

# Lezione n.5: Reference priors

**Roma, 20 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 (Poisson)

Siano  $X_1, \dots, X_n$  i.i.d.  $Po(\theta)$ .

La verosimiglianza è

$$L(\theta; \mathbf{x}) \propto \exp\{-n\theta\} \theta^{\sum x_i}$$

ma quella relativa ad una sola osservazione è

$$L(\theta; x_1) \propto \exp\{-\theta\} \theta^{x_i}$$

$$\ell(\theta) \propto -\theta + x_i \log \theta$$

$$\frac{\partial \ell(\theta)}{\partial \theta} = -1 + \frac{x_i}{\theta}$$

$$-\frac{\partial^2 \ell(\theta)}{\partial \theta^2} = \frac{x_i}{\theta^2}$$

Dunque

$$I(\theta) = \mathbb{E}_\theta \left( \frac{X_i}{\theta^2} \right) = \frac{1}{\theta^2} \mathbb{E}_\theta (X_i) = \frac{1}{\theta}$$

Perciò

$$h^J(\theta) \propto \frac{1}{\sqrt{\theta}}$$

La distribuzione finale è

$$h(\theta | \mathbf{x}) \propto \frac{1}{\sqrt{\theta}} \theta^t \exp\{-n\theta\} = \theta^{t-1/2} \exp\{-n\theta\},$$

ovvero

$$\theta | \mathbf{x} \sim \text{Gamma}(n, t + \frac{1}{2})$$

Il metodo di Jeffreys è ancora oggi quello più comunemente utilizzato quando la dimensione di  $\Theta$  è 1.

Tuttavia, lo stesso Jeffreys suggerì alcune modifiche alla sua regola generale nel caso di parametro multidimensionale. Soprattutto, egli considerava a parte eventuali parametri di posizione e scala presenti nel modello.

Esempio 2 Sia  $X_i \sim N(\mu_i, 1)$ ,  $i = 1, \dots, p$  indipendenti tra loro. Si vuole stimare

$$\theta = \frac{1}{p} \sum \mu_i^2$$

L'uso del metodo di Jeffreys condurrebbe a

$$\pi^J(\mu_1, \dots, \mu_p) \propto 1$$

Conseguentemente,

$$(\mu_1, \dots, \mu_p) = \boldsymbol{\mu} \sim N_p(\mathbf{x}, \mathbf{1})$$

e

$$p\theta \sim \chi_p^2\left(\sum x_i^2\right)$$

Questa NON è una buona distribuzione a posteriori per  $\theta$ .

Ad esempio,

$$E^\pi(\theta|\mathbf{x}) = 1 + \frac{\sum x_i^2}{p}.$$

E' facile dimostrare che tale stimatore è inconsistente, ovvero

$$\lim_{p \rightarrow \infty} E^\pi (\theta | \mathbf{x}) - \theta = 2$$

e lo stesso accade considerando la moda o la mediana a posteriori

Questo esempio mette in risalto uno dei problemi più frequenti nella selezione di "buone" distribuzioni non informative. Il metodo di Jeffreys cerca la distribuzione non informativa per l'intero vettore  $\omega$ . Se il parametro d'interesse è soltanto  $\theta$  questo introduce una "distorsione" nella procedura.

Questa osservazione è alla base del metodo delle reference priors.

Il concetto di informazione contenuta in una legge di probabilità

Una misura diretta del contenuto informativo di una legge di probabilità  $\pi$  è dato dall'Entropia  $\mathcal{E}$

$$\mathcal{E} = - \int_{\Omega} \pi(\omega) \log \pi(\omega) d\omega$$

Una misura della “distanza” tra due misure di probabilità è invece fornita dal *numero di Kullback-Leibler*,

$$K(p; q) = \int_{\Omega} q \log \frac{q}{p}$$

che vale zero se e solo se  $q = p$  quasi certamente (rispetto a  $q$ ).

