Sufficient statistics, Neyman-Fisher factorization

Thursday, 18 March 2021 13:56

$\underline{x}$ $\qquad$ $p(\underline{x};\theta)$

$$\text{mse}(\hat\theta) = E(|\theta-\hat\theta|^2) = \text{var}(\hat\theta) + b^2(\theta)$$

$$\hat\theta = g(\underline{x})$$

- MVU Estimators
  - → min. variance
  - → $E(\hat\theta)=\theta \quad \forall\theta$

- CRLB

$$\text{var}(\hat\theta) \geq \frac{1}{I(\theta)}$$

($\hat\theta$ → unbiased)

$I(\theta)$ → variance of the score function
or
average curvature of the likelihood function

$$\frac{\partial \ln p(\underline{x};\theta)}{\partial\theta} = I(\theta)\,[g(x)-\theta]$$

$g(x) = \hat\theta$ → MVU, $\text{var}(\hat\theta) = \frac{1}{I(\theta)}$

Sufficient Statistics:

$$\underline{x} = [x[0], x[1], \dots x[N-1]]^T$$

$\underline{x}$ is related to 'θ'

$T(\underline{x})$ captures all the information about 'θ'

then we may discard $\underline{x}$

→ "Sufficient Statistic" : compress raw observations to draw inference about 'θ'

$$T(x) = \sum_{n=0}^{N-1} x[n] \qquad g(T(x)) = \hat\theta = \frac{1}{N}T(x)$$

Example: $x = A + w[n]$ $\qquad n = 0, 1, \dots, N-1$

$$\hat A = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \qquad \text{var}(\hat A) = \frac{\sigma^2}{N}$$

$$\check A = x[0] \qquad \text{var}(\check A) = \sigma^2$$

data sets:

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

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

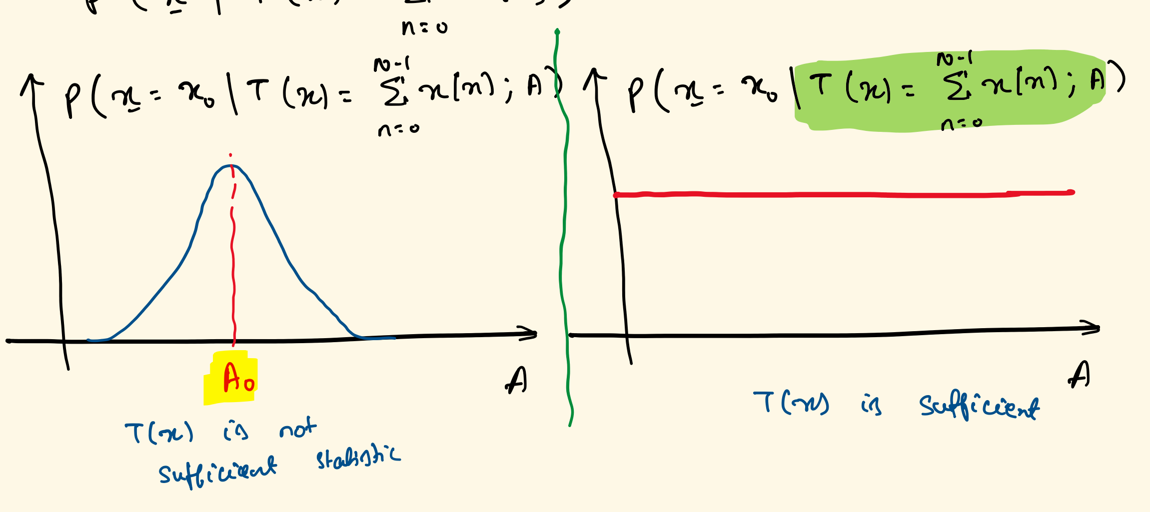
1 statistic $\quad S_3 = \{\sum_{n=0}^{N-1} x[n]\}$

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

$p(\underline{x};A)$: data PDF

Once we observe $T(x) = \sum_{n=0}^{N-1} x[n]$

$$p\left(\underline{x} \mid T(x) = \sum_{n=0}^{N-1} x[n]\right)$$



Example:

Recall:

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

To prove that $T(x) = \sum_{n=0}^{N-1} x[n]$ is a Sufficient statistic. (→ functionally depends on $\underline{x}$)

To do so,

$$p(\underline{x} \mid T(x) = T_0; A)$$

[Suppose we have 2 random variable x and y such that $x = y$.

$$p(x, y) = p(x)\,\delta(x-y)$$

and this collapses to a line in 2D.]

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

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

$$p(\underline{x};A)\,\delta(T(\underline{x})-T_0)$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n] - 2AT(\underline{x}) + NA^2\right]\delta(T(\underline{x})-T_0)$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n]\right] \exp\left[-\frac{1}{2\sigma^2}(-2AT_0 + NA^2)\right]\delta(T(x)-T_0)$$

$$p(\underline{x} \mid T(x) = T_0; A)$$

$$= \frac{\frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n]\right] \exp\left[-\frac{1}{2\sigma^2}(-2AT_0 + NA^2)\right]\delta(T(x)-T_0)}{\frac{1}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{1}{2N\sigma^2}(T_0 - NA)^2\right]}$$

$$-\frac{1}{2\sigma^2}\left[-2AT_0 + NA^2 - \frac{T_0^2}{N} - NA^2 + 2AT_0\right]$$

$$= \frac{\sqrt{N}}{(2\pi\sigma^2)^{\frac{N-1}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n]\right] \exp\left[\frac{T_0^2}{2N\sigma^2}\right]\delta(T(x)-T_0)$$

Does not depend on 'A'

Neyman-Fisher factorization:

If we can factorize the PDF $p(\underline{x};\theta)$ as

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

- $g(x)$ is depending on $\underline{x}$ through $T(x)$
- h is a function that depends only $\underline{x}$

Then, $T(\underline{x})$ is a sufficient statistic for θ.

Conversely, if $T(x)$ is a sufficient statistic, then the PDF can be factored as above.

Example: $x[n] = A + w[n]$ $\qquad n = 0, 1, \dots, N-1$

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

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

$$T_1(\underline{x}) = \sum_{n=0}^{N-1} x[n]$$

$$T_2(\underline{x}) = 2\sum_{n=0}^{N-1} x[n]$$

} unique only up to a one-to-one transformation

Example 2: $A = 0$ and find σ^2

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

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

Using T(x) (i.e., Sufficiency) to find MVU:

Rao - Blackwell - Lehmann - Scheffe theorem:

- Given $T(x)$; find some function $g(T(x))$ so that $g(T(x))$ is unbiased
  - Example: $T(x) = \sum_{n=0}^{N-1} x[n]$
  - $g(\cdot) = \frac{1}{N}$
  - $\hat\theta = \frac{1}{N}T(\underline{x})$: unbiased
- Given $T(x)$, let us take any unbiased estimator $\check\theta$,
  - $\hat\theta = E[\check\theta \mid T]$

If $\check\theta$ is an unbiased estimator of θ and $T(x)$ is a sufficient statistic, then $\hat\theta = E[\check\theta \mid T(x)]$ is

(a) unbiased, does not depend on θ

(b) $\text{var}(\hat\theta) \leq \text{var}(\check\theta) \quad \forall\theta$

Furthermore, if $T(x)$ is complete sufficient statistic, then $\hat\theta$ is the MVU estimator.