

430.714

Estimation Theory

Prof. Songhwai Oh

ECE, SNU

Sufficient Statistics

Sufficient Statistics

Is there a set of data that is sufficient for the estimation problem?

$$x[n] = A + w[n], \quad \text{where } w[n] \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

$$\text{MVU estimator: } \hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

$$S_1 = \{x[0], x[1], \dots, x[N-1]\}$$

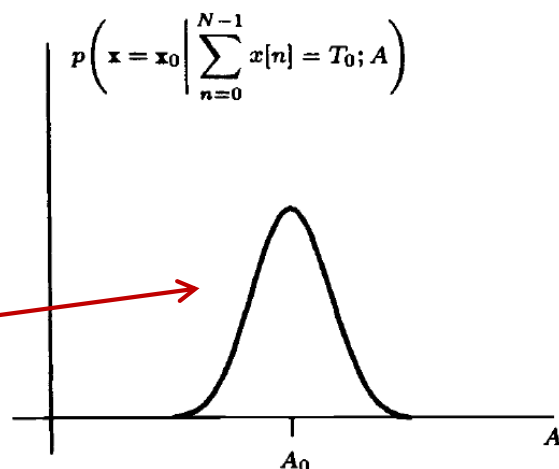
$$S_2 = \{x[0] + x[1], x[2], x[3], \dots, x[N-1]\}$$

$$S_3 = \left\{ \sum_{n=0}^{N-1} x[n] \right\}.$$

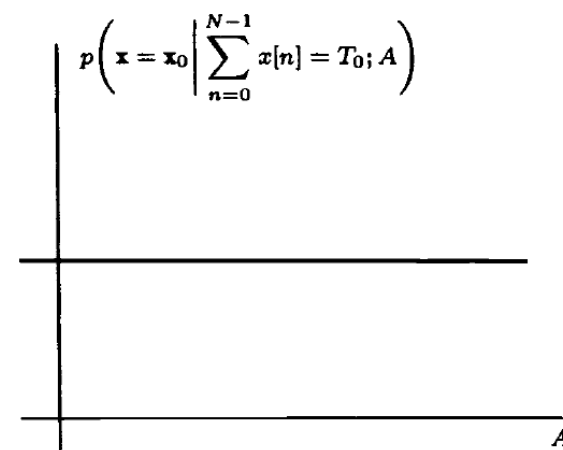
All sufficient. We are looking for the **minimal** sufficient statistic.

$T(\mathbf{x})$ is a sufficient statistic if $p(\mathbf{x}|T(\mathbf{x}); \theta)$ does not depend on θ .

There is more information in $\mathbf{x}$ which is not captured by $T(\mathbf{x})$.



(a) Observations provide information after $T(\mathbf{x})$ observed— $T(\mathbf{x})$ is not sufficient



(b) No information from observations after $T(\mathbf{x})$ observed— $T(\mathbf{x})$ is sufficient

$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$ is a sufficient statistic. $x[n] = A + w[n]$, where $w[n] \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$

Need to check if $p(\mathbf{x}|T(\mathbf{x}); A)$ does not depend on A .

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

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

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

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

- Does not depend on A .
- $T(\mathbf{x})$ is a sufficient statistic.

Neyman-Fisher Factorization

Theorem: $T(\mathbf{x})$ is a sufficient statistic for θ if and only if $p(\mathbf{x}; \theta)$ can be factored as follows

$$p(\mathbf{x}; \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x}),$$

where g is a function depending on $\mathbf{x}$ only through $T(\mathbf{x})$ and h is a function of only $\mathbf{x}$.

Example: DC level in WGN $x[n] = A + w[n]$, where $w[n] \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$

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

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

$$p(\mathbf{x}; A) = \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \left(NA^2 - 2A \sum_{n=0}^{N-1} x[n] \right) \right]}_{g(T(\mathbf{x}), A)} \underbrace{\exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right]}_{h(\mathbf{x})}.$$

- $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$ is a sufficient statistic for A .
- $T'(\mathbf{x}) = 2 \sum_{n=0}^{N-1} x[n]$ is also a sufficient statistic for A .

PrC • Sufficient statistics are unique only to within *one-to-one* transformations.

Proof of the Neyman-Fisher Factorization

If $p(\mathbf{x}; \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x})$, then $T(\mathbf{x})$ is a sufficient statistic.

