Lezione n.3:Analisi bayesiana

Roma, 11 marzo 2003

Brunero Liseo
Dipartimento di studi geoeconomici, linguistici, statistici e storici
per l'analisi regionale
Università di Roma "La Sapienza"
Rome, Italy
`brunero.liseo@uniroma1.it`
`tel. 06-49766110`

Esempio gaussiano

Si effettuano n replicazioni di una misurazione. Il modello è

$$X_i = \theta + \varepsilon_i, \qquad i = 1, \cdots n;$$

con $\varepsilon_i \sim N(0, \sigma^2)$ e gli errori sono indipendenti tra loro. La verosimiglianza è

$$L(\theta; z) \propto \exp\left\{-\frac{n}{\sigma^2}(\bar{x} - \theta)^2\right\}$$

Occorre adesso determinare una legge H. Partiamo da alcune considerazioni:

- siamo reagionevolmente sicuri che θ sia compreso tra due valori l_1 e l_2 che esprimiamo come

$$\mu_0 \pm 3\sigma_0$$

- sono più probabili i valori vicini a μ_0 che non quelli più lontani
- l'incertezza intorno a μ_0 è simmetrica

Una possibile (non certo l'unica) legge che soddisfa queste caratteristuche è una legge $N(\mu_0, \sigma_0^2)$, dove

$$\mu_0 - 3\sigma_0 = l_1, \qquad \mu_0 + 3\sigma_0 = l_2$$

cioè

$$\mu_0 = \frac{l_1 + l_2}{2}, \qquad \sigma_0 = \frac{l_1 - l_2}{6}$$

Dunque l'elicitazione di l_1 e l_2 ci conduce alla scelta della legge iniziale H. σ_0 è una misura della nostra incertezza, mentre μ_0 è la nostra **prior guess**. A questo punto il calcolo della legge finale è semplice

$$h(\theta \mid z) \propto h(\theta)L(\theta; z) =$$

$$\frac{1}{\sigma_0\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_0^2}(\theta - \mu_0)^2\right\} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x} - \theta)^2\right\}$$

Lemma 1 *Siano A e B reali positivi, e a e b reali qualunque. Allora*

$$A(\mu-a)^2+B(\mu-b)^2=(A+B)\left(\mu-\frac{aA+bB}{A+B}\right)^2+\frac{AB}{A+B}(a-b)^2 \tag{1}$$

Dimostrazione 1 *Semplici calcoli.*

Eliminando tutto quello che non dipende da θ, e tenendo conto che

$$A = \frac{1}{\sigma_0^2}, \quad a = \mu_0, \quad B = \frac{n}{\sigma^2}, \quad b = \bar{x}$$

si ottiene che

$$h(\theta \mid z) \propto \exp\left\{-\frac{\sigma^2/n + \sigma_0^2}{2\sigma^2\sigma_0^2}\left(\theta - \frac{\bar{x}\sigma_0^2 + \mu_0\sigma^2/n}{\sigma_0^2 + \sigma^2/n}\right)^2\right\}$$

In altri termini

$$\theta \mid z \sim N\left(\frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}, \frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}\right)$$

Commenti

La media della distribuzione finale può essere vista come media ponderata della media campionaria e di quella iniziale

$$\mathrm{E}\left(\theta \mid z\right) = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

- Al crescere di n diminuisce il peso della legge iniziale
- Al crescere di σ_0 stesse conseguenze....

Al limite, per $\sigma_0 \to \infty$, $\theta \mid z \sim N(\bar{x}, \sigma^2/n)$

Commenti

La varianza della distribuzione è

$$\mathrm{Var}\,(\theta \mid z) = \frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2}$$

- La varianza non dipende dal dato osservato

Dati dicotomici

$$X_1, X_2, \cdots, X_n \text{i.i.d.} \sim Be(\theta), \qquad\qquad \theta \sim Beta(\alpha, \beta), \qquad (2)$$

ovvero

$$h(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},$$

con α e $\beta > 0$

Notizie sulla Beta

$$\mathbf{E}\theta = \frac{\alpha}{\alpha+\beta}, \qquad \mathrm{Var}\,(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Se $\alpha = 1, \beta > 1$ densità decrescente

