STAT 5010: Advanced Statistical Inference

Lecturer: Tony Sit
Scribe: Dominic Leung and Bencong Zhu

Lecture 5

5 Anciallarity and Completeness

5.1 Recap: Minimal Sufficiency

Theorem 1 *Let $\{p(x;\theta) : \theta \in \Omega\}$ be a family of densities with respect to some measure μ (Lebesgue measure for continuous distributionm, counting measure for discrete distribution). Suppose that there exists a statistic T such that for every $x, y \in \mathcal{X}$*

$$p(x;\theta) = C_{x,y}p(y;\theta), \quad \Leftrightarrow \quad T(x) = T(y).$$

for every θ and some $C_{x,y} \in \mathbb{R}$. Then T is a minimal sufficient statistic.

Reference from books: Theorem 6.2.3 **?**, Theorem 3.11 **?**.

Example 1 (Normal minimal sufficient statistic) *Let $X_1, \dots, X_n$ be iid $N(\mu, \sigma^2)$, with μ and σ^2 unknown. Let x and y be two sample points, and let $(\overline{x}, S_x^2)$ and $(\overline{y}, S_y^2)$ be the sample means and variances corresponding to the x and y samples respectively.*

$$\begin{aligned}\frac{f(x;\mu,\sigma^2)}{f(y;\mu,\sigma^2)} &= \frac{(2\pi\sigma^2)^{n/2}\exp\left[-\left\{n(\overline{x}-\mu)^2 + (n-1)S_x^2\right\}/(2\sigma^2)\right]}{(2\pi\sigma^2)^{n/2}\exp\left[-\left\{n(\overline{y}-\mu)^2 + (n-1)S_y^2\right\}/(2\sigma^2)\right]} \\ &= \exp\left[\left\{-n(\overline{x}^2 - \overline{y}^2) + 2n\mu(\overline{x}-\overline{y}) - (n-1)(S_x^2 - S_y^2)\right\}/(2\sigma^2)\right].\end{aligned}$$

This ratio will be constant as a function of $\theta = (\mu, \sigma^2)$ if and only if $\overline{x} = \overline{y}$ and $S_x^2 = S_y^2$. Thus, by the above theorem, $(\overline{X}, S^2)$ is a minimum sufficient statistic for θ, where $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X})^2$.

Example 2 (Curved exponential family) *Let $X_1, \dots, X_N \overset{iid}{\sim} N(\sigma, \sigma^2), \sigma > 0$. Denote $\theta = \sigma$, then*

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

Hence, $T(X) = (T_1(X), T_2(X)) = \left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i\right)$ is minimal sufficient.

Remark 1 *You should be reminded that if $p(x;\theta) = C_{x,y}p(y;\theta)$, x and y must be supported by the same θ (support of X : $\{x \in \mathcal{X} : p(x;\theta) > 0\}$). Otherwise, the 'constant' $C_{x,y}$ will be θ-dependent.*

Example 3 *Let $X_1, \dots, X_n \overset{iid}{\sim}$ Uniform$(0, \theta)$ and $T(X) = \max_{1\le i\le n} X_i = X_{(n)}$. In that case for $x = (x_1, \dots, x_n)$ such that $x_i > 0$, $i = 1, \dots, n$,*

$$p(x;\theta) = \prod_{i=1}^n \frac{1}{\theta} I(x_i < \theta) = \frac{1}{\theta^n} I(T(X) < \theta).$$

If $T(x)$ and $T(y)$ equals, then $p(x;\theta) = 1 \times p(y;\theta)$. The ratio between the two distributions does not depend on θ, so T is sufficient.

Conversely, if $x, y > 0$ (i.e. $x_i, y_i > 0$, $i = 1, \ldots, n$) are supported by the same θ's, then

$$\{\theta \text{ supporting } x\} = (T(x), \infty) = (T(y), \infty) = \{\theta \text{ supporting } y\}.$$

Therefore, it implies $T(x) = T(y)$ and is a minimal sufficient statistic.

Theorem 2 *For any minimal, s-dimensional exponential family, the statistic $(\sum_{i=1}^n T_1(X_I), \ldots, \sum_{i=1}^n T_s(X_i))$ is a minimal sufficient statistic. [Example 3.12 **?**]*

Proof 1 *Let $p(x;\theta) = \exp\{\eta(\theta)T(x) - B(\theta)\}\, h(x)$ be the density of an s-dimensional exponential family, where $\theta \in \Omega$. By NFFC, T is sufficient.*

Suppose $p(x;\theta) \propto_\theta p(y;\theta)$, then

$$e^{\eta(\theta)T(x)} \propto_\theta e^{\eta(\theta)T(y)},$$

which imples that

$$\eta(\theta) \cdot T(x) = \eta(\theta) \cdot T(y) + C,$$

where the constant C may depend on both x and y (but is independent of θ).

