Sufficient Statistics

1. What is Sufficient Statistics?

$x[n] = A + w[n], \quad w[n] \sim N(0, \sigma^2)$

해당 데이터들에 대해서 MVUE는 $\frac{1}{N}\sum_{n=0}^{N-1} x[n]$와 같음을 확인했다.

그리고 해당 estimator은 최소 분산 $\frac{\sigma^2}{N}$을 가진다.

만약 추정값으로 다음의 값을 사용한다고 가정해보자.

$\breve{A} = x[0]$

해당 데이터는 unbiased 되어있지만 분산은 σ^2로 최소가 아니므로 좋은 추정값이라 할 수 없다.

그렇다면 어떤 데이터가 A의 정보를 충분히 가지고 있을까?

$S_1 = (x[0], x[1], \ldots, x[N-1])$

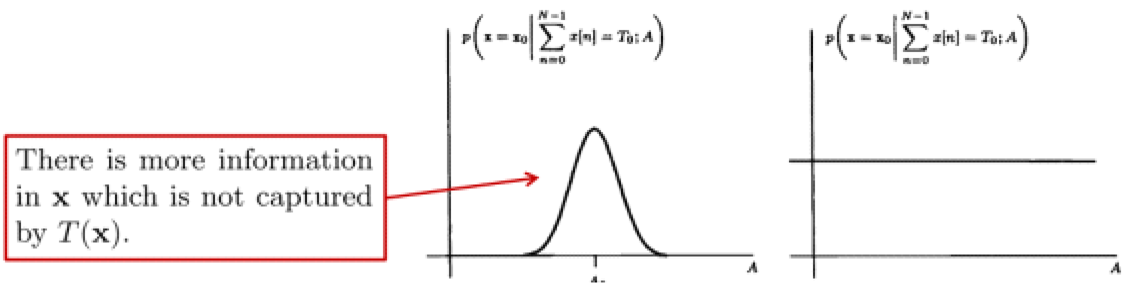
$S_2 = (x[0] + x[1], x[2], x[3], \ldots, x[N-1])$

$S_3 = (\sum_{n=0}^{N-1} x[n])$

위 데이터들은 모두 충분하다. 여기서 중요한 것은 최소로 충분한 데이터를 가지는 것이다. 위 데이터들 중에서는 S_3가 최소로 충분한 데이터를 가지고 있다고 볼 수 있다. 이러한 값들을 충분통계량(minimal sufficient statistics) 이라고 한다.

데이터가 충분한 지는 어떻게 알까?

임의의 통계량 $T(x)$가 주어질 때, $p(x|T(x), \theta)$가 θ에 종속되지 않을 때 해당 통계량 $T(x)$가 충분하다고 말한다.



$x[n] = A + w[n], \quad w[n] \sim N(0, \sigma^2)$일 때, $T(x) = \sum_{n=0}^{N-1} x[n]$이 충분한 지 알아보자.

$p(x|T(x), A)$이 A에 종속되지 않는지 확인하면 된다.

$$p(\mathbf{x}|T(\mathbf{x})=T_0;A)=\frac{p(\mathbf{x},T(\mathbf{x})=T_0;A)}{p(T(\mathbf{x})=T_0;A)}=\frac{p(\mathbf{x};A)\delta(T(\mathbf{x})-T_0)}{p(T(\mathbf{x})=T_0;A)}$$

$$\begin{aligned}
p(\mathbf{x};A)\delta(T(\mathbf{x})-T_0) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]\delta(T(\mathbf{x})-T_0)\\
&= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}x^2[n]-2AT(\mathbf{x})+NA^2\right)\right]\delta(T(\mathbf{x})-T_0)\\
&= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}x^2[n]-2AT_0+NA^2\right)\right]\delta(T(\mathbf{x})-T_0).
\end{aligned}$$

$$T(\mathbf{x})\sim\mathcal{N}(NA,N\sigma^2)$$

$$\begin{aligned}
p(\mathbf{x}|T(\mathbf{x})=T_0;A) &= \frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]\exp\left[-\frac{1}{2\sigma^2}(-2AT_0+NA^2)\right]}{\frac{1}{\sqrt{2\pi N\sigma^2}}\exp\left[-\frac{1}{2N\sigma^2}(T_0-NA)^2\right]}\delta(T(\mathbf{x})-T_0)\\
&= \frac{\sqrt{N}}{(2\pi\sigma^2)^{\frac{N-1}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]\exp\left[\frac{T_0^2}{2N\sigma^2}\right]\delta(T(\mathbf{x})-T_0)
\end{aligned}$$

해당 전개를 통해 $p(x|T(x),A)$가 A에 종속되지 않음을 알 수 있다.