I due concetti suddetti sono alla base della definizione di informazione contenuta in un esperimento, dovuto a Lindley (1956) che riprende e adatta idee di Shannon (1948)

Informazione di Shannon-Lindley Dato un esperimento

$$E_k = (\mathcal{X}_k, \Omega, \mathcal{P})$$

si definisce “*Informazione contenuta in  $E_k$* ”,  
relativamente ad una distribuzione a priori  $\pi$  la quantità

$$I_{E_k}(\pi) = \int_{\mathcal{X}_k} \int_{\Omega} m(\mathbf{x}_k) \pi(\omega | \mathbf{x}_k) \log \frac{\pi(\omega | \mathbf{x}_k)}{\pi(\omega)} d\omega d\mathbf{x}_k \quad (1)$$

$I_{E_k}(\pi)$  rappresenta il valore medio rispetto alla legge marginale  $m$  del numero di K-L della “a priori” rispetto alla “a posteriori”.

E’ allora ragionevole misurare il contributo informativo di una determinata  $\pi$  in termini della (2).

Ragionevole ma non obbligatorio:

Nelle espressioni  $I_{E_k}(\pi)$  e  $\mathcal{E}$  non si integra su valori di  $\omega$  ma solo di  $\pi(\omega)$ .

Le reference priors

Bernardo (1979) ha introdotto il metodo delle reference priors. Le due novità introdotte nella ricerca di  $\pi^r$  furono

- determinazione di  $\pi^r$  come argomento massimizzante  $I_{E_k}(\pi)$
- nel caso multiparametrico, distinzione esplicita tra parametro d'interesse e parametri di disturbo.

La tecnica è stata via via perfezionata nel corso degli anni;

Riferimenti principali:

Bernardo, J.M. Reference posterior distributions for Bayesian inference. With discussion. *JRSS B* 41 (1979), no. 2, 113-147

Berger, J.O.; Bernardo, J.M. On the development of reference priors. *Bayesian statistics, 4* 35-60, Oxford Univ. Press, New York, 1992.

Berger, J.O.; Bernardo, J.M. Ordered group reference priors with application to the multinomial problem. *Biometrika* 79 (1992), no. 1, 25-37.

Bernardo, J.M.; Ramon, J.M. An introduction to Bayesian reference analysis: inference on the ratio of multinomial parameters. *The Statistician* 47 no.1 (1998) 101-135.

## Determinazione delle reference priors

Il calcolo esatto di  $\pi^r$  comporta una serie di problemi tecnici non sempre risolvibili.

Iniziamo la trattazione considerando un modello statistico regolare, in cui

- esiste una statistica sufficiente della stessa dimensione del parametro
- lo stimatore di massima verosimiglianza ha distribuzione asintotica normale
- la distribuzione a posteriori è asintoticamente normale.

Il caso di un solo parametro reale. Supponiamo  $\omega \in \mathbb{R}$ . La quantità  $I_{E_k}(\pi)$  rappresenta l'incremento medio di informazione che l'esperimento fornisce quando la legge a priori è  $\pi(\omega)$ .

Per  $k \rightarrow \infty$ ,  $I_{E_k}(\pi)$  assume il significato di *ammontare complessivo d'informazione mancante sul parametro  $\omega$* , e la legge a priori che massimizza  $I_{E_\infty}(\pi)$  può a ben diritto essere definita come la “meno informativa”.

C'è però un problema:  $I_{E_k}(\pi)$  è in genere illimitato come funzione di  $k$ .

Il suggerimento di Bernardo.

1. Massimizzare  $I_{E_k}(\pi)$  per  $k$  fissato  
 $\implies k$ -reference prior  $\pi_k^r(\omega)$

2. Definire la reference prior come limite puntuale

$$\pi^r(\omega) \lim_{k \rightarrow \infty} \frac{\pi_k^r(\omega)}{\pi_k^r(\Omega_0)}$$

dove  $\Omega_0$  è uno specifico compatto.

