t)

$$S_{yy} = \lim \frac{1}{T} E\left[Y^*(f,T) \cdot Y(f,T)\right]$$

$$S_{xy} = \lim \frac{1}{T} E\left[X^*(f,T) \cdot Y(f,T)\right]$$

- Single sided functions



$G_{xx} = 2S_{xx}$

$G_{yy} = 2S_{yy}$

$G_{xy} = 2S_{xy}$

$G_{ij} = C_{ij} - iQ_{ij}$ $\quad$ $-C_{ij} = \mathrm{Re}\left(G_{ij}\right)$

$Q_{ij} = \mathrm{lm}\left(G_{ij}\right)$

Magnitude = $\left|G_{xy}\right| = \sqrt{C_{ij}^2 + Q_{ij}^2}$

Phase = $\theta_{xy} = \tan^{-1}\dfrac{Q_{ij}}{C_{ij}}$

- Transfer functions: x – input, y – output

$H_{yx}(f) * S_x(f) = S_y(f)$ $\quad$ $H(f)\cdot X(f) = Y(f)$

$S_{yy} = \left|H_{yx}(f)\right|^2 \cdot S_{xx} = \left|H_{yx}\right|^2 \cdot S_{xx}$

$S_{yx} = H_{yx} \cdot S_{xx}$

$S_{xy} = \dfrac{1}{T} E\left[X \cdot Y^x\right]$ $\quad$ $S_{xx} = \dfrac{1}{T} E\left[X^x \cdot X\right]$

- Coherence function

$$\gamma_{xy}^2 = \frac{S_{yx}^2}{S_{xx} S_{yy}} = \frac{G_{yx}^2}{G_{xx} G_{yy}}$$

$0 \le \gamma_{xy} \le 1.0$

[This is a frequency function $\gamma_{xy}(f)$]

γ_{xy}^{2} = real function (= real values).

Models with extraneous noise

$x(t) = u(t) + n_x(t)$

$y(t) = v(t) + n_y(t)$

$$G_{xx} = G_{uu} + G_{n_x n_x} + G_{un_x} + G_{n_x u}$$

$$G_{yy} = G_{vv} + G_{n_y n_y} + G_{vn_y} + G_{n_y v}$$

$$G_{xy} = G_{uv} + G_{un_x} + G_{n_y v} + G_{n_x n_y}$$

where:

$$G_{vv} = H^2 \cdot G_{uu} \qquad = G_{vv}, G_{uu} = \text{ideal functions}$$

$$G_{uv} = H \cdot G_{uu}$$

Example:

No input noise, $n_x(t)=0$; Uncorrelated output noise: $x = u$, $y = v + n$

$$G_{n_x n_x} = 0 \qquad G_{un_x} = G_{n_x u} = 0 \qquad G_{vn_x} = 0$$

$$\therefore G_{xx} = G_{uu} \qquad \Rightarrow G_{xx} \text{ is the ideal function}$$

$$\therefore G_{yy} = G_{vv} + G_{nn}$$

$$\therefore G_{xy} = G_{uv} = H \cdot G_{uu} = H \cdot G_{xx}$$

Transfer Function:

$$\left|\overline{H}\right|^2 = \frac{G_{yy}}{G_{xx}} = \frac{G_{vv}+G_{nn}}{G_{uu}} = \left|\overline{H}\right|^2 + \frac{G_{nn}}{G_{uu}}$$ = contaminated with noise

If the transfer function is determined from Cross Power Spectrum:

$$\overline{H} = \frac{G_{xy}}{G_{xx}} = \frac{G_{uv}}{G_{uu}} = H$$

then it is an ideal transfer function,

not influenced by noise

Important

For uncorrelated noise:

$$G_{ny} = 2 \cdot \lim_{T\to\infty} \frac{1}{T} E\left[Y(f) \cdot N(f)\right] \Rightarrow 0$$

[The mathematical expectation E[a·b] reduces the average of noise to zero.]

$$G_{nn} = \left[1-\gamma_{xy}^2\right] \cdot G_{yy} =$$ obtained from basic definitions

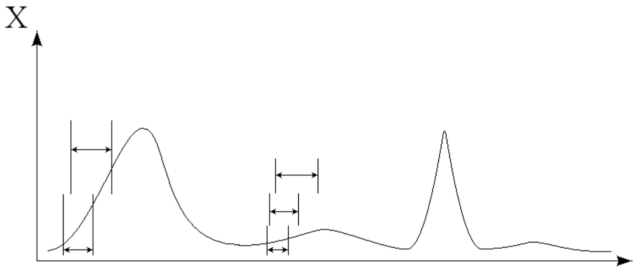
[Helps in eliminating noise.]