2. How to find Sufficient Statistics

Theorem : Neyman-Fisher Factorization
$T(x)$가 sufficient statistics임은 $p(x;\theta)=g(T(x),\theta)h(x)$로 표현 가능함과 동치이다. 이때, g는 $T(x)$를 포함하는 x에 종속되며, h는 x에만 종속적이다.

예를 들어,
$x[n]=A+w[n]$, $w[n]\sim N(0,\sigma^2)$에 대해,

$$\begin{aligned}
p(x;A) &= \frac{1}{(2\pi\sigma^2)^{N/2}}\exp[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2]\\
&= \frac{1}{(2\pi\sigma^2)^{N/2}}\exp[-\frac{1}{2\sigma^2}(NA^2-2A\sum_{n=0}^{N-1}x[n])]\exp[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]]\\
&= g(T(x),A)h(x)
\end{aligned}$$

가 되므로, $T(x)=\sum_{n=0}^{N-1}x[n]$에 대해 충분하며,

$T'(x)=2\sum_{n=0}^{N-1}x[n]$에 대해서도 충분하다.

따라서 Sufficient Statistics끼리는 one-to-one transformation이 가능하다.

*Proof of the Neyman-Fisher Factorization
1) <-
$p(x;\theta)=g(T(x),\theta)h(x)$

$$
\begin{aligned}
p(x|T(x)=T_0;\theta) &= \frac{p(x,T(x)=T_0;\theta)}{p(T(x)=T_0;\theta)} \\
&= \frac{p(x;\theta)\delta(T(x)-T_0)}{p(T(x)=T_0;\theta)}
\end{aligned}
$$

따라서, 대입하면

$$
p(x|T(x)=T_0;\theta) = \frac{g(T(x)=T_0,\theta)h(x)\delta(T(x)-T_0)}{p(T(x)=T_0;\theta)}
$$

가 된다.

이 때,

$$
\begin{aligned}
p(T(x)=T_0;\theta) &= \int p(x;\theta)\delta(T(x)-T_0)dx \\
&= \int g(T(x)=T_0,\theta)h(x)\delta(T(x)-T_0)dx \\
&= g(T(x)=T_0,\theta)\int h(x)\delta(T(x)-T_0)dx
\end{aligned}
$$

이므로 이를 위 식에 대입하여 약분하면,

$$
p(x|T(x)=T_0;\theta) = \frac{h(x)\delta(T(x)-T_0)}{\int h(x)\delta(T(x)-T_0)dx}
$$

가 되어 θ에 종속되지 않는다. 따라서 $T(x)$는 sufficient하다.

2) ->

$T(x)$가 sufficient하므로

$p(x|T(x)=T_0;\theta)=p(x|T(x)=T_0)$이다.

$p(x|T(x)=T_0)=w(x)\delta(T(x)-T_0)$, $where \int w(x)\delta(T(x)-T_0)dx=1$로 둔다면,

$$
\begin{aligned}
p(x,T(x)=T_0;\theta) &= p(x|T(x)=T_0;\theta)p(T(x)=T_0;\theta) \\
&= w(x)\delta(T(x)-T_0)p(T(x)=T_0;\theta)
\end{aligned}
$$

이 된다. 이 때,

$p(x,T(x)=T_0;\theta)=p(x;\theta)\delta(T(x)=T_0)$이므로,

$p(x,\theta)\delta(T(x)-T_0)=w(x)\delta(T(x)-T_0)p(T(x)=T_0;\theta)$이다.

그런데 여기서 $w(x)=\dfrac{h(x)}{\int h(x)\delta(T(x)-T_0)dx}$ 로 두고 대입하면,

$p(x,\theta)=\dfrac{p(T(x)=T_0;\theta)}{\int h(x)\delta(T(x)-T_0)dx}h(x)=g(T(x)=T_0;\theta)h(x)$로 정리된다.

***정리**

다음의 조건 중 하나만 만족하면 $T(x)$는 sufficient.

1. $p(x|T(x);\theta)$가 θ에 종속되지 않는다.
2. $p(x;\theta)=g(T(x),\theta)h(x)$로 표현 가능하다.

3. Sufficient Statistics로 MVU Estimator 구하기

$x[n] = A + w[n],\ \ w[n] \sim N(0, \sigma^2)$

해당 문제에서, sufficient statistics인 $T(x) = \sum_{n=0}^{N-1} x[n]$을 사용하여 두 가지 방법으로 MVUE를 구해보자.

Approach 1.

임의의 unbiased 추정값 $\breve{A} = x[0]$가 주어졌다고 하자.

