LMS (Least Mean Square)

Estimation.

Estimation criteria

- Probability of error $\mathbb{P}(\hat{\Theta} \neq \Theta)$ treats all wrong estimates as "equally bad".

- In some applications, "closer" estimates are better.

  eg: ① Infer the unknown bias for choosing lab problems.

  $$\theta = 0.1, \quad \hat{\theta}_1 = 0.11 \quad \text{vs} \quad \hat{\theta}_2 = 0.9.$$

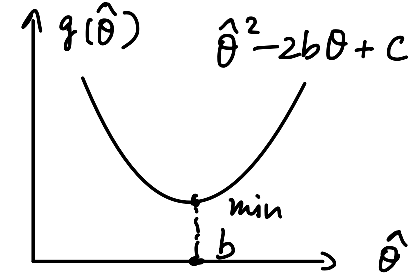
  ② Predict the stock price in December 2020.

  $$\theta = 1239 \quad \hat{\theta}_1 = 1200 \quad \text{vs.} \quad \hat{\theta}_2 = 2000$$

- A different estimation criterion

  Find $\hat{\Theta}$ that minimizes the mean square error, $\text{MSE} \triangleq \mathbb{E}[(\hat{\Theta} - \Theta)^2]$.

LMS in the absence of observations.



- Unknown Θ ; prior $p(\theta)$.
- Want to find $\hat{\theta}$ that minimizes $\underbrace{\mathbb{E}[(\Theta - \hat{\theta})^2]}_{\text{a function in } \hat{\theta}}$.

$$g(\hat{\theta}) = \mathbb{E}[\hat{\theta}^2 - 2\Theta\hat{\theta} + \Theta^2]$$

linearity of $\mathbb{E}$

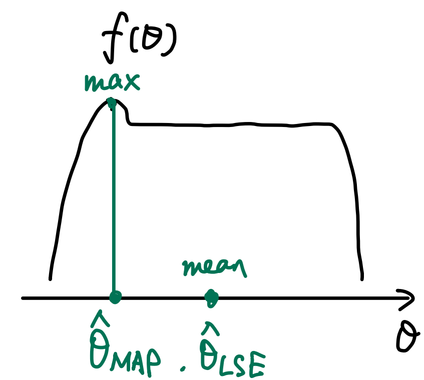
$$= \hat{\theta}^2 - 2\underbrace{\mathbb{E}[\Theta]}_{b}\hat{\theta} + \underbrace{\mathbb{E}[\Theta^2]}_{c}.$$

(numbers).

$$g'(\hat{\theta}) = 2\hat{\theta} - 2b \overset{\text{set}}{=} 0 \Rightarrow \hat{\theta} = b = \mathbb{E}[\Theta]$$

$$\boxed{\hat{\theta}_{LMS} = \mathbb{E}(\Theta)}$$

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\, f_{\Theta}(\theta)$$



LMS in the absence of observation.

Recall $Var(Y) \triangleq \mathbb{E}[(Y-\mathbb{E}Y)^2]$

$= \mathbb{E}(Y^2) - (\mathbb{E}Y)^2$

$Var(Y+c) = Var(Y)$

for any constant c.

Alternatively, $\mathbb{E}[(\Theta-\hat{\theta})^2] = Var(\Theta-\hat{\theta}) + (\mathbb{E}[\Theta-\hat{\theta}])^2$

$\hat{\theta}$ is a constant $\downarrow$ $= Var(\Theta) + (\mathbb{E}[\Theta]-\hat{\theta})^2$

$(\mathbb{E}[\Theta]-\hat{\theta})^2 \geq 0$ $\downarrow$ $\geq Var(\Theta)$

Equality is attained when $\hat{\theta}_{LSE} = \mathbb{E}(\Theta)$.

