METODI STATISTICI E COMPUTAZIONALI

Stefania Spagnolo

Dipartimento di Matematica e Fisica, Univ. del Salento

LEZIONE 7

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

Dispense L. Angelini Univesità di Bari P. 1-27

H.J.C. Berendens Data and Error Analysis p.53-58 e p.135-140

P.R. Bevington Data Reduction and Error Analysis for the Physical Sciences p. 56-65

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

Definizione di **variabile aleatoria**

$$X : \mathrm{B} \to \Re$$

Variabili **aleatorie discrete e continue finite o infinite**

$$X(w_i) = x_i \quad w_i \in \mathrm{B}, x_i \in \Re$$

Processi Discreti
La funzione che ad ogni valore dalla variabile aleatoria associa una probabilità è detta **distribuzione di probabilità**

$$f(x_i) = P(X = x_i) \quad i = 1,2,\ldots$$

Si definisce **funzione di ripartizione o funziona cumulativa** la funzione:

$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

Valgono le seguenti proprietà:

$$f(x_i) \ge 0 \quad \forall i$$

$$\sum_{x_i} f(x_i) = 1$$

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

Processi Continui

Se **X è una variabile aleatoria continua**, la probabilità che assuma uno specifico valore è nulla. In questo caso ha senso parlare di probabilità che la variabile aleatoria assuma valori compresi tra *a* e *b* con *a* ≤*b*

Se presuppone l'esistenza di una funzione detta **densità di probabilità** tale che:

$$f(x) \geq 0 \quad \forall x \in \Re$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Si definisce quindi probabilità che X sia compresa tra *a* e *b* come:

$$P(a < x < b) = \int_{b}^{a} f(x)dx$$

Anche nel caso della variabili aleatorie continue si definisce una **funzione di ripartizione o cumulativa** come:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

Processi Continui

Nel continuo ha significato calcolare solo la probabilità che una variabile aleatoria assuma valori all'interno di un intervallo. Questo comporta che la funzione f(x) **NON** rappresenta la probabilità, infatti

esistono x per cui $f(x)\neq 0$ ma
$P(X=x)=0$ per ogni x appartenente ai reali

Valgono quindi le seguenti identità:

$$P(X \leq a) = P(X < a)$$

$$P(X \geq a) = P(X > a)$$

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

Data una variabile aleatoria alla sua distribuzione di probabilità o densità di probabilità sono associati alcuni **parametri**.

Caso discreto:

Valor Medio o **valore di aspettazione**

$$\mu = E[X] = \sum_i x_i f(x_i)$$

Nel caso particolare di $f(x_i)$ tutte uguali (eventi equiprobabili) la precedente si trasforma nella media aritmetica

Varianza

$$\sigma^2 = \mathrm{var}(X) = E[(X-\mu)^2] = \sum_i (x_i - \mu)^2 f(x_i)$$

Deviazione standard o **scarto quadratico medio**

$$\sigma = \sqrt{\mathrm{var}(X)} = \sqrt{\sum_i (x_i - \mu)^2 f(x_i)}$$

<u>**Caso continuo:**</u>

Valor Medio o **valore di aspettazione**

$$\mu = E[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

Varianza

$$\sigma^2 = \mathrm{var}(X) = E[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx$$

Deviazione standard o **scarto quadratico medio**

$$\sigma = \sqrt{\mathrm{var}(X)} = \sqrt{\int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx}$$

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

Momenti di una distribuzione

I parametri che abbiamo trattato in precedenza (media, varianza, etc.) che caratterizzano una distribuzione (densità) di probabilità sono un caso particolare di una classe di parametri che danno informazioni sulle peculiarità di una distribuzione (densità) di probabilità.

In generale si dicono **momenti di ordine k rispetto al punto x_0 di una distribuzione** le grandezze:

$$M_k = E[(X - X_0)^k]$$

In statistica i momenti utilizzati sono i **momenti semplici**, cioè intorno allo 0 e i **momenti centrali**, cioè intorno alla media

$$\mu_k = E[X^k]$$

$$m_k = E[(X - \mu)^k]$$

Variabili Aleatorie e Distribuzioni di Probabilità

Alcuni già li conosciamo.
In particolare il momento di ordine 0 coincide con la normalizzazione della distribuzione (densità) di probabilità.

$$m_0 = \int_{-\infty}^{-\infty} (x-\mu)^0 f(x)dx = \int_{-\infty}^{-\infty} x^0 f(x)dx = \mu_0 = 1$$

Il momento semplice di ordine 1 è la media

$$\mu_1 = \int_{-\infty}^{-\infty} x f(x)dx = \mu$$

Il momento centrato di ordine 2 è la varianza

$$m_2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx = \mathrm{var}(x)$$

Notiamo inoltre che il momento centrato di ordine 1 è nullo

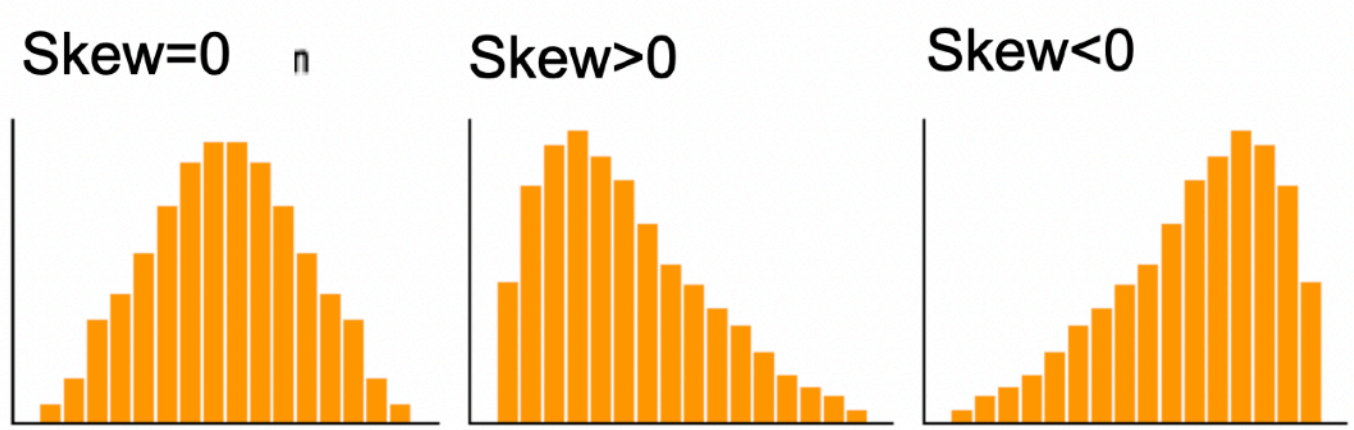
$$m_1 = \int_{-\infty}^{+\infty} (x-\mu) f(x)dx = 0$$

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

Anche i momenti di ordine superiore possono essere utili per definire grandezze utili a caratterizzare una distribuzione.
Si definisce **skewness** (asimmetria) di una distribuzione la grandezza:

$$Skew = \frac{m_3}{(m_2)^{3/2}} = \frac{E[(X-\mu)^3]}{E[(X-\mu)^2]^{3/2}}$$

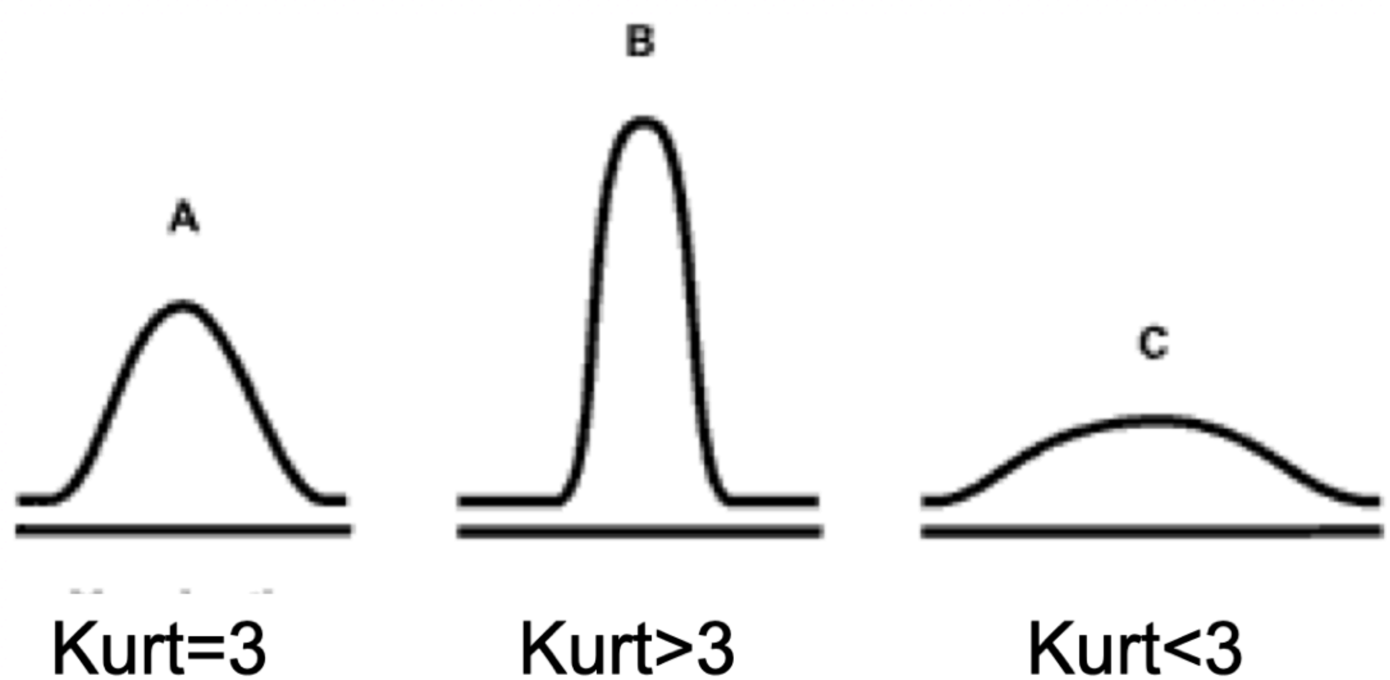
Questo momento tiene conto dell'asimmetria di una distribuzione.
Una distribuzione perfettamente simmetrica (normale) ha skewness=0.
Il denominatore ha solo lo scopo di rendere la grandezza adimensionale.



Si definisce **curtosi** di una distribuzione la grandezza:

$$kurt = \frac{m_4}{(m_2)^2} = \frac{E[(X-\mu)^4]}{E[(X-\mu)^2]^2}$$

La curtosi indica quanto la distribuzione è piccata.
Una distribuzione normale ha curtosi 3.



Altri parametri utili

Altri parametri che in alcune circostanze vengono usati per descrivere una distribuzione di probabilità sono:

La **mediana:** definita come il punto che divide a metà la distribuzione (densità) di probabilità. Si noti che la mediana coincide con la media solo nel caso di distribuzioni perfettamente simmetriche. La mediana è utilizzata in casi in cui la distribuzione degli eventi presenta lunghe code che alterano il valore della media.

$$\int_{-\infty}^{mediana} f(x)dx = 0.5$$

La **moda**: definita come il valore più probabile della distribuzione (densità). A differenza della media e della mediana una distribuzione può avere più di una moda o essere tale per cui non sia possibile individuare la moda.

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

Oltre alla mediana che divide in due una distribuzione si possono definire altri **indici di posizione** che dividono la distribuzione in parti uguali. Tali indici sono detti **quantili** o **percentili.**
Per esempio il percentile $\mathbf{P_{0.25}}$ identifica il punto che raccoglie il primo 25% della distribuzione.

$$\int_{-\infty}^{P_{0.25}} f(x)dx = 0.25$$

I percentili non sono facilmente calcolabili analiticamente e si ricorre a determinazione numerica o a valori tabulati per specifiche distribuzioni di probabilità.

2 O PIU' VARIABILI ALEATORIE

Probabilita' congiunta
Probabilita' marginale
Probabilita' condizionata
Variabili indipendenti

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ CONGIUNTA

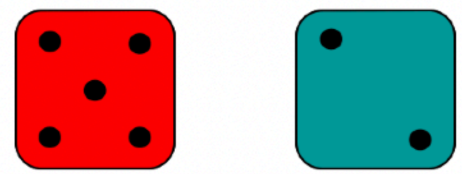
Dato uno spazio campione S, comunque esso sia fatto (finito o infinito, discreto o continuo), è possibile definire più variabili aleatorie che operano sullo stesso spazio.

Esempio
Consideriamo lo spazio campione costituito dal lancio di 3 monete. Ogni moneta può assumere due possibili valori T (testa) o C (croce).
Lo spazio S è discreto e finito.
Posso definire la variabile aleatoria risultato del lancio della prima moneta:
$X(T_1)=+1$ o $X(C_1)=-1$
Questa è una variabile aleatoria in quanto associa all' evento T nella prima moneta il numero reale +1, ecc.

Posso definire la variabile aleatoria associata agli eventi risultato del lancio delle prime due monete.
$X(T_1,T_2)=+2$; $X(T_1,C_2)=0$; $X(C_1,T_2)=0$; $X(C_1,C_2)=-2$
Anche questa è una variabile aleatoria. ecc.

Distribuzione di probabilità **congiunta**

Per chiarire il concetto di distribuzione di **probabilità congiunta** facciamo un esempio nel caso discreto.



