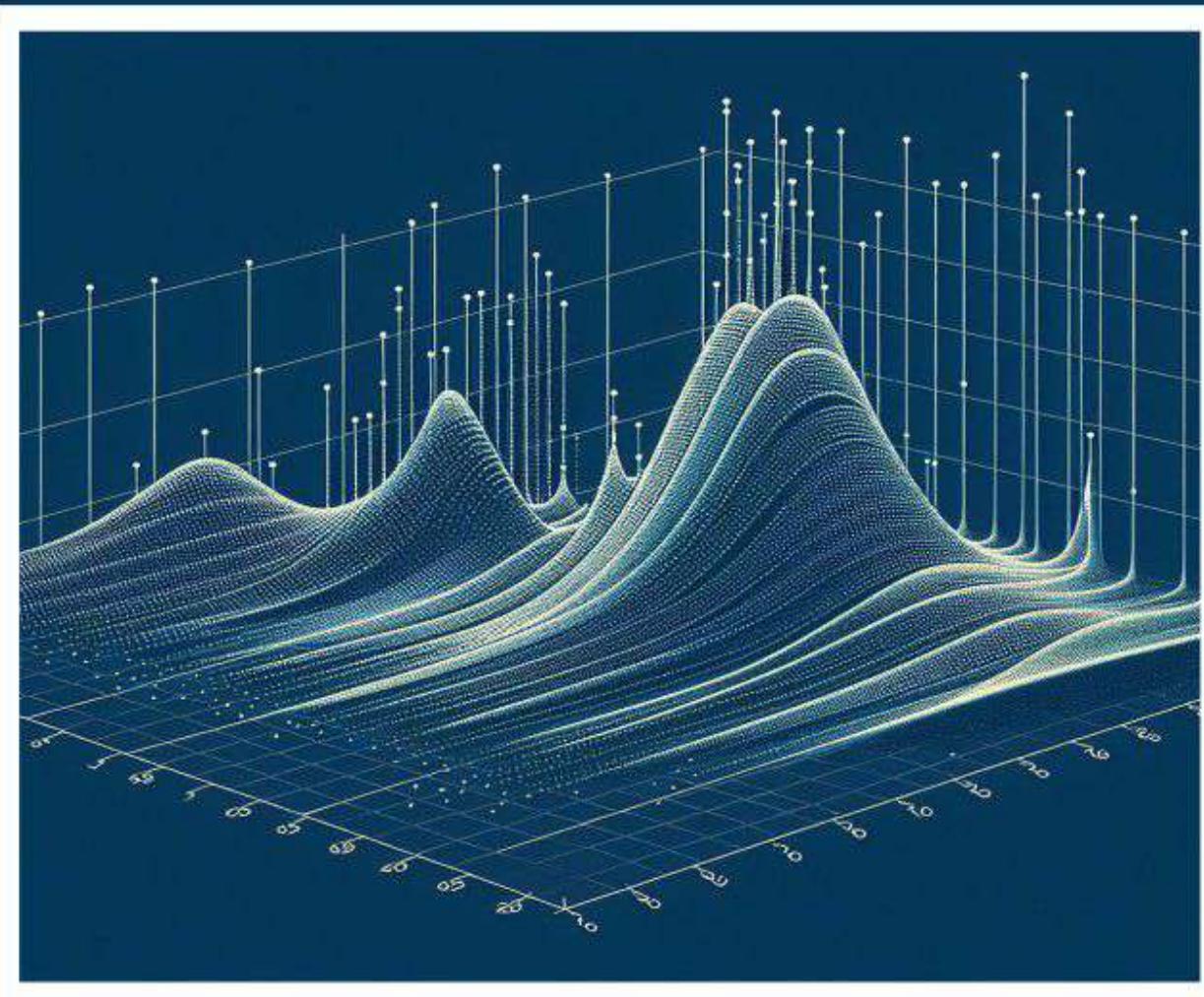


# *Statistica* *&* *Probabilità*



*Corso di laurea triennale in Fisica*

*Di Lernia Giulio*

---

---

## Contents

---

<b>1 La probabilità e la distribuzione di probabilità</b>	<b>9</b>
1.1 Assiomi di Kolgomorov . . . . .	9
1.1.1 Definizione di probabilitá frequentista . . . . .	9
1.1.2 Teorema di Bayes . . . . .	10
1.2 Variabili aleatorie . . . . .	11
1.3 Probability Distribution Function (P.d.f) . . . . .	12
1.3.1 Cumulative distribution function (C.d.f) . . . . .	12
1.3.2 Istogrammi . . . . .	13
1.3.3 Proprietà di una probability distribution function . . .	14
1.3.4 Valore di aspettazione di una probability distribution function . . . . .	15
1.3.5 Momenti principali di una pdf per la popolazione e per un campione . . . . .	15
1.4 Trasformazioni di distribuzioni di probabilità . . . . .	18
1.5 Assenza di memoria per una distribuzione di probabilità . . .	20
<b>2 Modelli di probabilità</b>	<b>23</b>
2.1 La distribuzione Uniforme . . . . .	23
2.1.1 Proprietà della pdf . . . . .	23
2.1.2 La distribuzione Gaussiana (o Normale) . . . . .	24
2.1.3 Distribuzione cumulativa di una Gaussiana . . . . .	25
2.1.4 Teorema Centrale del Limite . . . . .	26
2.2 Distribuzione di probabilità Log-normale . . . . .	27
2.2.1 Proprietà della log-normale . . . . .	27
2.3 Distribuzione di Probabilità di Bernoulli . . . . .	27
2.3.1 Proprietà della distribuzione Bernoulliana . . . . .	27

## CONTENTS

---

2.4	Distribuzione di Probabilità Binomiale . . . . .	28
2.4.1	Proprietà della distribuzione binomiale . . . . .	28
2.4.2	La densità di probabilitá di un’istogramma . . . . .	29
2.5	Distribuzione Multinomiale . . . . .	30
2.6	Distribuzione di Poisson . . . . .	31
2.6.1	Proprietà della distribuzione di Poisson . . . . .	31
2.6.2	Da eventi Poissoniani all’esponenziale . . . . .	32
2.7	Distribuzione Esponenziale . . . . .	33
2.7.1	Proprietá della distribuzione esponenziale . . . . .	35
2.7.2	Decadimento Radioattivo . . . . .	35
2.8	Comportamenti Asintotici delle distribuzioni . . . . .	36
2.8.1	Proprietà asintotiche della Binomiale e Poissoniana .	36
2.9	Distribuzione di $\chi^2$ . . . . .	38
<b>3</b>	<b>Distribuzioni di probabilità multidimensionali</b>	<b>41</b>
3.1	Probabilità per variabili aleatorie in più dimensioni . . . . .	41
3.2	Distribuzione di probabilità marginale . . . . .	42
3.3	Distribuzione di probabilità condizionata . . . . .	42
3.4	Valore di aspettazione di una joint pdf . . . . .	43
3.5	Varianza di una pdf multidimensionale . . . . .	43
3.6	Correlazione . . . . .	44
3.7	Variabili Statisticamente indipendenti . . . . .	46
3.8	Cambiamento di variabili per una pdf multidimensionale . . . . .	46
3.8.1	Propagazione degli errori . . . . .	47
3.8.2	Decorrelazione delle variabili . . . . .	48
3.9	Distribuzione di Gauss Multidimensionale . . . . .	48
<b>4</b>	<b>Stime di Parametri</b>	<b>51</b>
4.1	La statistica . . . . .	51
4.2	Stimatori . . . . .	51
4.3	Proprietà degli stimatori . . . . .	52
4.3.1	Precisione e Accuratezza . . . . .	54

---

## CONTENTS

4.3.2	Incertezze sulle stime . . . . .	55
4.3.3	La Varianza come stimatore . . . . .	56
4.3.4	Correzione di Bessel . . . . .	57
<b>5</b>	<b>Maximum Likelihood e Minimi Quadrati</b>	<b>61</b>
5.1	La Verosomiglianza - Likelihood . . . . .	61
5.2	Comportamento della likelihood . . . . .	62
5.3	Minimum Variance Bound . . . . .	63
5.3.1	Informazione di Fischer . . . . .	63
5.3.2	Teorema di Rao - Cramér . . . . .	64
5.4	Maximum Likelihood . . . . .	64
5.5	Varianza dello stimatore di ML - Metodo Grafico . . . . .	66
5.6	Extended likelihood . . . . .	68
5.7	Varianza di uno stimatore per più parametri . . . . .	68
5.8	Intervallo di confidenza . . . . .	68
5.9	Metodo dei minimi quadrati . . . . .	69
5.9.1	Funzionale $Q^2$ . . . . .	70
5.9.2	Valore di aspettazione di $\hat{\mu}_{MQ}$ . . . . .	70
5.9.3	Varianza dello stimatore $\hat{\mu}_{MQ}$ . . . . .	71
5.10	Varianza di un stimatore usando i MQ - Metodo grafico . . .	71
5.11	Modelli lineari nei parametri . . . . .	73
5.11.1	Stime dei parametri . . . . .	74
5.12	Sovrastima degli errori . . . . .	76
5.12.1	Relazione tra il numero di parametri e il campione . .	76
5.12.2	Incertezze sulla variabile indipendente . . . . .	76
5.12.3	Stima del fattore di sovrastima $\alpha$ . . . . .	77
5.13	Teorema di Gauss-Markov . . . . .	77
5.14	Interpolazione ed Estrapolazione . . . . .	78
5.15	Fit d'istogrammi . . . . .	80
5.15.1	Binned Data - Minimi Quadrati . . . . .	82
5.15.2	Binned Data - Maximum Likelihood . . . . .	83
<b>6</b>	<b>Test d'ipotesi</b>	<b>85</b>
6.1	Test del $\chi^2$ . . . . .	85

## CONTENTS

---

6.1.1	Applicazione del test di $\chi^2$	88
6.2	Errori di test statistici	88
6.2.1	Errori del $I^\circ$ tipo	88
6.2.2	Errori del $II^\circ$ tipo	89
6.3	Test di Kolgomorov-Smirnov	90
6.3.1	Costruzione del test	90
6.4	Confronto di una misura con il valore di riferimento	91
6.4.1	Distribuzione di t-student	91
6.5	Intervalli di confidenza	94
6.5.1	Metodo della cintura di confidenza	94
6.6	Discovery Significance	96

---

## Bibliography

---

- [1] Slide del corso di Laboratorio di calcolo e statistica del secondo anno di Fisica, università di Milano Bicocca, Maura Pavan, 2022/2023
- [2] Appunti manoscritti presi durante le lezioni frontali tenute dal prof. Pietro Govoni per il corso di Laboratorio di calcolo e statica del secondo anno di Fisica, Università di Milano Bicocca
- [3] Data Analysis Technique for Physical Scientist, Pruneau, Cambridge Univeristy Press, 2017
- [4] Statistical Data Analysis, Glen Cowan, Clarendon Press, 1998

## BIBLIOGRAPHY

---

# CHAPTER 1

---

## La probabilità e la distribuzione di probabilità

---

### 1.1 Assiomi di Kolgomorov

Come si definisce la probabilitá secondo Kolgomorov?

Consideriamo degli eventi  $E_1, E_2 \subset \Omega$  spazio campione, vogliamo costruire definiamo probabilitá una un funzione  $P : \Omega \rightarrow [0, 1]$  che soddisfa le seguenti proprietá:

- $P(\Omega) = 1$
- $\forall A \subset \Omega$  si ha che  $P(A) \geq 0$
- $\forall A, B \subset \Omega$ , si ha che  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . Se gli eventi sono indipendenti  $A \cap B = \emptyset$ , si ha che  $P(A \cup B) = P(A) + P(B)$ .

#### 1.1.1 Definizione di probabilitá frequentista

Come si definisce operativamente la probabilitá secondo la formulazione frequentista?

La probabilitá di un evento A é definita come il rapporto tra il numero di casi favorevoli (in cui avviene l'evento A) e il numero di casi possibili (la popolazione). Consideriamo di avere **N** eventi, e di contare il numero di volte in cui l'evento A, indicandolo come **n(A)**. Definiamo la probabilitá di

## CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI PROBABILITÀ

---

un evento come:

$$\lim_{N \rightarrow \infty} \frac{n(A)}{N} \quad (1.1)$$

dove il limite è inteso in senso probabilistico ovvero al crescere del numero degli eventi.

### Probabilità condizionata

Che cos'è la probabilità condizionata ?

Consideriamo una coppia di eventi  $A, B \subset \Omega$  definiamo **probabilità condizionata** la probabilità che si verifichi l'evento A al verificarsi previamente dell'evento B.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1.2)$$

si dimostra che la probabilità condizionata soddisfa gli assiomi di Kolgomorov.

Nel caso di due eventi indipendenti, ovvero che il verificarsi di B non influenza la probabilità che si verifichi A

$$P(A|B) = P(A|\Omega) = P(A)$$

per due eventi indipendenti possiamo riscrivere la probabilità condizionata come:

$$P(A \cap B) = P(A) \cdot P(B) \quad (1.3)$$

#### 1.1.2 Teorema di Bayes

Che cosa afferma il teorema di Bayes ?

Consideriamo due eventi  $A, B \subset \Omega$  la probabilità nel caso di due eventi non indipendenti possiamo riscrivere la probabilità condizionata come:

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

---

## CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI PROBABILITÀ

---

dunque l'equazione (1.2) puó essere riscritta come :

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (1.4)$$

tale espressione definisce il **teorema di Bayes**. La probabilitá  $P(A|B)$  viene detta "a posteriori" poiché permette di calcolare la probabilitá di ,sapendo che si é verificato (o si verificherá con certezza assoluta) B. La probabilitá  $P(A)$  si dice invece "a priori" poiché non condizionata da alcun altro evento o da alcuna conoscenza che potremmo avere sul suo verificarsi.

## 1.2 Variabili aleatorie

Che cos' una variabile aleatoria ?

Una **variabile aleatoria** é una funzione definita sullo spazio campione a cui viene associato un sottoinsieme misurabile.

$$X : \Omega \rightarrow E \quad (1.5)$$

dove  $\Omega$  é lo spazio di tutti gli esiti possibili, mentre  $E$  é l'insieme degli esiti al verificarsi di un determinato evento.

Nel caso in cui  $\Omega \subseteq \mathbb{N}$  si ha che  $x$  viene definita **variabile casuale discreta**.

Se  $\Omega \subseteq \mathbb{R}$ ,  $x$  viene definita **variabile aleatoria continua**, in questo caso la probabilitá viene misurata su intervalli  $P(a < x < b)$  e non pi sui singoli eventi come nel caso di quella discreta.

**Esempio:**

- Il conteggio del numero di volte in cui compare un certo valore di una misura é una variabile aleatoria discreta.
- La temperatura é una variabile aleatoria continua.

## CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI PROBABILITÀ

---

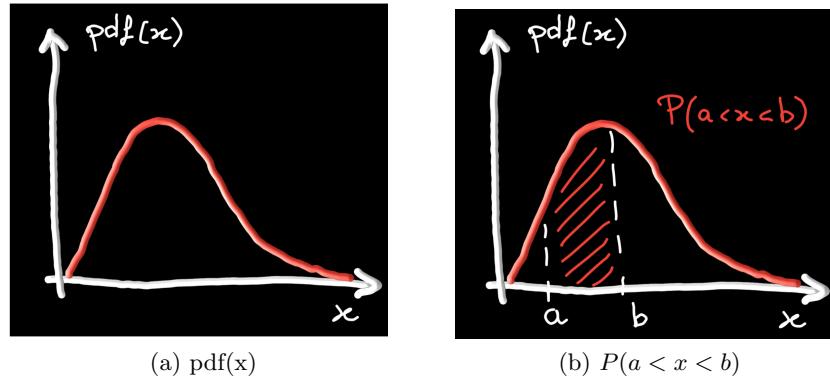
### 1.3 Probability Distribution Function (P.d.f)

Che cosa é una pdf e quali sono le sue proprietá ?

Si definisce densitá di probabilitá (p.d.f) una funzione che soddisfa le seguenti proprietá:

1.  $pdf(x) : \Omega \subseteq \mathbb{R} \rightarrow \mathbb{R}^+$
2.  $P(x < \hat{x} < x + dx) = pdf(x) \cdot dx$
3.  $P(a < x < b) = \int_a^b pdf(x) \cdot dx \leq 1$
4.  $P(x) = 0$ , la probabilitá in una singola misura é nulla.

per la definizione di porbabilitá nei punti 2) e 3) valgono gli assiomi di Kolgomorov.



#### 1.3.1 Cumulative distribution function (C.d.f)

Che cosé una c.d.f.?

Si definisce distribuzione cumulativa la primitiva dell'integrale della pdf di una variabile aleatoria continua  $x$ :

$$cdf(x) = \int_a^x pdf(x)dx \quad (1.6)$$

## CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI PROBABILITÀ

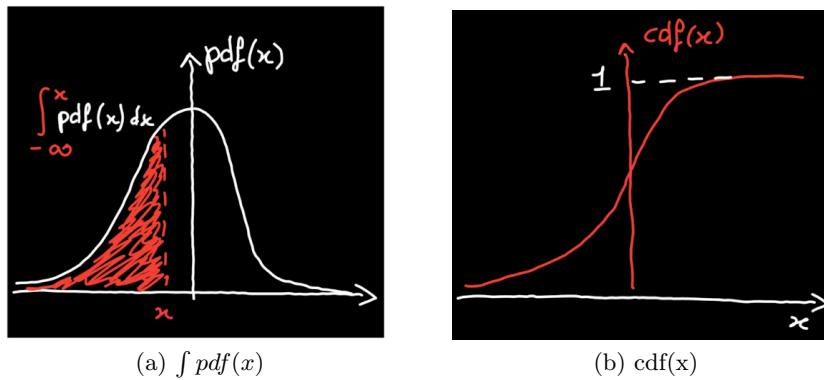
ovvero si ha che:

$$\frac{d}{dx} cdf(x) = pdf(x) \quad (1.7)$$

di conseguenza la  $cdf(x)$  descrive l'evolversi della probabilità di una variabile casuale.

### Corollario

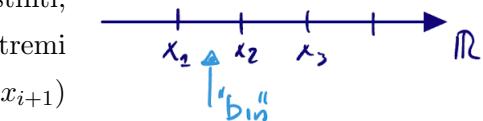
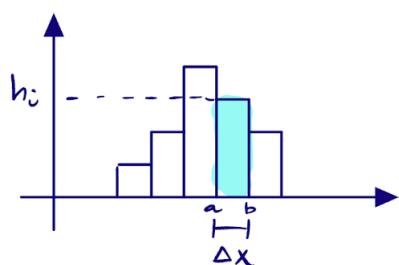
Esiste una forma analitica della  $cdf(x)$  di una r.v.  $\iff$  la pdf ammette primitiva.



### 1.3.2 Istogrammi

Che cos'è un istogramma ?

Consideriamo una variabile aleatoria  $x$ , di cui conosciamo  $\{x_i\}_i^N$  valori distinti, utilizziamo tali grandezze come estremi per definire degli intervalli  $\omega_k = [x_i, x_{i+1})$  disgiunti tra loro  $\omega_k \cap \omega_m = \emptyset$  per  $k \neq m$ , sull'asse reale. Tali intervalli vengono definiti **bin**.



Ripetendo le misure contiamo il numero di volte  $n_l$  che un valore compare e cade all'interno di un'intervalle  $\omega_k$ .

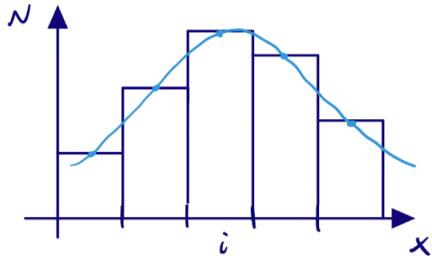
## CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI PROBABILITÀ

---

In questo modo posizionando sull'asse delle ordinate il numero di conteggi  $n_l(x)$  si costruisce un istogramma.

Come si può utilizzare un istogramma per rappresentare una pdf ?

All'aumentare del numero di elementi  $N \rightarrow \infty$  e diminuendo l'ampiezza dei bin  $\Delta x \rightarrow 0$ , ci si riconduce alla nozione di probabilità (1.1), in questo modo si passa a una  $pdf(x)$  continua. Al crescere del numero di eventi e il numero di eventi cambia e di conseguenza cambia la frequenza relativa.



$$\frac{f_{k+1} - f_k}{\Delta x} = \frac{df}{dx} \Rightarrow df = \frac{(f_{k+1} - f_k)}{\Delta x} dx \quad (1.8)$$

di conseguenza  $p \rightarrow \frac{d}{dx} f(x) dx$ .

### 1.3.3 Proprietà di una probability distribution function

Che cosa sono i momenti di una pdf?

Conoscere l'espressione analitica di una pdf è spesso poco significativo (soprattutto se la sua espressione non lascia intuire la forma della curva) in altri casi è impossibile definirla. Di conseguenza per studiare il comportamento di una pdf facciamo affidamento ad alcune grandezze che ne descrivono il comportamento, queste prendono il nome di **momenti**.

Tali momenti possiamo classificarli come:

- $E[x^m] = \int x^m \cdot pdf(x) dx$  prendono il nome di **momenti centrali di ordine m**.
- $E[(x - \mu)^m] = \int (x - \mu)^m \cdot pdf(x) dx$  prendono il nome di **momenti centrali**

## CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI PROBABILITÀ

---

il momento centrale di ordine 1 vale  $E[(x - \mu)] = 0$ .

### 1.3.4 Valore di aspettazione di una probability distribution function

Che cosa è il valore di aspettazione di una quantità rispetto ad una pdf ?

