Estimation Theory (in a nutshell)

Suppose, we "model" the data as

$$x[n] = A + w[n]$$ 'A' is unknown!

where $w[n]$ denotes some zero mean noise process. ($w[n] \sim N(0, \sigma^2)$)

Based on the data set $\{x[0], x[1], \cdots, x[N-1]\}$, we would like to estimate A. Intuitively, since A is the average level of $x[n]$, it would be "reasonable" to estimate A as

$$\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n].$$

Then, several questions come to mind:

1. How "close" will $\hat{A}$ be to A?
2. Are there better estimators than $\hat{A}$?
3. What about $\check{A} = x[0]$, which is the first sample,

· We first show the "unbiasedness" of the estimators.

· $\mathbb{E}[\hat{A}] = \mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right]$

$$= \frac{1}{N}\cdot\sum_{n=0}^{N-1}\mathbb{E}[x[n]] = \frac{1}{N}\cdot N\cdot A = A$$

· $\mathbb{E}[\check{A}] = \mathbb{E}[x[0]]$

$$= \mathbb{E}[A + w[0]] = A + \mathbb{E}[w[0]] = A$$

Hence, both $\hat{A}$ and $\check{A}$ are unbiased estimators of A.

· Then, we show the variances of the estimators.

· $\mathrm{Var}[\hat{A}] = \mathrm{Var}\left[\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right]$ $\quad * \ \mathrm{Var}(Ax) = A^2\,\mathrm{Var}(x)$

$$= \frac{1}{N^2}\sum_{n=0}^{N-1}\mathrm{Var}[x[n]] = \frac{1}{N^2}\cdot N\cdot \sigma^2 \leftarrow \quad \because w[n] \sim N(0, \sigma^2)$$

$$= \frac{1}{N}\sigma^2$$

· $\mathrm{Var}[\check{A}] = \sigma^2$

Hence, the variance of the first estimator ($\hat{A}$) decreases as N increases ($\frac{1}{N}$) while the variance of $\check{A}$ remains the same!

Note that 1. An estimator is a random variable, hence, its performance can only be evaluated statistically.

Minimum mean squared error (MMSE) estimator

Suppose, we would like to estimate the value of an "unobserved" random variable X, given that we have observed $Y=y$.
In general, we can write it with a function of y, i.e.,

$$\hat{X} = g(y).$$

Then, the error of the estimate is given by

$$\tilde{X} = X - \hat{X} = X - g(y).$$

Often, we are interested in the mean squared error (MSE) given by

$$\mathbb{E}[(X-\hat{X})^2 \mid Y=y] = \mathbb{E}[(X-g(y))^2 \mid Y=y].$$

We will show that $g(y) = \mathbb{E}[X \mid Y=y]$ has the lowest MSE among "all" possible estimators, hence, it is the MMSE estimator.

i) First, consider the case that we would like to estimate X "without" observing anything. Let α be the estimate of X. Then, the MSE is given by

depends on our selection of α.

$$\begin{aligned} h(\alpha) &= \mathbb{E}[(X-\alpha)^2] \\ &= \mathbb{E}X^2 - 2\alpha\mathbb{E}X + \alpha^2 \end{aligned}$$

$\cdot\ \frac{\partial}{\partial\alpha} h(\alpha) = 0 \Rightarrow -2\mathbb{E}X + 2\alpha = 0$, $\quad \therefore \alpha^* = \mathbb{E}X$

Therefore, we conclude the minimizing value of α is

$$\alpha^* = \mathbb{E}X. \leftarrow \text{expectation of } X \text{ is our best guess.}$$

ii) Now, suppose that we have observed $Y=y$. Then, the MSE is

$$\begin{aligned} h(\alpha; y) &= \mathbb{E}[(X-\alpha)^2 \mid Y=y] \\ &= \mathbb{E}[X^2 \mid Y=y] - 2\alpha\mathbb{E}[X \mid Y=y] + \alpha^2 \end{aligned}$$

$\cdot\ \frac{\partial}{\partial\alpha} h(\alpha) = 0 \Rightarrow \alpha^* = \mathbb{E}[X \mid Y=y]$.

