9.Famílies Exponencials, Suficiència i Propietats Assimptòtiques dels MLE

Estadística

Grau en Matemàtiques

Josep A. Sanchez
Dept. Estadística i I.O.(UPC)

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Família exponencial

- Definició: Una família de densitats de probabilitat amb vector de paràmetres θ s'anomena **família exponencial** si es pot escriure com:

$$f(x|\theta) = h(x)c(\theta)exp\left[\sum_{i=1}^{k} w_i(\theta)t_i(x)\right]$$

on,

- $h(x) \geq 0$ i $t_i(x) \quad \forall i = 1, \ldots, k$ are funcions que depenen només de x i no de θ
- $c(\theta) \geq 0$ i $w_i(\theta) \quad \forall i = 1, \ldots, k$ are funcions que depenen només de θ i no de x

Famílies exponencials

- Les distribucions de la família exponencial tenen bones propietats matemàtiques i estadístiques
- La família exponencial compleixen les hipòtesis del teorema de Cràmer-Rao
- Moltes de les distribucions més habituals (Normal, Poisson, Bernouilli, Gamma,...) pertanyen a aquesta família

Exemple: Distribució Poisson

$$P(X = x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!} \quad \lambda \in \mathbb{R}^+$$

$$P(X = x|\lambda) = \frac{1}{x!}e^{-\lambda}e^{x\log(\lambda)}\mathbb{I}_{\mathbb{N}}(x)$$

$$= h(x)c(\lambda)\exp\left[w(\lambda)t(x)\right]$$

on,

$$h(x) = \frac{1}{x!}\mathbb{I}_{\mathbb{N}}(x) \qquad c(\lambda) = e^{-\lambda}$$
$$w(\lambda) = \log(\lambda) \qquad t(x) = x$$

Exemple: Distribució Normal

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \mu \in \mathbb{R} \quad \sigma^2 \in \mathbb{R}^+$$

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathbb{I}_{\mathbb{R}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2-2\mu x+\mu^2}{2\sigma^2}\right) \mathbb{I}_{\mathbb{R}}(x)$$

$$= \frac{1}{\sqrt{2\pi}} \mathbb{I}_{\mathbb{R}}(x) \frac{1}{\sqrt{\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(\frac{\mu}{\sigma^2} x - \frac{1}{2\sigma^2} x^2\right) =$$

$$= h(x) c(\mu,\sigma^2) \exp\left[w_1(\mu,\sigma^2) t_1(x) + w_2(\mu,\sigma^2) t_2(x)\right]$$

on,

$$h(x) = \frac{1}{\sqrt{2\pi}} \mathbb{I}_{\mathbb{R}}(x) \qquad c(\mu,\sigma^2) = \frac{1}{\sqrt{\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right)$$

$$w_1(\mu,\sigma^2) = \frac{\mu}{\sigma^2} \quad t_1(x) = x \qquad w_2(\mu,\sigma^2) = -\frac{1}{2\sigma^2} \quad t_2(x) = x^2$$

Contraexemple: Distribució exponencial trasladada

$$f(x|\lambda, \mu) = \lambda e^{-\lambda(x-\mu)} \qquad \mu \in \mathbb{R} \quad \lambda \in \mathbb{R}^+ \quad x > \mu$$

$$f(x|\lambda, \mu) = \lambda e^{\lambda\mu} e^{-\lambda x} \mathbb{I}_{(\mu,+\infty)}(x)$$

$$w(\mu, \lambda) = -\lambda \qquad t(x) = x$$

$$c(\mu, \lambda) = e^{-\lambda\mu}? \qquad h(x) =?$$

Suficiència

- Un estadístic és **suficient** si captura tota la informació que la mostra poseeix respecte al paràmetre θ
- Si $T(\underset{\sim}{X})$ és un estadístic suficient i tenim dues mostres $\underset{\sim}{X}$ i $\underset{\sim}{Y}$, tal que $T(\underset{\sim}{X}) = T(\underset{\sim}{Y})$, llavors tota la inferència sobre θ serà la mateixa per ambdues mostres.
- La inferència que fem sobre el paràmetre θ depen de la mostra a través de l'estadístic suficient

