

Outline

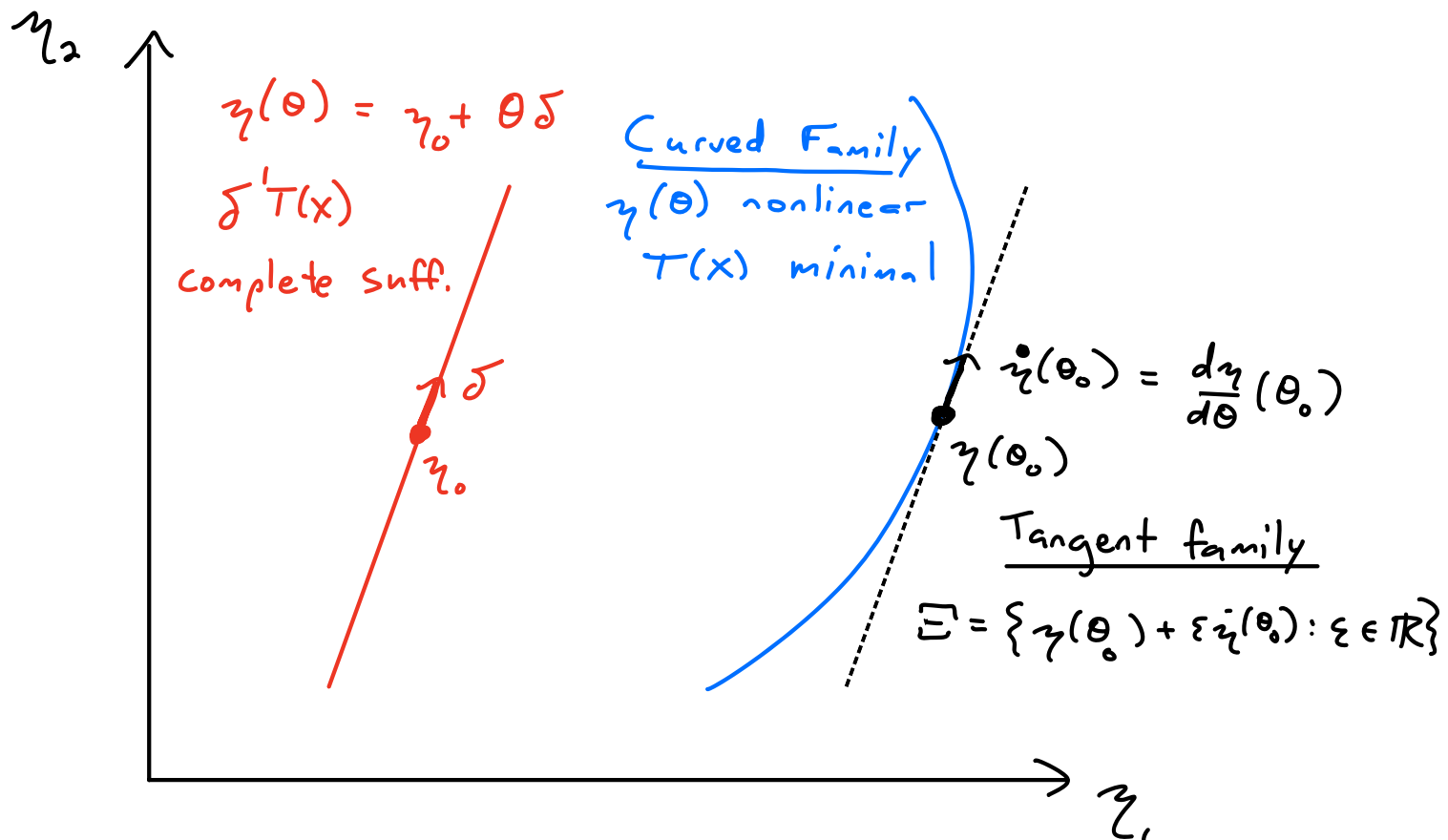
9/19/2023

- 1) Score function
- 2) Fisher information
- 3) Cramér-Rao Lower Bound
- 4) Examples

Motivation: Tangent family

$$p_{\theta}(x) = e^{\eta(\theta)'T(x) - A(\eta(\theta))} h(x) \quad \eta: \mathbb{R} \rightarrow \mathbb{R}^2$$

$$\Xi = \{\eta(\theta) : \theta \in \mathbb{R}\}$$



$$q_{\varepsilon}(x) = e^{(\eta(\theta_0) + \varepsilon \dot{\eta}(\theta_0))' T(x) - A(\dots)} h(x)$$

$$= e^{\underbrace{\varepsilon \dot{\eta}(\theta_0)' (T(x) - \mathbb{E}_{\theta_0} T)}_{S_{\theta_0}(x)} - B(\varepsilon)} k(x)$$