Se $\alpha = 1, \beta = 1$ densità uniforme

Se $\alpha > 1, \beta = 1$ densità crescente

Se $\alpha > 1, \beta > 1$ densità unimodale

Se $\alpha < 1, \beta < 1$ densità a forma di U

Calcolo della finale

Verosimiglianza:

$$L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^k(1-\theta)^{n-k},$$

Legge finale per θ:

$$\pi(\theta \mid \mathbf{x}) \propto \theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1},$$

ovvero $\theta \mid \mathbf{x} \sim Beta(\alpha+k, \beta+n-k)$

Sintesi a posteriori

$$E(\theta \mid \mathbf{x}) = \frac{\alpha + k}{\alpha + \beta + n};$$

che può scriversi come

$$E(\theta \mid \mathbf{x}) = \frac{\alpha + \beta}{\alpha + \beta + n} \frac{\alpha}{\alpha + \beta} + \frac{n}{\alpha + \beta + n} \frac{k}{n} =, \frac{\alpha + \beta}{\alpha + \beta + n} E(\theta) + \frac{n}{\alpha + \beta + n} \hat{\theta},$$

media ponderata della media di θ a priori ($\frac{\alpha}{\alpha+\beta}$) e della stima di massima verosimiglianza ($\hat{\theta} = k/n$).

