CORRETTEZZA ED EFFICIENZA

Correttezza di $\overline{X}$:

$$E(\overline{X}) = E\left[\frac{1}{n}(X_1 + \dots X_n)\right]$$

Le analisi di questo corso si baseranno su tre proprietà degli stimatori: correttezza, efficienza e consistenza. Le prime due, trattate in questa sequenza, si riferiscono all'analisi per campioni finiti: analisi dove il campione ha un numero finito di osservazioni.

CORRETTEZZA ED EFFICIENZA

Correttezza di $\overline{X}$:

$$E(\overline{X}) = E\left[\frac{1}{n}(X_1 + \dots X_n)\right]$$

La consistenza, una proprietà che si riferisce alle analisi in cui la dimensione campionaria tende all'infinito, sarà trattata successivamente.

CORRETTEZZA ED EFFICIENZA

Correttezza di $\bar{X}$:

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + \ldots X_n)\right]$$

Supponiamo che si desidera stimare la media della popolazione μ_X di una variabile casuale *X* dato un campione. Dimostreremo che la media campionaria è uno stimatore corretto, ma non il solo.

Correttezza di $\overline{X}$:

$$E(\overline{X}) = E\left[\frac{1}{n}(X_1 + \ldots X_n)\right] = \frac{1}{n}E(X_1 + \ldots + X_n)$$

$$= \frac{1}{n}\left[E(X_1) + \ldots + E(X_n)\right] = \frac{1}{n}n\mu_X = \mu_X$$

Ricordando i risultati della sequenza precedente, sappiamo che ogni componente del campione $\{X_1, \ldots, X_n\}$ ha media uguale a μ_X, e quindi il valore atteso della media campionaria è μ_X.

Correttezza di $\bar{X}$:

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + \ldots X_n)\right] = \frac{1}{n}E(X_1 + \ldots + X_n)$$

$$= \frac{1}{n}\left[E(X_1) + \ldots + E(X_n)\right] = \frac{1}{n}n\mu_X = \mu_X$$

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

Comunque, la media campionaria non è l'unico stimatore corretto della media della popolazione. Dimostreremo ciò supponendo di avere a disposizione soltanto due osservazioni (per semplicità).

Correttezza di $\bar{X}$:

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + \ldots X_n)\right] = \frac{1}{n}E(X_1 + \ldots + X_n)$$

$$= \frac{1}{n}\left[E(X_1) + \ldots + E(X_n)\right] = \frac{1}{n}n\mu_X = \mu_X$$

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

Definiamo uno stimatore generalizzato *Z* che è la somma pesata delle due osservazioni con pesi λ_1 e λ_2.

CORRETTEZZA ED EFFICIENZA

Correttezza di $\bar{X}$:

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + \dots X_n)\right] = \frac{1}{n}E(X_1 + \dots + X_n)$$

$$= \frac{1}{n}\left[E(X_1) + \dots + E(X_n)\right] = \frac{1}{n}n\mu_X = \mu_X$$

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$E(Z) = E(\lambda_1 X_1 + \lambda_2 X_2)$$

Dimostreremo che *Z* è uno stimatore corretto.

CORRETTEZZA ED EFFICIENZA

Correttezza di $\bar{X}$:

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + \dots X_n)\right] = \frac{1}{n}E(X_1 + \dots + X_n)$$

$$= \frac{1}{n}\left[E(X_1) + \dots + E(X_n)\right] = \frac{1}{n}n\mu_X = \mu_X$$

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$E(Z) = E(\lambda_1 X_1 + \lambda_2 X_2) = E(\lambda_1 X_1) + E(\lambda_2 X_2)$$

Iniziamo con l'usare la prima regola del valore atteso.

Correttezza di $\bar{X}$:

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + ... X_n)\right] = \frac{1}{n}E(X_1 + ... + X_n)$$

$$= \frac{1}{n}\left[E(X_1) + ... + E(X_n)\right] = \frac{1}{n} n\mu_X = \mu_X$$

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$E(Z) = E(\lambda_1 X_1 + \lambda_2 X_2) = E(\lambda_1 X_1) + E(\lambda_2 X_2)$$

$$= \lambda_1 E(X_1) + \lambda_2 E(X_2)$$

Adesso usiamo la seconda regola portando fuori dall'espressione del valore atteso λ_1 e λ_2 .

