

Basic Bayes

Introduzione all'inferenza bayesiana

Fulvio De Santis

8 novembre 2024

Basic Bayes – Introduzione all’inferenza bayesiana

Fulvio De Santis

Dipartimento di Scienze Statistiche – Sapienza Università di Roma

Indice

1	Modelli parametrici	3
1.1	Introduzione	3
1.2	Inferenza statistica: posizione del problema	3
1.3	Modello statistico di base	4
1.4	Famiglie posizione-scala	15
1.5	Modello statistico	18
1.6	Famiglie esponenziali	22
1.7	Problemi inferenziali	24
1.8	Impostazioni inferenziali	26
1.9	Esercizi	28
2	Verosimiglianza	29
2.1	Introduzione	29
2.2	Funzione di verosimiglianza	29
2.3	Stima puntuale	32
2.4	Stima mediante insiemi	35
2.5	Verifica di ipotesi	38
2.6	Informazione di Fisher osservata	39
2.7	Esempi notevoli	40
2.8	Principio di verosimiglianza	41
2.9	Esercizi	42
3	Sufficienza	43
3.1	Introduzione	43
3.2	Statistiche sufficienti	43
3.3	Statistiche sufficienti minimali	47
3.4	Statistiche sufficienti e famiglie esponenziali	52
3.5	Esercizi	54
4	Inferenza bayesiana	55
4.1	Introduzione	55
4.2	Elementi costitutivi	55
4.3	Analisi coniugata	58
4.3.1	Scelta dei valori degli iperparametri	65
4.4	Distribuzioni a priori mistura	66
4.5	Densità a posteriori di una funzione del parametro	68

4.6	Esercizi	71
5	Metodi per inferenze ipotetiche	75
5.1	Introduzione	75
5.2	Stima puntuale	75
5.3	Esempi notevoli	77
5.4	Stima mediante insiemi	85
5.5	Verifica di ipotesi	88
5.6	Fattore di Bayes	89
5.7	Esercizi	94
6	Analisi oggettiva	99
6.1	Introduzione	99
6.2	Metodo 1: casi limite di distribuzioni coniugate	100
6.3	Metodo 2: distribuzione a priori di Jeffreys	103
6.4	Esempi notevoli	105
6.5	Esercizi	107
7	Approssimazioni normali	109
7.1	Introduzione	109
7.2	Approssimazione normale della funzione di verosimiglianza	109
7.3	Approssimazioni normali della distribuzione a posteriori	112
7.4	Esercizi	117
8	Metodi per inferenze predittive	119
8.1	Introduzione	119
8.2	Distribuzioni predittive	119
8.3	Inferenza su funzioni scalari di dati futuri	121
8.4	Esempi notevoli	122
8.5	Esercizi	127

Premessa

Questo testo contiene una breve introduzione all'inferenza bayesiana. Il libro, rivolto primariamente, ma non esclusivamente, agli studenti dei corsi di laurea e laurea magistrale in Statistica, è il frutto delle lezioni tenute per l'insegnamento *Teoria statistica delle decisioni* presso l'Università Sapienza. I destinatari ideali del libro sono lettori che posseggono già nozioni di Probabilità e di Inferenza statistica. Tuttavia, i primi tre capitoli richiamano molte nozioni propedeutiche alla comprensione di ciò che segue: modelli statistici, verosimiglianza, sufficienza. I restanti ... capitoli sono dedicati rispettivamente a: elementi costitutivi dell'inferenza bayesiana (Cap. 4), metodi inferenziali per problemi ipotetici (Cap. 5), analisi oggettiva (Cap. 6), approssimazioni normali (Cap. 7), metodi per inferenze predittive (Cap. 8). Successivamente ...

Roma, ottobre 2024, fds

Capitolo 1

Modelli parametrici

1.1 Introduzione

Questo capitolo introduce il problema generale dell'inferenza statistica parametrica. L'esposizione è quindi dedicata primariamente al concetto di modello statistico di base e per n -ple campionarie indipendenti ed identicamente distribuite. Vengono poi introdotti i modelli posizione-scala e il concetto unificante di famiglia esponenziale. Infine, si schematizzano i principali problemi di inferenza ipotetica e predittiva e si fa cenno alle diverse possibili impostazioni inferenziali.

1.2 Inferenza statistica: posizione del problema

L'inferenza statistica è la disciplina che studia come utilizzare dei *dati* per ricostruire il meccanismo aleatorio che si suppone li abbia generati. Il problema si può formulare nel modo seguente. Si considera un fenomeno aleatorio osservabile di interesse, X , che può assumere valori nello spazio \mathcal{X} (*spazio campionario*). Se, ad esempio, il fenomeno aleatorio è il lancio di un dado, lo spazio campionario risulta essere $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$. Se invece X rappresenta il risultato di una misurazione di una grandezza che assume valori nell'intervallo di numeri reali compresi tra i valori a e b , lo spazio campionario coincide con l'intervallo di numeri reali $[a, b]$. Indichiamo con f_X la legge di probabilità di X . Se X è una v.a. discreta, come nel caso del lancio del dado, allora f_X è una *funzione di massa di probabilità*; se invece X è una v.a. assolutamente continua, come nel caso della misurazione sperimentale, f_X è una *funzione di densità di probabilità*. Assumiamo inoltre di non conoscere esattamente quale sia la legge di probabilità che ha generato i dati disponibili, ma di sapere solo che f_X appartiene ad una famiglia \mathcal{F} di leggi di probabilità. Ci poniamo il problema di individuare, tra gli elementi che costituiscono \mathcal{F} , la specifica legge f_X^* che regola il fenomeno aleatorio considerato. A tal fine si utilizzano i risultati osservati, denominati *dati campionari* (o osservazioni campionarie), ovvero un numero limitato di *realizzazioni* della v.a. X . Negli esempi più comuni, le leggi di probabilità $f_X \in \mathcal{F}$ dipendono da una quantità θ non nota, detta *parametro*, che può assumere valori nell'insieme Ω , detto *spazio parametrico*. In questo caso le leggi f_X le indichiamo con $f_X(\cdot; \theta)$ e l'insieme \mathcal{F} costituisce una famiglia *parametrica* di

leggi di probabilità:

$$\mathcal{F} = \{f_X(\cdot; \theta), \Omega\}.$$

Il parametro *indicizza* le leggi di probabilità in \mathcal{F} , nel senso che, al variare di θ , si ottengono tutti gli elementi della famiglia \mathcal{F} : in corrispondenza di ciascun elemento di Ω si ottiene un unico membro della famiglia \mathcal{F} . Si ha quindi una corrispondenza biunivoca tra elementi di Ω e leggi in \mathcal{F} ¹. L'obiettivo dell'inferenza statistica è di stabilire, sulla base delle osservazioni campionarie, quale elemento di \mathcal{F} (ovvero quale valore θ^* del parametro θ) individua la legge di X che ha generato i dati. In altre parole, si assume di conoscere l'espressione esplicita della legge di probabilità in \mathcal{F} della v.a. X che ha generato le osservazioni (ad esempio possiamo supporre che la funzione di densità sia quella di una v.a. normale, esponenziale, beta, oppure che la funzione di massa di probabilità sia quella di una v.a. bernoulliana o di Poisson), ma di non conoscere quale, tra gli elementi della famiglia, ha generato i dati.

1.3 Modello statistico di base

I tre elementi che caratterizzano dal punto di vista probabilistico una v.a. X la cui legge di probabilità dipende da un parametro θ sono:

- \mathcal{X} l'insieme dei valori che la v.a. X può assumere;
- $f_X(\cdot; \theta)$ la legge di probabilità per la v.a. X ;
- Ω lo spazio parametrico.

I tre elementi descritti si raccolgono nella terna

$$\{\mathcal{X}, f_X(\cdot; \theta), \Omega\},$$

che chiamiamo *modello statistico di base* o *modello probabilistico* per la v.a. X . Va osservato che, in generale, X può essere una variabile aleatoria anche multidimensionale e Ω uno spazio di dimensione superiore ad uno (problemi *multiparametrici*). Nella maggior parte degli esempi che seguono, X è una v.a. semplice e il parametro θ è uno scalare (Ω è un sottoinsieme di \mathbb{R}), oppure un vettore di due elementi, $\theta = (\theta_1, \theta_2)$, con Ω sottoinsieme di $\mathbb{R} \times \mathbb{R}$.

Illustriamo ora le caratteristiche dei modelli statistici per le principali variabili aleatorie discrete e assolutamente continue.

Modelli statistici parametrici per v.a. discrete

Le v.a. discrete assumono un numero finito o al più numerabile di valori, che sono quindi in corrispondenza biunivoca con l'insieme dei numeri naturali, o con un suo sottoinsieme. La funzione $f_X(\cdot; \theta)$ indica in questo caso la *funzione di massa di probabilità* della v.a. X . Nel caso di un numero k finito di valori², che indichiamo con x_1, x_2, \dots, x_k , la *distribuzione di probabilità* di X è data da

¹Si parla, in questo caso, di modello *identificato*.

²L'estensione al caso di una infinità numerabile di modalità è immediata.

x_1	x_2	\dots	x_k
$f_X(x_1; \theta)$	$f_X(x_2; \theta)$	\dots	$f_X(x_k; \theta)$

dove

$$f_X(x_i; \theta) = \mathbb{P}(X = x_i; \theta), \quad i = 1, 2, \dots, k,$$

è la probabilità che X sia uguale a x_i : Questa probabilità dipende dal parametro incognito θ . Per la v.a. X con distribuzione di probabilità data dalla tabella di cui sopra, *valore atteso* e *varianza*³ sono definiti rispettivamente da:

$$\mathbb{E}_\theta[X] = \sum_{i=1}^k x_i f_X(x_i; \theta), \quad \mathbb{V}_\theta[X] = \sum_{i=1}^k (x_i - \mathbb{E}_\theta[X])^2 f_X(x_i; \theta).$$

In questa notazione si esplicita la dipendenza del valore atteso e della varianza della v.a. X dal parametro incognito θ .

1.1 Esempio (modello bernoulliano). Per la v.a. discreta X *bernoulliana* il modello statistico di base è individuato dalla terna

$$\{\mathcal{X} = \{0, 1\}, \quad f_X(x; \theta) = \theta^x (1 - \theta)^{1-x} I_{\{0,1\}}(x), \quad \Omega = [0, 1]\},$$

dove $I_{\{0,1\}}(x) = \mathbb{I}_{\mathcal{X}}(x)$ denota la funzione indicatrice⁴ dell'insieme \mathcal{X} , supporto della v.a. X . Nel seguito indicheremo che una v.a. discreta X ha distribuzione bernoulliana di parametro θ con la notazione $X|\theta \sim \text{Ber}(\theta)$. Il valore atteso e la varianza della v.a. bernoulliana sono:

$$\mathbb{E}_\theta[X] = \theta, \quad \mathbb{V}_\theta[X] = \theta(1 - \theta).$$

1.2 Esempio (modello binomiale). Se X_1, \dots, X_k sono k v.a. bernoulliane indipendenti di parametro θ , la v.a.

$$X = \sum_{i=1}^k X_i$$

prende il nome di v.a. *binomiale* di parametro θ relativa a k prove. La v.a. binomiale emerge pertanto dal cosiddetto *schema delle prove ripetute* e rappresenta il numero aleatorio di successi che si possono avere in k prove bernoulliane indipendenti in cui

³Nel caso di v.a. discrete che possono assumere una infinità numerabile di valori, come la v.a. di Poisson, valore atteso e varianza esistono se le corrispondenti serie che li definiscono sono convergenti.

⁴Definiamo *funzione indicatrice* di un insieme A la seguente funzione:

$$\mathbb{I}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \in A^C \end{cases}.$$

la probabilità di successo in ciascuna prova è costante e pari a θ . Per questa v.a. discreta il modello statistico di base è individuato dalla terna:

$$\left\{ \mathcal{X} = \{0, 1, 2, \dots, k\}, \quad f_X(x; \theta) = \binom{k}{x} \theta^x (1 - \theta)^{k-x} \mathbb{I}_{\{0, \dots, k\}}(x), \quad \Omega = [0, 1] \right\}.$$

Nel seguito indicheremo che una v.a. discreta X ha distribuzione binomiale di parametro θ ed è relativa a k prove, con la notazione $X|\theta \sim \text{Bin}(k, \theta)$. Il valore atteso e la varianza della v.a. binomiale di parametro θ sono:

$$\mathbb{E}_\theta[X] = k\theta, \quad \mathbb{V}_\theta[X] = k\theta(1 - \theta).$$

1.3 Osservazione/i. Nei casi in cui non è noto neanche il valore k , il modello binomiale presenta due parametri incogniti.

1.4 Osservazione/i. Per $k = 1$, la v.a. binomiale di parametro θ corrisponde a una bernoulliana di parametro θ .

1.5 Esempio (modello geometrico). Per la v.a. discreta X *geometrica* il modello statistico di base è individuato dalla terna:

$$\left\{ \mathcal{X} = \overline{\mathbb{N}}, \quad f_X(x; \theta) = (1 - \theta)^x \theta \cdot \mathbb{I}_{\overline{\mathbb{N}}}(x), \quad \Omega = [0, 1] \right\},$$

dove $\overline{\mathbb{N}} = 0, 1, 2, \dots$. Nel seguito indicheremo che una v.a. discreta X ha distribuzione geometrica di parametro θ con la notazione: $X|\theta \sim \text{Geom}(\theta)$. Il valore atteso e la varianza della v.a. geometrica sono:

$$\mathbb{E}_\theta[X] = \frac{1 - \theta}{\theta}, \quad \mathbb{V}_\theta[X] = \frac{1 - \theta}{\theta^2}.$$

Anche la v.a. geometrica può essere derivata dallo schema delle prove ripetute. Se X_1, X_2, \dots sono una successione di v.a. bernoulliane indipendenti di parametro θ , la v.a. geometrica conta il numero totale di insuccessi (il numero totale di $X_i = 0$) prima che si abbia un successo. Se $Y|\theta \sim \text{Geom}(\theta)$, allora

$$\mathbb{P}(Y = k; \theta) = \mathbb{P}(X_1 = 0, \dots, X_k = 0, X_{k+1} = 1; \theta) = (1 - \theta)^k \theta.$$

La v.a. geometrica rappresenta quindi il numero aleatorio (tempo di attesa) di prove bernoulliane prima di ottenere il primo successo.

Per il modello si può considerare una formulazione alternativa. Supponendo sempre che θ indichi la probabilità di successo di una singola prova bernoulliana, possiamo indicare con X la variabile aleatoria che indica il numero di lanci da effettuare da effettuare (in condizioni di indipendenza e identica distribuzione) per avere il primo

successo. In questo caso abbiamo che la v.a. X ha funzione di massa di probabilità definita da

$$f(x|\theta) = \theta(1 - \theta)^{x-1}, \quad x = 1, 2, 3, \dots, \quad \theta \in [0, 1].$$

In questo caso $\mathbb{E}[X] = 1/\theta$ e $\mathbb{V}[X] = (1 - \theta)/\theta^2$.

1.6 Esempio (modello di Poisson). Per la v.a. discreta X di *Poisson* di parametro θ , il modello statistico di base è individuato dalla terna:

$$\left\{ \mathcal{X} = \overline{\mathbb{N}}, \quad f_X(x; \theta) = \mathbb{P}(X = x; \theta) = \frac{e^{-\theta} \theta^x}{x!} \cdot \mathbb{I}_{\overline{\mathbb{N}}}(x), \quad \Omega = \mathbb{R}^+ \right\}.$$

Come per la v.a. geometrica, anche la v.a. di Poisson può assumere una infinità numerabile di valori. Nel seguito indicheremo che una v.a. discreta X ha distribuzione di Poisson di parametro θ con la notazione $X|\theta \sim \text{Pois}(\theta)$. Il valore atteso e la varianza della v.a. di Poisson sono:

$$\mathbb{E}_\theta[X] = \mathbb{V}_\theta[X] = \theta.$$

Modelli statistici parametrici per v.a. assolutamente continue

Le v.a. assolutamente continue assumono una infinità non numerabile di valori. Gli elementi di \mathcal{X} sono quindi in corrispondenza biunivoca con i numeri reali, \mathbb{R} , o con un loro sottoinsieme. In questo caso $f_X(\cdot; \theta)$ indica la *funzione di densità di probabilità*⁵ della v.a. X . Valore atteso e varianza di X sono ora definiti dai seguenti integrali, che assumiamo esistere (finiti):

$$\mathbb{E}_\theta[X] = \int_{\mathbb{R}} x f_X(x; \theta) dx, \quad \mathbb{V}_\theta[X] = \int_{\mathbb{R}} (x - \mathbb{E}_\theta[X])^2 f_X(x; \theta) dx.$$

1.7 Esempio (modello esponenziale negativo). Per la v.a. assolutamente continua X *esponenziale negativa* di parametro θ , il modello statistico di base è individuato dalla terna

$$\left\{ \mathcal{X} = \mathbb{R}^+, \quad f_X(x; \theta) = \theta e^{-\theta x} \cdot \mathbb{I}_{\mathbb{R}^+}(x), \quad \Omega = \mathbb{R}^+ \right\}.$$

⁵Per la definizione formale di funzione di densità di probabilità si veda, ad esempio, Dall'Aglio G. (2003). *Calcolo delle Probabilità*, Zanichelli (pag. 93). Ricordiamo che $f: \mathbb{R} \rightarrow \mathbb{R}^+$ è una funzione di densità se: $f(x) \geq 0$, $x \in \mathbb{R}$ e $\int_{\mathbb{R}} f(x) dx = 1$. Ricordiamo inoltre che il *supporto* di una v.a. è costituito dall'insieme

$$\mathcal{X} = \{x \in \mathbb{R} : f(x) > 0\}.$$

Nel seguito indicheremo che una v.a. assolutamente continua X ha distribuzione esponenziale negativa di parametro θ con la notazione $X|\theta \sim \text{EN}(\theta)$. Il valore atteso e la varianza della v.a. esponenziale negativa sono rispettivamente uguali a:

$$\mathbb{E}_\theta[X] = \frac{1}{\theta}, \quad \mathbb{V}_\theta[X] = \frac{1}{\theta^2}.$$

1.8 Esempio (modello esponenziale). Una semplice variazione di parametrizzazione nel modello precedente dà luogo al modello esponenziale, che si ottiene sostituendo il parametro θ con il suo inverso $1/\theta$. Per la v.a. assolutamente continua X *esponenziale* di parametro θ il modello statistico di base è individuato dalla terna:

$$\left\{ \mathcal{X} = \mathbb{R}^+, \quad f_X(x; \theta) = \frac{1}{\theta} e^{-x/\theta} \cdot \mathbb{I}_{\mathbb{R}^+}(x), \quad \Omega = \mathbb{R}^+ \right\}.$$

Nel seguito indicheremo che una v.a. assolutamente continua X ha distribuzione esponenziale di parametro θ con la notazione: $X|\theta \sim \text{Esp}(\theta)$. Il valore atteso e la varianza della v.a. esponenziale sono rispettivamente uguali a:

$$\mathbb{E}_\theta[X] = \theta, \quad \mathbb{V}_\theta[X] = \theta^2.$$

La Figura (1.1) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori del parametro.

1.9 Osservazione/i Si osservi che i modelli esponenziale-negativo ed esponenziale si differenziano solo per la *parametrizzazione*: al variare di $\theta \in \mathbb{R}^+$ le due famiglie \mathcal{F} contengono gli stessi modelli. Con la parametrizzazione usata nel modello esponenziale, il parametro θ è detto parametro di **scala** (in inglese **scale**); con quella adottata nel modello esponenziale-negativo, θ è un parametro di **tasso** (di solito indicato con il termine anglosassone **rate**).

1.10 Esempio (modello normale). Per la v.a. assolutamente continua X *normale* di parametro vettoriale $\boldsymbol{\theta} = (\theta_1, \theta_2)$, il modello statistico di base è individuato dalla terna:

$$\left\{ \mathcal{X} = \mathbb{R}, \quad f_X(x; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\theta_2}} \exp \left\{ -\frac{1}{2\theta_2} (x - \theta_1)^2 \right\} \cdot \mathbb{I}_{\mathbb{R}}(x), \quad \Omega = \mathbb{R} \times \mathbb{R}^+ \right\}.$$

Nel seguito indicheremo che una v.a. assolutamente continua X ha distribuzione normale di parametri $\boldsymbol{\theta} = (\theta_1, \theta_2)$ con la notazione $X|\boldsymbol{\theta} \sim \text{N}(\theta_1, \theta_2)$. Il valore atteso e la varianza della v.a. normale sono:

$$\mathbb{E}_\theta[X] = \theta_1, \quad \mathbb{V}_\theta[X] = \theta_2.$$

Dal modello generale introdotto si ottengono due casi particolari. Il primo si ha quando è noto il parametro θ_2 e incognito il valore atteso θ_1 ; il secondo, viceversa,

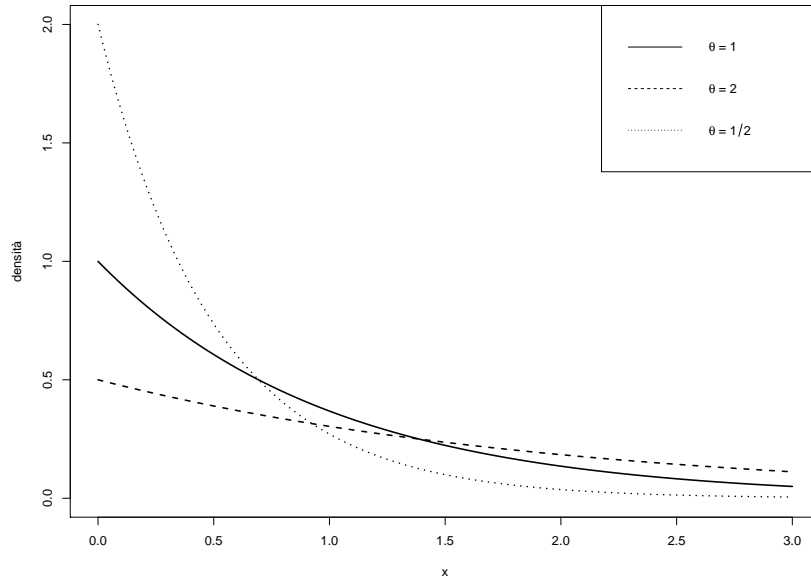


Figura 1.1: Grafico della funzione di densità di una v.a. esponenziale (per alcuni valori del parametro θ).

quando è noto il valore atteso θ_1 e incognita la varianza, θ_2 . Nel primo caso, posto $\theta_2 = \sigma_0^2$, il modello statistico di base è individuato dalla terna:

$$\left\{ \mathcal{X} = \mathbb{R}, \quad f_X(x; \theta) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left\{ -\frac{1}{2\sigma_0^2} (x - \theta)^2 \right\} \cdot \mathbb{I}_{\mathbb{R}}(x), \quad \Omega = \mathbb{R}. \right\}.$$

La Figura (1.2) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori del parametro θ (valore atteso della v.a. X).

Se invece il valore atteso di X è noto e pari a μ_0 , allora il solo parametro incognito è $\theta_2 = \theta$ e il modello statistico di base è individuato dalla terna:

$$\left\{ \mathcal{X} = \mathbb{R}, \quad f_X(x; \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp \left\{ -\frac{1}{2\theta} (x - \mu_0)^2 \right\} \cdot \mathbb{I}_{\mathbb{R}}(x), \quad \Omega = \mathbb{R}^+. \right\}.$$

La Figura (1.3) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori del parametro θ (varianza della v.a. X).

1.11 Esempio (modello gamma). Per la v.a. assolutamente continua X con distribuzione *gamma* di parametro $\theta = (\theta_1, \theta_2)$, il modello statistico di base è

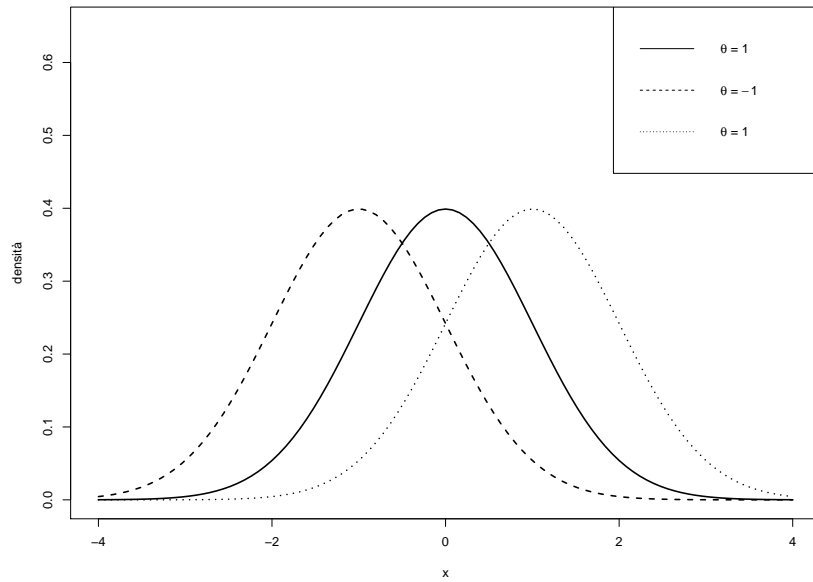


Figura 1.2: Grafico della funzione di densità di una v.a. normale per alcuni valori del valore atteso θ e con varianza $\sigma_0^2 = 1$.

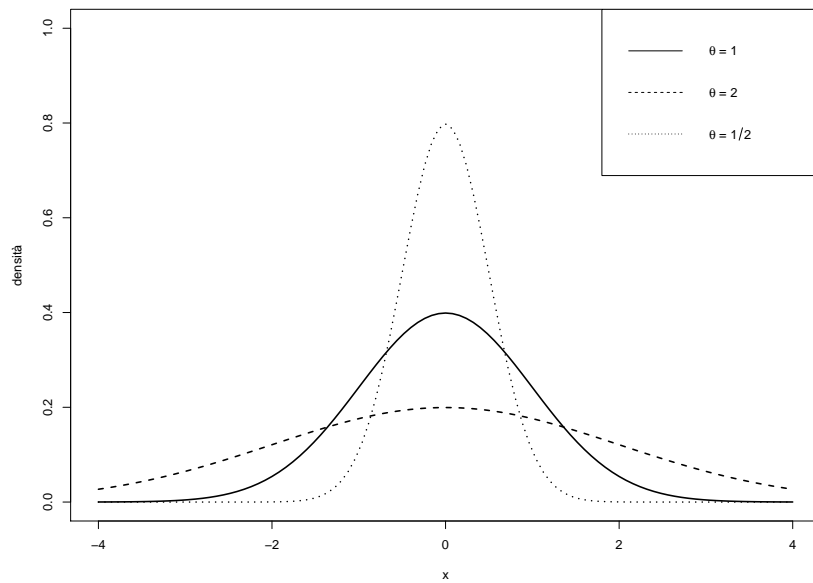


Figura 1.3: Grafico della funzione di densità di una v.a. normale per alcuni valori della varianza θ e con valore atteso $\mu_0 = 0$.

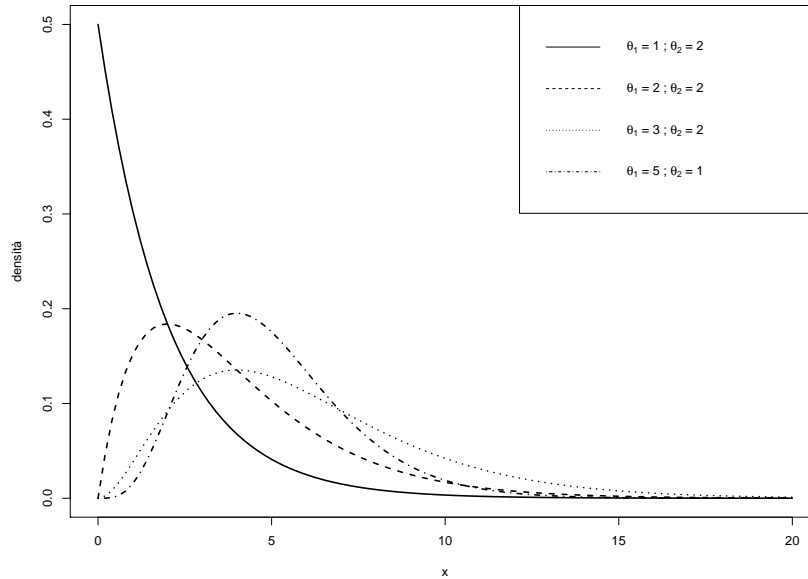


Figura 1.4: Grafico della funzione di densità di una v.a. gamma per alcuni valori dei parametri θ_1 e θ_2 (con θ_2 parametro di scala).

individuato dalla terna⁶:

$$\left\{ \mathcal{X} = \mathbb{R}^+, \quad f_X(x; \boldsymbol{\theta}) = \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} x^{\theta_1-1} e^{-\theta_2 x} \cdot \mathbb{I}_{\mathbb{R}^+}(x), \quad \Omega = \mathbb{R}^+ \times \mathbb{R}^+ \right\}.$$

Nel seguito indicheremo che una v.a. assolutamente continua X ha distribuzione gamma di parametri $\boldsymbol{\theta} = (\theta_1, \theta_2)$ con la notazione $X|\boldsymbol{\theta} \sim \text{Gamma}(\theta_1, \theta_2)$. Il valore atteso e la varianza della v.a. gamma sono:

$$\mathbb{E}_{\boldsymbol{\theta}}[X] = \frac{\theta_1}{\theta_2}, \quad \mathbb{V}_{\boldsymbol{\theta}}[X] = \frac{\theta_1}{\theta_2^2}.$$

La Figura (1.4) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori dei parametri θ_1 e θ_2 .

1.12 Osservazione/i.

1. Se X_1, \dots, X_k sono k v.a. $\text{Gamma}(\theta_{1,i}, \theta_2)$ indipendenti e non necessariamente somiglianti, ciascuna con parametro $\theta_{1,i}$, la v.a. $\sum_{i=1}^k X_i \sim \text{Gamma}(\sum_{i=1}^k \theta_{1,i}, \theta_2)$. Se le k v.a. sono anche somiglianti ($\theta_{1,i} = \theta_1, i = 1, \dots, k$), allora $\sum_{i=1}^k X_i \sim \text{Gamma}(k\theta_1, \theta_2)$.

⁶Si ricordi che la funzione gamma $\Gamma : \mathbb{R}^+ \mapsto \mathbb{R}^+$ è definita come segue: $\Gamma(y) = \int_0^{+\infty} t^{y-1} e^{-t} dt, y > 0$. La funzione gamma soddisfa la relazione ricorsiva $\Gamma(y+1) = y\Gamma(y)$. Nel caso in cui si consideri un numero intero n , si ha: $\Gamma(n) = (n-1)!$

2. Se X_1, \dots, X_k sono k v.a. $\text{EN}(\theta) = \text{Ga}(1, \theta)$ indipendenti, la v.a. $\sum_{i=1}^k X_i \sim \text{Gamma}(k, \theta)$.
3. A volte viene usata una parametrizzazione diversa da quella riportata. In particolare il parametro θ_2 è sostituito dal suo inverso. In questo caso la funzione di densità della v.a. $\text{Gamma}(\theta_1, \theta_2)$ risulta essere:

$$f_X(x; \boldsymbol{\theta}) = \frac{1}{\Gamma(\theta_1) \theta_2^{\theta_1}} x^{\theta_1-1} e^{-\frac{x}{\theta_2}}, \quad x > 0.$$

Inoltre

$$\mathbb{E}_\theta[X] = \theta_1 \theta_2, \quad \mathbb{V}_\theta[X] = \theta_1 \theta_2^2.$$

4. Nelle parametrizzazioni introdotte il parametro θ_1 è detto parametro di forma (oppure **shape**). Nella prima parametrizzazione il parametro θ_2 è detto parametro di tasso (oppure **rate**); nella seconda detto parametro di scala (oppure **scale**).
5. La distribuzione $\text{Gamma}(1, \text{rate} = \theta)$ coincide con la v.a. $\text{EN}(\theta)$.
6. La v.a. $\text{Gamma}(1, \text{scale} = \theta)$ coincide con la v.a. $\text{Exp}(\theta)$.
7. Utilizzando la parametrizzazione **scale** introdotta e ponendo $\theta_2 = 2$ e $\theta_1 = \frac{\nu}{2}$, $\nu > 0$, si ottiene la funzione di densità della v.a. Chi quadrato di parametro $\nu > 0$:

$$f_X(x; \nu) = \frac{1}{\Gamma(\frac{\nu}{2}) 2^{\nu/2}} x^{(\nu/2)-1} e^{-x/2} \cdot \mathbb{I}_{\mathbb{R}^+}(x).$$

Per il parametro ν si usa la denominazione di *gradi di libertà* e inoltre si ha:

$$E_\nu[X] = \nu, \quad V_\nu[X] = 2\nu.$$

1.13 Esempio (modello di Cauchy). Per la v.a. assolutamente continua X *Cauchy* di parametro $\boldsymbol{\theta} = (\theta_1, \theta_2)$, con $\theta_1 \in \mathbb{R}$ e $\theta_2 \in \mathbb{R}^+$, il modello statistico di base è individuato dalla terna:

$$\left\{ \mathcal{X} = \mathbb{R}, f_X(x; \boldsymbol{\theta}) = \frac{1}{\pi \theta_2 \left[1 + \left(\frac{x - \theta_1}{\theta_2} \right)^2 \right]} \cdot \mathbb{I}_{\mathbb{R}}(x), \Omega = \mathbb{R} \times \mathbb{R}^+ \right\}.$$

Nel seguito indicheremo che una v.a. assolutamente continua X ha distribuzione di Cauchy di parametri $\boldsymbol{\theta} = (\theta_1, \theta_2)$ con la notazione $X|\boldsymbol{\theta} \sim \text{Cau}(\theta_1, \theta_2)$. Il valore atteso e la varianza della v.a. di Cauchy non esistono. Il parametro θ_1 , coincidente con la moda e la mediana della v.a. di Cauchy, è un parametro di *posizione*. Il secondo parametro (o meglio, come vedremo nel seguito, la sua radice quadrata) è invece un parametro di *scala*, che controlla il grado di dispersione della densità rispetto al parametro θ_1 . La densità della v.a. di Cauchy è simmetrica rispetto al parametro

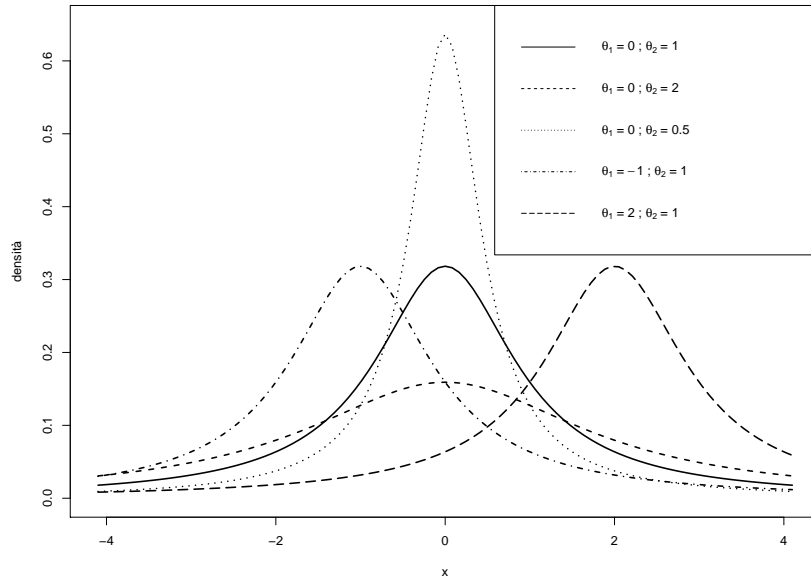


Figura 1.5: Grafico della funzione di densità di una v.a. Cauchy per alcuni valori dei parametri θ_1 e θ_2 .

θ_1 . Va inoltre ricordato che la v.a. di Cauchy è un caso speciale della distribuzione t di Student. La Figura (1.5) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori dei parametri θ_1 e θ_2 . Si noti che il parametro θ_1 controlla la *posizione* del del punto di massimo della funzione di densità, mentre il parametro θ_2 controlla il grado di dispersione (ovvero la *scala*) della densità intorno al suo punto di massimo.

1.14 Esempio (modello beta). Per la v.a. assolutamente continua X *beta* di parametro $\boldsymbol{\theta} = (\theta_1, \theta_2)$, con $\theta_1, \theta_2 \in \mathbb{R}^+$, il modello statistico di base è individuato dalla terna:

$$\left\{ \mathcal{X} = [0, 1], f_X(x; \boldsymbol{\theta}) = \frac{1}{B(\theta_1, \theta_2)} x^{\theta_1-1} (1-x)^{\theta_2-1} \cdot \mathbb{I}_{[0,1]}(x), \Omega_{\mathbb{R}^+ \times \mathbb{R}^+} \right\},$$

dove

$$B(\theta_1, \theta_2) = \int_0^1 y^{\alpha-1} (1-y)^{\beta-1} dy = \frac{\Gamma(\theta_1)\Gamma(\theta_2)}{\Gamma(\theta_1 + \theta_2)}.$$

Nel seguito indicheremo che una v.a. assolutamente continua X ha distribuzione beta di parametri $\boldsymbol{\theta} = (\theta_1, \theta_2)$ con la notazione $X|\boldsymbol{\theta} \sim \text{Beta}(\theta_1, \theta_2)$. Il valore atteso e la varianza della v.a. beta sono:

$$\mathbb{E}_{\boldsymbol{\theta}}[X] = \frac{\theta_1}{\theta_1 + \theta_2}, \quad \mathbb{V}_{\boldsymbol{\theta}}[X] = \frac{\theta_2 \theta_1}{(\theta_1 + \theta_2)^2 (\theta_1 + \theta_2 + 1)}.$$

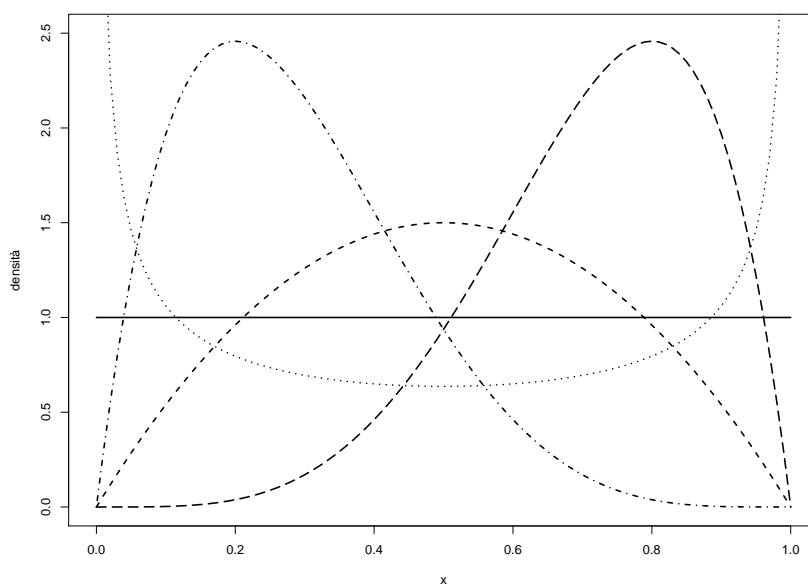


Figura 1.6: Grafico della funzione di densità di una v.a. beta per alcuni valori dei parametri θ_1 e θ_2 : densità uniforme ($\theta_1 = \theta_2 = 1$), con moda in $x = 1/2$ ($\theta_1 = \theta_2 = 2$), con moda in $x = 1/5$ ($\theta_1 = 2, \theta_2 = 5$), con moda in $x = 4/5$ ($\theta_1 = 5, \theta_2 = 2$) e con punto di minimo in $x = 1/2$ ($\theta_1 = 1/2, \theta_2 = 1/2$).

La Figura (1.6) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori dei parametri θ_1 e θ_2 .

Modelli con supporto dipendente dal parametro

Fin qui abbiamo considerato degli esempi in cui il supporto della v.a. X non dipende dal parametro incognito. Consideriamo ora un esempio in cui invece l'insieme dei valori in cui $f_X(\cdot; \theta) > 0$ dipende da θ . In questo caso indichiamo la dipendenza del supporto dal parametro con la notazione \mathcal{X}_θ .

1.15 Esempio (modello uniforme). Per la v.a. assolutamente continua X uniforme di parametro $\theta = (\theta_1, \theta_2)$, con $\theta_1, \theta_2 \in \mathbb{R}$ e $\theta_1 < \theta_2$, si ha che

$$f_X(x; \theta) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & x \in [\theta_1, \theta_2] \\ 0 & x \notin [\theta_1, \theta_2] \end{cases} = \frac{1}{\theta_2 - \theta_1} \cdot \mathbb{I}_{[\theta_1, \theta_2]}(x).$$

Il supporto della v.a. dipende dal parametro incognito θ . Il modello statistico di base è individuato dalla terna:

$$\left\{ \mathcal{X}_\theta = [\theta_1, \theta_2], f_X(x; \theta) = \frac{1}{\theta_2 - \theta_1} \cdot \mathbb{I}_{[\theta_1, \theta_2]}(x), \Omega = \{(\theta_1, \theta_2) \mathbb{R} \times \mathbb{R}, \theta_1 < \theta_2\} \right\}.$$

Nel seguito indicheremo che una v.a. assolutamente continua X ha distribuzione uniforme di parametri $\theta = (\theta_1, \theta_2)$ con la notazione $X|\theta \sim \text{Unif}(\theta_1, \theta_2)$. Il valore atteso e la varianza della v.a. uniforme sono:

$$\mathbb{E}_\theta[X] = \frac{\theta_2 + \theta_1}{2}, \quad \mathbb{V}_\theta[X] = \frac{(\theta_2 - \theta_1)^2}{12}.$$

Un caso particolarmente importante (frequentemente usato nel seguito) è quello del modello uniforme in $[0, \theta]$ che si ottiene ponendo $\theta_1 = 0$ e $\theta_2 = \theta > 0$.

1.4 Famiglie posizione-scala

Alcuni dei modelli per v.a. assolutamente continue considerati in questo capitolo appartengono a classi di distribuzioni denominate famiglie di *posizione*, famiglie di *scala*, famiglie *posizione-scala*. Tra questi ci sono i modelli esponenziale, esponenziale negativo, normale, gamma, Cauchy. Partendo da una funzione di densità $f(\cdot)$, denominata *standard*, con opportune trasformazioni è possibile definire tre famiglie parametriche utilizzando il seguente risultato: date due costanti $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}^+$, la funzione

$$g(x; \mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$$

è una funzione di densità di probabilità. Per verificare quanto affermato è sufficiente mostrare che $g(x; \mu, \sigma) \geq 0 \forall x \in \mathbb{R}$ e che $\int_{-\infty}^{+\infty} g(x; \mu, \sigma) dx = 1$. La positività di g discende da quella di f per ogni valore reale e da quella di σ . Inoltre, considerando

la sostituzione $y = (x - \mu)/\sigma$ (da cui $dx = \sigma dy$) si ha che $\int_{-\infty}^{+\infty} g(x; \mu, \sigma) dx = \int_{-\infty}^{+\infty} f(y) dy = 1$.

1.16 Definizione. Sia $f(x)$ una qualsiasi funzione di densità in \mathbb{R} .

1. La famiglia di funzioni di densità $\{g_p(\cdot; \mu), \mu \in \mathbb{R}\}$, definita ponendo

$$g_p(x; \mu) = f(x - \mu)$$

e ottenuta al variare di $\mu \in \mathbb{R}$ è denominata *famiglia di posizione con funzione di densità standard* f . Il parametro μ è denominato *parametro di posizione* della famiglia.

2. La famiglia di funzioni di densità $\{g_s(\cdot; \sigma), \sigma \in \mathbb{R}^+\}$, definita ponendo

$$g_s(x; \sigma) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$$

e ottenuta al variare di $\sigma \in \mathbb{R}^+$ è denominata *famiglia di scala con funzione di densità standard* f . Il parametro σ è denominato *parametro di scala* della famiglia.

3. La famiglia di funzioni di densità $\{g_{ps}(x; \mu, \sigma), \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$, definita ponendo

$$g_{ps}(x; \mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$$

e ottenuta al variare di $\mu \in \mathbb{R}$ e di $\sigma \in \mathbb{R}^+$ è denominata *famiglia di posizione-scala con funzione di densità standard* f (con μ e σ parametri di posizione e scala).

1.17 Osservazione/i

1. Il parametro di posizione μ ha l'effetto di traslare la funzione di densità standard sull'asse delle ascisse.
2. Il parametro di scala σ controlla il grado di dispersione della funzione f . In alcuni casi (ad esempio nel modello normale e nel modello Cauchy) le famiglie di scala si ottengono a partire da densità standard simmetriche rispetto a un valore x_0 , punto di massimo di f . In questi casi, introducendo il parametro di scala, la densità g_s resta simmetrica rispetto a x_0 ma valori di $\sigma > 1$ determinano (rispetto ad f) un "appiattimento" di g_s , mentre valori di $\sigma < 1$ determinano un addensamento di g_s intorno a x_0 .
3. La funzione di densità g_{ps} risulta traslata rispetto a f (con traslazione determinata da μ) e con un diverso grado di dispersione (determinato del valore di σ).

4. Le famiglie di posizione e di scala si possono ottenere da quella di posizione-scala ponendo, rispettivamente, $\sigma = 1$ e $\mu = 0$.
5. Se f è una funzione di densità simmetrica rispetto a zero, allora μ è la mediana delle densità g_p e g_{ps} .

1.18 Esempio (modello normale). Consideriamo la funzione di densità standard

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}, \quad x \in \mathbb{R}.$$

La corrispondente famiglia di posizione è l'insieme delle densità normali con valore atteso pari a μ e varianza unitaria:

$$g_p(x; \mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x - \mu)^2\right\}, \quad x \in \mathbb{R}, \quad \mu \in \mathbb{R}.$$

In questo caso il parametro di posizione coincide con il valore atteso μ della v.a. Partendo dalla stessa densità standard si mostra che il modello $N(0, \sigma^2)$ è una famiglia di scala (con parametro di scala σ) e che il modello $N(\mu, \sigma^2)$ costituisce la famiglia posizione-scala normale.

1.19 Esempio (modello di Cauchy). Se consideriamo la funzione di densità standard di Cauchy

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \quad x \in \mathbb{R},$$

la corrispondente famiglia di posizione di parametro μ risulta essere

$$g_p(x; \mu) = \frac{1}{\pi} \frac{1}{1 + (x - \mu)^2}, \quad x \in \mathbb{R}, \quad \mu \in \mathbb{R},$$

ovvero la famiglia delle densità di Cauchy con parametro di posizione μ (che però non coincide con il valore atteso della v.a. di Cauchy, che non esiste) e parametro di scala unitario. Analogamente, a famiglia di scala è

$$g_s(x; \sigma) = \frac{1}{\pi\sigma} \frac{1}{\left[1 + \left(\frac{x}{\sigma}\right)^2\right]}, \quad x \in \mathbb{R}, \quad \sigma \in \mathbb{R}^+.$$

Il modello $\text{Cau}(\mu, \sigma)$ è invece una famiglia posizione-scala.

1.20 Esempio (modello esponenziale). Considerando la densità standard $f(x) = e^{-x}$, $x \geq 0$, la famiglia di posizione esponenziale è

$$g_p(x; \mu) = e^{-(x-\mu)} I_{[0, \infty)}(x - \mu) = e^{-(x-\mu)} I_{[\mu, \infty)}(x), \quad \mu \in \mathbb{R}.$$

La famiglia di scala esponenziale risulta essere invece definita da

$$g_s(x; \sigma) = \frac{1}{\sigma} e^{-\frac{x}{\sigma}}, \quad x \in \mathbb{R}^+, \quad \sigma \in \mathbb{R}^+,$$

Si noti che g_s coincide con l'usuale f.ne di densità della v.a. esponenziale di parametro (di scala) σ .

1.21 Osservazione/i. Nel modello normale $N(\mu, \sigma^2)$, il valore atteso coincide con il parametro di posizione mentre la deviazione standard con il parametro di scala. Questo fatto non è vero in generale. Ad esempio, il modello posizione-scala $\text{Cau}(\mu, \sigma)$ non possiede né valore atteso né deviazione standard; nel modello $\text{Esp}(\sigma)$, il parametro di scala σ coincide con il valore atteso della variabile aleatoria.