## Qualche commento

- non è detto che il limite esista
- il limite è puntuale e non assicura una convergenza nella metrica indotta dalla misura d'informazione di K-L
- il limite risulterà spesso in una distribuzione impropria

Derivazione euristica di  $\pi^r$

$$\begin{aligned}
 I_{E_k}(\pi) &= \\
 &\int_{\Omega} \pi(\omega) \left[ \int_{\mathcal{X}_k} p(\mathbf{x}_k|\omega) (\log \pi(\omega|\mathbf{x}_k) - \log \pi(\omega)) \right] d\omega d\mathbf{x}_k \\
 &= \int_{\Omega} \pi(\omega) \log \frac{\exp\left\{ \int_{\mathcal{X}_k} p(\mathbf{x}_k|\omega) \log \pi(\omega|\mathbf{x}_k) d\mathbf{x}_k \right\}}{\pi(\omega)} d\omega
 \end{aligned}$$

Il problema è del tipo

$$\sup_f \int f \log g/f$$

che viene massimizzato per

$$f \propto g.$$

Ne segue che

$$\pi^r(\omega) \propto \exp\left\{ \int_{\mathcal{X}_k} p(\mathbf{x}_k|\omega) \log \pi(\omega|\mathbf{x}_k) d\mathbf{x}_k \right\}$$

che però fornisce la soluzione solo in maniera implicita.  
In casi regolari, tuttavia, esiste uno stimatore ML  $\hat{\omega}_k$ , tale che

$$\begin{aligned} \exp\left\{\int_{\mathcal{X}_k} p(\mathbf{x}_k|\omega) \log \pi(\omega|\mathbf{x}_k) d\mathbf{x}_k\right\} = \\ \exp\left\{\int_{\mathbf{R}} p(\hat{\omega}_k|\omega) \log \pi(\omega|\hat{\omega}_k) d\hat{\omega}_k\right\} \end{aligned} \quad (2)$$

Inoltre

$$\pi(\omega|\hat{\omega}_k) \sim N(\omega; \hat{\omega}_k, [kH(\hat{\omega}_k)]^{-1})$$

e la (??) risulta approssimativamente uguale a

$$\exp\left\{\int_{\mathbb{R}} p(\hat{\omega}_k|\omega) [\log H(\hat{\omega}_k)^{1/2} - \frac{k}{2}H(\hat{\omega}_k)(\omega - \hat{\omega}_k)^2] d\hat{\omega}_k\right\}$$

Assumendo che  $\hat{\omega}_k$  tende a concentrarsi attorno a  $\omega$  si avrà allora

$$\pi^r(\omega) \propto H(\omega)^{1/2}$$

## Commenti

- Nel caso univariato, sotto condizioni di regolarità, la reference prior coincide con la Jeffreys' prior.
- Una formalizzazione più rigorosa di questa tecnica condurrebbe a risultati poco rassicuranti. Si può dimostrare che, in certi casi, la distribuzione a priori che massimizza la  $I_{E_k}$  risulta concentrata su un numero finito di punti.

Un solo parametro di disturbo

Sia ora  $\omega = (\theta, \lambda)$  e sia  $\theta$  il solo parametro di interesse.

In questo caso la Jeffreys' prior è

$$\pi^J(\theta, \lambda) \propto \det(H(\theta, \lambda))^{1/2}$$

Nel caso delle reference prior, si cerca quella  $\pi(\theta, \lambda)$  che massimizza la distanza d'informazione tra  $\pi(\theta|\mathbf{x}_k)$  e  $\pi(\theta)$ , ovvero

$$I_{E_k}(\pi(\theta, \lambda)) = \int_{\Theta} \pi(\theta) \left[ \int_{\mathcal{X}_k} p(\mathbf{x}_k|\theta) (\log \pi(\theta|\mathbf{x}_k) - \log \pi(\theta)) \right] d\omega d\mathbf{x}_k \quad (4)$$

L'equazione non dipende direttamente da  $\lambda$  (è stato già integrato!) e, analogamente a quanto vista prima,

$$\pi_k^r(\theta) \propto \exp\left\{ \int_{\mathcal{X}_k} p(\mathbf{x}_k|\theta) \log \pi(\theta|\mathbf{x}_k) d\mathbf{x}_k \right\}$$

Tale risultato vale qualunque sia la scelta per  $\pi(\lambda|\theta)$ .

