Stima dei parametri

a cura del prof. Guida

- Precedentemente sono state analizzate le relazioni tra le statistiche campionarie e i parametri della popolazione:

$$\mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \;*$$

per la distribuzione campionaria delle <u>medie</u> nel caso di <u>estrazione bernoulliana</u>

**** (errore medio di campionamento)***

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \quad *$$

per la distribuzione campionaria delle <u>medie</u> nel caso di <u>estrazione in blocco</u>

**** (errore medio di campionamento)***

diap. 29 diap. 34 diap. 37

per la distribuzione campionaria della <u>varianza</u> nel caso di estrazione bernoulliana

$$\mu_{S^2} = M(S^2) = \frac{n-1}{n} \cdot \sigma^2$$

per la distribuzione campionaria della <u>varianza</u> nel caso di estrazione in blocco

$$\mu_{S^2} = M(S^2) = \frac{N}{N-1} \cdot \frac{n-1}{n} \cdot \sigma^2$$

per la distribuzione campionaria della frequenza dei successi

(nel caso di una popolazione con distribuzione binomiale)

$$\mu_{\left(\frac{k}{n}\right)} = M(F) = p$$

$$\sigma_{\left(\frac{k}{n}\right)} = \sqrt{\frac{pq}{n}}$$

diap. 45 diap. 77

I risultati ottenuti sulle statistiche campionarie saranno utilizzati per risolvere problemi di

Inferenza statistica

Rientrano in questa categoria tutti i problemi che possono ricondursi a:

- STIMA dei parametri
- Verifica di IPOTESI STATISTICHE

Le STIME dei parametri possono essere:

- STIME PUNTUALI
- STIME INTERVALLARI

Sono stime puntuali quelle definite da un unico numero, la statistica campionaria

Sono stime intervallari le stime espresse con due numeri che indicano un intervallo entro il quale cade, con una probabilità ben definita, un parametro della popolazione

L’inferenza statistica utilizza i risultati ottenuti dalle indagini campionarie per ottenere STIME relative ai parametri della popolazione dalla quale è stato estratto il campione

INDICHIAMO CON:

- θ (teta) o ϑ il <u>parametro della popolazione</u> che deve essere stimato
- T (tau) o $\hat{\vartheta}$ il <u>generico stimatore</u> (variabile aleatoria che dipende dai dati del campione e possiede una distribuzione di probabilità ben determinata)

AD ESEMPIO:

La media campionaria $\overline{X}$ è uno STIMATORE del parametro μ (media della popolazione).

Se si estrae un campione e si determina il valore $\overline{x}$, tale valore, che varia da campione a campione, è una STIMA di μ

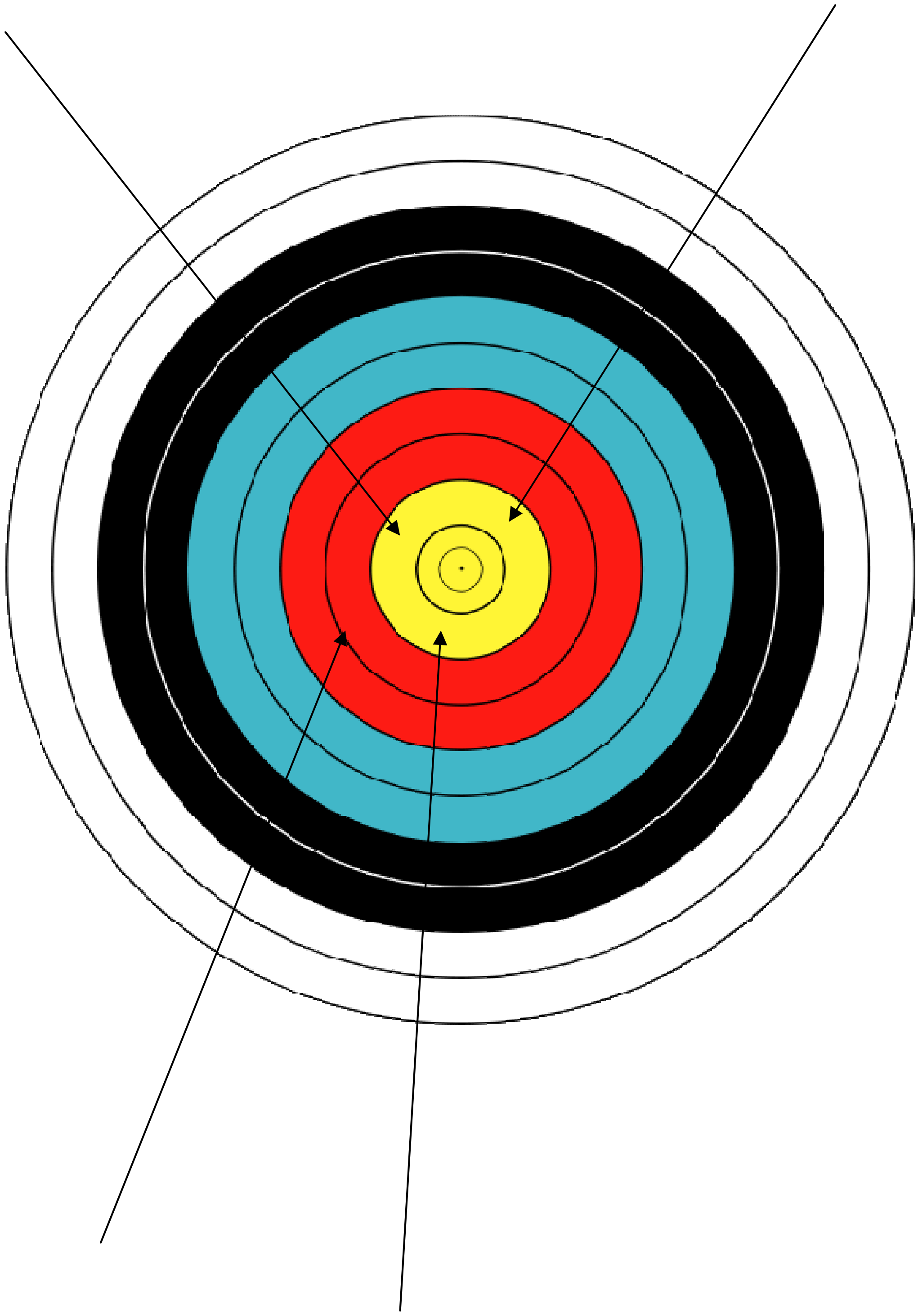
La frequenza campionaria relativa è uno STIMATORE del parametro p (probabilità di successo) di una popolazione con distribuzione bernoulliana.

Se si estrae un campione e si determina la frequenza relativa campionaria dei successi, tale valore, che varia da campione a campione, è una STIMA.

E' intuitivo ammettere che una stima è tanto migliore quanto più si avvicina al valore vero (incognito) del parametro da stimare.

Poiché T (tau) **è una variabile aleatoria, questo equivale a chiedere che la distribuzione di probabilità dello stimatore sia concentrata attorno al valore del parametro θ** (teta) **da stimare, ovvero che la dispersione dei valori di T attorno al suo valor medio sia la minima possibile.**

… un concorrente in una gara di tiro a bersaglio, desidera che i suoi colpi si distribuiscano *attorno* al centro del bersaglio!!!!