그러면 찾고자 하는 추정값을 $\hat{A} = E(\breve{A}|T)$로 정의한다. 따라서 이 경우에선

$\hat{A} = E(x[0]|\sum_{n=0}^{N-1} x[n])$

로 표현 가능하다.

이를 전개하기 위해선 조건부 pdf를 사용해야 하는데 이는 다음과 같이 유도된다.

두 개의 확률변수 $[x, y]^T$에서 둘의 결합 pdf의 평균은 $\mu = \begin{bmatrix} E(x) \\ E(y) \end{bmatrix}$이고, 분산은 $C = \begin{bmatrix} var(x) & cov(x,y) \\ cov(y,x) & var(y) \end{bmatrix}$이다.

$$
\begin{aligned}
E(x|y) &= \int_{-\infty}^{\infty} x p(x|y) dx \\
&= \int_{-\infty}^{\infty} x \frac{p(x,y)}{p(y)} dx \\
&= E(x) + \frac{cov(x,y)}{var(y)} (y - E(y))
\end{aligned}
$$

따라서 위 식에서 $x = x[0]$, $y = \sum_{n=0}^{N-1} x[n]$이므로

$$
\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x[0] \\ \sum_{n=0}^{N-1} x[n] \end{bmatrix} = \begin{bmatrix} 1\,0\,0\,\ldots\,0 \\ 1\,1\,1\,\ldots\,1 \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ \ldots \\ x[N-1] \end{bmatrix} = L \begin{bmatrix} x[0] \\ x[1] \\ \ldots \\ x[N-1] \end{bmatrix}
$$

로 표현된다.

$[x, y]^T \sim N(\mu, C)$일 때, 원래 $x \sim N(A, \sigma^2)$였으므로 여기에 선형변환을 취하면,

$\mu = LE(x) = LA1 = \begin{bmatrix} A \\ NA \end{bmatrix}$

$C = \sigma^2 L L^T = \sigma^2 \begin{bmatrix} 1 & 1 \\ 1 & N \end{bmatrix}$

이 된다.

따라서 구하고자 하는 추정값은 다음과 같다.

$$
\begin{aligned}
\hat{A} &= E(x|y) \\
&= E(x) + \frac{cov(x,y)}{var(y)}(y - E(y)) \\
&= A + \frac{\sigma^2}{N\sigma^2}\left(\sum_{n=0}^{N-1} x[n] - NA\right) \\
&= \frac{1}{N}\sum_{n=0}^{N-1} x[n]
\end{aligned}
$$

Approach 2.

충분통계량 $T(x)$에 대하여 $\hat{A} = g(T)$가 되는 g를 찾는다.

위 예제에 대해서는 $g(x) = \frac{1}{N}x$가 이를 만족하는 것을 알 수 있다.

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

Approach 2가 Approach 1보다 훨씬 빠르게 MVUE를 찾을 수 있다. 수학적으로는 Approach 1이 더 엄밀하지만 실제 예제에선 Approach 2를 많이 사용한다.

4. RBLS (Rao-Blackwell-Lehmann-Scheffe)

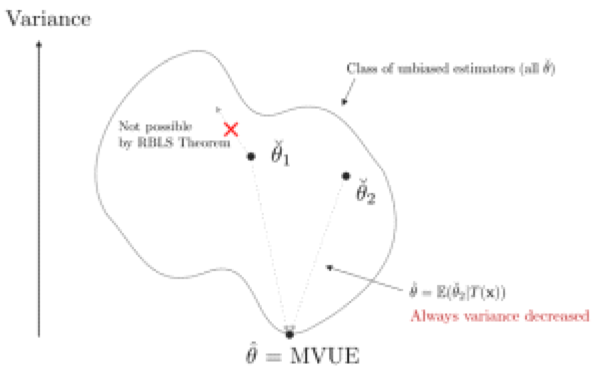
Theorem : Rao-Blackwell-Lehmann-Scheffe

$\check{\theta}$가 θ의 unbiased estimator이고, $T(x)$가 충분통계량일 때, $\hat{\theta} = E(\check{\theta}|T(x))$는 다음을 만족한다.

1. a valid estimator for θ (θ에 종속되지 않음)
2. unbiased
3. 모든 θ에 대해, $\check{\theta}$보다 작거나 같은 분산을 가짐

추가적으로, 충분통계량이 완비하다면, $\hat{\theta}$는 MVU estimator임

$\hat{\theta} = E(\check{\theta}|T(x)) = \int \check{\theta}(x)p(x|T(x);\theta)dx$인데, $T(x)$는 충분통계량이므로, $p(x|T(x);\theta)$는 θ에 종속되지 않는다. 따라서, $\hat{\theta} = \int \check{\theta}p(\check{\theta}|T(x))d\check{\theta} = g(T(x))$가 된다. 따라서 이 조건부 기댓값은 $T(x)$에 대한 단일함수며, 이 논리에 따라 위의 Approach 2가 성립된다.