Note that Spectral Density Functions are also named as “Power Spectral Density” (PSD) or Power Spectrum with definitions of <u>Auto</u> Power Spectrum (G_{xx}, G_{yy}) or <u>Cross</u> Power Spectrum (G_{xy}).

Numerical Analysis – Modal

(1) Determined transfer function from spectral densities
(2) Identify peak values using moving average window

(window size determined for desired resolution)



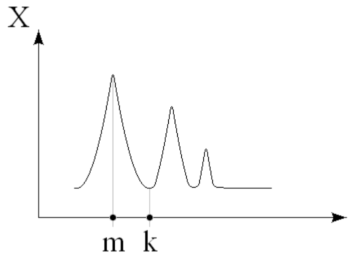
Determine peaks and valleys:

Peak = Max $(X_i)_{i=k+\ldots}$

k is the point of the identified valley.

Valley = Min $(X_i)_{i=m+\ldots}$

m is the point of the identified peak.



M_i = points determine identified frequencies

f_i = frequencies = $M_i \cdot \Delta f$

Δf = frequency resolution

T = time length of input record = $N \cdot \Delta t$

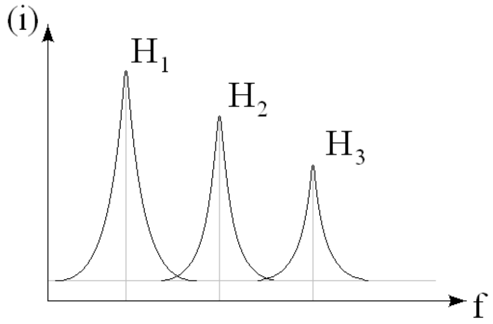
N = number of points in acquisitioned data used for analysis (FT)

Δt = time interval (sampling interval)

(3) Identify peak values from original record or averaged record using small window (of not more than three points).

(4) Analyze transfer function for modal shapes:

$$H_{ix}(f) = \sum_{1}^{N} \varphi_{ij}\Gamma_j H_j(f)$$



i = output location - for multiple output i > 1

- For well separated modes:

$$H_{ix}(f_j) = \varphi_{ij} \times \Gamma \times_j H_j(f_j)$$ where f_j is the resonant frequency of mode j

$$H_{kx}(f_j) = \varphi_{kj} \times \Gamma_j \times H_j(f_j)$$

$$\frac{H_{ix}}{H_{kx}} = \frac{\varphi_{ij}}{\varphi_{kj}} \times \frac{\Gamma_j \times H_j(f)}{\Gamma_j \times H_j(f)} = \frac{\varphi_{ij}}{\varphi_{kj}} = \overline{\varphi_{ij}}$$

(normalized to φ_{kj})

[Normalize to the largest value you have in the record! All other values produce errors. (Homework exercise; prove that using the lowest value of H_{kx} for normalization, produces the largest errors.)]

- For not well separated modes use two set of equations and solve

(5) Determine the magnitude and phase shifts of each identified mode:
[θ = 0° or 360° - in phase; θ = 90° or 360° - delayed response (no real shape; θ = 180° - out of phase.]

Structure Identification

For ideal normal modes:

$\Phi^T \cdot M \cdot \Phi = I$ for orthonormalized mode shapes.

Orthonormalization:

$$[\Phi_i]^*_{fromidentification} * \left[M^{-\frac{1}{2}} \right] = \Phi$$

$$\Phi^T K \Phi = \begin{bmatrix} \ddots & & \\ & \omega^2 & \\ & & \ddots \end{bmatrix} = \Omega^2$$

$K = \Phi^{-T} \Omega^2 \Phi^{-1}$ where: $\Phi^{-1} = \underline{\Phi}^T \underline{M}$ $\Phi^{-T} = \underline{M}\underline{\Phi}$

- Procedure

(a) Determine mode shapes (as number of DOF desired).

(b) Determine mass matrix in desired DOF.

(c) Determine stiffness matrix:

$K = M \Phi \Omega^2 \Phi^T M$

How do you determine the damping matrix?

Damping Identification

Identify component frequency functions by curve fitting.

