Completeness

Outline

1) Completeness
2) Ancillarity
3) Basu's Theorem

Completeness

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

if $\mathbb{E}_\theta f(T(X)) = 0 \quad \forall \theta$

$\Rightarrow \quad f(T) \overset{a.s.}{=} 0 \quad \forall \theta$

[Name comes from a prior notion that
$\mathcal{P}^T = \{P_\theta^T : \theta \in \Theta\}$ is "complete basis"
wrt inner product $\langle f, P_\theta^T \rangle = \int f(t)\, dP_\theta^T(t)$]
(see HW 3)

Ex. (Cont'd) Laplace location family has minimal suff stat. $S = (X_{(i)})_{i=1}^n$. Complete?

No: Let $M(S) = \text{median}(X)$

$\bar{X}(S) = \frac{1}{n} \sum X_i$

$\mathbb{E}_\theta \bar{X} = \mathbb{E}_\theta M = \theta$ (by symmetry)

$\mathbb{E}_\theta [\bar{X}(S) - M(S)] = 0 \quad \forall \theta$

$S(X)$ still has "a lot of extra fluff"

Ex $X_1, \ldots, X_n \overset{iid}{\sim} U[0, \theta] \quad \theta \in (0, \infty)$

Can show $T(X) = X_{(n)}$ min. suff. Complete?

Find density of $T(X)$:

$$\mathbb{P}_\theta(T \le t) = \left(\frac{t}{\theta} \wedge 1\right)^n = \left(\frac{t}{\theta}\right)^n \wedge 1$$

$$\Rightarrow p_\theta(t) = \frac{d}{dt}\mathbb{P}_\theta(T \le t)$$

$$= n\frac{t^{n-1}}{\theta^n} \, 1\{t \le \theta\}$$

Suppose $0 = \mathbb{E}_\theta f(T) \quad \forall \theta > 0$

$$= \frac{n}{\theta^n}\int_0^\theta f(t)\, t^{n-1}\, dt \quad \forall \theta > 0$$

$$\Rightarrow \int_0^\theta f(t)\, t^{n-1}\, dt = 0 \quad \forall \theta > 0$$

$$\Rightarrow f(t)\, t^{n-1} = 0 \quad \text{a.e. } t > 0$$

Def Assume $\mathcal{P} = \{P_\eta : \eta \in \Xi\}$ has densities

$$p_\eta(x) = e^{\eta' T(x) - A(\eta)} h(x)$$

If $T(X)$ satisfies no linear constraint $\left(\begin{array}{l}\nexists\ \beta \neq 0, \alpha: \\ \beta' T(X) \overset{a.s.}{=} \alpha\end{array}\right)$ and Ξ contains an open set, we say $\mathcal{P}$ is <u>full-rank</u>

If $\mathcal{P}$ is not full-rank we say it is <u>curved</u>

[Note: If $T(x)$ satisfies linear constraint, then $\mathcal{P}$ might still be full-rank for a lower-dim. sufficient statistic]