위 그림과 같이 unbiased estimator들의 집합이 있다고 가정할 때, $E(\check{\theta}|T(x))$ 값은 항상

$\check{\theta}$보다 작거나 같은 분산값을 가진다. 또한 $E(\check{\theta}|T(x))$가 MVUE가 되려면 그림에서와 같이, 그리고 앞서 언급했듯이 $g(T(x))$는 유일한 함수여야 한다. 즉, $T(x)$가 완비통계량이 되기 위해서는 $g(T(x))$가 유일해야 한다.

Ex1) 앞의 문제에서, $T(x) = \sum_{n=0}^{N-1} x[n]$가 완비충분통계량으로써, $g(x) = \frac{1}{N}x$일 때 유일하게 $\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$가 MVUE가 됨을 확인했다. g함수 외에 $E(h(\sum_{n=0}^{N-1} x[n])) = A$를 만족하는 함수 h가 존재한다고 가정하자.

우선,
$E(g(T) - h(T)) = A - A = 0$
을 만족한다.

충분통계량은 $T \sim N(NA, N\sigma^2)$를 만족하기 때문에 $v(T) = g(T) - h(T)$일 때

$$\int_{-\infty}^{\infty} v(T) \frac{1}{\sqrt{2\pi N\sigma^2}} \exp[-\frac{1}{2N\sigma^2}(T - NA)^2] dT = 0 \text{ for } all\ A$$

을 만족한다.

$T = N\tau$로 치환하면,

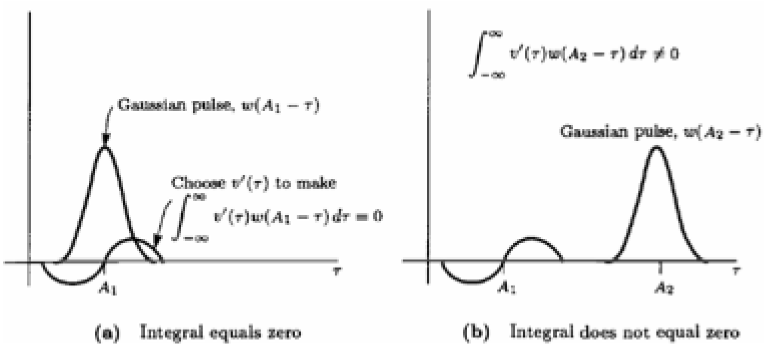
$$\int_{-\infty}^{\infty} v'(\tau) \frac{N}{\sqrt{2\pi N\sigma^2}} \exp[-\frac{N}{2\sigma^2}(A - \tau)^2] d\tau = 0 \text{ for } all\ A$$

을 만족한다.

뒤의 가우시안 펄스를 $w(\tau)$라 하면,

$$\int_{-\infty}^{\infty} v'(\tau) w(A - \tau) d\tau = 0 \text{ for } all\ A$$

를 만족해야 한다.



(a) Integral equals zero (b) Integral does not equal zero

그런데 (a)와 같이 $A = A_1$일 때 위 식을 만족하기 위해서는 $v'(\tau)$는 (a)와 같이 그려져야 한다. 하지만 위 식은 모든 A에 대해서 성립해야 하므로 (b)에서처럼 $A = A_2$에서도 성립해야 하는데 그렇지 못한다. 따라서 $v(T) = g(T) - h(T) = 0$인 경우에만 위 식이 성립한다. 그러므로 $g(T)$는 유일하다.

Ex2) 이번에는 $x[0] = A + w[0]$의 데이터가 주어지고, $w[0] \sim u[-\frac{1}{2}, \frac{1}{2}]$의 분포가 주어졌다. 이 때 충분통계량은 $T = x[0]$이고 A의 unbiased estimator은 $x[0]$이다. 이번에도 $g(T) = g(x[0]) = x[0]$외에 다른 함수 h가 있는 지 확인해보자.

$v(T) = g(T) - h(T)$라고 하면,

$$\int_{-\infty}^{\infty} v(T)p(x;A)dx = 0 \text{ for } all\ A$$

가 성립한다. 그런데 여기서 $x = x[0] = T$이므로,

$$\int_{-\infty}^{\infty} v(T)p(T;A)dT = 0 \text{ for } all\ A$$

가 성립한다.