Un estadístic $T(\underset{\sim}{X})$ és suficient pel paràmetre θ si la distribució condicional de la mostra $\underset{\sim}{X}$ donat el valor de $T(\underset{\sim}{X})$ no depen del paràmetre θ

- $P(\underset{\sim}{X} = \underset{\sim}{x} | T(\underset{\sim}{X}) = T(\underset{\sim}{x}))$ no és funció de θ

$$P(\underset{\sim}{X} = \underset{\sim}{x} | T(\underset{\sim}{X}) = T(\underset{\sim}{x})) = \frac{P(\underset{\sim}{X} = \underset{\sim}{x} \cap T(\underset{\sim}{X}) = T(\underset{\sim}{x}))}{P(T(\underset{\sim}{X}) = T(\underset{\sim}{x}))} =$$

$$= \frac{P(\underset{\sim}{X} = \underset{\sim}{x})}{P(T(\underset{\sim}{X}) = T(\underset{\sim}{x}))} = \frac{p(\underset{\sim}{x}|\theta)}{q(T(\underset{\sim}{x})|\theta)}$$

Teorema: Si $p(\underset{\sim}{x}|\theta)$ és la densitat conjunta de $\underset{\sim}{X}$ and $q(t|theta)$ és la densitat de $T(\underset{\sim}{X})$ llavors:

$T(\underset{\sim}{X})$ és un estadístic suficient per $\theta \quad \Leftrightarrow \quad \frac{p(\underset{\sim}{x}|\theta)}{q(T(\underset{\sim}{x})|\theta)}$ és constant com a funció de θ

Exemple: Distribució Bernouilli

Sigui $\underset{\sim}{X} = (X_1, \ldots, X_n)$ m.a.s. amb $X_i \sim Bern(\theta)$ on $\theta \in [0,1]$

L'estatistic $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ és suficient pel paràmetre θ:

Notem que $T(\underset{\sim}{X}) \sim B(n, \theta)$ i sigui $T(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i = t$

$$\frac{f(\underset{\sim}{x}|\theta)}{q(T(\underset{\sim}{x})|\theta)} = \frac{\prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{\theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}}$$

Teorema de factorització

Sigui $f(\underset{\sim}{x}|\theta)$ la densitat conjunta de $\underset{\sim}{X}$

- L'estadístic $T(\underset{\sim}{X})$ és suficient per $\theta \quad \Leftrightarrow \quad$ existeixen funcions $g(t|\theta)$ i $h(x)$ tal que per tots els punts $\underset{\sim}{x}$ de l'espai mostral i per tots els punts θ de l'espai de paràmetres tenim,

$$f(\underset{\sim}{x}|\theta) = g(T(\underset{\sim}{x})|\theta)h(\underset{\sim}{x})$$

- Demostració:

Suposem $T(\underset{\sim}{X})$ és suficient, llavors escollim $g(t|\theta) = P(T(\underset{\sim}{X}) = t|\theta))$ i $h(\underset{\sim}{x}) = P(\underset{\sim}{X} = \underset{\sim}{x}|T(\underset{\sim}{X}) = T(\underset{\sim}{x}))$.