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

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

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

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

Not a function of θ . Hence, $T(\mathbf{x})$ is a sufficient statistic.

Proof of the Neyman-Fisher Factorization

If $T(\mathbf{x})$ is a sufficient statistic, then $p(\mathbf{x}; \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x})$.

$$p(\mathbf{x}, T(\mathbf{x}) = T_0; \theta) = p(\mathbf{x}|T(\mathbf{x}) = T_0; \theta)p(T(\mathbf{x}) = T_0; \theta)$$

- $p(\mathbf{x}|T(\mathbf{x}) = T_0; \theta) = p(\mathbf{x}|T(\mathbf{x}) = T_0)$ since $T(\mathbf{x})$ is a sufficient statistic.
- Let $p(\mathbf{x}|T(\mathbf{x}) = T_0) = w(\mathbf{x})\delta(T(\mathbf{x}) - T_0)$, where $\int w(\mathbf{x})\delta(T(\mathbf{x}) - T_0)d\mathbf{x} = 1$.

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

$$p(\mathbf{x}; \theta)\delta(T(\mathbf{x}) - T_0) = w(\mathbf{x})\delta(T(\mathbf{x}) - T_0)p(T(\mathbf{x}) = T_0; \theta)$$

Let
$$w(\mathbf{x}) = \frac{h(\mathbf{x})}{\int h(\mathbf{x})\delta(T(\mathbf{x}) - T_0) d\mathbf{x}}$$

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

$$p(\mathbf{x}; \theta) = g(T(\mathbf{x}) = T_0; \theta)h(\mathbf{x}), \quad \text{where} \quad g(T(\mathbf{x}) = T_0; \theta) = \frac{p(T(\mathbf{x}) = T_0; \theta)}{\int h(\mathbf{x})\delta(T(\mathbf{x}) - T_0) d\mathbf{x}}$$

Sufficient Statistics

We can show that $T(\mathbf{x})$ is a sufficient statistics for θ by checking one of the following conditions:

1. $p(\mathbf{x}|T(\mathbf{x}); \theta)$ does not depend on θ .
2. $p(\mathbf{x}; \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x})$ (Neyman-Fisher factorization).

Finding the MVU estimator using sufficient statistics

1. Find *any* unbiased estimator of θ , say $\check{\theta} = x[0]$, and determine $\hat{\theta} = \mathbb{E}(\check{\theta}|T)$, where T is a sufficient statistic for θ . The expectation is taken with respect to $p(\check{\theta}|T)$; or
2. Find some function g so that $\hat{\theta} = g(T)$ is an unbiased estimator of θ .

Approach 1

Example: DC level in WGN $x[n] = A + w[n]$, where $w[n] \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$

Let $\check{A} = x[0]$. With $T = \sum_{n=0}^{N-1} x[n]$, determine $\hat{A} = \mathbb{E}(\check{A}|T) = \mathbb{E}(x[0] | \sum_{n=0}^{N-1} x[n])$.

Fact: For a multivariate Gaussian r.v. $[x \ y]^T$, its mean is $\mu = [\mathbb{E}(x) \ \mathbb{E}(y)]^T$ and its covariance matrix is $C = \begin{bmatrix} \mathbf{var}(x) & \mathbf{cov}(x, y) \\ \mathbf{cov}(y, x) & \mathbf{var}(y) \end{bmatrix}$. Then

$$\mathbb{E}(x|y) = \mathbb{E}(x) + \frac{\mathbf{cov}(x, y)}{\mathbf{var}(y)}(y - \mathbb{E}(y)).$$

Approach 1

Let $x = x[0]$ and $y = \sum_{n=0}^{N-1} x[n]$.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x[0] \\ \sum_{n=0}^{N-1} x[n] \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}}_{\mathbf{L}} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}$$

$$[x \ y]^T \sim \mathcal{N}(\mu, C)$$

$$\begin{aligned} \mu &= \mathbf{L}E(\mathbf{x}) = \mathbf{L}A\mathbf{1} = \begin{bmatrix} A \\ NA \end{bmatrix} \\ \mathbf{C} &= \sigma^2 \mathbf{L}\mathbf{L}^T = \sigma^2 \begin{bmatrix} 1 & 1 \\ 1 & N \end{bmatrix}. \end{aligned}$$

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

Approach 2

$$\hat{A} = g \left(\sum_{n=0}^{N-1} x[n] \right)$$

- With $g(x) = x/N$, $\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$ is an unbiased estimate of A .
- $\hat{A}$ is the MVU estimator.