Consideriamo due dadi, ma in questo caso invece di interessarci alla somma dei due siamo interessati al valore che assumono singolarmente.
Ho quindi due variabili aleatorie e la mia distribuzione di probabilità e' una funzione del tipo:

$$S \times T \rightarrow \Re$$

I possibili valori delle variabili aleatorie sono
(1,1), (1,2), (2,1), ….. (6,6)
$f(x_i,y_j) = 1/36$ positivo e in questo caso uniforme

$$\sum_{i=1}^{6} \sum_{j=1}^{6} p(x_i, y_j) = 1$$

$f(x_i,y_j)$ è un esempio di distribuzione di probabilità congiunta

La **densità di probabilità congiunta** rappresenta la densità di probabilità di più di una variabile aleatoria **continua** (nel nostro caso 2).
Anche per la densità di probabilità valgono le stesse regole del caso a una dimensione:

$$f(x,y) \geq 0 \quad \forall x, y \in \Re^2$$

Definita positiva

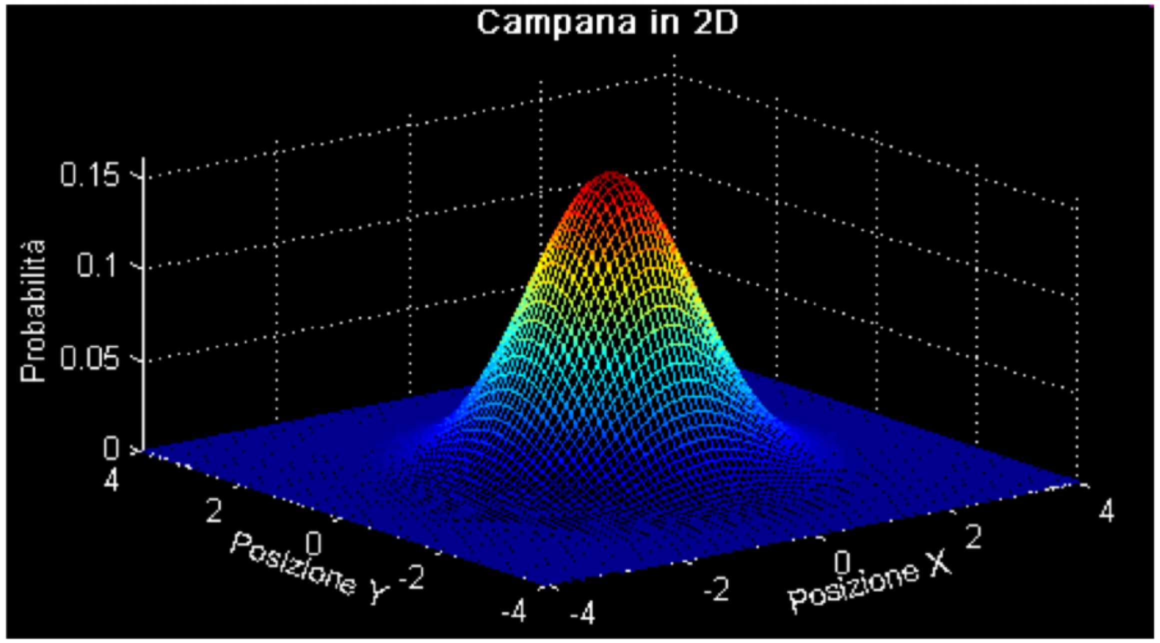
$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$$

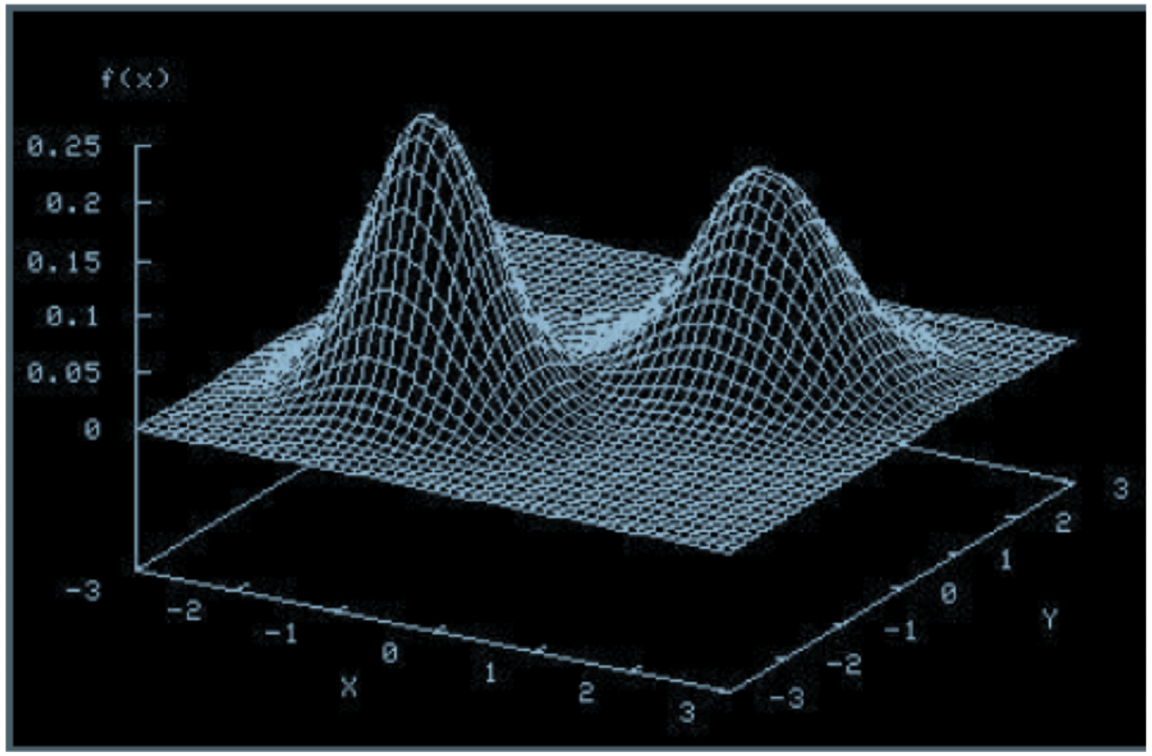
Normalizzazione

$$P(a < x < b, c < y < d) = \int_a^b\int_c^d f(x,y)dxdy$$

Probabilità di un evento

Gaussiana bi-dimensionale





Combinazione di Gaussiane bi-dimensionale

probabilità marginale

Si dice **distribuzione di probabilità marginale** la distribuzione di probabilità di una sola delle variabili aleatori. Nel nostro caso per esempio la probabilità che il dado rosso assuma un certo valore indipendentemente dal valore assunto dal dato blu.

$$p_X(x_i) = \sum_j p(x_i, y_j)$$

La distribuzione di probabilità marginale coincide con la distribuzione di probabilità della singola variabile aleatoria.
Esempio: Voglio la probabilità che sul dado rosso esca il numero 1 indipendentemente da ciò che succede sul dato blu.

$$p(1) = p(1,1) + p(1,2) + p(1,3) + p(1,4) + p(1,5) + p(1,6) =$$
$$= \frac{6}{36} = \frac{1}{6}$$

Dalla congiunta posso ottenere la marginale ma non vale il viceversa.

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

probabilità condizionata

Caso di variabili discrete

Si definisce probabilità condizionata del caso di due variabili aleatorie

$$p(x_i \mid y_j) = \frac{p(x_i, y_j)}{p_Y(y_j)}$$

Probabilità congiunta

Probabilità marginale

Notiamo il parallelismo tra queste due formulazioni di probabilità condizionata che abbiamo incontrato:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Sono diverse?

Interpretazione di probabilità congiunta

Cosa significa: "probabilità congiunta di due variabili aleatorie"?
Con questo termine si intende che i due eventi a cui le due variabili sono associate debbano verificarsi simultaneamente. Quindi:

$$p(x_i, y_j) = P(A \cap B)$$

Dove x_i è la variabile aleatoria associata all'evento A e y_j quella associata all'evento B.

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

Che significa che due variabili aleatorie sono **INDIPENDENTI**???

Dire che due variabili aleatorie sono indipendenti equivale a dire che gli eventi a cui esse sono associate sono indipendenti. Il verificarsi di uno dei due eventi non altera la probabilità associata al verificarsi dell'altro evento.

$$P(A\,|\,B) = \frac{P(A\cap B)}{P(B)} = P(A)$$

La probabilità che si verifichi A non dipende dal verificarsi di B

$$P(B\,|\,A) = \frac{P(B\cap A)}{P(A)} = P(B)$$

La probabilità che si verifichi B non dipende dal verificarsi di A

Da questo consegue che la probabilità che si verifichino entrambi è data dal prodotto delle probabilità

$$P(A\cap B) = P(A)P(B)$$

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

probabilità condizionata

Caso di variabili discrete

$$p(x_i \mid y_j) = \frac{p(x_i, y_j)}{p_Y(y_j)}$$

Probabilità congiunta

Probabilità marginale

Le variabili aleatorie discrete X e Y **si dicono indipendenti (per definizione)** se si verificano le relazioni

$$\boxed{\begin{aligned} p(x_i \mid y_i) &= p_X(x_i) \\ p(y_i \mid x_i) &= p_Y(y_i) \end{aligned}}$$

Dalle quali segue che

$$p(x_i, y_i) = p_X(x_i) \times p_Y(y_i)$$

la prob congiunta e' il prodotto delle prob marginali probabilità condizionata

Caso di variabili continue

Se si passa a variabili aleatorie continue si può dimostrare che quanto detto prima vale anche in questo caso solo che si utilizza la densità di probabilità al posto della probabilità:

Infatti, in termini di prob. abbiamo

$$\underbrace{dx\, f(x \mid y)}_{\text{Prob(x in [x,x+dx] | y)}} = \frac{\overbrace{f(x,y)\,dxdy}^{\text{Prob(x in [x, x+dx] e y in [y, y+dy])}}}{\underbrace{f_Y(y)\,dy}_{\text{Prob(y in [y, y+dy])}}}$$

Semplificando, si ha …

probabilità condizionata

Caso di variabili continue

Se si passa a variabili aleatorie continue si può dimostrare che quanto detto prima vale anche in questo caso solo che si utilizza la densità di probabilità al posto della probabilità:

$$f(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

$f(x, y)$ Densità di probabilità congiunta. Funzione di due variabili aleatorie.

$f(x \mid y)$ Densità di probabilità condizionata. Funzione di una sola variabile aleatoria (x). y è un parametro.

$f_Y(y)$ Densità di probabilità marginale.

$$\boxed{f_y(y) = \int f(x, y)\,dx}$$

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

probabilità condizionata

Caso di variabili continue

densita' di prob condizionata

$$f(x \mid y) = \frac{f(x,y)}{f_Y(y)}$$

densita' di prob congiunta

densita' di prob marginale

Le variabili aleatorie continue *x* e *y* **si dicono indipendenti (per definizione)** se si verificano le relazioni

$$\boxed{\begin{aligned} f(x \mid y) &= f_X(x) \\ f(y \mid x) &= f_Y(y) \end{aligned}}$$

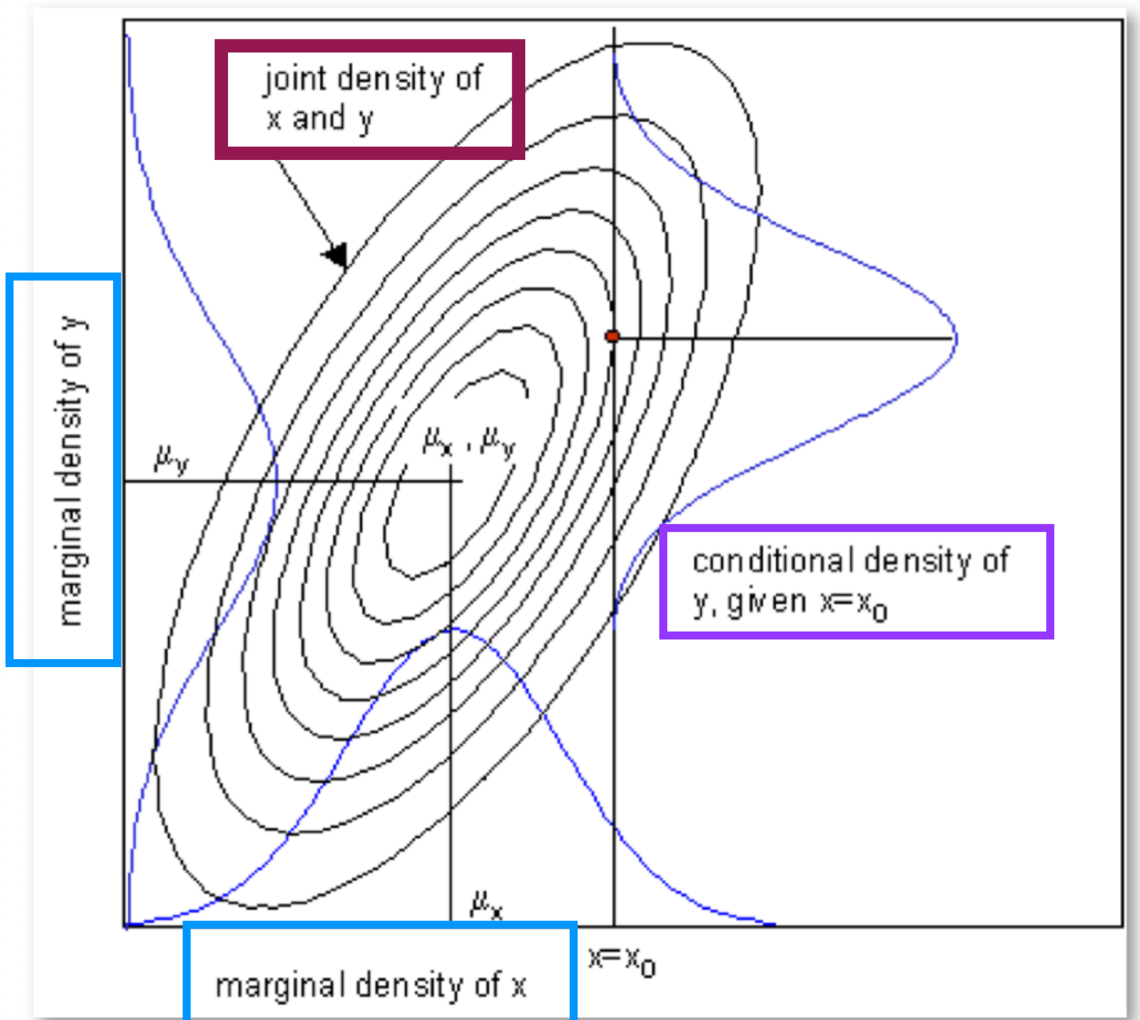
Dalle quali segue che

$$f(x,y) = f_X(x) \times f_Y(y)$$

la densita' di prob congiunta e' il prodotto delle densita' di prob marginali

Caso di variabili continue

$$f(x,y) = \frac{e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$



Distribuzione gaussiana bidimensionale nel caso di variabili aleatorie correlate con ρ coefficiente di correlazione.
Ponendo ρ=0 si ottiene l'espressione analoga nel caso di variabili non correlate.
Le distribuzioni marginali delle due variabili aleatorie sono delle gaussiane unidimensionali.