- Corresponding $\boxed{MSE = \mathbb{E}[(\Theta-\mathbb{E}[\Theta])^2] = Var(\Theta)}$

LSM estimation based on observation $X=x$

- Unknown Θ ; prior $p(\theta)$,
  - want to find an estimator $\hat{\Theta}$ that minimizes MSE
- Observation X ; model $p(x|\theta)$
  - Observe that $X=x$
- For each x, find $\hat{\theta}(x)$ that minimizes conditional MSE

$$\mathbb{E}[(\Theta-\hat{\theta}(x))^2 | X=x] = \mathrm{Var}(\Theta | X=x) + (\mathbb{E}[\Theta|X=x]-\hat{\theta}(x))^2$$
$$\geq \mathrm{Var}(\Theta|X=x)$$

Equality is attained when $\boxed{\hat{\Theta}_{LMS}(x) = \mathbb{E}[\Theta|X=x]. \text{ conditional MSE} = \mathrm{Var}(\Theta|X=x)}$

LMS estimation based on observation X (cont.)

a function of x

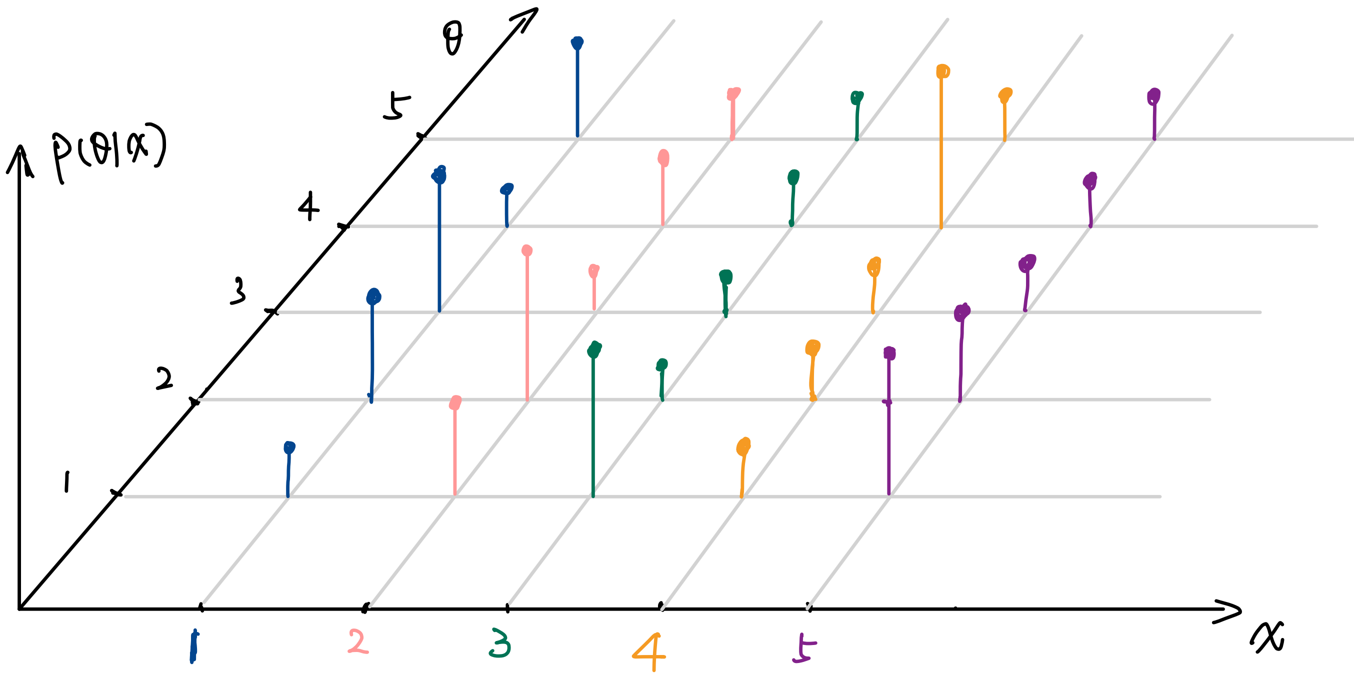
$$\xrightarrow{\ \Theta\ } \boxed{\quad p_{X|\Theta} \quad} \xrightarrow{\ X\ } \boxed{\quad \hat{\theta}_{LMS}(x) \quad} \xrightarrow{\ \hat{\Theta}\ }$$

a random variable

- LMS estimator $\boxed{\hat{\Theta} = \hat{\theta}_{LMS}(X) = \mathbb{E}[\Theta \mid X].}$