If θ_0 and θ_1 are any two points in Ω,

$$\{\eta(\theta_0) - \eta(\theta_1)\} \cdot T(x) = \{\eta(\theta_0) - \eta(\theta_1)\} \cdot T(y)$$

if and only if

$$\{\eta(\theta_0) - \eta(\theta_1)\} \cdot \{T(x) - T(y)\} = 0 \tag{1}$$

This shows that $T(x) - T(y)$ is orthogonal to eery vector in

$$\eta(\Omega) \ominus \eta(\Omega) \equiv \{\eta(\theta_0) - \eta(\theta_1) : \theta_0 \in \Omega, \theta_1 \in \Omega\},$$

so it must lie in the orthogonal complement of the linear span of $\eta(\Omega) \ominus \eta(\Omega)$. In particular, if the linear span of $\eta(\Omega) \ominus \eta(\Omega)$ is all of $\mathbb{R}^s$, then $T(x)$ must equal $T(y)$ in which case T is minimal sufficient.

5.2 Anciallarity and Completeness

Illustration:

Exponential families $\quad\rightarrow\quad$ significant data compression (without losing any information about θ).

Example 4 *Consider $X_1, \ldots, X_n \overset{iid}{\sim}$ Cauchy(θ), with densities*

$$p(x;\theta) = \frac{1}{\pi}\frac{1}{1 + (x-\theta)^2} \equiv f(x - \theta).$$

*Based on **?** §1.5 result, we know that $(X_{(1)}, \ldots, X_{(n)})$ is minimal sufficient. The similar conclusion/observation can also be found for the double exponential location model: $p(x;\theta) \propto \exp(|x - \theta|)$. [The density is $f(x;\theta) = e^{|x-\theta|}/2$].*

IDEA: Determine the amount of 'ancillarity' information stored in its minimal sufficient statistics.

Definition 3 *A statistic A is ancillary for $X \sim p_\theta \in \mathcal{P}$ if the distribution of $A(X)$ does not depend on θ.*

Example 4 (Continued) *Again $X_1, \ldots, X_n \overset{iid}{\sim}$ Cauchy(θ), then*

$$A(X) = X_{(n)} - X_{(1)} \text{ is ancillary},$$

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

To see this, observe that $X_i = Z_i + \theta$ for $Z_i \overset{iid}{\sim}$ Cauchy(0), so $X_{(i)} = Z_{(i)} + \theta$ and $A(X) = A(Z)$, which does not depend on θ.

Definition 4 (First-order ancillary statistic) *A statistic A is first-order ancillary for $X \sin p_\theta \in \mathcal{P}$ if $E_\theta(A(X))$ does not depend on θ.*

Definition 5 (Complete statistic) *A statistic T is complete for $X \sim p_\theta \in \mathcal{P}$ if no non-constant function of T is first-order ancillary. In other words, if $E_\theta(f(T(X)) = 0$ for all θ, then $f(T(X)) = 0$ with probability 1 for all θ.*

Remark 2

1. *If T is complete sufficient, then T is minimal sufficient. [Bahadur's theeorem].*
2. *Complete sufficient statistic yield optimal unbiased estimators.*

Example 5 (Discrete Case) *Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(θ), $\theta \in (0, 1)$. Then $T(X) = \sum_{i=1}^n X_i$ is sufficient.*

Suppose that $E_\theta[f(T(X))] = 0$ for all $\theta \in (0, 1)$, then

$$\sum_{j=0}^{n} f(j) \binom{n}{j} \theta^j (1-\theta)^{n-j} = 0, \qquad \forall \theta \in [0, 1]. \tag{2}$$

Dividing both sides by θ^n and substituting $\beta = \theta/(1-\theta)$, we can rewrite (2) as

$$\sum_{j=0}^{n} f(j) \binom{n}{j} \beta^j = 0 \qquad \forall \beta > 0.$$

If f are non-zero, then the quantity on the LHS is a polynomial of degree at most n. However, an nth-degree polynomial can have at most n roots. Hence, it is impossible for the LHS equals 0 for every $\beta > 0$ unless $f = 0$. So, T is complete.