Correttezza di $\bar{X}$:

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + \dots X_n)\right] = \frac{1}{n}E(X_1 + \dots + X_n)$$

$$= \frac{1}{n}\left[E(X_1) + \dots + E(X_n)\right] = \frac{1}{n} n\mu_X = \mu_X$$

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$E(Z) = E(\lambda_1 X_1 + \lambda_2 X_2) = E(\lambda_1 X_1) + E(\lambda_2 X_2)$$

$$= \lambda_1 E(X_1) + \lambda_2 E(X_2) = (\lambda_1 + \lambda_2)\mu_X$$

Il valore atteso di *X* per ogni osservazione è μ_X.

CORRETTEZZA ED EFFICIENZA

Correttezza di $\bar{X}$:

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + \dots X_n)\right] = \frac{1}{n}E(X_1 + \dots + X_n)$$

$$= \frac{1}{n}\left[E(X_1) + \dots + E(X_n)\right] = \frac{1}{n}n\mu_X = \mu_X$$

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

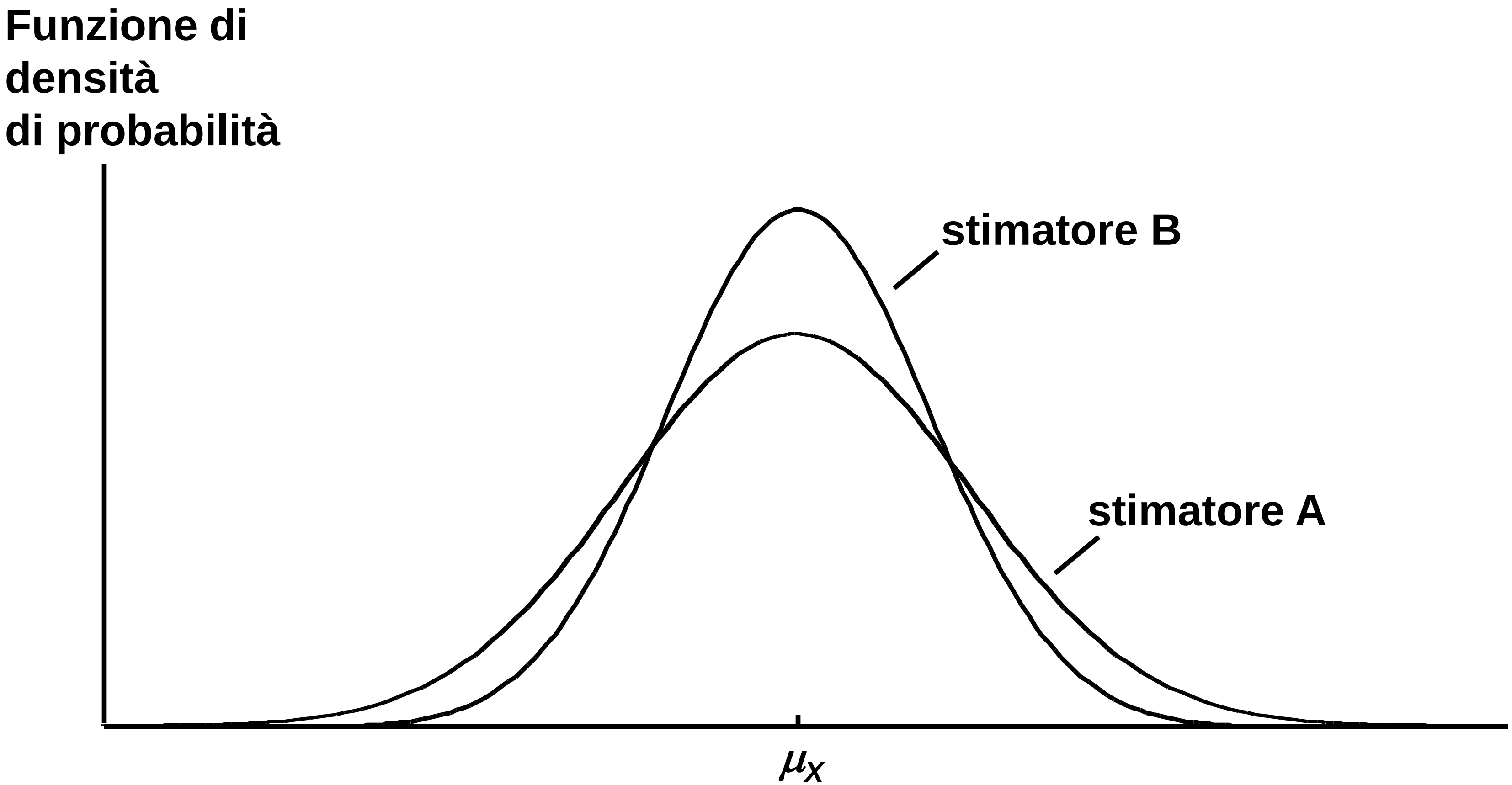
$$E(Z) = E(\lambda_1 X_1 + \lambda_2 X_2) = E(\lambda_1 X_1) + E(\lambda_2 X_2)$$