1.22 Osservazione/i. I seguenti due risultati mettono in relazione tra loro le v.a. le cui funzioni di densità sono membri di una famiglia posizione-scala (che indichiamo con Y) con la v.a. (indicata con Z) associata alla funzione di densità standard della stessa famiglia. Per la dimostrazione si rimanda a Casella-Berger (2002, pp. 120-121).

1. Risultato 1. Sia $f(\cdot)$ una qualsiasi funzione di densità, μ un qualsiasi numero reale e σ un qualsiasi valore reale positivo. La v.a. Y ha funzione di densità $1/\sigma f((y - \mu)/\sigma)$ se e solo se esiste una v.a. Z con funzione di densità $f(\cdot)$ e tale che $Y = \sigma Z + \mu$.

Il risultato afferma sostanzialmente che la v.a. Y (la cui densità è un membro della famiglia di posizione-scala con parametri μ e σ) si ottiene per trasformazione lineare della v.a. Z (la cui densità coincide con la densità standard).

2. Risultato 2. Sia Z una v.a. con funzione di densità f . Supponiamo che esistano $\mathbb{E}(Z)$ e $\mathbb{V}(Z)$. Se Y è una v.a. con funzione di densità $1/\sigma f((y - \mu)/\sigma)$, allora $\mathbb{E}(Y) = \sigma \mathbb{E}(Z) + \mu$ e $\mathbb{V}(Y) = \sigma^2 \mathbb{V}(Z)$. Se $\mathbb{E}(Z) = 0$ e $\mathbb{V}(Z) = 1$, allora $\mathbb{E}(Y) = \mu$ e $\mathbb{V}(Y) = \sigma^2$.

1.5 Modello statistico

Si è detto che l'idea alla base dell'inferenza statistica è l'individuazione - nel modo più accurato possibile - del valore di θ , che indichiamo con θ^* , al quale corrisponde il membro $f_X(\cdot; \theta^*)$ di \mathcal{F} che regola il meccanismo aleatorio che ha generato i dati. L'individuazione (approssimativa) di θ^* avviene attraverso l'uso di *dati campionari*. Per questo scopo si effettuano n prove (n lanci della moneta, n lanci del dado, la somministrazione del farmaco a n soggetti, la rilevazione di un carattere su n individui estratti da una popolazione...) e si utilizzano i risultati osservati per l'operazione di inferenza su θ . Prima di effettuare l'esperimento, i risultati potenziali delle n

prove sono assimilabili a n variabili aleatorie X_1, X_2, \dots, X_n , in cui la generica X_i rappresenta la v.a. “risultato della prova i -esima”. Il vettore aleatorio

$$\mathbf{X}_n = (X_1, X_2, \dots, X_i, \dots, X_n)$$

costituisce il *campione*, di *numerosità* (o *ampiezza*) n . L'insieme delle possibili realizzazioni del campione \mathbf{X}_n , denominato *spazio campionario* e indicato con \mathcal{X}^n , è costituito da tutte le possibili n -ple

$$\mathbf{x}_n = (x_1, x_2, \dots, x_n)$$

che possiamo osservare. Una realizzazione \mathbf{x}_n di \mathbf{X}_n si chiama *campione osservato* (o *realizzato*). In pratica quindi, ciascun valore x_i rappresenta il valore osservato della v.a. X_i , relativa alla prova i -esima. Poichè la legge di probabilità di ciascuna v.a. X_i , $f_{X_i}(\cdot; \theta)$, dipende dal parametro incognito θ , anche la legge di probabilità della v.a. multipla X_1, \dots, X_n , che indichiamo con

$$f_{X_1 \dots X_n}(\cdot, \dots, \cdot; \theta)$$

dipende dallo stesso parametro incognito $\theta \in \Omega$. Per semplicità, indichiamo la legge $f_{X_1 \dots X_n}$ anche con le notazioni $f_{\mathbf{X}_n}$ e f_n e con

$$f_n(\mathbf{x}_n; \theta) = f_n(x_1, \dots, x_n; \theta) = f_{X_1 \dots X_n}(x_1, \dots, x_n; \theta)$$

il valore della funzione in corrispondenza del campione \mathbf{x}_n . Si noti che f_n indica la funzione di densità di \mathbf{X}_n , quando le v.a. X_i di \mathbf{X}_n sono assolutamente continue; oppure la funzione di massa di probabilità, nel caso di v.a. discrete⁷.

Così come abbiamo fatto per la v.a. di base X , possiamo ora definire, con riferimento al campione \mathbf{X}_n e alla sua legge di probabilità, il *modello statistico*. Questo è costituito dalla terna:

$$\{\mathcal{X}^n, f_n(\cdot; \theta), \Omega\}.$$

Useremo anche la notazione

$$\{\mathcal{X}^n, \mathcal{F}_n\},$$

dove

$$\mathcal{F}_n = (f_n(\cdot; \theta), \Omega)$$

rappresenta la famiglia parametrica di leggi di probabilità della v.a. multipla X_1, \dots, X_n .

Campioni casuali

Una situazione sperimentale particolare, ma molto rilevante, si ha quando si assume che le n v.a. X_1, \dots, X_n siano *indipendenti* (per ogni valore di θ) e *somiglianti* (o *identicamente distribuite*, sinteticamente: i.i.d.). In questo caso il campione

⁷Nel caso di v.a. assolutamente continue $f_n : \mathcal{X}^n \rightarrow \mathbb{R}^+$, mentre per le v.a. discrete $f_n : \mathcal{X}^n \rightarrow [0, 1]$.

$\mathbf{X}_n = (X_1, \dots, X_n)$ prende il nome di *campione casuale*. L'assunzione di somiglianza implica che le leggi di probabilità $f_{X_i}(\cdot; \theta)$ delle v.a. X_i siano tutte uguali tra loro, e uguali alla legge $f_X(\cdot; \theta)$ della v.a. di base, X (qualunque sia il vero valore di θ):

$$f_{X_1}(\cdot; \theta) = f_{X_2}(\cdot; \theta) \dots = f_{X_n}(\cdot; \theta) = f_X(\cdot; \theta).$$

Per l'assunzione di indipendenza, la probabilità (nel caso discreto) o la densità di probabilità (caso continuo) di osservare un generico campione (x_1, \dots, x_n) , elemento di \mathcal{X}^n , risulta uguale al prodotto delle singole funzioni di massa di probabilità o di densità di probabilità delle n v.a. aleatorie, nei punti x_1, \dots, x_n :

$$f_n(\mathbf{x}_n; \theta) = f_n(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_{X_i}(x_i; \theta) = \prod_{i=1}^n f_X(x_i; \theta). \quad (1.1)$$

La seconda uguaglianza è la conseguenza dell'ipotesi di indipendenza delle n v.a.; la terza della loro identica distribuzione. Si noti che, nell'ultima produttoria, l'indice i agisce unicamente sui valori x_i in cui la legge $f_X(\cdot; \theta)$ viene valutata, dal momento che il parametro delle leggi di probabilità delle n v.a. è, in virtù dell'ipotesi di uguale distribuzione, sempre lo stesso. Dall'ipotesi di indipendenza e identica distribuzione discende quindi una notevole semplificazione del problema. Infatti, la legge di probabilità della v.a. *moltiplica* n -dimensionale \mathbf{X}_n , (definita in \mathbb{R}^n o in $\overline{\mathbb{N}}^n$ e che assume valori in \mathbb{R}^+) può essere scritta come prodotto della legge di probabilità della v.a. *semplice* di base, X (definita in \mathbb{R} o in $\overline{\mathbb{N}}$ e a valori in \mathbb{R}^+), valutata negli n punti x_1, \dots, x_n di un campione.

Modelli statistici per campioni casuali

Con riferimento al campione casuale \mathbf{X}_n e alla sua legge di probabilità, il *modello statistico* è dato dalla terna

$$\left\{ \mathcal{X}^n, \prod_{i=1}^n f_X(x_i; \theta), \Omega \right\}. \quad (1.2)$$

Possiamo ora descrivere alcuni dei modelli statistici per campioni casuali relativi alle principali variabili aleatorie introdotte in precedenza⁸.

1.23 Esempio (modello bernoulliano). Supponiamo che $X_i | \theta \sim \text{Ber}(\theta)$, i.i.d. $i = 1, \dots, n$. Lo spazio \mathcal{X}^n è quindi costituito da tutte le n -ple distinte di 0 e di 1: $\mathcal{X}^n = \{0, 1\}^n$. Dalla (1.1) si ottiene:

$$f_n(\mathbf{x}_n; \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1 - x_i} = \theta^{n\bar{x}_n} (1 - \theta)^{n - n\bar{x}_n},$$

⁸Per semplicità di notazione, quando considereremo modelli per v.a. con supporto che non dipende dal parametro, nelle espressioni delle funzioni $f_n(\mathbf{x}_n; \theta)$ verrà omessa la funzione indicatrice dell'insieme \mathcal{X}^n .

dove $\bar{x}_n = \sum_{i=1}^n x_i/n$ è la media campionaria.

Il modello statistico per un campione casuale bernoulliano è individuato da:

$$\{\mathcal{X}^n = \{0, 1\}^n, \quad f_n(\mathbf{x}_n; \theta) = \theta^{n\bar{x}_n} (1 - \theta)^{n - n\bar{x}_n}, \quad \Omega = [0, 1]\}.$$

1.24 Esempio (modello esponenziale negativo). Supponiamo che $X_i|\theta \sim \text{EN}(\theta)$, i.i.d. $i = 1, \dots, n$. Ricordando che, per una v.a. esponenziale negativa di parametro θ , $f_X(x; \theta) = \theta e^{-\theta x}$, $x > 0$ si avrà

$$f_n(\mathbf{x}_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta) = \prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^n x_i}, \quad x_i > 0.$$

Il modello statistico per un campione casuale esponenziale negativo è quindi

$$\{\mathcal{X}^n = (\mathbb{R}^+)^n, \quad f_n(\mathbf{x}_n; \theta) = \theta^n e^{-\theta n\bar{x}_n}, \quad \Omega = \mathbb{R}^+\}.$$

1.25 Esempio (modello normale). Supponiamo che $X_i|\theta \sim \text{N}(\theta_1, \theta_2)$, i.i.d. $i = 1, \dots, n$. Dall'espressione della funzione di densità normale, per la (1.1) abbiamo che

$$f_n(\mathbf{x}_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta) = \frac{1}{(2\pi\theta_2)^{n/2}} \exp \left\{ -\frac{1}{2\theta_2} \sum_{i=1}^n (x_i - \theta_1)^2 \right\}, \quad x_i \in \mathbb{R}.$$

Il modello statistico per un campione casuale normale è:

$$\left\{ \mathcal{X}^n = \mathbb{R}^n, \quad f_n(\mathbf{x}_n; \theta) = (2\pi\theta_2)^{-n/2} \exp \left\{ -\frac{1}{2\theta_2} \sum_{i=1}^n (x_i - \theta_1)^2 \right\}, \quad \Omega = \mathbb{R} \times \mathbb{R}^+ \right\}.$$

In modo del tutto analogo si ottengono i casi in cui uno dei due parametri del modello è noto.

Modelli con supporto dipendente dal parametro

Nel caso di campioni casuali per v.a. con supporto \mathcal{X}_θ dipendente dal parametro incognito, è consigliabile ricorrere all'uso delle funzioni indicatrici del supporto stesso. Illustriamo come procedere con il seguente esempio.

1.26 Esempio (modello uniforme). Nel caso di un campione casuale estratto da una v.a. $X \sim \text{Unif}[\theta_1, \theta_2]$ con supporto della v.a. di base dipendente da $\boldsymbol{\theta} = [\theta_1, \theta_2]$. Ogni campione potenzialmente osservabile è costituito da n numeri che appartengono

ad un ipercubo nello spazio \mathbb{R}^n che ha per spigoli gli intervalli $[\theta_1, \theta_2]$ e cioè: $\mathcal{X}_\theta^n = \{[\theta_1, \theta_2]\}^n$. Ricordando che, per una v.a. uniforme di parametro $\theta = (\theta_1, \theta_2)$,

$$f_X(x; \theta) = \frac{1}{\theta_2 - \theta_1} \cdot \mathbb{I}_{[\theta_1, \theta_2]}(x),$$

abbiamo

$$f_n(\mathbf{x}_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta) = (\theta_2 - \theta_1)^{-n} \prod_{i=1}^n \mathbb{I}_{[\theta_1, \theta_2]}(x_i).$$

Osservando inoltre che $\prod_{i=1}^n \mathbb{I}_{[\theta_1, \theta_2]}(x_i) = \mathbb{I}_{(-\infty, x_{(1)}]}(\theta_1) \times \mathbb{I}_{[x_{(n)}, \infty)}(\theta_2)$, dove $x_{(1)} = \min\{x_1, \dots, x_n\}$ e $x_{(n)} = \max\{x_1, \dots, x_n\}$, il modello statistico risulta essere

$$\left\{ \mathcal{X}_\theta^n = \{[\theta_1, \theta_2]\}^n, f_n(\mathbf{x}_n; \theta) = \frac{1}{(\theta_2 - \theta_1)^n} \mathbb{I}_{(-\infty, x_{(1)}]}(\theta_1) \mathbb{I}_{[x_{(n)}, +\infty)}(\theta_2), \theta_1, \theta_2 \in \mathbb{R}, \theta_1 < \theta_2 \right\}.$$

1.6 Famiglie esponenziali

1.27 Definizione. Una famiglia di leggi di probabilità $\mathcal{F} = \{f_X(x; \theta), \Omega\}$ è detta *famiglia esponenziale uniparametrica* se $\Omega \subseteq \mathbb{R}$ e se esistono delle funzioni $h(x)$, $\eta(\theta)$, $T(x)$ e $B(\theta)$ tali che:

$$f_X(x; \theta) = h(x) \exp\{\eta(\theta)T(x) - B(\theta)\}. \quad (1.3)$$

La funzione $\eta(\theta)$ si chiama *parametro naturale* della famiglia esponenziale.

Tutti i modelli parametrici (con supporto indipendente da θ) considerati nei precedenti paragrafi, eccetto il modello di Cauchy, sono famiglie esponenziali. A titolo di esempio, illustriamo come si verifica l'appartenenza alla famiglia esponenziale del modello di Poisson.

1.28 Esempio (modello di Poisson). Sia $X|\theta \sim \text{Pois}(\theta)$, allora

$$f_X(x; \theta) = \frac{\theta^x}{x!} e^{-\theta} = \frac{1}{x!} \exp\{x \log \theta - \theta\}.$$

Ponendo

$$h(x) = \frac{1}{x!}, \quad \eta(\theta) = \log \theta, \quad T(x) = x, \quad B(\theta) = \theta,$$

si verifica che il modello $\text{Pois}(\theta)$ è una famiglia esponenziale con parametro naturale $\eta(\theta) = \log \theta$.

La definizione di famiglia esponenziale è facilmente estesa al caso in cui il parametro del modello sia vettoriale: $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$.

1.29 Definizione. La famiglia parametrica $\mathcal{F} = \{f_X(x; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Omega\}$, è detta *famiglia esponenziale k -parametrica* se $\Omega \subseteq \mathbb{R}^k, k > 1$, e se esistono delle funzioni di x e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ $h(x), \eta_1(\boldsymbol{\theta}), \dots, \eta_k(\boldsymbol{\theta}), T_1(x), \dots, T_k(x)$ e $B(\boldsymbol{\theta})$ tali che:

$$f_X(x; \boldsymbol{\theta}) = h(x) \exp \left\{ \sum_{j=1}^k \eta_j(\boldsymbol{\theta}) T_j(x) - B(\boldsymbol{\theta}) \right\}.$$

Le funzioni $\eta_1(\boldsymbol{\theta}), \dots, \eta_k(\boldsymbol{\theta})$ si chiamano *parametri naturali* della famiglia esponenziale.

1.30 Esempio (modello normale). Sia $X|\boldsymbol{\theta} \sim N(\theta_1, \theta_2)$, con entrambi i parametri incogniti. Si ha:

$$f_X(x; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\theta_2}} \exp \left\{ -\frac{1}{2\theta_2} x^2 + \frac{\theta_1}{\theta_2} x - \frac{2\theta_1^2}{\theta_2} \right\} = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\theta_2} x^2 + \frac{\theta_1}{\theta_2} x - \frac{\theta_1^2}{2\theta_2} - \frac{1}{2} \log \theta_2 \right\}.$$

Ponendo quindi:

$$h(x) = \frac{1}{\sqrt{2\pi}} \mathbb{I}_{\mathbb{R}}(x), \quad \eta_1(\boldsymbol{\theta}) = -\frac{1}{2\theta_2}, \quad \eta_2(\boldsymbol{\theta}) = \frac{\theta_1}{\theta_2}, \quad T_1(x) = x^2, \quad T_2(x) = x, \quad B(\boldsymbol{\theta}) = \frac{\theta_1^2}{2\theta_2} + \frac{1}{2} \log \theta_2,$$

si verifica che la famiglia $N(\theta_1, \theta_2)$ è una famiglia esponenziale con $k = 2$.

Modelli statistici per campioni casuali di famiglie esponenziali

Ricordando l'espressione (1.3) che definisce una famiglia esponenziale, considerando un campione casuale $\mathbf{X}_n = (X_1, \dots, X_n)$, proveniente dalla variabile di base X con legge di probabilità $\mathcal{F} = \{f_X(x; \boldsymbol{\theta}) = h(x) \exp\{\eta(\boldsymbol{\theta})T(x) - B(\boldsymbol{\theta})\}, \Omega\}$, si verifica immediatamente che

$$\begin{aligned} f_n(\mathbf{x}_n; \boldsymbol{\theta}) &= \prod_{i=1}^n f_X(x_i; \boldsymbol{\theta}) = \prod_{i=1}^n [h(x_i) \exp\{\eta(\boldsymbol{\theta})T(x_i) - B(\boldsymbol{\theta})\}] \\ &= \prod_{i=1}^n h(x_i) \times \exp\{\eta(\boldsymbol{\theta}) \sum_{i=1}^n T(x_i) - nB(\boldsymbol{\theta})\}. \end{aligned}$$

Pertanto

$$f_n(\mathbf{x}_n; \boldsymbol{\theta}) = h_n(\mathbf{x}_n) \exp\{\eta(\boldsymbol{\theta})T_n(\mathbf{x}_n) - B_n(\boldsymbol{\theta})\}, \quad (1.4)$$

dove

$$h_n(\mathbf{x}_n) = \prod_{i=1}^n h(x_i), \quad T_n(\mathbf{x}_n) = \sum_{i=1}^n T(x_i), \quad B_n(\boldsymbol{\theta}) = nB(\boldsymbol{\theta}).$$

Per la (1.3) possiamo quindi affermare che la famiglia $\mathcal{F}_n = \{f_n(\mathbf{x}_n; \boldsymbol{\theta}) = \prod_{i=1}^n f_X(x_i; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Omega\}$ è una famiglia esponenziale, caratterizzata dalle funzioni $h_n(\mathbf{x}_n), \eta(\boldsymbol{\theta}), T_n(\mathbf{x}_n)$ e $B_n(\boldsymbol{\theta})$.

1.31 Esempio (modello di Poisson). Sia X_1, \dots, X_n , un campione casuale, con $X_i|\theta \sim \text{Pois}(\theta)$, $i = 1, \dots, n$. È semplice verificare che, in questo caso, $f_n(\mathbf{x}_n; \theta)$ è una famiglia esponenziale uniparametrica con

$$h_n(\mathbf{x}_n) = \frac{1}{\prod_{i=1}^n x_i!}, \quad \eta(\theta) = \log \theta, \quad T_n(\mathbf{x}_n) = \sum_{i=1}^n x_i, \quad B_n(\theta) = n\theta.$$

Nel caso k -parametrico, in modo analogo, è immediato verificare che:

$$\begin{aligned} f_n(\mathbf{x}_n; \theta) &= \prod_{i=1}^n [h(x_i)] \exp\left\{ \sum_{j=1}^k \eta_j(\theta) T_j(x_i) - B(\theta) \right\} \\ &= h_n(\mathbf{x}_n) \exp\left\{ \sum_{j=1}^k \eta_j(\theta) \sum_{i=1}^n T_j(x_i) - nB(\theta) \right\} = h_n(\mathbf{x}_n) \exp\left\{ \sum_{j=1}^k \eta_j(\theta) S_j(\mathbf{x}_n) - B_n(\theta) \right\}, \end{aligned}$$

ovvero che $f_n(\mathbf{x}_n; \theta)$ è una famiglia esponenziale caratterizzata dalle funzioni:

$$h_n(\mathbf{x}_n), \quad \eta_1(\theta), \dots, \eta_k(\theta), \quad S_1(\mathbf{x}_n), \dots, S_k(\mathbf{x}_n), \quad B_n(\theta),$$

con

$$S_j(\mathbf{x}_n) = \sum_{i=1}^n T_j(x_i), \quad j = 1, \dots, k.$$

1.32 Esempio (modello normale). Consideriamo il modello $N(\theta_1, \theta_2)$. Per quanto mostrato nell'Esempio 1.30, $f_n(\mathbf{x}_n; \theta)$ può essere scritto come famiglia esponenziale con $k = 2$ e, nel caso di un campione casuale di dimensione n , si ha che

$$\left(\sum_{i=1}^n T_1(x_i), \sum_{i=1}^n T_2(x_i) \right) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2 \right).$$

1.7 Problemi inferenziali

La classificazione dei problemi inferenziali non è univoca. Ne diamo qui una non necessariamente esaustiva ma abbastanza generale e sufficiente utile ai fini didattici di questo testo. Distinguiamo innanzitutto i problemi *post-sperimentali* da quelli *pre-sperimentali*. I problemi post-sperimentali sono quelli che prevedono l'uso di dati campionari. Rientrano tra questi i più usuali problemi, come ad esempio la stima del parametro incognito di un modello. I problemi pre-sperimentali sono quelli che si affrontano prima ancora di osservare i dati e consistono essenzialmente nel *disegno dell'esperimento*. Lo scopo è scegliere in modo ottimale gli elementi di un esperimento su cui lo sperimentatore e lo statistico hanno la possibilità di intervenire. L'ottimalità della scelta si riferisce qui all'informatività e all'efficienza statistiche dell'esperimento stesso. Un tipico problema di scelta dell'esperimento

consiste nello stabilire il numero di osservazioni da includere nel campione. In questo caso, ad esempio, si vuole scegliere un numero di osservazioni adeguato a garantire, entro determinati limiti di costo, sufficiente accuratezza della procedura inferenziale adottata.

Consideriamo ora con maggiore dettaglio i problemi post-sperimentali. Dato un modello statistico parametrico $\{\mathcal{X}^n, f_n(\cdot|\Omega), \Omega \in \Omega\}$, possiamo individuare due classi di problemi inferenziali: problemi *ipotetici* e problemi *predittivi*.

Problemi ipotetici. Si tratta della classe di problemi a cui abbiamo già accennato, riguardanti l'uso dei dati campionari al fine di acquisire informazioni sul parametro incognito del modello. L'obiettivo è quindi valutare, con la maggiore accuratezza possibile, il valore vero θ^* che individua il meccanismo aleatorio che ha generato le osservazioni. Tradizionalmente i problemi ipotetici si articolano in tre sottogruppi:

- (a) *Stima puntuale.* I dati campionari vengono sintetizzati attraverso una opportuna funzione T in un valore numerico $T(\mathbf{x}_n) = t$, chiamato *stima* del parametro. Evidentemente è opportuno richiedere che, per ogni campione osservato, il valore t sia un punto dello spazio parametrico Ω .
- (b) *Stima mediante insiemi.* Si ricerca un insieme $S(\mathbf{x}_n) \subseteq \Omega$ (con S funzione dei dati campionari) che contenga il valore incognito θ^* . L'obiettivo della stima per insiemi è quindi meno ambizioso di quello della stima puntuale, anche se la maggiore approssimazione della conclusione è compensata da una valutazione della affidabilità che l'insieme di stima contenga θ^* . Nei più comuni problemi uniparametrici in cui $\Omega \subset \mathbb{R}$, gli insiemi di stima S sono spesso degli intervalli. In questi casi, si ha che $S(\mathbf{x}_n) = [L(\mathbf{x}_n), U(\mathbf{x}_n)]$, dove L ed U sono due opportune funzioni dei dati campionari tali che, per ogni possibile campione osservato, $L(\mathbf{x}_n) \leq U(\mathbf{x}_n)$.
- (c) *Verifica di ipotesi.* In questi problemi si considera una opportuna partizione dello spazio Ω , ovvero una suddivisione in due sottospazi Ω_0 e Ω_1 , la cui intersezione è vuota e la cui unione coincide con l'intero spazio Ω . Sulla base del campione osservato si deve quindi decidere se più plausibile l'ipotesi che $\theta^* \in \Omega_0$ oppure l'ipotesi complementare che $\theta^* \in \Omega_1$. Il caso più elementare consiste nel confronto tra *ipotesi semplici*, in cui Ω_0 e Ω_1 consistono rispettivamente in due punti, θ_0 e θ_1 .

Problemi predittivi. Nei problemi predittivi si utilizza l'informazione fornita dal campione \mathbf{x}_n osservato in un esperimento statistico per prevedere il valore del risultato di un esperimento futuro governato da una legge di probabilità che condivide lo stesso parametro θ^* con la legge che governa l'esperimento che ha generato \mathbf{x}_n . Anche quelli di tipo predittivo si possono suddividere in problemi di stima puntuale, stima mediante insiemi e verifica di ipotesi, riferiti in questo caso al risultato dell'esperimento futuro. È importante osservare che la differenza principale tra problemi ipotetici e predittivi consiste nel fatto che, nel primo caso, l'oggetto di interesse (θ) non è osservabile, mentre nel secondo caso è una variabile effettivamente osservabile.

I problemi appena illustrati costituiscono la quasi totalità degli argomenti trattati in questo testo. Esistono tuttavia altri relevantissimi problemi statistici a cui accenniamo brevemente.

- (a) *Scelta tra modelli.* Le considerazioni precedenti presuppongono la scelta preliminare di un determinato modello statistico. Ad esempio, si dà per scontato il fatto che i dati provengano da una popolazione normale o da una popolazione di Cauchy. Un problema centrale dell'inferenza statistica consiste proprio nella scelta di uno tra più possibili modelli alternativi, sulla base dei dati osservati. Scelto il modello, si possono poi affrontare i problemi ipotetici e/o predittivi eventuali.
- (b) *Controllo del modello.* In questo caso si è interessati a verificare la plausibilità di un modello adottato alla luce dei dati effettivamente osservati. Ad esempio, attraverso opportuni strumenti, può essere necessario valutare e quantificare la plausibilità che i dati provengano da una popolazione normale. Nel caso in cui i dati osservati facciano sorgere dubbi sul modello assunto, è in genere necessario mettere in discussione tale assunzione prima di affrontare ogni problema ipotetico e/o predittivo.

1.8 Impostazioni inferenziali

La teoria dell'inferenza statistica non è unitaria. Esistono infatti diverse impostazioni o logiche inferenziali che, pur affrontando gli stessi problemi (pre- e post-sperimentali) si basano su presupposti diversi e propongono, spesso, soluzioni che divergono anche dal punto di vista operativo. Le tre principali impostazioni inferenziali, che sono quelle a cui ci riferiamo in questo testo, sono: impostazione basata sull'analisi della funzione di verosimiglianza; impostazione frequentista; impostazione bayesiana.

- *Impostazione basata sull'analisi della funzione di verosimiglianza.* Si basa su elaborazioni di una funzione matematica, la *funzione di verosimiglianza* (fdv) che quantifica il grado di plausibilità che ciascun valore possibile del parametro incognito θ assume alla luce dei dati campionari effettivamente osservati. Opportuni usi della fdv consentono di affrontare i principali problemi inferenziali.
- *Impostazione frequentista.* Le procedure inferenziali frequentiste vengono selezionate in funzione del loro comportamento valutato complessivamente sull'intero spazio dei campioni, e non in riferimento a un campione effettivamente osservato. Ad esempio, nel caso della stima puntuale, si cercano funzioni dei dati campionari che con buona probabilità (nello spazio dei campioni) danno luogo a valori di stima non lontani da θ^* . Si tratta dell'impostazione inferenziale attualmente più diffusa e più comunemente utilizzata, alla quale viene dedicata ampia parte di questo testo.

- *Impostazione bayesiana.* Lo schema logico bayesiano presuppone che l'incertezza sul parametro incognito del modello sia rappresentata attraverso una distribuzione di probabilità. Tutte le informazioni pre-sperimentali su Ω concorrono pertanto a precisare questa legge di probabilità, detta *distribuzione a priori del parametro*. Tale fonte di informazioni su θ viene combinata con l'informazione che il campione di dati osservato fornisce su θ attraverso la funzione di verosimiglianza. Lo strumento per effettuare tale sintesi è la formula di Bayes, con la quale si ottiene la *distribuzione a posteriori del parametro*. Da opportune sintesi della distribuzione a posteriori si ottengono gli strumenti per affrontare i principali problemi inferenziali (ad esempio i problemi ipotetici).

1.9 Esercizi

Capitolo 2

Verosimiglianza

2.1 Introduzione

Il capitolo è dedicato all'illustrazione di uno strumento centrale nell'inferenza statistica: la funzione di verosimiglianza (funzione di verosimiglianza). Dopo avere introdotto il concetto di funzione di verosimiglianza, nel capitolo viene mostrato come utilizzarla per risolvere i tre canonici problemi ipotetici (stima puntuale, stima intervallare e verifica di ipotesi). Nella seconda parte del capitolo si introduce un altro concetto fondamentale nella teoria dell'inferenza statistica: la sufficienza. A conclusione del capitolo viene discusso il cosiddetto Principio di Verosimiglianza, il fondamento logico dell'impiego dei metodi basati sulla funzione omonima.

2.2 Funzione di verosimiglianza

Lo scopo principale di un esperimento statistico è quello di fornire informazioni sul meccanismo generatore dei dati, in parte non noto. Nei problemi parametrici, l'incertezza riguarda il parametro incognito θ del modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Omega)$. La funzione di verosimiglianza svolge esattamente questo ruolo: formalizza matematicamente l'informazione che i dati osservati danno sul parametro incognito di un modello. Infatti, in corrispondenza di un campione osservato, \mathbf{x}_n^o , la funzione di verosimiglianza assegna ad ogni possibile valore di $\theta \in \Omega$ una misura di plausibilità (verosimiglianza) basata sui dati. Si tratta, come vedremo, di una quantificazione della compatibilità tra il campione che si è osservato e ciascuno dei possibili valori che θ può assumere.

2.1 Definizione (Funzione di verosimiglianza). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Omega)$ si chiama *funzione di verosimiglianza* associata al campione osservato $\mathbf{x}_n^o = (x_1^o, \dots, x_n^o)$, la funzione $L : \Omega \mapsto \mathbb{R}^+$ definita ponendo

$$L(\theta; \mathbf{x}_n^o) = f_n(\mathbf{x}_n^o; \theta), \quad \theta \in \Omega.$$

La funzione di verosimiglianza di θ si ottiene quindi calcolando il valore che $f_n(\cdot; \mathbf{x}_n^o)$ assume nel particolare campione osservato \mathbf{x}_n^o . Poichè i valori x_1^o, \dots, x_n^o sono dei numeri reali, $f_n(\mathbf{x}_n^o; \theta)$ è una funzione di θ . Per semplicità di notazione, nel seguito indicheremo il campione osservato con \mathbf{x}_n , ovvero con la stessa simbologia utilizzata per un generico campione di \mathcal{X}^n . La definizione introdotta è valida sia nel caso in cui si considerano v.a. discrete che in quelli per v.a. assolutamente continue. Se le v.a. X_i sono discrete, la funzione di verosimiglianza fornisce, per ogni valore di θ , la probabilità del campione realizzato (e quindi assume valori nell'intervallo $[0, 1]$). Se le v.a. X_i sono assolutamente continue, la funzione di verosimiglianza fornisce, per ogni valore di θ , la densità di probabilità del campione realizzato, \mathbf{x}_n . L'interpretazione resta comunque inalterata rispetto al caso discreto. I valori di θ per i quali L è più elevata sono i valori del parametro più compatibili con il campione dato. Per questo motivo i valori che $L(\theta; \mathbf{x}_n)$ assegna ai diversi valori di θ sono detti verosimiglianze e, complessivamente, la funzione di verosimiglianza quantifica l'*evidenza sperimentale* che il campione osservato fornisce per ciascun valore di θ .

La funzione di verosimiglianza assume, per definizione, valori nello spazio \mathbb{R}^+ . La definizione che segue introduce una misura relativa (compresa tra 0 e 1) dell'evidenza che \mathbf{x}_n assegna ai valori di $\theta \in \Omega$.

2.2 Definizione (Funzione di verosimiglianza relativa). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Omega)$, si definisce funzione di verosimiglianza relativa associata al campione osservato \mathbf{x}_n la funzione $\bar{L} : \Omega \mapsto [0, 1]$ definita ponendo

$$\bar{L}(\theta; \mathbf{x}_n) = \frac{L(\theta; \mathbf{x}_n)}{\sup_{\theta \in \Omega} L(\theta; \mathbf{x}_n)}, \quad \theta \in \Omega. \quad (2.1)$$

La funzione di verosimiglianza relativa $\bar{L}(\theta; \mathbf{x}_n)$ è proporzionale a $L(\theta; \mathbf{x}_n)$, dal momento che $\sup_{\theta \in \Omega} L(\theta; \mathbf{x}_n)$ è una costante rispetto a θ . Nei problemi più semplici $L(\theta; \mathbf{x}_n)$ ha un unico punto di massimo (che, vedremo, si chiama *stima di massima verosimiglianza*) $\theta = \hat{\theta}_{mv}$ appartenente a Ω . Si ha quindi che

$$\bar{L}(\theta; \mathbf{x}_n) = \frac{L(\theta; \mathbf{x}_n)}{\max_{\theta \in \Omega} L(\theta; \mathbf{x}_n)} = \frac{L(\theta; \mathbf{x}_n)}{L(\hat{\theta}_{mv}; \mathbf{x}_n)}, \quad \theta \in \Omega.$$

La funzione di verosimiglianza relativa consente di confrontare direttamente i diversi possibili valori di θ tra loro e con l'ipotesi $\theta = \hat{\theta}_{mv}$ privilegiata dal campione osservato.

2.3 Osservazione/i. La funzione di verosimiglianza è definita a meno di costanti che non dipendono da θ . Per una costante arbitraria $c > 0$, eventualmente dipendente dai dati campionari, la funzione $cL(\theta; \mathbf{x}_n)$ fornisce sul parametro incognito le stesse informazioni che otteniamo considerando la funzione $L(\theta; \mathbf{x}_n)$. In generale, se esistono due funzioni h (di \mathbf{x}_n) e g (di \mathbf{x}_n e θ) tali che

$$L(\theta; \mathbf{x}_n) = h(\mathbf{x}_n)g(\theta; \mathbf{x}_n), \quad \forall \mathbf{x}_n \in \mathcal{X}^n, \quad \forall \theta \in \Omega,$$

la funzione $g(\theta; \mathbf{x}_n)$ rappresenta il *nucleo* della funzione di verosimiglianza di θ . La funzione di verosimiglianza è identificata dal suo nucleo.

2.4 Osservazione/i. I valori della funzione di verosimiglianza non possono essere interpretati come probabilità assegnate a θ , alla luce del risultato \mathbf{x}_n . Il parametro θ non è infatti una v.a. e il valore $L(\theta; \mathbf{x}_n)$ rappresenta come varia la probabilità o la densità di probabilità di $\mathbf{X}_n = \mathbf{x}_n$ al variare di θ in Ω . In altre parole: in $L(\theta, \mathbf{x}_n)$, dobbiamo considerare i valori di \mathbf{x}_n fissati e i valori di θ variano in Ω .

Funzione di verosimiglianza associata a campioni casuali

La definizione di funzione di verosimiglianza vale per campioni generici (non necessariamente i.i.d.). Si ottiene una notevole semplificazione analitica quando \mathbf{X}_n è un campione casuale. In questo caso, infatti, si ha che

$$L(\theta; \mathbf{x}_n) = \prod_{i=1}^n f_X(x_i; \theta).$$

2.5 Esempio (modello bernoulliano). Riprendiamo in esame il modello bernoulliano e consideriamo ora un generico campione osservato di dimensione n . In questo caso

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}.$$

Se consideriamo il campione di dimensione $n = 10$ in cui $\sum_{i=1}^n x_i = 8$ e $(n - \sum_{i=1}^n x_i = 2)$, la funzione di verosimiglianza di θ è

$$L(\theta; \mathbf{x}_n) = \theta^8 (1 - \theta)^2, \quad \theta \in [0, 1].$$

Con semplici calcoli si mostra che l'unico punto di massimo assoluto di questa funzione è nel punto $\theta = 0.8$. La funzione di verosimiglianza relativa è quindi

$$\bar{L}(\theta; \mathbf{x}_n) = \left(\frac{\theta}{0.8}\right)^8 \left(\frac{1 - \theta}{0.2}\right)^2, \quad \theta \in [0, 1].$$

La Figura 2.1 mostra il grafico di $\bar{L}(\theta; \mathbf{x}_n)$ per $\theta \in [0, 1]$ (linea continua). Consideriamo ora un nuovo esperimento in cui $n = 50$ e il numero di successi osservati nel campione è pari a 40. La funzione di verosimiglianza relativa è ora

$$\bar{L}(\theta; \mathbf{x}_n) = \left(\frac{\theta}{0.8}\right)^{40} \left(\frac{1 - \theta}{0.2}\right)^{10}, \quad \theta \in [0, 1].$$

Si verifica facilmente che anche in questo caso $\hat{\theta}_{mv} = 0.8$. La linea tratteggiata nella Figura 2.1 corrisponde al grafico di $\bar{L}(\theta; \mathbf{x}_n)$ in questo secondo caso. Va osservato

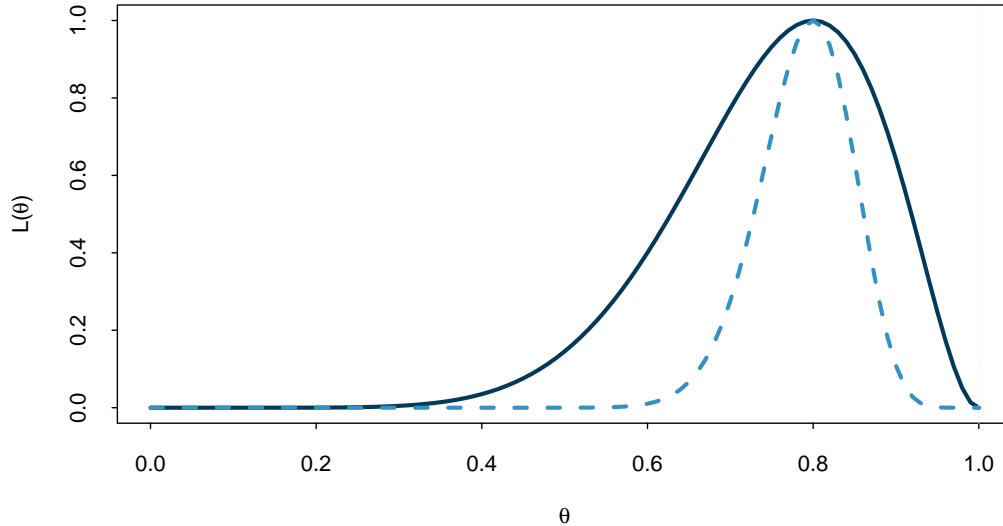


Figura 2.1: Grafico della funzione di verosimiglianza relativa di θ , \bar{L} , associata a due campioni provenienti da una popolazione bernoulliana di numerosità $n = 10$ (linea continua) e $n = 50$ (linea tratteggiata), entrambi con $\bar{x}_n = 0.8$.

che, nonostante il punto di massimo resti inalterato, quando $n = 50$ la funzione di verosimiglianza relativa risulta ben più concentrata intorno al suo punto di massimo che nel caso $n = 10$. All'aumentare del del numero di osservazioni, si riduce la lunghezza dell'intervallo di valori del parametro che ricevono dal campione una misura di evidenza superiore a un livello prescelto. L'aumento della dimensione campionaria consente quindi alla funzione di verosimiglianza di discernere meglio tra i possibili valori di θ .

Nei prossimi paragrafi mostriamo l'uso della funzione di verosimiglianza per la risoluzione dei tre problemi inferenziali riguardanti il parametro incognito del modello: stima puntuale, stima mediante regioni e verifica di ipotesi. A tale fine si considerano opportune sintesi della funzione di verosimiglianza che, nel suo insieme, rappresenta la totalità dell'informazione sperimentale sul parametro incognito del modello.

2.3 Stima puntuale

Premettiamo una definizione.

2.6 Definizione (Stima di un parametro). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Omega\}$ e il campione osservato \mathbf{x}_n , si chiama *stima* di θ una funzione dei dati campionari (non dipendente da θ), $\hat{\theta}(\mathbf{x}_n)$, che utilizziamo come valutazione numerica di θ .

Si osservi che la definizione data è piuttosto generica e vale sia nel caso scalare che vettoriale: la dimensione della stima coincide con quella del vettore dei parametri.

2.7 Definizione (Stima di massima verosimiglianza). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Omega\}$ e il campione osservato \mathbf{x}_n , si chiama *stima di massima verosimiglianza* (smv) di θ un valore $\hat{\theta}_{mv}(\mathbf{x}_n) \in \Omega$ tale che $L(\hat{\theta}_{mv}; \mathbf{x}_n) \geq L(\theta; \mathbf{x}_n), \forall \theta \in \Omega$:

$$\hat{\theta}_{mv}(\mathbf{x}_n) = \arg \max_{\theta \in \Omega} L(\theta; \mathbf{x}_n).$$

L'esistenza della smv non è garantita per tutti i modelli statistici. Inoltre esistono modelli in cui la smv esiste ma non è unica. Tuttavia, nei più comuni casi uniparametrici, la smv esiste ed è unica.

Consideriamo ora il problema della ricerca della smv per due classi di problemi rilevanti (problemi regolari di stima e problemi con supporto della v.a. dipendente dal parametro incognito).

Stima di massima verosimiglianza in problemi regolari di stima

Per semplicità consideriamo in questa sezione il caso uniparametrico e assumiamo quindi che Ω sia un sottoinsieme di \mathbb{R} . Quando il modello statistico soddisfa le proprietà della definizione che segue, la ricerca della smv può essere effettuata con le usuali tecniche con cui si determinano i punti di massimo di una funzione reale di variabile reale derivabile nel dominio.

2.8 Definizione (Problema regolare di stima). Dati una v.a. X di base, il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Omega\}$, un campione osservato \mathbf{x}_n e la funzione di verosimiglianza ad esso associata, $L(\theta; \mathbf{x}_n)$, si ha un *problema regolare di stima* se:

- il supporto della v.a. di base X non dipende da θ .
- esistono finite le derivate prima e seconda di $L(\theta; \mathbf{x}_n)$ rispetto a θ .

In questi casi il punto $\hat{\theta}_{mv}$ è un massimo per la funzione $L(\theta; \mathbf{x}_n)$ se sono verificate le seguenti due condizioni:

$$\left. \frac{d}{d\theta} L(\theta; \mathbf{x}_n) \right|_{\theta=\hat{\theta}_{mv}} = 0 \quad \text{e} \quad \left. \frac{d^2}{d\theta^2} L(\theta; \mathbf{x}_n) \right|_{\theta=\hat{\theta}_{mv}} < 0. \quad (2.2)$$

In molti degli esempi che tratteremo, è più semplice calcolare le derivate della *funzione di log-verosimiglianza*:

$$\ell(\theta; \mathbf{x}_n) = \ln L(\theta; \mathbf{x}_n).$$

Dal momento che la funzione logaritmo (con base maggiore di uno) è una funzione crescente, i punti di massimo di $L(\theta; \mathbf{x}_n)$ e $\ell(\theta; \mathbf{x}_n)$ coincidono. Pertanto, il punto $\hat{\theta}_{mv}$ è un massimo per la funzione $L(\theta; \mathbf{x}_n)$ se sono verificate le seguenti due condizioni:

$$\left. \frac{d}{d\theta} \ell(\theta; \mathbf{x}_n) \right|_{\theta=\hat{\theta}_{mv}} = 0 \quad \text{e} \quad \left. \frac{d^2}{d\theta^2} \ell(\theta; \mathbf{x}_n) \right|_{\theta=\hat{\theta}_{mv}} < 0.$$

Le equazioni $d/d\theta L(\theta; \mathbf{x}_n) = 0$ e $d/d\theta \ell(\theta; \mathbf{x}_n) = 0$ prendono rispettivamente il nome di *equazione di verosimiglianza* e *equazione di log-verosimiglianza*.

Nell'esempio che segue illustriamo la determinazione della smv nel caso del modello bernoulliano.

2.9 Esempio (modello bernoulliano). Consideriamo ancora una volta il modello bernoulliano e assumiamo che lo spazio Ω sia l'intervallo chiuso $[0, 1]$. La funzione di log-verosimiglianza è

$$\ell(\theta; \mathbf{x}_n) = \ln L(\theta; \mathbf{x}_n) = \left(\sum_{i=1}^n x_i \right) \ln \theta + \left(n - \sum_{i=1}^n x_i \right) \ln(1 - \theta).$$

L'equazione di log-verosimiglianza

$$\frac{d}{d\theta} \ell(\theta; \mathbf{x}_n) = \frac{d}{d\theta} \ln L(\theta; \mathbf{x}_n) = \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1 - \theta} = 0$$

ha un'unica soluzione nel punto $\hat{\theta}_{mv} = \bar{x}_n = \sum x_i/n$. Si tratta della smv in quanto la derivata seconda di $\ell(\theta; \mathbf{x}_n)$ calcolata in $\theta = \bar{x}_n$ è

$$\left. \frac{d^2}{d\theta^2} \ell(\theta; \mathbf{x}_n) \right|_{\theta=\bar{x}_n} = - \frac{n}{\bar{x}_n(1 - \bar{x}_n)} < 0.$$

Stima di massima verosimiglianza nei modelli con supporto dipendente da θ

Quando il supporto della v.a. dipende dal parametro del modello, il problema di stima è non regolare e non si può ricorrere al metodo appena illustrato. In questi casi, infatti, la funzione di verosimiglianza presenta punti di discontinuità (in cui la funzione di verosimiglianza non è derivabile) che sono proprio i punti di massima verosimiglianza.

2.10 Esempio (modello uniforme). Consideriamo di avere un campione casuale da v.a. uniforme in $[0, \theta]$. La funzione di densità della variabile aleatoria di base $X \sim \text{Unif}[0, \theta]$ è $f_X(x; \theta) = \theta^{-1} I_{[0, \theta]}(x)$. La funzione di verosimiglianza per il campione osservato \mathbf{x}_n è quindi

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \frac{1}{\theta^n} \prod_{i=1}^n I_{[0, \theta]}(x_i) = \frac{1}{\theta^n} I_{[x_{(n)}, \infty]}(\theta),$$

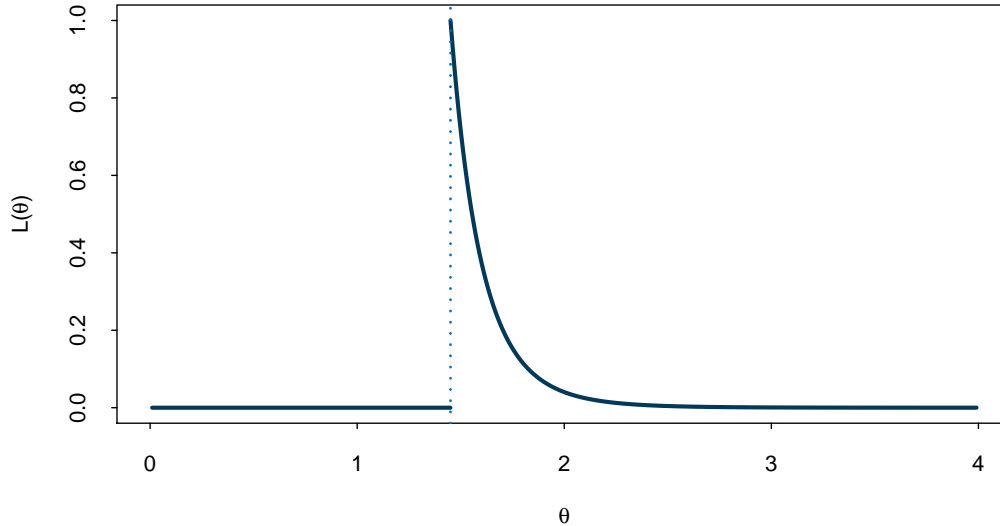


Figura 2.2: Grafico delle funzione di verosimiglianza relativa di θ , \bar{L} , per un campione di dimensione estratto da una popolazione $Unif[0, \theta]$, con $n = 10$, $x_{(10)} = 1.45$.

dove $x_{(n)} = \max\{x_1, \dots, x_n\}$ indica il massimo valore tra le osservazioni campionarie osservate. L'ultima uguaglianza si giustifica osservando che

$$\prod_{i=1}^n I_{[0, \theta]}(x_i) = 1 \Leftrightarrow x_i \in [0, \theta], i = 1, \dots, n \Leftrightarrow 0 \leq x_{(n)} \leq \theta \Leftrightarrow x_{(n)} \leq \theta < +\infty.$$

La funzione di verosimiglianza ha quindi un unico punto di massimo in $\theta = x_{(n)}$, punto di discontinuità di prima specie per $L(\theta; \mathbf{x}_n)$ (ove la funzione non è derivabile e vale $[x_{(n)}]^{-n}$). Abbiamo quindi che $\hat{\theta}_{mv} = x_{(n)}$. La Figura 2.2 mostra il grafico di $\bar{L}(\theta; \mathbf{x}_n)$ nel caso di un campione osservato in cui $n = 10$ e $x_{(10)} = 1.45$.