이때, $w[0] \sim u[-\frac{1}{2}, \frac{1}{2}]$이므로,

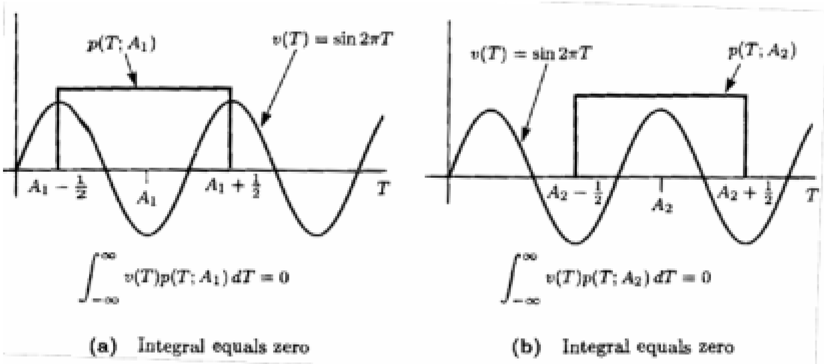
$$p(T;A) = \begin{cases} 1 & A - \frac{1}{2} \le T \le A + \frac{1}{2} \\ 0 & otherwise \end{cases}$$

이다.

따라서 이를 대입하면,

$$\int_{A-\frac{1}{2}}^{A+\frac{1}{2}} v(T)dT = 0$$

가 성립한다. 그런데 이는 아래 그림과 같이 $v(T) = \sin 2\pi T$일 때 모든 A에 대해서 성립하므로 완비통계량이 complete하진 않으며 해당 추정값도 MVUE가 아니다.



*Definition

$$\int_{-\infty}^{\infty} v(T)p(T;A)dT = 0 \text{ for } all\ A$$

위 식이 오직 $v(T) = 0$일 때만 성립하면 충분통계량 T는 완비하다.

5. 정리

MVUE를 찾는 과정은 다음과 같다.

1. Neyman-Fisher factorization theorem으로 충분통계량을 찾는다.
2. 찾은 충분통계량이 완비한지 확인한다.
3. $\hat{\theta}= E(\check{\theta}|T)$(approach 1) 혹은 $\hat{\theta}= g(T)$(approach 2)로 MVUE를 찾는다.

6. Vector parameter case

$x[n] = A + w[n],\ where\ w[n] \sim N(0,\sigma^2)$

에서 A와 σ^2를 추정해보자.

우선 충분통계량은

$$T(x) = \begin{bmatrix} T_1(x) \\ T_2(x) \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \\ \sum_{n=0}^{N-1} x^2[n] \end{bmatrix}$$

이다.

위 충분통계량에 대해 기댓값을 구해보면,

$$E(T(x)) = \begin{bmatrix} NA \\ NE(x^2[n]) \end{bmatrix} = \begin{bmatrix} NA \\ N(\sigma^2 + A^2) \end{bmatrix}$$

가 된다. 첫 번째 항은 $\frac{1}{N}$배 해주면 쉽게 bias가 제거되지만 두 번째 항은 복잡하다. 위 bias들을 제거해주기 위해 식을 다음과 같이 변형한다.

$$\begin{aligned} g(T(x)) &= \begin{bmatrix} \frac{1}{N} T_1(x) \\ \frac{1}{N-1}[T_2(x) - N(\frac{1}{N} T_1(x))^2] \end{bmatrix} \\ &= \begin{bmatrix} \bar{x} \\ \frac{1}{N-1}[\sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2] \end{bmatrix} \\ &= \begin{bmatrix} \bar{x} \\ \frac{1}{N-1} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 \end{bmatrix} \end{aligned}$$

이와 같이 변형하면 두 번째 항의 기댓값은

$$\frac{1}{N-1}(E(T_2) - NE(\bar{x}^2)) = \frac{1}{N-1}(N(\sigma^2 + A^2) - N(A^2 + \sigma^2/N)) = \sigma^2$$

가 되어 bias가 제거된다.

따라서 MVUE는

$\hat{\theta}= \begin{bmatrix} \bar{x} \\ \frac{1}{N-1} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 \end{bmatrix}$가 되지만 분산을 구해보면

$$C = \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N-1} \end{bmatrix}$$

이 되어 CRLB로 구한 분산인

$$I^{-1}(\theta) = \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix}$$

보다 작다.
따라서 해당 MVUE는 efficient하지 않다.

이와 같이 RBLS를 이용하면 CRLB로 구할 수 없는 efficient하지 않은 MVUE도 구할 수 있다.