Appunti di Statistica C.P.

Alberto Cavalin × Francesco Ragazzo

2006/2008

Hall of fame:

►	*Francesco R.*	:	⋆ *strategical idea hunter* ⋆ *official licensed motivator* ⋆ *coffee break promoter* ⋆ *Marlesia's food feeder*
►	*Laura P.*	:	⋆ *sergeant major general* ⋆ *official study promoter* ⋆ *pomodorini & peperoni feeder*
►	*Silvia B.*	:	⋆ *house, food & cat provider* ⋆ *natural pheromone dispenser* ⋆ *master Cek distracter*
►	*Stefano C.*	:	⋆ *battlefield master chef* ⋆ *house, food & sm0ke provider* ⋆ *natural fun dispenser* ⋆ *"mele yeah"™*

. . . e tutte le persone che mi hanno supportato finora ;-)

NB: Nel titolo il simbolo "×" indica "*congiuntamente a*" e non "*dedicato a*"!

Indice

1 Inferenza frequentista **1**

1.1 Modello statistico 1

1.2 Stime e stimatori 1

1.2.1 Proprietà degli stimatori 1

1.3 Verifica d'ipotesi/Test 2

1.3.1 Livello di significativià 2

1.3.2 Funzione di potenza 2

1.3.3 Proprietà di un test 2

1.3.4 Regioni di confidenza 2

1.4 La verosimiglianza 3

1.4.1 Stima di massima verosimiglianza 4

1.5 Statistiche sufficienti 4

1.5.1 Statistiche sufficienti e criterio di Neyman-Fisher 4

1.5.2 Statistiche sufficienti minimali 4

1.6 Famiglie esponenziali 5

1.6.1 Famiglie esponenziali regolari 5

1.7 Efficienza di uno stimatore 6

1.7.1 Problemi regolari di stima 6

1.7.2 Disuguaglianza di Cramer-Rao 6

1.7.3 Teorema di Rao-Blackwell 7

1.7.4 Teorema di Lehman-Scheffé 7

1.8 Test di verosimiglianza (Wilks) e regioni di confidenza 7

1.9 Verosimiglianza profilo 8

1.10 Test ottimi 8

1.10.1 Test rapporto di verosimiglianza 9

1.10.2 Lemma di Neyman-Pearson . . . 9
1.10.3 Ipotesi composite . . . 10

2 Inferenza robusta **11**
2.1 Equazioni di stima . . . 11
2.2 Proprietà degli stimatori di tipo M . . . 11
2.2.1 Normalità asintotica . . . 12
2.3 Quasi-verosimiglianza . . . 13
2.4 Funzionali statistici e funzione di influenza . . . 13
2.4.1 Stimatore di Huber . . . 14

3 Inferenza bayesiana **15**
3.1 Il teorema di Bayes . . . 15
3.1.1 Versione generale . . . 15
3.1.2 Versione nel continuo . . . 15
3.2 Procedure inferenziali . . . 16
3.2.1 Stima puntuale . . . 16
3.2.2 Regioni di confidenza HPD . . . 16
3.2.3 Verifica d'ipotesi . . . 17
3.3 Coniugate naturali . . . 17

A Richiami **19**
A.1 Distribuzioni note . . . 19
A.1.1 Alcune nozioni fondamentali . . . 19
A.1.2 Distribuzione di media e varianza campionaria . . . 19
A.2 Funzioni indicatrici . . . 19
A.3 Disuguaglianza di Jensen . . . 19
A.4 Newton-Raphson . . . 20
A.5 Cambio di variabile . . . 20
A.6 Metodo delta . . . 20
A.7 Quantità pivotale . . . 21
A.8 Partizione di un insieme . . . 21
A.9 Promemoria per la risoluzione di un problema . . . 21
A.9.1 Inferenza frequentista . . . 22

A.9.2 Inferenza robusta 22
A.9.3 Inferenza bayesiana 22

B Notazioni, tabelle, e grafici **23**

B.1 Notazioni 23

Capitolo 1

Inferenza frequentista

1.1 Modello statistico

Def: $\mathscr{F} = \{\mathbb{P}_\theta; \theta \in \Theta\} = \{f_Y(y;\theta); \theta \in \Theta\}$	Y = variabile aleatoria, θ = parametro $\mathscr{Y}$ = spc campionario / supporto di Y Θ = spc parametrico / supporto di f_Y

- $\mathscr{F}$ non cambia se soggetto a riparametrizzazioni biunivoche
- $\hat{\theta}(\underline{y})$ = stima di θ, $\underline{y}$ = vettore oss camp
- $\hat{\theta}(\underline{Y})$ = stimatore di θ, $\underline{Y}$ = vettore aleatorio

1.2 Stime e stimatori

Metodi di stima:

- sostituzione ($\tau = g(\theta) \;\rightarrow\; \hat{\tau} = g(\hat{\theta})$)
- analogia (la qt della pop viene stimata con la qt campionaria)
- momenti = metodo dell'analogia applicato ai momenti
- massima verosimiglianza

1.2.1 Proprietà degli stimatori

- non distorto (corretto) $\Leftrightarrow$ $\mathbb{E}(\hat{\theta}) = \theta$
- consistente $\Leftrightarrow$ $\hat{\theta} \xrightarrow{p} \theta \;\Leftrightarrow\; \mathbb{E}(\hat{\theta}) = \theta \wedge \lim_{n\to+\infty} \mathbb{V}(\hat{\theta}) = 0$
- asintoticamente non distorto $\Leftrightarrow$ $\lim_{n\to+\infty} \mathbb{E}(\hat{\theta}) = \theta$
- debolmente consistente $\Leftrightarrow$ $\lim_{n\to+\infty} \mathbb{E}(\hat{\theta}) = \theta \wedge \lim_{n\to+\infty} \mathbb{V}(\hat{\theta}) = 0$

<u>NB:</u> Le proprietà di $\hat{\theta}$ dipendono dalla parametrizzazione.

1.3 Verifica d'ipotesi/Test

$$\text{Sistema d'ipotesi: } \begin{cases} H_0: & \theta \in \Theta_0 \\ H_1: & \theta \in \Theta \setminus \Theta_0 \end{cases} \quad \left| \quad \begin{aligned} t(y) &= \text{statistica test} \\ &= \text{funzione dei dati indipendente da } \theta \end{aligned} \right.$$

- regione di accettazione : $A = \{y \in \mathscr{Y} : t(y) \text{ porta verso } H_0\}$
- regione di rifiuto : $R = \{y \in \mathscr{Y} : t(y) \text{ porta verso } H_1\}$

- errore di I tipo : rifiutare H_0 quando è vera H_0
- errore di II tipo : rifiutare H_1 quando è vera H_1

1.3.1 Livello di significativià

Def: Livello di significativià fissato α

$$\begin{aligned} \alpha &= \sup_{H_0} \mathbb{P}_\theta(Y \in R) \\ &= \max(\text{prob di errore del I tipo}) \end{aligned}$$

Def: Livello di significativià osservato α^{oss}

- unilaterale dx : $\alpha^{oss} = \sup_{H_0} \mathbb{P}_\theta(T \geq t^{oss})$
- unilaterale sx : $\alpha^{oss} = \sup_{H_0} \mathbb{P}_\theta(T \leq t^{oss})$
- bilaterale : $\alpha^{oss} = 2\sup_{H_0} \min\left[\mathbb{P}_\theta(T \leq t^{oss}), \mathbb{P}_\theta(T \geq t^{oss})\right]$

1.3.2 Funzione di potenza

$$\pi(\theta) = \mathbb{P}_\theta(Y \in R) = \begin{cases} \alpha(\theta) & , \theta \in \Theta_0 \\ 1 - \beta(\theta) & , \theta \in \Theta \setminus \Theta_0 \end{cases} \quad \left| \quad \begin{aligned} \alpha(\theta) &= \mathbb{P}(\text{err I tipo}) &&= \mathbb{P}(\text{rif. } H_0) \\ 1 - \beta(\theta) &= 1 - \mathbb{P}(\text{err II tipo}) &&= \mathbb{P}(\text{rif. } H_0) \end{aligned} \right.$$

1.3.3 Proprietà di un test

Def: Test non distorto $\Leftrightarrow$ $\pi(\theta) \geq \alpha, \ \forall \theta \in \Theta \setminus \Theta_0$ $\Leftrightarrow$ $\mathbb{P}(\text{rif } H_0) \geq \alpha, \ \forall \theta \in \Theta \setminus \Theta_0$

Def: Test consistente $\Leftrightarrow$ $\lim_{n \to +\infty} \pi(\theta) = 1, \ \forall \theta \in \Theta \setminus \Theta_0$

