STATS 300A - Lecture 3

Dominik Rothenhaeusler
Scribed by Michael Howes

9/27/2021

Contents

1 Announcements

- HW1 due Wednesday.

- Sign up on gradescope

- Form study groups

2 Recap

- Decision theoretic framework

- Sufficiency and NFFC

- Exponential families

- Minimal sufficiency and optimal data reduction.

We will soon see a relationship between optimal data reduction and optimal estimation.

3 Minimal Sufficiency

Definition 1. A sufficient statistic T is minimal if for every sufficient statistic T', if $T'(x) = T'(y)$, then $T(x) = T(y)$.

That is, T is the coarsest sufficient statistic. Last time we stated the following theorem

Theorem 1. *Let $p(x; \theta)$ be the density of X (w.r.t. to μ) and let T be a statistic such that for all $x, y \in \mathcal{X}$,*

$$\text{there exists } C_{x,y} \text{ such that } p(x; \theta) = C_{x,y} p(y; \theta) \text{ for all } \theta \iff T'(x) = T'(y).$$

Then T is a minimal sufficient statistic.

In the case when $p(x; \theta) > 0$ for all x and θ, the condition on the left is equivalent to the ratio $\frac{p(x;\theta)}{p(y;\theta)}$ being constant as a function of θ.

Proof. [Discrete case where the densities are strictly positive] We will first show that T is sufficient. To do this we will use NFFC and show that for some h and g_θ we have

$$p(y; \theta) = h(y)g_\theta(T(y)),$$

for all $y \in \mathcal{X}$. For each t in the range of T, define $A_t = \{x \in \mathcal{X} : T(x) = t\}$ and choose $x_t \in A_t$. Then for every y we have $T(y) = T(x_{T(y)})$. Thus there exists $C_{y,x_{T(y)}}$ such that $p(y; \theta) = C_{y,x_{T(y)}} p(x_{T(y)}; \theta)$. Hence we can define $h(y) = C_{y,x_{T(y)}}$ and $g_\theta(t) = p(x_t; \theta)$ giving us our factorisation

$$p(y; \theta) = C_{y,x_{T(y)}} p(x_{T(y)}; \theta) = h(y)g_\theta(T(y)).$$

Thus T is sufficient.

Now suppose that T' is another sufficient statistic. Then there exists a factorisation $p(x; \theta) = \widetilde{h}(x)\widetilde{g}_\theta(T'(x))$. Now suppose that $T'(x) = T'(y)$ for some $x, y \in \mathcal{X}$. Then

$$\frac{p(x;\theta)}{p(y;\theta)} = \frac{\widetilde{h}(x)\widetilde{g}_\theta(T'(x))}{\widetilde{h}(y)\widetilde{g}_\theta(T'(y))} = \frac{\widetilde{h}(x)}{\widetilde{h}(y)},$$

and thus $T(x) = T(y)$ since the above ratio does not depend on θ. $\square$

Example 1. A natural question is: what are the minimal sufficient statistics for exponential families? The answer (Keener 3.17) is that for n iid samples from a minimal s-dimensional exponential family the statistics

$$\left(\sum_{i=1}^n T_1(X_i), \sum_{i=1}^n T_2(X_i), \ldots, \sum_{i=1}^n T_s(X_i) \right),$$

are minimal sufficient for $(X_i)_{i=1}^n$.

Example 2. Consider now the case $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\sigma, \sigma^2)$ where $\theta = \sigma > 0$. This is a curved exponential family and so the above result does not apply. We will show that $T(x) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2 \right)$ is minimal sufficient. Note that

$$\frac{p(x;\theta)}{p(x;\theta)} = \exp\left\{ \frac{1}{2\sigma^2}\left(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2 \right) + \frac{1}{\sigma^2}\left(\sum_{i=1}^n x_i - \sum_{i=1}^n y_i \right) \right\}.$$

Thus if $T(x) = T(y)$, then the above does not depend on θ. Suppose now that the above ratio is constant. Then

$$0 = \lim_{\sigma \to 0} \sigma^2 \log\frac{p(x;\theta)}{p(y;\theta)} = \frac{1}{2}\left(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2 \right) + \lim_{\sigma \to 0} \sigma\left(\sum_{i=1}^n x_i - \sum_{i=1}^n y_i \right) = \frac{1}{2}\left(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2 \right).$$

Thus $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_I^2$. Using this we can see that

$$0 = \lim_{\sigma \to 0} \sigma \log\frac{p(x;\theta)}{p}(y;\theta) = \sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2,$$

and thus $T(x) = T(y)$.