- $$\boxed{\text{MSE}} = \mathbb{E}[(\Theta - \hat{\Theta})^2] = \sum_x p(x)\, \mathbb{E}[(\Theta - \hat{\Theta})^2 \mid X = x]$$
$$= \sum_x p(x)\, \mathrm{Var}(\Theta \mid X = x)$$
$$\boxed{= \mathbb{E}[\mathrm{Var}(\Theta \mid X)].}$$

Illustration of LMS estimation.



$$\hat{\theta}_{LMS}(x) = \mathbb{E}[\Theta \mid X = x]$$

$$MSE(x) = Var[\Theta \mid X = x]$$

$$\hat{\theta}_{MAP}(x) = \arg\max_{\theta} p(\theta|x)$$

$$\mathbb{P}(\hat{\Theta} \neq \Theta \mid X = x) = 1 - p_{\Theta|X}(\hat{\theta}_{MAP}(x) \mid x)$$

Review of conditional expectation and conditional variance.

- $\mathbb{E}X = \sum_x x\, p(x) \quad \left(= \int x f(x)\, dx\right)$

- $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2]$
$$= \mathbb{E}(X^2) - (\mathbb{E}X)^2$$

- $\mathbb{E}[X^2] = \sum_x x^2 p(x) \quad \left(= \int x^2 f(x)\, dx\right)$

$\mathbb{E}[X \mid Y = y] = \sum_x x\, p(x \mid y) \quad \left(= \int x\, f(x \mid y)\, dx\right)$

$\mathrm{Var}(X \mid Y = y) = \mathbb{E}[(X - \mathbb{E}[X \mid Y = y])^2 \mid Y = y]$
$$= \mathbb{E}[X^2 \mid Y = y] - (\mathbb{E}[X \mid Y = y])^2$$

$\mathbb{E}[X^2 \mid Y = y] = \sum_x x^2 p(x \mid y) \quad \left(= \int x^2 f(x \mid y)\, dx\right)$

- Law of total expectation

$$\mathbb{E}X = \sum_y p(y)\, \mathbb{E}[X \mid Y = y] = \mathbb{E}_Y[\mathbb{E}[X \mid Y]]$$

- Law of total variance

$$\mathrm{Var}(X) = \mathbb{E}_Y[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y])$$

Example: Infer the unknown bias in choosing lab problems.

- Θ is drawn uniform at random in $[0,1]$.
- Given $\Theta=\theta$, each problem is chosen with prob. θ, indep. of each other.
- n problems in total (fixed). K problems were chosen
- Q: What is LMS estimator? What are the conditional MSE and MSE?
- posterior $f(\theta|k) = \dfrac{\binom{n}{k}\theta^k(1-\theta)^{n-k}}{\int_0^1 \binom{n}{k}\tilde{\theta}^k(1-\tilde{\theta})\,d\tilde{\theta}}$
- $\hat{\theta}_{LMS}(k) = \mathbb{E}[\Theta|K=k] = \int_0^1 \theta f(\theta|k)\,d\theta = \dfrac{\int_0^1 \binom{n}{k}\theta^{k+1}(1-\theta)^{n-k}\,d\theta}{\int_0^1 \binom{n}{k}\tilde{\theta}^k(1-\tilde{\theta})^{n-k}\,d\tilde{\theta}}$

$$= \frac{(k+1)!\,(n-k)!}{(n+2)!}\cdot\frac{(n+1)!}{k!\,(n-k)!} = \frac{k+1}{n+2}$$

$\hat{\theta}_{MAP}(k) = \dfrac{k}{n}$

Formula: for integers $\alpha\geq 0, \beta\geq 0$.

$$\int_0^1 \theta^\alpha(1-\theta)^\beta\,d\theta = \frac{\alpha!\,\beta!}{(\alpha+\beta+1)!}$$

Example: Infer the unknown bias in choosing lab problems.

