Inference 2, 2023, lecture 4

Rolf Larsson

November 14, 2023

Today

Chap. 3.3. Inference Principles (continued):

- Sufficiency
- Minimal sufficiency

Sufficiency

Recall:

Theorem (3.5)

Let $\mathbf{X}$ *be a random variable with* $\mathbf{X} \sim \mathrm{P}_\theta$. *Let* $T(\mathbf{X})$ *be any statistic. Suppose that regularity conditions 1-3 hold. Then,*

$$I_{T(\mathbf{X})}(\theta) \leq I_{\mathbf{X}}(\theta),$$

where $I_{T(\mathbf{X})}$ *is computed with respect to the distribution of* $T(\mathbf{X})$

- Any statistic $T : \mathbf{x} \in \mathcal{X} \to T(\mathbf{x}) = t \in \mathcal{T}$ generates a partition of the sample space

$$\mathcal{X} = \bigcup_{T(\mathbf{x})=t} \mathcal{X}_t, \quad \mathcal{X}_t = \{\mathbf{x} : T(\mathbf{x}) = t\}.$$

- What characterizes those partitions that don't make us loose any information about the parameters?

Sufficiency

Example 1:

- A coin has probability p of coming up 'heads'.
- For $i = 1, 2$, let $X_i = 1$ if it comes up 'heads' in the ith throw, and 0 otherwise.
- Let $P(X_i = 1) = p$, let X_1 and X_2 be independent.
- Define the statistic $T(X_1, X_2) = X_1 + X_2$.

1. Write down the induced partition of the sample space.
2. Does this partition cause any loss of information regarding p?

Sufficiency

Definition (3.7)

A statistic T is said to be **sufficient** for the statistical model $\{\mathrm{P}_\theta : \theta \in \Theta\}$ of $\mathbf{X}$ if the conditional distribution of $\mathbf{X}$ given $T = t$ is independent of θ for all t.

- In short: For sufficient statistics T,

$$\mathbf{X} \sim \mathrm{P}_\theta, \quad T \sim \mathrm{P}_\theta^T, \quad \mathbf{X}|T \sim \mathrm{P}.$$

- Is the statistic of example 1 sufficient?

Sufficiency

Is there any simple criterion that shows that a statistic is sufficient?

Theorem (3.7)

The Factorization criterion: *Let $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$ be a statistical model with probability function $p(\cdot; \theta)$.*
A statistic T is sufficient for $\mathcal{P}$ if and only if there exist nonnegative functions $g(\cdot; \theta)$ and h such that

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

Sufficiency

Suppose we have an *i.i.d.* sample from any of the following distributions. Derive sufficient statistics for the parameters.

1. Exponential with intensity β
2. Uniform on the interval $[0, \theta]$
3. $N(\mu, \sigma^2)$

Sufficiency

Corollary (3.2)

a) *A statistic T is sufficient for $\theta \in \Theta$ if and only if*

$$L(\theta; \mathbf{x}) \propto g\{T(\mathbf{x}), \theta\} \quad \textit{for all } \mathbf{x}.$$

b) *Let T be a sufficient statistic and suppose that the Fisher information can be computed for $\{\mathrm{P}_\theta : \theta \in \Theta\}$ and for $\{\mathrm{P}_\theta^T : \theta \in \Theta\}$. Then,*

$$I_{\mathbf{X}}(\theta) = I_{T(\mathbf{X})}(\theta).$$

Corollary (3.3)

Suppose $\mathbf{X} = (X_1, ..., X_n)$ is a sample of i.i.d. random variables with distribution F. Then the order statistic $(X_{[1]}, ..., X_{[n]})$ is sufficient for F.

Minimal sufficiency

Example 2:

- Suppose $\mathbf{X} = (X_1, ..., X_n)$ where the X_i are independent and distributed as X.
- We have a sample $\mathbf{x} = (x_1, ..., x_n)$.
- X is multinomial with probabilities $P(X = j) = p_j$ for $j = 1, 2, 3$ and 0 otherwise.
- $p_1 + p_2 + p_3 = 1$.
- Let n_1, n_2, n_3 be the observed number of ones, twos and threes in the sample.
- $\mathrm{P}(\mathbf{x}; p_1, p_2, p_3) = p_1^{n_1} p_2^{n_2} p_3^{n_3} = g(T_1; p_1, p_2, p_3),\ T_1 = (n_1, n_2, n_3)$.
- $\mathrm{P}(\mathbf{x}; p_1, p_2) = p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n - n_1 - n_2} = g(T_2; p_1, p_2)$, $T_2 = (n_1, n_2)$.
- By the factorization theorem, T_1 and T_2 are both sufficient, but T_2 'condenses' the information of the sample the most.

Minimal sufficiency

Definition (3.8)

A sufficient statistic T is **minimal sufficient** if.f. it is a function of any other sufficient statistic.

Intuition:

- A sufficient statistic S is a function of the sample. It induces a partition of the sample space.
- A sufficient statistic $T = H(S)$ for some function H induces a partition of the sample space that may be even 'rougher'. (It might be that $S(\mathbf{x}_1) = s_1 \neq s_2 = S(\mathbf{x}_2)$ but $T(\mathbf{x}_1) = T(\mathbf{x}_2)$.)

Minimal sufficiency

Theorem (3.8)

Let T be a sufficient statistic for $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$.
Let $\mathcal{K}$ be the set of all pairs $(\mathbf{x}, \mathbf{y})$ for which there is a function $k(\mathbf{x}, \mathbf{y}) > 0$ (not dependent on θ) such that

$$L(\theta; \mathbf{x}) = k(\mathbf{x}, \mathbf{y}) L(\theta; \mathbf{y}) \quad \textit{for all } \theta \in \Theta.$$

If for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{K}$ the statistic T satisfies $T(\mathbf{x}) = T(\mathbf{y})$ then T is minimal sufficient.

In other words:
If $\frac{L(\theta;\mathbf{x})}{L(\theta;\mathbf{y})}$ is no function of θ if.f. $T(\mathbf{x}) = T(\mathbf{y})$, then T is minimal sufficient.

Minimal sufficiency

Recall:
If $\frac{L(\theta;\mathbf{x})}{L(\theta;\mathbf{y})}$ is no function of θ if.f. $T(\mathbf{x}) = T(\mathbf{y})$, then T is minimal sufficient.

Suppose we have an *i.i.d.* sample from any of the following distributions. Derive minimal sufficient statistics for the parameters.

1. Exponential with intensity β
2. Uniform on the interval $[0, \theta]$
3. $N(\mu, \sigma^2)$

News of today

- Sufficient statistic. (Conditioning on it, no information about the parameter remains.)
- The factorization criterion. (The parameters enter the likelihood only together with the sufficient statistic.)
- Minimal sufficient statistic. (No "rougher" partitioning of the sample space is possible.)
- A sufficient statistic that is constant on parts of the sample space with "proportional" likelihoods is minimal sufficient.