Hence, the conditional expectation of X given $Y=y$, $\mathbb{E}[X \mid Y=y]$ is the MMSE estimate of X.

$\therefore$ Conditional expectation is the MMSE estimate.

- **Linear MMSE estimate**

We have shown that $g(y) = \mathbb{E}[X|Y=y]$ is the MMSE estimate of X given $Y=y$.

In practice, however, $g(y) = \mathbb{E}[X|Y=y]$ might have a complicated form.

To mitigate this, we might want $g(y)$ to be a **linear function** of y.

Suppose that we would like to have an estimator for X of the form

$$\hat{X}_L = g(y) = ay + b$$

where a and b are some real numbers to be determined.

More specifically, our goal is to choose a and b such that the MSE of the above estimator $\hat{X}_L$

$$MSE = \mathbb{E}[(X-\hat{X}_L)^2]$$

is minimized. We call the resulting estimator the linear MMSE (LMMSE) estimator.

Theorem

Let X and Y be two random variables with finite means and variances. Also, let ρ be the correlation coefficient of X and Y. Consider the following error function of (a,b):

$$h(a,b) = \mathbb{E}[(X-aY-b)^2].$$

Then,

1. The MSE $h(a,b)$ is minimized if

$$\begin{cases} a = a^* = \dfrac{Cov(X,Y)}{Var(Y)} \\ b = b^* = \mathbb{E}X - a\mathbb{E}Y \end{cases}$$

2. The minimum MSE is $h(a^*, b^*) = (1-\rho^2)\, Var(X)$
3. $\mathbb{E}[(X-a^*Y-b^*)Y] = 0$ (aka orthogonality principle)

proof)

$$\begin{aligned} h(a,b) &= \mathbb{E}[(X-aY-b)^2] \\ &= \mathbb{E}[X^2 + a^2Y^2 + b^2 - 2aXY - 2bX + 2abY] \\ &= \mathbb{E}X^2 + a^2\mathbb{E}Y^2 + b^2 - 2a\mathbb{E}XY - 2b\mathbb{E}X + 2ab\mathbb{E}Y \end{aligned}$$

$\downarrow$

i) $\frac{\partial}{\partial a} h(a,b) = 0 \Rightarrow 2a\mathbb{E}Y^2 - 2\mathbb{E}XY + 2b\mathbb{E}Y = 0$

ii) $\frac{\partial}{\partial b} h(a,b) = 0 \Rightarrow 2b - 2\mathbb{E}X + 2a\mathbb{E}Y = 0$

$$\begin{cases} \mathbb{E}Y^2 \cdot a + \mathbb{E}Y \cdot b = \mathbb{E}XY \\ \mathbb{E}Y \cdot a + 1 \cdot b = \mathbb{E}X \end{cases}$$

First, solve it for a,

$$\begin{array}{rl} & \mathbb{E}Y^2 \cdot a + \mathbb{E}Y \cdot b = \mathbb{E}XY \\ - & (\mathbb{E}Y)^2 \cdot a + \mathbb{E}Y \cdot b = \mathbb{E}X \cdot \mathbb{E}Y \\ \hline \end{array}$$

$$(\mathbb{E}Y^2 - (\mathbb{E}Y)^2)\, a = \mathbb{E}XY - \mathbb{E}X\,\mathbb{E}Y$$

$$\rightarrow a^* = \frac{\mathbb{E}XY - \mathbb{E}X\,\mathbb{E}Y}{\mathbb{E}Y^2 - (\mathbb{E}Y)^2} = \frac{Cov(X,Y)}{Var(Y)} \qquad \left(\because \begin{array}{l} Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 \\ Cov(X,Y) = \mathbb{E}XY - \mathbb{E}X\,\mathbb{E}Y \end{array} \right)$$

$$b^* = \mathbb{E}X - \mathbb{E}Y \cdot a^* \quad \text{———} \quad (*)$$

Also, by substituting (a^*, b^*) to $h(a,b)$,

$$h(a^*, b^*) = \mathbb{E}[(X - a^*Y - b^*)^2]$$

$$\downarrow \; \because \mathbb{E}[(X - a^*Y - b^*)] = \mathbb{E}X - a^*\,\mathbb{E}Y - b^* = 0$$