$$= \lambda_1 E(X_1) + \lambda_2 E(X_2) = (\lambda_1 + \lambda_2)\mu_X$$

$$= \mu_X \quad \text{if} \quad (\lambda_1 + \lambda_2) = 1$$

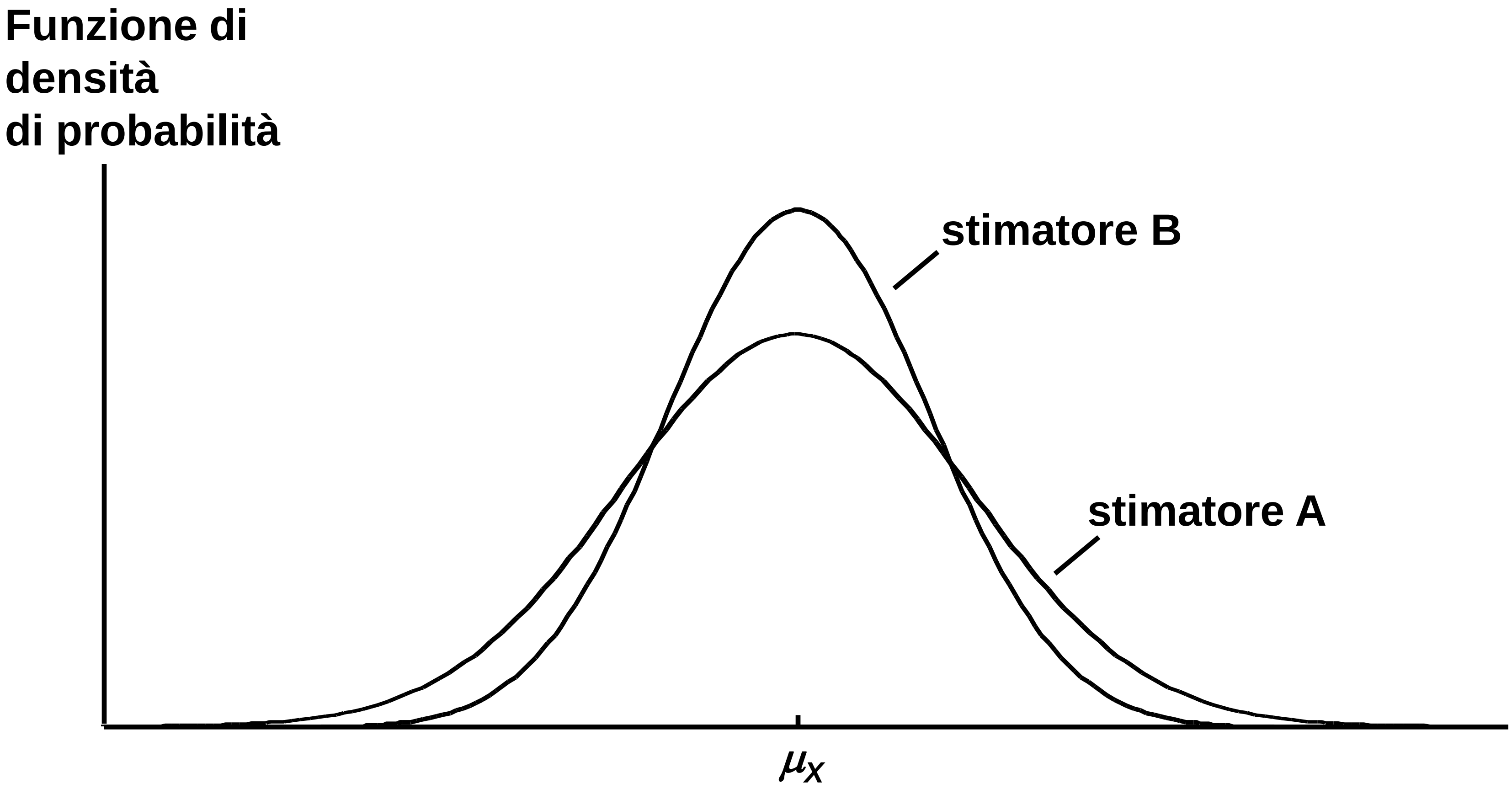
Quindi *Z* è uno stimatore corretto (o non distorto) di μ_X se la somma dei pesi è uguale a 1. Possiamo scegliere un numero infinito di combinazioni di λ_1 e λ_2 che soddisfano questa condizione (e non solo la media campionaria).

