§4 Sufficiency and Likelihood

§4.1 Motivating example

Let $X \in \{1, 2, 3, 4\}$ be a discrete random variable with mass function $f(x|\theta)$ tabulated below:

$\theta \setminus x$	1	2	3	4
0	0.25	0.15	0.45	0.15
1	0.5	0.1	0.3	0.1

Suppose $X = x$ is observed. We would like to learn from x as much as possible about the unknown parameter $\theta \in \{0, 1\}$. This is a typical statistical inference problem.
More generally, we shall discuss the following important questions concerning statistical inference:

- Does x contain any information relevant to θ?
- Does x contain information more than necessary for learning about θ?
- If x contains information more than necessary for learning about θ, how can we compress x "*economically*" (i.e. make it simpler) without losing any useful information about θ?

§4.2 Statistics as carriers of information

4.2.1 Sample data: $\boldsymbol{X} \in \mathcal{S}$ = sample space
Postulated parametric family of probability functions for $\boldsymbol{X}$: $\{f(\cdot|\theta) : \theta \in \Theta\}$

A function $T(\boldsymbol{X})$ of $\boldsymbol{X}$ is generally known as a *statistic*. If the sampling distributions of $T(\boldsymbol{X})$ under different $\theta \in \Theta$ are not all identical, then observation of $T(\boldsymbol{X})$ provides useful information **relevant** to the unknown true value of θ.

4.2.2 If two statistics T_1, T_2 satisfy:

$$T_1(\boldsymbol{X}) = T_1(\boldsymbol{X}') \;\Leftrightarrow\; T_2(\boldsymbol{X}) = T_2(\boldsymbol{X}') \qquad \forall\, \boldsymbol{X}, \boldsymbol{X}' \in \mathcal{S},$$

then T_1 and T_2 carry the **same** information about θ.

Information about θ carried by a statistic $T(\boldsymbol{X})$ is determined essentially by the *partition* of $\mathcal{S}$ induced by distinct values of $T(\boldsymbol{X})$. It does not matter how we define those distinct values of $T(\boldsymbol{X})$ as long as they produce the same partition.

4.2.3 If $T_2(\boldsymbol{X}) = \psi\big(T_1(\boldsymbol{X})\big)$ for some function ψ, then the statistic T_2 carries **no more** information about θ than T_1.

Trivial special case: $T(\boldsymbol{X})$ carries no more information about θ than the raw sample data $\boldsymbol{X}$.

§4.3 Sufficiency

4.3.1 **Definition.** Statistic $T = T(\boldsymbol{X})$ is *sufficient* for θ if the conditional distribution of $\boldsymbol{X}$ given T is **free** of θ.

4.3.2 A sufficient statistic carries **all** the information contained in data $\boldsymbol{X}$ that is relevant to θ.

4.3.3 Suppose samples $\boldsymbol{X}_A$ and $\boldsymbol{X}_B$ are generated from the same probability function of the same statistical model.

Sufficiency principle —

Sample $\boldsymbol{X}_A$ $\rightarrow$ sufficient statistic $T(\boldsymbol{X}_A) = t_A$ $\rightarrow$ inference A
Sample $\boldsymbol{X}_B$ $\rightarrow$ sufficient statistic $T(\boldsymbol{X}_B) = t_B$ $\rightarrow$ inference B
If $t_A = t_B$, then *inference A must be the same as inference B*.

4.3.4 **Definitions.** The *likelihood* and *loglikelihood* functions of θ, given $\boldsymbol{X}$, are defined by

$$\ell_{\boldsymbol{X}}(\theta) \propto f(\boldsymbol{X}|\theta) \text{ (regarded as function of } \theta) \quad \text{and} \quad S_{\boldsymbol{X}}(\theta) = \ln \ell_{\boldsymbol{X}}(\theta), \quad \text{respectively.}$$

The likelihood or loglikelihood function summarises all information available in $\boldsymbol{X}$ which is relevant to θ. It measures the "likelihood" of each $\theta \in \Theta$ being the true θ that generates $\boldsymbol{X}$.

Common special case —

$\boldsymbol{X} = (X_1, \ldots, X_n)$ independent with $X_i \sim p_i(x|\theta)$:

$$\ell_{\boldsymbol{X}}(\theta) = \prod_{i=1}^{n} \ell_{X_i}(\theta) \propto \prod_{i=1}^{n} p_i(X_i \mid \theta) = f(\boldsymbol{X}|\theta), \qquad S_{\boldsymbol{X}}(\theta) = \sum_{i=1}^{n} S_{X_i}(\theta).$$

4.3.5 $\ell_{\boldsymbol{X}}$ and $S_{\boldsymbol{X}}$ vary from sample to sample and are **random**.

4.3.6 **Theorem.** The following three statements are equivalent:

(i) Statistic $T = T(\boldsymbol{X})$ is sufficient for θ.

(ii) For any samples $\boldsymbol{X}, \boldsymbol{X}'$, $\quad T(\boldsymbol{X}) = T(\boldsymbol{X}') \Rightarrow \ell_{\boldsymbol{X}}(\theta) \propto \ell_{\boldsymbol{X}'}(\theta)$.

(iii) [*Factorization Criterion*] There exists a function $g(\cdot)$ such that for any sample $\boldsymbol{X}$, $\ell_{\boldsymbol{X}}(\theta) \propto g\big(T(\boldsymbol{X}), \theta\big)$.