Data  $x \in \Omega$  variabile aleatoria e  $u(x) : \mathbb{R} \rightarrow \mathbb{R}$ , il valore di aspettazione di  $u(x)$  è definito come:

$$E[u(x)] = \int u(x) \cdot pdf(x) dx \quad (1.9)$$

dove  $E[ ]$  è un operatore lineare:

- $E[x + y] = E[x] + E[y]$
- $E[\alpha x] = \alpha E[x] \quad \forall \alpha \in \mathbb{R}$

### 1.3.5 Momenti principali di una pdf per la popolazione e per un campione

Che cosa sono e che cosa rappresentano la media, la varianza, l'assimetria e la curtosi di un campione ? e di una popolazione?

Sia  $x \in \Omega$  una variabile aleatoria continua definiamo valore di aspettazione **media della popolazione** la grandezza:

$$\mu \equiv E[x] = \int x \cdot pdf(x) dx \quad (1.10)$$

Essa rappresenta il centro della distribuzione di probabilità .

Per un campione di  $N$  misure **la media** viene definita come:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (1.11)$$

Si definisce **varianza della popolazione**:

$$\sigma^2 \equiv V[x] = E[(x - \mu)^2] = \int (x - \mu)^2 \cdot pdf(x) dx \quad (1.12)$$

## CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI PROBABILITÀ

---

la sua radice prende il nome di **deviazione standard** e definisce l'ampiezza della pdf. La varianza può essere riscritta come:

$$V[x] = E[x^2] + E[x]^2 = E[x^2] + \mu^2 \quad (1.13)$$

Per un campione di N misure la **varianza** è data da:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (1.14)$$

mentre la **deviazione standard** è di un campione è data da:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (1.15)$$

Tra i momenti centrali sono anche presenti la **skewness** per un campione di N misure:

$$\gamma_1 = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - \mu)^3}{\sigma^3} \quad (1.16)$$

e per la popolazione:

$$\gamma_1 = \frac{E[(x - \mu)^3]}{\sigma^3} \quad (1.17)$$

tale grandezza definisce la simmetria di una distribuzione di probabilità rispetto al valore centrale dato dal valore di aspettazione  $\mu$  per una popolazione. Se una distribuzione è simmetrica  $\gamma_1 = 0$ .

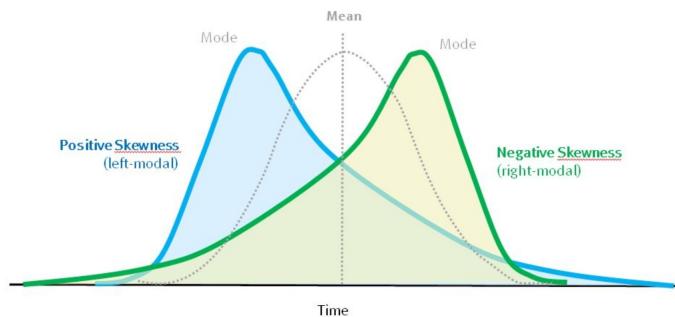


Figure 1.1: Simmetria di una distribuzione di probabilità

## CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI PROBABILITÀ

---

Si definisce indice di **kurtosi** di un campione di N elementi:

$$\gamma_2 = \frac{1}{N} \sum_{i=1}^N \frac{(x - \mu)^4}{\sigma^4} - 3 \quad (1.18)$$

mentre per la popolazione è data da:

$$\gamma_2 = \frac{E[(x - \mu)^4]}{\sigma^4} - 3 \quad (1.19)$$

tale stima fornisce informazioni sul peso delle code della distribuzione, nel senso che per code significative la distribuzione risulterà più schiacciata nell'intorno del valore  $\mu$ , mentre per code meno incidenti sarà più piccata.

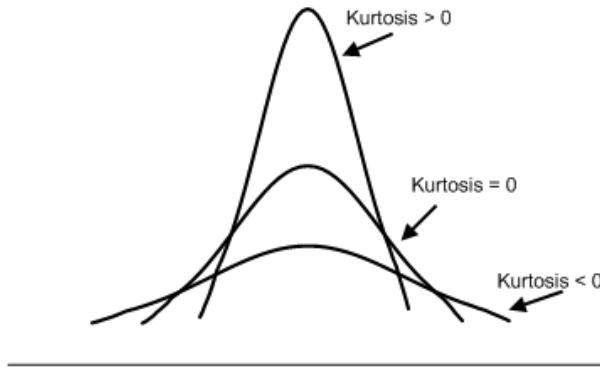
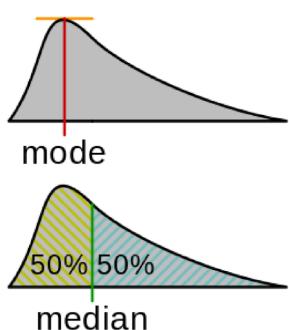


Figure 1.2: Kurtosi di una distribuzione di probabilità

### Stime di tendenza centrale

Che cosa sono moda, media e mediana di un campione? e di una popolazione?

Definiamo **moda** di una popolazione il valore che definisce il massimo della pdf.



Mentre si definisce **mediana** il valore che divide in due parti uguali la pdf. Nel caso dei campioni le grandezze diventano rispettivamente il valore che compare con maggior frequenza e la misura che divide in due parti eguali il campione.

## CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI PROBABILITÀ

---

Per media e mediana di una distribuzione sono unici mentre per una distribuzione può avere più mode, in questi casi si parla di distribuzione **multimodale**.

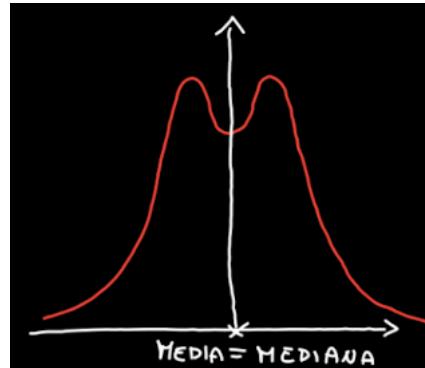


Figure 1.3: Distribuzione multimodale

## 1.4 Trasformazioni di distribuzioni di probabilità

Sia  $x$  una variabile aleatoria continua che segue una  $pdf(x)$  e ipotizziamo di voler costruire una nuova variabile aleatoria  $y$  legandola ad  $x$  mediante una funzione  $f(x)$ , come definiamo la  $pdf(y)$ ?

Sappiamo che la probabilità che una misura cada in un intervallo  $[a,b]$  è data da

$$P(a < x < b) = \int_a^b pdf(x)dx$$

Se effettuiamo un cambio di variabili  $y=f(x)$  vogliamo inanzitutto che la trasformazione sia biunivoca dunque  $f(x)$  deve essere monotona e continua. In questo modo applicando il teorema per il cambio di coordinate sotto segno d'integrale si ha che:

$$\frac{dy}{dx} = f'(x) \Rightarrow dx = f'(x)^{-1} dy \quad (1.20)$$

di conseguenza:

$$P(a < x < b) = \int_a^b pdf(x) dx = \int_{f(a)}^{f(b)} pdf(y) |f'(x)| dy$$

per comodità si è posto un modulo affinchè la funzione sia sempre positiva. In conclusione si ha che:

$$pdf(y) = pdf(x) \left| \frac{d}{dy} f^{-1}(y) \right| \quad (1.21)$$

Nel caso in cui abbia più di una dimensione è necessario che la funzione sia differenziabile e monotona e il modulo della derivata prima della funzione di cambio delle variabili viene sostituita dal modulo del determinante dell'inversa della matrice Jacobiana.

$$pdf(\mathbf{y}) = pdf(\mathbf{x}) \cdot |\det J| \quad (1.22)$$

Come si osserva dalle equazioni (1.21) e (1.22) la nuova  $pdf(y)$  aumenta o diminuisce di un fattore rispetto alla  $pdf(x)$ . Comunque sia la probabilità in un intervallo resta invariata.

Resta da discutere come cambiano i momenti della distribuzione, per farlo suddividiamo i risultati in due casi:

**Caso in cui il cambio di variabile è lineare  $y = ax+b$**

Il valore di aspettazione per una  $pdf(y)$  dove  $y = ax+b$  diventa:

$$E[y] = \int y \cdot pdf(y) dy = \int (ax + b) \cdot pdf(x) dx = y(\mu_x) \quad (1.23)$$

analogamente la varianza sarà data da:

$$V[y] = E[y^2] - E[y]^2 = a^2 \cdot V[x] \quad (1.24)$$

## CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI PROBABILITÀ

---

### Caso in cui il cambio di variabile è non lineare $y = f(x)$

il valore di aspettazione e la varianza di una  $pdf(y)$  per  $y = f(x)$  non lineare si stima approssimando con il polinomio di Taylor la funzione  $y(x)$  in un'intorno di  $\mu_x$  al secondo ordine:

$$y(x) \approx y(\mu_x) + \frac{dy}{dx} \Big|_{x=\mu_x} (x - \mu_x) + \frac{1}{2} \frac{dy^2}{dx^2} \Big|_{x=\mu_x} (x - \mu_x)^2$$

di conseguenza il valore di aspettazione è:

$$\begin{aligned} E[y] &= \int y \cdot pdf(y) dy = y(\mu_x) + \frac{dy}{dx} \Big|_{x=\mu_x} \int (x - \mu_x) \cdot pdf(x) dx + \\ &+ \frac{1}{2} \frac{dy^2}{dx^2} \Big|_{x=\mu_x} \int (x - \mu_x)^2 \cdot pdf(x) dx = \\ &= y(\mu_x) + \frac{1}{2} \frac{dy^2}{dx^2} \Big|_{x=\mu_x} \cdot \sigma_x^2 \end{aligned} \quad (1.25)$$

Mentre la varianza viene si ottiene dalla relazione:

$$V[y] = E[y^2] - E[y]^2 = \left( \frac{dy}{dx} \Big|_{x=\mu_x} \right)^2 \cdot \sigma_x^2 \quad (1.26)$$

L'espressione (1.26) è alla base della formula di propagazione degli errori.

## 1.5 Assenza di memoria per una distribuzione di probabilità

Gli eventi Poissoniani sono indipendenti l'uno dall'altro e la loro probabilità di decadimento è costante, dunque definita  $q(t)$  la probabilità che non si verifichi un evento in un intervallo di tempo  $[0,t]$  si ha che:

$$q(t + \delta t) = q(t)q(\Delta t) \quad (1.27)$$

## CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI PROBABILITÀ

---

la (1.27) definisce l'assenza di memoria di una distribuzione. Si può dimostrare che una distribuzione ha assenza di memoria  $\iff$  è esponenziale.

CHAPTER 1. LA PROBABILITÀ E LA DISTRIBUZIONE DI  
PROBABILITÀ

---

# CHAPTER 2

---

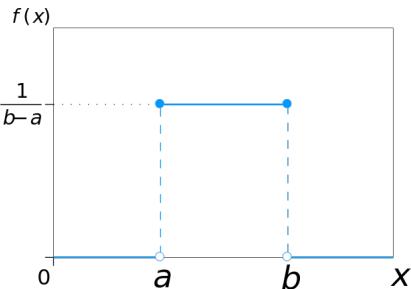
## Modelli di probabilità

---

### 2.1 La distribuzione Uniforme

La distribuzione di probabilità uniforme per una variabile aleatoria  $x$  continua è definita come:

$$pdf(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & altrimenti \end{cases}$$



Tale distribuzione definisce una distribuzione di probabilità uguale per tutti le variabili aleatorie  $x \in \Omega$ .

#### 2.1.1 Proprietà della pdf

Il valore atteso della distribuzione uniforme e la sua varianza sono date da:

$$E[x] = \int_{\alpha}^{\beta} \frac{x}{b-a} dx = \frac{b+a}{2} \quad (2.1)$$

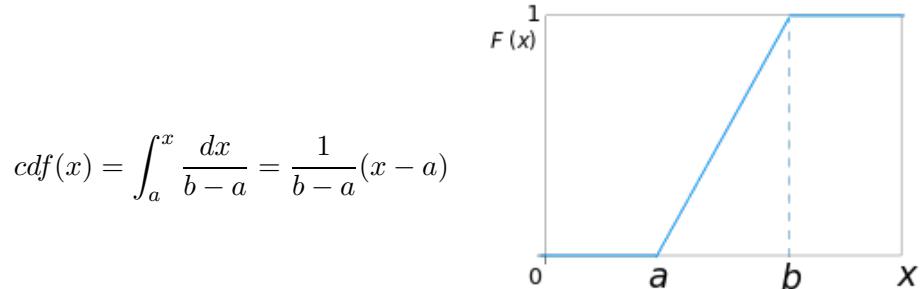
$$V[x] = \int_{\alpha}^{\beta} \frac{1}{b-a} \left(x - \frac{b+a}{2}\right)^2 dx = \frac{(b-a)^2}{12} \quad (2.2)$$

Per una distribuzione di questo tipo si ha che la simmetria e la kurtosi rispettivamente hanno valore nullo.

Inoltre tale pdf gode della proprietà di **riproduttività** ovvero date due variabili casuali  $x$  e  $y$  che seguono la pdf uniforme si ha che la nuova variabile aleatoria  $z = x + y$  segue una distribuzione di probabilità  $\text{pdf}(z)$  con media  $\mu_z = \mu_x + \mu_y$  e varianza  $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$ .

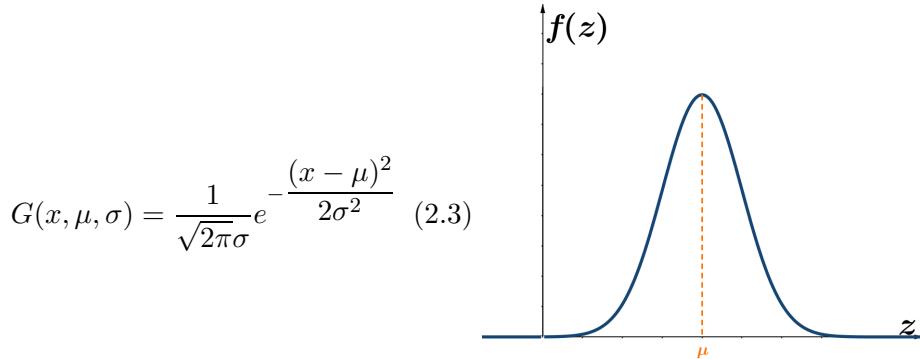
Un'altra proprietà importante della pdf è data dal fatto che se si ha una variabile aleatoria  $x$  che segue una distribuzione uniforme, la variabile casuale  $y = f(x)$  seguirà anch'essa la medesima distribuzione di probabilità.

Poichè la pdf possiede una forma analitica è possibile definirne la cdf:



### 2.1.2 La distribuzione Gaussiana (o Normale)

Data una variabile aleatoria  $x \in \Omega$  continua la distribuzione di probabilità Gaussiana è definita come:



Per una distribuzione Gaussiana si ha che la media, la moda e la mediana coincidono essendo una distribuzione simmetrica ovvero  $\gamma_1, \gamma_2 = 0$ . Media, Varianza e deviazione standard coincidono con quelle definite al capitolo precedente. Gode della proprietà di **riproduttività** definita nella sezione della distribuzione uniforme.

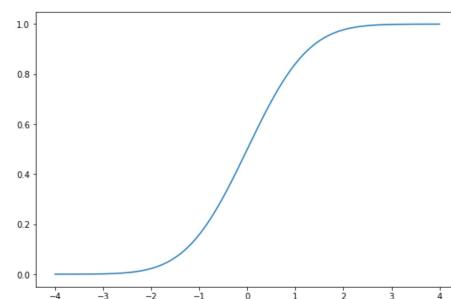
Una caso speciale di Gaussiana è quella **standardizzata** che è espressa come:

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

### 2.1.3 Distribuzione cumulativa di una Gaussiana

La pdf Gaussiana non ammette primitiva e la sua cdf viene calcolata numericamente e prende il nome di **funzione d'errore**, abbreviata con  $\text{erf}(x)$ .

$$\text{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}}$$



### 2.1.4 Teorema Centrale del Limite

La distribuzione Gaussiana è particolarmente importante poichè per essa vale un importante risultato.

#### Teorema

Si ipotizzi di avere N variabili aleatorie  $\{x_i\}_i^N$  indipendenti tra loro e ciascuna di esse segue una distribuzione di probabilità  $f_i(x)$ , e definiamo una nuova variabile aleatoria  $s = \sum_{i=1}^N x_i$  che segue una distribuzione pdf(s), allora si avrà che al crescere del numero di variabili:

$$\lim_{N \rightarrow \infty} \text{pdf}(s) \rightarrow G(s, \mu_s, \sigma_s)$$

dove  $\mu_s = \sum_{i=1}^N \mu_i$  e  $\sigma_s^2 = \sum_{i=1}^N \sigma_i^2$ .

#### Corollario della media

Date N variabili aleatorie  $\{x_i\}_i^N$  indipendenti ed identicamente distribuite, ovvero che per ogni variabile si ha la stessa distribuzione di probabilità  $f(x)$  e ipotizzando che  $\forall i$  si ha che  $\mu_i$  e  $\sigma_i^2$  siano finite, definita una nuova variabile aleatoria  $\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$  che segue una pdf  $f(\bar{x})$ , si ha che al crescere del numero di variabili:

$$\lim_{N \rightarrow \infty} f(\bar{x}) \rightarrow G(\bar{x}, \mu, \frac{\sigma}{\sqrt{N}})$$

dunque si ha che la pdf delle medie campionarie di un campione che cresce converge alla distribuzione Gaussiana.

#### Conseguenze del TCL

Quando effettuiamo una misura in un esperimento in generale possiamo definirla come la somma del valore vero  $x_0$  e di un errore  $\epsilon$  che è una variabile aleatoria.

$$x = x_0 + \epsilon$$

la misura così ottenuta segue una  $\text{pdf}(x)$ , il che equivale a dire che segue la  $\text{pdf}(\epsilon)$ , quando si effettua la misurazione di una grandezza possiamo pensare ad  $\epsilon$  che l'incertezza di stima come il contributo di molteplici errori, dunque al crescere di tale numero per il TCL si avrà che  $\text{pdf}(\epsilon) \rightarrow G(\epsilon, \mu, \frac{\sigma}{\sqrt{N}})$ .

## 2.2 Distribuzione di probabilità Log-normale

Presi una variabile aleatoria  $x \in \Omega$  che segue una distribuzione Gaussiana, costruiamo una nuova variabile aleatoria  $y$  legata alla precedente dalla relazione  $y = e^x$ , in questo modo si definisce la distribuzione log-normale applicando un cambio di variabile alla  $G(x)$  di partenza, il metodo con cui si effettua un cambio di variabile verrà presentato nei capitoli successivi.

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

### 2.2.1 Proprietà della log-normale

Il valore di aspettazione e varianza della log-normale sono dati da:

$$E[x] = e^{(\mu + \frac{1}{2}\sigma^2)} \quad (2.4)$$

$$V[x] = e^{(2\mu + \sigma^2)} e^{(\sigma^2 - 1)} \quad (2.5)$$

## 2.3 Distribuzione di Probabilità di Bernoulli

Sia  $x \in \Omega = \{0, 1\}$  una variabile aleatoria discreta, a cui è possibile associare una probabilità  $p \in [0, 1]$ . Definiremo che l'evento  $x$  si è realizzato con successo quando  $P(x = 1) = p$ , mentre l'insuccesso con  $P(x = 0) = (1-p)$ . Dunqua la  $\text{pdf}(x)$  è data da:

$$f(x, p) = p^x (1-p)^{1-x}$$

### 2.3.1 Proprietà della distribuzione Bernoulliana

Per una distribuzione di Bernoulli media e varianza sono date:

$$E[x] = \sum_{\{0,1\}} xf(x, p) = p \quad (2.6)$$

$$V[x] = \sum_{\{0,1\}} (x - p)^2 f(x, p) = p(1 - p) \quad (2.7)$$

## 2.4 Distribuzione di Probabilità Binomiale

Si consideri un insieme di  $N$  osservazioni indipendenti, che seguono una distribuzione di Bernoulli. Contiamo il numero di volte  $k$  in cui l'evento ha successo rispetto al numero di tentativi effettuati, tale insieme è il nostro campione. Se si ripete l'esperimento effettuando sempre  $N$  tentativi, il numero di successi  $n$  sarà distribuita secondo la distribuzione binomiale:

$$B(k|N, p) = \binom{N}{k} p^k (1 - p)^{N-k}$$

notare che a differenza delle distribuzioni viste fino ad ora, la distribuzione binomiale in realtà rappresenta la probabilità stessa di una variabile aleatoria  $k$ , come anche la Bernoulliana.

### 2.4.1 Proprietà della distribuzione binomiale

Per una distribuzione binomiale il valore di aspettazione e la varianza sono definite da :

$$E[k] = \sum_k^\infty k \binom{N}{k} p^k (1 - p)^{N-k} = Np \quad (2.8)$$

$$V[k] = E[k^2] - E[k]^2 = Np(1 - p) \quad (2.9)$$

la simmetria è data da:

$$\gamma_1 = \frac{1 - 2p}{\sqrt{Np(1 - p)}} \quad (2.10)$$

dove  $\gamma_1 \rightarrow 0$  per  $N \rightarrow \infty$  o per  $p = \frac{1}{2}$ .

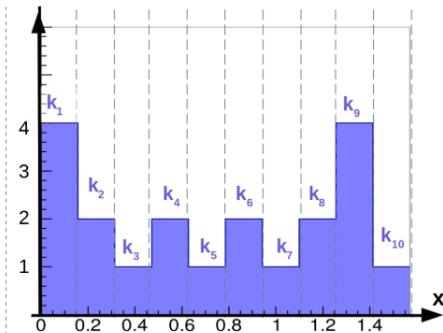
la kurtosi invece si esprime come:

$$\gamma_2 = \frac{1 - 6p}{Np(1 - p)} \quad (2.11)$$

dove  $\gamma_2 \rightarrow 0$  per  $N \rightarrow \infty$  e  $p = \frac{1}{6}$ . Inoltre la distribuzione binomiale gode della proprietà di **riproduuttività**

### 2.4.2 La densità di probabilità di un'istogramma

Consideriamo di avere una variabile aleatoria  $x$  di cui effettuiamo  $N$  misurazioni, ciascuna misurazione segue una  $\text{pdf}(x)$ .