CORRETTEZZA ED EFFICIENZA



Come facciamo a scegliere tra un'infinità di stimatori? La risposta è: usare lo stimatore più efficiente, ovvero quello che presenti la più bassa varianza, in quanto tende ad essere più accurato.

CORRETTEZZA ED EFFICIENZA



Nel diagramma, A e B sono entrambi stimatori non distorti ma B è preferibile in quanto è più efficiente.

CORRETTEZZA ED EFFICIENZA

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$\sigma_Z^2 = \mathrm{var}(\lambda_1 X_1 + \lambda_2 X_2)$$

Analizzeremo adesso la varianza dello stimatore generalizzato e troveremo una condizione tale che la varianza sia minima.

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$\sigma_Z^2 = \mathrm{var}(\lambda_1 X_1 + \lambda_2 X_2)$$

$$= \mathrm{var}(\lambda_1 X_1) + \mathrm{var}(\lambda_2 X_2) + 2\,\mathrm{cov}(\lambda_1 X_1, \lambda_2 X_2)$$

La prima regola della varianza viene usata per decomporre la varianza.

CORRETTEZZA ED EFFICIENZA

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$\sigma_Z^2 = \text{var}(\lambda_1 X_1 + \lambda_2 X_2)$$

$$= \text{var}(\lambda_1 X_1) + \text{var}(\lambda_2 X_2) + 2\,\text{cov}(\lambda_1 X_1, \lambda_2 X_2)$$

$$= \lambda_1^2 \sigma_{X_1}^2 + \lambda_2^2 \sigma_{X_2}^2$$

Assumiamo che X_1 e X_2 siano indipendenti e quindi la loro covarianza è zero. La seconda regola della varianza viene applicata per portare fuori delle espressioni delle varianze λ_1 e λ_2.

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$\begin{aligned}\sigma_Z^2 &= \mathrm{var}(\lambda_1 X_1 + \lambda_2 X_2) \\ &= \mathrm{var}(\lambda_1 X_1) + \mathrm{var}(\lambda_2 X_2) + 2\,\mathrm{cov}(\lambda_1 X_1, \lambda_2 X_2) \\ &= \lambda_1^2 \sigma_{X_1}^2 + \lambda_2^2 \sigma_{X_2}^2 \\ &= (\lambda_1^2 + \lambda_2^2)\sigma_X^2\end{aligned}$$

La varianza di X_1 è σ_X^2, così come per X_2.