. . . .

Proof of "(i) ⇒ (ii)" — [Assume $\boldsymbol{X}$ is discrete. A more general proof is outlined in Appendix A4.1.]

Let $f_\theta^{(T)}$ be the probability function of T. Assume T is sufficient. Then the conditional probability function of $\boldsymbol{X}$ given $T(\boldsymbol{X}) = t$ does not depend on θ and has an expression $\psi(\boldsymbol{x}, t)$ free of θ, for $\boldsymbol{x} \in \mathcal{S}_t \triangleq \big\{\boldsymbol{x}' \in \mathcal{S} : T(\boldsymbol{x}') = t\big\}$. Note that

$$\psi(\boldsymbol{x}, t) = \frac{f(\boldsymbol{x}|\theta)}{f_\theta^{(T)}(t)}, \quad \boldsymbol{x} \in \mathcal{S}_t.$$

For any samples $\boldsymbol{X}, \boldsymbol{X}'$ with $T(\boldsymbol{X}) = T(\boldsymbol{X}') = t$, we have

$$\frac{\ell_{\boldsymbol{X}}(\theta)}{\ell_{\boldsymbol{X}'}(\theta)} \propto \frac{f(\boldsymbol{X}|\theta)}{f(\boldsymbol{X}'|\theta)} = \frac{\psi(\boldsymbol{X}, t) f_\theta^{(T)}(t)}{\psi(\boldsymbol{X}', t) f_\theta^{(T)}(t)} = \frac{\psi(\boldsymbol{X}, t)}{\psi(\boldsymbol{X}', t)},$$

which does not depend on θ. Hence, $\ell_{\boldsymbol{X}}(\theta) \propto \ell_{\boldsymbol{X}'}(\theta)$.

Proof of "(ii) ⇒ (iii)" —

For any $t \in \big\{T(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{S}\big\}$, there exists $\tilde{\boldsymbol{x}}(t) \in \mathcal{S}$ such that $T(\tilde{\boldsymbol{x}}(t)) = t$, which enables us to define $g(t, \theta) \triangleq \ell_{\tilde{\boldsymbol{x}}(t)}(\theta)$. Thus, for any sample $\boldsymbol{X}$, we have $T(\boldsymbol{X}) = T\big(\tilde{\boldsymbol{x}}(T(\boldsymbol{X}))\big)$ so that, using (ii),

$$\ell_{\boldsymbol{X}}(\theta) \propto \ell_{\tilde{\boldsymbol{x}}(T(\boldsymbol{X}))}(\theta) = g\big(T(\boldsymbol{X}), \theta\big).$$

Proof of "(iii) ⇒ (i)" — [Assume $\boldsymbol{X}$ is discrete. A more general proof is outlined in Appendix A4.1.]

By (iii), the probability function of $\boldsymbol{X}$ has the form $f(\boldsymbol{x}|\theta) = g(T(\boldsymbol{x}), \theta)h(\boldsymbol{x})$, for some positive function $h(\cdot)$. The conditional probability function of $\boldsymbol{X}$ given $T(\boldsymbol{X}) = t$ is

$$f_\theta^{(\boldsymbol{X}|T)}(\boldsymbol{x}|t) = \frac{f(\boldsymbol{x}|\theta)}{f_\theta^{(T)}(t)} = \frac{g(t, \theta)\, h(\boldsymbol{x})}{\displaystyle\sum_{\boldsymbol{y} \in \mathcal{S}_t} g(T(\boldsymbol{y}), \theta)\, h(\boldsymbol{y})} = \frac{h(\boldsymbol{x})}{\displaystyle\sum_{\boldsymbol{y} \in \mathcal{S}_t} h(\boldsymbol{y})}, \quad \boldsymbol{x} \in \mathcal{S}_t,$$

which does not depend on θ. ∎

4.3.7 Examples.

(i) Given exponential family indexed by natural parameter $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k) \in \Pi$, i.e. $f(\boldsymbol{x}|\boldsymbol{\pi}) = C(\boldsymbol{\pi})h(\boldsymbol{x}) \exp\big\{\sum_{j=1}^k \pi_j t_j(\boldsymbol{x})\big\}\mathbf{1}\{\boldsymbol{x} \in \mathcal{S}\}$,

$$\ell_{\boldsymbol{X}}(\boldsymbol{\pi}) \propto C(\boldsymbol{\pi}) \exp\Big\{\sum_{j=1}^k \pi_j t_j(\boldsymbol{X})\Big\} = g(T(\boldsymbol{X}), \boldsymbol{\pi}),$$

where $T(\boldsymbol{X}) = (t_1(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X}))$ and $g(\boldsymbol{t}, \boldsymbol{\pi}) = C(\boldsymbol{\pi}) e^{\sum_{j=1}^k \pi_j t_j}$ for $\boldsymbol{t} = (t_1, \ldots, t_k)$

$\longrightarrow$ natural statistic $T(\boldsymbol{X})$ is sufficient for natural parameter $\boldsymbol{\pi}$.