Example 6 (Continuous Case) *Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ with unknown $\theta \in \mathbb{R}$ and $\sigma^2 > 0$. We can verify if $T(X) = \overline{X}_n = n^{-1} \sum_{i=1}^n X_i$ is complete for the model. [Note that T is minimal sufficient.]*

Let's consider the case wtih $n = 1$ and assume WLOG $\sigma^2 = 1$, $T(X) \sim N(\theta, 1)$. Suppose

$$E_\theta(f(X)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \exp\left\{ -\frac{(x-\theta)^2}{2} \right\} dx = 0, \qquad \forall \theta \in \mathbb{R}. \tag{†}$$

We then decompose f into its positive and negative parts as

$$f(x) = f_+(x) - f_-(x),$$

where $f_+(x) = \max(f(x), 0)$ and $f_-(x) = \max(-f(x), 0)$. Then $f_+(x) \geq 0$ and $f_-(x) \geq 0$ for all $x \in \mathbb{R}$.

Observation: *$f_+(x) = f_-(x)$ if and only if $f_+(x) = f_-(x) = 0$.*

1. *If $f(x) \geq 0$ almost everywhere (a.e.) or $f(x) \leq 0$ a.e., then (†) implies that $f(x) = 0$ a.e. because setting $\theta = 0$ because setting $\theta = 0$ gives us an integral of a nonnegative (resp. non-positive) function of zero. This gives/shows completeness.*

2. *Suppose f_+ and f_- have non-zero components, we may write*

$$\frac{\int_{-\infty}^{\infty} f_+(x) e^{-\frac{x^2}{2}} e^{\theta x} dx}{\int_{-\infty}^{\infty} f_+(x) e^{-\frac{x^2}{2}} dx} = \frac{\int_{-\infty}^{\infty} f_-(x) e^{-\frac{x^2}{2}} e^{\theta x} dx}{\int_{-\infty}^{\infty} f_-(x) e^{-\frac{x^2}{2}} dx}, \tag{††}$$

since (†) shows that the denominator of (††) are both equal. The quantity

$$\frac{f_+(x) e^{-\frac{x^2}{x}}}{\int_{-\infty}^{\infty} f_+(x) e^{-x^2} x dx}$$

defines a probability density and the LHS of (††) is the moment generating function of this density. Similarly, the RHS is the moment generating function of the density

$$\frac{f_-(x) e^{-\frac{x^2}{x}}}{\int_{-\infty}^{\infty} f_-(x) e^{-x^2} x dx}$$

It implies that $f_+(x) = f_-(x)$ a.e.. Then $f_+(x) = f_-(x) = 0$ a.e., or in other words, $f(x) = 0$ a.e..

Hence T is copmlete (and sufficient).

Example 7 *(Example 3.16 of Keenen (2010))*

Exercise 1 *If $X_1, \ldots, X_n \overset{iid}{\sim} p(x, \theta) \propto h(x) e^{\theta x}$, then the Statistics $T(x) = X$ is complete*
$\rightarrow$ Suppose $\int f(x) h(x) e^{\theta x} dx = 0$ for all $\theta \in \Omega$
$\rightarrow$ decompose $f(x) = f_+(x) - f_-(x)$ with $f_+ \geq 0, f_- \geq 0$
$\rightarrow$ f_+ and f_- can be viewed as unnormalised densities $p_+(x)$ and $p_-(x)$, respectively.
$\rightarrow$ argue that the $m.g.f.$'s of p_+ and p_- are equal

Theorem 6 (Theorem 4.3.1 ?) *$(T_1, \ldots, T_n)$ is complete for any s-dimensional full rank exponential family. [see P. 117 of TSH]*

Theorem 7 (Basu's Theorem) *If T is complete and sufficient for $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ and V is ancillary, then $T(X) \perp\!\!\!\perp V$.*

Proof 2 *Define $q_A(t) = P_\theta(V \in A \mid T = t)$ or $q_A(T) = P_\theta(V \in A \mid T)$ and $p_A = P_\theta(V \in A)$. By sufficiency and ancillarity, neither p_A nor $q_A(t)$ depends θ. By smoothing,*

$$(P_A = P_\theta(V \in A) = E_\theta \left(P_\theta(V \in A \mid T) \right) = E_\theta(q_A(T))$$

and so by completeness, $q_A(T) = p_A$ a.e. for $\mathcal{P}$. Again, by smoothing/tower expecation,

$$\begin{aligned} P_\theta(T \in B, V \in A) &= E_\theta \left(1_B(T) 1_A(V) \right) \\ &= E_\theta \left(E_\theta \left(1_B(T) 1_A(v) \mid T \right) \right) \\ &= E_\theta \left(1_B(T) E_\theta \left(1_A(v) \mid T \right) \right) \\ &= E_\theta \left(1_B(T) q_A(T) \right) \\ &= E_\theta \left(1_B(T) \cdot p_A \right) \\ &= P_\theta(T \in B) \cdot P_\theta(V \in A) \end{aligned}$$

Hence, T and V ane independent as A and B are arbitrary Borel sets.

