Lecture 2

Wednesday, January 12, 2022 10:05 AM

Admin.

① HW 1 due Feb 2nd

② 01 OH today.

objectives

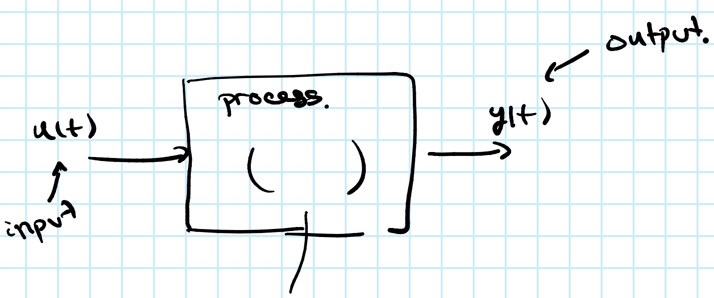
① overview of parameter estimation

② Random variables and probability/density funcns.

③ Formulate student's t-test as a parameter estimation problem.

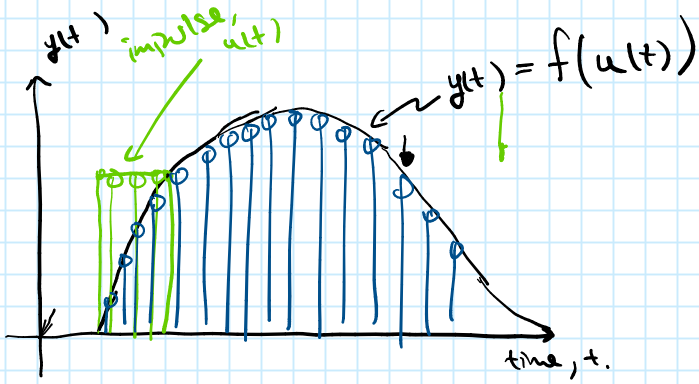
overview of parameter estimation => "system identification"

- Goals of model & parameter estimation.
  - mathematical model to predict future behaviors
  - Ask if two pieces of data define distinct processes.
  - Extract processes or signatures from the data that are non-obvious.
    - discerning fetal ECG signals
    - eliminating artifacts from noisy signals.
    - genes related to disease
    - many others.
- Consider a deterministic process



- Car response to accelerator

- Car response to accelerator
- robot arm response to motor
- immune cell response to a pathogen.



- Finite impulse response model:

$$y(t) = b_1 u(t-1) + b_2 u(t-2) + \cdots$$

parameters to be identified

$$\Rightarrow y(t) = \varphi^T(t)\,\theta \quad \leftarrow \text{deterministic}$$

where

$$\theta = [b_1 \quad b_2 \quad b_3 \quad \cdots]^T$$

$$\varphi(t) = [u(t-1) \quad u(t-2) \quad \cdots]^T$$

- Noise corrupted:

$$\hat{y}(t) = \varphi^T(t)\,\theta + e(t)$$

noise. — measurement noise, biological, wear.

- We can estimate parameters, θ, by using $u(t)$, $y(t)$.

$$V(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[\left(y(t) - \hat{y}(t)\right)^2\right]$$

cost function — $V(\theta)$; actual outputs — $y(t)$; modeled outputs — $\hat{y}(t)$

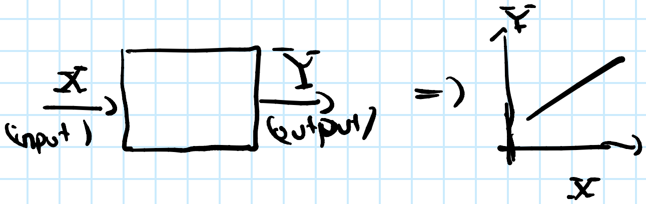
Random variables & probability Density Funcns.

- Deterministic vs. Random

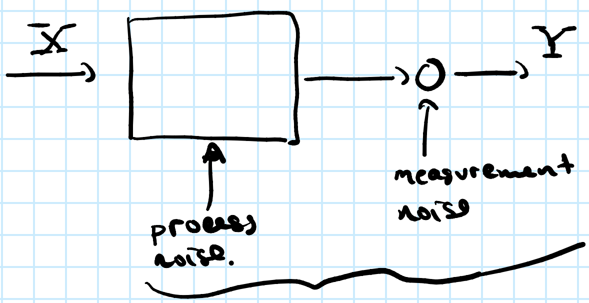
Funcns.

- Deterministic vs. Random processes.

Deterministic



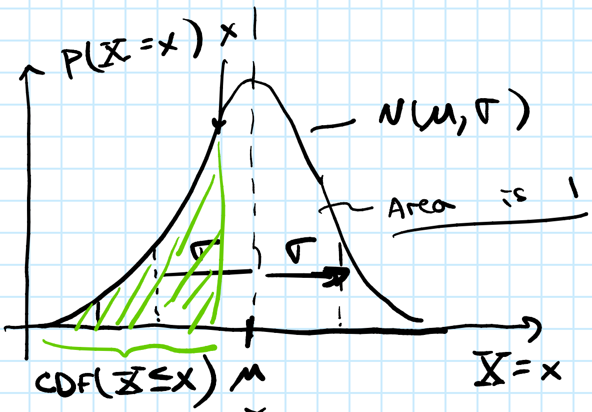
Random process



- In real systems, we need to understand the properties of noise to estimate system parameters.

- Random variables.

  - A random variable is a funcn. that maps every point in sample space onto the the real number line:



← we frequently assume this type of noise

$$p(X = x) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$CDF(x) = \int_{-\infty}^{x} p(x')\,dx'$$

MATLAB:	cdf(x)
R :	ecdf(x).

$$\mathrm{CDF}(x) = \int_{-\infty} p(x')\,dx'$$

MATLAB: cdf(x)

R : ecdf(x).

- $$E[X] = \int_{-\infty}^{\infty} X \cdot p(x)\,dx \quad \text{(continuous)}$$

or

$$E[X] = \Sigma p_i x_i \quad \text{(discrete)}$$

- $$\mathrm{Var}[X] = E[\underbrace{X - E[X]}_{\text{deviation from mean.}}]^2 \leftarrow E[X^2] - 2E[X]E[X] + (E[X])^2$$

- $$\mathrm{std}(X) = \sigma_X = \sqrt{\mathrm{Var}(X)}$$

- Statistic independence.

  -> consider two random vars, X, Y: statistically independent.

  $$p_{XY}(x,y) = p_X(x)\, p_Y(y)$$

- Conditional probability density:

  $$p_{X|Y} = p(X \mid \text{given } Y=y) = \frac{p_{XY}(x,y)}{p_Y(y)}$$

- If X & Y are independent, then

  $$p_{X|Y} = p_X(x)$$

- Correlation

  The expectation of the product of 2 random variables, X & Y is called "correlation":

  $$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, p_{XY}(x,y)\,dx\,dy$$

  - If independent: $E[XY] = E[X]E[Y]$

  - Note: correlation $=0$ does not imply independence.

- Note: correlation $=0$ does not imply independence.

- orthogonality:

  X & Y are said to be orthogonal if correlation is zero: $E[XY]=0$.

- Covariance:

  $$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$