… analogamente, chi effettua una stima statistica desidera avere la massima probabilità di avvicinarsi al vero valore da stimare.

Questo può essere espresso in termini matematici in base alle proprietà degli stimatori; tali proprietà sono:

PROPRIETA' DEGLI STIMATORI

- Correttezza

- Efficienza

- Consistenza

Correttezza

- Uno stimatore T del parametro θ si dice corretto se il suo valor medio M(T) coincide con θ per qualsiasi valore di T

$$M(T) = \theta$$

Ad esempio $\overline{X}$ è uno stimatore corretto della media μ della popolazione: infatti

$$M(\overline{X}) = \mu$$

…..anche la frequenza relativa dei successi in un campione estratto da una popolazione di tipo binomiale è uno STIMATORE corretto del parametro p della popolazione

INFATTI

$$M\left(\frac{k}{n}\right)=\sum_{k=0}^{n}\left(\frac{k}{n}\right)\cdot P(k)=\frac{1}{n}\sum_{k=0}^{n}\ k\cdot P(k)$$

Essendo

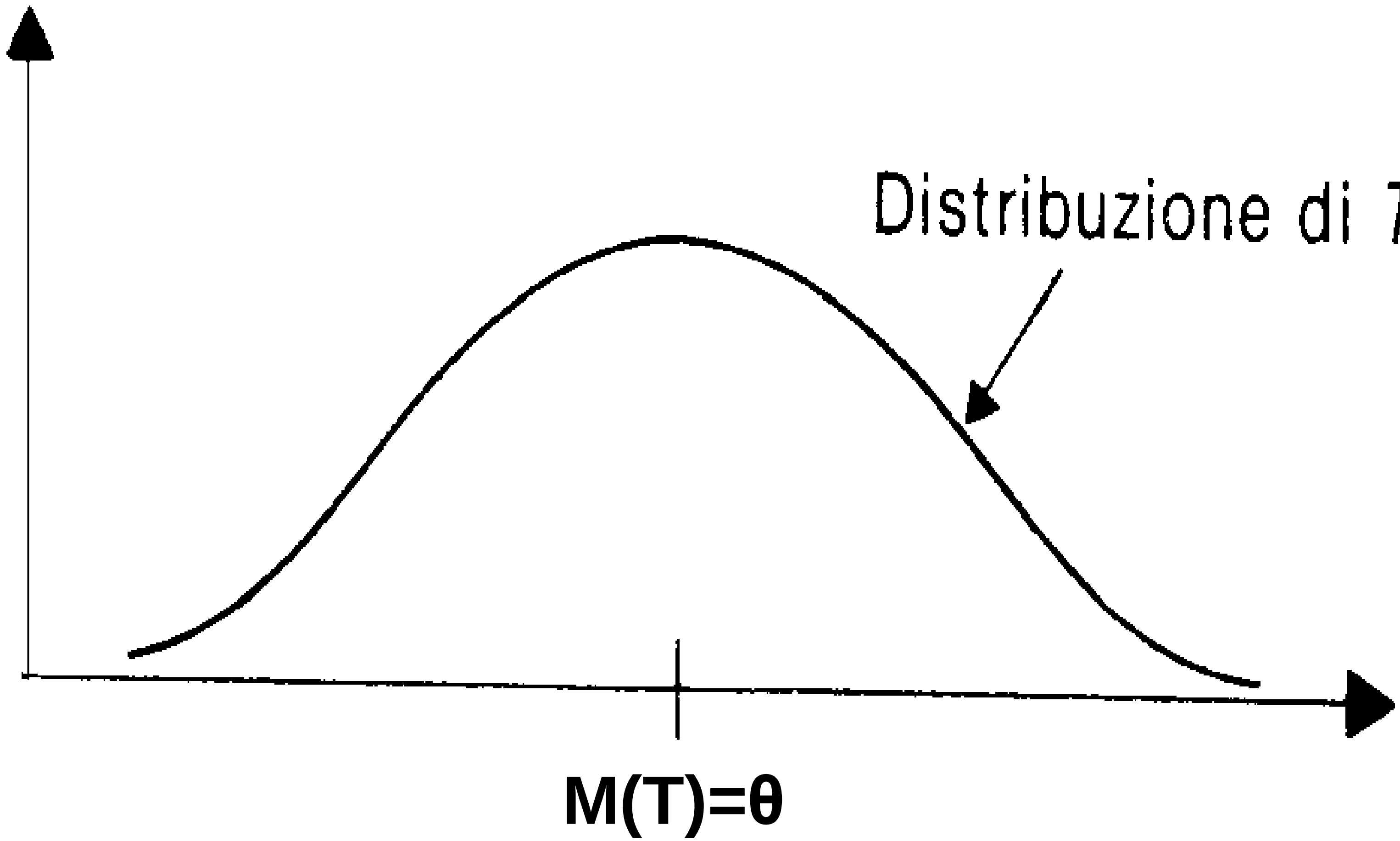
$$\sum_{k=0}^{n}k\cdot P(k)=np$$

si ha:

$$M\left(\frac{k}{n}\right)=\frac{1}{n}np=\ p$$

Si può concludere che la frequenza campionaria è uno STIMATORE corretto del parametro p della popolazione con distribuzione binomiale

