1 Problemi sollevati dagli studenti

1. Come cambiano gli intervalli di confidenza qualora lo stimatore NON sia corretto?

Assumiamo sempre l'ipotesi che lo stimatore abbia densità Gaussiana (altrimenti i conti cambiano per forza. La deduzione dell'intervallo a 2 SD per avere probabilià del 95% è stata fatta nell'ipotesi Gaussiana. Cambiando la forma della densità cambia per forza l'intervallo), quindi che sia

$$\hat{\theta}(y) \sim \mathcal{N}(\mu, \sigma^2), \ b(\theta) = \mu - \theta \ (\mu = E[\hat{\theta}(y)])$$

dove $b(\theta)$ rappresenta il BIAS. In tal caso vale

$$P[\mu - 2\sigma \leq \hat{\theta}(y) \leq \mu + 2\sigma] = 95\%$$

Semplici conti trasformano l'intervallo citato in

$$\theta \in [\ \hat{\theta}(y) - b(\theta) - 2\sigma \ \ \hat{\theta}(y) - b(\theta) + 2\sigma\ ]$$

Quindi l'intervallo ha sempre ampiezza 4σ, ma è CENTRATO su un diverso punto, che è $\hat{\theta}(y) - b(\theta)$ (che coincide con $\hat{\theta}(y)$ solo se il BIAS è nullo). Se conosciamo il BIAS allora, in corrispondenza ad una realizzazione del vettore aleatorio y (cioè all'effettuazione di una misura), possiamo costruire comunque un intervallo di confidenza al 95%, ma se non conosciamo il BIAS allora è il gatto che si mangia la coda: θ_0 viene (con una certa probabilità) ad essere contenuto in un intervallo che ... dipende da θ_0 stesso! Tuttavia spesso il problema può essere risolto risolvendo delle disequazioni. Segue un ESEMPIO: assumiamo che il BIAS valga $\frac{\theta}{2}$. Allora in tal caso l'appartenenza all'intervallo si traduce in

$$\hat{\theta}(y) - \frac{\theta}{2} - 2\sigma \leq \theta \leq \hat{\theta}(y) - \frac{\theta}{2} + 2\sigma \ \Rightarrow \ \theta \in [\ \frac{2}{3}\hat{\theta}(y) + \frac{4}{3}\sigma \ \ \frac{2}{3}\hat{\theta}(y) + \frac{4}{\sigma}\ ]$$

e come si vede l'intervallo è centrato ora in $\frac{2}{3}\hat{\theta}(y)$ ed ha ampiezza inferiore! Ma l'ampiezza diventerebbe invece superiore se il BIAS avesse segno opposto. Non ci sono quindi regole generali, tutto dipende dall'espressione, più o meno complicata, del BIAS. Solo se esso fosse COSTANTE (non dipendente da θ), si otterrebbe un intervallo simile al caso di stimatore corretto (semplicemente traslato di $-b$).

2. Supponiamo che vengano fatte delle misure y_1 (vettoriali) e che poi l'esperimento venga ripetuto, ottenendo delle misure y_2. È possibile che gli intervalli di confidenza risultino DISGIUNTI?

La risposta è SI! Anche se poco **intuitivo** (nel senso che la prima volta otteniamo che θ_0 sta con probabilià del 95% in un certo intervallo I_1, la seconda volta in un intervallo I_2, eppure gli intervalli non hanno punti in comune ... dove cavolo sta θ_0?), questo dipende dal fatto che siamo in ambito probabilistico, e TUTTO PUÒ ACCADERE, anche se è estremamente improbabile! Ad esempio, in una sequenza di testa e croce, posso ottenere 100 teste? SI, è possibile, ma nella pratica non avverrà MAI (la probabilità che avvenga é $\left(\frac{1}{2}\right)^{100} \simeq 10^{-30}$!). Dimostriamo che può accadere (con riferimento a stimatori corretti e Gaussiani come usuale): si avrebbe che gli intervalli di confidenza sarebbero