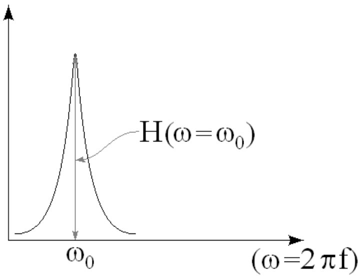
$$H_{ix}^{(\rho)} = \sum \varphi_{ij} \cdot \Gamma_j \cdot H_j(f) \quad ; \quad H_j(f) = \frac{1}{\left(1-\frac{\omega^2}{\omega_j^2}\right) + 2\xi_j \frac{\omega}{\omega_j} i}$$

Determine φ_{ij}, ω_j, ξ_j from functions.

(Suggested home exercise: Develop a technique to identify modes too closely spaced or highly damped.)

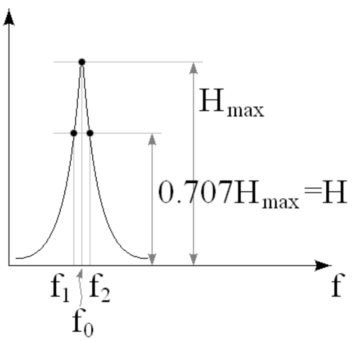
Identification of damping

1.



$$\xi \cong \frac{1}{2H(f=f_o)}$$

2.



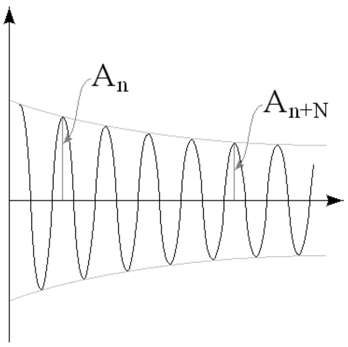
$$\left(H^2 = H_{max}^2 / 2\right)$$

$$\xi = \frac{f_1 - f_2}{f_1 + f_2} = \frac{f_1 - f_2}{2f_0}$$

$$f_0 = \frac{f_1 + f_2}{2}$$

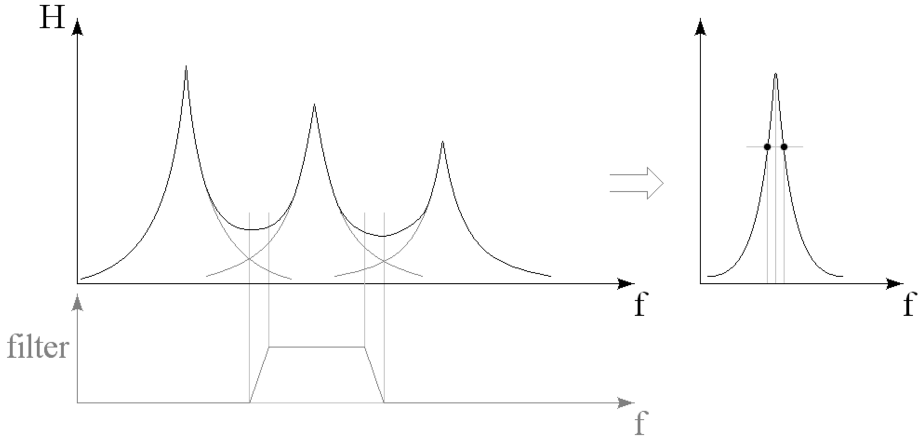
(true for symmetric functions)

3.



$$\xi = \frac{1}{2\pi N} \ln \frac{A_n}{A_{n+N}} \left(\sqrt{1-\xi^2}\right)$$

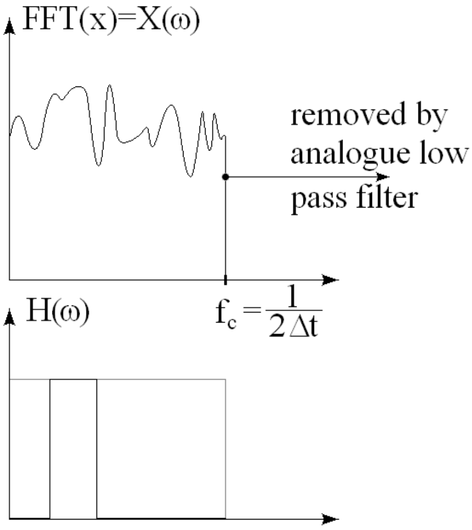
where $\sqrt{1-\xi^2} \cong 1$


H
f
filter
f
f

Digital Filtering

Frequency domain filtering

- Prepare FFT. = X(ω)
- Choose H(ω) for filter
- Make transform. IFT[H(ω) · X(ω)]