Stimatore corretto M(T)=θ

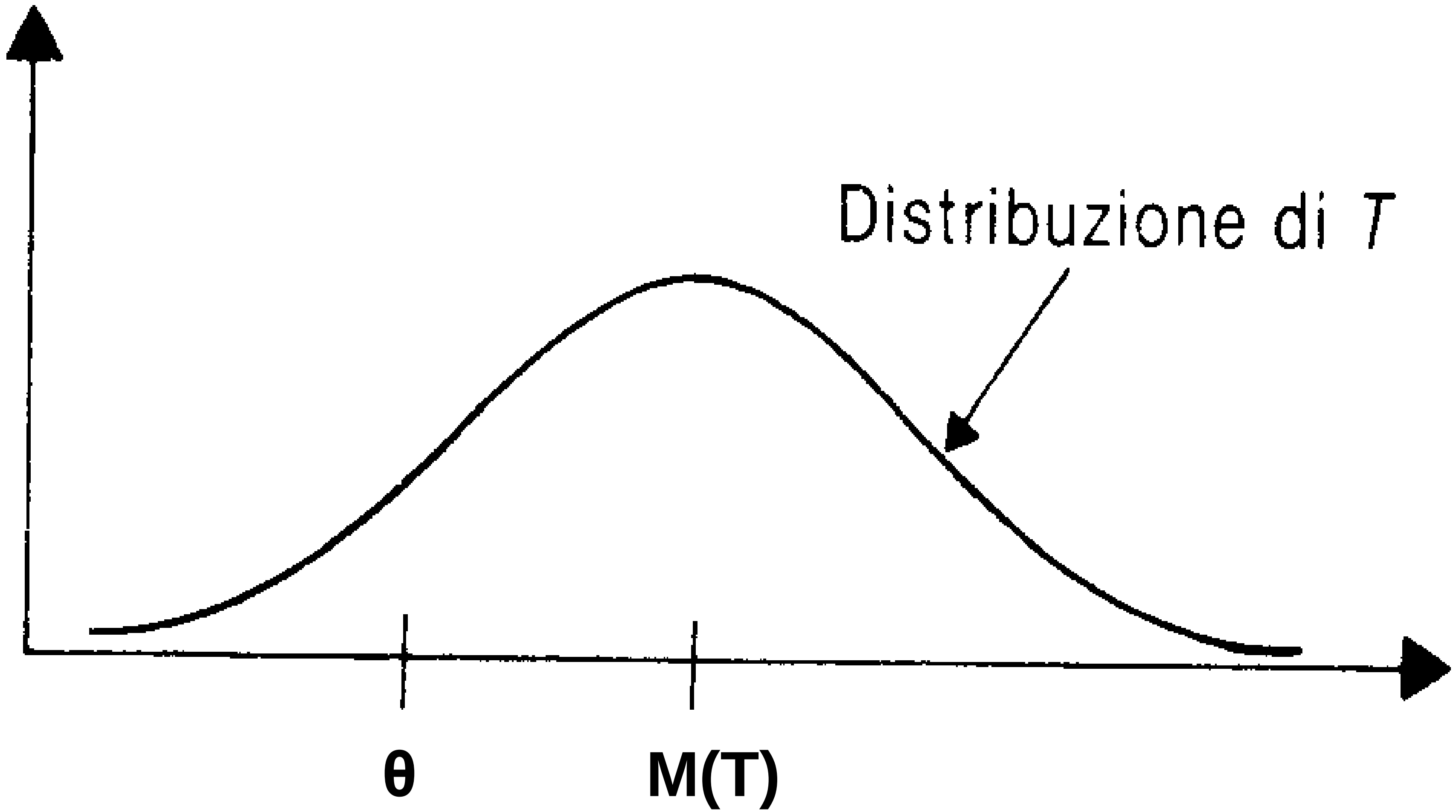


La varianza campionaria S^2 risulta, invece, uno stimatore **<u>distorto</u>** della varianza σ^2 della popolazione

infatti: $$\mu\left(S^2\right) \neq \sigma^2$$

risultando: $$\mu\left(S^2\right) = \frac{n-1}{n}\sigma^2$$

Stimatore distorto M(T) ≠θ



Efficienza

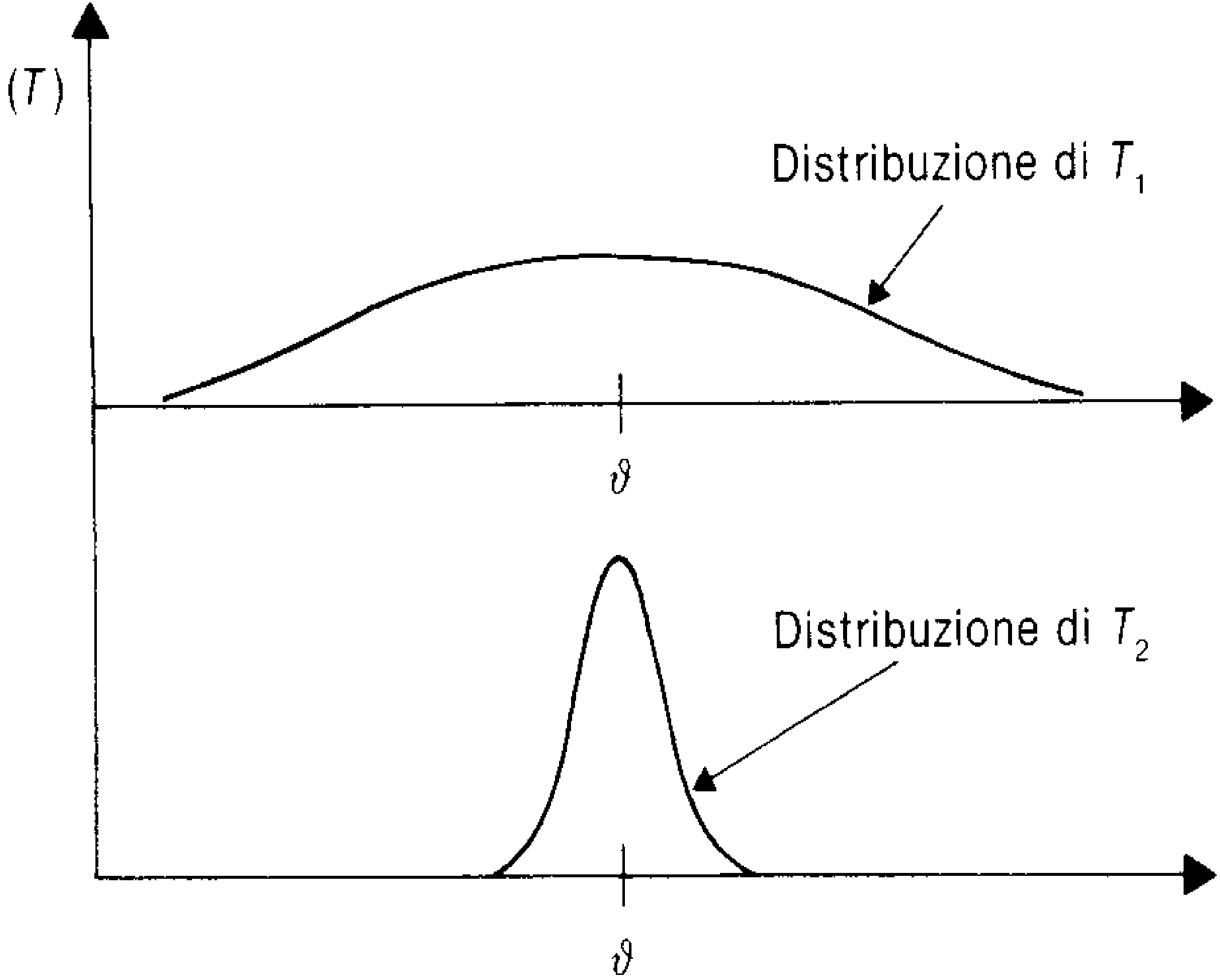
- La bontà di uno stimatore non può essere valutata solo in base alla correttezza.

Si consideri un parametro θ che può essere valutato mediante due stimatori corretti, T_1 e T_2;

