Stima puntuale

Cosa vuol dire stimare?

Stimare: attribuire un valore plausibile a un parametro incognito

Quando un parametro della popolazione è stimato attraverso un singolo valore tale valore viene chiamato **stima puntuale** del parametro

Stima puntuale - Caso di studio

Un’azienda deve tenere sotto controllo la qualità del processo di produzione di un macchinario che produce pezzi di una certa lunghezza. Se si tenessero sotto controllo tutti i pezzi prodotti e se ne calcolasse la lunghezza si potrebbe ottenere senza difficoltà la lunghezza media (parametro).

Supponiamo che non si possano misurare le lunghezze di tutti i pezzi prodotti ma solo quelle relative ad un campione casuale di pezzi di numerosità n. Allora si può calcolare solo la lunghezza media del campione di pezzi (statistica). *In tal caso la lunghezza media della popolazione di pezzi è ignota.*

Caso di studio

La lunghezza è la v.c. che indichiamo con X dove la popolazione corrisponde all'insieme di pezzi che quel macchinario è in grado di produrre

Si può assumere ad esempio che $X \sim N(\mu;\sigma^2)$ dove media e varianza siano entrambe ignote

Quindi la forma della distribuzione della popolazione è nota a meno dei due parametri ignoti che devono essere stimati (stima puntuale)

Generalizziamo

X è una v.c. che rappresenta il carattere osservato sulla popolazione d'interesse la cui distribuzione di probabilità è di forma nota a meno di qualche parametro θ che la caratterizza

Se X è discreta la sua f. di probabilità è indicata con $p(x; \theta)$

Se X è continua la sua funzione di densità è indicata con $f(x; \theta)$

Stimatore

Dato che il parametro θ non è noto, si vogliono trarre delle conclusioni su di esso sulla base di un campione estratto dalla popolazione, cioè si vuole *fare inferenza* su θ

Si definisce **stimatore** del parametro θ ogni statistica (funzione dei dati campionari)

$$T = t(X_1, \ldots X_n)$$

utilizzata per stimare θ

Uno stimatore è quindi una v.c.

Stima

La **stima** del parametro di interesse θ

$$t(x_1, \ldots x_n)$$

è il valore assunto dallo stimatore
su un particolare campione osservato

$$(x_1, \ldots x_n)$$

Quindi una stima è una determinazione della v.c. stimatore

Scelta dello stimatore

Se esistono diversi stimatori per θ, come si fa a scegliere tra di essi?

Esistono dei criteri (**proprietà degli stimatori**) in base ai quali è possibile scegliere tra i diversi stimatori quello più appropriato alla situazione in esame

Errore quadratico medio

L'errore quadratico medio di uno stimatore T di un parametro θ è dato da:

$$MSE(T) = E[(T - \theta)^2]$$

Quindi MSE(T) è una misura sintetica della prossimità di uno stimatore al parametro ignoto θ ovvero misura l'errore che si commette in media utilizzando T per stimare θ

Stimatore più efficiente

Dati due stimatori T_1 e T_2 del parametro θ,
T_1 è più **efficiente** di T_2
se e solo se $MSE(T_1) < MSE(T_2)$ per ogni θ
(uniformemente)

In generale non esiste uno stimatore con MSE *uniformemente minimo* quindi usualmente ci si restringe alla classe degli stimatori non distorti (o stimatori corretti)

Stimatore corretto

Uno stimatore T del parametro θ si dice **corretto** (o non distorto) se per ogni valore di θ vale che