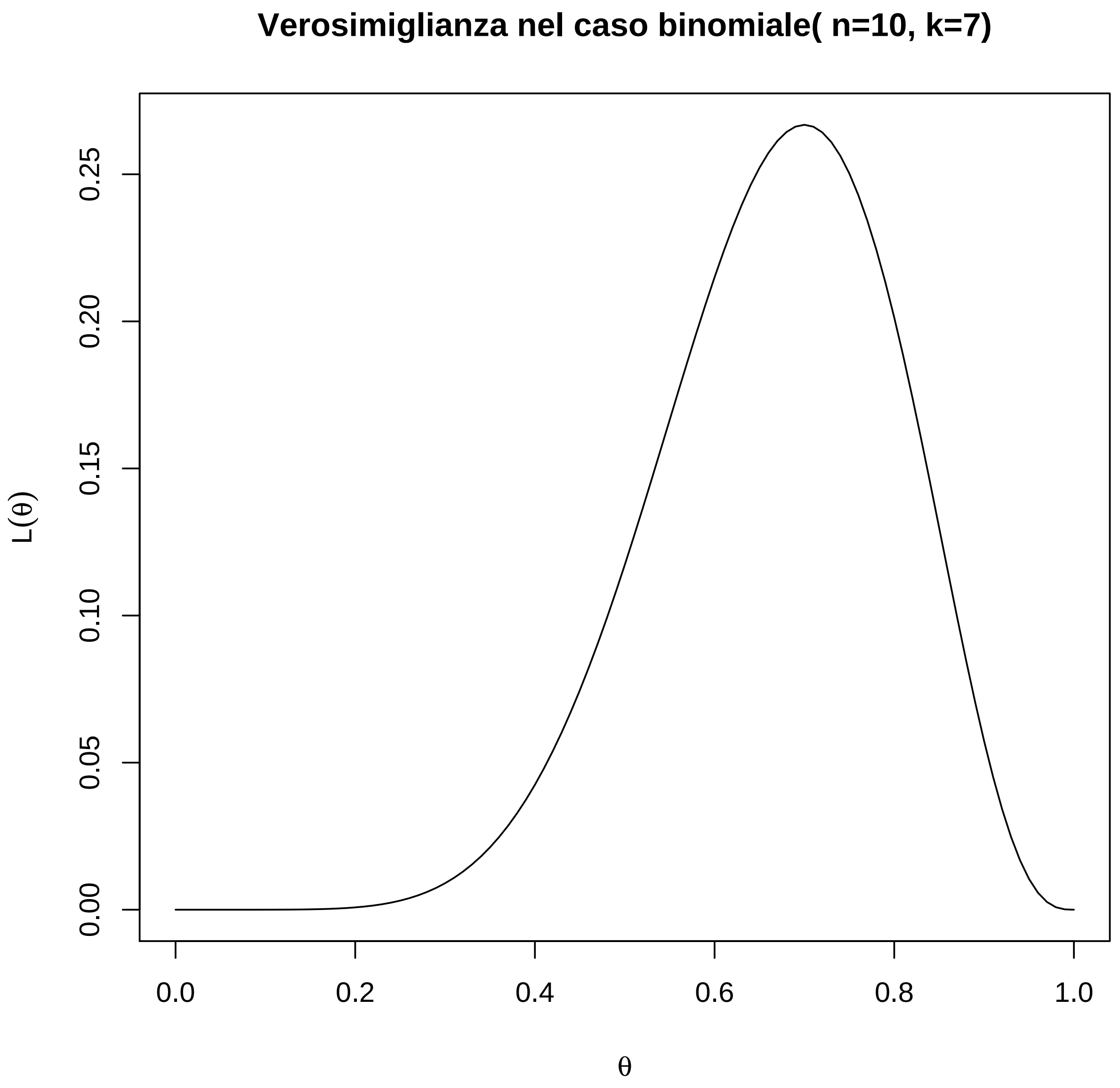
- ruolo della dimensione campionaria

$$V(\theta \mid \mathbf{x}) = \frac{(a + k)(b + n - k)}{(a + b + n)^2 (a + b + n + 1)}$$