VARIABILI ALEATORIE E DISTRIBUZIONI DI PROBABILITÀ

Caso di variabili discrete

Esempio: X=dado rosso, Y= somma dei due dadi
(1,2),(1,3),(1,4),(1,5),(1,6),(1,7),(2,3),……..(6,12)
Anche in questo caso la $f(x_i,y_j)$ è uniforme =1/36

$$\sum_{i=1}^{6}\sum_{j=i+1}^{i+6} p(x_i, y_j) = 1$$

Etc.

Anche in questo caso posso ottenere la distribuzione marginale a partire dalla congiunta ma le due variabili sono chiaramente NON indipendenti.

Integro su x $\quad p(y=2) = 1/36$ $\quad$ *prob. marginale*

Integro su y $\quad p(x=1) = p(1,2) + p(1,3) + p(1,4) + p(1,5) + p(1,6) + p(1,7) =$

$$= \frac{6}{36} = \frac{1}{6}$$

Probabilità marginale

p(x,y) != p(x) x p(y)

p(1,2) = 1/36 != 1/6 x 1/36 => Eventi NON indipendenti

Valori di aspettazione di una variabile aleatoria a partire da una distribuzione congiunta

$$\mu_X = \int x f_X(x) dx = \iint x f(x,y) dx dy$$

Valore di aspettazione della variabile X

$$\mu_Y = \int y f_Y(y) dy = \iint y f(x,y) dx dy$$

Valore di aspettazione della variabile Y

Analogo per variabili discrete

$$var(x) = \iint (x - \mu_x)^2 f(x,y) dx dy = \int (x - \mu_x)^2 f_x(x) dx$$

RIASSUNTO

- Variabile aleatoria: $X : B \rightarrow \mathbb{R}, \quad X(w) = x$, con $w \in B$, $x \in \mathbb{R}$
    - Dominio = campo di Borel, Insieme di arrivo => numeri reali
    - Discreta o continua
  - Caso discreto: la distribuzione di probabilita' $f(x_i)$ quantifica la probabilita' di del valor x_i
  - Caso continuo: la densità di probabilita' (moltiplicata per dx) $f(x)\,dx$ quantifica la probabilita' che la variabile sia compresa tra il valore x e x+dx
    - Definite positive e normalizzate a 1; **momenti della distribuzioni**
- Distribuzioni di probabilita' congiunte, descrivono la probabilita' di evenienza contemporanea di alcuni valori specifici di due [x,y] (o piu') variabili aleatorie
  - Probabilita' marginale di una variabile aleatoria: prob di x per ogni possibile valore di y
  - Probabilita' condizionata: prob del valore x se si e' verificato y
  - Variabili indipendenti

ALCUNE DISTRIBUZIONI DI PROBABILITÀ

DISTRIBUZIONI DI PROBABILITÀ NOTEVOLI

Distribuzioni continue: **distribuzione uniforme**
Una variabile aleatoria X si dice che segue una distribuzione di probabilità uniforme tra i valori a e b se la sua funzione densità di probabilità assume la forma:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & altrove \end{cases}$$

La corrispondente distruzione cumulativa è data da:

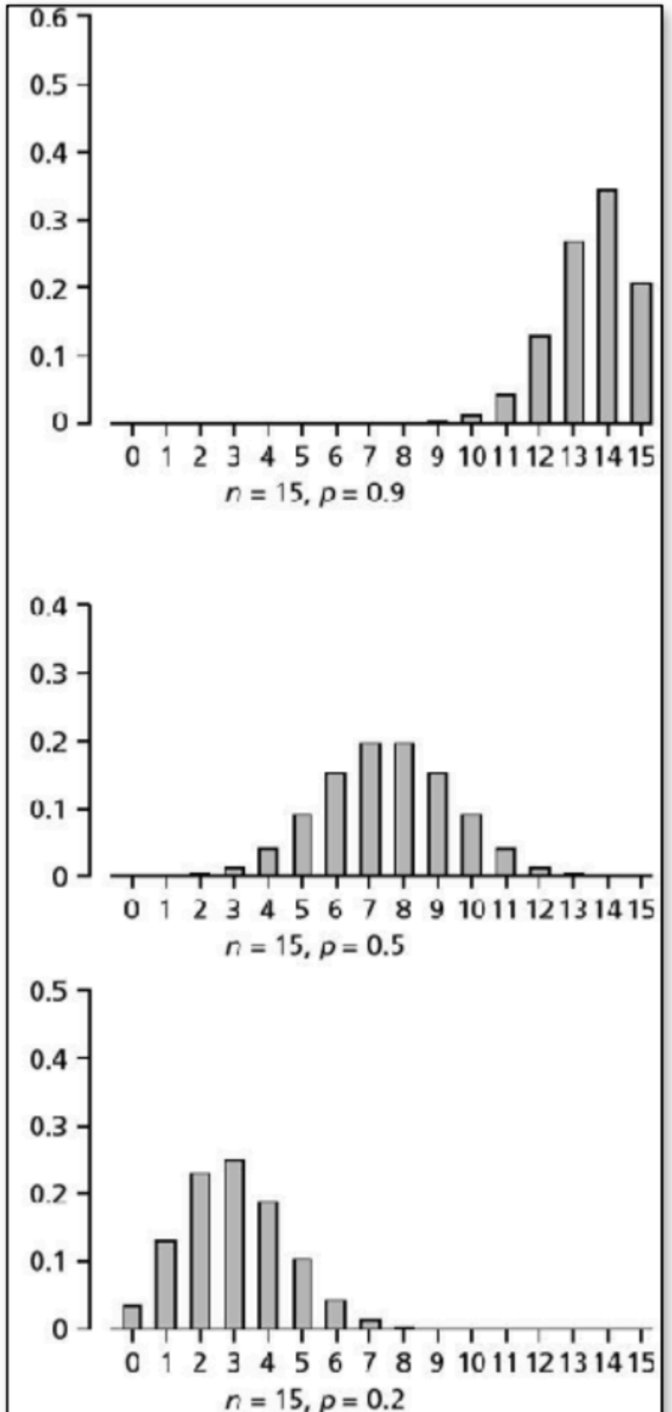
$$F(x) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \ge b \end{cases}$$

Si ottiene inoltre:

$$E[X] = \frac{a+b}{2} \qquad Skew = 0$$

$$\mathrm{var}(X) = \frac{(b-a)^2}{12} \qquad Kurt = \frac{9}{5}$$

DISTRIBUZIONI DI PROBABILITÀ NOTEVOLI



Distribuzioni discrete: **distribuzione Binomiale o di Bernoulli** {J. Bernoulli 1654-1705}
Dato un processo in cui sono possibili solo due risultati ("successo" o "insuccesso") mutuamente esclusivi e esaustivi, Se la probabilità di successo p è la stessa per ogni prova e se tutte le prove sono indipendenti allora la variabile aleatoria X che conta il numero di successi x su N prove assume distribuzione di probabilità data da:

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x} \quad x=0,1,\dots,n$$

X successi (ciascuno con prob p)
e n-x insuccessi (ciascuno relativo a prob 1-p)
Non conta l'ordine => **combinazioni di x tra n**

Si dimostra inoltre che :

$$E[X]=pn$$

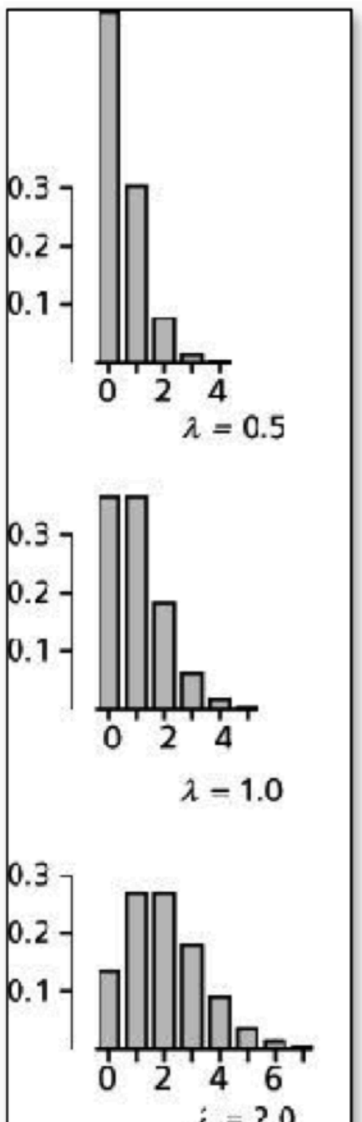
$$\mathrm{var}(X)=np(1-p)$$

$$Skew=\frac{1-2p}{\sqrt{p(1-p)}}$$

$$Kurt=\frac{1-3p(1-p)}{p(1-p)}$$

DISTRIBUZIONI DI PROBABILITÀ NOTEVOLI

Nell'intervallo il numero di prove e' estremante alto, ma la prob di successo e' molto bassa, Np = λ finito



Distribuzione discrete: **distribuzione di Poisson** {S. D. Poisson 1781-1840}

Dato un processo in cui in un intervallo di tempo o di spazio possa verificarsi un numero arbitrario di eventi caratterizzati, però, da una probabilità molto bassa. Ipotizziamo inoltre che la probabilità di verificarsi di un evento non influenzi la probabilità che si verifichi un secondo evento nello stesso intervallo e che la probabilità che un evento si verifichi dipenda solo dall'ampiezza dell'intervallo.

Allora la variabile aleatoria X che indica il numero di volte in cui l'evento si e' verificato in un dato intervallo segue la distribuzione di probabilità data da:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0,1,\ldots$$

Si noti che a differenza del caso binomiale in cui x va da 0 ad n, in questo caso x va da 0 ad infinito. Il parametro λ>0 indica il numero medio di realizzazione di eventi nell'intervallo.

Si trova quindi che:

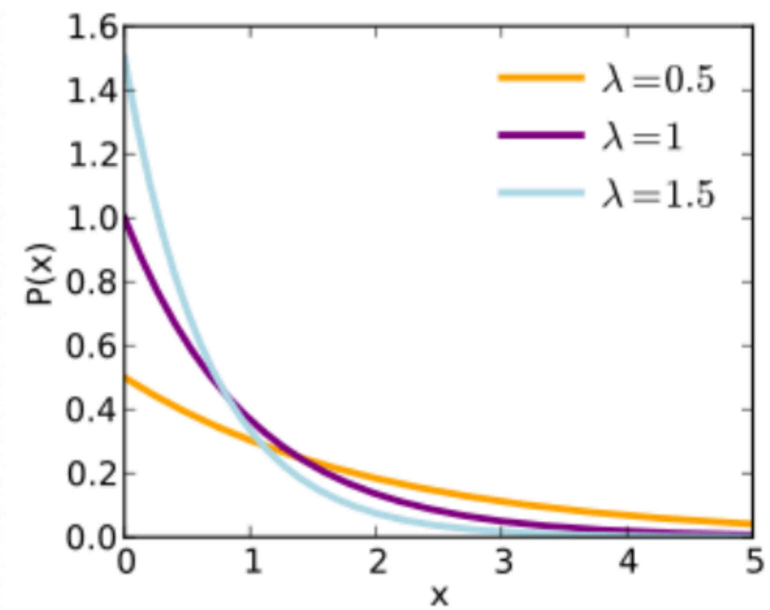
$$E[X] = \lambda \qquad Skew = \frac{1}{\sqrt{\lambda}}$$

$$\mathrm{var}[X] = \lambda \qquad Kurt = \frac{1+3\lambda}{\lambda}$$

DISTRIBUZIONI DI PROBABILITÀ NOTEVOLI

Distribuzioni continue: **distribuzione esponenziale**
Un'altra distribuzione di probabilità molto usata in fisica è la distribuzione esponenziale la cui densità di probabilità è data da:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq o \\ 0 & altrimenti \end{cases}$$



Si dimostra che

$$E[X] = \frac{1}{\lambda}$$

$$\mathrm{var}(X) = \frac{1}{\lambda^2}$$

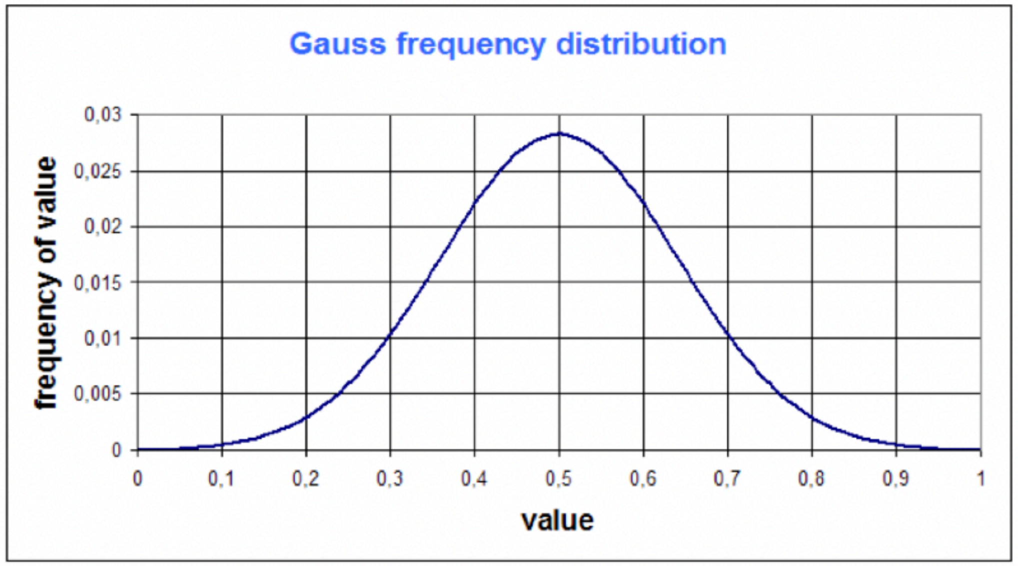
$$Skew = 2$$

$$Kurt = 9$$

DISTRIBUZIONI DI PROBABILITÀ NOTEVOLI

Distribuzione continue: **distribuzione normale o di Gauss** {C.F. Gauss 1777-1855}
Detta anche legge degli errori
La distribuzione normale è alla base dell'inferenza statistica grazie al teorema del limite centrale. Si dimostra che le distribuzioni Binomiale e di Poisson tendono ad una distribuzione normale al crescere del loro valore medio.

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



Si dimostra che

$$E[X] = \mu \qquad Skew = 0$$

$$\mathrm{var}(X) = \sigma^2 \qquad Kurt = 3$$

LEZIONE 8

FUNZIONI DI UNA VARIABILE ALEATORIE

Nota la pdf di una variabile aleatoria X,
e data Y=g(X), Y e' un variabile aleatoria.

