

Lezione 5

Stima Puntuale dei Parametri

Stima Puntuale

- ❑ Supponiamo di aver fatto n misure di una variabile casuale X : $x_1, x_2, \dots, x_n$.
Noi assumiamo che le n misure siano **indipendenti** tra di loro e **distribuite allo stesso modo** (i.i.d.)

- ❑ Noi vogliamo determinare la p.d.f. della variabile casuale X a partire dalle n misure fatte (il nostro campione di dati). Questa p.d.f. in generale non è conosciuta.

- ❑ Spesso è noto a priori che la p.d.f. della variabile studiata appartiene ad una famiglia parametrica (per esempio una gaussiana) con uno o più parametri liberi θ

$$f(x) = f(x | \theta) \quad \theta \in \Theta$$

- ❑ A partire dalle n misure fatte noi vogliamo determinare i parametri θ (inferenza statistica). Per stimare questi parametri devo introdurre gli stimatori. Si dice fit di un parametro il processo che porta alla stima di un parametro.

Stimatori Puntuali

- ❑ Supponiamo di dover stimare un parametro per esempio il valore medio dell'altezza di n studenti. Posso fare questa stima in vari modi. Per esempio:
 - 1) Sommo le altezze degli n studenti e divido per n (media aritmetica);
 - 2) Sommo le altezze degli n studenti e divido per $n-2$;
 - 3) Sommo le altezze dei primi dieci studenti e faccio la media aritmetica;
 - 4) Sommo le altezze dei due studenti più alti e dei due più bassi e divido per 4;
 - 5) Tolgo i 3 studenti più alti ed i 3 più bassi e faccio la media aritmetica dei rimanenti;
 - 6) Moltiplico le n altezze tra di loro e poi estraggo la radice n -esima (media geometrica)
 - 7) Sommo le altezze del secondo studente, del quarto, ecc e poi divido per $n/2$;
- ❑ Ognuno dei 7 procedimenti rappresenta uno stimatore della media. Questi sette valori medi sono in generale diversi.
- ❑ Si noti che ognuno di questi stimatore non sono altro che una funzione delle misure fatte

Statistica

- ❑ Una qualunque funzione delle n misure effettuate

$$t = t(x_1, x_2, \dots, x_n)$$

è detta **statistica**

- ❑ Uno stimatore puntuale è una statistica introdotta allo scopo di stimare il parametro θ (o i parametri θ) di una p.d.f.

- ❑ Lo stimatore del parametro θ noi lo indicheremo con $\hat{\theta}$

- ❑ L'uguaglianza tra stimatore e valore θ da stimare $\hat{\theta} = \theta$ è solo approssimata: Lo stimatore darà un valore il più vicino possibile al valore vero θ .

- ❑ L'insieme di n misure posso pensarlo come un unico esperimento :

$$x_1, x_2, \dots, x_n = \bar{x}$$

- ❑ Posso pensare di rifare tante volte l'esperimento. In ogni esperimento avrò un determinato punto nello spazio ad n dimensioni.

Statistica

- ❑ Lo stimatore $\hat{\theta}(\bar{x})$ assumerà nello spazio ad n dimensioni valori diversi distribuiti secondo una p.d.f. $g(\hat{\theta}; \theta)$ che dipende dal valore vero θ del parametro
- ❑ La p.d.f. di uno stimatore è detta p.d.f. di campionamento
- ❑ Tutte le misure sono supposte i.i.d. e quindi se $f(x_i)$ è la p.d.f. della i-sima misura allora la p.d.f. congiunta delle n misure è:

$$f_{\text{campionamento}}(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \cdots f(x_n)$$

- ❑ Il valore di aspettazione di uno stimatore $\hat{\theta}$ con una p.d.f. di campionamento $g(\hat{\theta}; \theta)$ è dato da:

$$E[\hat{\theta}(\vec{x})] = \int \hat{\theta} g(\hat{\theta}; \theta) d\hat{\theta} = \iint \theta(\vec{x}) f(x_1; \theta) \cdots f(x_n; \theta) dx_1 \cdots dx_n$$

Proprietà degli Stimatori

- ❑ Noi diciamo che uno stimatore è buono se è:
non distorto, consistente ed efficiente
- ❑ Si dice distorsione (o bias) b la differenza tra il valore di un parametro stimato dallo stimatore ed il suo valore vero: $b = E[\hat{\theta}] - \theta$ Lo stimatore è **non distorto** se $b = 0$
- ❑ Lo stimatore media aritmetica è non distorto. Infatti

$$E[\hat{a}] = E\left[\frac{a_1 + a_2 + \dots + a_n}{n}\right] = \frac{E[a] + E[a] + \dots + E[a]}{n} = \frac{nE[a]}{n} = a$$