$$E(T) = \theta$$

La distorsione di uno stimatore T (o BIAS) è uguale a

$$B(T) = E(T) - \theta$$

Se $B(T)>0$ allora in media T sovrastima θ

Se $B(T)<0$ allora in media T sottostima θ

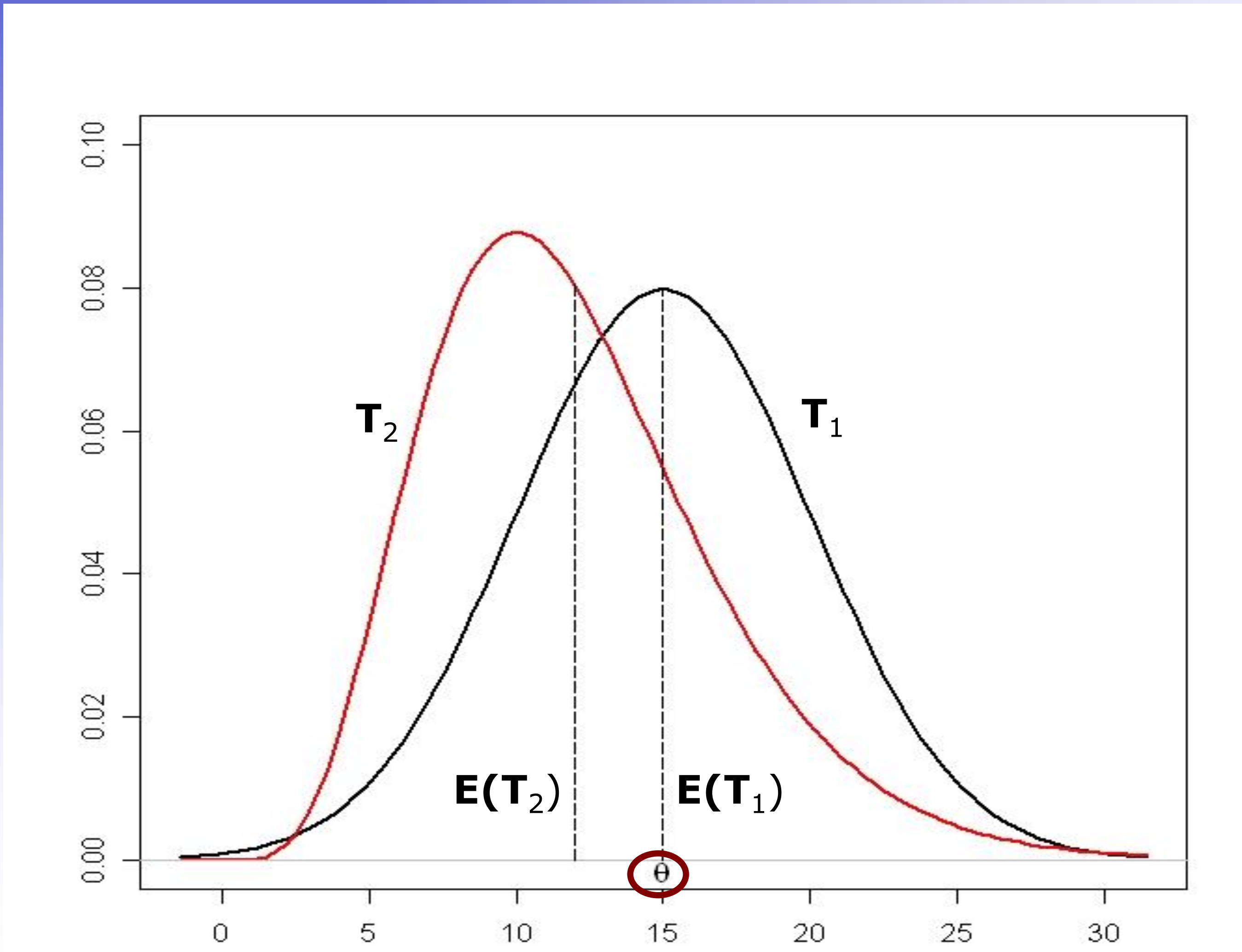
Scomposizione dell’errore quadratico medio

$$MSE(T) = Var(T) + [B(T)]^2$$

Se lo stimatore T è corretto, allora $B(T)=0$ quindi $MSE(T) = Var(T)$

Dati due stimatori corretti, T_1 e T_2 del parametro θ, T_1 è più efficiente di T_2 se $Var(T_1) < Var(T_2)$ per ogni θ