Qual e' il valore di aspettazione, qual e' la varianza di Y ?

FUNZIONI DI VARIABILI ALEATORIE

Nota la pdf di una variabile aleatoria X,
e Y=g(X), Y e' un variabile aleatoria.

Qual e' il valore di aspettazione, qual e' la varianza di Y ?

$$E[Y] = E[g(x)] = \int y f_Y(y) dy$$

law of the unconscious statistician

LOTUS stabilisce che

$E[Y] = E[g(x)] = \int y f_Y(y) dy = \int g(x) f_X(x) dx$ ===>>>> non occorre conoscere la pdf della variabile y per calcolare il valore di aspettazione di Y, basta calcolare il valore di aspettazione della g(x) secondo la pdf della variabile aleatoria X.

DOMANDA:

$E[Y] = E[\ g(x)\ ] = g(\ E[X]\ )$??? **NON SEMPRE**

FUNZIONI DI VARIABILI ALEATORIE

DOMANDA:
$E[Y] = E[g(x)] = g(E[X])$??? **NON SEMPRE**

Osserviamolo empiricamente su un esempio

Esempio:
Supponiamo di chiedere ad una classe di ragazzini di 5 elementare di tagliarci dei dischi di carta di 10 centimetri di diametro.
Alla fine del lavoro constatiamo che **la variabile aleatoria raggio dei dischi** segue una distribuzione gaussiana con media 5 cm e sigma 1 cm.

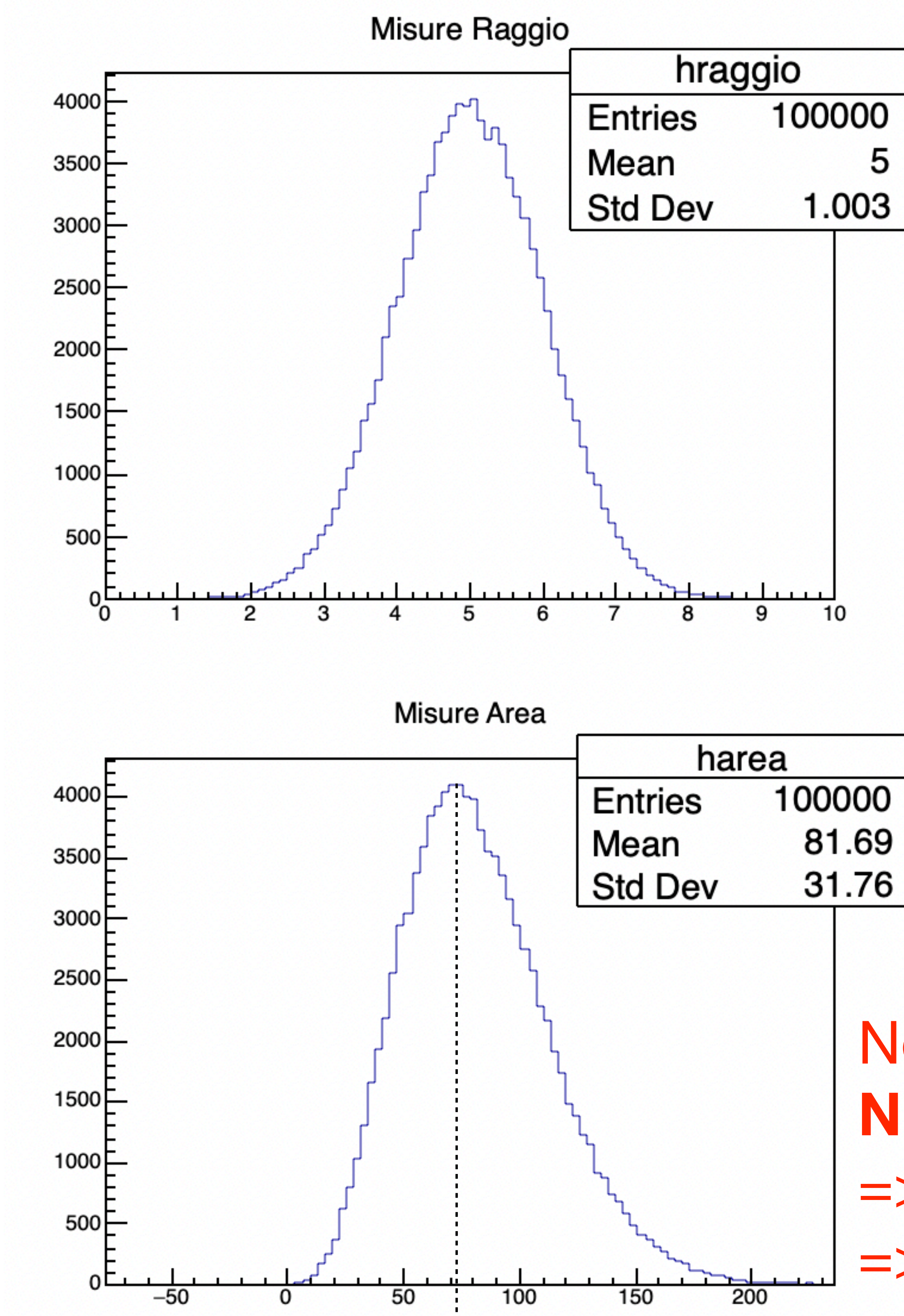
1) Che distribuzione di probabilità seguirà la variabile aleatoria superficie dei dischi?

2) Quanto varranno la media e la varianza della variabile aleatoria superficie dei dischi

Simulazione dell'esperimento.
La variabile aleatoria che rappresenta il raggio dischi di carta può essere simulata estraendo N valori da una distribuzione gaussiana con media 5 e varianza 1.

L'utilizzo della simulazione mi permette di determinare per il mio campione di dischi la loro superficie e quindi di studiare direttamene la densità di probabilità associata alla variabile aleatoria superficie del disco

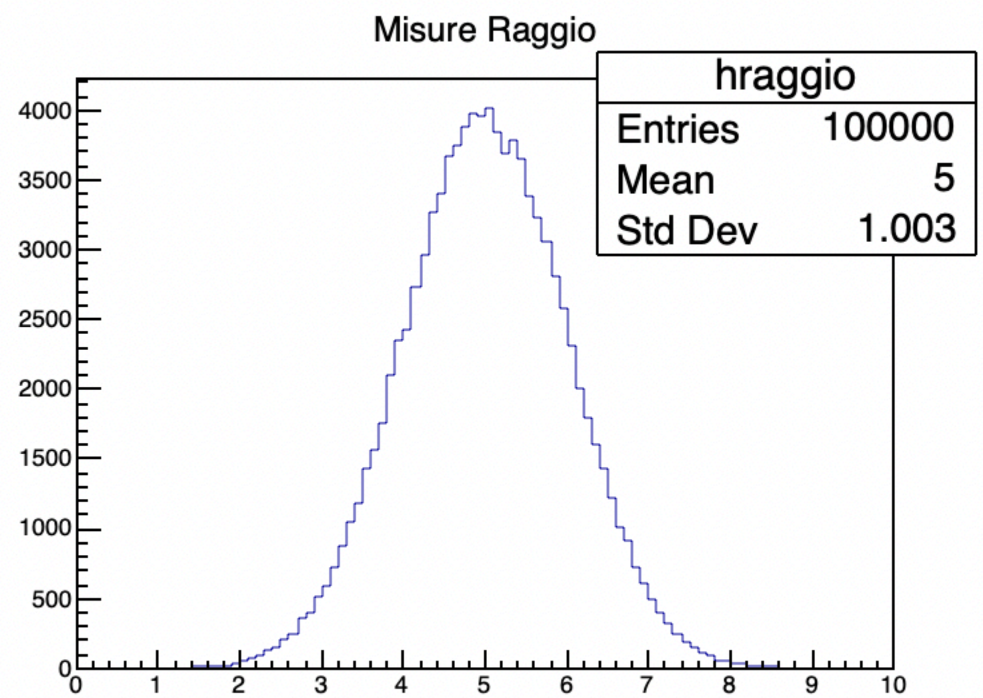
FUNZIONI DI VARIABILI ALEATORIE

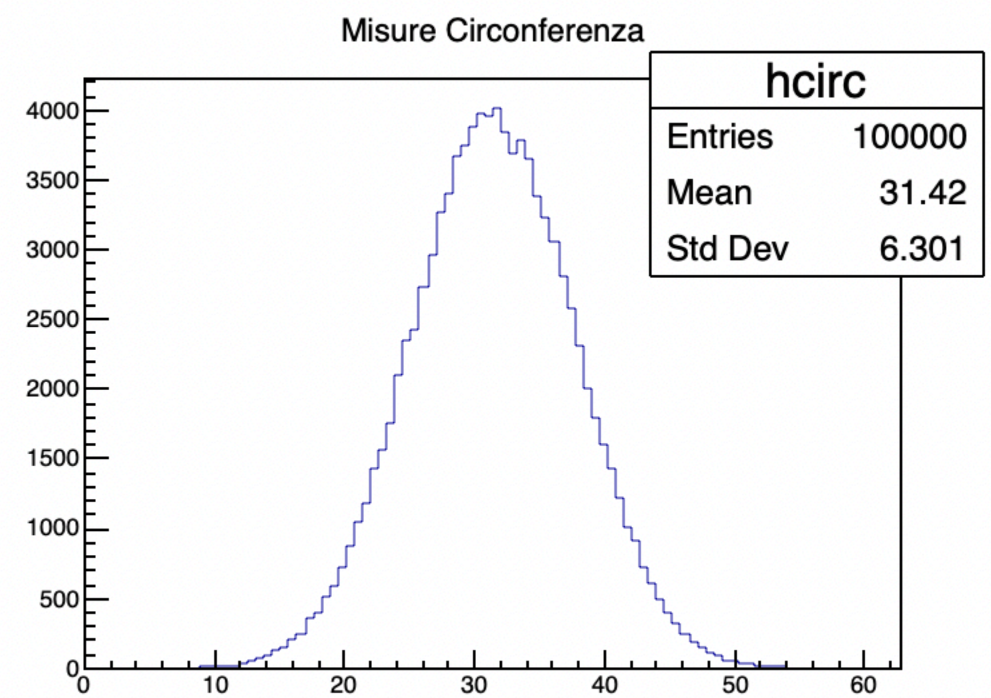


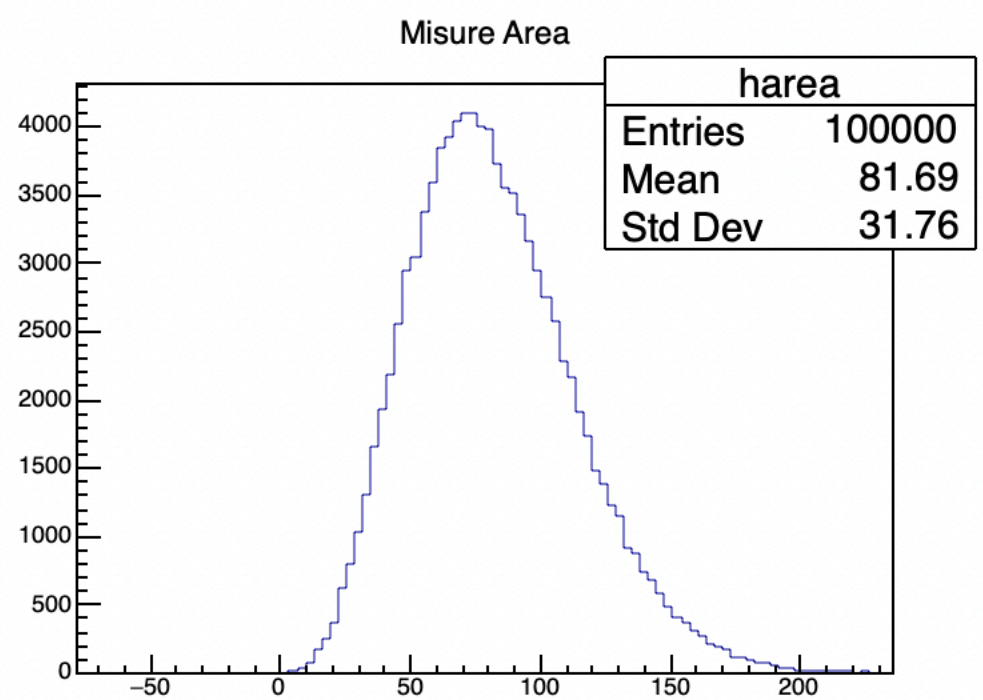
Nota: circonferenza = 2πR
E' distribuita gaussianamente come R
=> Media 31.42 = 2π<R>

Nota: **area** = πR^2
Non e' distribuita gaussianamente come R
=> Media 81.69 > Moda ~ 72
=> Media != $\pi\langle R\rangle^2$ ~ 78,5

FUNZIONI DI VARIABILI ALEATORIE








```
Raggio, raggioMedio     5 5.0002 devStd_raggio 1.00284

              Raggio,  Circ, Area      5      31.4159 78.5398 valori esatti
da simulazione: Raggio,  Circ, Area      5.0002 31.4172 81.7054
da simulazione:eRaggio, eCirc,eArea      0.00317126 0.0993498 0.258375
da f(R_medio) :          Circ, Area      5.0002 31.4172 78.5459
simul-f(<R>)  :          Circ, Area      2.45137e-13 3.15946
[simul-f(<R>)]/err:      Circ, Area      2.46742e-12 12.2282
Correzione per Area = 3.15946   Area media attesa(post correzione) = 81.7054
```

FUNZIONI DI VARIABILI ALEATORIE

Nota la pdf di una variabile aleatoria X,
e Y=g(X), Y e' un variabile aleatoria.

Qual e' il valore di aspettazione, qual e' la varianza di Y ?

$$E[Y] = E[g(x)] = \int y f_Y(y) dy$$

law of the unconscious statistician

LOTUS stabilisce che

$$E[Y] = E[g(x)] = \int y f_Y(y) dy = \int g(x) f_X(x) dx$$ ===>>>> non occorre conoscere la pdf della variabile y per calcolare il valore di aspettazione di Y, basta calcolare il valore di aspettazione della g(x) secondo la pdf della variabile aleatoria X.