CORRETTEZZA ED EFFICIENZA

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$\begin{aligned}
\sigma_Z^2 &= \mathrm{var}(\lambda_1 X_1 + \lambda_2 X_2) \\
&= \mathrm{var}(\lambda_1 X_1) + \mathrm{var}(\lambda_2 X_2) + 2\,\mathrm{cov}(\lambda_1 X_1, \lambda_2 X_2) \\
&= \lambda_1^2 \sigma_{X_1}^2 + \lambda_2^2 \sigma_{X_2}^2 \\
&= (\lambda_1^2 + \lambda_2^2)\sigma_X^2 \\
&= (\lambda_1^2 + [1-\lambda_1]^2)\sigma_X^2 \qquad \text{if } \ (\lambda_1 + \lambda_2) = 1
\end{aligned}$$

Considerando la condizione imposta per la correttezza, possiamo riscrivere la varianza di *Z*, sostituendo λ_2 con 1- λ_1.

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$\begin{aligned}
\sigma_Z^2 &= \mathrm{var}(\lambda_1 X_1 + \lambda_2 X_2) \\
&= \mathrm{var}(\lambda_1 X_1) + \mathrm{var}(\lambda_2 X_2) + 2\,\mathrm{cov}(\lambda_1 X_1, \lambda_2 X_2) \\
&= \lambda_1^2 \sigma_{X_1}^2 + \lambda_2^2 \sigma_{X_2}^2 \\
&= (\lambda_1^2 + \lambda_2^2)\sigma_X^2 \\
&= (\lambda_1^2 + [1 - \lambda_1]^2)\sigma_X^2 \qquad \text{if} \;\; (\lambda_1 + \lambda_2) = 1 \\
&= (2\lambda_1^2 - 2\lambda_1 + 1)\sigma_X^2
\end{aligned}$$

Sviluppiamo il quadrato. Per minimizzare la varianza di *Z*, occorre scegliere quel valore di λ_1 tale da minimizzare l'espressione finale.

Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$\begin{aligned}
\sigma_Z^2 &= \operatorname{var}(\lambda_1 X_1 + \lambda_2 X_2) \\
&= \operatorname{var}(\lambda_1 X_1) + \operatorname{var}(\lambda_2 X_2) + 2\operatorname{cov}(\lambda_1 X_1, \lambda_2 X_2) \\
&= \lambda_1^2 \sigma_{X_1}^2 + \lambda_2^2 \sigma_{X_2}^2 \\
&= (\lambda_1^2 + \lambda_2^2)\sigma_X^2 \\
&= (\lambda_1^2 + [1-\lambda_1]^2)\sigma_X^2 \qquad \text{if } (\lambda_1 + \lambda_2) = 1 \\
&= (2\lambda_1^2 - 2\lambda_1 + 1)\sigma_X^2
\end{aligned}$$

$$\frac{d\sigma_Z^2}{d\lambda_1} = 0 \;\Rightarrow\; 4\lambda_1 - 2 = 0$$

Deriviamo rispetto λ_1.

