
METODI STATISTICI E COMPUTAZIONALI

Stefania Spagnolo

Dipartimento di Matematica e Fisica, Univ. del Salento



LEZIONE 11

INTERVALLI DI CONFIDENZA

e limiti inferiori o superiori

[Dispense B. Chiandotto \(Università di Firenze\) Cap 4](#)
[Particle Data Book \(Statistics\)](#)

Finora

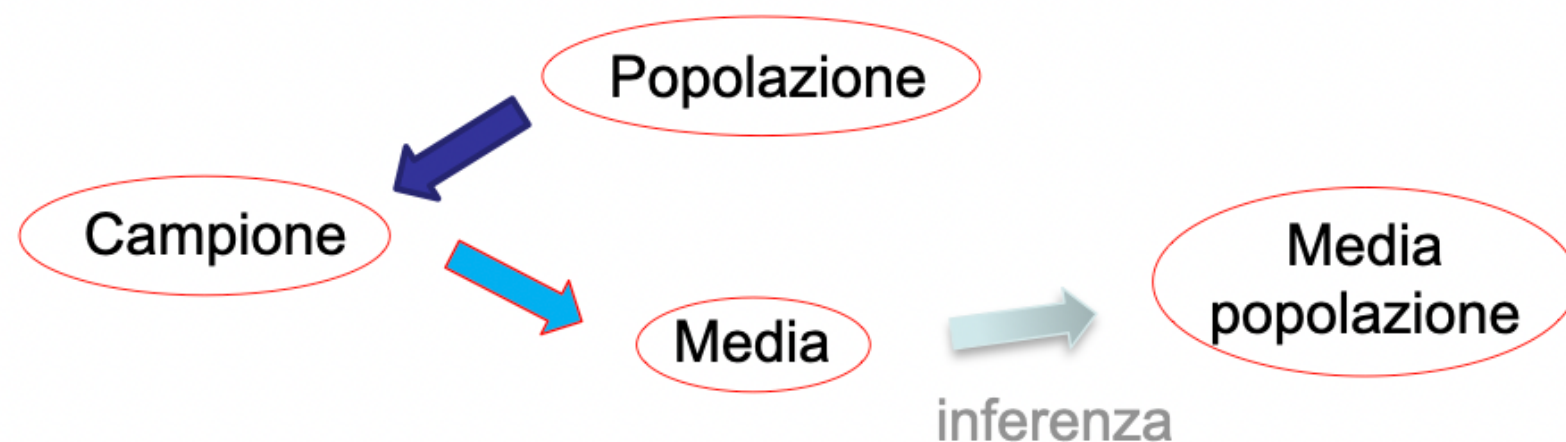
Se dispongo di un insieme (campione) di n misure x_i della grandezza fisica μ , quale è la miglior stima ($\hat{\mu}$) di μ ?

METODO DELLA MASSIMA VEROSIMIGLIANZA per stimare i parametri di una popolazione non nota a partire da un campione

Stimati valori
puntuali

INTERVALLI DI CONFIDENZA

- La stima dei parametri di una popolazione non nota a partire da un campione vista in precedenza è una **stima puntuale**. Si cerca, cioè, il miglior valore che mi approssima il parametro incognito.



- Oltre alla stima puntuale esiste un altro metodo di stima è la stima per intervallo.
- Scopo di questo metodo di stima è identificare un **intervallo** [denominato **di confidenza**] all'interno del quale con una certa probabilità sia presente il valore vero della popolazione.
- Normalmente **si dispone di un solo campione di dati estratto da una popolazione** e non si ha la possibilità di estrarne altri.
 - Come faccio ad individuare una regione che con una certa confidenza (o fiducia) contenga il valore vero del parametro della popolazione?

INTERVALLI DI CONFIDENZA

- Consideriamo il **caso** in cui ci interessi **determinare l'intervallo in cui si trova la media di una popolazione** con un certo livello di confidenza
- La teoria dei campioni ci ha permesso di ottenere questi risultati:
 - se si estraggono campioni casuali di ampiezza n da una popolazione con media μ e varianza σ^2 , la media campionaria sarà distribuita con
 - media pari a $\mu_{\langle X \rangle} = \mu$ e deviazione standard $\sigma_{\langle X \rangle} = \frac{\sigma}{\sqrt{n}}$
 - inoltre, il base al teorema del limite centrale ci dice che per n grandi la p.d.f della media campionaria tende ad una distribuzione gaussiana
 - quindi la variabile standardizzata Z tende ad essere gaussiana normale
 - se la varianza della popolazione non è nota, e *la popolazione è gaussiana*, la variabile T è distribuita come una t-di Student con $n-1$ gradi di libertà

INTERVALLI DI CONFIDENZA

Dalla teoria dei campioni è noto con quale probabilità troverò la media di un campione in un certo intervallo

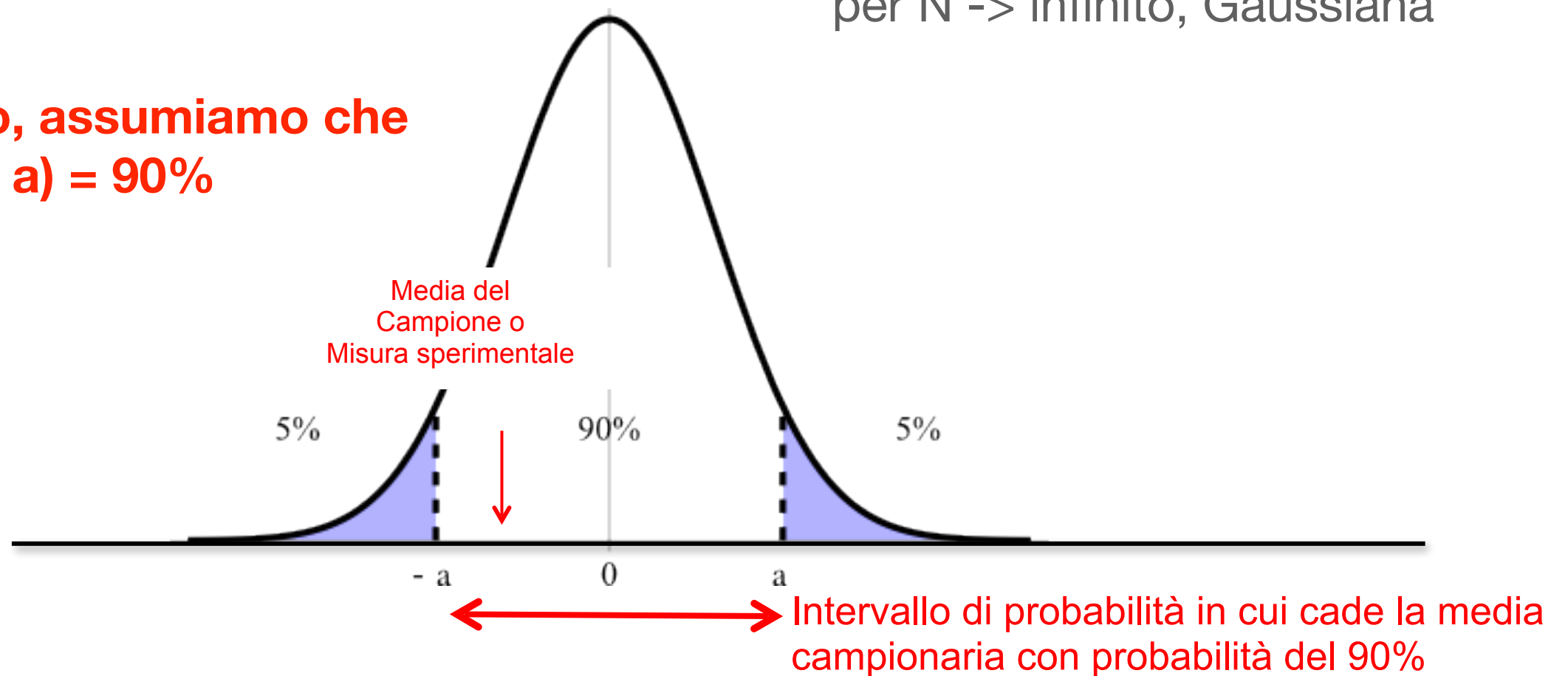
La p.d.f. della media campionaria ha

$$\mu_{\langle x \rangle} = \mu$$

$$\sigma_{\langle x \rangle} = \sigma / \sqrt{N}$$

per $N \rightarrow \text{infinito}$, Gaussiana

Per esempio, assumiamo che
 $P(-a < \langle x \rangle < a) = 90\%$



$$\mu_{\langle X \rangle} = \mu$$

La media della distribuzione delle medie campionarie COINCIDE con μ

Valore vero Media della popolazione μ

PROBABILITA' GAUSSIANA NORMALE

<https://www.math.arizona.edu/~jwatkins/normal-table.pdf>

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641



Se $N \rightarrow \infty$, possiamo assumere pdf Gaussiana, Quindi $a = 1.65 \sigma_{<x>}$

INTERVALLI DI CONFIDENZA

Dalla teoria dei campioni è noto con quale probabilità troverò la media di un campione in un certo intervallo

La p.d.f. della media campionaria ha

$$\mu_{\langle x \rangle} = \mu$$

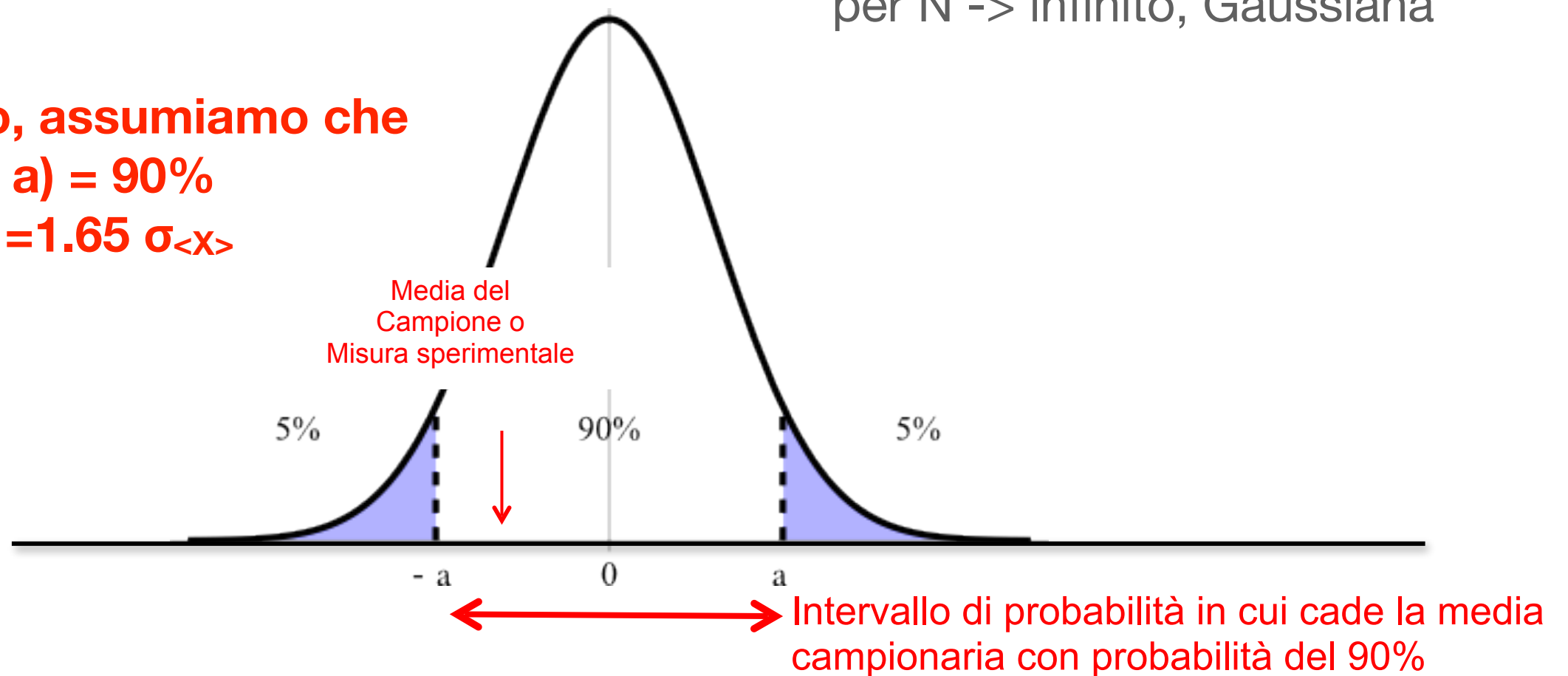
$$\sigma_{\langle x \rangle} = \sigma / \sqrt{N}$$

per $N \rightarrow \infty$, Gaussiana

Per esempio, assumiamo che

$$P(-a < \langle x \rangle < a) = 90\%$$

$$\text{Se } N \rightarrow \infty, a = 1.65 \sigma_{\langle x \rangle}$$



$$\mu_{\langle X \rangle} = \mu$$

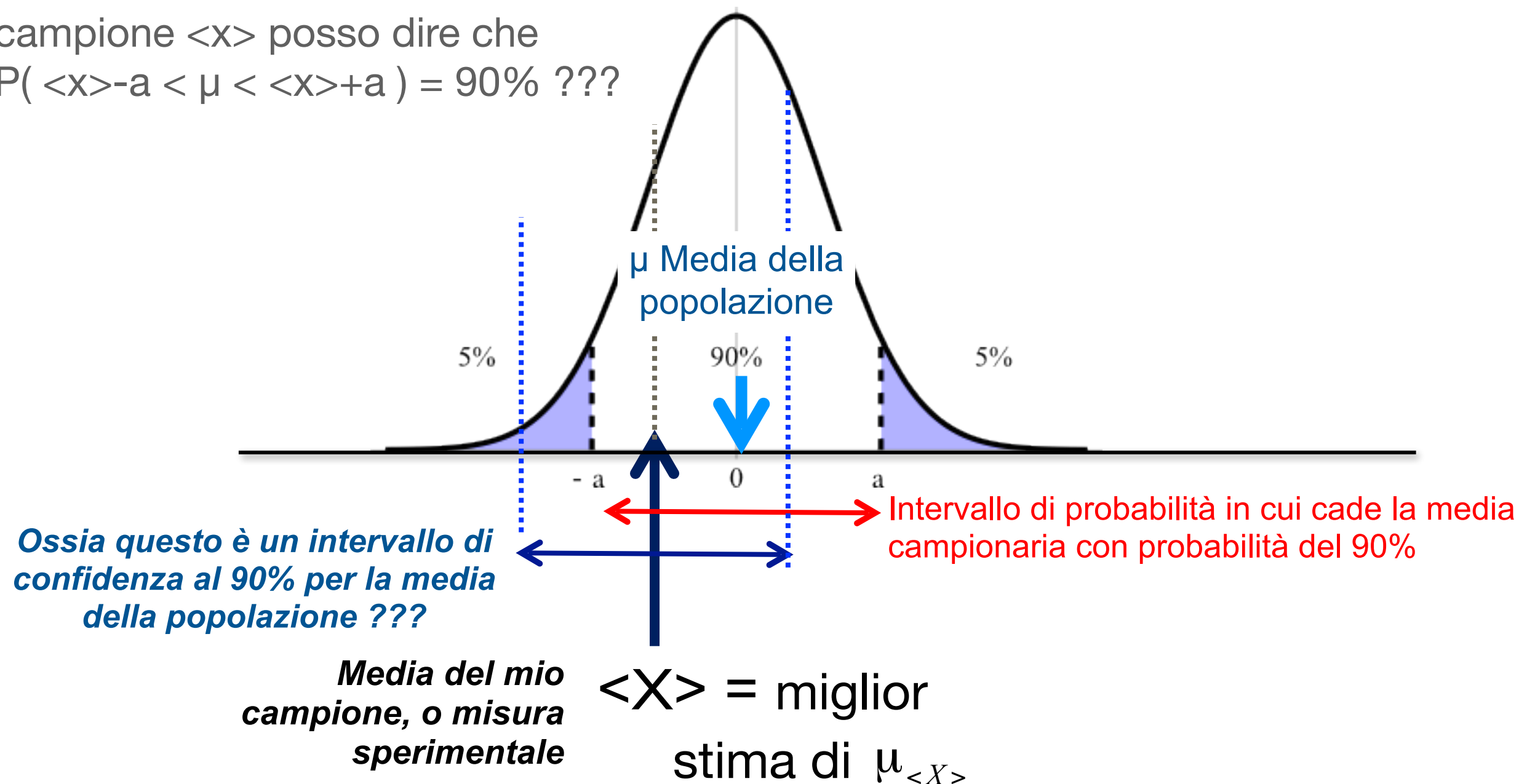
La media della distribuzione delle medie campionarie COINCIDE con μ

Valore vero Media della popolazione μ

INTERVALLI DI CONFIDENZA

Invertiamo il problema

Se conosco la media di un UNICO campione $\langle x \rangle$ posso dire che $P(\langle x \rangle - a < \mu < \langle x \rangle + a) = 90\%$???