Distorsione di uno stimatore



T_1 e T_2 stimatori di θ

$E(T_1) = \theta$

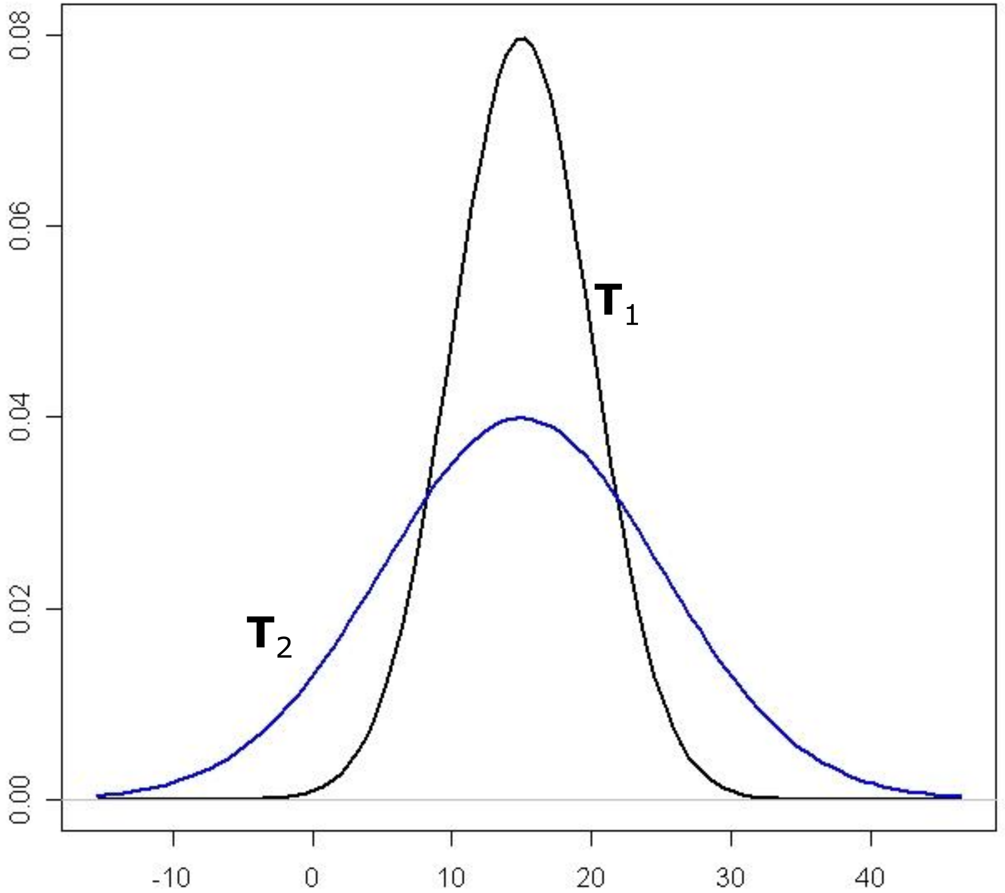
$E(T_2) < \theta$

T_1 è uno stimatore corretto di θ

T_2 è uno stimatore distorto di θ

$B(T_2)=E(T_2) - \theta$ è la distorsione di T_2

Efficienza

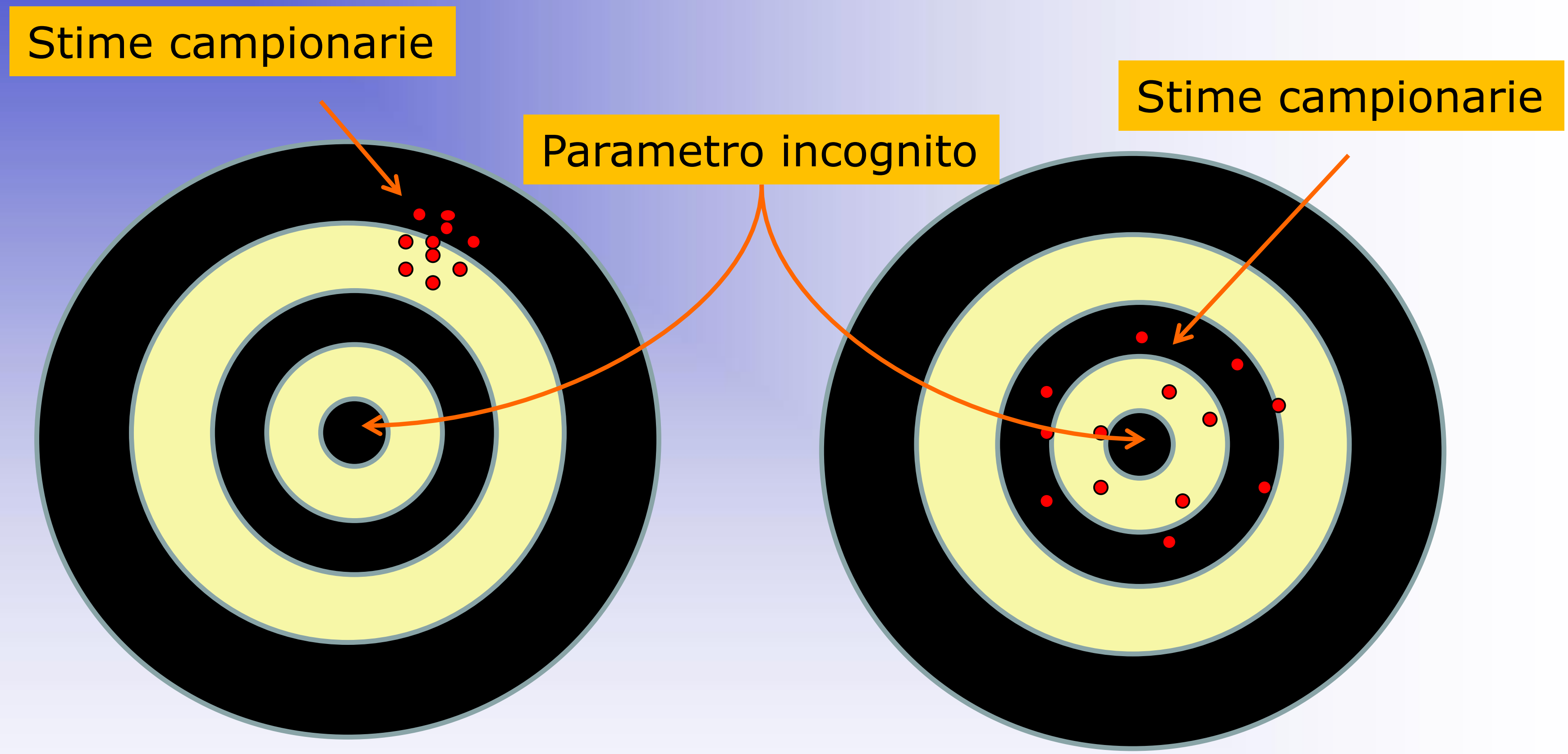


T_1 e T_2 stimatori corretti di θ

$$Var(T_1) < Var(T_2)$$

T_1 è più efficiente di T_2

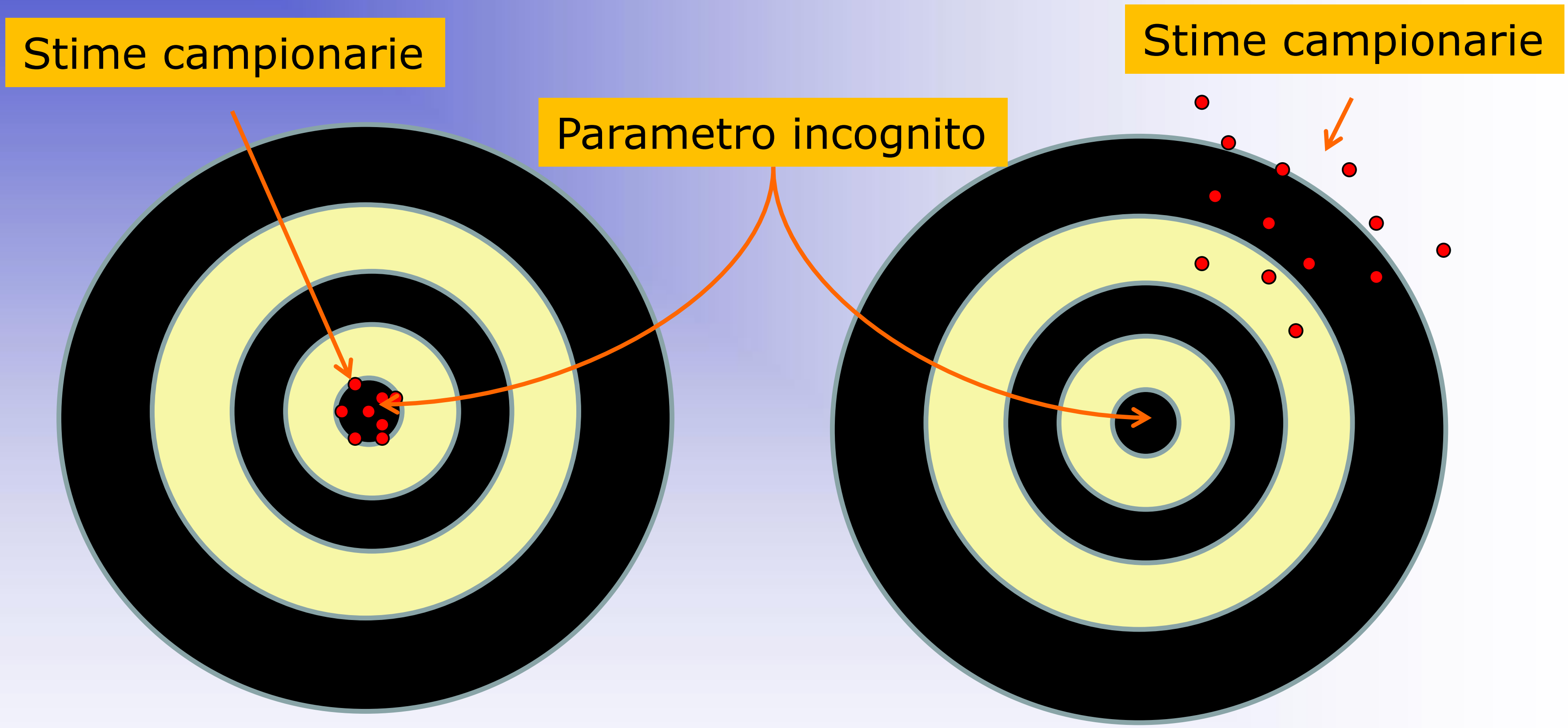
Stima come tiro al bersaglio



Alta distorsione, bassa variabilità. Bassa distorsione, alta variabilità.

Distorsione: lanci sistematicamente fuori mira.
Variabilità: dispersione dei lanci sul bersaglio.

Stima come tiro al bersaglio



Bassa distorsione, bassa variabilità

Alta distorsione, alta variabilità