2.4 Stima mediante insiemi

La procedura di stima di massima verosimiglianza seleziona il valore o i valori di θ , se esistenti, massimamente compatibili con il campione osservato. Sembra tuttavia ragionevole non scartare i valori di θ che ricevono dal campione osservato un supporto superiore ad una soglia adeguatamente elevata.

2.11 Definizione (Insieme di verosimiglianza). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta)\}$,

$\theta \in \Omega$ }, il campione osservato \mathbf{x}_n e un valore $q \in [0, 1]$, si definisce *insieme di verosimiglianza di livello q* l'insieme di valori di θ

$$L_q(\mathbf{x}_n) = \{\theta \in \Omega : \bar{L}(\theta; \mathbf{x}_n) \geq q\}.$$

La definizione data, valida per uno spazio parametrico di dimensione arbitraria, non garantisce che, nel caso in cui $\Omega \subseteq \mathbb{R}$, l'insieme L_q sia un intervallo, anche se questo è il caso di molti dei più comuni modelli che consideriamo in questo testo. Si noti inoltre che, quanto più elevato è il valore di q , tanto minore risulta la dimensione di L_q . Al contrario, quanto minore q , tanto maggiore la dimensione di L_q . Come casi limite abbiamo che

$$q \rightarrow 0 \Rightarrow L_q(\mathbf{x}_n) \rightarrow \Omega \quad \text{e che} \quad q \rightarrow 1 \Rightarrow L_q(\mathbf{x}_n) \rightarrow \hat{\theta}_{mv}.$$

Consideriamo un esempio in cui è possibile determinare esplicitamente gli estremi dell'intervallo L_q per un parametro scalare.

2.12 Esempio (modello normale). Consideriamo per la v.a. di base un modello normale con valore atteso incognito θ e varianza nota, pari a σ^2 . Per questo modello gli estremi dell'intervallo di log-verosimiglianza di livello q possono essere determinati analiticamente. La funzione di verosimiglianza associata ad un campione casuale osservato è

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2} \propto e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2}.$$

La stima di massima verosimiglianza per questo modello (vedi Esempio ??) è $\hat{\theta}_{mv} = \bar{x}_n$, la media campionaria di \mathbf{x}_n . La funzione di verosimiglianza relativa è:

$$\bar{L}(\theta; \mathbf{x}_n) = \frac{L(\theta; \mathbf{x}_n)}{L(\hat{\theta}_{mv}; \mathbf{x}_n)} = \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \theta)^2 - \sum_{i=1}^n (x_i - \bar{x}_n)^2 \right] \right\}.$$

La quantità $\sum (x_i - \theta)^2$ può essere scomposta nel modo seguente:

$$\begin{aligned} \sum_{i=1}^n (x_i - \theta)^2 &= \sum_{i=1}^n (x_i - \bar{x}_n + \bar{x}_n - \theta)^2 = \\ &= \sum_{i=1}^n (x_i - \bar{x}_n)^2 + n(\bar{x}_n - \theta)^2, \end{aligned}$$

dove l'ultima uguaglianza sussiste ricordando che $\sum (x_i - \bar{x}_n) = 0$. La funzione di verosimiglianza relativa diventa

$$\begin{aligned} \bar{L}(\theta; \mathbf{x}_n) &= \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \bar{x}_n)^2 + n(\bar{x}_n - \theta)^2 - \sum_{i=1}^n (x_i - \bar{x}_n)^2 \right] \right\} \\ &= \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x}_n - \theta)^2 \right\}. \end{aligned}$$

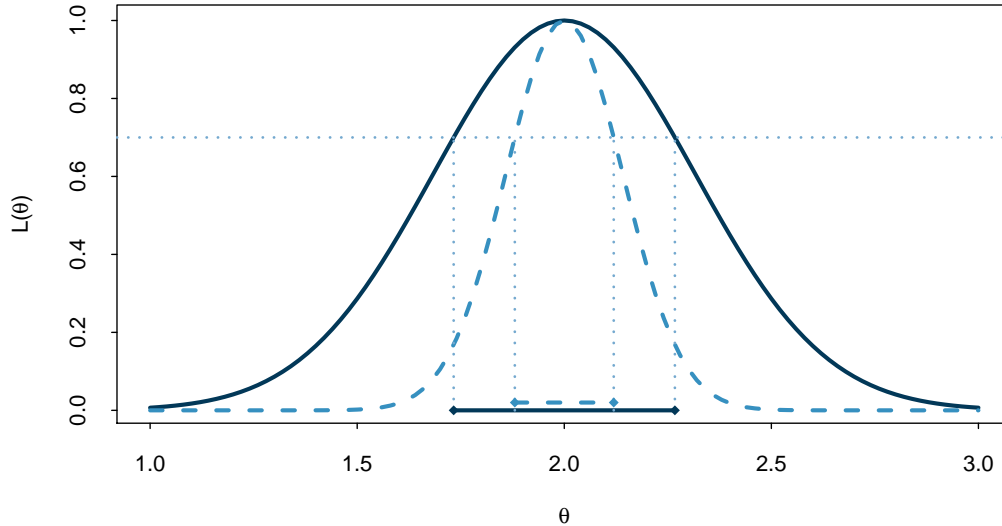


Figura 2.3: Insiemi L_q per il parametro θ del modello $N(\theta, \sigma^2)$.

L'intervallo di verosimiglianza per θ di livello q è dato dall'insieme di valori di θ tali che:

$$\begin{aligned}
 L_q &= \{\theta \in \Omega : \bar{L}(\theta; \mathbf{x}_n) \geq q\} = \left\{ \theta \in \Omega : \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}_n - \theta)^2\right\} \geq q \right\} \\
 &= \left\{ \theta \in \Omega : -\frac{n}{2\sigma^2}(\bar{x}_n - \theta)^2 \geq \ln q \right\} \\
 &= \left\{ \theta \in \Omega : \bar{x}_n - k_q \frac{\sigma}{\sqrt{n}} \leq \theta \leq \bar{x}_n + k_q \frac{\sigma}{\sqrt{n}} \right\} \\
 &= \left[\bar{x}_n - k_q \frac{\sigma}{\sqrt{n}}, \bar{x}_n + k_q \frac{\sigma}{\sqrt{n}} \right], \quad k_q = \sqrt{-2 \ln q}.
 \end{aligned}$$

Si noti che l'ipotesi $q \in (0, 1)$ garantisce l'esistenza di k_q . La Figura 2.3 riporta il grafico della funzione di verosimiglianza relativa associata a due campioni di diversa numerosità ma con stessa media aritmetica ($\bar{x}_n = 2$) (la linea tratteggiata corrisponde alla funzione di verosimiglianza associata al campione di dimensione maggiore). Si osservi anche la simmetria delle funzione di verosimiglianza e la diversa lunghezza degli insiemi di verosimiglianza.

Per molti modelli non è possibile risolvere analiticamente la disequazione che definisce l'insieme $L_q(\mathbf{x}_n)$ e quindi determinare le formule che definiscono gli estremi

dell'intervallo di verosimiglianza. In questi casi tali valori si possono tuttavia ottenere numericamente. In alternativa, come vedremo, è anche possibile ricorrere ad approssimazioni della funzione di verosimiglianza per ottenere formule esplicite degli estremi (approssimati) degli intervalli di livello q .

2.5 Verifica di ipotesi

Nei problemi di verifica di ipotesi siamo interessati a confrontare la plausibilità di congetture alternative sul parametro θ , utilizzando i dati a disposizione. Consideriamo due sottoinsiemi dello spazio dei parametri Ω , che indichiamo con Ω_1 e Ω_2 e supponiamo che $\Omega_1 \cap \Omega_2 = \emptyset$. Indichiamo con H_i l'ipotesi che $\theta \in \Omega_i$, $i=1,2$, ovvero che il valore del parametro che individua il modello che ha generato i dati sia un valore in Ω_i . Un caso particolare si ha quando si assume che gli insiemi Ω_i siano costituiti da punti dello spazio parametrico, ovvero che $\Omega_1 = \{\theta_1\}$ e $\Omega_2 = \{\theta_2\}$. In questo caso il confronto tra le due ipotesi (dette *ipotesi semplici*) può essere effettuato attraverso il *rapporto delle verosimiglianze*:

$$\lambda_{12}(\mathbf{x}_n) = \frac{L(\theta_1; \mathbf{x}_n)}{L(\theta_2; \mathbf{x}_n)} = \frac{\bar{L}(\theta_1; \mathbf{x}_n)}{\bar{L}(\theta_2; \mathbf{x}_n)}$$

Evidentemente, l'ipotesi H_i è più verosimile dell'ipotesi H_j se $\lambda_{ij}(\mathbf{x}_n) > 1$ ($i, j = 1, 2$ e $i \neq j$); le ipotesi sono equivalenti se $\lambda_{12}(\mathbf{x}_n) = 1$.

Il rapporto delle verosimiglianze, oltre a indicare se una ipotesi sia più verosimile di un'altra, misura il *grado di evidenza* che i dati assegnano a una ipotesi rispetto ad un'altra. Se, ad esempio, in corrispondenza di \mathbf{x}_n si ha che $\lambda_{ij}(\mathbf{x}_n) = 8$, alla luce dei dati a disposizione l'ipotesi H_j risulta 8 volte più plausibile dell'ipotesi H_i .

2.13 Esempio (modello normale). Supponiamo che $X_i|\theta \sim N(\theta, \sigma^2)$, i.i.d. $i = 1, \dots, n$. In questo caso

$$\lambda_{12}(\bar{x}_n) = \frac{\phi(\bar{x}_n; \theta_1, \sigma^2/n)}{\phi(\bar{x}_n; \theta_2, \sigma^2/n)},$$

dove $\phi(z; a, b)$ indica il valore della funzione di densità della v.a. $N(a, b)$ calcolata in $z \in \mathbb{R}$. Con semplici calcoli si mostra che

$$\lambda_{12}(\bar{x}_n) = \exp \left\{ -\frac{n(\theta_2 - \theta_1)}{\sigma^2} (\bar{x}_n - \bar{\theta}) \right\}$$

dove $\bar{\theta} = (\theta_1 + \theta_2)/2$.

Quando Ω_i non è un singolo punto di Ω , l'ipotesi H_i si dice *composta*. La funzione di verosimiglianza è una funzione di punto definita in Ω e non consente di assegnare valutazioni di verosimiglianza ad insiemi non puntuali dello spazio Ω . Questo fatto determina una difficoltà nel caso di confronto tra ipotesi composte. La procedura più comune è quella di utilizzare come valutazione del supporto inferenziale di un sottoinsieme Ω_i di Ω la quantità

$$\sup_{\theta \in \Omega_i} L(\theta; \mathbf{x}_n).$$

In questo caso il confronto tra due ipotesi composte H_1 e H_2 si effettua con il *rapporto di verosimiglianze massimizzate*

$$\lambda_{12}^m(\mathbf{x}_n) = \frac{\sup_{\theta \in \Omega_1} L(\theta; \mathbf{x}_n)}{\sup_{\theta \in \Omega_2} L(\theta; \mathbf{x}_n)}.$$

Nell'impostazione frequentista (si veda il Capitolo 6), si assume di solito che $\Omega = \Omega_1 \cup \Omega_2$ e, per il confronto tra ipotesi composte, si usa comunemente la seguente definizione di rapporto di verosimiglianze massimizzate:

$$\lambda_{12}^m(\mathbf{x}_n) = \frac{\sup_{\theta \in \Omega_1} L(\theta; \mathbf{x}_n)}{\sup_{\theta \in \Omega} L(\theta; \mathbf{x}_n)}.$$

Si noti tuttavia che, in questo secondo caso, $0 \leq \lambda_{12}^m \leq 1$ e quindi il valore di indifferenza tra le due ipotesi non coincide con l'unità.

2.6 Informazione di Fisher osservata

Nel caso di problemi regolari di stima (Definizione 2.8), il grado di concentrazione della funzione di verosimiglianza è quantificato dalla nozione che segue.

2.14 Definizione (Informazione osservata). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Omega)$ la cui stima di massima verosimiglianza $\hat{\theta}_{mv}$ è un punto interno all'insieme $\Omega \subseteq \mathbb{R}^1$ e la cui funzione di log-verosimiglianza è derivabile due volte in $\theta = \hat{\theta}_{mv}$ (problema regolare di stima), si definisce *informazione di Fisher osservata* la quantità:

$$\mathcal{I}_n(\hat{\theta}_{mv}; \mathbf{x}_n) = - \left. \frac{d^2}{d\theta^2} \ln L(\theta; \mathbf{x}_n) \right|_{\theta=\hat{\theta}_{mv}} = - \left. \frac{d^2}{d\theta^2} \ell(\theta; \mathbf{x}_n) \right|_{\theta=\hat{\theta}_{mv}}.$$

La funzione

$$\mathcal{I}_n(\theta; \mathbf{x}_n) = - \frac{d^2}{d\theta^2} \ell(\theta; \mathbf{x}_n)$$

si chiama *funzione di informazione* (o *informazione di Fisher*).

2.15 Osservazione/i. La precedente definizione richiede la derivabilità della funzione di log-verosimiglianza rispetto a θ . Pertanto la funzione di informazione (e quindi l'informazione osservata) non possono essere determinate nei modelli con supporto dipendente dal parametro incognito.

La funzione di informazione dipende dalla coppia (θ, \mathbf{x}_n) mentre l'informazione di Fisher osservata è una funzione dei dati campionari e non dipende dal parametro incognito. Per semplicità, quando useremo la notazione \mathcal{I}_n intenderemo che si tratta dell'informazione osservata. La funzione di informazione può anche essere espressa come

$$\mathcal{I}_n(\theta; \mathbf{x}_n) = - \frac{d}{d\theta} S(\theta; \mathbf{x}_n)$$

dove $S(\theta; \mathbf{x}_n) = \frac{d}{d\theta} \ln L(\theta; \mathbf{x}_n)$ si chiama *funzione di punteggio* (in inglese *score function*).

Ricordando che, in un punto di massimo di $\ell(\theta; \mathbf{x}_n)$, la derivata seconda $\ell''(\hat{\theta}_{mv}; \mathbf{x}_n)$ è negativa, necessariamente \mathcal{I}_n è un numero reale positivo $\forall \mathbf{x}_n \in \mathcal{X}^n$. In condizioni di regolarità, la derivata seconda di $\ell(\theta)$ in $\hat{\theta}_{mv}$ è negativa ed è una misura della concavità della log-verosimiglianza nel punto di massimo. Quanto maggiore è il valore assoluto della derivata seconda (e quindi di $-\ell''(\hat{\theta}_{mv})$), tanto più accentuata è la concavità della log-verosimiglianza in $\hat{\theta}_{mv}$. Quanto più accentuata è la concavità in $\ell(\hat{\theta}_{mv})$, tanto più rapidamente decresce la funzione $\ell(\theta)$ in un intorno di $\hat{\theta}_{mv}$. Così, anche per valori di θ in un intorno piccolo del punto di massimo, la differenza $\ell(\hat{\theta}_{mv}) - \ell(\theta)$ può essere rilevante. Del resto, osservando che

$$\ell(\hat{\theta}; \mathbf{x}_n) - \ell(\theta; \mathbf{x}_n) = \ln \frac{L(\hat{\theta}_{mv}; \mathbf{x}_n)}{L(\theta; \mathbf{x}_n)} = \ln \frac{\bar{L}(\hat{\theta}_{mv}; \mathbf{x}_n)}{\bar{L}(\theta; \mathbf{x}_n)}, \quad (2.3)$$

si deduce che valori elevati di \mathcal{I}_n corrispondono a valori elevati del rapporto tra le verosimiglianze in $\theta = \hat{\theta}_{mv}$ e punti di θ in un intorno di $\hat{\theta}_{mv}$.

2.16 Esempio (modello bernoulliano). Abbiamo visto in questo caso che $\hat{\theta}_{mv} = \bar{x}_n$. Inoltre,

$$\frac{d}{d\theta} \ln L(\theta; \mathbf{x}_n) = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n - \sum_{i=1}^n x_i}{1 - \theta} \quad \text{e} \quad \frac{d^2}{d\theta^2} \ln L(\theta; \mathbf{x}_n) = -\frac{\sum_{i=1}^n x_i}{\theta^2} - \frac{n - \sum_{i=1}^n x_i}{(1 - \theta)^2}$$

Pertanto con semplici calcoli si verifica che

$$\mathcal{I}_n = -\frac{d^2}{d\theta^2} \ln L(\theta; \mathbf{x}_n) \Big|_{\theta=\bar{x}_n} = \frac{n}{\bar{x}_n(1 - \bar{x}_n)}.$$

2.7 Esempi notevoli

Riportiamo di seguito una tabella riassuntiva relativa ai modelli più comuni contenente: (a) nome del modello; (b) espressione della funzione di verosimiglianza (a meno di costanti ininfluenti); (c) spazio dei parametri; (d) stima di massima verosimiglianza di θ ; (e) informazione osservata. Tutte le quantità di cui sopra possono essere ricavate seguendo le procedure descritte in questo capitolo.

Modello	$L(\theta; \mathbf{x}_n)$	Ω	$\hat{\theta}_{mv}$	\mathcal{I}_n
Bernoulli	$\theta^{\bar{x}_n} (1 - \theta)^{n - \bar{x}_n}$	$[0, 1]$	\bar{x}_n	$\frac{n}{\bar{x}_n(1 - \bar{x}_n)}$
Poisson	$e^{-n\theta} \theta^{n\bar{x}_n}$	\mathbb{R}^+	\bar{x}_n	$\frac{n}{\bar{x}_n}$
Esponenziale	$\frac{1}{\theta^n} e^{-\frac{n\bar{x}_n}{\theta}}$	\mathbb{R}^+	\bar{x}_n	$\frac{n}{\bar{x}_n^2}$
Esponenziale negativo	$\theta^n e^{-\theta \sum_{i=1}^n x_i}$	\mathbb{R}^+	$\frac{1}{\bar{x}_n}$	$n\bar{x}_n^2$
Norm 1 (σ^2 noto)	$e^{-\frac{n}{2\sigma^2}(\theta - \bar{x}_n)^2}$	\mathbb{R}	\bar{x}_n	$\frac{n}{\sigma^2}$
Norm 2 (μ_0 noto)	$\frac{1}{\theta^{\frac{n}{2}}} e^{-\frac{nS_0^2}{2\theta}}$	\mathbb{R}^+	$S_0^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$	e
a	b	c	d	e
Uniforme in $[0, \theta]$	$\frac{1}{\theta^n} \cdot \mathbb{I}_{[x_{(n)}, +\infty)}(\theta)$	\mathbb{R}^+	$x_{(n)}$	non esiste

2.8 Principio di verosimiglianza

Il concetto di verosimiglianza, dovuto allo statistico inglese R.A. Fisher (anni 20) formalizza matematicamente, attraverso un sistema di pesi su tutti i valori del parametro, l'informazione che i dati osservati forniscono su un parametro. Questo ruolo è stato formalizzato da A. Birnbaum (1962) nel cosiddetto Principio di verosimiglianza (PdV).

Principio di verosimiglianza (PdV). Siano

$$\{\mathcal{X}^n, f_n(\cdot; \theta), \theta \in \Theta\} \quad \text{e} \quad \{\mathcal{Y}^m, p_m(\cdot; \theta), \theta \in \Theta\}$$

due modelli statistici per i quali lo spazio dei parametri è lo stesso. Se le funzioni di verosimiglianza $L(\theta; \mathbf{x}_n)$ e $L(\theta; \mathbf{y}_m)$ associate a due campioni osservati $\mathbf{x}_n \in \mathcal{X}^n$ e $\mathbf{y}_m \in \mathcal{Y}^m$ sono proporzionali tra loro, ovvero se

$$L(\theta, \mathbf{x}_n) = c \cdot L(\theta; \mathbf{y}_m), \quad \forall \theta \in \Theta,$$

dove $c > 0$ è una costante rispetto a θ (che può dipendere da \mathbf{x}_n e \mathbf{y}_m), i risultati campionari \mathbf{x}_n e \mathbf{y}_m forniscono la stessa informazione sul parametro incognito.

Il PdV formalizza il fatto che la fdv incorpora tutta l'informazione campionaria riguardo al parametro. Afferma infatti che se due campioni distinti, provenienti anche da due modelli diversi (ma che condividono lo spazio dei parametri), danno luogo a verosimiglianze equivalenti, le conclusioni inferenziali su θ che si hanno utilizzando i due campioni devono essere le stesse. Ad esempio, si devono ottenere con i due campioni stesse stime puntuali, intervallari e stesse scelte nei problemi di verifica di ipotesi.

La formulazione data è a volte indicata come versione *forte* del PdV, mentre per la versione *debole* si assume che lo spazio campionario e la legge di probabilità corrispondenti ai due campioni osservati siano gli stessi.

Le procedure inferenziali basate sulle sintesi della fdv obbediscono automaticamente al PdV. Non tutte le impostazioni inferenziali rispettano però il PdV. In particolare, le procedure ispirate al Principio del campionamento ripetuto non necessariamente rispettano il PdV, che invece è pienamente osservato dall'inferenza bayesiana.

2.9 Esercizi

Capitolo 3

Sufficienza

3.1 Introduzione

In questo capitolo viene introdotto un altro concetto centrale per l'inferenza statistica in tutte le sue impostazioni: la sufficienza. Nei paragrafi che seguono vengono fornite le definizioni di sufficienza conformi alle impostazioni non bayesiane (Paragrafo 3.2); la definizione bayesiana è contenuta nel Capitolo 4. Viene poi illustrato il concetto di minimalità per statistiche sufficienti (Paragrafo 3.3) e infine si illustra la struttura delle statistiche sufficienti nei modelli che sono famiglie esponenziali.

3.2 Statistiche sufficienti

Siamo interessati a verificare se, per un determinato modello statistico, esistano delle funzioni in grado di sintetizzare i dati campionari senza perdita di informazione. L'obiettivo è identificare funzioni T di \mathbf{x}_n (statistiche campionarie) tali che la conoscenza del valore $T(\mathbf{x}_n) = t$ sia equivalente, ai fini dell'inferenza su θ , della conoscenza degli n valori (x_1, \dots, x_n) . Per queste funzioni – denominate *statistiche sufficienti* – la parte di informazione che necessariamente si perde passando da \mathbf{x}_n a t è inessenziale ai fini inferenziali. Dal momento che la funzione di verosimiglianza formalizza tutta l'informazione che i dati campionari forniscono sul parametro incognito di un modello, il valore osservato t di una statistica T ha lo stesso valore informativo dell'intero campione \mathbf{x}_n se per ottenere $L(\theta; \mathbf{x}_n)$ è sufficiente conoscere il valore $T(\mathbf{x}_n) = t$. Poiché \mathbf{x}_n è un oggetto n -dimensionale (un punto di \mathcal{X}^n), si ha un vantaggio operativo se la statistica sufficiente T assume valori in uno spazio \mathcal{T} di dimensioni inferiori ad n .

3.1 Esempio (modello bernoulliano). Nel caso di un modello statistico bernoulliano e di un campione casuale osservato \mathbf{x}_n la funzione di verosimiglianza è

$$\begin{aligned} L(\theta; \mathbf{x}_n) &= \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \\ &= \theta^t (1 - \theta)^{n-t}, \end{aligned}$$

dove $t = \sum_{i=1}^n x_i$. In questo caso, per ottenere funzione di verosimiglianza è sufficiente conoscere il valore della somma delle osservazioni campionarie e non è necessario disporre dei singoli valori delle x_i , $i = 1, \dots, n$.

L'esempio precedente ci consente di comprendere il senso della definizione che segue, nota anche come *criterio di fattorizzazione di Fisher*.

3.2 Definizione (Sufficienza). Dato il modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta)$, una statistica $T : \mathcal{X}^n \rightarrow \mathcal{T}$ è *sufficiente* se esistono due funzioni $h : \mathcal{X}^n \rightarrow \mathbb{R}^+$ (funzione dei dati campionari \mathbf{x}_n) e $g : \Theta \times \mathcal{T} \rightarrow \mathbb{R}^+$ (di θ e di $T(\mathbf{x}_n)$) tali che

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = h(\mathbf{x}_n) \cdot g(\theta, T(\mathbf{x}_n)), \quad \forall \theta \in \Theta, \quad \forall \mathbf{x}_n \in \mathcal{X}^n.$$

In base alla definizione, una statistica T è quindi sufficiente per un modello se il *nucleo* della funzione di verosimiglianza, g , dipende dai dati esclusivamente attraverso il valore t assunto dalla statistica T nel campione osservato \mathbf{x}_n .

3.3 Esempio (modello di Poisson). Consideriamo un campione casuale estratto da una popolazione di Poisson. In questo caso

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \frac{1}{\prod_{i=1}^n x_i!} \theta^{\sum_{i=1}^n x_i} e^{-n\theta} = h(\mathbf{x}_n) \cdot g(\theta, T(\mathbf{x}_n)),$$

dove $h(\mathbf{x}_n) = (\prod_{i=1}^n x_i!)^{-1}$ e $g(\theta; T(\mathbf{x}_n)) = \theta^{\sum_{i=1}^n x_i} e^{-n\theta}$. La statistica $T(\mathbf{x}_n) = \sum_{i=1}^n x_i$, è quindi una statistica sufficiente per il modello considerato.

3.4 Esempio (modello esponenziale negativo). Per il modello esponenziale negativo, la funzione di verosimiglianza associata ad un campione casuale di dimensione n è

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i} \prod_{i=1}^n \mathbb{I}_{(0,+\infty)}(x_i) \propto \theta^n e^{-\theta n \bar{x}_n}. \quad (3.1)$$

da cui, per il criterio di fattorizzazione, si vede facilmente che $\sum_{i=1}^n x_i$ è una statistica sufficiente.

3.5 Esempio (modello normale). Il modello normale dipende da due parametri. A seconda di quale parametro viene considerato incognito (valore atteso, varianza, entrambi), possiamo distinguere tre casi.

1. **Valore atteso incognito e varianza nota.** Per il modello normale $N(\theta, \sigma^2)$, in cui θ è il parametro incognito e la varianza σ^2 è nota, la funzione di

verosimiglianza ottenuta in (2.3) è:

$$\begin{aligned} L(\theta; \mathbf{x}_n) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{n}{2\sigma^2} \sum (x_i - \theta)^2\right\} \prod_{i=1}^n \mathbb{I}_{\mathbb{R}}(x_i) \\ &\propto \exp\left\{-\frac{n}{2\sigma^2} \sum (x_i - \theta)^2\right\} \\ &\propto \underbrace{\exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x}_n)^2\right\}}_{h(\mathbf{x}_n)} \cdot \underbrace{\exp\left\{-\frac{n}{2\sigma^2} (\bar{x}_n - \theta)^2\right\}}_{g(\bar{x}_n, \theta)} \end{aligned}$$

Pertanto

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

da cui si evince che \bar{x}_n è una statistica sufficiente per il modello. Si noti che, in questo caso, è sufficiente anche la statistica $\sum_{i=1}^n x_i$.

2. **Valore atteso noto e varianza incognita.** Per il modello $N(\mu_0, \theta)$, con μ_0 noto, si ha che

$$L(\theta; \mathbf{x}_n) = \frac{1}{(2\pi\theta)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\theta} \sum_{i=1}^n (x_i - \mu)^2\right\} \prod_{i=1}^n \mathbb{I}_{\mathbb{R}^+}(x_i).$$

Se si pone $h(\mathbf{x}_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \prod_{i=1}^n \mathbb{I}_{\mathbb{R}^+}(x_i)$ e $g(t(\mathbf{x}_n; \theta)) = \theta^{-n/2} \exp\left\{-\frac{1}{2\theta} \sum_{i=1}^n (x_i - \mu)^2\right\}$, si verifica che la statistica $\sum_{i=1}^n (x_i - \mu)^2$ è sufficiente per il modello considerato.

3. **Valore atteso e varianza incogniti.** Per il modello $N(\theta_1, \theta_2)$ (entrambi i parametri incogniti) abbiamo mostrato in precedenza che

$$\sum (x_i - \theta_1)^2 = n\hat{\sigma}^2 + n(\bar{x}_n - \theta_1)^2,$$

dove $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$. In questo caso la funzione di verosimiglianza è

$$\begin{aligned} L(\theta_1, \theta_2; \mathbf{x}_n) &= \left(\frac{1}{\sqrt{2\pi\theta_2}}\right)^n \exp\left\{-\frac{n\hat{\sigma}_n}{2\theta_2}\right\} \cdot \exp\left\{-\frac{n}{2\theta_2} (\bar{x}_n - \theta_1)^2\right\} \cdot \prod_{i=1}^n \mathbb{I}_{\mathbb{R}}(x_i) \\ &\propto \frac{1}{(\theta_2)^{\frac{n}{2}}} \cdot \exp\left\{-\frac{n\hat{\sigma}_n}{2\theta_2}\right\} \cdot \exp\left\{-\frac{n}{2\theta_2} (\bar{x}_n - \theta_1)^2\right\}. \end{aligned}$$

Nell'espressione precedente possiamo ignorare il fattore $h(\mathbf{x}_n) = (\sqrt{2\pi})^{-n} \prod_{i=1}^n \mathbb{I}_{\mathbb{R}}(x_i)$ che non dipende dai parametri incogniti. L'espressione restante dipende dai dati attraverso la coppia $(\bar{x}_n, \hat{\sigma}_n^2)$, che quindi è una statistica sufficiente per il modello. La dimensione della statistica sufficiente vettoriale coincide con quella del parametro incognito, (θ_1, θ_2) .

Statistiche sufficienti in modelli con supporto dipendente dal parametro

Nei modelli con supporto dipendente dal parametro incognito, il fattore contenente il prodotto delle funzioni indicatrici $\prod_{i=1}^n \mathbb{I}_{\mathcal{X}_\theta}(x_i)$ che appare in $f_n(\mathbf{x}_n; \theta)$ non può essere trascurato quando si considera la funzione di verosimiglianza, e fa quindi parte della funzione $g(t; \theta)$.

3.6 Esempio (modello uniforme). Per il modello uniforme nell'intervallo $[0, \theta]$ si ha che

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{I}[0, \theta](x_i) = \frac{1}{\theta^n} \mathbb{I}_{[x_{(n)}, +\infty)}(\theta).$$

Il massimo valore del campione $x_{(n)}$ è quindi una statistica sufficiente.

Nell'Esempio 3 abbiamo osservato che, per il modello normale con entrambi i parametri incogniti, la dimensione della statistica sufficiente vettoriale coincide con il numero di parametri incogniti. Tuttavia, in generale, la dimensione della statistica sufficiente non coincide necessariamente con quella del vettore θ .

3.7 Esempio (modello uniforme). Supponiamo di estrarre un campione di ampiezza n da una popolazione uniforme sull'intervallo $[\theta - 1, \theta + 1]$, $\theta \in \mathbb{R}$. La funzione di verosimiglianza associata al campione è

$$L(\theta; \mathbf{x}_n) = \frac{1}{2^n} \prod_{i=1}^n \mathbb{I}[\theta - 1, \theta + 1](x_i) = \frac{1}{2^n} \mathbb{I}[x_{(n)} - 1, x_{(1)} + 1](\theta).$$

In questo caso $h(\mathbf{x}_n) = 1/2^n$ e $g(T(\mathbf{x}_n); \theta) = \mathbb{I}[x_{(n)} - 1, x_{(1)} + 1](\theta)$ dipende dalle osservazioni campionarie attraverso la coppia di valori $(x_{(1)}, x_{(n)})$, che quindi costituisce una statistica sufficiente bidimensionale anche se il parametro incognito è uno solo.

Già dai primi esempi considerati possiamo trarre tre considerazioni rilevanti:

- (i) la statistica sufficiente per un modello *non è unica* (possono esistere infatti statistiche diverse ma del tutto equivalenti in termini informativi su θ);
- (ii) esistono statistiche che sono sufficienti ma non essenziali (ovvero contenenti informazioni superflue);
- (iii) due distinte statistiche sufficienti possono sintetizzare in misura diversa le informazioni campionarie.

Illustriamo queste tre considerazioni con un esempio.

3.8 Esempio (modello di Poisson). Per il modello di Poisson è semplice verificare che, oltre alla statistica $T(\mathbf{x}_n) = \sum_{i=1}^n x_i$, sono sufficienti anche (e non solo) le statistiche $T_1(\mathbf{x}_n) = \bar{x}_n$, $T_2(\mathbf{x}_n) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ e la statistica $T_3(\mathbf{x}_n) =$

$(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ (vettore contenente le osservazioni ordinate in senso crescente, detto *campione ordinato*).

(i) La statistica T_1 è del tutto equivalente alla statistica T ai fini della ricostruzione di $L(\theta; \mathbf{x}_n)$. Più in generale, ogni funzione biunivoca di una statistica sufficiente è ancora una statistica sufficiente.

(ii) La statistica T_2 è certamente sufficiente, ma contiene una componente inessenziale ($\sum x_i^2$) per la ricostruzione della funzione di verosimiglianza. Si noti che, per ogni campione osservato \mathbf{x}_n , il valore della statistica $T_1(\mathbf{x}_n)$ si può ottenere come funzione di $T_2(\mathbf{x}_n)$, ma che non è vero il viceversa.

(iii) La statistica T_3 è sufficiente per il modello, ma contiene informazioni inutili rispetto alla determinazione della funzione di verosimiglianza. Infatti $\sum_{i=1}^n x_{(i)} = \sum_{i=1}^n x_i$: la conoscenza degli n valori di del campione ordinato consentono quindi di ricostruire la funzione di verosimiglianza, ma la conoscenza dei singoli valori $x_{(i)}, i = 1, \dots, n$ è del tutto superflua al tale scopo.

La statistica T , che sintetizza in un numero naturale l'informazione dell'intero campione casuale, riassume quindi in modo più efficiente l'informazione campionaria su θ di quanto faccia T_3 , che è invece un vettore di n numeri. L'aspetto cruciale consiste anche in questo caso nel fatto che la statistica T può essere ottenuta come funzione di T_3 , ma non viceversa.

3.3 Statistiche sufficienti minimali

Tra tutte le statistiche sufficienti di un modello, siamo interessati ad individuare quelle che comportano la massima riduzione dimensionale possibile rispetto al campione osservato.

3.9 Definizione (Sufficienza minimale I). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Omega)$, una statistica sufficiente T si dice *minimale* se, per ogni altra statistica sufficiente T' per il modello, T è una funzione di T' .

Questa definizione formalizza quanto osservato nel caso specifico del modello di Poisson nell'esempio 3.8. La comprensione di questa definizione necessita tuttavia un approfondimento che riguarda il concetto di partizione dello spazio dei campioni indotto da una statistica.

Partizioni sufficienti

Ricordiamo che una statistica T è per definizione un'applicazione da \mathcal{X}^n (spazio campionario) in \mathcal{T} (insieme dei valori che può assumere T). Indichiamo con A_t l'insieme degli elementi di \mathcal{X}^n in corrispondenza dei quali la statistica T assume il valore $t \in \mathcal{T}$:

$$A_t = \{\mathbf{x}_n \in \mathcal{X}^n : T(\mathbf{x}_n) = t\}.$$

Poichè per ogni coppia $(t, t') \in \mathcal{T} : t \neq t'$

$$A_t \cap A_{t'} = \emptyset, \quad \text{e} \quad \bigcup_{t \in \mathcal{T}} A_t = \mathcal{X}^n,$$

possiamo affermare che la statistica T induce una partizione dello spazio campionario, costituito dagli insiemi A_t , $t \in \mathcal{T}$. La partizione indotta da una statistica T sufficiente si chiama *partizione sufficiente*. Dato un valore t che può assumere T , tutti i campioni che appartengono al medesimo insieme A_t della partizione sufficiente, danno luogo alla stessa funzione di verosimiglianza. Infatti, fissato un valore t in \mathcal{T} , per tutti i campioni $\mathbf{x}_n \in A_t$ si ha che

$$L(\theta; \mathbf{x}_n) \propto g(\theta, T(\mathbf{x}_n)) = g(\theta; t).$$

Se consideriamo due statistiche sufficienti T_1 e T_2 non in corrispondenza biunivoca, indicati con \mathcal{T}_1 e \mathcal{T}_2 gli insiemi dei valori che possono assumere, ad esse corrispondono due distinte partizioni sufficienti:

$$\{A_{t_i}, t_i \in \mathcal{T}_i\}, \quad \text{dove} \quad A_{t_i} = \{\mathbf{x}_n \in \mathcal{X}^n : T(\mathbf{x}_n) = t_i\}, \quad i = 1, 2.$$

Dire che T_1 è funzione di T_2 vuol dire che, per ogni generica coppia di campioni \mathbf{x}_n e \mathbf{y}_n , il fatto che $T_2(\mathbf{x}_n) = T_2(\mathbf{y}_n)$ implica che $T_1(\mathbf{x}_n) = T_1(\mathbf{y}_n)$ (e non viceversa). Possiamo, ad esempio, ottenere il valore di $t_1 = \sum_{i=1}^n x_i$ conoscendo i valori $t_2 = (x_{(1)}, \dots, x_{(n)})$ ma non possiamo ricostruire i valori del campione ordinato, t_2 , partendo dalla conoscenza di t_1 (e infatti possiamo ottenere $\sum_{i=1}^n x_i$ facendo la somma di $[x_{(1)}, \dots, x_{(n)}]$ ma non possiamo fare il viceversa). In termini di partizioni, ciò vuol dire che ogni insieme A_{t_2} della partizione indotta da T_2 è contenuto in un insieme A_{t_1} della partizione indotta da T_1 : la partizione indotta da T_1 risulta quindi meno fine di quella indotta da T_2 . Da quanto detto, discende quindi che la partizione indotta da una statistica sufficiente minimale è la partizione *meno fine* tra tutte le partizioni sufficienti per un dato modello.

Le precedenti considerazioni implicano che la *sufficienza* di una statistica T richiede che la funzione di verosimiglianza sia la stessa per tutti i campioni appartenenti ad un elemento A_t della partizione che induce; la *minimalità* richiede che ciò si verifichi per *tutti e i soli* campioni di A_t . Questo vuol dire che, se T è sufficiente e minimale, a campioni appartenenti ad insiemi distinti della partizione indotta da T corrispondono funzioni di verosimiglianza non equivalenti. Le precedenti considerazioni si riassumono nella seguente definizione, che esplicita la precedente:

3.10 Definizione (Sufficienza minimale II). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Omega)$, una statistica sufficiente T si dice *minimale* se la funzione di verosimiglianza è la stessa, a meno di fattori costanti, per tutti e i soli campioni \mathbf{x}_n tali che $T(\mathbf{x}_n) = t$, per ogni $t \in \mathcal{T}$.

3.11 Osservazione/i (partizione di verosimiglianza). Anche la funzione di verosimiglianza, benchè non sia una statistica (in quanto dipende da θ , induce una

partizione dello spazio dei campioni, detta *partizione di verosimiglianza*. In ciascun elemento della partizione di verosimiglianza si trovano tutti e i soli campioni che danno luogo a verosimiglianze equivalenti (ovvero, proporzionali tra loro). Una statistica sufficiente è quindi minimale se la partizione che induce coincide con la partizione di verosimiglianza.

3.12 Osservazione/i (unicità sostanziale della statistica sufficiente minimale). Dalla precedente osservazione discende che tutte le statistiche sufficienti minimali inducono la stessa partizione, che coincide con quella di verosimiglianza. Le statistiche sufficienti minimali sono quindi legate tra loro da relazioni biunivoche. Possiamo allora affermare che le statistiche sufficienti minimali costituiscono una classe di equivalenza, ovvero una classe di funzioni dei dati diverse tra loro (come, ad esempio, somma e media campionaria nel modello di Poisson e ogni loro funzione biunivoca) alle quali però corrisponde la stessa partizione dello spazio dei campioni. In questo senso possiamo parlare di unicità sostanziale della statistica sufficiente minimale.

3.13 Esempio (modello di Poisson). Consideriamo il modello di Poisson e le statistiche sufficienti $T_1(\mathbf{x}_n) = \sum_{i=1}^n x_i$ e $T_2(\mathbf{x}_n) = (x_{(1)}, \dots, x_{(n)})$. Risulta evidente che T_1 è funzione di T_2 : presi due campionari distinti di dimensione n che danno luogo alla stessa campione ordinato, questi danno luogo anche alla stessa somma ($T_2(\mathbf{x}_n) = T_2(\mathbf{y}_n) \Rightarrow T_1(\mathbf{x}_n) = T_1(\mathbf{y}_n)$ e non viceversa). La partizione indotta da T_1 è quindi meno fine di quella indotta da T_2 . Infatti, se ad esempio per $n = 3$ consideriamo i due campioni $(5, 3, 2)$ e $(1, 7, 2)$, questi appartengono a due insiemi A_{t_2} distinti della partizione indotta da T_2 (perchè danno luogo a campioni ordinati distinti), ma sono nello stesso insieme A_{t_1} della partizione indotta da T_1 (poichè producono lo stesso valore della somma campionaria). Poichè però i due campioni danno luogo alla stessa verosimiglianza, la statistica T_2 non è minimale: nella partizione sufficiente che induce, campioni distinti ma equivalenti (a cui è associata una stessa funzione di verosimiglianza) appartengono a elementi distinti della partizione. La statistica sufficiente T_1 è invece minimale: a campioni che appartengono a insiemi distinti della partizione indotta da T_1 corrispondono funzioni di verosimiglianza di θ non equivalenti.

Le considerazioni precedenti si concretizzano in un criterio operativo di verifica di sufficienza minimale di una statistica¹

3.14 Teorema (Criterio di Lehmann-Scheffè). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta))$, $\theta \in \Omega$, supponiamo che esista una statistica T tale che, per qualsiasi coppia di campioni \mathbf{x}_n e \mathbf{y}_n , il rapporto delle verosimiglianze $L(\theta; \mathbf{x}_n)/L(\theta; \mathbf{y}_n)$ è costante rispetto a θ se e solo se $T(\mathbf{x}_n) = T(\mathbf{y}_n)$. Allora T è una statistica sufficiente minimale per il modello.

¹Per la dimostrazione si rimanda a Casella-Berger (2002, p. 281).

3.15 Esempio (modello normale). Consideriamo il modello $N(\theta, 1)$. Il rapporto delle verosimiglianze per due campioni casuali \mathbf{x}_n e \mathbf{y}_n è

$$\frac{L(\theta; \mathbf{x}_n)}{L(\theta; \mathbf{y}_n)} = \frac{\exp\{-(\theta - \bar{x}_n)^2/2\}}{\exp\{-(\theta - \bar{y}_n)^2/2\}}$$

Se assumiamo che il rapporto sia costante rispetto a θ (ovvero se i due campioni danno luogo a stessa verosimiglianza), necessariamente $\bar{x}_n = \bar{y}_n$ (e i due campioni \mathbf{x}_n e \mathbf{y}_n appartengono allo stesso insieme della partizione indotta dalla statistica media campionaria). Se invece $\bar{x}_n = \bar{y}_n$ (i due campioni \mathbf{x}_n e \mathbf{y}_n appartengono allo stesso insieme della partizione indotta dalla statistica media campionaria), il rapporto $L(\theta; \mathbf{x}_n)/L(\theta; \mathbf{y}_n)$ non dipende da θ (i due campioni producono la stessa verosimiglianza). La media campionaria soddisfa la condizione necessaria e sufficiente del criterio di Lehmann-Scheffè ed è quindi una statistica sufficiente minimale per il modello considerato.

3.16 Esempio (modello uniforme). In questo caso, dati due campioni casuali \mathbf{x}_n e \mathbf{y}_n , si ha che

$$\frac{L(\theta; \mathbf{x}_n)}{L(\theta; \mathbf{y}_n)} = \frac{\mathbb{I}[x_{(n)}, +\infty)(\theta)}{\mathbb{I}[y_{(n)}, +\infty)(\theta)}.$$

Questo rapporto non dipende da θ se e solo se $x_{(n)} = y_{(n)}$. La statistica *massimo campionario* è quindi sufficiente minimale per il modello uniforme in $[0, \theta]$.

3.17 Osservazione/i. La statistica sufficiente minimale T di un modello non è unica: ogni statistica T' , funzione biunivoca di T , è infatti sufficiente minimale, dal momento che alle due statistiche corrisponde la stessa partizione (minimale) di \mathcal{X}^n .

Definizione frequentista di sufficienza

La definizione del concetto di sufficienza adottata in questo capitolo è coerente con il ruolo centrale della funzione di verosimiglianza nel formalizzare l'informazione che i dati forniscono sui parametri dei modelli. Nella maggior parte dei testi, la definizione che noi abbiamo utilizzato costituisce un criterio pratico per individuare statistiche sufficienti (da cui il nome di Criterio di fattorizzazione). La seguente caratterizzazione della sufficienza è in genere utilizzata come definizione².

3.18 Teorema. Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Omega)$, la statistica T è sufficiente se e solo se la distribuzione di probabilità di \mathbf{X}_n condizionata a $T = t$ (per qualunque $t \in \mathcal{T}$) non dipende da θ .

Dimostrazione. Per semplicità consideriamo il caso in cui $f_T(\cdot; \theta) = \mathbb{P}_\theta(\cdot)$ sia una funzione di massa di probabilità (ovvero T una v.a. discreta). Per dimostrare il teorema dobbiamo mostrare che valgono le due implicazioni.

²La cui dimostrazione è tratta da veda Piccinato (2009), pp. 136-7.

a) Supponiamo che T sia sufficiente. Per il criterio di fattorizzazione esistono allora due funzioni h e g tali che

$$\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n) = h(\mathbf{x}_n) \cdot g(\theta, t),$$

dove $t = T(\mathbf{x}_n)$. Si ha quindi che (per la definizione di probabilità condizionata)

$$\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n | T = t) = \frac{\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n, T = t)}{\mathbb{P}_\theta(T = t)} = \frac{\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n)}{\mathbb{P}_\theta(T = t)},$$

dove l'ultima uguaglianza si ottiene osservando che l'evento $(\mathbf{X}_n = \mathbf{x}_n)$ è contenuto nell'evento $(T = t)$ e quindi che l'intersezione dei due eventi coincide con l'evento $(\mathbf{X}_n = \mathbf{x}_n)$. Osserviamo che il denominatore dell'ultima espressione può essere riscritto come

$$\mathbb{P}_\theta(T = t) = \sum_{\mathbf{x}_n: T(\mathbf{x}_n)=t} h(\mathbf{x}_n) \cdot g(\theta, t) = g(\theta, t) \sum_{\mathbf{x}_n: T(\mathbf{x}_n)=t} h(\mathbf{x}_n) = g(\theta) \cdot \bar{h}(\mathbf{x}_n).$$

Sostituendo in (3.2) si ha quindi

$$\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n | T = t) = \frac{h(\mathbf{x}_n) \cdot g(\theta, t)}{\bar{h}(\mathbf{x}_n) \cdot g(\theta, t)} = \frac{h(\mathbf{x}_n)}{\bar{h}(\mathbf{x}_n)}$$

che non dipende da θ .

b) Supponiamo ora invece che $\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n | T = t)$ non dipenda da θ , ovvero che sia funzione dei soli dati campionari. Indicando con $h(\mathbf{x}_n)$ tale funzione e ricorrendo alla definizione di probabilità condizionata abbiamo che

$$\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n) = \mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n | T = t) \cdot \mathbb{P}_\theta(T = t) = h(\mathbf{x}_n) \cdot \mathbb{P}_\theta(T = t),$$

ovvero che $\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n)$ può essere fattorizzata nel prodotto di una funzione dei soli dati, $h(\mathbf{x}_n)$, e una funzione del parametro, $\mathbb{P}_\theta(T = t)$, che dipende dai dati solo attraverso la statistica $T(\mathbf{x}_n)$.

