Statistics 210A Lecture 5 Notes

Daniel Raban

September 9, 2021

1 Minimal Sufficient and Complete Statistics

1.1 Recap: sufficient statistics

Last time, we talked about sufficient statistics. We said that $T(X)$ is **sufficient** for $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ if the distribution of $X \mid T(X)$ does not depend on θ. We encountered the **sufficiency principle**, which said that we should only attend to sufficient statistics T in our statistical analysis, rather than the whole data.

The factorization theorem says that if $\mathcal{P}$ has densities $p_\theta(x)$ with respect to μ, then $T(X)$ is sufficient iff there exist functions g_θ, h such that $p_\theta(x) = g_\theta(T(x))h(x)$. For exponential families, we have

$$p_\theta(x) = \underbrace{e^{\eta(\theta)^\top T(x) - B(\theta)}}_{g_\theta(T(x))} h(x).$$

Here are a few examples we saw last time:

Example 1.1 (Order statistics)**.** If $X_1, \dots, X_n \in \mathbb{R}$, $X_{(k)}$ is the k-th smallest value (including repeats). Then if $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} P^{(1)}$ with any model for $P^{(1)}$ on $\mathbb{R}$, then $S(X) = (X_{(1)}, \dots, X_{(n)})$ is sufficient.

Example 1.2. If $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\theta, 1)$, then we have the following hierarchy of sufficient statistics:

$$\begin{tikzcd}
& X \arrow[d] \arrow[ddl] \arrow[ddr] & \\
& S(X) \arrow[dl] \arrow[dr] & \\
\sum_i X_i & & \overline{X}
\end{tikzcd}$$

The higher up statistics in this diagram can be "compressed" more to get the ones at the bottom, which we may think of as minimal sufficient (or the most compressed).

1.2 Minimal sufficient statistics

Proposition 1.1. *If $T(X)$ is sufficient and $T(X) = f(S(X))$, then $S(X)$ is sufficient.*

So statistics with more information than sufficient statistics are also sufficient.

Proof. Using the factorization theorem,

$$\begin{aligned} p_\theta(x) &= g_\theta(T(x))h(x) \\ &= (g_\theta \circ f)(S(x))h(x). \end{aligned}$$

□

Here are the sufficient statistics with the least information.

Definition 1.1. $T(X)$ is **minimal sufficient** if

1) $T(X)$ is sufficient.
2) For any other sufficient statistic $S(X)$, $T(X) = f(S(X))$ for some f (a.s. in $\mathcal{P}$).

1.3 Likelihood functions

We will see that the shape of all likelihood ratios will be minimal sufficient, so any statistic that has the same information will be minimal sufficient.

Definition 1.2. If $\mathcal{P}$ has densities $p_\theta(x)$ with respect to μ the **likelihood function** (resp. **log-likelihood**) is the density (resp. **log-density**), reframed as a *random function* of θ.

$$\mathrm{Lik}(\Theta; X) = p_\theta(X), \qquad \ell(\theta; X) = \log \mathrm{Lik}(\theta; X).$$

If T is sufficient, then

$$\mathrm{Lik}(\theta; x) = \underbrace{g_\theta(T(x))}_{\text{determines shape}} \cdot \underbrace{h(x)}_{\text{scalar multiple}}.$$

Theorem 1.1. *Assume $\mathcal{P}$ has densities p_θ and $T(X)$ is sufficient for $\mathcal{P}$. If*

$$\mathrm{Lik}(\theta; x) \propto_\theta \mathrm{Lik}(\theta; y) \implies T(x) = T(y),$$

then $T(x)$ is minimal sufficient.