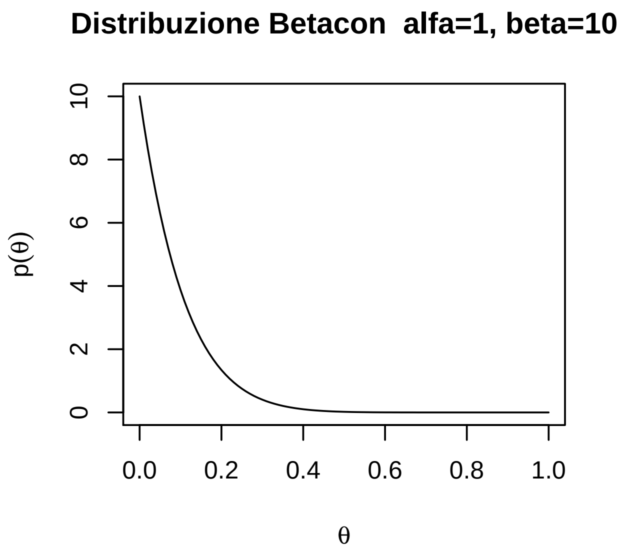
Esempio

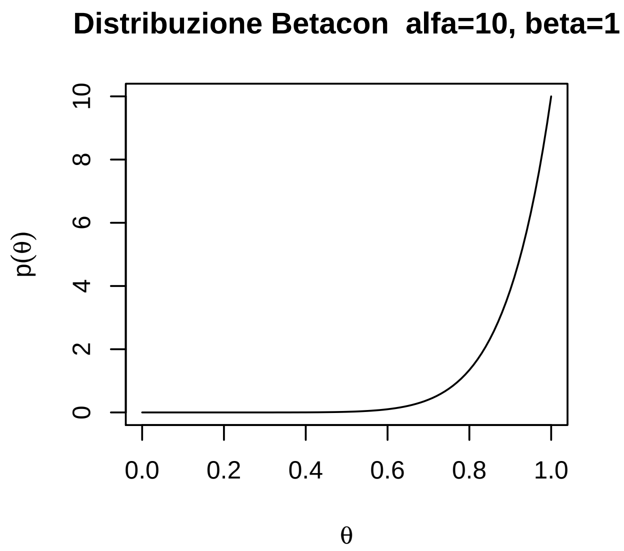
Diverse distribuzione a priori

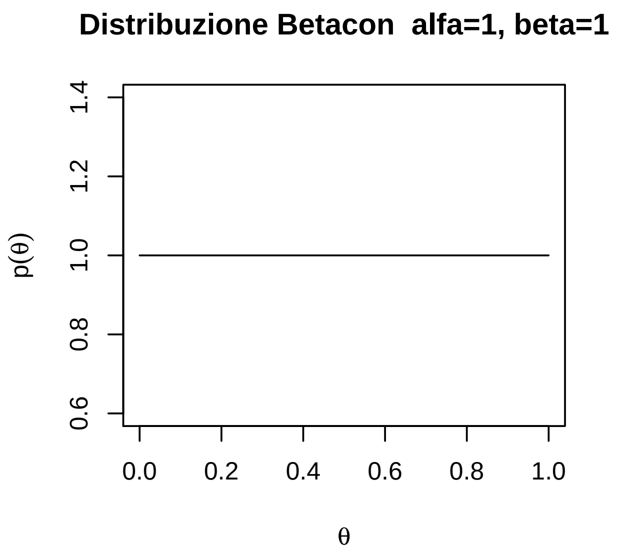
- $\alpha = 1, \beta = 10$
- $\alpha = 10, \beta = 1$
- $\alpha = 5, \beta = 5$
- $\alpha = 1, \beta = 1$

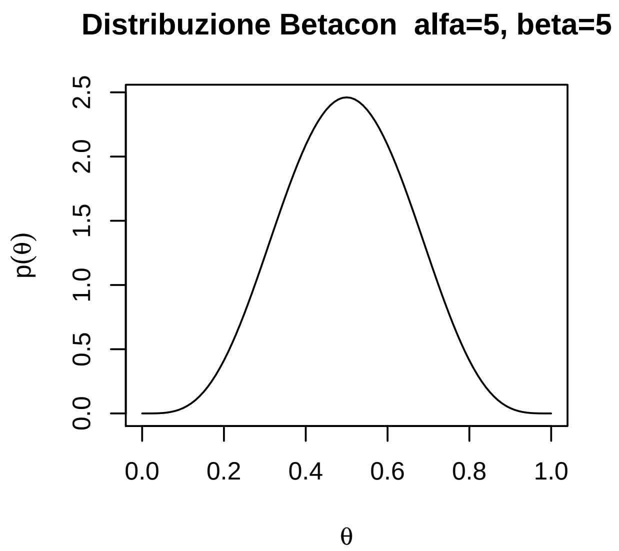


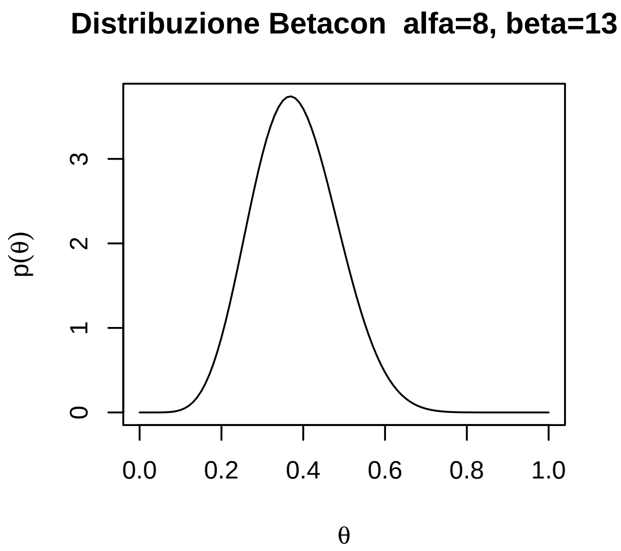
Verosimiglianza nel caso binomiale(n=10, k=7)
L(θ)
0.00
0.05
0.10
0.15
0.20
0.25
0.0
0.2
0.4
0.6
0.8
1.0
θ

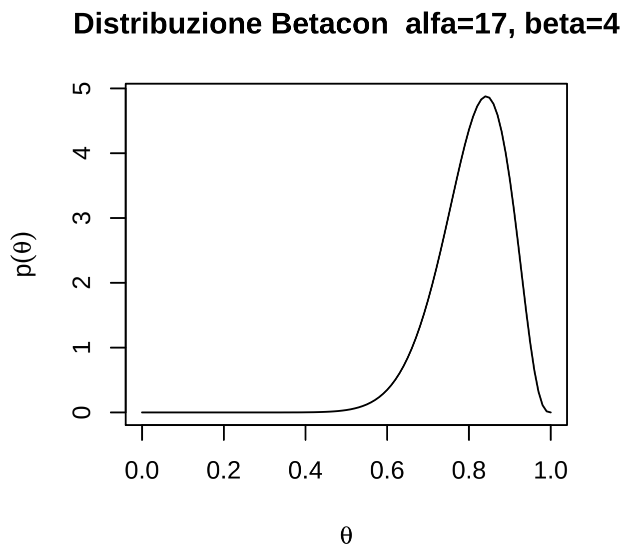

Distribuzione Betacon alfa=1, beta=10
p(θ)
θ
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0