INTERVALLI DI CONFIDENZA

- *Se conosco la media di un UNICO campione $\langle x \rangle$ posso dire che*
 - *$P(\langle x \rangle - a < \mu < \langle x \rangle + a) = 90\% ???$*
- *posso ricorrere alla variabile standardizzata Z (o T se la sigma della popolazione non è nota) per stimare l'intervallo di confidenza entro il quale mi aspetto di trovare il valore vero ????*

$$Z = \frac{\langle X \rangle - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$T = \frac{\langle X \rangle - \mu}{\frac{S}{\sqrt{n}}}$$

- ***Siamo abituati a farlo, infatti ...***

INFATTI USO DELLA VARIABILE STANDARDIZZATA Z

Dire che voglio che il mio valore vero μ sia contenuto con una certa **confidenza (1- α)** in prossimità del mio valore misurato $\langle X \rangle$ equivale a dire che la variabile Z sia prossima allo zero con la stessa confidenza (probabilità). Cioè

$$P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

1- α mi rappresenta il livello di confidenza richiesto

Quindi i valori $\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$ identificano una regione entro la quale Z si trova con probabilità 1- α

ossia
$$-z_{\frac{\alpha}{2}} < \frac{\langle X \rangle - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}} \quad \text{con CL} = 90\%$$

Successivamente diciamo che con il CL del 90%

$$\langle X \rangle - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \langle X \rangle + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

INTERVALLI DI CONFIDENZA

Risolvendo per μ

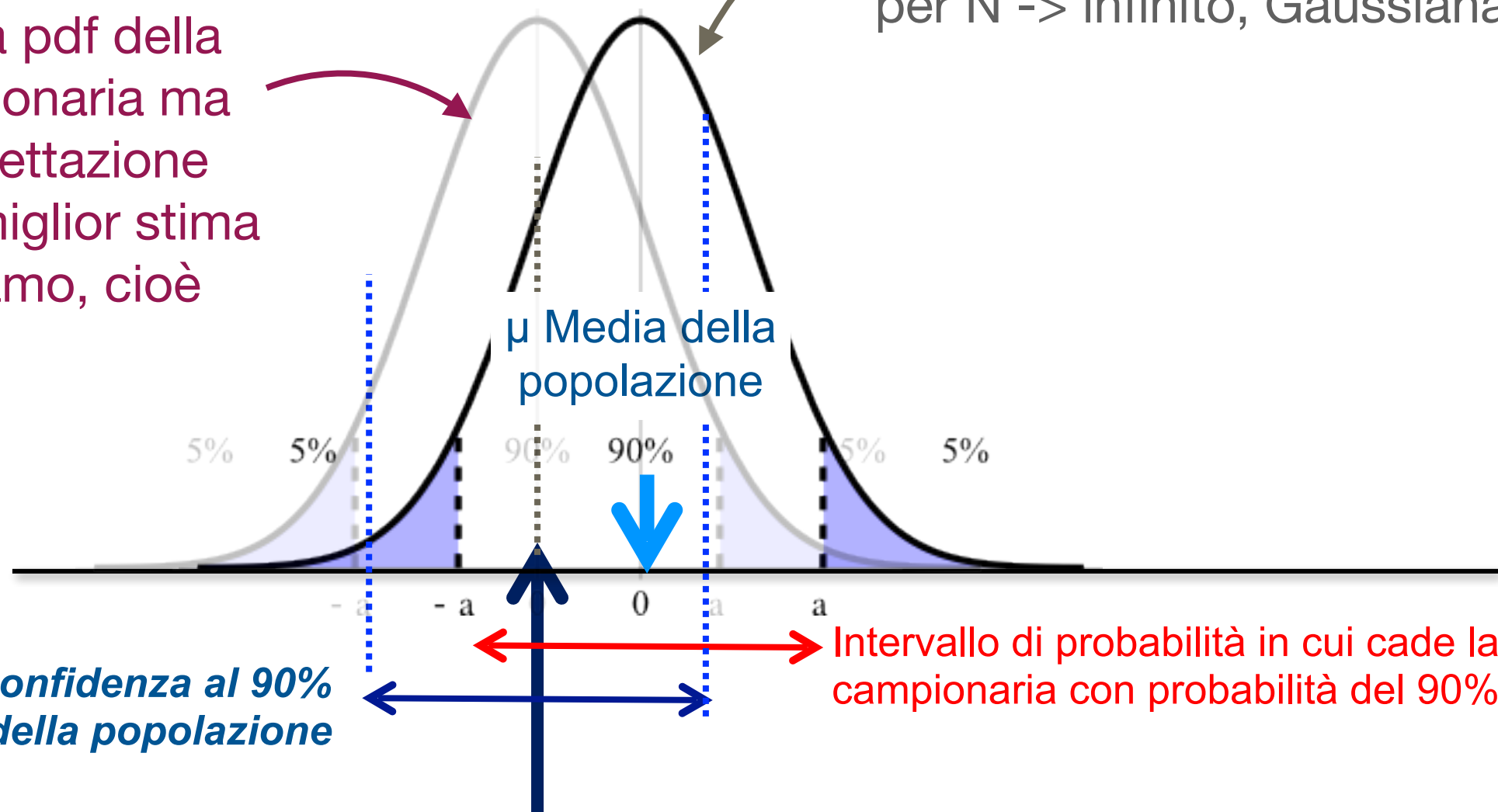
Stiamo trattando μ come una variabile aleatoria con la stessa pdf della media campionaria ma valore di aspettazione uguale alla miglior stima che ne abbiamo, cioè $\langle X \rangle$

La p.d.f. della media campionaria ha

$$\mu_{\langle X \rangle} = \mu$$

$$\sigma_{\langle X \rangle} = \sigma / \sqrt{N}$$

per $N \rightarrow \text{infinito}$, Gaussiana



Intervallo di confidenza al 90% per la media della popolazione

Intervallo di probabilità in cui cade la media campionaria con probabilità del 90%

Media del mio campione, o misura sperimentale

$\langle X \rangle = \text{miglior stima di } \mu_{\langle X \rangle}$

USO DELLA VARIABILE STANDARDIZZATA Z

■ *Esempio:*

- Ipotizziamo che la variabile aleatoria X rappresenti una certa popolazione con varianza $\sigma^2 = 25$. Supponendo di estrarre un campione di 100 elementi di calcolarne la media e di ottenere 5.4.

- Costruiamo l'intervallo di confidenza al 90% per la media μ della popolazione, ossia:

- identifico la regione che contiene il 90% della distribuzione per una distribuzione normale unitaria $\Rightarrow 90\% \rightarrow z(\alpha/2) = 1.65$

$$P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = \int_{-z_{\frac{\alpha}{2}}}^{z_{\frac{\alpha}{2}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x)^2}{2}} = 0.9 \quad z_{\frac{\alpha}{2}} = 1.65$$

- Affermo che la variabile standardizzata (Z) appartiene a quella regione se
 - $Z = (5.4 - \mu) / \sigma / \text{sqrt}(100) = (5.4 - \mu) / (5/10)$ appartiene a $[-1.65, 1.65]$ al 90% di C.L.
 - Ricavo gli estremi dell'intervallo di confidenza per μ

$$5.4 - 1.65 \cdot 0.5 < \mu < 5.4 + 1.65 \cdot 0.5$$

$$4.575 < \mu < 6.225$$

Al 90% di C.L.

USO DELLA VARIABILE STANDARDIZZATA T

Esempio.

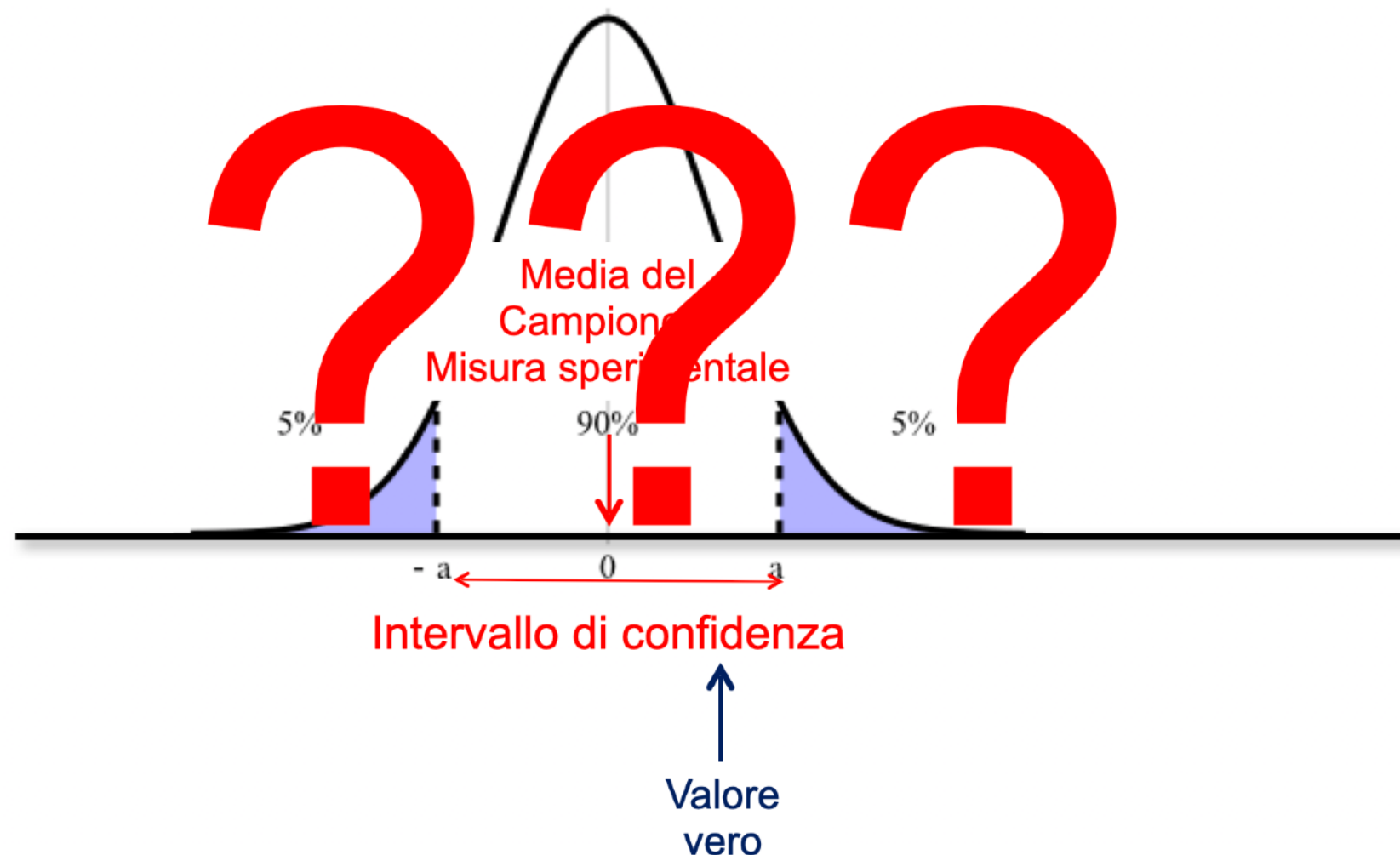
Le misure in Kg di 10 studenti maschi del primo anno di università sono:
60, 63, 60, 68, 70, 72, 65, 61, 69, 67.

Trovare l'intervallo di confidenza con grado di fiducia pari al 90% della popolazione studenti maschi del primo anno.

In questo caso non conosco la varianza della popolazione e il mio campione è limitato. Per la stima è più corretta una distribuzione t di Student

$$T = \frac{\langle X \rangle - \mu}{\frac{S}{\sqrt{n}}}$$

INTERVALLI DI CONFIDENZA



Notate che la distribuzione di probabilità è centrata intorno al valore sperimentale o stima del campione e non nel valore vero.
IL PASSAGGIO CONCETTUALE E' LEGITTIMO ?

E' utile introdurre la variabile aleatoria **media campionaria standardizzata**

$$Z = \frac{\langle X \rangle - \mu}{\frac{\sigma}{\sqrt{n}}}$$

La densità di probabilità della variabile Z sarà normale (media 0, varianza 1) se la popolazione di partenza segue una distribuzione di probabilità normale.

Nel caso in cui la popolazione iniziale non segua una distribuzione normale la Z per n grandi tenderà a una distribuzione normale.

Questo è una conseguenza del teorema del Limite Centrale.

Z mi permette di rispondere alla domanda

Come faccio a valutare con che probabilità, dato un campione di dimensione nota n , la sua media si possa discostare dalla media della popolazione?

Come faccio a valutare con che probabilità, dato un campione di dimensione nota n , la sua media si possa discostare dalla media della popolazione?

Si dimostra che se la popolazione segue una distribuzione normale allora la variabile

$$T = \frac{\langle X \rangle - \mu}{\frac{S}{\sqrt{n}}}$$

È una variabile aleatoria che segue la distribuzione di probabilità detta: t di Student.

$$f(t) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

Dove la funzioni gamma sono definite come

$$\Gamma(\nu) = \int_0^{\infty} e^{-x} x^{\nu-1} dx$$

E ν (unico parametro) è detto numero di gradi di libertà e vale (nel nostro caso)

$$\nu = n - 1$$

INTERVALLO DI CONFIDENZA ALLA NEYMAN

Introduciamo ora una definizione generale di intervallo di fiducia che giustifica l'approccio utilizzato finora per la media della popolazione.

Consideriamo il caso di un parametro incognito (μ) e una sola grandezza misurata (x). Data la grandezza misurata vogliamo stimare (inferire) il parametro incognito.

La p.d.f. del processo sarà una $f(x; \mu)$ (che possiamo considerare come una densità di probabilità di x condizionata a μ). Facciamo l'esempio di una p.d.f. gaussiana di cui assumiamo di conoscere con esattezza la varianza e immaginiamo che sia uguale a 1. Per un certo valore vero μ scriviamo:



Jerzy Neyman (16 aprile 1894 – 5 agosto 1981)
statistico polacco.

$$f(x, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

INTERVALLO DI CONFIDENZA ALLA NEYMAN

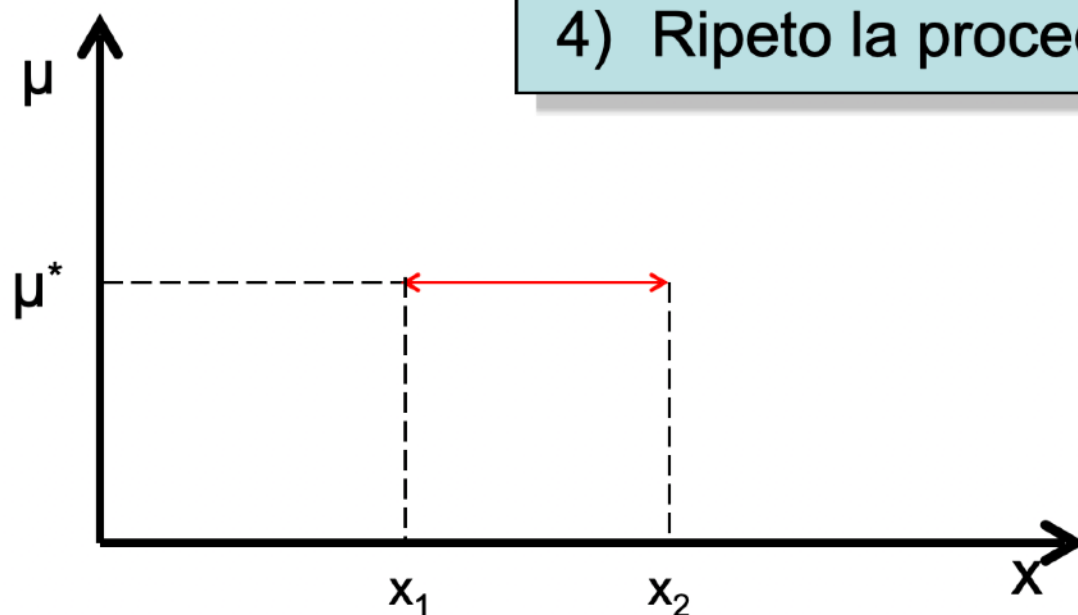
La tecnica di Neyman consiste nell'individuare nel piano μ, x la regione che identifica per ogni μ i valori che la x può assumere affinché si trovi intorno a μ con probabilità $1-\alpha$

Operativamente:

- 1) Prendo un valore di $\mu = \mu^*$
- 2) Determino gli estremi x_1 e x_2 tali che:

$$\int_{x_1}^{x_2} f(x, \mu^*) dx = 1-\alpha$$

- 3) Traccio una "linea" nel mio piano da (x_1, μ^*) a (x_2, μ^*)
- 4) Ripeto la procedura per ogni μ .