$$= Var[X - a^*Y - b^*]$$

$$= Var[X - a^*Y]$$

$$= Var[X] + a^{*2}\, Var[Y] - 2a^*\, Cov(X,Y)$$

$$= Var[X] + \frac{Cov(X,Y)^2}{Var[Y]^{\cancel{2}}} \cancel{Var[Y]} - 2 \cdot \frac{Cov(X,Y)}{Var[Y]} Cov(X,Y)$$

$$= Var[X] - \frac{Cov(X,Y)^2}{Var[Y]}$$

$$= Var[X] \left(1 - \frac{Cov(X,Y)^2}{Var[X]\, Var[Y]} \right)$$

$$= Var[X] \left(1 - \rho_{XY}^2 \right) \qquad \because \rho_{XY} \triangleq \frac{Cov(X,Y)}{\sqrt{Var[X]\,Var[Y]}}$$

It means that the performance of the estimator increases as

1. $Var[X]$ is small.
2. ρ_{XY} is high.

∘ Kalman Filter

(Linear) Kalman filter assumes the following linear system.

$$\begin{cases} x_k = F x_{k-1} + G u_{k-1} + w & , \quad w \sim N(0, Q) \\ y_k = H x_k + v & , \quad v \sim N(0, R) \end{cases}$$

Kalman filter is a **sequential linear MMSE estimator.** Hence, we will be using the conditional expectation of x given y.
One useful fact is, for both x and y be Gaussians, then the conditional distribution of x given y is

$$p(x|y) = N(\mu_{x|y}, \Sigma_{x|y}) \quad \text{where}$$

$$\begin{cases} \mu_{x|y} = \mu_x + C_{xy} C_{yy}^{-1} (y - \mu_y) \\ \Sigma_{x|y} = C_{xx} - C_{xy} C_{yy}^{-1} C_{yx} \end{cases}$$

* We can derive this with the matrix inversion lemma.

$\hat{x}_k^- \triangleq \mathbb{E}[x_k | y_1, \cdots, y_{k-1}]$: dynamic update (before observation)

$\hat{x}_k^+ \triangleq \mathbb{E}[x_k | y_1, \cdots, y_k]$: measurement update (after observation)

$\hat{P}_k^- \triangleq \mathbb{E}[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^T]$: dynamic update

$\hat{P}_k^+ \triangleq \mathbb{E}[(x_k - \hat{x}_k^+)(x_k - \hat{x}_k^+)^T]$: measurement update

1. Dynamic update

- $$\begin{aligned} \hat{x}_k^- &= \mathbb{E}[x_k | y_1, \cdots, y_{k-1}] \\ &= \mathbb{E}[F x_{k-1} + G u_{k-1} + w | y_1, \cdots, y_{k-1}] \\ &= F \cdot \mathbb{E}[x_{k-1} | y_1, \cdots y_{k-1}] + G u_{k-1} \\ &= F \cdot \hat{x}_{k-1}^+ + G u_{k-1} \qquad \text{(recursive equation)} \end{aligned}$$

- $$\begin{aligned} \hat{P}_k^- &= \mathbb{E}[(x_k - \hat{x}_k^-)(\sim)^T] \\ &= \mathbb{E}[(F x_{k-1} + \cancel{G u_{k-1}} + w - F \hat{x}_{k-1}^+ - \cancel{G u_{k-1}})(\sim)^T] \\ &= \mathbb{E}[(F(x_{k-1} - \hat{x}_{k-1}^+) + w)(\sim)^T] \\ &= F \cdot \mathbb{E}[(x_{k-1} - \hat{x}_{k-1}^+)(\sim)^T] F^T + Q \\ &= F \hat{P}_{k-1}^+ F^T + Q \qquad \text{(recursive equation)} \end{aligned}$$