$$\theta_0 \in [\ \hat{\theta}(y_1) - 2\sigma \ \ \hat{\theta}(y_1) + 2\sigma\ ], \ \theta_0 \in [\ \hat{\theta}(y_2) - 2\sigma \ \ \hat{\theta}(y_2) + 2\sigma\ ]$$

ed assumiamo che, ad esempio, sia $\hat{\theta}(y_2) - 2\sigma > \hat{\theta}(y_1) + 2\sigma$, il che renderebbe disgiunti gli intervalli. L'unica domanda sensata da porsi è: quanto è probabile che accada questo fatto? Essa equivale a

$$P[\hat{\theta}(y_2) - \hat{\theta}(y_2) > 4\sigma]$$

ed assumendo l'ipotesi ragionevole che le misure y_2 siano IID con quelle y_1, abbiamo che $\hat{\theta}(y_2) - \hat{\theta}(y_2) \in \mathcal{N}(0, 2\sigma^2)$, da cui

$$P[\hat{\theta}(y_2) - \hat{\theta}(y_2) > 4\sigma] = \int_{4\sigma}^{+\infty} \frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)} e^{-\frac{y^2}{4\sigma^2}} dy = \int_{2\sqrt{2}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1 - erf(2)}{2} \simeq 0.23\%$$

Quindi, seppur estremamente improbabile, può accadere! In tal caso ovviamente la cosa più sensata da fare è quella di riprogettare uno stimatore che TENGA CONTO sia di y_1 che di y_2, cioè progettare un nuovo

$\hat{\theta}(y_1, y_2)$ e considerare l'intervallo di confidenza da questo generato! Più semplicemente, essendo caduti in un *caso disperato* (molto improbabile), se possibile meglio rifare l'esperimento trovando un y_3 che, con elevata probabilità, porgerà un intervallo di confidenza *simile* ad uno solo dei due iniziali disponibili.

3. Nel caso NON IID, esistono risultati generali sulle proprietà asintotiche della Stima ML?

In generale la risposta è NO, tuttavia ci sono casi in cui le proprietà permangono anche nel caso non-IID, cioè esistono teoremi che garantiscono la consistenza anche in ipotesi più deboli. Cominciamo ad esaminare meglio casi in cui la risposta è NO. L'esempio fatto a lezione (con θ_1, θ_2) può sembrare **ovvio**, in quanto θ_2 era coinvolto in UNA SOLA MISURA, ma possiamo fare anche esempi con un solo parametro! Un primo esempio è il seguente. Siano $y_i = \theta + e_1, \; i = 1, 2, \ldots, n$, con e_1 a media nulla e varianza σ^2, e NON GAUSSIANO. È evidente come le n relazioni conducano allo STESSO VALORE per y_i, cioè rappresentano la STESSA formula, in quanto lo STESSO e_1 è coinvolto in tutte le misure! Di conseguenza le misure dopo la prima non portano NESSUNA INFO aggiuntiva, e tutto è contenuto nell'UNICA equazione $y_1 = \theta + e_1$, che conduce ad uno stimatore ML in generale non Gaussiano, non corretto, non efficiente (dipende tutto dalla densità di e_1). Ma quel che è peggio, lo stimatore NON CAMBIA con n, quindi le proprietà asintotiche non ci sono proprio, nel senso che è sempre e soltanto come aver effettuato un'unica misura! Tuttavia tale esempio può sembrare talmente banale da rappresentare un *trucco* per eludere la domanda. Conviene quindi fare un esempio più articolato. Assumiamo il modello LINEARE GAUSSIANO

$$y_i = \theta + e_i, \; e_i \sim \mathcal{N}(0, \sigma_i^2), \; \text{(indipendenti, ma non IID)}, \; i = 1, 2, \ldots, n$$