Per tant,

$$f(\underset{\sim}{x}|\theta) = P(\underset{\sim}{X} = \underset{\sim}{x} \cap T(\underset{\sim}{X}) = T(\underset{\sim}{x})) =$$

$$= P(\underset{\sim}{X} = \underset{\sim}{x}|T(\underset{\sim}{X}) = T(\underset{\sim}{x}))P(T(\underset{\sim}{X}) = T(\underset{\sim}{x})|\theta)) = h(\underset{\sim}{x})g(T(\underset{\sim}{x})|\theta)$$

- Demostració (cont.):

Penem $t = T(\underset{\sim}{x})$ i calculem la distribució de $\underset{\sim}{X}$ condicionada a t

$$P(\underset{\sim}{X} = \underset{\sim}{x} | T(\underset{\sim}{X}) = t) = \frac{P(\underset{\sim}{X} = \underset{\sim}{x})}{P(T(\underset{\sim}{X}) = t)} =$$

$$= \frac{P(\underset{\sim}{X} = \underset{\sim}{x})}{\sum_{\underset{\sim}{y}: T(\underset{\sim}{y}) = t} P(\underset{\sim}{X} = \underset{\sim}{y})} = \frac{g(t|\theta)h(\underset{\sim}{x})}{\sum_{\underset{\sim}{y}: T(\underset{\sim}{y}) = t} g(t|\theta)h(\underset{\sim}{y})} = \frac{h(\underset{\sim}{x})}{\sum_{\underset{\sim}{y}: T(\underset{\sim}{y}) = t} h(\underset{\sim}{y})}$$

que no depen de θ

Exemple: Distribució Normal amb σ^2 coneguda

Sigui $\underset{\sim}{X} = (X_1, \ldots, X_n)$ m.a.s. amb $X_i \sim N(\mu, \sigma^2)$ on $\mu \in \mathbb{R}$ i σ^2 coneguda.

$$f(\underset{\sim}{x}|\mu) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right) =$$

$$= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{2\sigma^2}\right) \exp\left(-\frac{n(\bar{X} - \mu)^2}{2\sigma^2}\right) =$$

i per tant, si considerem l'estadístic $T(\underset{\sim}{X}) = \bar{X}$, la factorització s'obté:

$$h(x) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{2\sigma^2}\right) \qquad g(t|\mu) = \exp\left(-\frac{n(t-\mu)^2}{2\sigma^2}\right)$$

Que prova que la mitjana mostral $\bar{X}$ en aquest cas és suficient pel paràmetre μ

Sigui $\underset{\sim}{X} = (X_1, \dots, X_n)$ m.a.s. amb $X_i \sim U\{1, \dots, \theta\}$ on $\theta \in \mathbb{N}$

$$P(X = x|\theta) = \frac{1}{\theta} \quad x \in \{1, \dots, \theta\}$$

$$P(\underset{\sim}{X} = \underset{\sim}{x}|\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbb{I}_{\{1,\dots,\theta\}}(x_i) = \frac{1}{\theta^n} \mathbb{I}_{\{1,\dots,\theta\}}(x_{(n)}) \prod_{i=1}^{n} \mathbb{I}_{\mathbb{N}}(x_i)$$

En aquest cas, l'estadístic suficient és el màxim de la mostra, $T(\underset{\sim}{X}) = X_{(n)}$:

$$h(x) = \prod_{i=1}^{n} \mathbb{I}_{\mathbb{N}}(x_i) \qquad g(t|\mu) = \frac{1}{\theta^n} \mathbb{I}_{\{1,\dots,\theta\}}(t)$$

Exemple: Distribución Normal biparamètrica

Sigui $\underset{\sim}{X} = (X_1, \ldots, X_n)$ m.a.s. amb $X_i \sim N(\mu, \sigma^2)$ on $\mu \in \mathbb{R}$ i $\sigma^2 \in \mathbb{R}^+$

$$f(\underset{\sim}{x}|\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right) =$$

$$= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{2\sigma^2}\right) \exp\left(-\frac{n(\bar{X} - \mu)^2}{2\sigma^2}\right) =$$