Example 3. Suppose $(X_i)_{i=1}^n \overset{\text{iid}}{\sim} \text{Unif}([0, \theta])$ this is not an exponential family and the support of $p(x; \theta)$ depend on θ. In this case, $T(x) = \max(X_1, \ldots, X_n)$ is minimal sufficient.

Example 4. Sometimes data reduction is not possible. If $(X_i)_{i=1}^n \overset{\text{iid}}{\sim} \text{Cauchy}$ with mean θ, then the order statistics $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ are minimal sufficient.

4 "Useless" data

The next definition captures the idea of useless information

Definition 2. A statistic A is *ancillary* for $X \sim P_\theta$ if the distribution of $A(X)$ does not depend on θ.

Example 5. Continuing the Cauchy example. The order statistics are minimal sufficient but the differences $X_{(i)} - X_{(j)}$ are ancillary for θ. This is because we can write $X_i = Z_i + \theta$ where Z_i is Cauchy with mean 0. This implies that $X_{(i)} = Z_{(i)} + \theta$ and thus $X_{(i)} - X_{(j)} = Z_{(i)} - Z_{(j)}$ which has no θ dependence. Thus we can see that sometimes minimal sufficient statistics contain "useless" information.

Later we will see the following general stratergy for finding an optimal estimator and proving optimality. We want to use the following two part stratergy

(a) Show that the optimal unbiased estimator is a function of a minimal statistic.

(b) Show there exists only one function of a minimal statistic that is unbiased.

We won't be able to show (b) if our minimal statistics contain an ancillary part and thus the next definitions will give us a new notion of minimal sufficiency. We first need a weaker notion of ancillary.

Definition 3. A statistic A is *first order ancillary* for $X \sim P_\theta$ if $\mathbb{E}_\theta[A(X)]$ does not depend on θ.

Definition 4. A statistic T is complete if any non-constant function of T is *not* first order ancillary.

Equivalenty a statistic T is complete if for all functions f if $\mathbb{E}_\theta[f(T(X))] = c$ for all θ, then $f \equiv c$. Note that since $\mathbb{E}_\theta[c] = c$ and $\mathbb{E}_\theta$ is linear this is equivalent to stating that for all functions f, if $\mathbb{E}_\theta[f(T(X))] = 0$ for all θ, then $f \equiv 0$. Complete statistics contain no useless info.

Bahadur's theorem states that every complete sufficient statistic is minimal sufficient.

Example 6. Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(\theta)$. We have seen that

$$T(X) = \sum_{i=1}^{n} X_i$$

is sufficient for $(X_i)_{i=1}^n$. We know that $T \sim \text{Binom}(n, \theta)$. Suppose f is a function of T satisfying $\mathbb{E}_\theta[f(T)] = 0$. Then, since T is discrete

$$
\begin{aligned}
0 &= \mathbb{E}_\theta[f(T)] \\
&= \sum_{k=0}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k} f(k) \\
&= (1-\theta)^n \sum_{k=0}^{n} \binom{n}{k} \left(\frac{\theta}{1-\theta} \right)^k f(k) \\
\therefore 0 &= \sum_{k=0}^{n} \binom{n}{k} f(k) \beta^k.
\end{aligned}
$$

Where $\beta = \frac{\theta}{1-\theta} \in [0, \infty)$. Thus the above polynoimal is β is zero on an uncountable set and hence $\binom{n}{k} f(k) = 0$ for all k and so $f \equiv 0$. Thus T is complete.

Another example when $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} N(\theta, \sigma^2)$ with σ^2 known. Then $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is a complete statistic (proof in scribed notes).