L'operazione è equivalente a campionare  $N$  volte una  $\text{pdf}(x)$ . Posto che le  $\{x_i\}_i^N$  misure siano il campione di riferimento, la statistica si pone obiettivo quello di ricostruire partendo dal campione le caratteristiche di una  $\text{pdf}(x)$ .

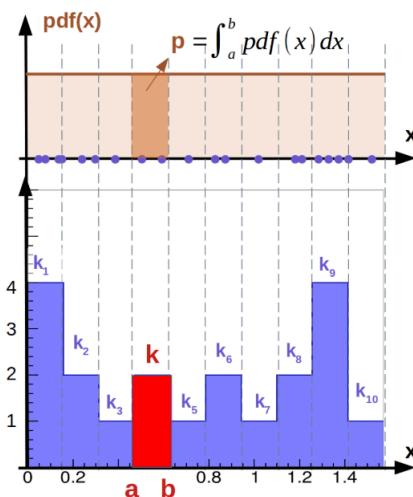
Per farlo si procede a costruire un istogramma e a contare il numero di eventi  $k_i$  che cadono all'interno di

un bin e rappresentare tale numero con una barra. Il numero di misure  $k_i$  che cade all'intero di un bin è una variabile aleatoria.

Di conseguenza studiando il comportamento di un singolo bin  $[a,b]$ , il numero di misure  $k_i$  è descritto da una distribuzione di probabilità, poiché una misura cade o non cade all'interno dell'intervallo la  $\text{pdf}(k_i)$  è quella di Bernoulli.

Reiterando l'esperimento sullo stesso bin il numero di successi  $k_i$  seguirà una  $\text{pdf}(k)$  di una binomiale  $B(k|N, p)$  dove la probabilità è  $p = \int_a^b \text{pdf}(x) dx$  sull'intervallo del bin associato.

Di conseguenza un conteggio di misure è soggetto a fluttuazioni statistiche, e al crescere del numero delle



misure l’istogramma assumerà la distribuzione di probabilità  $\text{pdf}(x)$  della variabile aleatoria  $X$  misurata.

Se ipotizziamo che  $X$  abbia una distribuzione uniforme, valore di aspettazione e varianza di un conteggio coincideranno con quella della distribuzione Binomiale.

I metodi statistici che consentono di utilizzare il metodo dell’istogramma richiedono che i conteggi  $k_i$  siano variabili aleatorie indipendenti, si hanno due strategie affinchè questo sia realizzabile:

- Se la dimensione dei bin è piccola e la probabilità è sufficientemente piccola allora la covarianza è pressoché nulla. Di conseguenza le variabili  $k_i$  sono indipendenti tra loro e dunque ciascuna distribuita come una binomiale.
- Se il numero di eventi  $N$  è molto grande e la probabilità  $p_i$  è piccola la distribuzione di probabilità dei conteggi è approssimabile a quella di una distribuzione Poissoniana.

## 2.5 Distribuzione Multinomiale

Considerando quanto discusso nella sezione precedente per ogni singolo bin che si è costruito nell’istogramma, si definisce una probabilità multinomiale:

$$\text{Multi}(k_1, k_2, \dots, p_1, p_2, \dots) = \frac{N!}{k_1! \cdot k_2! \dots} p_1^{k_1} p_2^{k_2} \dots \quad (2.12)$$

Valore di aspettazione e varianza rimangono le medesime per ogni singolo bin, solo che i conteggi  $k_i$  non sono variabili aleatorie indipendenti tra loro, in questo caso si ha la **covarianza**:

$$\text{cov}[k_i, k_j] = -N \cdot p_i \cdot p_j \quad (2.13)$$

## 2.6 Distribuzione di Poisson

La distribuzione di Poisson è una distribuzione di probabilità discreta che esprime la probabilità che un numero di eventi indipendenti si verifichi in un certo intervallo di spazio o tempo fissato con una frequenza media costante. Il verificarsi dell'evento a un tempo  $t$  non influenza il verificarsi del medesimo in un intervallo di tempo successivo o precedente al tempo  $t$ . La probabilità di contare  $k$  eventi in un intervallo di tempo unitario è:

$$Poiss(k, \lambda) = \frac{e^{-\lambda} \cdot \lambda^k}{k!} \quad (2.14)$$

lo spazio campione rispetto alla quale è definita è tutto  $\mathbb{N}$

### 2.6.1 Proprietà della distribuzione di Poisson

Il valore di aspettazione e la varianza della distribuzione di Poisson sono:

$$E[k] = \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda} \cdot \lambda^k}{k!} = \lambda \quad (2.15)$$

$$V[k] = \sum_{k=0}^{\infty} (k - \lambda)^2 \cdot \frac{e^{-\lambda} \cdot \lambda^k}{k!} = \lambda \quad (2.16)$$

si nota che all'aumentare della media aumenta anche la varianza.

L'indice di simmetria e kurtosi sono dati da:

$$\gamma_1 = \frac{1}{\sqrt{\lambda}} \quad (2.17)$$

$$\gamma_2 = \frac{1}{\lambda} \quad (2.18)$$

per  $\lambda \rightarrow \infty$  la distribuzione di Poisson diventa più simmetrica. Inoltre la distribuzione di Poisson gode della proprietà di **riproduttività**.

### 2.6.2 Da eventi Poissoniani all'esponenziale

Il modello su cui si basa la distribuzione di Poisson è dato dalle seguenti assunzioni:

- la probabilità che un evento avvenga in un intervallo di tempo  $[t, t+dt]$  è data da  $p \cdot dt$ , dove  $p$  è un valore costante e indipendente da  $t$ ;
- la probabilità di avere più di un evento nell'intervallo  $dx$  è nulla.
- gli eventi sono indipendenti tra loro.

Definiamo  $E = \{ \text{nessun evento avvenuto in } [0, t+dt] \}$  e la funzione  $q(t) = \{\text{probabilità di nessun evento in } [0, t]\}$ , posta  $I$  come l'informazione a priori (ex: tempo di decadimento medio) si ha che la probabilità che si verifichi  $E$  essendosi già verificato  $I$  è:

$$P(E|I) \equiv q(t + dt) \quad (2.19)$$

Dove  $q$  ha come argomento la lunghezza degli intervalli di tempo; possiamo riformulare  $E$  come l'intersezione di due eventi  $E_1 = \{0 \text{ eventi in } [0, t]\}$  ed  $E_2 = \{0 \text{ eventi in } [t, t+dt]\}$ , dunque 2.19 può essere riscritta come:

$$P(E|I) = P(E_1, E_2|I) = P(E_1|I)P(E_2|E_1, I)$$

poichè per ipotesi  $E_1$  ed  $E_2$  sono indipendenti dato che il verificarsi di un evento in un intervallo di tempo non influenza il verificarsi del medesimo evento in un intervallo successivo si ha che:

$$P(E|I) = P(E_1|I)P(E_2|I) \quad (2.20)$$

dove  $P(E|I) = q(t + dt)$ ,  $P(E_1|I) = q(t)$  e  $P(E_2|I) = 1 - pdt$ . Dunque:

$$q(t + dt) = q(t) \cdot (1 - pdt) \Rightarrow \frac{q(t + dt) - q(t)}{dt} = -q(t)p \quad (2.21)$$

questo è equivalente all'equazione differenziale al primo ordine:

$$\frac{dq}{q} = -p \cdot dt \iff q(t) = q(0)e^{-pt}$$

dove  $q(0)$  è determinata normalizzando la distribuzione di probabilità.  
Per eventi che occorrono in modo Poissoniano la probabilità di non avere eventi in un intervallo  $[0,x]$  è data da una esponenziale.

Per comodità poniamo  $q(0)=1$ . Definiamo  $A = \{1\text{ evento in }[t_1, t_1+dt_1]\}$  ed  $B = \{0\text{ eventi in }[0,t_1]\}$  definiamo:

$$P(A, E_1|I) = P(A|I) \cdot P(E_1|I) = e^{-pt}(p \cdot dt) \quad (2.22)$$

Ipotizzando di costruire una catena di eventi in cui si realizza e non si realizza un evento per tempi ordinati si ha che:

$$P(A, E|I) = e^{-pt_1} \cdot p \cdot dt_1 \cdot e^{-p(t_2-t_1)} \cdot p \cdot dt_2 \cdots = e^{-pt} p^n dt_1 \cdots dt_n \quad (2.23)$$

Per definire una probabilità che si realizzino  $n$  eventi in un intervallo di tempo  $[0,t]$ , a prescindere dal tempo preciso in cui avvengono, dobbiamo sommare le probabilità definite in (2.23):

$$P(n|p, t, I) = e^{-pt} p^n \int_0^{t_2} dt_1 \cdot \int_0^{t^3} dt_2 \cdots \int_0^t dt_n \quad (2.24)$$

il prodotto d'integrali ha come soluzione  $\frac{t^n}{n!}$  e dunque:

$$P(n|p, t, I) = \frac{(pt)^n}{n!} e^{-pt} = \frac{\lambda^n}{n!} e^{-\lambda} \quad (2.25)$$

## 2.7 Distribuzione Esponenziale

La distribuzione esponenziale è la distribuzione dei tempi che intercorrono tra due eventi Poissoniani. Infatti consideriamo  $k$  il numero di eventi in un processo Poissoniano con frequenza  $p$ , avvenuti in un intervallo di tempo  $[0,t]$ . Allora si ha:

$$P(k, p) = \frac{(pt)^k}{k!} e^{-pt}$$

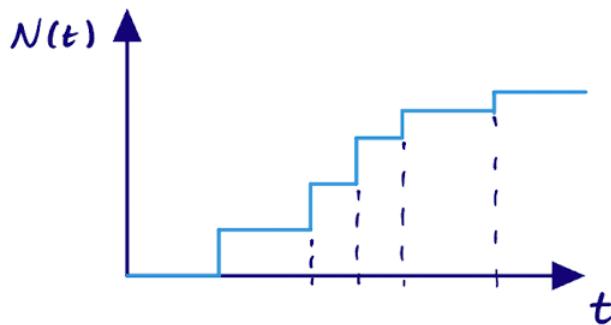
Sia  $T > t$  il tempo in cui effettuiamo la prima misura, poichè la frequenza  $p$  è una costante indipendente dal tempo, se misuriamo la probabilità per 0 eventi in un tempo  $t < T$  si avrà:

$$P(0|p, t) = \frac{(pt)^0}{0!} e^{-pt} = e^{-pt}$$

Se rileviamo il medesimo evento per  $T > t$  avrà la stessa probabilità dell'evento avvenuto al tempo  $t$ :

$$P(0|p, T > t) = P(0|p, t) = e^{-pt}$$

Dunque la distribuzione di probabilità che intercorre da tra due eventi Poissoniani è esponenziale.



ed è definita da:

$$pdf(t, \lambda) = \lambda \cdot e^{-\lambda t} \quad (2.26)$$

dove  $\tau = E[t]$  è il tempo medio tra due eventi consecutivi, e vale la relazione  $\lambda = \frac{1}{\tau}$  dove  $\lambda$  è il parametro della Poissoniana che determina la frequenza media di eventi.

### 2.7.1 Proprietà della distribuzione esponenziale

Il valore di aspettazione e varianza sono date dalle quantità:

$$E[t] = \tau = \frac{1}{\lambda} \quad (2.27)$$

$$V[t] = \tau^2 = \frac{1}{\lambda^2} \quad (2.28)$$

mentre i parametri di simmetria della distribuzione dati da simmetria e kurtosi:

$$\gamma_1 = 2 \quad (2.29)$$

$$\gamma_2 = 6 \quad (2.30)$$

come si osserva dai momenti essendo  $\gamma_1$  e  $\gamma_2$  costanti la distribuzione esponenziale non può tendere ad una Gaussiana al crescere del numero die venti.

### 2.7.2 Decadimento Radioattivo

Si consideri un campione di materiale radioattivo contenente  $N_0$  nuclei al tempo  $t=0$  per un singolo nucleo padre la probabilità di decadimento segue una distribuzione esponenziale il cui  $\tau$  tempo medio di decadimento è dato dalla meccanica quantistica. Gli eventi di decadimento sono **indipendenti** tra di loro.

La variazione del numero dei nuclei in un intervallo di tempo  $\Delta t$  sarà proporzionale al numero di nuclei restanti:

$$-\frac{dN}{dt} = \lambda N \quad (2.31)$$

dove  $\lambda$  è la frequenza media di decadimento. Di conseguenza il decadimento radioattivo è definito dalla funzione:

$$N(t) = N_0 e^{-\lambda t} \quad (2.32)$$

che possiamo riscrivere come :

$$N(t) = N_0 e^{-\frac{t}{\tau}} \quad (2.33)$$

Se  $\tau \gg t$  tempo di osservazione, la frequenza media dei decadimenti è costante nel tempo e di conseguenza la probabilità di osservare  $k$  decadimenti in un campione di  $N$  nuclei è data dalla distribuzione di Poisson con  $\lambda = \frac{1}{\tau}$ .

## 2.8 Comportamenti Asintotici delle distribuzioni

### 2.8.1 Proprietà asintotiche della Binomiale e Poissoniana

$$\text{Binom}(k|N,p) \rightarrow \text{Poisson}(k,\mu = N \cdot p)$$

La relazione tra distribuzione Binomiale e Poissoniana si costruisce assumendo di prendere un intervallo  $\Delta t$  e di dividerlo in  $N$  intervalli di dimensione  $dt$  tali per cui:

- in ciascun  $dt$  c'è uno solo evento (tale ipotesi è verificata solo se il numero di eventi è molto grande);
- la probabilità in un singolo intervallo  $dt$  è  $p = \frac{\lambda}{N}$  ed è finita;
- gli eventi che cadono in ogni intervallo  $dt$  sono indipendenti tra loro.

Essendo gli eventi indipendenti tra loro per ipotesi ciascuno di essi segue una distribuzione di Bernoulli. Tali condizioni consentono di usare una distribuzione binomiale per descrivere eventi Poissoniani.

Per induzione si ha che per  $k=0$  posto  $\lambda = N \cdot p$

$$B(0|N, \frac{\lambda}{N}) = \left(1 - \frac{\lambda}{N}\right)^N =_{N \rightarrow \infty} \frac{\lambda^0}{0!} e^{-\lambda}$$

dunque la relazione per  $k-1$  sarà verificata e:

$$B(k-1|N, \frac{\lambda}{N}) =_{N \rightarrow \infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}$$

Allora si dimostra che dall'espressione binomiale:

$$\frac{B\left(k|N, \frac{\lambda}{N}\right)}{B\left(k-1|N, \frac{\lambda}{N}\right)} = \frac{Np - (k-1)p}{k(1-p)} \underset{p \rightarrow 0}{\approx} \frac{\lambda}{k}$$

dunque:

$$B\left(k|N, \frac{\lambda}{N}\right) = \frac{\lambda}{k} \cdot B\left(k-1|N, \frac{\lambda}{N}\right) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

Si dimostra inoltre che data una distribuzione Binomiale  $B(k|N, p) \rightarrow G(k, \mu, \sqrt{\mu})$  al crescere del numero di misure  $N \rightarrow \infty$  per  $\mu = N \cdot p$ .

Vale anche il fatto che per una Poissoniana si ha  $Poiss(k, \lambda) \rightarrow G(k, \lambda, \sqrt{\lambda})$  per  $\lambda \rightarrow \infty$ .

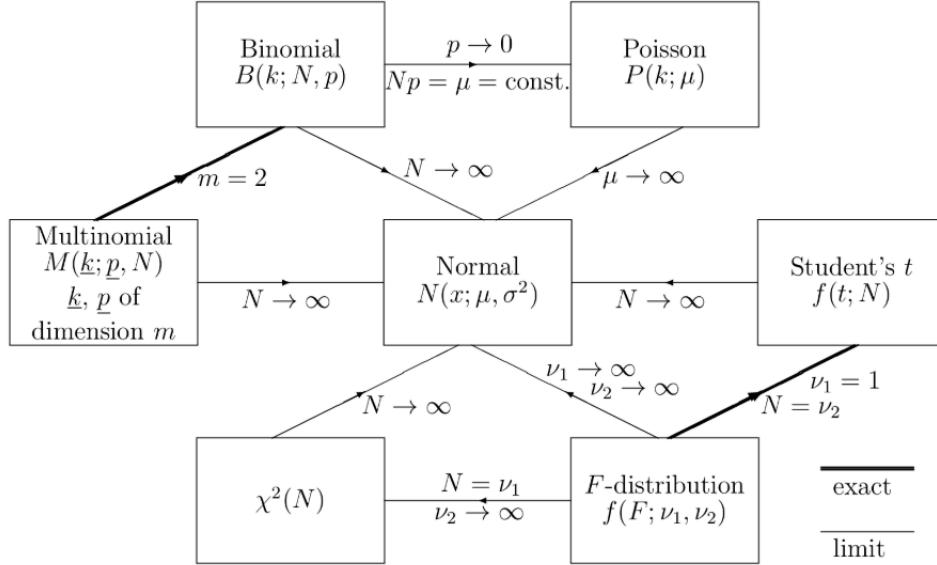


Figure 2.1: Mappa delle distribuzioni asintotiche

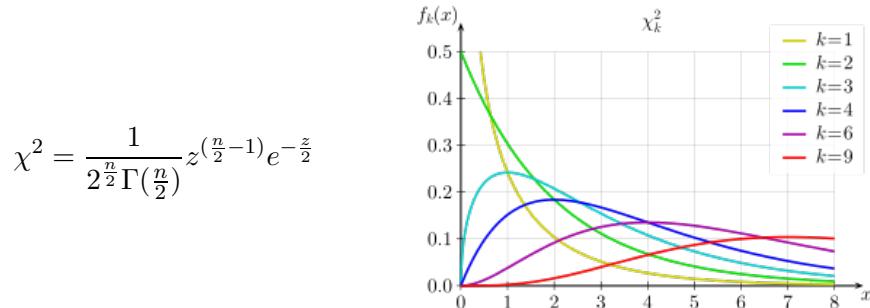
## 2.9 Distribuzione di $\chi^2$

La distribuzione di  $\chi^2$  nasce nel contesto della modellazione di dati per la stima di parametri usando i metodi di maximum likelihood e minimi quadrati. Dato un campione di  $n$  punti  $\{(x_i, y_i)\}_i^N$  e un modello  $f(x, \underline{\theta})$ , dipendente da  $m$  parametri  $\underline{\theta}$  si valuta la bontà del modello nell'approssimare i dati usando la funzione di  $Q^2$  definita come:

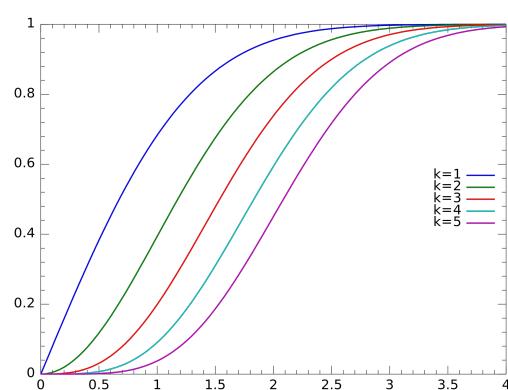
$$Q^2 = \sum_{i=1}^n \frac{[y_i - f(x_i, \underline{\theta})]}{\sigma_i^2} \quad (2.34)$$

dove le  $\sigma_i^2$  sono gli errori sulle  $n$  misure di  $y_i$ .

Per le misure  $y$  che seguono una pdf Gaussiana, la pdf del  $\chi^2$  è data dalla distribuzione:



dove **n** prende il nome di **numero di gradi di libertà** e  $\Gamma(x)$  è la funzione gamma.



La distribuzione di  $\chi^2$  possiede una distribuzione cumulativa e gode delle seguenti proprietà:

- **media:**  $E[n] = n$
- **Varianza:**  $V[n] = 2n$
- è riproduttiva
- **moda:**  $n-2$

### $\chi^2$ ridotto

Si definisce  $\chi^2$  ridotto data dal rapporto  $\frac{\chi^2_N}{n}$  e la sua media è pari a 1.

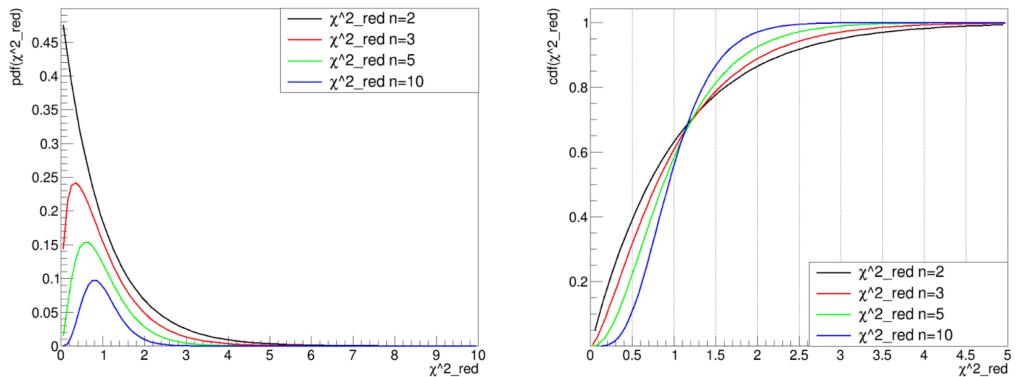


Figure 2.2: Distribuzione di  $\chi^2$  ridotto e rispetti distribuzione cumulativa.