Approach 2 is simpler.

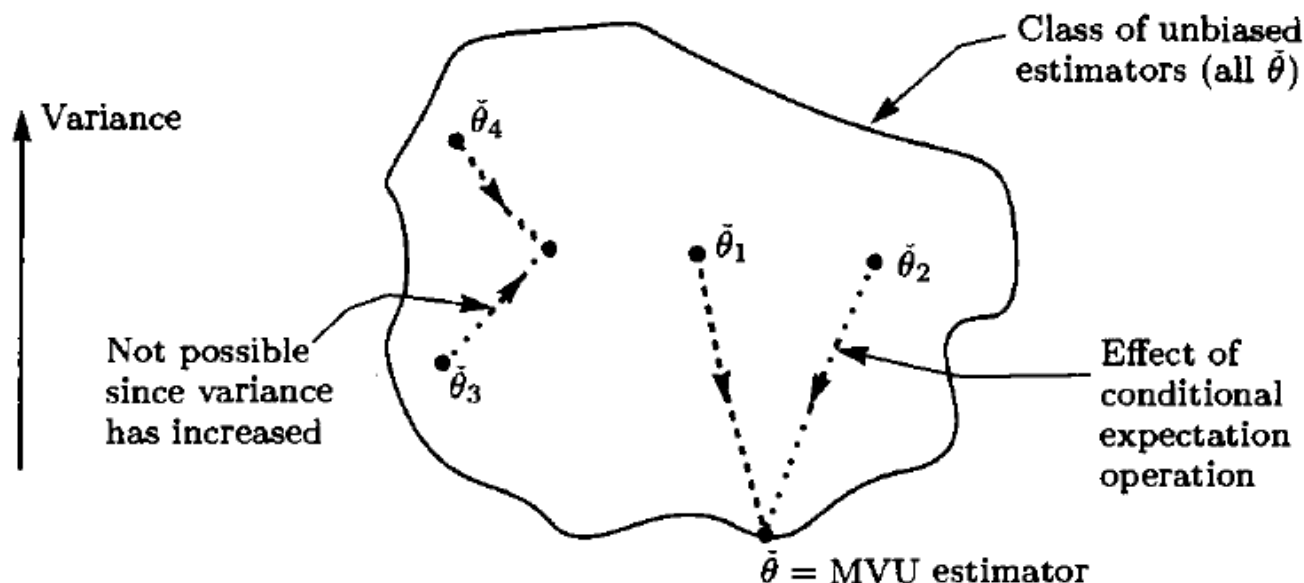
Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ , then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is

1. a valid estimator for θ (not dependent on θ)
2. unbiased
3. of lesser or equal variance than that of $\check{\theta}$, for all θ .

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

$T(\mathbf{x})$ is *complete*, if there is only one function of T that is an unbiased estimator.

$$\begin{aligned}\hat{\theta} &= E(\check{\theta}|T(\mathbf{x})) = \int \check{\theta} p(\check{\theta}|T(\mathbf{x})) d\check{\theta} \\ &= g(T(\mathbf{x})).\end{aligned}$$



Rao-Blackwell-Lehmann-Scheffe Theorem

$$\begin{aligned} (1) \quad \hat{\theta} &= E(\check{\theta}|T(\mathbf{x})) \\ &= \int \check{\theta}(\mathbf{x})p(\mathbf{x}|T(\mathbf{x}); \theta) d\mathbf{x} \end{aligned} \quad \begin{array}{l} p(\mathbf{x}|T(\mathbf{x}); \theta) \text{ does not depend on } \theta. \\ \text{Hence, } \hat{\theta} \text{ is not dependent on } \theta. \end{array}$$

$$\begin{aligned} (2) \quad E(\hat{\theta}) &= \int \int \check{\theta}(\mathbf{x})p(\mathbf{x}|T(\mathbf{x}); \theta) d\mathbf{x} p(T(\mathbf{x}); \theta) dT \\ &= \int \check{\theta}(\mathbf{x}) \int p(\mathbf{x}|T(\mathbf{x}); \theta)p(T(\mathbf{x}); \theta) dT d\mathbf{x} \\ &= \int \check{\theta}(\mathbf{x})p(\mathbf{x}; \theta) d\mathbf{x} \\ &= E(\check{\theta}). \end{aligned} \quad \mathbb{E}(\hat{\theta}) = \mathbb{E}(\check{\theta}) = \theta$$

Completeness of a sufficient statistic

Definition: A sufficient statistic T is *complete* if the condition

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

is satisfied only by the zero function, i.e., $v(T) = 0$ for all T .