1.3.4 Regioni di confidenza

Def: Una regione di confidenza $\hat{\Theta}(y) \subset \Theta$ di livello $(1-\alpha)$, viene ricavata fissando la sua probabilità di copertura pari a $(1 - \alpha)$:

$$\mathbb{P}(\theta \in \hat{\Theta}(y)) = 1 - \alpha, \forall \theta \in \Theta \qquad \rightsquigarrow \qquad \hat{\Theta}(y)$$

oppure, quando è possibile, tramite una quantità pivotale (§A.7):

$$\mathbb{P}(a(Y,\theta) \in E) = 1 - \alpha,\, \forall \theta \in \Theta, E \subset \mathbb{R} \qquad \rightsquigarrow \qquad \hat{\Theta}(y)$$

Oss: Nel caso del p-value si ha: $\hat{\Theta}(y) = \{\theta \in \Theta : \alpha^{oss}(y,\theta) \geq \alpha\}$

1.4 La verosimiglianza

Possibili applicazioni della verosimiglianza:

- stima puntuale : $\hat{\theta} : L(\hat{\theta}) \geq L(\theta),\ \forall \theta \in \Theta$
- stima intervallare : $\hat{\Theta}(y) = \left\{\theta \in \Theta : L(\theta) \geq L(\hat{\theta}) \cdot c,\ 0 \leq c \leq 1\right\}$
- test TRV : $L(\hat{\theta})/L(\theta_0)$

Def: Funzione di verosimiglianza $L(\theta)$, e log-verosimiglianza $l(\theta)$:

$$\begin{aligned}
L(\theta) &= \Pi_{i=1}^{n} f_{X_i|X_{i-1},\dots,X_1}(x_i|x_{i-1},\,\dots,\,x_1;\theta)\,, \quad X = (X_1,\,\dots,\,X_n) \\
l(\theta) &= \log L(\theta) \\
l^*(\theta) &= \nabla l(\theta) = [\partial l(\theta)/\partial\theta_i] \qquad i,j = 1,\,\dots,\,p = \text{dim}(\theta) \\
l^{**}(\theta) &= \mathbb{J}l(\theta) = [\partial^2 l(\theta)/\partial\theta_i\partial\theta_j] \qquad \nabla = \text{gradiente}\,,\ \mathbb{J} = \text{jacobiano}
\end{aligned}$$

Proprietà della verosimiglianza:

- $L(\theta)$ è invariante rispetto a riparametrizzazioni biunivoche
- $L(\theta) \propto L(\theta) \cdot c\,, \quad c =$ costante

  $\Rightarrow \quad L(\theta)$ è equivalente a $L(\theta) \cdot c \quad \Rightarrow \quad l(\theta)$ è equivalente a $l(\theta) + c$
- $\mathbb{E}(l^*(\theta)) = 0, \quad \mathbb{V}(l^*(\theta)) = \mathbb{E}(l^*(\theta) \cdot l^*(\theta)^T) = I(\theta) \quad \rightsquigarrow \quad l^*(\theta) \dot{\sim} N(0, I(\theta))$
- Informazione osservata: $J(\theta) = -l^{**}(\theta) = -\mathbb{J}l(\theta)$
- Informazione attesa: $I(\theta) = \mathbb{E}(J(\theta)) = -\mathbb{E}(l^{**}(\theta))$

  Oss: Se gli X_i sono i.i.d. si ha $I(\theta) = ni(\theta)$ dove $i(\theta)$ è l'informazione osservata per X_1.

1.4.1 Stima di massima verosimiglianza

Def: È definita stima di max verosimiglianza (o SMV) di θ la quantità[1]:

$$\hat{\theta} := \left\{ \hat{\theta} \in \Theta : L(\hat{\theta}) \geq L(\theta),\ \forall \theta \in \Theta \setminus \{\hat{\theta}\} \right\}$$

$\hat{\theta} \,\dot{\sim}\, N_p\big(\theta,\, 1/J(\hat{\theta})\big)$	distrib. approx	, $p = \dim(\theta)$	$\rightarrow$	più fedele al campione
$\hat{\theta} \,\dot{\sim}\, N_p\big(\theta,\, 1/I(\hat{\theta})\big)$	distrib asintotica		$\rightarrow$	più fedele alla distrib. asintotica

Def: Equivarianza (od ~~Invarianza~~) della SMV:

Data $\hat{\theta}$ SMV per $L(\theta)$, e $\psi(\cdot) : \Theta \rightarrow \Psi$ biunivoca, allora la SMV per $\psi(\theta)$ è $\hat{\psi} = \psi(\hat{\theta})$.

1.5 Statistiche sufficienti

1.5.1 Statistiche sufficienti e criterio di Neyman-Fisher

Def: Una statistica $t(\cdot)$ è sufficiente se

$$t(y) = t(z) \quad \Rightarrow \quad L(\theta, y) \propto L(\theta, z)\,, \quad \forall \theta \in \Theta\,, \quad \forall y, z \in \mathscr{Y}$$

o equivalentemente se e solo se

$$L(\theta) \propto h(y) k\,(t(y);\, \theta)$$

Oss: Se $t(\cdot)$ è sufficiente, allora $L(\theta)$ dipende da y solo tramite $t(y)$, cioè $L(\theta, y) = L(\theta, t(y))$.

Oss: Esiste sempre una statistica sufficiente e coincide con l'intero campione di dati y.

1.5.2 Statistiche sufficienti minimali

Def: Una statistica $t(\cdot)$ è sufficiente minimale se

$$t(y) = t(z) \quad \Leftrightarrow \quad L(\theta, y) \propto L(\theta, z)\,, \quad \forall \theta \in \Theta\,, \quad \forall y, z \in \mathscr{Y}$$

o equivalentemente se

$$L(y; \theta)/L(z; \theta) = h(y, z) \quad \Leftrightarrow \quad t(y) = t(z)$$

[1] Per una soluzione numerica si veda il metodo di Newton-Raphson in §A.4

NB: Considerando le curve di livello $\mathscr{Y}_V = \{y \in \mathscr{Y} : \text{le } L(\theta, y) \text{ sono proporzionali tra loro } \}$ e $\mathscr{Y}_c = \{y \in \mathscr{Y} : t(y) = c\}$, si ottengono P_V e P_t, due partizioni[2] di $\mathscr{Y}$.

Nel caso delle stat. suff. si ha che $\#P_t > \#P_V$ in quanto si ha almeno un caso in cui $\mathscr{Y}_c \subset \mathscr{Y}_V$, mentre in quelle minimali $\#P_t = \#P_V$ poiché $P_V \equiv P_t$ perché in ogni caso $\mathscr{Y}_c \equiv \mathscr{Y}_V$.

Oss: Se la stat. suff. min. esiste, essa è essenzialmente unica, ovvero le stat. suff. min. sono tutte funzioni l'una dell'altra, in quanto condividono la stessa partizione indotta.

1.6 Famiglie esponenziali

Def: In una famiglia esponenziale $L(\theta)$ assume la forma:

$$L(\theta) \propto h(y) \exp\left\{\sum_{i=1}^{k} \psi_i(\theta) t_i(y) - c(\theta)\right\}, \quad \dim(\theta) = p$$

dove k è detto ordine, e $\psi = \psi(\theta)$ parametro canonico.

1.6.1 Famiglie esponenziali regolari

Def: Una famiglia esponenziale è detta regolare se:

- Θ coincide con l'intero insieme per cui $L(\theta)$ è integrabile
- la dimensione di θ coincide con l'ordine: $p = k$
- le funzioni $\psi(\theta) : \Theta \to \Psi$ sono invertibili
- le funzioni $\psi(\theta)$ sono infinitamente derivabili in θ

Parametrizzazione canonica

La funzione $\psi(\theta)$ è una riparametrizzazione, è quindi possibile scrivere:

$$\left.\begin{array}{l} l(\psi) = \sum_{i=1}^{p} \psi_i t_i - k(\psi), \quad k(\psi) = c(\theta(\psi)) \\ l^*(\psi) = \left[t_i - \dfrac{\partial k(\psi)}{\partial \psi_i}\right] \\ J(\psi) = -l^{**}(\psi) = \left[\dfrac{\partial^2 k(\psi)}{\partial \psi_r \partial \psi_s}\right] \end{array}\right| \begin{array}{l} \\ i = 1, \ldots, p \\ r, s = 1, \ldots, p \end{array}$$

Oss: $J(\psi)$ non stocastica $\Rightarrow I(\psi) \equiv J(\psi)$

[2] Si veda la definizione e relativi esempi di partizione di un insieme in §A.8

NB: Da qui possiamo ricavare valore atteso e varianza di T_i:

- $\mathbb{E}(l^*(\psi)) = 0 \quad \Rightarrow \quad \mathbb{E}(T_i) = \mathbb{E}\left(\frac{\partial k(\psi)}{\partial \psi_i}\right) = \frac{\partial k(\psi)}{\partial \psi_i}$
- $\mathbb{E}(l^*(\psi) l^*(\psi)^T) = I(\psi) \quad \Rightarrow \quad \mathbb{E}(l_r^*(\psi; T_r) l_s^*(\psi; T_s)) = \mathbb{C}\text{ov}(T_r, T_s) = \frac{\partial^2 k(\psi)}{\partial \psi_r \partial \psi_s}$