L'algoritmo delle reference priors suggerisce di

- scegliere  $\pi_k^r(\lambda|\theta) \propto H_{22}(\theta, \lambda)^{1/2}$   
(la Jeffreys prior per  $\theta$  noto)
- massimizzare la (??) con  $\pi_k^r(\lambda|\theta)$ .

Dunque,

$$\begin{aligned} \pi_k^r(\theta) &\propto \exp\left\{ \int_{\hat{\theta}} \int_{\hat{\lambda}} p(\hat{\theta}, \hat{\lambda}|\theta) \right. \\ &\quad \times \log N(\theta; \hat{\theta}, S_{11}(\hat{\theta}, \hat{\lambda})) d\hat{\theta} d\hat{\lambda} \end{aligned}$$

dove  $S = H^{-1}$ .

$$\begin{aligned} \pi_k^r(\theta) &\propto \exp\left\{ \int_{\hat{\theta}} \int_{\hat{\lambda}} \int_{\Lambda} p(\hat{\theta}, \hat{\lambda}|\theta, \lambda) \pi^r(\lambda|\theta) \right. \\ &\quad \times \log N(\theta; \hat{\theta}, S_{11}(\hat{\theta}, \hat{\lambda})) d\lambda d\hat{\theta} d\hat{\lambda} \end{aligned}$$

$$= \exp\left\{\int_{\hat{\theta}} \int_{\hat{\lambda}} p(\hat{\theta}, \hat{\lambda}|\theta, \lambda) \int_{\Lambda} \pi^r(\lambda|\theta) \right. \\ \left. \times \log N(\theta; \hat{\theta}, S_{11}(\hat{\theta}, \hat{\lambda})) d\lambda d\hat{\theta} d\hat{\lambda}\right\}$$

$$\cong \exp\left\{\frac{1}{2} \int_{\Lambda} \pi^r(\lambda|\theta) \log S_{11}^{-1}(\theta, \lambda) d\lambda\right\}$$

Poiché  $S_{11} = \frac{H_{22}}{\det(H)}$ ,

$$\pi^r(\theta, \lambda) = \pi(\lambda|\theta) \exp\left\{\frac{1}{2} \int_{\Lambda} \pi(\lambda|\theta) \log \frac{\det(H)}{H_{22}} d\lambda\right\}.$$

Fin qui abbiamo trascurato il problema della non integrabilità delle leggi a priori. All'interno dell'algoritmo tale problema si aggira considerando una successione di compatti che "invadono"  $\Theta$  e sui quali definiamo una successione di reference priors.

## Algoritmo

Passo 1  $\pi^*(\lambda|\theta) \propto \sqrt{H_{22}(\theta, \lambda)}$

Passo 2 Normalizzazione di  $\pi^*(\lambda|\theta)$

- $\pi^*(\lambda|\theta)$  è propria:  
 $\pi(\lambda|\theta) = \pi^*(\lambda|\theta)k(\theta)$
- $\pi^*(\lambda|\theta)$  è impropria:

Si determina una successione

$$\Lambda_1(\theta), \Lambda_2(\theta), \dots, \Lambda_m(\theta), \dots \rightarrow \Lambda, \forall \theta$$

sui quali è possibile definire

$$\pi_m(\lambda|\theta) = \pi^*(\lambda|\theta)k_m(\theta)$$

Passo 3 Distribuzione marginale di  $\theta$ .

$$\pi_m(\theta) \propto \exp\left\{\frac{1}{2}\pi_m(\lambda|\theta) \log \frac{\det H(\theta, \lambda)}{H_{22}(\theta, \lambda)} d\lambda\right\}$$