L'idea intuitiva di questo risultato (o di questa definizione) è che il campione fornisce informazioni su θ grazie al fatto che la sua distribuzione di probabilità dipende da θ . Se il condizionamento a $T = t$ cancella il legame tra campione e parametro, ciò vuol dire che tutta l'informazione su θ offerta dal campione è riassunta dalla conoscenza di $T = t$.

3.19 Osservazione/i Ricordando che campioni diversi che danno luogo allo stesso valore t di una statistica T si trovano tutti nello stesso insieme A_t della partizione indotta da T , il risultato precedente implica che la probabilità dei campioni che appartengono allo stesso A_t , subordinatamente al fatto di osservare uno dei campioni in tale insieme, non dipende da θ .

3.20 Esempio (modello di Poisson). Riprendendo ancora il modello di Poisson, mostriamo che $\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n | \sum X_i = t)$ non dipende da θ . Si osservi innanzitutto che

$$\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n | \sum X_i = t) = \frac{\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n, \sum X_i = t)}{\mathbb{P}_\theta(\sum X_i = t)} = \frac{\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n)}{\mathbb{P}_\theta(\sum X_i = t)},$$

dove l'ultima uguaglianza deriva dal fatto che l'evento $(\mathbf{X}_n = \mathbf{x}_n)$ è contenuto nell'evento $(\sum_{i=1}^n X_i = t)$. Osservando che

$$\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \frac{1}{x_1! \dots x_n!} e^{-n\theta} \theta^{\sum_{i=1}^n x_i}$$

e ricordando che la somma di n v.a. di Poisson indipendenti ha distribuzione di Poisson di parametro $n\theta$,

$$\mathbb{P}_\theta \left(\sum X_i = t \right) = \frac{1}{t!} e^{-n\theta} (n\theta)^t,$$

si ottiene la probabilità

$$\mathbb{P}_\theta \left(\mathbf{X}_n = \mathbf{x}_n \mid \sum_{i=1}^n X_i = t \right) = \frac{t!}{n^t x_1! \dots x_n!},$$

è indipendente da θ .

3.4 Statistiche sufficienti e famiglie esponenziali

Per un generico modello che forma una famiglia esponenziale è particolarmente semplice individuare una statistica sufficiente minimale. Nel caso uniparametrico, si è visto nella (1.4) che

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = h_n(\mathbf{x}_n) \exp\{\eta(\theta)T_n(\mathbf{x}_n) - B_n(\theta)\}.$$

Per il criterio di fattorizzazione, la statistica $T_n(\mathbf{x}_n) = \sum_{i=1}^n T(x_i)$, è pertanto sufficiente per il modello considerato. È semplice verificare che, in base al criterio di Lehman-Scheffè, la statistica $\sum_{i=1}^n T(x_i)$ è anche minimale. Infatti, dati due campioni \mathbf{x}_n e \mathbf{y}_n , si ha che:

- (i) se $T_n(\mathbf{x}_n) = T_n(\mathbf{y}_n)$, allora $L(\theta; \mathbf{x}_n)/L(\theta; \mathbf{y}_n) = h_n(\mathbf{x}_n)/h_n(\mathbf{y}_n)$, che è costante rispetto a θ ;
- (ii) se $L(\theta; \mathbf{x}_n)/L(\theta; \mathbf{y}_n)$ è costante rispetto a θ , necessariamente si deve avere che $T_n(\mathbf{x}_n) = T_n(\mathbf{y}_n)$.

Il risultato si estende al caso di modelli k -parametrici ($\Omega \subset \mathbb{R}^k, k > 1$) che sono famiglie esponenziali. Ricordiamo che, in questo caso, la funzione di verosimiglianza è

$$L(\boldsymbol{\theta}; \mathbf{x}_n) = f_n(\mathbf{x}_n; \boldsymbol{\theta}) = h_n(\mathbf{x}_n) \exp \left\{ \sum_{j=1}^k \eta_j(\boldsymbol{\theta}) T_j(\mathbf{x}_n) - B_n(\boldsymbol{\theta}) \right\}.$$

Per il criterio di fattorizzazione, il vettore

$$\mathbf{T}(\mathbf{x}_n) = (T_1(\mathbf{x}_n), \dots, T_j(\mathbf{x}_n), \dots, T_k(\mathbf{x}_n)),$$

in cui $T_j(\mathbf{x}_n) = \sum_{i=1}^n T_j(x_i)$, $j = 1, \dots, k$, è una statistica sufficiente per il modello corrispondente. La minimalità si verifica in modo analogo al caso $k = 1$ con il criterio di Lehman-Scheffè.

3.21 Esempio Ricordando che, per i modelli che seguono, le leggi di probabilità sono famiglie esponenziali, possiamo affermare che:

- per il modello Bernuoliano, di Poisson, esponenziale negativo e normale con varianza nota, la statistica $T(\mathbf{x}_n) = \sum_{i=1}^n x_i$ è una statistica sufficiente minimale;
- per il modello normale con valore atteso noto, μ , la statistica $T(\mathbf{x}_n) = \sum_{i=1}^n (x_i - \mu)^2$ è sufficiente minimale;
- per il modello $N(\theta_1, \theta_2)$ il vettore $\mathbf{T}(\mathbf{x}_n) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ è sufficiente minimale.

3.5 Esercizi

Capitolo 4

Inferenza bayesiana

4.1 Introduzione

In questo capitolo vengono introdotti i principi e gli elementi costitutivi dell'inferenza statistica bayesiana. Si tratta di una impostazione alternativa a quella frequentista (o classica), ispirata al cosiddetto *principio del campionamento ripetuto*. L'impostazione bayesiana ripetta invece il *principio di verosimiglianza*, ma i metodi che propone si distinguono da quelli basati sulle sintesi della funzione di verosimiglianza. Nell'approccio classico e in quello basato sulla funzione di verosimiglianza il parametro θ di un modello statistico è una quantità incognita ma fissa. La logica bayesiana si caratterizza invece con l'assegnare una probabilità a ogni evento incerto di un problema. Pertanto, nel caso dei modelli parametrici, la quantità incognita θ viene trattata come una variabile aleatoria che, da ora in poi, indicheremo con Θ . Denoteremo invece con Ω lo spazio dei parametri, ovvero l'insieme di valori che la variabile aleatoria Θ può assumere. Alla v.a. Θ viene quindi assegnata una distribuzione di probabilità – la distribuzione *iniziale* – che descrive l'incertezza e le informazioni che si hanno sul parametro del modello prima di effettuare l'esperimento che genera i dati. L'idea di base dell'inferenza bayesiana consiste nel combinare le due fonti di informazione, quella sperimentale, fornita dai dati attraverso la funzione di verosimiglianza, e quella extra-sperimentale (o pre-sperimentale), espressa dalla distribuzione iniziale. La sintesi si attua attraverso il teorema di Bayes, da cui trae il nome l'impostazione inferenziale in esame, che produce la distribuzione *finale* del parametro. I metodi bayesiani si basano su elaborazioni effettuate su tale distribuzione.

In questo capitolo, dopo aver delineato lo schema probabilistico generale di un modello bayesiano, si esaminano alcuni esempi notevoli in cui si mostra come ottenere la distribuzione finale.

4.2 Elementi costitutivi

Nell'analisi bayesiana si assume che sia le osservazioni \mathbf{X}_n che il parametro Θ siano variabili aleatorie. L'assegnazione della distribuzione di probabilità congiunta alla coppia (\mathbf{X}_n, Θ) avviene in genere fissando due leggi di probabilità: la legge condizio-

nata di \mathbf{X}_n a un valore θ di Θ , $f_{\mathbf{X}_n|\theta}(\mathbf{x}_n|\theta)$, e la distribuzione marginale di θ , $f_{\Theta}(\theta)$. La distribuzione condizionata $f_{\mathbf{X}_n|\theta}(\mathbf{x}_n|\theta)$ coincide con la distribuzione campionaria di \mathbf{X}_n che, nei capitoli precedenti, abbiamo indicato con la notazione $f_n(\mathbf{x}_n; \theta)$: da questo capitolo in poi utilizziamo il simbolo di condizionamento (a θ) che ha senso in quanto ora θ indica un valore per la variabile aleatoria Θ . La distribuzione $f_{\Theta}(\theta)$ viene denominata *distribuzione iniziale* (o *a priori*) di θ . Le due distribuzioni $f_{\mathbf{X}_n|\theta}(\mathbf{x}_n|\theta)$ e $f_{\Theta}(\theta)$ determinano in modo univoco la distribuzione congiunta di (\mathbf{X}_n, Θ) :

$$f_{\mathbf{X}_n, \Theta}(\mathbf{x}_n, \theta) = f_{\mathbf{X}_n|\theta}(\mathbf{x}_n|\theta) \times f_{\Theta}(\theta).$$

Obiettivo dell'analisi bayesiana è l'aggiornamento della distribuzione iniziale alla luce dei risultati sperimentali, \mathbf{x}_n , e la determinazione della *distribuzione finale* (o *a posteriori*) di θ , ovvero della legge $f_{\theta|\mathbf{x}_n}(\theta|\mathbf{x}_n)$. Questo risultato viene conseguito attraverso il teorema di Bayes. Assumendo, per semplicità, che Θ sia una v.a. assolutamente continua¹, il teorema ci permette di dire che la distribuzione² finale è:

$$f_{\theta|\mathbf{x}_n}(\theta|\mathbf{x}_n) = \frac{f_{\mathbf{X}_n|\theta}(\mathbf{x}_n|\theta)f_{\Theta}(\theta)}{f_{\mathbf{X}_n}(\mathbf{x}_n)},$$

dove

$$f_{\mathbf{X}_n}(\mathbf{x}_n) = \int_{\Omega} f_{\mathbf{X}_n|\theta}(\mathbf{x}_n|\theta)f_{\Theta}(\theta)d\theta$$

indica la *distribuzione marginale* (o *distribuzione predittiva a priori*) di \mathbf{X}_n calcolata in \mathbf{x}_n . Per seguire la notazione più comunemente utilizzata, nel seguito indicheremo con $f_n(\mathbf{x}_n|\theta)$ la distribuzione campionaria di \mathbf{X}_n , con $\pi(\theta)$ e $\pi(\theta|\mathbf{x}_n)$ rispettivamente la distribuzione iniziale e finale di θ e con $m(\mathbf{x}_n)$ la densità marginale di \mathbf{X}_n . In questo modo possiamo scrivere che

$$\pi(\theta|\mathbf{x}_n) = \frac{f_n(\mathbf{x}_n|\theta)\pi(\theta)}{m_n(\mathbf{x}_n)}, \quad (4.1)$$

dove

$$m_n(\mathbf{x}_n) = \int_{\Omega} f_n(\mathbf{x}_n|\theta)\pi(\theta)d\theta.$$

Nella formula precedente i dati si assumono noti: la logica bayesiana è pertanto post-sperimentale. In (4.1), $f_n(\mathbf{x}_n|\theta)$ coincide quindi con la funzione di verosimiglianza di θ associata al campione \mathbf{x}_n , mentre il denominatore – $m_n(\mathbf{x}_n)$ – non dipende da θ ma è funzione dai dati osservati: gli elementi di \mathbf{x}_n sono infatti valori numerici, in quanto realizzazioni della v.a. \mathbf{X}_n . Possiamo quindi scrivere che

$$\pi(\theta|\mathbf{x}_n) = c \times f_n(\mathbf{x}_n|\theta) \times \pi(\theta),$$

dove $c = 1/m_n(\mathbf{x}_n)$ è anche denominata *costante di normalizzazione* della distribuzione a posteriori. Nel seguito indicheremo che, a posteriori, la v.a. ha distribuzione

¹Il caso in cui Θ è una v.a. discreta si ottiene in modo del tutto analogo (con sommatorie che sostituiscono gli integrali).

²Con un leggero abuso, utilizziamo, qui e nel seguito, il termine *distribuzione* anche come sinonimo di *densità*.

coincidente con quella di una v.a. W con la notazione

$$\Theta|\mathbf{x}_n \sim W.$$

Inoltre, supponendo che gli integrali coinvolti esistano, indicheremo con

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \int_{\Omega} \theta \pi(\theta|\mathbf{x}_n) d\theta \quad \text{e} \quad \mathbb{V}_{\theta}[\Theta|\mathbf{x}_n] = \int_{\Omega} (\theta - \mathbb{E}[\theta|\mathbf{x}_n])^2 \pi(\theta|\mathbf{x}_n) d\theta$$

il valore atteso e la varianza della distribuzione a posteriori di θ .

Possiamo riassumere quanto detto sopra osservando che l'analisi bayesiana consiste essenzialmente in un processo di aggiornamento dell'informazione espressa dalla distribuzione a priori alla luce dei dati campionari. Tale aggiornamento si realizza con la determinazione della distribuzione a posteriori (4.1), che sintetizza le due fonti di informazioni a disposizione sul parametro incognito:

1. informazione pre-sperimentale, rappresentata dalla distribuzione iniziale, $\pi(\theta)$
2. informazione post-sperimentale, rappresentata dalla funzione di verosimiglianza, $L(\theta; \mathbf{x}_n) \propto f_n(\mathbf{x}_n|\theta)$.

4.1 Osservazione/i.

- a) Spesso si usano distribuzioni a priori per Θ che dipendono da ulteriori parametri, detti *iperparametri*. Nei casi più semplici (quelli considerati in questo testo), si assume che gli iperparametri siano noti.
- b) L'analisi bayesiana ha carattere dinamico: se si esegue un secondo esperimento e si ottengono nuovi dati \mathbf{x}_{n_2} , la distribuzione a posteriori ottenuta dopo il primo esperimento (che fornisce i dati \mathbf{x}_{n_1}) viene utilizzata come distribuzione iniziale per il parametro, mentre le osservazioni relative al secondo esperimento vengono utilizzate per ottenere la funzione di verosimiglianza e calcolare una nuova distribuzione finale:

$$\pi(\theta|\mathbf{x}_{n_1}, \mathbf{x}_{n_2}) \propto f_{n_2}(\mathbf{x}_{n_2}|\theta) \times \pi(\theta|\mathbf{x}_{n_1}).$$

- c) Nel contesto probabilistico che stiamo considerando è necessario distinguere con attenzione la distribuzione *marginale* del vettore aleatorio $\mathbf{X}_n = (X_1, \dots, X_n)$, $m(\cdot)$, dalla distribuzione di \mathbf{X}_n *condizionata* a $\Theta = \theta$, $f_n(\cdot|\theta)$. L'usuale ipotesi di indipendenza e identica distribuzione riguarda quest'ultima distribuzione. Diciamo infatti che le v.a. X_1, \dots, X_n sono indipendenti (e identicamente distribuite) *condizionatamente a θ* se

$$f_n(\mathbf{x}_n|\theta) = \prod_{i=1}^n f_{X_i}(x_i|\theta).$$

L'indipendenza condizionata a θ di X_1, \dots, X_n non implica però l'indipendenza (marginale) delle stesse X_1, \dots, X_n , ovvero non implica che

$$m_n(\mathbf{x}_n) = \prod_{i=1}^n m_1(x_i),$$

dove (nel caso di v.a. assolutamente continue) $m_1(x_i) = \int_{\Omega} f_{X_i}(x_i|\theta)\pi(\theta)d\theta$, indica la funzione di densità (o di massa di probabilità, nel caso di v.a. discrete) marginale di una singola v.a. X_i .

4.3 Analisi coniugata

L'implementazione del metodo bayesiano richiede una distribuzione iniziale per Θ , ovvero di una distribuzione di probabilità che in generale ha supporto Ω . Il problema della scelta di tale distribuzione costituisce uno degli aspetti più importanti e critici dell'analisi bayesiana. La letteratura sull'argomento è estremamente vasta. Ci limitiamo in questa sede alla descrizione di due approcci: (1) analisi coniugata (in questo capitolo); (2) analisi non-informativa (vedi Cap. 5). Nel caso dell'analisi coniugata, la scelta della distribuzione iniziale è fortemente condizionata dall'esigenza di ottenere una distribuzione finale in modo agevole dal punto di vista matematico. Si tiene conto sia della trattabilità matematica che della necessità che la distribuzione a priori sia sufficientemente flessibile per rappresentare le informazioni pre-sperimentali sul parametro.

4.2 Definizione (Famiglia coniugata di distribuzioni iniziali). Data la famiglia di distribuzioni $\mathcal{F}_n = (f_n(\cdot|\theta), \theta \in \Omega)$, una classe \mathcal{P} di distribuzioni iniziali per Θ è una *famiglia coniugata* a \mathcal{F}_n se la distribuzione a posteriori di θ , $\pi(\theta|\mathbf{x}_n) \propto f_n(\mathbf{x}_n|\theta) \times \pi(\theta)$, appartiene a \mathcal{P} , per ogni $\pi(\theta) \in \mathcal{P}$ e per ogni $\mathbf{x}_n \in \mathcal{X}^n$.

I principali modelli statistici considerati in questo testo (bernoulliano, binomiale, Poisson, esponenziale, normale, gamma) posseggono una famiglia di distribuzioni coniugate.

4.3 Esempio (modello bernoulliano). Sia X_1, \dots, X_n un campione casuale con $X_i|\theta \sim \text{Ber}(\theta)$, $\theta \in [0, 1]$. Supponiamo che l'informazione disponibile sul parametro incognito del modello sia rappresentata da una densità³ $\text{Beta}(\alpha, \beta)$, $\alpha, \beta > 0$:

$$\pi(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1 - \theta)^{\beta-1}, \quad \theta \in [0, 1], \quad \alpha, \beta > 0.$$

In questo caso $\mathbb{E}[\Theta] = \alpha/(\alpha + \beta)$, $\text{Mo}[\Theta] = (\alpha - 1)/(\alpha + \beta - 2)$, $\mathbb{V}[\Theta] = \alpha\beta/[(\alpha + \beta)^2(\alpha + \beta + 1)]$ (vedi Esempio 5.1). Si assume che i valori di (α, β) – gli iperparametri – siano noti e scelti in modo tale che la densità beta corrispondente traduca nel modo migliore possibile l'informazione e l'incertezza su θ . Prendendo, ad esempio, $\alpha = \beta = 1$, $\pi(\theta)$ coincide con la distribuzione uniforme sull'intervallo $[0, 1]$: tale scelta di α e β è quindi ragionevole quando, a priori, non si vogliono privilegiare particolari valori per il parametro incognito. Se esistono invece dei sottoinsiemi di $[0, 1]$ che si ritengono più plausibili per θ , è allora opportuno effettuare scelte diverse.

³Si noti che la v.a. beta ha come supporto l'intervallo $[0, 1]$ che, per il modello statistico considerato, coincide con lo spazio dei parametri, Ω .

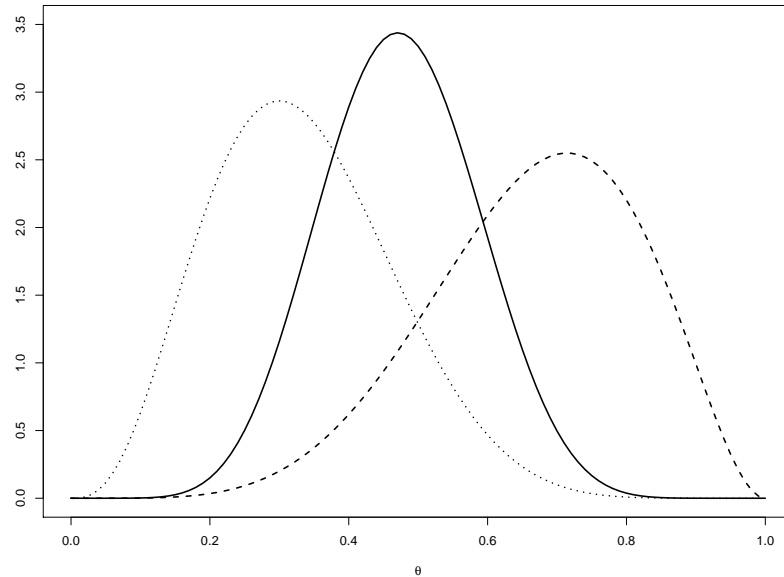


Figura 4.1: Grafici di $\pi(\theta)$ (linea tratteggiata), $L(\theta; \mathbf{x}_n)$ (linea a puntini) e $\pi(\theta|\mathbf{x}_n)$ (linea continua) per $n = 10$, $\sum_{i=1}^n x_i = 3$, $\alpha = 6$, $\beta = 3$.

Supponiamo di avere osservato il campione $\mathbf{x}_n = (x_1, \dots, x_n)$. La funzione di verosimiglianza di θ è quindi $L(\theta; \mathbf{x}_n) \propto f_n(\mathbf{x}_n|\theta) = \theta^{y_n}(1-\theta)^{n-y_n}$, dove $y_n = \sum_{i=1}^n x_i$. Si ottiene allora che

$$\begin{aligned} \pi(\theta|\mathbf{x}_n) &\propto \underbrace{[\theta^{y_n}(1-\theta)^{n-y_n}]}_{L(\theta;\mathbf{x}_n)} \times \underbrace{[\theta^{\alpha-1}(1-\theta)^{\beta-1}]}_{\pi(\theta)} \\ &\propto \theta^{(\alpha+y_n)-1}(1-\theta)^{(\beta+n-y_n)-1} \\ &\propto \theta^{\bar{\alpha}-1}(1-\theta)^{\bar{\beta}-1}, \text{quad } \theta \in [0, 1] \end{aligned}$$

con $\bar{\alpha} = \alpha + y_n$ e $\bar{\beta} = \beta + n$. Nell'ultima espressione di $\pi(\theta|\mathbf{x}_n)$ si riconosce il nucleo di una densità beta di parametri $\bar{\alpha}$ e $\bar{\beta}$. Pertanto $\Theta|\mathbf{x}_n \sim \text{Beta}(\bar{\alpha}, \bar{\beta})$. In questo caso risulta quindi nota la distribuzione di probabilità a posteriori del parametro: si tratta ancora di una densità beta (che pertanto è coniugata al modello bernoulliano) con parametri diversi da quelli della distribuzione iniziale e dipendenti sia da (α, β) che da (n, y_n) . Osserviamo che la sintesi tra il contributo della distribuzione iniziale e quello della funzione di verosimiglianza si attua semplicemente attraverso l'aggiornamento dei parametri della distribuzione iniziale e non attraverso un cambiamento della famiglia di distribuzioni di probabilità.

4.4 Osservazione/i. Per disegnare i grafici di $\pi(\theta)$, $L(\theta; \mathbf{x}_n)$ e $\pi(\theta|\mathbf{x}_n)$ si può ricorrere alle opportune funzioni R. Ponendo $\alpha = \text{alpha}$, $\beta = \text{beta}$, $y_n = \text{yn}$, $\bar{\alpha} = \text{alpha.p}$ e $\bar{\beta} = \text{beta.p}$ e osservando che $L(\theta; \mathbf{x}_n)$ (pur non essendo una funzione di densità) è proporzionale a una densità beta di parametri $(y_n + 1, n - y_n + 1)$ si ha:

$\pi(\theta)$	<code>curve(dbeta(x, shape1=alpha, shape2=beta), from=0, to=1)</code>
$L(\theta; \mathbf{x}_n)$	<code>curve(dbeta(x, shape1=yn+1, shape2=n-yn+1), add=T, lty=2)</code>
$\pi(\theta \mathbf{x}_n)$	<code>curve(dbeta(x, shape1=alpha.p, shape2=beta.p), add=T, lty=3)</code>

Esempio numerico. La Figura 4.1 mostra i grafici di $\pi(\theta)$, $L(\theta; \mathbf{x}_n)$ e $\pi(\theta|\mathbf{x}_n)$ per $n = 10$, $\sum_{i=1}^n x_i = 3$, $\alpha = 6$, $\beta = 3$. Si osservi che la distribuzione a posteriori si situa in posizione intermedia tra la funzione di verosimiglianza e la distribuzione a priori. Il punto di massimo di $\pi(\theta|\mathbf{x}_n)$ (pari a 0.47) risulta compreso tra il punto di massimo di $L(\theta; \mathbf{x}_n)$ (la stima di massima verosimiglianza di θ , pari a 0.3) e il punto di massimo di $\pi(\theta)$ (pari a 0.71). Si noti inoltre che la concentrazione della funzione di verosimiglianza e quella della distribuzione a priori sono molto simili (come vedremo, questo dipende dal fatto che $n \simeq \alpha + \beta$).

4.5 Esempio (modello di Poisson). Mostriamo che la famiglia delle densità gamma è coniugata al modello di Poisson. Consideriamo a tal fine una generica densità gamma⁴ di parametri α, β come distribuzione a priori per θ :

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \quad \theta \in \Omega = \mathbb{R}^+.$$

In questo caso $\mathbb{E}[\Theta] = \alpha/\beta$, $\text{Mo}[\Theta] = (\alpha - 1)/(\beta)$, $\mathbb{V}[\Theta] = \alpha/\beta^2$ (vedi Esempio 5.4). Dato un campione x_1, \dots, x_n da una $\text{Pois}(\theta)$, la funzione di verosimiglianza di θ è $L(\theta; \mathbf{x}_n) \propto \theta^{y_n} e^{-n\theta}$, dove $y_n = \sum_{i=1}^n x_i$. La distribuzione a posteriori si ottiene quindi dal prodotto di $\pi(\theta)$ e $L(\theta; \mathbf{x}_n)$ che risulta uguale a

$$\pi(\theta|\mathbf{x}_n) \propto \theta^{\alpha+y_n-1} e^{-(\beta+n)\theta} = \theta^{(\alpha+y_n)-1} e^{-(\beta+n)\theta}, \quad \theta \in \mathbb{R}^+$$

e che coincide con il nucleo della densità di una v.a. gamma di parametri $(\bar{\alpha}, \bar{\beta})$, con $\bar{\alpha} = \alpha + y_n$ e $\bar{\beta} = \beta + n$.

Al medesimo risultato si giunge, ovviamente, anche se si procede tenendo conto delle costanti (che nei passaggi precedenti abbiamo invece ignorato e avendo individuato la densità a posteriori gamma analizzando il nucleo della densità stessa). Abbiamo infatti che:

$$\begin{aligned} \pi(\theta|\mathbf{x}_n) &= \frac{\frac{1}{\prod_{i=1}^n x_i!} e^{-n\theta} \theta^{y_n} \times \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\theta} \theta^{\alpha-1}}{\frac{1}{\prod_{i=1}^n x_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \theta^{\alpha+y_n-1} e^{-(\beta+n)\theta} d\theta} \\ &= \frac{\bar{\beta}}{\Gamma(\bar{\alpha})} \theta^{\bar{\alpha}-1} e^{-\bar{\beta}\theta}, \quad \theta \in \mathbb{R}^+ \end{aligned}$$

dove l'ultima uguaglianza si ottiene ricordando che (vedi Esempio 5.4)

$$\int_0^\infty t^{\alpha-1} e^{-t} dt = \Gamma(\alpha) \quad \text{e} \quad \int_0^\infty t^{\alpha-1} e^{-\beta t} dt = \frac{\Gamma(\alpha)}{\beta^\alpha}.$$

⁴Usiamo per la densità gamma la parametrizzazione in cui in β è parametro **rate**.

4.6 Osservazione/i. Per disegnare i grafici di $\pi(\theta)$, $L(\theta; \mathbf{x}_n)$ e $\pi(\theta|\mathbf{x}_n)$ nell'intervallo $[0, A]$, si può ricorrere alle opportune funzioni R. Ponendo $\alpha = \text{alpha}$, $\beta = \text{beta}$, $y_n = \text{yn}$, $\bar{\alpha} = \text{alpha.p}$ e $\bar{\beta} = \text{beta.p}$ e osservando che $L(\theta; \mathbf{x}_n)$ risulta proporzionale a una densità gamma di parametri ($y_n + 1, \text{rate} = n$) si ha:

$\pi(\theta)$	<code>curve(dgamma(x, shape=alpha, rate=beta), from=0, to=A)</code>
$L(\theta; \mathbf{x}_n)$	<code>curve(dgamma(x, shape=yn+1, rate=n), add=T, lty=2)</code>
$\pi(\theta \mathbf{x}_n)$	<code>curve(dgamma(x, shape=alpha.p, rate=beta.p), add=T, lty=3)</code>

4.7 Esempio(modello geometrico). Sia X_1, \dots, X_n un campione casuale con $X_i|\theta \sim \text{Geom}(\theta)$. Utilizziamo la formulazione del modello per la quale

$$f(x|\theta) = \theta(1 - \theta)^{x-1}, \quad x = 1, 2, 3, \dots, \quad \theta \in [0, 1].$$

In questo caso $L(\theta; \mathbf{x}_n) = \theta^n(1 - \theta)^{y_n - n}$, con $y_n = \sum_{i=1}^n x_i$. È semplice verificare che il modello Beta è coniugato a quello geometrico. Si ha infatti che

$$\begin{aligned} \pi(\theta|\mathbf{x}_n) &\propto \theta^n(1 - \theta)^{y_n - n} \times \theta^{\alpha-1}(1 - \theta)^{\beta-1} \\ &\propto \theta^{\bar{\alpha}-1}(1 - \theta)^{\bar{\beta}-1} \quad \theta \in [0, 1] \end{aligned}$$

con $\bar{\alpha} = \alpha + n$ e $\bar{\beta} = \beta + y_n - n$. Abbiamo quindi che $\Theta|\mathbf{x}_n \sim \text{Beta}(\bar{\alpha}, \bar{\beta})$. I grafici di si ottengono ricorrendo alla funzione `dbeta()` di R.

4.8 Esempio (modello esponenziale negativo). Sia X_1, \dots, X_n un campione casuale con $X_i|\theta \sim \text{EN}(\theta) = \text{Ga}(1, \text{rate} = \theta)$, $\theta > 0$. Consideriamo per Θ la densità $\text{Ga}(\theta|\alpha, \text{rate} = \beta)$:

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \quad \theta \in \mathbb{R}^+, \quad \alpha, \beta > 0.$$

Per un campione osservato \mathbf{x}_n , la funzione di verosimiglianza di θ risulta essere $L(\theta|\mathbf{x}_n) = \theta^n e^{-y_n\theta}$ con $y_n = \sum_{i=1}^n x_i$. Abbiamo quindi che

$$\pi(\theta|\mathbf{x}_n) \propto \theta^{\bar{\alpha}-1} e^{-\bar{\beta}\theta}, \quad \theta \in \mathbb{R}^+$$

con $\bar{\alpha} = \alpha + n$ e $\bar{\beta} = \beta + y_n$, e quindi possiamo dire che $\Theta|\mathbf{x}_n \sim \text{Gamma}(\bar{\alpha}, \text{rate} = \bar{\beta})$.

4.9 Osservazione/i. Per disegnare i grafici di $\pi(\theta)$, $L(\theta; \mathbf{x}_n)$ e $\pi(\theta|\mathbf{x}_n)$ nell'intervallo $[0, A]$, si può ricorrere alle opportune funzioni R. Ponendo $\alpha = \text{alpha}$, $\beta = \text{beta}$, $y_n = \text{yn}$, $\bar{\alpha} = \text{alpha.p}$ e $\bar{\beta} = \text{beta.p}$, e osservando che la funzione di verosimiglianza

è proporzionale al nucleo di una densità gamma di parametri $(n + 1, \text{rate} = y_n)$, si ha:

$\pi(\theta)$	<code>curve(dgamma(x, shape=alpha, rate=beta), from=0, to=A)</code>
$L(\theta; \mathbf{x}_n)$	<code>curve(dgamma(x, shape=n+1, rate=yn), add=T, lty=2)</code>
$\pi(\theta \mathbf{x}_n)$	<code>curve(dgamma(x, shape1=alpha.p, rate=beta.p), add=T, lty=3)</code>

4.10 Esempio (modello esponenziale). Sia X_1, \dots, X_n un campione casuale con $X_i|\theta \sim \text{Esp}(\theta) = \text{Ga}(1, \text{scale} = \theta)$, $x \geq 0$, $\theta > 0$. Consideriamo per la v.a. Θ la densità *gamma inversa* di parametri α e β :

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\theta^{\alpha+1}} \exp\left\{-\frac{\beta}{\theta}\right\}, \quad \theta > 0, \quad \alpha, \beta > 0.$$

Indichiamo che la v.a. Θ ha distribuzione gamma-inversa di parametri (α, β) con la notazione $\Theta \sim \text{InvGa}(\alpha, \beta)$. Per questa v.a.⁵ sappiamo che $\mathbb{E}(\Theta) = \beta/(\alpha - 1)$, $\mathbb{M}_0[\Theta] = \beta/(\alpha + 1)$, $\mathbb{V}[\Theta] = \beta^2/[(\alpha - 1)^2(\alpha - 2)]$ (vedi Esempio 5.7). In questo caso abbiamo che $L(\theta; \mathbf{x}_n) = \theta^{-n} \exp\{-y_n/\theta\}$, dove $y_n = \sum_{i=1}^n x_i$. Pertanto

$$\pi(\theta|\mathbf{x}_n) \propto \frac{1}{\theta^{\bar{\alpha}} + 1} e^{-\bar{\beta}/\theta},$$

ovvero possiamo dire che $\Theta|\mathbf{x}_n \sim \text{InvGa}(\bar{\alpha}, \bar{\beta})$ con $\bar{\alpha} = \alpha + n$ e $\bar{\beta} = \beta + y_n$. I grafici di $\pi(\theta)$, $L(\theta; \mathbf{x}_n)$ e $\pi(\theta|\mathbf{x}_n)$ possono essere tracciati usando la funzione `dinvgamma()` del pacchetto `invgamma` di R.

4.11 Esempio (modello normale – varianza nota). Sia X_1, \dots, X_n un campione casuale con $X_i|\theta \sim \text{N}(\theta, \sigma^2)$, $\theta \in \mathbb{R}$ (σ^2 noto). Supponiamo che l'informazione disponibile sul parametro incognito θ sia rappresentata da una densità normale di parametri μ_0 (che rappresenta anche valore atteso, moda e mediana di Θ) e τ_0^2 (varianza di Θ). Per comodità analitiche e interpretative, poniamo $\tau_0^2 = \sigma^2/n_0$, $n_0 > 0$. Abbiamo così che

$$\pi(\theta) = \left(\frac{n_0}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{n_0}{2} \left(\frac{\theta - \mu_0}{\sigma}\right)^2\right\}.$$

⁵Si osservi che, se una v.a. $X \sim \text{Ga}(\alpha, \text{rate} = \beta)$, allora $Y = X^{-1} \sim \text{InvGa}(\alpha, \text{rate} = \beta)$. Ponendo infatti $y = g(x) = 1/x$, si ottiene che $g^{-1}(y) = 1/y$ e che $|\frac{d}{dy} g^{-1}(y)| = 1/y^2$. Pertanto la funzione di densità f_Y di Y si ottiene dalla densità f_X :

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{y^{\alpha-1}} \exp\left\{-\frac{\beta}{y}\right\} \frac{1}{y^2} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{y^{\alpha+1}} \exp\left\{-\frac{\beta}{y}\right\}, \quad y \geq 0. \end{aligned}$$

In corrispondenza di un campione di dati osservato x_1, \dots, x_n , la funzione di verosimiglianza di θ risulta essere

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

Pertanto

$$\pi(\theta | \mathbf{x}_n) \propto \underbrace{\exp \left\{ -\frac{n}{2} \left(\frac{\theta - \bar{x}_n}{\sigma} \right)^2 \right\}}_{L(\theta; \mathbf{x}_n)} \times \underbrace{\exp \left\{ -\frac{n_0}{2} \left(\frac{\theta - \mu_0}{\sigma} \right)^2 \right\}}_{\pi(\theta)}.$$

Osservando che⁶

$$n(\theta - \bar{x}_n)^2 + n_0(\theta - \mu_0)^2 = (n_0 + n) \left(\theta - \frac{n_0\mu_0 + n\bar{x}_n}{n_0 + n} \right)^2 + \left(\frac{1}{n_0} + \frac{1}{n} \right)^{-1} (\bar{x}_n - \mu_0)^2$$

e che il secondo addendo della precedente espressione non dipende da θ , si ha che

$$\pi(\theta | \mathbf{x}_n) \propto \exp \left\{ -\frac{n_0 + n}{2\sigma^2} \left(\theta - \frac{n_0\mu_0 + n\bar{x}_n}{n_0 + n} \right)^2 \right\}, \quad \theta \in \mathbb{R}$$

e quindi che

$$\Theta | \mathbf{x}_n \sim N(\bar{\mu}, \bar{\sigma}^2),$$

dove

$$\bar{\mu} = \mathbb{E}[\theta | \mathbf{x}_n] = \frac{n_0\mu_0 + n\bar{x}_n}{n_0 + n} \quad \text{e} \quad \bar{\sigma}^2 = \mathbb{V}[\theta | \mathbf{x}_n] = \frac{\sigma^2}{n_0 + n}. \quad (4.2)$$

Anche in questo esempio, come nei precedenti, distribuzione iniziale e finale appartengono allo stesso modello parametrico (gaussiano) e la distribuzione a posteriori si ottiene con un semplice aggiornamento dei parametri della distribuzione a priori, che tiene conto dei dati osservati.

4.12 Osservazione/i. Per disegnare i grafici della distribuzione a priori, della funzione di verosimiglianza e della distribuzione a posteriori nell'intervallo $[A, B]$, si può ricorrere alle opportune funzioni R. Ponendo $\mu_0 = \text{mu.0}$, $\sigma = \text{sigma}$, $\bar{x}_n = \text{xmed}$, $\bar{\mu} = \text{mu.p}$, $\bar{\sigma} = \text{sigma.p}$ e $n_0 = \text{n.0}$ e osservando che $L(\theta; \mathbf{x}_n)$ è proporzionale a una densità normale di parametri $(\bar{x}_n, \sigma^2/n)$, si ha:

$\pi(\theta)$	curve(dnorm(x, mean=mu.0, sd=sigma/n.0), from=A, to=B)
$L(\theta; \mathbf{x}_n)$	curve(dnorm(x, mean=xmed, sd=sigma/n), add=T, lty=2)
$\pi(\theta \mathbf{x}_n)$	curve(dnorm(x, mean=mu.p, sd=sigma.p), add=T, lty=3)

⁶Dati $x, a, b \in \mathbb{R}$ e $A, B \in \mathbb{R}^+$, vale in generale la seguente uguaglianza:

$$A(x - a)^2 + B(x - b)^2 = (A + B)(x - c)^2 + \frac{AB}{A + B}(a - b)^2,$$

dove $c = \frac{Ab + Bb}{A + B}$.

4.13 Esempio (modello normale – valore atteso noto). Sia X_1, \dots, X_n un campione casuale con $X_i|\theta \sim N(\mu_0, \theta)$, $\theta \in \mathbb{R}^+$ (μ_0 noto). Consideriamo per la v.a. Θ la densità *gamma-inversa* di parametri α e β (vedi Esempio 4.10). Poichè in questo caso la funzione di verosimiglianza di θ associata a un campione osservato di dimensione n risulta essere

$$L(\theta; \mathbf{x}_n) = \left(\frac{1}{2\pi\theta}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\theta} \sum_{i=1}^n (x_i - \mu_0)^2\right\} = \left(\frac{1}{2\pi\theta}\right)^{\frac{n}{2}} \exp\left\{-\frac{nS_0^2}{2\theta}\right\},$$

dove $S_0^2 = n^{-1} \sum_{i=1}^n (x_i - \mu_0)^2$, si ha che la densità a posteriori di Θ è proporzionale a

$$\frac{1}{\theta^{(\alpha+\frac{n}{2})+1}} \exp\left\{-\frac{1}{\theta} \left(\beta + \frac{n}{2}S_0^2\right)\right\}$$

e quindi che

$$\pi(\theta|\mathbf{x}_n) = \frac{\bar{\beta}^{\bar{\alpha}}}{\Gamma(\bar{\alpha}) \theta^{\bar{\alpha}+1}} \exp\left\{-\frac{\bar{\beta}}{\theta}\right\}, \quad \theta \in \mathbb{R}^+$$

con $\bar{\alpha} = \alpha + n/2$ e $\bar{\beta} = \beta + nS_0^2/2$. Pertanto $\Theta|\mathbf{x}_n \sim \text{InvGa}(\bar{\alpha}, \bar{\beta})$. I grafici di $\pi(\theta)$, $L(\theta; \mathbf{x}_n)$ e $\pi(\theta|\mathbf{x}_n)$ possono essere tracciati usando la funzione `dinvgamma()` del pacchetto `invgamma` di R.

4.14 Esempio (modello uniforme). Sia X_1, \dots, X_n un campione casuale con $X_i|\theta \sim \text{Unif}[0, \theta]$ i.i.d., $\theta > 0$. Supponiamo che a priori Θ abbia distribuzione di Pareto di parametri (α, β) (ovvero: $\Theta \sim \text{Pa}(\alpha, \beta)$), con funzione di densità

$$\pi(\theta) = \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}} \mathbb{I}_{[\beta, +\infty)}(\theta), \quad \alpha, \beta > 0.$$

Dalla precedente espressione si evince che, essendo $\pi(\theta)$ decrescente in $[\beta, +\infty)$, la moda della densità a priori è $\text{Mo}[\Theta] = \beta$. Inoltre abbiamo che $\mathbb{E}[\Theta] = \alpha\beta/(\alpha - 1)$ e che $\text{Me}[\Theta|\mathbf{x}_n] = \beta 2^{\frac{1}{\alpha}}$ (vedi Esempio 5.11).

Possiamo verificare che il modello di Pareto è coniugato a quello uniforme. Infatti, ricordando che

$$L(\theta; \mathbf{x}_n) = \frac{1}{\theta^n} \cdot \mathbb{I}_{[x_{(n)}, +\infty)}(\theta),$$

abbiamo che

$$\begin{aligned} \pi(\theta|\mathbf{x}_n) &\propto L(\theta; \mathbf{x}_n) \times \pi(\theta) \\ &\propto \frac{1}{\theta^n} \cdot \mathbb{I}_{[x_{(n)}, +\infty)}(\theta) \times \frac{1}{\theta^{\alpha+1}} \cdot \mathbb{I}_{[\beta, +\infty)}(\theta) \\ &\propto \frac{1}{\theta^{\bar{\alpha}+1}} \cdot \mathbb{I}_{[\bar{\beta}, +\infty)}(\theta), \end{aligned}$$

con $\bar{\alpha} = \alpha + n$ e $\bar{\beta} = \max\{\beta, x_{(n)}\}$. Abbiamo quindi che $\Theta|\mathbf{x}_n \sim \text{Pa}(\bar{\alpha}, \bar{\beta})$. I grafici di $\pi(\theta)$, $L(\theta; \mathbf{x}_n)$ e $\pi(\theta|\mathbf{x}_n)$ possono essere tracciati usando la funzione `dpareto()` del pacchetto `EnvStats` di R.

4.15 Esempio (modello bernoulliano, parametrizzazione odds). Sia X_1, \dots, X_n un campione casuale con $X_i|\theta \sim \text{Ber}[\theta]$ i.i.d., $\theta \in (0, 1)$. Scriviamo la funzione di massa di probabilità nella parametrizzazione *odds*, ovvero in funzione del parametro

$$\psi = \frac{\theta}{1 - \theta}.$$

Poichè $\theta = \psi/(1 + \psi)$, si ottiene facilmente che

$$f(x|\psi) = \left(\frac{\psi}{1 + \psi}\right)^x \left(\frac{1}{1 + \psi}\right)^{1-x} = \frac{\psi^x}{(1 + \psi)}, \quad \psi > 0.$$

La funzione di verosimiglianza associata a un campione causale osservato è quindi

$$L(\psi|\mathbf{x}_n) = \frac{\psi^{y_n}}{(1 + \psi)^n}, \quad \psi \in \mathbb{R}^+$$

dove $y_n = \sum_{i=1}^n x_i$. Per il parametro ψ possiamo ora utilizzare una densità (detta *Beta del secondo tipo*)

$$\pi(\psi) = \frac{1}{B(\alpha, \beta)} \frac{\psi^{\alpha-1}}{(1 + \psi)^{\alpha+\beta}}, \quad \psi > 0, \quad \alpha, \beta > 0,$$

dove $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$. È possibile mostrare che $\mathbb{E}[\Theta] = \alpha/(\beta - 1)$, $\text{Mo}[\Theta] = (\alpha - 1)/(\beta + 1)$, $\text{V}[\Theta] = \alpha(\alpha + \beta - 1)/[(\beta - 2)(\beta - 1)^2]$. La densità $\pi(\psi)$ è coniugata al modello in esame. Infatti la densità a posteriori di ψ risulta essere

$$\pi(\psi|\mathbf{x}_n) = \frac{1}{B(\bar{\alpha}, \bar{\beta})} \frac{\psi^{\bar{\alpha}-1}}{(1 + \psi)^{\bar{\alpha}+\bar{\beta}}}, \quad \psi > 0,$$

ovvero una densità Beta-del secondo tipo di parametri $(\bar{\alpha}, \bar{\beta})$, con $\bar{\alpha} = \alpha + y_n$ e $\bar{\beta} = \beta + n - y_n$.

4.3.1 Scelta dei valori degli iperparametri

Negli esempi concreti, la scelta del valore degli iperparametri deve essere effettuata in modo tale che la corrispondente distribuzione $\pi(\cdot)$ per Θ rappresenti l'informazione pre-sperimentale di cui si vuole tenere conto. Spesso questa informazione giunge in termini di alcuni valori sintetici (posizione, scala, quantili, asimmetria...) che possono essere usati per impostare un sistema con tante equazioni quanti sono gli iperparametri della distribuzione a priori (incognite del sistema). La soluzione di questo sistema (che assumiamo esistere ed essere unica) ci fornisce i valori degli iperparametri compatibili con le caratteristiche strutturali che abbiamo su $\pi(\cdot)$. Supponiamo ad esempio che la distribuzione iniziale dipenda da due soli iperparametri, (a, b) e che l'informazione iniziale su Θ giunga in termini di un valore sintetico puntuale della *posizione* della distribuzione a priori (m) e di un valore che indica la variabilità della stessa (v). Per trovare a e b possiamo, ad esempio, richiedere

che valore atteso e varianza di Θ (che dipendono da a e da b) siano rispettivamente uguali a m e v :

$$\begin{cases} \mathbb{E}[\Theta] = m \\ \mathbb{V}[\Theta] = v \end{cases} . \quad (4.3)$$

Si osservi che in genere il valore di v non si fissa direttamente ma si ricava richiedendo che un intervallo di valori per Θ ritenuto a priori plausibile abbia una probabilità elevata. Un modo semplice per procedere, che di fatto assume una certa simmetria di $\pi(\cdot)$, consiste nel richiedere che l'intervallo

$$C = [\ell, u] = [m - 2\sqrt{v}, m + 2\sqrt{v}]$$

contenga il 95% della probabilità assegnata da $\pi(\cdot)$. Pertanto, se fissiamo due valori numerici per ℓ e u , possiamo determinare il valore di v tale che, ad esempio, $m + 2\sqrt{v} = u$, ovvero $v = \frac{(u-m)^2}{4}$.

4.16 Esempio(modello beta). Sia $\Theta \sim \text{Beta}(a, b)$. Vogliamo trovare i valori di a e b tali che siano soluzioni del sistema (4.3), che in questo caso diventa:

$$\begin{cases} \frac{a}{a+b} = m \\ \frac{ab}{(a+b+1)(a+b)^2} = v \end{cases} .$$

Con alcuni passaggi si verifica che

$$\begin{cases} a = \frac{m[m(1-m)-v]}{v} \\ b = \frac{(1-m)[m(1-m)-v]}{v} \end{cases} .$$

Se, ad esempio, vogliamo che $\pi(\cdot)$: (i) abbia punti di massimo approssimativamente pari a $m = 0.2$; (ii) assegni probabilità pari a 0.95 all'insieme $[0.2, 0.6]$; possiamo porre $m = 0.2$ e $v^2 = (0.6 - 0.2)^2/4 = 0.01$. $v = 0.01$. Risolvendo il sistema otteniamo $a = 9.2$ e $b = 13.8$.