2. Measurement update

- $\hat{x}_k^+ = \mathbb{E}[x_k | y_1, \cdots, y_k]$

$\because P(x|y) = \mathcal{N}(\mu_{x|y}, \Sigma_{x|y})$ where

$$\begin{cases} \mu_{x|y} = \mu_x + C_{xy} C_{yy}^{-1} (y - \mu_y) \\ \Sigma_{x|y} = C_{xx} - C_{xy} C_{yy}^{-1} C_{yx} \end{cases}$$

$$= \mathbb{E}x_k + C_{xy} C_{yy}^{-1} (y_k - \mathbb{E}y_k)$$

- $\hat{P}_k^+ = C_{xx} - C_{xy} C_{yy}^{-1} C_{yx}$

$\Rightarrow$ i) $\mathbb{E}x_k = \hat{x}_k^-$ ← Since, we do not know x_k, use $\hat{x}_k^-$ & $\hat{P}_k^-$. (LMMSE estimate)

ii) $C_{xx} = \hat{P}_k^-$

iii) $\mathbb{E}y_k = \mathbb{E}[Hx_k + v] = H\hat{x}_k^-$

iv) $C_{xy} = \mathbb{E}[(x_k - \hat{x}_k^-)(y_k - \mathbb{E}y_k)^T]$

$$= \mathbb{E}[(x_k - \hat{x}_k^-)(H \cdot x_k + \cancel{v} - H\hat{x}_k^-)^T]$$

$$= \mathbb{E}[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^T] H^T$$

$$= \hat{P}_k^- H^T$$

v) $C_{yy} = \mathbb{E}[(Hx_k + v - H\hat{x}_k^-)(\sim)^T]$

$$= H\hat{P}_k^- H^T + R$$

Kalman gain $K_k = \hat{P}_k^- H^T (H\hat{P}_k^- H^T + R)^{-1}$

$$\therefore \hat{x}_k^+ = \hat{x}_k^- + \hat{P}_k^- H^T (H\hat{P}_k^- H^T + R)^{-1} (y_k - H\hat{x}_k^-)$$

$$\hat{P}_k^+ = \hat{P}_k^- - \hat{P}_k^- H^T (H\hat{P}_k^- H^T + R)^{-1} H\hat{P}_k^-$$

To summarize,

- Initialize, $\hat{x}_0^+$, u_0, $\hat{P}_0^+$
- Dynamic update : step $k-1$ to k.

$$\begin{cases} \hat{x}_k^- = F\hat{x}_{k-1}^+ + G u_{k-1} \\ \hat{P}_k^- = F\hat{P}_{k-1}^+ F^T + Q \end{cases}$$

- Kalman Gain
  - $K_k = \hat{P}_k^- H^T (H\hat{P}_k^- H^T + R)^{-1}$
- Measurement update (given observation y_k)

$$\begin{cases} \hat{x}_k^+ = \hat{x}_k^- + K_k (y_k - H\hat{x}_k^-) \\ \hat{P}_k^+ = \hat{P}_k^- - K_k H\hat{P}_k^- \end{cases}$$

Extended Kalman Filter (EKF)

Now, we have nonlinear models,

$$\begin{cases} x_k = f(x_{k-1}, u_{k-1}) + w & , \; w \sim N(0, Q) \\ y_k = h(x_k) + v & , \; v \sim N(0, R) \end{cases}$$

- Initialize : $\hat{x}_0^+$, u_0, $\hat{P}_0^+$
- Dynamic update — use the nonlinear model

$$\begin{cases} \hat{x}_k^- = f(\hat{x}_{k-1}^+, u_{k-1}) \\ \hat{P}_k^- = F_{k-1} \hat{P}_{k-1}^+ F_{k-1}^T + Q \end{cases}$$

F_{k-1}: linearization of $f(\cdot,\cdot)$ at $\hat{x}_{k-1}^+$

- Kalman Gain

H_k: linearization of $h(\cdot)$ at $\hat{x}_k^-$

$$K_k = \hat{P}_k^- H_k^T (H_k \hat{P}_k^- H_k^T + R)^{-1}$$

- Measurement update (given y_k) — use the nonlinear model

$$\begin{cases} \hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - h(\hat{x}_k^-)) \\ \hat{P}_k^+ = \hat{P}_k^- - K_k H_k \hat{P}_k^- \end{cases}$$

Unscented Kalman Filter (UKF)

UKF approximates the propagated Gaussian using (2D+1) sigma points.

⇒ the update rule becomes very simple!