Outline

1) Score function
2) Fisher information
3) Cramér-Rao Lower Bound
4) Examples
Motivation: Tangent family

\[ p_{\theta}(x) = e^{\gamma(\theta)'T(x) - A(\gamma(\theta))} \quad h(x) \quad \gamma: \mathbb{R} \to \mathbb{R}^2 \]

\[ \Xi = \{ \gamma(\theta): \theta \in \mathbb{R}^2 \} \]

\[ \eta(\theta) = \eta_0 + \theta \delta \]

\[ \delta' T(x) \text{ complete suff.} \]

\[ \gamma(\theta) \text{ non-linear} \]

\[ T(x) \text{ minimal} \]

\[ \gamma(\theta_0) = \frac{d\gamma}{d\theta}(\theta_0) \]

\[ \Xi = \{ \gamma(\theta_0) + \varepsilon \gamma'(\theta_0): \varepsilon \in \mathbb{R} \} \]

\[ q_\varepsilon(x) = e^{(\gamma(\theta_0) + \varepsilon \gamma'(\theta_0))'T(x) - A(\cdots) - h(x)} \]

\[ = e^{\varepsilon \gamma'(\theta_0)'(T(x) - \mathbb{E}T)} - k(x) \]

Complete sufficient for tangent family at \( \theta_0 \)

Called Score function
Score function

Assume \( \mathcal{P} \) has densities \( p_\theta \) wrt \( \mu \), \( \Theta \subset \mathbb{R}^d \)
Common support: \( \{ x : p_\theta(x) > 0 \} \) same \( \forall \theta \)
Recall \( l(\theta; x) = \log p_\theta(x) \),
Thought of as random function of \( \Theta \)

**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":
\[
p_{\theta + \eta}(x) = e^{l(\theta + \eta; x)} p_\theta(x) = e^{\eta^T \nabla l(\theta; x)} p_\theta(x) \quad \text{for } \eta \approx 0
\]

**Differential identities** (assuming enough regularity)
\[
1 = \int x e^{l(\theta; x)} \, d\mu(x)
\]
\[
\frac{\partial}{\partial \theta_j} \Rightarrow 0 = \int \frac{\partial}{\partial \theta_j} l(\theta; x) \, e^{l(\theta; x)} \, d\mu(x)
\]
\[
\Rightarrow \quad E_\theta \left[ \nabla l(\theta; x) \right] = 0
\]

\* Only true if these are the same value of \( \theta \)!
\[ \frac{\partial}{\partial \theta_k} = 0 = \int \left( \frac{\partial^2 l}{\partial \theta_j \partial \theta_k} + \frac{\partial l}{\partial \theta_j} \cdot \frac{\partial l}{\partial \theta_k} \right) e^l \, dm \]

\[ = E_\theta \left[ \frac{\partial^2 l}{\partial \theta_j \partial \theta_k} \right] + E_\theta \left[ \frac{\partial l}{\partial \theta_j} \frac{\partial l}{\partial \theta_k} \right] \]

\[ \Rightarrow \quad \text{Var}_\theta \left[ \nabla l(\theta; x) \right] = E_\theta \left[ -\nabla^2 l(\theta; x) \right] \]

\[ J(\theta) = \begin{cases} \text{same } \theta & \text{same } \theta \end{cases} \]

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

\[ g(\theta) = E_\theta [\delta(X)] \quad \text{("unbiased estimator")} \]

\[ g(\theta) = \int \delta e^l \, dm \]

\[ \Rightarrow \quad \nabla g(\theta) = \int \delta \nabla l e^l \, dm = E_\theta [\delta(X) \nabla l(\theta; x)] \]

\[ \Rightarrow \quad \nabla g(\theta) = E_\theta [\delta(X) \nabla l(\theta; x)] = \text{Cov}_\theta (\delta(X), \nabla l(\theta; x)) \]

Since \( E l = 0 \)
Combining these results with Cauchy–Schwarz gives us the Cramér–Rao Lower Bound or Information Lower Bound:

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

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

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

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

\[ \text{Proof:} \]

\( \text{Var}_\theta(\delta) \cdot a' J(\theta) a = \text{Var}_\theta(\delta) \text{Var}(a' V l(\theta)) \)