Proprietà asintotiche degli stimatori

Gli stimatori hanno anche delle proprietà asintotiche ovvero proprietà che valgono solo per campioni grandi,

ovvero per $n \to \infty$

Si usa la notazione T_n

per indicare la dipendenza dello stimatore dalla numerosità campionaria

Consistenza in media quadratica

Uno stimatore T_n è consistente in media quadratica se:

$$\lim_{n \to \infty} MSE(T_n) = 0$$

La consistenza in media quadratica sia ha se e solo se la varianza e la distorsione dello stimatore tendono a zero al crescere di n

Correttezza asintotica

Uno stimatore T_n è asintoticamente corretto se

$$\lim_{n \to \infty} E(T_n) = \theta$$

per ogni θ

I parametri della popolazione che più frequentemente interessa stimare sono:

1) La media della popolazione μ

2) La proporzione π della popolazione di unità statistiche che presentano una certa modalità di interesse

3) La varianza della popolazione σ^2

Stima puntuale della media di una popolazione

La media campionaria $\overline{X}$

è uno stimatore corretto e consistente della media μ della popolazione: Infatti:

$$E\left(\overline{X}\right) = \mu$$

$$\lim_{n\to\infty} MSE(\overline{X}) = \lim_{n\to\infty} Var\left(\overline{X}\right) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

Quindi la precisione dello stimatore aumenta al crescere di n

Stima puntuale della proporzione di una popolazione

Sia X una popolazione distribuita come una Bernoulli con parametro π (proporzione di successi = media della popolazione)