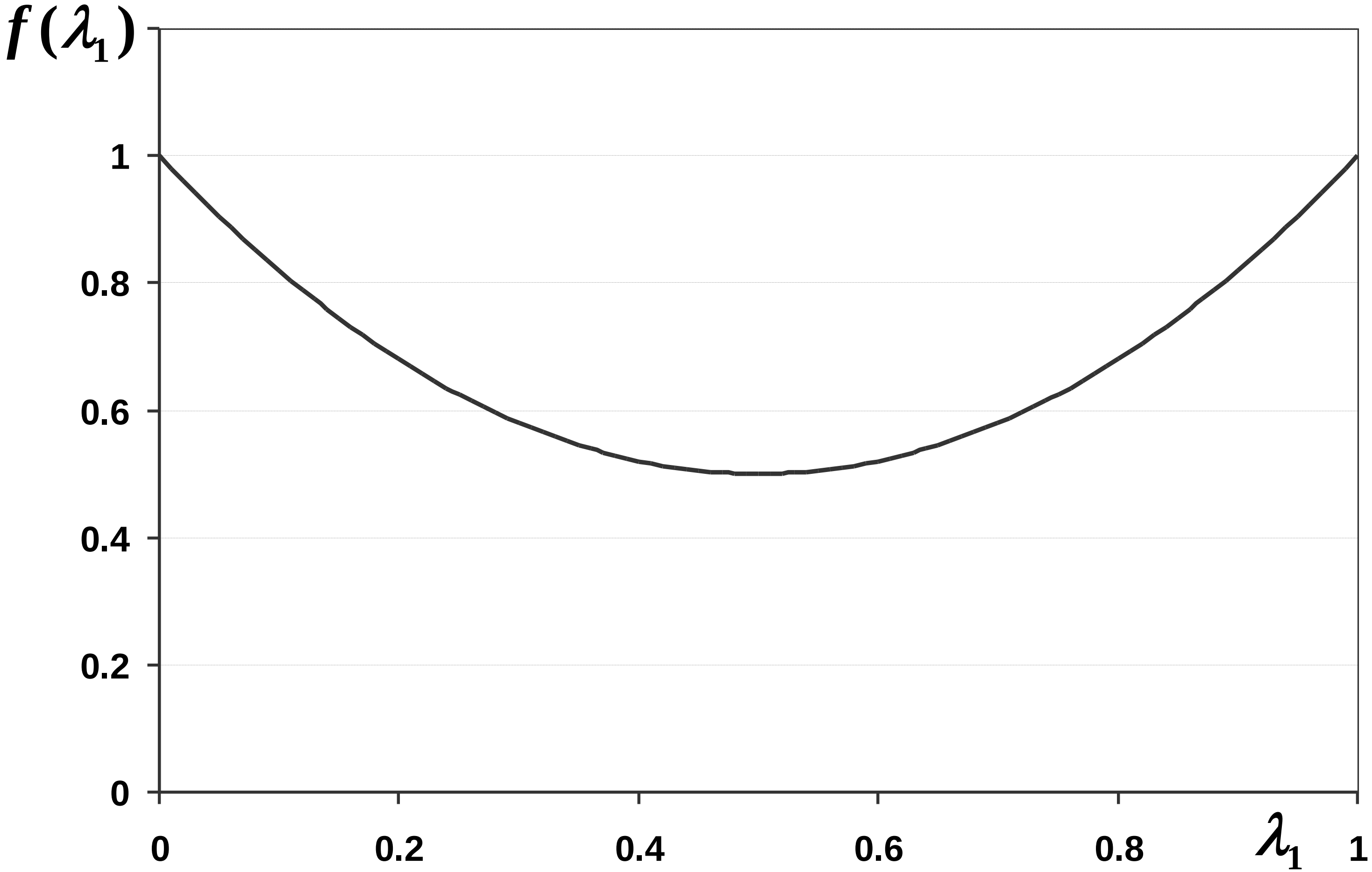
Stimatore generalizzato $Z = \lambda_1 X_1 + \lambda_2 X_2$

$$\begin{aligned}
\sigma_Z^2 &= \mathrm{var}(\lambda_1 X_1 + \lambda_2 X_2) \\
&= \mathrm{var}(\lambda_1 X_1) + \mathrm{var}(\lambda_2 X_2) + 2\,\mathrm{cov}(\lambda_1 X_1, \lambda_2 X_2) \\
&= \lambda_1^2 \sigma_{X_1}^2 + \lambda_2^2 \sigma_{X_2}^2 \\
&= (\lambda_1^2 + \lambda_2^2)\sigma_X^2 \\
&= (\lambda_1^2 + [1-\lambda_1]^2)\sigma_X^2 \qquad \text{if } (\lambda_1 + \lambda_2) = 1 \\
&= (2\lambda_1^2 - 2\lambda_1 + 1)\sigma_X^2
\end{aligned}$$

$$\frac{d\sigma_Z^2}{d\lambda_1} = 0 \;\Rightarrow\; 4\lambda_1 - 2 = 0 \;\Rightarrow\; \lambda_1 = \lambda_2 = 0.5$$

