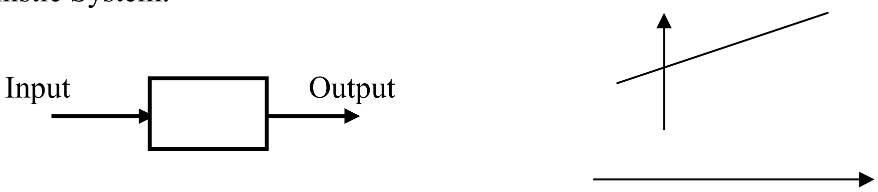
2.160 Identification, Estimation, and Learning
Lecture Notes No. 4
February 17, 2006

3. Random Variables and Random Processes

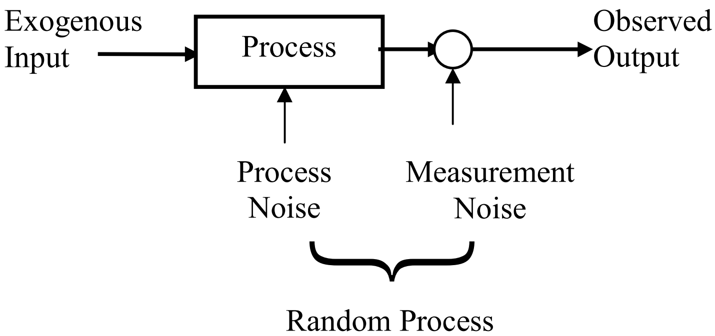
Deterministic System:



In realty, the observed output is noisy and does not fit the model perfectly. In the deterministic approach we treat such discrepancies as “error’.

Random Process:

An alternative approach is to explicitly model the process describing how the “error” is generated.



Objective of modeling random processes:

- Use stochastic properties of the process for better estimating parameters and state of the process, and
- Better understand, analyze, and evaluate performance of the estimator.

3.1 Random Variables

You may have already learned probability and stochastic processes in some subjects. The following are fundamentals that will be used regularly for the rest of this course. Check if you feel comfortable with each of the following definitions and terminology. (Check the box of each item below.) If not, consult standard textbooks on the subject. See the references at the end of this chapter.

1) Random Variable

X: Random variable is a function that maps every point in the sample space to the real axis line.

2) Probability distribution $F_x(x)$ and probability density function $f_x(x)$; PDF

$$F_X(x) = \text{Prob}(X \le x)$$

$$f_x(x) = \frac{d}{dx} F_x(x)$$

In the statistics and probability literature, the convention is that capital X represents a random variable while lower-case x is used for an instantiation/realization of the random variable.

3) Joint probability densities
Let X and Y be two random variables

$f_{XY}(x,y) =$ Prob (X = x and Y = y, simultaneously)

4) Statistically independent (or simply Independent) random variables

$$f_{XY}(x,y) = f_X(x)\, f_Y(y)$$

5) Conditional probability density

$$f_{X|Y} = \Pr ob\left(X \middle| \; given \quad Y = y\right) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

= (Joint probability density divided by Probability density of Y=y)

If X and Y are independent,

$$f_{X|Y} = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x)$$

Occurrence of $Y = y$ does not influence the occurrence of $X = x$.

6) Bayes Rule

$$f_{X|Y} = \frac{f_{Y|X}(x)f_X(x)}{f_Y(y)} \qquad \because f_{X|Y}f_Y(y) = f_{XY}(x,y) = f_{Y|X}(x)f_X(x)$$

7) Expectation

Expected value of X $\Longrightarrow$ $E[X] = \int_{-\infty}^{\infty} xf_X(x)dx$

mean, average

For discrete random variables,

$$E[X] = \sum_i p_i x_i$$

8) Variance

$$VarX = E[(X - E(X))^2] = E[X^2 - 2XE(X) + \left(E(X)\right)^2] = E[X^2] - \left(E(X)\right)^2$$