4.17 Esempio (modello gamma). Sia $\Theta \sim \text{Ga}(a, \text{rate} = b)$. Il sistema (4.3) diventa

$$\begin{cases} \frac{a}{a} = m \\ \frac{a}{b^2} = v \end{cases} , \quad (4.4)$$

da cui si ottiene subito $a = m^2/v$ e $b = m/v$.

4.4 Distribuzioni a priori mistura

È possibile che all'interno delle classi coniugate non vi siano distribuzioni adatte a rappresentare l'informazione iniziale. Ad esempio, la famiglia normale non contiene densità bimodali (o, in generale, con più mode). Possiamo però generare densità multimodali attraverso *mixture* di densità unimodali.

4.18 Definizione (distribuzione a priori mistura). Siano $\pi_1(\theta), \dots, \pi_r(\theta)$ r distribuzioni iniziali per Θ (densità o funzioni di massa di probabilità) e $\kappa_1, \dots, \kappa_r$ r valori tali che $\kappa_j \in [0, 1]$, $j = 1, \dots, r$ e $\sum_{j=1}^r \kappa_j = 1$. La distribuzione $\pi(\theta)$ definita ponendo

$$\pi(\theta) = \sum_{j=1}^r \kappa_j \pi_j(\theta) \quad (4.5)$$

si chiama *distribuzione a priori mistura* di Θ .

È semplice verificare che, in effetti, $\pi(\theta)$ è una funzione di massa/densità di probabilità su Ω . Se, per semplicità, consideriamo il caso assolutamente continuo (le $\pi_j(\theta)$ sono quindi funzioni di densità), abbiamo infatti che

$$\int_{\Omega} \pi(\theta) d\theta = \sum_{j=1}^r \kappa_j \int_{\Omega} \pi_j(\theta) d\theta = \sum_{j=1}^r \kappa_j = 1.$$

Utilizzando la distribuzione mistura (4.5), per il teorema di Bayes otteniamo che

$$\begin{aligned} \pi(\theta|\mathbf{x}_n) &= \frac{f_n(\mathbf{x}_n|\theta)\pi(\theta)}{m(\mathbf{x}_n)} \\ &= \frac{\sum_{j=1}^r \kappa_j \pi_j(\theta) f_n(\mathbf{x}_n|\theta)}{m(\mathbf{x}_n)}. \end{aligned}$$

Osservando che, in virtù del teorema di Bayes, per ogni $j = 1, \dots, r$, vale l'uguaglianza $\pi_j(\theta) f_n(\mathbf{x}_n|\theta) = m_j(\mathbf{x}_n) \pi_j(\theta|\mathbf{x}_n)$, e che $m(\mathbf{x}_n) = \sum_{h=1}^r \kappa_h m_h(\mathbf{x}_n)$, dove $m_h(\mathbf{x}_n) = \int_{\Omega} f_n(\mathbf{x}_n|\theta) \pi_h(\theta) d\theta$, possiamo scrivere

$$\pi(\theta|\mathbf{x}_n) = \frac{\sum_{j=1}^r \kappa_j m_j(\mathbf{x}_n) \pi_j(\theta|\mathbf{x}_n)}{m(\mathbf{x}_n)} = \sum_{j=1}^r \kappa_j^* \pi_j(\theta|\mathbf{x}_n), \quad (4.6)$$

dove

$$\kappa_j^* = \frac{\kappa_j m_j(\mathbf{x}_n)}{m(\mathbf{x}_n)} = \frac{\kappa_j m_j(\mathbf{x}_n)}{\sum_{h=1}^r \kappa_h m_h(\mathbf{x}_n)} \quad (4.7)$$

Poichè $\kappa_j^* \in [0, 1]$, $j = 1, \dots, r$ e $\sum_{h=1}^r \kappa_h^* = 1$, anche $\pi(\theta|\mathbf{x}_n)$ ha la struttura di una mistura - delle r distribuzioni a posteriori $\pi_j(\theta|\mathbf{x}_n)$ -, con pesi κ_j^* dipendenti dai dati osservati.

4.19 Esempio (modello bernoulliano). Consideriamo il modello bernoulliano, un campione \mathbf{x}_n di dimensione n e la mistura di densità beta di parametri (α_j, β_j) :

$$\pi(\theta) = \sum_{j=1}^r \kappa_j \pi_j(\theta), \quad \pi_j(\theta) = \text{Beta}(\theta|\alpha_j, \beta_j), \quad j = 1, \dots, r.$$

Con queste ipotesi si ottiene che

$$\pi(\theta|\mathbf{x}_n) = \sum_{j=1}^r \kappa_j^* \cdot \pi_j(\theta|\mathbf{x}_n), \quad \pi_j(\theta|\mathbf{x}_n) = \text{Beta}(\theta|\bar{\alpha}_j, \bar{\beta}_j),$$

dove $\bar{\alpha}_j = \alpha + \sum_{i=1}^n x_i$ e $\bar{\beta}_j = \beta + n - \sum_{i=1}^n x_i$ e dove le espressioni κ_j^* si ottengono dalla (4.7) osservando che, per ogni $j = 1, \dots, r$, si ha

$$m_j(\mathbf{x}_n) = \frac{B(\bar{\alpha}_j, \bar{\beta}_j)}{B(\alpha_j, \beta_j)} = \frac{\Gamma(\bar{\alpha}_j)\Gamma(\bar{\beta}_j)}{\Gamma(\bar{\alpha}_j + \bar{\beta}_j)} \times \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)}.$$

Questo risultato si giustifica osservando che

$$m_j(\mathbf{x}_n) = \int_0^1 f_n(\mathbf{x}_n|\theta)\pi_j(\theta)d\theta = \frac{1}{B(\alpha, \beta)} \int_0^1 \theta^{\bar{\alpha}_j-1} \times (1-\theta)^{\bar{\beta}_j-1} d\theta$$

e ricordando che, in generale, $\int_0^1 y^{\alpha-1}(1-y)^{\beta-1}dy = B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

4.20 Esempio (modello normale). Consideriamo il modello normale $N(\theta, \sigma^2)$ (varianza nota), un campione \mathbf{x}_n di dimensione n e la mistura di densità normali di parametri $(\mu_j, \sigma^2/n_j)$, $j = 1, \dots, r$:

$$\pi(\theta) = \sum_{j=1}^r \kappa_j \pi_j(\theta), \quad \pi_j(\theta) = N\left(\theta \middle| \mu_j, \frac{\sigma^2}{n_j}\right), \quad j = 1, \dots, r.$$

Per comodità, possiamo formulare il problema in funzione della statistica sufficiente, \bar{x}_n , sostituendo nelle formule (4.6) e (4.7) \bar{x}_n a \mathbf{x}_n . Abbiamo quindi che

$$\pi(\theta|\bar{x}_n) = \sum_{j=1}^r \kappa_j^* \cdot \pi_j(\theta|\bar{x}_n), \quad \pi_j(\theta|\bar{x}_n) = N(\theta|\bar{\mu}_j, \bar{\sigma}_j^2),$$

dove $\bar{\mu}_j = (n_j\mu_j + n\bar{x}_n)/(n_j + n)$ e $\bar{\sigma}_j^2 = \sigma^2/(n_j + n)$. Per ottenere le espressioni dei pesi a posteriori κ_j^* attraverso la (4.7) si osservi che, per un risultato che verrà illustrato nel paragrafo 8.3, si ha che la distribuzione predittiva di \bar{X}_n è $N(\mu_j, \sigma_j^2)$, con $\sigma_j^2 = \sigma^2(n^{-1} + n_j^{-1})$. Si ha quindi che

$$m_j(\bar{x}_n) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{1}{2\sigma_j}(\bar{x}_n - \mu_j)^2\right\} = \frac{1}{\sigma_j} \phi\left(\frac{\bar{x}_n - \mu_j}{\sigma_j}\right),$$

dove $\phi(\cdot)$ è la funzione di densità della v.a. $N(0, 1)$.

4.5 Densità a posteriori di una funzione del parametro

In molti casi l'interesse inferenziale verte su funzioni del parametro del modello piuttosto che sul parametro stesso. Sia, in generale

$$\psi = g(\theta)$$

una funzione scalare di interesse del parametro (scalare) del modello in esame. Per semplicità consideriamo il caso in cui Θ sia una v.a. assolutamente continua. In

questo caso, se $g(\cdot)$ è una funzione invertibile, per la regola di trasformazione di variabili unidimensionali, la funzione di densità $\pi_{\Psi}(\psi)$ di $\Psi = g(\Theta)$ si ottiene dalla densità della v.a. Θ :

$$\pi_{\Psi}(\psi) = \pi(g^{-1}(\psi)) \left| \frac{d}{d\psi} g^{-1}(\psi) \right|,$$

dove $\left| \frac{d}{d\psi} g^{-1}(\psi) \right|$ indica lo jacobiano della trasformazione. Nel seguito utilizzeremo questa regola per ottenere la densità a posteriori di Ψ a partire da $\pi(\theta|\mathbf{x}_n)$.

4.21 Esempio (modello bernoulliano). Consideriamo un campione casuale di dimensione n da v.a. bernoulliana di parametro θ e $\pi(\theta) = \text{Beta}(\theta|\alpha, \beta)$, $\alpha, \beta > 0$. Sappiamo allora che

$$\pi(\theta|\mathbf{x}_n) = \text{Beta}(\theta|\bar{\alpha}, \bar{\beta}), \quad \bar{\alpha} = \alpha + \sum_{i=1}^n x_i, \bar{\beta} = \beta + n - \sum_{i=1}^n x_i.$$

Possiamo ora determinare la funzione di densità a posteriori della v.a. log-odds:

$$\Psi = \ln \frac{\Theta}{1 - \Theta}.$$

È semplice verificare che i valori ψ di Ψ appartengono a \mathbb{R}^+ e che

$$\psi = g(\theta) = \ln \frac{\theta}{1 - \theta} \implies g^{-1}(\psi) = \frac{e^{\psi}}{1 + e^{\psi}} \implies \frac{d}{d\psi} g^{-1}(\psi) = \frac{e^{\psi}}{(1 + e^{\psi})^2}.$$

Si ottiene quindi

$$\begin{aligned} \pi_{\Psi}(\psi|\mathbf{x}_n) &= \frac{1}{B(\bar{\alpha}, \bar{\beta})} \left(\frac{e^{\psi}}{1 + e^{\psi}} \right)^{\bar{\alpha}-1} \left(\frac{1}{1 + e^{\psi}} \right)^{\bar{\beta}-1} \frac{e^{\psi}}{(1 + e^{\psi})^2} \\ &= \frac{1}{B(\bar{\alpha}, \bar{\beta})} \frac{e^{\bar{\alpha}\psi}}{(1 + e^{\psi})^{\bar{\alpha}+\bar{\beta}}}, \quad \psi > 0. \end{aligned}$$

4.22 Esempio (modello di Poisson). Consideriamo un campione casuale di dimensione n da v.a. di Poisson di parametro θ e $\pi(\theta) = \text{Gamma}(\theta|\alpha, \beta)$, $\alpha, \beta > 0$. Dall'Esempio 4.5 sappiamo allora che

$$\pi(\theta) = \text{Gamma}(\theta|\bar{\alpha}, \bar{\beta}), \quad \bar{\alpha} = \alpha + \sum_{i=1}^n x_i, \bar{\beta} = \beta + n.$$

Possiamo ora determinare la densità a posteriori della funzione parametrica

$$\Psi = g(\Theta) = e^{-\Theta}$$

che rappresenta probabilità che la v.a. X_i (condizionatamente a θ) assuma il valore zero. Abbiamo infatti che

$$\psi = g(\theta) = e^{-\theta} \implies g^{-1}(\psi) = -\ln \psi \implies \frac{d}{d\psi} g^{-1}(\psi) = -\frac{1}{\psi}$$

da cui discende che

$$\pi_{\Psi}(\psi|\mathbf{x}_n) = \frac{\bar{\beta}^{\bar{\alpha}}}{\Gamma(\bar{\alpha})\psi} (-\ln \psi)^{\bar{\alpha}-1} \exp\{\beta \ln \psi\}, \quad \psi \in [0, 1].$$

Il grafico della densità $\pi_{\Psi}(\psi|\mathbf{x}_n)$ può essere ottenuto scrivendo una apposita funzione **R**. È inoltre possibile effettuare inferenza su ψ ricorrendo a tecniche numeriche (metodo di Monte Carlo).

4.6 Esercizi

1. Sia X_1, \dots, X_n un campione casuale da una va. bernoulliana di parametro incognito θ e considerare per Θ una distribuzione a priori di tipo $\text{Beta}(\alpha, \beta)$.
 - (a) Determinare le espressioni della funzione di verosimiglianza e della stima di massima verosimiglianza per θ .
 - (b) Supponendo che $\alpha = 9.2$ and $\beta = 13.8$ e di avere osservato un campione $\sum_{i=1}^n x_i = 15$ e $n = 20$, tracciare nello stesso grafico $\pi(\theta)$, $L(\theta; \mathbf{x}_n)$ e $\pi(\theta|\mathbf{x}_n)$.
 - (c) Calcolare le probabilità a priori e a posteriori dei seguenti che $\Theta > 0.2$ e che $\Theta \in [0.2, 0.6]$.
 - (d) Calcolare i valori numerici di $\mathbb{E}[\Theta|\mathbf{x}_n]$, $\text{Mo}[\Theta|\mathbf{x}_n]$, $\text{Me}[\Theta|\mathbf{x}_n]$.
 - (e) Calcolare il valore dei quantili di livello 0.025 and 0.975 usando sia $p_i(\theta)$ che $\pi(\theta|\mathbf{x}_n)$.
 - (f) Ripetere l'esercizio assumendo ora che $\sum_{i=1}^n x_i = 30$ e $n = 40$.
 - (g) Ripetere l'esercizio assumendo ora che $\alpha = 1$ e $\beta = 1$.
2. Siano $X_1, \dots, X_n|\theta \sim \text{Ber}(\theta)$ i.i.d., $\theta \in [0, 1]$; $\Theta \sim \text{Beta}(\alpha, \beta)$, $\alpha, \beta > 0$; $\mathbf{x}_n = (x_1, \dots, x_n)$ un campione osservato. Determinare:
 - (a) la funzione di densità $\pi_\Psi(\psi)$, dove $\Psi = \frac{1}{\Theta}$;
 - (b) la distribuzione a posteriori di Θ e di Ψ ;
 - (c) la moda $\text{Mo}(\Psi|\mathbf{x}_n)$ della densità a posteriori di Ψ ;
 - (d) il valore atteso $\mathbb{E}[\Psi|\mathbf{x}_n]$ delle densità a posteriori di Ψ [sugg.: integrare rispetto a θ];
 - (e) i valori numerici delle due stime ottenute per $n = 5$, $\sum_{i=1}^n x_i = 3$, $\alpha = 2$, $\beta = 3$.
3. Siano $X_1, \dots, X_n|\theta \sim \text{Geom}(\theta)$ i.i.d., con valore atteso $\mathbb{E}_\theta[X] = \frac{1}{\theta}$, $\text{V}_\theta[X] = \frac{1-\theta}{\theta^2}$ e funzione di massa di probabilità

$$f_X(x|\theta) = \theta(1 - \theta)^{x-1}, \quad x = 1, 2, 3, \dots \quad \theta \in [0, 1].$$

- (a) Determinare la funzione di verosimiglianza e $\hat{\theta}_{mv}(\mathbf{x}_n)$, stima di massima verosimiglianza del parametro θ .
- (b) Verificare che la famiglia delle densità $\text{Beta}(\alpha, \beta)$, con $\alpha, \beta > 0$, è coniugata al modello in esame e determinare i parametri $(\bar{\alpha}, \bar{\beta})$ della distribuzione a posteriori di Θ , $\pi(\theta|\mathbf{x}_n)$.
- (c) Determinare di $\hat{\theta}_B(\mathbf{x}_n) = \mathbb{E}[\Theta|\mathbf{x}_n]$.
- (d) Determinare analiticamente l'espressione di $\text{Mo}[\Theta|\mathbf{x}_n]$, moda della distribuzione a posteriori di Θ .
- (e) Verificare che, per valori elevati di n , si ha che $\text{Mo}[\Theta|\mathbf{x}_n] \approx \hat{\theta}_{mv}(z)$.

- (f) Supponendo che il campione abbia dimensione $n = 5$, che $\alpha = \beta = 2$ e che $\sum_{i=1}^n x_i = 15$, calcolare il valore numerico di $\widehat{\theta}_B(\mathbf{x}_n)$, $\text{Mo}[\Theta|\mathbf{x}_n]$ e di $\widehat{\theta}_{mv}(\mathbf{x}_n)$.
4. Sia $\mathbf{X}_n = (X_1, \dots, X_n)$ un campione in cui le v.a X_i sono, condizionatamente a θ , i.i.d. con distribuzione $\text{EN}(\theta)$, ovvero con funzione di densità

$$f_{X_i}(x; \theta) = \theta e^{-\theta x}, \quad x \geq 0, \quad \theta > 0.$$

Si consideri inoltre per Θ la famiglia delle densità $\text{Gamma}(\alpha, \beta)$ coniugata al modello, $\alpha, \beta > 0$.

- (a) Determinare le distribuzioni a posteriori di Θ e della v.a. $\Lambda = \Theta^{-1}$.
- (b) Considerare due costanti reali a, b e calcolare $\mathbb{E}[a\Lambda + b|\mathbf{x}_n]$.
- (c) Determinare l'espressione di $\text{Mo}[\Theta|\mathbf{x}_n]$, moda della distribuzione a posteriori di Θ .
- (d) Determinare l'espressione di $\mathbb{P}[\Theta > \delta|\mathbf{x}_n]$ in funzione di $F(\cdot|\mathbf{x}_n)$, funzione di ripartizione a posteriori di Θ .
5. Sia $X_1, \dots, X_n|\theta \sim \text{N}(\mu_0, \theta)$ (μ_0 noto) e $\Theta \sim \text{InvGa}(\alpha, \beta)$. Determinare quanto richiesto di seguito.
- (a) Densità a posteriori di Θ .
- (b) Moda della densità a posteriori di Θ .
- (c) Espressione della funzione di densità, il valore atteso e la varianza di $\Psi = \Theta^{-1}$.
6. Sia $X_1, \dots, X_n|\theta$ i.i.d. con

$$f_X(x|\theta) = \theta x^{\theta-1}, \quad x \in (0, 1), \quad \theta > 0.$$

Considerare per Θ una distribuzione a priori $\text{Ga}(\alpha, \beta)$ ($\alpha > 0, \beta > 0$). Determinare la distribuzione a posteriori di Θ , il suo valore atteso e la sua varianza.

7. Sia X_1, \dots, X_n un campione casuale proveniente da una popolazione di Rayleigh di parametro θ , con funzione di densità

$$f_X(x|\theta) = \frac{x}{\theta^2} \exp\left\{-\frac{x^2}{2\theta^2}\right\}, \quad x \geq 0, \quad \theta > 0.$$

Sia inoltre

$$\pi(\theta) = \frac{\alpha^\beta}{2^{\beta-1}\Gamma(\beta)} \frac{1}{\theta^{2\beta+1}} \exp\left\{-\frac{\alpha}{2\theta^2}\right\}, \quad \alpha, \beta > 0, \quad \theta > 0.$$

- (a) Determinare $\pi(\theta|\mathbf{x}_n)$ e verificare che $\pi(\theta)$ è coniugata al modello considerato.

- (b) Individuare le espressioni di $\bar{\alpha}$ e $\bar{\beta}$ (parametri della densità a posteriori di Θ).
- (c) Tracciare in una stessa figura i grafici di $\pi(\theta)$, $L(\theta; \mathbf{x}_n)$ e $\pi(\theta|\mathbf{x}_n)$ assumendo che $n = 10$, $\sum_{i=1}^n x_i^2 = 5$, $\alpha = 2$, $\beta = 3$.
8. Mostrare con il seguente esempio che i concetti di indipendenza e *indipendenza condizionata* di variabili aleatorie non coincidono. Assumere che: $X_i|\theta \sim N(\theta, \sigma^2)$, $i = 1, \dots, n$ (i.i.d.) e che $\Theta \sim N(\mu_0, \frac{\sigma^2}{n_0})$.
- (a) Determinare $m_X(x_i) = \int_{\mathbb{R}} f_X(x|\theta)\pi(\theta)d\theta$ e $m_{\mathbf{X}_n}(\mathbf{x}_n) = \int_{\mathbb{R}} f_n(\mathbf{x}_n|\theta)\pi(\theta)d\theta$.
- (b) Mostrare che $m_{\mathbf{X}_n}(\mathbf{x}_n) \neq \prod_{i=1}^n m_X(x_i)$.

Capitolo 5

Metodi per inferenze ipotetiche

5.1 Introduzione

In ottica bayesiana, il massimo contenuto informativo disponibile sul parametro incognito di un modello parametrico è fornito dall'intera distribuzione a posteriori. Tuttavia anche in questa impostazione inferenziale, come in quella frequentista, esistono procedure inferenziali per stima puntuale, stima mediante regioni e la scelta tra ipotesi. Tali procedure si concretizzano in opportune sintesi della distribuzione (funzione di massa o densità di probabilità) a posteriori del parametro, ovvero in opportuni funzionali di $\pi(\theta|\mathbf{x}_n)$. Nei prossimi paragrafi consideriamo i principali metodi per i tre maggiori problemi inferenziali di tipo ipotetico, con riferimento ai principali esempi notevoli trattati con l'analisi coniugata.

5.2 Stima puntuale

I metodi più semplici e comuni per sintetizzare la distribuzione a posteriori in una stima puntuale di θ sono: (a) valore atteso; (b) moda; (c) mediana della distribuzione a posteriori di Θ . Tali procedure hanno una giustificazione formale nell'ambito della teoria statistica delle decisioni. In quest'ottica, una stima del parametro incognito θ , che indichiamo con $\hat{\theta}$, viene valutata attraverso una funzione, detta funzione di perdita, che valuta quanto si "perde" stimando la quantità θ con il numero $\hat{\theta}$. Formalmente la funzione di perdita è quindi una funzione $\mathbb{L}(\theta, \hat{\theta})$ a valori reali (che, per semplicità assumiamo maggiori o uguali a zero) tale che, per ogni coppia $(\theta, \hat{\theta})$, fornisce la valutazione della perdita associata alla stima di θ effettuata con $\hat{\theta}$: maggiore il valore assunto dalla perdita, peggiore la qualità della stima. In genere per i problemi di stima puntuale si usano, come funzioni di perdita, delle distanze (o funzioni monotone di distanze) tra θ e $\hat{\theta}$. Poiché però Θ è una variabile aleatoria, la perdita $\mathbb{L}(\Theta, \hat{\theta})$ è essa stessa una quantità aleatoria. Per avere una valutazione sintetica (un numero reale) della qualità di $\hat{\theta}$ si ricorre quindi, in genere, al valore atteso di $\mathbb{L}(\Theta, \hat{\theta})$ effettuato utilizzando la distribuzione a posteriori di Θ . Nel caso assolutamente continuo la quantità che si ottiene - denominata *perdita finale attesa*

- è definita quindi da

$$\rho(\hat{\theta}, \mathbf{x}_n) = \mathbb{E} \left[\mathbb{L}(\Theta, \hat{\theta}) | \mathbf{x}_n \right] = \int_{\Omega} \mathbb{L}(\theta, \hat{\theta}) \pi(\theta | \mathbf{x}_n) d\theta.$$

Poichè la funzione ρ assume valori in \mathbb{R}^+ , date due stime $\hat{\theta}_1$ e $\hat{\theta}_2$, diciamo che la prima è migliore della seconda se $\rho(\hat{\theta}_1, \mathbf{x}_n) < \rho(\hat{\theta}_2, \mathbf{x}_n)$ (l'uguaglianza implica equivalenza delle due stime). Vogliamo quindi determinare, la o le stime che minimizzano $\rho(\cdot, \mathbf{x}_n)$, ovvero le stime ottimali, secondo la funzione di perdita in uso. Si tratta allora di determinare, tra tutte le possibili stime di θ , la/e seguenti:

$$\hat{\theta}^* = \arg \min_{\theta \in \Omega} \rho(\hat{\theta}, \mathbf{x}_n).$$

La perdita finale attesa minima raggiungibile in un problema di stima con funzione di perdita assegnata è quindi $\rho(\hat{\theta}^*, \mathbf{x}_n)$. Le tre principali funzioni di perdita, nel caso di stima puntuali, sono le funzioni di perdita quadratica, assoluta e 0 – 1. Nel seguito ci poniamo, per semplicità, nel caso uniparametrico e assolutamente continuo e supponiamo che media, moda e mediana di $\pi(\theta | \mathbf{x}_n)$ esistano e siano uniche. In particolare, è possibile verificare quanto segue.

1. Se si considera la funzione di perdita *quadratica* $\mathbb{L}(\Theta, \hat{\theta}) = (\theta - \hat{\theta})^2$, allora si ha che

$$\hat{\theta}^* = \hat{\theta}_B = \mathbb{E}(\Theta | \mathbf{x}_n) = \int_{\Omega} \theta \pi(\theta | \mathbf{x}_n) d\theta.$$

In questo caso è semplice allora verificare che $\rho(\hat{\theta}^*, \mathbf{x}_n) = \mathbb{V}(\Theta | \mathbf{x}_n)$, (ovvero: la perdita finale attesa minima coincide con la varianza a posteriori di Θ).

2. Consideriamo la perdita *zero-uno*, definita per $\epsilon > 0$ da

$$\mathbb{L}(\theta, \hat{\theta}) = \begin{cases} 0, & |\theta - \hat{\theta}| \leq \epsilon \\ 1, & |\theta - \hat{\theta}| > \epsilon \end{cases}.$$

In tal caso abbiamo che

$$\hat{\theta}^* = \text{Mo}[\Theta | \mathbf{x}_n] = \arg \max_{\theta \in \Theta} \pi(\theta | \mathbf{x}_n).$$

Se $\pi(\theta | \mathbf{x}_n)$ è una funzione derivabile di θ , la ricerca del punto di massimo si effettua con le derivate prima e seconda rispetto a θ . In genere conviene determinare il punto di massimo di $\pi(\theta | \mathbf{x}_n)$ utilizzando la funzione $\ln \pi(\theta | \mathbf{x}_n)$. La moda (che indichiamo con $\tilde{\theta}$) deve quindi soddisfare le due seguenti condizioni:

$$\frac{d}{d\theta} \ln \pi(\theta | \mathbf{x}_n) |_{\theta=\tilde{\theta}} = 0; \quad \text{e} \quad \frac{d^2}{d\theta^2} \ln \pi(\theta | \mathbf{x}_n) |_{\theta=\tilde{\theta}} < 0.$$

Se invece la densità a posteriori non è derivabile per ogni valore in Ω , si devono utilizzare altre tecniche analitiche (si veda l'Esempio 5.11, relativo al modello uniforme in $[0, \theta]$).

3. Se si considera la funzione di perdita *assoluta* $\mathbb{L}(\Theta, \hat{\theta}) = |\theta - \hat{\theta}|$, allora si ha che

$$\hat{\theta}^* = \mathbb{M}e[\Theta|\mathbf{x}_n].$$

Si tratta quindi del quantile di livello 0.5 di $\pi(\theta|\mathbf{x}_n)$. Poichè qui stiamo assumendo che $\pi(\theta|\mathbf{x}_n)$ sia una densità, $\mathbb{M}e[\Theta|\mathbf{x}_n]$ è il valore di $\theta \in \Theta$ tale che

$$\int_{-\infty}^{\mathbb{M}e[\Theta|\mathbf{x}_n]} \pi(\theta|\mathbf{x}_n) d\theta = \frac{1}{2}.$$

5.3 Esempi notevoli

5.1 Esempio (modello bernoulliano). Se si considera un campione casuale di n osservazioni dal modello bernoulliano e la densità coniugata $\text{Beta}(\theta|\alpha, \beta)$ per Θ , abbiamo verificato nell'esempio 4.3 che $\Theta|\mathbf{x}_n \sim \text{Beta}(\alpha, \beta)$ ovvero che $\pi(\theta|\mathbf{x}_n) = \text{cost} \times \theta^{\bar{\alpha}-1}(1-\theta)^{\bar{\beta}-1}$, con $\bar{\alpha} = \alpha + y_n$, $\bar{\beta} = \beta + n - y_n$ e $y_n = \sum_{i=1}^n x_i$.

Valore atteso a posteriori. L'espressione di $\mathbb{E}[\Theta|\mathbf{x}_n]$ nel caso del modello in esame si ottiene facilmente da quello di $\mathbb{E}[\Theta]$ sostituendo i valori di α e β con quelli di $\bar{\alpha}$ e $\bar{\beta}$. Ricordando che $\Gamma(y+1) = y\Gamma(y)$ e che

$$B(u, v) = \int_0^1 y^{u-1}(1-y)^{v-1} dy = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)},$$

abbiamo che

$$\begin{aligned} \mathbb{E}[\Theta] &= \int_{\Omega} \theta \times \pi(\theta) d\theta = \int_0^1 \theta \times \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta \\ &= \frac{1}{B(\alpha, \beta)} \int_0^1 \theta^{(\alpha+1)-1}(1-\theta)^{\beta-1} d\theta \\ &= \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} = \frac{\alpha}{\alpha+\beta}. \end{aligned}$$

Pertanto

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\bar{\alpha}}{\bar{\alpha} + \bar{\beta}} = \frac{\alpha + \sum_{i=1}^n x_i}{\alpha + \beta + n} = \frac{\alpha + n\bar{x}_n}{\alpha + \beta + n}.$$

Ossevando che

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{(\alpha + \beta) \frac{\alpha}{\alpha + \beta} + n\bar{x}_n}{(\alpha + \beta) + n}$$

e ponendo $n_0 = \alpha + \beta$ si ottiene che

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{n_0}{n_0 + n} \mathbb{E}[\Theta] + \frac{n}{n_0 + n} \hat{\theta}_{mv}$$

5.2 Osservazione/i

1. La stima ottima bayesiana $\hat{\theta}^*$ che si ottiene con la funzione di perdita quadratica è una media ponderata dei valori assunti da $\mathbb{E}[\theta]$ (valore atteso della distribuzione a priori) e di $\hat{\theta}_{mv} = \bar{x}_n$ (stima di massima verosimiglianza di θ), con pesi rispettivamente proporzionali a $n_0 = \alpha + \beta$ e n . Il valore di $\mathbb{E}[\Theta|\mathbf{x}_n]$ è quindi compreso tra il valore della stima di massima verosimiglianza (\bar{x}_n) di θ e la media della distribuzione a priori (μ_0):

$$\min\{\bar{x}_n, \mathbb{E}[\Theta]\} \leq \hat{\theta}^* \leq \max\{\bar{x}_n, \mathbb{E}[\Theta]\}$$

2. La vicinanza di $\mathbb{E}[\Theta|\mathbf{x}_n]$ a $\hat{\theta}_{mv}$ e a $\mathbb{E}[\Theta]$ dipende dal rapporto n/n_0 . Il valore di $\mathbb{E}[\Theta|\mathbf{x}_n]$ risulta tanto più vicino a $\hat{\theta}_{mv}$ quanto più elevato è il rapporto n/n_0 e tanto più vicino a $\mathbb{E}[\Theta]$ quanto minore risulta n/n_0 .
3. Come casi limite, si ha che:
 - per valori elevati di n/n_0 , $\mathbb{E}[\Theta|\mathbf{x}_n] \approx \hat{\theta}_{mv}$;
 - per valori bassi di n/n_0 , $\mathbb{E}[\Theta|\mathbf{x}_n] \approx \mathbb{E}[\Theta]$.

Questo vuole semplicemente dire che, tanto maggiore è l'informazione campionaria rispetto a quella della distribuzione a priori, tanto maggiore il peso di $\hat{\theta}_{mv}$ nella media ponderata che definisce $\mathbb{E}[\Theta|\mathbf{x}_n]$; tanto maggiore è la precisione della distribuzione a priori, tanto maggiore il peso di $\mathbb{E}[\Theta]$ nella media ponderata che definisce $\mathbb{E}[\Theta|\mathbf{x}_n]$.

4. La quantità n_0 ha nel valore atteso $\mathbb{E}[\Theta|\mathbf{x}_n]$ un ruolo simmetrico a quello di n e controlla il peso della distribuzione a priori in quella a posteriori. Per questo motivo è denominata *prior sample size*.
5. Se poniamo $n_0 = \alpha + \beta = 0$, ovvero $\alpha = \beta = 0$, otteniamo che $\mathbb{E}[\Theta|\mathbf{x}_n] = \hat{\theta}_{mv} = \bar{x}_n$. In questo caso la stima bayesiana (media a posteriori) coincide esattamente con la stima frequentista più comune (media campionaria delle osservazioni). Con questa scelta si assume che la distribuzione a priori non porti informazioni su θ . Come vedremo meglio nel Cap. 6, si tratta di un modo per effettuare una analisi bayesiana non informativa. Si osservi che con tale scelta “estrema” per i valori di α e β (che dovrebbero essere strettamente maggiori di zero) la distribuzione iniziale diventa:

$$\pi(\theta) = \frac{1}{\theta(1-\theta)}, \quad \theta \in (0, 1)$$

Questa funzione non integra a uno in Ω e non è quindi una funzione di densità in Ω . Si parla allora di distribuzione iniziale *impropria*. Tuttavia, anche in questo caso estremo, la distribuzione a posteriori è ben definita (purchè non tutte le osservazioni siano uguali tra loro) e si ha che

$$\pi(\theta|\mathbf{x}_n) = \text{Beta}(\theta|y_n, n - y_n), \quad y_n = \sum_{i=1}^n x_i.$$

Come vedremo meglio in seguito, possono esistere più distribuzioni iniziali di tipo non informativo.

Moda a posteriori Dall'espressione di $\pi(\theta|\mathbf{x}_n)$ si ottiene che

$$\begin{aligned}\frac{d}{d\theta} \ln \pi(\theta|\mathbf{x}_n) &= \frac{d}{d\theta} [\ln \text{cost} + (\bar{\alpha} - 1) \ln \theta + (\bar{\beta} - 1) \log(1 - \theta)] \\ &= \frac{\bar{\alpha} - 1}{\theta} - \frac{\bar{\beta} - 1}{1 - \theta}.\end{aligned}$$

Il punto candidato ad essere un massimo per $\ln \pi(\theta|\mathbf{x}_n)$ – e quindi per $\pi(\theta|\mathbf{x}_n)$ – è $\theta = \frac{\bar{\alpha} - 1}{\bar{\alpha} + \bar{\beta} - 2}$. Osservando che, per ogni $\theta \in [0, 1]$

$$\frac{d^2}{d\theta^2} \ln \pi(\theta|\mathbf{x}_n) = -\frac{\bar{\alpha} - 1}{\theta^2} - \frac{\bar{\beta} - 1}{(1 - \theta)^2} < 0,$$

possiamo concludere che

$$\text{Mo}[\Theta|\mathbf{x}_n] = \tilde{\theta} = \frac{\bar{\alpha} - 1}{\bar{\alpha} + \bar{\beta} - 2}$$

è effettivamente un punto di massimo per la densità a posteriori di Θ .

5.3 Osservazione/i. Dalle precedenti espressioni discende che

$$\text{Mo}[\Theta|\mathbf{x}_n] = \frac{\alpha + y_n - 1}{\alpha + \beta + n - 2}.$$

Pertanto

$$\text{Mo}[\Theta|\mathbf{x}_n] = \hat{\theta}_{mv} = \bar{x}_n \iff \begin{cases} \alpha = 1 \\ \beta = 1 \end{cases} \iff \pi(\theta) = I_{[0,1]}(\theta).$$

Per ottenere quindi uguaglianza tra moda a posteriori e stima frequentista si deve quindi utilizzare, come distribuzione iniziale per θ , la densità uniforme in $[0, 1]$.

Mediana a posteriori. Come nella maggior parte degli altri esempi che considereremo, non esiste una formula esplicita per la mediana delle densità a priori e a posteriori. Il calcolo numerico è comunque possibile utilizzando la funzione **R** che fornisce i quantili delle densità beta: $\text{Me}[\Theta|\mathbf{x}_n] = \text{qbeta}(0.5, \bar{\alpha}, \bar{\beta})$

5.4 Esempio (modello di Poisson). Abbiamo visto nell'Esempio 4.5 che se $X_i|\theta \sim \text{Pois}(\theta)$ i.i.d. e se $\Theta \sim \text{Gamma}(\alpha, \text{rate} = \beta)$, allora $\Theta|\mathbf{x}_n \sim \text{Gamma}(\bar{\alpha}, \text{rate} = \bar{\beta})$, ovvero che $\pi(\theta|\mathbf{x}_n) = c \cdot \theta^{\bar{\alpha}-1} \exp\{-\bar{\beta}\theta\}$, con $\bar{\alpha} = \alpha + y_n$ e $\bar{\beta} = \beta + n$, dove $y_n = \sum_{i=1}^n x_i$.

Media a posteriori In questo caso, utilizzando per $\pi(\theta)$ la densità $\text{Ga}(\alpha, \beta)$, abbiamo che

$$\begin{aligned}\mathbb{E}[\Theta] &= \int_{\Omega} \theta \times \pi(\theta) d\theta \\ &= \int_0^{\infty} \theta \times \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} \theta^{(\alpha+1)-1} e^{-\beta\theta} d\theta.\end{aligned}$$

Con la sostituzione $\beta\theta = y$ e osservando che $d\theta = \frac{1}{\beta} dy$, otteniamo

$$\begin{aligned}\mathbb{E}[\Theta] &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} \frac{1}{\beta^\alpha} y^{(\alpha+1)-1} e^{-y} \frac{1}{\beta} dy \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^{\infty} y^{(\alpha+1)-1} e^{-y} dy \\ &= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\beta} \\ &= \frac{\alpha}{\beta},\end{aligned}$$

dove la penultima uguaglianza si ottiene ricordando che $\int_0^{\infty} t^{y-1} e^{-t} dt = \Gamma(y)$, mentre l'ultima uguaglianza è giustificata dal fatto che $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$. Per il valore atteso a posteriori di Θ abbiamo quindi

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\bar{\alpha}}{\beta} = \frac{\alpha + n\bar{x}_n}{\beta + n} = \frac{\beta \times \frac{\alpha}{\beta} + n\bar{x}_n}{\beta + n}.$$

Ponendo $\beta = n_0$ (prior sample size) e ricordando che nel modello in esame la stima di massima verosimiglianza di θ è $\hat{\theta}_{mv} = \bar{x}_n$, si verifica facilmente che

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{n_0}{n_0 + n} \mathbb{E}(\Theta) + \frac{n}{n_0 + n} \hat{\theta}_{mv}.$$

Valgono anche per questo caso considerazioni analoghe a quelle riportate nell'Osservazione 5.2 - 5 per il modello Beta-Bernoulliano.

Moda a posteriori Consideriamo ora direttamente la densità a posteriori di Θ . Abbiamo che $\pi(\theta|\mathbf{x}_n) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\bar{\alpha}-1} \exp\{-\bar{\beta}\theta\}$, e quindi

$$\begin{aligned}\frac{d}{d\theta} \ln \pi(\theta|\mathbf{x}_n) &= \frac{d}{d\theta} [\ln(\beta^\alpha/\Gamma(\alpha)) + (\bar{\alpha}-1) \ln \theta - \bar{\beta}\theta] \\ &= 0 + \frac{\bar{\alpha}-1}{\theta} - \bar{\beta}.\end{aligned}$$

Il valore candidato ad essere punto di massimo di $\pi(\theta|\mathbf{x}_n)$ è quindi $\tilde{\theta} = \frac{\bar{\alpha}-1}{\bar{\beta}}$. Per ogni $\theta > 0$ si ha che $\frac{d^2}{d\theta^2} \ln \pi(\theta|\mathbf{x}_n) = -\frac{\bar{\alpha}-1}{\theta^2} < 0$. Pertanto

$$\text{Mo}[\Theta|\mathbf{x}_n] = \frac{\bar{\alpha}-1}{\bar{\beta}}.$$

Osservando che $\text{Mo}[\Theta|\mathbf{x}_n] = \frac{(\alpha-1)+n\bar{x}_n}{\beta+n}$ si verifica che, se si pone $\alpha = 1$ e $\beta = 0$, si ottiene $\text{Mo}[\Theta|\mathbf{x}_n] = \hat{\theta}_{mv}$ (caso non informativo).

Mediana a posteriori Il valore della mediana a posteriori si ottiene con `qgamma(0.5, $\alpha + y_n$, rate = $\beta + n$)`.
inserire

5.5 Esempio (modello geometrico). Nell'Esempio 4.7 abbiamo visto che, se $X_i|\theta \sim \text{Geom}(\theta)$ i.i.d. e se $\Theta \sim \text{Beta}(\alpha, \beta)$, allora $\Theta|\mathbf{x}_n \sim \text{Beta}(\bar{\alpha}, \bar{\beta})$ con $\bar{\alpha} = \alpha + n$ e $\bar{\beta} = \beta + y_n - n$, $y_n = \sum_{i=1}^n x_i$. Pertanto, ricordando anche quanto visto nell'Esempio 5.1, abbiamo che

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\bar{\alpha}}{\bar{\alpha} + \bar{\beta}} = \frac{\alpha + n}{\alpha + \beta + y_n}$$

Inoltre, sempre per le proprietà delle densità Beta, abbiamo che

$$\text{Mo}[\Theta|\mathbf{x}_n] = \frac{\bar{\alpha} - 1}{\bar{\alpha} + \bar{\beta} - 1} = \frac{\alpha + n - 1}{\alpha + \beta + y_n - 2}.$$

Infine abbiamo che la mediana a posteriori risulta essere

$$\text{Me}[\Theta|\mathbf{x}_n] = \text{qbeta}(0.5, \alpha + n, \beta + y_n - n).$$

Dall'espressione di $\mathbb{E}[\Theta|\mathbf{x}_n]$, ricordando che in questo modello $\hat{\theta}_{mv} = 1/\bar{x}_n$, si ottiene facilmente che

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\alpha + n}{\alpha \frac{\alpha + \beta}{\alpha} + n \frac{y_n}{n}} = \frac{1}{\frac{\alpha}{\alpha + n} \mathbb{E}[\Theta] + \frac{n}{\alpha + n} \frac{1}{\hat{\theta}_{mv}}}.$$

Ponendo $\alpha = n_0$ (prior sample size) abbiamo quindi che

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{1}{\frac{n_0}{n_0 + n} \mathbb{E}[\Theta] + \frac{n}{n_0 + n} \frac{1}{\hat{\theta}_{mv}}},$$

ovvero che $\mathbb{E}[\Theta|\mathbf{x}_n]$ è la media armonica ponderata di $\mathbb{E}[\Theta]$ e $\hat{\theta}_{mv}$ con pesi rispettivamente uguali a $n_0/(n_0 + n)$ e $n/(n_0 + n)$.

5.6 Esempio (modello esponenziale negativo). Nell'Esempio 4.8 abbiamo visto che, se $X_i|\theta \sim \text{EN}(\theta)$ i.i.d. e se $\Theta \sim \text{Gamma}(\alpha, \beta)$, allora $\Theta|\mathbf{x}_n \sim \text{Gamma}(\bar{\alpha}, \text{rate} = \bar{\beta})$ con $\bar{\alpha} = \alpha + n$ e $\bar{\beta} = \beta + y_n$, $y_n = \sum_{i=1}^n x_i$. Pertanto, ricordando quanto visto nell'Esempio 5.4 a proposito di valore atteso e moda di una v.a. gamma (parametrizzazione `rate`), abbiamo che

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\bar{\alpha}}{\bar{\beta}} = \frac{\alpha + n}{\beta + y_n}$$

e

$$\tilde{\theta} = \text{Mo}[\Theta|\mathbf{x}_n] = \frac{\bar{\alpha} - 1}{\bar{\beta}} = \frac{\alpha + n - 1}{\beta + y_n}.$$

Infine abbiamo che

$$\text{Me}[\Theta|\mathbf{x}_n] = \text{qgamma}(0.5, \alpha + n, \text{rate} = \beta + y_n).$$

In modo del tutto analogo a quanto visto nell'Esempio 5.5 Dall'espressione di $\mathbb{E}[\Theta|\mathbf{x}_n]$ si ottiene facilmente che

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\alpha + n}{\alpha \frac{\beta}{\alpha} + n \frac{y_n}{n}} = \frac{1}{\frac{\alpha}{\alpha+n} \frac{1}{\mathbb{E}[\Theta]} + \frac{n}{\alpha+n} \frac{1}{\hat{\theta}_{mv}}}.$$

Ponendo $\alpha = n_0$ (prior sample size) abbiamo quindi che

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{1}{\frac{n_0}{n_0+n} \frac{1}{\mathbb{E}[\Theta]} + \frac{n}{n_0+n} \frac{1}{\hat{\theta}_{mv}}},$$

ovvero che $\mathbb{E}[\Theta|\mathbf{x}_n]$ è la media armonica ponderata di $\mathbb{E}[\Theta]$ e $\hat{\theta}_{mv}$ con pesi rispettivamente uguali a $n_0/(n_0 + n)$ e $n/(n_0 + n)$.

Si noti che, affinché $\mathbb{E}[\Theta|\mathbf{x}_n] = \hat{\theta}_{mv} = n/y_n$, si deve porre $\alpha = \beta = 0$ mentre per ottenere $\text{Mo}[\Theta|\mathbf{x}_n] = \hat{\theta}_{mv}$ si deve porre $\alpha = 1$ e $\beta = 0$. In entrambi i casi le densità iniziali di Θ sono improprie mentre le corrispondenti densità a posteriori sono ben definite (proprie).

5.7 Esempio (modello esponenziale). Nell'Esempio 4.10 abbiamo visto che se X_1, \dots, X_n è un campione casuale con $X_i|\theta \sim \text{Esp}(\theta) = \text{Ga}(1, \text{scale} = \theta)$, $x \geq 0$, $\theta > 0$. e se $\Theta \sim \text{InvGa}(\alpha, \beta)$, allora $\Theta|\mathbf{x}_n \sim \text{InvGa}(\bar{\alpha}, \bar{\beta})$ con $\bar{\alpha} = \alpha + n$ e $\bar{\beta} = \beta + y_n$. Per ottenere valore atteso e moda a posteriori, determiniamo prima $\mathbb{E}[\Theta]$ e $\text{Mo}[\Theta]$ e poi sostituiamo nelle espressioni trovate (α, β) con $\bar{\alpha}, \bar{\beta}$. Per il valore atteso a priori, osservando che, per $a, b > 0$ si ha che

$$\int_0^\infty \frac{1}{y^{a+1}} e^{-\frac{b}{y}} dy = \frac{\Gamma(a)}{b^a},$$

abbiamo

$$\begin{aligned} \mathbb{E}[\Theta] &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{1}{\theta^{(\alpha-1)+1}} e^{-\frac{\beta}{\theta}} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-1)}{\beta^{\alpha-1}} \\ &= \frac{\beta}{\alpha-1}. \end{aligned}$$

Pertanto

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\beta + y_n}{\alpha - 1 + n}.$$

Ponendo $n_0 = \alpha - 1$ e osservando che $\hat{\theta}_{mv} = \bar{x}_n$, si verifica facilmente che anche in questo caso abbiamo che $\mathbb{E}[\Theta|\mathbf{x}_n] = (n_0\mathbb{E}[\Theta] + n\hat{\theta}_{mv})/(n_0 + n)$. Per ottenere la moda di $\pi(\theta)$ basta osservare che

$$\frac{d}{d\theta} \ln \pi(\theta) = -\frac{\alpha + 1}{\theta} + \frac{\beta}{\theta^2} =$$

ha come radice il valore $\tilde{\theta} = \beta/(\alpha + 1)$ e che la derivata seconda di $\ln \pi(\theta)$ in $\tilde{\theta}$ è minore di zero. Segue quindi che

$$\text{Mo}[\Theta|\mathbf{x}_n] = \frac{\tilde{\beta}}{\tilde{\alpha} + 1} = \frac{\beta + y_n}{\alpha + n + 1}.$$

Infine, la mediana a posteriori si ottiene dalla funzione `qinvgamma()` del pacchetto `invgamma` di R.