Funzioni di Variabili Aleatorie

LOTUS Dimostrazione non rigorosa

$$E[Y] = E[g(x)] = \int y f_Y(y)dy = \int g(x) f_X(x) dx$$

Variabile Y=g(X), dove la g(X) è una funzione **strettamente monotona crescente**. Supponiamo nota la densità di probabilità che descrive la variabile X ($f_X(X)$) che espressione avrà la distribuzione di probabilità associata alla Y ($f_Y(Y)$)?

Ricaviamo $f_Y(y)$ con una dimostrazione non rigorosa e limitata al caso di una g(x) strettamente monotona e pertanto invertibile

Parto dall'identità:

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1 = \int_{-\infty}^{+\infty} f_Y(y)dy$$

Essendo y=g(x) strettamente monotona allora è invertibile per cui posso determinare $x=g^{-1}(y)$, da cui:

$$dx = \frac{d(g^{-1}(y))}{dy} dy$$

Eseguendo il cambiamento di variabili al primo membro della precedente identità ottengo

$$\int_{-\infty}^{+\infty} f_X(g^{-1}(y)) \frac{d(g^{-1}(y))}{dy} dy = \int_{-\infty}^{+\infty} f_Y(y)dy$$

Non rigoroso….

$$f_Y(y) = f_X(g^{-1}(y)) \frac{d(g^{-1}(y))}{dy}$$

LOTUS Dimostrazione non rigorosa

$$E[Y] = E[g(x)] = \int y f_Y(y) dy = \int g(x) f_X(x) dx$$

Se Y=g(X), con g(X) **strettamente monotona decrescente**.

$$\int_{-\infty}^{+\infty} f_X(g^{-1}(y)) \frac{d(g^{-1}(y))}{dy} dy = -\int_{-\infty}^{+\infty} f_Y(y) dy$$

$$f_Y(y) = -f_X(g^{-1}(y)) \frac{d(g^{-1}(y))}{dy}$$

Da cui genericamente:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d(g^{-1}(y))}{dy} \right|$$

Funzioni di Variabili Aleatorie

LOTUS Dimostrazione non rigorosa

$$E[Y] = E[g(x)] = \int y f_Y(y) dy = \int g(x) f_X(x) dx$$

Calcolo della media della variabile aleatoria Y=g(X)

Per calcolare il valore di aspettazione della variabile aleatoria Y occorre fare l'integrale

$$\mu_y = \int_{-\infty}^{+\infty} y f_Y(y) dy$$

ma

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d(g^{-1}(y))}{dy} \right|$$

Relazione valida sempre anche se g(x) non e' strettamente monotona

$$x = g^{-1}(y); \qquad dx = \left| \frac{d(g^{-1}(y))}{dy} \right| dy$$

$$\Rightarrow \boxed{\mu_y = \int_{-\infty}^{+\infty} g(x) f_X(x) dx}$$

L'ultima espressione mi permette di determinare, quindi, il valore di aspettazione della variabile Y=g(X) nota la densità di probabilità che caratterizza la variabile aleatoria X.

Funzioni di Variabili Aleatorie

Calcolo della varianza della variabile aleatorie Y=g(X)

Analogamente al caso della media, per calcolare la varianza della Y dovrei calcolare l'integrale

$$\mathrm{var}(Y) = \int_{-\infty}^{+\infty} (y - \mu_y)^2 f_Y(y)dy$$

ma

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d(g^{-1}(y))}{dy} \right|$$

$$x = g^{-1}(y) \qquad ; \qquad dx = \left| \frac{d(g^{-1}(y))}{dy} \right| dy$$

quindi

$$\mathrm{var}(Y) = \int_{-\infty}^{+\infty} (g(x) - \mu_y)^2 f_X(x)dx$$

Relazione valida sempre anche se g(x) non e' strettamente monotona

Quest'ultima espressione mi permette di determinare la varianza utilizzando la densità di probabilità della variabile X

FUNZIONI DI VARIABILI ALEATORIE

Media e varianza di una funzione lineare di X

Alcuni casi particolari

Per valor medio e varianza valgono le seguenti proprietà:

$$Y = aX + b$$

$$E[Y] = E[aX + b] = aE[X] + b \quad \forall a,b \in \Re$$

DOMANDA:
$E[Y] = E[g(x)] = g(E[X])$?
SI per funzioni lineari

Dimostrabile dalla definizione di valore di aspettazione.
Qualche attenzione in più merita il calcolo della varianza.

$$\mathrm{var}(Y) = \mathrm{var}(aX + b) = E[(aX + b - E[aX + b])^2] =$$

$$= E[(aX + b - aE[X] - b)^2] =$$

$$= E[(a(X - E[X]))^2] = E[a^2(X - \mu)^2] =$$

$$= a^2\,\mathrm{var}(X)$$

Per cui

$$\mathrm{var}(aX + b) = a^2\,\mathrm{var}(X) \quad \forall a,b \in \Re$$

FUNZIONI DI VARIABILI ALEATORIE

Media e varianza di una funzione ***generica*** di X

Caso generale

$$\mu_y = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

Ipotizziamo ora che la variabile aleatoria X sia caratterizzata da una densità di probabilità di tipo gaussiano con standard deviation in valore assoluto piccola rispetto al valore medio. Per calcolare l'integrale precedente possiamo sviluppare in serie la funzione g(x) intorno alla media della variabile aleatoria X

$$g(x) = g(\mu_X) + (x - \mu_X) g'(\mu_X) + \frac{1}{2}(x - \mu_X)^2 g''(\mu_X) + \dots$$

$$\mu_Y = \int g(\mu_X) f(x) dx + \int (x - \mu_X) g'(\mu_X) f(x) dx + \int \frac{1}{2}(x - \mu_X)^2 g''(\mu_X) f(x) dx + \dots$$

Se mi fermo ai primi due termini dello sviluppo in serie ottengo che

$g(\mu_X) \int f(x) dx = g(\mu_X)$ Termine di ordine 0

$g'(\mu_X) \int (x - \mu_X) f(x) dx = 0$ Termine di ordine 1

Per cui posso concludere che:

$$\mu_Y = g(\mu_X)$$

Se mi fermo al primo ordine

FUNZIONI DI VARIABILI ALEATORIE

Ma attenzione. Non sempre i termini di ordine superiore possono essere trascurati!
Se aggiungessi il termine del secondo ordine dello sviluppo precedente

$$\frac{1}{2}\int (x-\mu_X)^2 g''(\mu_X) f(x)dx = \frac{1}{2} g''(\mu_X)\int (x-\mu_X)^2 f(x)dx =$$

$$= \frac{1}{2} g''(\mu_X)\,\mathrm{var}(X)$$ Termine di ordine 2

Otterrei che

$$\mu_Y = g(\mu_X) + \frac{1}{2} g''(\mu_X)\,\mathrm{var}(x)$$

DOMANDA:
$E[Y] = E[g(x)] = g(E[X])$?
NO in generale

Questo termine aggiuntivo risulta trascurabile solo se il prodotto $g''(\mu_X)\mathrm{var}(X)$ è piccolo rispetto a $g(\mu_X)$. Quindi solo nel caso di funzioni g(x) lentamente variabili ($g''(\mu_X) \cong 0$) nella regione di interesse posso trascurare il termine correttivo.

FUNZIONI DI VARIABILI ALEATORIE

- **Varianza**
  - Come prima sviluppiamo in serie di Taylor g(x) attorno a μ_X
    - $$g(x) = g(\mu_X) + (x-\mu_X)g'(\mu_X) + \frac{1}{2}(x-\mu_X)^2 g''(\mu_X) + \ldots$$
  - *var(Y)* = $\int [\ g(\mu_X) - \mu_Y + (x-\mu_X)g'(\mu_X) + \frac{1}{2}(x-\mu_X)^2 g''(\mu_X) + \ldots\ ]^2\ f(x)dx$
    - ***ci arrestiamo al primo ordine***, quindi per consistenza usiamo $\mu_Y = g(\mu_X)$ quindi
  - = $\int [\ g(\mu_X) - \mu_Y\ ]^2 f(x)dx + \int [(x-\mu_X)g'(\mu_X)]^2 f(x)dx + \int 2[\ g(\mu_X) - \mu_Y\ ][(x-\mu_X)g'(\mu_X)]\ f(x)dx$
    - Il primo termine e' nullo perché $g(\mu_X) - \mu_Y$ e' circa zero
    - Il terzo e e' nullo perché si tratta di un momento centrato del primo ordine, inoltre ricompare $g(\mu_X) - \mu_Y$

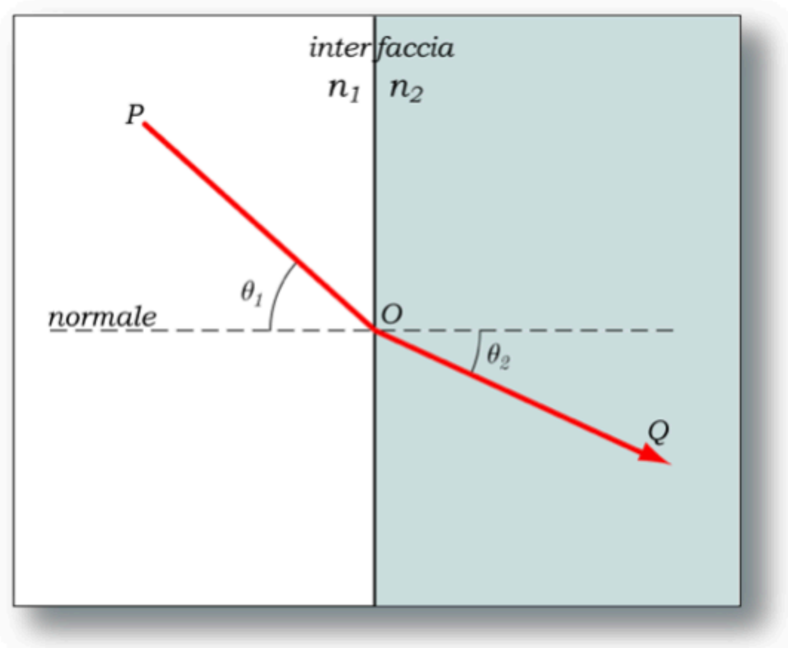
$$\mathrm{var}(Y) = g'^2(\mu_X)\,\mathrm{var}(X)$$

Formula "standard" alla base della propagazione degli errori statistici.
… ma come è facile immaginare la cosa non funziona perfettamente …..

Perche' e' ottenuta con uno sviluppo fino al primo ordine

FUNZIONI DI VARIABILI ALEATORIE

Esempio: La legge di Snell
La legge di Snell descrive le modalità di rifrazione di un raggio luminoso quando passa tra due mezzi di indice di rifrazione diverso



$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2)$$

Dove n_1 e n_2 sono detti indici di rifrazione rispettivamente del mezzo 1 e 2.
Se il mezzo 1 è il vuoto allora l'indice di rifrazione è pari ad 1. (Nel caso dell'aria l'indice di rifrazione è quasi 1 e spesso viene approssimato con 1)

Supponiamo di avere uno strumento che mi permette di misurare l'indice di rifrazione di un mezzo facendo incidere un raggio luminoso in modo quasi radente sul mezzo. Ipotizziamo che lo strumento misuri l'angolo di incidenza con un errore relativo del 5%, mentre misuri l'angolo di rifrazione con precisione molto maggiore. Possiamo ritenere che al calcolo dell'errore su n contribuisca solo la misura dell'angolo di incidenza.

FUNZIONI DI VARIABILI ALEATORIE

Supponiamo che si effettui una serie di misure e si ottenga che la variabile aleatoria *angolo di incidenza* segua una distribuzione di tipo gaussiano con media pari a 1.56 radianti e deviazione standard di 0.08 radianti.
Se l'angolo di rifrazione è pari a 0.6283185 radianti (con errore trascurabile) quanto vale l'indice di rifrazione e qual è l'errore associato a questa misura?

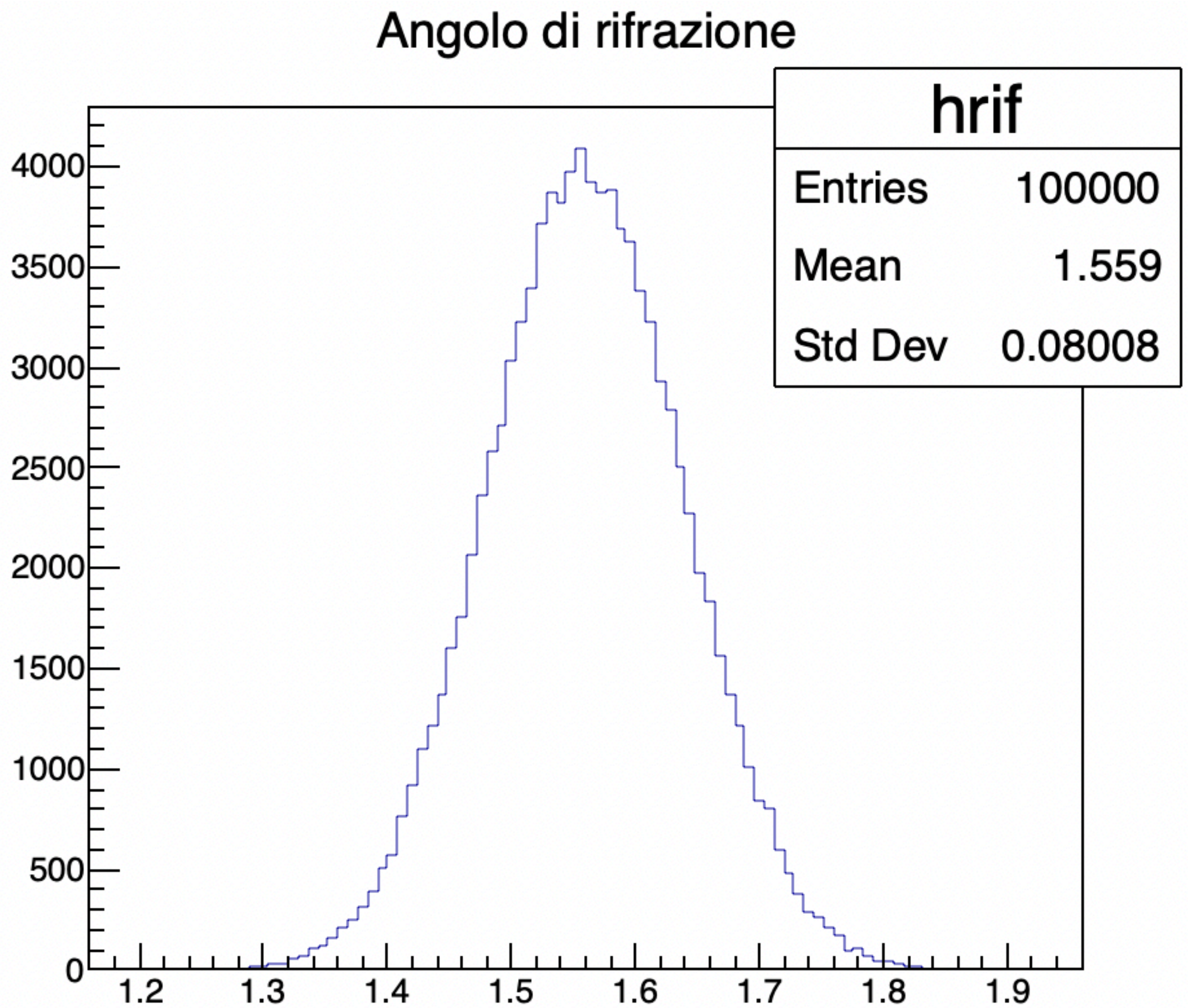
Calcolo analitico con formule approssimate:

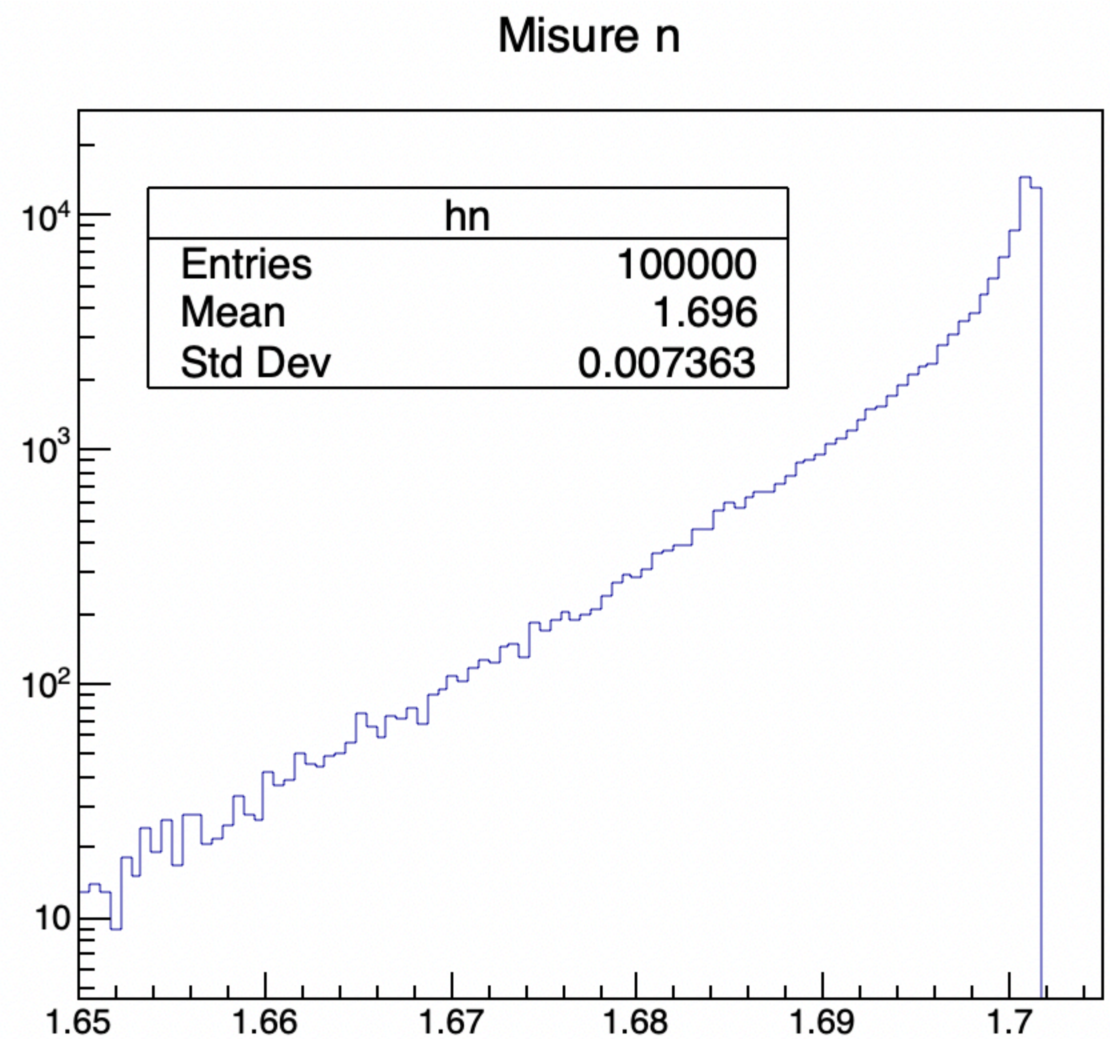
$$n = \frac{\sin(\theta_i)}{\sin(\theta_r)} = 1.7012$$

$$\mathrm{var}(n) = \left(\frac{\cos(\theta_i)}{\sin(\theta_r)}\right)^2 \mathrm{var}(X) = 0.000002$$

Simuliamo l'esperimento per valutare correttamente la propagazione degli errori ….

FUNZIONI DI VARIABILI ALEATORIE





Risultati attesi: n=1.70119 errore=4.90223e-06
Risultati ottenuti con la simulazione: indice di rifrazione ed errore 1.69575 2.47642e-05
Correction (for second order term)= -0.00545502
Difference Simulation - Theory(1st order) = 0.0054463
Valore atteso + correzione = 1.69574

FUNZIONI DI PIU' VARIABILI ALEATORIE

FUNZIONI DI DUE VARIABILI ALEATORIE

Media

Consideriamo il caso di una variabile aleatoria derivata funzione di due variabili aleatorie z=g(x,y), dove g(x,y) è una funzione qualunque.
Si può dimostrare che calcolare

$$\mu_Z = \int z f_Z(z) dz$$

Equivale a calcolare

$$\mu_Z = \int\int g(x,y) f(x,y) dxdy$$

E ricorrendo allo sviluppo in serie di Taylor della funzione g(x,y) intorno al punto $P_\mu=(\mu_X, \mu_Y)$ fermandosi al primo ordine si ottiene.

$$\mu_Z = g(\mu_X, \mu_Y)$$

Dove μ_X e μ_Y sono i valori di aspettazione della distribuzioni marginali.

FUNZIONI DI DUE VARIABILI ALEATORIE

Dimostrazione

$$\mu_X = \int x f_X(x)dx = \iint x f(x,y)dxdy$$ Valore di aspettazione della variabile X

$$\mu_Y = \int y f_Y(y)dy = \iint y f(x,y)dxdy$$ Valore di aspettazione della variabile Y

$$g(x,y) = g(\mu_X,\mu_Y) + (x-\mu_X)\left.\frac{\partial g}{\partial x}\right|_{\mu_X,\mu_Y} + (y-\mu_Y)\left.\frac{\partial g}{\partial y}\right|_{\mu_X,\mu_Y} + \boxed{?}$$

$$\mu_Z = \iint g(x,y) f(x,y)dxdy = g(\mu_X,\mu_Y)\underbrace{\iint f(x,y)dxdy}_{1} +$$

$$\iint \underbrace{(x-\mu_X) f(x,y)}_{0}\left.\frac{\partial g}{\partial x}\right|_{\mu_X,\mu_Y} dxdy + \iint \underbrace{(y-\mu_Y) f(x,y)}_{0}\left.\frac{\partial g}{\partial y}\right|_{\mu_X,\mu_Y} dxdy + \boxed{?}$$

$$\mu_Z = g(\mu_X,\mu_Y)$$

Ma in questo modo, come nel caso ad una singola variabile, è possibile introdurre sistematiche se le derivare seconde della g(x,y) nel punto P_μ sono molto diverse da zero.

FUNZIONI DI DUE VARIABILI ALEATORIE

Casi semplici: media e varianza di funzioni lineari

$$E[aX + bY] = aE[X] + bE[Y] \quad \forall a,b \in \Re$$

Dimostrabile dalla definizione di valore di aspettazione

Uso E[X+Y] = E[X]+E[Y]

$$\mathrm{var}(X+Y) = E[((X+Y) - E[X+Y]))^2] =$$

$$= E[((X+Y) - E[X] - E[Y])^2] = E[\,(\,(X-E[X]) + (Y-E[Y])\,)^2\,] =$$

$$= E[((X-E[X])^2 + (Y-E[Y])^2 + 2(X-E[X])(Y-E[Y])] =$$

$$= \mathrm{var}(X) + \mathrm{var}(Y) + 2\,\mathrm{cov}(X,Y)$$

$$\mathrm{cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(y-\mu_y)(x-\mu_x)f(x,y)\,dxdy$$

Covarianza

Dove f(x,y) è la densità di probabilità congiunta delle due variabili aleatorie e μ_x μ_y rappresentano rispettivamente il valore di aspettazione della variabile X e della variabile Y.

Covarianza

$\boxed{\text{Se le variabili sono indipendenti}}$ allora

$$f(x,y) = f_X(x) f_Y(y)$$

Da cui segue che la covarianza è nulla essendo il valore di aspettazione intorno alla media di una qualunque variabile aleatoria nullo.

$$E[(x-\mu)] = 0$$

$$\mathrm{cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (y-\mu_y)(x-\mu_x) f(x,y)\,dx\,dy =$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (y-\mu_y)(x-\mu_x) f_X(x) f_Y(y)\,dx\,dy =$$

$$= \int_{-\infty}^{+\infty} (y-\mu_y) f_Y(y)\,dy \int_{-\infty}^{+\infty} (x-\mu) f_X(x)\,dx = 0$$

Covarianza

Nel caso di variabili aleatorie indipendenti

$$E[aX+b] = aE[X]+b \quad \forall a,b \in \Re$$

$$\mathrm{var}(aX+b) = a^2\,\mathrm{var}(X) \quad \forall a,b \in \Re$$

$$E[aX+bY] = aE[X]+bE[Y] \quad \forall a,b \in \Re$$

$$\mathrm{var}(aX+bY) = a^2\,\mathrm{var}(X)+b^2\,\mathrm{var}(Y) \quad \forall a,b \in \Re$$

Da cui consegue che:

$$E[X-Y] = E[X]-E[Y]$$

$$\mathrm{var}(X-Y) = \mathrm{var}(X)+\mathrm{var}(Y)$$

Utilizzato nella stima della propagazione degli errori sperimentali su grandezze derivate

Si noti che la covarianza nulla è condizione necessaria ma non sufficiente affinché due variabili aleatorie siano indipendenti.

Covarianza e coefficiente di correlazione

Spesso invece della covarianza si usa il coefficiente di correlazione

$$\rho = \frac{\mathrm{cov}(x,y)}{\sqrt{\mathrm{var}(x)\,\mathrm{var}(y)}}$$

Se le variabili sono indipendenti la correlazione è nulla.
Il coefficiente di correlazione, data la sua definizione, assume valori compresi tra -1 e 1 ed è adimensionale (non lo sono media e varianza!)

FUNZIONI DI DUE VARIABILI ALEATORIE

Varianza di Z=g(X,Y) in generale

In termini più generali, limitandosi al primo ordine si trova che

$$\mathrm{var}(Z) = \left(\frac{\partial g}{\partial x}\right)^2_{|P_\mu} \mathrm{var}(X) + \left(\frac{\partial g}{\partial y}\right)^2_{|P_\mu} \mathrm{var}(Y) + 2\,\mathrm{cov}(x,y)\left(\frac{\partial g}{\partial x}\frac{\partial g}{\partial y}\right)_{|P_\mu}$$

Si noti la presenza del termine di covarianza già approssimando al primo ordine lo sviluppo in serie.
Se le variabili sono indipendenti allora il termine di covarianza scompare.

Nel caso più generale di $Z=g(X_1,X_2, \ldots ,X_n)$ l'espressione della varianza assume la forma.

$$\mathrm{var}(Z) = \sum_{i,j} \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\mathrm{cov}(x_i, x_j)$$

Dove

$$\mathrm{cov}(x_i, x_i) = \mathrm{var}(x)$$

In generale per una stima accurata dei momenti delle variabili aleatorie derivate è conveniente ricorrere a tecniche di simulazione.

Stima della Varianza e Covarianza da un campione

Per calcolare la covarianza tra due grandezze misurate consideriamo (per ora) il parallelismo con le stime per una singola variabile aleatoria.

Stima del valore di aspettazione della variabile aleatoria associata al lancio di un dado.