Passo 4

$$\pi^r(\theta, \lambda) = \lim_{m \rightarrow \infty} \frac{k_m(\theta)\pi_m(\theta)}{k_m(\theta_0)\pi_m(\theta_0)} \frac{\pi_m(\lambda|\theta)}{\pi(\lambda|\theta_0)}$$

Esempio 1 (continua) La funzione di verosimiglianza è

$$L(\omega) = \omega^t (1 - \omega)^{k-t}, \quad t = \sum x_i,$$

L'informazione di Fisher vale  $H(\omega) = \frac{1}{\omega(1-\omega)}$  e, di conseguenza

$$\pi^J(\omega) = \pi^r(\omega) = \frac{1}{\pi} \omega^{-1/2} (1 - \omega)^{-1/2}$$

Commento:  $\pi^J$  e  $\pi^r$  sono distribuzioni proprie ma non uniformi.

Esempio 3: Modello Trinomiale) Riconsideriamo ora l'esempio precedente ma suddividiamo i risultati possibili non più in due categorie bensì in tre, ovvero

$$X_i = S, N, F$$

con prob.  $\omega_1, \omega_2, 1 - \omega_1 - \omega_2$ .

Il parametro d'interesse (lo stesso di prima) è ora  $\theta = \omega_1$  ma nel modello è presente anche  $\lambda = \omega_2$ .

L'informazione di Fisher vale ora

$$H(\theta, \lambda) = \frac{1}{1 - \theta - \lambda} \times \begin{pmatrix} \frac{1 - \lambda}{\theta} & 1 \\ 1 & \frac{1 - \theta}{\lambda} \end{pmatrix}$$

La Jeffreys prior è dunque

$$\pi^J(\theta, \lambda) \propto \frac{1}{\sqrt{\theta\lambda(1 - \lambda - \theta)}}$$

Calcolo della reference prior

$$1. \pi^*(\lambda|\theta) \propto \sqrt{H_{22}(\theta, \lambda)} = \frac{1}{\sqrt{\lambda(1-\lambda-\theta)}}$$

$$2. \pi(\lambda|\theta) = k(\theta) \frac{1}{\sqrt{\lambda(1-\lambda-\theta)}} I_{[0,1-\theta]}(\lambda)$$

$$3. \pi(\theta) = \exp \left\{ \frac{1}{2} \int_{\Lambda(\theta)} k(\theta) \frac{1}{\sqrt{\lambda(1-\lambda-\theta)}} \log \frac{1}{\theta(1-\theta)} d\lambda \right\}$$

$$= \frac{1}{\sqrt{\theta(1-\theta)}}$$

$$4. \pi^r(\theta, \lambda) \propto \frac{1}{\sqrt{\theta\lambda(1-\theta)(1-\theta-\lambda)}}$$

Confronto tra  $\pi^r(\theta, \lambda)$  e  $\pi^J(\theta, \lambda)$

La natura differente delle due distribuzioni si può notare considerando le corrispondenti marginali per  $\theta$ .

- $\pi^r(\theta) = \frac{1}{\pi} \frac{1}{\sqrt{\theta(1-\theta)}}$   
 $\Rightarrow E(\theta|\pi^r) = 1/2$
- $\pi^J(\theta) = \frac{1}{2}\theta^{-1/2}$   
 $\Rightarrow E(\theta|\pi^J) = 1/3$

Questa differenza è ancor più accentuata nel caso generale con  $h$  possibili risultati.

