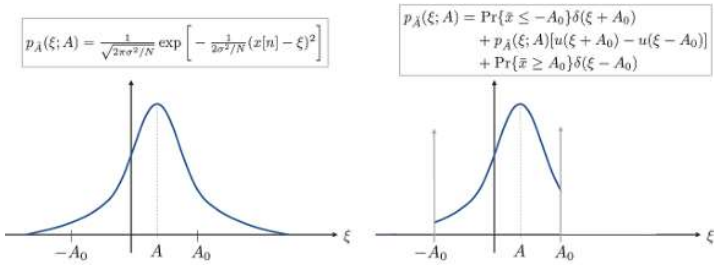
Bayesian Approach

$x[n] = \theta + w[n]$
위 데이터에서 고전적인 추론 방법에서는 θ가 결정되어 있었지만 bayesian approach에서는 확률 변수로 두고 해석한다. Bayesian 추정을 통해 θ에 대한 사전 정보가 주어진다면 더 나은 추정이 가능하며 MVUE를 찾는 것이 불가능한 경우 차선의 추정값을 얻을 수 있다.

1. Example about Prior Knowledge



위와 같이 데이터가 제한된 영역인 $-A_0 \le A \le A_0$에 위치한다는 사전 정보가 있다고 가정하자. 이 경우에 원래의 추정값인 $\hat{A} = \bar{x}$가 부적절할 수 있다.

$$\check{A} = \begin{cases} -A_0 & (\bar{x} \leftarrow A_0) \\ \bar{x} & (-A_0 < \bar{x} < A_0) \\ A_0 & (\bar{x} > A_0) \end{cases}$$

다음 추정값에 대해 $\hat{A} = \bar{x}$와 mse를 비교해보자.

$$\begin{aligned} mse(\hat{A}) &= \int_{-\infty}^{\infty} (\xi - A)^2 p_{\hat{A}}(\xi;A)d\xi \\ &= \int_{-\infty}^{-A_0} (\xi - A)^2 p_{\hat{A}}(\xi;A)d\xi + \int_{-A_0}^{A_0} (\xi - A)^2 p_{\hat{A}}(\xi;A)d\xi + \int_{A_0}^{\infty} (\xi - A)^2 p_{\hat{A}}(\xi;A)d\xi \\ &> \int_{-\infty}^{-A_0} (-A_0 - A)^2 p_{\hat{A}}(\xi;A)d\xi + \int_{-A_0}^{A_0} (\xi - A)^2 p_{\hat{A}}(\xi;A)d\xi + \int_{A_0}^{\infty} (A_0 - A)^2 p_{\hat{A}}(\xi;A)d\xi \\ &= mse(\check{A}) \end{aligned}$$

따라서 $\check{A}$가 $\hat{A}$보다 mse측면에서 더 좋다.

2. Minimum MSE(MMSE)

A가 가지는 특정 분포에 대한 설정이 없으므로 A는 균일 분포를 따른다고 가정한다.

$$A \sim u[-A_0, A_0]$$

그리고 베이지안 추정 관점에서 Bayesian MSE를 최소로 가지는 추정값을 찾도록 한다.

$$Bmse(\hat{A}) = E[(A - \hat{A})^2]$$

Bmse는 mse에서와 달리 A도 확률 변수로 보기에 그 차이를 인식해야 한다.

$$mse(\hat{A}) = \int (\hat{A} - A)^2 p(x;A)dx$$

$$\begin{aligned} Bmse(\hat{A}) &= \int\int (A - \hat{A})^2 p(x, A)dx dA \\ &= \int [\int (A - \hat{A})^2 p(A|x)dA] p(x)dx \end{aligned}$$

확률의 정의에 따라 모든 x에 대해 $p(x) > 0$이 성립한다. 따라서 괄호 안의 적분식만 최소가 되면 된다. 그러므로 괄호 안의 적분식을 미분하면 다음과 같다.

$$\begin{aligned}\frac{\partial}{\partial A}\int (A-\hat{A})^2 p(A|x)dA &= \int \frac{\partial}{\partial A}(A-\hat{A})^2 p(A|x)dA \\ &= \int -2(A-\hat{A})p(A|x)dA \\ &= -2\int Ap(A|x)dA + 2\hat{A}\int \hat{A}p(A|x)dA \\ &= -2\int Ap(A|x)dA + 2\hat{A} \\ &= 0\end{aligned}$$

따라서 다음과 같은 추정값을 얻을 수 있다.

$$\hat{A} = \int Ap(A|x)dA = E(A|x)$$

위 값은 posterior pdf $p(A|x)$의 기댓값과 동일하다.