  $\Rightarrow \mathbb{V}(T_i) = \mathbb{C}\text{ov}(T_i, T_i) = \frac{\partial^2 k(\psi)}{\partial^2 \psi_i}$

SMV per famiglie esponenziali regolari

In pratica, risolvere il sistema:

$$l^*(\psi) = \frac{\partial l(\psi)}{\partial \psi_i} = 0 \leadsto \hat{\psi}_i \leadsto \hat{\theta}_i\,, \quad i = 1, \ldots, p$$

equivale a risolvere questo altro sistema:

$$t_i - \frac{\partial k(\psi)}{\partial \psi_i} = 0 \quad \Leftrightarrow \quad t_i = \mathbb{E}(T_i) \leadsto \hat{\psi}_i \leadsto \hat{\theta}_i\,, \quad i = 1, \ldots, p$$

1.7 Efficienza di uno stimatore

1.7.1 Problemi regolari di stima

Un problema regolare di stima è caratterizzato dalle seguenti proprietà:

- il modello statistico è identificabile secondo la definizione al §1.1
- Θ è un aperto di $\mathbb{R}^p$, con $p = \dim(\theta)$
- le funzioni di densità f_Y hanno tutte lo stesso supporto
- per le f_Y si può scambiare due volte il simbolo di integrale e derivata rispetto a θ

1.7.2 Disuguaglianza di Cramer-Rao

In un problema regolare di stima, dato uno stimatore $\tilde{\theta}$ con $\mathbb{E}(\tilde{\theta}) = a(\theta)$, si ha che:

$$\mathbb{V}(\tilde{\theta}) \geq [a'(\theta)]^2 / I(\theta) \quad , \tilde{\theta} \text{ distorto}$$
$$\mathbb{V}(\tilde{\theta}) \geq 1/I(\theta) \quad , \tilde{\theta} \text{ non distorto}$$

Def: Stimatore $\tilde{\theta}$ efficiente $\quad \Leftrightarrow \quad \mathbb{E}(\tilde{\theta}) = \theta \wedge \mathbb{V}(\tilde{\theta}) = 1/I(\theta)$

NB: Non sempre è possibile reperire un $\tilde{\theta}$ che sia efficiente.

1.7.3 Teorema di Rao-Blackwell

$$\left.\begin{array}{l}\hat{\theta}(T) \text{ funzione dei dati solo tramite T stat. suff.}\\ \tilde{\theta}(Y) \text{ basato sui dati non ridotti}\end{array}\right\} \quad \Rightarrow \quad \mathbb{V}(\hat{\theta}) \leq \mathbb{V}(\tilde{\theta})$$

<u>Oss:</u> Se T è stat. suff. min., e $\mathbb{E}\left(\hat{\theta}(T)\right) = \theta$ con $\hat{\theta}(T)$ unico $\quad \Rightarrow \quad \hat{\theta}(T)$ è efficiente.

<u>NB:</u> Per dimostrare l'unicità di $\hat{\theta}(T)$ suppongo che esista un altro stimatore $\hat{\theta}^*(T)$ non distorto, e poi verifico che $\mathbb{E}\left(\hat{\theta}^*(T) - \hat{\theta}(T)\right) = 0 \Leftrightarrow \hat{\theta}^*(T) = \hat{\theta}(T)$.

1.7.4 Teorema di Lehman-Scheffé

<u>Thm:</u> Se $\exists T$ stat. suff. completa ed $\exists \tilde{\theta}(Y)$ non distorto $\quad \Rightarrow \quad \exists!\, \tilde{\theta}^* = \mathbb{E}(\tilde{\theta}|T = t)$ ed è efficiente.

<u>Def:</u> Una statistica T è detta <u>completa</u> se $\mathbb{E}\left(f(T)\right) = 0,\ \forall \theta \in \Theta \quad \Leftrightarrow \quad f(T)$ è la v.a. degenere 0.

<u>Oss:</u> Alcune importanti considerazioni:

- Se t è stat. suff. completa $\quad \Rightarrow \quad T$ è l'unico stimatore efficiente per $\mathbb{E}(T)$
- Se $\tilde{\theta}(T)$ è non distorto e funzione biunivoca di $T \quad \Rightarrow \quad \tilde{\theta}(T)$ è efficiente
- Tipicamente la SMV $\hat{\theta}$ è distorta per rispettare l'equivarianza rispetto a riparametrizzazioni. Se quindi essa non è efficiente ma esiste un $\tilde{\theta}$ che lo è, si ha che $\hat{\theta}$ è una buona approssimazione di $\tilde{\theta}$ perchè $\mathbb{E}(\hat{\theta}) \doteq \theta$ per n sufficientemente grande.

1.8 Test di verosimiglianza (Wilks) e regioni di confidenza

$H_0 : \theta = \theta_0\ vs\ H_1 : \theta \neq \theta_0$	$H_0 : \theta = \theta_0\ vs\ H_1 : \theta <> \theta_0$
$W(\theta_0) = -2 \log \dfrac{\max_{H_0} L(\theta)}{\max_{H_1} L(\theta)} = 2\left[l(\hat{\theta}) - l(\theta_0)\right]$	$r(\theta_0) = \text{sgn}(\hat{\theta} - \theta_0)\left\|\sqrt{W(\theta_0)}\right\|$
$W(\theta_0) \dot{\sim} \chi^2_p,\ p = \text{dim}(\theta)$	$r(\theta_0) \dot{\sim} N(0,1)$
$\Theta(y) \doteq \left\{\theta \in \Theta : W(\theta) < \chi^2_{p;1-\alpha}\right\}$	$\Theta(y) \doteq \left\{\theta \in \Theta : r(\theta) < \quad z_{1-\alpha}\right\},\ H_1 : \theta > \theta_0$
$\doteq \left\{\theta \in \Theta : l(\theta) > l(\hat{\theta}) - \frac{1}{2}\chi^2_{p;1-\alpha}\right\}$	$\Theta(y) \doteq \left\{\theta \in \Theta : r(\theta) > -z_{1-\alpha}\right\},\ H_1 : \theta < \theta_0$
$\doteq \left\{\theta \in \Theta : -z_{1-\alpha} < r(\theta) < z_{1-\alpha}\right\}$	

<u>Oss:</u> Se W (oppure r) è funzione monotona crescente di una statistica T avente distribuzione nota sotto $H_0 \quad \Rightarrow \quad$ è possibile effettuare un test equivalente <u>esatto</u> tramite T:

$\alpha^{oss} = \mathbb{P}\left(W(\theta) \geq W^{oss}\right) = \mathbb{P}\left(T \geq t^{oss}\right)$	$\alpha^{oss} = \mathbb{P}\left(r(\theta) \leq\geq r^{oss}\right) = \mathbb{P}\left(T \leq\geq t^{oss}\right)$

<u>NB:</u> Se $r(\theta)$ è funzione monotona decrescente di $\theta \quad \Rightarrow \quad$ è possibile utilizzare anche <u>ipotesi nulle composite</u> ($H_0 : \theta \leq\geq \theta_0$) in modo del tutto identico a quelle semplici.

1.9 Verosimiglianza profilo

Def: Data $L(\theta)$ con $\theta = (\tau, \xi)$, dim$(\theta) = p$, dim$(\tau) = k$, dim$(\xi) = p - k$, dove τ è il vettore di componenti d'interesse, mentre ξ è quello di disturbo, è detta verosimiglianza profilo:

$$L_p(\tau) = L(\tau, \hat{\xi}_\tau)$$

dove $\hat{\xi}_\tau$ è la SMV di ξ fatta considerando τ fissato/costante.

Oss: Si può trattare $L_p(\tau)$ come una comune verosimiglianza; è infatti possibile utilizzare i test $W_p(\tau)$ e $r_p(\tau)$ in modo del tutto analogo ai $W(\theta)$ e $r(\theta)$.

NB: Nel caso di un test con ipotesi "multiparametriche" basta applicare una riparametrizzazione biunivoca e poi ci si riconduce ad una verosimiglianza profilo. Ad esempio:

$$\begin{cases} H_0 : \alpha \leq \beta \\ H_1 : \alpha > \beta \end{cases} \rightsquigarrow \begin{cases} H_0 : \tau \leq 0 \\ H_1 : \tau > 0 \end{cases}, \begin{cases} \tau = (\alpha - \beta) \\ \beta = \beta \end{cases} \rightsquigarrow L_p(\tau) = L(\tau, \hat{\beta}_\tau)$$

Oss: Nel caso di un test bilaterale è inoltre possibile usare il test $W(\theta)$ come segue:

$$W(H_0) = 2\left[\max\left(l(\theta)|_{H_1}\right) - \max\left(l(\theta)|_{H_0}\right)\right]$$

dove le $l(\theta)$ sono calcolate sotto i vincoli delle ipotesi. In pratica questo metodo viene impiegato quando sotto H_0 si hanno $k \leq p = \dim(\theta)$ vincoli.