Proof in Lehmann & Romano, Thm. 4.3.1

Theorem If $\mathcal{P}$ is full rank then $T(X)$ is complete sufficient

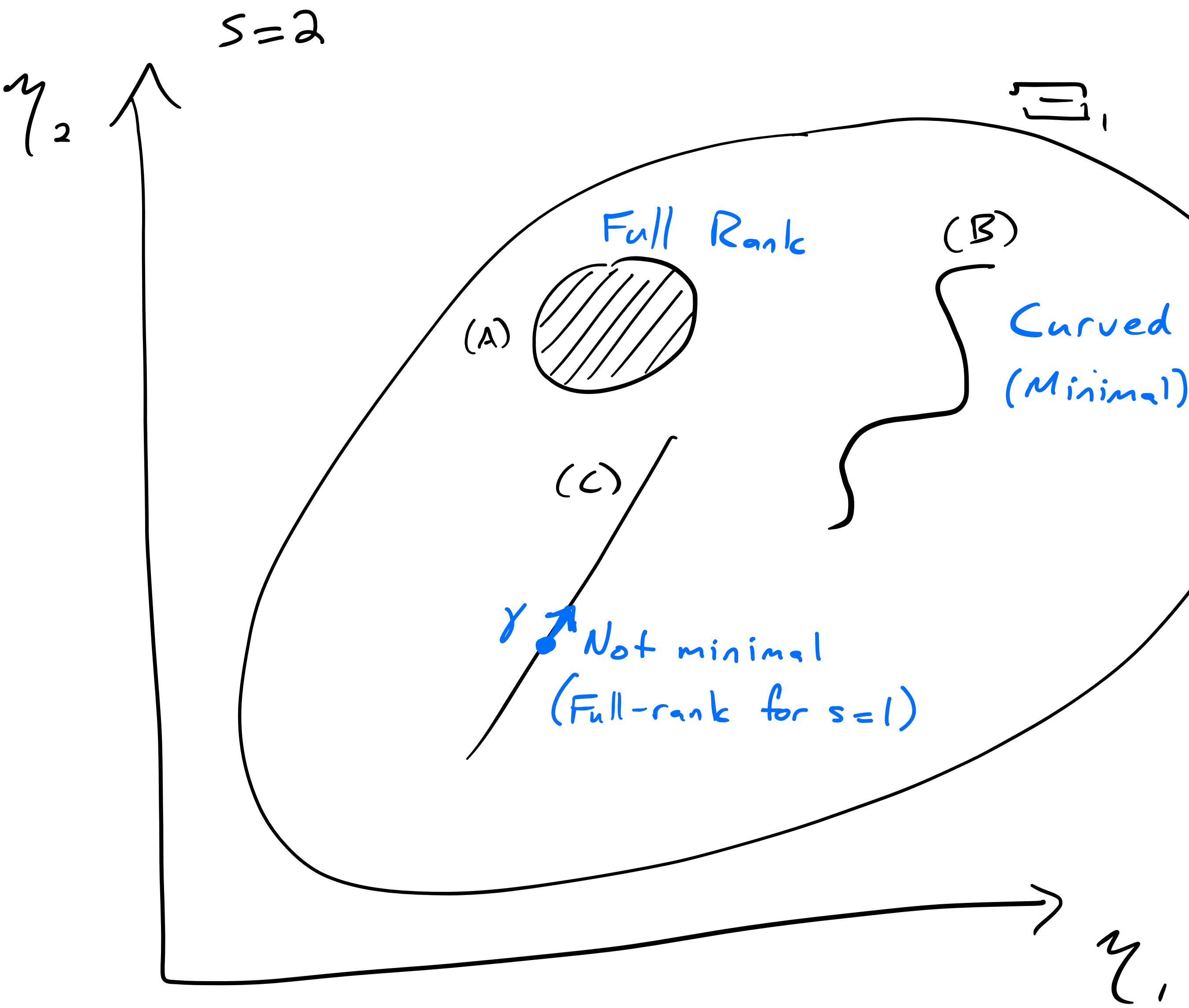
Proof idea wlog $T(X) = X$, $p_\eta(x) = e^{\eta' x - A(\eta)}$, $0 \in \overset{\circ}{\Xi}$ (interior)

Write $f(x) = f^+(x) - f^-(x)$, for $f^+, f^- \geq 0$

$$\underbrace{\int e^{\eta' x} f^+(x)\, d\mu(x)}_{\text{MGF for } Y^+ \sim \frac{f^+(x)}{\int f^+ d\mu}} = \underbrace{\int e^{\eta' x} f^-(x)\, d\mu(x)}_{\text{MGF for } Y^- \sim \frac{f^-(x)}{\int f^- d\mu}}$$

Uniqueness of MGFs $\Rightarrow Y^+ \overset{D}{=} Y^- \Rightarrow f^+ \overset{a.s.}{=} f^-$

Diagram again



$T(X)$ definitely complete for (A)

Maybe not for (B), (C)

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

Game plan for completeness proofs: show two things are a.s. equal by showing they have = expectation.

Proof Assume $S(X)$ is minimal suff

Let $\bar{T}(S(X)) = \mathbb{E}_{\cancel{\theta}}\left[T(X) \mid S(X)\right]$ ← S suff.