- ❑ Per lo stimatore n. 2 (tre slide precedenti) si ha invece: $E[\hat{a}] = \frac{n}{n-2}a$
Questo stimatore è distorto. Questo stimatore è **asintoticamente non distorto**

Proprietà degli Stimatori

- Uno stimatore $\hat{a}$ di un parametro a è detto **consistente** se il valore stimato all'aumentare delle dimensioni del campione tende al valore vero:

$$\lim_{n \rightarrow \infty} \hat{a} = a$$

- Lo stimatore media aritmetica è uno stimatore consistente così come il secondo stimatore della altezza media appena visto. Il terzo stimatore non è consistente
- Lo stimatore più **efficiente** tra quelli che stimano un parametro è quello che stima il parametro con la minore varianza. Lo stimatore della media n. 7 visto prima è meno efficiente dello stimatore media aritmetica perché ha una varianza $\sqrt{2}$ volte maggiore.
- Un'altra misura della qualità di uno stimatore è l'errore quadratico medio (MSE) così definito: $MSE = E[(\hat{\theta} - \theta)^2]$

Proprietà degli Stimatori

- Si ha l'identità :

$$E[(\hat{\theta} - \theta)^2] = E[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2] = E[(\hat{\theta} - E[\hat{\theta}])^2] + (E[\hat{\theta}] - \theta)^2$$

- La seconda uguaglianza sussiste perché

$$E[(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta)] = (E[\hat{\theta}] - \theta)E[\hat{\theta} - E[\hat{\theta}]] = 0$$

- Dall'identità si ha che: $MSE = V[\hat{\theta}] + b^2$ (somma della varianza e del quadrato del bias). Si può interpretare MSE come somma in quadratura degli errori statistici e sistematici
- Un'altra caratteristica dello stimatore è la **sufficienza**. Uno stimatore è detto sufficiente se dal campione di dati estrae sul parametro tutta l'informazione possibile presente nelle misure. La media aritmetica è uno stimatore sufficiente
- Nelle situazioni pratiche spesso bisogna accontentarsi di compromessi a seconda della situazione particolare in esame

Stimatore della Media Aritmetica

□ Sia μ (o $E[x]$) il valore di aspettazione di un campione di n dati $x_1, x_2, \dots, x_n$.

□ Si ha: $E[x_i] = \int \int x_i f(x_1) f(x_2) \dots f(x_n) dx_1 dx_2 \dots dx_n = \mu$

□ e di conseguenza $E[\bar{x}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \sum_{i=1}^n \mu = \mu$ (stimatore non distorto)

□ La varianza sullo stimatore della media aritmetica è $V[\hat{\mu}] = \frac{\sigma^2}{n}$
Questa è la varianza sul valore di aspettazione (valore medio) della variabile casuale X mentre σ^2 è la varianza di X .

□ La radice quadrata della varianza sul valore medio è nota come **errore standard della media** σ_μ

$$\sigma_\mu = \frac{\sigma}{\sqrt{n}}$$

Stimatore della Varianza

- ❑ Supponiamo di voler stimare la varianza in un campione di n dati di una variabile casuale X . Possiamo avere due situazioni:

- 1) è noto il valore vero μ della variabile X (fatto raro)
- 2) non è noto il valore vero

- ❑ Nel primo caso stimiamo la varianza con lo stimatore $\hat{V}[x] = \frac{1}{n} \sum (x_i - \mu)^2$

- ❑ Questo stimatore è **consistente e non distorto**. Infatti:

$$E[\hat{V}[x]] = \frac{nE[(x - \mu)^2]}{n} = E[(x - \mu)^2] = V[x]$$

- ❑ Nel secondo caso non essendo noto μ usiamo la media aritmetica (che calcoliamo dai dati) :

$$\hat{V}[x] = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{n} \sum (x_i^2 - \bar{x}^2)$$

- ❑ Questo stimatore è distorto. Infatti se prendiamo il valore di aspettazione

Stimatore della Varianza

$$E[\hat{V}[x]] = n \frac{E[x^2 - \bar{x}^2]}{n} = E[x^2] - E[\bar{x}^2]$$

- essendo anche $E[x] = E[\bar{x}]$ si ottiene:

$$E[\hat{V}[x]] = E[x^2] - E[x]^2 - (E[\bar{x}^2] - E[\bar{x}]^2) = V[x] - V[\bar{x}] \quad \text{e quindi}$$

$$E[\hat{V}[x]] = \left(1 - \frac{1}{n}\right) V[x] = \frac{n-1}{n} V[x] \neq V[x]$$

- Questo stimatore è perciò **distorto**. Possiamo però correggere la distorsione, introducendo il seguente stimatore della varianza:

$$\hat{V}[x] = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

- Si verifica che:

$$E[\hat{V}[x]] = \frac{n}{n-1} E[x^2 - \bar{x}^2] = \frac{n}{n-1} (E[x^2] - E[\bar{x}^2]) = \frac{n}{n-1} V[x] \left(1 - \frac{1}{n}\right) = V[x]$$