Tra i due risulta migliore quello con varianza minore

PERTANTO

- Uno stimatore è tanto più efficiente quanto minore è la sua varianza



(T)
Distribuzione di T_1
ϑ
Distribuzione di T_2
ϑ

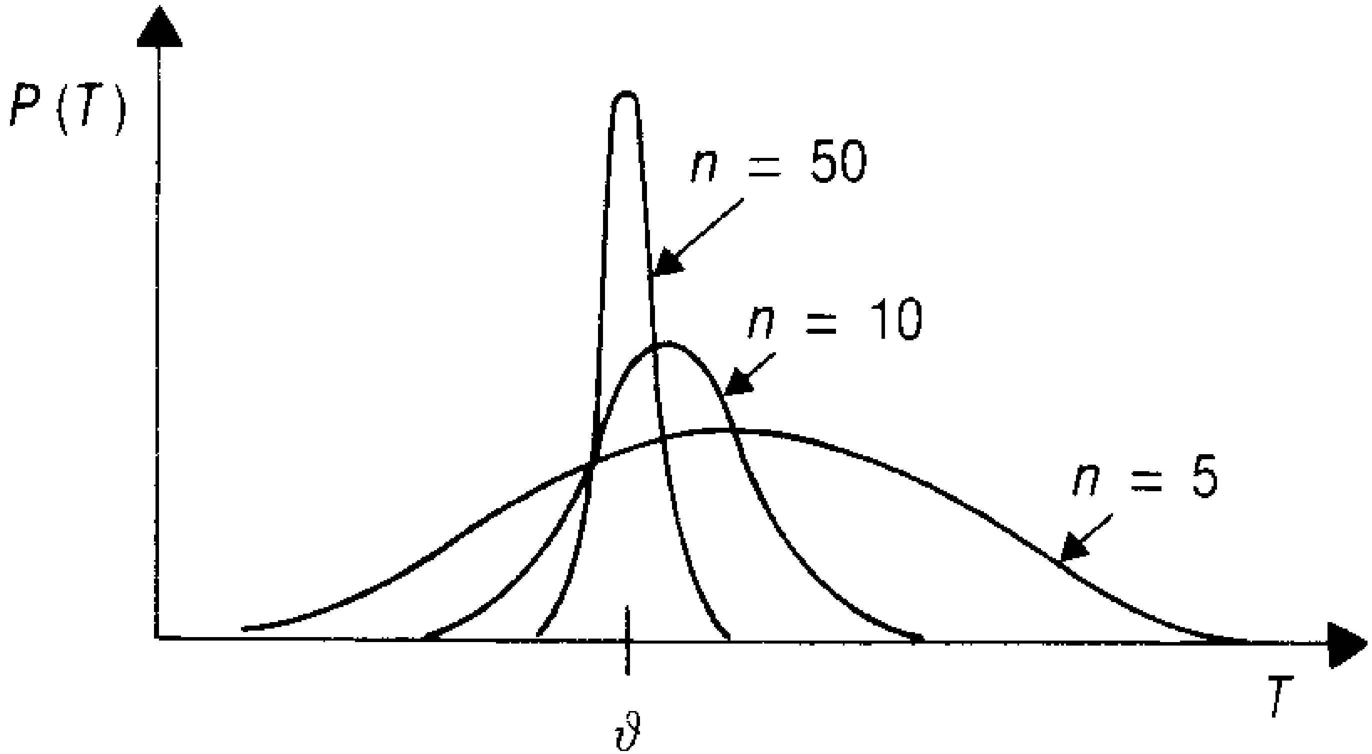
Ad esempio la media μ di una popolazione può essere valutata sia mediante la media campionaria $\overline{X}$ sia mediante la mediana campionaria $\tilde{X}$ ma tra i due stimatori la media campionaria risulta più EFFICIENTE

Consistenza

- Uno stimatore si dice consistente quando, al crescere delle dimensioni del campione, converge verso il parametro della popolazione da stimare; in formula:

$$\lim_{n \to \infty} P(|T - \vartheta| < \varepsilon) = 1$$

Distribuzione di T per campioni di ampiezza n



Stime puntuali e stime per intervallo

Una volta scelto lo stimatore che risulta migliore per stimare un parametro della popolazione

Si possono effettuare due tipi di stime

- **STIME PUNTUALI**
- **STIME INTERVALLARI**

Stime puntuali

Le stime puntuali di un parametro, sono quelle che si esprimono con un valore numerico e sono più semplici da calcolare:

si assume come stima il valore calcolato dal campione associandogli lo scarto quadratico medio detto *<u>errore medio di campionamento</u>* ***che indica l'errore mediamente commesso stimando, per mezzo del campione, il parametro incognito della popolazione***

ATTIVITA' DI LABORATORIO INFORMATICA

Per chiarire quanto detto finora, svolgere in laboratorio la seguente attività:

Si consideri una popolazione formata dalle misure: 130, 140, 180, 110.

Ci proponiamo di stimare il parametro μ utilizzando campioni bernoulliani di dimensione n=3, mediante due stimatori: lo stimatore media e lo stimatore mediana.

…….*fare riferimento al file prodotto*

Stime intervallari

Le stime intervallari di un parametro, sono quelle che si esprimono con un intervallo (<u>detto fiduciario</u>) che contenga con una certa probabilità il parametro incognito della popolazione.

La stima per intervallo <u>si applica se è nota la legge di distribuzione dello stimatore</u> e permette di valutare meglio l'errore che si può commettere.

diap. 50

....esamineremo quanto segue:

STIME PUNTUALI

- STIMA PUNTUALE DI UNA MEDIA
- STIMA PUNTUALE DI UNA FREQUENZA

STIME PER INTERVALLO

- INTERVALLO FIDUCIARIO PER LA MEDIA
- INTERVALLO FIDUCIARIO PER LE FREQUENZE RELATIVE

Per quanto riguarda le STIME PUNTUALI, ci rifaremo a quanto precedentemente studiato riguardo alla valutazione della media e della frequenza campionaria unitamente alle rispettive relazioni con i parametri della popolazione

Stima puntuale di una media

Il problema della stima puntuale di una media consiste nel valutare da un campione il valore μ della popolazione; come precedentemente visto tale valore (stima) corrisponde alla media della popolazione;

a tale stima si dovrà associare lo scarto quadratico medio delle medie campionarie (errore medio di campionamento), che, come precedentemente visto, NON corrisponde allo scarto quadratico medio della popolazione

Per questo motivo si assume come stimatore di σ^2 la

VARIANZA CAMPIONARIA CORRETTA $\widehat{S}^2$

$$\widehat{s}^2 = \frac{n}{n-1} s^2$$

DEVIAZIONE STANDARD CAMPIONARIA CORRETTA

$$\widehat{s} = \sqrt{\frac{n}{n-1}} s$$

diap. 71

(nel caso di estrazione bernoulliana)

Per questo motivo si assume come stimatore di σ^2 la

VARIANZA CAMPIONARIA CORRETTA $\widehat{S}^2$

$$\hat{s}^2 = s^2 \cdot \frac{n}{n-1} \cdot \frac{N}{N-1}$$

DEVIAZIONE STANDARD CAMPIONARIA CORRETTA

$$\hat{s} = s \cdot \sqrt{\frac{n}{n-1}} \cdot \sqrt{\frac{N}{N-1}}$$

diap. 71

(nel caso di estrazione senza ripetizione)

- un elemento alla volta -

Per questo motivo si assume come stimatore di σ² la

VARIANZA CAMPIONARIA CORRETTA $\widehat{S}^2$

$$\hat{s}^2 = \frac{N-1}{N} \cdot \frac{n}{n-1} \cdot s^2$$

DEVIAZIONE STANDARD CAMPIONARIA CORRETTA

$$\hat{s} = \sqrt{\frac{N-1}{N} \cdot \frac{n}{n-1}} \cdot s$$

diap. 71

(nel caso di estrazione senza ripetizione)

- estrazione in blocco -

PERTANTO, nel caso in cui non conosciamo σ, per determinare l'errore di campionamento

sarà sufficiente sostituire a σ nelle precedenti relazioni il valore di $\hat{S}$

$$\sigma_{\bar{x}} = \frac{\hat{s}}{\sqrt{n}} \qquad\qquad \sigma_{\bar{x}} = \frac{\hat{s}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

estrazione bernoulliana | **estrazione in blocco**

Per rendere minore l'errore di campionamento e avere quindi una stima più precisa, occorre aumentare il numero *n* degli elementi del campione

esercizi…….

