Inference

- Let y_i denote observations from some process, & let $f(\underline{x}_i ; \vec{\theta})$, where $\underline{x}_i$ is independent variable & $\vec{\theta}$ parameters.
  - We assume f approximates the true process. Then

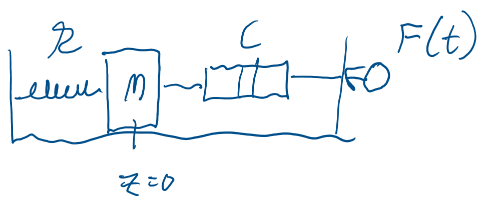
$$Y_i = f(\underline{x}_i ; \vec{\theta}) + \varepsilon_i , \quad i = 1, \dots, N$$

where ε_i accounts for difference between y_i & f.

- Y_i: random observations (real. y_i)
- ε_i: random measurement errors (real ε_i)
- $\vec{\theta}$: Parameters
  - Frequentist: fixed, unknown
  - Bayesian: random variables

Observation Models

- Let us consider $f(\underline{x}_i ; \vec{\theta})$ where $\underline{x}_i$ represents ind. variable. ($x, \vec{x}, t$, or combo)
  - Y_i denotes observations.



Ex) Spring eq. $F = ma$

z: Displacement of mass

$$\Rightarrow \quad m \frac{d^2 z}{dt^2} + c \frac{dz}{dt} + k z = F(t) = f_0 \cos(\omega_F t) \qquad (\mathrm{I})$$

$$z(0) = z_0 , \quad \left.\frac{dz}{dt}\right|_{t=0} = v_0 , \quad t \in [0, T]$$

- Recall that we can rewrite (I) as

- Recall that we can rewrite (I) as

$$\frac{d\vec{u}}{dt} = \underset{\sim}{A}\vec{u} + \vec{F}, \quad \vec{u} = \begin{bmatrix} z \\ v \end{bmatrix}, \quad \vec{u}_0 = \begin{bmatrix} z_0 \\ v_0 \end{bmatrix},$$

$$\underset{\sim}{A} = \begin{bmatrix} 0 & 1 \\ -k/m & -c/m \end{bmatrix}, \quad \vec{F} = \begin{bmatrix} 0 \\ f_0/m \cos(\omega_F t) \end{bmatrix}$$

- Note $\vec{\theta} = [m, c, k]$ are parameters if $\vec{u}_0$ + $F(t)$ are known.

- Suppose we have a sensor that provides measurements at discrete times t_j. Then:

$$f(t_j; \vec{\theta}) = \vec{C} \cdot \vec{u}(t_j; \vec{\theta})$$

- We call $\vec{C}$ the observation operator
  - $\vec{C} = [0, 1] \Rightarrow$ Velocity measurements
  - $\vec{C} = [1, 0] \Rightarrow$ Displacement measurements
  - $\vec{C} = [1, 1] \Rightarrow$ Full System

$\Bigg\}\quad C: \mathbb{R}^{n_u} \to \mathbb{R}^{n_y}$

Statistical Observation Models

- Here, we relate $f(t_i; \vec{\theta}) = C(u(t_j; \vec{\theta}))$, to Y_i + its possible measurement error.

Additive Noise Models

- For additive measurement error, we have:

$$Y_i = f(t_i; \vec{\theta}) + \varepsilon_i, \quad i = 1, \dots, n_y$$

$$Y_i = f(x_i; \bar{\theta}) + \varepsilon_i, \quad i = 1, \ldots, n_y$$

- we typically assume $\varepsilon_i \sim N(\mu_\varepsilon, \sigma^2)$, $\mu_\varepsilon \equiv$ bias
- we also like to assume $\mu_\varepsilon = 0$

- If ε_i are additive with $\varepsilon_i \sim N(0, \sigma^2)$, we call these errors "independent + identically distributed", $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

- under these assumptions, the sampling distribution is given by a Gaussian pdf

$$P(y|\theta) = \frac{1}{(2\pi\sigma^2)^{n_y/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n_y} (y_i - f(x_i; \vec{\theta}))^2\right)$$

Inference with noise models

- once we determine C + ε distribution, we can begin inferring (estimating/learning) the model parameters, $\vec{\theta}$.

- under the assumption of $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, the ordinary least squares (OLS) estimator + estimate for $\vec{\theta}$ are

(estimator) $$\tilde{\vec{\Theta}}_{OLS} = \underset{\vec{\theta} \in \Gamma}{\arg\min} \sum_{i=1}^{n_y} \left[Y_i - f(x_i; \vec{\theta}) \right]^2$$

(estimate) $$\vec{\theta}_{OLS} = \underset{\vec{\theta} \in \Gamma}{\arg\min} \sum_{i=1}^{n_y} \underbrace{\left[y_i - f(x_i; \vec{\theta}) \right]^2}$$

cost function
Loss

- Note that $\vec{\theta}_{OLS}$ can be obtained without ε_i information.
- However, without ε_i or observation error assumptions, we cannot analyze $\tilde{\vec{\Theta}}$ (i.e., sampling distribution)

Maximum Likelihood estimation (MLE)

- We can also consider the likelihood function for a random sample $\vec{Y} = [Y_1, \dots, Y_{n_y}]$ $(\vec{\theta} \in \Gamma)$

$$L(\vec{\theta} \mid \vec{y}) = f_Y(\vec{y} \mid \theta), \quad L : \Gamma \to [0, \infty)$$

where $f_Y(\vec{y} \mid \theta)$ is the parameter-dependent joint pdf for $\vec{Y}$.

- Note that if Y_i are iid, then

$$L(\vec{\theta} \mid \vec{y}) = f_Y(\vec{y} \mid \vec{\theta}) = \prod_{i=1}^{n_y} f_{Y_i}(y_i \mid \vec{\theta})$$

- Note: since $L : \Gamma \to [0, \infty)$, we often look at the log-likelihood,

$$\ln(L(\vec{\theta} \mid \vec{y})) = LL(\vec{\theta} \mid \vec{y})$$