Claim: $\bar{T}(S(X)) \overset{a.s.}{=} T(X)$

We have $S(X) \overset{a.s.}{=} f(T(X))$ (S minimal suff)

Let $g(t) = t - \bar{T}(f(t))$

$$\mathbb{E}_\theta\left[g(T(X))\right] = \mathbb{E}_\theta T(X) - \mathbb{E}_\theta\left[\bar{T}(S(X))\right]$$

$$= \mathbb{E}_\theta T(X) - \mathbb{E}_\theta\left[\mathbb{E}[T \mid S]\right]$$

$$= 0$$

$$\Rightarrow g(T(X)) \overset{a.s.}{=} 0 \quad \text{(completeness)}$$

$\boxtimes$

Ancillarity

Two reasons to care about completeness:

1) Uniqueness of unbiased estimators using T

If $\mathbb{E}_\theta \delta_1(T) = \mathbb{E}_\theta \delta_2(T) = g(\theta), \forall \theta \in \Theta$

Then $\mathbb{E}_\theta[\delta_1 - \delta_2] = 0 \Rightarrow \delta_1 \overset{a.s.}{=} \delta_2$

[We will explore this further next time]

2) Basu's theorem: neat way to show independence

<u>Def</u> $V(X)$ is <u>ancillary</u> for $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ if its distribution does not depend on θ. (V carries no info. about θ)

(Aside:) Conditionality Principle

If $V(X)$ is ancillary then all inference should be conditional on $V(X)$

[will return to this in testing & CI unit

Basu's Theorem

Theorem (Basu)

If $T(X)$ is complete sufficient and $V(X)$ is ancillary for $\mathcal{P}$, then

$$V(X) \perp\!\!\!\perp T(X) \quad \text{for all } \theta \in \Theta$$

Proof

Want $\mathbb{P}_\theta(V \in A, T \in B) = \mathbb{P}_\theta(V \in A)\, \mathbb{P}_\theta(T \in B)$ all A, B, θ

$$\text{Let} \quad q_A(T(X)) = \mathbb{P}_{\cancel{\theta}}(V \in A \mid T) \quad \leftarrow T \text{ suff.}$$

$$p_A = \mathbb{P}_{\cancel{\theta}}(V \in A) \quad \leftarrow V \text{ ancillary}$$

$$\mathbb{E}_\theta\left[q_A(T) - p_A\right] = p_A - p_A = 0, \quad \forall \theta$$

$$\Longrightarrow \quad q_A(T) \overset{a.s.}{=} p_A \qquad \forall \theta$$

$$\begin{aligned}
\mathbb{P}_\theta(V \in A, T \in B) &= \int q_A(t)\, 1\{t \in B\}\, dP_\theta^T(t) \\
&= p_A \int 1\{t \in B\}\, dP_\theta^T(t) \\
&= \mathbb{P}(V \in A)\, \mathbb{P}_\theta(T \in B) \quad \boxtimes
\end{aligned}$$

Using Basu's Theorem

Ancillarity, Completeness, Sufficiency are all properties wrt a family $\mathcal{P}$

Independence is a property of a distribution

If you can't verify the thm's hypotheses for one family, try a different family!

Ex. $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ $\mu \in \mathbb{R}, \sigma^2 > 0$

Sample mean $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

Sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

Want to show $\bar{X} \perp\!\!\!\perp S^2$

But neither stat. is ancillary or sufficient in the full family with μ, σ^2 unknown

To apply Basu, use family with σ^2 known:

$$\mathcal{P} = \{ N(\mu, \sigma^2)^n : \mu \in \mathbb{R} \}$$

In $\mathcal{P}$, $\bar{X}$ is complete sufficient

and S^2 is ancillary since

$$S^2 = \sum (Z_i - \bar{Z})^2 \quad \text{for } Z_i = X_i - \mu \overset{iid}{\sim} N(0, \sigma^2)$$

not statistics
but doesn't matter

Therefore $\bar{X} \perp\!\!\!\perp S^2$

[Conclusion has nothing to do with "known" or "unknown" parameters]