Esercizi

1

I pesi degli studenti di un istituto si distribuiscono con legge normale. Vogliamo stimare il peso medio della popolazione costituita da 1650 studenti, utilizzando il seguente campione di 10 individui con il peso espresso in Kg:

45; 42,5; 43,8; 47; 51; 54,5; 60; 49,7; 44; 55.

Valutare i risultati ottenuti.

2

Una popolazione è composta dai numeri 1,2,3,4. Estrarre tutti i possibili campioni di dimensione n=2, dapprima secondo lo schema della estrazione senza ripetizione e successivamente utilizzando lo schema della estrazione in blocco; valutare i risultati ottenuti.

S

Esercizi

3 (stima puntuale, nota la deviazione standard della popolazione)

Un campione casuale di pile estratte da un lotto di 4 000 pile e sottoposto alla prova di durata ha presentato una durata media di 25 ore.
Da controlli precedenti si ritiene che la deviazione standard della durata delle pile prodotte sia pari a 4 ore.
Determinare nel caso sia di estrazione bernoulliana sia di estrazione in blocco (se la prima fosse possibile):

a) *la stima puntuale della durata media di tutta la popolazione;*

b) *l'errore medio di campionamento se la dimensione del campione è* $n = 10$, $n = 100$, $n = 200$.

4 (stima puntuale, senza la deviazione standard della popolazione)

Un campione di 100 lampadine, estratto da un lotto di 10 000 lampadine, presenta una durata media di 900 ore con una deviazione standard di 98 ore.
Determinare, nel caso sia di estrazione bernoulliana sia di estrazione in blocco (se la prima fosse possibile):

a) *la stima puntuale della durata media della popolazione;*

b) *l'errore medio di campionamento.*

S

Esercizi

5

Il diametro medio di alcuni cuscinetti a sfera è di 0,60 cm con una deviazione standard di 0,01 cm; effettuare una stima puntuale del diametro medio dei cuscinetti e determinare l'errore medio di campionamento. Inoltre:

a) qual è la probabilità che selezionando un campione casuale di 50 elementi, si ottenga un diametro medio del campione superiore a 0,601 cm?

b) qual è la probabilità che selezionando un campione casuale di 50 elementi, si ottenga un diametro medio del campione inferiore a 0,598 cm?

S

Esercizi

6

Una ditta applica il seguente criterio per decidere se accettare o meno i pezzi meccanici di un fornitore.

Un campione casuale di 40 pezzi forniti viene selezionato e misurato. Se il diametro medio del campione è tra 14.5 e 15.0 cm la fornitura viene accettata, in caso contrario viene respinta.

Qual è la probabilità di accettare la fornitura avente un diametro medio di:

a) 14 cm e una deviazione standard di 1 cm?

b) 14.7 cm e deviazione standard 0.8?

S

Esercizi

7

In un'indagine sociale per determinare il numero medio di ore lavorative giornaliere vengono selezionati casualmente un gruppo di 100 lavoratori in una città e determinato il numero medio di ore di lavoro effettuate in un periodo di tre giorni lavorativi.

La media e la deviazione standard campionarie sono risultate essere rispettivamente 8.40 ore e 1.10 ore.

Effettuare una stima puntuale della:

— media della popolazione;
— varianza della popolazione;
— deviazione standard della popolazione;
— varianza della distribuzione campionaria delle medie.

S

Stima puntuale di una frequenza

Il problema della stima puntuale di una frequenza consiste nel valutare quante unità di una popolazione posseggono una certa caratteristica; come precedentemente visto, per stimare il parametro della popolazione (cioè la probabilità che un certo attributo della popolazione ha di presentarsi), prenderemo in considerazione l'estrazione bernoulliana che dà luogo ad una distribuzione binomiale

Ricordando che la frequenza relativa del campione è:

$$f = \frac{x}{n}$$

- **x** numero degli elementi del campione che posseggono l'attributo

- ***n*** la dimensione del campione

e che tale stimatore oltre ad essere corretto è anche consistente

ASSOCIAMO alla stima f l'errore medio di campionamento che dalle formule precedenti è dato da:

$$\sigma_{\left(\frac{k}{n}\right)} = \sqrt{\frac{pq}{n}}$$

$$\sigma_{\left(\frac{k}{n}\right)} = \sqrt{\frac{pq}{n}}$$

essendo *p* ignoto si stima l'errore medio di campionamento sostituendo a *p* e *q* rispettivamente *f* ed (1-*f*)

$$s_f = \sqrt{\frac{f(1-f)}{n}}$$

che si può rendere piccolo, e quindi ottenere una stima più accurata, aumentando il numero *n* degli elementi del campione

nel caso di estrazione bernoulliana

Esempio

1. Si vuole stimare la percentuale delle famiglie di una regione che posseggono il televisore a colori. Si estrae un campione casuale di 200 famiglie e si ricava che 120 di esse posseggono il televisore a colori. Determinare la percentuale della popolazione che possiede il televisore a colori.

   Si ricava:

$$f = \frac{120}{200} = 0{,}60$$

$$s_f = \sqrt{\frac{0{,}60 \cdot 0{,}40}{200}} = 0{,}034$$

Quindi dal campione si stima che il 60% della popolazione possiede il televisore a colori, con scarto quadratico medio della stima del 3,4%.

Esercizi

8

In un referendum una proposta di legge ha ottenuto il 60 % dei voti; determinare la probabilità che un'indagine campionaria, su 100 votanti scelti a caso, mostri una maggioranza di voti a favore della proposta di legge.

9

Un campione di 40 elementi è estratto da una popolazione binomiale con parametro p=0,9.

Determinare la probabilità che nel campione si verifichi una percentuale di successi maggiore dell'88%.

S

Esercizio con attività di lab. inf.

10

Fare riferimento al <u>file.xls</u> relativamente alla scheda

"Distribuzione campionaria delle frequenze"

Stime intervallari

Rispetto a quanto detto precedentemente (diap. 10 e diap.31)

indichiamo con:

$$\vartheta - \varepsilon \leq \hat{\vartheta} \geq \vartheta + \varepsilon$$ **l'intervallo fiduciario**

α **il rischio di errore o livello di significatività**

(generalmente prefissato)

$1 - \alpha$ **il livello fiduciario o probabilità di confidenza**

PERTANTO

$$P(\vartheta - \varepsilon \leq \hat{\vartheta} \geq \vartheta + \varepsilon) = 1 - \alpha$$

indica che il valore dello stimatore $\hat{\vartheta}$ **è contenuto <u>nell'intervallo di confidenza</u>**

$$[\vartheta - \varepsilon ; \vartheta + \varepsilon]$$

con probabilità $1 - \alpha$

MENTRE

$$P(\hat{\vartheta}-\varepsilon \leq \vartheta \leq \hat{\vartheta}+\varepsilon) = 1-\alpha$$

indica che il valore del parametro ϑ è contenuto <u>nell'intervallo di confidenza</u>

$$\left[\hat{\vartheta}-\varepsilon ; \hat{\vartheta}+\varepsilon\right]$$

con probabilità $1-\alpha$

a chiarimento consideriamo la <u>seguente esemplificazione</u>:

Stima per intervallo di una media

Si consideri il caso assai ricorrente in cui il parametro da stimare è la media di una popolazione;

si vuole determinare un intervallo fiduciario attorno alla media campionaria, tale che, con una probabilità prefissata (pari al livello fiduciario), includa il valore della MEDIA della popolazione.

