Outline

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

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

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

Curved Family

\[ \eta(\theta) \text{ nonlinear} \]

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

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

\[ \eta_0 \]

\[ \delta \]

Complete Sufficient

Score function

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

\[ = e^{\varepsilon \dot{\eta}(\theta_0)'(T(x) - \bar{E}T) - B(\varepsilon)} k(x) \]

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

Called Score function
**Score function**

Assume $\mathcal{F}$ has densities $p_\theta$ wrt $\mu$, $\Theta \subseteq \mathbb{R}^d$

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

Recall $l(\theta; x) = \log p_\theta(x)$, thought of as random function of $\theta$

**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":

$$p_{\theta + \varepsilon}(x) = e^{l(\theta + \varepsilon; x)}$$

$$\approx e^{\varepsilon' \nabla l(\theta; x)} p_\theta(x) \text{ for } \varepsilon \approx 0$$

**Differential identities:** (assuming enough regularity)

$$1 = \int x e^{l(\theta; x)} \, d\mu(x)$$

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

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

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

\[
= \mathbb{E}_\theta \left[ \frac{\partial l}{\partial \theta} \right] + \mathbb{E}_\theta \left[ \frac{\partial l}{\partial \theta_i} \frac{\partial l}{\partial \theta_k} \right]
\]

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

\[
J(\theta) = \text{same } \theta \quad \text{same } \theta
\]

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

\[
g(\theta) = \mathbb{E}_\theta [ \delta(X) ] \quad \text{ (“unbiased estimator”)}
\]

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

\[
\Rightarrow \quad \nabla g(\theta) = \int \delta \nabla l e^l \, dm = \mathbb{E}_\theta \left[ \delta(X) \nabla l(\theta; x) \right]
\]

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

Since \( \mathbb{E} \nabla l = 0 \)
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: \( \Theta \in \mathbb{R}^d, g(\theta), \delta(x) \in \mathbb{R} \)

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

**Proof:**

\[ \text{Var}_\theta(\delta) \cdot a' J(\theta) a = \text{Var}_\theta(\delta) \text{Var}(a' \ell(\theta)) \]

\[ \geq \text{Cov}_\theta(\delta, a' \ell(\theta))^2 \]

\[ = a' g' g'a , \text{ for all } a \in \mathbb{R}^d \]

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

**Exercise**

\[ \Rightarrow g(\theta)' J(\theta)^{-1} g(\theta) \]

**Interpretation:** If \( g(\theta) \) is an estimand, no unbiased estimator can have smaller variance than \( g(\theta)' J(\theta)^{-1} g(\theta) \)
Ex. (i.i.d. sample)

\[ X_1, \ldots, X_n \overset{i.i.d.}{\sim} p_\theta(x) \quad \Theta \in \Theta \subseteq \mathbb{R}^d \]

\( p_\theta \) "regular": common support, finite derivative wrt \( \Theta \)

\[ X \sim p_\theta(x) = \prod_i p_\theta(x_i) \]

Let \( l_i(\theta; x_i) = \log p_\theta(x_i) \)

\[ l(\theta; x) = \sum_i l_i(\theta; x_i) \]

\[ J(\theta) = \text{Var}_\theta(\nabla l(\theta; x)) \]

\[ = \text{Var}_\theta(\sum_i \nabla l_i(\theta; x_i)) \]

\[ = n J_1(\theta) \quad \text{where } J_1(\theta) \text{ is Fisher info} \]

\[ \Rightarrow \text{Lower bound scales like } n^{-1} \quad (\text{s.e.} = n^{-1/2} \text{ for "regular" families}) \]
Efficiency

CRLB is not nec. attainable.

We define the efficiency of an unbiased estimator as:

\[
\text{eff}_\theta(\delta) = \frac{\text{CRLB}}{\text{Var}_\theta(\delta)} = \frac{1/\mathcal{I}(\theta)}{\text{Var}_\theta(\delta)} \quad \text{if } g(\theta) = \theta \in \mathbb{R}
\]

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

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

Depends on \( \text{Corr}_\theta(\delta(X), \mathbb{E}[\ell(\theta; X)]) \):

\[
\text{eff}_\theta(\delta) = \frac{\text{Cov}_\theta(\delta(X), \mathbb{E}[\ell(\theta; X)])}{\text{Var}_\theta(\delta) \cdot \text{Var}_\theta(\mathbb{E}[\ell(\theta)])} = \text{Corr}_\theta(\delta, \mathbb{E}[\ell(\theta)])
\]

\[
\leq 1
\]

\( \delta(X) \) is \underline{efficient} \( \iff \text{Corr}_\theta^2(\delta, \mathbb{E}[\ell(\theta)]) = 1 \ \forall \theta \)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

\[ = \dot{\gamma}(\theta)' (T(x) - E_\theta T(x)) \]

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