(ii) $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid from $U[0, \theta]$:

$$\ell_{\boldsymbol{X}}(\theta) \propto \prod_{i=1}^n \theta^{-1} \mathbf{1}\{0 \le X_i \le \theta\} \propto \theta^{-n} \mathbf{1}\{\max_i X_i \le \theta\} = g(T(\boldsymbol{X}), \theta),$$

where $T(\boldsymbol{X}) = \max_i X_i$ and $g(t, \theta) = \theta^{-n} \mathbf{1}\{t \le \theta\}$

$\longrightarrow$ $T(\boldsymbol{X}) = \max_i X_i$ is sufficient for θ.

4.3.8 A sufficient statistic is NOT unique in general: e.g.

- raw dataset $\boldsymbol{X}$ is always sufficient,
- if a sufficient statistic is a function of another statistic V, then V is sufficient.

§4.4 Minimal sufficiency

4.4.1 Among plausible choices of sufficient statistics, we wish to find the "most economical" function of $\boldsymbol{X}$ that remains sufficient for θ, i.e. a *minimal sufficient statistic*.

Definition. $T(\boldsymbol{X})$ is *minimal sufficient* for θ if it is sufficient and is a function of every other sufficient statistic.

4.4.2 **Theorem.** $T(\boldsymbol{X})$ is *minimal sufficient* for θ if and only if

$$\text{for any samples } \boldsymbol{X}, \boldsymbol{X}', \quad T(\boldsymbol{X}) = T(\boldsymbol{X}') \;\Leftrightarrow\; \ell_{\boldsymbol{X}}(\theta) \propto \ell_{\boldsymbol{X}'}(\theta).$$

. . . .

Proof:

(i) For the "if" part —

Theorem §4.3.6 implies that $T(\boldsymbol{X})$ is sufficient. Suppose that $V(\boldsymbol{X})$ is another sufficient statistic. Fix two samples $\boldsymbol{X}$ and $\boldsymbol{X}'$ with $V(\boldsymbol{X}) = V(\boldsymbol{X}')$. Then we have, by Theorem §4.3.6 again, $\ell_{\boldsymbol{X}}(\theta) \propto \ell_{\boldsymbol{X}'}(\theta)$. Hence $T(\boldsymbol{X}) = T(\boldsymbol{X}')$ also, so that T is a function of V. This proves minimality of T.

(ii) For the "only if" part —

Suppose that T is minimal sufficient. We need only prove that for any $\boldsymbol{X}, \boldsymbol{X}'$,

$$\ell_{\boldsymbol{X}}(\theta) \propto \ell_{\boldsymbol{X}'}(\theta) \;\Rightarrow\; T(\boldsymbol{X}) = T(\boldsymbol{X}'),$$

since the converse follows immediately from Theorem §4.3.6.

Consider samples $\boldsymbol{X}, \boldsymbol{X}'$ satisfying $\ell_{\boldsymbol{X}}(\theta) \propto \ell_{\boldsymbol{X}'}(\theta)$. Define statistic

$$U(\boldsymbol{x}) = \begin{cases} T(\boldsymbol{X}') & \textit{if } T(\boldsymbol{x}) = T(\boldsymbol{X}), \\ T(\boldsymbol{x}) & \textit{otherwise.} \end{cases}$$

For any $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ with $U(\boldsymbol{X}_1) = U(\boldsymbol{X}_2)$, we have either

(a) neither $T(\boldsymbol{X}_1)$ nor $T(\boldsymbol{X}_2)$ equals $T(\boldsymbol{X})$; or

(b) $T(\boldsymbol{X}_1) = T(\boldsymbol{X}_2) = T(\boldsymbol{X})$; or

(c) one and only one of $\{T(\boldsymbol{X}_1), T(\boldsymbol{X}_2)\}$ equals $T(\boldsymbol{X})$.

Both cases (a) and (b) imply that $T(\boldsymbol{X}_1) = T(\boldsymbol{X}_2)$ and hence $\ell_{\boldsymbol{X}_1}(\theta) \propto \ell_{\boldsymbol{X}_2}(\theta)$ by sufficiency of T and Theorem §4.3.6. For case (c), assume w.l.o.g. $T(\boldsymbol{X}_1) = T(\boldsymbol{X})$. Then $U(\boldsymbol{X}_1) = T(\boldsymbol{X}') = U(\boldsymbol{X}_2) = T(\boldsymbol{X}_2)$. Sufficiency of T, Theorem §4.3.6 and the assumptions on $\boldsymbol{X}$, $\boldsymbol{X}'$ imply that $\ell_{\boldsymbol{X}_1}(\theta) \propto \ell_{\boldsymbol{X}}(\theta) \propto \ell_{\boldsymbol{X}'}(\theta) \propto \ell_{\boldsymbol{X}_2}(\theta)$. Thus we conclude that U is sufficient by Theorem §4.3.6, so that T is a function of U by minimality. Since $U(\boldsymbol{X}) = T(\boldsymbol{X}') = U(\boldsymbol{X}')$, we must have $T(\boldsymbol{X}) = T(\boldsymbol{X}')$. This completes the proof. ▌

4.4.3 *Likelihood principle* —

Sample $\boldsymbol{X}_A$ $\rightarrow$ likelihood function $\ell_{\boldsymbol{X}_A}$ $\rightarrow$ inference A
Sample $\boldsymbol{X}_B$ $\rightarrow$ likelihood function $\ell_{\boldsymbol{X}_B}$ $\rightarrow$ inference B
If $\ell_{\boldsymbol{X}_A}(\theta) \propto \ell_{\boldsymbol{X}_B}(\theta)$, then *inference A must be the same as inference B*.