- Θ is drawn uniform at random in $[0,1]$.
- Given $\Theta = \theta$, each problem is chosen with prob. θ, indep. of each other.
- n problems in total (fixed). K problems were chosen
- Q1: What is LMS estimator?
    - $\hat{\theta}_{LMS}(k) = \frac{k+1}{n+2}$

Q2: What is the conditional MSE given $K=k$?

What is the (overall) MSE?

- Conditional MSE

$$\mathbb{E}[(\Theta - \hat{\Theta})^2 \mid K = k] = \mathrm{Var}[\Theta \mid K = k] = \mathbb{E}[\Theta^2 \mid K = k] - (\mathbb{E}[\Theta \mid K = k])^2$$

$$= \int_0^1 \theta^2 f(\theta \mid k)\, d\theta - \left(\frac{k+1}{n+2}\right)^2$$

$$= \frac{\int_0^1 \theta^{k+2}(1-\theta)^{n-k}\, d\theta}{\int_0^1 \tilde{\theta}^k (1-\tilde{\theta})^{n-k}\, d\tilde{\theta}} - \left(\frac{k+1}{n+2}\right)^2$$

$$= \frac{(k+2)!\,(n-k)!}{(n+3)!} \frac{(n+1)!}{k!\,(n-k)!} - \left(\frac{k+1}{n+2}\right)^2$$

$$= \frac{(k+2)(k+1)}{(n+3)(n+2)} - \left(\frac{k+1}{n+2}\right)^2.$$

$$= \frac{kn - k^2 + n + 1}{(n+3)(n+2)^2}$$

$$\boxed{\text{Formula: for integers } \alpha \geq 0, \beta \geq 0. \quad \int_0^1 \theta^\alpha (1-\theta)^\beta\, d\theta = \frac{\alpha!\,\beta!}{(\alpha+\beta+1)!}}$$

- $\text{MSE} = \mathbb{E}[\text{Var}(\Theta|K)] = \mathbb{E}\left[\frac{Kn - K^2 + n + 1}{(n+3)(n+2)^2}\right]$

linearity of $\mathbb{E}$

$$= \frac{n\,\mathbb{E}[K]}{(n+3)(n+2)^2} - \frac{\mathbb{E}[K^2]}{(n+3)(n+2)^2} + \frac{n+1}{(n+3)(n+2)^2}.$$

$\Rightarrow$ Want $\mathbb{E}[K]$ and $\text{Var}[K]$. $\left(\mathbb{E}[K^2] = \text{Var}[K] + (\mathbb{E}[K])^2\right)$.

Recall $K \mid \Theta = \theta \sim \text{Binom}(n, \theta)$. $\mathbb{E}[K \mid \Theta = \theta] = n\theta$, $\text{Var}(K \mid \Theta = \theta) = n\theta(1-\theta)$.

law of total expectation

$$\mathbb{E}[K] = \mathbb{E}_\Theta[\mathbb{E}[K|\Theta]] = \int_0^1 f(\theta)\,\mathbb{E}[K|\Theta = \theta]\,d\theta = \int_0^1 1 \cdot n\theta\,d\theta = \frac{n}{2}$$

law of total variance

$$\text{Var}[K] = \mathbb{E}_\Theta[\text{Var}(K|\Theta)] + \text{Var}(\mathbb{E}[K|\Theta])$$

$\text{Var}(cY) = c^2\,\text{Var}(Y)$

$$= \int_0^1 n\theta \cdot (1-\theta)\,d\theta + \text{Var}(n\Theta) = \frac{n}{6} + n^2\,\text{Var}(\Theta) = \frac{n}{6} + \frac{n^2}{12}.$$

$$\Rightarrow \mathbb{E}[K^2] = \text{Var}[K] + (\mathbb{E}[K])^2 = \frac{(2n+1)n}{6}.$$

$$\Rightarrow \text{MSE} = \frac{1}{6(n+2)}$$

Exercise: Derive $p(k)$ and compute $\mathbb{E}[K]$, $\mathbb{E}[K^2]$ from $p(k)$.