PRENDIAMO IN ESAME DUE CASI:

- **la popolazione è distribuita *normalmente* con varianza nota**

- **la popolazione è distribuita *normalmente* con varianza incognita**

I^ CASO: la popolazione è distribuita *normalmente* con varianza nota

In questo caso la media campionaria $\overline{X}$

è una variabile aleatoria con distribuzione

esattamente normale con

$$\mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

nel caso di estrazione bernoulliana

$$\mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

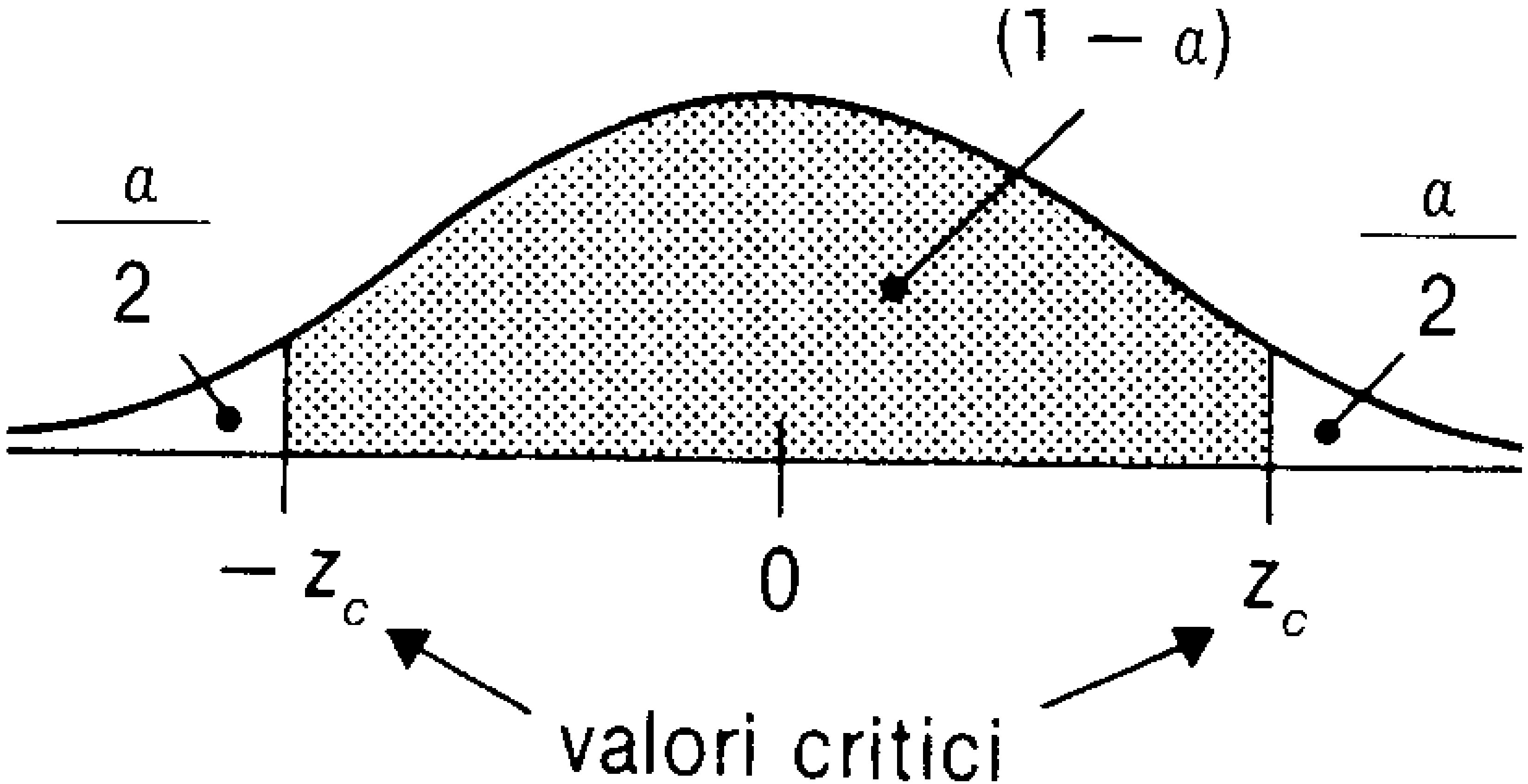
caso di estrazione senza ripetizione (in blocco)

Di conseguenza la v.c. $$Z = \frac{\bar{X} - \mu}{\sigma}$$

ha distribuzione normale standardizzata

Fissato il valore di α (o il livello di confidenza $1-\alpha$), è possibile determinare, per mezzo della tavola delle aree delle superfici sotto la curva normale standardizzata, i due valori z_1 e z_2 detti valori critici, tali che:

$$P(z_1 \leq Z \leq z_2) = 1 - \alpha$$



$$P(-z_c \leq Z \leq z_c) = 1 - \alpha$$

$$P(-z_c \le Z \le z_c) = 1 - \alpha$$

Sostituendo a Z il valore $\dfrac{\bar{X} - \mu}{\sigma}$ **si ha:**

$$P\left(-z_c \le \frac{\bar{X} - \mu}{\sigma} \le z_c \right) = 1 - \alpha$$

Moltiplicando ciascun termine della disuguaglianza per $\dfrac{\sigma}{\sqrt{n}}$ **e aggiungendo μ si ha:**

$$P\left(\mu - z_c \frac{\sigma}{\sqrt{n}} \le \bar{X} \le \mu + z_c \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

$$P\left(\mu - z_c \frac{\sigma}{\sqrt{n}} \le \bar{X} \le \mu + z_c \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