# CHAPTER 3

## Distribuzioni di probabilità multidimensionali

### 3.1 Probabilità per variabili aleatorie in più dimensioni

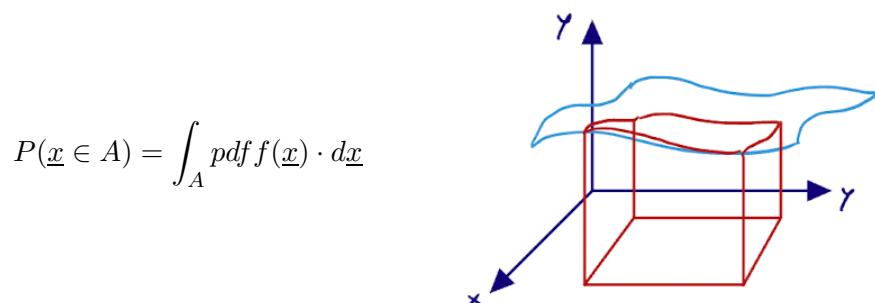
Quando un evento è identificato da un vettore  $\underline{x}$ , si parla di distribuzioni di probabilità multidimensionali o **joint pdf (probabilità congiunta)**.

$$pdf(\underline{x}) : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^+$$

e la cdf per estensione della definizione mono-dimensionale è:

$$cdf(\underline{x}) = \int_A pdf(\underline{x}) \cdot d\underline{x}$$

la probabilità di un evento per una variabile continua  $\underline{x}$  è definita su una regione di spazio  $A \subseteq \mathbb{R}^n$ :



Tale definizione di probabilità verifica gli assiomi di Kolgomorov.

### 3.2 Distribuzione di probabilità marginale

Data una distribuzione di probabilità multidimensionale  $\text{pdf}(x_1, \dots, x_n)$  si definisce **distribuzione di probabilità marginale**:

$$f_{x_i}(x_1, \dots, x_i, \dots, x_n) = \int \text{pdf}(x_1, \dots, x_i, \dots, x_n) dx_1 \cdots dx_n \quad (3.1)$$

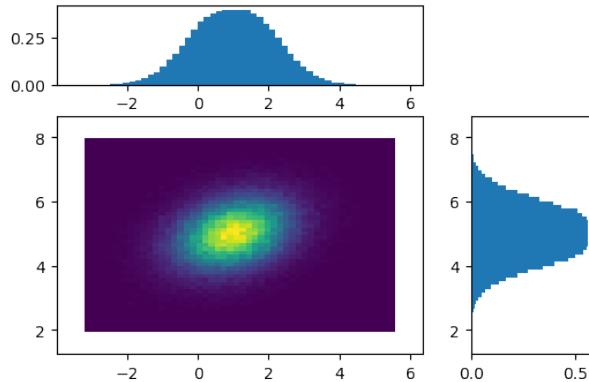


Figure 3.1: Distribuzione di probabilità marginale per una  $\text{pdf}(x,y)$

#### Esempio

Nel caso bi-dimensionale si ha che le rispettive distribuzioni marginali di una joint  $\text{pdf}(x_1, x_2)$  sono date da:

$$f_x(y) = \int \text{pdf}(x, y) dy \quad f_y(x) = \int \text{pdf}(x, y) dx \quad (3.2)$$

### 3.3 Distribuzione di probabilità condizionata

Per semplicità consideriamo una distribuzione di probabilità rispetto a due variabili aleatorie  $x$  ed  $y$  e le rispettive distribuzioni marginali  $f_x$  e  $f_y$ . Vogliamo determinare la probabilità che  $P(x|y = y_0)$  o  $P(y|x = x_0)$ . Consideriamo due eventi  $A, B \subset \Omega$  disgiunti. Possiamo identificare le pdf come:

- $P(A \cap B) \rightarrow$  joint pdf

- $P(A) \rightarrow$  pdf marginale
- $P(A|B) \rightarrow$  pdf condizionata

Definiamo L'evento  $A = \{y \mid x \in [x_0, x_0 + dx], y \in \mathbb{R}\}$  e l'evento  $B = \{x \mid y \in [y_0, y_0 + dy], x \in \mathbb{R}\}$ . Le probabilità associate ai singoli eventi sono  $P(A) = f_x$  e  $P(B) = f_y$ . Mentre la probabilità congiunta è  $P(A \cap B) = \int pdf(x_0, y_0) dx dy$ .

Di conseguenza possiamo scrivere la probabilità condizionata come:

$$pdf(x|y = y_0) = P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{pdf(x, y_0)}{f_x(y_0)} \quad (3.3)$$

e analogamente:

$$pdf(y|x = x_0) = \frac{pdf(x_0, y)}{f_y(x_0)} \quad (3.4)$$

### 3.4 Valore di aspettazione di una joint pdf

Per una distribuzione di probabilità congiunta abbiamo che il valore di aspettazione è definito da un vettore:

$$E[\underline{x}] = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} \int \int x \cdot pdf(x, y) dx dy \\ \int \int y \cdot pdf(x, y) dx dy \end{bmatrix} \quad (3.5)$$

### 3.5 Varianza di una pdf multidimensionale

Nel caso della Varianza, le componenti di una variabile aleatoria multidimensionale  $\underline{x}$  possono essere legate tra loro, ovvero avere una relazione nel modo in cui variano, tale legame prende il nome di **covarianza**. Per due componenti  $x_i$  e  $x_j$  con  $i \neq j$  si ha che:

$$\sigma_{i,j}^2 = E[(x_i - \mu_i)(x_j - \mu_j)] = E[x_i x_j] - E[x_i]E[x_j] \quad (3.6)$$

mentre

$$\sigma_{i,i}^2 = E[(x_i - \mu_i)^2] \quad (3.7)$$

---

CHAPTER 3. DISTRIBUZIONI DI PROBABILITÀ  
MULTIDIMENSIONALI

---

di conseguenza la varianza di una pdf( $\underline{x}$ ) multidimensionale è rappresentata da una matrice che prende il nome di **matrice di covarianza** ed ha dimensione  $n \times n$  nel caso  $\underline{x}$  sia un vettore dimensione n.

$$V[\underline{x}] = \begin{bmatrix} V[x_1] & \cdots & Cov[x_1, x_n] \\ \vdots & \ddots & \vdots \\ Cov[x_n, x_1] & \cdots & V[x_n] \end{bmatrix} \quad (3.8)$$

Nel caso in cui le componenti della variabile aleatoria  $\underline{x}$  siano indipendenti tra loro si ha che la covarianza è nulla e dunque la (3.3) diventa una matrice diagonale.

$$V[\underline{x}] = \begin{bmatrix} V[x_1] & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V[x_n] \end{bmatrix} \quad (3.9)$$

La covarianza gode delle seguenti proprietà:

- Avere  $Cov[x_i, x_j] = 0$  non implica necessariamente che le due variabili aleatorie siano statisticamente indipendenti.
- Se due variabili aleatorie sono statisticamente indipendenti  
 $\Rightarrow Cov[x_i, x_j] = 0$
- Se due variabili aleatorie sono linearmente dipendenti  
 $\Rightarrow Cov[x_i, x_j] = 0$
- La matrice di covarianza è simmetrica  $\Rightarrow$  è diagonalizzabile.

### 3.6 Correlazione

Gli elementi  $V_{ij}$  della matrice di covarianza misurano il grado di correlazione tra le variabili  $x_i$  e  $x_j$ . Dato che ogni variabile mostra una varianza finita e

## CHAPTER 3. DISTRIBUZIONI DI PROBABILITÀ MULTIDIMENSIONALI

---

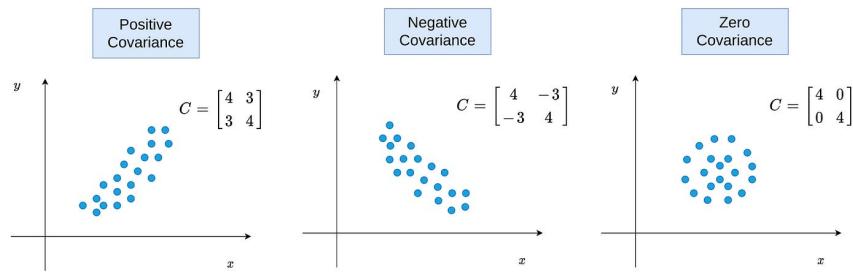


Figure 3.2: Esempi di come la covarianza si presenta in una distribuzione

positiva è utile confrontare la covarianza rispetto alle loro rispettive varianze, per farlo si introduce il coefficiente di correlazione:

$$\rho_{ij} = \frac{\text{Cov}[x_i, x_j]}{\sigma_i \sigma_j} \quad (3.10)$$

tale grandezza  $\rho_{ij} \in [-1, 1]$ .

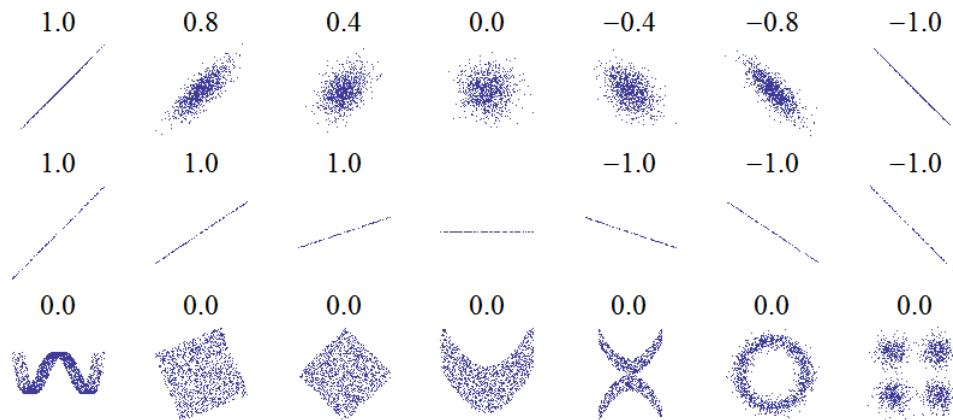


Figure 3.3: Indice di correlazione di Pearson per un campione di misure.

### 3.7 Variabili Statisticamente indipendenti

Applichiamo la definizione di probabilità congiunta data dall'equazione (3.3), poichè per ipotizziamo che  $P(x_2|x_1)$  e  $x_2$  non dipenda da  $x_1$  avremo che:

$$pdf(x_2|x_1) = \frac{pdf(x_1, x_2)}{f_{x_1}} \Rightarrow pdf((x_1, x_2)) = f_{x_1} \cdot pdf(x_2|x_1)$$

si ha che

$$f_{x_2}(x_1) = \int pdf(x_1, x_2) dx_1 = pdf(x_2|x_1) \int f_{x_1} dx_1 = pdf(x_2|x_1)$$

dunque si ha che:

$$pdf(x_1, x_2) = f_{x_1} \cdot f_{x_2}$$

#### Definizione d'indipendenza

Due variabili aleatorie  $x_1$  ed  $x_2$  descritte da una joint  $pdf(x_1, x_2)$  si definiscono **indipendenti**  $\iff$  la joint  $pdf(x_1, x_2) = pdf(x_1) \cdot pdf(x_2)$  dove  $pdf(x_1)$  e  $pdf(x_2)$  coincidono con le marginali.

#### Teorema

Due variabili aleatorie  $x_1$  ed  $x_2$  descritte da una joint  $pdf(x_1, x_2)$  sono indipendenti  $\Rightarrow$  la covarianza può essere scritta come:

$$E[(x_1 - \mu_{x_1})(x_2 - \mu_{x_2})] = E[(x_1 - \mu_{x_1})] \cdot E[(x_2 - \mu_{x_2})] \quad (3.11)$$

### 3.8 Cambiamento di variabili per una pdf multidimensionale

La matrice della varianza (3.6) è diagonalizzabile, questo vuol dire che esiste un cambio di base che la diagonalizza, in fisica è equivalente ad avere un cambio di variabile. Sperimentalmente raccolto un campione di misure esiste sempre un cambio di variabile tale per cui la matrice di covarianza è diagonalizzabile. Anche se con un cambio di variabili le grandezze diventano

### CHAPTER 3. DISTRIBUZIONI DI PROBABILITÀ MULTIDIMENSIONALI

---

decorrelate  $Cov[x_i, x_j] = 0$  questo non vuol dire che siano statisticamente indipendenti.

Analogamente al caso mono-dimensionale per il cambio di variabili si ha che date delle funzioni:

$$\begin{cases} x = u(\alpha, \beta) \\ y = w(\alpha, \beta) \end{cases}$$

la join pdf nelle nuove coordinate sarà data da:

$$pdf(\alpha, \beta) = pdf(x, y) \cdot |detJ| \quad (3.12)$$

dove J è la matrice Jacobiana associata alla trasformazione.

#### 3.8.1 Propagazione degli errori

Ipotizziamo di avere un insieme di N misure  $\{x_i\}_i^N$  e usiamo tali valori come componenti di un vettore  $\underline{x} \in \mathbb{R}^N$ , descritto da una  $pdf(\underline{x})$  di cui conosciamo  $\underline{\mu}$  e matrice di covarianza  $V[\underline{x}]$ , vogliamo calcolare  $y = u(\underline{x})$ . Per Farlo approssimiamo  $u(\underline{x})$  con un sviluppo di Taylor al primo ordine in un intorno di  $\underline{\mu}$ :

$$u(\underline{x}) \approx u(\underline{\mu}) + \nabla u|_{x=\underline{\mu}} \cdot (\underline{x} - \underline{\mu})$$

Dunque i momenti della  $pdf(y)$  sono:

- $E[y] = E[u(\underline{\mu})] + \sum_{i=1}^n \frac{\partial}{\partial x_i} u(\underline{\mu}) E[(x_i - \mu_i)] = u(\underline{\mu})$
- $\sigma_y^2 = E[y^2] - E[y]^2 = E[(\underline{x} - \underline{\mu})^T H(u(\underline{x}))|_{x=\underline{\mu}} (\underline{x} - \underline{\mu})] =$   
 $= \sum_{i,j=1}^n \frac{\partial u(\underline{x})}{\partial x_i} \cdot \frac{\partial u(\underline{x})}{\partial x_j}|_{x=\underline{\mu}} \cdot V_{ij}$

Per un caso bidimensionale l'incertezza su  $z = f(x,y)$  è data da:

$$\sigma_z^2 = \left( \frac{\partial f(x, y)}{\partial x} \right)^2|_{x=\mu} \sigma_x^2 + \left( \frac{\partial f(x, y)}{\partial y} \right)^2|_{x=\mu} \sigma_y^2 + 2 \frac{\partial f(x, y)}{\partial x} \cdot \frac{\partial f(x, y)}{\partial y} \Big|_{x=\mu} Cov[x, y] \quad (3.13)$$

### 3.8.2 Decorrelazione delle variabili

Consideriamo un vettore  $\underline{x}$  le cui componenti sono variabili aleatorie dipendenti tra loro. Poiché la matrice di covarianza è simmetrica possiamo definire una matrice  $U$  che la diagonalizza  $D[\underline{y}] = U^T V[\underline{x}] U$ , tale matrice definisce un cambio di coordinate  $\underline{y} = U \underline{x}$  dove le componenti di  $\underline{y}$  risultano essere decorrelate tra loro.

## 3.9 Distribuzione di Gauss Multidimensionale

L'espressione di una Gaussiana in più dimensioni è data da:

$$f(\underline{x} | \underline{\mu}, V) = \frac{1}{(2\pi)^{\frac{n}{2}} \det V^{\frac{1}{2}}} \cdot \exp \left[ -\frac{1}{2} \langle (\underline{x} - \underline{\mu}), V^{-1}(\underline{x} - \underline{\mu}) \rangle \right] \quad (3.14)$$

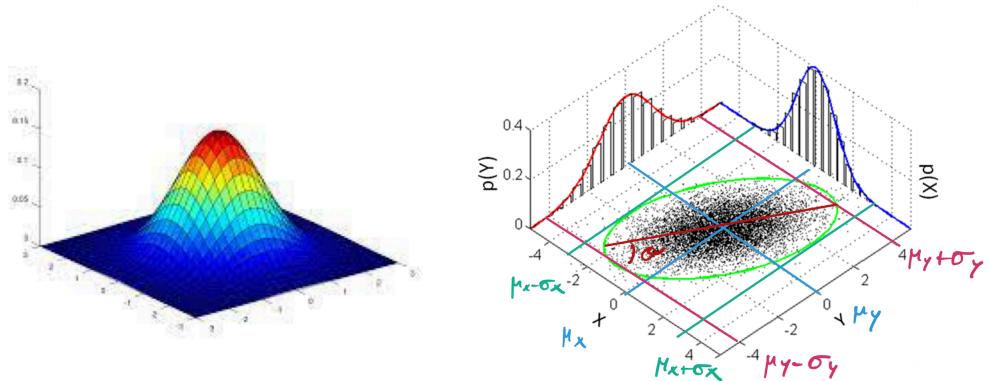


Figure 3.4: Gaussiana in due variabili (x,y) e le sue distribuzioni marginali

La gaussiana in due dimensioni in generale ha un profilo ellittico per  $1\sigma$ . L'angolo d'inclinazione del semiasse maggiore è legato al coefficiente di correlazione:

$$\theta = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \quad (3.15)$$

---

CHAPTER 3. DISTRIBUZIONI DI PROBABILITÀ  
MULTIDIMENSIONALI

---

Se esplicitiamo l'equazione (3.11) per due variabili si ha:

$$f(x, y \mid \mu_x, \mu_y, V) = \\ = \frac{1}{(2\pi)\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{x-\mu_x}{\sigma_x} \right)^2 - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left( \frac{y-\mu_y}{\sigma_y} \right)^2 \right] \right\}$$

Se il coefficiente di correlazione di Pearson  $\rho = 0$  possiamo riscrivere l'equazione precedente come:

$$f(x, y \mid \mu_x, \mu_y, V) = \frac{1}{(2\pi)\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left[ \left( \frac{x-\mu_x}{\sigma_x} \right)^2 + \left( \frac{y-\mu_y}{\sigma_y} \right)^2 \right] \right\} = \\ = G(x, \mu_x, \sigma_x) \cdot G(y, \mu_y, \sigma_y)$$

In generale è solo vero che due variabili statisticamente indipendenti sono anche decorrelate, nel caso Gaussiano vale anche il viceversa, ovvero se due variabili sono decorrelate allora sono anche statisticamente indipendenti.

CHAPTER 3. DISTRIBUZIONI DI PROBABILITÀ  
MULTIDIMENSIONALI

---

# CHAPTER 4

---

## Stime di Parametri

---

### 4.1 La statistica

La statistica studia il problema di inferire da un campione i parametri e/o i modelli che descrivono la popolazione dalla quale il campione è stato estratto. In particolare possiamo dividerla in due categorie:

- Stima dei parametri, misura di una quantità fisica;
- Test d'ipotesi, ovvero la prova della validità di un modello.

Una funzione dipendente da N misure di un campione  $f(x_1, \dots, x_n)$  si chiama **statistica**, essa è una **variabile aleatoria**. Quindi segue una sua distribuzione di probabilità  $pdf_f$  derivabile dalla joint-pdf dei campionamenti e dalla forma della funzione f.

Complessivamente si hanno 3 pdf:

- la  $pdf_x(x, \theta)$  delle singole misure campionate;
- la  $pdf_{set}(x_1, \dots, x_n, \theta)$  dei campionamenti (che è multidimensionale);
- la  $pdf_f$  della statistica dei campionamenti (dipende dalla forma funzionale di f).

### 4.2 Stimatori

Sia data una p.d.f. (probability distribution function),  $f(x, \theta)$  di una variabile x aleatoria continua e dipendente da un parametro  $\theta$ , di cui non conos-

ciamo il vero valore  $\theta_{true}$ .

Se si possiede un insieme  $\{x_i\}_i^N$  di N misure della variabile x, possiamo chiederci se sia possibile determinare una stima del parametro  $\theta_{true}$  in funzione di tali misure,  $\hat{\theta} = \hat{\theta}(x_1, \dots, x_N)$ , le funzioni di questo tipo prendono il nome di **stimatori**. Uno stimatore è una statistica opportunamente scelta. Con **stima** s'intende il valore assunto  $\hat{\theta}^*$  dallo stimatore per uno specifico campione.



Figure 4.1: N misure

Poichè lo stimatore  $\hat{\theta}$  è dipende da variabili aleatorie è anch'esso una variabile aleatoria e dunque si può parlare di valore medio  $E[\hat{\theta}]$  e varianza  $V[\hat{\theta}]$  di una particolare stima, oltre ad avere una sua pdf( $\hat{\theta}$ ).

$$\hat{\theta} \pm \sigma_{\theta} \quad (4.1)$$

Di conseguenza un insieme di misure restituirà un solo valore appartenente ad una popolazione ottenuta da campione fatto di misure della variabile aleatoria presa in considerazione.