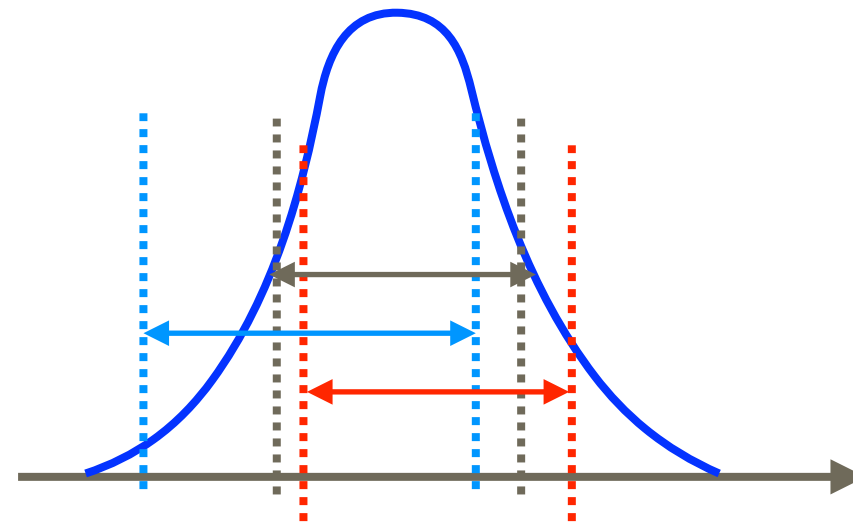
OSSERVA: ESISTE UNA AMBIGUITÀ

The Neyman confidence interval construction does not specify how you should draw, at fixed μ , the interval over the measured value that contains 90% of the probability content.

There are various different prescriptions:

- 1) add all x values greater than or less than a given value (upper limit or lower limit)
- 2) draw a central region with equal probability of the measurement falling above the region as below
- 3) starting with the parameter value which has maximum probability, keep adding points from more probable to less probable until the region contains 90% of the probability
- 4) The Feldman-Cousins prescription

OSSERVA: ESISTE UNA AMBIGUITÀ



Per esempio

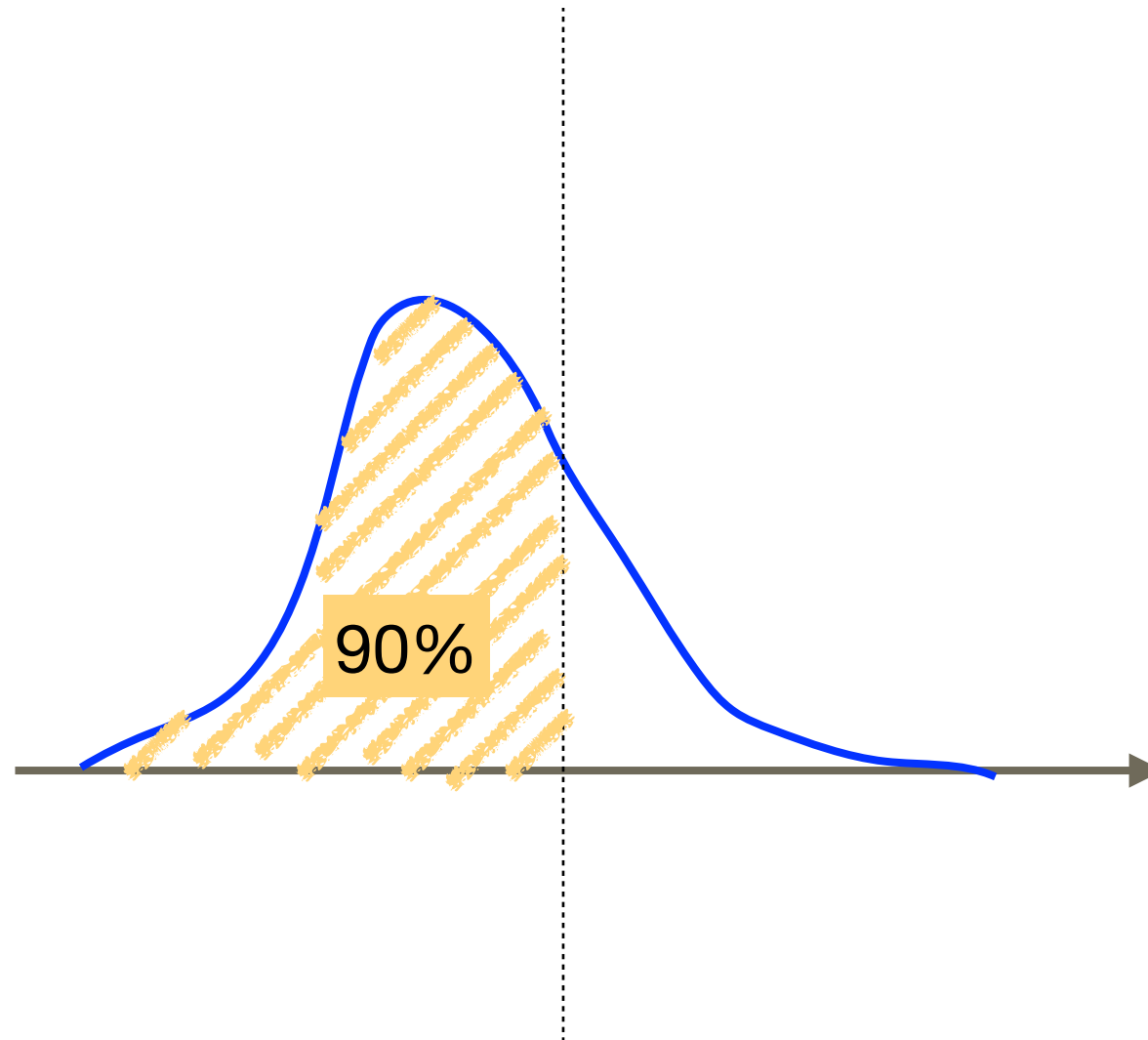
$$P(x_{1,1} < x < x_{1,2}) = P(x_{2,1} < x < x_{2,2}) = P(x_{3,1} < x < x_{3,2})$$

Quale di questi intervalli scegliamo ?

SOLUZIONE 1

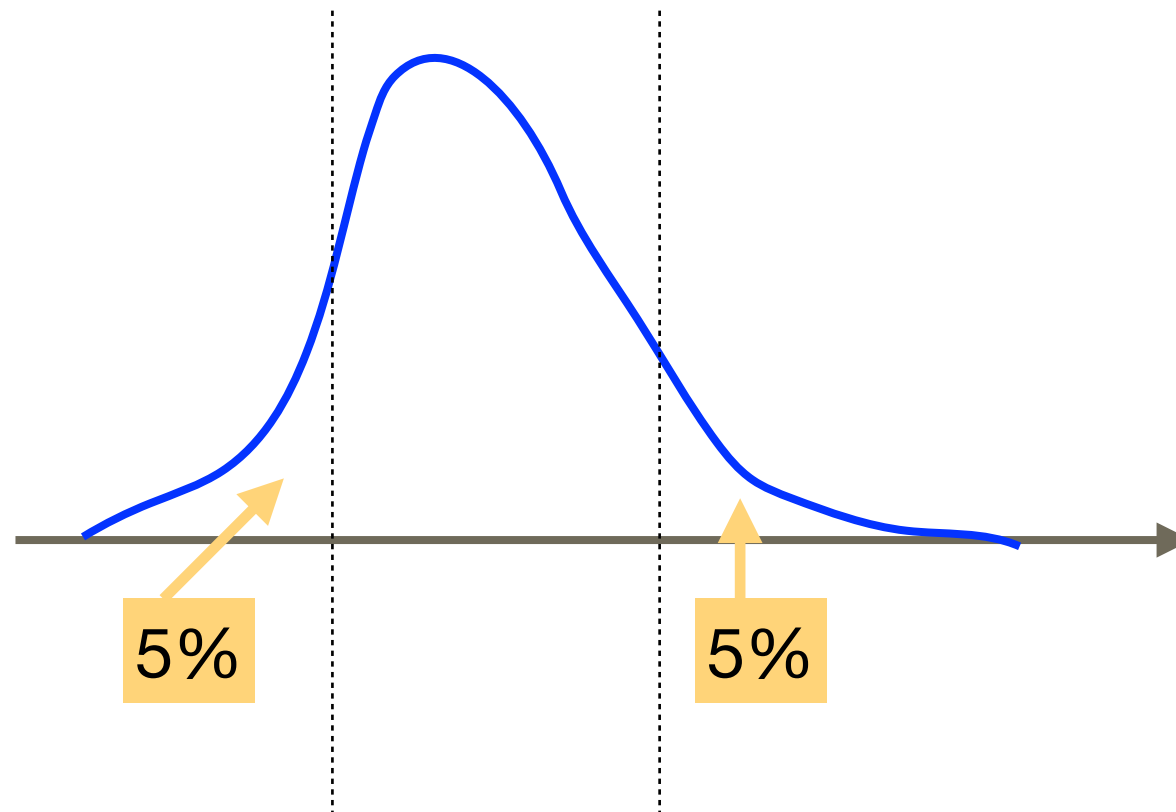
- 1) add all x values greater than or less than a given value (upper limit or lower limit)

2)



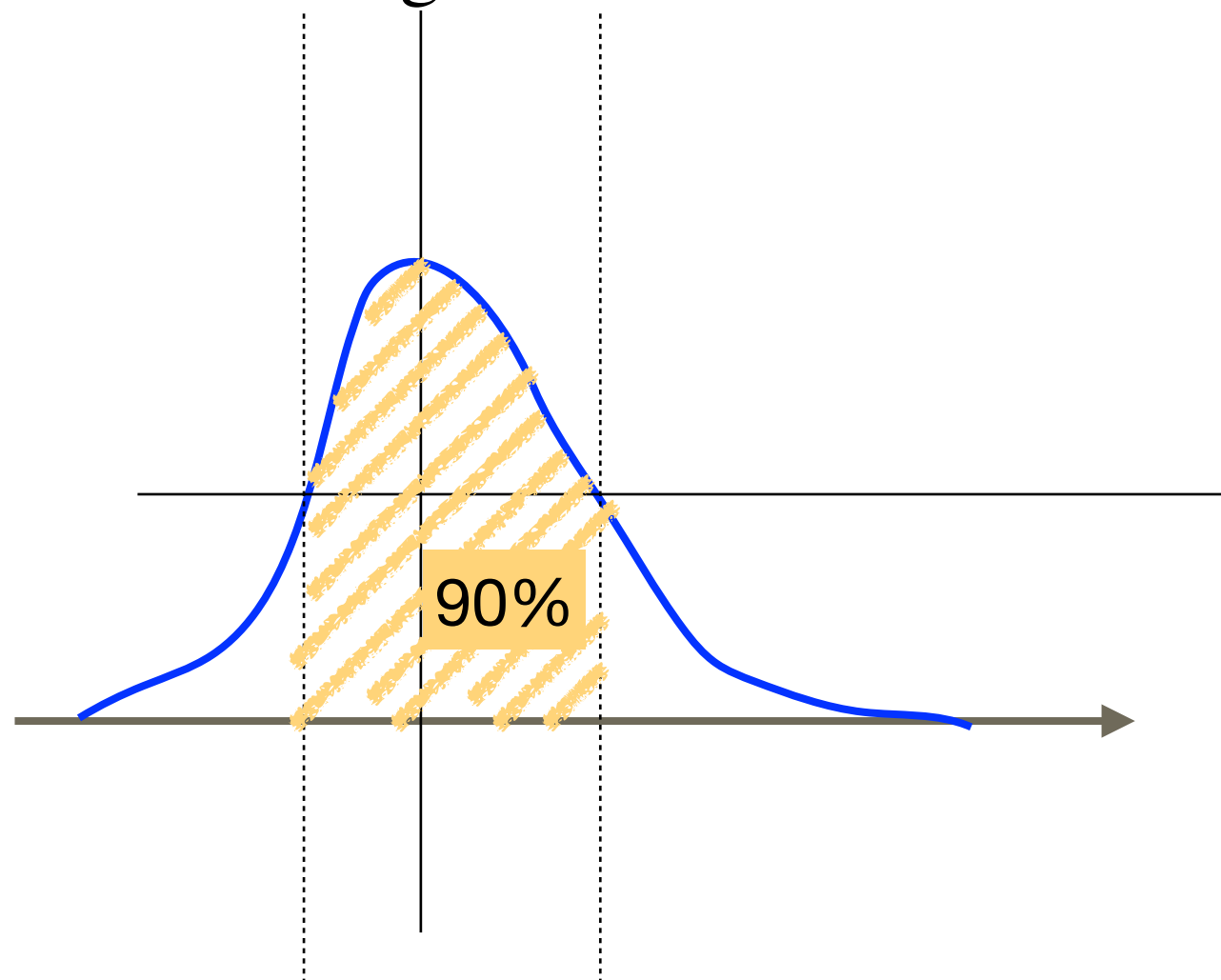
SOLUZIONE 2

- 2) draw a central region with equal probability of the measurement falling above the region as below

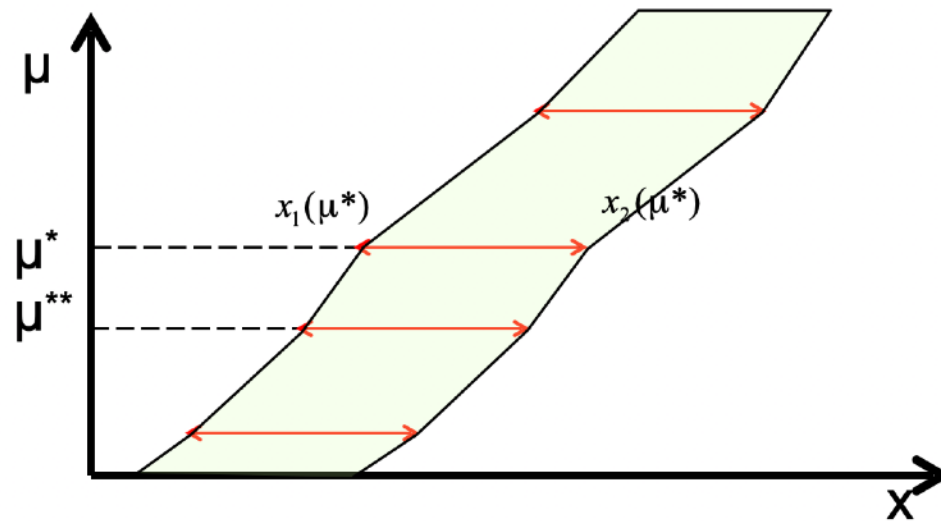


SOLUZIONE 3

- 3) starting with the parameter value which has maximum probability, keep adding points from more probable to less probable until the region contains 90% of the probability



INTERVALLO DI CONFIDENZA ALLA NEYMAN



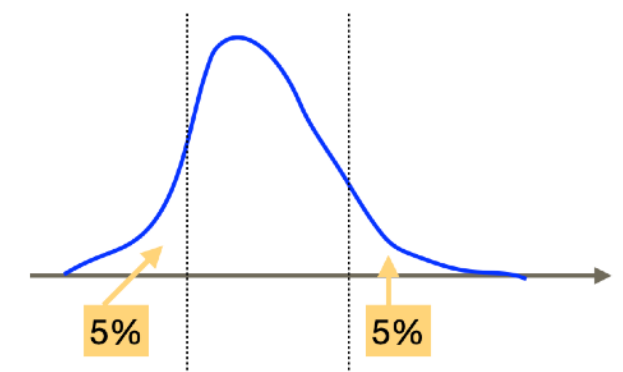
In questo modo costruisco nel piano μ, x una "banda di confidenza" costituita dalla ricopertura del piano con tutte le "linee" tracciate nella procedura.
Gli estremi della banda sono dati dai quantili:

$$\int_{-\infty}^{x_1(\mu)} f(x, \mu) dx = \frac{\alpha}{2}$$

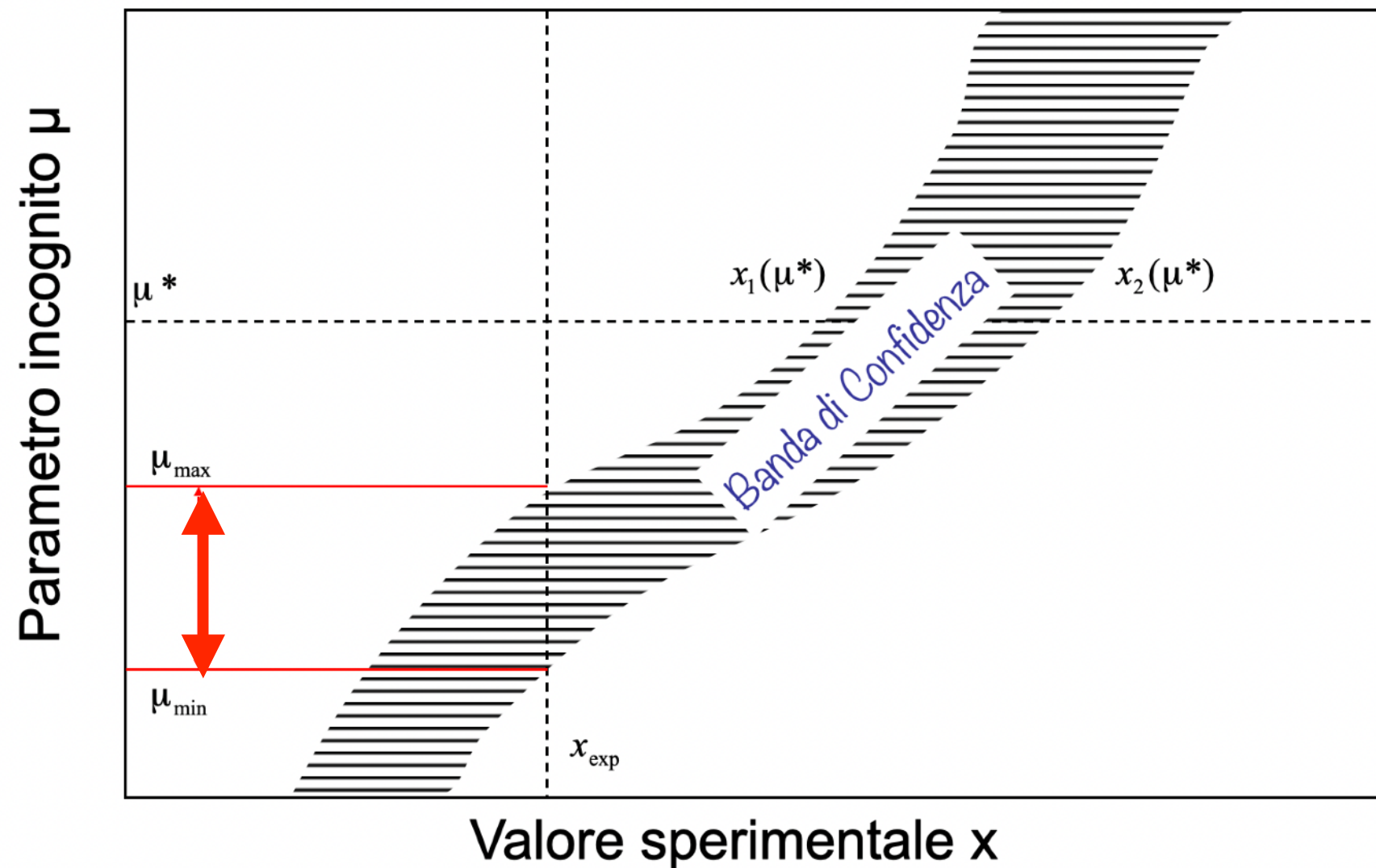
$$\int_{x_2(\mu)}^{+\infty} f(x, \mu) dx = 1 - \int_{-\infty}^{x_2(\mu)} f(x, \mu) dx = \frac{\alpha}{2}; \quad \int_{-\infty}^{x_2(\mu)} f(x, \mu) dx = 1 - \frac{\alpha}{2}$$