- Questo stimatore della varianza è non distorto e consistente. È noto come **varianza del campione**

Stimatore della Covarianza

- Anche per gli elementi della matrice di covarianza si può introdurre uno stimatore non distorto:

$$\hat{V}_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{n}{n-1} (\overline{xy} - \bar{x}\bar{y})$$

- Da questo si ottiene uno stimatore per il coefficiente di correlazione :

$$r_{xy} = \frac{\hat{V}_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left(\sum_{j=1}^n (x_j - \bar{x})^2 \cdot \sum_{k=1}^n (y_k - \bar{y})^2 \right)^{1/2}} =$$
$$\frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}}$$

Limite Inferiore della Varianza

- ❑ Dati la dimensione di un campione ed il tipo di distribuzione, la varianza di uno stimatore **consistente e non distorto** non può essere inferiore ad un certo limite
- ❑ Come facciamo a scegliere lo stimatore più efficiente (cioè quello più vicino al limite inferiore della varianza?)
- ❑ Si abbia un campione di n misure $x_1, x_2, \dots, x_n$ i.i.d.. La p.d.f. congiunta calcolata nei punti misurati è

$$L(\theta) = L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$
- ❑ Questa quantità è detta funzione di **verosimiglianza** (**likelihood**) del campione
- ❑ Considerato un generico stimatore $\hat{\theta}(x_1, x_2, \dots, x_n)$ se la funzione di likelihood è sufficientemente regolare allora la varianza dello stimatore non può essere inferiore a

$$V[\hat{\theta}] \geq \frac{1}{E \left(\left[\frac{\partial \log L}{\partial \theta} \right]^2 \right)_{\theta}}$$
 con θ valore vero del parametro

Limite Inferiore della Varianza

- ❑ Sotto le stesse condizioni di regolarità della likelihood, si può dimostrare che:

$$E \left(\left[\frac{\partial \log L}{\partial \theta} \right]^2 \right)_\theta = -E \left[\frac{\partial^2 \log L}{\partial \theta^2} \right]_\theta$$

- ❑ E quindi si ha che $V[\hat{\theta}] \geq -\frac{1}{E \left[\frac{\partial^2 \log L}{\partial \theta^2} \right]_\theta}$
- ❑ Questo limite inferiore della varianza è detto **limite di Cramer-Rao**
- ❑ L'inverso di questo limite è detto **informazione di Fisher** del campione

$$I(\theta) = -E \left[\frac{\partial^2 \log L}{\partial \theta^2} \right]_\theta$$

- ❑ Noi siamo interessati in modo particolare agli stimatori che raggiungono il limite di Cramer-Rao (MVB). Come facciamo a sapere quali sono questi stimatori?

Limite Inferiore della Varianza

- Uno stimatore di θ raggiunge il limite di Cramer-Rao se e solo se la derivata parziale di $\log L$ rispetto a θ fattorizza nel modo seguente:

$$l' = \frac{\partial \log L}{\partial \theta} = A(\theta)(\hat{\theta} - \theta)$$

con $A(\theta)$ che non dipende dalle misure. $A(\theta)$ è l'informazione di Fisher e vale la relazione:

$$V[\hat{\theta}] = \frac{1}{A(\theta)}$$

- Stimatori MVB esistono solo per particolari classi di distribuzioni (alcune molto importanti nelle applicazioni pratiche)
- In genere uno stimatore ha varianza V maggiore di quelli MVB. In questi casi si definisce efficienza ϵ dello stimatore:

$$\epsilon = \frac{MVB}{V}$$

Limite Inferiore della Varianza

- Stima della vita media τ di una particella. Segue una distribuzione esponenziale:

$$f(t) = \frac{1}{\tau} e^{-t/\tau}$$

- In un campione di n misure si ha:

$$l' = \frac{\partial \log L}{\partial \tau} = -n/\tau + \sum t_i/\tau^2 = (n/\tau^2) \left(\sum t_i/n - \tau \right)$$

- Quindi $\bar{t}_n$ è uno stimatore che raggiunge il limite di Cramer-Rao; inoltre $A(\tau) = n/\tau^2$ e $V[\bar{t}_n] = \tau^2/n$

- Per una distribuzione di Cauchy si ha $l' = \frac{\partial \log L}{\partial \theta} = 2 \sum \frac{x_i - \theta}{1 + (x_i - \theta)^2}$ e quindi non esiste uno stimatore MVB

- Quando il MVB non esiste si può cercare lo stimatore non distorto che ha minima varianza (MV). Si dimostra che se esiste uno stimatore MV, questo è unico.