1.10 Test ottimi

Def: Un test per $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0$ con α fissato, avente regione di rifiuto R^*, è detto ottimo (od uniformemente più potente) se:

$$\mathbb{P}(Y \in R^*) \geq \mathbb{P}(Y \in R), \quad \forall \theta \in \Theta_1, \quad \forall \text{ altro test con regione di rif. R}$$

ovvero se e solo se:

$$\pi^*(\theta) \geq \pi(\theta), \quad \forall \theta \in \Theta_1 \quad \left| \quad \begin{array}{ll} \pi^*(\theta) & = \text{f.d.p. del test ottimo} \\ \pi(\theta) & = \text{f.d.p. di un qualsiasi altro test} \end{array} \right.$$

1.10.1 Test rapporto di verosimiglianza

Def: Considerando il seguente sistema di ipotesi, e la statistica test associata:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}, \quad \Theta = \{\theta_0,\, \theta_1\} \quad \Bigg| \quad t^* = t^*(y) = \frac{L(\theta_1; y)}{L(\theta_0; y)}$$

la statistica t^* è detta rapporto di verosimiglianza. Il test omonimo che ne deriva (TRV) avrà perciò come regione di rifiuto $R = \{y : t^*(y) \geq c_\alpha\}$, dove c_α soddisfa l'equazione $\mathbb{P}\left(t^*(Y) \geq c_\alpha; \theta_0\right) = \alpha$.

Varianti del TRV

- Test di Wilks: $W(\theta_0) = 2\left[l(\hat{\theta}) - l(\theta_0)\right]$

  $r(\theta_0) = \text{sgn}(\hat{\theta} - \theta_0)\left|\sqrt{W(\theta_0)}\right| , \quad p = 1$

  È il più affidabile ed usato.

- Test di Wilks con correzione di Bartlett: $W^*(\theta_0) = [p/\mathbb{E}(W(\theta_0))] \cdot W(\theta_0)$

  $r^*(\theta_0) = [r(\theta_0) - \mathbb{E}(r(\theta_0))/\mathbb{V}(r(\theta_0))] , \quad p = 1$

  Ha distribuzione nulla più vicina a quella asintotica.

- Test di Wald: $W_e(\theta_0) = (\hat{\theta} - \theta_0)^T I(\theta_0)(\hat{\theta} - \theta_0)$

  $r_e(\theta_0) = (\hat{\theta} - \theta_0)\sqrt{I(\theta_0)} , \quad p = 1$

  Buona interpretabilità, minor affidabilità, dipende dalla parametrizzazione, e può generare regioni di confidenza errate.

- Test di Rao: $W_u(\theta_0) = l^*(\theta_0)^T I(\theta_0)^{-1} l^*(\theta_0)$

  $r_u(\theta_0) = \text{sgn}(\hat{\theta} - \theta_0)\sqrt{W_u(\theta_0)} , \quad p = 1$

  Non richiede $\hat{\theta}$ ma può generare risultati imprecisi se θ_0 è in prossimità della frontiera di Θ.

$W_\bullet$ e $W^* \dot{\sim} \chi^2_p, \quad p = \dim(\theta)$ NB: $I(\theta)$ può essere stimata con $I(\theta_0),\, I(\hat{\theta}),\, J(\hat{\theta})$

$r_\bullet$ e $r^* \dot{\sim} N(0,\, 1)$

1.10.2 Lemma di Neyman-Pearson

Def: Data una densità $f(y; \theta)$, e fissato un sistema di ipotesi:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}, \quad \Theta = \{\theta_0,\, \theta_1\} \quad \text{oppure} \quad \begin{cases} H_0 : f = f_0 \\ H_1 : f = f_1 \end{cases}, \quad \mathscr{F} = \{f_0(\theta_0),\, f_1(\theta_1)\}$$

allora il TRV $t^*(y) = L(\theta_1; y)/L(\theta_0; y) = f_1(\theta_1; y)/f_0(\theta_0)$ risulta essere il più potente tra quelli di livello α fissato.

1.10.3 Ipotesi composite

Nel caso in cui il TRV $t^*(y) = h(t(y))$, con $h(\cdot)$ funzione monotona; conoscendo la distribuzione di $t(y)$ sotto H_0, allora:

- $R^* = \{y : t^*(y) > c\} = \{y : t(y) > t_{1-\alpha}\}\,, \quad h(\cdot)$ monotona crescente
- $R^* = \{y : t^*(y) > c\} = \{y : t(y) < t_{\alpha} \quad \}\,, \quad h(\cdot)$ monotona decrescente

Risulta quindi possibile utilizzare ipotesi composite del tipo:

- $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases} , \quad \theta_1 > \theta_0 \quad \rightsquigarrow \quad \begin{cases} H_0 : \theta = \theta_0 \quad \vee \quad \theta \leq \theta_0 \\ H_1 : \theta \geq \theta_1 \end{cases}$
- $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases} , \quad \theta_1 < \theta_0 \quad \rightsquigarrow \quad \begin{cases} H_0 : \theta = \theta_0 \quad \vee \quad \theta \geq \theta_0 \\ H_1 : \theta \leq \theta_1 \end{cases}$

NB: Le famiglie esponenziali monoparametriche ($p = 1$) hanno sempre TRV monotono rispetto a $t(y)$ stat. suff. min.. Il verso della monotonia del TRV dipende solamente da quello della monotonia di $\psi(\theta)$.

Oss: I test $W(\theta)$ e $r(\theta)$ sono funzioni monotone del TRV, e quindi sono anch'essi dei test ottimi.

Riassumendo: una trasformazione monotona del TRV non altera la bipartizione $(\mathscr{Y}_0, \mathscr{Y}_1)$ dello spazio campionario, quindi $\mathbb{P}(Y \in R) = \pi(\theta)$ non viene alterata, e l'ottimalità permane:

$$\mathbb{P}(Y \in \{y \in \mathscr{Y} : t^*(y) > c\}) = \mathbb{P}(Y \in \{y \in \mathscr{Y} : t(y) \rightsquigarrow H_1\}) = \mathbb{P}(Y \in \{y \in \mathscr{Y}_1)\})$$

Capitolo 2

Inferenza robusta

L'obiettivo è modificare l'equazione di verosimiglianza del modello $\mathscr{F}$ in modo tale da renderla robusta rispetto a dati anomali.

2.1 Equazioni di stima

Def: Un'equazione in θ del tipo $q(\theta;\, y) = 0$, è detta equazione di stima e $q(\theta;\, y)$ funzione di stima.

Def: Una funzione di stima è detta non distorta $\quad \Leftrightarrow \quad \mathbb{E}(q(\theta;\, Y)) = 0,\ \forall\theta \in \Theta$

NB: Nella quasi totalità dei casi $q(\theta;\, y) = \sum_{i=1}^{n} g(\theta;\, y_i)$, dove $g(\theta;\, y_i)$ è la funzione di stima per l'osservazione i-esima; perchè $q(\theta;\, y)$ sia non distorta deve anche esserlo ogni singola $g(\theta;\, y_i)$, cioè:

$$\mathbb{E}(q(\theta;\, Y)) = 0,\ \forall\theta \in \Theta \quad \Leftrightarrow \quad \mathbb{E}(g(\theta;\, Y_i)) = 0,\ \forall\theta \in \Theta,\ i = 1,\, \dots,\, n$$

Oss: Un esempio di equazione di stima non distorta è $l^*(\theta;\, y) = \sum_{i=1}^{n} \ell(\theta;\, y_i) = 0$.

Def: Gli stimatori $\tilde{\theta}$ definiti come soluzione di $q(\theta;\, y) = 0$ sono detti stimatori di tipo M (ovvero stimatori di massima verosimiglianza).

2.2 Proprietà degli stimatori di tipo M

Def: Preso $\hat{\theta}$ soluzione di $l^*(\theta;\, y) = 0$ ed un qualsiasi altro stimatore $\tilde{\theta}$ soluzione di una $q(\theta;\, y) = 0$, si ha che $\mathbb{V}(\hat{\theta}) \leq \mathbb{V}(\tilde{\theta}),\ \forall\theta \in \Theta$, cioè $\hat{\theta}$ è lo stimatore a varianza minima tra i non distorti.