### 4.3 Proprietà degli stimatori

Consideriamo un campione di N misure  $\{x_i\}_i^N$  vengono definite IID (Independent Identically Distributed) quando sono:

- **indipendenti:** l'esito di una misura non è influenzato dalle misure precedenti;
- **identiche:** Delle misure vengono definite identiche quando tutte quante seguono la stessa distribuzione di probabilità

Nella statistica alle stime si possono associare diverse caratteristiche :

1. **Consistenza:** una stima si dice consistente quando all'aumentare del numero di misure (convergenza probabilistica) si converge al valore vero del parametro. Ossia quando:

$$\lim_{N \rightarrow \infty} \hat{\theta}(x_1, \dots, x_N) = \theta_{true} \quad (4.2)$$

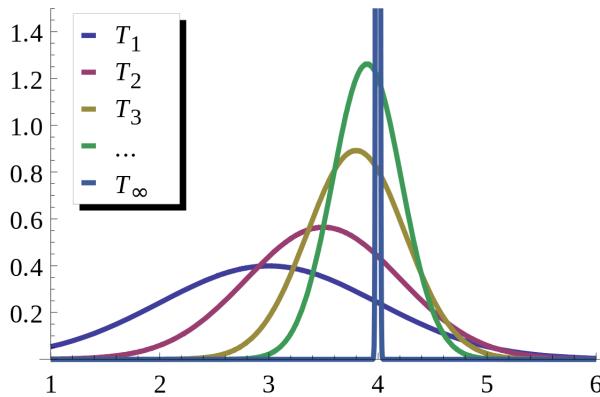


Figure 4.2: Proprietà di consistenza di uno stimatore

2. **Biased:**

- (a) una stima si dice unbiased o imparziale, se mediamente coincide con il valore vero del parametro, ovvero

$$b_n(\hat{\theta}) = E(\hat{\theta}_n - \theta_{true}) = E(\hat{\theta}_n) - E(\theta_{true}) = 0 \iff E(\hat{\theta}_n) = \theta_{true} \quad (4.3)$$

- (b) Una stima si dice asintoticamente unbiased se  $b_n(\hat{\theta}) \rightarrow 0$  per  $n \rightarrow \infty$ .

Si osserva che  $b_n(\hat{\theta})$  è uno stimatore lineare, dunque se  $\hat{\theta}$  è stimatore di  $\theta_{true}$  questo non vuol dire che  $\hat{\theta}^2$  è stimatore di  $\theta_{true}^2$ .

3. **Efficienza:** si dice che una stima è più efficiente di un'altra se la sua

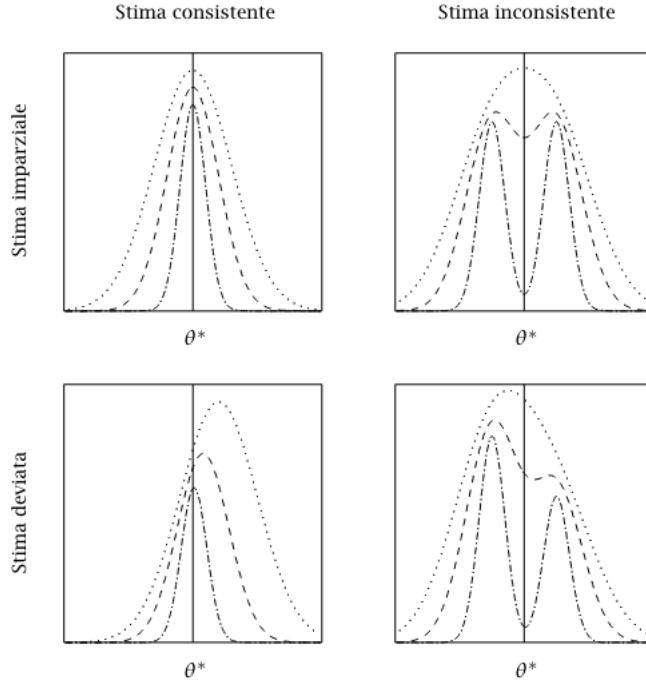


Figure 4.3: Consistenza e Bias di uno stimatore

varianza è inferiore, quindi se mediamente essa è più vicina al valore centrale  $E(\hat{\theta})$ , che coincide con  $\theta_{true}$  se la stima è anche imparziale (unbiased).

4. **Varianza:** desideriamo che ripetendo i campionamenti le stime ottenute siano tutte vicine tra loro, ovvero la varianza della  $pdf(\hat{\theta})$  sia il più piccola possibile.

### 4.3.1 Precisione e Accuratezza

Per uno stesso parametro si possono in generale definire tanti estimatori diversi tra loro, ma non tutti hanno le proprietà desiderate. Da notare che non è detto che esista (o sia possibile trovare) uno stimatore che soddisfi contemporaneamente tutte le proprietà richieste.

### Esempio

La media delle misure è uno stimatore non distorto.

Dim.

Definito come stimatore la media aritmetica delle misure di un campione IID:

$$\hat{\mu} = \frac{1}{N} \sum x_i$$

si ha che il valore di aspettazione dello stimatore è:

$$E(\hat{\mu}) = E\left(\frac{1}{N} \sum x_i\right) = \frac{1}{N} \sum E(x_i) = \frac{1}{N} \cdot N \cdot \mu_t = \mu_t$$

quindi la media aritmetica è uno stimatore **non distorto** poichè:

$$b_n(\hat{\mu}) = E(\hat{\mu}) - \mu_t = \mu_t - \mu_t = 0$$

Se la pdf(x) delle misure soddisfa le ipotesi del TCL, la pdf( $\hat{\mu}$ ) per  $N \rightarrow \infty$  tende a una **gaussiana** con media  $\mu$  e varianza  $\frac{\sigma^2}{N}$  si ha che  $\hat{\mu}$  è uno stimatore **consistente**. Poichè  $V[\hat{\mu}] = \frac{\sigma^2}{N}$  al crescere del numero di campionamenti la varianza si riduce e dunque le stime ottenute con diversi set di dati sono tutte vicine tra loro.

#### 4.3.2 Incertezze sulle stime

Uno stimatore come ogni altra variabile aleatoria è soggetto a due tipi d'incertezze:

1. **Incertezza sistematica:** nel caso di misure biased esiste una differenza sistematica fra la misura sperimentale ottenuta e il valor vero, ed è uguale per tutte le misure e non è possibile determinarlo essendo una proprietà intrinseca.
2. **Incertezza statistica:** è associata alla precisione, e la si può ridurre aumentando il numero di misure o cambiando l'apparato sperimentale.

$$MSE = E[(\hat{\theta} - \theta_t)^2] = Var(\hat{\theta}) + b_n^2(\hat{\theta}) \quad (4.4)$$

Definisce l'errore quadratico medio e tiene conto sia dell'errore statistico misurato dalla varianza che dell'errore sistematico misurato dal bias.

### 4.3.3 La Varianza come stimatore

Consideriamo di avere un insieme di  $N$  misure,  $\{x_i\}_i^N$  di cui conosciamo il valore medio  $\mu$  della popolazione e di volerne determinare la varianza, poichè essa dipende dalle misure del campione è una variabile aleatoria a sua volta e dunque da un campione definiamo una stima del valore reale del parametro  $\sigma_t$ . Di conseguenza possiamo domandarci le proprietà che tale stimatore possiede.

Definiamo lo stimatore varianza come:

$$\hat{\sigma}_{\mu}^2(x) = \frac{1}{N} \sum (x_i - \mu)^2 \quad \text{oppure} \quad \hat{\sigma}^2(x) = E(x^2) - E(x)^2$$

Verifichiamo che la varianza sia uno stimatore non distorto ovvero che:

$$E(\hat{\sigma}^2) = \sigma_t$$

Per farlo sfruttiamo la proprietà di linearità del valore atteso.

Dim.

$$\begin{aligned} E(\hat{\sigma}_{\mu}^2) &= E\left(\frac{1}{N} \sum (x_i - \mu)^2\right) = \frac{1}{N} \sum (E(x_i^2) - 2\mu E(x_i) + \mu^2) = \\ &= \frac{1}{N} \sum (E(x_i^2) - 2\mu E(x_i) + \mu^2) = \frac{1}{N} \sum (E(x_i^2) - \mu^2) = \frac{1}{N} \cdot N \cdot \sigma_t = \sigma_t \end{aligned}$$

Dunque la varianza è uno stimatore non distorto nel caso in cui si conosca il valore medio della popolazione. Raramente si conosce  $\mu$  della popolazione, dunque consideriamo come stimatore la varianza per un campione di  $N$  misure IID.

$$\hat{\sigma}_{\bar{x}}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

Sappiamo che la varianza determina la dispersione di un campione di mis-

ure attorno alla sua media. Ipotizziamo di conoscere il valore medio del campione  $\bar{x}$ , per costruzione risulta essere il valore più vicino alle misure dell'insieme. La media della popolazione  $\mu$  non necessariamente coincide con  $\bar{x}$  del campione, e dunque può non essere il valore attorno al quale si distribuiscono le misure del campione; infatti nel caso in cui non lo sia al crescere del numero di misure questo può diventare il valore più distante rispetto a  $\bar{x}$  stimato dal campione iniziale. Le distanze quadratiche da  $\bar{x}$  saranno quindi una sottostima di  $\mu$  e quindi anche  $\hat{\sigma}$  sarà uno sottostima di  $\sigma_t$ .

$$\begin{aligned} E[\hat{\sigma}_{\bar{x}}^2] &= \frac{1}{N} \sum E(x_i^2) - E([\frac{1}{N} \sum x_i]^2) = \sigma_t(x)^2 + \mu^2 - \frac{1}{N^2} [\sigma_t^2(\sum x_i) + E(\sum x_i)^2] = \\ &= \sigma_t(x)^2 + \mu^2 - \frac{1}{N^2} [N\sigma_t^2(x) + N^2\mu^2] = \sigma_t^2(x) \left[ \frac{N-1}{N} \right] \end{aligned}$$

Di conseguenza la varianza di un campione  $\hat{\sigma}_{\bar{x}}$  è uno stimatore distorto, infatti:

$$b_n[\hat{\sigma}_{\bar{x}}^2] = E[\hat{\sigma}_{\bar{x}}^2] - \sigma_t^2 = \sigma_t^2(x) \left[ \frac{N-1}{N} \right] - \sigma_t^2$$

ma **asintoticamente non distorto** poiché per  $N \rightarrow \infty$  si ha  $b_n[\hat{\sigma}_{\bar{x}}^2] \rightarrow 0$ .

Notare che quest'ultima definizione è quella operativa per verificare che la varianza sia uno stimatore non ubiased in quanto difficilmente si conosce il valore medio  $\mu$  della popolazione.

#### 4.3.4 Correzione di Bessel

Si può definire un terzo stimatore, che introduca una correzione a  $\hat{\sigma}_{\bar{x}}^2$  tale da cancellare il bias. La correzione del bias è applicabile tutte le colte in cui il bias è precisamente noto. Il nuovo stimatore della varianza sarà dato da:

$$s^2 \equiv \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (4.5)$$

e prende il nome di **correzione di Bessel**.

Lo stimatore è unbiased  $E[s^2] = \sigma_t^2$ , ma la varianza di tale stimare non può

essere determinata per il caso generale

$$V[s^2] = V\left[\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2\right] = \frac{1}{(N-1)^2} \sum_{i=1}^N V[(x_i - \bar{x})^2]$$

ma è possibile farlo solo nel caso in cui il campione di misura segue una pdf(x) **Gaussiana**.

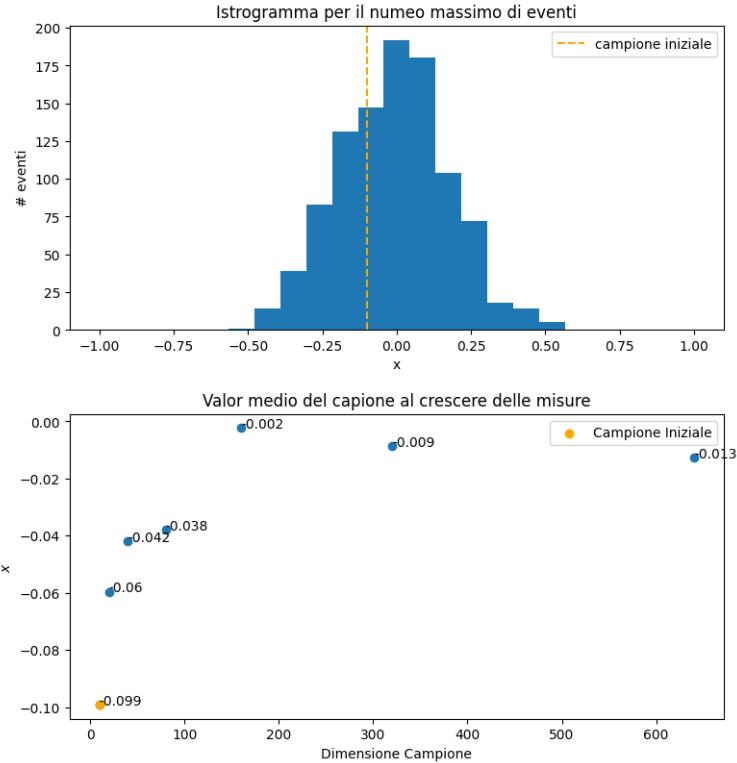


Figure 4.4: Misure casuali che seguono una pdf di Gauss tra -1 e 1 in cui  $\bar{x}$  del campione iniziale non coincide con il valore medio  $\mu = 0$  della popolazione. Dunque  $\bar{x}$  non è più il centro del campione al crescere delle misure.

Se riscriviamo lo stimatore  $s^2$  nel seguente modo:

$$s^2 \equiv \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{\sigma_t^2}{N-1} \sum_{i=1}^N \frac{(x_i - \bar{x})^2}{\sigma_t^2}$$

possiamo introdurre una variabile aleatoria ausiliaria definita come  $\chi^2$  e

riscrivere  $s^2$  come:

$$s^2 \equiv \frac{\sigma_t^2}{N-1} \chi^2$$

di conseguenza  $V[s^2]$  è legata alla  $V[\chi^2]$ . Nell'ipotesi in cui le misure raccolte seguano una distribuzione di probabilità Gaussiana la variabile  $\chi^2$  segue la distribuzione del chi-quadro. Tale pdf è descritta da un solo parametro definito gradi di libertà e nel nostro caso tale parametro vale  $N-1$ .

Di conseguenza con questa nuova distribuzione possiamo dimostrare che  $s^2$  è uno stimatore non distorto.

Dim.

$$E[s^2] = E\left[\frac{\sigma_t^2}{N-1} \chi^2\right] = \frac{\sigma_t^2}{N-1} E[\chi^2] = \frac{\sigma_t^2}{N-1} (N-1) = \sigma_t^2$$

La sua varianza è data da:

$$V[s^2] = V\left[\frac{\sigma_t^4}{(N-1)^2} \chi^2\right] = \frac{\sigma_t^4}{(N-1)^2} V[\chi^2] = \frac{2\sigma_t^4}{(N-1)^2}$$



# CHAPTER 5

---

## Maximum Likelihood e Minimi Quadrati

---

### 5.1 La Verosomiglianza - Likelihood

Date N misure  $\{x_i\}_i^N$  queste si definiscono IID quando sono:

- **indipendenti:** l'esito di un campionamento non è influenzato da nessuno degli altri
- **identicamente distribuiti:** Tutte le misure seguono la stessa funzione di distribuzione di probabilità

$$pdf_x(x, \underline{\theta}) : R \rightarrow R^+$$

La funzione di probabilità congiunta (joint-pdf) di N campionamenti IID è il prodotto delle singole probabilità (poichè indipendenti tra loro):

$$pdf_{set}(x_1, \dots, x_N, \underline{\theta}) = \prod_i^N pdf_x(x_i, \underline{\theta}_t)$$

essa rappresenta la densità di probabilità da associare all'evento casuale consistente nell'estrarre un particolare set di dati, ed è una funzione definita su uno spazio N-dimensionale.

Se si sostituisce al valore vero  $\underline{\theta}_t$  il generico valore  $\hat{\underline{\theta}}$  stimato dalle N misure e se esse vengono considerate non più variabili casuali, ma costanti che sono state determinate dalle operazioni di misura, la precedente funzione prende il nome di *funzione di verosomiglianza* o **likelihood**.

$$L(\underline{x}, \hat{\theta}) = \prod_i^N pdf_x(x_i, \hat{\theta})$$

Rappresenta la densità di probabilità da associare all'evento casuale consistente nell'essere un certo  $\hat{\theta}$  il valore vero del nostro parametro, nell'ipotesi di avere già ottenuto un set di N misure.

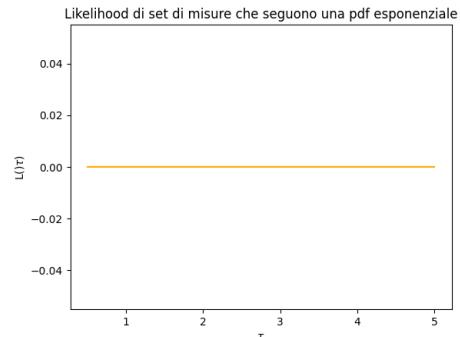
È possibile definire anche la **loglikelihood** ovvero:

$$l(\theta) = \log(L(\underline{x}, \hat{\theta})) = \log\left(\prod_i^N pdf_x(x_i, \hat{\theta})\right) = \sum_i \log(pdf_x(x_i, \theta))$$

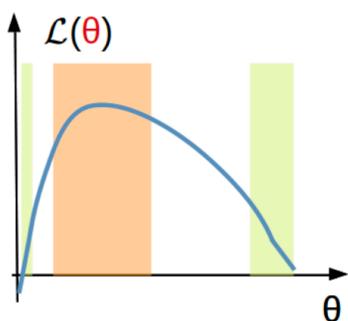
## 5.2 Comportamento della likelihood

Raccolte delle misure  $\{x_i\}_i^N$  IID, la funzione di likelihood  $L(\theta)$  restituisce la probabilità che si aveva di raccogliere tale campione assunta la conoscenza del parametro  $\theta$  (che per comodità in questo capitolo assumiamo in una sola dimensione).

Se la  $L(\theta)$  è all'incirca piatta vuol dire che si ha la stessa probabilità di ottenere un campionamento  $\{x_1, \dots, x_N\}$  per ogni valore di  $\theta$ , ciò significa che i campionamenti non forniscono molte informazioni sul parametro  $\theta$ .



Se la  $L(\theta)$  è una campana vuol dire si avrà una probabilità:



- Alta se il  $\theta_t$  del campione è nell'area arancione in figura
- bassa se  $\theta_t$  è nella regione verde
- massima se  $\theta_t$  coincide con il massimo della funzione di likelihood

di aver raccolto il set di dati in esame. Dunque i campionamenti forniscono informazioni su  $\theta$ . Più è stretta la campana e più piccolo è il range dei valori che massimizzano la probabilità di ottenere il campione raccolto. La larghezza della campana è legata all'incertezza con cui è possibile determinare il valore vero di  $\theta$ .

Quando definiamo una  $pdf(x, \theta)$  senza conoscere  $\theta$  si descrive un fascio di distribuzioni di probabilità e dunque un'infinità di modelli diversi. Cambiando la varianza di una distribuzione di cambia anche la funzione di likelihood.

### 5.3 Minimum Variance Bound

La funzione di likelihood  $L(\theta)$  può essere utilizzata per misurare l'informazione che i campionamenti contengono relativamente al parametro  $\theta$  che descrive il modello. L'informazione così ottenuta consente di valutare la minima varianza raggiungibile di uno stimatore di  $\hat{\theta}$ , ovvero dato un insieme di misure e un modello esiste un limite inferiore alla varianza raggiungibile.

#### 5.3.1 Informazione di Fischer

Vogliamo costruire una metrica che  $I(\theta)$  che fornisca l'ammontare d'informazione contenuta in una variabile casuale osservabile  $x$ , relativamente a un parametro non osservabile  $\theta$ , da cui dipende la  $pdf(x)$ . Tale funzione prende il nome d'**informazione di Fischer**.

$$I(\theta) = E[(\frac{\partial}{\partial \theta} \log(L(\theta)))^2] = -E[\frac{\partial^2}{\partial \theta^2} \log(L(\theta))]$$

essa gode delle seguenti proprietà:

- Associa un valore nullo se i dati sono irrilevanti per la stima di  $\theta$ .
- Aumenta al crescere della dimensione del campione ( purchè i dati siano rilevanti per la stima di  $\theta$ ).
- È legata alla precisione della stima: se l'informazione aumenta la varianza minima raggiungibile da uno stimatore  $\hat{\theta}$  diminuisce.

Più i valori assunti dalla likelihood risultano essere sparpagliati peggio è l'informazione che si ha a disposizione e dunque la precisione di stima di uno stimatore ne risentirà in egual modo. Viceversa più i valori risultano essere raggruppati in un certa regione di spazio e migliore sarà la precisione con cui si ottiene lo stimatore.

### 5.3.2 Teorema di Rao - Cramér

Per uno stimatore  $\hat{\theta}(x)$  consistente, bias  $b_n(\hat{\theta})$ ,  $V[\hat{\theta}] < \infty$  e non dipendente da  $x$ , definito su un dominio  $\Omega_\theta$  si ha che il **MVB (Minimum Variance Bound)** definisce una relazione tra la varianza di un qualsiasi stimato  $\hat{\theta}$  e la likelihood  $L(\underline{X}, \theta)$  nel seguente modo:

$$V[\hat{\theta}] \geq \frac{(1 - \frac{\partial}{\partial \theta} b_n(\theta))^2}{E[(\frac{\partial}{\partial \theta} \log(L(\theta)))^2]} = V[\hat{\theta}]_{min} \quad (5.1)$$

Al numeratore si ha una quantità che dipende dallo specifico stimatore  $\hat{\theta}$ , ma solo se questo è biased, mentre a denominatore si ha l'informazione di Fischer, dunque una quantità che non dipende dallo specifico stimatore ma solo da modello e dati.

