Stimatori. Propagazione degli errori

Laboratorio di Metodi Computazionali e Statistici (2022/2023)

16 Novembre 2022

Statistica

- Per caratterizzare la pdf di un set di misure sperimentali (campione) occorre costruire una funzione (delle nostre misure sperimentali) detta **stimatore**. Lo stimatore, come suggerisce il nome, fornisce una stima di uno dei parametri della distribuzione.
- Ad esempio lo stimatore per la media è la media aritmetica delle misure

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Se immagino di ripetere più volte l'esperimento il valore dello stimatore cambierà. Posso quindi considerare la distribuzione di probabilità dello stimatore e considerare quindi il comportamento asintotico (per grandi n) dello stimatore.

Proprietà degli stimatori

Chiamo $t_n(x)$ un stimatore generico (basato su n misure) di una quantità θ

- Consistenza
  - nel limite $n \to \infty$ lo stimatore deve convergere, in probabilità, al valore vero (θ):

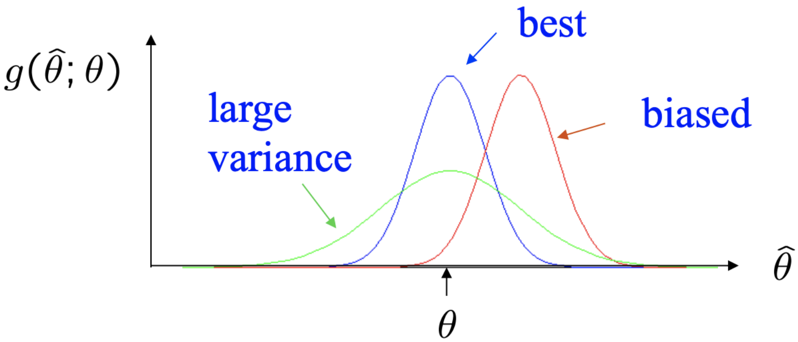
$$\lim_{n\to\infty} P(|t_n(x) - \theta| < \varepsilon) = 1 \ \ \forall\varepsilon$$

  - Proprietà irrinunciabile.
- Unbiased
  - Definizione bias: $b = E[t_n(x) - \theta]$
  - $t_n(x)$ è unbiased se il suo valore di aspettazione coincide con θ

$$E[t_n] = \theta$$

- Efficienza: un estimatore è tanto più efficiente quanto la sua varianza è piccola.

Proprietà degli stimatori



Esempi

- La media aritmetica ha tutte le proprietà elencate come stimatore della media
- Al contrario per la stima della varianza, lo stimatore

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

è consistente e unbiased solo se si conosce il valore medio vero μ. Se invece si usa il la media aritmetica ($\bar{x}$) occorre riscriverlo come

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

perché sia unbiased (verificare).

Stimatore per la covarianza

- Si verifica che per due variabili generiche x e y lo stimatore consistente ed unbiased per la covarianza è

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

- Il corrispondente stimatore per il coefficiente di correlazione è

$$R_{xy} = \frac{E_{xy}}{\sqrt{s_x^2 s_y^2}}$$

che, tuttavia non è unbiased (anche se di uso comune per la sua semplicità)

Propagazione degli errori

- Supponiamo di avere un set di variabili aleatorie x_1, x_2, ... , x_n
- Ogni coppia di variabili x_i, x_j è caratterizzata da $V_{ij} = cov[x_i, x_j]$
- Consideriamo una funzione $f(x_1, x_2, ..., x_n)$ e vogliamo calcolare la varianza di f
- Consideriamo ora il caso in cui f è una combinazione lineare di variabili

$$f(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} a_i x_i$$

$$\sigma_f^2 = Cov[f, f] =$$

Propagazione degli errori in casi particolari

Supponiamo di avere due variabili aleatorie: x_1 e x_2 con σ_1, σ_2e consideriamo una funzione $f(x1, x2)$
Propaghiamo l'errore su $f(x1, x2)$:

$$\sigma_f^2 = a_1\sigma_1^2 + a_2 * \sigma_2^2 + 2\rho a_1 a_2 \sigma_1 \sigma_2$$

dove ρ è il coefficiente di correlazione ($\rho = \frac{cov[x_1, x_2]}{\sigma_1\sigma_2}$). Prendiamo $x_1 = x_2 = 10$, $\sigma_1 = \sigma_2 = 1$ e consideriamo il caso $\rho = 0$ e $\rho = 1$. Consideriamo una funzione $f(x1, x2)$ ed in particolare consideriamo $f = x_1 - x_2$
Propaghiamo l'errore sulla differenza f:

$\rho = 0$

Propagazione degli errori in casi particolari

Supponiamo di avere due variabili aleatorie: x_1 e x_2 con σ_1, σ_2e consideriamo una funzione $f(x1, x2)$
Propaghiamo l'errore su $f(x1, x2)$:

$$\sigma_f^2 = a_1\sigma_1^2 + a_2 * \sigma_2^2 + 2\rho a_1 a_2 \sigma_1 \sigma_2$$

dove ρ è il coefficiente di correlazione ($\rho = \frac{cov[x_1, x_2]}{\sigma_1 \sigma_2}$). Prendiamo $x_1 = x_2 = 10$, $\sigma_1 = \sigma_2 = 1$ e consideriamo il caso $\rho = 0$ e $\rho = 1$. Consideriamo una funzione $f(x1, x2)$ ed in particolare consideriamo $f = x_1 - x_2$
Propaghiamo l'errore sulla differenza f:

$\rho = 0$

$$\sigma_f^2 = \sigma_1^2 + \sigma_2^2 = 2\sigma^2$$

Propagazione degli errori in casi particolari

Supponiamo di avere due variabili aleatorie: x_1 e x_2 con σ_1, σ_2e consideriamo una funzione $f(x1, x2)$
Propaghiamo l'errore su $f(x1, x2)$:

$$\sigma_f^2 = a_1\sigma_1^2 + a_2 * \sigma_2^2 + 2\rho a_1 a_2 \sigma_1 \sigma_2$$

dove ρ è il coefficiente di correlazione ($\rho = \frac{cov[x_1, x_2]}{\sigma_1 \sigma_2}$). Prendiamo $x_1 = x_2 = 10$, $\sigma_1 = \sigma_2 = 1$ e consideriamo il caso $\rho = 0$ e $\rho = 1$. Consideriamo una funzione $f(x1, x2)$ ed in particolare consideriamo $f = x_1 - x_2$
Propaghiamo l'errore sulla differenza f:
$\rho = 1$

Propagazione degli errori in casi particolari

Supponiamo di avere due variabili aleatorie: x_1 e x_2 con σ_1, σ_2e consideriamo una funzione $f(x1, x2)$
Propaghiamo l'errore su $f(x1, x2)$:

$$\sigma_f^2 = a_1\sigma_1^2 + a_2 * \sigma_2^2 + 2\rho a_1 a_2 \sigma_1\sigma_2$$

dove ρ è il coefficiente di correlazione ($\rho = \frac{cov[x_1, x_2]}{\sigma_1\sigma_2}$). Prendiamo $x_1 = x_2 = 10$, $\sigma_1 = \sigma_2 = 1$ e consideriamo il caso $\rho = 0$ e $\rho = 1$. Consideriamo una funzione $f(x1, x2)$ ed in particolare consideriamo $f = x_1 - x_2$
Propaghiamo l'errore sulla differenza f:
$\rho = 1$

$$\sigma_f^2 = \sigma_1^2 + \sigma_2^2 + 2(-1)\sigma_1\sigma_2 = 2\sigma^2 - 2\sigma^2 = 0$$

Il risultato riflette il fatto che se due grandezze sono completamente correlare ($\rho = 1$) queste di muovono, rispetto al valore centrale, coerentemente (uno spostamento di σ_1 di x_1 corrisponde allo spostamento di σ_2 di x_2) e la differenza rimane invariata.

Analogo risultato con $f = x_1 + x_2$ e $\rho = -1$.