Example: Romeo and Juliet.

Romeo and Juliet start dating, but Juliet will be late on any date by a random amount X uniformly distributed over the interval $[0, \theta]$. The parameter θ is unknown and is the realization of a r.v. $\Theta \sim Unif[0,1]$. Assuming that Juliet was late by an amount x on their first date, how should Romeo use this information to estimate Θ using MAP & LMS rule? Compute conditional MSE for both MAP & LMS estimators

- Unknown: $\Theta \sim Unif[0,1]$
- Observation model: $f(x|\theta) = \frac{1}{\theta}, \quad 0 \le x \le \theta \le 1$
- posterior $f(\theta|x) = \dfrac{f(\theta) f(x|\theta)}{f(x)} = \dfrac{\frac{1}{\theta}}{\int_x^1 \frac{1}{\theta} d\theta} = \dfrac{\frac{1}{\theta}}{\ln\theta|_x^1} = \dfrac{\frac{1}{\theta}}{-\ln x}. \quad 0 \le x \le \theta \le 1.$
- $\hat{\Theta}_{MAP}(x) = \arg\max_\theta f(\theta|x) = x$
- $\hat{\Theta}_{LMS}(x) = \mathbb{E}[\Theta | X = x] = \int \theta f(\theta|x) d\theta = \int_x^1 \dfrac{\theta \cdot \frac{1}{\theta}}{-\ln x} d\theta = \dfrac{1-x}{-\ln x}$

- To compute the MSE, we first write a general expression for any $\hat{\theta}(x)$

cond. $MSE(x) = \mathbb{E}[(\Theta - \hat{\Theta})^2 | X = x] = \mathbb{E}[(\Theta - \hat{\theta}(x))^2 | X = x]$

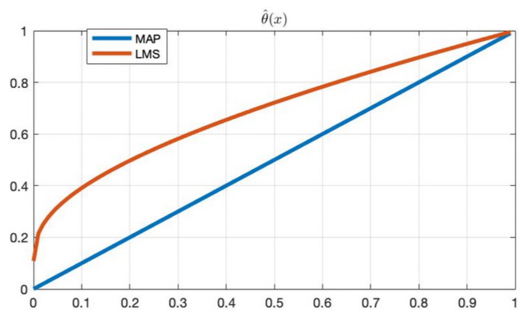
$$= \mathbb{E}[\Theta^2 | X = x] - 2\hat{\theta}(x)\,\mathbb{E}[\Theta | X = x] + (\hat{\theta}(x))^2 .$$

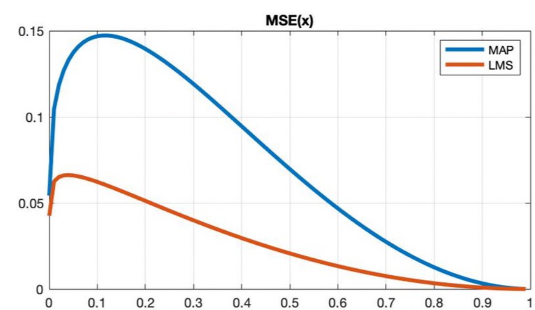
$$= \int_x^1 \frac{\theta^2 \frac{1}{\theta}}{-\ln x} d\theta - 2\hat{\theta}(x) \cdot \frac{1-x}{-\ln x} + (\hat{\theta}(x))^2$$

$$= \frac{1-x^2}{-2\ln x} - 2\hat{\theta}(x)\frac{(1-x)}{-\ln x} + (\hat{\theta}(x))^2 .$$

Plug in $\hat{\theta}_{MAP}(x) = x$. $MSE_{MAP}(x) = x^2 + \frac{3x^2 - 4x + 1}{-2\ln x}$.