5.8 Esempio (modello normale – varianza nota). Nell'esempio 4.11 (modello normale, varianza nota, distribuzione iniziale coniugata) abbiamo verificato che

$$\Theta|\mathbf{x}_n \sim N\left(\frac{n_0\mu_0 + n\bar{x}_n}{n_0 + n}, \frac{\sigma^2}{n_0 + n}\right).$$

In questo caso, la moda, la mediana e il valore atteso della distribuzione a posteriori coincidono:

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \text{Mo}[\Theta|\mathbf{x}_n] = \text{Me}[\Theta|\mathbf{x}_n] = \hat{\theta}^* = \frac{n_0\mu_0 + n\bar{x}_n}{n_0 + n}.$$

5.9 Osservazione/i

1. Si osservi che, come nel caso delle analisi coniugate già esaminate, la stima bayesiana $\hat{\theta}^*$ è una media ponderata dei valori assunti da $\mathbb{E}[\theta] = \mu_0$ e da $\hat{\theta}_{mv} = \bar{x}_n$, con pesi rispettivamente proporzionali all'inverso della varianza della distribuzione a priori (σ^2/n_0) e all'inverso della varianza della distribuzione campionaria di \bar{X}_n (σ^2/n). Il valore di $\hat{\theta}^*$ è quindi compreso tra il valore della stima di massima verosimiglianza (\bar{x}_n) di θ e la media della distribuzione a priori (μ_0): $\min\{\bar{x}_n, \mu_0\} \leq \hat{\theta}^* \leq \max\{\bar{x}_n, \mu_0\}$
2. Di nuovo, la vicinanza di $\mathbb{E}[\Theta|\mathbf{x}_n]$ a \bar{x}_n e a μ_0 dipende dal rapporto n/n_0 . Il valore di $\hat{\theta}_B$ risulta tanto più vicino a \bar{x}_n quanto più elevato è il rapporto n/n_0 e tanto più vicino a μ_0 quanto minore risulta n/n_0 . Si noti che la quantità n_0 *prior sample size* è qui direttamente introdotta come iperparametro, della distribuzione iniziale di θ .
3. La precisione della distribuzione a posteriori, definita come inverso della varianza della stessa distribuzione (pari a $(n_0 + n)/\sigma^2$) risulta uguale alla somma della precisione associata alla distribuzione a priori (pari a n_0/σ^2) e a quella della distribuzione campionaria di \bar{X}_n (uguale a n/σ^2). La combinazione delle due fonti di informazione determina pertanto un aumento di precisione, ovvero una riduzione di incertezza (variabilità) rispetto alle due fonti informative considerate separatamente.

5.10 Esempio (modello normale - valore atteso noto). Nell'Esempio 4.13 abbiamo visto che se X_1, \dots, X_n è un campione casuale con $X_i|\theta \sim N(\mu_0, \theta)$, $\theta \in \mathbb{R}^+$ (μ_0 noto) e se si considera per per la v.a. Θ la densità *gamma-inversa* di parametri α e β , si ottiene che $\Theta|\mathbf{x}_n \sim \text{InvGa}(\bar{\alpha}, \bar{\beta})$, con $\bar{\alpha} = \alpha + n/2$ e $\bar{\beta} = \beta + nS_0^2/2$ e con funzione di densità $\pi(\theta|\mathbf{x}_n) = \bar{\beta}^{\bar{\alpha}}/\Gamma(\bar{\alpha})\theta^{-(\bar{\alpha}+1)}e^{-\bar{\beta}/\theta}$. Per il calcolo del valore atteso e della moda a posteriori, come nei casi precedenti possiamo sfruttare le espressioni delle corrispondenti quantità ottenute dalla distribuzioni a priori, sostituendo (α, β) con $(\bar{\alpha}, \bar{\beta})$. Ricordando che $\int_0^\infty \frac{1}{\theta^{\alpha+1}}e^{-\frac{\beta}{\theta}}d\theta = \frac{\Gamma(\alpha)}{\beta^\alpha}$, si verifica che

$$\mathbb{E}[\Theta] = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{1}{\theta^{(\alpha-1)+1}}e^{-\frac{\beta}{\theta}}d\theta = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-1)}{\beta^{\alpha-1}} = \frac{\beta}{\alpha-1}.$$

Pertanto

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\bar{\beta}}{\bar{\alpha}-1} = \frac{\beta + \frac{n}{2}S_0^2}{\alpha + \frac{n}{2} - 1},$$

dove $S_0^2 = \hat{\theta}_{mv} = \sum_{i=1}^n (X_i - \mu_0)^2/n$. È quindi semplice verificare che

$$\begin{aligned} \mathbb{E}[\Theta|\mathbf{x}_n] &= \frac{(\alpha-1)\frac{\beta}{\alpha-1} + \frac{n}{2}S_0^2}{(\alpha-1) + \frac{n}{2}} \\ &= \frac{2(\alpha-1)\mathbb{E}[\Theta] + n\hat{\theta}_{mv}}{2(\alpha-1) + n} \\ &= \frac{n_0}{n_0 + n}\mathbb{E}[\Theta] + \frac{n}{n_0 + n}\hat{\theta}_{mv}, \end{aligned}$$

dove l'ultima uguaglianza si ottiene ponendo $n_0 = 2(\alpha-1)$ (prior sample size). Per quanto riguarda la moda a priori, osservando che

$$\frac{d}{d\theta} \ln \pi(\theta|\mathbf{x}_n) = -\frac{\bar{\alpha}+1}{\theta} + \frac{\bar{\beta}}{\theta^2},$$

il valore che annulla tale derivata è $\tilde{\theta} = \bar{\beta}/(\bar{\alpha}+1)$. Inoltre, poichè

$$\left. \frac{d^2}{d\theta^2} \ln \pi(\theta|\mathbf{x}_n) \right|_{\theta=\tilde{\theta}} = -\frac{(\bar{\alpha}+1)^3}{\bar{\beta}^2} < 0,$$

possiamo concludere che

$$\text{Mo}[\Theta|\mathbf{x}_n] = \frac{\bar{\beta}}{\bar{\alpha}+1} = \frac{2\beta + nS_0^2}{2\alpha + 1}.$$

Infine, la mediana a posteriori si ottiene dalla funzione `qinvgamma()` del pacchetto `invgamma` di R.

5.11 Esempio (modello uniforme). Nell'esempio 4.14 abbiamo visto che la densità posteriori è $\pi(\theta|\mathbf{x}_n) = \frac{\bar{\alpha}\bar{\beta}^{\bar{\alpha}}}{\theta^{\bar{\alpha}+1}}\mathbb{I}_{[\bar{\beta}, +\infty)}(\theta)$ con $\bar{\alpha} = \alpha + n$ e $\bar{\beta} = \max\{\beta, x_{(n)}\}$.

Media a posteriori È semplice determinare il valore atteso a priori di Θ . Si ha infatti che

$$\begin{aligned}\mathbb{E}[\Theta] &= \alpha\beta^\alpha \int_\beta^\infty \theta \cdot \frac{1}{\theta^{\alpha+1}} d\theta = \alpha\beta^\alpha \int_\beta^\infty \frac{1}{\theta^{(\alpha-1)+1}} d\theta \\ &= \frac{\alpha\beta^\alpha}{(\alpha-1)\beta^{\alpha-1}} \int_\beta^\infty \text{Pa}(\theta|\alpha-1, \beta) d\theta = \frac{\alpha\beta}{\alpha-1}.\end{aligned}$$

Pertanto a posteriori si ha che $\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\bar{\alpha}\bar{\beta}}{\bar{\alpha}-1}$.

Moda a posteriori Il supporto della distribuzione di $\Theta|\mathbf{x}_n$ dipende dal parametro e il punto di massimo di $\pi(\theta|\mathbf{x}_n)$ risulta uguale a $\tilde{\theta} = \bar{\beta}$. Naturalmente tale determinazione non avviene con il metodo delle derivate della densità $\pi(\theta|\mathbf{x}_n)$ dal momento che questa, proprio nel punto di massimo, presenta una discontinuità.

Mediana a posteriori Si può verificare che il quantile di livello ϵ di una v.a. Pareto di parametri (α, β) è uguale a $q_\epsilon = \beta(1-\epsilon)^{-\frac{1}{\alpha}}$. Si ha quindi che $\text{Me}[\Theta|\mathbf{x}_n] = \bar{\beta}2^{\frac{1}{\bar{\alpha}}}$.

5.4 Stima mediante insiemi

Le stime intervallari bayesiane prendono il nome di *insiemi di credibilità*. Forniamo prima la definizione nel caso di Ω sottoinsieme di \mathbb{R}^k , $k \geq 1$, per poi concentrarci sul caso uniparametrico ($k = 1$). Per semplicità, ci limitiamo al caso in cui Θ sia una v.a. assolutamente continua.

5.12 Definizione (Insiemi di credibilità). Sia $\Omega \subseteq \mathbb{R}^k$, $k \geq 1$ e $\pi(\theta|\mathbf{x}_n)$ la funzione di densità posteriori di Θ associata al campione osservato \mathbf{x}_n . Un insieme $C \subset \Omega$ è un *insieme di credibilità* di livello $1 - \gamma$ per il parametro del modello se

$$\mathbb{P}[\Theta \in C|\mathbf{x}_n] = \int_C \pi(\theta|\mathbf{x}_n) d\theta = 1 - \gamma, \quad \gamma \in (0, 1).$$

Un insieme di credibilità dipende dai dati osservati, dalla distribuzione iniziale e dal livello $1 - \gamma$ prescelto. e dovrebbe quindi essere indicato con la notazione $C(\mathbf{x}_n, \pi, \gamma)$. Per rendere la notazione meno pesante nel seguito indicheremo in genere l'insieme semplicemente con la notazione C . Se consideriamo il caso in cui Θ sia una v.a. unidimensionale ($k = 1$), un *intervallo di credibilità* di livello $1 - \gamma$ è un intervallo

$$C = [l_n(\mathbf{x}_n), u_n(\mathbf{x}_n)]$$

di $\Omega \subseteq \mathbb{R}^1$ definito da due opportuni quantili, $l_n(\mathbf{x}_n) \leq u_n(\mathbf{x}_n)$, della densità a posteriori di Θ tali che

$$\int_{l_n(\mathbf{x}_n)}^{u_n(\mathbf{x}_n)} \pi(\theta|\mathbf{x}_n) d\theta = 1 - \gamma.$$

Per semplicità usiamo la notazione $l_n(\mathbf{x}_n) = l_n$ e $u_n(\mathbf{x}_n) = u_n$.

Consideriamo due possibili metodi per la stima mediante regioni: (a) insiemi di massima densità (o probabilità) finale (insiemi HPD: *highest posterior density intervals*); (b) insiemi a code-equiprobabili (insiemi ET: *equal-tails intervals*).

a) Insiemi HPD. Dato lo spazio parametrico Ω e la funzione di densità $\pi(\theta|\mathbf{x}_n)$ a posteriori della v.a. Θ , un sottoinsieme C di Ω definito ponendo

$$C : \{\theta \in \Omega : \pi(\theta|\mathbf{x}_n) \geq h\}, \quad h > 0,$$

è un insieme HPD. L'insieme C si dice di livello $1 - \gamma$ se la sua probabilità a posteriori risulta uguale a $1 - \gamma \in (0, 1)$. L'insieme C è quindi, come un insieme di verosimiglianza, un *insieme di sopravello*: contiene tutti e i soli valori del parametro in corrispondenza dei quali la densità finale risulta maggiore o uguale al valore h . Nel caso di densità a posteriori unimodali è possibile verificare (vedi, ad esempio, Casella-Berger (2002), pp. 441-2) che l'insieme HPD di livello $1 - \gamma$ è l'intervallo i cui estremi l_n e u_n soddisfano le due seguenti condizioni:

a) esiste un valore h_γ in corrispondenza biunivoca con γ , tale che

$$\pi(l_n|\mathbf{x}_n) = \pi(u_n|\mathbf{x}_n) = h_\gamma;$$

b) l'integrale definito sull'intervallo $[l_n, u_n]$ della distribuzione finale è uguale a $1 - \gamma$:

$$\int_{l_n}^{u_n} \pi(\theta|\mathbf{x}_n) d\theta = 1 - \gamma.$$

b) Insiemi ET. Data la funzione di densità finale $\pi(\theta|\mathbf{x}_n)$, l'insieme $[l_n, u_n] \in \Omega$ è un intervallo ET di livello $1 - \gamma$ se

$$l_n = q_{\frac{\gamma}{2}} \quad \text{e} \quad u_n = q_{1-\frac{\gamma}{2}},$$

dove q_ϵ indica il quantile di livello ϵ di $\pi(\theta|\mathbf{x}_n)$. Si ha quindi che

$$\int_{-\infty}^{q_{\frac{\gamma}{2}}} \pi(\theta|\mathbf{x}_n) d\theta = \int_{q_{1-\frac{\gamma}{2}}}^{+\infty} \pi(\theta|\mathbf{x}_n) d\theta = \frac{\gamma}{2}$$

ovvero che

$$\int_{q_{\frac{\gamma}{2}}}^{q_{1-\frac{\gamma}{2}}} \pi(\theta|\mathbf{x}_n) d\theta = 1 - \gamma.$$

5.13 Osservazione/i

- Se $\pi(\theta|\mathbf{x}_n)$ è una densità con moda unica interna a Ω , l'insieme HPD di livello $1 - \gamma$ è quello di lunghezza minima tra tutti gli insiemi di credibilità di uguale livello (vedi, ad esempio, Casella-Berger (2002), pp. 441-2).

- Nel caso di una distribuzione a posteriori simmetrica rispetto al punto di massimo, gli insiemi HPD e ET di livello $1 - \gamma$ coincidono.
- Per la maggior parte dei modelli considerati in questo testo (ovvero: bernouliano, Poisson, geometrico, esponenziale negativo, esponenziale, normale con valore atteso noto) gli insiemi HPD (sia nel caso di analisi coniugata che non informativa) non possono essere determinati in forma analitica, in quanto non è possibile risolvere in forma esplicita l'equazione $\pi(\theta|\mathbf{x}_n) = h_\gamma$. In questi casi si può ricorrere, ad esempio, a funzioni R per la determinazione numerica di tali insiemi. Due eccezioni notevoli, considerate in due esempi che seguono, sono il modello normale (varianza nota) e il modello uniforme in $[0, \theta]$, per i quali esistono formule esplicite degli insiemi HPD di livello $1 - \gamma$.
- Per numerosità campionarie piuttosto elevate, la normalità asintotica della distribuzione a posteriori (conseguenza della normalità asintotica della funzione di verosimiglianza) garantisce che, in un determinato problema inferenziale, a parità di distribuzione iniziale e considerando lo stesso campione di osservazioni, insiemi HPD e insiemi ET di stesso livello $1 - \gamma$ tendano a diventare sempre più vicini.

5.14 Esempio (modello normale – varianza nota). In questo caso la distribuzione a posteriori di θ è simmetrica rispetto al suo valore atteso (che è anche mediana e moda) e quindi si verifica agevolmente che l'insieme HPD (che è anche insieme ET) di livello $1 - \gamma$ è dato dall'intervallo

$$C = \left[\frac{n_0\mu_0 + n\bar{x}_n}{n_0 + n} - z_{1-\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n_0 + n}}, \frac{n_0\mu_0 + n\bar{x}_n}{n_0 + n} + z_{1-\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n_0 + n}} \right].$$

Osservazioni

1. L'insieme di credibilità HPD coincide con intervallo ET, che ha come estremi i percentili di livello $\frac{\gamma}{2}$ e $1 - \frac{\gamma}{2}$ della densità a posteriori di θ .
2. Per ogni $n_0 > 0$ la lunghezza dell'insieme HPD è inferiore a quella dei corrispondenti insiemi di verosimiglianza e di confidenza¹.
3. Per $n_0 = 0$ (caso non informativo) l'intervallo di credibilità coincide formalmente con l'intervallo di confidenza di livello $1 - \gamma$:

$$C_F = \bar{x}_n \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

Si tratta di una (importante) coincidenza formale: le interpretazioni restano diverse. Si osservi inoltre che la lunghezza di entrambi gli intervalli dipende

¹Quanto affermato ha senso a parità di livello di credibilità e di confidenza, concetti che vanno comunque tenuti distinti nell'interpretazione, e supponendo di considerare un livello $q \in (0, 1)$ per l'insieme di verosimiglianza tale che $\sqrt{-2 \ln q}$ coincida con il percentili $z_{1-\frac{\alpha}{2}}$, $\alpha > 1/2$.

dai dati solo attraverso la dimensione campionaria n e che, fissato il valore di α , la lunghezza dell'intervallo frequentista C_F , pari a $2z_{1-\frac{\alpha}{2}}\sigma/\sqrt{n}$, è sempre maggiore di quella dell'intervallo bayesiano ottenuto con distribuzione iniziale propria ($n_0 > 0$).

Nella maggior parte degli esempi considerati nel Cap. 4 (nel caso di analisi coniugata: modelli beta-bernoulliano, Poisson-gamma, esponenziale negativo-gamma, esponenziale-gamma inversa, normale con valore atteso noto-gamma inversa), le densità a posteriori dei parametri dei modelli sono continue nel supporto, unimodali con moda strettamente interna al supporto di $\pi(\theta|\mathbf{x}_n)$ e asimmetriche. In questi casi:

1. gli insiemi ET di livello $1-\gamma$ hanno come estremi inferiore e superiore i quantili di livello $\gamma/2$ e $1-\gamma/2$ delle distribuzioni a posteriori, che si determinano con le opportune funzioni R (ad esempio, `qbeta`, `qgamma` etc.);
2. gli insiemi HPD di livello $1-\gamma$ si determinano numericamente, utilizzando, ad esempio, la funzione `hdi` del pacchetto `HDInterval` di R.

Nel caso di modelli con supporto dipendente da parametro e densità a posteriori con discontinuità di primo tipo in corrispondenza della moda, consideriamo unicamente gli insiemi a coda unica di tipo HPD per il calcolo dei quali si procede come nell'esempio che segue.

5.15 Esempio (modello uniforme). Nell'Esempio 4.14 abbiamo visto che se $X_i|\theta \sim \text{Unif}[0, \theta]$, $i = 1, \dots, n$ i.i.d., $\theta > 0$ e se $\Theta \sim \text{Pa}(\alpha, \beta)$, allora $\Theta|\mathbf{x}_n \sim \text{Pa}(\bar{\alpha}, \bar{\beta})$, con $\bar{\alpha} = \alpha + n$ e $\bar{\beta} = \max\{\beta, x_{(n)}\}$. Poichè $\pi(\theta|\mathbf{x}_n)$ vale zero per $\theta < \bar{\beta}$, ha una moda (unica) in $\theta = \bar{\beta}$ e decresce a zero per $\theta > \bar{\beta}$, necessariamente l'insieme HPD di livello $1-\gamma$ coincide con l'intervallo $[\bar{\beta}, q_{1-\gamma}(\mathbf{x}_n)]$, dove $q_{1-\gamma}(\mathbf{x}_n)$ è il quantile di livello $1-\gamma$ di $\pi(\theta|\mathbf{x}_n)$. Ricordando l'espressione del quantile di livello ϵ di una v.a. Pareto (vedi Esempio 5.11), abbiamo che $q_{1-\gamma}(\mathbf{x}_n) = \bar{\beta}\gamma^{-\frac{1}{\bar{\alpha}}}$.

5.5 Verifica di ipotesi

Come nel caso dei due precedenti problemi inferenziali (stima puntuale e intervallare), la scelta tra ipotesi in ambito bayesiano si basa su elaborazioni che coinvolgono la distribuzione a posteriori. Sia $\{\Omega_0, \Omega_1\}$ una partizione di Ω e supponiamo di dover scegliere tra le due ipotesi $H_1 : \theta \in \Omega_0$ vs. $\theta \in \Omega_1$. Indichiamo con $\mathbb{P}(H_i)$ le probabilità a priori delle ipotesi, $i = 1, 2$. In ottica bayesiana il metodo più intuitivo per confrontare due ipotesi si basa sul confronto delle corrispondenti probabilità a posteriori $\mathbb{P}(H_i|\mathbf{x}_n) = \mathbb{P}(\Theta \in \Omega_i|\mathbf{x}_n)$, $i = 0, 1$. Tali probabilità si ottengono con il teorema di Bayes. Si sceglie l'ipotesi H_0 se il rapporto

$$\frac{\mathbb{P}(H_0|\mathbf{x}_n)}{\mathbb{P}(H_1|\mathbf{x}_n)}$$

risulta maggiore di uno; si sceglie invece l'ipotesi H_1 se il rapporto è inferiore a uno. Le formule esplicite per ottenere le probabilità $P(H_i|\mathbf{x}_n)$ e il rapporto $\mathbb{P}(H_0|\mathbf{x}_n)/\mathbb{P}(H_1|\mathbf{x}_n)$ dipendono dal tipo di ipotesi a confronto. Nei paragrafi che seguono vedremo che, ad esempio, nel caso di ipotesi semplici² $H_1 : \theta = \theta_0$ vs. $\theta = \theta_1$ (dove θ_i è un punto in Ω , $i = 0, 1$), il rapporto delle probabilità a posteriori diventa

$$\frac{f_n(\mathbf{x}_n|\theta_0)\mathbb{P}(H_0)}{f_n(\mathbf{x}_n|\theta_1)\mathbb{P}(H_1)} = \lambda_{01}(\mathbf{x}_n) \times \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)},$$

dove $\lambda_{01}(\mathbf{x}_n)$ indica il rapporto delle verosimiglianze di θ_0 e θ_1 . Se consideriamo invece il sistema di ipotesi $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ il rapporto delle probabilità a posteriori delle ipotesi diventa:

$$\frac{\int_{-\infty}^{\theta_0} \pi(\theta|\mathbf{x}_n)d\theta}{\int_{\theta_0}^{+\infty} \pi(\theta|\mathbf{x}_n)d\theta} = \frac{F(\theta_0|\mathbf{x}_n)}{1 - F(\theta_0|\mathbf{x}_n)},$$

dove $F(\cdot|\mathbf{x}_n)$ indica la funzione di ripartizione a posteriori di θ . Nel paragrafo che segue mostriamo nel dettaglio come procedere.

5.16 Esempio (esempio bernoulliano).

5.17 Esempio (esempio normale – varianza nota).

5.6 Fattore di Bayes

Uno metodo bayesiano spesso utilizzato per il confronto tra ipotesi è il fattore di Bayes. Per introdurlo, consideriamo il sistema di ipotesi $H_0 : \theta \in \Omega_0$ vs. $H_1 : \theta \in \Omega_1$ e, come sopra, indichiamo con $\mathbb{P}(H_i)$ e $\mathbb{P}(H_i|\mathbf{x}_n)$ le probabilità a priori e a posteriori delle ipotesi H_i , $i = 1, 2$. Per semplicità consideriamo, per il momento, il caso in cui X_1, \dots, X_n sono variabili aleatorie discrete. Per il teorema di Bayes

$$\mathbb{P}(H_i|\mathbf{x}_n) = \frac{\mathbb{P}(\mathbf{X}_n = \mathbf{x}_n|H_i)\mathbb{P}(H_i)}{\mathbb{P}(\mathbf{X}_n = \mathbf{x}_n)}$$

dove, per la legge delle probabilità totali, $\mathbb{P}(\mathbf{X}_n = \mathbf{x}_n) = \sum_{j=1}^2 \mathbb{P}(\mathbf{X}_n = \mathbf{x}_n|H_j)\mathbb{P}(H_j)$. Nelle formule precedenti $\mathbb{P}(\mathbf{X}_n = \mathbf{x}_n|H_i)$ indica la probabilità che la v.a. \mathbf{X}_n assuma i valori \mathbf{x}_n condizionatamente all'ipotesi che $\theta \in \Omega_i$; $\mathbb{P}(\mathbf{X}_n = \mathbf{x}_n)$ è invece la probabilità marginale (rispetto alle due ipotesi H_0 e H_1) che $\mathbf{X}_n = \mathbf{x}_n$. Possiamo allora scrivere che

$$m_i(\mathbf{x}_n) = \mathbb{P}(\mathbf{X}_n = \mathbf{x}_n|H_i), \quad i = 1, 2.$$

Si ha quindi che

$$\begin{aligned} \frac{\mathbb{P}(H_0|\mathbf{x}_n)}{\mathbb{P}(H_1|\mathbf{x}_n)} &= \frac{\mathbb{P}(\mathbf{X}_n = \mathbf{x}_n|H_0)}{\mathbb{P}(\mathbf{X}_n = \mathbf{x}_n|H_1)} \times \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} \\ &= B_{01}(\mathbf{x}_n) \times \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}. \end{aligned}$$

²In questo caso la v.a. Θ è discreta e può assumere solo i valori θ_0 e θ_1 .

dove

$$B_{01}(\mathbf{x}_n) = \frac{\mathbb{P}(\mathbf{X}_n = \mathbf{x}_n | H_0)}{\mathbb{P}(\mathbf{X}_n = \mathbf{x}_n | H_1)} = \frac{m_0(\mathbf{x}_n)}{m_1(\mathbf{x}_n)} \quad (5.1)$$

è il fattore di Bayes di H_0 rispetto H_1 associato al campione \mathbf{x}_n . Dalle precedenti formule si evince che $B_{01}(\mathbf{x}_n)$ coincide con il rapporto tra le odds³ a posteriori e le odds a priori di H_0 e H_1 :

$$B_{01}(\mathbf{x}_n) = \frac{\mathbb{P}(H_0 | \mathbf{x}_n) / \mathbb{P}(H_1 | \mathbf{x}_n)}{\mathbb{P}(H_0) / \mathbb{P}(H_1)} \quad (5.2)$$

Il fattore di Bayes è quindi un numero reale positivo che esprime la variazione che si ha nell'odds di H_0 rispetto a H_1 dopo aver osservato i dati \mathbf{x}_n . Per questo motivo possiamo affermare che $B_{01}(\mathbf{x}_n)$ quantifica l'*evidenza* che i dati osservati \mathbf{x}_n portano a favore delle ipotesi a confronto e che può essere interpretato nel modo seguente:

- $B_{01}(\mathbf{x}_n) > 1 \quad \implies$ i dati forniscono evidenza a favore di H_0 ;
- $B_{01}(\mathbf{x}_n) < 1 \quad \implies$ i dati forniscono evidenza a favore di H_1 ;
- $B_{01}(\mathbf{x}_n) = 1 \quad \implies$ i dati forniscono stessa evidenza per le due ipotesi.

Il fattore di Bayes può quindi essere utilizzato come statistica test per il confronto tra le ipotesi H_0 e H_1 . Si osservi che $B_{01}(\mathbf{x}_n)$ non dipende dalle probabilità a priori delle ipotesi, $\mathbb{P}(H_i)$, $i = 1, 2$, e quindi che l'evidenza fornita dai dati che esso esprime non dipende dall'assegnazione delle *probabilità* iniziali assegnate a H_0 e H_1 . Il valore che assume $B_{01}(\mathbf{x}_n)$ dipende invece dalle *distribuzioni* di probabilità iniziali assegnate a Θ .

Per comprendere quanto affermato, esplicitiamo le espressioni di $m_i(\mathbf{x}_n) = \mathbb{P}(\mathbf{X}_n = \mathbf{x}_n | H_i)$ della distribuzioni a priori di Θ , condizionatamente agli eventi $\theta \in \Omega_i$, $i = 1, 2$. Assumendo che Θ sia una v.a. assolutamente continua, indichiamo, al solito, con $\pi(\theta)$ la densità iniziale e con

$$\tilde{\pi}_i(\theta) = \frac{\pi(\theta)}{\int_{\Omega_i} \pi(\theta) d\theta} \cdot \mathbb{I}_{\Omega_i}(\theta) = \frac{1}{\Pi_i} \cdot \pi(\theta) \cdot \mathbb{I}_{\Omega_i}(\theta) \quad (5.3)$$

la funzione di densità di Θ con supporto Ω_i che si ottiene per restrizione della funzione di densità di Θ a Ω_i , dove $\Pi_i = \mathbb{P}(H_i)$. Abbiamo allora che

$$m_i(\mathbf{x}_n) = \int_{\Omega_i} f_n(\mathbf{x}_n | \theta) \tilde{\pi}_i(\theta) d\theta, \quad i = 1, 2 \quad (5.4)$$

indica la probabilità marginale di \mathbf{x}_n , condizionata all'evento che $\theta \in \Omega_i$. Per la (5.1) possiamo quindi scrivere che

$$B_{01}(\mathbf{x}_n) = \frac{m_0(\mathbf{x}_n)}{m_1(\mathbf{x}_n)} = \frac{\int_{\Omega_0} f_n(\mathbf{x}_n | \theta) \tilde{\pi}_0(\theta) d\theta}{\int_{\Omega_1} f_n(\mathbf{x}_n | \theta) \tilde{\pi}_1(\theta) d\theta}. \quad (5.5)$$

³Considerato un evento A , si definisce odds il rapporto $\frac{\mathbb{P}(A)}{1 - \mathbb{P}(A)}$.

La precedente espressione mostra la dipendenza del fattore di Bayes dalle *distribuzioni* iniziali $\pi_i(\cdot)$ dei parametri e non dipendenza dalle *probabilità* iniziali Π_i delle ipotesi, $i_0, 1$.

5.18 Osservazione/i

1. Si noti che, se X_1, \dots, X_n sono v.a. assolutamente continue, tutte le formule precedenti sono ancora valide e $m_i(\mathbf{x}_n)$ indica semplicemente la densità marginale (5.4) di \mathbf{X}_n calcolata in \mathbf{x}_n .
2. La (5.4) mostra che $B_{01}(\mathbf{x}_n)$ è costituito dal rapporto delle verosimiglianze integrate $m_i(\mathbf{x}_n)$, dove l'integrazione viene effettuata rispetto alle densità a priori $\tilde{\pi}_i(\theta)$, $i = 1, 2$ e può quindi essere interpretato come una estensione bayesiana della statistica test rapporto delle verosimiglianze (con massimizzazione sostituita da integrazione).
3. La misura di evidenza di H_1 rispetto a H_0 si ottiene semplicemente invertendo il valore di $B_{01}(\mathbf{x}_n)$, ovvero: $B_{10}(\mathbf{x}_n) = [B_{01}(\mathbf{x}_n)]^{-1}$.
4. Se T è una statistica sufficiente per il modello e $T(\mathbf{x}_n) = t$, allora $B_{01}(\mathbf{x}_n) = B_{01}(t) = m_0^T(t)/m_1^T(t)$, dove $m_i^T(\cdot) = \int_{\mathcal{T}} f_T(t|\theta)\tilde{\pi}_i(\theta)d\theta$ indica la distribuzione marginale di T (con \mathcal{T} spazio campionario di T).

Fattore di Bayes per verifica di ipotesi su parametri scalari

Esaminiamo ora il caso dei tre sistemi di ipotesi più comuni utilizzati ($\Omega \subseteq \mathbb{R}$).

A – Ipotesi puntuali. Consideriamo il sistema di ipotesi

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1$$

dove $\theta_0, \theta_1 \in \mathbb{R}$. In questo caso $\Omega = \{\theta_0, \theta_1\}$, $\Omega_i = \{\theta_i\}$, $i = 0, 1$, e $\pi(\theta)$ è una legge che assegna probabilità Π_0 e $\Pi_1 = 1 - \Pi_0$ a θ_0 e θ_1 . Poniamo allora

$$\tilde{\pi}_i(\theta) = \delta_{\theta_i}(\theta), \quad i = 1, 2$$

dove $\delta_{\theta_i}(\theta)$ è il *delta di Dirac* di θ_i , ovvero la funzione di θ tale che,

$$\int_{\Omega} f_n(\mathbf{x}_n|\theta)\delta_{\theta_i}(\theta)d\theta = f_n(\mathbf{x}_n|\theta_i).$$

Di conseguenza il fattore di Bayes coincide con il rapporto delle verosimiglianze $\lambda_{01}(\mathbf{x}_n)$:

$$B_{01}(\mathbf{x}_n) = \frac{f_n(\mathbf{x}_n|\theta_0)}{f_n(\mathbf{x}_n|\theta_1)} = \frac{L(\theta_0; \mathbf{x}_n)}{L(\theta_1; \mathbf{x}_n)} = \lambda_{01}(\mathbf{x}_n). \quad (5.6)$$

B – Ipotesi nulla puntuale e ipotesi alternativa bilaterale. Per il sistema di ipotesi

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$$

con $\theta_0 \in \mathbb{R}$, abbiamo che $\Omega_0 = \{\theta_0\}$, $\Omega_1 = \Omega - \{\theta_0\}$. Inoltre $\tilde{\pi}_0(\theta) = \delta_{\theta_0}(\theta)$ mentre $\tilde{\pi}_1(\theta)$, coincide con $\pi(\theta)$ per ogni $\theta \in \theta$ eccetto che in θ_0 , dove non è definita. Quindi, per le proprietà della funzione delta di Dirac e le proprietà degli integrali⁴ abbiamo che

$$m_0(\mathbf{x}_n) = f_n(\mathbf{x}_n|\theta_0) \quad \text{e} \quad m_1(\mathbf{x}_n) = \int_{\Omega_1} f_n(\mathbf{x}_n|\theta)\tilde{\pi}(\theta)d\theta = \int_{\Omega} f_n(\mathbf{x}_n|\theta)\pi(\theta)d\theta.$$

Pertanto

$$B_{01}(\mathbf{x}_n) = \frac{f_n(\mathbf{x}_n|\theta_0)}{m(\mathbf{x}_n)}.$$

Il fattore di Bayes risulta quindi uguale al rapporto tra la verosimiglianza di θ_0 e la verosimiglianza integrata rispetto a $\pi(\theta)$ in Ω .

5.19 Osservazione/i Per questo sistema di ipotesi, quando Ω è illimitato non possono generalmente essere utilizzate le usuali distribuzioni non informative dal momento che, con tali distribuzioni, l'integrale che definisce $m(\cdot)$ non esiste (vedi Esempio 5.20).

C – Ipotesi unilaterali. Consideriamo ora il sistema di ipotesi

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

dove $\theta_0 \in \mathbb{R}$. Se indichiamo con $F_{\Theta}(\cdot)$ e $F_{\Theta|\mathbf{x}_n}(\cdot)$ rispettivamente le funzioni di ripartizioni a priori e a posteriori di Θ , utilizzando la (5.2) si ottiene che

$$B_{01}(\mathbf{x}_n) = \frac{F_{\Theta|\mathbf{x}_n}(\theta_0)/[1 - F_{\Theta|\mathbf{x}_n}(\theta_0)]}{F_{\Theta}(\theta_0)/[1 - F_{\Theta}(\theta_0)]}. \quad (5.7)$$

Si giunge alla medesima espressione anche utilizzando la (5.4). Abbiamo infatti che

$$\begin{aligned} m_i(\mathbf{x}_n) &= \int_{\Omega} f_n(\mathbf{x}_n|\theta)\tilde{\pi}_i(\theta)d\theta \\ &= \frac{1}{\Pi_i} \int_{\Omega_i} f_n(\mathbf{x}_n|\theta)\pi(\theta)\mathbb{I}_{\Omega_i}(\theta)d\theta \\ &= \frac{1}{\Pi_i} \int_{\Omega_i} m(\mathbf{x}_n)\pi(\theta|\mathbf{x}_n)d\theta \\ &= m(\mathbf{x}_n) \cdot \frac{\mathbb{P}(H_i|\mathbf{x}_n)}{\Pi_i}, \end{aligned}$$

dove la penultima uguaglianza si ottiene ricordando che $f_n(\mathbf{x}_n|\theta)\pi(\theta) = m(\mathbf{x}_n)\pi(\theta|\mathbf{x}_n)$. Ne segue che l'espressione di $B_{01}(\mathbf{x}_n)$ ottenuta con la (5.5) coincide esattamente con la (5.7).

5.20 Esempio (modello normale). Supponiamo che $X_i|\theta \sim N(\theta, \sigma^2)$, che Θ abbia distribuzione $N(\mu_0, \sigma^2/n_0)$ e che il valore di σ^2 sia noto. Il calcolo del

⁴Ricordare che $\int_{\mathbb{R}} g(x)dx = \int_{\mathbb{R}-\{x_0\}} h(x)dx$ se $g(\cdot)$ e $h(\cdot)$ coincidono per ogni $x \in \mathbb{R}$ eccetto che in x_0 e se $g(x_0), h(x_0) \in \mathbb{R}$.

fattore di Bayes può essere effettuato considerando la distribuzione campionaria della statistica sufficiente, ovvero il fatto che $\bar{X}_n|\theta \sim N(\theta, \sigma^2/n)$ e ricordando che la densità marginale $m(\cdot)$ di \bar{X}_n è quella di una v.a. normale di parametri μ_0 e $\sigma^2(n_0^{-1} + n^{-1})$.

- A. *Ipotesi puntuali.* In questo caso, ricordando quanto mostrato nell'esempio 2.13, indicando con $\phi(z; a, b)$ il valore della funzione di densità della v.a. $N(a, b)$ calcolata in $z \in \mathbb{R}$, abbiamo che

$$B_{01}(\bar{x}_n) = \lambda_{01}(\mathbf{x}_n) = \frac{\phi(\bar{x}_n; \theta_0, \sigma^2/n)}{\phi(\bar{x}_n; \theta_1, \sigma^2/n)}.$$

Con semplici calcoli si mostra che

$$B_{01}(\mathbf{x}_n) = \exp \left\{ -\frac{n(\theta_1 - \theta_0)}{\sigma^2} (\bar{x}_n - \bar{\theta}) \right\} \quad (5.8)$$

dove $\bar{\theta} = (\theta_0 + \theta_1)/2$.

- B. *Ipotesi nulla puntuale e ipotesi alternativa bilaterale.* In questo caso è semplice verificare che $m_0(\bar{x}_n) = \phi(\bar{x}_n; \theta_0, \sigma^2/n)$ e $m_1(\bar{x}_n) = m(\bar{x}_n) = \phi(\bar{x}_n; \mu_0, \sigma^2(1/n_0 + 1/n))$. Con alcuni calcoli si ottiene

$$B_{01}(\bar{x}_n) = \left(\frac{n_0 + n}{n_0} \right)^{1/2} \exp \left\{ -\frac{u^2}{2} \frac{n}{n_0 + n} \right\}, \quad (5.9)$$

dove $u = \sqrt{n}(\bar{x}_n - \theta_0)/\sigma$ è la statistica test comunemente usata nell'impostazione frequentista per la scelta tra le ipotesi in esame. Si noti che la precedente espressione richiede che $n_0 > 0$, ovvero che la distribuzione $\pi(\theta) = N(\theta|\mu_0, \sigma^2/n_0)$ sia propria. Infatti, se $n_0 \rightarrow 0$, $B_{01}(\mathbf{x}_n) \rightarrow +\infty$, qualunque sia il campione osservato \mathbf{x}_n . Questo comportamento è comunemente chiamato *paradosso di Lindley*.

- C. *Ipotesi unilaterali.* Si verifica facilmente che

$$B_{01}(\mathbf{x}_n) = \frac{\Phi\left(\frac{\theta_0 - \bar{\mu}}{\bar{\sigma}}\right) / \left[1 - \Phi\left(\frac{\theta_0 - \bar{\mu}}{\bar{\sigma}}\right)\right]}{\Phi\left(\frac{\theta_0 - \mu_0}{\sigma/\sqrt{n_0}}\right) / \left[1 - \Phi\left(\frac{\theta_0 - \mu_0}{\sigma/\sqrt{n_0}}\right)\right]} \quad (5.10)$$

dove, ricordando la 4.2, $\bar{\mu} = \mathbb{E}[\Theta|\mathbf{x}_n] = (n_0\mu_0 + n\bar{x}_n)/(n_0 + n)$, $\bar{\sigma}^2 = \mathbb{V}[\Theta|\mathbf{x}_n] = \sigma^2/(n_0 + n)$ e dove $\Phi(\cdot)$ indica la funzione di ripartizione della v.a. $N(0, 1)$.

5.7 Esercizi

1. Siano X_1, \dots, X_n le risposte quantitative (aleatorie) a un trattamento medico somministrato a n pazienti e sia θ l'effetto del trattamento medio nella popolazione da cui provengono i dati. Supporre che, condizionatamente alla conoscenza di θ , le risposte al trattamento siano v.a. indipendenti con distribuzione **normale** centrata sul valore vero dell'effetto del trattamento e varianza unitaria. Si dispone di informazione pre-sperimentale e sperimentale sull'effetto del trattamento.

Informazione sperimentale: su 10 osservazioni sperimentali, la somma delle risposte al trattamento è uguale a 20 (in opportuna unità di misura).

Informazione pre-sperimentale: basandoci su dati storici è lecito supporre che il vero valore del trattamento θ abbia distribuzione **normale** con valore atteso pari a $1/2$ e varianza pari a $1/4$.

- (a) Formalizzare opportunamente il problema per mezzo di un modello bayesiano, ovvero: indicare quali sono le distribuzioni di $X_i|\theta$ e di Θ da utilizzare e i valori da associare ai parametri noti e agli e iperparametri dei modelli considerati.
 - (b) Con riferimento al precedente esercizio: scrivere le espressioni generiche (formule) e specifiche (sostituendo i valori numerici) della distribuzione a posteriori di Θ .
 - (c) Con riferimento ai precedenti due esercizi, fornire una stima puntuale e una intervallare di θ basata sulla distribuzione a posteriori di Θ [riportare espressioni generiche (formule) e specifiche (sostituendo i valori numerici)]. Per la stima intervallare utilizzare un livello di credibilità $\gamma = 0.95$ e ricordare che $u_{0.975} = 1.96$.
2. I seguenti valori numerici indicano la concentrazione di acido lattico (misurata in una opportuna unità di misura) di alcuni campioni di confezioni di formaggio di una determinata marca:

1.5, 1, 0.5, 2, 2, 1, 2.5, 1, 0.5, 1.

I ricercatori ipotizzano che i valori campionari osservati siano realizzazioni di v.a. che tendono a concentrarsi intorno al valore vero incognito di concentrazione di acido lattico con una varianza pari a $1/9$. In uno studio precedente, basato su soli tre campioni di formaggio, la concentrazione media di acido lattico era risultata pari a 2. Siamo interessati a determinare la distribuzione di probabilità del valore vero di concentrazione di acido lattico in una confezione di formaggio.

- (a) Formalizzare quanto sopra in termini di un problema di inferenza bayesiana [sugg.: associare le quantità in gioco a variabili, parametro, dimensione campionaria, iperparametri...].
- (b) Scegliere le distribuzioni di probabilità per variabili e parametro incognito.

- (c) Determinare una stima puntuale della concentrazione di acido lattico in una confezione, utilizzando tutte le informazioni disponibili [sugg.: i valori numerici possono essere espressi in frazioni].
 - (d) Determinare una stima intervallare al 95% della concentrazione di acido lattico sulla base di tutte le informazioni disponibili [sugg.: i valori numerici possono essere espressi in frazioni].
 - (e) Sappiamo che, in una confezione formaggio, la concentrazione di acido lattico non debba superare il valore 2. Calcolare la probabilità che ciò avvenga, sulla base dell'informazione disponibile [sugg.: i valori numerici possono essere espressi in frazioni].
3. Sia X_1, \dots, X_n un campione casuale proveniente da una popolazione di Rayleigh di parametro θ , con funzione di densità

$$f_X(x|\theta) = \frac{x}{\theta^2} \exp\left\{-\frac{x^2}{2\theta^2}\right\}, \quad x \geq 0, \quad \theta > 0.$$

Sia inoltre

$$\pi(\theta) = \frac{\alpha^\beta}{2^{\beta-1}\Gamma(\beta)} \frac{1}{\theta^{2\beta+1}} \exp\left\{-\frac{\alpha}{2\theta^2}\right\}, \quad \alpha, \beta > 0, \quad \theta > 0.$$

- (a) Determinare $\pi(\theta|\mathbf{x}_n)$, verificare che $\pi(\theta)$ è coniugata al modello considerato e determinare le espressioni di $\bar{\alpha}$ e $\bar{\beta}$.
- (b) Sapendo che una stima bayesiana ha la seguente espressione

$$\xi[\Theta|\mathbf{x}_n] = \left[\frac{\Gamma(n + \beta - 1)}{\Gamma(n + \beta)} \left(\frac{t + \alpha}{2} \right) \right]^{1/2}, \quad \text{dove} \quad t = \sum_{i=1}^n x_i^2,$$

determinare la corrispondente espressione $\xi[\Theta]$ ottenuta con la distribuzione a priori di Θ .

- (c) Determinare $\hat{\theta}_{mv}$, stima di massima verosimiglianza di θ .
 - (d) Sapendo che $n = 10$, $t = 5$, $\alpha = 2$, $\beta = 3$, determinare i valori di $\hat{\theta}_{mv}$ e $\xi[\Theta|\mathbf{x}_n]$.
4. Dimostrare le formule 5.8, 5.9 e 5.10.
5. Determinare l'espressione del fattore di Bayes $B_{01}(z_n)$ per i tre sistemi di ipotesi (ipotesi puntuali, puntuale vs. bilaterale, unilaterali) per i seguenti modelli: (a) Bernoulli-Beta; (b) Poisson-Gamma; (c) Esponenziale negativo-Gamma.
6. Sia $X_1, \dots, X_n|\theta \sim N(\theta, 1)$ i.i.d. e $\Theta \sim N(0, n_0^{-1})$. Considerare il sistema di ipotesi $H_0 : \theta = 0$ vs. $H_1 : \theta \neq 0$. Calcolare l'espressione di $B_{01}(\bar{x}_n)$ per $\bar{x}_n = 0$.
7. Siano:

- $X_1, \dots, X_n | \theta \sim \text{Pois}(\theta)$ i.i.d., $\theta > 0$;
- $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$;
- $\theta_0 > \theta_1 > 0$;
- \mathbf{x}_n un campione osservato.

- (a) Determinare l'espressione di $B_{01}(\mathbf{x}_n)$.
 (b) Verificare che, preso un valore $k > 0$, si ha

$$B_{01}(\mathbf{x}_n) > k \quad \Leftrightarrow \quad \bar{x}_n > \xi(k),$$

dove $\xi(k)$ è un valore che dipende da k .

- (c) Determinare l'espressione di $\xi(k)$ (funzione di k, n, θ_0, θ_1).
 (d) Supponendo ora che $\Theta \sim \text{Ga}(\alpha, \beta)$, calcolare il valore atteso predittivo a priori di

$$Y = \ln B_{01}(\mathbf{x}_n).$$

[Sugg.: utilizzare la distribuzione predittiva di $\sum_{i=1}^n X_i$].

- (e) Determinare il valore numerico di $\mathbb{E}[Y]$ in corrispondenza dei seguenti valori: $\alpha = \beta = 2$, $n = 10$, $\theta_0 = e$, $\theta_1 = 1$ (e indica il numero di Nepero).

8. Sia $X_1, \dots, X_n | \theta \sim \text{Esp}(\theta)$.

- (a) Determinare $\pi_j(\theta)$, ovvero la distribuzione a priori di Jeffreys.
 (b) Determinare $\pi_j(\theta | \mathbf{x}_n)$ [distribuzione a posteriori per Θ trovata usando $\pi_j(\theta)$] e verificare che coincide con la funzione di densità di una v.a. Gamma-Inversa(n, t), $t = \sum_{i=1}^n x_i$.
 (c) Verificare che, utilizzando $\pi_j(\theta)$, si ottiene

$$m(\mathbf{x}_n) = \frac{\Gamma(n)}{t^n}, \quad \text{con} \quad t = \sum_{i=1}^n x_i.$$

- (d) Determinare l'espressione di $B_{01}(\mathbf{x}_n)$ per il sistema di ipotesi $H_0 : \theta = \theta_0$, vs. $H_1 : \theta \neq \theta_0$.
 (e) Calcolare il valore di $B_{01}(\mathbf{x}_n)$ e stabilire quale ipotesi si sceglie se $\theta_0 = 1$, $t = 2$ e $n = 3$.
 (f) Determinare $\mathbb{E}_j[\Theta | \mathbf{x}_n]$ e determinarne il valore atteso predittivo che si ottiene considerando come nuova distribuzione a priori una Gamma-Inversa(3,3).

9. Sia $X_1, \dots, X_n | \theta \sim \text{Exp}(\theta)$ e $\Theta \sim \text{InvGamma}(\alpha, \beta)$, ovvero:

$$f_X(x|\theta) = \frac{1}{\theta} \exp\left\{-\frac{x}{\theta}\right\}, \quad x > 0, \quad \theta \geq 0,$$

e

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\theta^{\alpha+1}} \exp\left\{-\frac{\beta}{\theta}\right\}, \quad \alpha, \beta > 0.$$

- (a) Determinare $\pi(\theta|\mathbf{x}_n)$ e $\mathbb{E}[\Theta|\mathbf{x}_n]$, sapendo che $\mathbb{E}[\Theta] = \frac{\beta}{\alpha-1}$.
- (b) Si consideri il sistema di ipotesi semplici $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ ($\theta_0 < \theta_1$). Siano p_0 e p_1 le probabilità a priori delle due ipotesi. Determinare l'espressione del fattore di Bayes $B_{01}(\mathbf{x}_n)$.