MMSE를 구하기 위해 우선 posterior pdf를 구해보자. $p(x|A)$는 $x[n] = A + w[n]$에서 $p(x;A)$를 구하는 것과 동일하다.

$$p(x|A) = \frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]$$

그런데 posterior pdf는 bayesian rule에 따라 다음과 같다.

$$p(A|x) = \frac{p(x|A)p(A)}{p(x)} = \frac{p(x|A)p(A)}{\int p(x|A)p(A)dA}$$

따라서 구하고자 하는 posterior pdf는 다음과 같다.

$$p(A|x) = \begin{cases}\dfrac{1}{c\sqrt{2\pi\dfrac{\sigma^2}{N}}}\exp\left[-\dfrac{1}{2\dfrac{\sigma^2}{N}}(A-\overline{x})^2\right] & |A| \le A_0 \\ 0 & |A| > A_0\end{cases}$$

여기서 c는 정규화를 위한 값으로 다음과 같다.

$$c = \int_{-A_0}^{A_0}\frac{1}{2\pi\dfrac{\sigma^2}{N}}\exp\left[-\frac{1}{2\dfrac{\sigma^2}{N}}(A-\overline{x})^2\right]dA$$

따라서 MMSE는 다음과 같이 구할 수 있다.

$$\hat{A} = E(A|x) = \int_{-\infty}^{\infty}Ap(A|x)dA$$

3. Choosing Prior PDF

이번에는 prior pdf를 가우시안 분포로 가정하자.

$$p(A) = \frac{1}{\sqrt{2\pi\sigma_A^2}}\exp\left[-\frac{1}{2\sigma_A^2}(x-\mu_A)^2\right]$$

다음으로 likelihood는 다음과 같다.

$$p(x|A)=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2]$$

$$=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]]\exp[-\frac{1}{2\sigma^2}(NA^2-2NA\bar{x})]$$

따라서 posterior pdf는 prior pdf와 likelihood를 대입하여 다음과 같이 계산된다.

$$p(A|x)=\frac{p(x|A)p(A)}{\int p(x|A)p(A)dA}$$

$$=\frac{\frac{1}{(2\pi\sigma^2)^{N/2}\sqrt{2\pi\sigma_A^2}}\exp[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]]\exp[-\frac{1}{2\sigma^2}(NA^2-2NA\bar{x})]\exp[-\frac{1}{2\sigma_A^2}(A-\mu_A)^2]}{\int_{-\infty}^{\infty}\frac{1}{(2\pi\sigma^2)^{N/2}\sqrt{2\pi\sigma_A^2}}\exp[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]]\exp[-\frac{1}{2\sigma^2}(NA^2-2NA\bar{x})]\exp[-\frac{1}{2\sigma_A^2}(A-\mu_A)^2]dA}$$

$$=\frac{\exp[-\frac{1}{2\sigma^2}(NA^2-2NA\bar{x})-\frac{1}{2\sigma_A^2}(A-\mu_A)^2]}{\int_{-\infty}^{\infty}\exp[-\frac{1}{2\sigma^2}(NA^2-2NA\bar{x})-\frac{1}{2\sigma_A^2}(A-\mu_A)^2]dA}$$

$$=\frac{\exp[-\frac{1}{2}Q(A)]}{\int_{-\infty}^{\infty}\exp[-\frac{1}{2}Q(A)]dA}$$

위 식에서 $Q(A)$는 다음과 같이 전개된다.

$$Q(A)=\frac{N}{\sigma^2}A^2-\frac{2NA\bar{x}}{\sigma^2}+\frac{A^2}{\sigma_A^2}-\frac{2\mu_A A}{\sigma_A^2}+\frac{\mu_A^2}{\sigma_A^2}$$

$$=\left(\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}\right)A^2-2(\frac{N\bar{x}}{\sigma^2}+\frac{\mu_A}{\sigma_A^2})A+\frac{\mu_A^2}{\sigma_A^2}$$