1. Per calcolare il valore di aspettazione (media) eseguo N lanci e calcolo la media.

   N=21; 1 + 4 + 4 + 6 + 2 + 3 + 2 + 5 + 1 + 4 + 6 + 5 + 3 + 1 + 6 + 5 + 2 + 3 + 6 + 1 + 2 = 72; 72/21=3.42

2. Per calcolare il valore di aspettazione calcolo le frequenze di uscita dei singoli valori negli N lanci.

   N=21; f_1 =4/21; f_2 =4/21; f_3 =3/21; f_4 =3/21; f_5 =3/21; f_6 =4/21; 1* f_1 + 2* f_2 + 3* f_3 + 4* f_4 + 5* f_5 + 6* f_6 = 3.42

3. Per calcolare il valore di aspettazione assumo nota la distribuzione di probabilità del processo.

$$\mu_x = \sum x_i p_i;$$

$$p_i = \frac{1}{6}$$

$$\mu_x = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

FUNZIONI DI DUE VARIABILI ALEATORIE

Stima della Varianza e Covarianza da un campione

Varianza di un campione

$$\mathrm{var}(x) = \int_{-\infty}^{+\infty} (x-\mu_x)^2 f(x)dx$$

$$\mathrm{var}(x) = \frac{\sum (x_i-\mu_x)^2}{N}$$

Covarianza tra due grandezze.

$$\mathrm{cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (y-\mu_y)(x-\mu_x) f(x,y)dxdy$$

$$\mathrm{cov}(X,Y) = \frac{\sum (y_i-\mu_y)(x_i-\mu_x)}{N}$$

TEORIA DEGLI ERRORI DI MISURA

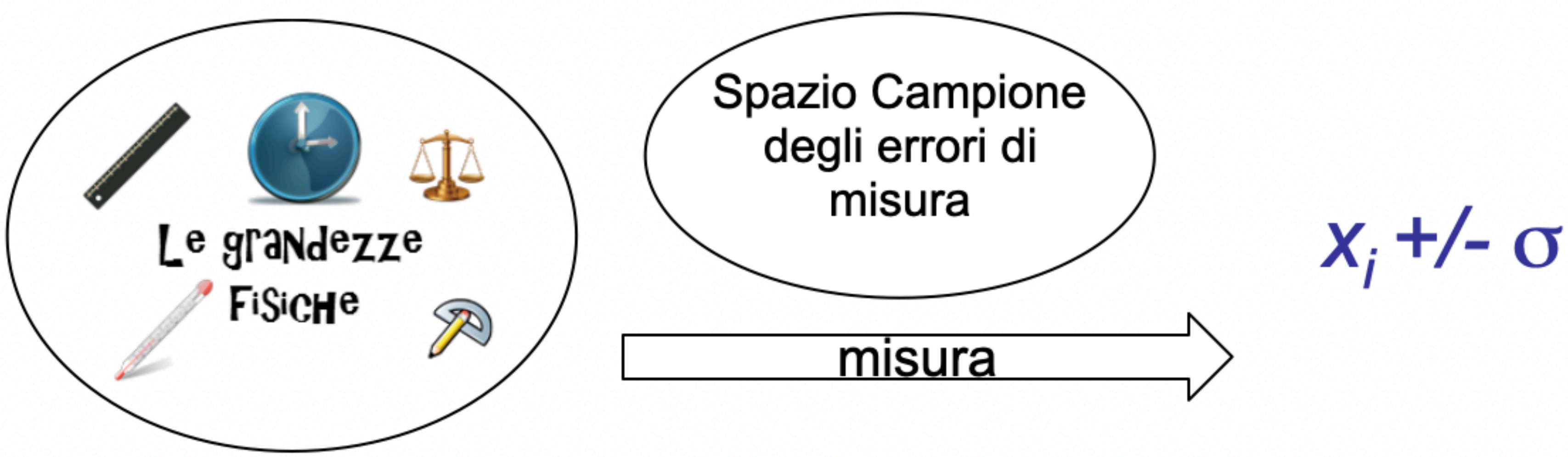
Nella teoria degli errori si assume che la misura
i-esima di una grandezza fisica il cui valore vero è μ, in conseguenza di indeterminabili errori di misura casuali dovuti al metodo di misura stesso e al processo fisico in studio, assuma il valore xi=μ+δi, dove δi è l'errore associato alla misura i-esima.

Quando eseguo una misura x_i di una grandezza fisica con valore vero μ so che in conseguenza del processo di misura questa grandezza sarà affetta da un errore δ_i il quale è un elemento dello spazio campione degli errori le cui caratteristiche dipendono dal metodo di misura adottato.

TEORIA DEGLI ERRORI DI MISURA

Quando dico che il valore misurato della grandezza fisica è x_i +/- σ *con* σ *errore statistico sto facendo alcune assunzioni.*

Sto assumendo che ***lo spazio campione degli errori di misura segue una distribuzione di probabilità nota*** *e che in un qualche modo* ***sono in grado di conoscerne la sigma****. Sto inoltre implicitamente assumendo che tale spazio campione abbia* ***valore di aspettazione nullo****.*



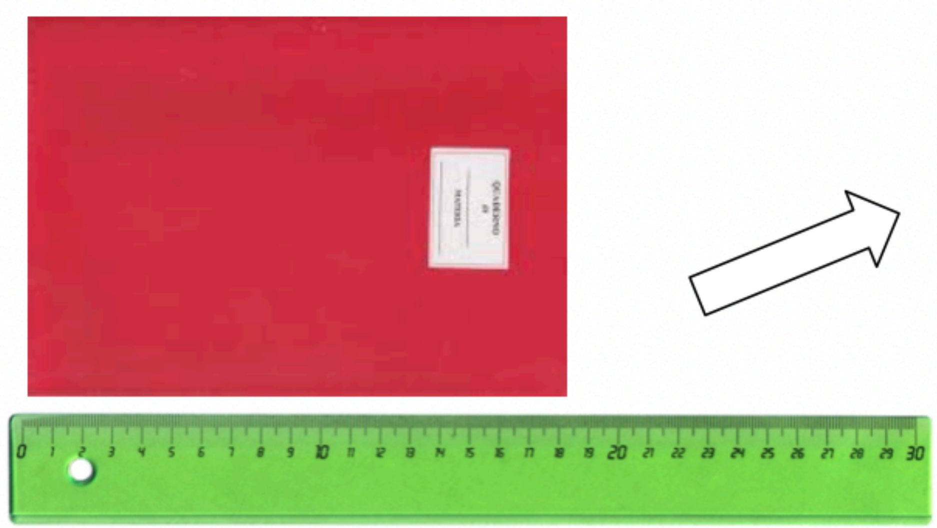
L'assumere che il valore di aspettazione della popolazione "*errori di misura*" associata al processo di misura in studio sia a valore di aspettazione nullo equivale a dire che la misura non sia affetta da "*errori sistematici*".

TEORIA DEGLI ERRORI DI MISURA

Se la misura è unica significa che ho assunto la misura stessa come stima del valore di aspettazione della popolazione.

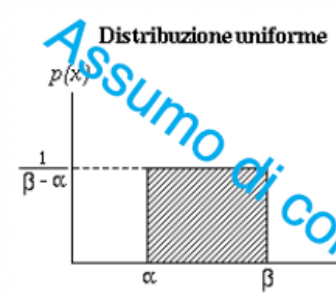
$x=\mu+\delta$

Posso considerare x *come la variabile aleatoria associata al processo di misura dipendente da un'altra variabile aleatoria* (δ) *attraverso una relazione lineare.*
$E(x)=E(\mu+\delta)=\mu+E(\delta)=\mu$ se la popolazione errori di misura ha valore di aspettazione nullo.



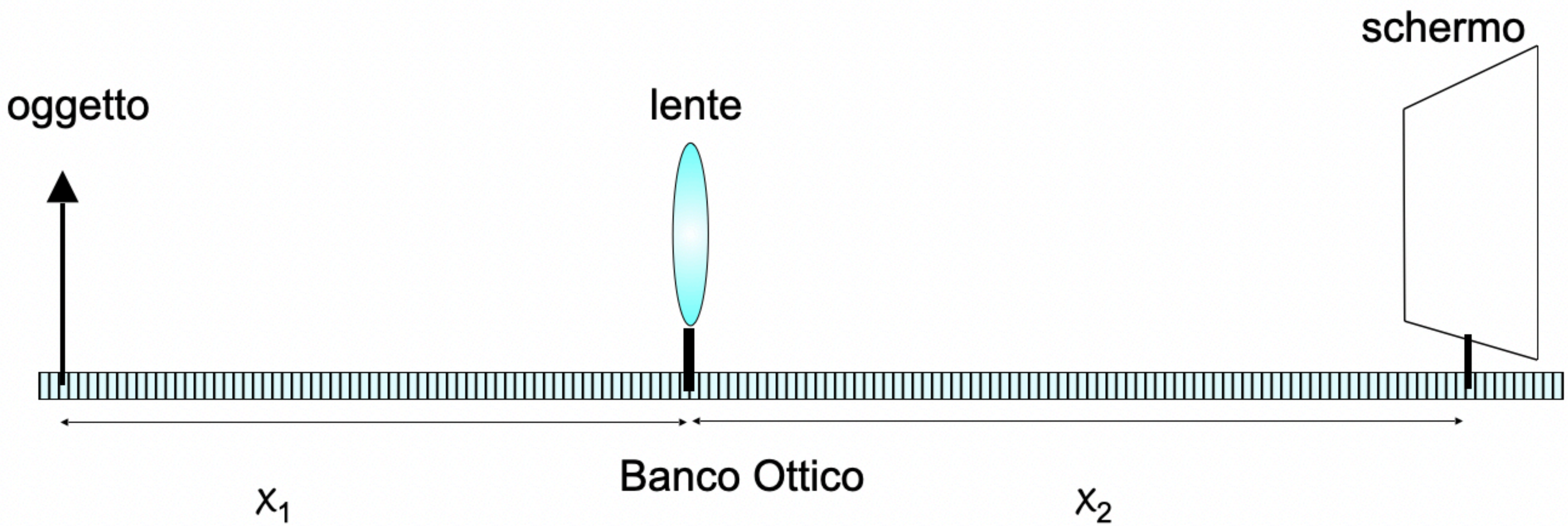
Spazio Campione degli errori di misura

18.2 +/- 0.3 cm



Assumo di conoscerlo

Esempio: Misura della lunghezza focale di una lente.



Per determinare la lunghezza focale della lente occorre misurare simultaneamente la distanza dalla lente dell'oggetto (x_1) e la distanza dalla lente dell'immagine (x_2).

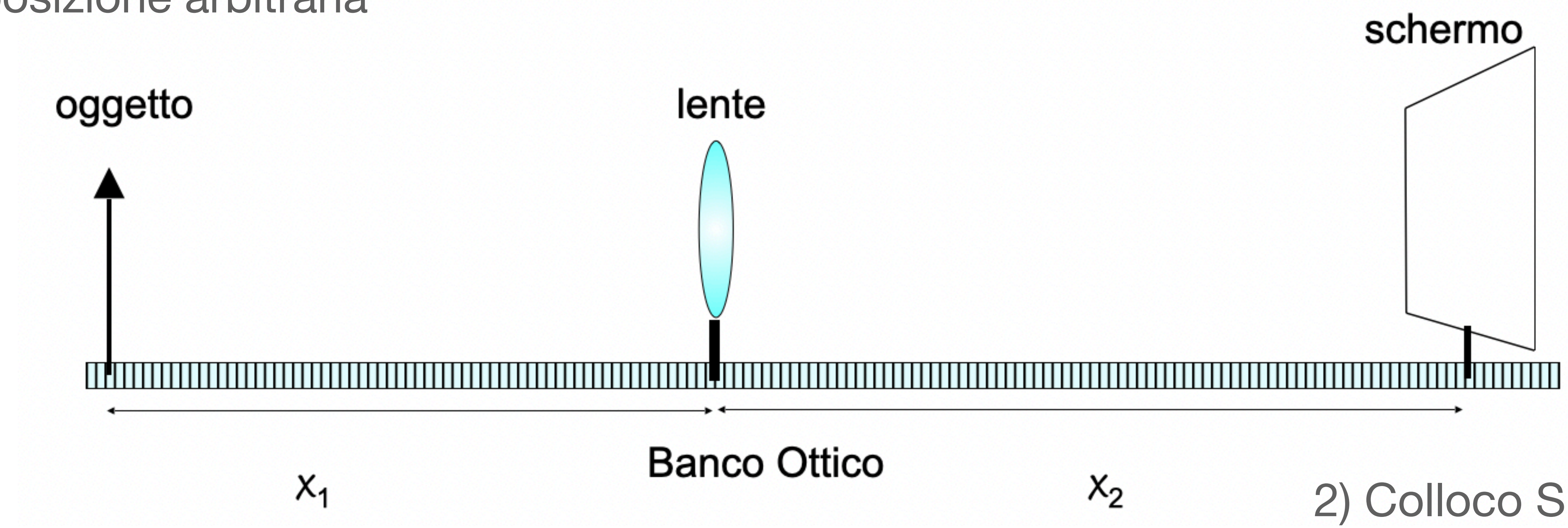
$$\frac{1}{f} = \frac{1}{x_1} + \frac{1}{x_2}$$

f lunghezza focale

Dipartimento di Matematica e Fisica "Ennio De Giorgi"

Procedura

1) Colloco O in posizione arbitraria



2) Colloco S in posizione arbitraria

3) Colloco L al centro dell'intervallino (ampio 2 cm) che consente di focalizzare I sullo schermo

TEORIA DEGLI ERRORI DI MISURA

Errori su x_1 e x_2:

Determino la posizione dell'oggetto P_o, la posizione della lente P_l e la posizione dello schermo P_s.
Calcolo x_1 e x_2 per differenza tra le posizioni.
Ogni posizione ha un suo errore di misura. Calcolo con la propagazione l'errore su le distanza oggetto-lente e lente-schermo.

$$P_0 = 0.00 \pm 0.05 \quad cm$$

$$P_l = 30 \pm 1 \quad cm$$

$$P_s = 72.00 \pm 0.05 \quad cm$$

$$x_1 = 30.0 \pm 0.3$$

$$x_2 = 42.0 \pm 0.3$$

$$f = \frac{x_1 x_2}{x_1 + x_2} = 17.5$$

$$\sigma_f^{\,2} = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma^2_{x_1} + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma^2_{x_2} \qquad \boxed{???}$$

Errori massimi ed errori statistici

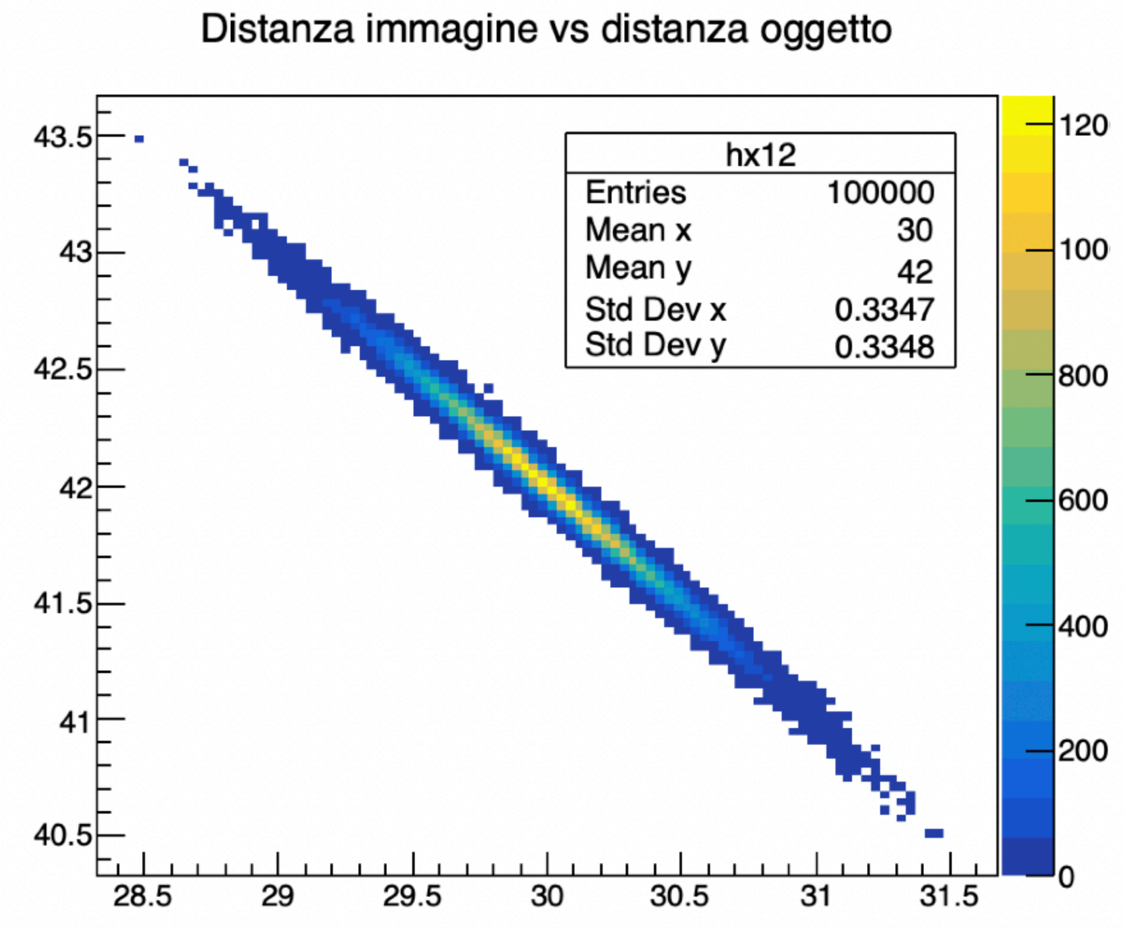
$$\sigma_{P_o} = \frac{0.05}{\sqrt{3}} = 0.03 \qquad \sigma_{P_s} = \frac{0.05}{\sqrt{3}} = 0.03 \qquad \sigma_{P_l} = \frac{1}{3}$$