Dove $1-\alpha$ è il livello di fiducia voluto.

Sto adottando la soluzione 2) draw a central region with equal probability of the measurement falling above the region as below



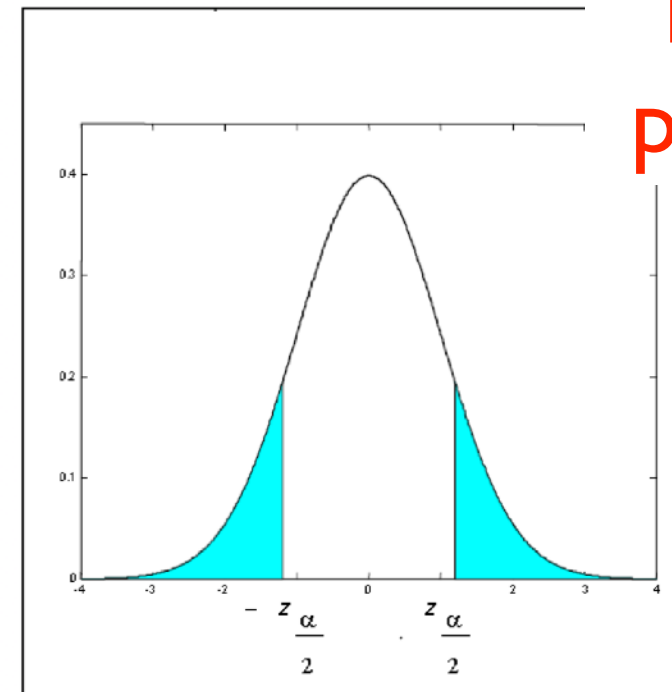
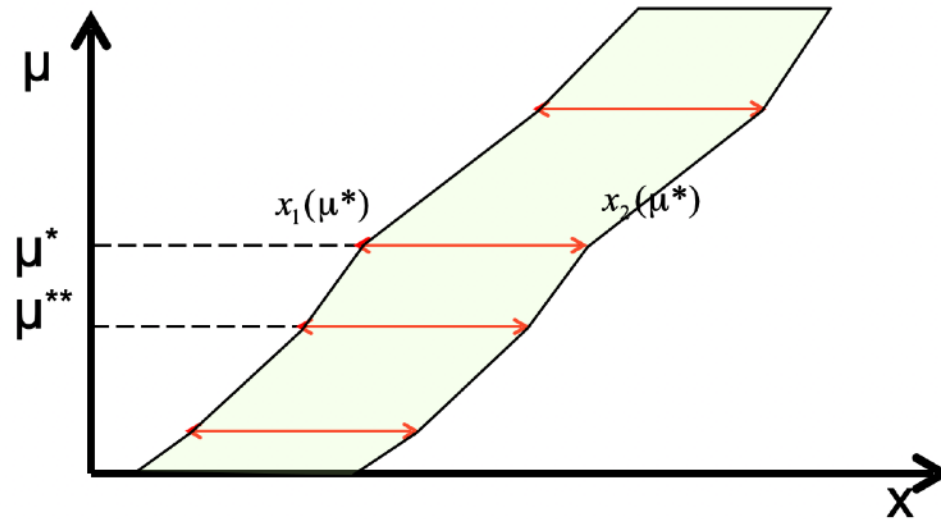
INTERVALLO DI CONFIDENZA ALLA NEYMAN



- 5) Dato un valore sperimentale misurato x_{exp} si traccia la linea verticale passante per questo valore e si identificano i due estremi in μ della banda; $[\mu_{\min}, \mu_{\max}]$ rappresenta l'intervallo di confidenza con fiducia $1-\alpha$ per il parametro μ

INTERVALLO DI CONFIDENZA ALLA NEYMAN

Nel caso di
pdf gaussiana



In questo modo costruisco nel piano μ, x una “banda di confidenza” costituita dalla ricopertura del piano con tutte le “linee” tracciate nella procedura. Gli estremi della banda sono dati dai quantili:

$$\int_{-\infty}^{x_1(\mu)} f(x, \mu) dx = \frac{\alpha}{2}$$

$$\int_{x_2(\mu)}^{+\infty} f(x, \mu) dx = 1 - \int_{-\infty}^{x_2(\mu)} f(x, \mu) dx = \frac{\alpha}{2}; \quad \int_{-\infty}^{x_2(\mu)} f(x, \mu) dx = 1 - \frac{\alpha}{2}$$

$$f(x, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

Dove $1-\alpha$ è il livello di fiducia voluto.

INTERVALLO DI CONFIDENZA ALLA NEYMAN

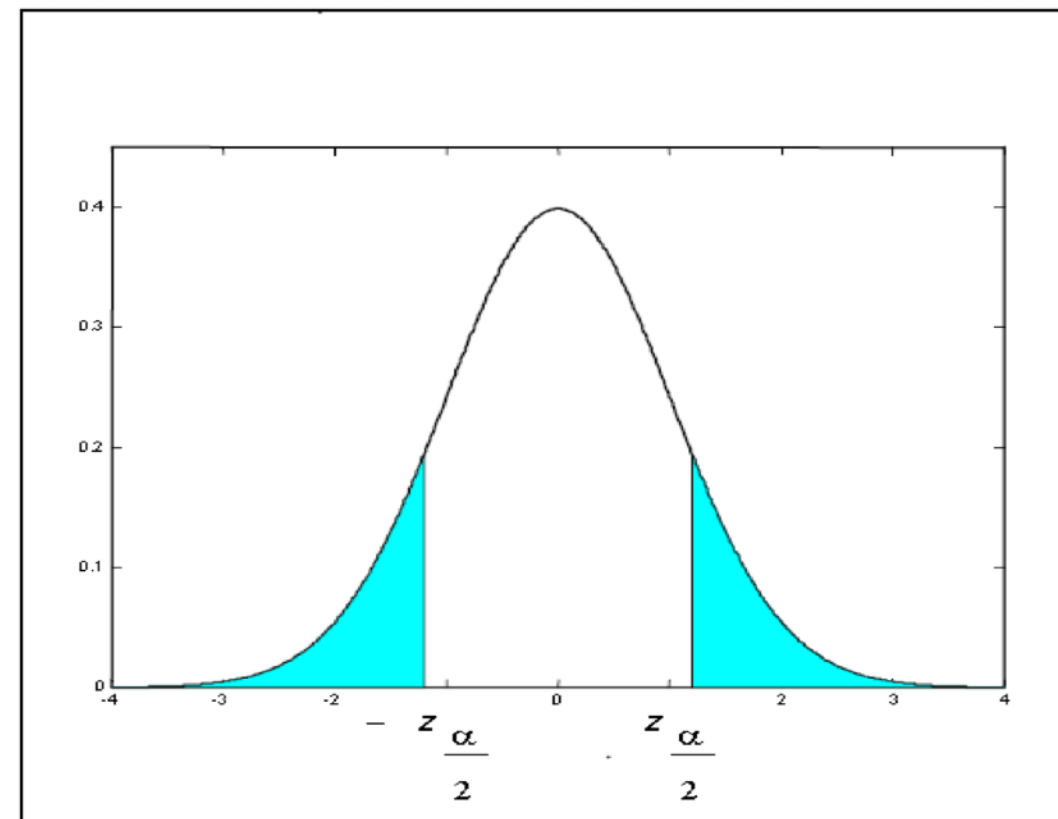
Nel caso di
pdf gaussiana

Esempio:

consideriamo una popolazione con p.d.f. gaussiana
In questo caso i quantili x_1 e x_2 che identificano gli estremi della banda di
confidenza sono facilmente calcolabili.

$$\int_{-\infty}^{x_1(\mu)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{\alpha}{2} \Rightarrow x_1(\mu) = \mu - N(\alpha)\sigma$$

$$\int_{-\infty}^{x_2(\mu)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 - \frac{\alpha}{2} \Rightarrow x_2(\mu) = \mu + N(\alpha)\sigma$$



Dove il “numero di sigma” ($N(\alpha)$) dipende dalla scelta di α .

INTERVALLO DI CONFIDENZA ALLA NEYMAN

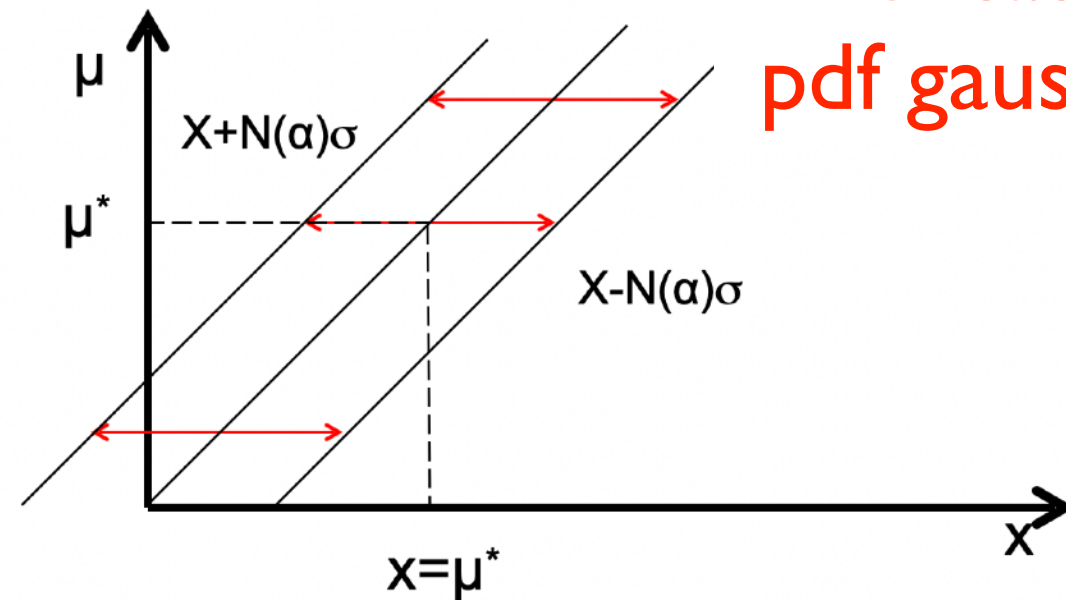
La banda di confidenza consiste in due rette parallele equidistanti dalla retta bisettrice $\mu=x$ e l'intervallo di confidenza si calcola facilmente

$$x_1 = \mu - N(\alpha)\sigma$$

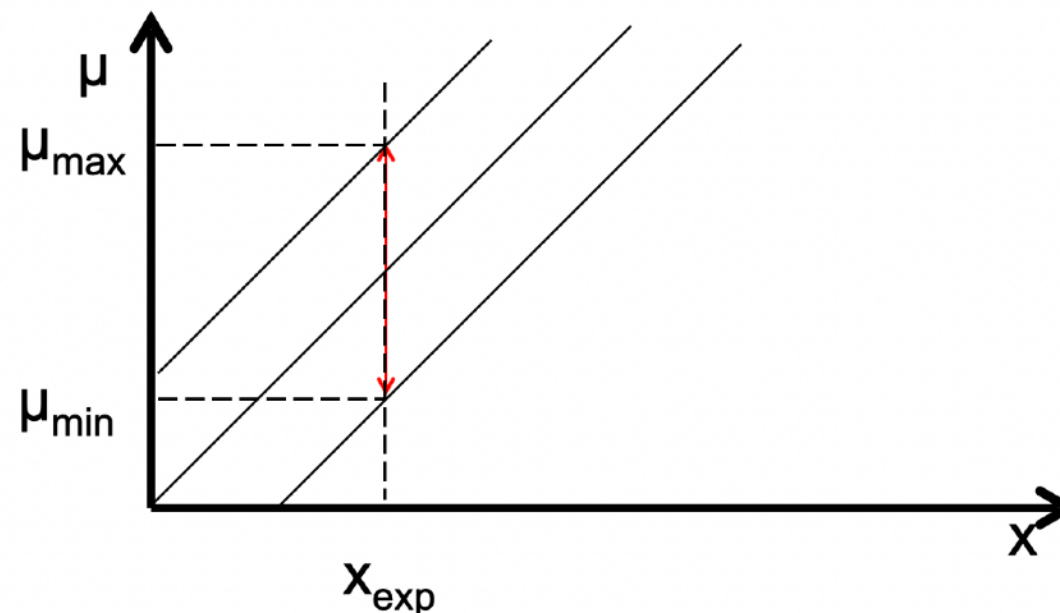
$$x_2 = \mu + N(\alpha)\sigma$$

$$\begin{cases} \mu_{\max} = x_{\text{exp}} + N(\alpha)\sigma \\ \mu_{\min} = x_{\text{exp}} - N(\alpha)\sigma \end{cases} \Rightarrow x_{\text{exp}} - N(\alpha)\sigma < \mu < x_{\text{exp}} + N(\alpha)\sigma$$

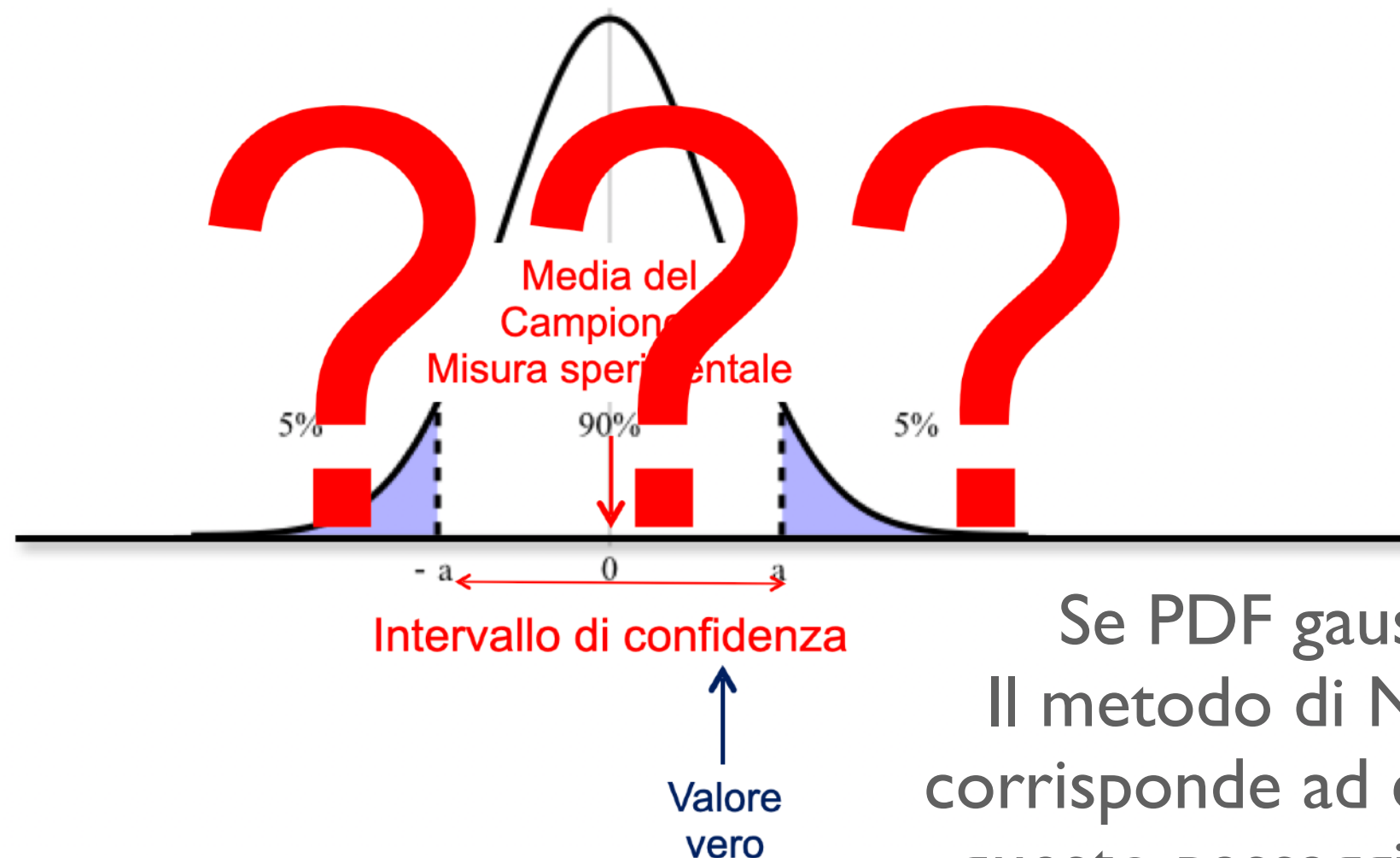
Nel caso di pdf gaussiana



Ritroviamo la solita regola



INTERVALLI DI CONFIDENZA



Se PDF gaussian,
Il metodo di Neyman
corrisponde ad effettuare
questo passaggio logico

Notate che la distribuzione di probabilità è centrata intorno al
valore sperimentale o stima del campione e non nel valore vero.
IL PASSAGGIO CONCETTUALE E' LEGITTIMO ?