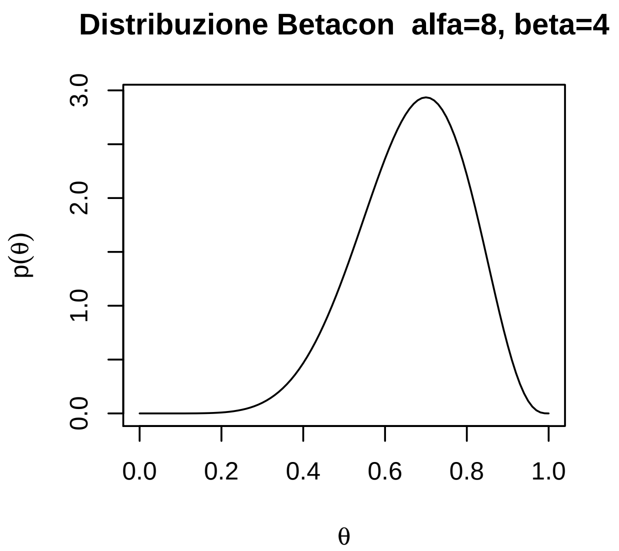

Distribuzione Betacon alfa=10, beta=1
p(θ)
θ
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0

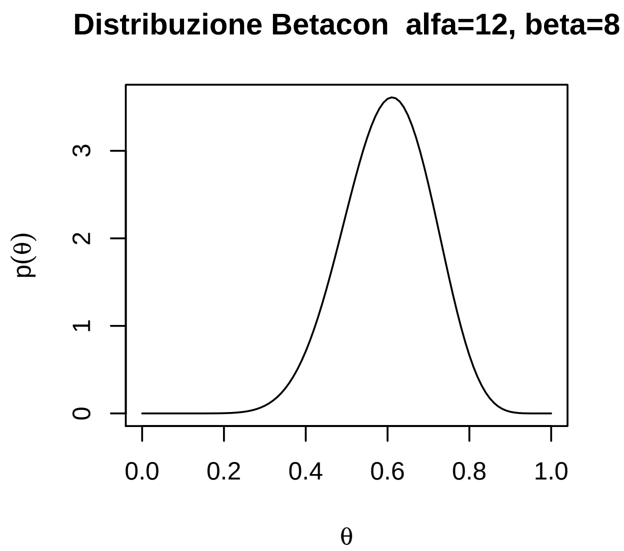

Distribuzione Betacon alfa=1, beta=1
p(θ)
θ
0.6
0.8
1.0
1.2
1.4
0.0
0.2
0.4
0.6
0.8
1.0


Distribuzione Betacon alfa=5, beta=5
p(θ)
θ
0.0
0.5
1.0
1.5
2.0
2.5
0.0
0.2
0.4
0.6
0.8
1.0


Distribuzione Betacon alfa=8, beta=13
p(θ)
θ


Distribuzione Betacon alfa=17, beta=4
p(θ)
θ


Distribuzione Betacon alfa=8, beta=4
p(θ)
θ


Distribuzione Betacon alfa=12, beta=8
p(θ)
θ

Asintotica della distribuzione finale

Come si modifica la distribuzione finale al crescere di n?
Consideriamo un caso semplice

$$\Theta = \{\theta_0, \theta_1, \theta_2, \cdots, \theta_m\}$$

con legge a priori

$$h(\theta_1), h(\theta_2), \cdots, h(\theta_m), \quad \text{TUTTE} > 0$$

e supponiamo che θ_0 sia il vero valore di θ. In quest'ottica $X_1, \cdots, X_n$ sono v.a. i.i.d. con legge $p(x \mid \theta_0)$.