9) Moment

k-th moment of X

$$E[x^k] = \int_{-\infty}^{\infty} x^k f_X(x)dx$$

k=1 mean, k=2, k=3 higher moments

10) Normal (Gaussian) Random Variable

$X \sim N(m_x, \sigma^2)$: Random variable X has a normal distribution with mean m_x and variance σ^2

The first and second moments completely characterize the distribution.

$$f_X(x) = \frac{1}{2\pi\sigma}\exp[-\frac{1}{2\sigma^2}(x - m_x)]$$

$$\sigma^2 = \int_{-\infty}^{\infty}(x - m_x)^2 f_X(x)dx$$

11) ***Correlation***

The expectation of the **product** of two random variables, X and Y

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\, dxdy$$

Joint probability

If X and Y are independent

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_X(x) f_Y(y)\, dxdy \\ &= \int_{-\infty}^{\infty} x\, f_X(x)\, dx \cdot \int_{-\infty}^{\infty} y f_Y(y)\, dy \\ &= E[X]E[Y] \end{aligned}$$

Note that, although the correlation is zero, the two random variables are not necessarily independent.

12) ***Orthogonality***

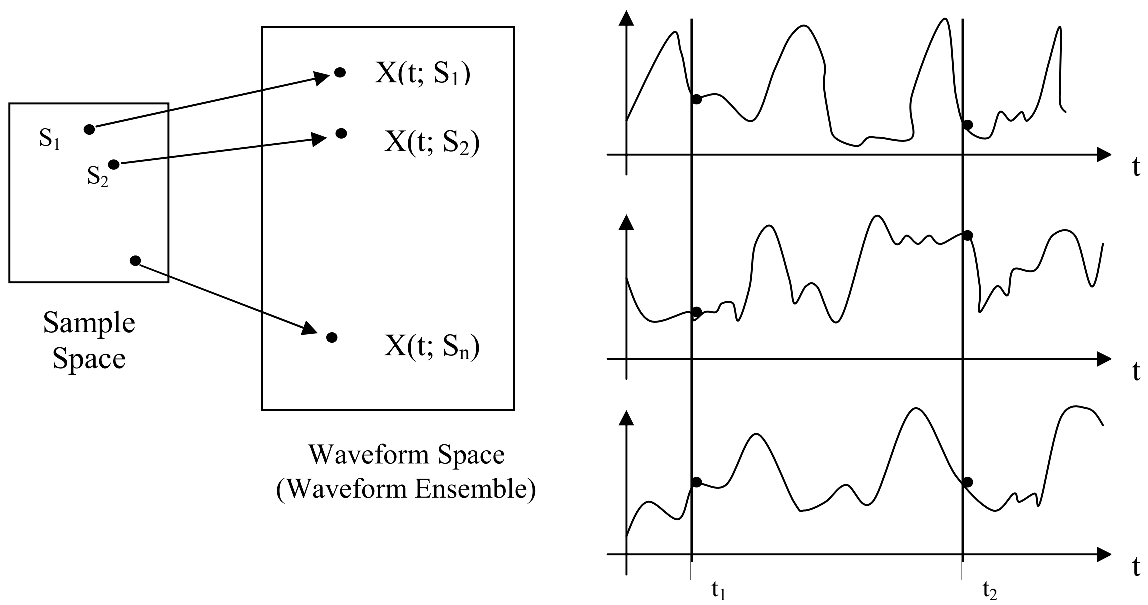
X and Y are said to be orthogonal, if $E[XY] = 0$