2.2.1 Normalità asintotica

Lo score $l^*(\theta;\, y)$ è un caso particolare di $q(\theta;\, y)$, cerchiamo quindi di generalizzarne i risultati. Sviluppando $q(\theta;\, y)$ in serie di Taylor e con opportune considerazioni, si ottiene:

$$q(\theta;\, y) = \sum_{i=1}^{n} g(\theta;\, y_i)$$
$$\sqrt{n}(\tilde{\theta} - \theta) = \frac{\sqrt{n} q(\theta;\, y)}{-\mathbb{E}(\frac{\partial}{\partial\theta} q(\theta;\, Y))}$$

$$l^*(\theta;\, y) = \sum_{i=1}^{n} \ell(\theta;\, y_i)$$
$$\sqrt{n}(\hat{\theta} - \theta) = \frac{\sqrt{n} l^*(\theta;\, y)}{-\mathbb{E}(l^{**}(\theta;\, Y))}$$

entrambi i membri dell'equazione hanno distribuzione asintotica normale centrata nello zero; definite le seguenti quantità:

$$V_g(\theta) = \mathbb{V}(g(\theta;\, Y_i)) = \mathbb{E}(g(\theta;\, Y_i) g(\theta;\, Y_i)^T)$$
$$D_g(\theta) = -\mathbb{E}(\tfrac{\partial}{\partial\theta} g(\theta;\, Y_i))$$

$$\mathbb{V}(\ell(\theta;\, Y_i)) = \mathbb{E}(\ell(\theta;\, Y_i)\ell(\theta;\, Y_i)^T) = i(\theta)$$
$$-\mathbb{E}(\ell(\theta;\, Y_i)) = i(\theta)$$

dobbiamo calcolare la varianza del membro di destra:

$$\mathbb{V}\left(\frac{\sqrt{n} q(\theta;\, y)}{-\mathbb{E}(\frac{\partial}{\partial\theta} q(\theta;\, Y))}\right) = \frac{n(nV_g(\theta))}{(nD_g(\theta))^2}$$

$$\mathbb{V}\left(\frac{\sqrt{n} l^*(\theta;\, y)}{-\mathbb{E}(l^{**}(\theta;\, Y))}\right) = \frac{nI(\theta)}{I(\theta)^2} = \frac{1}{i(\theta)}$$

ora possiamo scrivere la distribuzione dello stimatore:

$$\tilde{\theta} \,\dot{\sim}\, N\left(\theta;\, \frac{V_g(\theta)}{nD_g(\theta)^2}\right),\ p = 1$$
$$\tilde{\theta} \,\dot{\sim}\, N_p\left(\theta;\, \tfrac{1}{n}[(D_g(\theta)^{-1})^T\, V_g(\theta)\, D_g(\theta)^{-1}]\right),\ p \geq 1$$

$$\hat{\theta} \,\dot{\sim}\, N_p\left(\theta;\, \frac{1}{I(\theta)}\right),\ p = 1$$

<u>Oss:</u> In generale $V_g(\theta) \neq D_g(\theta)$, ma se si considera l'equazione di stima modificata:

$$\tilde{q} = D_g(\theta)^T V_g(\theta)^{-1} q(\theta;\, y)$$

essa ha la stessa soluzione $\tilde{\theta}$ di $q(\theta;\, y)$, è non distorta, e soddisfa anche $\mathbb{V}\left(\tilde{q}(\theta;\, Y)\right) = -\mathbb{E}\left(\frac{\partial}{\partial\theta}\tilde{q}(\theta;\, Y)\right)$.

Stimatore sandwitch

Una scorretta specificazione del modello $\mathscr{F}$ può far cadere la proprietà $\mathbb{V}(l^*(\theta;\, Y)) = I(\theta)$, perciò una stima consistente della varianza asintotica di $\hat{\theta}$ è fatta considerando $l^*(\theta;\, y)$ come una generica

$q(\theta;\, y)$, cioè senza effettuare le usuali semplificazioni:

$$\begin{aligned}\hat{\mathbb{V}}(\hat\theta) &= \tfrac{1}{n}[(D_g(\hat\theta)^{-1})^T\, V_g(\hat\theta)\, D_g(\hat\theta)^{-1}] \\ &= \mathbb{E}(l^{**}(\hat\theta;\, Y))^{-1}\mathbb{V}(l^{*}(\hat\theta;\, Y))(\mathbb{E}(l^{**}(\hat\theta;\, Y))^{-1})^T \\ &= J(\hat\theta)^{-1}\left[\textstyle\sum_{i=1}^n \ell(\hat\theta;\, y_i)\ell(\hat\theta;\, y_i)^T\right] J(\hat\theta)^{-1}\end{aligned}$$

dove $\hat{\mathbb{V}}(\hat\theta)$ è detto stimatore sandwich.

Oss: Se $\mathscr{F}$ è correttamente specificato si avrà $J(\hat\theta)^{-1}\left[\sum_{i=1}^n \ell(\hat\theta;\, y_i)\ell(\hat\theta;\, y_i)^T\right] = \mathbb{1}_n$, dove $\mathbb{1}_n$ è la matrice identità di dimensione nxn.

2.3 Quasi-verosimiglianza

Se è possibile reperire una $l_Q(\theta)$ primitiva di $q(\theta;\, y)$, utilizzabile per definire test e regioni di confidenza nel modo usuale, allora essa è detta funzione di quasi-verosimiglianza basata sullo quasi-score $q(\theta;\, y)$.

Se $p = 1$ e $\Theta \subset \mathbb{R}$, preso un $\theta_0 \in \Theta$ arbitrario , si ha $l_Q(\theta) = \int_{\theta_0}^{\theta} q(t,\, y)\mathrm{d}t$.

Se $p > 1$, $\Theta \subset \mathbb{R}^p$, e $q \in \mathrm{C}^1(\theta)$, è necessario per l'esistenza di l_Q che $J_Q(\theta) = -\frac{\partial}{\partial\theta}q(t;\, y)dt$ sia una matrice simmetrica.

NB: $W_Q(\theta) = 2\left[l_Q(\tilde\theta) - l_Q(\theta)\right] \dot\sim \chi^2_p \quad \Leftrightarrow \quad V_g(\theta) = D_g(\theta)$.

2.4 Funzionali statistici e funzione di influenza

Si vuole un metodo che indipendentemente dal tipo di modello $\mathscr{F}$ specificato, generi uno stimatore robusto rispetto a dati anomali.

Una soluzione è data dall'utilizzo della funzione di ripartizione empirica $\hat F$ per stimare il parametro ignoto.

Def: Poichè si può considerare $\theta = T(F)$, allora diventa possibile stimare θ con $\hat\theta = T(\hat F)$, dove $T(\cdot)$ è detto funzionale statistico. È infatti possibile ricavare $T(\cdot)$ come:

$$\mathbb{E}\left\{q(\theta;\, Y_i)\right\} = 0 \;\Leftrightarrow\; \int g(\theta;\, y)\, \mathrm{d}F(\theta;\, y) = 0 \;\Leftrightarrow\; \int g(T(\theta);\, y)\, \mathrm{d}T(\theta) = 0 \quad \rightsquigarrow \quad T(\theta)$$

oppure in base alla funzione d'influenza.

Def: La funzione d'influenza $IF(c; T, F)$ rappresenta l'effetto su $T(F)$ di una contaminazione infinitesimale nel punto c; per uno stimatore definito da una $q(\theta;\, y) = \sum_{i=1}^n g(\theta;\, y_i) = 0$ la IF assume la forma:

$$IF(c; g, F) = \frac{g(\theta;\, c)}{-\mathbb{E}_F\left(\frac{\partial}{\partial\theta} g(\theta;\, Y_i)\right)}$$

lo stimatore $\tilde{\theta}$ è detto robusto $\quad\Leftrightarrow\quad$ la funzione $g(\theta;\, y_i)$ è limitata.

2.4.1 Stimatore di Huber

Si ottengono stimatori robusti con proprietà di ottimalità utilizzando funzioni di stima tali che:

$$g^*(\theta;\, y_i) = [\ell(\theta;\, y_i) - a]_{-b}^{b} = \begin{cases} -b & ,\ [\cdot] < -b \\ [\cdot] & ,\ -b \leq [\cdot] \leq b \\ b & ,\ [\cdot] > b \end{cases}$$

dove $a = \mathbb{E}(l^*(\theta;\, y))$ per garantire la non distorsione di $q(\theta;\, y)$, mentre b viene generalmente scelto in modo tale che l'efficienza di $\tilde{\theta}$ sia inferiore di un 5% rispetto quella di $\hat{\theta}$.

Capitolo 3

Inferenza bayesiana

3.1 Il teorema di Bayes

Dati due eventi A e B, la regola del prodotto è definita come:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Applicando più volte la regola precedente si ottiene la formula di Bayes:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B \cup A) + \mathbb{P}(B \cup A^c)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)}$$

3.1.1 Versione generale

In generale, dati gli eventi $A_1, \ldots, A_n$ a due a due disgiunti di Ω:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{k=1}^{n} \mathbb{P}(B|A_k)\mathbb{P}(A_k)}$$

$\mathbb{P}(A_i) =$ prob a priori

$\mathbb{P}(A_i|B) =$ prob a posteriori

$\mathbb{P}(B|A_i) =$ verosimiglianza

3.1.2 Versione nel continuo

$$\pi(\theta|x) = \frac{\pi(\theta)f(x|\theta)}{\int_\Theta f(x|\theta)\pi(\theta)d\theta} \propto \pi(\theta)f(x|\theta)$$

$\pi(\theta) =$ densità di θ a priori

$\pi(\theta|x) =$ densità di θ a posteriori

$f(x|\theta) =$ verosimiglianza

NB: Per ottenere la $\pi(\theta|x)$ esatta, occorre prima calcolare $\pi(\theta)f(x|\theta)$ e successivamente dividerlo per la costante di normalizzazione $\int_\Theta \pi(\theta)f(x|\theta)d\theta$.

