Theorem: Let A1-2-3 hold, and assume that

$$\frac{\partial^2}{\partial\vartheta_j\partial\vartheta_i}\int \log\frac{p_\vartheta(x)}{p_{\vartheta'}(x)}\, p_\vartheta(x)\,\mu(dx) = $$

$$= \int \frac{\partial^2}{\partial\vartheta_j\partial\vartheta_i}\left(\log\frac{p_\vartheta(x)}{p_{\vartheta'}(x)}\, p_\vartheta(x)\right)\mu(dx).$$

Then,

$$I(\vartheta)_{ij} = \frac{\partial^2}{\partial\vartheta_i'\partial\vartheta_j'} D_{KL}(\vartheta \,\|\, \vartheta')\Big|_{\vartheta=\vartheta'}$$

Note: ① At $\vartheta = \vartheta'$ KL is 0

② The second derivative tells us how fast the KL divergence grows as we move away from ϑ.

③ Fisher information is the local curvature of KL divergence at ϑ.

Proof:

$$D_{KL}(\vartheta|\vartheta') = \int [\log p_\vartheta(x) - \underbrace{\log p_{\vartheta'}(x)}_{\text{depends on } \vartheta'}]\, p_\vartheta(x)\, \mu(dx)$$

$$\frac{\partial}{\partial \vartheta_j'} D_{KL}(\vartheta|\vartheta') = -\int \frac{\partial}{\partial \vartheta_j'} \log p_{\vartheta'}(x) \cdot p_\vartheta(x)\, \mu(dx)$$

$$\frac{\partial}{\partial \vartheta_i' \partial \vartheta_j'} D_{KL}(\vartheta|\vartheta') = -\int \frac{\partial^2}{\partial \vartheta_i' \partial \vartheta_j'} \log p_{\vartheta'}(x) \cdot p_\vartheta(x)\, \mu(dx)$$

$$\frac{\partial}{\partial \vartheta_i' \partial \vartheta_j'} D_{KL}(\vartheta|\vartheta')\Big|_{\vartheta=\vartheta'} = -\int \frac{\partial^2}{\partial \vartheta_i' \partial \vartheta_j'} \log p_{\vartheta'}(x)\Big|_{\vartheta=\vartheta'} p_\vartheta(x)\, \mu(dx)$$

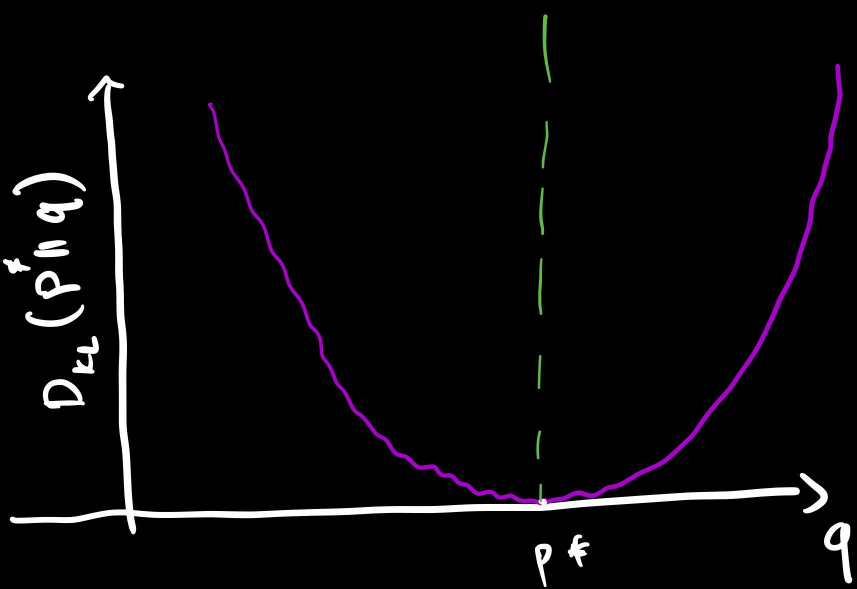
By Fisher Information matrix.

$$I(\theta)_{ij} = E_\theta\left[\frac{\partial}{\partial \vartheta_i} \log p_\vartheta(x) \cdot \frac{\partial}{\partial \vartheta_j} \log p_\theta(x)\right] = -E_\theta\left(\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \log p_\vartheta(x)\right)$$

So we conclude that:

$$I(\theta)_{ij} = \frac{\partial^2}{\partial \vartheta_i' \partial \vartheta_j'} D_{KL}(\vartheta \,|\, \vartheta')\Big|_{\vartheta' = \vartheta}$$

Ex:



p^* fixed, true

q varies

Asymptotic Statistics:

Q: what happens when $n \to \infty$?

