# 4 – Metodi statistici

M. Bertero – DISI – Università di Genova

- Metodo di massima verosimiglianza
- Il caso di noise Gaussiano
- Il caso di noise Poissoniano

Se si conoscono le proprietà statistiche del noise, è possibile scrivere esplicitamente la distribuzione di probabilità dell'immagine, per un dato oggetto f. Pertanto l'oggetto gioca il ruolo di un insieme di parametri da cui dipende la distribuzione di probabilità. Questa proprietà suggerisce la possibiltà di formulare il problema di deconvoluzione come un problema di massima verosimiglianza. Nel caso di noise Gaussiano si ritrova il problema ai minimi quadrati, mentre si giunge ad un nuovo problema nel caso di noise Poissoniano. Infine si considera la possibilità di regolarizzare il problema in ambito probabilistico con i cosiddetti metodi Bayesiani.

# 4.1 – Metodi di massima verosimiglianza

Come si è visto nel primo capitolo, il valore misurato g(n) dell'immagine nel pixel n, può essere considerato come la realizzazione di una variabile aleatoria $\eta$(n), il cui valore atteso è dato dalle equazioni (1.5) e (1.4). Usando notazioni successivamente introdotte, possiamo condensare queste equazioni nella forma seguente:

$$E\left\{\eta\left(\underline{n}\right)\right\} = \left(A\overline{f}\right)\left(\underline{n}\right) \qquad . \qquad (4.1)$$

Al fine di introdurre i metodi di massima verosimiglianza nella loro generalità, è opportuno analizzare in dettaglio i due modelli di noise introdotti nel primo capitolo.

Incominciamo dal caso più semplice, cioè quello di noise Gaussiano bianco . In questo caso, la variabile aleatoria $\eta$ (n) è data dal seguente modello:

$$\eta\left(\underline{n}\right) = \left(Af\right)\left(\underline{n}\right) + \varsigma \qquad , \qquad (4.2)$$

dove $\zeta$ è una variabile aleatoria, con valore atteso nullo, che non dipende da n. La relazione è stata scritta per un generico oggetto f, non necessariamente quello che corrisponde ai dati osservati.

Nel caso di noise Gaussiano, la densità di probabilità della variabile aleatoria z è data da:

$$P_{\varsigma}(w) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{w^2}{2\sigma^2}} \quad , \qquad (4.3)$$

e pertanto, dalla (4.2) segue che la densità di probabilità della variabile $\eta$(n) ha l'espessione seguente:

$$P_{\eta(\underline{n})}(g(\underline{n}) \mid f) = P_{\zeta}(g(\underline{n}) - (Af)(\underline{n})) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|g(\underline{n}) - (Af)(\underline{n})|^2}{2\sigma^2}} \quad .$$

Un'ipotesi del tutto naturale, che è anche ben verificata in pratica, consiste nell'ammettere che le variabili aleatorie associate a pixel diversi siano tra di loro statisticamente indipendenti. In questo caso, la densità di probabilità congiunta delle variabili aleatorie associate ai vari pixel è data da:

$$P_{\eta}(g \mid f) = \prod_{\underline{n}} P_{\eta(\underline{n})}(g(\underline{n}) \mid f) \quad , \qquad (4.4)$$

dove si è indicato con $\eta$ l'insieme delle variabili aleatorie $\eta$(n). Si è inoltre indicato esplicitamente che tali densità di probabilità dipendono da f.

Infine, combinando le varie equazioni e ricordando che si hanno $N^2$ pixel, si ottiene:

$$P_\eta(g \mid f) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{N^2} e^{-\frac{\|Af - g\|^2}{2\sigma^2}} \quad . \qquad (4.5)$$

Nel caso di noise Poissoniano il valore g(n) dell'immagine al pixel n è la realizzazione di una variabile aleatoria $\eta(n)$ con distribuzione di probabilità:

$$P_{\eta(\underline{n})}(g(\underline{n}) \mid f) = \frac{e^{-(Af)(\underline{n})}(Af)(\underline{n})^{g(\underline{n})}}{g(\underline{n})!} \quad . \qquad (4.6)$$

Anche in questo caso, se si ammette che i processi associati a pixel diversi sono statisticamente indipendenti, la distribuzione di probabilità di un'immagine può essere ottenuta combinando la (4.4) con la (4.6).

Sia nel caso Gaussiano che nel caso Poissoniano si ha una esplicita dipendenza della distribuzione di probabilità da oggetto e immagine.

