Maximum Likelihood Estimation

MVUE가 존재하지 않거나 찾기 힘든 경우, 그 대안에 해당함
데이터 수가 커지면 근사적으로 MVUE에 해당하므로 믿을만 함

1. Example - DC Level in White Gaussian Noise
   $x[n] = A + w[n],\ w[n] \sim N(0, A)$

   1) CRLB
   우선 CRLB를 살펴보면,

$$\frac{\partial \ln p(x;A)}{\partial A} = -\frac{N}{2A} + \frac{1}{A}\sum_{n=0}^{N-1}(x[n]-A) + \frac{1}{2A^2}\sum_{n=0}^{N-1}(x[n]-A)^2$$
$$= I(A)(\hat{A}-A)?$$

   위 식을 만족시키는 $I(A)$가 없고, efficient한 추정값을 찾을 수 없다.

   2) RBLS - simple
   우선,

$$p(x;A) = \frac{1}{(2\pi A)^{N/2}}\exp[-\frac{1}{2A}\sum_{n=0}^{N-1}(x[n]-A)^2]$$
$$= \frac{1}{(2\pi A)^{N/2}}\exp[-\frac{1}{2A}\sum_{n=0}^{N-1}(x^2[n]-2N\bar{x}+NA)]$$
$$= \frac{1}{(2\pi A)^{N/2}}\exp[-\frac{1}{2}(\frac{1}{A}\sum_{n=0}^{N-1}x^2[n]+NA)]\exp(N\bar{x})$$
$$= g(\sum_{n=0}^{N-1}x^2[n];A)h(x)$$

   위와 같이 전개게 되므로 Neymann-Fisher factorization에 따라 충분통계량은 $T(x) = \sum_{n=0}^{N-1}x^2[n]$이 된다. 그 다음으로 알맞은 $g(T(x))$를 찾아야 하는데 이 때 g는 다음을 만족해야 할 것이다.

$$E[g(T(x))] = A \text{ for all } A > 0$$

   이를 위해 $E(\sum_{n=0}^{N-1}x^2[n])$을 전개해보면,

$$E(\sum_{n=0}^{N-1}x^2[n]) = \neq (x^2[n])$$
$$= N[var(x[n]) + E^2(x[n])]$$
$$= N(A + A^2)$$

   다음과 같이 되는데, 해당 값을 A로 만들 함수 g를 찾을 수 없다.

   3) RBLS - complicate
   불편추정값 $\hat{A} = x[0]$에 대해 $E(\hat{A}|\sum_{n=0}^{N-1}x^2[n]) = E(x[0]|\sum_{n=0}^{N-1}x^2[n])$을 찾을 수 있지만 이를 계산하기 쉽지 않고, 2)에서 보였듯 충분통계량이 완비성을 갖추지 못한다.

4) MLE

log 가능도 함수를 미분해 0인 값을 찾아 최댓값을 찾는다.

$$\frac{\partial \ln p(x;A)}{\partial A} = -\frac{N}{2A} + \frac{1}{A}\sum_{n=0}^{N-1}(x[n]-A) + \frac{1}{2A^2}\sum_{n=0}^{N-1}(x[n]-A)^2 = 0$$

$$-\frac{N}{2A} - N + \frac{1}{2A^2}\sum_{n=0}^{N-1}x^2[n] + \frac{N}{2} = 0$$

$$\hat{A}^2 + \hat{A} - \frac{1}{N}\sum_{n=0}^{N-1}x^2[n] = 0$$

$$\hat{A} = -\frac{1}{2} \pm \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x^2[n] + \frac{1}{4}}$$

여기서 DC Level이기에 $A > 0$이므로

$$\hat{A} = -\frac{1}{2} + \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x^2[n] + \frac{1}{4}}$$

MLE 추정값의 기대값을 계산해보면

$$\begin{aligned} E(\hat{A}) &= E(-\frac{1}{2} + \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x^2[n] + \frac{1}{4}}) \\ &\neq -\frac{1}{2} + \sqrt{E(\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]) + \frac{1}{4}} \quad \text{for } all\ A \\ &= -\frac{1}{2} + \sqrt{A + A^2 + \frac{1}{4}} \\ &= A \end{aligned}$$

편향되어 있음을 알 수 있다.

다만 N이 무한히 커지면 MLE는 점근적으로 편향성이 없어진다.

2. Theorem for MLE

Theorem 7.1 (Asymptotic Properties of the MLE) *If the PDF $p(\mathbf{x};\theta)$ of the data $\mathbf{x}$ satisfies some "regularity" conditions, then the MLE of the unknown parameter θ is asymptotically distributed (for large data records) according to*

$$\hat{\theta} \overset{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta)) \tag{7.11}$$

where $I(\theta)$ is the Fisher information evaluated at the true value of the unknown parameter.

MLE is asymptotically efficient.

Regularity conditions:
- log-likelihood function is differentiable
- Fisher information is nonzero.

Theorem 7.2 (Invariance Property of the MLE) *The MLE of the parameter $\alpha = g(\theta)$, where the PDF $p(\mathbf{x};\theta)$ is parameterized by θ, is given by*

$$\hat{\alpha} = g(\hat{\theta})$$

where $\hat{\theta}$ is the MLE of θ. The MLE of $\hat{\theta}$ is obtained by maximizing $p(\mathbf{x};\theta)$. If g is not a one-to-one function, then $\hat{\alpha}$ maximizes the modified likelihood function $\bar{p}_T(\mathbf{x};\alpha)$, defined as

$$\bar{p}_T(\mathbf{x};\alpha) = \max_{\{\theta:\alpha=g(\theta)\}} p(\mathbf{x};\theta).$$

3. For Linear Model

$$x = H\theta + w, \text{ where } w \sim N(0, C)$$

위 linear model에 대해,

$$p(x;\theta) = \frac{1}{(2\pi)^{N/2}\det^{1/2}(C)}\exp[-\frac{1}{2}(x-H\theta)^T C^{-1}(x-H\theta)]$$

이므로

$$\frac{\partial \ln p(x;\theta)}{\partial \theta_k} = -\frac{1}{2}tr(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_k}) + \frac{\partial \mu(\theta)^T}{\partial \theta_k}C^{-1}(\theta)(x-\mu(\theta)) - \frac{1}{2}(x-\mu(\theta))^T\frac{\partial C^{-1}(\theta)}{\partial \theta_k}(x-\mu(\theta))$$

이고, 위 경우에 대해선 $\mu(\theta) = H\theta$이므로 미분값을 0으로 두면,

$$\frac{\partial \ln p(x;\theta)}{\partial \theta} = \frac{\partial (H\theta)^T}{\partial \theta}C^{-1}(x-H\theta) = H^T C^{-1}(x-H\hat{\theta}) = 0$$

가 되어 MLE는

$\hat{\theta} = (H^T C^{-1} H)^{-1} H^T C^{-1} x$

가 된다. 따라서 앞서 계산한 MVUE와 동일하므로 linear model에서 MLE는 MVUE가 된다.