Consider a family $\{P_\theta\}_{\theta \in \Theta}$ of probability measures on $(X, \mathcal{G})$ and the statistical model

$$(\mathcal{X}, \mathcal{F}, \{P_\theta\}_{\theta \in \Theta}) = (X^N, \mathcal{G}^{\otimes N}, \{P_\theta^{\otimes n}\}_{\theta \in \Theta})$$

allowing us to define the sequence:

$$\{\underline{X}_n\}_{n=1}^{\infty} = \{(X_1, \ldots, X_n)\}_{n=1}^{\infty}$$

Theorem (Ionescu-Tulcea): for $\mathcal{X} = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, fixed P_ϑ, there exists unique $\mathbb{P}_\vartheta = P_\vartheta^{\otimes n}$ with

$$\forall \{A_j \in \mathcal{F}\}_{j=1}^{n} \quad \mathbb{P}_\vartheta(A_1 \times \ldots \times A_n \times \mathbb{R} \ldots) = P_\vartheta(A_1) \ldots P_\vartheta(A_n)$$

Notation: Let $g : \Theta \longrightarrow \mathbb{R}^p$,

$$\Gamma = \{g(\vartheta) : \vartheta \in \Theta\}.$$

Definition: We call $\{T_n\}_{n=1}^{\infty}$ for $T_n : \mathcal{X}^{\otimes n} \longrightarrow \Gamma$ a sequence of estimators of $g(\vartheta)$ if T_n is an estimator of $g(\vartheta)$ $\forall n \in \mathbb{N}$.

Example: $X_j \sim \mathrm{Pois}(\vartheta)$ $g(\vartheta) = \vartheta$

$$T_n(\underline{X}_n) = \frac{1}{n} \sum_{j=1}^{n} X_j$$

Definition: An estimator T_n of $g(\vartheta) \in \mathbb{R}^p$ is an M-estimator if there exists $\rho : \mathcal{X} \times \mathbb{R}^p \longrightarrow \mathbb{R}$ such that

$$T_n(\underline{X}) = \arg\min_{\gamma} \sum_{j=1}^{n} \rho(X_j, \gamma)$$

Ex (MLE): if $g(\vartheta) = \vartheta$ $\rho(x, \vartheta) = -\log p_\vartheta(x)$

we use that MLE is a special case of M-Estimator.

Ex (Huber) consider the loss

$$\rho_k(x, \gamma) = \begin{cases} \frac{1}{2} \| x_j - \gamma \|^2 & \| x - \gamma \| < k \\ k \| x_j - \gamma \|^2 & \| x - \gamma \| \geq k \end{cases}$$

Then the huber estimator

$$T_H(x) = \arg\min_{\mu} \sum_{j=1}^{n} \rho_k(x_j, \gamma)$$

Note that if $\rho(x, \cdot) \in C^1$ for all $x \in X$ we find the minimum by differentiation. $T_n(\underline{X})$ solves:

$$\sum_{j=1}^{n} \underbrace{D_\theta \rho(x_j, \theta)}_{\psi(x,\theta)} = 0$$

Definition: An estimator T_n of $g(\vartheta) \in \mathbb{R}^p$ is a Z-estimator if there exists $\psi : X \times \mathbb{R}^p \to \mathbb{R}$ such that $T_n(\underline{X})$ solves:

$$\sum_{j=1}^{n} \psi(x_j, T_n(\underline{X})) = 0$$

we would like to have

$$T_n \longrightarrow g(\vartheta) \text{ as } n \to \infty.$$

<u>Definition</u> : a sequence of r.v $\{T_n\}$ converges :

- in probability towards the r.v. T^* ($T_n \xrightarrow[n\to\infty]{\mathbb{P}} T_*$) if

$$\lim_{n\to\infty} \mathbb{P}(|T_n - T| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

- almost surely towards the r.v. T^* ($T_n \xrightarrow[n\to\infty]{a.s} T_*$) if

$$\mathbb{P}\left(\lim_{n\to\infty} T_n = T_*\right) = 1$$

<u>Note</u>: $T_n \xrightarrow{a.s} T_* \overset{\text{fatou lemma}}{\Longrightarrow} T_n \longrightarrow T_*$

<u>Definition</u>: A sequence T_n of estimators of $g(\theta)$ is called <u>consistent</u> / <u>strongly consistent</u> if

$T_n \xrightarrow[n\to\infty]{\mathbb{P}} g(\theta)$ / $T_n \xrightarrow[n\to\infty]{a.s} g(\theta) \quad \forall \theta \in \Theta$

Ex (Poisson) $X_j \overset{i.i.d}{\sim} \mathrm{Pois}(\vartheta)$

$g(\vartheta) = \vartheta \qquad T_n(\underline{X}) = \frac{1}{n}\sum_j X_j$

Prove that T_n is consistent.

chebyshev

$$\mathbb{P}\left[\, \| T_n(x) - \vartheta \| > \varepsilon \right] \le \frac{\mathrm{Var}_\vartheta(T_n)}{\varepsilon^2} \xrightarrow{n \to \infty} 0$$

$\mathrm{Var}_\vartheta(T_n) = \frac{\vartheta}{n}$

$\Rightarrow T_n$ is consistent.