Complete sufficient statistic: An example

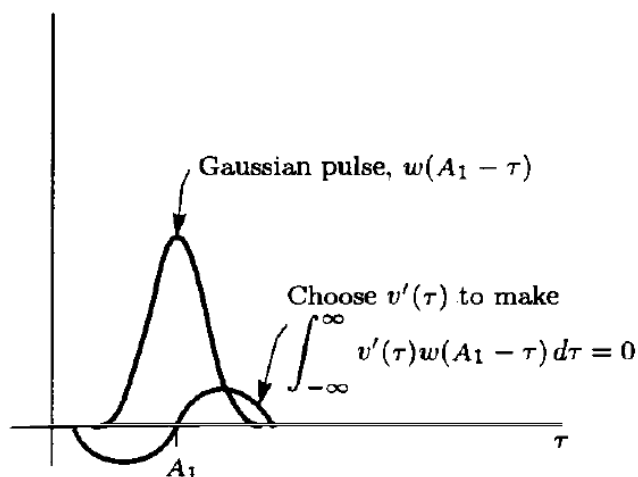
- $\sum_{n=0}^{N-1} x[n]$ is complete ($\exists!$ function g for which $\mathbb{E}(g(\sum_{n=0}^{N-1} x[n])) = A$).
- Suppose there is another function h such that $\mathbb{E}(h(\sum_{n=0}^{N-1} x[n])) = A$.
- $\mathbb{E}(g(T) - h(T)) = A - A = 0$ for all A .
- Since $T \sim \mathcal{N}(NA, N\sigma^2)$, ($v(T) = g(T) - h(T)$, $\tau = T/N$, $v'(\tau) = v(N\tau)$)

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

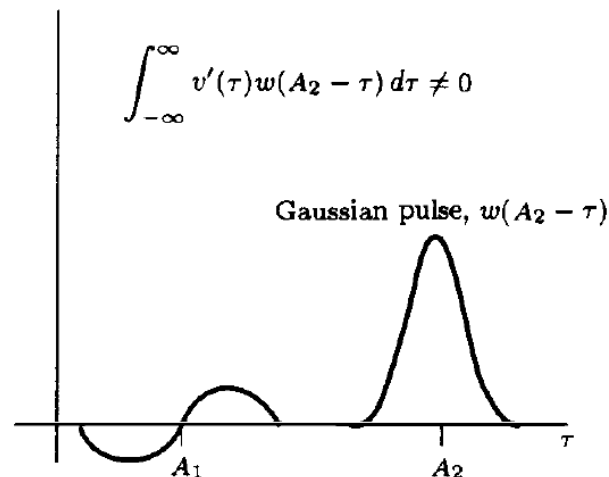
$$\int_{-\infty}^{\infty} v'(\tau) \frac{N}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{N}{2\sigma^2}(A - \tau)^2\right] d\tau = 0 \quad \text{for all } A$$

Convolution of $v'(\tau)$ with a Gaussian pulse $w(\tau)$.

This expression is true for all A if and only if $v'(\tau)=0$ for all τ .



(a) Integral equals zero



(b) Integral does not equal zero

Incomplete sufficient statistic: An example

- $x[0] = A + w[0]$, where $w[0] = \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$.
- $T = x[0]$ is a sufficient statistic. $x[0]$ is an unbiased estimator of A .
- Consider two unbiased estimators $g(x[0]) = x[0]$ and $h(x[0])$, and let $v(T) = g(T) - h(T)$.
- If T is a complete sufficient statistic,