3.2 Procedure inferenziali

La densità ottenuta $\pi(\theta|x)$, rappresenta la distribuzione di probabilità del parametro d'interesse θ, condizionatamente al campione x rilevato. Grazie a questa quantità è possibile effettuare stime puntuali/intervallari, verifiche d'ipotesi, e previsioni.

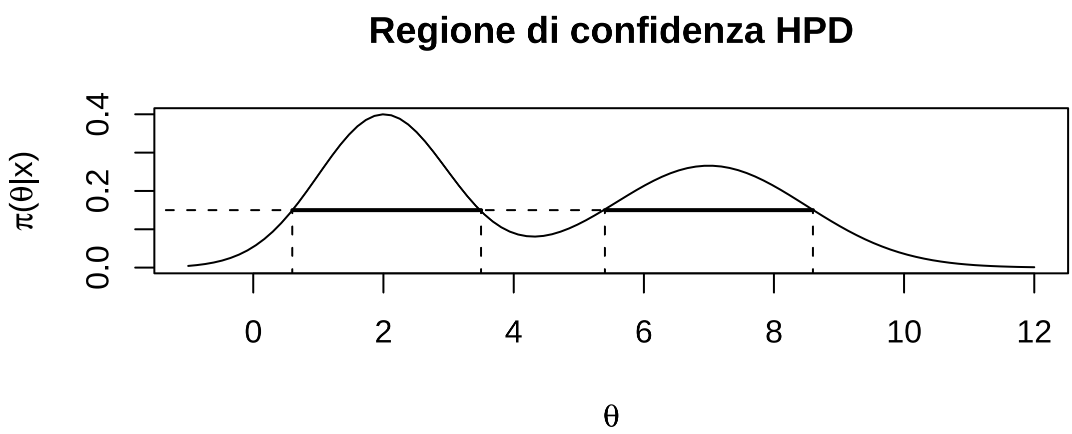
3.2.1 Stima puntuale

- valore atteso : $\hat{\theta} = \mathbb{E}(\theta|x)$
- moda (discreto) : prendo il valore con maggior densità di prob
- mediana (discreto, valori ordinati) : prendo il valore k-mo t.c. $\sum_{i=0}^{k} p(x_i) \simeq 0.5$

3.2.2 Regioni di confidenza HPD

Una regione di confidenza viene calcolata nel seguente modo:

1. si fissa una soglia c pari al valore massimo di $\pi(\theta|x)$, e la si fa decrescere progressivamente
2. in corrispondeza dei punti di intersezione della retta $\pi(\theta|x) = c$ con la curva $\pi(\theta|x)$, si identifica una regione per θ
3. si interrompe la diminuzione di c non appena la somma delle densità, dei valori appartenenti alla regione identificata, è pari ad $1 - \alpha$
4. la regione di livello $1 - \alpha$ cercata è quella ottenuta in quest'ultimo passo



La regione trovata include quindi quei valori di θ ai quali corrisponde una densità a posteriori più elevata (High Posterior Density).

NB: Se nel continuo capita di avere una $\pi(\theta|x)$ monotona crescente, basta integrarla tra a e $+\inf$ (oppure tra $-\inf$ e b se decrescente); in questo modo si ottiene una funzione in a (oppure b), e non resta che porla pari ad $1-\alpha$ per ricavare l'estremo cercato.

3.2.3 Verifica d'ipotesi

Il fattore di Bayes viene utilizzato per verificare il sistema ipotesi $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta \setminus \Theta_0$; esso consiste nel valutare il rapporto tra le densità a posteriori calcolate sotto le due diverse ipotesi:

$$B = \begin{cases} \dfrac{\pi(\theta_{H0}|x)}{\pi(\theta_{H1}|x)} & \text{per ipotesi alternative semplici } (<>) \\ \dfrac{\pi(\theta_{H0}|x)}{1-\pi(\theta_{H0}|x)} & \text{per ipotesi alternative composte } (\neq) \end{cases}$$

$B > 1$ porta all'accettazione di $H0$, mentre $B < 1$ porta al suo rifiuto.

3.3 Coniugate naturali

Fissata la densità a priori e la verosimiglianza, la a posteriori risulta già nota:

A priori $\pi(\theta)$	**Verosimiglianza** $f(x|\theta)$	**A posteriori** $\pi(\theta|x)$
Beta(a,b)	Bin(m,θ)	Beta$(a+\sum x_i, b+m\cdot n-\sum x_i)$
$\Gamma(a,b)$	Exp(θ)	$\Gamma(a+n, b+\sum x_i)$
N(η,τ^2)	N(μ,σ_0^2)	N$\left(\dfrac{\sigma_0^2\eta+\tau^2 n\overline{x}}{\sigma_0^2+\tau^2 n}, \dfrac{\tau^2\sigma^2}{\sigma_0^2+\tau^2 n}\right)$
$\Gamma(a,b)$	Pois(θ)	$\Gamma(a+\sum x_i, b+n)$
Pareto(a,b)	U$(0,\theta)$	Pareto$\big(a+n, \max(b, x_{(n)})\big)$

NB: La numerosità del campione x è n.

NB: In N(μ,σ_o^2), il parametro d'interesse è μ con σ_0^2 noto.

Appendice A

Richiami

A.1 Distribuzioni note

A.1.1 Alcune nozioni fondamentali

$Z \sim N(0,\, 1)$ $\qquad Z^2 \sim \chi^2_1$ $\qquad \sum_{i=1}^{v} \chi^2_1 \sim \chi^2_v$ $\qquad Z/\sqrt{\chi^2_v/v} \sim t_v$ $\qquad \dfrac{\chi^2_{v_1}/v_1}{\chi^2_{v_2}/v_2} \sim F_{v_1,\, v_2}$

A.1.2 Distribuzione di media e varianza campionaria

Sia $X = (X_1,\, \dots,\, X_n)$ vettore aleatorio con $X_i \sim N(\mu,\, \sigma^2),\ i = 1,\, \dots,\, n$, allora:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \bar{X} \sim N(\mu,\, \tfrac{\sigma^2}{n})$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad S^2 \sim \frac{\sigma^2}{n-1}\chi^2_{n-1}$$

Oss: Grazie al teorema del limite centrale, questi risultati valgono anche asintoticamente.

A.2 Funzioni indicatrici

- $\mathbb{1}_{[a,\, b]}(x) = \mathbb{1}_{[-\infty,\, x]}(a) \cdot \mathbb{1}_{[x,\, +\infty]}(b)$
- $\mathbb{1}_{[a,\, b]}(x) = \mathbb{1}_{\left[\frac{1}{b},\, \frac{1}{a}\right]}(\frac{1}{x})$
- $\mathbb{1}_{[a,\, b]}(x) = \mathbb{1}_{[ac,\, bc]}(xc)$
- $\mathbb{1}_{[a,\, b]}(x) = \mathbb{1}_{[a+c,\, b+c]}(x + c)$

A.3 Disuguaglianza di Jensen

- $g(\cdot)$ convessa "$\bigcup$" $\Rightarrow \mathbb{E}\left(g(X)\right) \geq g\left(\mathbb{E}(X)\right)$
- $g(\cdot)$ concava "$\bigcap$" $\Rightarrow \mathbb{E}\left(g(X)\right) \leq g\left(\mathbb{E}(X)\right)$

A.4 Newton-Raphson

Quando non è possibile calcolare in modo esplicito la SMV, occorre allora procedere tramite metodi numerici. Viene spesso usato il metodo iterativo di Newton-Raphson il quale, dato un punto iniziale di partenza, genera una successione di punti sempre più prossimi al valore cercato.

L'equazione di aggiornamento di indice $s = 0,\, 1,\, \ldots$ è la seguente:

$$\begin{aligned} \hat{\theta}_{s+1} &= \hat{\theta}_s - \frac{l'(\hat{\theta}_s)}{l''(\hat{\theta}_s)} & , \quad \dim(\theta) = 1 \\ \hat{\theta}_{s+1} &= \hat{\theta}_s - \Big(\mathbb{J}l(\hat{\theta}_s)\Big)^{-1} \cdot \nabla l(\hat{\theta}_s) & , \quad \dim(\theta) > 1 \end{aligned}$$

Il punto iniziale di partenza viene scelto tramite un metodo alternativo (ad es. quello dei momenti).