Si vede facilmente che, in questo caso

- $E(\theta_i|\pi^J) = \frac{1}{h} \quad i = 1, \dots, h.$
- $E(\theta_i|\pi^r) = \frac{1}{2^i} \quad i = 1, \dots, h.$

## Il problema di Fieller Siano

$$X \sim N(\omega_1, 1) \quad Y \sim N(\omega_2, 1)$$

- Parametro di interesse  $\theta = \frac{\omega_1}{\omega_2}$
- Parametro di disturbo  $\lambda = \text{sgn}(\omega_2)\sqrt{\omega_1 + \omega_2}$   
(ortogonale)

## Matrice d'informazione

$$H(\theta, \lambda) = \begin{pmatrix} \frac{\lambda^2}{(1 + \theta^2)^2} & 0 \\ 0 & 1 \end{pmatrix}$$

## Jeffreys prior

$$\pi^J(\theta, \lambda) \propto \frac{|\lambda|}{1 + \theta^2}$$

## Reference Prior

$$A_m = (-a_m < \omega_1 < a_l) \times (-b_m < \omega_2 < b_l)$$

che diventa

- $\frac{a_m}{\theta} \sqrt{(1 + \theta^2)} < \lambda < -\frac{a_m}{\theta} \sqrt{(1 + \theta^2)}$  per  $\theta < -\frac{a_m}{b_m}$
- $-b_m \sqrt{1 + \theta^2} < \lambda < b_m \sqrt{1 + \theta^2}$  per  $|\theta| < \frac{a_m}{b_m}$
- $-\frac{a_m}{\theta} \sqrt{(1 + \theta^2)} < \lambda < \frac{a_m}{\theta} \sqrt{(1 + \theta^2)}$  per  $\theta > \frac{a_m}{b_m}$

Ne segue che

- $\pi_m^*(\lambda|\theta) \propto 1$
- $\pi_m(\lambda|\theta) = k_m(\theta)$

La reference prior è dunque

$$\pi_m(\theta, \lambda) \propto k_m(\theta) \times \exp \left\{ \frac{1}{2} k_m(\theta) \int_{A_m} (\log \lambda^2 - \log(1 + \theta^2))^2 d\lambda \right\}$$

$$= k_m(\theta) \exp \left\{ k_m(\theta) \int_{A_m} \log |\lambda| d\lambda \right\} \frac{1}{1 + \theta^2}$$

$$\rightarrow \pi^r(\theta, \lambda) \propto \frac{1}{1 + \theta^2}$$

Va notato che questa è la reference prior quando il parametro d'interesse è  $\theta = \omega_1/\omega_2$ . Se ad esempio fossimo interessati a  $\xi = \omega_1\omega_2$  il risultato sarebbe differente mentre la Jeffreys prior per  $\xi$  si otterrebbe attraverso un cambio di variabile da  $\pi^J(\omega_1, \omega_2) \propto 1$ .

Esempio 2 (continua) Sia  $X_i \sim N(\mu_i, 1)$ ,  $i = 1, \dots, p$ . Si vuole stimare

$$\theta = \frac{1}{p} \sum \mu_i^2 = \frac{1}{p} \|\boldsymbol{\mu}\|^2$$

Una scelta opportuna per il parametro di disturbo è

$$\lambda = \boldsymbol{\mu} / \|\boldsymbol{\mu}\|$$

ovvero la “direzione” del vettore  $\boldsymbol{\mu}$  sulla superficie dell’ipersfera di raggio unitario. E’ naturale, e l’algoritmo delle reference priors lo conferma, assegnare una distribuzione a priori uniforme per  $\lambda|\theta$ , e poi determinare la marginale per  $\theta$ .

Ragionare nella parametrizzazione  $(\theta, \lambda)$  ha permesso di “scoprire” il problema emerso con la Jeffreys prior.

L'uso di  $\pi(\boldsymbol{\mu}) = 1$  implicava, infatti,

$$\pi(\theta, \lambda) = \left(\frac{1}{\theta}\right)^{\frac{p-2}{2}} \pi(\lambda|\theta)$$

Il caso multiparametrico Consideriamo il caso

$\omega = (\omega_1, \omega_2, \dots, \omega_k)$ . Il metodo si generalizza in modo ovvio.