- (c) Mostrare che

$$B_{01}(\mathbf{x}_n) > 1 \iff \bar{x}_n < k, \quad k > 0$$

e fornire l'espressione esplicita di k in funzione di p_0, p_1, n .

- (d) Calcolare la probabilità

$$\mathbb{P}[B_{01}(\mathbf{x}_n) > k]$$

in funzione della funzione di ripartizione di una v.a. Gamma.

- (e) Determinare la distribuzione a priori di Jeffreys e dire se può essere considerata un membro (proprio o limite) della famiglia delle densità Gamma inverse.

10. Sia $X|\theta \sim f_X(x|\theta)(x) = \theta x^{\theta-1}$, $x \in (0, 1)$ e $\theta > 0$. Considerare: $H_0 : \theta = \theta_0$ e $H_1 : \theta = \theta_1$, con $\theta_0 > \theta_1 > 0$.

- (a) Determinare l'espressione di $B_{01}(x)$, fattore di Bayes per il confronto tra le ipotesi in esame basato su una singola osservazione x .
- (b) Verificare che $B_{01}(x) > 1 \iff x > A$ e determinare l'espressione di A (in funzione di θ_0 e θ_1).
- (c) Determinare la probabilità $\mathbb{P}_{\theta_0}[B_{01}(X) > 1]$ [**Sugg.:** usare la distribuzione di $X|\theta_0$];
- (d) Calcolare il valore numerico che si ottiene ponendo $\theta_0 = 2$ e $\theta_1 = 1$.

Capitolo 6

Analisi oggettiva

6.1 Introduzione

Nei capitoli precedenti abbiamo più volte sottolineato che una caratteristica dell'analisi bayesiana risiede nella possibilità di sfruttare l'informazione pre-sperimentale disponibile sul parametro incognito e formalizzata attraverso la distribuzione a priori. È evidente però anche l'interesse per i casi in cui l'informazione pre-sperimentale sia assente o, se esistente, non la si voglia utilizzare. Ci si chiede cioè se l'analisi bayesiana ammetta la possibilità di essere “non informativa”, o meglio, se sia possibile definire distribuzioni a priori che non portino informazioni aggiuntive rispetto a quelle che si ottengono dai dati campionari, in modo tale che la distribuzione a posteriori sia unicamente determinata dalla funzione di verosimiglianza. Tranne casi molto elementari, tuttavia, il concetto di non informatività di una distribuzione di probabilità è purtroppo piuttosto controverso. Nella letteratura più recente si preferisce infatti parlare di distribuzioni iniziali “oggettive”, basate cioè su regole formali di qualche tipo, da contrapporre alle distribuzioni “soggettive”, ovvero quelle che si basano su informazioni disponibili.

In generale, come vedremo, i principali problemi legati alla definizione di distribuzioni oggettive si presentano nei casi (tutt'altro che insoliti) in cui Ω non è limitato, mentre sono (almeno apparentemente) più semplici i casi il cui lo spazio dei parametri è limitato.

6.1 Esempio (modello bernoulliano). Abbiamo visto che la famiglia coniugata nel modello bernoulliano è la famiglia delle densità Beta di parametri (α, β) :

$$\pi(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}, \quad \theta \in [0, 1], \quad \alpha, \beta > 0.$$

Se si scelgono i valori $\alpha = 1$ e $\beta = 1$, la distribuzione a priori coincide con la distribuzione uniforme in $[0, 1]$. Tale distribuzione traduce una nozione intuitiva di non informatività. La distribuzione a posteriori risultante è una densità Beta di parametri $\bar{\alpha} = 1 + s_n$ e $\bar{\beta} = 1 + n - s_n$, con $s_n = \sum_{i=1}^n x_i$. Una stima puntuale bayesiana di θ è, ad esempio, $\mathbb{E}[\Theta | \mathbf{x}_n] = \bar{\alpha} / (\bar{\alpha} + \bar{\beta}) = (s_n + 1) / (n + 2)$. Per n sufficientemente elevato abbiamo quindi che $\mathbb{E}[\Theta | \mathbf{x}_n] \approx \bar{x}_n = \hat{\theta}_{mv}(\mathbf{x}_n)$. Per ottenere

l'uguaglianza esatta dobbiamo porre $\alpha = 0$ e $\beta = 0$. Si ricordi, però, che il modello Beta richiede che α e β siano strettamente maggiori di zero. Ciò vuol dire che possiamo ottenere $\mathbb{E}[\Theta|\mathbf{x}_n] \approx \bar{x}_n = \hat{\theta}_{mv}(\mathbf{x}_n)$ solo come caso limite, per α, β tendenti entrambi a zero. Per ulteriori sviluppi si veda l'Esempio 6.3.

La semplicità dell'esempio precedente risiede nel fatto che lo spazio dei parametri è un sottoinsieme limitato di \mathbb{R} . In molti casi, tuttavia, lo spazio dei parametri Ω non è limitato, e non è possibile assegnare a Θ una distribuzione uniforme, dal momento che non sarebbe una distribuzione propria, ovvero tale che $\int_{\Omega} \pi(\theta) d\theta = 1$. Tuttavia, come vedremo, anche nel caso apparentemente semplice del modello Bernoulliano permane un problema, ovvero l'esistenza di più proposte sensate di distribuzioni non informative e la conseguente necessità di sceglierne una.

L'approccio oggettivo per la costruzione di distribuzioni iniziali costituisce ormai un capitolo molto vasto dell'analisi bayesiana¹. In questa sede ci occupiamo unicamente di due metodi per definire distribuzioni oggettive. Il primo metodo (prg. 6.2), molto empirico, si sviluppa a partire dall'analisi coniugata di un dato modello. Il secondo metodo (prg. 6.3) si basa invece sulla nozione di invarianza per trasformazioni del parametro.

6.2 Metodo 1: casi limite di distribuzioni coniugate

L'idea alla base del metodo è individuare, attraverso l'analisi della distribuzione a posteriori ottenuta considerando una a priori coniugata, i valori (esatti o limite) degli iperparametri che rendano una procedura inferenziale bayesiana formalmente coincidente con una procedura frequentista proposta per lo stesso problema. Nei problemi di stima puntuale, ad esempio, determiniamo, se possibile, i valori dei parametri della distribuzione a priori che rendano uguali una stima bayesiana, $\hat{\theta}_B(\mathbf{x}_n)$, a una stima frequentista $\hat{\theta}_F(\mathbf{x}_n)$, ovvero tali che

$$\hat{\theta}_B(\mathbf{x}_n) = \hat{\theta}_F(\mathbf{x}_n). \quad (6.1)$$

Nei casi più comuni si pone $\hat{\theta}_B(\mathbf{x}_n) = \mathbb{E}[\Theta|\mathbf{x}_n]$ e $\hat{\theta}_F(\mathbf{x}_n) = \hat{\theta}_{mv}(\mathbf{x}_n)$. In alcuni casi i valori degli iperparametri della distribuzione a priori coniugata necessari a ottenere la corrispondenza sono ammissibili, ma nella maggior parte dei casi la (6.1) si può ottenere solo come caso limite. I valori necessari alla perfetta identità tra stima frequentista e stima bayesiana fanno perdere di significato alla distribuzione a priori, ma resta invece ben definita la distribuzione a posteriori. Questo fatto dimostra, in un certo senso, da un lato l'incompatibilità formale (esatta) tra le due procedure (bayesiana e frequentista), dall'altro il fatto che le procedure inferenziali frequentiste possono essere interpretate come casi particolari (spesso come casi *limite*) di procedure bayesiane.

+++

6.2 Esempio (modello normale-normale). Nel caso del modello normale con valore atteso incognito e varianza nota, considerato nell'Esempio 5.8, se $\Theta \sim N(\mu_0, \sigma^2/n_0)$,

¹Si veda, ad esempio ... per approfondimenti

allora il valore atteso della distribuzione a posteriori, ovvero la stima puntuale bayesiana più comunemente utilizzata, può essere scritto seguente forma:

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{n_0}{n_0 + n} \mathbb{E}(\Theta) + \frac{n}{n_0 + n} \hat{\theta}_{mv}(\mathbf{x}_n), \quad (6.2)$$

dove n indica la dimensione del campione e $\hat{\theta}_{mv}(\mathbf{x}_n) = \bar{x}_n$ indica la stima di massima verosimiglianza di θ e $\mathbb{E}[\Theta] = \mu_0$. Il valore atteso a posteriori è quindi uguale a una media ponderata ottenuta considerando il valore atteso della distribuzione a priori e la stima di massima verosimiglianza, con pesi rispettivamente proporzionali a n_0 e n . Risulta evidente, nella precedente formula, il ruolo simmetrico di n_0 e n . Inoltre,

$$n_0 \rightarrow 0 \quad \implies \quad \hat{\theta}_B(\mathbf{x}_n) \rightarrow \hat{\theta}_{mv}(\mathbf{x}_n).$$

Per $n_0 \rightarrow 0$, la distribuzione iniziale perde di significato (la varianza a priori tende a $+\infty$), ma la distribuzione finale resta bene definita:

$$\Theta|\mathbf{x}_n \sim N\left(\bar{x}_n, \frac{\sigma^2}{n}\right).$$

È semplice verificare che, per $n_0 \rightarrow 0$, le tre procedure bayesiane di stima puntuale, intervallare e di verifica di ipotesi sul parametro θ coincidono con quelle frequentiste. Infatti:

- Stima puntuale: $\hat{\theta}_B(\mathbf{x}_n) = \mathbb{E}[\Theta|\mathbf{x}_n] = \bar{x}_n$ (stima di massima verosimiglianza);
- Stima intervallare: $C_{1-\gamma}^{et} = C_{1-\gamma}^{hpd} = \bar{x}_n \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ (intervallo di confidenza);
- Verifica di ipotesi: $\mathbb{P}[\Theta < \delta|\mathbf{x}_n] = 1 - \Phi\left[\frac{\sqrt{n}(\bar{x}_n - \delta)}{\sigma}\right]$ (p-value per il test unilaterale).

Da tutte le considerazioni fatte, risulta naturale assegnare in questo modello alla quantità n_0 il ruolo di *prior sample size*. \square

L'individuazione della prior sample size, automatico nel modello appena considerato, richiede maggiore attenzione negli altri esempi rilevanti.

6.3 Esempio (modello bernoulliano). Abbiamo visto nell'Esempio che in questo caso, utilizzando la famiglia delle densità Beta di parametri (α, β) , si ottiene

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\bar{\alpha}}{\bar{\alpha} + \bar{\beta}} = \frac{\alpha + \sum_{i=1}^n x_i}{\alpha + \beta + n}.$$

Pertanto, ricordando che $\hat{\theta}_{mv}(\mathbf{x}_n) = \bar{x}_n$, $\mathbb{E}[\Theta|\mathbf{x}_n] = \hat{\theta}_{mv}(\mathbf{x}_n) \Leftrightarrow \alpha = \beta = 0$. La distribuzione a priori oggettiva dobbiamo è quindi quella che si ottiene dalla famiglia delle densità Beta per α e β che tendono a zero:

$$\pi(\theta) = \frac{1}{\theta(1-\theta)} \mathbb{I}_{[0,1]}(\theta).$$

Questa densità non è propria (le densità beta richiedono infatti che α e β siano strettamente positive), ma la distribuzione a posteriori corrispondente è invece una densità ben definita, ovvero una densità di tipo $\text{Beta}(\bar{\alpha}, \bar{\beta}) = \text{Beta}(\sum_{i=1}^n x_i, n - \sum_{i=1}^n x_i)$ che produce, appunto, un valore atteso pari a $\mathbb{E}[\Theta|\mathbf{x}_n] = \hat{\theta}_{mv}(\mathbf{x}_n) = \bar{x}_n$. Vale anche la pena osservare che, in questo modello, la prior sample size è $n_0 = \alpha + \beta$. La distribuzione iniziale non informativa ha quindi prior sample size uguale a zero.

6.4 Esempio (modello di Poisson). Abbiamo visto nell'Esempio che in questo caso, utilizzando la famiglia coniugata delle densità Gamma di parametri (α, β) , si ottiene

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\bar{\alpha}}{\bar{\beta}} = \frac{\alpha + \sum_{i=1}^n x_i}{\beta + n}.$$

Pertanto, di nuovo, $\mathbb{E}[\Theta|\mathbf{x}_n] = \hat{\theta}_{mv}(\mathbf{x}_n) \Leftrightarrow \alpha = \beta = 0$. In questo caso la distribuzione iniziale non informativa che si ottiene come caso limite delle densità gamma al tendere a zero di α e β risulta essere

$$\pi(\theta) \propto \frac{1}{\theta}, \quad \theta > 0$$

che è una distribuzione impropria in \mathbb{R}^+ . Anche in questo modello, la prior sample size diventa $n_0 = \beta = 0$. In questo caso otteniamo che $\Theta|\mathbf{x}_n \sim \text{Gamma}(\sum x_i, \text{rate} = n)$. \square

6.5 Esempio (modello normale-gamma inversa). Abbiamo visto nell'Esempio 4.13 che in questo caso, utilizzando la famiglia coniugata delle densità Gamma inversa di parametri (α, β) , ricordando che $\mathbb{E}[\Theta] = \beta/(\alpha - 1)$, si ottiene

$$\mathbb{E}[\Theta|\mathbf{x}_n] = \frac{\bar{\beta}}{\bar{\alpha} - 1} = \frac{\beta + \frac{n}{2} S_0^2}{\alpha + \frac{n}{2} - 1}.$$

Pertanto, $\mathbb{E}[\Theta|\mathbf{x}_n] = \hat{\theta}_{mv}(\mathbf{x}_n) \Leftrightarrow \alpha = 1, \beta = 0$. La distribuzione iniziale non informativa che si ottiene come caso limite delle densità gamma per $\alpha = 1$ e al tendere di β a zero risulta essere

$$\pi(\theta) \propto \frac{1}{\theta^2}, \quad \theta > 0$$

che è una densità propria. Per la distribuzione a posteriori abbiamo che $\Theta|\mathbf{x}_n \sim \text{InvGa}(n/2, \text{rate} = nS_0^2/2)$.

6.6 Esempio (modello uniforme). Abbiamo visto nell'Esempio (4.14) che se X_1, \dots, X_n un campione casuale con $X_i|\theta \sim \text{Unif}[0, \theta]$ i.i.d., $\theta > 0$ e se $\pi(\theta) = \text{Pa}(\theta|\alpha, \beta) = \frac{\alpha^\beta}{\theta^{\alpha+1}} I_{[\beta, +\infty)}(\theta)$, $\alpha, \beta > 0$, allora $\pi(\theta|\mathbf{x}_n) = \text{Pa}(\bar{\alpha}, \bar{\beta})$, con $\bar{\alpha} = \alpha + n$ e $\bar{\beta} = \max\{\beta, x_{(n)}\}$. Si noti innanzitutto che, in questo caso, ponendo $\beta = 0$ si ottiene che, $\forall \alpha > 0$ la moda della densità risulta essere

$$\tilde{\theta} = \hat{\theta}_{mv}(\mathbf{x}_n) = x_{(n)}.$$

Se invece consideriamo lo stimatore

$$\hat{\theta}_B = \mathbb{E}[\Theta | \mathbf{x}_n] = \frac{\bar{\alpha}}{\bar{\alpha} - 1} \bar{\beta},$$

possiamo proporre diverse possibili scelte di α, β compatibili con l'idea di non-informatività di $\pi(\theta)$.

1. ($\alpha = \beta = 0$). In questo caso si ottiene la distribuzione impropria $\pi(\theta) = \theta^{-1} \times \mathbb{I}_{[0, \infty)}(\theta)$, con la quale si ottiene

$$\hat{\theta}_B = \frac{n}{n-1} x_{(n)}.$$

2. ($\alpha = 1, \beta = 0$). In questo caso si ottiene la distribuzione propria $\pi(\theta) = \theta^{-2} \times \mathbb{I}_{[0, \infty)}(\theta)$, con la quale si ottiene

$$\hat{\theta}_B = \frac{n+1}{n} x_{(n)}.$$

Il corrispondente stimatore $\hat{\Theta}_B(\mathbf{X}_n)$ coincide con lo stimatore frequentista non distorto di varianza minima (UMVUE).

6.3 Metodo 2: distribuzione a priori di Jeffreys

La proposta di Jeffreys si basa sull'idea intuitiva che una distribuzione non informativa debba essere invariante rispetto a cambiamenti nella parametrizzazione. L'idea è la seguente. Consideriamo una funzione g della v.a. Θ e indichiamo con $\Lambda = g(\Theta)$ la variabile aleatoria definita con tale trasformazione. Indichiamo con $\pi_{\Theta}^o(\theta)$ la distribuzione assegnata a Θ attraverso un metodo (o *regola*) che genera distribuzioni *oggettive* e con $\pi_{\Lambda}^o(\lambda)$ la distribuzione che lo stesso metodo assegna alla trasformazione $\Lambda = g(\Theta)$ di Θ . Sia Λ_0 un qualsiasi sottoinsieme di valori che può assumere Λ e sia

$$\Omega_0 = \{\theta \in \Omega : g(\theta) \in \Lambda_0\}$$

l'antimmagine di Λ_0 . Il metodo che genera le distribuzioni a priori è allora invariante se

$$\mathbb{P}(\Lambda \in \Lambda_0) = \mathbb{P}(\Theta \in \Omega_0),$$

ovvero se, utilizzando per semplicità la notazione per variabili assolutamente continue

$$\int_{\Lambda_0} \pi_{\Lambda}^o(\lambda) d\lambda = \int_{\Omega_0} \pi_{\Theta}^o(\theta) d\theta.$$

Sempre nel caso assolutamente continuo, se g è monotona, il metodo che genera le distribuzioni considerate è *invariante se*

$$\pi_{\Lambda}^o(\lambda) = \pi_{\Theta}^o(g^{-1}(\lambda)) \times \left| \frac{d}{d\lambda} g^{-1}(\lambda) \right|, \quad (6.3)$$

ovvero se la legge che troviamo applicando il metodo al modello espresso nella parametrizzazione in λ coincide con la legge che troviamo applicando la trasformazione di funzioni di densità alla distribuzione $\pi_{\Theta}^o(\cdot)$ ottenuta con lo stesso metodo applicato al modello parametrizzato in θ .

Tra le regole invarianti, la più nota è quella di Jeffreys, che si ottiene ponendo

$$\pi_{\Theta}^J(\theta) \propto \sqrt{I_n(\theta)}, \quad (6.4)$$

dove $I_n(\theta) = \mathbb{E}_{\theta} \left[\left(\frac{d}{d\theta} \ln f_n(\mathbf{X}_n|\theta) \right)^2 \right]$ indica l'informazione attesa di Fisher che, nel caso di osservazioni i.i.d. e problemi regolari diventa $I_n(\theta) = -n \mathbb{E}_{\theta} \left(\frac{d^2}{d\theta^2} \ln f_X(X|\theta) \right)$. L'invarianza della regola di Jeffreys discende da quella dell'informazione attesa di Fisher. Si può infatti verificare ² che, se $\theta = g^{-1}(\lambda)$, per l'informazione attesa di Fisher $I_n^{\lambda}(\lambda)$ nella nuova parametrizzazione è legato all'informazione attesa $I_n(\theta)$ nella parametrizzazione originale in θ dalla relazione

$$I_n^{\lambda}(\lambda) = I_n[g^{-1}(\lambda)] \times \left[\frac{d}{d\lambda} g^{-1}(\lambda) \right]^2$$

e che quindi la 6.3 è rispettata.

6.7 Osservazione/i

1. Il metodo può essere utilizzato solo se il modello è regolare e quindi se esiste l'informazione attesa di Fisher. Ad esempio, per il caso del modello uniforme in $[0, \theta]$ dell'Esempio 6.6, non è possibile definire la distribuzione di Jeffreys.
2. Quando Ω non è un intervallo di \mathbb{R} , in genere (ma non sempre) π_{Θ}^J è impropria ma la distribuzione a posteriori risultante è ben definita e quindi utilizzabile per ottenere metodi inferenziali per θ .
3. Nella maggior parte dei modelli più comuni, l'espressione di $\pi^J(\theta)$ si può ottenere come caso particolare (in genere come caso limite) della distribuzione a priori coniugata del modello (vedi esempio anseguire).
4. Nel caso multiparametrico in cui $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$, $k > 1$ la definizione si generalizza ponendo $\pi_{\Theta}^J(\boldsymbol{\theta}) = \sqrt{|I_n(\boldsymbol{\theta})|}$, dove $|I_n(\boldsymbol{\theta})|$ indica il determinante della matrice di informazione di Fisher.

6.8 Esempio (modello di Poisson). Consideriamo il modello di Poisson con

funzione di massa di probabilità

$$f_X(x|\theta) = \frac{\theta^x}{x!} \exp(-\theta), \quad x = 0, 1, 2, \dots, \quad \theta > 0.$$

²Si veda, ad esempio, pag. 121 di Schervish M.J. (1995). *Theory of Statistics*. Springer-Verlag, New York.

In questo caso abbiamo che, per un campione causale di dimensione n , l'informazione attesa è $I_n(\theta) = n/\theta$ e quindi

$$\pi^J(\theta) \propto \frac{1}{\sqrt{\theta}}, \quad \theta > 0$$

è la distribuzione di Jeffreys per θ (impropria). Si osservi che tale distribuzione non coincide con quella considerata nell'Esempio 6.4. Consideriamo ora lo stesso modello nella parametrizzazione $\lambda = g(\theta) = 1/\theta$:

$$f_X(x|\lambda) = \frac{1}{x!\lambda^x} \exp\left(-\frac{1}{\lambda}\right), \quad x = 0, 1, 2, \dots, \quad \lambda > 0.$$

Osservando che

$$\frac{d}{d\lambda} g^{-1}(\lambda) = -\frac{1}{\lambda^2},$$

la distribuzione a priori che troviamo da $\pi^J(\theta)$ con la 6.3 è quindi

$$\pi(\lambda) \propto \lambda^{-3/2}, \quad \lambda > 0$$

che, come la precedente, è una densità impropria. Mostriamo ora che, se applichiamo direttamente la regola 6.4 al modello nella nuova parametrizzazione in λ , otteniamo la stessa legge appena determinata. Infatti è semplice verificare che l'informazione di Fisher nella nuova parametrizzazione è

$$I_n^\lambda(\lambda) = \frac{1}{\lambda^3}$$

e quindi

$$\pi^J(\lambda) \propto \sqrt{I(\lambda)} = \text{cost} \lambda^{-3/2}.$$

La regola di Jeffreys risulta quindi effettivamente invariante per la trasformazione in esame.

6.4 Esempi notevoli

La determinazione dell'espressione di $\pi^J(\theta)$ e di $\pi^J(\theta|\mathbf{x}_n)$ risulta particolarmente semplice per i modelli regolari più comuni, ovvero per le famiglie esponenziali uniparametriche. Per questi casi, infatti, basta ricordare che

$$I_n(\theta) = \text{cr}(\theta)^{-1} = [\mathbb{V}_\theta[\hat{\theta}_u(\mathbf{X}_n)]]^{-1},$$

dove $\hat{\theta}_u(\mathbf{X}_n)$ indica lo stimatore UMVUE di θ e $\text{cr}(\theta)$ il limite inferiore di Cramer-Rao.

6.9 Esempio

1. Modello bernoulliano. In questo caso $\hat{\theta}_u(\mathbf{X}_n) = \bar{X}_n$ e quindi $\text{cr}(\theta) = \mathbb{V}_\theta(\bar{X}_n) = I_n^{-1}(\theta) = \frac{\theta(1-\theta)}{n} \Rightarrow \pi^J(\theta) \propto \frac{1}{\sqrt{\theta(1-\theta)}}$ (propria). Pertanto $\pi^J(\theta|\mathbf{x}_n) = \text{Beta}(\theta|\bar{\alpha}, \bar{\beta})$, con $\bar{\alpha} = s_n + 1/2, \bar{\beta} = n - s_n + 1/2, s_n = \sum_{i=1}^n x_i$. La densità a posteriori $\pi^J(\theta|\mathbf{x}_n)$ coincide quindi con quella ottenuta usando una densità a priori coniugata Beta di parametri per $\alpha = \beta = 1/2$ (propria).
2. Modello di Poisson. In questo caso $\hat{\theta}_u(\mathbf{X}_n) = \bar{X}_n$ e quindi $\text{cr}(\theta) = \mathbb{V}_\theta(\bar{X}_n) = I_n^{-1}(\theta) = \frac{\theta}{n} \Rightarrow \pi^J(\theta) \propto \frac{1}{\sqrt{\theta}}$ (impropria). Pertanto $\pi^J(\theta|\mathbf{x}_n) = \text{Ga}(\theta|\bar{\alpha}, \text{rate} = \bar{\beta})$, con $\bar{\alpha} = s_n + 1/2, \bar{\beta} = n, s_n = \sum_{i=1}^n x_i$. La densità a posteriori $\pi^J(\theta|\mathbf{x}_n)$ coincide quindi con quella che si ottiene usando una densità a priori Beta per $\alpha = 1/2$ e con $\beta \rightarrow 0$.
3. Modello esponenziale. In questo caso $\hat{\theta}_u(\mathbf{X}_n) = \bar{X}_n$ e quindi $\text{cr}(\theta) = \mathbb{V}_\theta(\bar{X}_n) = I_n^{-1}(\theta) = \frac{\theta^2}{n} \Rightarrow \pi^J(\theta) \propto \frac{1}{\theta}$ (impropria). Pertanto $\pi^J(\theta|\mathbf{x}_n) = \text{InvGa}(\theta|\bar{\alpha}, \text{rate} = \bar{\beta})$, con $\bar{\alpha} = n, \bar{\beta} = s_n, s_n = \sum_{i=1}^n x_i$. La densità a posteriori $\pi^J(\theta|\mathbf{x}_n)$ coincide quindi con quella che si ottiene usando una densità a priori InvGa per $\alpha \rightarrow 0$ e $\beta \rightarrow 0$.
4. Modello normale $N(\theta, \sigma^2)$ (N1). Anche in questo caso $\hat{\theta}_u(\mathbf{X}_n) = \bar{X}_n$ e quindi $\text{cr}(\theta) = \mathbb{V}_\theta(\bar{X}_n) = I_n^{-1}(\theta) = \frac{\sigma^2}{n} \Rightarrow \pi^J(\theta) \propto 1$ (impropria). Quindi $\pi^J(\theta|\mathbf{x}_n) = N\left(\bar{x}_n, \frac{\sigma^2}{n}\right)$, La densità a posteriori $\pi^J(\theta|\mathbf{x}_n)$ coincide quindi con quella che si ottiene usando una densità a priori $N(\mu_0, \sigma^2/n_0)$ per $n_0 \rightarrow 0$.
5. Modello normale $N(\mu_0, \theta)$ (N3). Per questo modello $\hat{\theta}_u(\mathbf{X}_n) = S_0^2$ e quindi $\text{cr}(\theta) = \mathbb{V}_\theta(S_0^2) = I_n^{-1}(\theta) = \frac{2\theta^2}{n} \Rightarrow \pi^J(\theta) \propto \frac{1}{\theta}$ (impropria). Si ha allora che $\pi^J(\theta|\mathbf{x}_n) = \text{InvGa}(\theta|\bar{\alpha}, \text{rate} = \bar{\beta})$, con $\bar{\alpha} = \frac{n}{2}, \bar{\beta} = \frac{n}{2}S_0^2$. La densità a posteriori $\pi^J(\theta|\mathbf{x}_n)$ coincide quindi con quella che si ottiene usando una densità a priori InvGa per $\alpha \rightarrow 0$ e $\beta \rightarrow 0$ (impropria).

6.5 Esercizi

1. Sia $X_1, \dots, X_n | \theta \sim \text{Pois}(\theta)$ i.i.d., $\theta > 0$.
 - (a) Determinare $\pi^J(\theta)$ (Jeffreys prior).
 - (b) Verificare che, utilizzando $\pi^J(\theta)$, si ottiene che $\pi^J(\theta | \mathbf{x}_n)$ è una densità Gamma($s_n + 1/2, n$).
 - (c) Determinare $\mathbb{E}[\Theta | \mathbf{x}_n]$ e $\mathbb{V}[\Theta | \mathbf{x}_n]$.
2. Sia $X_1, \dots, X_n | \theta \sim \text{EN}(\theta)$ i.i.d., $\theta > 0$.
 - (a) Determinare $\pi^J(\theta)$ (Jeffreys prior).
 - (b) Verificare che, utilizzando $\pi^J(\theta)$, si ottiene che $\pi^J(\theta | \mathbf{x}_n)$ è ancora una densità gamma e determinarne i parametri.
 - (c) Determinare $\mathbb{E}[\Theta | \mathbf{x}_n]$ e $\mathbb{V}[\Theta | \mathbf{x}_n]$.
3. Sia $\mathbf{X}_n = (X_1, \dots, X_n)$ un campione in cui le v.a X_i sono, condizionatamente a θ , i.i.d. con valore atteso $\mathbb{E}_\theta[X] = 1/\theta$ e funzione di massa di probabilità

$$f_X(x|\theta) = (1 - \theta)^{x-1}\theta, \quad x = 1, 2, 3, \dots \quad \theta \in [0, 1].$$
 - (a) Verificare che la famiglia delle densità Beta(α, β), con $\alpha, \beta > 0$, è coniugata al modello in esame (per v.a. geometriche) e determinare i parametri della distribuzione a posteriori di Θ , $\pi(\theta | \mathbf{x}_n)$.
 - (b) Determinare $\pi^J(\theta)$, la distribuzione a priori di Jeffreys per θ .
 - (c) Determinare la distribuzione a posteriori di Θ che si ottiene utilizzando $\pi^J(\theta)$.
 - (d) Determinare a quali valori devono tendere α e β affinché la distribuzione a posteriori ottenuta con la distribuzione beta coincida con quella ottenuta con $\pi^J(\theta)$.
4. Sia $X_1, \dots, X_n | \theta \sim \text{Ber}(\theta)$ i.i.d.
 - (a) Determinare $\pi^J(\theta)$, distribuzione a priori di Jeffreys per Θ .
 - (b) Determinare la distribuzione a posteriori $\pi^J(\theta | \mathbf{x}_n)$.
 - (c) Utilizzando la regola di trasformazione delle funzioni di densità per trasformazioni monotone di variabili aleatorie, determinare $\pi_\Psi^J(\psi | \mathbf{x}_n)$ per $\psi = g(\theta) = \frac{\theta}{1-\theta}$.
5. Sia $X_1, \dots, X_n | \theta \sim \text{Beta}(\theta, 1)$ i.i.d., con $f_X(x|\theta) = \theta x^{\theta-1}$, $x \in (0, 1)$, $\theta > 0$.
 - (a) Determinare $\pi^J(\theta)$, distribuzione a priori di Jeffreys per Θ .
 - (b) Determinare la densità a posteriori $\pi^J(\theta | \mathbf{x}_n)$.
 - (c) Determinare $\tilde{\theta}$ (moda a posteriori).
6. Sia $X_1, \dots, X_n | \theta \sim \text{N}(\mu_0, \theta)$ i.i.d. (μ_0 noto), $\theta > 0$.

- (a) Determinare $\hat{\theta}_{mv}(\mathbf{x}_n)$ e la sua varianza.
 - (b) Determinare (eseguendo i calcoli necessari) $I_n(\theta)$ e verificare che il suo inverso coincide con la varianza di $\hat{\theta}_{mv}(\mathbf{X}_n)$.
 - (c) Determinare $\pi^J(\theta)$, distribuzione a priori di Jeffreys per Θ .
 - (d) Determinare la distribuzione a posteriori $\pi^J(\theta|\mathbf{x}_n)$.
 - (e) Stabilire a quale famiglia di densità appartiene $\pi^J(\theta|\mathbf{x}_n)$ (dire per quali valori dei parametri).
7. Si considerino i modelli statistici per le seguenti v.a. (a) $\text{Ber}(\theta)$; (b) $\text{Pois}(\theta)$, $\text{Exp}(\theta)$, $\text{N}(\theta, \sigma_0^2)$. Per ciascuno dei modelli considerati determinare quanto richiesto.
- (a) Determinare informazione attesa di Fisher e limite inferiore di Cramer Rao ($\text{cr}(\theta)$) per gli stimatori di massima verosimiglianza di θ , che sono anche stimatori UMVUE dei parametri dei singoli modelli (con varianza esattamente uguale a $\text{cr}(\theta)$).
 - (b) Determinare $\pi^J(\theta)$, la distribuzione a priori di Jeffreys per θ .
 - (c) Determinare la distribuzione a posteriori di θ utilizzando la distribuzione di Jeffreys. Indicare la famiglia coniugata per il modello e a quali valori devono essere uguali o devono tendere gli iperparametri in modo tale che le distribuzioni a posteriori risultanti coincidano con le distribuzioni a posteriori ottenute dalle distribuzioni di Jeffreys.

Capitolo 7

Approssimazioni normali

7.1 Introduzione

Sotto alcune condizioni di regolarità del modello di base e per una dimensione campionaria adeguata, la densità a posteriori di un parametro è ben approssimata da una densità normale. Esistono due approssimazioni normali per una densità: la prima basata sulla sola funzione di verosimiglianza, la seconda dipendente da verosimiglianza e densità a priori. Le approssimazioni normali sono utili, ad esempio, per trovare delle valide approssimazioni per gli insiemi di credibilità.

7.2 Approssimazione normale della funzione di verosimiglianza

Sia $\Theta \subseteq \mathbb{R}$ e siano $L(\theta)$ ed $\ell(\theta)$ le funzioni di verosimiglianza e di log-verosimiglianza associate ad un campione osservato. Supponendo che il supporto di X sia indipendente da θ e che le funzioni $L(\theta)$ ed $\ell(\theta)$ siano derivabili (almeno due volte) rispetto a θ , sia \mathcal{I}_n l'informazione osservata di Fisher (omettiamo per semplicità la dipendenza dal campione osservato). Se $\ell(\theta)$ ammette un punto di massimo $\hat{\theta}_{mv}$ interno a Ω , sfruttando lo sviluppo in serie di Taylor fino al secondo grado intorno a $\hat{\theta}_{mv}$, possiamo mostrare che la funzione di verosimiglianza di un parametro unidimensionale può essere approssimata, in un intorno di $\hat{\theta}_{mv}$, da una funzione proporzionale a una densità $N(\hat{\theta}_{mv}, \mathcal{I}_n^{-1})$. Possiamo infatti scrivere:

$$\ell(\theta) = \ell(\hat{\theta}_{mv}) + (\theta - \hat{\theta}_{mv}) \ell'(\theta)|_{\theta=\hat{\theta}_{mv}} + \frac{1}{2} (\theta - \hat{\theta}_{mv})^2 \ell''(\theta)|_{\theta=\hat{\theta}_{mv}} + R_n$$

dove il resto R_n è una funzione dei dati osservati e del parametro che assumiamo essere tanto più trascurabile quanto più elevata la numerosità campionaria. Dato che $\ell'(\hat{\theta}_{mv}) = 0$, trascurando il resto, si ottiene:

$$\ell(\theta) - \ell(\hat{\theta}_{mv}) = \log \bar{L}(\theta) \approx \frac{1}{2} (\theta - \hat{\theta}_{mv})^2 \ell''(\theta)|_{\theta=\hat{\theta}_{mv}}.$$

Pertanto

$$\bar{L}(\theta; \mathbf{x}_n) \approx \exp \left\{ \frac{1}{2} (\theta - \hat{\theta})^2 \ell''(\theta)|_{\theta=\hat{\theta}} \right\} = \exp \left\{ -\frac{1}{2} (\theta - \hat{\theta}_{mv})^2 \mathcal{I}_n(\hat{\theta}_{mv}) \right\}.$$

L'ultima eguaglianza è dovuta al fatto che $\mathcal{I}_n = -\ell''(\theta)|_{\theta=\hat{\theta}_{mv}}$. Possiamo quindi utilizzare, come approssimazione di $\bar{L}(\theta; \mathbf{x}_n)$, la funzione $\tilde{L} : \Omega \rightarrow [0, 1]$, definita da

$$\tilde{L}(\theta; \mathbf{x}_n) = \exp\left\{-\frac{1}{2}(\theta - \hat{\theta}_{mv})^2 \mathcal{I}_n\right\}, \quad \theta \in \Omega. \quad (7.1)$$

L'approssimazione \tilde{L} per \bar{L} è proporzionale a una densità gaussiana di parametri $(\hat{\theta}_{mv}, \mathcal{I}_n^{-1})$. Il grafico di \tilde{L} in Ω ha quindi la forma tipica di una densità normale. Questo non vuol dire però che \tilde{L} sia interpretabile come se fosse funzione di densità, dal momento che, come si è più volte sottolineato, il parametro θ non è una variabile aleatoria.

Da quanto affermato discende che, nei modelli regolari, l'informazione osservata \mathcal{I}_n controlla sia la curvatura di ℓ in $\hat{\theta}_{mv}$ che la dispersione dell'approssimazione normale intorno allo stesso punto. L'approssimazione normale di \bar{L} è tanto migliore quanto più piccolo è l'intorno di $\hat{\theta}_{mv}$ che si considera e, a parità di altre condizioni, quanto più elevata è la dimensione campionaria.

Insiemi di verosimiglianza approssimati

La (7.1) consente di determinare intervalli di verosimiglianza di livello q approssimati per il parametro di un modello statistico regolare. L'utilità di questo risultato risiede nel fatto che molto spesso, anche per i più comuni modelli, le soluzioni della disequazione che definisce l'insieme L_q non sono determinabili in forma esplicita. Ciò vuol dire che non siamo in grado (o comunque non è agevole) determinare gli estremi dell'insieme di verosimiglianza. In questi casi è quindi necessario ricorrere a soluzioni numeriche. Un'alternativa è appunto rappresentata dagli insiemi di verosimiglianza che si possono ottenere dall'approssimazione normale, definiti come segue:

$$\tilde{L}_q = \left\{ \theta \in \Omega : \tilde{L}(\theta; \mathbf{x}_n) \geq q \right\}, \quad q \in (0, 1).$$

Dalla (7.1) discende infatti che $L_q \approx \tilde{L}_q$ dove

$$\tilde{L}_q = \left\{ \theta \in \Omega : \exp\left\{-\frac{1}{2}(\theta - \hat{\theta}_{mv})^2 \mathcal{I}_n\right\} \geq q \right\}, \quad q \in (0, 1).$$

Con calcoli del tutto analoghi a quelli svolti per determinare l'insieme L_q per il modello $N(\theta, \sigma^2)$, si verifica facilmente che

$$\tilde{L}_q = \left[\hat{\theta}_{mv} - k_q \sqrt{[\mathcal{I}_n(\hat{\theta}_{mv})]^{-1}}, \hat{\theta}_{mv} + k_q \sqrt{[\mathcal{I}_n(\hat{\theta}_{mv})]^{-1}} \right], \quad k_q = \sqrt{-2 \ln q}. \quad (7.2)$$

Si noti per prima cosa che per ogni $q \in (0, 1)$, $-2 \ln q > 0$ e quindi che $k_q \in \mathbb{R}^+$. Inoltre, si osservi che, maggiore è $\mathcal{I}_n(\hat{\theta}_{mv})$ (ovvero l'informazione campionaria), minore è la lunghezza dell'intervallo di verosimiglianza per θ .

La perdita di precisione nel calcolo dell'insieme di verosimiglianza che si ha ricorrendo all'approssimazione \tilde{L}_q risulta compensata, in particolar modo per n adeguatamente elevato, dalla disponibilità di formule esplicite per il calcolo degli estremi degli intervalli di stima.

7.1 Esempio (Modello bernoulliano). Ricordando che, per il modello in esame,

$$\hat{\theta}_{mv} = \bar{x}_n \quad \text{e che} \quad \mathcal{I}_n(\hat{\theta}_{mv}) = \frac{n}{\bar{x}_n(1 - \bar{x}_n)}$$

si ha che

$$\tilde{L}_q = \left[\bar{x}_n - k_q \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}}, \bar{x}_n + k_q \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}} \right].$$

Se ad esempio si considera un campione di dimensione $n = 10$ in cui $\bar{x}_n = 5$ e se si assume che $\sigma^2 = 1$, scegliendo il valore $q = 0.147$ si ottiene

$$\left[5 - \sqrt{-2 \ln(0.147)} \frac{1}{\sqrt{10}}, 5 + \sqrt{-2 \ln(0.147)} \frac{1}{\sqrt{10}} \right] = [4.38, 5.62].$$

7.2 Esempio (Modello di Poisson). Ricordando che, per il modello in esame,

$$\hat{\theta}_{mv} = \bar{x}_n \quad \text{e che} \quad \mathcal{I}_n(\hat{\theta}_{mv}) = \frac{n}{\bar{x}_n}$$

si ha che

$$\tilde{L}_q = \left[\bar{x}_n - k_q \sqrt{\frac{\bar{x}_n}{n}}, \bar{x}_n + k_q \sqrt{\frac{\bar{x}_n}{n}} \right].$$

7.3 Esempio (Modello normale, valore atteso noto). Abbiamo visto in precedenza che, per un modello normale con media nota μ e varianza incognita $\theta = \sigma^2$, si ha:

$$\hat{\theta}_{mv} = \frac{1}{n} \sum (x_i - \mu)^2 \quad \text{e} \quad \mathcal{I}_n(\hat{\theta}_{mv}) = \frac{n}{2\hat{\theta}_{mv}^2}.$$

Gli estremi dell'intervallo \tilde{L}_q per $\theta = \sigma^2$ sono dati da:

$$\hat{\theta}_{mv} \pm k_q \sqrt{[\mathcal{I}_n(\hat{\theta}_{mv})]^{-1}} = \hat{\theta}_{mv} \pm \hat{\theta}_{mv} k_q \sqrt{\frac{2}{n}}, \quad q \in (0, 1).$$

La figura 7.1 mostra la fdv esatta (linea continua) e l'approssimazione normale (linea tratteggiata) nel caso di un campione di piccola dimensione ($n = 10$). L'approssimazione risulta accurata solo per un intorno molto piccolo di $\hat{\theta}_{mv}$. L'approssimazione dell'insieme di verosimiglianza esatta con quello approssimato risulta abbastanza grossolana. Al crescere di n l'approssimazione tende a migliorare sensibilmente.

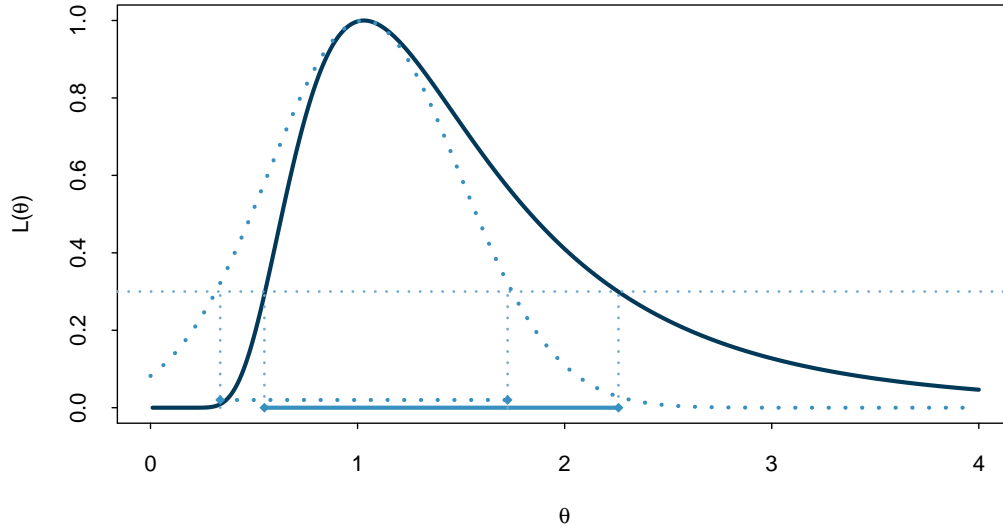


Figura 7.1: Insieme \bar{L}_q e approssimazione \tilde{L}_q per il parametro del modello $N(\mu, \theta)$.

7.3 Approssimazioni normali della distribuzione a posteriori

Prendiamo in considerazione due possibili approssimazioni normali per la distribuzione a posteriori. La prima si basa sull'approssimazione della funzione di verosimiglianza, la seconda sullo sviluppo in serie di Taylor della densità a posteriori stessa.

Approssimazione basata sulla normalità asintotica della funzione di verosimiglianza

La distribuzione a posteriori di θ è proporzionale alla funzione di verosimiglianza. Sotto opportune condizioni di regolarità (vedi Cap. 2), al crescere della numerosità campionaria la funzione di verosimiglianza tende a concentrarsi intorno al suo punto di massimo (la stima di massima verosimiglianza) e ad assumere la forma di una gaussiana. Di conseguenza anche la densità a posteriori tende a una densità normale. Le ipotesi di regolarità sono le seguenti:

- supporto delle v.a. X_i indipendente dal parametro incognito;
- funzione di verosimiglianza derivabile almeno due volte rispetto a θ ;
- stima di massima verosimiglianza di θ è unica e interna al supporto della distribuzione a priori.

Sotto queste condizioni è possibile mostrare che, per numerosità campionaria sufficientemente elevate, la densità posteriori di θ è ben approssimata da una densità normale di con valore atteso e varianza rispettivamente uguali alla stima di massima verosimiglianza e all'inverso dell'informazione osservata di Fisher, ovvero

$$\Theta|\mathbf{x}_n \sim N\left(\hat{\theta}_{mv}, \mathcal{I}_n(\mathbf{x}_n)^{-1}\right)$$

, dove

$$\mathcal{I}_n(\mathbf{x}_n) = -\left(\frac{d^2}{d\theta^2} \ln L(\theta; \mathbf{x}_n)\right)\Big|_{\theta=\hat{\theta}_{mv}}.$$

Possiamo quindi procedere alla stima puntuale, alla stima intervallare e ai test utilizzando le seguenti approssimazioni:

- Stima puntuale: $\hat{\theta}_B(\mathbf{x}_n) \simeq \hat{\theta}_{mv}$;
- Stima intervallare: $\tilde{C}_{1-\gamma}^{hpd}(\mathbf{x}_n) \simeq \left[\hat{\theta}_{mv} - z_{1-\frac{\gamma}{2}} \sqrt{\mathcal{I}_n^{-1}}, \hat{\theta}_{mv} + z_{1-\frac{\gamma}{2}} \sqrt{\mathcal{I}_n^{-1}}\right]$;
- Verifica di ipotesi: $\mathbb{P}[\Theta < \delta|\mathbf{x}_n] \simeq \Phi\left[\frac{\delta - \hat{\theta}_{mv}}{\sqrt{\mathcal{I}_n^{-1}}}\right]$.

Approssimazione normale della densità a posteriori

Supponiamo che $\forall \theta \in \Omega$ la funzione di densità $\pi(\theta|\mathbf{x}_n)$ sia derivabile rispetto a θ almeno due volte e che

$$\tilde{\theta} = \text{Mo}[\Theta|\mathbf{x}_n]$$

sia la moda (unica e interna a Ω) della densità a posteriori. Dallo sviluppo in serie di Taylor di $\pi(\theta|\mathbf{x}_n)$ per i punti di un intorno di $\tilde{\theta}$ possiamo scrivere che:

$$\ln \pi(\theta|\mathbf{x}_n) = \ln \pi(\tilde{\theta}) + (\theta - \tilde{\theta}) \frac{d}{d\theta} \ln \pi(\theta|\mathbf{x}_n)\Big|_{\theta=\tilde{\theta}} + \frac{1}{2}(\theta - \tilde{\theta})^2 \frac{d^2}{d\theta^2} \ln \pi(\theta|\mathbf{x}_n)\Big|_{\theta=\tilde{\theta}} + R_n,$$

dove ora con R_n indichiamo una successione numerica che tende a zero al crescere di n . Osservando che $\frac{d}{d\theta} \ln \pi(\theta|\mathbf{x}_n)\Big|_{\theta=\tilde{\theta}} = 0$ e che $\ln \pi(\tilde{\theta})$ è costante rispetto a θ , abbiamo che, per n sufficientemente elevato),

$$\pi(\theta|\mathbf{x}_n) \simeq \exp\left\{-\frac{1}{2I_B^{-1}}(\theta - \tilde{\theta})^2\right\},$$

dove

$$I_B = -\frac{d^2}{d\theta^2} \ln \pi(\theta|\mathbf{x}_n)\Big|_{\theta=\tilde{\theta}}$$

è denominata *informazione bayesiana*. Si ha quindi che

$$\Theta|\mathbf{x}_n \simeq N(\tilde{\theta}, I_B^{-1}).$$

7.4 Osservazione/i.

1. A differenza dell'approssimazione di verosimiglianza, l'approssimazione bayesiana tiene conto anche della distribuzione a priori di θ e risulta quindi in genere più accurata, a parità di numerosità campionaria.
2. L'interpretazione dell'informazione bayesiana, I_B è simile a quella dell'informazione osservata: I_B misura il grado di concavità (o vvero di concentrazione) della densità $\pi(\theta|\mathbf{x}_n)$ intorno al suo punto di massimo, $\tilde{\theta}$; maggiore il grado di concentrazione, maggiore l'informazione fornita da $\pi(\theta|\mathbf{x}_n)$ sui valori plausibili per il parametro.