La proporzione campionaria P (media campionaria) è uno stimatore corretto e consistente di π perché:

$$E(P) = E(\overline{X}) = \pi$$

$$\lim_{n\to\infty} MSE(P) = \lim_{n\to\infty} MSE(\overline{X}) = \lim_{n\to\infty} Var(\overline{X}) =$$

$$= \lim_{n\to\infty} \frac{\pi(1-\pi)}{n} = 0$$

Stima puntuale della varianza di una popolazione

Lo stimatore corretto e consistente della varianza di una popolazione X è dato da:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

(si divide per n-1 invece che per n)

S^2 varianza campionaria corretta $E\left(S^2\right) = \sigma^2$

È uno stimatore consistente

$$\lim_{n\to\infty} MSE\left(S_n^2\right) = 0$$

Ancora sulla varianza

- Varianza della popolazione σ^2

parametro che misura la variabilità del carattere nella popolazione

- Varianza campionaria S^2

stimatore corretto della varianza della popolazione

- Varianza della media campionaria $\dfrac{\sigma^2}{n}$
- Varianza della proporz. campionaria $\dfrac{\pi \cdot (1-\pi)}{n}$

misurano la variabilità delle distribuzioni campionarie

Esempio 1

Dato un campione di 5 misurazioni della lunghezza in cm di scatole regalo:

10 12 13 16 15

trovare gli stimatori corretti e consistenti della media e della varianza della popolazione.

Lo stimatore corretto e consistente del parametro μ è la media campionaria $\overline{X}$ la cui stima sulla base del campione osservato è uguale a 13,2

Lo stimatore corretto e consistente del parametro σ^2 è la varianza campionaria S^2 la cui stima sulla base del campione osservato è uguale a 5,7

Esempio 2

Su un campione di 5 individui si rileva se essi sono intenzionati ad acquistare un certo prodotto. Le risposte sono elencate di seguito:

SI NO NO SI NO

trovare una stima corretta e consistente della media (proporzione di chi è intenzionato all'acquisto) della popolazione

La stima corretta e consistente della media della popolazione è la media campionaria (proporzione di successi) nel campione che è uguale a 2/5

Esperimenti Monte Carlo

Per verificare le proprietà degli stimatori (e quindi per verificare la validità dei metodi inferenziali) possiamo usare la metodologia della simulazione attraverso esperimenti di tipo Monte Carlo.

Per gli esperimenti Monte Carlo si utilizzano dati artificiali creati attraverso un processo di generazione specificato dall'utente.

Ad esempio, si generano molteplici campioni di numeri casuali che simulano l'estrazione da una popolazione con una data distribuzione

Confronti di efficienza tra 3 stimatori

Vogliamo confrontare l'efficienza di 3 stimatori per la stima della media della popolazione

-Media campionaria

-Mediana campionaria

-Valore centrale (media tra il valore minimo e il valore massimo)

A tale scopo, possiamo generare, ad esempio, 10 campioni di numerosità 50 da una popolazione Normale con media 5 e varianza 4

Confronti di efficienza tra 3 stimatori

Consideriamo la distribuzione di ciascuno stimatore (al variare dei campioni)

	Stimatore:		
	media	**mediana**	**valore centrale**
Valore atteso	5.033	5.181	4.785
Deviazione standard	0.287	0.399	0.525
Errore quadratico medio	0.084	0.192	0.322

La media campionaria è lo stimatore più efficiente (il suo errore quadratico medio è il più piccolo rispetto agli altri stimatori)

Proprietà asintotiche degli stimatori

Con 10 campioni di numerosità 100 (anziché 50)

	Stimatore:		
	media	**mediana**	**valore centrale**
Valore atteso	4.997	4.992	5.051
Deviazione standard	0.142	0.305	0.383
Errore quadratico medio	0.020	0.093	0.150

Al crescere di n, tutti i 3 stimatori sono: asintoticamente corretti e consistenti

Proprietà asintotiche degli stimatori

Con 200 campioni di numerosità 250

	Stimatore:		
	media	**mediana**	**valore centrale**
Valore atteso	5.000	4.999	5.007
Deviazione standard	0.129	0.154	0.595
Errore quadratico medio	0.017	0.024	0.352

Al crescere di n e del numero dei campioni, diminuisce la distorsione e aumenta la precisione