Example 8 *Let* $X_1, \ldots, X_n \overset{iid}{\sim} N\left(\mu, \sigma^2\right)$ *, where both of* μ *and* σ^2 *are unknown. Then* $\bar{X}_n \perp\!\!\!\perp n^{-1} \sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$ *with* $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$.

Fix any $\sigma > 0$ and consider the submodel $\mathcal{P}_\sigma = \left\{N\left(\mu, \sigma^2\right) : \mu \in \mathbb{R}\right\}$ In each submodel, $\bar{X}_n$ is complete and sufficient, and $n^{-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ is ancillary

$$X_i = Z_i + \mu \qquad X_i - \bar{X}_n \to Z_i - \bar{Z}$$

By Basu's theorem, $\bar{X}_n \perp\!\!\!\perp \sum_{i=1}^{n}\left(x_i - \bar{x}_n\right)^2$ under $N\left(\mu, \sigma^2\right)$ for any μ. Since σ is arbitrary, we can conclude that $\bar{X}_n \perp\!\!\!\perp \sum_{i=1}^{n}\left(x_i - \bar{x}_n\right)^2$ hold for the full model $\mathcal{P} = \left\{N\left(\mu, \sigma^2\right) : \mu \in \mathbb{R}, \sigma^2 > 0\right\}$

From data reduction to optimal Inference

Lossless data reduction (sufficiency)	→	Optimal data reduction (minimal sufficiency + completeness)	→	Optimal inference estimator

Definition 8 *A function* $f : C \to \mathbb{R}$ *with* C *convex is a convex function if* $x \neq y \in C$ *and* $\gamma \in (0,1)$*:*

$$f(\gamma x + (1-\gamma)y) \leqslant \gamma f(x) + (1-\gamma) f(y)$$

The function f *is said to be strictly convex if the above inequality holds strictly (ie. "* $<$ *")*

Example 9 *For any* $\theta \in \Omega$*, the function* $f(d) = (d-\theta)^2$ *is strictly convex on* $\mathbb{R}$.

Example 10 *For any* $\theta \in \Omega$*, the function* $f(d) = |d - \theta|$ *is convex, but not strictly convex.*

Theorem 9 (Jensen's Inequality) *If* $f : C \to \mathbb{R}$ *is convex on any open set* C*,* $P(x \in C) = 1$ *and* $E(X)$ *exists, then*

$$f(E(x)) \leqslant E(f(x))$$

If f *is strictly convex, then the above inequality holds strictly unless* $X = E(X)$ $w.p.1$

Theorem 10 *Suppose that* T *is sufficient for* $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$*, that* $\delta(X)$ *is an estimator for* $g(\theta)$ *for which* $E(\delta(x))$ *exists and that* $R(\theta, \delta) = E_\theta L(\theta, \delta(x)) < \infty$*. If, in particular,* $L(\theta, \cdot)$ *is convex (as a function of* $d \in \mathcal{D}$*), then*

$$R(\theta, \eta) \leqslant R(\theta, \delta) \quad for \quad \eta(T(x)) = E(\delta(x) \mid T(x))$$

If $L(\theta, \cdot)$ *is strictly convex, then* $R(\theta, \eta) < R(\theta, \delta)$ *for any* θ *unless* $\eta(T^{'}(x)) = \delta(x)$

Example 11 *Let* $X_1, \ldots, X_n \overset{iid}{\sim} Bernoulli(\theta), \theta \in (0,1)$*, and consider* $L(\theta, d) = (\theta - d)^2$*. Suppose we start with an unresaonable estimator* $\delta(X) = X_1$*. We know that* $T(X) = \bar{x}_n$ *is sufficient, so we can apply Rao-Blackwell theorem to improve our eatimator* δ

$$\begin{aligned}
\underline{\eta(T(X)} &= E(\delta(X) \mid T(X)) \\
&= \frac{1}{n}\sum_{i=1}^{n} E(X_1 \mid \bar{X}_n) \\
&= E(\bar{X}_n \mid \bar{X}_n) = \bar{X}_n
\end{aligned}$$

Recall that in lecture I, we showed already that $R(\theta, \eta) = \frac{\theta(1-\theta)}{n} < \theta(1-\theta) = R(\theta, \delta)$.

$$(\delta_{naive}(X) = 1/2) \qquad R(1/2, \delta_{naive}) < R(1/2, \delta^{'})$$

References