INTERVALLI DI CONFIDENZA

Nel caso di
pdf t-di
Student

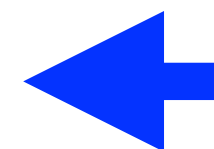
- Nel caso in cui non si conosce la varianza della distribuzione gaussiana della popolazione, per stimare la probabilità della media campionaria si utilizza la variabile T distribuita come la t -di Student
- Anche in questo caso, se si applica il metodo di Neyman per costruire una banda di confidenza e determinare intervalli di confidenza per la media della popolazione sulla base di una misura $\langle x \rangle$, si ottiene una banda di confidenza delimitata da due rette parallele ed equidistanti dalla bisettrice (corrispondente a $\mu = \langle x \rangle = x_{exp}$);
- il risultato sarà nuovamente
 - $x_{exp} - N(\alpha)\sigma_s < \mu < x_{exp} + N(\alpha)\sigma_s$ al CL $1-\alpha$
 - dove $\sigma_s = S/\sqrt{n}$ e' la deviazione standard (stimata dal campione di dimensione n) della distribuzione di probabilità della media campionaria

INTERVALLI DI CONFIDENZA

- Nel caso in cui non si conosce la varianza della distribuzione gaussiana della popolazione, per stimare la probabilità della media campionaria si utilizza la variabile T distribuita come la t-di Student
- Anche in questo caso, se si applica il metodo di Neyman per costruire una banda di confidenza e determinare intervalli di confidenza per la media della popolazione sulla base di una misura $\langle x \rangle$, si ottiene una banda di confidenza delimitata da due rette parallele ed equidistanti dalla bisettrice (corrispondente a $\mu = \langle x \rangle = x_{exp}$);
- il risultato sarà nuovamente
 - $x_{exp} - N(\alpha)\sigma_s < \mu < x_{exp} + N(\alpha)\sigma_s$ al CL $1-\alpha$
 - dove $\sigma_s = S/\sqrt{n}$ e' la deviazione standard (stimata dal campione di dimensione n) della distribuzione di probabilità della media campionaria

Nel caso di
pdf t-di
Student

E' chiaro che il
metodo di Neyman
si riconduce a
questa regola tutte
le volte in cui la
p.d.f. è simmetrica



PROBABILITA' GAUSSIANA NORMALE

<https://www.math.arizona.edu/~jwatkins/normal-table.pdf>

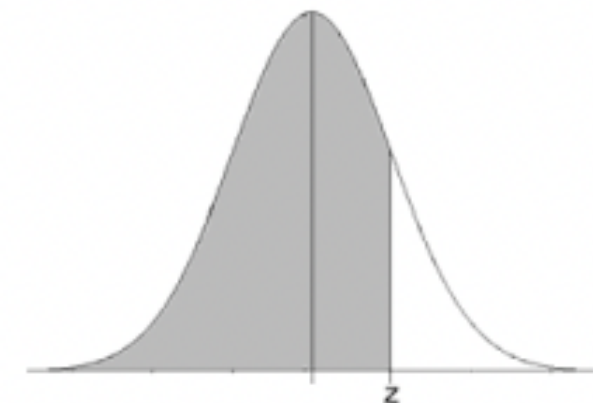
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641



PROBABILITA' GAUSSIANA NORMALE

<https://www.math.arizona.edu/~jwatkins/normal-table.pdf>

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998



INTERVALLO DI CONFIDENZA ALLA NEYMAN

Oltre il paradigma gaussiano

Esempio: supponiamo che la $f(x, \mu)$ sia una distribuzione esponenziale di cui μ sia il parametro.

$$f(x, \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \quad \text{definita } [0, \infty)$$

Supponiamo di avere un'unica misura sperimentale x_{exp} e da questa vogliamo identificare un intervallo di confidenza all'interno del quale cada μ al 68%.

Nella distribuzione esponenziale la varianza è pari a μ^2

Avendo una sola misura la miglior stima di μ che possiamo fare è $\mu = x_{\text{exp}}$ dato che μ coincide con la media (o valore di aspettazione) dell'esponenziale. In questo esempio la media delle misure coincide con il valore dell'unica misura.

Utilizzando l'approccio che non tenga conto della costruzione dell'intervallo di confidenza della tecnica di Neyman avrei concluso che al 68%

$$x_{\text{exp}} - x_{\text{exp}} < \mu < x_{\text{exp}} + x_{\text{exp}}$$

(valore +/- una sigma)

INTERVALLO DI CONFIDENZA ALLA NEYMAN

Oltre il paradigma gaussiano

Calcoliamo, invece, l'intervallo di confidenza con la tecnica di Neyman.

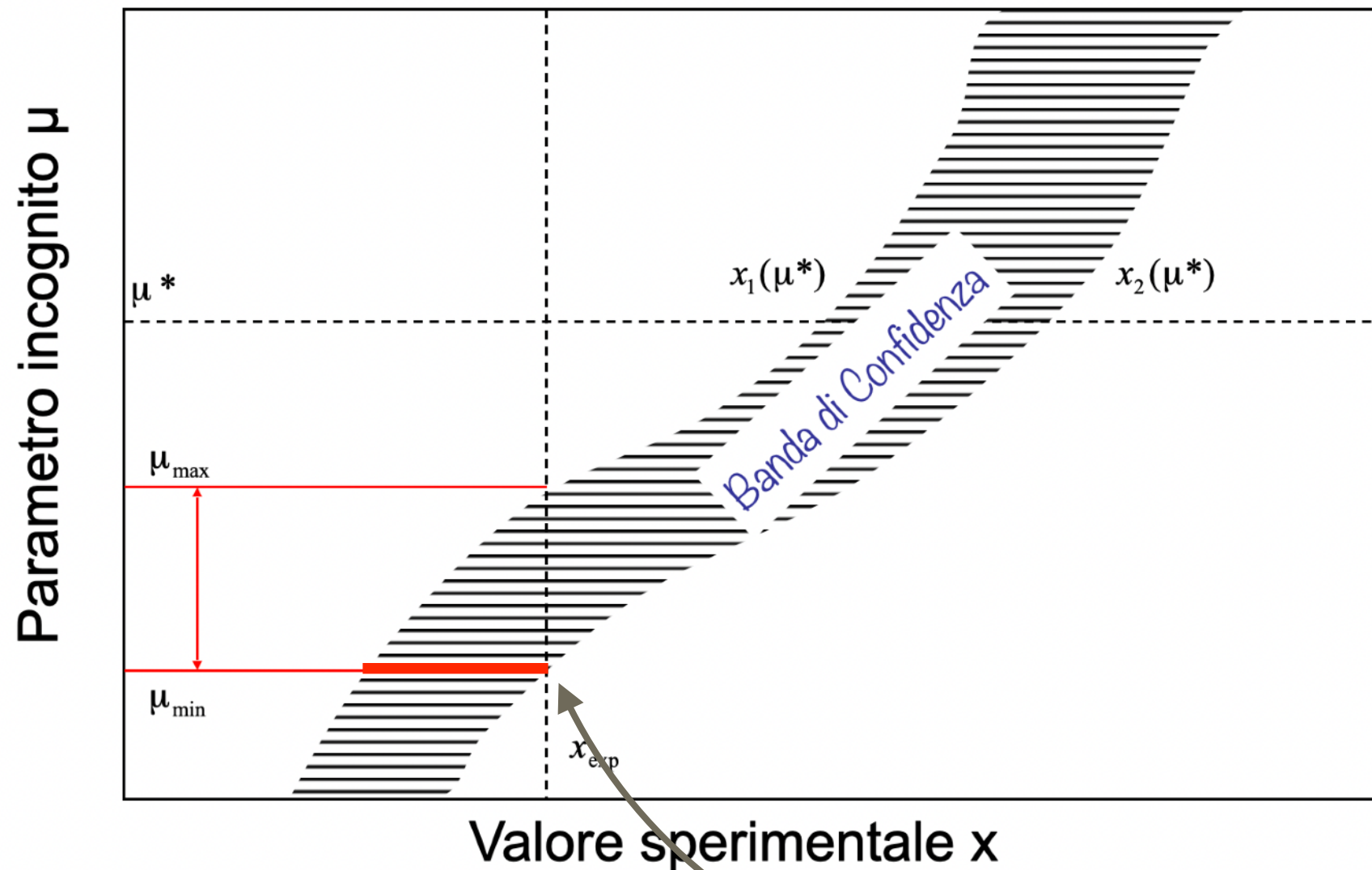
Per un
certo
valore di μ

$$\int_0^{x_1(\mu)} \frac{1}{\mu} e^{-\frac{x}{\mu}} dx = \frac{\alpha}{2} \Rightarrow 1 - e^{-\frac{x_1(\mu)}{\mu}} = \frac{\alpha}{2} \Rightarrow x_1(\mu) = -\mu \ln\left(\frac{2-\alpha}{2}\right)$$

$$\int_0^{x_2(\mu)} \frac{1}{\mu} e^{-\frac{x}{\mu}} dx = 1 - \frac{\alpha}{2} \Rightarrow 1 - e^{-\frac{x_2(\mu)}{\mu}} = 1 - \frac{\alpha}{2} \Rightarrow x_2(\mu) = -\mu \ln\left(\frac{\alpha}{2}\right)$$

Individuo gli estremi
dell'intervallo di
confidenza per x

INTERVALLO DI CONFIDENZA ALLA NEYMAN

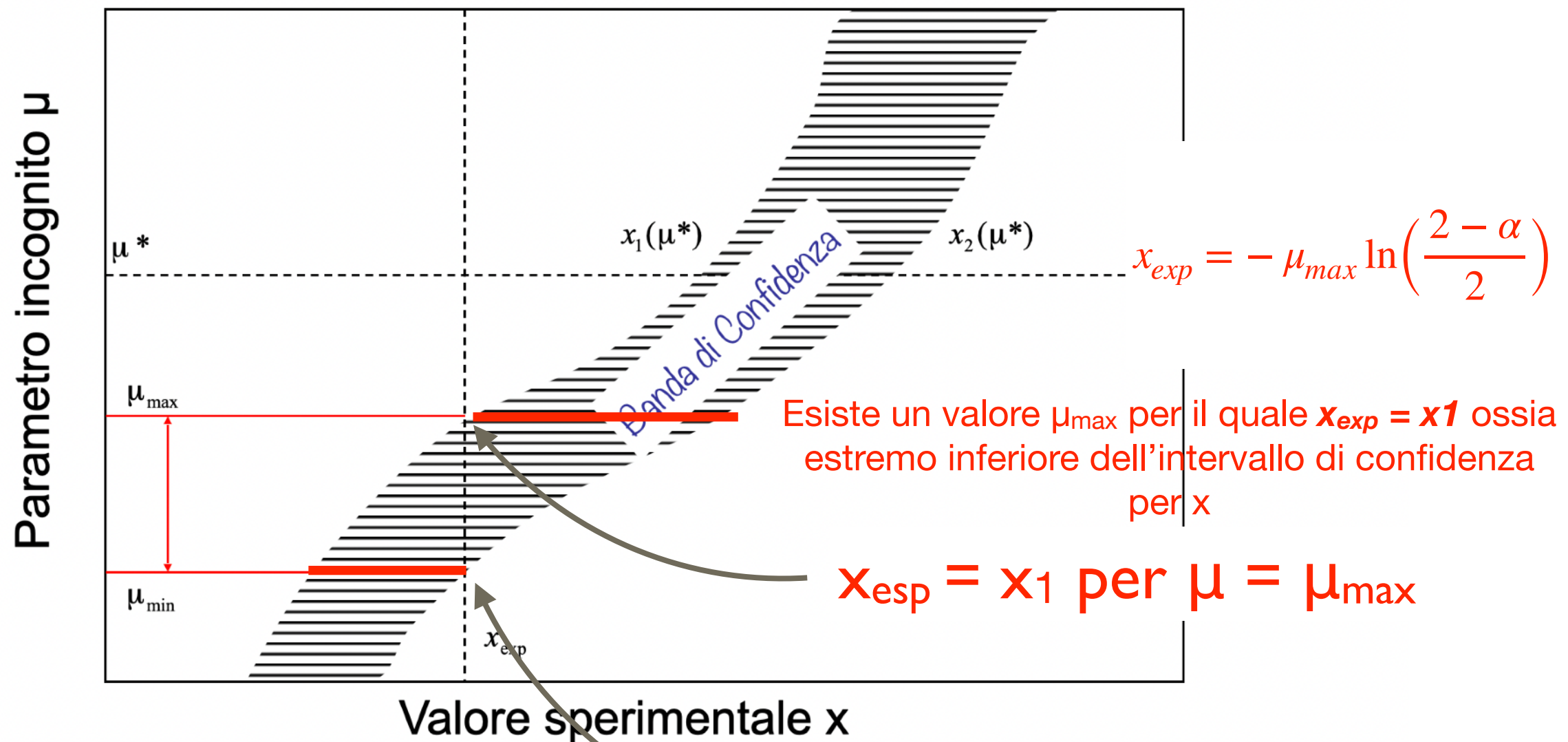


Esiste un valore μ_{min} per il quale $x_{exp} = x_2$
ossia estremo superiore dell'intervallo di
confidenza per x

$$x_{exp} = x_2 \text{ per } \mu = \mu_{min}$$

$$x_{exp} = -\mu_{min} \ln\left(\frac{\alpha}{2}\right)$$

INTERVALLO DI CONFIDENZA ALLA NEYMAN



Esiste un valore μ_{max} per il quale $x_{exp} = x_1$ ossia
estremo inferiore dell'intervallo di confidenza
per x

$$x_{exp} = x_1 \text{ per } \mu = \mu_{max}$$

Esiste un valore μ_{min} per il quale $x_{exp} = x_2$
ossia estremo superiore dell'intervallo di
confidenza per x

$$x_{exp} = x_2 \text{ per } \mu = \mu_{min}$$

$$x_{exp} = -\mu_{min} \ln\left(\frac{\alpha}{2}\right)$$

INTERVALLO DI CONFIDENZA ALLA NEYMAN

Oltre il paradigma gaussiano

Calcoliamo, invece, l'intervallo di confidenza con la tecnica di Neyman.

$$\int_0^{x_1(\mu)} \frac{1}{\mu} e^{-\frac{x}{\mu}} dx = \frac{\alpha}{2} \Rightarrow 1 - e^{-\frac{x_1(\mu)}{\mu}} = \frac{\alpha}{2} \Rightarrow x_1(\mu) = -\mu \ln\left(\frac{2-\alpha}{2}\right)$$

$$\int_0^{x_2(\mu)} \frac{1}{\mu} e^{-\frac{x}{\mu}} dx = 1 - \frac{\alpha}{2} \Rightarrow 1 - e^{-\frac{x_2(\mu)}{\mu}} = 1 - \frac{\alpha}{2} \Rightarrow x_2(\mu) = -\mu \ln\left(\frac{\alpha}{2}\right)$$