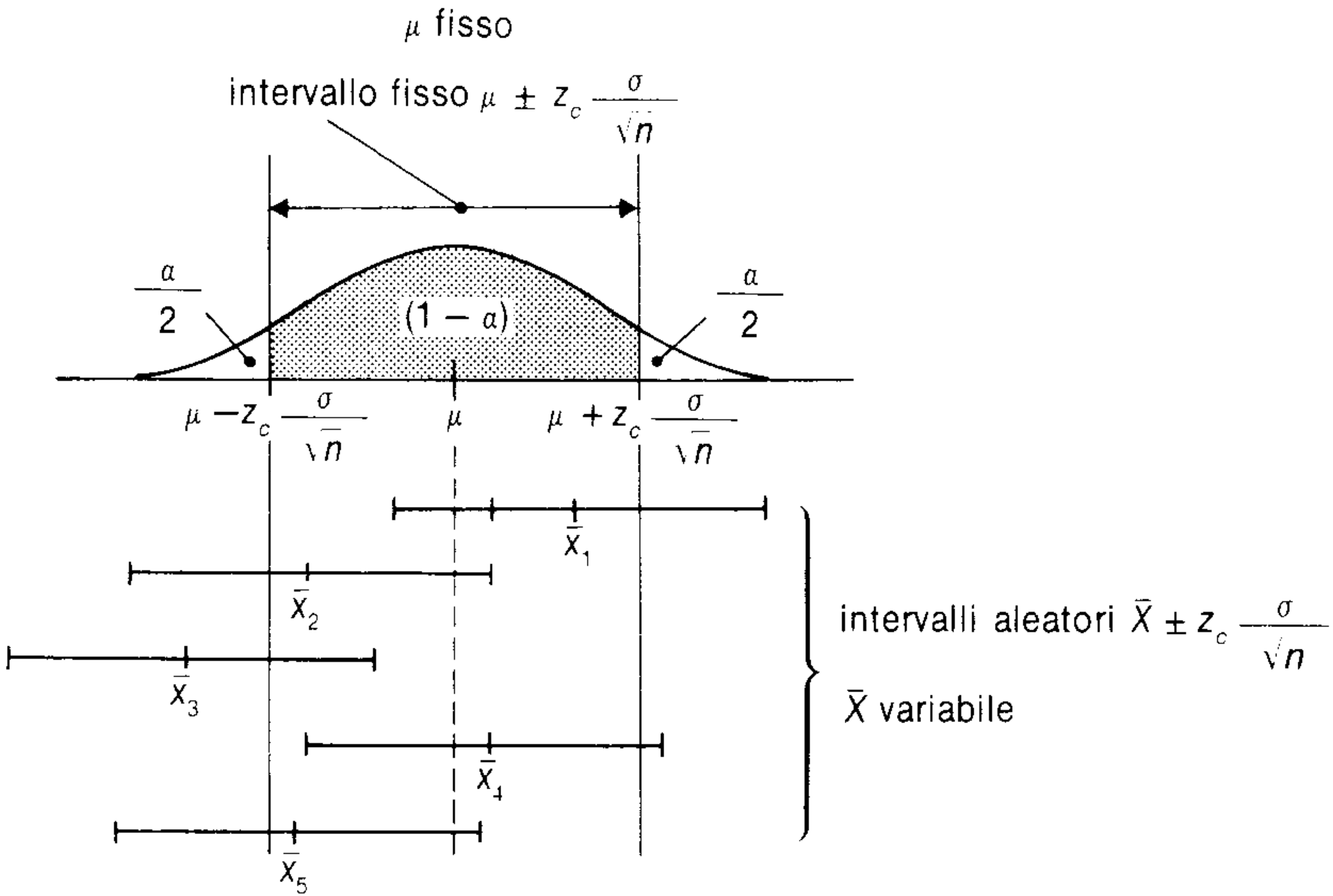
che esprime la probabilità che la media campionaria appartenga all'intervallo

$$\mu - z_c \frac{\sigma}{\sqrt{n}} \le \bar{X} \le \mu + z_c \frac{\sigma}{\sqrt{n}}$$

risolvendo la disuguaglianza in funzione di μ si ottiene:

$$P\left(\bar{X} - z_c \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_c \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

che esprime la probabilità che la media della popolazione appartenga all'*intervallo aleatorio*



Questa figura mette in evidenza che, per ogni campione estratto (con *n* costante), si ottiene una stima $\bar{X}$ **mentre la probabilità che sia esclusa è α**

Esempio 1 (costruzione intervalli di confidenza)

Scelto il livello di confidenza 1-α = 0,8664, vogliamo determinare l'intervallo

$$-z_c \le Z \le z_c$$

tale che $P(-z_c \le Z \le z_c) = 0{,}8664$

per la simmetria di $f(Z)$ **si ha:** $P(0 \le Z \le z_c) = 0{,}4332$

Utilizzando la tavola perveniamo al valore z_c=1,5

Pertanto l'intervallo richiesto è: $-1{,}5 \le Z \le 1{,}5$

Esempio 2 (costruzione intervalli di confidenza)

Costruire l'intervallo di confidenza della variabile standardizzata Z, al livello di significatività:

α= 31,73% ;

Per quanto riguarda α= 31,73% cioè α= 0,3173 si ha:

1-α=0,6827 cioè: $P(-z_c \leq Z \leq z_c) = 0,6827$

Essendo per la simmetria di $f(Z)$ $P(0 \leq Z \leq z_c) = 0,3413$

in corrispondenza di questo valore, sulla tavola troviamo che $|z_c| \cong 1$

Per cui possiamo essere fiduciosi che il 68,27% dei valori assunti dalla v.c. Z sia compreso fra -1 e +1; parimenti ci aspettiamo che il rimanente 31,73% sia esterno a questi limiti

Esempio 3 (costruzione intervalli di confidenza)

Un campione casuale bernoulliano di dimensione n=100 ha come media $\bar{x} = 50$ e proviene da una popolazione distribuita normalmente con σ=10. Fornire una stima della media per intervallo, con livello di confidenza del 95%.

Essendo (1-α)=0,95 cioè $P(-z_c \leq Z \leq z_c) = 0{,}95$

Per la simmetria di $f(Z)$ $P(0 \leq Z \leq z_c) = 0{,}475$

in corrispondenza di questo valore, sulla tavola troviamo che $|z_c| = 1{,}96$

Quindi la probabilità che la media della popolazione appartenga all'*intervallo aleatorio è:*

$$P\left(\overline{X} - z_c \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_c \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(50 - 1{,}96 \cdot \frac{10}{\sqrt{100}} \leq \mu \leq 50 + 1{,}96 \cdot \frac{10}{\sqrt{100}}\right) = 0{,}95$$

$$P(48{,}04 \leq \mu \leq 51{,}96) = 0{,}95$$

Siamo così fiduciosi che con probabilità del 95% il valore della media dell'universo sia compresa nell'intervallo trovato.

Esempio 4 (costruzione intervalli di confidenza)

Da un universo costituito da 2500 studenti, con varianza $\sigma^2 = 12,25.cm^2$ è stato estratto un campione casuale senza ripetizione con estraz. in blocco di 150 studenti. Determinare l'intervallo di confidenza per la stima dell'altezza media con probabilità di confidenza pari al 98%, sapendo che la statistica $\overline{X}$ assume nel campione il valore

$$\overline{x} = 169,5.cm$$

Trattandosi di campione estratto SENZA RIPETIZIONE (estraz. in blocco), si tratta di determinare la seguente probabilità:

$$P\left(\bar{X} - z_c \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \le \mu \le \bar{X} + z_c \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha$$

Essendo (1-α)=0,98 si ha: $P(-z_c \le Z \le z_c) = 0{,}98$

Per la simmetria di $f(Z)$ $P(0 \le Z \le z_c) = 0{,}49$

in corrispondenza di questo valore, sulla tavola troviamo mediante procedimento di interpolazione che $|z_c| = 2{,}326$ **per cui:**

$$P\left(169{,}5 - 2{,}326 \frac{3{,}5}{\sqrt{150}} \cdot \sqrt{\frac{2500-150}{2500-1}} \le \mu \le 169{,}5 + 2{,}326 \frac{3{,}5}{\sqrt{150}} \cdot \sqrt{\frac{2500-150}{2500-1}}\right) = 0{,}98$$

Cioè:

$$P(168{,}86 \le \mu \le 170{,}14) = 0{,}98$$

Possiamo concludere dicendo che l'altezza media dei 2500 studenti è compresa nel suddetto intervallo, consapevoli che il rischio massimo di errore è del 2%

n.b. ***procedimento di interpolazione*** :

$$(2{,}33 - 2{,}32) : (0{,}4901 - 0{,}4898) = (z_c - 2{,}32) : (0{,}49 - 0{,}4898)$$

Esercizi (costruzione intervalli di confidenza)

Costruire gli intervalli di confidenza della variabile standardizzata Z, ai livelli di significatività:

$$\alpha = 4{,}55\% \; ; \quad \alpha = 0{,}27\%$$

II^ CASO: la popolazione è distribuita *normalmente* con varianza incognita

Cosa mettere, nelle formule per il calcolo degli estremi di un intervallo di confidenza al posto di σ ?