Possiamo quindi procedere alla stima puntuale, alla stima intervallare e ai test utilizzando le seguenti approssimazioni:

- Stima puntuale: $\hat{\theta}_B(\mathbf{x}_n) \simeq \tilde{\theta}$;
- Stima intervallare: $\tilde{C}_{1-\gamma}^{hpd}(\mathbf{x}_n) \simeq \left[\tilde{\theta} - z_{1-\frac{\gamma}{2}} \sqrt{I_B^{-1}}, \tilde{\theta} + z_{1-\frac{\gamma}{2}} \sqrt{I_B^{-1}} \right]$;
- Verifica di ipotesi: $\mathbb{P}[\Theta < \delta|\mathbf{x}_n] \simeq \Phi \left[\frac{\delta - \tilde{\theta}}{\sqrt{I_B^{-1}}} \right]$.

7.5 Esempio. (Modello beta-bernoulli). Nell'esempio (5.1) abbiamo verificato che, nel caso di un campione casuale di dimensione n da una popolazione bernoullina a di parametro θ , se si considera una distribuzione a priori coniugata Beta(α, β) si ha che $\pi(\theta|\mathbf{x}_n) = \text{cost} \times \theta^{\bar{\alpha}-1}(1-\theta)^{\bar{\beta}-1}$, e quindi

$$\frac{d}{d\theta} \ln \pi(\theta|\mathbf{x}_n) = \frac{\bar{\alpha} - 1}{\theta} - \frac{\bar{\beta} - 1}{1 - \theta},$$

da cui si ottiene che il punto candidato ad essere punto di massimo per $\ln \pi(\theta|\mathbf{x}_n)$ e per $\pi(\theta|\mathbf{x}_n)$ è $\tilde{\theta} = (\bar{\alpha} - 1)/(\bar{\alpha} + \bar{\beta} - 2)$. Per ottenere I_B si osservi che

$$\frac{d^2}{d\theta^2} \ln \pi(\theta|\mathbf{x}_n) = -\frac{\bar{\alpha} - 1}{\theta^2} - \frac{\bar{\beta} - 1}{(1 - \theta)^2} < 0,$$

il valore $\tilde{\theta}$ è effettivamente un punto di massimo per la densità a posteriori di Θ . Inoltre

$$\begin{aligned} I_B &= \frac{d^2}{d\theta^2} \ln \pi(\theta|\mathbf{x}_n) \\ &= \frac{\bar{\alpha} - 1}{\theta^2} + \frac{\bar{\beta} - 1}{(1 - \theta)^2} \Big|_{\theta=\tilde{\theta}} \\ &= \frac{(\bar{\alpha} + \bar{\beta} - 2)^2}{\bar{\alpha} - 1} + \frac{(\bar{\alpha} + \bar{\beta} - 2)^2}{\bar{\beta} - 1} \\ &= \frac{(\bar{\alpha} + \bar{\beta} - 1)^3}{(\bar{\alpha} - 1)(\bar{\beta} - 1)}. \end{aligned}$$

Abbiamo quindi che

$$\Theta|\mathbf{x}_n \sim N\left(\frac{\bar{\alpha}-1}{\bar{\alpha}+\bar{\beta}-2}, \frac{(\bar{\alpha}-1)(\bar{\beta}-1)}{(\bar{\alpha}+\bar{\beta}-1)^3}\right) \quad (7.3)$$

Si osservi che $\bar{\alpha}$ e $\bar{\beta}$ sono entrambi infiniti dell'ordine di n . Pertanto I_B^{-1} tende a zero come la successione di termine generico n^{-1} . Dall'approssimazione normale della densità a posteriori di Θ possiamo trovare agevolmente formule approssimate per stime puntuali, stime intervallari e probabilità di ipotesi.

7.6 Esempio. (Modello poisson-gamma). Poichè, in questo caso, $\pi(\theta|\mathbf{x}_n) = \text{cost} \cdot \theta^{\bar{\alpha}-1} \exp\{-\bar{\beta}\theta\}$, allora

$$\begin{aligned} \frac{d}{d\theta} \log \pi(\theta|\mathbf{x}_n) &= \frac{d}{d\theta} [\log \text{cost} + (\bar{\alpha}-1) \log \theta - \bar{\beta}\theta] \\ &= -\frac{\bar{\alpha}-1}{\theta} - \bar{\beta}. \end{aligned}$$

Il punto di massimo candidato di $\pi(\theta|\mathbf{x}_n)$ è quindi

$$\tilde{\theta} = \frac{\bar{\alpha}-1}{\bar{\beta}}.$$

Per ogni $\theta > \tilde{\theta}$ si ha che

$$\frac{d^2}{d\theta^2} \log \pi(\theta|\mathbf{x}_n) = -\frac{\bar{\alpha}-1}{\theta^2} < 0$$

Pertanto $\tilde{\theta}$ è un punto di massimo per la densità a posteriori e

$$I_B = \frac{\bar{\alpha}-1}{\theta^2} \Big|_{\tilde{\theta}} = \frac{\bar{\beta}^2}{\bar{\alpha}-1}$$

e

$$\Theta|\mathbf{x}_n \sim N\left(\frac{\bar{\alpha}-1}{\bar{\beta}}, \frac{\bar{\alpha}-1}{\bar{\beta}^2}\right). \quad (7.4)$$

Di nuovo, dall'approssimazione normale della densità a posteriori di Θ possiamo trovare agevolmente formule approssimate per stime puntuali, stime intervallari e probabilità di ipotesi.

7.7 Esempio (Modello normale-gamma inversa) Nell'Esempio 5.10 abbiamo verificato che, se X_1, \dots, X_n è un campione casuale con $X_i|\theta \sim N(\mu_0, \theta)$, $\theta \in \mathbb{R}^+$ (μ_0 noto) e se si considera per per la v.a. Θ la densità *gamma-inversa* di parametri α e β , si ottiene che $\Theta|\mathbf{x}_n \sim \text{InvGa}(\bar{\alpha}, \bar{\beta})$, con $\bar{\alpha} = \alpha + n/2$ e $\bar{\beta} = \beta + nS_0^2/2$ e con funzione di densità $\pi(\theta|\mathbf{x}_n) = \bar{\beta}^{\bar{\alpha}}/\Gamma(\bar{\alpha})\theta^{-(\bar{\alpha}+1)}e^{-\bar{\beta}/\theta}$. Abbiamo inoltre verificato che

$$\frac{d}{d\theta} \ln \pi(\theta|\mathbf{x}_n) = -\frac{\bar{\alpha}+1}{\theta} + \frac{\bar{\beta}}{\theta^2},$$

e che

$$\frac{d^2}{d\theta^2} \ln \pi(\theta|\mathbf{x}_n) \Big|_{\theta=\tilde{\theta}} = -\frac{(\alpha+1)^3}{\beta^2} < 0,$$

Pertanto

$$\tilde{\theta} = \bar{\beta}/(\bar{\alpha}+1) \quad \text{e} \quad I_B = \frac{(\alpha+1)^3}{\bar{\beta}^2},$$

e quindi

$$\Theta|\mathbf{x}_n \sim N\left(\frac{\bar{\beta}}{\bar{\alpha}+1}, \frac{\bar{\beta}^2}{(\bar{\alpha}+1)^3}\right).$$

Anche in questo caso la varianza a posteriori tende a zero al crescere di n e anche in questo caso possiamo trovare procedure inferenziali per il parametro incognito basate sull'approssimazione asintotica della densità a posteriori.

7.4 Esercizi

1. Sia $\mathbf{X}_n = (X_1, \dots, X_n)$ un campione in cui le v.a X_i sono, condizionatamente a θ , i.i.d. con distribuzione $\text{EN}(\theta)$. Si consideri inoltre per Θ la famiglia delle densità $\text{Gamma}(\alpha, \beta)$ coniugata al modello, $\alpha, \beta > 0$.

- (a) Determinare le distribuzioni a posteriori di Θ e della v.a. $\Lambda = \Theta^{-1}$.
 (b) Verificare che

$$\mathbb{E}[\Lambda|\mathbf{x}_n] = \frac{\beta + s_n}{\alpha + n - 1}, \quad \text{dove} \quad s_n = \sum_{i=1}^n x_i.$$

- (c) Calcolare l'informazione bayesiana I_B e verificare che (asintoticamente)

$$\Theta|\mathbf{x}_n \sim \text{N}\left(\frac{\bar{\alpha} - 1}{\bar{\beta}}, \frac{\bar{\alpha} - 1}{\bar{\beta}^2}\right). \quad (7.5)$$

- (d) Utilizzando (7.5), determinare le generiche espressioni di $C_{1-\gamma}(\mathbf{x}_n)$, insieme HPD di livello $1 - \gamma$ per θ , e della probabilità $\mathbb{P}[\Theta > \delta|\mathbf{x}_n]$.
 (e) Utilizzando (7.5) e assumendo $\alpha = 2$, $\beta = 3$, $n = 20$, $s_n = 2$, $\gamma = 0.05$ e $\delta = 5$, calcolare i valori numerici di $\mathbb{E}[\Theta|\mathbf{x}_n]$, $C_{1-\gamma}(\mathbf{x}_n)$ e $\mathbb{P}[\Theta > \delta|\mathbf{x}_n]$.

2. Sia $X_1, \dots, X_n|\theta \sim \text{Pois}(\theta)$ e $\Theta \sim \text{Gamma}(\alpha, \beta)$.

- (a) Determinare la distribuzione a posteriori di Θ , valore atteso e varianza.
 (b) Determinare l'espressione dell'insieme ET (equal-tails) per Θ di livello $1 - \gamma$.
 (c) Calcolare esplicitamente $\tilde{\theta}$, il punto di massimo (moda) della densità a posteriori.
 (d) Determinare l'informazione bayesiana I_B e l'approssimazione normale per la densità a posteriori di Θ .
 (e) Determinare l'espressione generica della probabilità che Θ appartenga all'intervallo $[\theta_L, \theta_U]$ (usare approssimazione normale).
 (f) Determinare \tilde{C} , l'insieme HPD di livello γ per θ (usare approssimazione normale).
 (g) Sapendo che $\alpha = 3$, $\beta = 4$, $n = 20$, $\sum_{i=1}^n x_i = 10$, $\gamma=0.95$, calcolare $\tilde{\theta}$, I_B e gli estremi dell'insieme \tilde{C} .
 (h) Determinare l'espressione della funzione di densità di $\Psi = \Theta^{-1}$.

3. Sia $X_1, \dots, X_n|\theta \sim \text{N}(\mu_0, \theta)$ (μ_0 noto) e $\Theta \sim \text{InvGa}(\alpha, \beta)$. Determinare quanto richiesto di seguito.

- (a) Distribuzione a posteriori di Θ .

- (b) Espressione della funzione di densità, il valore atteso e la varianza di $\Psi = \Theta^{-1}$.
 - (c) Il punto di massimo (moda) $\tilde{\theta}$ della densità a posteriori di Θ (calcolare esplicitamente $\tilde{\theta}$ utilizzando la densità a posteriori di Θ).
 - (d) Informazione bayesiana I_B e l'approssimazione normale per la densità a posteriori di Θ .
 - (e) Espressione generica della probabilità che Θ sia minore di θ_0 (usare approssimazione normale e assumere che $\theta_0 > 0$).
 - (f) Espressione di \tilde{C} , l'insieme HPD di livello $1 - \gamma$ per θ (usare approssimazione normale).
4. Siano $X_1, \dots, X_n | \theta \sim \text{Geom}(\theta)$ i.i.d., con valore atteso $\mathbb{E}_\theta[X] = \frac{1}{\theta}$, $\mathbb{V}_\theta[X] = \frac{1-\theta}{\theta^2}$ e funzione di massa di probabilità

$$f_X(x|\theta) = \theta(1-\theta)^{x-1}, \quad x = 1, 2, 3, \dots \quad \theta \in [0, 1].$$

Per la distribuzione a priori considerare la famiglia delle densità gamma di parametri α e β .

- (a) Determinare l'espressione di $\tilde{\theta}$ e di $I_B(z)$.
 - (b) Verificare che, per n sufficientemente elevato, si ha che

$$\Theta | z \sim \text{N} \left(\frac{\bar{\alpha} - 1}{\bar{\alpha} + \bar{\beta} - 2}, \frac{(\bar{\alpha} - 1)(\bar{\beta} - 1)}{(\bar{\alpha} + \bar{\beta} - 2)^3} \right).$$
 - (c) Basandosi sull'approssimazione normale della distribuzione a posteriori, determinare l'espressione di $\tilde{C}(\mathbf{x}_n)$ di livello $1 - \gamma$, insieme HDP per θ .
 - (d) Basandosi sull'approssimazione normale della distribuzione a posteriori, calcolare l'espressione della probabilità che $\Theta < 1/2$, in funzione di $\Phi(\cdot)$, funzione di ripartizione della v.a. $\text{N}(0, 1)$.
 - (e) Supponendo che il campione abbia dimensione $n = 20$, che $\alpha = \beta = 2$ e che $\sum_{i=1}^n x_i = 50$, determinare gli estremi di $\tilde{C}(\mathbf{x}_n)$ (assumere $\gamma = 0.05$).
 - (f) Con gli stessi dati del punto precedente, calcolare $\mathbb{P}[\Theta < 1/2 | \mathbf{x}_n]$.
5. Sia $X_1, \dots, X_n | \theta \sim \text{Beta}(\theta, 1)$ i.i.d., con $f_X(x|\theta) = \theta x^{\theta-1}$, $x \in (0, 1)$, $\theta > 0$.
- (a) Determinare $\pi^J(\theta)$, distribuzione a priori di Jeffreys per Θ e la corrispondente densità a posteriori $\pi^J(\theta | \mathbf{x}_n)$.
 - (b) Determinare $\tilde{\theta}$ (moda a posteriori) e \tilde{I}_B (informazione bayesiana).
 - (c) Scrivere l'espressione dell'insieme *equal-tails* (ET) di livello $1 - \gamma$ per θ , basato sull'approssimazione normale della distribuzione a posteriori di Θ .
6. Sia $X_1, \dots, X_n | \theta \sim \text{Pois}(\theta)$ i.i.d., $\theta > 0$.
- (a) Determinare $\pi^J(\theta | \mathbf{x}_n)$, ovvero la densità a posteriori ottenuta con distribuzione a priori di Jeffreys. Determinare l'approssimazione normale per $\pi^J(\theta | \mathbf{x}_n)$.

Capitolo 8

Metodi per inferenze predittive

8.1 Introduzione

Nei capitoli precedenti abbiamo parlato di problemi ipotetici, ovvero problemi di inferenza sul parametro incognito di un modello. Quando l'oggetto di inferenza è costituito dai dati futuri, o da una loro funzione, parliamo di problemi predittivi. In questo capitolo introduciamo le distribuzioni predittive a priori e a posteriori ed esaminiamo i principali problemi di inferenza predittiva per funzioni unidimensionali di dati futuri, quali sono, ad esempio, somma e media campionaria.

8.2 Distribuzioni predittive

Nei problemi predittivi l'oggetto di interesse inferenziale è rappresentato da dati futuri $\mathbf{X}_{n_f} = (X_i^f, \dots, X_{n_f}^f)$ - un campione di dimensione n_f - o da opportune funzioni di dati futuri, $Y_f = T(\mathbf{X}_{n_f})$. A differenza dei problemi ipotetici, nel caso predittivo l'inferenza riguarda quindi v.a. osservabili. Per formalizzare, consideriamo un esperimento non realizzato

$$\{\mathcal{X}^{n_f}, f_{n_f}(\cdot|\theta), \theta \in \Omega\},$$

in cui il parametro incognito è lo stesso dell'esperimento corrente $\{\mathcal{X}^n, f_n(\cdot|\theta), \theta \in \Omega\}$. Dal punto di vista probabilistico dobbiamo considerare una variabile aleatoria tripla

$$(\Theta, \mathbf{X}_n, \mathbf{X}_{n_f})$$

e assumiamo che, fissato $\Theta = \theta$, le v.a. \mathbf{X}_n e \mathbf{X}_{n_f} siano indipendenti. Per semplificare la notazione omettiamo qui il suffisso per le leggi di probabilità delle variabili aleatorie coinvolte: l'argomento delle leggi stesse chiarisce quali sono le variabili aleatorie coinvolte. La legge congiunta di $(\Theta, \mathbf{X}_n, \mathbf{X}_{n_f})$ è quindi data da

$$f(\theta, \mathbf{x}_n, \mathbf{x}_{n_f}) = f(\mathbf{x}_n, \mathbf{x}_{n_f}|\theta) \cdot \pi(\theta) = f(\mathbf{x}_n|\theta) \cdot f(\mathbf{x}_{n_f}|\theta) \cdot \pi(\theta),$$

dove per l'ultima uguaglianza si deve tenere conto che, per l'indipendenza tra \mathbf{X}_n e \mathbf{X}_{n_f} condizionatamente a θ si ha

$$f(\mathbf{x}_n, \mathbf{x}_{n_f}|\theta) = f(\mathbf{x}_{n_f}|\mathbf{x}_n, \theta) \cdot f(\mathbf{x}_n|\theta) = f(\mathbf{x}_{n_f}|\theta) \cdot f(\mathbf{x}_n|\theta).$$

I problemi in esame si affrontano utilizzando le distribuzioni marginali (rispetto a Θ) delle quantità di interesse, ovvero le distribuzioni dei dati futuri \mathbf{X}_{n_f} o di loro funzioni $T(\mathbf{X}_{n_f})$. Per semplicità assumiamo che $\Omega \subseteq \mathbb{R}$ e che Θ sia dotata di funzione di densità. Nello schema che stiamo considerando, se abbiamo a disposizione i dati dell'esperimento \mathbf{x}_n , possiamo considerare la legge (funzione di massa o di densità di probabilità) *predittiva a posteriori* (o *finale*) $m(\mathbf{x}_{n_f}|\mathbf{x}_n)$, ovvero la distribuzione marginale (rispetto a Θ) di $\mathbf{X}_{n_f}|\mathbf{X}_n = \mathbf{x}_n$

$$m(\mathbf{x}_{n_f}|\mathbf{x}_n) = \int_{\Omega} f(\mathbf{x}_{n_f}, \theta|\mathbf{x}_n) d\theta = \int_{\Omega} f(\mathbf{x}_{n_f}|\mathbf{x}_n, \theta) \pi(\theta|\mathbf{x}_n) d\theta = \int_{\Omega} f(\mathbf{x}_{n_f}|\theta) \pi(\theta|\mathbf{x}_n) d\theta, \quad (8.1)$$

dove l'ultima uguaglianza dipende dal fatto che

$$f(\mathbf{x}_{n_f}|\mathbf{x}_n, \theta) = f(\mathbf{x}_{n_f}|\theta).$$

I problemi predittivi vengono quindi affrontati in modo analogo a quelli ipotetici, utilizzando la distribuzione $m(\mathbf{x}_{n_f}|\mathbf{x}_n)$ al posto di $\pi(\theta|\mathbf{x}_n)$.

In assenza dei dati \mathbf{x}_n , la previsione su \mathbf{X}_{n_f} si effettua sfruttando la distribuzione predittiva a priori, la cui legge di probabilità è $m(\mathbf{x}_{n_f})$.

$$m(\mathbf{x}_{n_f}) = \int_{\Omega} f(\mathbf{x}_{n_f}|\theta) \pi(\theta) d\theta. \quad (8.2)$$

La distribuzione predittiva a posteriori va quindi considerata come aggiornamento della distribuzione *predittiva a priori* che si ha alla luce dei dati \mathbf{x}_n .

8.1 Osservazione/i

1. La distribuzione predittiva a priori $m(\mathbf{x}_{n_f})$ è una mistura delle distribuzioni campionarie $f(\mathbf{x}_{n_f}|\theta)$ di \mathbf{X}_n rispetto alla distribuzione a priori $\pi(\theta)$; la distribuzione predittiva a posteriori $m(\mathbf{x}_{n_f}|\mathbf{x}_n)$ è una mistura delle distribuzioni campionarie $f(\mathbf{x}_{n_f}|\theta)$ di \mathbf{X}_{n_f} rispetto alla distribuzione a posteriori $\pi(\theta|\mathbf{x}_n)$.
2. La distribuzione predittiva a priori $m(\mathbf{x}_{n_f}|\mathbf{x}_n)$ consente di tenere conto della variabilità di Θ nella distribuzione dei dati, che invece viene ignorata dalla distribuzione campionaria $f(\mathbf{x}_{n_f}|\theta)$.
3. La distribuzione predittiva a posteriori $m(\mathbf{x}_{n_f})$ consente di tenere conto della variabilità di Θ nella distribuzione dei dati, ignorata dalla distribuzione campionaria $f(\mathbf{x}_{n_f}|\theta)$, e delle informazioni che i dati \mathbf{x}_n forniscono sul parametro.
4. Nel caso di analisi coniugata, poichè $\pi(\theta)$ e $\pi(\theta|\mathbf{x}_n)$ appartengono alla stessa famiglia di distribuzioni, l'integrale che definisce $m(\mathbf{x}_{n_f})$ e quello che definisce $m(\mathbf{x}_{n_f}|\mathbf{x}_n)$ sono strutturalmente uguali (vedi gli esempi che seguono).
5. Quasi sempre le distribuzioni oggettive $\pi^o(\theta)$ sono improprie. In questi casi le corrispondenti distribuzioni predittiva iniziale non sono definite in quanto l'integrale della funzione $f(\mathbf{x}_n|\theta)\pi^o(\theta)$ in Ω non esiste. Tuttavia, se la distribuzione a posteriori $\pi^o(\theta|\mathbf{x}_n)$ esiste, risulta ben definita anche la distribuzione $m(\mathbf{x}_{n_f}|\mathbf{x}_n)$

6. È importante distinguere tra *indipendenza* (o *indipendenza marginale*) delle variabili aleatorie X_1, \dots, X_n e *indipendenza condizionata a θ* delle stesse. L'indipendenza (marginale) di X_1, \dots, X_n riguarda, appunto, la distribuzione marginale delle X_i : si ha se e solo se $m(\mathbf{x}_n) = \prod_{i=1}^n m_i(x_i)$, dove $m_i(x_i) = \int_{\Omega} f(x_i|\theta)\pi(\theta)d\theta$ indica la distribuzione marginale della v.a. unidimensionale X_i . L'indipendenza condizionata a θ riguarda invece la distribuzione campionaria $f(x_1, \dots, x_n|\theta)$ e si ha se e solo se $f(\mathbf{x}_n) = \prod_{i=1}^n f(x_i|\theta)$, dove $f(\cdot|\theta)$ è la distribuzione campionaria (dipendente da θ) della variabile unidimensionale X_i . Si osservi che, in generale, l'indipendenza condizionata non implica l'indipendenza marginale delle X_i .
7. Le v.a. X_1, \dots, X_n si dicono *scambiabili* se

$$m(\mathbf{x}_n) = m(x_1, \dots, x_n) = \int_{\Omega} \prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta,$$

ovvero se le X_1, \dots, X_n sono indipendenti e identicamente distribuite condizionatamente a θ , tutte con funzione di densità $f(\cdot|\theta)$.

8.3 Inferenza su funzioni scalari di dati futuri

In genere siamo interessati a effettuare inferenza su funzioni scalari $Y_f = T(\mathbf{X}_{n_f})$ di dati futuri, quali la somma o oppure la media aritmetica, rispettivamente definite da $\sum_{i=1}^{n_f} X_i^f$ e da \bar{X}_{n_f} . In questo caso si utilizza la legge

$$m(y_f|\mathbf{x}_n) = \int_{\Omega} f(y_f|\theta)\pi(\theta|\mathbf{x}_n)d\theta,$$

dove $f(y_f|\theta)$ rappresenta la distribuzione campionaria di Y_f e $\pi(\theta|\mathbf{x}_n)$ è la densità a posteriori di θ . In assenza dei dati \mathbf{x}_n si usa la distribuzione predittiva a priori di Y_f :

$$m(y_{n_f}) = \int_{\Omega} f(y_{n_f}|\theta)\pi(\theta)d\theta.$$

Possiamo quindi procedere in modo del tutto analogo a quanto fatto per l'inferenza ipotetica: le stesse elaborazioni di sintesi che venivano effettuate sulla densità $\pi(\theta|\mathbf{x}_n)$ possono ora essere effettuate su $m(y_f|\mathbf{x}_n)$. Di seguito riportiamo i metodi inferenziali basati sulla distribuzione predittiva a posteriori; analoghe elaborazioni possono essere effuate con la disistribuzione predittiva a priori.

Stima puntuale

Per la stima puntuale di Y_f si considerano le usuali sintesi di una funzione di densità, in questo caso di $m(y_f|\mathbf{x}_n)$:

- $\mathbb{E}[Y_f|\mathbf{x}_n] = \int_{\Omega} y_f m(y_f|\mathbf{x}_n) dy_f$ valore atteso di $m(y_f|\mathbf{x}_n)$

- $\mathbb{M}_0[Y_f|\mathbf{x}_n] = \arg \max_{\mathbf{x}_n \in m(y_f|\mathbf{x}_n)}$ moda di $m(y_f|\mathbf{x}_n)$
- $\mathbb{M}_e[Y_f|\mathbf{x}_n] = q_m(0.5|\mathbf{x}_n)$ mediana di $m(y_f|\mathbf{x}_n)$,

dove $q_m(\epsilon|\mathbf{x}_n)$ indica il quantile di livello ϵ della densità predittiva $m(y_f|\mathbf{x}_n)$.

Stima intervallare

Per la stima intervallare di Y_f si possono considerare insiemi di previsione equal-tail (ET) oppure di massima densità predittiva finale (HD=highest density):

- $C_{1-\gamma}^{et} = [q_m(\gamma/2|\mathbf{x}_n), q_m(1 - \gamma/2|\mathbf{x}_n)]$;
- $C_{1-\gamma}^{hd} = \{y_f : m(y_f|\mathbf{x}_n) \geq h\}$ con $\int_{C_{1-\gamma}^{hd}} = 1 - \gamma$.

Verifica di ipotesi

Per la scelta tra le ipotesi $H_0 : Y_f \leq \delta$ vs. $H_1 : Y_f > \delta$ si considerano le probabilità predittive a posteriori:

- $\mathbb{P}[Y_f \leq \delta] = F_m(\delta|\mathbf{x}_n)$ e $\mathbb{P}[Y_f > \delta] = 1 - F_m(\delta|\mathbf{x}_n)$,

dove $F_m(\cdot|\mathbf{x}_n)$ rappresenta la funzione di ripartizione predittiva a posteriori di Y_f .

8.2 Osservazione/i Per le elaborazioni che seguono si sfruttano ampiamente le formule di calcolo di valore atteso e varianza della distribuzione marginale di Y_f . Nel caso di valore atteso e varianza rispetto alla distribuzione predittiva a priori abbiamo che

$$\mathbb{E}_m[Y_f] = \mathbb{E}_\Theta[\mathbb{E}_{Y_f|\theta}(Y_f|\theta)], \quad (8.3)$$

e

$$\mathbb{V}_m[Y_f] = \mathbb{V}_\Theta[\mathbb{E}_{Y_f|\theta}(Y_f|\theta)] + \mathbb{E}_\Theta[\mathbb{V}_{Y_f|\theta}(Y_f|\theta)], \quad (8.4)$$

In modo analogo si calcolano valore atteso e varianza di Y_f rispetto alla distribuzione predittiva a posteriori: in questo caso \mathbb{E}_Θ e \mathbb{V}_Θ sono sostituiti rispettivamente da $\mathbb{E}_{\Theta|\mathbf{x}_n}$ e $\mathbb{V}_{\Theta|\mathbf{x}_n}$.

8.4 Esempi notevoli

8.3 Esempio (modello normale). Consideriamo un campione casuale di dimensione n da v.a. $N(\theta, \sigma^2)$ (con σ^2 noto). Indichiamo inoltre con $N(u|m, v)$ la funzione di densità di una v.a. normale di parametri (m, v) calcolata in u e assumiamo che $\pi(\theta) = N(\theta|\mu_0, \sigma^2/n_0)$, $n_0 > 0$. Sappiamo allora che $\pi(\theta|\mathbf{x}_n) = N(\theta|\bar{\mu}, \bar{\sigma}^2)$, $\bar{\mu} = (n_0\mu_0 + n\bar{x}_n)/(n_0 + n)$, $\bar{\sigma}^2 = \sigma^2/(n_0 + n)$. Supponiamo di voler effettuare previsione sulla media aritmetica \bar{X}_{n_f} delle osservazioni di un campione casuale futuro

di dimensione n_f . Ricordiamo che $\bar{X}_{n_f}|\theta \sim N(\theta, \sigma^2/n_f)$ e poniamo $Y_f = \bar{X}_{n_f}$ e $y_f = \bar{x}_{n_f}$. Abbiamo allora che

$$\begin{aligned} m(y_f) &= \int_{\mathbb{R}} f(y_f|\theta)\pi(\theta)d\theta \\ &= \int_{\mathbb{R}} N(y_f|\theta, \sigma^2/n_f) \cdot N(y_f|\mu_0, \sigma^2/n_0)d\theta \\ &= N\left(y_f\left|\mu_0, \sigma^2\left[\frac{1}{n_0} + \frac{1}{n_f}\right]\right.\right). \end{aligned}$$

Il risultato precedente si ottiene poichè, in generale, per $y, c \in \mathbb{R}$ e $b, d \in \mathbb{R}^+$ si ha

$$\int_{\mathbb{R}} N(x|y, b) N(y|c, d) dy = N(x|c, b + d).$$

Analogamente, dato un campione \mathbf{x}_n di dati osservati relativi a un esperimento realizzato, è semplice verificare che

$$Y_f|\mathbf{x}_n \sim N(\mu_{pr}, \sigma_{pr}^2), \quad \text{con} \quad \mu_{pr} = \bar{\mu} \quad \text{e} \quad \sigma_{pr}^2 = \sigma^2 \left[\frac{1}{n_0 + n} + \frac{1}{n_f} \right].$$

Per l'inferenza a posteriori su Y_f abbiamo quindi:

1. stima puntuale: $\mu_{pr} = \frac{n_0\mu_0 + n\bar{x}_n}{n_0 + n}$;
2. stima intervallare: $\mu_{pr} \pm z_{1-\frac{\gamma}{2}}\sigma_{pr}$ (intervallo di credibilità predittiva di livello $1 - \gamma$);
3. probabilità a posteriori: $\mathbb{P}[Y_f < \delta] = \Phi\left(\frac{\delta - \mu_{pr}}{\sigma_{pr}}\right)$.

Caso non informativo Si ottiene per $n_0 = 0$. In questo caso $m(y_f)$ non esiste mentre $\mu_{pr} = \bar{x}_n$, $\sigma_{pr}^2 = \sigma^2\left(\frac{1}{n_f} + \frac{1}{n}\right)$ e quindi $m(y_f|\mathbf{x}_n) = N\left(y_f|\bar{x}_n, \sigma^2\left[\frac{1}{n_f} + \frac{1}{n}\right]\right)$, da cui si ottiene, per l'inferenza su Y_f : (1) stima puntuale: \bar{x}_n ; (2) stima intervallare: $\bar{x}_n \pm z_{1-\frac{\gamma}{2}}\sigma\sqrt{\frac{1}{n_f} + \frac{1}{n}}$; (3) probabilità a posteriori: $\mathbb{P}[Y_f < \delta] = \Phi\left(\sqrt{\frac{n_f n}{n_f + n}}\frac{\delta - \bar{x}_n}{\sigma}\right)$.

8.4 Osservazione/i. In presenza di dati passati, la più elementare distribuzione predittiva proposta in ambito frequentista si ottiene dalla distribuzione campionaria di $Y_f|\theta$ con un metodo *plug-in*, ovvero sostituendo a θ una sua stima puntuale basata su \mathbf{x}_n ; in questo caso si pone $\theta = \mathbf{x}_n$ nella distribuzione campionaria di Y_f e si usa a fini predittivi la densità $N\left(y_f|\bar{x}_n, \frac{\sigma^2}{n_f}\right)$. Questa distribuzione ha il difetto riconosciuto di sottostimare la variabilità di Y_f , dal momento che non tiene conto della variabilità campionaria associata alla stima \bar{x}_n di θ . È utile confrontare questa densità con la densità predittiva a posteriori che si ottiene nel caso non informativo, ovvero con la densità $N\left(y_f|\bar{x}_n, \sigma^2\left[\frac{1}{n_f} + \frac{1}{n}\right]\right)$. Quest'ultima densità ha varianza

superiore di quella che si ottiene con il metodo *plug-in*. Il metodo bayesiano non-informativo incorpora infatti in modo naturale nella varianza della densità predittiva la variabilità associata ai dati campionari, pari a σ^2/n , al contrario del metodo *plug-in* che, come abbiamo detto, sottostima la variabilità di Y_f .

Applicazione: esame diagnostico. Lesaffre e Lawson (2012) propongono un esempio per l'uso del metodo predittivo bayesiano che si basa su una ricerca di Total et al. (2003). Il problema consiste nel determinare un intervallo di valori normali per

8.5 Esempio (modello bernoulliano). Consideriamo un campione casuale di dimensione n da v.a. bernoulliana di parametro θ e $\pi(\theta) = \text{Beta}(\theta|\alpha, \beta)$, $\alpha, \beta > 0$. Sappiamo allora che $\pi(\theta|\mathbf{x}_n) = \text{Beta}(\theta|\bar{\alpha}, \bar{\beta})$, $\bar{\alpha} = \alpha + y_n$, $\bar{\beta} = \beta + n - y_n$, dove $y_n = \sum_{i=1}^n x_i$. Supponiamo di essere interessati ad effettuare previsione su $Y_f = \sum_{i=1}^{n_f} X_i^f$, somma delle osservazioni di un campione casuale futuro di dimensione n_f . Ricordando che $Y_f|\theta \sim \text{Binom}(n_f, \theta)$ e ricorrendo all'integrale Beta, per la distribuzione predittiva a priori abbiamo che

$$\begin{aligned} m(y_f) &= \int_0^1 f(y_f|\theta)\pi(\theta)d\theta \\ &= \int_0^1 \binom{n_f}{y_f} \theta^{y_f} (1-\theta)^{n_f-y_f} \times \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta \\ &= \frac{1}{B(\alpha, \beta)} \int_0^1 \binom{n_f}{y_f} \theta^{\alpha+y_f-1} (1-\theta)^{\beta+n_f-y_f-1} d\theta \\ &= \binom{n_f}{y_f} \frac{B(\alpha+y_f, \beta+n_f-y_f)}{B(\alpha, \beta)} \quad y_f = 0, 1, \dots, n_f. \end{aligned}$$

Si tratta della funzione di massa di probabilità di una v.a. Beta-Binomiale di parametri (n_f, α, β) ; possiamo quindi scrivere che

$$Y_f \sim \text{BB}(n_f, \alpha, \beta)$$

Il valore atteso predittivo a priori di Y_f si calcola con la (8.3):

$$\mathbb{E}_m[Y_f] = \mathbb{E}_\Theta[\mathbb{E}_{Y_f|\theta}(Y_f|\theta)] = \mathbb{E}_\Theta[n_f\Theta] = \frac{n_f\alpha}{\beta}.$$

In modo analogo si può ottenere la varianza predittiva a priori.

La distribuzione predittiva a posteriori si ottiene da quella appena determinata, sostituendo (α, β) con $(\bar{\alpha}, \bar{\beta})$, ovvero:

$$Y_f|\mathbf{x}_n \sim \text{BB}(n_f, \bar{\alpha}, \bar{\beta}), \quad y_f = 0, 1, \dots, n_f$$

Il valore atteso predittivo a posteriori di Y_f sarà quindi:

$$\mathbb{E}_m[Y_f|\mathbf{x}_n] = \mathbb{E}_{\Theta|\mathbf{x}_n}[\mathbb{E}_{Y_f|\theta}(Y_f|\theta)] = \mathbb{E}_{\Theta|\mathbf{x}_n}[n_f\Theta|\mathbf{x}_n] = \frac{n_f\bar{\alpha}}{\bar{\beta}} = \frac{n_f(\alpha + y_n)}{\alpha + \beta + n}.$$

Casi non informativi Se consideriamo il caso $\alpha = \beta = 1$ (corrispondente alla distribuzione uniforme in $[0, 1]$), è possibile determinare sia la distribuzione predittiva a priori che quella a posteriori. Se invece si considera la distribuzione a priori che si ottiene per $\alpha = \beta = 0$, $\pi(\theta)$ perde di significato mentre abbiamo che $Y_f | \mathbf{x}_n \sim \text{BB}(n_f, y_f, n)$. In questo caso $\mathbb{E}[Y_f | \mathbf{x}_n] = n_f y_n / n = n_f \bar{x}_n$: la stima bayesiana oggettiva della somma di n_f dati futuri coincide quindi una stima del tutto intuitiva di Y_f .

8.6 Osservazione/i Per le applicazioni numeriche è utile sapere che le funzioni R per la distribuzione beta-binomiale si trovano nel pacchetto `Betabinom` della libreria `VGAM`.

8.7 Esempio (modello Poisson). Consideriamo un campione casuale di dimensione n da v.a. di Poisson di parametro θ e $\pi(\theta) = \text{Ga}(\theta | \alpha, \text{rate} = \beta)$, $\alpha, \beta > 0$. Sappiamo allora che $\pi(\theta) = \text{Ga}(\theta | \bar{\alpha}, \text{rate} = \bar{\beta})$, $\bar{\alpha} = \alpha + y_n$, $\bar{\beta} = \beta + n$, dove $y_n = \sum_{i=1}^n x_i$. Supponiamo anche in questo caso di voler effettuare previsione su Y_f , somma delle osservazioni di un campione casuale futuro di dimensione n_f . Ricordando che $Y_f | \theta \sim \text{Pois}(n_f, \theta)$, abbiamo che

$$\begin{aligned} m(y_f) &= \int_0^\infty f(y_f | \theta) \pi(\theta) d\theta \\ &= \int_0^\infty \frac{n_f^{y_f}}{y_f!} e^{-n_f \theta} (n_f \theta)^{y_f} \times \frac{\beta^\alpha}{\Gamma(\alpha)} e^{\beta \theta} \theta^{\alpha-1} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{n_f^{y_f}}{y_f!} \int_0^\infty \theta^{(\alpha+y_f)-1} e^{(\beta+n_f)\theta} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{n_f^{y_f}}{y_f!} \frac{\Gamma(\alpha + y_f)}{(\beta + n_f)^{\alpha+y_f}} \\ &= \frac{\Gamma(\alpha + y_f)}{\Gamma(\alpha) y_f!} \left(\frac{\beta}{\beta + n_f} \right)^\alpha \left(\frac{n_f}{\beta + n_f} \right)^{y_f} \quad y_m = 0, 1, \dots \end{aligned}$$

La v.a. Y_f ha quindi la funzione di massa di probabilità di una binomiale negativa di parametri $\left(\alpha, \frac{\beta}{\beta + n_f} \right)$:

$$Y_f \sim \text{BinNeg} \left(\alpha, \frac{\beta}{\beta + n_f} \right).$$

Se α è un numero intero, abbiamo che

$$\frac{\Gamma(\alpha + y_f)}{\Gamma(\alpha) y_f!} = \binom{\alpha + y_n - 1}{\alpha - 1}$$

e quindi

$$m(y_f) = \binom{\alpha + y_n - 1}{\alpha - 1} \left(\frac{\beta}{\beta + n_f} \right)^\alpha \left(\frac{n_f}{\beta + n_f} \right)^{y_f}, \quad y_m = 0, 1, \dots$$

Si verifica facilmente che valore atteso e varianza di Y_f sono quindi:

$$\mathbb{E}_m[Y_f] = \mathbb{E}_\Theta[n_f\Theta] = \frac{n_f\alpha}{\beta}$$

e

$$\mathbb{V}_m[Y_f] = \mathbb{E}_\Theta[n_f\Theta] + \mathbb{V}_\Theta[n_f\Theta] = \frac{n_f\alpha(\beta + n_f)}{\beta^2}.$$

Analogamente, a posteriori, si ha che $Y_f|\mathbf{x}_n \sim \text{BinNeg}\left(\bar{\alpha}, \frac{\bar{\beta}}{\bar{\beta}+n_f}\right)$, e quindi

$$\mathbb{E}_m[Y_f|\mathbf{x}_n] = \frac{n_f\bar{\alpha}}{\bar{\beta}}, \quad \mathbb{V}_\Theta[n_f\Theta] = \frac{n_f\bar{\alpha}(\bar{\beta} + n_f)}{\bar{\beta}^2}.$$

con $\bar{\alpha} = \alpha + y_n$, $\bar{\beta} = \beta + n$.

Caso non informativo Il caso non informativo si pu'ò ottenere per $\alpha = \beta = 0$; anche in questo caso, come visto in precedenza, $\pi(\theta)$ perde di significato, mentre abbiamo che $Y_f|\mathbf{x}_n \sim \text{BinNeg}\left(y_n, \frac{n}{n+n_f}\right)$. Anche in questo caso $\mathbb{E}[Y_f|\mathbf{x}_n] = n_f y_n/n = n_f \bar{x}_n$ e quindi la stima bayesiana oggettiva della somma di n_f dati futuri coincide quindi la stima pi'ù intuitiva che possiamo proporre sapendo che la media aritmetica dei dati passati è pari a \bar{x}_n .

8.8 Osservazione/i. In R si possono usare le funzioni `rnbinom()`: ad esempio per la funzione di massa di probabilità, usiamo la funzione `dnbinom(x,size=alpha, prob=beta/(beta+n))`

8.5 Esercizi

- Sia $X_1, \dots, X_n | \theta \sim N(\theta, \sigma^2)$ i.i.d. e $\Theta \sim N\left(\mu_0, \frac{\sigma^2}{n_0}\right)$. Considerare un campione futuro di dimensione n_f da $N(\theta, \sigma^2)$ e sia Y_f la corrispondente media campionaria. Determinare quanto richiesto di seguito:
 - Intervallo di previsione a priori per Y_f , assumendo che $\sigma^2 = 1$, $n_0 = 5$, $n_f = 3$, $\mu_0 = 2$;
 - intervallo di previsione a posteriori per Y_f , assumendo che $n = 20$ e $\bar{x}_n = 0.5$;
 - intervallo di previsione a posteriori per Y_f nel caso di distribuzione iniziale non informativa, assumendo che $n = 20$ e $\bar{x}_n = 0.5$.
- Si assuma che $X_i | \theta \sim N(\theta, \sigma^2)$, $i = 1, \dots, n$ (i.i.d.) e che $\Theta \sim N\left(\mu_0, \frac{\sigma^2}{n_0}\right)$.
 - Determinare $f_X(x_i) = \int_{\mathbb{R}} f_X(x | \theta) \pi(\theta) d\theta$.
 - Determinare $f_{\mathbf{X}_n}(\mathbf{x}_n) = \int_{\mathbb{R}} f_n(\mathbf{x}_n | \theta) \pi(\theta) d\theta$.
 - Mostrare che $X_1 \dots X_n$ non sono variabili aleatorie i.i.d.

[Oss.: questo esercizio mostra che *indipendenza* e *indipendenza condizionata* (a $\Theta = \theta$) sono due proprietà distinte.]

- Assumiamo che: $X_i | \theta \sim \text{EN}(\theta)$, $i = 1, \dots, n_f$ (i.i.d.), $Y_f = \sum_{i=1}^{n_f} X_i | \theta \sim \text{Gamma}(n_f, \text{rate} = \theta)$, $\Theta \sim \text{Gamma}(\alpha, \beta)$. Mostrare che la distribuzione predittiva a priori di Y_f è una gamma-gamma di parametri (α, β, n_f) [in simboli: $Y_f \sim \text{GG}(\alpha, \beta, n_f)$], ovvero che

$$f(y_f) = \frac{\beta^\alpha}{B(\alpha, n_f)} \frac{y_f^{n_f-1}}{(\beta + y_f)^{\alpha+n_f}}, \quad \text{dove} \quad B(\alpha, n_f) = \frac{\Gamma(\alpha)\Gamma(n_f)}{\Gamma(\alpha + n_f)}.$$

- Sia $X | \theta \sim \text{EN}(\theta)$ e $\Theta \sim \text{Gamma}(\alpha, \beta)$.
 - Determinare l'espressione di $m(x)$, distribuzione predittiva a priori della singola v.a. X .
 - Calcolare il valore atteso di X e stabilire per quali valori di α e β tale valore esiste.
 - Calcolare il valore della densità marginale $m(x)$ di X in $x = 1$ che si ottiene ponendo $\alpha = \beta = 1$.
 - Calcolare con $m(x)$ la probabilità dell'evento $(X < 1)$ che si ottiene ponendo $\alpha = \beta = 1$.
- Per n_f v.a. X_i^f di un esperimento futuro, indipendenti ed identicamente distribuite condizionatamente a θ , si assuma che $X_i^f | \theta \sim N(\theta, \sigma^2)$. Il successo dell'esperimento consiste nel rifiutare l'ipotesi nulla del sistema di ipotesi $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$, utilizzando il test basato sulla regione di rifiuto

$$R = \left\{ \mathbf{x}_{n_f} : \frac{\sqrt{n_f} |y_f - \theta_0|}{\sigma} > z_{1-\frac{\alpha}{2}} \right\},$$

dove y_f è la media campionaria delle osservazioni future. Si supponga che $\Theta \sim N\left(\mu_0, \frac{\sigma^2}{n_0}\right)$. Assumere $\theta_0 = 3$ e $\sigma^2 = 1$.

- (a) Fornire l'espressione (non serve fare i calcoli) della distribuzione predittiva *a priori* di Y_f e della corrispondente probabilità che l'esperimento sia un successo.
 - (b) Supponendo di avere osservato in passato un campione di dimensione n di v.a. X_i con distribuzione $N(\theta, \sigma^2)$ e i.i.d. condizionatamente a θ , fornire l'espressione della distribuzione predittiva *a posteriori* di Y_f e della corrispondente probabilità che l'esperimento sia un successo.
 - (c) Calcolare la probabilità predittiva a priori di successo per $n_0 = 5$, $\mu_0 = 2$, $n_f = 6$, $\alpha = 0.05$.
 - (d) Calcolare la probabilità predittiva a posteriori di successo per $n_0 = 5$, $\mu_0 = 2$, $n_f = 6$, $\alpha = 0.05$, $n = 10$, $\bar{x}_n = 4$.
 - (e) Con i dati assegnati, calcolare l'intervallo di previsione a posteriori per Y_f di livello $1 - \gamma = 0.95$.
6. Sia \mathbf{X}_{n_f} un campione casuale di dati futuri di dimensione n_f da $N(\theta, \sigma^2)$, σ^2 noto; e sia $\Theta \sim N(\mu_0, \sigma^2 n_0^{-1})$, $\mu_0 \in \mathbb{R}$, $n_0 \in \mathbb{R}^+$. Riteniamo che l'esperimento abbia **successo** se

$$\mathbb{P}(\Theta < \delta | \mathbf{x}_{n_f}) > \gamma, \quad \gamma \in (0, 1).$$

- (a) Verificare che si ha un **successo** se si osserva un campione \mathbf{x}_{n_f} che appartiene all'insieme

$$S = \{\mathbf{x}_{n_f} : y_f = \bar{x}_{n_f} < k\},$$

dove $k = k(\delta, \pi, \gamma)$ è una costante che dipende da δ , γ e dagli iperparametri della distribuzione a priori.

- (b) Fornire l'espressione di $k(\delta, \pi, \gamma)$. [**Sugg.:** ricordare che, se $\mathbb{F}(\cdot)$ è la funzione di ripartizione della v.a. X , allora $\mathbb{F}(\cdot)^{-1}(\epsilon) = q_\epsilon$, $\epsilon \in (0, 1)$.]
- (c) Determinare la probabilità predittiva *a priori* di osservare un campione in S .