I metodi di massima verosimiglianza sono basati sulleseguenti ipotesi e definizioni:

$i)$ si ammette che l'immagine sia la realizzazione di una variabile aleatoria a piu' valori $\eta$, che dipende dall'oggetto incognito $f$ e che la distribuzione di probabilità di $\eta$ abbia una dipendenza funzionale nota $P_\eta(g \mid f)$ sia da $g$ che da $f$;

*ii)* per una data immagine $g$ si definisce *funzione di verosimiglianza* associata a $g$ la funzione di $f$ che si ottiene sostituendo i valori di $g$ in $P_\eta(g \mid f)$:

$$L_g(f) = P_\eta(g \mid f) \quad . \qquad (4.7)$$

$iii)$ il *principio di massima verosimiglianza* consiste nel scegliere come valutazione dell'oggetto incognito $\overline{f}$ un oggetto che renda massima la funzione verosimiglianza.

Si ha dunque:

$$f_{ML} = \arg\max_f L_g(f) \quad . \qquad (4.8)$$

Poichè, nel caso di immagini, la funzione di verosimiglianza è il prodotto di un gran numero di funzioni è conveniente considerare il logaritmo di tale funzione, la cosiddetta funzione di verosimiglianza logaritmica. Poichè il logaritmo è una funzione concava e crescente, i massimi della funzione logaritmica coincidono con quelli della funzione di verosimiglianza e dunque, posto:

$$l_g(f) = \ln L_g(f) \quad , \qquad (4.9)$$

si ha anche:

$$f_{ML} = \arg\max_f l_g(f) \quad . \qquad (4.10)$$

Tuttavia, per ragioni che saranno chiare tra poco, può essere opportuno trattare i problemi di massima verosimiglianza come problemi di minimo, introducendo funsionali del tipo:

$$J(f \mid g) = -B \ln L_g(f) + C \quad . \qquad (4.11)$$

# 4.2 – Il caso di noise Gaussiano

Nel caso di noise Gaussiano addittivo e bianco, la funzione di verosimiglianza è data dalla (4.5). Pertanto, se si sostituisce nella (4.11), ponendo:

$$B = 2\sigma^2 \quad , \quad C = -N^2 \ln\left(\sqrt{2\pi}\sigma\right) \quad ,$$

si ottiene:

$$J\left(f \mid g\right) = \left\| Af - g \right\|^2 \quad .$$

Pertanto, in questo caso, il metodo della massima verosimiglianza coincide con il metodo ai minimi quadrati, anzi, ne fornisce la giustificazione statistica.

Questo fatto mette anche in luce che, verosimilmente, i metodi di massima verosimiglianza sono mal posti nel caso di problemi di deconvoluzione.

# 4.3 – Il caso di noise Poissoniano

Dalle equazioni (4.4) e (4.6), sostituendo nella (4.11) con:

$$B = 1 \quad , \quad C = g(\underline{n}) \ln g(\underline{n}) - g(\underline{n}) - \ln[g(\underline{n})!] \ ,$$

si ottiene:

$$J(f \mid g) = \sum_{\underline{n}} \left\{ g(\underline{n}) \ln \frac{g(\underline{n})}{(Af)(\underline{n})} + (Af)(\underline{n}) - g(\underline{n}) \right\} \quad . \quad (4.12)$$

Questo funzionale viene detto divergenza di Csiszàr, è definito sul cono degli oggetti non-negativi e gode delle seguenti proprietà:

$i)$ $\quad J(f \mid g) \geq 0 \quad ;$

$ii)$ $\quad$ per ogni fissato $g$, $J(f \mid g)$ è una funzione convessa di $f$ .

Per dimostrare la i) si osservi che, per ogni fissato a, la funzione:

$$\psi(x) = a(\ln a - \ln x) + x - a \quad ,$$

definita per x > 0, è una funzione convessa con un unico punto di minimo per x=a; quindi:

$$\psi(x) \geq \psi(a) = 0 \quad .$$

Confrontando con la (4.12), ne segue che ogni termine della somma è non-negativo; più esattamente ogni termine è strettamente positivo, dato che, in generale, a causa della presenza del noise, non esiste una f non-negativa tale che Af=g.

Per dimostrare la ii), occorre calcolare esplicitamente il gradiente e la matrice hessiana del funzionale (4.12). A tal scopo si osservi che:

$$\frac{\partial}{\partial f(\underline{n})}(Af)(\underline{n}') = A(\underline{n}', \underline{n}) \ ,$$

dove A(n’,n) è l’elemento di matrice n’,n della matrice A.