\( \geq \text{Cov}_\theta(\delta, a' V l(\theta))^2 \)

\( = a' \nabla g \nabla g a, \text{ for all } a \in \mathbb{R}^d \)

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

\( \nabla g \quad J(\theta)^{-1} \quad \nabla g \)

Interp: If \( g(\theta) \) is estimand, no unbiased estimator can have smaller variance than \( \nabla g(\theta)' J(\theta)^{-1} \nabla g(\theta) \)
**Example:** (i.i.d. sample)
\[ X_1, \ldots, X_n \overset{i.i.d.}{\sim} \rho_\theta^{(i)}(x) \quad \Theta \in \Theta \subseteq \mathbb{R}^d \]

\( \rho_\theta \) "regular": common support, finite derivative w.r.t. \( \Theta \)

\[ X \sim \rho_\theta(x) = \prod_i \rho_\theta^{(i)}(x_i) \]

Let \( l_1(\theta; x_i) = \log \rho_\theta^{(i)}(x_i) \)
\[ l(\theta; x) = \sum_i l_i(\theta; x_i) \]
\[ \bar{J}(\theta) = \text{Var}_\theta(\nabla l(\theta; x)) \]
\[ = \text{Var}_\theta(\sum_i \nabla l_i(\theta; x_i)) \]
\[ = n \bar{J}_1(\theta) \quad \text{where } \bar{J}_1(\theta) \text{ is Fisher info} \]

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

CRLB is not necessarily attainable.

We define the efficiency of an unbiased estimator as:

\[ \text{eff}_T(\delta) = \frac{\text{CRLB}}{\text{Var}_\theta(\delta)} \left( = \frac{1}{\text{Var}_\theta(\delta)} \text{ if } g(\theta) = \Theta e^R \right) \]

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

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

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

\[ \text{eff}_\theta(\delta) = \frac{\text{Cov}_\theta^2(\delta(X), \nabla \ell(\theta; X))}{\text{Var}_\theta(\delta) \cdot \text{Var}_\theta(\nabla \ell(\theta))} \]

\[ = \text{Corr}_\theta^2(\delta, \nabla \ell(\theta)) \leq 1 \]

\( \delta(X) \) is efficient \( \iff \text{Corr}_\theta(\delta, \nabla \ell(\theta)) = 1 \ \forall \theta \)

Rarely achieved in finite samples but we can approach it asymptotically as \( n \to \infty \)
Ex. Exponential Families

\[ \rho_\eta(x) = e^{\eta^T T(x) - A(\eta)} h(x) \]

\[ l(\eta; x) = \eta^T T(x) - A(\eta) + \log h(x) \]

\[ \nabla l(\eta; x) = T(x) - \nabla A(\eta) \]

\[ = T(x) - \mathbb{E}_\eta T(x) \]

\[ \text{Var}_\eta(\nabla l(\eta)) = \text{Var}_\eta(T(x)) = \nabla^2 A(\eta) \]

\[ \nabla^2 l(\eta; x) = -\nabla^2 A(\eta) \]

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

So any unbiased est. of \( \eta \) has

\[ \text{Var}_\eta(\hat{\eta}) \geq \nabla^2 A(\eta)^{-1} \]
Curved family:  \[ p_\theta(x) = e^{\gamma(\theta)'T(x) - B(\theta)} h(x), \quad \theta \in \mathbb{R} \]

\[ B(\theta) = A(\gamma(\theta)) \]

\[ l(\theta; x) = \gamma(\theta)' T(x) - B(\theta) + \log h(x) \]

\[ \dot{l}(\theta; x) = \dot{\gamma}(\theta)' T(x) - \dot{\gamma}(\theta)' \nabla_{\gamma} A(\gamma(\theta)) \]

\[ = \dot{\gamma}(\theta)' \left( T(x) - \nabla_{\gamma} A(\gamma(\theta)) \right) \]

\[ = \dot{\gamma}(\theta)' \left( T(x) - \mathbb{E}_\theta T(x) \right) \]

\[ \Rightarrow \dot{\gamma}(\theta)' T(x) \text{ is "locally complete suff. stat."} \]