Definim els estadístics $T_1(\underset{\sim}{X}) = \bar{X}$ i $T_2(\underset{\sim}{X}) = S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$, llavors $T(\underset{\sim}{X}) = (T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X}))$ es suficient per (μ, σ^2) in the Normal model

$$f(\underset{\sim}{x}|\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{(n-1)S^2}{2\sigma^2}\right) \exp\left(-\frac{n(\bar{X} - \mu)^2}{2\sigma^2}\right) = g(t_1, t_2|\mu, \sigma^2)$$

$$h(x) = 1$$

Consistència

Definició: Una seqüència de estimadors $W_n = W_n(X_1, \ldots, X_n)$ és un estimador **consistent** del paràmetre θ si per cada $\epsilon > 0$ i cada $\theta \in \Theta$,

$$\lim_{n \to \infty} P(|W_n - \theta| > \epsilon) = 0$$

Sigui $B_\theta(W)$ el biaix de l'estimador W pel paràmetre θ

- Teorema: Si W_n és una seqüència d'estimadors de θ que satisfa:
  - $\lim_{n \to \infty} V(W_n) = 0$
  - $\lim_{n \to \infty} B_\theta(W_n) = 0$

  llavors, W_n és una seqüència d'estimadors consistent

La intrepretació pràctica és que quan augmenta la mida mostral léstimador s'apropa al paràmetre amb alta probabilitat.

- Demostració: Hem vist que

$$E\left[(W_n-\theta)^2\right]=V(W_n)+B_\theta(W_n)^2$$

Per tant, si es compleixen les condicions,

$$\lim_{n\to\infty}E\left[(W_n-\theta)^2\right]=0$$

Fent servir la desigualtat de Chevyshev tenim

$$P(|W_n-\theta|>\epsilon)=P((W_n-\theta)^2\geq\epsilon^2)\leq\frac{E\left[(W_n-\theta)^2\right]}{\epsilon^2}\quad\forall\epsilon>0,\forall\theta\in\Theta$$

i per tant la seqüència compleix la propietat de consistència,

$$\lim_{n\to\infty}P(|W_n-\theta|>\epsilon)=0$$

Nota: Els estimadors ML sota condicions de regularitat són consistents

Propietats assimptòtiques dels estimadors ML

Sigui $X_1, \ldots, X_n$ una m.a.s amb densitat $f(x|\theta)$ i $\hat{\theta}$ l'estimador pel mètode de màxima versemblança de θ. Sota condicions de regularitat,

$$\hat{\theta} \to N\left(\theta, \frac{1}{I_n(\theta)}\right)$$

on $1/I_n(\theta)$ és la cota de Cràmer-Rao del model.

Més correctament,

$$\sqrt{nI_X(\theta)}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{D} N(0,1)$$

Per tant, els estimadors ML, asimptòticament són:

- Sense biaix
- Eficients
- Amb distribució Normal

Si necessitem estimar una transformació $\tau(\theta)$ de θ tenim:

Sigui $X_1, \ldots, X_n$ una m.a.s amb densitat $f(x|\theta)$ i $\tau(\hat{\theta})$ l'estimador pel mètode de màxima versemblança de $\tau(\theta)$. Sota condicions de regularitat,

$$\tau(\hat{\theta}) \rightarrow N\left(\tau(\theta), \frac{\tau'(\theta)^2}{I_n(\theta)}\right)$$

on $\tau'(\theta)^2/I_n(\theta)$ és la cota de Cràmer-Rao del model.

Propietats assimptòtiques dels estimadors ML

Per fer inferència fent servir aquest resultat, enlloc de la **informació (esperada) de Fisher** que sol ser funció del paràmetre θ desconegut, es fa servir la **informació observada de Fisher** en el valor de l'estimador $\hat{\theta}_n$

$$IObs(\hat{\theta}_n) = -\frac{\partial^2}{\partial\theta^2} \log L(\theta; \underset{\sim}{X})\Big|_{\theta=\hat{\theta}_n}$$

I per tant, els resultats queden:

$$\sqrt{IObs(\hat{\theta}_n)}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{D} N(0,1)$$

$$\sqrt{\frac{IObs(\hat{\theta}_n)}{\tau'(\hat{\theta}_n)^2}}(\tau(\hat{\theta}_n) - \tau(\theta)) \xrightarrow[n\to\infty]{D} N(0,1)$$