Complete sufficient for tangent family at θ_0
 Called Score function

Score function

Assume \mathcal{P} has densities p_θ wrt μ , $\Theta \subseteq \mathbb{R}^d$

Common support: $\{x: p_\theta(x) > 0\}$ same $\forall \theta$

Recall $l(\theta; x) = \log p_\theta(x)$,

Thought of as random function of θ

Def The score is $\nabla l(\theta; x)$; plays a key role in many areas of statistics, esp. asymptotics.

Can think of as "local complete sufficient statistic":

$$\begin{aligned} p_{\theta_0 + \eta}(x) &= e^{l(\theta_0 + \eta; x)} \\ &\approx e^{\eta' \nabla l(\theta_0; x)} p_{\theta_0}(x) \quad \text{for } \eta \approx 0 \end{aligned}$$

Differential identities: (assuming enough regularity)

$$1 = \int_{\mathcal{X}} e^{l(\theta; x)} d\mu(x)$$

$$\frac{\partial}{\partial \theta_j} \Rightarrow 0 = \int \frac{\partial}{\partial \theta_j} l(\theta; x) e^{l(\theta; x)} d\mu(x)$$

$$\Rightarrow \mathbb{E}_\theta [\nabla l(\theta; x)] = 0$$

↑
only true if these are the same value of θ !

$$\begin{aligned} \frac{\partial}{\partial \theta_k} \Rightarrow 0 &= \int \left(\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_k} + \frac{\partial \ell}{\partial \theta_i} \cdot \frac{\partial \ell}{\partial \theta_k} \right) e^\ell d\mu \\ &= \mathbb{E}_\theta \left[\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_k} \right] + \mathbb{E}_\theta \left[\frac{\partial \ell}{\partial \theta_i} \frac{\partial \ell}{\partial \theta_k} \right] \end{aligned}$$

$$\Rightarrow J(\theta) = \text{Var}_\theta [\nabla \ell(\theta; x)] = \mathbb{E}_\theta [-\nabla^2 \ell(\theta; x)]$$

← same θ
← same θ

Called "Fisher Information"

[It is possible to extend this definition to certain cases where ℓ is not even differentiable, e.g. Laplace location family, but for our purposes we can just assume "sufficient regularity."]

Try with another statistic $\delta(x)$, let $g(\theta) = \mathbb{E}_\theta[\delta(x)]$ ("unbiased estimator")

$$g(\theta) = \int \delta e^\ell d\mu$$

$$\Rightarrow \nabla g(\theta) = \int \delta \nabla \ell e^\ell d\mu = \mathbb{E}_\theta [\delta(x) \nabla \ell(\theta; x)]$$

$$\stackrel{\text{Since } \mathbb{E} \nabla \ell = 0}{=} \text{Cov}_\theta(\delta(x), \nabla \ell(\theta; x))$$

Combining these results with Cauchy-Schwarz gives us the Cramér-Rao Lower Bound or Information Lower Bound:

1-param: $\text{Var}_{\theta}(\delta) \cdot \text{Var}_{\theta}(\dot{\ell}(\theta; X)) \geq \text{Cov}_{\theta}(\delta, \dot{\ell}(\theta; X))^2$

$$\Rightarrow \text{Var}_{\theta}(\delta) = \dot{g}(\theta)^2 / J(\theta)$$

Multivariate: $\theta \in \mathbb{R}^d$, $g(\theta)$, $\delta(x) \in \mathbb{R}$

$$\text{Var}_{\theta}(\delta) \geq \nabla g(\theta)' J(\theta)^{-1} \nabla g(\theta)$$

Proof:

$$\begin{aligned} \text{Var}_{\theta}(\delta) \cdot a' J(\theta) a &= \text{Var}_{\theta}(\delta) \text{Var}(a' \nabla \ell(\theta)) \\ &\geq \text{Cov}_{\theta}(\delta, a' \nabla \ell(\theta))^2 \\ &= a' \nabla g \nabla g' a, \text{ for all } a \in \mathbb{R}^d \end{aligned}$$

$$\Rightarrow \text{Var}_{\theta}(\delta) \geq \max_{a \neq 0} \frac{a' \nabla g \nabla g' a}{a' J(\theta) a} \quad \text{Exercise} \quad (= \nabla g' J(\theta)^{-1} \nabla g)$$