Propagazione degli errori

- Si consideri ora invece una funzione generica $f(\boldsymbol{x})$ di n variabili.
- Vogliamo trovare la varianza di f: $V[f]$
- Assumiamo che la funzione sia approssimativamente lineare nelle variabili e quindi prendiamo solo i primi due termini dell'espansione in serie di Taylor

$$f(\boldsymbol{x}) = f(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial f}{\partial x_i} \right] \Bigg|_{\boldsymbol{x}=\boldsymbol{\mu}} (x_i - \mu_i) + \mathcal{O}((x_i - \mu_i)^2$$

- Potremo quindi trovare l'errore su f come

$$V[f] = E[f^2] - (E[f])^2$$

Propagazione degli errori

$$f(\boldsymbol{x}) = f(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial f}{\partial x_i} \right] \Bigg|_{\boldsymbol{x}=\boldsymbol{\mu}} (x_i - \mu_i) + \mathcal{O}(x_i - \mu_i)^2$$

- Devo calcolare $E[f]$:

Propagazione degli errori

- Calcoliamo $E[f^2]$

Propagazione degli errori

- Quindi otteniamo per $V[f]$:

$$\sigma_f^2 = f^2(\boldsymbol{\mu}) + \sum_{i,j=1}^{n} \left[\frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \right]_{\boldsymbol{x}=\boldsymbol{\mu}} V_{ij} - f^2(\boldsymbol{\mu})$$

$$\sigma_f^2 = \sum_{i,j=1}^{n} \left[\frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \right]_{\boldsymbol{x}=\boldsymbol{\mu}} V_{ij} = \sum_{i=1}^{n} \left[\frac{\partial f}{\partial x_i} \right]_{\boldsymbol{x}=\boldsymbol{\mu}}^2 \sigma_i^2 + 2 \sum_{i,j>i}^{n} \left[\frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \right]_{\boldsymbol{x}=\boldsymbol{\mu}} V_{ij}$$

Se le misure sono scorrelate $V_{ij} = \sigma_i^2 \delta_{ij}$

$$\sigma_f^2 = \sum_{i=1}^{n} \left[\frac{\partial f}{\partial x_i} \right]_{\boldsymbol{x}=\boldsymbol{\mu}}^2 \sigma_i^2$$

Propagazione degli errori

Altro modo di ragionare

$$f(\boldsymbol{x}) = f(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial f}{\partial x_i} \right] \bigg|_{\boldsymbol{x}=\boldsymbol{\mu}} (x_i - \mu_i) + \mathcal{O}((x_i - \mu_i)^2$$

è equivalente a

$$f(\boldsymbol{x}) = c + \sum_{i=1}^{n} (a_i x_i + b_i)$$

con $a_i = |\partial f / \partial x_i|_{\boldsymbol{x}=\boldsymbol{\mu}}$. I termini costanti (senza errore) non contribuiscono alla covarianza (momento centrato) e quindi si può ri-applicare la formula ottenuta per

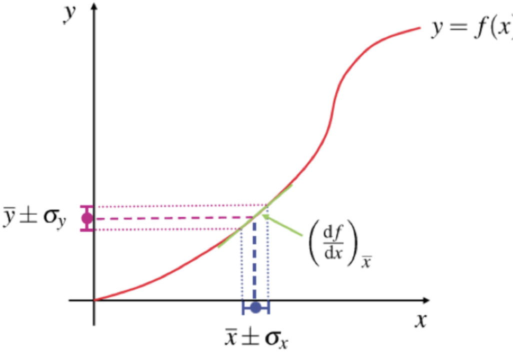
$$f(\boldsymbol{x}) = \sum_{i=1}^{n} a_i x_i$$

$$V[f] = \sigma_f^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j V[x_i, x_j] = \sum_{i,j=1}^{n} \left[\frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \right]_{\boldsymbol{x}=\boldsymbol{\mu}} V_{ij}$$