Plug in $\hat{\theta}_{LMS}(x) = \frac{1-x}{-\ln x}$, $MSE_{LMS}(x) = \frac{1-x^2}{-2\ln x} - \left(\frac{1-x}{-\ln x}\right)^2$





Example: Multiple independent observations of a signal

- $X_1 = \Theta + W_1$ $\quad\quad \Theta \sim N(x_0, \sigma_0^2) \quad W_i \sim N(0, \sigma_i^2) \quad i=1,2,\cdots,n$

  $X_2 = \Theta + W_2$ $\quad\quad \Theta, W_1, W_2, \cdots W_n$ independent

  $\vdots$

  $X_n = \Theta + W_n$ $\quad$ Find MAP & LMS estimators of Θ given $(x_1,\cdots,x_n)$ and corresponding MSE.

- Unknown $\Theta \sim N(x_0, \sigma_0^2)$. Data: a vector $\underline{x} = (x_1, \cdots, x_n)$

- Model: $f(\underline{x}|\theta) = f(x_1,\cdots,x_n|\theta) = f_{W_1,\cdots,W_n}(x_1-\theta, \cdots, x_n-\theta) \overset{\text{indep. of } W_i}{=} \prod_{i=1}^{n} f_{W_i}(x_i-\theta)$

- posterior $f(\theta|\underline{x}) = \dfrac{1}{f(\underline{x})} c_0 e^{-\frac{(x_0-\theta)^2}{2\sigma_0^2}} \cdot \prod_{i=1}^{n} c_i e^{-\frac{(x_i-\theta)^2}{2\sigma_i^2}} \triangleq c \cdot e^{-g(\theta)}$

  where $g(\theta) = \dfrac{(x_0-\theta)^2}{2\sigma_0^2} + \dfrac{(x_1-\theta)^2}{2\sigma_1^2} + \cdots + \dfrac{(x_n-\theta)^2}{2\sigma_n^2}$.

$$\frac{dg(\theta)}{d\theta} \triangleq 0 \Rightarrow \sum_{i=0}^{n} \frac{(\theta - x_i)}{\sigma_i^2} = 0 \Rightarrow \sum_{i=0}^{n} \frac{\theta}{\sigma_i^2} = \sum_{i=0}^{n} \frac{x_i}{\sigma_i^2} \Rightarrow \hat{\theta}_{MAP}(x_1,\cdots,x_n) = \frac{\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

Recognizing Gaussian pdf.

- $X \sim N(\mu, \sigma^2)$ $\quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\mathbb{E}X = \mu$, $\operatorname{Var}X = \sigma^2$.

- $f(x) = c \cdot e^{-8(x-3)^2}$ $\Rightarrow$ $\mu = 3$; $\frac{1}{2\sigma^2} = 8$ $\Rightarrow$ $\sigma^2 = \frac{1}{16}$. ; $c = \frac{1}{\frac{1}{4}\sqrt{2\pi}}$; $X \sim N(3, \frac{1}{16})$

- $f(x) = c \cdot e^{-(\alpha x^2 + \beta x + \gamma)}$, $\alpha > 0$.

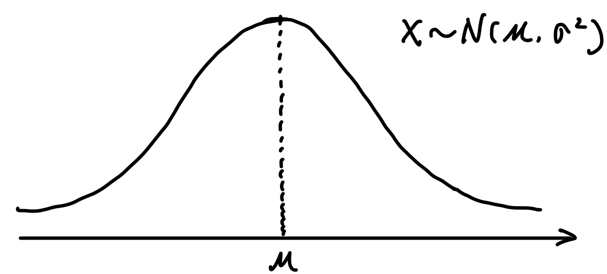
$$\alpha\left(x^2 + \frac{\beta}{\alpha}x + \frac{\beta^2}{4\alpha^2}\right) - \frac{\beta^2}{4\alpha} + \gamma = \alpha\left(x + \frac{\beta}{2\alpha}\right)^2 - \frac{\beta^2}{4\alpha} + \gamma$$

