Sufficiency and Data Reduction

Andrew Nobel

September, 2024

How a Function Partitions its Domain

An Equivalence Relation

Setting. Let $f : \mathcal{X} \to \mathcal{Y}$ be a function. Define a relation $\overset{f}{\sim}$ on $\mathcal{X} \times \mathcal{X}$ by

$$x \overset{f}{\sim} x' \quad \text{if} \quad f(x) = f(x')$$

- Easy to see that $\overset{f}{\sim}$ is an equivalence relation
- Equivalence classes $C_f(y) = \{x \in \mathcal{X} : f(x) = y\}$ partition $\mathcal{X}$
- If f is 1:1 each class $C_f(y)$ is empty or contains a single point of $\mathcal{X}$

Idea: Function f groups points in $\mathcal{X}$, may result in loss of information

- If $C_f(y)$ contains a single point, we can recover x from $f(x) = y$
- If $C_f(y)$ has two or more points we can't recover x from $f(x) = y$

Coarser and Finer Functions

Fact: Let $f : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{X} \to \mathcal{Z}$ be known functions. The following are equivalent

- $g(x) = g(x')$ implies $f(x) = f(x')$
- $x \overset{g}{\sim} x'$ implies $x \overset{f}{\sim} x'$
- Each class $C_g(z)$ is contained in a single, unique class $C_f(y)$
- From the value z of $g(x)$ we can recover the value y of $f(x)$
- There is a function $h : \mathcal{Z} \to \mathcal{Y}$ such that $f(x) = h(g(x))$

Under the conditions above, g is finer than f, f is coarser than g

Sufficiency and Minimal Sufficiency

Orientation

- Family $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Theta\}$ of distributions on sample space $\mathcal{X}$
- Observation $X \in \mathcal{X}$ with $X \sim f(\cdot|\theta) \in \mathcal{P}$, where θ is unknown
- Goal: Inference about θ

Definition: A *statistic* is a function $T : \mathcal{X} \to \mathcal{Y}$ of the data that facilitates inference

- Data summary or reduction
- Point estimate or test statistic

Note: If $X \sim f(\cdot|\theta)$ then $T(X)$ is random, its distribution may depend on θ

Example

Observe $X = X_1, \ldots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$, make inferences about μ, σ.

Here $\mathcal{X} = \mathbb{R}^n$. Potential statistics

- $T(x) = x$
- $T(x) = (x_{(1)}, \ldots, x_{(n)})$ (order statistics)
- $T(x) = \max\{x_1, \ldots, x_n\}$
- $T(x) = x_1 + \cdots + x_n$
- $T(x) = x_1^2 + \cdots + x_n^2$
- $T(x) = (x_1, x_2)$
- $T(x) = 34$

NB: Value of statistics changes if we wish to estimate θ from $X_1, \ldots, X_n$ iid $\mathrm{U}(0, \theta)$

Sufficient Statistics

Question: Given family $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Theta\}$ of distributions on $\mathcal{X}$ what does a statistic $T : \mathcal{X} \to \mathcal{Y}$ tell us about θ?

Competing goals

- Simplifying and summarizing data. Favors courser statistics T
- Capturing information needed for estimation. Favors finer statistics T

Definition: A statistic $T : \mathcal{X} \to \mathcal{Y}$ is *sufficient* for $\mathcal{P}$ (or θ) if the conditional distribution of X given $T(X)$ does not depend on θ

Idea: Once $T(X)$ is known, remaining uncertainty in X is independent of θ

Two Statisticians

Suppose T sufficient for family $\mathcal{P}$ and $X \sim f(\cdot|\theta) \in \mathcal{P}$. Consider two statisticians.

- Amy: Observes X. Makes inference about θ based on X.
- Bob: Observes $T(X)$. Simulates $\tilde{X}$ from conditional distribution of X given $T(X)$. Makes inference about θ based on $\tilde{X}$.

Fact: For each $x \in \mathcal{X}$ and each $\theta \in \Theta$, $\mathbb{P}_\theta(X = x) = \mathbb{P}_\theta(\tilde{X} = x)$

Corollary: For any loss function $\ell(\theta, a)$ and every decision rule $d(x)$

$$R_A(\theta, d) \;=\; \mathbb{E}_\theta \ell(\theta, d(X)) \;=\; \mathbb{E}_\theta \ell(\theta, d(\tilde{X})) \;=\; R_B(\theta, d)$$

So Amy and Bob are in equivalent positions in regards to inference about θ

Criteria for Sufficiency

Factorization Theorem: A statistic $T : \mathcal{X} \to \mathcal{Y}$ is sufficient for $\mathcal{P}$ if and only if there exist functions $g : \mathcal{Y} \times \Theta \to \mathbb{R}$ and $h : \mathcal{X} \to \mathbb{R}$ such that

$$f(x|\theta) = g(T(x)|\theta)\, h(x) \text{ for each } x \in \mathcal{X} \text{ and } \theta \in \Theta$$