Diciamo che uno stimatore è **efficiente** se la grandezza:

$$\xi(\hat{\theta}) = \frac{V[\hat{\theta}]_{min}}{V[\hat{\theta}]} = 1 \quad (5.2)$$

## 5.4 Maximum Likelihood

Considerato un campione di  $N$  misure IID, si definiscono estimatori di **Maximum Likelihood (ML)** dei parametri quei valori che massimizzano la funzione di likelihood. Data una likelihood differenziabile rispetto  $\underline{\theta}$  e il cui punto di massimo non è ai margini del range dei parametri, gli estimatori  $\hat{\theta}$  sono dati dalla soluzione delle equazioni differenziali:

$$\frac{\partial L}{\partial \theta_i} = 0 \quad i = 1, \dots, N \quad (5.3)$$

Se esiste più di un massimo, viene considerato il maggiore.

Con questa definizione di ML si considera il valore dello stimatore in corrispondenza del quale la probabilità associata al campionamento è la massima possibile.

Le proprietà degli estimatori ottenute con il metodo di ML valgono asintoticamente, ovvero con un numero sufficientemente elevato di misure. Alcune **proprietà notevoli** sono:

1. Consistente
2. Asintoticamente efficiente
3. Asintoticamente non distorto
4. Proprietà d'invarianza:  $\hat{\theta}_{ML}$  stimatore di  $\theta_t \Rightarrow g(\hat{\theta}_{ML})$  è stimatore di  $g(\theta_t)$ .

### Esempio proprietà d'invarianza

Consideriamo la distribuzione esponenziale  $f(x, \tau) = \frac{1}{\tau} e^{-\frac{x}{\tau}}$  e si abbiano N misure IID la likelihood rispetto a  $\tau$  è  $L(\tau) = \prod_i^N \frac{1}{\tau} e^{-\frac{t_i}{\tau}}$ .

Applichiamo il metodo di ML alla loglikelihood di  $\tau$ , dunque si ha che:

$$\sum_i^N \left( -\frac{1}{\tau} + \frac{t_i}{\tau^2} \right) = 0 \iff N = \sum_i^N \frac{t_i}{\tau} \iff \tau_{ML} = \frac{1}{N} \sum_i^N t_i$$

Posto  $\lambda(\tau) = \frac{1}{\tau}$  possiamo riscrivere la pdf come  $f(x, \lambda) = \lambda e^{-\lambda t}$  data la sostituzione vogliamo verificare che  $\lambda(\tau_{ML}) = \frac{1}{\tau_{ML}}$ .

Dim.

$$\frac{\partial l}{\partial \theta} = 0 \iff \sum_i^N \left( \frac{1}{\lambda} - t_i \right) = 0 \iff \frac{N}{\lambda} = \sum_i^N t_i \iff \lambda = \frac{1}{\tau}$$

Se si dimostra che  $\hat{\tau}$  è uno stimatore non distorto, è anche vero che  $\hat{\lambda}$  è anch'esso non distorto?

Poichè la relazione  $\hat{\lambda}(\hat{\tau})$  può essere non lineare non è detto che anche  $\hat{\lambda}$  sia non distorto.

Si può dimostrare che  $\hat{\lambda}$  è uno stimatore distorto di  $\lambda$  e dunque solo asintoticamente non lo è.

## 5.5 Varianza dello stimatore di ML - Metodo Grafico

Espandiamo fino al secondo ordine con un polinomio di Taylor la loglikelihood in un'intorno del parametro  $\hat{\theta}$  ottenuto con il metodo di ML.

$$\log(L(\theta)) \approx \log(L(\hat{\theta})) + \frac{1}{2} \frac{d^2}{d\theta^2} (\log(L(\theta)))|_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 \quad (5.4)$$

Se assumiamo che lo stimatore sia unbiased (il che per le proprietà precedente asintoticamente può esserlo) ed efficiente si ha dal teorema di Rao-Cramer che il secondo addendo è equivalente a:

$$\frac{d^2}{d\theta^2} (\log(L(\theta)))|_{\theta=\hat{\theta}} = -\frac{1}{\sigma_{\hat{\theta}_{ML}}^2}$$

E dunque possiamo riscrivere l'equazione (8.6) come:

$$\log(L(\theta)) \approx \log(L(\hat{\theta})) - \frac{1}{2\sigma_{\hat{\theta}_{ML}}^2} (\theta - \hat{\theta})^2 \quad (5.5)$$

Consideriamo un cambio di variabile da  $\hat{\theta}$  a  $\hat{\theta} \pm \sigma_{\hat{\theta}}$  dunque possiamo riscrivere la (8.7) come:

$$\log(L(\hat{\theta} \pm \sigma_{\hat{\theta}})) = \log(L(\hat{\theta})) - \frac{1}{2} \quad (5.6)$$

dunque da tale cambio di variabile si ha che la loglikelihood diminuisce della metà del suo valore di massimo. Si può dimostrare che la funzione di loglikelihood diventa una parabola (e la funzione di likelihood diventa una Gaussiana) per valori grandi del campione. Anche se la  $\log(L)$  non è parabolica si può utilizzare l'equazione (8.8) come la definizione di errore

## CHAPTER 5. MAXIMUM LIKELIHOOD E MINIMI QUADRATI

statistico. Determinando i punti d'intersezione con quanto definito in (8.8) si ottiene una stima della varianza dello stimatore  $\hat{\theta}$  usando il metodo di ML.

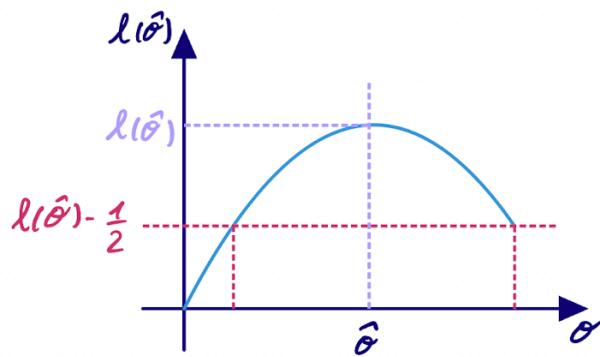


Figure 5.1: funzione di loglikelihood, stima della varianza

Con un numero finito di misure  $l(\theta)$  non è simmetrico e al crescere del numero di misure la forma funzionale assume un aspetto parabolico. Per simmetria la likelihood  $L(\theta)$  diventa una Gaussiana.

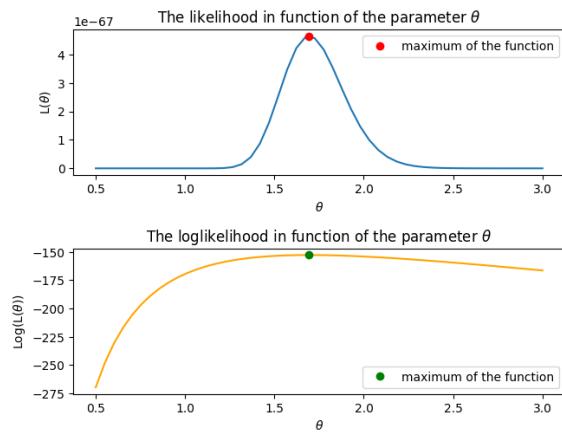


Figure 5.2: likelihood per un campione di 800 misure

## 5.6 Extended likelihood

Consideriamo una variabile aleatoria  $x$  distribuita secondo una pfd  $f(x, \underline{\theta})$ , dove i parametri  $\underline{\theta}$  non li conosciamo e ipotizziamo di avere un campione di  $N$  misure  $\{x_1, \dots, x_N\}$ . Spesso si ha che il numero di osservazioni  $n$  del campione è una variabile aleatoria che segue una pdf Poissoniana con frequenza media  $\lambda$ . Il risultato di un'esperimento può essere definito dalla variabile  $n$  (che è la dimensione del campione) e dalle  $n$  misure raccolte  $\{x_1, \dots, x_n\}$ .

Dunque possiamo definire la likelihood dei parametri  $\underline{\theta}$ , dato un numero di venti medio  $\lambda$  rispetto ad un campione di  $n$  misure.

$$L(\lambda, \theta) = \frac{\lambda^{-n}}{n!} e^{-\lambda} \prod_{i=1}^n f(x_i, \underline{\theta}) = \frac{e^{-\lambda}}{n!} \prod_{i=1}^n \lambda_i f(x_i, \underline{\theta}) \quad (5.7)$$

Tale funzione prende il nome di **extended likelihood**.

## 5.7 Varianza di uno stimatore per più parametri

La varianza di  $V[\hat{\theta}]$  di un vettore di parametri è definita dalla matrice di covarianza dove le sue componenti sono date da:

$$V_{ij} = \left[ -\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\hat{\theta}) \right]^{-1} \quad \text{per } i, j = 1, \dots, n \quad (5.8)$$

dove i termini per  $i \neq j$  sono dati da  $Cov[\theta_i, \theta_j]$ .

## 5.8 Intervallo di confidenza

Si è costruito uno stimatore  $\hat{\theta}$  per un parametro  $\theta$  che è una variabile aleatoria di conseguenza esiste una distribuzione  $pdf(\hat{\theta})$  che è caratterizzata da una media  $E[\hat{\theta}]$  e una varianza  $V[\hat{\theta}]$ . La **stima** è il valore che lo stimatore assume in corrispondenza di un campione di dati e lo definiamo  $\hat{\theta}^*$ .

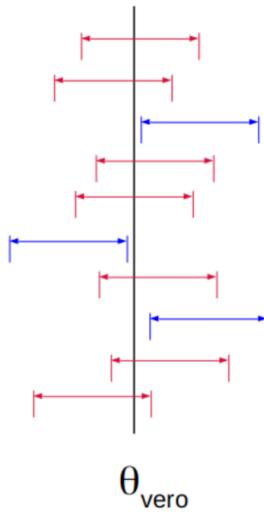
L'intervallo di confidenza è solitamente indicato come:

$$\hat{\theta}^* \pm \sigma \quad \text{oppure} \quad \hat{\theta}^* \begin{matrix} +\sigma_1 \\ -\sigma_1 \end{matrix} \quad (5.9)$$

la seconda notazione viene usata quando l'intervallo è asimmetrico.

In generale il risultato della stima

è un intervallo  $[a,b]$ , poiché è una variabile aleatoria si ha che gli estremanti sono variabili aleatorie e dunque per diversi campioni si hanno diversi intervalli di confidenza. A ciascun intervallo viene associata una probabilità (**livello di confidenza**) che misura quanto è buona la stima del valore vero.



$$P(\theta_t \in [a, b])$$

dove essa rappresenta la frazione delle volte in cui ripetendo l'esperimento, la stima restituisce un intervallo che contiene il valore vero.

In generale si vuole trovare un metodo che individui un intervallo  $[a,b]$  tale per cui la probabilità che  $\theta_t \in [a, b]$  sia pari a un certo valore  $\beta$  detto **livello di confidenza**. Un intervallo di confidenza così definito prende il nome di **intervallo di confidenza di Neyman**. Un intervallo particolare è quello dato da  $[\hat{\theta}^* - \sigma, \hat{\theta}^* + \sigma]$  e si definisce **intervallo centrale**.

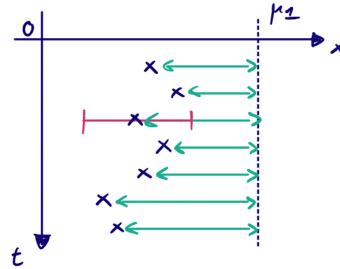
## 5.9 Metodo dei minimi quadrati

Si parte da un modello  $y = f(x, \underline{\theta})$  che definisce una relazione tra due grandezze fisiche  $x$  ed  $y$  e che dipende da  $N$  parametri  $\underline{\theta}$ , ipotizziamo di avere un campione di misure formato da punti  $\{(x_i, y_i)\}_i^N$  ciascuno di essi segue una  $\text{pdf}(x)$  e una  $\text{pdf}(y)$  di conseguenza sono variabili aleatorie che

hanno una loro incertezza  $\sigma_x$  e  $\sigma_y$ . Vogliamo trovare un metodo che ci permetta di stimare i parametri ignoti partendo dai dati sperimentali. Tale metodo prende il nome di **interpolazione** o **fit** dei dati. Per semplicità consideriamo il caso il cui l'incertezza sulle misure  $x$  sia trascurabile.

### 5.9.1 Funzionale $Q^2$

Partiamo considerando solo la variabile aleatoria  $x$  e ci domandiamo per quali valori del campione, si minimizza la distanza dalla media della popolazione  $E[x] = \mu$ ; per rispondere a tale domanda definiamo la misura della distanza definendo un funzionale:



$$Q^2(\{x_i\}_1^N, \mu) = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - \mu)^2}{\sigma_i^2} \quad (5.10)$$

Dunque ciascuna distanza è normalizzata alla larghezza della  $pdf(x_i)$  e quindi tiene conto delle fluttuazioni statistiche di quel determinato campionamento.

Per determinare il parametro cerchiamo quel valore  $\hat{\mu}_{MQ}$  per cui  $Q^2$  ammette un minimo assoluto nello spazio dei parametri (che nel nostro caso ha dimensione 1):

$$\frac{d}{d\mu} Q^2(\{x_i\}_1^N, \mu) = 0 \quad (5.11)$$

### 5.9.2 Valore di aspettazione di $\hat{\mu}_{MQ}$

Derivando la funzione di  $Q^2$  si ha:

$$\frac{d}{d\mu} Q^2 = -2 \sum_{i=1}^N \frac{(x_i - \hat{\mu}_{MQ})}{\sigma_i^2} = 0 \iff \hat{\mu}_{MQ} = \frac{\sum_{i=1}^N \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} = \bar{x}$$

dunque per un campione di  $N$  misure la stima del valore atteso della popolazione coincide con la media aritmetica pesata del campione.

### 5.9.3 Varianza dello stimatore $\hat{\mu}_{MQ}$

Consideriamo  $\hat{\mu}_{MQ} = \phi(x_1, \dots, x_N)$  e che il campione sia costituito da misure statisticamente indipendenti rispetto la variabile aleatoria  $x$ , la varianza dello stimatore sarà data da:

$$V[\hat{\mu}_{MQ}] = \sum_{i=1}^N \left( \frac{\partial \phi}{\partial x_i} \right)^2 \sigma_i^2 = \left[ \sum_{i=1}^N \frac{1}{\sigma_i^2} \right]^{-2} \sum_{i=1}^N \left( \frac{1}{\sigma_i^2} \right)^2 \sigma_i^2 = \frac{1}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} < \sigma_i^2 \quad \forall i$$

L'incertezza è dominata dalle misure con le  $\sigma_i$  più piccole.

## 5.10 Varianza di un stimatore usando i MQ - Metodo grafico

Con il metodo di maximum likelihood si sono cercati i valori dei parametri che rendevano massima la probabilità, dato un modello, di osservare i dati campionati. Con il metodo dei minimi quadrati invece si cerca il valore dei parametri che rende minima la distanza tra i dati campionati e il modello. Quando le  $pdf(x_i)$  sono Gaussiane i due stimatori coincidono.

$$L(\hat{\theta}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[ -\frac{(x_i - \hat{\theta})^2}{2\sigma_i^2} \right] = \text{cost} \cdot e^{-\frac{Q^2}{2}}$$

se passiamo alla loglikelihood

$$\log(L(\hat{\theta})) = \log(\text{cost}) - \frac{Q^2(\hat{\theta})}{2}$$

Stimando  $\hat{\theta}$  con il metodo di Maximum Likelihood si ha

$$\frac{\partial \log(L(\hat{\theta}))}{\partial \hat{\theta}} = -\frac{1}{2} \frac{\partial Q^2(\hat{\theta})}{\partial \hat{\theta}} \tag{5.12}$$

Quando le misure seguono una pdf Gaussiana e lo stimatore  $\hat{\theta}$  è efficiente e non distorto sappiamo che il MVB coincide con la varianza dello stimatore

ottenuto con il metodo di ML.

$$V[\hat{\theta}_{ML}] = -\frac{1}{\left.\frac{\partial^2 \log(L(\theta))}{\partial \theta^2}\right|_{\theta=\hat{\theta}_{ML}}} \quad (5.13)$$

Dalla relazione (5.12) possiamo riscrivere l'uguaglianza (5.13) come:

$$V[\hat{\theta}_{MQ}] = \frac{2}{\left.\frac{\partial^2 Q^2(\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}_{MQ}}} \quad (5.14)$$

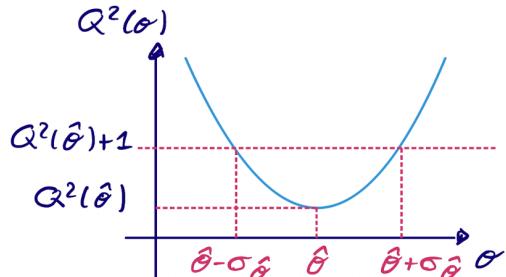
Sviluppando con Taylor al secondo ordine la funzione  $Q^2$  in un intorno di  $\hat{\theta}_{MQ}$  si ha:

$$Q^2(\theta) \approx Q^2(\hat{\theta}_{MQ}) + \frac{1}{2} \left. \frac{d^2 Q^2(\theta)}{d\theta^2} \right|_{\theta=\hat{\theta}_{MQ}} (\theta - \hat{\theta}_{MQ})^2 = Q^2(\hat{\theta}_{MQ}) + \frac{(\theta - \hat{\theta}_{MQ})^2}{V[\theta]}$$

Effettuando un cambio di coordinate  $\theta = \hat{\theta}_{MQ} \pm \sigma_{\hat{\theta}}$  si ha:

$$Q^2(\hat{\theta}_{MQ} \pm \sigma_{\hat{\theta}}) = Q^2(\hat{\theta}_{MQ}) + 1$$

Dunque considerando graficamente l'intersezione con  $Q^2(\theta)$  si determinano  $\hat{\theta}_{MQ} \pm \sigma_{\hat{\theta}}$ . Poichè per ipotesi le misure  $\{x_i\}_i^N$  seguono una distribuzione di probabilità Gaussiana e sono IID si ha che  $Q^2(\theta)$  segue la distribuzione di probabilità del chi quadro  $\chi^2(N - k)$  per  $E[Q^2] = N - k$  gradi di libertà. Se questo non avviene è perchè:



- si è avuta una fluttuazione statistica sfavorevole
- il modello non descrive i dati
- il modello è corretto, ma i dati sono stati raccolti in modo errato

- il modello è corretto, ma le incertezze attribuite ai dati sono sbagliate (sono sopra o sottovalutate).

## 5.11 Modelli lineari nei parametri

Un modello si definisce **lineare** se due grandezza fisiche  $x$  ed  $y$  sono legate da una relazione lineare rispetto a parametr  $\underline{\theta}$ .

$$y_0^i = \sum_{j=1}^N \theta_j h_j(x_i) \quad (5.15)$$

Assumiamo che l'incertezza sulla variabile  $x$  sia ininfluente, mentre  $y$  è una variabile aleatoria le cui misure sono IID con varianza  $V[y_i] = \sigma_i^2$ . Definiamo una variabile aleatoria asuiliaria  $\epsilon$  che possiede la stessa distribuzione di probabilità di  $y$ . Riscriviamo (5.15) come:

$$y_i = \sum_{j=1}^N \theta_j h_j(x_i) + \epsilon_i \quad (5.16)$$

Assumiamo che per le  $\epsilon$  valgano le seguenti proprietà:

$$E[\epsilon_i] = 0 \quad V[\epsilon_i] = V[y_i] = \sigma_i^2$$

In questo modo la pdf( $y$ ) coincide con la pdf( $\epsilon$ ). La variabile  $\epsilon$  rappresenta l'errore statistico sulla misura, mentre le  $y_0$  sono il valore vero della misura restituito dal modello.

Applichiamo il metodo dei minimi quadrati, andando a stimare quei valori dei parametri  $\underline{\theta}$  che minimizzano la distanza tra il valore vero restituito dal modello lineare e i dati campionati sperimentalmente, pesando gli scarti quadratici rispetto la varianza delle distribuzioni di probabilità delle singole  $y_i$ .

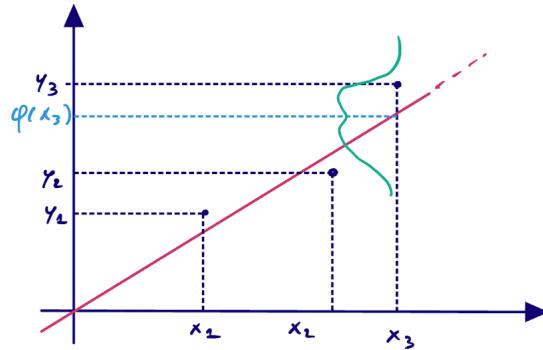


Figure 5.3:  $\psi(x_3)$  rappresenta il valore vero della misura rispetto al modello, mentre  $y_3$  il dato campionario su cui è presente un errore statistico.

Per farlo utilizziamo il funzionale  $Q^2$  riscrivendolo come:

$$Q^2 = \sum_{i=1}^N \frac{(y_i - y_0(x_i, \theta))^2}{\sigma_i^2} = \sum_{i=1}^N \frac{(y_i - \sum_{j=1}^N \theta_j h_j(x_i))^2}{\sigma_i^2} = \sum_{i=1}^N \frac{\epsilon_i^2}{\sigma_i^2} \quad (5.17)$$

### 5.11.1 Stime dei parametri

Per un vettore di parametri  $\underline{\theta}$  di dimensione N, avremo un sistema di K equazioni:

$$\nabla Q^2(\underline{x}, \underline{\theta}) = \begin{cases} \frac{\partial Q^2}{\partial \theta_1} = 0 \\ \vdots \\ \frac{\partial Q^2}{\partial \theta_n} = 0 \end{cases} \quad (5.18)$$

Per una singola riga si ha:

$$\frac{\partial Q^2}{\partial \theta_j} = \sum_{i=1}^N \frac{-2}{\sigma_i^2} \left( \frac{y_i - \sum_{j=1}^N \theta_j h_j(x_i)}{\sigma_i} \right)^2 h_j(x_i) = 0 \quad \forall i = 1, \dots, N \quad (5.19)$$

Per comodità di esposizione riscriviamo l'espressione (5.16) in forma vettore-

riale:

$$\underline{y} = H\underline{\theta} + \underline{\epsilon} \quad (5.20)$$

ed essendo una modello a più parametri si ha la matrice di covarianza  $V[\epsilon]$ , possiamo considerare la matrice in funzione di  $\epsilon$  poichè  $V[y]$  e  $V[\epsilon]$  sono legate da una traslazione rigida e dunque risultano essere identiche. In generale tale matrice non è diagonale, ma poichè si è assunto che le misure siano IID la matrice è diagonale.

$$V[\epsilon] = \begin{bmatrix} \sigma_i^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

Riscrivendo  $Q^2 = \underline{\epsilon}^T V^{-1} \underline{\epsilon}$  in forma matriciale e risolvendo il sistema (5.19) si stimano i parametri come:

$$\hat{\underline{\theta}}_{MQ} = (H^T V^{-1} H)^{-1} \cdot H^T V^{-1} \underline{y} \quad (5.21)$$

dove  $V(\hat{\underline{\theta}}) = (H^T V^{-1} H)^{-1}$  dove nella matrice  $H$  è contenuta l'informazione delle misure.