여기서 posterior의 평균과 분산을 다음과 같이 정의하자.

$$\sigma_{A|x}^2=\frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}$$

$$\mu_{A|x}=(\frac{N}{\sigma^2}\bar{x}+\frac{\mu_A}{\sigma_A^2})\sigma_{A|x}^2$$

위 식을 $Q(A)$에 대입하면 다음과 같다.

$$Q(A)=\frac{1}{\sigma_{A|x}^2}(A^2-2\mu_{A|x}A+\mu_{A|x}^2)-\frac{\mu_{A|x}^2}{\sigma_{A|x}^2}+\frac{\mu_A^2}{\sigma_A^2}$$

$$=\frac{1}{\sigma_{A|x}^2}(A-\mu_{A|x})^2-\frac{\mu_{A|x}^2}{\sigma_{A|x}^2}+\frac{\mu_A^2}{\sigma_A^2}$$

따라서 $p(A|x)$는 다음과 같이 쓸 수 있다.

$$p(A|x) = \frac{\exp[-\frac{1}{2\sigma_{A|x}^2}(A-\mu_{A|x})^2]\exp[-\frac{1}{2}(\frac{\mu_A^2}{\sigma_A^2}-\frac{\mu_{A|x}^2}{\sigma_{A|x}^2})]}{\int_{-\infty}^{\infty}\exp[-\frac{1}{2\sigma_{A|x}^2}(A-\mu_{A|x})^2]\exp[-\frac{1}{2}(\frac{\mu_A^2}{\sigma_A^2}-\frac{\mu_{A|x}^2}{\sigma_{A|x}^2})]dA}$$
$$= \frac{1}{\sqrt{2\pi\sigma_{A|x}^2}}\exp[-\frac{1}{2\sigma_{A|x}^2}(A-\mu_{A|x})^2]$$

posterior 역시 가우시안 분포를 가진다는 것을 알 수 있고, MMSE는 다음과 같다.

$$\hat{A} = E(A|x)$$
$$= \mu_{A|x}$$
$$= \frac{\frac{N}{\sigma^2}\bar{x}+\frac{\mu_A}{\sigma_A^2}}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}$$
$$= \frac{\sigma_A^2}{\sigma_A^2+\frac{\sigma^2}{N}}\bar{x}+\frac{\frac{\sigma^2}{N}}{\sigma_A^2+\frac{\sigma^2}{N}}\mu_A$$
$$= \alpha\bar{x}+(1-\alpha)\mu_A$$

데이터가 거의 없는 경우에는 α가 0에 인접하여 $\hat{A}\approx\mu_A$가 되고, 데이터가 많은 경우에는 α가 1에 근접하여 $\hat{A}\approx\bar{x}$가 된다. 또한 분산을 살펴보면,

$$var(A|x) = \sigma_{A|x}^2 = \frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}$$

이므로 데이터 수가 클수록 분산이 작아져 pdf가 뾰족한 형태를 가짐을 알 수 있다. 만약 prior pdf를 설정하지 않았다면 σ_A^2가 무한대가 됨과 동일하기 때문에 $\hat{A}\approx\bar{x}$가 될 것이다.

BMSE는 정의에 따라 다음과 같이 전개할 수 있다.

$$\begin{aligned}
Bmse(\hat{A}) &= \int\int (A-\hat{A})^2 p(x, A)\,dx\,dA \\
&= \int [\int (A-\hat{A})^2 p(A|x)\,dA] p(x)\,dx \\
&= \int [\int (A-E(A|x))^2 p(A|x)\,dA] p(x)\,dx \\
&= \int var(A|x) p(x)\,dx \\
&= \int \sigma_{A|x}^2 p(x)\,dx \\
&= \sigma_{A|x}^2 \int p(x)\,dx \\
&= \frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}} \\
&= \frac{\sigma^2}{N}\frac{\sigma_A^2}{\sigma_A^2+\frac{\sigma^2}{N}} < \frac{\sigma^2}{N}
\end{aligned}$$

따라서 prior pdf가 있을 때의 BMSE 값은 사전 지식이 없는 MMSE 값보다 항상 작음을 알 수 있다. 따라서 prior pdf는 추정값의 성능을 향상시킨다.