[Note: Samples $\boldsymbol{X}_A, \boldsymbol{X}_B$ may arise from different statistical models sharing the same parameter space Θ and same true $\theta \in \Theta$. Clearly, likelihood principle $\Rightarrow$ sufficiency principle.]

4.4.4 **Example.** $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid from $U[0, \theta]$: $\ell_{\boldsymbol{X}}(\theta) \propto \theta^{-n}\mathbf{1}\{\max_i X_i \le \theta\}$.

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{X}' = (X_1', \ldots, X_n')$ be two iid samples drawn from $U[0, \theta]$. Assume without loss of generality that $\max_i X_i \le \max_i X_i'$. To examine proportionality between $\ell_{\boldsymbol{X}}$ and $\ell_{\boldsymbol{X}'}$, it suffices to consider the subset of Θ on which $\ell_{\boldsymbol{X}}$ and $\ell_{\boldsymbol{X}'}$ are not both 0, i.e. on $\theta \ge \max_i X_i$. Consider

$$\frac{\ell_{\boldsymbol{X}}(\theta)}{\ell_{\boldsymbol{X}'}(\theta)} \propto \frac{\mathbf{1}\{\max_i X_i \le \theta\}}{\mathbf{1}\{\max_i X_i' \le \theta\}} = \begin{cases} 1/0 = \infty, & \max_i X_i \le \theta < \max_i X_i', \\ 1/1 = 1, & \theta \ge \max_i X_i', \end{cases}$$

which does not depend on θ if and only if $\max_i X_i = \max_i X_i'$.

Thus $T(\boldsymbol{X}) = \max_{1 \le i \le n}\{X_i\}$ *is minimal sufficient for θ.*

4.4.5 Theorem. *(Exponential family)* $\boldsymbol{X} \sim f(\boldsymbol{x}|\boldsymbol{\pi}) \propto h(\boldsymbol{x}) \exp\left\{\sum_{j=1}^k \pi_j t_j(\boldsymbol{x})\right\}$.

Assume that the natural parameter space Π is **not** contained in an *affine hyperplane* of the form $\left\{\boldsymbol{\pi} : \sum_{j=1}^k c_j \pi_j = b\right\} \subset \mathbb{R}^k$, where $(c_1, \ldots, c_k) \in \mathbb{R}^k$ is a non-zero vector. Then

$T(\boldsymbol{X}) = \big(t_1(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X})\big)$ *is minimal sufficient for* $\boldsymbol{\pi}$.

. . . .

Proof: *Note that for any samples* $\boldsymbol{X}, \boldsymbol{X}'$,

$$\ell_{\boldsymbol{X}}(\boldsymbol{\pi}) \propto \ell_{\boldsymbol{X}'}(\boldsymbol{\pi}) \quad \Leftrightarrow \quad \exp\left\{\sum_{j=1}^k \pi_j t_j(\boldsymbol{X})\right\} \propto \exp\left\{\sum_{j=1}^k \pi_j t_j(\boldsymbol{X}')\right\}$$

$$\Leftrightarrow \quad \exp\left\{\sum_{j=1}^k \pi_j \big(t_j(\boldsymbol{X}) - t_j(\boldsymbol{X}')\big)\right\} \textit{ does not depend on } \boldsymbol{\pi} \quad \Leftrightarrow \quad t_j(\boldsymbol{X}) = t_j(\boldsymbol{X}') \ \forall j.$$

To prove the last equivalence, the "$\Leftarrow$" part is trivial. For the "$\Rightarrow$" part, assume that there exists some function $b(\boldsymbol{X}, \boldsymbol{X}')$*, independent of* $\boldsymbol{\pi}$*, such that* $\sum_{j=1}^k \pi_j \big(t_j(\boldsymbol{X}) - t_j(\boldsymbol{X}')\big) = b(\boldsymbol{X}, \boldsymbol{X}') \ \forall \boldsymbol{\pi} \in \Pi$*. Since* Π *is not contained in any affine hyperplane, we must have* $t_j(\boldsymbol{X}) - t_j(\boldsymbol{X}') = 0 \ \forall j$*. Theorem §4.4.5 then follows from the above and Theorem §4.4.2.* ∎

4.4.6 Examples.

(i) $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid from $N(\mu, \sigma^2) \longrightarrow$ natural parameter $\boldsymbol{\pi} = (\pi_1, \pi_2) = \left(-\dfrac{1}{2\sigma^2}, \dfrac{\mu}{\sigma^2}\right)$
$\Pi = (-\infty, 0) \times (-\infty, \infty) \not\subset$ affine hyperplane (i.e. a straight line) in $\mathbb{R}^2$.
Thus $T(\boldsymbol{X}) = \left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i\right)$ *is minimal sufficient for* (μ, σ^2), or equivalently,
$\left(\dfrac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}, \bar{X}\right)$ *is minimal sufficient for* (μ, σ^2)*, etc.*