Interp: If $g(\theta)$ is estimand, no unbiased estimator can have smaller variance than $\nabla g(\theta)' J(\theta)^{-1} \nabla g(\theta)$

Ex.: (i.i.d. sample)

$$X_1, \dots, X_n \stackrel{\text{iid}}{\sim} p_{\theta}^{(1)}(x) \quad \theta \in \Theta \subseteq \mathbb{R}^d$$

p_{θ} "regular": common support, finite derivative wrt θ

$$X \sim p_{\theta}(x) = \prod_i p_{\theta}^{(1)}(x_i)$$

$$\text{Let } \ell_1(\theta; x_i) = \log p_{\theta}^{(1)}(x_i)$$

$$\ell(\theta; x) = \sum_i \ell_1(\theta; x_i)$$

$$J(\theta) = \text{Var}_{\theta}(\nabla \ell(\theta; x))$$

$$= \text{Var}_{\theta}(\sum_i \nabla \ell_1(\theta; x_i))$$

$$= n J_1(\theta)$$

where $J_1(\theta)$ is Fisher info
in single observation

\Rightarrow Lower bound scales like n^{-1} (SD $\asymp n^{-1/2}$ for "regular" families)

Efficiency

CRLB is not nec. attainable.

We define the efficiency of an unbiased estimator as:

$$\text{eff}_{\theta}(\delta) = \frac{\text{CRLB}}{\text{Var}_{\theta}(\delta)} \quad (= \frac{1/J(\theta)}{\text{Var}_{\theta}(\delta)} \text{ if } g(\theta) = \theta \in \mathbb{R})$$

$$\text{eff}_{\theta}(\delta) \leq 1$$

We say $\delta(x)$ is efficient if $\text{eff}_{\theta}(\delta) = 1 \quad \forall \theta$

Depends on $\text{Corr}_{\theta}(\delta(x), \nabla \ell(\theta; x))$:

$$\begin{aligned} \text{eff}_{\theta}(\delta) &= \frac{\text{Cov}_{\theta}^2(\delta(x), \dot{\ell}(\theta; x))}{\text{Var}_{\theta}(\delta) \cdot \text{Var}_{\theta}(\dot{\ell}(\theta))} \\ &= \text{Corr}_{\theta}^2(\delta, \dot{\ell}(\theta)) \end{aligned}$$

$$\leq 1$$

$$\delta(x) \text{ is efficient} \Leftrightarrow \text{Corr}_{\theta}^2(\delta, \dot{\ell}(\theta)) = 1 \quad \forall \theta$$

Rarely achieved in finite samples but we can approach it asymptotically as $n \rightarrow \infty$

Ex. Exponential Families

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

$$\ell(\eta; x) = \eta' T(x) - A(\eta) + \log h(x)$$

$$\begin{aligned}\nabla \ell(\eta; x) &= T(x) - \nabla A(\eta) \\ &= T(x) - \mathbb{E}_{\eta} T(x)\end{aligned}$$

$$\text{Var}_{\eta}(\nabla \ell(\eta)) = \text{Var}_{\eta}(T(x)) = \nabla^2 A(\eta)$$

$$\nabla^2 \ell(\eta; x) = -\nabla^2 A(\eta)$$

$$\mathbb{E}_{\eta}[-\nabla^2 \ell(\eta; x)] = \nabla^2 A(\eta) \quad \checkmark$$

So any unbiased est. of η has

$$\text{Var}_{\eta}(\delta) \geq \nabla^2 A(\eta)^{-1}$$

Curved family: $p_{\theta}(x) = e^{\eta(\theta)'T(x) - B(\theta)} h(x)$, $\theta \in \mathbb{R}$
 $B(\theta) = A(\eta(\theta))$

$$l(\theta; x) = \eta(\theta)'T(x) - B(\theta) + \log h(x)$$

$$\begin{aligned} \dot{l}(\theta; x) &= \dot{\eta}(\theta)'T(x) - \dot{\eta}(\theta)'\nabla_{\eta}A(\eta(\theta)) \\ &= \dot{\eta}(\theta)'(T(x) - \nabla_{\eta}A(\eta(\theta))) \\ &= \dot{\eta}(\theta)'(T(x) - E_{\theta}T(x)) \end{aligned}$$

$\Rightarrow \dot{\eta}(\theta)'T(x)$ is "locally complete suff. stat."