Esiste un valore $\mu_{\max}$ per il quale $x_{\exp} = x_1$ ossia estremo inferiore dell'intervallo di confidenza per x

Esiste un valore $\mu_{\min}$ per il quale $x_{\exp} = x_2$ ossia estremo superiore dell'intervallo di confidenza per x

$$x_{\exp} = -\mu_{\min} \ln\left(\frac{\alpha}{2}\right)$$

$$x_{\exp} = -\mu_{\max} \ln\left(\frac{2-\alpha}{2}\right)$$

Quindi

$$-\frac{x_{\exp}}{\ln\left(\frac{\alpha}{2}\right)} < \mu < -\frac{x_{\exp}}{\ln\left(\frac{2-\alpha}{2}\right)} \Rightarrow -\frac{x_{\exp}}{\ln(0.16)} < \mu < -\frac{x_{\exp}}{\ln(0.84)} \Rightarrow 0.55 x_{\exp} < \mu < 5.75 x_{\exp}$$

Dove si è posto $\alpha=1-0.68$. Si noti che l'intervallo ottenuto NON è simmetrico intorno a $x_{\exp}$ che continua ad essere il valore di aspettazione di μ

INTERVALLO DI CONFIDENZA ALLA NEYMAN

Oltre il paradigma gaussiano

- Calcoliamo l'intervallo di confidenza con la tecnica di Neyman

$$\int_0^{x_1(\mu)} \frac{1}{\mu} e^{-\frac{x}{\mu}} dx = \frac{\alpha}{2} \rightarrow -e^{-\frac{x_1(\mu)}{\mu}} + 1 = \frac{\alpha}{2} \rightarrow x_1(\mu) = -\mu \ln\left(\frac{2-\alpha}{2}\right)$$

$$\int_0^{x_2(\mu)} \frac{1}{\mu} e^{-\frac{x}{\mu}} dx = 1 - \frac{\alpha}{2} \rightarrow -e^{-\frac{x_2(\mu)}{\mu}} + 1 = 1 - \frac{\alpha}{2} \rightarrow x_2(\mu) = -\mu \ln\left(\frac{\alpha}{2}\right)$$

- Assumendo che il mio risultato sperimentale sia x_{exp} determino l'intervallo su μ ponendo

$$x_{exp} = -\mu_2 \ln\left(\frac{2-\alpha}{2}\right) \qquad x_{exp} = -\mu_1 \ln\left(\frac{\alpha}{2}\right)$$

$$\mu_1 = -\frac{x_{exp}}{\ln\left(\frac{\alpha}{2}\right)} < \mu < \mu_2 = -\frac{x_{exp}}{\ln\left(\frac{2-\alpha}{2}\right)}$$

- Ponendo $\alpha = 1-0.68$ si ha

$$-\frac{x_{exp}}{\ln(0.16)} < \mu < = -\frac{x_{exp}}{\ln(0.84)}$$

NOTA: intervallo non simmetrico
tornando a x_{exp} che è sempre il valore
di aspettazione per μ

INTERVALLO DI CONFIDENZA ALLA NEYMAN

Oltre il paradigma gaussiano

$$-\frac{x_{\text{exp}}}{\ln(0.16)} < \mu < -\frac{x_{\text{exp}}}{\ln(0.84)}$$
$$x_{\text{exp}} - x_{\text{exp}} < \mu < x_{\text{exp}} + x_{\text{exp}}$$

Come valutare quale delle due scelte è più corretta?

Il significato frequentista di intervallo di confidenza è il seguente:

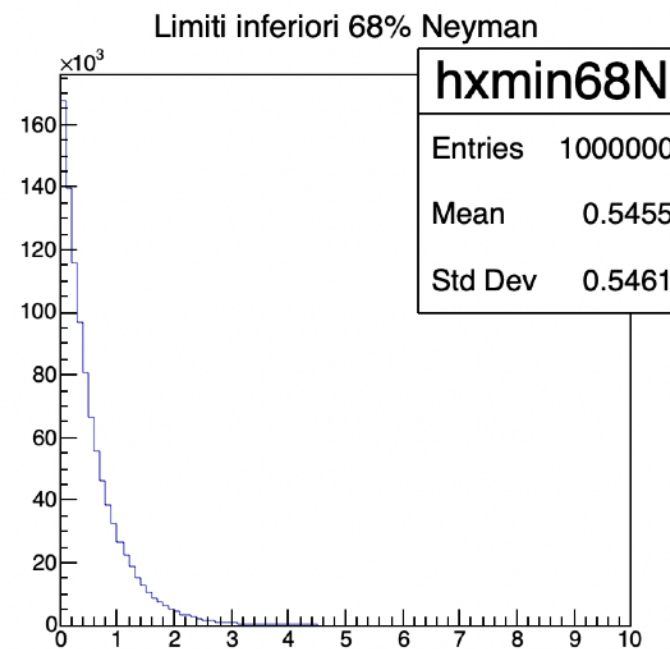
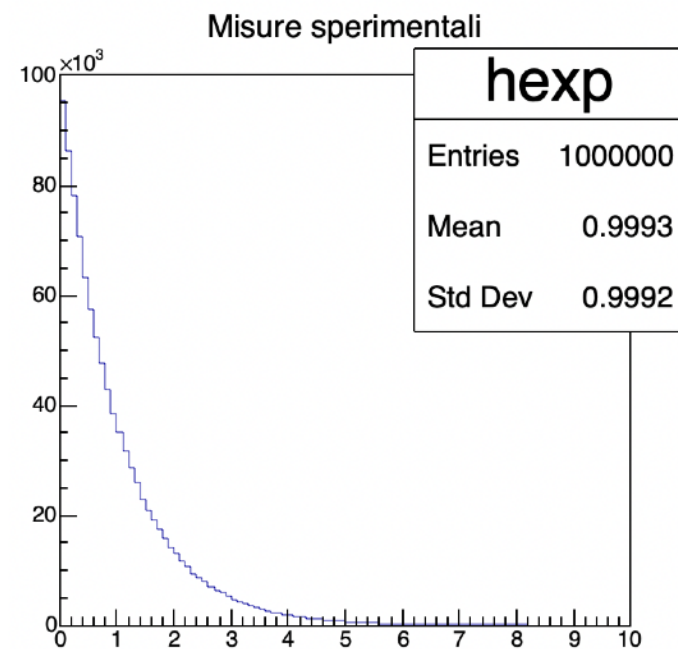
Se ripetessi l'esperimento per un numero N grande di volte allora in una frazione $1 - \alpha$ dei casi (68%) l'intervallo determinato conterrebbe il valore vero

Possiamo verificare quale dei due intervalli rispetta la definizione di intervallo di confidenza.

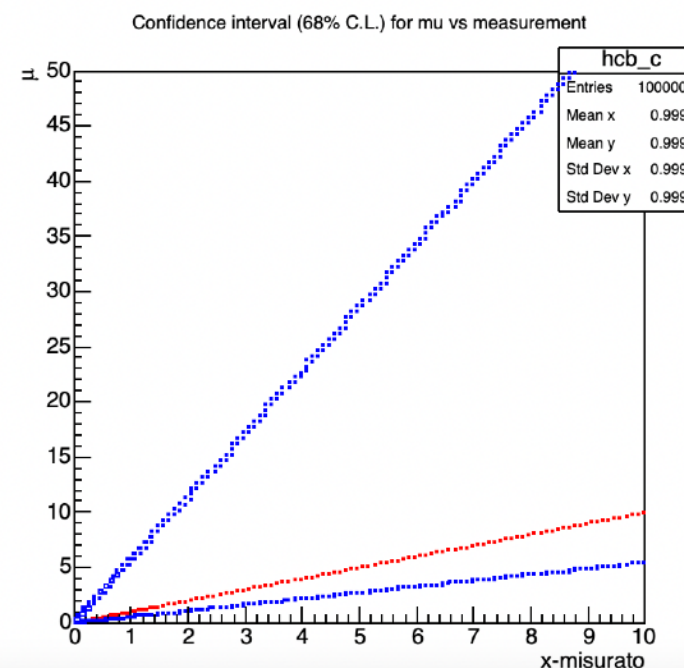
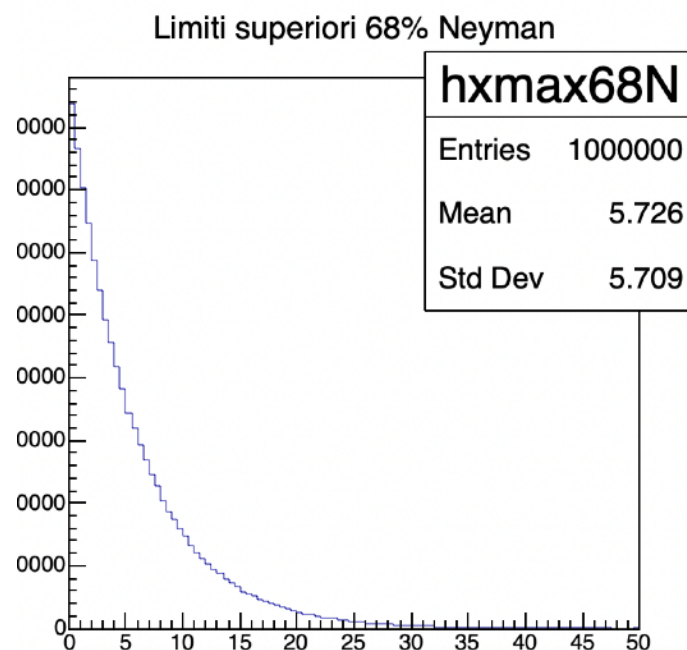
... codice... **Neyman.C**