By the law of large numbers we could expect that

$$\frac{1}{n}\sum \rho(x_j, \gamma) \xrightarrow{P/a.s} E_\vartheta(\rho(x, \gamma))$$

and therefore $T_n \xrightarrow[n \to \infty]{} \arg\min_\gamma E_\vartheta(\rho(x,\gamma)) \quad \forall \vartheta \in \Theta$

Consistency of MLE estimators:

Theorem: Let $X_1, \ldots, X_n \overset{i.i.d}{\sim} P_\theta(x)$ for $\vartheta \in \Theta \subseteq \mathbb{R}^k$ and let:

(A1) identifiability

$\forall \vartheta \neq \vartheta_0, \; P_\vartheta(x) \neq P_{\vartheta_0}(x)$ on a set of positive measures

(A2) Parameter space $\Theta \subseteq \mathbb{R}^k$ compact.

(A3) The map $\vartheta \longrightarrow \log P_\vartheta(x)$ is continuous for all x.

(A4) There exists a function $M(x) \geq 0$

s.t. $|\log P_\vartheta(x)| \leq M(x) \;\; \forall \vartheta \in \Theta$.

and $\mathbb{E}_{\vartheta_0}[M(X)] < \infty$

(A5) For each $\vartheta \in \Theta$, $X_1, \ldots, X_n$ are i.i.d. from P_{ϑ_0}.

where ϑ_0 is the true parameter

Then the MLE estimator T_n is consistent i.e.:

$$T_n \xrightarrow[n\to\infty]{\mathbb{P}} \vartheta_0$$

Proof: we want to show that

$$T_n = \arg\max_{\vartheta \in \Theta} \ell_n(\vartheta) \xrightarrow{\mathbb{P}} \vartheta_0$$

Let $\ell(\vartheta) = \mathbb{E}_{\vartheta_0}\left[\log p_\vartheta(x)\right] = \int \log p_\vartheta(x)\, p_{\vartheta_0}(x)\, \mu(dx)$

By (A1) $\ell(\vartheta)$ is uniquely maximized at $\vartheta = \vartheta_0$

By LLN (Pointwise convergence) for fixed $\vartheta \in \Theta$ the log likelihood satisfies:

$$\frac{1}{n}\ell_n(\vartheta) = \frac{1}{n}\sum \log p_\vartheta(x_i) \xrightarrow{a.s.} \ell(\vartheta)$$

$\log p_\vartheta(x)$ is integrable under ϑ_0 (A4)

Uniform Convergence over Θ.

By A2 - A3 - A4 and the Uniform law of large numbers ULLN

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} l_n(\theta) - l(\theta) \right| \xrightarrow{\mathbb{P}} 0$$

From argmax consistency

$$T_n = \arg\max_{\vartheta} \frac{1}{n} l_n(\vartheta)$$

$$\vartheta_0 = \arg\max_{\vartheta} l(\vartheta)$$

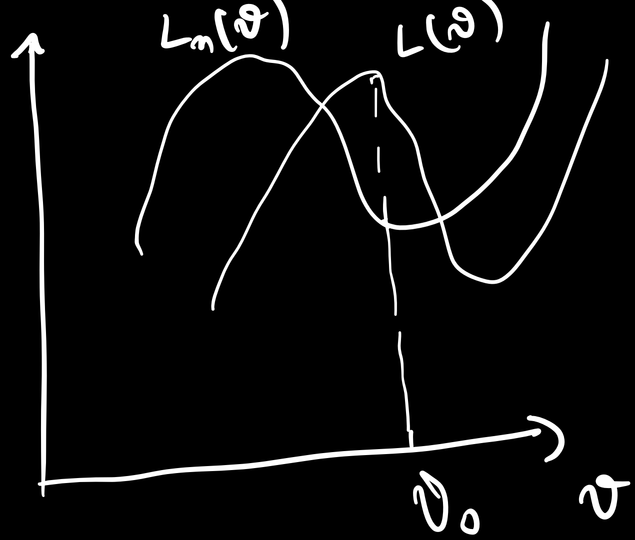
Then if Θ is compact

$$\frac{l_n(\vartheta)}{n} \longrightarrow l(\vartheta)$$

$l(\vartheta)$ has a unique maximizer at ϑ_0.

$$\Rightarrow \quad T_n \xrightarrow{\mathbb{P}} \vartheta_0 .$$

$\blacksquare$



$L_n(\vartheta)$
$L(\vartheta)$
ϑ_0
ϑ