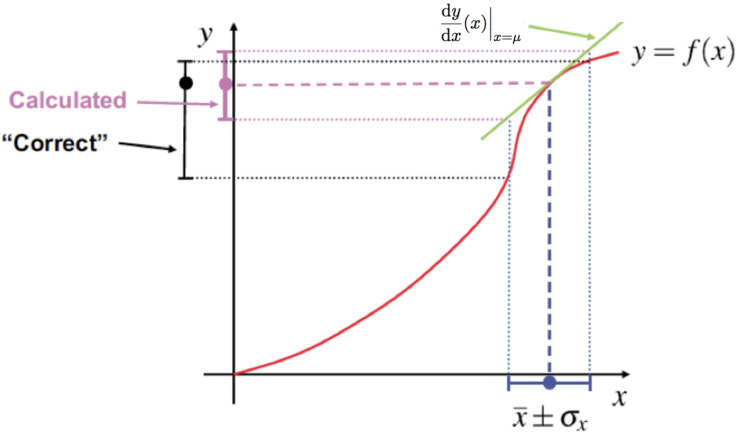
Limitazioni

- Propagazione dell'errore basata su un'approssimazione lineare
- Supponiamo di avere $f(x)$ (curva rossa) come in figura
    - L'intervallo blu corrisponde a $\bar{x} \pm \sigma_x$
    - Approssimo la $f(x)$ con una funzione lineare nell'intervallo (curva verde)
    - Se approssimazione lineare è valida in un intervallo $\pm 1\sigma$:

$$\sigma_y = \frac{\partial f}{\partial x} \sigma_x$$



Limitazioni



Guida alla propagazione degli errori (I)

- funzione lineare nei parametri affetti da errore $\rightarrow$ propagazione con derivate
- funzione lineare nei parametri affetti da errore (che sono inoltre indipendenti). Il calcolo di $\frac{\partial f}{\partial x_i}\sigma_i$ può essere sostituito dalla relazione (sviluppo in serie di Taylor al prim'ordine in $x_i + \sigma_i$)

$$f(x_1, x_2, ..., x_i + \sigma_i, ..x_n) = f(x_1, x_2, ..., x_i, ..x_n) + \frac{\partial f}{\partial x_i}\sigma_i$$

$$\frac{\partial f}{\partial x_i}\sigma_i = f(x_1, x_2, ..., x_i + \sigma_i, ..x_n) - f(x_1, x_2, ..., x_i, ..x_n)$$

  - Vario ogni singola variabile x_i della sua σ_i e calcolo la differenza $f(x_i + \sigma_i) - f(x_i)$: errore dovuto a x_i
  - Se $f(x_1, x_2, ..., x_i, ..x_n)$ vario in maniera indipendente ogni singola x_i

La relazione presenta vantaggi: non richiede di conoscere f in forma analitica e permette di verificare se la relazione di linearità è verificata. Se infatti

$$|f(..., x_i + \sigma_i, ...) - f(..., x_i, ...)| \neq |f(..., x_i - \sigma_i, ...) - f(..., x_i, ...)|$$

sicuramente la funzione non è lineare in $[x_i - \sigma_i, x_i + \sigma_i]$.

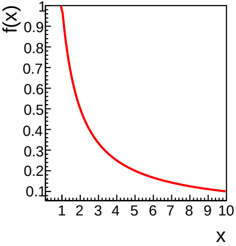
Guida alla propagazione degli errori

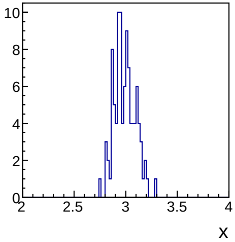
- Se la funzione non è lineare (o approssimabile linearmente) nei parametri affetti da errori l'unico metodo possibile è la "simulazione" con metodi Monte Carlo della distribuzione di f:
  - si generano N esperimenti, in ciascun di essi si estraggono valori per le variabili casuali x_i (seguendo le rispettive distribuzioni di probabilità)
  - per ciascuno esperimento e, quindi, per ciascun set di valori x_i si calcola f
  - si crea un istogramma di f e, da esso, si calcolano i momenti rilevanti (tra questi il valor medio e la varianza).