$\boxed{\mu = -\frac{\beta}{2\alpha}}$; $\frac{1}{2\sigma^2} = \alpha \Rightarrow \boxed{\sigma^2 = \frac{1}{2\alpha}}$; $c \cdot e^{-(\gamma - \frac{\beta^2}{4\alpha})} = \frac{1}{\frac{1}{\sqrt{2\alpha}} \cdot \sqrt{2\pi}} \Rightarrow c = \sqrt{\frac{\alpha}{\pi}}\, e^{\gamma - \frac{\beta^2}{4\alpha}}$

$X \sim N(-\frac{\beta}{2\alpha}, \frac{1}{2\alpha})$

- Property:

$\boxed{\text{For Gaussian pdf } f(x):\ \underset{x}{\operatorname{argmax}}\, f(x) = \mathbb{E}X}$



Example: Multiple independent observations of a signal

- $X_1 = \Theta + W_1$ $\quad \Theta \sim N(x_0, \sigma_0^2) \quad W_i \sim N(0, \sigma_i^2) \quad i = 1, 2, \cdots, n$

  $X_2 = \Theta + W_2$ $\quad \Theta, W_1, W_2, \cdots W_n$ independent

  $\vdots$

  $X_n = \Theta + W_n$ $\quad$ Find MAP & LMS estimators of Θ given $(x_1, \cdots, x_n)$ and corresponding MSE.

- Unknown $\Theta \sim N(x_0, \sigma_0^2)$. Data: a vector $\underline{x} = (x_1, \cdots, x_n)$

- Model: $f(\underline{x}|\theta) = f(x_1, \cdots, x_n|\theta) = f_{W_1, \cdots, W_n}(x_1 - \theta, \cdots, x_n - \theta) \overset{\text{indep. of } W_i}{=} \prod_{i=1}^{n} f_{W_i}(x_i - \theta)$

- $f(\theta|\underline{x}) = c\, e^{-g(\theta)} = c \cdot e^{-\left[\frac{(\theta - x_0)^2}{2\sigma_0^2} + \frac{(\theta - x_1)^2}{2\sigma_1^2} + \cdots + \frac{(\theta - x_n)^2}{2\sigma_n^2}\right]}$

- $g(\theta) = \underbrace{\left(\sum_{i=0}^{n} \frac{1}{2\sigma_i^2}\right)}_{\alpha} \theta^2 - \underbrace{\left(\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}\right)}_{\beta} \theta + \sum_{i=0}^{n} \frac{x_i^2}{2\sigma_i^2}$, quadratic in θ, $\alpha > 0 \Rightarrow$ Gaussian pdf. $N(\mu, \sigma^2)$

- $\mathbb{E}[\Theta|X=x] = \mu = -\frac{\beta}{2\alpha} = \frac{\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$, $\quad \mathrm{Var}[\Theta|X=x] = \sigma^2 = \frac{1}{2\alpha} = \frac{1}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$

Gaussian pdf

$$\Rightarrow \hat{\theta}_{MAP}(x) = \underset{\theta}{\arg\max}\, \underbrace{f(\theta|x)}_{\text{Gaussian pdf}} = \mathbb{E}[\Theta | X = x] = \hat{\theta}_{LMS}(x) = \mu = \frac{\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

$$\Rightarrow MSE_{MAP}(x) = MSE_{LMS}(x) = Var(\Theta | X = x) = \sigma^2 = \frac{1}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}.$$

$$\Rightarrow MSE_{MAP} = MSE_{LMS} = \frac{1}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

<u>Remarks</u>:
- For Gaussian $f(\theta|x)$, $\underset{\theta}{\arg\max}\, f(\theta|x) = \mathbb{E}[\Theta | X = x]$, thus,

$$\hat{\Theta}_{MAP} = \hat{\Theta}_{LMS}.$$

- When $f(\theta|x)$ is Gaussian, $\hat{\Theta}_{LMS}$ is <u>linear</u> in X, i.e.

$$\hat{\Theta}_{LMS} = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + b.$$