Definition: For $x \in \mathcal{X}$ and $\theta_1, \theta_2 \in \Theta$ define the likelihood ratio of θ_1 vs θ_2

$$\Lambda_x(\theta_1, \theta_2) = \frac{f(x|\theta_1)}{f(x|\theta_2)}$$

Proposition: A statistic $T : \mathcal{X} \to \mathcal{Y}$ is sufficient for $\mathcal{P}$ if and only if

$$T(x) = T(x') \;\to\; \Lambda_x(\theta_1, \theta_2) = \Lambda_{x'}(\theta_1, \theta_2) \text{ for all } \theta_1, \theta_2 \in \Theta$$

Examples

Ex. $X_1, \ldots, X_n$ iid $\sim f(\cdot|\theta) \in \mathcal{P}$. Then $T(x) = (x_{(1)}, \ldots, x_{(n)})$ sufficient for θ

Ex. $X_1, \ldots, X_n$ iid Bern(θ) with $0 < \theta < 1$. Then $T(x) = \sum_{i=1}^{n} x_i$ sufficient for θ

Ex. $X_1, \ldots, X_n$ iid $\mathcal{N}(\theta, \sigma^2)$ with $\theta \in \mathbb{R}$, σ^2 known. Then $T(x) = \overline{x}$ sufficient for θ

Ex. $X_1, \ldots, X_n$ iid U$(0, \theta)$ with $\theta > 0$. Then $T(x) = \max\{x_i\}$ sufficient for θ

Sufficiency and Exponential Families

Recall: Exponential family $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Theta\}$ has densities of the form

$$f(x|\theta) \;=\; c(\theta)\, h(x)\, \exp\left\{ \sum_{j=1}^{k} \eta_j(\theta)\, T_j(x) \right\}$$

Fact

(i) $T(x)$ is sufficient for the family $\mathcal{P}$

(ii) $\tilde{T}(x) = \sum_{i=1}^{n} T(x_i)$ is sufficient for the product family $\mathcal{P}^n$

Example: Normal Family

Consider $X_1, \ldots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma > 0$

- $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}$ is an EF with sufficient statistic $T(x) = (x, x^2)$
- $\mathcal{P}^n$ is an EF with sufficient statistic $\tilde{T}(x) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2\right)$
- By Fact, $\tilde{T}(x)$ is sufficient for $\mathcal{P}^n$
- There is a 1:1 relationship between $\tilde{T}(x)$ and sample mean/variance

$$\overline{x} = n^{-1} \sum_{i=1}^n x_i \qquad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x})^2$$

Upshot: Statistic $(\overline{x}, s^2)$ is sufficient for (μ, σ)

Minimal Sufficiency, Background

Fact: Let $T : \mathcal{X} \to \mathcal{Y}$ be sufficient for $\mathcal{P}$. If T is a function of another statistic $S : \mathcal{X} \to \mathcal{Z}$ then S is also sufficient for $\mathcal{P}$

Ex. Observe $X_1, \ldots, X_n$ iid $\sim \mathcal{N}(0, \sigma^2)$, estimate σ. Consider statistics

- $T_0(x) = x_1^2 + \cdots + x_n^2$ easily seen to be sufficient
- $T_1(x) = (x_1^2, \ldots, x_n^2)$, $T_2(x) = (x_1^2 + x_2^2, x_3^2 + \cdots + x_n^2)$, $T_3(x) = 2$

Note

- T_0 is a function of T_1, T_2, so these are also sufficient for σ
- T_0, T_1, T_2 not functions of T_3, easy to see T_3 not sufficient for σ

Minimal Sufficiency

Idea: Find statistic T giving greatest reduction of data while still being sufficient

Definition: A statistic $T : \mathcal{X} \to \mathcal{Y}$ is *minimal sufficient* for $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ if

(i) T is sufficient for $\mathcal{P}$

(ii) If S is any other sufficient statistic for $\mathcal{P}$ then T is a function of S

Conditions (i) and (ii) say that T is a coarsest function of the data that captures all the information needed for inference about θ

Fact: If T and S are minimal sufficient and surjective, then there are 1:1 functions F and G such that $T = F(S)$ and $S = G(T)$

Sufficient Condition for Minimal Sufficiency

Proposition: A statistic $T : \mathcal{X} \to \mathcal{Y}$ is minimal sufficient for $\mathcal{P}$ if and only if

$$T(x) = T(x') \;\leftrightarrow\; \frac{f(x|\theta_1)}{f(x|\theta_2)} = \frac{f(x'|\theta_1)}{f(x'|\theta_2)} \;\; \forall \theta_1, \theta_2$$

Corollary: A statistic $T : \mathcal{X} \to \mathcal{Y}$ is minimal sufficient for $\mathcal{P}$ if and only if

$$T(x) = T(x') \;\leftrightarrow\; \frac{f(x|\theta)}{f(x'|\theta)} \text{ constant in } \theta$$

Example: If $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma > 0$ then

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

is minimal sufficient for (μ, σ)