Se conosciamo la varianza S^2 di un campione, la varianza corretta $\hat{S}^2$ può essere calcolata direttamente così, come si evince anche dalle precedenti formule:

$$\hat{s}^2 = s^2 \cdot \frac{n}{n-1}$$

Per campioni con ripetizione (bernoulliani)

$$\hat{s}^2 = s^2 \cdot \frac{n}{n-1} \cdot \frac{N}{N-1}$$

Per campioni senza ripetizione (un elemento alla volta)

$$\hat{s}^2 = s^2 \cdot \frac{N-1}{N} \cdot \frac{n}{n-1}$$

Per campioni senza ripetizione (estraz. In blocco)

Esempio 1 (costruzione intervalli di confidenza)

Misurando la lunghezza di un tavolo 65 volte, si è trovata una media $\overline{x} = 1{,}853.m$ *ed uno s.q.m.* $s = 0{,}012.m$

Stimare per intervallo la lunghezza reale μ *del tavolo con grado di fiducia del 95%, considerando che il campione sia di tipo bernoulliano.*

Possiamo utilizzare la nota relazione:

$$P\left(\overline{X} - z_c \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_c \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

$$P\left(\overline{X} - z_c \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_c \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Ma non abbiamo lo s.q.m. reale (σ), conoscendo soltanto quello del campione ($s = 0{,}012.m$); pertanto possiamo utilizzare la relazione:

$$P\left(\overline{X} - z_c \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \overline{X} + z_c \frac{\hat{s}}{\sqrt{n}}\right) = 1 - \alpha$$

Sostituendo a $\hat{S}$

$$\hat{s} = s\sqrt{\frac{n}{n-1}} = 0{,}012 \cdot \sqrt{\frac{65}{64}} = 0{,}01209$$

Essendo (1-α)=0,95 cioè $P(-z_c \leq Z \leq z_c) = 0{,}95$

Per la simmetria di $f(Z)$ $P(0 \leq Z \leq z_c) = 0{,}475($

in corrispondenza di questo valore, sulla tavola troviamo che

$$|z_c| = 1{,}96$$

Pertanto si avrà:

$$P\left(1{,}853 - 1{,}96\frac{0{,}01209}{\sqrt{65}} \leq \mu \leq 1{,}853 + 1{,}96\frac{0{,}01209}{\sqrt{65}}\right) = 0{,}95$$

$$P(1{,}85006 \leq \mu \leq 1{,}85594) = 0{,}95$$

Possiamo concludere dicendo che la lunghezza del tavolo è compresa nel suddetto intervallo, consapevoli di un margine di errore pari al 5%

Esercizio

11

Un campione di 100 lampadine, estratto da un lotto di 10.000 lampadine, presenta una durata media di 900 ore con una deviazione standard di 98 ore. Calcolare i limiti entro i quali è contenuta la media della popolazione al livello di confidenza del:

a) 90% **b) 95%** **c) 99%**

S

Stima per intervallo della frequenza relativa

Si consideri il caso di una popolazione con distribuzione binomiale.

Si vuole determinare un intervallo fiduciario per la stima del parametro p della popolazione, cioè della probabilità che si abbiano k successi su N osservazioni.

Dalla nota relazione

$$P\left(-z_c \leq \frac{\overline{X}-\mu}{\sigma} \leq z_c\right) = 1-\alpha$$

$$P\left(-z_c \leq \frac{\bar{X}-\mu}{\sigma} \leq z_c\right) = 1-\alpha$$

Trattandosi di distribuzione campionaria delle frequenze, la variabile standardizzata Z potrà scriversi: $Z = \dfrac{f_r - p}{\sigma}$

per cui $$P\left(-z_c \leq \frac{f_r - p}{\sigma} \leq z_c\right) = 1-\alpha$$

risolvendo la disuguaglianza in funzione di p si ottiene:

$$P\left(f_r - z_c \cdot \sigma \leq p \leq f_r + z_c \cdot \sigma\right) = 1-\alpha$$

Dalle formule precedenti, sostituendo σ = $\sqrt{\dfrac{f_r(1-f_r)}{n}}$ **si ha:**

$$P\left(f_r - z_c \cdot \sqrt{\frac{f_r(1-f_r)}{n}} \leq p \leq f_r + z_c \cdot \sqrt{\frac{f_r(1-f_r)}{n}}\right) = 1-\alpha$$

Esempio 1 (costruzione intervalli di confidenza)

Estraendo un campione di 80 elementi da una popolazione di tipo binomiale si sono ottenuti 20 successi; determinare una stima intervallare del parametro p della popolazione ad un livello fiduciario del 95%.

In questo caso: f_r =20/80=0,25

Essendo (1-α)=0,95 cioè $P(-z_c \leq Z \leq z_c) = 0{,}95$

Per la simmetria di $f(Z)$ $P(0 \leq Z \leq z_c) = 0{,}475($

in corrispondenza di questo valore, sulla tavola troviamo che

$$|z_c| = 1{,}96$$

Applicando la relazione:

$$P\left(f_r - z_c \cdot \sqrt{\frac{f_r(1-f_r)}{n}} \le p \le f_r + z_c \cdot \sqrt{\frac{f_r(1-f_r)}{n}} \right) = 1-\sigma$$

si ha: $P\left(0{,}25 - 1{,}96 \cdot \sqrt{\frac{0{,}25 \cdot 0{,}75}{80}} \le p \le 0{,}25 + 1{,}96 \cdot \sqrt{\frac{0{,}25 \cdot 0{,}75}{80}} \right) = 0{,}95$

cioè: $$P(0{,}155 \le p \le 0{,}345) = 0{,}95$$

Possiamo concludere dicendo il parametro p della popolazione è compreso nel suddetto intervallo, consapevoli di un margine di errore pari al 5%

Esercizi

12

E' stata condotta un'indagine su tutte le persone di maggiore età allo scopo di accertare la frequenza relativa dei possessori di patente. A tale scopo è stato estratto un campione di 2000 persone ed è stato accertato che 1850 di esse sono in possesso di patente. Vogliamo determinare la frequenza relativa dei possessori di patente dell'universo delle persone di maggiore età con un livello di confidenza del 95%.

13

Su un campione di 1800 abbonati alla televisione la percentuale di coloro che hanno dichiarato di avere seguito un dato programma è risultata pari al 27%. Entro quali valori sarà compresa, con probabilità uguale a 99%, la percentuale di coloro che hanno dichiarato di avere visto quel programma nell'universo di tutti gli abbonati?

S

RIFERIMENTO LIBRO DI TESTO

pag. 46-74

FINE DELLA LEZIONE

prof. C. Guida