(ii) $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid from $N(1+\sigma^2, \sigma^2) \longrightarrow$ natural parameter $(\pi_1, \pi_2) = \left(-\dfrac{1}{2\sigma^2}, 1+\dfrac{1}{\sigma^2}\right)$
$\Pi \subset \{(\pi_1, \pi_2) : 2\pi_1 + \pi_2 = 1\}$, an affine hyperplane in $\mathbb{R}^2$.
In fact, we have in this case,

$$f(\boldsymbol{x}|\boldsymbol{\pi}) \propto h(\boldsymbol{x}) \exp\left\{\pi_1 \sum_i x_i^2 + (1 - 2\pi_1) \sum_i x_i\right\} = \left\{h(\boldsymbol{x})\, e^{\sum_i x_i}\right\} e^{\pi_1 \sum_i x_i (x_i - 2)},$$

which has exponential family form with natural parameter space $(-\infty, 0) \not\subset$ affine hyperplane (i.e. a single point) in $\mathbb{R}$. Thus

$T(\boldsymbol{X}) = \sum_{i=1}^n X_i \big(X_i - 2\big)$ *is minimal sufficient for* σ^2.

(iii) $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid from $N(\mu, \mu^2) \longrightarrow$ natural parameter $(\pi_1, \pi_2) = \Big(-\dfrac{1}{2\mu^2}, \dfrac{1}{\mu} \Big)$

$\Pi = \{(\pi_1, \pi_2) : 2\pi_1 + \pi_2^2 = 0\} \setminus \{(0,0)\} \not\subset$ affine hyperplane (i.e. a straight line) in $\mathbb{R}^2$. Thus

$$T(\boldsymbol{X}) = \big(\textstyle\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i \big) \textit{ is minimal sufficient for } \mu.$$

§4.5 Completeness

4.5.1 Sample data: $\boldsymbol{X} \sim$ probability function $f(\boldsymbol{x}|\theta)$

Definition. A sufficient statistic $T = T(\boldsymbol{X})$ is *complete* for θ if and only if, for any real function $g(\cdot)$ not depending on θ,

$$\mathbb{E}_\theta\big[g(T(\boldsymbol{X}))\big] = 0 \ \ \forall\,\theta \quad \Rightarrow \quad \mathbb{P}_\theta\big\{g(T(\boldsymbol{X})) = 0\big\} = 1 \ \ \forall\,\theta.$$

The latter condition says that $g(\cdot)$ is a zero function *almost surely* under any θ.

4.5.2 **Examples.**

1. $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid from $U[0, \theta]$

   $f(\boldsymbol{X}|\theta) = \theta^{-n}\mathbf{1}\{\max_i X_i \le \theta\}\mathbf{1}\{\min_i X_i \ge 0\}$

   We have shown in Example §4.4.4 that $T(\boldsymbol{X}) = \max_i X_i$ is minimal sufficient for θ. Note that $T(\boldsymbol{X})$ has pdf

   $$f_T(t|\theta) = \frac{d}{dt}\mathbb{P}_\theta(T(\boldsymbol{X}) \le t) = \frac{d}{dt}\left(\frac{t}{\theta}\right)^n = n\theta^{-n}t^{n-1}, \ \ t \in [0, \theta].$$

   For any function g such that $\mathbb{E}_\theta[g(T(\boldsymbol{X}))] = 0\ \forall\,\theta$, we have

   $$n\theta^{-n}\int_0^\theta g(t)\, t^{n-1} dt = 0\ \forall\,\theta \ \ \Rightarrow \ \ g(t) = 0 \text{ almost everywhere in } [0, \infty).$$

   Thus $\mathbb{P}_\theta\big(g(T) = 0\big) = 1\ \forall\,\theta$, and hence $T(\boldsymbol{X}) = \max_i X_i$ is complete for θ.

   [A more formal proof of the above is given in Appendix A4.2.]

2. $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid from $U[\theta, \theta + 1]$

   $f(\boldsymbol{X}|\theta) = \mathbf{1}\big\{\max_i X_i \le \theta + 1\big\}\mathbf{1}\big\{\min_i X_i \ge \theta\big\} = \mathbf{1}\big\{\max_i X_i - 1 \le \theta \le \min_i X_i\big\}$

   Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{X}' = (X_1', \ldots, X_n')$ be two iid samples drawn from $U[\theta, \theta + 1]$.

If $\big[\max_i X_i - 1,\, \min_i X_i\big] \backslash \big[\max_i X'_i - 1,\, \min_i X'_i\big] \neq \emptyset$, then

$$\ell_{\boldsymbol{X}}(\theta)/\ell_{\boldsymbol{X}'}(\theta) \propto \mathbf{1}\big\{\max_i X_i - 1 \le \theta \le \min_i X_i\big\} \big/ \mathbf{1}\big\{\max_i X'_i - 1 \le \theta \le \min_i X'_i\big\}$$
$$\begin{cases} = 1/0 = \infty, & \theta \in \big[\max_i X_i - 1,\, \min_i X_i\big] \backslash \big[\max_i X'_i - 1,\, \min_i X'_i\big], \\ \le 1, & \theta \in \big[\max_i X'_i - 1,\, \min_i X'_i\big], \end{cases}$$

which depends on θ. The same conclusion holds if $\big[\max_i X'_i - 1,\, \min_i X'_i\big] \backslash \big[\max_i X_i - 1,\, \min_i X_i\big] \neq \emptyset$.