13) ***Covariance***

Covariance of X and $Y = E[(X-m_X)(Y-m_Y)]$

3.2 Random Processes

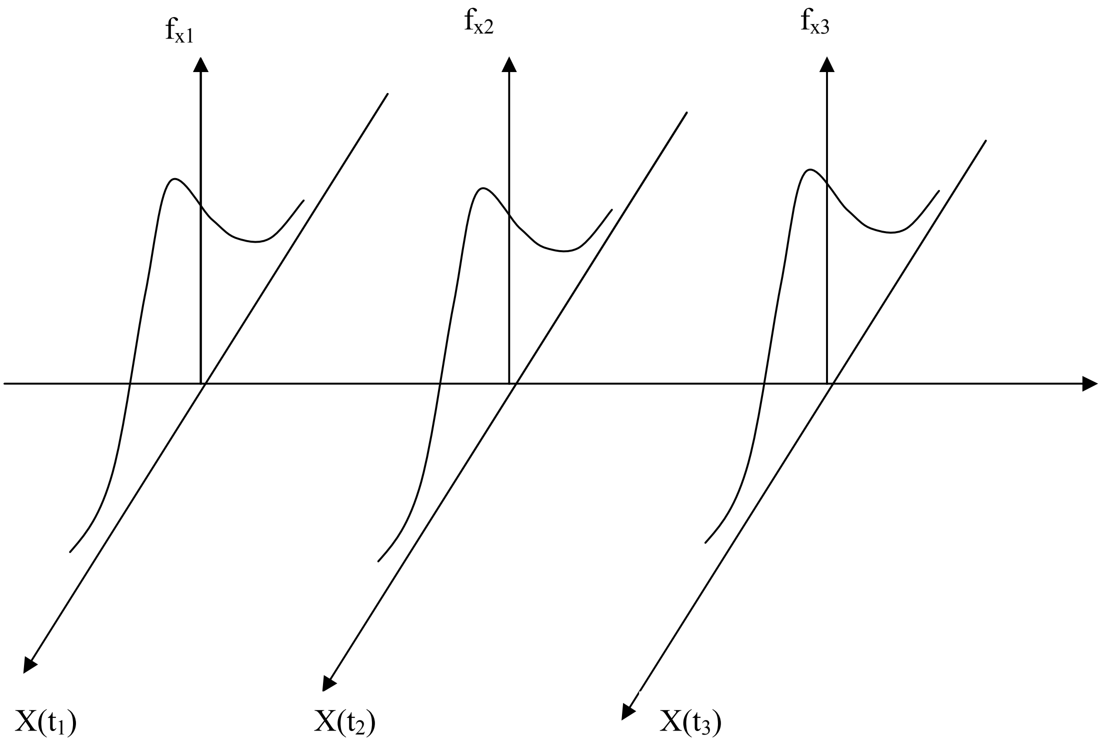
Random Variable X ⟶ Random Process $X(t)$



A Random Process = a family (ensemble) of time functions having a probability measure

Characterization of a random process

- First-order Densities



- Second-order densities

Consider a joint probability density $f_{X1X2}(x_1,x_2)$

If the random process has some correlation between two random variables $X(t_1)$ and $X(t_2)$, it can be captured by the following autocorrelation: ("auto" means correlation within the single same random process)

$$R_{XX}(t_1,t_2) = E[X(t_1)\,X(t_2)] = \iint_{-\infty}^{+\infty} x_1 x_2\, f_{X_1X_2}(x_1,x_2)\,dx_1 dx_2$$

If the auto-correlation function depends only on the time difference $\tau = t_1 - t_2$, it reduces to

$$R_{XX}(\tau) = E[X(t+\tau)X(t)] \qquad\qquad R_{XX}(\tau) = R_{XX}(-\tau) \text{ even function}$$

Then the process is called "Wide Sense Stationary".

Auto-covariance:

$$C_{XX}(t_1,t_2) = E[(X(t_1) - m_X(t_1)) \cdot (X(t_2) - m_X(t_2))]$$
$$= R_{XX}(t_1,t_2) - m_X(t_1)m_X(t_2)$$

- Higher-order densities

  Joint density $f_{X1X2....Xn}(x_1, x_2, ..., x_n)$
  $E[X(t_1)X(t_2)\ldots X(t_n)]$ n-th order

  Total characterization: If joint densities of $X(t_1)$, $X(t_2)$, ….., $X(t_n)$ for all n are known, the random process is said to be totally (completely) characterized. … Unrealistic.