Resta da verificare che lo stimatore sia non distorto ovvero  $E(\hat{\underline{\theta}}) = \underline{\theta}_t$ :

$$E[\hat{\underline{\theta}}_{MQ}] = (H^T V^{-1} H)^{-1} \cdot H^T V^{-1} E[\underline{y}] = (H^T V^{-1} H)^{-1} \cdot (H^T V^{-1} H) \underline{\theta}_t$$

dove:

$$E[\underline{y}] = E[H\hat{\underline{\theta}} + \underline{\epsilon}] = E[H\hat{\underline{\theta}}] + E[\underline{\epsilon}] = H E[\hat{\underline{\theta}}] = H \underline{\theta}_t$$

duque lo stimatore ottenuto è non distorto.

A differenza del metodo di ML che gode di buone proprietà asintoticamente si ha che quello dei MQ le ha per un numero finito di misure.

## 5.12 Sovrastima degli errori

Ipotizziamo che le incertezze siano sovrastimate di un fattore  $\alpha$  ovvero  $\sigma_i^* = \alpha \cdot \sigma_i$ , di conseguenza possiamo riscrivere la matrice di covarianza come:

$$W[y^*] = \alpha^2 \cdot \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix} = \alpha^2 \cdot V[y]$$

Se sostituiamo tale ipotesi all'interno dell'equazione (5.21) si ha:

$$\begin{aligned} \hat{\underline{\theta}} &= (H^T W^{-1} H)^{-1} \cdot H^T W^{-1} \underline{y} = \alpha^2 \cdot \frac{1}{\alpha^2} \cdot (H^T V^{-1} H)^{-1} \cdot H^T V^{-1} \underline{y} = (5.22) \\ &= (H^T V^{-1} H)^{-1} \cdot H^T V^{-1} \underline{y} \end{aligned}$$

dunque la stima del parametro  $\hat{\underline{\theta}}_{MQ}$  rimane invariata anche se si sono sovrastimate le incertezze. Quella che cambia è la varianza del parametro stimato infatti:

$$V[\hat{\underline{\theta}}_{MQ}] = (H^T W^{-1} H)^{-1} = \alpha^2 (H^T V^{-1} H)^{-1} \quad (5.23)$$

e dunque viene sovrastimata di un fattore  $\alpha^2$ .

### 5.12.1 Relazione tra il numero di parametri e il campione

Se si hanno K parametri e campione di dimensione N dove K=N si ha che:

- il sistema ammette soluzione, e si ha una curva che passa per tutti i punti;
- se il sistema non ammette soluzione allora i dati falsificano il modello.

### 5.12.2 Incertezze sulla variabile indipendente

Nel caso in cui il modello sia lineare rispetto ai parametri, per esempio una retta  $y(x, a, b) = a + bx$  e siano presenti delle incertezze sulla variabile indipendente x, si propaga l'errore sulle  $y_i$  l'errore di  $\sigma_{x_i}$  ottenendo un errore

complessivo  $\sigma_i^2 = \sigma_{y_i}^2 + b\sigma_{x_i}^2$  e dunque il funzionale  $Q^2$  si riscrive come:

$$Q^2 = \frac{1}{N} \sum_{i=1}^N \frac{(y_i - \bar{y})^2}{\sigma_{y_i}^2 + b\sigma_{x_i}^2} \quad (5.24)$$

### 5.12.3 Stima del fattore di sovrastima $\alpha$

Ipotizziamo che gli  $\epsilon$  seguano una distribuzione di probabilità Gaussiana e di avere N misure indipendenti allora il funzione  $Q^2$  definito come l'ultima uguaglianza della (5.17) è segue la distribuzione di  $\chi^2(N - K)$  per N-K gradi di libertà dove K è il numero di parametri da stimare. In forma matriciale possiamo riscrivere rispetto alla matrice di sovrastima W come:

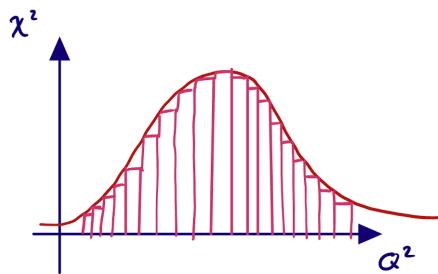
$$\hat{Q}^2 = \epsilon^T W^{-1} \epsilon = \frac{1}{\alpha^2} \cdot \epsilon^T W^{-1} \epsilon = \alpha^{-2} \cdot Q^2 \quad (5.25)$$

Vogliamo determinare  $\alpha^2$  affinché  $\hat{Q}_{min}^2 = E[\hat{Q}^2] = N - K$  e dunque:

$$\alpha^2 = \frac{\hat{Q}_{min}^2}{N - K} \quad (5.26)$$

Il fattore di scala così determinato non costituisce la sovrastima "reale" di cui si sono sbagliate le misure, poiché  $\hat{Q}_{min}^2(\underline{x}, \hat{\theta}_{MQ})$  per  $\hat{\theta}_{MQ}$  fissato è anch'esso una variabile aleatoria che segue la distribuzione di  $\chi^2(N - K)$  di conseguenza la (5.26) stima un solo valore rispetto

al campione a disposizione, ma  $\alpha^2$  è una variabile aleatoria che segue la medesima distribuzione di  $\hat{Q}_{min}^2(\underline{x}, \hat{\theta}_{MQ})$  e dunque per determinarlo bisogna ricostruire la distribuzione del  $\chi^2$ .



## 5.13 Teorema di Gauss-Markov

Si consideri un insieme di variabili aleatorie statisticamente indipendenti  $\{(x_i, y_i)\}_i^N$  tali che sono legate tra loro da una relazione lineare rispetto ai parametri  $y_i = \psi(x_i, \underline{\theta}) + \epsilon_i$  dove  $E[\epsilon_i] = 0$  e  $V[\epsilon_i]$  finita  $\forall i$  (proprietà di

**omoschedasticità**) ed inoltre  $y_i$  indipendenti dai parametri  $\underline{\theta} \Rightarrow$  si ha che lo stimatore  $\hat{\underline{\theta}}$  ottenuto con il metodo dei minimi quadrati è non distorto e ha varianza minima fra tutti gli estimatori lineari.

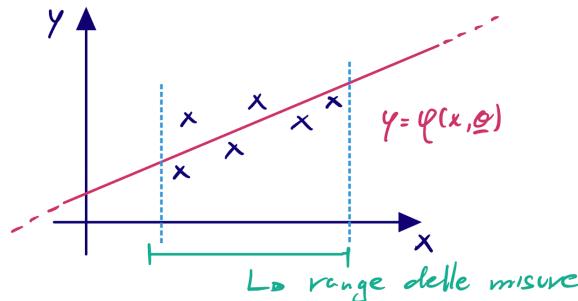
### Osservazioni

- $\epsilon_i$  non è necessario che siano distribuite secondo una Gaussiana
- il teorema non ci dice che si è trovato lo stimatore più **efficiente** fra tutti gli estimatori possibili, ma sicuramente quello piú efficiente tra quelli lineari.

È possibile trovare uno stimatore più efficiente per i parametri  $\underline{\theta}$  però dovremmo determinarlo con una funzione non lineare e ammesso che questa esista non è garantito che lo stimatore ottenuto sia non distorto.

Gli estimatori descritti dal teorema di Gauss-Markov vengono definiti **B.L.U.E** (**Best Linear Unbiased Estimator**).

## 5.14 Interpolazione ed Estrapolazione



Assumendo che i parametri  $\underline{\theta}$  siano già stati determinati. **L'interpolazione** è la determinazione del valore di  $y$  mediante la funzione  $\psi(x, \underline{\theta})$  per una misura  $x$  contenuta all'interno del range delle misure, ovvero l'intervallo con-

tente i dati del campione. Si definisce **estrapolazione** il valore di  $y$  per qualsiasi altro valore di  $x$  non compreso all'interno del range delle misure definito sperimentalmente. Per una valore ottenuto da una retta di regressione lineare, l'errore è legato all'incertezza nella stima dei parametri di  $\underline{\theta}$ . Di conseguenza:

$$V[y]_i = \nabla\psi(x, \underline{\theta})^T \cdot \text{Cov}[\hat{\underline{\theta}}_{MQ}] \cdot \nabla\psi(x, \underline{\theta}) \quad (5.27)$$

Per esempio in un caso a due parametri si avrà:

$$V[y]_i = \begin{bmatrix} \frac{\partial \psi}{\partial \theta_1} & \frac{\partial \psi}{\partial \theta_2} \end{bmatrix} \cdot \begin{bmatrix} \sigma_{\theta_1}^2 & Cov(\theta_1, \theta_2) \\ Cov(\theta_2, \theta_1) & \sigma_{\theta_2}^2 \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial \psi}{\partial \theta_1} \\ \frac{\partial \psi}{\partial \theta_2} \end{bmatrix}$$

che coincide con la formula di propagazione degli errori presente al capitolo 3.  
In generale l'errore aumenta allontanandosi dalla regione di campionamento  
( quindi quando si passa dall'interpolazione all'estrapolazione).

### Esempio di fit lineare

Consideriamo di avere un set di misure  $\{(x_i, y_i)\}_i^N$  IID con incertezza  $\sigma_{y_i}$  sulla variabile aleatoria y, e che il modello è quello di una retta  $y = a + bx$ , stimiamo i parametri a e b usando il metodo dei minimi quadrati. Avremo che:

$$\vec{\theta} = \begin{bmatrix} a \\ b \end{bmatrix} \quad \vec{h} = \begin{bmatrix} 1 \\ x \end{bmatrix}$$

e poichè le misure sono indipendenti tra loro, la matrice di covarianza è diagonale. Applicando il metodo dei MQ si stimano:

$$\hat{a} = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \sum_{i=1}^N \frac{y_i - \hat{b}x_i}{\sigma_i^2} = \bar{y} - \hat{b}\bar{x}$$

$$\hat{b} = \frac{1}{\sum_{i=1}^N \left(\frac{x_i}{\sigma_i}\right)^2} \sum_{i=1}^N \frac{y_i x_i - \hat{a}x_i}{\sigma_i^2} = \frac{\bar{xy} - \bar{y}\bar{x}}{\bar{x^2} - \bar{x}^2}$$

e le incertezze associate alle stime sono date dalla matrice di covarianza associate ai parametri:

$$V[\hat{\theta}] = \begin{bmatrix} \sigma_a^2 & Cov[a, b] \\ Cov[b, a] & \sigma_b^2 \end{bmatrix} = \frac{\sigma^2}{N(\bar{x^2} - \bar{x}^2)} \cdot \begin{bmatrix} \bar{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

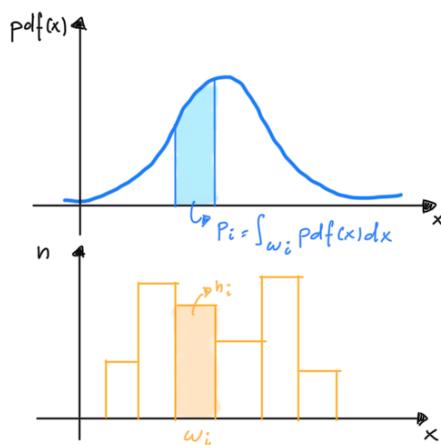
I termini in  $\boxed{\text{ }}$  viene definito **dispersione delle misure**. Si osserva che non necessariamente i parametri sono scorrelati tra loro. Più le misure sono sparpagliate meglio è la precisione con cui si stimano i parametri, poiché la banda che definisce le possibili rette che fittano i dati diventa più sottile, inoltre anche il numero di misure fornisce un contributo alla precisione con cui si determinano gli stimatori.

Se calcoliamo il valore assunto dalla funzione in punto  $x_0$  rispetto ai parametri  $\hat{a}$  e  $\hat{b}$  si ricava che l'errore su  $y_0$  è dato da:

$$V[y_0] = \frac{\sigma^2}{N} \left[ 1 + \frac{(x_0 - \bar{x})^2}{(\bar{x}^2 - \bar{x}^2)^2} \right]$$

Più  $x_0$  si allontana da  $\bar{x}$  più è grande la varianza su  $y_0$ .

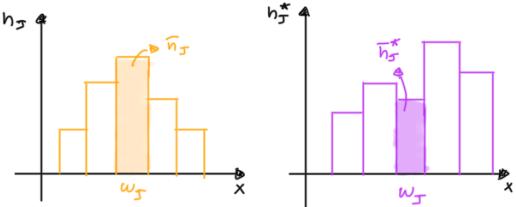
## 5.15 Fit d'istogrammi



$\sum_{j=1}^N n_j = N$  numero totale di misure. Poiché le misure sono IID a ciascuna bin associamo una probabilità:

$$p_j = \int_{\omega_j} pdf(x, \vec{\theta}) dx \quad \forall j = 1, \dots, k$$

Partendo dalla probabilità associata a ciascun bin  $\omega_j$  costruiamo un istogramma i cui conteggi attesi sono dati da  $E[n_j] = N \cdot p_j(\vec{\theta})$ . Le misure di conteggio dell'istogramma



definito dai dati del campione seguono una joint-pdf  $pdf(n_1, \dots, n_k, p_1, \dots, p_k)$  che definisce una distribuzione multinomiale. Per  $p \rightarrow 0$  le misure di conteggio sono statisticamente indipendenti tra loro di conseguenza la multinomiale può essere riscritta come prodotto di distribuzioni di probabilità binomiali.

$$pdf(n_1, \dots, n_k, p_1, \dots, p_k) = \prod_{i=1}^k pdf(n_i, p_i)$$

Nel caso in cui il valore di aspettano del conteggio dei bin sia costante possiamo approssimare le singole pdf binomiali come delle Poissoniane dove  $\lambda_i(\vec{\theta}) = N * p_i(\vec{\theta})$  rappresenta la frequenza media di eventi della distribuzione di Poisson di conseguenza la joint-pdf diventa :

$$\prod_{i=1}^k pdf(n_i, p_i) = \prod_{i=1}^k \frac{\lambda_i(\vec{\theta})^{n_i}}{n_i!} e^{-\lambda_i} = L(n_1, \dots, n_k, \theta)$$

Che coincide con la extended likelihood.

Non solo si conosce la pdf associata ai dati ma anche il loro valore di aspettazione e varianza di conseguenza è possibile applicare il metodo di ML e dei MQ.

## Osservazioni

Utilizzare le misure binate ci permette di ridurre le dimensioni dei dati con cui si lavora, anche se si ha una perdita d'informazione questa viene bilanciata da una maggiore semplicità e dalla possibilità di usare differenti tecniche per la stima dei parametri, come per esempio quella di ML che per campioni molto grandi risulta essere da un punto di vista computazionale dispendiosa.

### 5.15.1 Binned Data - Minimi Quadrati

Applichiamo il metodo dei minimi quadrati confrontando i conteggi dell'istogramma dei dati con l'istogramma dei valori attesi. Dato un campione  $\{x_i\}_i^N$  IID dopo averli rappresentati in un istogramma, associamo al valore atteso di conteggi per ciascun bin l'espressione:

$$\mu_i = E[n_j] = N \cdot \int_{\omega_j} pdf(x, \vec{\theta}) dx = N \cdot p_j(\vec{\theta})$$

Scrivendo il funzionale  $Q^2$  come:

$$Q^2 = \sum_{i=1}^N \frac{(n_i - E[n_i])^2}{V[n_i]} = \sum_{i=1}^N \frac{(n_i - Np_i(\vec{\theta}))^2}{Np_i(\vec{\theta})}$$

Assumendo che i conteggi  $n_i$  seguano una pdf Poiss( $n_i, \mu_i$ ) (la discussione che segue risulterebbe verificata ugualmente anche se le pdf fossero binomiali). Per  $n_i \rightarrow \infty$  possiamo approssimare la poissoniana a una gaussiana con valore atteso e varianza pari a  $\mu_i$  di conseguenza  $Q^2$  dipendendo da variabili aleatorie distribuite secondo una gaussiana, ed essendo lui stesso una variabile casuale si ha che segue la distribuzioni di  $\chi^2(k - s)$  dove  $s$  è il numero di parametri. Definiamo dunque il  $\chi^2$  nella forma di **Neyman**:

$$\chi^2_{Neyman} = \sum_{i=1}^N \frac{(n_i - \mu_i(\vec{\theta}))^2}{n_i} = \sum_{i=1}^N \frac{(O_i - E_i)^2}{O_i} \quad (5.28)$$

Dove per  $n_i$  grande si è approssimata la  $V[n_i] \approx n_i$ . La necessità di approssimare i dati a una Gaussiana e la varianza al numero di conteggi di un bin, è determinata dal fatto che così facendo il teorema di Gauss-Markov risulta soddisfatto poichè tra le ipotesi è necessario che i momenti non dipendano dai parametri. In conclusione si procede a minimizzare il  $\chi^2_{Neyman}$  per determinare i parametri  $\vec{\theta}$ .

#### Procedura di fit di un istogramma

La procedura di fit di un istogramma con il metodo del  $\chi^2$  di Neyman prevede quindi di:

- verificare che sia rispetta l'ipotesi  $\text{pdf}(n_i)$  gaussiana;
- associare a  $n_i$  un errore  $\sqrt{n_i}$
- calcolare per ogni valore del parametro  $\theta_i$  i valori attesi  $\mu_i$  dei conteggi in ciascun bin;
- minimizzare il  $\chi^2_{Neyman}$  trovando il valore di  $\theta$  per cui l'accordo valori misurati - valori attesi è il migliore;
- all'occorrenza effettuare un test del  $\chi^2$

Se  $N \cdot p_i$  è piccolo, la distribuzione binomiale/poissoniana non è approssimabile a una Gaussiana. Ne consegue che  $Q^2$  non è approssimabile ad un  $\chi^2$  e la varianza non è approssimabile con  $n_i$ , in questi casi il metodo dei MQ non è utilizzabile.

### 5.15.2 Binned Data - Maximum Likelihood

Fare il fit di un'istogramma quando non vale l'approssimazione Gaussiana prevede necessariamente l'utilizzo della binned Maximum Likelihood. Analogamente a quanto discusso a inizio sezione si procede a passare discretizzare il modello. Dove la funzione di likelihood associata al campione è data da

$$L(n_1, \dots, n_k, \vec{\theta}) = \prod_i^k \text{Bin}(n_i, p_i(\vec{\theta}, N))$$

e si procede a massimizzare la funzione di likelihood:

$$\frac{\partial L(\vec{n}, \vec{\theta})}{\partial \theta_j} = 0 \quad \forall j = 1, \dots, k$$

Ad un bin-size deve essere associata una probabilità  $p_i(\vec{\theta})$  piccola. Non è necessario che i bin abbiano tutti la stessa dimensione, possiamo sceglierle in modo che  $N \cdot p_i$  sia grande e quindi valga l'approssimazione Gaussiana.

L'operazione di binning, fa perdere informazione, questo si può riflettere sull'incertezza del parametro stimato , ma anche sulla capacità di verificare

## CHAPTER 5. MAXIMUM LIKELIHOOD E MINIMI QUADRATI

che i dati siano ben descritti dal modello. In questi casi è utile studiare la dipendenza del risultato dal binning.

# CHAPTER 6

---

## Test d'ipotesi

---

### 6.1 Test del $\chi^2$

Consideriamo un campione di  $N$  misure  $\{(x_i, y_i)\}_i^N$  IID legate tra loro da una funzione  $y = \psi(x, \vec{\theta})$  avremo che le misure campionate rispetto alla variabile aleatoria  $y$ , possono essere riscritte come  $y_i = \psi(x, \vec{\theta}) + \epsilon_i$  dove  $\epsilon$  ipotizziamo essere una variabile aleatoria la cui  $pdf(\epsilon)$  segue una distribuzione di probabilità Gaussiana. Nell'ipotesi in cui valga il TCL per le  $\epsilon_i$ , si ha che il  $Q^2$  associato alle misure e il modello segue la distribuzione di  $\chi^2(N - k)$ .