INTERVALLO DI CONFIDENZA ALLA NEYMAN

Oltre il paradigma gaussiano



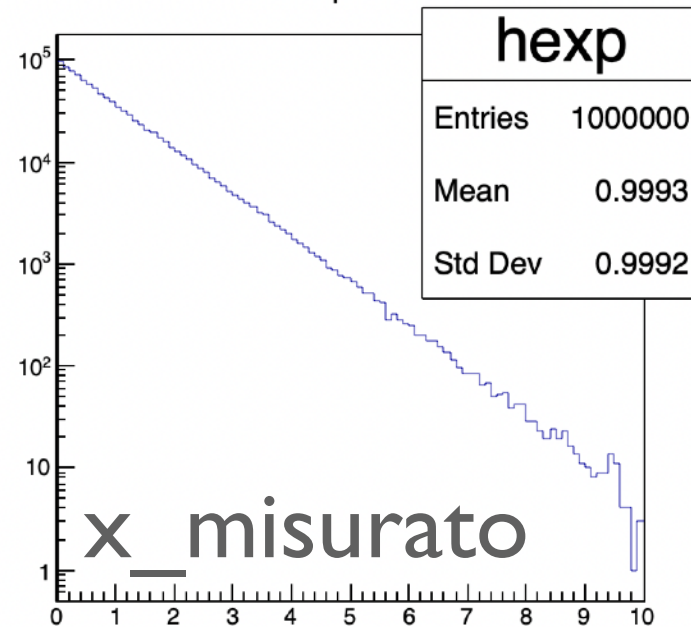
Neyman.C



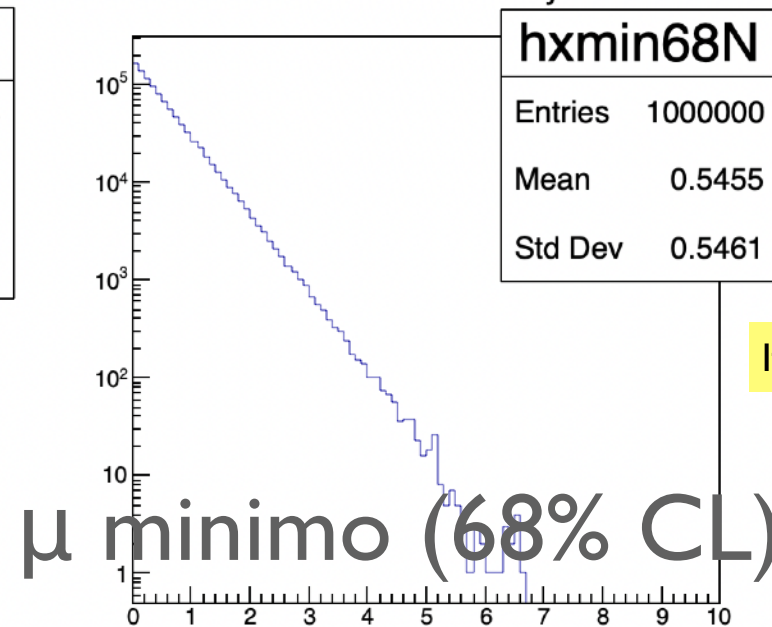
INTERVALLO DI CONFIDENZA ALLA NEYMAN

Oltre il paradigma gaussiano

Misure sperimentali



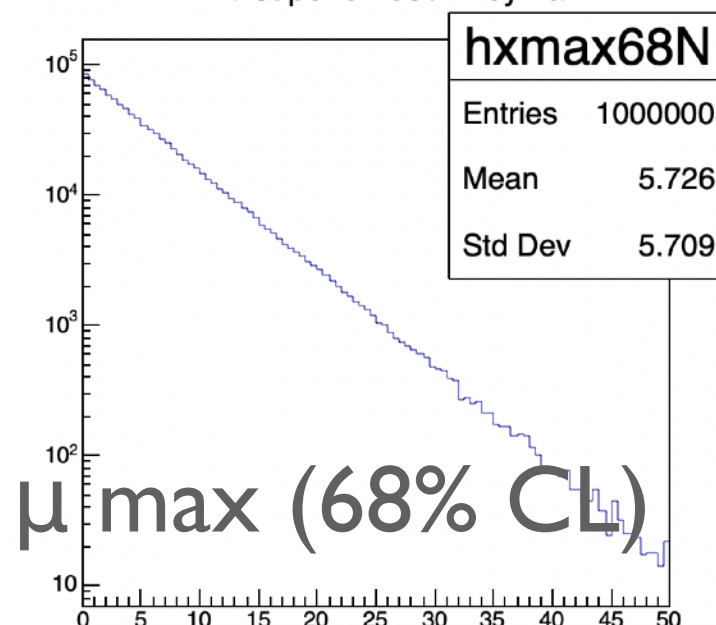
Limiti inferiori 68% Neyman



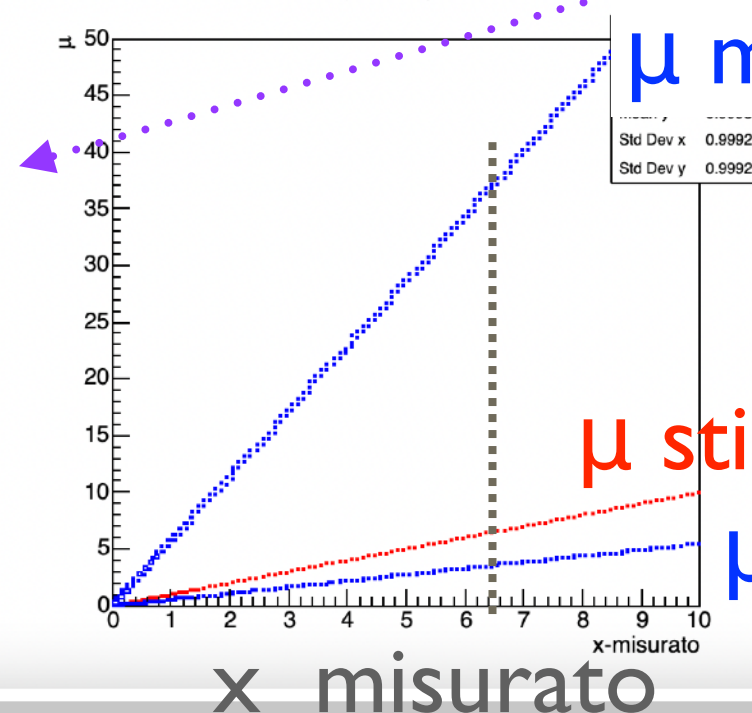
In log scale

If $x_{\text{exp}} = \mu = 1$, 68% CL (Neyman) = **[0.545678, 5.73548]**

Limiti superiori 68% Neyman

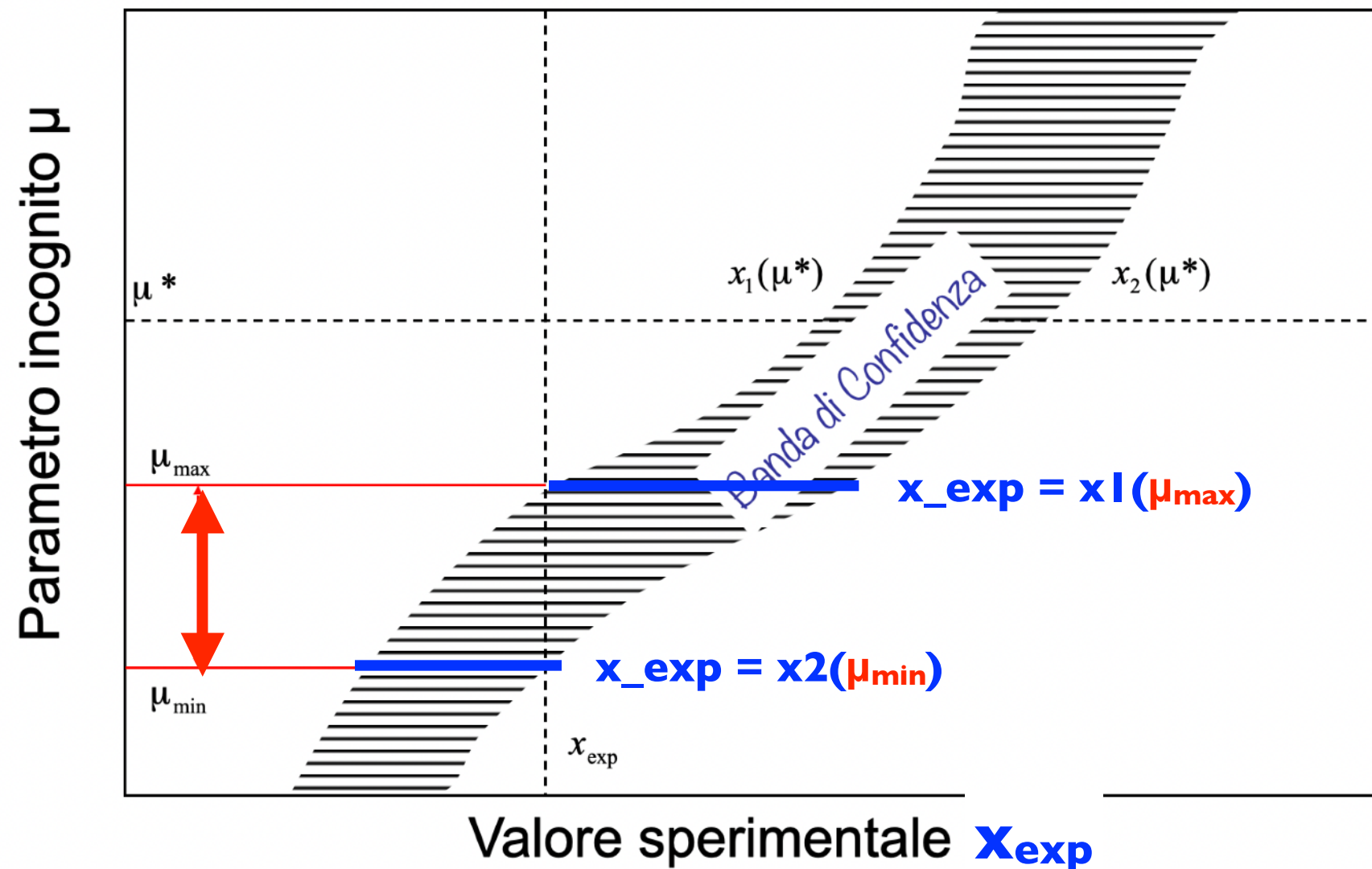


Confidence interval (68% C.L.) for μ vs measurement



NOTA:
intervallo
fortemente
asimmetrico

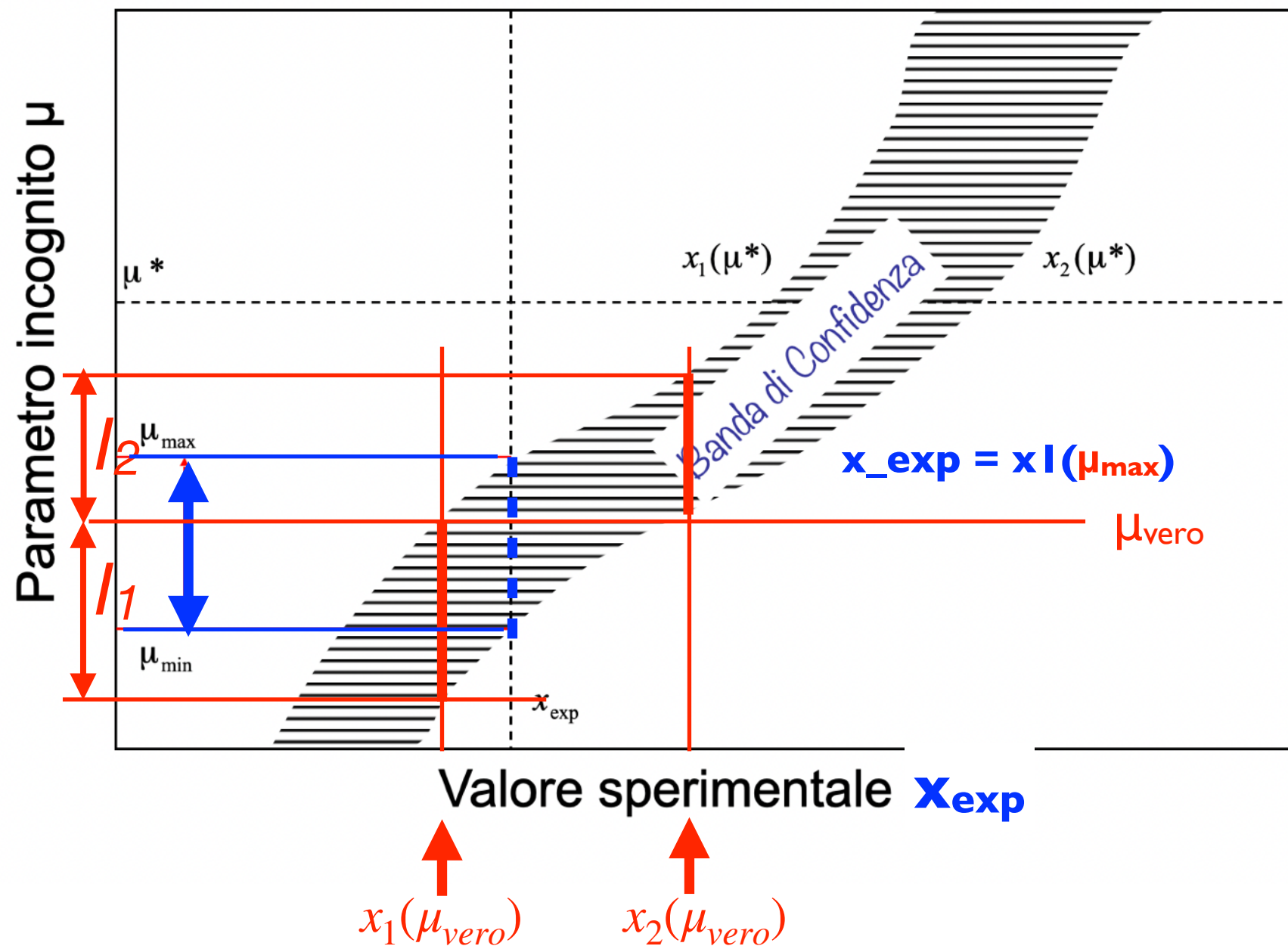
INTERVALLO DI CONFIDENZA ALLA NEYMAN



- 5) Dato un valore sperimentale misurato $x_{\exp}$ si traccia la linea verticale passante per questo valore e si identificano i due estremi della banda di confidenza che identificano l'intervallo di confidenza con fiducia $1-\alpha$, **per esempio 90%**

Osserviamo che la costruzione dell'intervallo di confidenza con il metodo di Neyman è intrinsecamente coerente con l'approccio frequentista, **infatti ...**

INTERVALLO DI CONFIDENZA ALLA NEYMAN



NOTA:

Per μ_{vero} ho ottenuto una misura X_{exp} che con 90% di probabilità si trova tra $x_1(\mu_{vero})$ e $x_2(\mu_{vero})$

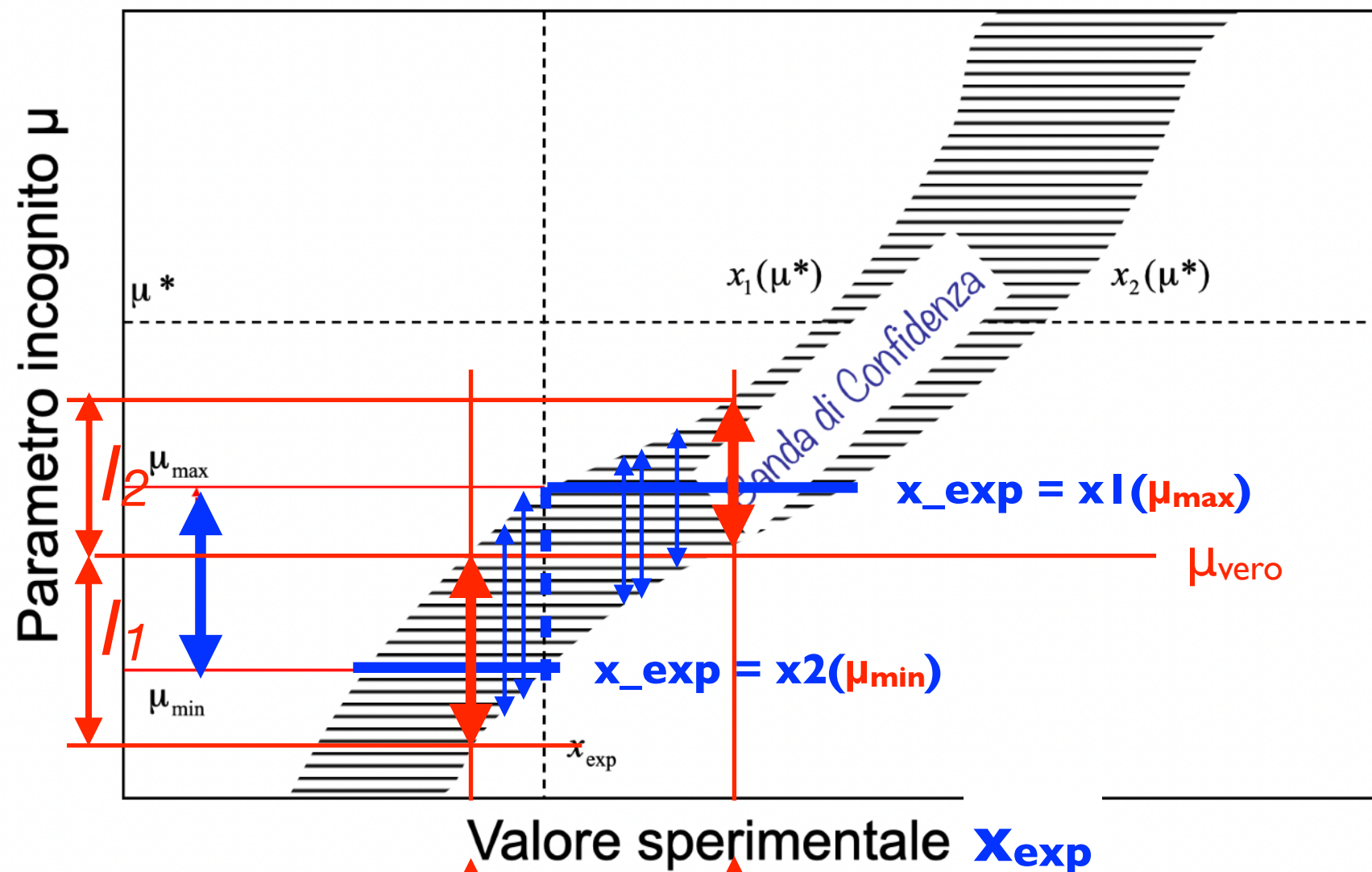
Sulla base di X_{exp} determino l'intervallo $I = [\mu_{min}, \mu_{max}]$ che dichiaro intervallo di confidenza al 90% per μ

Se $x_{exp} = x_1(\mu_{vero})$ allora $I = I_1$, μ_{vero} è al bordo di I (estremo superiore)

Se $x_{exp} = x_2(\mu_{vero})$ allora $I = I_2$, μ_{vero} è al bordo di I (estremo inferiore)

Per ogni $x_1(\mu_{vero}) < x_{exp} < x_2(\mu_{vero})$ μ_{vero} è interno a I

INTERVALLO DI CONFIDENZA ALLA NEYMAN



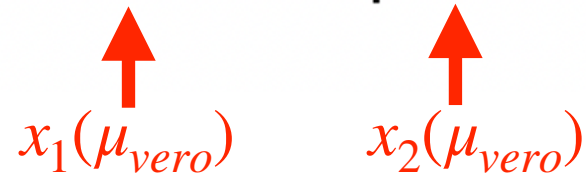
NOTA:

Per μ_{vero} ho ottenuto una misura $X_{\exp}$ che con 90% di probabilità si trova tra $x_1(\mu_{\text{vero}})$ e $x_2(\mu_{\text{vero}})$

Sulla base di $X_{\exp}$ determino l'intervallo $I = [\mu_{\min}, \mu_{\max}]$ che dichiaro intervallo di confidenza al 90% per μ

Per ogni $x_1(\mu_{\text{vero}}) < x_{\exp} < x_2(\mu_{\text{vero}})$, μ_{vero} è interno a I

La condizione $x_1(\mu_{\text{vero}}) < x_{\exp} < x_2(\mu_{\text{vero}})$ si verifica con frequenza 90% di conseguenza, la condizione μ_{vero} è interno a I si verifica con frequenza 90%



cui μ_{vero} sarà contenuto

Sulla base di **X_{exp}**
determino l'intervallo
 $[\mu_{min}, \mu_{max}]$
che dichiaro intervallo
di confidenza al 90%
per μ **perché**

Approccio frequentista

INTERVALLO DI CONFIDENZA ALLA NEYMAN

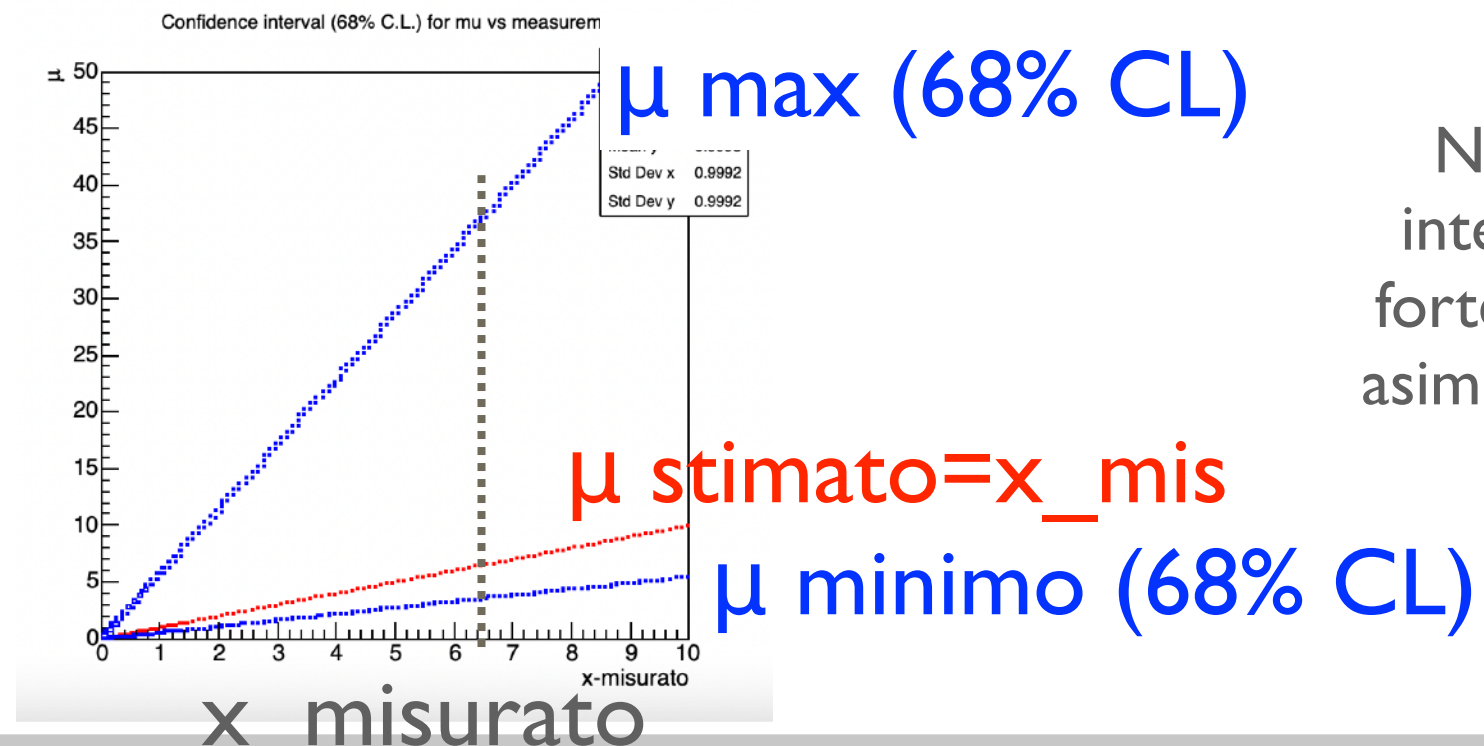
Nel nostro esercizio

C.L. 0.68 def. Standard: frequenza di μ_{vero} in intervallo 0.606061

C.L. 0.68 def. Neyman: frequenza di μ_{vero} in intervallo 0.679564

C.L. 0.9544 def. Standard: frequenza di μ_{vero} in intervallo 0.715712

C.L. 0.9544 def. Neyman: frequenza di μ_{vero} in intervallo 0.954477



NOTA:
intervallo
fortemente
asimmetrico

INTERVALLO DI CONFIDENZA ALLA NEYMAN

Il metodo di Neyman è intrinsecamente basato sull'approccio **frequentista**

Dato un risultato sperimentale X_{exp}
Si costruisce un intervallo di confidenza al CL desiderato $(1-\alpha)$ in modo tale che, ripetuto l'esperimento numerose volte, l'intervallo determinato, diverso ogni volta, contenga il valore effettivo di μ in una frazione degli esperimenti pari al CL.