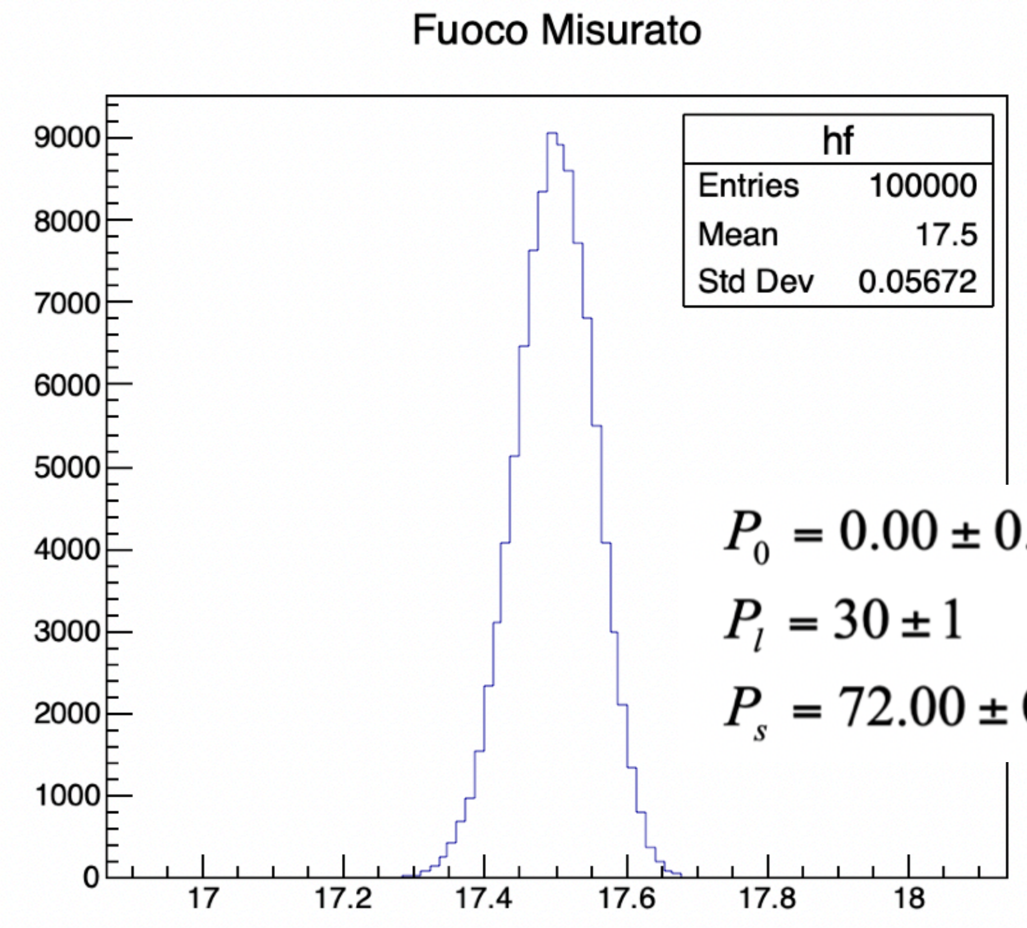
$$\sigma_{x_1} = 0.3 \qquad \sigma_{x_2} = 0.3$$

$$\sigma_f^2 = \left(\frac{x_2(x_1+x_2) - x_1 x_2}{(x_1+x_2)^2}\right)^2 \sigma_{x_1}^2 + \left(\frac{x_1(x_1+x_2) - x_1 x_2}{(x_1+x_2)^2}\right)^2 \sigma_{x_2}^2$$

$$\sigma_f \cong 0.13$$

TEORIA DEGLI ERRORI DI MISURA





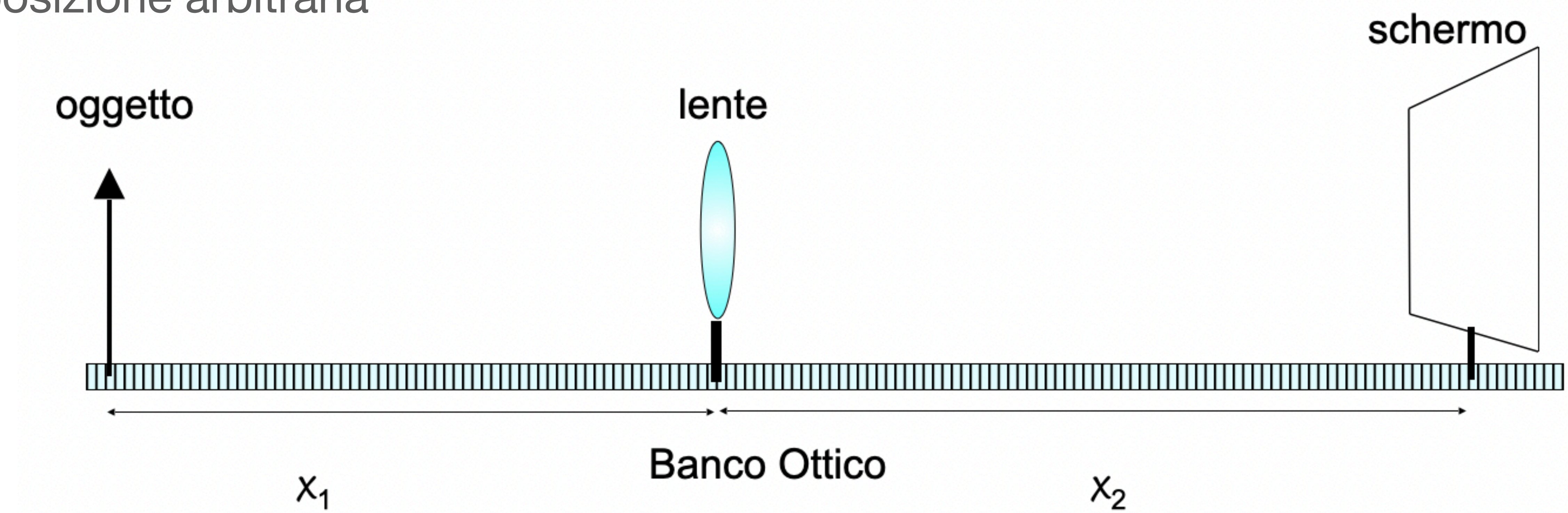
$$P_0 = 0.00 \pm 0.05 \quad cm$$
$$P_l = 30 \pm 1 \quad cm$$
$$P_s = 72.00 \pm 0.05 \quad cm$$

Calcolo Analitico
x1=30.000000 +/- 0.334581
x2=42.000000 +/- 0.334581
fuoco=17.500000
sigma=0.127812 var(f)=Df_x1*var(x1)+Df_x2*var(x2))=**0.016336** <<<<<———— ignorando la correlazione
Verifica Simulazione
media di X1=29.998597, sigma(X1)=0.334730 e Var(x1)=0.112044
media di X2=42.001437, sigma(X2)=0.334760 e Var(x2)=0.112064
covarianza -0.111218
Calcolo della varianza e sigma di F tenendo conto delle correlazione
sigma=0.056528 var(f)=Df_x1*var(x1)+Df_x2*var(x2)+2Df_x1*Df_x2*cov(X1,X2)=**0.003195**
Variabile aleatoria derivata. Risultati Simulazione
media di f=17.498222, sigma(f)=**0.056716** e Var(f)=**0.003217**

TEORIA DEGLI ERRORI DI MISURA

Una diversa procedura

1) Colloco O in posizione arbitraria



2) Colloco L in posizione arbitraria

3) Colloco S al centro dell'intervallino (ampio 2 cm) che consente di focalizzare I sullo schermo

NOTA: ora $\sigma(x_2) >> \sigma(x_1) >> \sigma(P_L)$

Errori massimi ed errori statistici

σ(P_L) σ(P_S)

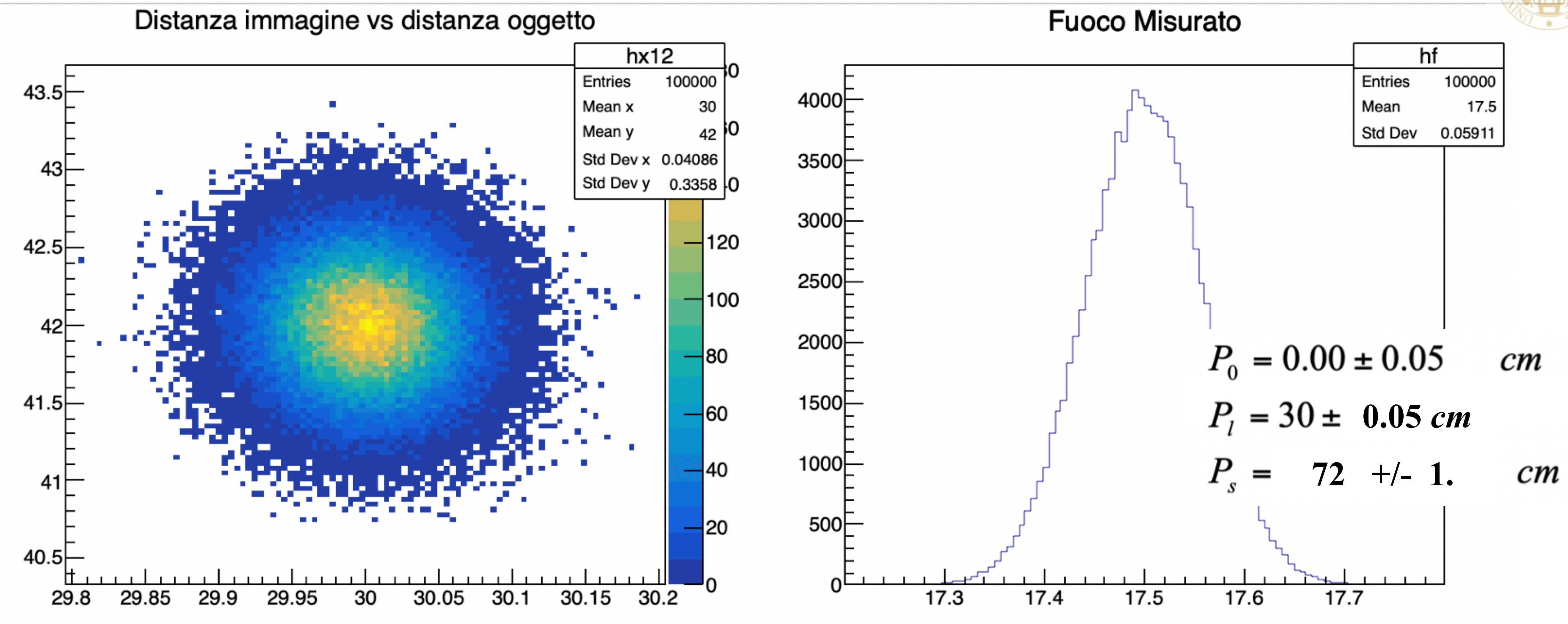
$$\sigma_{P_o} = \frac{0.05}{\sqrt{3}} = 0.03 \qquad \sigma_{P_s} = \frac{0.05}{\sqrt{3}} = 0.03 \qquad \sigma_{P_l} = \frac{1}{3}$$

$$\sigma_{x_1} = 0.045 \qquad \sigma_{x_2} = 0.3$$

$$\sigma_f^2 = \left(\frac{x_2(x_1+x_2) - x_1 x_2}{(x_1+x_2)^2}\right)^2 \sigma_{x_1}^2 + \left(\frac{x_1(x_1+x_2) - x_1 x_2}{(x_1+x_2)^2}\right)^2 \sigma_{x_2}^2$$

$$\sigma_f \cong 0.06$$

TEORIA DEGLI ERRORI DI MISURA



Calcolo Analitico
x1=30.000000 +/- 0.040825
x2=42.000000 +/- 0.334581
fuoco=17.500000
sigma=**0.059725** var(f)=Df_x1*var(x1)+Df_x2*var(x2))=***0.003567***
Verifica Simulazione
media di X1=29.999951, sigma(X1)=0.040861 e Var(x1)=0.001670
media di X2=41.999606, sigma(X2)=0.335764 e Var(x2)=0.112738
covarianza -0.000833 <<<<———————————————————— *piccola covarianza*
Calcolo della varianza e sigma di F tenendo conto delle correlazione
sigma=0.058896 var(f)=Df_x1*var(x1)+Df_x2*var(x2)+2Df_x1*Df_x2*cov(X1,X2)=***0.003469***
Variabile aleatoria derivata. Risultati Simulazione
media di f=17.499630, sigma(f)=**0.059107** e Var(f)=***0.003494***