Vogliamo studiare il comportamento asintotico di $h(\theta \mid \mathbf{x})$ al crescere di n secondo la legge $p(x \mid \theta_0)$.

Poiché

$$h(\theta_0 \mid \mathbf{x}) = \frac{h(\theta_0)}{\sum_{j=0}^{m} h(\theta_j) \prod_{i=1}^{n} \frac{p(x_i|\theta_j}{p(x_i|\theta_0)}} =$$

$$h(\theta_0) \left[h(\theta_0) + \sum_{j=1}^{m} h(\theta_j) \prod_{i=1}^{n} \frac{p(x_i \mid \theta_j}{p(x_i \mid \theta_0} \right]^{-1}$$

Teorema

Se θ_0 *è il vero valore di* θ*, allora*

$$h(\theta_0 \mid \mathbf{x}) \stackrel{q.c.}{\to} 1$$

e

$$h(\theta_j \mid \mathbf{x}) \stackrel{q.c.}{\to} 0, \qquad j = 1, \cdots m.$$

Dimostrazione: alla lavagna.

Commenti:

1. Il teorema valer sotto condizioni più generali ($\Theta \subset \mathrm{R}_k$). Ad esempio che la legge a priori ponga una massa di probabilità positiva intorno al vero valore θ_0

2. Il teorema precedente ci dà la convergenza delle singole probabilità senza dire nulla sulla velocità di convergenza. A tal proposito pu'ò essere utile un risultato generale, basato sul teorema del limite centrale.

Lemma

Siano $\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}$ vettori di R_k e $\mathbf{A}, \mathbf{B}$ matrici simmetriche $k \times k$ tali che $(\mathbf{A}+\mathbf{B})^{-1}$ esiste. Allora,

$$(\boldsymbol{\theta}-\mathbf{a})'\mathbf{A}(\boldsymbol{\theta}-\mathbf{a}) + (\boldsymbol{\theta}-\mathbf{b})'\mathbf{B}(\boldsymbol{\theta}-\mathbf{b})$$

$$= (\boldsymbol{\theta}-\mathbf{c})'(\mathbf{A}+\mathbf{B})(\boldsymbol{\theta}-\mathbf{c}) + (\mathbf{a}-\mathbf{b})'\mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}(\mathbf{a}-\mathbf{b})$$

dove

$$c = (\mathbf{A}+\mathbf{B})^{-1}(\mathbf{A}\mathbf{a}+\mathbf{B}\mathbf{b})$$

Trattazione euristica

Consideriamo il caso generico con $\theta \in \mathrm{R}_k$, con legge $h(\theta)$ e n osservazioni $X_1, \cdots, X_n$ i.i.d. con legge $p(x \mid \theta)$.

Allora

$$h(\theta \mid \mathbf{x}) \propto \exp\{\log h(\theta) + \log p(\mathbf{x} \mid \theta)\}$$

Sviluppando i due logaritmi in serie di Taylor intorno ai due massimi m_0 e $\hat{\theta}_n$, si ottiene

$$\log h(\theta) = \log h(m_0) - \frac{1}{2}(\theta - m_0)'\mathbf{H}_0^{-1}(\theta - m_0) + R_0$$

e

$$\log p(\mathbf{x} \mid \theta) = \log p(\mathbf{x} \mid \hat{\theta}_n) - \frac{1}{2}(\theta - \hat{\theta}_n)'\mathbf{H}(\hat{\theta}_n)^{-1}(\theta - \hat{\theta}_n) + R_n$$

dove

$$\mathbf{H}_0^{-1} = \left(-\frac{\partial^2 \log h(\theta)}{\partial\theta_i \partial\theta_j}\right) |_{\theta=m_0}\,, \qquad \mathbf{H}_{(}\hat{\theta}_n)^{-1} = \left(-\frac{\partial^2 \log p(\mathbf{x} \mid \theta)}{\partial\theta_i \partial\theta_j}\right) |_{\theta=\hat{\theta}_n}$$