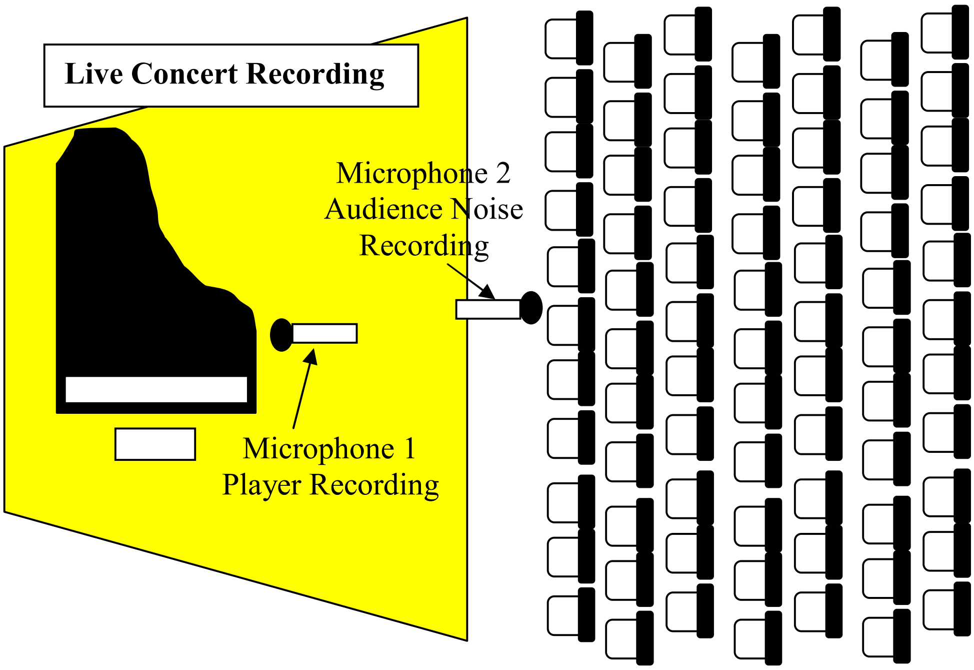
Reference Textbook on Random processes

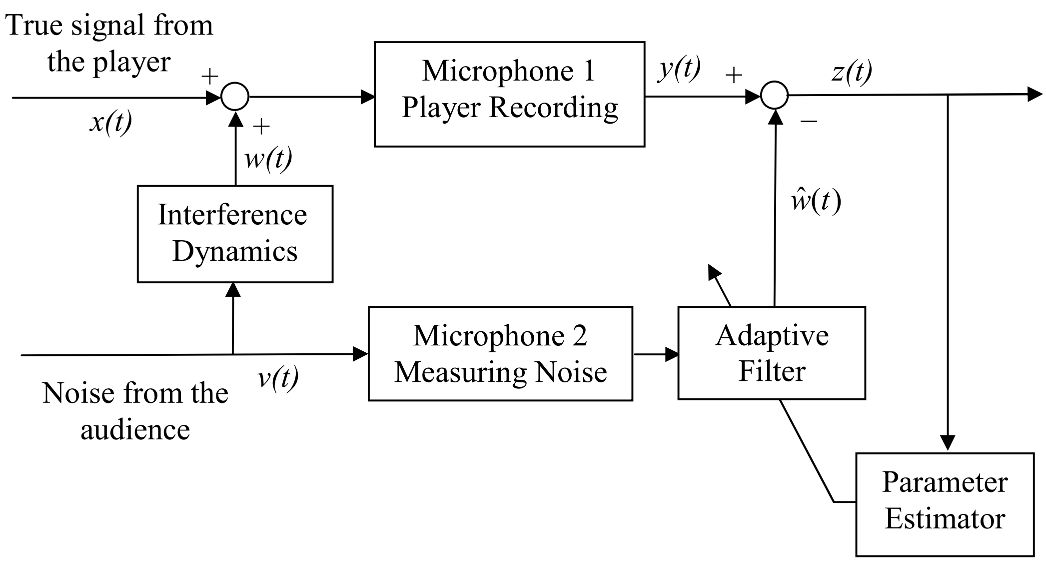
Lonnie Ludeman , "Random Processes - Filtering, Estimation, and Detection", Wiley 2003, ISBN 0-471-25975-6