Time domain filtering

- Non-recursive filters $y_i = \sum_{k=0}^{M} h_k \cdot x_{i-k}$

convolution integral $\equiv \text{IFT}(......) \equiv \int_0^\infty h(\tau)\, x(t-\tau)\, d\tau$

- Recursive filters $y_n = \sum_{k=1}^{M} t_k \cdot y_{n-k} = cx_n + by_{n-1}$

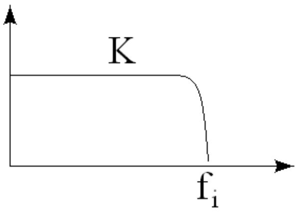
Non-recursive digital filters to use

1. $$|H(f)|^2 = \frac{1}{1+\left(\frac{\sin\pi f/\Delta t}{\sin\pi f_0 \Delta t}\right)^{2M}} \quad \ldots\ 0 \le f \le \frac{1}{2\Delta t} = \text{Nyquist Freq.}$$

2. Butterworth filter

$$|H(f)|^2 = \frac{1}{1+\left(f/f_0\right)^k}$$

here K determines the rolloff rate



Recursive filters

$$y_n = (1\text{-}a)\,x_n + a y_{n\text{-}1} \text{ for time domain use}$$

$$a = \exp\left(\frac{-\Delta t}{RC}\right)$$

$$RC = \frac{1}{2\pi f_0}\left[\frac{1}{|H(\omega)|^2} - 1\right]^{\frac{1}{2}} \left(-\Delta t/RC\right) = \text{time constant of filter}$$

Δt is the sampling rate

f_0 is the cutoff frequency with specified reduction

Example:

1000 sps (Δt = 1 msec)
f_0 = 10 Hz, 0.5 half power reduction:

$$\Delta t = 10^{-3}$$

$$RC = \frac{1}{2\pi f_0}\left[\frac{1}{|H(\omega)|^2} - 1\right]^{\frac{1}{2}} = 0.016$$

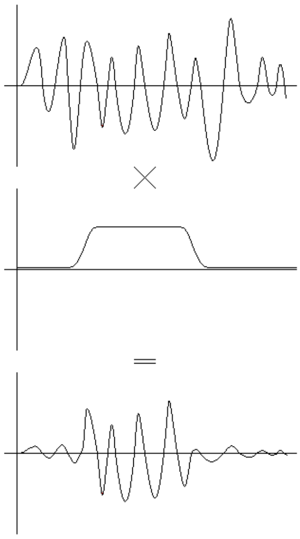
$$a = \exp\left(\frac{10^{-3}}{0.016}\right) = 0.94$$

so $$y_n = 0.06x_n + 0.94y_{n-1}$$

Windowing – zooming

(a) Time domain

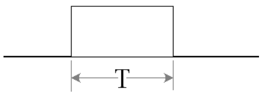
(b) Frequency domain



(a) Time Domain

- required for FFT
- reduce signal

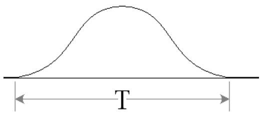
Rectangular



$$\begin{cases} a & t_1<t_i<t_2 \\ 0 & \text{otherwise} \end{cases}$$

and $t_2 - t_1 = T$

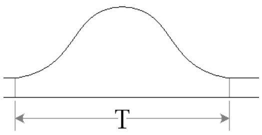
Hanning



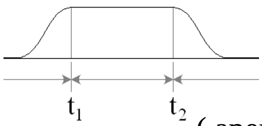
$$\begin{cases} a\left(1-\cos^2\frac{t}{T}\right) \\ 0 & \text{otherwise} \end{cases}$$

and $t_1 < t < t_2$

Hamming



Flat top / Gaussian



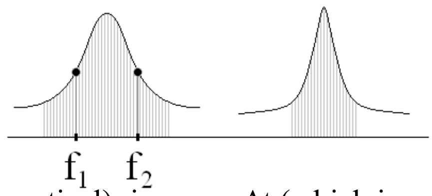
$$\begin{cases} ap(t) & -\infty<t<t_1 \\ a & t_1<t<t_2 \\ ap(t) & t_2<t<\infty \end{cases}$$

(apex of Gaussian function is at t_1 and t_2)

Filters do not distort amplitude, and introduce errors at only two frequencies.