It follows that $\ell_{\boldsymbol{X}}(\theta) \propto \ell_{\boldsymbol{X}'}(\theta)$ if and only if the intervals $\big[\max_i X_i - 1,\, \min_i X_i\big]$ and $\big[\max_i X'_i - 1,\, \min_i X'_i\big]$ are identical, i.e. $\big(\min_i X_i, \max_i X_i\big) = \big(\min_i X'_i, \max_i X'_i\big)$. Thus $T(\boldsymbol{X}) = \big(\min_i X_i,\, \max_i X_i\big)$ is minimal sufficient for θ.

Note that

$\min_i X_i$ has pdf $f_{min}(t|\theta) = n(\theta + 1 - t)^{n-1}$, for $t \in [\theta, \theta + 1]$,
$\max_i X_i$ has pdf $f_{max}(t|\theta) = n(t - \theta)^{n-1}$, for $t \in [\theta, \theta + 1]$.

Consider a non-zero function $g(t_1, t_2) = t_1 - t_2 + \dfrac{n-1}{n+1}$. Then

$$\begin{aligned} \mathbb{E}_\theta\big[g(T(\boldsymbol{X}))\big] &= \mathbb{E}_\theta[\min_i X_i] - \mathbb{E}_\theta[\max_i X_i] + \frac{n-1}{n+1} \\ &= \int_\theta^{\theta+1} t f_{min}(t|\theta)\, dt - \int_\theta^{\theta+1} t f_{max}(t|\theta)\, dt + \frac{n-1}{n+1} \\ &= n \int_\theta^{\theta+1} t\big\{(\theta + 1 - t)^{n-1} - (t - \theta)^{n-1}\big\}\, dt + \frac{n-1}{n+1} \\ &= n \int_0^1 (u + \theta)\big\{(1 - u)^{n-1} - u^{n-1}\big\}\, du + \frac{n-1}{n+1} = 0 \ \ \forall\, \theta. \end{aligned}$$

Thus $T(\boldsymbol{X}) = \big(\min_i X_i,\, \max_i X_i\big)$ is <u>not</u> complete for θ.

4.5.3 **Theorem.** *(Lehmann–Scheffé)*

If $T(\boldsymbol{X})$ is complete sufficient for θ, then $T(\boldsymbol{X})$ is minimal sufficient.

. . . .

Proof:

Let $U(\boldsymbol{X})$ be a minimal sufficient statistic for θ. Then $U(\boldsymbol{X}) = g_1(T(\boldsymbol{X}))$ for some function $g_1(\cdot)$ by minimality of U. Define $g(t) = t - \mathbb{E}\big[T(\boldsymbol{X})\big|U(\boldsymbol{X}) = g_1(t)\big]$, which does not depend on θ by sufficiency of $U(\boldsymbol{X})$. Then we have

$$\mathbb{E}_\theta\big[\, g(T(\boldsymbol{X}))\big] = \mathbb{E}_\theta[T(\boldsymbol{X})] - \mathbb{E}_\theta\Big[\,\mathbb{E}\big[T(\boldsymbol{X})\big|U(\boldsymbol{X})\big]\,\Big] = \mathbb{E}_\theta[T(\boldsymbol{X})] - \mathbb{E}_\theta[T(\boldsymbol{X})] = 0 \ \textit{for all } \theta.$$

Hence $g(T(\boldsymbol{X})) = 0$ with probability 1 by completeness of T. This implies that with probability 1,

$$T(\boldsymbol{X}) = \mathbb{E}\big[T(\boldsymbol{X})\big|U(\boldsymbol{X})\big], \quad \textit{which is a function of } U(\boldsymbol{X}).$$

Since $U(\boldsymbol{X})$ is minimal sufficient, it is a function of any sufficient statistic, and so is $T(\boldsymbol{X})$. Thus $T(\boldsymbol{X})$ is minimal sufficient. ▮

4.5.4 **Theorem.** *(Exponential family)* $\boldsymbol{X} \sim f(\boldsymbol{x}|\boldsymbol{\pi}) \propto h(\boldsymbol{x}) \exp\big\{ \sum_{j=1}^{k} \pi_j t_j(\boldsymbol{x})\big\}$.

If the natural parameter space Π contains an open rectangle, i.e. a nonempty set of the form $(a_1, b_1) \times \cdots \times (a_k, b_k) \subset \mathbb{R}^k$, then the natural statistic $T(\boldsymbol{X}) = \big(t_1(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X})\big)$ is *complete* for the natural parameter $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$.

. . . .

Proof:

Recall that $T(\boldsymbol{X})$ has probability function also of exponential family form $C(\boldsymbol{\pi})h^*(\boldsymbol{t}) \exp\Big(\sum_{j=1}^{k} \pi_j t_j\Big)$. Consider a function $g(\cdot)$ such that

$$\mathbb{E}_{\boldsymbol{\pi}}\big[\, g(T(\boldsymbol{X}))\big] = C(\boldsymbol{\pi}) \int g(\boldsymbol{t})h^*(\boldsymbol{t}) \exp\Big(\sum_{j=1}^{k} \pi_j t_j\Big) d\boldsymbol{t} = 0, \quad \textit{for all } \boldsymbol{\pi} \in \Pi.$$