Robert Brown and Patrick Hwang, "Introduction to Random Signals and Applied Kalman Filthering, Third Edition", Wiley 1997, ISBN 0-471-12839-2, TK5102.9.B75

Application: Adaptive Noise Cancellation

Let us consider a simple example of the above definitions and properties of random processes. Active noise cancellation is a technique for canceling unwanted noise by measuring the noise source as well as the signal source. It was originated by Widrow's research in the 60's, but recently this technique was used for advanced headsets, like "Bose". So, people are using it daily, without knowing it. Here is how it works.





Process:

The recorded signal

$$y(t) = x(t) + w(t)$$

Assumed interference dynamics (FIR):

$$w(t) = b_1 v(t-1) + b_2 v(t-2) + \cdots + b_m v(t-m)$$

Interference dynamics model (FIR):

$$\hat{w}(t;\hat{\theta}) = \hat{b}_1 v(t-1) + \hat{b}_2 v(t-2) + \cdots + \hat{b}_m v(t-m) = \varphi^T(t) \cdot \hat{\theta}$$

Noise cancellation:

$$z(t:\hat{\theta}) = y(t) - \hat{w}(t:\hat{\theta})$$

Problem: Tune the FIR parameters $\hat{\theta}$ so that the recovered signal $z(t:\hat{\theta})$ is as close to the original true signal *x(t)* as possible. Assume that the interference *w(t)* is strongly correlated to noise *v(t)* and that the true signal *x(t)* is uncorrelated with the noise *v(t)*. Consider the expectation of the squared output $z(t:\hat{\theta})$, i.e. the average power of signal $z(t:\hat{\theta})$,

$$\begin{aligned} E[z(t:\hat{\theta})^2] &= E[\{x(t) + w(t) - \varphi^T(t) \cdot \hat{\theta}\}^2] \\ &= E[x^2(t)] + 2E[x(t)\{w(t) - \varphi^T(t) \cdot \hat{\theta}\}] + E[\{w(t) - \varphi^T(t) \cdot \hat{\theta}\}^2] \end{aligned}$$

Our objective is to find the parameter vector $\hat{\theta}$ that minimizes the mean squared error $E[\{w(t) - \varphi^T(t) \cdot \hat{\theta}\}^2]$. Examining the second term:

$$E[x(t)w(t)] = E[x(t)b_1 v(t-1)] + E[x(t)b_2 v(t-2)] + \cdots + E[x(t)b_m v(t-m)] = 0$$

$$E[x(t) \cdot \varphi^T(t)\hat{\theta}] = 0$$

Therefore, minimizing the average power of $z(t:\hat{\theta})$ with respect to parameter vector $\hat{\theta}$ is equivalent to minimizing $E[\{w(t) - \varphi^T(t) \cdot \hat{\theta}\}^2]$,

$$\hat{\theta} = \arg\min_{\theta} E[\{w(t) - \varphi^T(t) \cdot \theta\}^2] = \arg\min_{\theta} E[z(t:\theta)^2]$$

since $E[x(t)^2]$ is not a function of the parameter vector and is not relevant to minimization of the squared error.

We can use the Recursive Least Squares algorithm with forgetting factor α ($0 < \alpha \le 1$):

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P_{t-1}\varphi(t)}{\alpha + \varphi^T(t)P_{t-1}\varphi(t)}\{y(t) - \varphi^T(t)\hat{\theta}(t-1)\}$$

$$P_t = \frac{1}{\alpha}\left(P_{t-1} - \frac{P_{t-1}\varphi(t)\varphi^T(t)P_{t-1}}{\alpha + \varphi^T(t)P_{t-1}\varphi(t)}\right)$$