A.5 Cambio di variabile

Def: Una funzione g è detta diffeomorfismo $\quad \Leftrightarrow \quad \exists\, h = g^{-1} \;\wedge\; g,\, h \in \mathrm{C}^1$

Def: Data una funzione $f(x) : \mathscr{X} \to \mathbb{R}$ ed un diffeomorfisfmo $g(x) : \mathscr{X} \to \mathscr{Y}$, si effettua un cambio di variabile in f ponendo $y = g(x)$, perciò $x = g^{-1}(y) = h(y)$, e:

$$f(y) = \Big|\det\big(\mathbb{J}h(y)\big)\Big| \cdot f(h(y))$$

NB: Si effettua un cambio di variabile, non di parametro!

A.6 Metodo delta

Def: Dato un $\hat{\theta} \xrightarrow{d} N_p\big(\theta,\, 1/I(\theta)\big)$, se si effettua una riparametrizzazione $\psi = g(\theta),\ g \in \mathrm{C}^2$ allora:

$$\hat{\psi} = g(\hat{\theta}) \xrightarrow{d} N_p\Big(\psi,\, \big(\Delta^T \cdot I(\theta)|_{\psi=\psi(\theta)} \cdot \Delta\big)^{-1}\Big) \,, \qquad \Delta = \begin{pmatrix} \nabla g^{-1}(\psi_1) \\ \vdots \\ \nabla g^{-1}(\psi_p) \end{pmatrix}$$

dove $p = \dim(\theta) = \dim(\psi)$, e Δ è la matrice delle derivate parziali prime delle componenti di $\theta = g^{-1}(\psi)$ (cioè si derivano le componenti di θ dopo averle esplicitate in funzione di ψ).

A.7 Quantità pivotale

Data una funzione $a : \mathscr{Y} \times \Theta \to \mathbb{R}$, la v.c. $a(Y, \theta)$ è detta quantità pivotale se, per ogni fissato $\theta \in \Theta$, essa ha distribuzione di probabilità che non dipende da θ.

Oss: Il p-value α^{oss} è per costruzione una qt. pivotale, in quanto $\alpha^{oss}(Y, \theta) \overset{H_0}{\sim} U(0, 1)$.

A.8 Partizione di un insieme

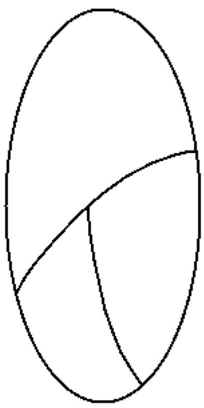
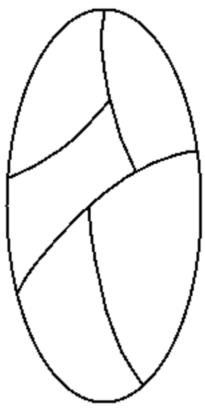
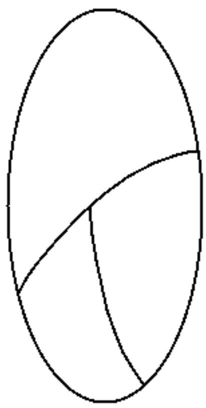
Una partizione di un insieme $\mathscr{Y}$ è definita come una collezione di sottoinsiemi di $\mathscr{Y}$ non vuoti, mutuamente disgiunti, e tali che la loro unione è l'insieme $\mathscr{Y}$ stesso.

I sottoinsiemi che costituiscono una partizione vengono detti parti, orbite o curve di livello.

Es. 1:

$$\mathscr{Y} = \{1, 2, 3, 4, 5, 6, 7, 8\} \quad \rightsquigarrow \quad P = \{\{1, 3, 5, 7\}, \{2, 6\}, \{4, 8\}\}, \quad \#P = 3$$

Es. 2:

partizioni indotte su $\mathscr{Y}$		
verosim.	*stat. suff.*	*stat. suff. min.*
P_V	P_{ss}	P_{ssm}
		
$\#P_V = 3$	$\#P_{ss} = 5$	$\#P_{ss} = 3$

A.9 Promemoria per la risoluzione di un problema

Per qualunque problema si devono effettuare i seguenti passi:

- identificare il vettore aleatorio relativo all'esperimento
- identificare il parametro d'interesse ($p \geq 1$)

mentre i restanti variano a seconda del tipo di quesito posto.

A.9.1 Inferenza frequentista

- durante la scrittura della verosimiglianza, ricordarsi di usare la formula estesa (quella con le densità condizionate), che al più verrà ridotta se sussiste l'indipendenza tra le osservazioni

- la funzione di potenza $\pi(\theta)$ è sempre definita come la probabilità di rifiutare H_0 quando H_0 è vera; e la porzione di funzione d'interesse è sempre quella definita nell'intervallo di Θ indotto da H_1 (cioè in $\Theta \setminus \Theta_0$)

- per calcolare la SMV bisogna trovare il punto di massimo della verosimiglianza $L(\theta)$; non sempre questo si ricava tramite $l^*(\theta) = 0$ e bisogna quindi usare altri metodi alternativi

- un test del tipo definito in §1.10.1, avente una statistica T funzione monotona del TRV, è anch'esso un test ottimo (altresì detto uniformemente più potente)

A.9.2 Inferenza robusta

- lo stimatore sandwich stima la varianza dello stimatore di tipo M considerato

- uno stimatore è robusto se tutte le $g(\theta; y_i)$ (le generiche componenti di $q(\theta; y)$) sono limitate

A.9.3 Inferenza bayesiana

- nella densità a priori la variabile d'interesse coincide col parametro!!
  Es: $\theta \sim Exp(\lambda) \quad \rightsquigarrow \quad \pi(\theta) = \lambda \exp(-\lambda\theta)$

- quando necessario, per semplificare i calcoli, basta identificare solamente il nucleo di una distribuzione, vengono cioè eliminate tutte le costanti proporzionali che non dipendono dal parametro