Thus the k-dimensional Laplace transform of the function $g(\boldsymbol{t})h^*(\boldsymbol{t})$ vanishes on Π. Uniqueness of inverse of Laplace transform which exists in an open rectangle implies that $g(\boldsymbol{t})h^*(\boldsymbol{t}) = 0$ for all $\boldsymbol{t}$. Since $h^*(\boldsymbol{t}) > 0$ for $\boldsymbol{t}$ in the support of $T(\boldsymbol{X})$, we have $g(T(\boldsymbol{X})) = 0$ with probability 1. ▮

4.5.5 **Examples.**

(i) $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid from $N(\mu, \sigma^2) \longrightarrow$ natural parameter $\boldsymbol{\pi} = (\pi_1, \pi_2) = \Big(-\dfrac{1}{2\sigma^2}, \dfrac{\mu}{\sigma^2}\Big)$
$\Pi = (-\infty, 0) \times (-\infty, \infty)$ contains an open rectangle in $\mathbb{R}^2$.
Thus $T(\boldsymbol{X}) = \big(\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i\big)$ *is complete sufficient for* (μ, σ^2).

(ii) $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid from $N(1 + \sigma^2, \sigma^2) \longrightarrow$ natural parameter $\pi_1 = -\dfrac{1}{2\sigma^2}$
$\Pi = (-\infty, 0)$ contains an open rectangle (interval) in $\mathbb{R}$.
Thus $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i\big(X_i - 2\big)$ *is complete sufficient for* σ^2.

(iii) $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid from $N(\mu, \mu^2) \longrightarrow$ natural parameter $(\pi_1, \pi_2) = \Big(-\dfrac{1}{2\mu^2}, \dfrac{1}{\mu}\Big)$
$\Pi = \big\{(\pi_1, \pi_2) : 2\pi_1 + \pi_2^2 = 0\big\} \setminus \big\{(0, 0)\big\}$ does not contain an open rectangle in $\mathbb{R}^2$.

In fact, if we define $g(t_1, t_2) = (n+1)t_1 - 2t_2^2$, we have

$$\mathbb{E}_\mu\Big[\, g\big(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i\big)\Big] = (n+1)(2n\mu^2) - 2\left\{n\mu^2 + (n\mu)^2\right\} = 0,$$

implying that $T(\boldsymbol{X}) = \big(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i\big)$ is <u>not</u> complete for μ.

§4.6 Exercise: motivating example §4.1

(a) Construct a minimal sufficient statistic for θ.

(b) Is the minimal sufficient statistic constructed in (a) complete for θ?

Appendix

A4.1 A more formal treatment of §4.3.6...

Proof of "(i) ⇒ (ii)" —

Assume T is sufficient. For any $\boldsymbol{X}$-measurable set A, define $\psi_A^*(t) = \mathbb{P}(\boldsymbol{X} \in A \mid T = t)$, which does not depend on θ. It can be shown that there exists $\boldsymbol{X}^*$ with probability function f^* defined on the sample space of $\boldsymbol{X}$, such that for all $\boldsymbol{X}$-measurable set A,

$$\mathbb{P}(\boldsymbol{X}^* \in A \mid T(\boldsymbol{X}^*) = t) = \psi_A^*(t) \ \ \forall\, t,$$

and that

$$\mathbb{P}(\boldsymbol{X}^* \in A) = 0 \quad \textit{iff} \quad \mathbb{P}(\boldsymbol{X} \in A \mid \theta) = 0 \,\forall\, \theta.$$

Let $T^* = T(\boldsymbol{X}^*)$. Denote by $f_T(t|\theta)$ and $f_{T^*}(t)$ the probability functions of T and T^*, respectively. By definition of conditional probability, we have, for any $\boldsymbol{X}$-measurable set A,

$$\begin{aligned}\mathbb{P}_\theta(\boldsymbol{X} \in A) = \int \psi_A^*(t)\, f_T(t|\theta)\, dt &= \int \psi_A^*(t) \left\{\frac{f_T(t|\theta)}{f_{T^*}(t)}\right\} f_{T^*}(t)\, d\boldsymbol{x} \\ &= \mathbb{E}\left[\mathbf{1}\{\boldsymbol{X}^* \in A\} \left\{\frac{f_T(T^*|\theta)}{f_{T^*}(T^*)}\right\}\right] = \int_A \left\{\frac{f_T(T(\boldsymbol{x})|\theta)}{f_{T^*}(T(\boldsymbol{x}))}\right\} f^*(\boldsymbol{x})\, d\boldsymbol{x},\end{aligned}$$

so that $\boldsymbol{X}$ has probability function $f(\boldsymbol{x}|\theta) = f_T(T(\boldsymbol{x})|\theta) f^*(\boldsymbol{x})/f_{T^*}(T(\boldsymbol{x}))$. For $\boldsymbol{X}, \boldsymbol{X}' \sim f(\cdot|\theta)$ with $T(\boldsymbol{X}) = T(\boldsymbol{X}')$, we have

$$\ell_{\boldsymbol{X}}(\theta) \propto f_T(T(\boldsymbol{X})|\theta) = f_T(T(\boldsymbol{X}')|\theta) \propto \ell_{\boldsymbol{X}'}(\theta).$$