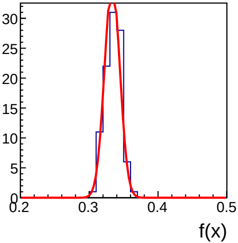
Esempi di propagazione degli errori

$$f(x) = \frac{1}{x} \qquad\qquad x_m = (3.0 \pm 0.1)$$

- f è lineare in $x_0 \pm \sigma$? $|f(x_0+\sigma) - f(x_0)| \simeq |f(x_0-\sigma) - f(x_0)| \simeq 0.011$: sì .
- Posso usare tutte e tre i metodi
- si noti che uno (variazione della funzione), di fatto, l'ho già calcolato: $\sigma(f(x)) = 0.011$
- Posso usare il metodo delle derivate: $\sigma(f(x)) = \dfrac{\partial f}{\partial x}\sigma(x)$: $\sigma(f) = \sigma(x)/x^2 = 0.011$
- Posso usare il metodo MC



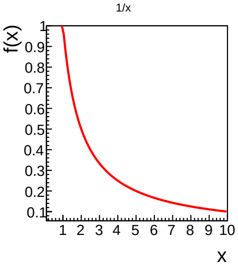


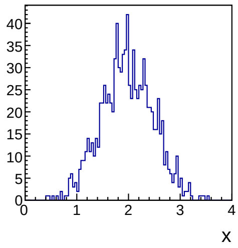


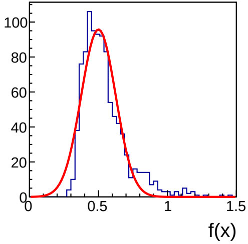
Esempi di propagazione degli errori

$$f(x) = \frac{1}{x} \qquad\qquad x_m = (2.0 \pm 0.5)$$

- f è lineare in $x_0 \pm \sigma$? $|f(x_0+\sigma) - f(x_0)| \simeq 0.10$ $|f(x_0-\sigma) - f(x_0)| \simeq 0.17$: no .
- Posso usare solo Montecarlo (disegno però il confronto con la prop. con le derivate)







La propagazione con Montecarlo trova

$$0.54 \pm 0.17$$

mentre quella con le derivate avrebbe detto 0.50 ± 0.13.

Lezione in guscio di noce

- Gli stimatori sono funzioni delle misure sperimentali il cui scopo è valutare i parametri della distribuzione di probabilità.
- Proprietà fondamentali degli stimatori sono: consistenza, assenza di "bias", efficienza
- Il problema della propagazione degli errori da una serie di misure con errore ad una loro funzione può essere risolta sempre esattamente se la funzione delle misure è lineare
- Se la funzione non è lineare si può:
    - linearizzare la funzione e
        - propagare l'errore con le derivate
        - propagare l'errore calcolando $|f(x+\sigma) - f(x)|$ e/o $|f(x-\sigma) - f(x)|$ (il calcolo di queste quantità permette anche di verificare la linearità)
    - costruire la distribuzione di $f(x)$ con il metodo Montecarlo (e da questa calcolare i momenti di interesse).

Appendice: stimatori per distribuzioni 1D: media

Media aritmetica $\bar{x}$ (stimatore di μ)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \mu = \int x p(x) dx$$

- È unbiased

$$E[\bar{x}] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n} n\mu = \mu$$

- Valutiamone la varianza per valutare l'incertezza dello stimatore e per vedere se è consistente

$$\begin{aligned}\sigma_{\bar{x}}^2 &= E[(\bar{x}-\mu)^2] \\ &= E\left[\left(\frac{x_1+x_2+\ldots+x_n}{n}-\mu\right)^2\right] = \frac{1}{n^2}E\left[(x_1+x_2+\ldots+x_n-n\mu)^2\right] \\ &= \frac{1}{n^2}E\left[((x_1-\mu)+(x_2-\mu)+\ldots+(x_n-\mu))^2\right]\end{aligned}$$

siccome tutti gli x_i sono indipendenti tutti i termini incrociati del tipo $E[(x_i-\mu)(x_j-\mu)]$ con $i \neq j$ sono nulli e quindi si ottiene

$$\sigma_{\bar{x}}^2 = \frac{1}{n^2}E\left[(x_1-\mu)^2+(x_2-\mu)^2+\ldots+(x_n-\mu)^2\right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Appendice: stimatori per distribuzioni 1D: media