Exemple: estimació del *odds* i la seva variància

Sigui $\underset{\sim}{X} = (X_1, \dots, X_n)$ m.a.s. amb $X_i \sim Bern(p)$ on $p \in [0,1]$ i volem estimar l'*odds* ($\tau(p) = \frac{p}{1-p}$)

L'estimador de màxima versemblança (MLE) de p és $\hat{p} = \bar{X}$ i la variància de l'estimador és $V(\hat{p}) = \frac{p(1-p)}{n}$

El principi d'invariància del MLE ens garanteix que l'estimador ML de $\tau(p)$ és $\tau(\hat{p}) = \frac{\hat{p}}{1-\hat{p}}$ i, intuitivament, $\hat{V}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n}$

$$L(p; \underset{\sim}{X}) = \prod_{i=1}^{n} p^{x_i}(1-p)^{(1-x_i)} = p^{\sum_{i=1}^n x_i}(1-p)^{(n-\sum_{i=1}^n x_i)}$$

$$\log L(p; \underset{\sim}{X}) = \log(p)\sum_{i=1}^{n} x_i + log(1-p)(n - \sum_{i=1}^{n} x_i)$$

$$\frac{\partial}{\partial p}\log L(p; \underset{\sim}{X}) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p}$$

$$\frac{\partial^2}{\partial p^2}\log L(p; \underset{\sim}{X}) = -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n - \sum_{i=1}^n x_i}{(1-p)^2}$$

Exemple: estimació del *odds* i la seva variància

Si calculem la Informació de Fisher continguda en la mostra, pel paràmetre p, aquesta expressió és funció del paràmetre p:

$$I_{\underset{\sim}{X}}(p) = E\left[-\frac{\partial^2}{\partial p^2}\log L(p;\underset{\sim}{X})\right] = \frac{n}{p(1-p)}$$

Però si fem servir la Informació de Fisher observada, hem de calcular:

$$IObs(\hat{p}) = -\frac{\partial^2}{\partial p^2}\log L(p;\underset{\sim}{X})\Big|_{p=\hat{p}} = \frac{\sum_{i=1}^n x_i}{\hat{p}^2} - \frac{n-\sum_{i=1}^n x_i}{(1-\hat{p})^2} = \frac{n}{\hat{p}} - \frac{n}{1-\hat{p}} = \frac{n}{\hat{p}(1-\hat{p})}$$

Per tant, la variància aproximada de l'estimador $\hat{p}$ és $\hat{V}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n}$

La distribució assimptòtica és: $\sqrt{\frac{n}{\hat{p}(1-\hat{p})}}(\hat{p}-p) \approx N(0,1)$

Exemple: estimació del *odds* i la seva variància

Volem estimar $\tau(p) = \frac{p}{1-p}$

Estimador puntual: $\tau(\hat{p}) = \frac{\hat{p}}{1-\hat{p}} = \frac{\sum_{i=1}^{n} x_i}{n-\sum_{i=1}^{n} x_i}$

$\frac{\partial}{\partial p}\tau(p) = \frac{1-p-(-1)p}{(1-p)^2} = \frac{1}{(1-p)^2} = \tau'(p)$

$$\hat{V}(\tau(\hat{p})) = \hat{V}\left(\frac{\hat{p}}{1-\hat{p}}\right) = \frac{\tau'(\hat{p})^2}{IObs(\hat{p})} = \frac{\frac{1}{(1-\hat{p})^4}}{\frac{n}{\hat{p}(1-\hat{p})}} = \frac{\hat{p}}{n(1-\hat{p})^3}$$

La distribució assimptòtica és:

$$\sqrt{\frac{n(1-\hat{p})^3}{\hat{p}}}\left(\frac{\hat{p}}{1-\hat{p}} - \frac{p}{1-p}\right) \approx N(0,1)$$