Proof. Proceed by contradiction. Suppose S is sufficient and there does not exist some f such that $f(S(x)) = T(x)$. Then there exist x, y with $S(x) = S(y)$ but $T(x) \neq T(y)$. Then

$$\begin{aligned} \mathrm{Lik}(\theta; x) &= g_\theta(S(x))h(x) \\ &\propto_\theta g_\theta(S(y))h(y) \\ &= \mathrm{Lik}(\theta; y), \end{aligned}$$

which is a contradiction. □

1.4 Minimal sufficiency in exponential families

Example 1.3. For an exponential family,

$$p_\theta(x) = e^{\eta(\theta)^\top T(x) - B(\theta)} h(x).$$

Is $T(X)$ minimal? Assume $\mathrm{Lik}(\theta; x) \propto_\theta \mathrm{Lik}(\theta; y)$, We want to show that $T(x) = T(y)$.

$$\begin{aligned}
\mathrm{Lik}(\theta; x) \propto_\theta \mathrm{Lik}(\theta; y) &\iff e^{\eta(\theta)^\top T(x) - B(\theta)} h(x) \propto_\theta e^{\eta(\theta)^\top T(y) - B(\theta)} h(y) \qquad \forall \theta \\
&\iff e^{\eta(\theta)^\top T(x)} = e^{\eta(\theta)^\top T(y)} c(x, y) \qquad \forall \theta \\
&\iff \eta(\theta)^\top T(x) = \eta(\theta)^\top T(y) + a(x, y) \qquad \forall \theta \\
&\iff \eta(\theta)^\top (T(x) - T(y)) = a(x, y) \qquad \forall \theta.
\end{aligned}$$

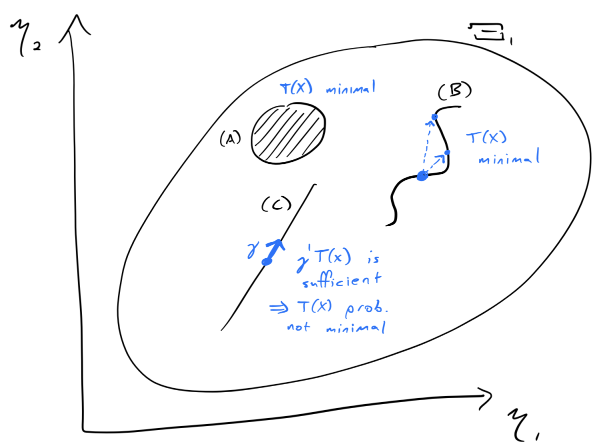
Plug in θ_1 and θ_2 to get 2 different equations and subtract:

$$\begin{aligned}
&\implies (\eta(\theta_1) - \eta(\theta_2))^\top (T(x) - T(y)) = 0 \qquad \forall \theta_1, \theta_2 \\
&\iff T(x) - T(y) \perp \mathrm{span}\{\eta(\theta_1) - \eta(\theta_2) : \theta_1, \theta_2 \in \Theta\}
\end{aligned}$$

If $\mathrm{span}\{\eta(\theta_1) - \eta(\theta_2) : \theta_1, \theta_2 \in \Theta\} = \mathbb{R}^s$, then we will get $T(x) = T(y)$.

Suppose $\eta(\theta) = \begin{bmatrix} \theta \\ 0 \end{bmatrix}$. Then $T_1(x)$ is sufficient. Does this mean that T cannot be minimal sufficient? In a $N(\mu, \sigma^2)$ family with $n = 1$, then $T(X) = \begin{bmatrix} X \\ X^2 \end{bmatrix}$. But if $n = 10$, then $T(x) = \begin{bmatrix} \sum_i X_i \\ \sum_i X_i^2 \end{bmatrix}$ in which case T cannot be recovered from T_1. So in general, it is possible that $T(X)$ may not be sufficient.

Here is a picture of exponential families A, B, and C in the natural parameter space Ξ.



- In exponential family A, the parameter space is locally 2-dimensional, so we get the whole span. Thus, $T(X)$ will be minimal.
- In exponential family B, we still get two vectors that span $\mathbb{R}^2$, so $T(X)$ is still minimal.
- In exponential family C, $\gamma^\top T(x)$ is minimal, where γ lies along the line. But $T(X)$ may not be minimal. If we say $\eta(\theta) = a + \theta\gamma$ with $\theta \in \mathbb{R}$, then $\eta^\top T(x) = a^\top T(x) + \theta\gamma^\top T(x)$.

