Sufficiency and minimal sufficiency

Where we are and where we're going

So far:

- we know how to calculate MLEs
- we know that in terms of MSE, MLEs are asymptotically the "best" we can do (under regularity conditions)
    - asymptotically unbiased
    - asymptotic variance is CRLB
- we also know how to Rao-Blackwellize an unbiased estimator to improve MSE
    - condition on sufficient statistics
- sufficient statistics contain all the info we need to calculate MLEs

Currently:

- how do we find sufficient statistics?

Up next:

- Are there more "compact" sufficient statistics?
- when do sufficient statistics give me a best unbiased estimator?

Recap: factorization theorem

Let $X_1, \dots, X_n$ be a sample with joint probability function $f(x_1, \dots, x_n \mid \theta)$. A statistic $T = T(x_1, \dots, x_n)$ is sufficient for θ if and only if there exist function $g(t \mid \theta)$ and $h(x_1, \dots, x_n)$ such that, for all possible $x_1, \dots, x_n$ and all possible θ,

$$f(x_1, \dots, x_n \mid \theta) = \underbrace{g(T(x_1, \dots, x_n) \mid \theta)}_{\text{joint distribution only depends on } \theta \text{ through the sufficient statistic}} h(x_1, \dots, x_n)$$

Pf: (we will prove for discrete distributions)

① Suppose the functions g & h exist. WTS T is sufficient.

$$f(x_1, \dots, x_n \mid T, \theta) = \frac{f(x_1, \dots, x_n, T(x_1, \dots, x_n) \mid \theta)}{f(T(x_1, \dots, x_n) \mid \theta)}$$

$$= \frac{f(x_1, \dots, x_n \mid \theta)}{f(T(x_1, \dots, x_n) \mid \theta)} = \frac{g(T(x_1, \dots, x_n) \mid \theta)\, h(x_1, \dots, x_n)}{\sum\limits_{\substack{y_1, \dots, y_n \\ T(y_1, \dots, y_n) = T(x_1, \dots, x_n)}} f(y_1, \dots, y_n \mid \theta)}$$

$$\frac{g(T(x_1,\dots,x_n)\mid\theta)\; h(x_1,\dots,x_n)}{\displaystyle\sum_{\substack{y_1,\dots,y_n\\ T(y_1,\dots,y_n)=T(x_1,\dots,x_n)}} f(y_1,\dots,y_n\mid\theta)} = \frac{g(T(x_1,\dots,x_n)\mid\theta)\, h(x_1,\dots,x_n)}{\displaystyle\sum_{\substack{y_1,\dots,y_n\\ T(y_1,\dots,y_n)=T(x_1,\dots,x_n)}} g(T(y_1,\dots,y_n)\mid\theta)\, h(y_1,\dots,y_n)}$$

$$= \frac{\cancel{g(T(x_1,\dots,x_n)\mid\theta)}\, h(x_1,\dots,x_n)}{\cancel{g(T(x_1,\dots,x_n)\mid\theta)}\displaystyle\sum_{\substack{y_1,\dots,y_n\\ T(y_1,\dots,y_n)=T(x_1,\dots,x_n)}} h(y_1,\dots,y_n)} = \frac{h(x_1,\dots,x_n)}{\displaystyle\sum_{\substack{y_1,\dots,y_n\\ T(y_1,\dots,y_n)=T(x_1,\dots,x_n)}} h(y_1,\dots,y_n)}$$

(does not depend on θ)

✓

(2) Suppose T is a sufficient statistic.

Let $g(t\mid\theta) = P_\theta(T(X_1,\dots,X_n) = t)$

$h(x_1,\dots,x_n) = P(X_1,\dots,X_n = x_1,\dots,x_n \mid T(X_1,\dots,X_n) = T(x_1,\dots,x_n))$ (does not depend on θ)

$$\begin{aligned} f(x_1,\dots,x_n\mid\theta) &= P_\theta(X_1,\dots,X_n = x_1,\dots,x_n) \\ &= P_\theta(X_1,\dots,X_n = x_1,\dots,x_n \text{ and } T(X_1,\dots,X_n) = T(x_1,\dots,x_n)) \\ &= P_\theta(T(X_1,\dots,X_n) = T(x_1,\dots,x_n))\, P(X_1,\dots,X_n = x_1,\dots,x_n \mid T(X_1,\dots,X_n) = T(x_1,\dots,x_n)) \\ &= g(T(x_1,\dots,x_n)\mid\theta)\, h(x_1,\dots,x_n) \end{aligned}$$

//

Example

Suppose $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with μ and σ^2 unknown. Let $\theta = (\mu, \sigma^2)$.