Passo 1 Dividi i  $k$  parametri in  $p$  gruppi  $(\theta_{(1)}, \dots, \theta_{(p)})$

Passo 2 Determina una successione di compatti

$$\Omega_1 \subseteq \Omega_2 \subseteq \dots \rightarrow \Omega$$

Passo 3 Calcola sul generico compatto  $\Omega_m$ , la reference prior per  $\theta_{(p)}$  dati gli altri, ovvero

$$\pi_m(\theta_{(p)} | \theta_{(1)}, \dots, \theta_{(p-1)})$$

Passo 4 Elimina il parametro  $\theta_{(p)}$  per integrazione e considera il modello marginale con  $p - 1$  gruppi di parametri

## Algoritmo 2

Passo 5 Ripeti i passi 3 e 4 per  $\theta_{(j)}$ , per  $j = p - 1, \dots, 2$ .

Passo 6 Definisci

$$\begin{aligned} \pi_m(\boldsymbol{\theta}) &= \pi_m(\theta_{(p)} | \theta(1), \dots, \theta_{(p-1)}) \\ &\times \pi_m(\theta_{(p-1)} | \theta(1), \dots, \theta_{(p-2)}) \\ &\times \dots \times \pi_m(\theta_1) \end{aligned}$$

Passo 7 Normalizzazione di  $\pi_m$

$$\pi^r(\boldsymbol{\theta}) = \lim_{m \rightarrow \infty} \frac{\pi_m(\boldsymbol{\theta})}{\pi_m(\boldsymbol{\theta}_0)},$$

dove  $\boldsymbol{\theta}_0$  è un opportuno punto interno di  $\Omega$ .

Passo 8 Verifica che

$$E_{\theta} (KL(\pi_m(\boldsymbol{\theta}|\mathbf{x}), \pi(\boldsymbol{\theta}|\mathbf{x}))) \rightarrow 0$$

Note tecniche Nel calcolo effettivo di  $\pi^r$  gli aspetti più complessi riguardano

A Calcolo di

$$\exp\left\{\frac{1}{2} \int_{\Lambda_m(\theta)} \pi_m(\lambda|\theta) \log \frac{\det(H)}{H_{22}} d\lambda\right\}$$

B

$$\pi^r(\theta, \lambda) = \lim_{m \rightarrow \infty} \frac{k_m(\theta) \pi_m(\theta)}{k_m(\theta_0) \pi_m(\theta_0)} \frac{\pi_m(\lambda|\theta)}{\pi(\lambda|\theta_0)}$$

Il più delle volte il calcolo di [B] semplifica il passo [A].

Infatti

$$[A] \approx K_m + C_m \Psi(\theta) + D_m(\theta)$$

dove  $K_m \rightarrow \infty$ ,  $C_m \rightarrow C$ ,  $D_m \rightarrow 0$

Ne segue che la parte relativa ad [A] del limite [B] vale

$$\exp\left\{\frac{1}{2} C \Psi(\theta)\right\}$$

Matching Priors Una interpretazione del termine “non informativa” per una data  $\pi$  è che le inferenze conseguenti l’uso di  $\pi$  abbiano un buon comportamento frequentista. In particolare si guarda alla **Probabilità Frequentista di Ricoprimento** (PFR)

Data una legge a priori  $\pi$ , si considera

$$\pi(\cdot) \longrightarrow \pi(\cdot|X) \longrightarrow C_\pi(X, 1 - \alpha)$$

dove  $C$  è l’insieme di credibilità ad una coda.  
Se la PFR è tale che

$$P(\theta \in C_\pi(X, 1 - \alpha)|\theta) = 1 - \alpha + O(n^{-\frac{\gamma}{2}}),$$

allora  $\pi$  è una matching priors di ordine  $\gamma$ .

Tibshirani (1989) ha dimostrato che nel caso di parametro

reale d'interesse, ortogonale a tutti i parametri di disturbo,

$$\pi^T(\theta, \lambda) \propto g(\lambda) \sqrt{H_{11}(\theta, \lambda)}$$

è una matching prior ( $\forall g$ ) di ordine 1.

- Legami con  $\pi^J$  e con  $\pi^r$