La relazione

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

assicura che l'estimatore è consistente perché la sua varianza tende a 0 per $n \to \infty$.
La deviazione standard di $\sigma_{\bar{x}}$ (spesso chiamato errore standard, denominato ε) è dato da:

$$\sigma_{\bar{x}} = \varepsilon = \frac{\sigma}{\sqrt{n}}$$

Dato un campione di misure la miglior stima della misura è quindi data $(\bar{x} \pm \varepsilon)$ o $(\bar{x} \pm \sigma_{\bar{x}})$ (dove a volte si omette la barra). Resta da stimare σ (che non è in genera nota) lo vedremo nelle prossime pagine.

N.B. Si ricordi che sono espressioni equivalenti: $\sigma_x^2 = V[x] = E[x^2] - E[x]^2 = E[(x - E[x])^2]$

Appendice: stimatori per distribuzione 1D: varianza

Varianza s (stimatore di σ^2)

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \qquad \sigma^2 = \int (x-\mu)^2 p(x) dx$$

- È consistente (non lo dimostriamo ma ragionevole in quanto è la discretizzazione della definizione di momento di ordine 2)
- È unbiased

$$E[s^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}E(x_i - \mu)^2 = \frac{1}{n}n\sigma^2 = \sigma^2$$

Appendice: stimatori per distribuzione 1D: varianza

Siccome nella maggior parte dei casi μ non è noto ma stimato a partire da $\bar{x}$ è necessario modificare lo stimatore s^2 affinché resti unbiased.
Varianza empirica s_n (stimatore di σ^2) quando non si conosce μ

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- È unbiased

$$\begin{aligned}
E[s_n^2] &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(x_i^2 - 2\bar{x}x_i + \bar{x}^2)\right] \\
&= \frac{1}{n-1}E\left[\sum_{i=1}^{n}x_i^2 - 2\bar{x}\sum_{i=1}^{n}x_i + n\bar{x}^2\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right] \\
&= \frac{1}{n-1}E\left[\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right] = \frac{1}{n-1}\sum_{i=1}^{n}(E(x_i^2) - E(n\bar{x}^2)) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(E(x_i^2) - nE(\bar{x}^2)) = \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right) \\
&= \frac{1}{n-1}\left((n-1)\sigma^2\right) = \sigma^2
\end{aligned}$$

Appendice: riassunto stimatori per distribuzione 1D

Media:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\sigma_{\bar{x}} = \frac{s_n}{\sqrt{n}}$$

Varianza:

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Si può dimostrare che

$$V[s_n^2] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right)$$ $\mu_{4/2}$ momenti centrati di ordine 4 e 2

(che mostra, tra l'altro, che s_n è consistente).

Appendice: riassunto stimatori per distribuzione 1D

Si dimostra che per la stima di un momento di ordine k generico si può usare lo stimatore unbiased

$$m_k = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^k$$

Appendice: stimatori per distribuzioni 2D

Per media e varianza non cambia nulla.

Stimatore per covarianza (unbiased e consistente)

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

Stimatore per coefficiente di correlazione

$$R_{xy} = \frac{E_{xy}}{\sqrt{s_x^2 s_y^2}}$$

Si dimostra che per una gaussiana 2D si ha:

$$E[R_{xy}] \sim \rho - \rho \frac{(1-\rho^2)}{2n}$$
$$V[R_{xy}] \sim \frac{1}{n}(1-\rho^2)^2$$

quindi lo stimatore R_{xy} del coefficiente di correlazione non è unbiased ma è consistente (tende a ρ con varianza nulla per $n \to \infty$)