Esquema de la demostració de la distribució assipmtòtica del MLE

Sigui X variable aleatòria amb funció de densitat $f(x|\theta)$,$\theta \in \Theta$. Sigui $L(\theta; \underset{\sim}{X}) = \prod_{i=1}^{n} f(x_i|\theta)$ la funció de versemblança d'una m.a.s. $X_1, \dot{;} X_n$

Teorema I: Suposem que es verifiquen les següents condicions:

- C1: El paràmetre θ és identificable (diferents valors de θ suposen diferents distribucions de X)
- C2: El conjunt $\{x : f(x|theta) > 0\}$ és el mateix per a tot $\theta \in \Theta$
- C3: La següent quantitat existeix per a tot parell $\theta, \theta_0 \in \Theta$

$$e(\theta_0, \theta) = E_{\theta_0}\left[\log\left(\frac{f(X|\theta)}{f(X|\theta_0)}\right)\right]$$

Llavors per a tot $\theta \neq \theta_0$ es verifica que:

a)

$$E_{\theta_0}\left[\log\left(\frac{L(\theta|\underset{\sim}{X})}{L(\theta_0|\underset{\sim}{X})}\right)\right] < 0$$

b)

$$\lim_{n\to\infty} P_{\theta_0}(L(\theta_0|\underset{\sim}{X}) > L(\theta|\underset{\sim}{X}) = 1$$

Esquema de la demostració de la distribució assipmtòtica del MLE

L'anterior teorema estableix que amb una mostra prou gran, la versemblança assoleix el màxim en el veritable valor del paràmetre.

Teorema II:Suposem que a mes de C1, C2 i C3 es verifiquen:

C4: Θ és un conjunt obert C5: $\frac{\partial}{\partial\theta} f(x|\theta)$ és continua en θ

Llavors, amb probabilitat que tendeix a 1 quan n tendeix a infinit, existeix una successió $\{\hat{\theta}_n\}_n$ d'arrels de l'equació de l'Score

$$\frac{\partial}{\partial\theta} \log L(\theta|\underset{\sim}{X}_n) = 0$$

que convergeix al veritable valor del paràmetre θ_0 en probabilitat.

Aquest teorema demostra que sempre existeix una arrell de l'equació de l'score que si é única, és màxim local i consistent.

Esquema de la demostració de la distribució assipmtòtica del MLE

Teorema III: Suposem que a més de C1, C2, C3, C4 i C5 es verifiquen H2 i H3 del teorema de Cramér-Rao i la següent hipòtesi:

C6: Existeix $\frac{\partial^3}{\partial\theta^3} \log f(x|\theta)$ i el seu valor absolut està afitat per una unció $K(\theta)$ tal que $E[K(\theta)] \leq k$

Sigui $\{\hat{\theta}_n\}_n$ una successió consistent d'arrels de l'equació de l'score: $\theta_n \to_P \theta_0$, si θ_0 és el veritable valor del paràmetre.

Llavors,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_D N\left(0, \frac{1}{I(\theta_0)}\right)$$

on

$$I(\theta_0) = \lim_{n\to\infty} \frac{1}{n} I_{\underset{\sim}{X}_n}(\theta_0) = I_X(\theta_0)$$

Esquema de la demostració de la distribució assipmtòtica del MLE

Teorema IV: Sota les condicions de l'anterior teorema (C1 a C6, H1 i H2), els estadístics definits com

$$O_n = -\frac{\partial^2}{\partial\theta^2} \log L(\theta|\underset{\sim}{X})\bigg|_{\theta=\hat{\theta}_n}$$

$$E_n = I_{\underset{\sim}{X}_n}(\hat{\theta}_n)$$

dividits per n són estimadors consistents de $I_X(\theta_0)$. És a dir, tant la Informació Observada (O_n) com la informació Esperada (E_n) evaluades en el màxim són estimadors consistents de la informació de Fisher pel paràmetre.