Tenendo conto della precedente relazione ed indicando con n'' l'indice di somma nella (4.12), si ottiene:

$$\frac{\partial}{\partial f(\underline{n})} J(f \mid g) = \sum_{\underline{n}''} \left\{ -A(\underline{n}'', \underline{n}) \frac{g(\underline{n}'')}{(Af)(\underline{n}'')} + A(\underline{n}'', \underline{n}) \right\} \ .$$

Poichè A è una matrice di convoluzione, dalla periodicità e dalla normalizzazione della PSF segue che:

$$\sum_{\underline{n}''} A(\underline{n}'', \underline{n}) = \sum_{\underline{n}''} K(\underline{n}'' - \underline{n}) = 1 \ .$$

Pertanto, si può scrivere il gradiente del funzionale nella forma seguente:

$$\nabla_f J(f \mid g) = -A^T \frac{g}{Af} + 1 \ , \qquad (4.13)$$

dove 1 è la tabella i cui elementi sono tutti eguali a 1 ed il quoziente di due tabelle è definito come quoziente elemento per elemento (pixel a pixel).

Quanto alla matrice Hessiana, derivando ulteriormente la formula della derivata prima si ottiene:

$$H\left(\underline{n},\underline{n}'\right)=\frac{\partial^2}{\partial f\left(\underline{n}\right)\partial f\left(\underline{n}'\right)}J\left(f\mid g\right)=$$

$$=\sum_{\underline{n}''}A\left(\underline{n}'',\underline{n}\right)\frac{g\left(\underline{n}''\right)}{\left[\left(Af\right)\left(\underline{n}''\right)\right]^2}A\left(\underline{n}'',\underline{n}'\right)\ . \qquad (4.14)$$

Pertanto, per ogni arbitraria tabella h, si ha:

$$\left(Hh\ ,h\right)=\sum_{\underline{n}}\sum_{\underline{n}'}H\left(\underline{n},\underline{n}'\right)h\left(\underline{n}\right)h\left(\underline{n}'\right)=$$

$$=\sum_{\underline{n}}\sum_{\underline{n}'}\sum_{\underline{n}''}A\left(\underline{n}'',\underline{n}\right)\frac{g\left(\underline{n}''\right)}{\left[\left(Af\right)\left(\underline{n}''\right)\right]^2}A\left(\underline{n}'',\underline{n}'\right)h\left(\underline{n}\right)h\left(\underline{n}'\right)=$$

$$=\sum_{\underline{n}''}\frac{g\left(\underline{n}''\right)}{\left[\left(Af\right)\left(\underline{n}''\right)\right]^2}\left[\left(Ah\right)\left(\underline{n}''\right)\right]^2\ \geq\ 0\quad ,$$

cioè la matrice Hessiana è semi-definita positiva e quindi il funzionale è convesso.

Poichè il funzionale è convesso e limitato inferiormente, tutti i suoi minimi sono globali; inoltre, condizioni necessarie e sufficienti per un minimo nel cono chiuso delle immagini non negative sono fornite dalle cosiddette condizioni di Kuhn-Tucker:

$$i) \quad f(\underline{n}) \frac{\partial}{\partial f(\underline{n})} J(f \mid g) = 0$$

$$ii) \quad f(\underline{n}) \geq 0 \quad , \quad \frac{\partial}{\partial f(\underline{n})} J(f \mid g) \geq 0 \ .$$

Dalla prima di queste condizioni, utilizzando l'espressione (4.13) del gradiente, si ottiene:

$$f\left( - A^T \frac{g}{Af} + 1 \right) = 0 \quad ,$$

cioè un'equazione di punto fisso del tipo:

$$f = T(f) \ , \ T(f) = f \ A^T \frac{g}{Af} \quad .$$

All'equazione di punto fisso si può applicare il metodo delle approssimazioni successive. Si ottiene cosi' il metodo iterativo detto metodo di Richardson-Lucy :

$$i)\ \text{inizializzare con } f^{(0)} > 0\,;$$

$$ii)\ \text{dato } f^{(k)}, \text{calcolare :}$$

$$f^{(k+1)} = f^{(k)} A^T \frac{g}{Af^{(k)}} \quad .$$

Si verifica facilmente che ogni iterata è automaticamente non negativa; inoltre, sommando su tutti i pixels:

$$\sum_{\underline{n}} f^{(k+1)}(\underline{n}) = \sum_{\underline{n}} f^{(k)}(\underline{n}) \sum_{\underline{n}'} A(\underline{n}', \underline{n}) \frac{g(\underline{n}')}{\left(Af^{(k)}\right)(\underline{n}')} =$$

$$= \sum_{\underline{n}'} \frac{g(\underline{n}')}{\left(Af^{(k)}\right)(\underline{n}')} \sum_{\underline{n}} A(\underline{n}', \underline{n}) f^{(k)}(\underline{n}) = \sum_{\underline{n}'} g(\underline{n}') \ .$$

Pertanto tutte le iterate hanno lo stesso flusso dell'immagine.