> Use the factorization theorem to find a sufficient statistic for θ. (The dimension of the statistic will be greater than 1).

$T = (\bar{x}, s)$

$T = (\bar{x}, s^2)$

$T = (\sum_i x_i, \sum_i x_i^2)$

$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$$\sum_i (x_i^2 - 2\mu x_i + \mu^2) = \sum_i x_i^2 - 2\mu \sum_i x_i + \sum_i \mu^2$$

$$f(x_1, \dots, x_n \mid \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \left(\sum_i x_i^2 - 2\mu \sum_i x_i + n\mu^2\right)\right\}$$

$$= g\left(\sum_i x_i, \sum_i x_i^2 \mid \mu, \sigma^2\right)$$

$$= g\left(\sum_i x_i, \sum_i x_i^2 \mid \mu, \sigma^2\right) \underbrace{h(x_1, \dots, x_n)}_{=1}$$

Sufficient statistic $= \left(\sum_i X_i, \sum_i X_i^2\right)$

$$\sum_i (X_i - \bar{X})^2 = \sum_i X_i^2 - n(\bar{X})^2$$

also so: $\left(\frac{1}{n} \sum_i X_i, \frac{1}{n-1} \sum_i (X_i - \bar{X})^2\right)$

Is the dimension of a sufficient statistic always = dimension of Θ?

NO: (a) $(X_1, X_2, X_3, \dots, X_n)$ is sufficient

(b) EX: $X_1, \dots, X_n \overset{iid}{\sim}$ Uniform$[\Theta, \Theta+1]$

minimal sufficient statistic: $(X_{(1)}, X_{(n)})$

Sufficient statistics as partitions

Suppose $X_1, X_2, X_3 \overset{iid}{\sim} \text{Bernoulli}(p)$, we know $T = X_1 + X_2 + X_3$ is sufficient

(x_1, x_2, x_3)	$T(x_1, x_2, x_3)$	$P(x_1, x_2, x_3 \mid T)$
$(0,0,0)$	0	1
$(1,0,0)$	1	$1/3$
$(0,1,0)$	1	$1/3$
$(0,0,1)$	1	$1/3$
$(1,1,0)$	2	$1/3$
$(1,0,1)$	2	$1/3$
$(0,1,1)$	2	$1/3$
$(1,1,1)$	3	1

equivalence classes

$T(x_1, x_2, x_3)$ partitions space (x_1, x_2, x_3) into sets B_1, B_2, B_3, B_4

$B_1 = \{(x_1, x_2, x_3) : T = 0\}$, $B_2 = \{(x_1, x_2, x_3) : T = 1\}$, etc.

T is sufficient $\iff$ $f((x_1, x_2, x_3) \mid (x_1, x_2, x_3) \in B_i)$ does not depend on p

Another partition $X_1, X_2, X_3 \overset{iid}{\sim} \text{Bernoulli}(p)$ $T = (X_1, X_2, X_3)$

(x_1, x_2, x_3)	$T(x_1, x_2, x_3)$	$P(x_1, x_2, x_3 \mid T)$
$(0,0,0)$	$(0,0,0)$	1
$(1,0,0)$	$(1,0,0)$	1
$(0,1,0)$	$(0,1,0)$	1

⋮

etc.

T partitions sample space into 8 groups (one per (x_1, x_2, x_3))

<u>Question</u>: which statistic more efficiently summarizes the data? (i.e., results in fewer partitions?)

Minimal sufficient statistics

Def: A statistic $T(X_1, \dots, X_n)$ is a minimal sufficient statistic if for any other sufficient statistic $T^*(X_1, \dots, X_n)$, $T(X_1, \dots, X_n)$ is a function of $T^*(X_1, \dots, X_n)$

Ex: $X_1, X_2, X_3 \overset{iid}{\sim} \text{Bernoulli}(p)$

$T = \sum_i X_i \qquad T^* = (X_1, X_2, X_3)$

T is a function of T^*, but T^* is not a function of T