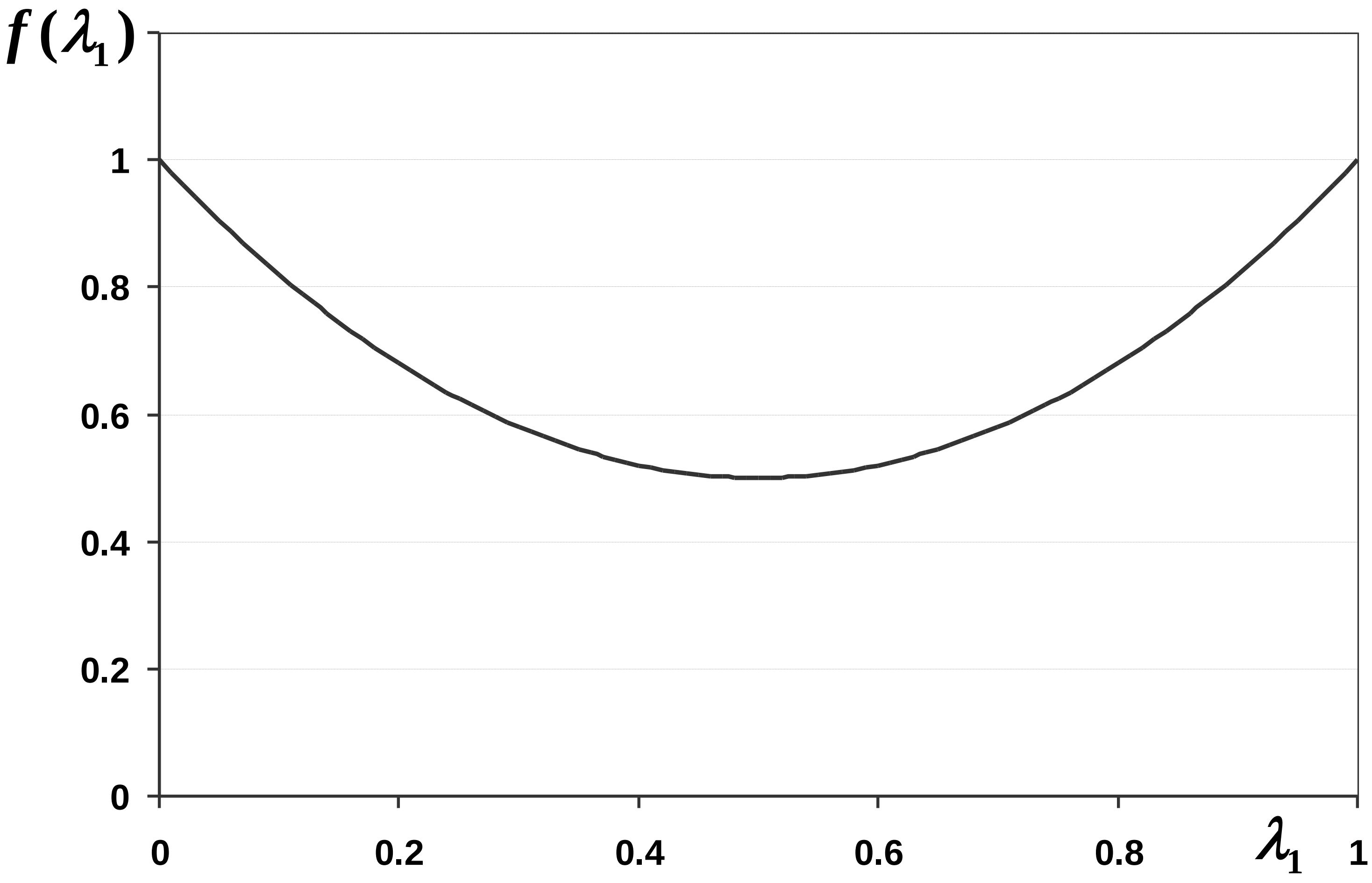
L'espressione viene minimizzata per λ_1 = 0.5. Quindi λ_2 = 0.5. Quindi abbiamo dimostrato che la media campionaria è lo stimatore corretto più efficiente, per lo meno nel nostro esempio (Nota che occorre calcolare la derivata seconda, ed essa è positiva, confermando che stiamo minimizzando.)

CORRETTEZZA ED EFFICIENZA



Alternativamente, possiamo trovare il minimo graficamente. Abbiamo rappresentato l'espressione in funzione di λ_1.

CORRETTEZZA ED EFFICIENZA



Possiamo vedere che la varianza viene minimizzata per λ_1 = 0.5 e quindi la media campionaria è lo stimatore corretto più efficiente.