Zoom transform procedures

For damping identification using the half bandwidth method or using peak filtered techniques better frequency resolution is required:

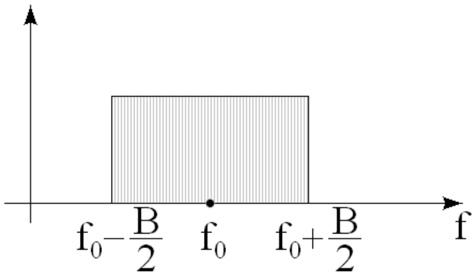


- Regular frequency resolution:

$\Delta f = \dfrac{1}{N\Delta t}$: increase N (which is impractical); increase Δt (which is also impractical).

- Band filter data (i.e. uniform):

$$y(t) = \begin{cases} x(t) \\ \varphi \end{cases} \qquad f_0 - \left(\frac{B}{2}\right) \le f \le f_0 + \left(\frac{B}{2}\right)$$

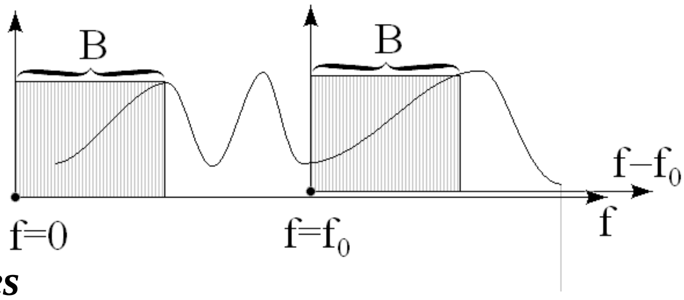


- Modulate signal:

$\bar{y}(t) = y(t) \cdot e^{i2\pi f_0 t}$ where f_0 is an arbitrary frequency

$\bar{Y}(f) = \int_0^T y(t) \cdot e^{i2\pi f_1 t} \cdot e^{-i2\pi ft} dt = Y(f - f_1) \equiv$ shift of frequency

- Use frequency transform and obtain frequency band desired.



Advantages

If for original record we have $f_{max}=nB$ windows:

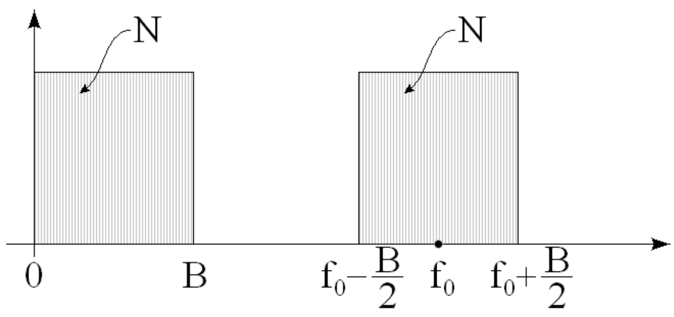
$\Delta t = \dfrac{1}{Nf_{max}} = \dfrac{1}{N \cdot nB}$ (small time step) or $(N \cdot n)_{samples} = \dfrac{1}{\Delta t B}$

For a zoomed window (f_{max}=B)

$$\Delta t = \frac{1}{Nf_{max}} = \frac{1}{NB}$$; No reduction of time step or increase in number of samples is required

- Zooming

Using modulation no increase in number of samples or reduction of samples rate is required.



Example:

$\Delta f = 1$ Hz (desired frequency resolution)

$N = 1024$ (total number of samples per block)

$f_{max} = 5120$ Hz (desired measurement bandwidth)

1. Sample rate: $$\Delta t = \frac{1}{N\Delta f} = 0.000977$$

2. Nyquist Frequency: $$f_c = \frac{1}{2\Delta t} = 512\text{Hz}$$ (≡ B for window).

3. To analyse data over 5120 Hz it is required to have 10 sequential zoomed windows of 512 Hz.

4. Maximum error in a record is related to the number of averages used in the computation of the expectation as follows:

$$\varepsilon_r = \frac{1}{\sqrt{n_d}}$$

where

$\varepsilon_r \equiv$ normalized random error

n_d = number of distinct records of length T.

For example: $\varepsilon_{r = 0.10}$ → $n_{d = 100}$ number of distinct

$$\text{Time} = n_d \cdot N \cdot \Delta t = 100 \times 1024 \times 0.000977 = 100\,\text{sec}$$

Without zooming technique: