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

Motivation: Tangent family

$$\rho(x) = e^{\eta(\theta) T(x) - A(\gamma(\theta))} h(x) \qquad \gamma: \mathbb{R} \to \mathbb{R}^{2}$$

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

$$\gamma(\theta) = \gamma_{0} + \theta \delta \qquad \text{Curved } F_{amily}$$

$$\gamma(\theta) = \eta_{0} + \theta \delta \qquad \text{Curved } F_{amily}$$

$$\gamma(\theta) = \frac{d\eta}{d\theta}(\theta_{0})$$

$$\gamma(\theta_{0}) = \frac$$

Called Score function

Score function

Assume \mathcal{F} has densities \mathcal{F}_{θ} with \mathcal{M} , $\mathcal{G} \subseteq \mathbb{R}^d$ 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 = key role in many areas of statistics, esp. asymptotics.

Differential identities: (assuming enough regularity)

$$1 = \int_{\chi} e^{\ell(\Theta_{j_x})} du(x)$$

$$\frac{\partial}{\partial \theta_{j}} \Rightarrow O = \int \frac{\partial}{\partial \theta_{j}} \ell(\theta_{j} \times) e^{\ell(\theta_{j} \times)} d\mu(x)$$

$$E_{\Theta}[Vl(\Theta; X)] = 0$$
only true if these are the same value of $\Theta!$

$$\frac{\partial}{\partial \theta_{k}} \Rightarrow 0 = \int \left(\frac{\partial^{2} \mathcal{Q}}{\partial \theta_{i} \partial \theta_{k}} + \frac{\partial \mathcal{L}}{\partial \theta_{i}} \cdot \frac{\partial \mathcal{L}}{\partial \theta_{k}}\right) e^{i} dm$$

$$= \mathbb{E}_{\theta} \left(\frac{\partial^{2} \mathcal{L}}{\partial \theta_{i} \partial \theta_{k}}\right) + \mathbb{E}_{\theta} \left(\frac{\partial \mathcal{L}}{\partial \theta_{i}} \cdot \frac{\partial \mathcal{L}}{\partial \theta_{k}}\right)$$

$$\Rightarrow Var_{\theta} \left[\nabla \mathcal{L}(\theta; X)\right] = \mathbb{E}_{\theta} \left[-\nabla^{2} \mathcal{L}(0; X)\right]$$

$$\int (0) \qquad \text{Same } \theta \qquad \text{Same } \theta$$

$$C. \text{Med "Fisher Information"}$$

$$\text{It is possible to extend this definition to certain the extend this definition the extend the extend this definition the extend the$$

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

Try with another statistic J(x), let $g(0) = \mathbb{E}_0[S(x)]$ ("unbiased extimator") $g(0) = \int Je^{\ell} dn$

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

 $A = Cov_{\theta}(S(x), \Delta f(\theta; x))$

Since EVI = 0

Combining these results with Cauchy-Schwarz gives us the Cramér-Rao Lower Bound or Information Lower Bound: 1-param: $V_{ar_{\theta}}(\delta) \cdot V_{ar_{\theta}}(\hat{I}(\theta;x)) \geq C_{ov_{\theta}}(\delta, \hat{I}(\theta;x))^{2}$ $\Rightarrow Var_{\theta}(\delta) = \frac{\dot{g}(\theta)}{J(\theta)}$ Multivariate: $\theta \in \mathbb{R}^d$, $g(\theta)$, $\delta(x) \in \mathbb{R}$ $Var_{\theta}(\delta) \geq \nabla_{g}(\theta)^{T} \mathcal{T}(\theta) \nabla_{g}(\theta)$ $\frac{P_{roof}:}{Var_{\theta}(\delta) \cdot a' J(\theta) a} = Var_{\theta}(\delta) Var(a' Vl(\theta))$ $\geq Cov_{\theta}(\delta, a' Vl(\theta))^{2}$ > Cove (5, a' 7/(0)) = a' Vg Vg'a, for all a ∈ IRd $\Rightarrow V_{\alpha(G)}(J) \geq \max_{\alpha \neq 0} \frac{a' \nabla_{g} \nabla_{g'} \alpha}{a' J(\theta) \alpha} \stackrel{\text{Exercise}}{=} \nabla_{g'} J(\theta)^{-1} \nabla_{g'}$

Interp: If g(0) is estimand, no unbiased estimator can have smaller variance than $\nabla g(0)' J(0)' \nabla g(0)$

Ex.: (i.i.d. sample) BEB ER \times , , , \times $\stackrel{iid}{\sim} \rho_{\theta}^{(n)}(x)$ support, finite derivative Po "regular": common $\times \sim \rho_{\Theta}(x) = \prod_{i} \rho_{\Theta}^{(i)}(x_{i})$ Let $l_i(\theta; x_i) = log \rho_0^{(i)}(x_i)$ $l(0;x) = \sum_{i} l_i(0;x_i)$ $T(0) = V_{ef}(\nabla \ell(0; X))$ $= V_{ro}(\Sigma V l(o; X_i))$ = nJ,(0) where J,(0) is Fisher into in single observation

> Lower bound scales like no (SD = n'/e for "regular" families)

Efficiency

CRLB is not nec. attainable.

We define the efficiency of an unbiased estimator as: $eff(\delta) = \frac{CRLB}{Var_{\theta}(\delta)} \left(= \frac{1/J(\theta)}{Var_{\theta}(\delta)} \text{ if } g(\theta) = \Theta e/R \right)$ $eff_{x}(\delta) \leq 1$

We say $\delta(x)$ is efficient if $eff_0(a) = 1 \ \forall \theta$

Depends on Corro (5(x), 7/(0; X)):

eff_o(δ) = $\frac{Cov_{\theta}^{2}(\delta(x), \hat{l}(\theta; x))}{Var_{\theta}(\delta) \cdot Ver_{\theta}(\hat{l}(\theta))}$ = $Corr_{\theta}^{2}(\delta, \hat{l}(\theta))$

41

J(x) is efficient \Longrightarrow $(orrac{1}{6}(5, l(0)) = 1 \forall 0$ Rarely achieved in finite samples but we can

approach it symptotically as n>00

Ex. Exponential Families
$$f_{\gamma}(x) = e^{\gamma' T(x)} - A(\gamma) h(x)$$

$$f_{\gamma}(x) = \gamma' T(x) - A(\gamma) h(x)$$

$$f_{\gamma}(x) = T(x) - \nabla A(\gamma) + \log h(x)$$

$$f_{\gamma}(x) = T(x) - \nabla A(\gamma)$$

$$f_{\gamma}(x) = T(x) - \nabla A(\gamma)$$

$$f_{\gamma}(x) = T(x) - \nabla A(\gamma)$$

$$f_{\gamma}(x) = \nabla^{2}A(\gamma)$$

$$f_{\gamma}(x) = -\nabla^{2}A(\gamma)$$

$$f_{\gamma}(x) = -\nabla^{2}A(\gamma)$$

$$f_{\gamma}(x) = -\nabla^{2}A(\gamma)$$

$$f_{\gamma}(x) = \nabla^{2}A(\gamma)$$

$$f_{\gamma}(x) = \nabla$$

Curved family:
$$\rho(x) = e^{\gamma(\theta)' T(x) - \beta(\theta)} h(x), \quad \theta \in \mathbb{R}$$

$$\beta(\theta) = A(\gamma(\theta))$$

$$\lambda(\theta; x) = \gamma(\theta)' T(x) - \beta(\theta) + \log h(x)$$

$$\lambda(\theta; x) = \gamma(\theta)' T(x) - \gamma(\theta)' \gamma \lambda(\gamma(\theta))$$

$$= \gamma(\theta)' (T(x) - \gamma A(\gamma(\theta)))$$

$$= \gamma(\theta)' (T(x) - E_{\theta} T(x))$$