Assumendo condizioni di regolarità che rendono R_0 e R_n trascurabili si ottiene, per n grande e trascurando costanti che non dipendono da θ

$$h(\theta \mid \mathbf{x}) \propto \exp\left\{-\frac{1}{2}(\theta - m_0)'\mathbf{H}_0^{-1}(\theta - m_0) - \frac{1}{2}(\theta - \hat{\theta}_n)'\mathbf{H}(\hat{\theta}_n)^{-1}(\theta - \hat{\theta}_n)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(\theta - m_n)'\mathbf{H}_n^{-1}(\theta - m_n)\right\}$$

dove

$$H_n^{-1} = \mathbf{H}_0^{-1} + \mathbf{H}(\hat{\theta}_n)^{-1}$$

e

$$m_n = \mathbf{H}_n \left(\mathbf{H}_0^{-1} m_0 + \mathbf{H}(\hat{\theta}_n)^{-1} \hat{\theta}_n \right)$$

(in virtù del lemma precedente)

Questa espressione permette dunque di dire che, al crescere di n la distribuzione finale assume una forma gaussiana intorno a m_n con parametro di scala $\mathbf{H}_n$. A loro volta,

$$\frac{m_n}{\hat{\theta}_n}\ n \overset{\approx}{\rightarrow} \infty\ 1$$

e, per la legge forte dei grandi numeri, e per ogni $i, j = 1, \cdots, k$,

$$\lim_{n\to\infty} \left\{ \frac{1}{n} \left(-\frac{\partial^2 \log p(\mathbf{x} \mid \theta)}{\partial\theta_i \partial\theta_j} \right) \right\} = \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{k=1}^{n} \left(-\frac{\partial^2 \log p(x_k \mid \theta)}{\partial\theta_i \partial\theta_j} \right) \right\}$$

$$= \int p(x \mid \theta) \left(-\frac{\partial^2 \log p(x \mid \theta}{\partial\theta_i \partial\theta_j} \right) dx$$

Perciò $\mathbf{H}(\hat{\theta}_n) \to nI(\hat{\theta}_n)$ dove $I(\theta)$ è la ben nota informazione attesa di Fisher relativa ad una osservazione.

Possiamo allora concludere che, sotto condizioni di regolarità (da precisare)

$$\theta \mid \mathbf{x} \approx N\left(\hat{\theta}_n, \mathbf{H}(\hat{\theta}_n)^{-1}\right)$$

oppure

$$\theta \mid \mathbf{x} \approx N\left(\hat{\theta}_n, \frac{1}{n} I(\hat{\theta}_n)^{-1}\right)$$

Esempio: $X_1, \cdots, X_n$ i.i.d. $N\left(\theta, \sigma^2\right)$ entrambi incogniti.
La verosimiglianza è

$$L(\theta, \sigma^2) \propto \frac{1}{\sigma^n} \exp\left\{-\frac{n}{2\sigma^2}\left[s^2 + (\bar{x} - \theta)^2\right]\right\}$$

Si dimostra che

$$\hat{\theta} = \bar{x}, \qquad \hat{\sigma^2} = s^2;$$

Calcoliamo la matrice di Fisher nel punto di MV.

- $\frac{\partial^2}{\partial\theta\partial\sigma^2} \log L(\theta,\sigma^2) \mid_{MV} = 0$
- $\frac{\partial^2}{\partial\theta^2} \log L(\theta,\sigma^2) \mid_{MV} = -\frac{n}{s^2}$
- $\frac{\partial^2}{\partial\sigma^4} \log L(\theta,\sigma^2) \mid_{MV} = -\frac{n}{2s^4}$

cioè

$$nI(\theta,\sigma^2) \mid_{MV} = \begin{vmatrix} \frac{n}{s^2} & 0 \\ 0 & \frac{n}{2s^4} \end{vmatrix},$$

da cui l'inversa è

$$\frac{1}{n}I^{-1}(\theta,\sigma^2) \mid_{MV} = \begin{vmatrix} \frac{s^2}{n} & 0 \\ 0 & \frac{2s^4}{n} \end{vmatrix},$$

cosicché, per grandi valori di n e senza dipendere dalla scelta della a priori $h(\theta, \sigma^2)$ (purché sia ragionevole!!) si avrà

$$\left(\theta, \sigma^2\right) \sim N\left[\left(\bar{x}, s^2\right), \frac{1}{n} I(\hat{\theta}, \hat{\sigma^2})\right]$$