Appendice B

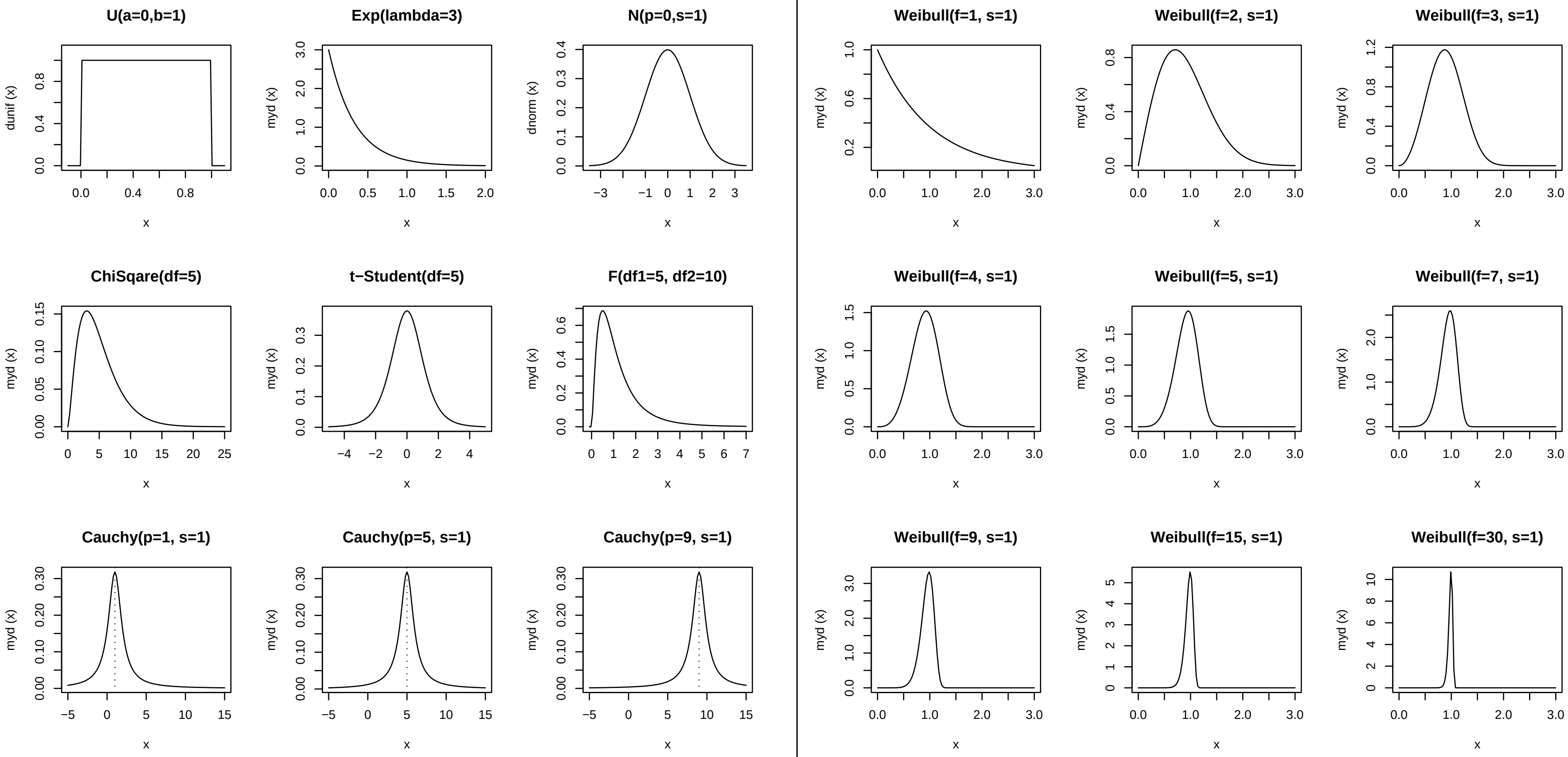
Notazioni, tabelle, e grafici

B.1 Notazioni

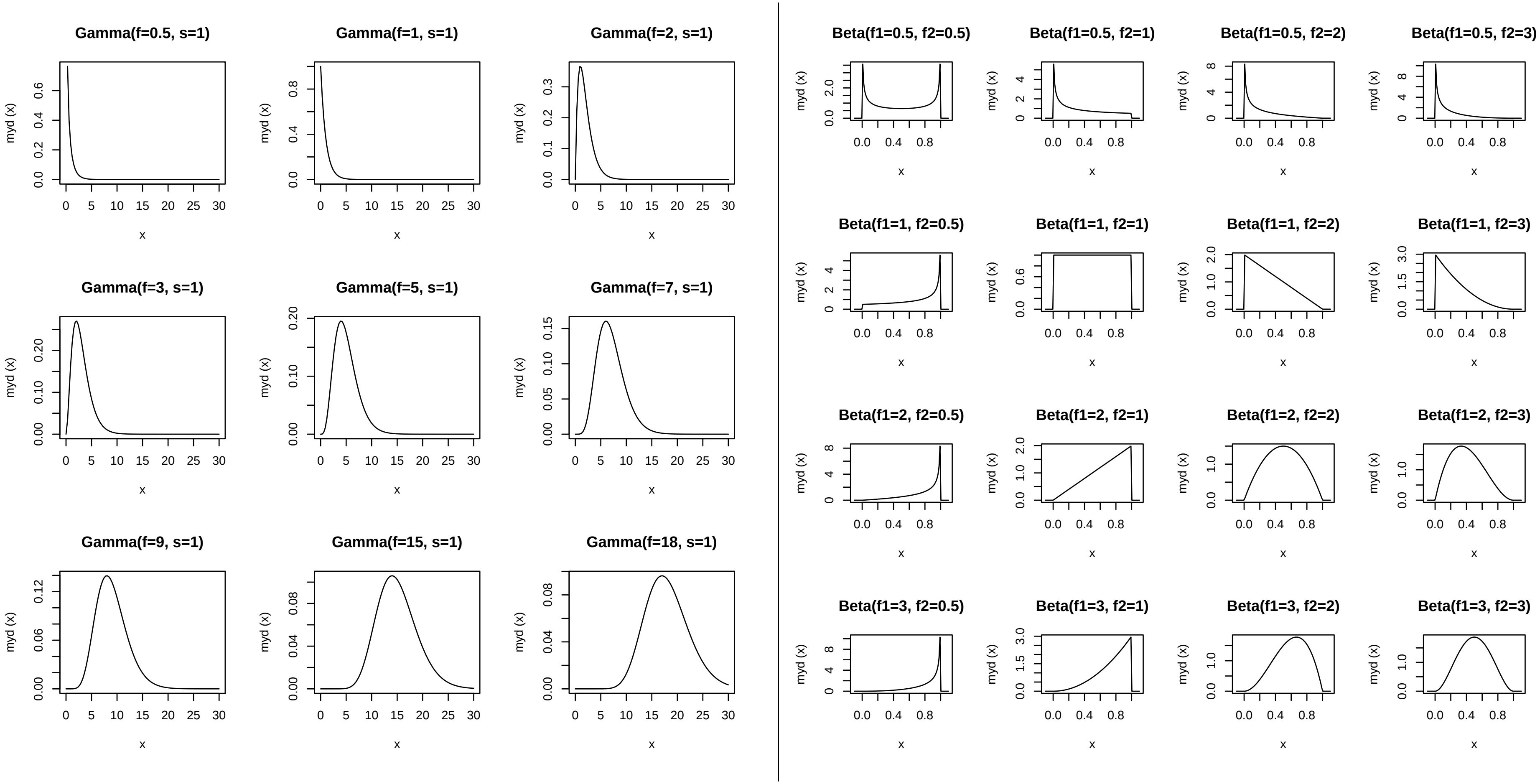
v.a.	:	variabile aleatoria
f.d.p.	:	funzione di potenza
f.d.r.	:	funzione di ripartizione
$\mathbb{1}_{[a,\ b]}(x)$	:	funzione indicatrice per l'intervallo $[a,\ b]$
$\ell(\theta,\ y_i)$	:	componente i-esima dello score di verosimiglianza $l^*(\theta,\ y) = \sum_{i=1}^{n} \ell(\theta,\ y_i)$
$\exists\cdot\ ,\ \exists!\cdot$	:	esiste un $\cdot$, esiste un unico (uno ed uno solo) $\cdot$
$\rightsquigarrow$	:	"porta a"
$\vee,\ \wedge$	:	"o, oppure", "e, congiuntamente"
$f \in$ C^n	:	f è continua e derivabile assieme alle sue derivate fino all'ordine n
$\mathbb{1}_n$	:	matrice identità di dimensioni nxn
$J(\theta)$	:	informazione osservata
$I(\theta)$	:	informazione attesa
$i(\theta)$	:	informazione attesa per una singola componente del vettore aleatorio
$t = t(y)$	:	una statistca test con le parentesi omesse per brevità
$T = t(Y)$		
∇	:	simbolo del gradiente
$\mathbb{J}$	:	simbolo del jacobiano
$X \perp\!\!\!\perp Y$	:	X è indipendente da Y
$X \not\perp\!\!\!\perp Y$	:	X è dipendente da Y
A^c	:	indica l'evento complementare di A

Tipo	Θ	$\mathscr{Y}$	**Densità**	$\mathbb{E}$	$\mathbb{V}$	**Proprietà**
$\mathrm{U}(a,\, b)$	$a < b \in \mathbb{R}$	$[a,\ b]$	$\dfrac{1}{b-a}\mathbb{I}_{[a,b]}(y)$	$\dfrac{(a+b)}{2}$	$\dfrac{(b-a)^2}{12}$	—
$\mathrm{Exp}(\lambda)$	$\lambda > 0$	$\mathbb{R}^+$	$\lambda \exp\{-\lambda y\}$	$1/\lambda$	$1/\lambda^2$	—
$\mathrm{N}(\mu,\, \sigma^2)$	$\mu \in \mathbb{R},\ \sigma > 0$	$\mathbb{R}$	$\dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\dfrac{1}{2\sigma^2}(y-\mu)^2\right\}$	μ	σ^2	$\mathrm{Bi}(n,\, \pi) \,\dot{\sim}\, \mathrm{N}(n\pi,\, n\pi(1-\pi))$ $\mathrm{Pois}(\lambda) \,\dot{\sim}\, \mathrm{N}(\lambda,\, \lambda),\ \lambda > 10$ $\chi^2_v \,\dot{\sim}\, \mathrm{N}(v,\, 2v),\ v > 30$
$\Gamma(a,\, b)$	$a,\, b > 0$	$\mathbb{R}^+$	$\dfrac{b^a}{\Gamma(a)} y^{a-1} \exp\{-by\}$	a/b	a/b^2	$\mathrm{Exp}(b) \sim \Gamma(1,\, b)$ $\chi^2_v \sim \Gamma(v/2,\, 1/2)$ $\Gamma(n) = (n-1)!\,,\ n \in \mathbb{N} \setminus \{0\}$
$\mathrm{Beta}(a,\, b)$	$a,\, b > 0$	$[0,\ 1]$	$\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1}$	$\dfrac{a}{a+b}$	$\dfrac{ab}{(a+b)^2(a+b+1)}$	$\mathrm{Beta}(1,\, 1) \sim \mathrm{U}(0,\, 1)$

Tabella B.1: Principali densità e relative proprietà



$f =$ forma, $s =$ scala, $p =$ posizione, $df =$ gradi di libertà



f = forma, s = scala, p = posizione, df = gradi di libertà

Si ricorda che questo documento rappresenta degli appunti personali e che gli autori, nonostante i loro sforzi, non ne garantiscono in alcun modo la correttezza dei contenuti.

Per errori e/o commenti mandate una mail con oggetto "*statcp*" all'indirizzo `steelgunblade@yahoo.it`

Versione 1.3 – Finito di stampare il 23 giugno 2008 con LaTeX