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

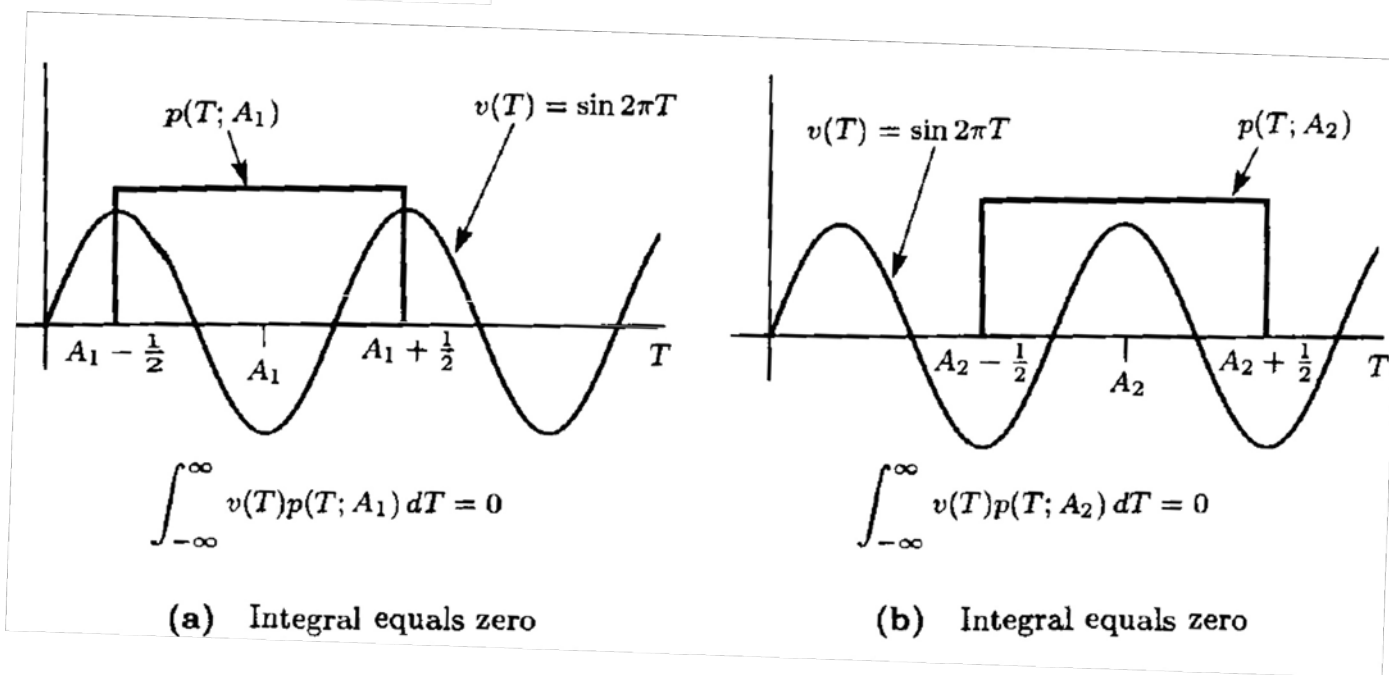
$$\int_{A-\frac{1}{2}}^{A+\frac{1}{2}} v(T) dT = 0 \quad \text{for all } A \quad p(T; A) = \begin{cases} 1 & A - \frac{1}{2} \leq T \leq A + \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

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

$$v(T) = \sin 2\pi T$$

$$g(T) = T \text{ and } h(T) = T - \sin 2\pi T$$

$$\hat{A} = x[0] - \sin 2\pi x[0]$$



Unbiased but
not complete

A procedure for finding the MVU estimator:

Use Neyman-Fisher factorization theorem to find sufficient statistic

$T(\mathbf{x})$

Determine if $T(\mathbf{x})$ is complete
See (5.8)

Find function of $T(\mathbf{x})$ that is unbiased

1. Evaluate $\hat{\theta} = \mathbb{E}(\check{\theta}|T)$, where $\check{\theta}$ is any unbiased estimator.
2. Find a function g so that $\hat{\theta} = g(T)$ is an unbiased estimator of θ .

$\hat{\theta} = g(T(\mathbf{x})) = \text{MVU estimator}$

$$p(\mathbf{x}; \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x}).$$

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

is satisfied only by the zero function,
i.e., $v(T) = 0$ for all T .

Vector parameter case

$$x[n] = A + w[n], \quad \text{where } w[n] \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

Unknowns: A and σ^2 ; $\theta = [A, \sigma^2]^T$

Sufficient statistics:

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

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

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

Unbiased

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

MVU estimator:

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

But not efficient.

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