Osservate che non si parla mai di probabilità del parametro μ .

Tornando al nostro esempio di pdf gaussiana con media incognita μ e con $\sigma=1$ per l'osservabile x , $f(x, \mu)$, osserviamo che è corretto dire che essa rappresenta una densità di probabilità condizionata $f(x | \mu)$, perché è la p.d.f. della variabile aleatoria x dato il valore μ del parametro.

Se μ cambia, cambia anche la f .



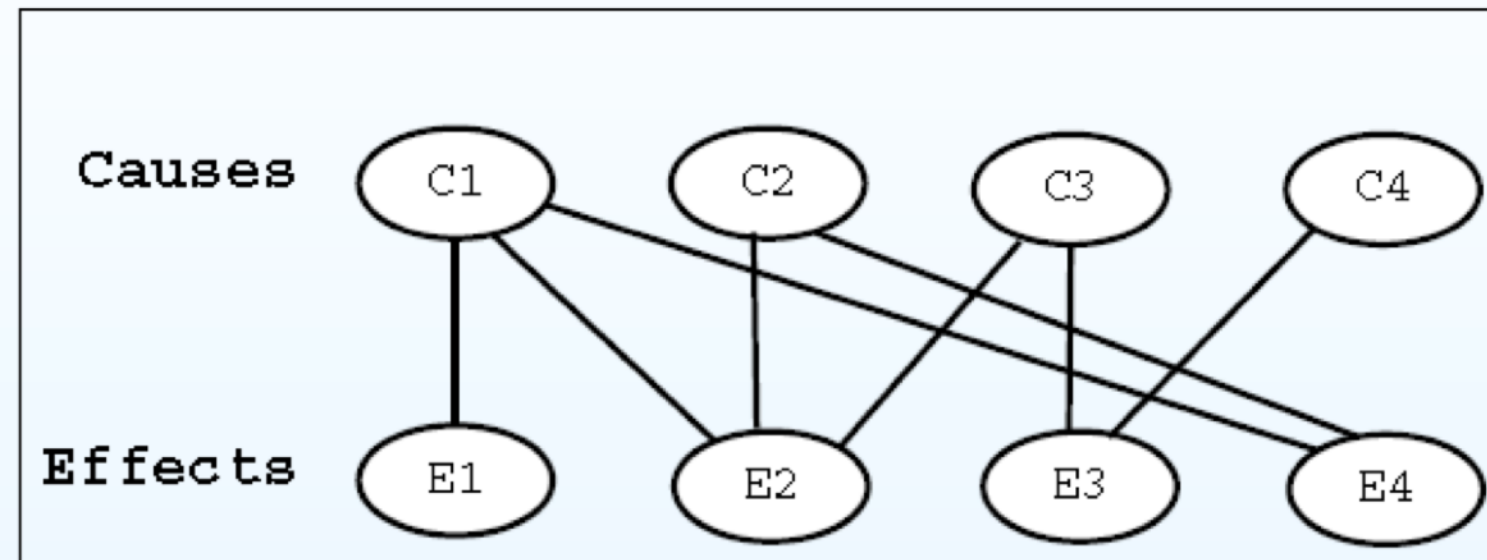
Jerzy Neyman (16 aprile 1894 – 5 agosto 1981)
statistico polacco.

$$f(x, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

RICORDIAMO IL TEOREMA DI BAYES

Analizziamo la nostra misura in termini di causa ed effetto

Our original problem:



Our conditional view of probabilistic causation

$$P(E_i | C_j)$$

Our conditional view of probabilistic inference

$$P(C_j | E_i)$$

TEOREMA DI BAYES

$$P(B_k | A) = \frac{P(A | B_k)P(B_k)}{\sum_{k=1}^n P(A | B_k)P(B_k)}$$

Vediamo come riformulare il teorema di Bayes nel caso di p.d.f. di variabili continue. Consideriamo come primo caso la situazione in cui ho una misura x e devo inferire un parametro μ entrambi variabili reali definite in $(-\infty, +\infty)$.

Non ho più le n famiglie di eventi B_k . Il mio spazio campione è diviso in maniera continua dal valore del parametro μ che svolge il ruolo di B_k .

La sommatoria va quindi sostituita da un integrale.

$$p(\mu | x) = \frac{p(x | \mu)\pi(\mu)}{\int_{-\infty}^{+\infty} p(x | \mu^*)\pi(\mu^*)d\mu^*}$$

Dove con $p(\mu|x)$ identifico la p.d.f. che segue la "causa" μ una volta osservato "l'effetto" x . Mentre le $p(x | \mu)$ sono le p.d.f. che segue la variabile x una volta noto μ . La funzione $\pi(\mu)$ è la p.d.f. *a priori* che seguirebbe il parametro incognito μ . In molti casi questa $\pi(\mu)$ viene assunta uniforme, io a priori non so qual'è il possibile valore di μ per cui assumo tutti i valori equiprobabili.

TEOREMA DI BAYES E INTERVALLI DI CONFIDENZA

- Se abbiamo n misure $\vec{x} = \{x_1, x_2, \dots, x_n\}$ indipendenti l'una dall'altra e tutte relative allo stesso parametro μ si ha

- $$p(\vec{x} | \mu) = \prod_{i=1}^n p(x_i | \mu) = \mathcal{L}(\vec{x} | \mu)$$

- cioe' la probabilità del set di misure è uguale alla likelihood.

- Quindi la probabilità a posteriori di μ dato il risultato $\vec{x}$ dell'esperimento e'

- $$p(\mu | \vec{x}) = \frac{\mathcal{L}(\vec{x} | \mu) \pi(\mu)}{\int_{-\infty}^{\infty} d\mu^* \mathcal{L}(\vec{x} | \mu^*) \pi(\mu^*)}$$

- Nell'approccio Bayesiano si calcola esplicitamente la probabilità a posteriori del parametro μ , pertanto la definizione di intervallo di confidenza $I = [\mu_1, \mu_2]$ al CL desiderato, $1-\alpha$, segue in modo concettualmente ovvio***

- $$1 - \alpha = \int_{\mu_1}^{\mu_2} d\mu p(\mu | \vec{x})$$

INTERVALLI DI CONFIDENZA BAYESIANI

Identificare un intervallo di credibilità con il metodo Bayessiano equivale quindi a risolvere rispetto a μ_1 e μ_2 l'equazione

$$1 - \alpha = CL = \frac{\int_{\mu_1}^{\mu_2} d\mu \mathcal{L}(\vec{x} | \mu) \pi(\mu)}{\int_{-\infty}^{\infty} d\mu \mathcal{L}(\vec{x} | \mu) \pi(\mu)}$$

Si noti come il risultato dipende dalla scelta della probabilità a priori che segue il parametro μ . Diverse scelte di questa p.d.f. possono portare a risultati diversi. La p.d.f. della μ gioca il ruolo di “qual’è la mia conoscenza del problema”.

Come detto spesso si assumono distribuzioni uniformi o quasi uniformi (questo per problemi di normalizzazione del calcolo)

INTERVALLI DI CONFIDENZA BAYESIANI

Proviamo a valutare che p.d.f. a posteriori ha la stima della media di una gaussiana a partire da una misura sperimentale.

So che i miei dati si distribuiscono in maniera Gaussiana ma non conosco la media della distribuzione. Ipotizzo nota la sigma = 1.

Assumo di fare una sola misura e di voler stimare un intervallo di confidenza per la media.

$$p(x | \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

$$p(\mu | x) = \frac{p(x | \mu)}{\int_{-\infty}^{+\infty} p(x | \mu) \pi(\mu) d\mu} \pi(\mu)$$

Chi è nel nostro caso $\pi(\mu)$? L'ipotesi meno informativa che posso avanzare è che possa assumere un qualsiasi valore tra meno e più infinito (distribuzione uniforme). Per cui in questo caso la distribuzione di probabilità a posteriori diventa:

$$p(\mu | x) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}}{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu')^2}{2}} d\mu'} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

INTERVALLI DI CONFIDENZA BAYESIANI

$$p(\mu | x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

Si noti che qui la variabile aleatoria è μ e non x che è il dato noto!
Dato che la distribuzione della probabilità a posteriori che si ottiene è sempre una Gaussiana l'intervallo di credibilità che si ottiene è quello tradizionale.

Quindi, al livello di confidenza $1 - \alpha = CL$, si ha

$$1 - \alpha = \int_{\mu_1}^{\mu_2} d\mu \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} \rightarrow x - N(\alpha) < \mu < x + N(\alpha)$$

Qualsiasi approccio statistico quando la distribuzione della popolazione segue un andamento gaussiano porta alla stessa definizione di intervallo di credibilità/confidenza/fiducia.

BAYESIAN VS FREQUENTIST

Discussione sulle criticità dell'approccio Bayessiano

- 1) La probabilità a priori non è definita univocamente
- 2) Non è sempre invariante per trasformazione dei parametri
- 3) È spesso estremamente complesso anche numericamente ottenere il risultato
- 4) Intuitivamente può essere complesso da accettare il concetto di soggettività nella definizione della probabilità intrinseco nella necessità di definire una probabilità a priori.

Lo scontro tra l'approccio frequentista e l'approccio Bayessiano è in alcuni casi molto acceso e tutt'ora in corso.

Vedi per esempio: G. D'Agostini "*Probably a discovery: Bad mathematics means rough scientific communication*" (2011) (<http://arxiv.org/abs/1112.3620>)

Uno suggerimento può essere quello di pubblicare un risultato utilizzando entrambe le tecniche in modo da permettere al lettore di considerare la soluzione che a lui sembra più corretta (... troppo Bayessiano?)

ESERCIZI

ESEMPIO

Neyman_esperimento.C

- Supponiamo di avere una legge fisica che predice che alcuni processi si verifichino con una abbondanza nel tempo descritta da:
 - $N(t) \sim \exp(-t/\tau)$
 - Vogliamo determinare τ sulla base della misura della frazione degli eventi F_{exp} osservati nell'intervallo Δt tra $t_1=10\text{ns}$ e $t_2=60\text{ns}$
 - Immaginiamo che $\tau=220\text{ ns}$ (valore vero)
 - $$\Rightarrow F(\Delta t) = \frac{\int_{t_1}^{t_2} e^{-t/\tau} dt}{\int_0^{\infty} e^{-t/\tau} dt} = \frac{-\tau e^{-t_2/\tau} + \tau e^{-t_1/\tau}}{\tau} = -e^{-60/220} + e^{-10/220} =$$

$$-0.7613 + 0.9556 = 0.1942$$
- Esperimento: Osserviamo 1850 (valore vero 1942) eventi tra t_1 e t_2 dei 10000 complessivamente prodotti
 - Determiniamo l'intervallo di confidenza al 90% per τ , ossia per $N(\Delta t)$

ESEMPIO

Neyman_esperimento.C

- Consideriamo tau [tra 50 e 500 ns]
- Per ogni valore calcoliamo $F(\Delta t) = -e^{-60/\tau} + e^{-10/\tau}$
 - $N(\Delta t) = 10000 * F(\Delta t)$; x fluttuera' come una poissoniana con media $N(\Delta t)$
 - Calcoliamo x1 e x2 che delimitano l'intervallo di confidenza al 90 % se la media di una poissoniana e' $N(\Delta t)$
 - Costruiamo la banda di confidenza
- Per il risultato del nostro esperimento, 1850, cerchiamo gli estremi τ_1 e τ_2 che corrispondono agli estremi della banda
- Verifichiamo che ripetendo la procedura n volte (estraiemo N_{exp} da una poissoniana con media = valore vero 1942) il valore vero tau =220 ns, ossia $N(\Delta t) = 1942$, si trova nell'intervallo il 90% delle volte.

ESEMPIO

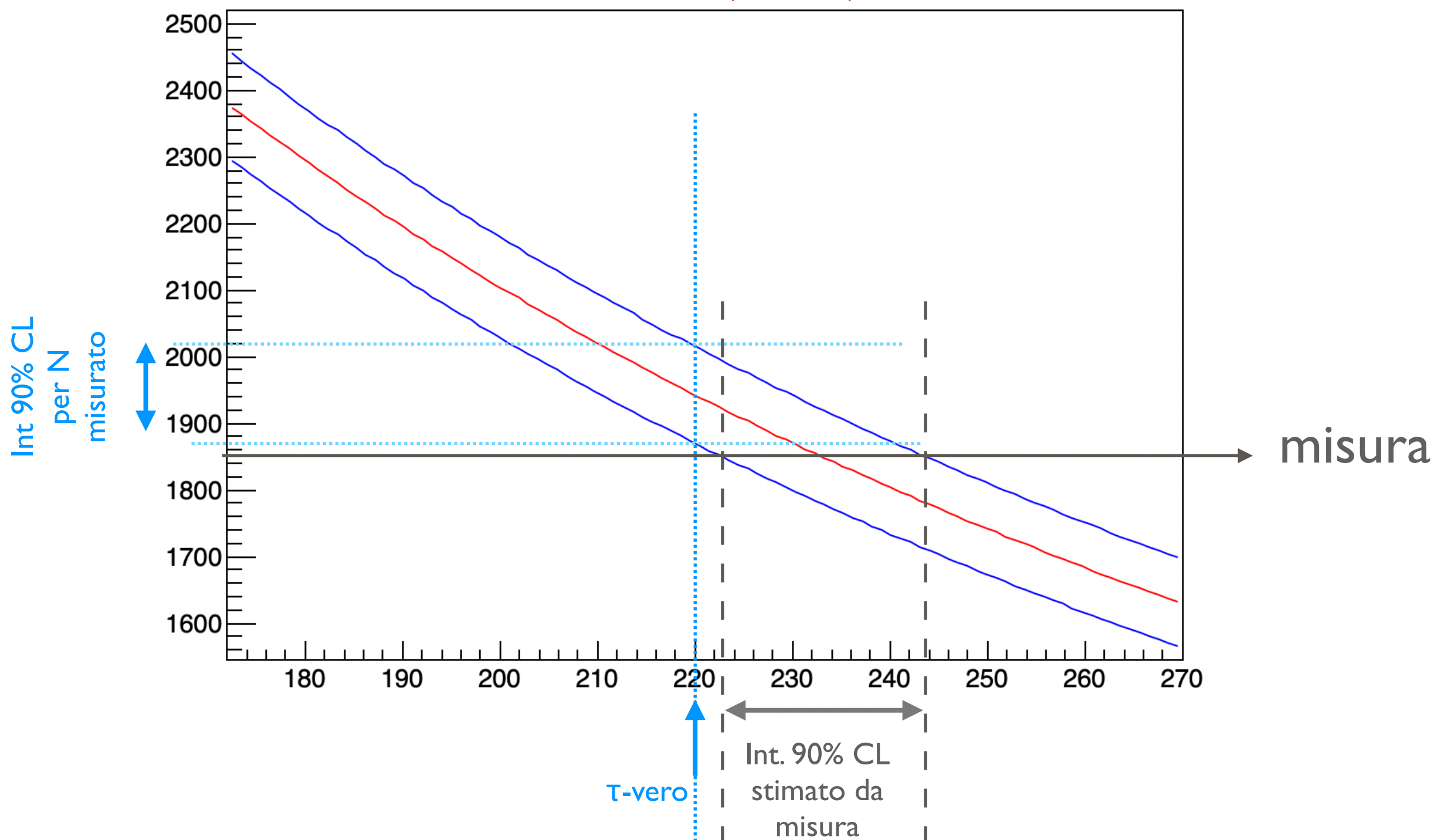
Neyman_esperimento.C



$\tau = 220$ Frazioni di eventi tra $t_1=10$ e $t_2=60 \rightarrow 0.194263$

Eventi attesi tra 10 e 60 s (per τ vero = 220 s) = 1942.63

min, max (90% CL) $\rightarrow$ 1870.12 2015.13



ESEMPIO

Neyman_esperimento.C

$\tau = 220$ Frazioni di eventi tra $t_1=10$ e $t_2=60 \rightarrow 0.194263$

Eventi attesi tra 10 e 60 s (per τ vero = 220 s) = 1942.63

min, max (90% CL) $\rightarrow$ 1870.12 2015.13