La matrice Σ è diagonale ed ha σ_i^2 sulla diagonale ($i = 1, 2, \ldots, n$), mentre $\phi = [\,1 \quad 1 \quad \ldots \quad 1\,]^T$, quindi vale facilmente

$$(\phi^T \Sigma^{-1} \phi)^{-1} = \frac{1}{\Sigma_1^n \frac{1}{\sigma_i^2}}$$

Se scegliamo una qualsiasi sequenza σ_i^2 che renda convergente $\Sigma_1^{+\infty} \frac{1}{\sigma_i^2}$, ad esempio $\sigma_i^2 = i^2$, otteniamo che per $n \to +\infty$ la varianza $(\phi^T \Sigma^{-1} \phi)^{-1}$ tende ad un valore finito ma non-nullo. Di conseguenza lo stimatore (ML ed LS) di θ non ha asintoticamente varianza nulla, anche se in tal caso si ha correttezza, efficienza e Normalità per ogni n e quindi a maggior ragione asintoticamente. Quindi questo stimatore NON tende in probabilità a θ_0! La spiegazione è semplice: ogni successiva equazione, avendo varianze di errore sempre maggiori, porta sempre meno INFO (cioè le misure successive sono sempre meno significative), e tale limitato incremento di INFO impedisce di stimare θ_0 oltre una massima accuratezza, cioè la varianza non può scendere sotto un valore finito e non-nullo ottenuto con le operazioni di limite sopra effettuate.

Cerchiamo ora di capire quando e perchè la risposta potrebbe essere SI. Nel caso IID, l'inversa di Fisher ha la forma $I_n^{-1}(\theta) = \frac{I^{-1}(\theta)}{n}$, e quindi tende a zero. Ma siccome è essa stessa che rappresenta asintoticamente la varianza di $\hat{\theta}_n^{ML}(y)$, ecco che ciò implica la consistenza. Tuttavia, l'ipotesi IID **non è necessaria** per garantire che $I_n(\theta)$ tenda a zero, che è l'unica cosa che ci serve per la consistenza, quando sono verificate opportune ipotesi che garantiscono che lo stimatore sia asintoticamente corretto ed efficiente! Ciò accade in particolare per sistemi dinamici usuali (lineari, tempo-invarianti e stabili), dove gli errori mantengono costante nel tempo la loro varianza, ma esiste una CORRELAZIONE che tende a zero asintoticamente. La correlazione è una possibile causa di non-consistenza, ma se essa è *sufficientemente debole* tra campioni distanti nel tempo, $I_n^{-1}(\theta)$ tende a zero anche se non ha la forma del caso IID. Chiariamo l'importanza della correlazione con un altro esempio lineare Gaussiano.

Esempio. Sia

$$y(t) = \theta + e(t), \; t = 1, 2, \ldots, n, \; e \sim \mathcal{N}(0, \Sigma), \; \Sigma_{ij} = a^{|i-j|}, \; |a| < 1, \; a \neq 0, \; \phi = [\,1 \quad 1 \quad \ldots \quad 1\,]^T$$

Trattasi di errori e_i correlati, con la correlazione $E[e(i)e(j)] = a^{|i-j|}$ che decresce (tende a zero) al crescere della *distanza temporale* tra i vari errori. In questo caso si dimostra per induzione che vale

$$(\phi^T \Sigma^{-1} \phi)^{-1} = \frac{1 + a}{n - (n-2)a}$$

e quindi anche se le v.a. non sono IID (sono correlate, quindi non indipendenti, anche se identicamente distribuite), la varianza tende a zero per $n \to +\infty$. Modifichiamo ora semplicemente Σ, rendendo $\Sigma_{ii} = 1, \; \Sigma_{ij} = a, \; |a| < 1, \; a \neq 0, \; i \neq j$, così la correlazione tra gli $e(t)$ rimane costante e quindi non tende a zero. In tal caso si troverebbe $(\phi^T \Sigma^{-1} \phi)^{-1} = \frac{n}{1+(n-1)a}$ che tende a $\frac{1}{a}$, quindi in tal caso la varianza asintotica dello stimatore non sarebbe più nulla.