Proof of "(iii) ⇒ (i)" —

By (iii), the probability function of $\boldsymbol{X}$ has the form $f(\boldsymbol{x}|\theta) = g(T(\boldsymbol{x}), \theta) h(\boldsymbol{x})$, for some positive function $h(\cdot)$. Denote by $\boldsymbol{X}^*$ a sample with probability function $f^*(\boldsymbol{x}) = g^*(T(\boldsymbol{x}))\, h(\boldsymbol{x})$, which satisfies, for any $\boldsymbol{X}$-measurable set A,

$$\mathbb{P}(\boldsymbol{X}^* \in A) = 0 \quad \textit{iff} \quad \mathbb{P}_\theta(\boldsymbol{X} \in A) = 0 \,\forall\, \theta.$$

Let $T^* = T(\boldsymbol{X}^*)$. Define $\psi_A^*(t) = \mathbb{P}\left(\boldsymbol{X}^* \in A \mid T^* = t\right)$ for any T-measurable set B and $\boldsymbol{X}$-measurable set A, and consider

$$\begin{aligned}&\mathbb{P}_\theta(\boldsymbol{X} \in A,\, T \in B) \\ &\quad = \int \mathbf{1}\{\boldsymbol{x} \in A,\, T(\boldsymbol{x}) \in B\}\, f(\boldsymbol{x}|\theta)\, d\boldsymbol{x} = \int \mathbf{1}\{\boldsymbol{x} \in A,\, T(\boldsymbol{x}) \in B\} \left\{\frac{g(T(\boldsymbol{x}), \theta)}{g^*(T(\boldsymbol{x}))}\right\} f^*(\boldsymbol{x})\, d\boldsymbol{x} \\ &\quad = \mathbb{E}\left[\mathbf{1}\{\boldsymbol{X}^* \in A,\, T^* \in B\} \left\{\frac{g(T^*, \theta)}{g^*(T^*)}\right\}\right] = \mathbb{E}\left[\mathbf{1}\{T^* \in B\} \left\{\frac{g(T^*, \theta)}{g^*(T^*)}\right\} \psi_A^*(T^*)\right] \\ &\quad = \int \mathbf{1}\{T(\boldsymbol{x}) \in B\} \psi_A^*(T(\boldsymbol{x})) \left\{\frac{g(T(\boldsymbol{x}), \theta)}{g^*(T(\boldsymbol{x}))}\right\} f^*(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\{\boldsymbol{x}:\, T(\boldsymbol{x}) \in B\}} \psi_A^*(T(\boldsymbol{x}))\, f(\boldsymbol{x}|\theta)\, d\boldsymbol{x}.\end{aligned}$$

By definition of the conditional probability $\psi_A(t|\theta) \equiv \mathbb{P}_\theta(\boldsymbol{X} \in A \mid T = t)$, *we have*

$$\mathbb{P}_\theta(\boldsymbol{X} \in A,\, T \in B) = \int_{\{\boldsymbol{x}\colon T(\boldsymbol{x}) \in B\}} \psi_A(T(\boldsymbol{x})|\theta)\, f(\boldsymbol{x}|\theta)\, d\boldsymbol{x}.$$

Uniqueness of conditional probability implies that

$$\mathbb{P}_\theta(\boldsymbol{X} \in A \mid T(\boldsymbol{X})) = \psi_A(T(\boldsymbol{X})|\,\theta) = \psi_A^*(T(\boldsymbol{X})),$$

which is independent of θ. *Thus* $T = T(\boldsymbol{X})$ *is sufficient for* θ. ∎

A4.2 A more formal treatment of a result used in Example 1 in §4.5.2:

$$n\theta^{-n}\!\int_0^\theta g(t)\, t^{n-1} dt = 0\; \forall\, \theta \;\;\Rightarrow\;\; \mathbb{P}_\theta\big(g(T) = 0\big) = 1\; \forall\, \theta.$$

. . . .

Outline of proof:

Note that

$$\int_{\theta_1}^{\theta_2} t^{n-1} g(t)\, dt = \int_0^{\theta_2} t^{n-1} g(t)\, dt - \int_0^{\theta_1} t^{n-1} g(t)\, dt = 0 \;\; \forall\, \theta_2 > \theta_1 > 0$$

$$\Rightarrow \quad \int_{\mathcal{B}} t^{n-1} g(t)\, dt = 0 \;\; \forall \text{ Borel-measurable } \mathcal{B} \subset (0, \infty).$$

For any $\epsilon > 0$,

$$\mathbb{P}_\theta\big(g(T) \geq \epsilon\big) = n\theta^{-n} \int_{(0,\theta) \cap \{g(t) \geq \epsilon\}} t^{n-1} dt \leq n\theta^{-n} \int_{\underbrace{(0,\theta)\, \cap\, \{g(t) \geq \epsilon\}}_{\text{Borel-measurable}}} t^{n-1} \left(\frac{g(t)}{\epsilon}\right) dt = 0.$$

Thus

$$\mathbb{P}_\theta\big(g(T) > 0\big) = \lim_{\epsilon \downarrow 0} \mathbb{P}_\theta\big(g(T) \geq \epsilon\big) = 0.$$

Similar arguments show that $\mathbb{P}_\theta\big(g(T) < 0\big) = 0$. *It then follows that* $\mathbb{P}_\theta\big(g(T) = 0\big) = 1 \;\; \forall\, \theta > 0$. ∎