Example 1.4. If $X \sim N_2(\mu(\theta), I_2) = e^{\mu(\theta)^\top x - B(\theta)} e^{-(1/2)x^\top x}$ with $\theta \in \mathbb{R}$. If $\Theta = \mathbb{R}$, $\mu(\theta) = a + \theta b$ with $a, b \in \mathbb{R}^2$, then

$$p_\theta(x) = e^{\theta(b^\top x) - B(\theta)} e^{-(1/2)(x-2a)^\top x}.$$

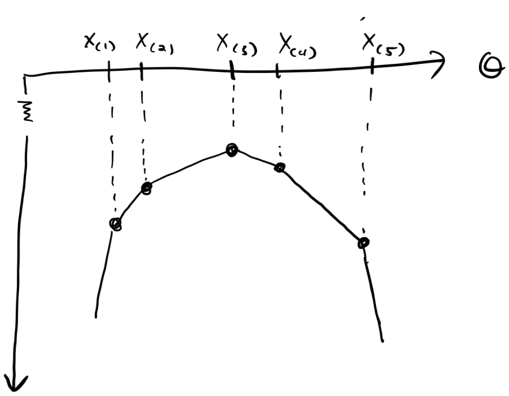
Because $b^\top x$ is sufficient, X is not minimal sufficient.

Example 1.5 (Laplace location family)**.** Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} p_\theta^{(1)}(x) = \frac{1}{2} e^{-|x-\theta|}$. Then

$$p_\theta(x) = \frac{1}{2^n} \exp\left(-\sum_{i=1}^n |x_i - \theta| \right),$$

so

$$\ell(\theta; x) = -\sum_{i=1}^n |x_i - \theta| - n \log 2.$$



Here, $(X_{(1)}, \ldots, X_{(n)})$ is minimal sufficient.

In many examples beyond exponential families, there aren't any useful sufficient statistics.

1.5 Complete statistics

Definition 1.3. $T(X)$ is **complete** for $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ if

$$\mathbb{E}_\theta[f(T)] = 0 \quad \forall\theta \implies f(T) \overset{\text{a.s.}}{=} 0 \quad \forall\theta$$

You should think of this as an upgrading of minimality.

Example 1.6. In the Laplace location family, are there any complete statistics? Let $f(S(X)) = \text{Med}(X) - \overline{X}$. Then $\mathbb{E}_\theta[f(S(x))] = \theta - \theta = 0$, but $\text{Med}(X) \neq \overline{X}$ a.s.

Definition 1.4. Let $\mathcal{P}$ be an exponential family with $p_\theta(x) = e^{\eta(\theta)^\top T(x) - B(\theta)} h(x)$. If $\Xi_0 = \eta(\Theta) = \{\eta(\theta) : \theta \in \Theta\}$ contains an open set, then we say $\mathcal{P}$ is **full-rank**. Otherwise, $\mathcal{P}$ is called **curved**.

Theorem 1.2. *If $\mathcal{P}$ is a full-rank exponential family, then $T(X)$ is complete sufficient.*

For a proof, see Lehmann and Romano Theorem 4.3.1.

Going back to our previous examples, in family A, T will be complete, whereas in families B and C, T will probably not be complete.

Theorem 1.3. *If $T(X)$ is complete sufficient for $\mathcal{P}$, then $T(X)$ is minimal.*

Proof. Assume $S(X)$ is minimal sufficient; we will recover T from S. Then $S(X) = f(T(X))$. Define

$$m(S(X)) = \mathbb{E}_\theta[T(X) \mid S(X)].$$

This is a proper statistic (not depending on θ) due to conditioning on the sufficiency of the statistic S. Then let $g(t) = t - m(f(t))$. Then

$$\mathbb{E}_\theta[g(T)] = \mathbb{E}_\theta[T] - \mathbb{E}_\theta[\mathbb{E}_\theta[T \mid S]] = 0 \qquad \forall\theta,$$

so $g(T) \overset{\text{a.s.}}{=} 0$ by completeness. This says that $T \overset{\text{a.s.}}{=} m(S(X))$. ☐

1.6 Ancillary statistics

Definition 1.5. $V(X)$ is **ancillary** for $\mathcal{P}$ if its distribution doesn't depend on θ.

This is a statistic that we already know without knowing θ.

Theorem 1.4 (Basu)**.** *If $T(X)$ is complete sufficient and $V(X)$ is ancillary, then $T \perp\!\!\!\perp V$ for all θ.*

Remark 1.1. Completeness is a property of the model, whereas independence is just a property of the distributions.