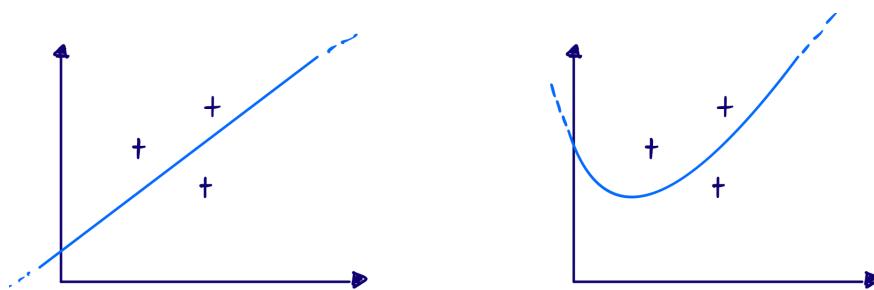
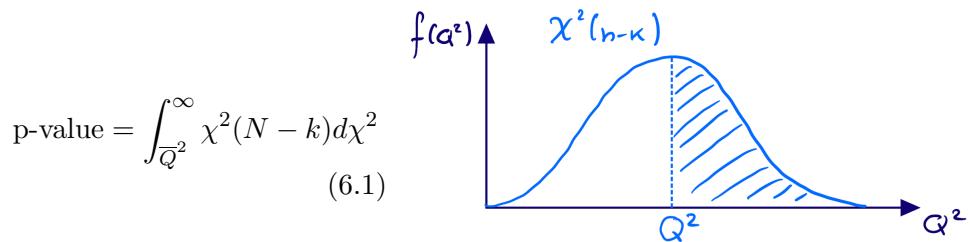


Figure 6.1: Due modelli differenti che interpolano lo stesso campione di dati.

Nel caso di destra in figura 6.1 gli scarti quadratici sono minori, mentre in quello di sinistra sono più grandi, di conseguenza possiamo aspettarci che il valore di aspettazione della distribuzione di  $\chi^2$  del modello di sinistra sia più grande di quello di destra. Definiamo il modello non corretto (quello

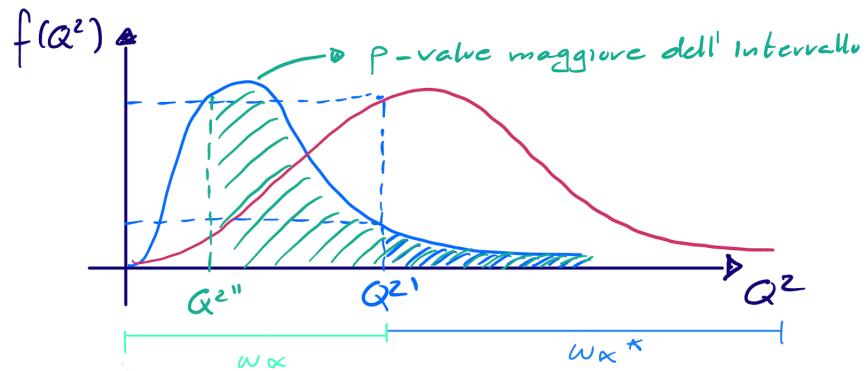
di sinistra)  $H_1$  e il modello corretto (quello di destra)  $H_0$ , la domanda che possiamo porci è: " Se partiamo da due ipotesi  $H_1$  e  $H_0$  e non sappiamo quale delle due sia corretta, come facciamo a determinare quella che descrive meglio la realtà sperimentale? ".

Introduciamo una nuova quantità definita **p-value** che ha la seguente espressione :



la quantità così definita risulta essere una misura di probabilità. Per rispondere alla domanda precedente fissiamo una soglia di tolleranza del p-value oltre alla quale i valori ottenuti risultano essere dei **falsi negativi**. Riprendendo i modelli  $H_0$  e  $H_1$  che definiamo rispettivamente **null hypothesis** e **alternative hypothesis** poichè  $H_1$  ha un valore di aspettazione più grande di  $H_0$ , fissata una soglia del p-value, e definita una statistica  $x$  associata al  $\chi^2(x|N - k)$  avremo che:

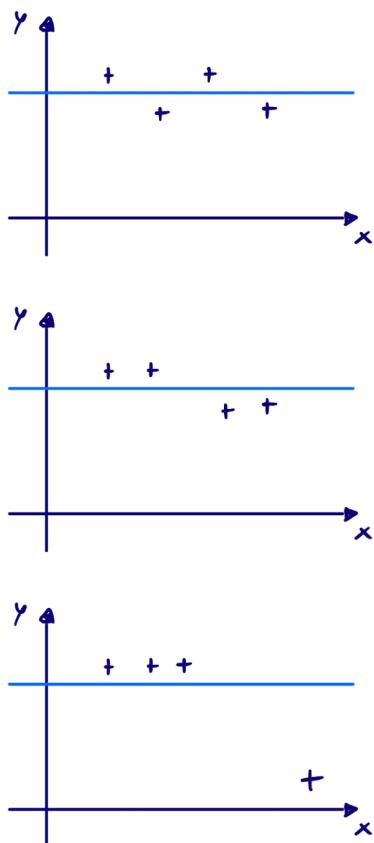
- $H_0$  è rigettata se  $x$  cade nella regione in azzurro  $\omega_\alpha^*$  in figura 6.2
- $H_0$  è accettata se  $x$  cade nella regione in verde  $\omega_\alpha$  nella figura sottostante.



Su quanto discusso fino ad ora possiamo fare le seguenti osservazioni:

- I minimi quadrati calcolano il  $Q^2$  (e anche la ML calcola i parametri  $\hat{\theta}$  da cui si pu calcolare il  $Q^2$ );
- Test del  $\chi^2$  funziona perch conosciamo la p.d.f. del  $Q^2 \Rightarrow \sigma^2$  deve essere nota e ben stimata;
- Test del  $\chi^2$  un test integrale  $\Rightarrow$  Somma gli scarti su tutti gli eventi.

### Esempio



Assumiamo che il p-value sia accettabile ovvero maggiore dell'intervallo di confidenza. Consideriamo gli stessi punti riorganizzati in un modo diverso, ma con stesso scarto quadratico. I due set di dati con y differenti hanno lo stesso p-value.

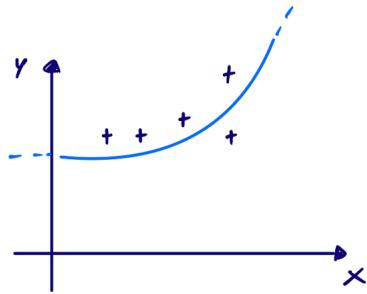
Essendo che il test del  $\chi^2$  ha una forma integrale, se graficamente si osserva una distribuzione differente delle misure, il test non lo tiene in considerazione.

Se consideriamo una dispersione delle misure come nella terza figura e assumiamo che abbia il medesimo p-value delle altre due, si ha che il test non considera che i dati diminuiscono di valore lungo l'asse delle ordinate e dunque si ottiene un **falso positivo**, ovvero p-value è verificato, ma il modello non descrive adeguatamente il comportamento

dei dati sperimentali.

Il fatto che il test del  $\chi^2$  abbia forma integrale limita la generalità con cui possiamo decidere se il risultato ottenuto sia affidabile o meno.

Ipotizziamo di avere un fit che ha  $Q^2 = 0$  e p-value-1, ovvero i dati vengono interpolati perfettamente, questo non è un buon risultato. Si ha un caso di overfitting, dove si sono introdotti così tanti parametri che il risultato del fit si è completamente adattato alle misure, perdendo qualsiasi capacità di generalizzare il modello.



### 6.1.1 Applicazione del test di $\chi^2$

Se il modello è corretto  $y = \psi(x, \vec{\theta})$ , allora il metodo dei MQ fornisce una stima dei parametri  $\vec{\theta}$  che lo descrivono. Il valore stimato dei parametri rappresenta il punto di minimo del funzione di  $Q_{min}^2 = Q^2(x, \vec{\theta}_{MQ})$  rispetto al campione sperimentale, tale punto di minimo coincide anche con il massimo della distribuzione di  $\chi^2$  se  $pdf(\epsilon)$  seguono una distribuzione gaussiana, che è dato da  $E[Q^2] = N - k$  e quindi  $Q_{min}^2 = N - k$ .

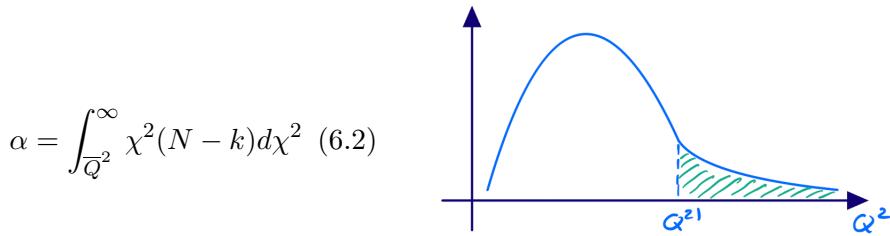
Il  $\chi^2$  ridotto è definito come  $\chi_0^2 = \frac{\chi^2}{ndof}$  ciò implica che per  $\chi_{min}^2 = Q_{min}^2$  il ridotto è  $\chi_0^2 \sim 1$ .

Se il valore del  $\chi_{min}^2$  è lontano dal suo valore di aspettazione  $N-k$ , possiamo concludere che alcune delle ipotesi precedenti non sia corrette e dunque **i dati non confermano il modello**.

## 6.2 Errori di test statistici

### 6.2.1 Errori del I° tipo

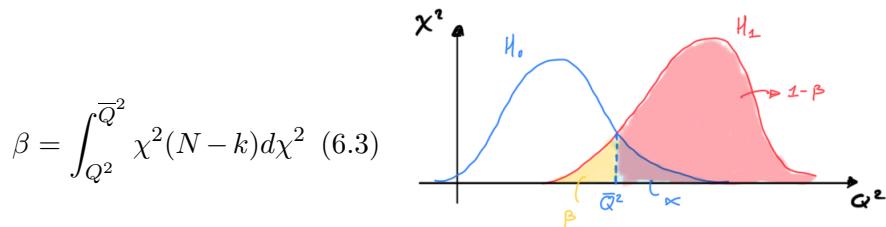
Un errore del primo tipo rappresenta il numero di casi veri per la null hypothesis  $H_0$  che scartiamo fissata una soglia del p-value.



Il termine  $\alpha$  prende il nome di **size del test**.

### 6.2.2 Errori del II° tipo

Un errore del secondo tipo rappresenta la probabilità di accettare  $H_0$  quando è vera  $H_1$ , in questo caso si parla di **falsi positivi**.



Si sta commettendo un errore poichè se  $H_1$  è la forma funzionale sbagliata il fit dei dati supera ugualmente il test del  $\chi^2$ .

Il termine  $1 - \beta$  prendere il nome di **power del test** e restituisce la probabilità di rifiutare  $H_0$  quando  $H_1$  è vera.

Fissate le due ipotesi alternative e definiti gli intervalli di confidenza, se si assume che l'errore di tipo uno sia quello più grave si procede scegliendo la percentuale di falsi negativi che si reputa accettabile e si cerca di definire gli intervalli  $\omega_\alpha$  e  $\omega_\alpha^*$  in modo tale che  $\beta$  sia il minore possibile (minor caso di falsi positivi). Il test così descritto viene definito il più potente per un determinato valore di soglia del p-value.

## 6.3 Test di Kolgomorov-Smirnov

Consideriamo un insieme di  $N$  misure della stessa grandezza fisica  $X$  vogliamo testare la null hypothesis  $H_0$  che siano campionamenti di una determinata pdf che prendiamo come riferimento. Una possibilità è di usare il test del  $\chi^2$  applicandolo agli istogrammi costruiti con le misure raccolte e la pdf-modello. Tale procedura è corretta, ma richiedere di binnare i dati e dunque si ha una perdita d'informazione, inoltre l'esito del test può dipendere dal binning scelto per la costruzione degli istogrammi.

L'alternativa è data dal test di **Kolgomorov-Smirnov** che confronta le due distribuzioni cumulative (dati - pdf-modello) e in questo modo sfrutta tutta l'informazione contenuta nei dati. Tale test è di tipo non parametrico ovvero non richiede la costruzione di stimatori rispetto ai dati raccolti sperimentalmente e viene utilizzato per variabili aleatorie continue.

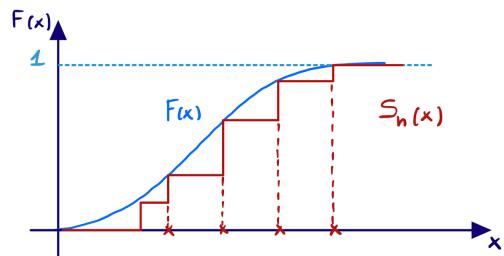
### 6.3.1 Costruzione del test

Date  $N$  misure ordinate in senso crescente, ricostruiamo la distribuzione cumulativa della pdf-modello, definendo una funzione a gradini  $S_n(x)$  rispetto ai dati del campione, tale funzione prende il nome di **EDF (Empirical Distribution Function)**, la scelta ricade su una funzione di questo tipo poichè sono presenti dei buchi nell'informazione (campione) raccolta. La forma dell'EDF è data:

$$S_n(x) = \sum_{i=1}^n I(x_i \leq x) \quad (6.4)$$

dove la funzione  $I(x_i \leq x)$  prende il nome di **indicator function** ed è espressa come:

$$I(x) = \begin{cases} 1 & x \leq x_{i+1} \\ 0 & \text{altrimenti} \end{cases}$$



Ci si domanda quanto bene  $S_n$  approssimi la cumulativa  $F(x)$ , la risposta è che dipende dal numero di misure contate prima del gradino successivo e quindi da come si scelgono gli intervalli. Per valutare l'approssimazione per ogni punto distinto che costituisce un estremante degli intervalli  $\tilde{x}$  valutiamo l'estremo superiore della differenza tra la EDF e la pdf con cui vogliamo confrontarla:

$$D_n = \sup_x |S_n(x) - F(x)|$$

la distanza  $d_n \equiv \sqrt{n}D_n$  definisce il valore di riferimento per il test di KS che consiste nel confrontare tale numero con una grandezza di riferimento  $d_0$ , che costituisce la quantità di soglia rispetto alla quale rigettare la null hypothesis  $H_0$ . Se  $d > d_0$  l'ipotesi di compatibilità viene rigettata. Il valore di  $\delta_0$  è scelto in base alla probabilità che la variabile casuale  $\delta$  sia maggiore di  $\delta_0$  quando il modello è corretto.

$$P(\delta > \delta_0 | H_0) = \alpha$$

Tale metodo non parametrico è anche utile per confrontare due campioni di dati al fine di determinare se provengono dalla stessa popolazione. Si noti anche che la **EDF(x)** costruita è anch'essa una distribuzione cumulativa di probabilità.

## 6.4 Confronto di una misura con il valore di riferimento

### 6.4.1 Distribuzione di t-student

Si consideri un campione di  $N$  misure di cui si è calcolata la media campionaria  $\bar{x} = \frac{1}{N} \sum x_i$  e si supponga di conoscere  $\sigma_i$  delle singole misure e  $E[x] = \mu$ , allora la media aritmetica  $\bar{x}$  è per il TCL è distribuita come una gaussian  $G(\bar{x}, \mu, \frac{\sigma}{\sqrt{N}})$ . Come facciamo a dire che  $\bar{x}$  e  $\mu$  sono sufficientemente vicine tra loro rispetto alle incertezze ?

Per determinare la distanza tra le grandezze definiamo la distribuzione t-student data dalla variabile aleatoria:

$$t = \frac{|\bar{x} - \mu|}{\frac{\sigma}{\sqrt{N}}}$$

definita rispetto al caso descritto nelle righe precedenti. In generale la t-student per un parametro è data da:

$$t = \frac{|\hat{\theta}^* - \theta_t|}{\sigma_{\theta^*}} \quad (6.5)$$

Notare che nel caso in cui si conoscano a priori l'incertezza della misura che si sta confrontando, dunque non si è ottenuta mediante un processo statistico si ha che la pdf(t) è Gaussiana. Se  $\bar{x}$  segue una pdf Gaussiana e  $Q^2 \sim \chi^2(N-1)$  la distibuzione di t-student ha la seguente forma funzionale.

$$f(t, \nu = N-1) = \frac{1}{\sqrt{n\nu}} \cdot \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \cdot \left(1 + \frac{t^2}{\nu}\right)^{-(\frac{\nu+1}{2})} \quad (6.6)$$

### Proprietà della distribuzione

$$\mu = E[t] = 0 \quad \gamma_1 = 0$$

$$\sigma^2 = V[t] = \frac{\nu}{\nu-2} \quad \nu > 2 \quad \gamma_2 = \frac{6}{\nu-4} \quad \nu > 4$$

Per  $\nu \rightarrow \infty$  la distribuzione diventa Gaussiana. Si osserva che la pdf della t-student è un po' più larga della distribuzione Gaussiana (fig 6.2).

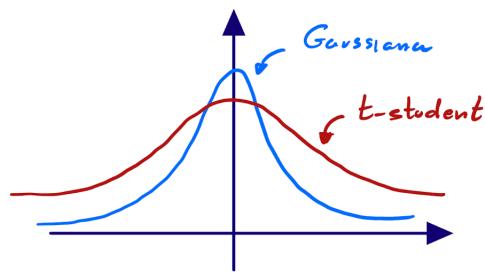


Figure 6.2: Confronto distribuzioe di Gauss e t-student

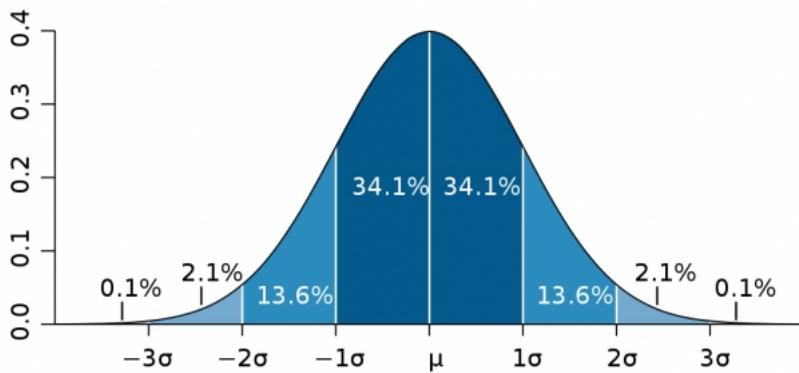
In generale possiamo vedere la distribuzione della t-student come il rapporto tra una Gaussiana normalizzata e la radice di  $\frac{\chi^2}{N-1}$ .

$$t = \frac{|\bar{x} - \mu|}{\left[\frac{\hat{\sigma}^2}{N}\right]^{\frac{1}{2}}} = \frac{|\bar{x} - \mu|}{\left[\frac{\sigma^2}{N}\right]^{\frac{1}{2}}} \cdot \frac{\left[\frac{\sigma^2}{N}\right]^{\frac{1}{2}}}{\left[\frac{\hat{\sigma}^2}{N}\right]^{\frac{1}{2}}} = \frac{|\bar{x} - \mu|}{\left[\frac{\sigma^2}{N}\right]^{\frac{1}{2}}} \cdot \left[\frac{\chi^2}{N-1}\right]^{\frac{1}{2}}$$

la parte in azzurro segue la pdf di una Gaussiana normalizzata  $N(0,1)$ , mentre la parte in rosso segue  $\frac{\chi^2}{N-1}$ . Dove:

$$\frac{1}{\left[\frac{\sigma^2}{\hat{\sigma}^2}\right]^{\frac{1}{2}}} = \frac{1}{\left[\frac{N-1}{\chi^2}\right]^{\frac{1}{2}}} = \left[\frac{\chi^2}{N-1}\right]^{\frac{1}{2}}$$

Per la t-student fissato un valore di soglia  $t_0$ , ci permette di determinare la compatibilità tra il valore stimato e quello atteso, se  $t > t_0$  allora le due misure risultano essere non compatibili tra loro. Dove gli intervalli di compatibilità risultano essere in multipli di deviazioni standard.



### Confronto tra la stima di due parametri

La distribuzione di t-student può essere utilizzata non solo per confrontare una misura con un valore atteso, ma anche due stime del medesimo parametro

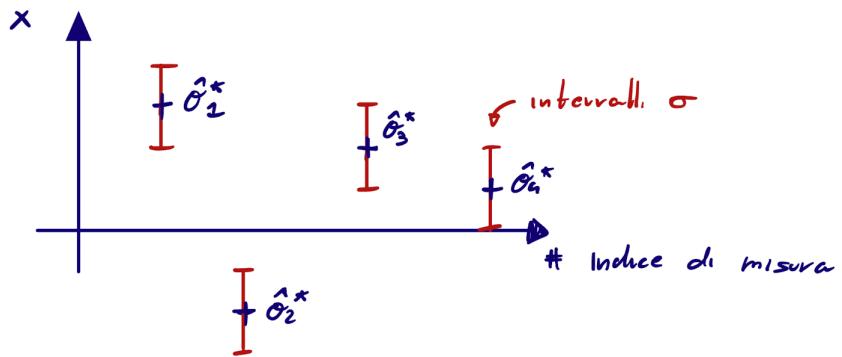
tra loro.

$$z = \frac{|\theta_1^* - \theta_2^*|}{\left[ \frac{\hat{\sigma}_1}{N} + \frac{\hat{\sigma}_2}{N} \right]^{\frac{1}{2}}} \quad (6.7)$$

dove la pdf(z) è Gaussiana se  $\sigma_1$  e  $\sigma_2$  sono senza errore, mentre altrimenti segue una distribuzione di t-student con N-M-2 gradi di libertà.

## 6.5 Intervalli di confidenza

Ipotizziamo che lo stimatore  $\hat{\theta}$  segue una distribuzione gaussiana  $G(\hat{\theta}, \theta_t, \sigma)$  determinato il suo valore  $\theta^*$  ci domandiamo quale sia la probabilità che tale valore disti  $\sigma$  dal valore vero  $\theta_t$ , ovvero  $P[\theta^* \in (\theta_t - \sigma, \theta_t + \sigma)] = 0,68$ . Essendo  $\theta^*$  una variabile aleatoria dipendente dal campione mentre  $\theta_t$  no, l'affermazione precedente non è corretta, in quanto  $\theta_t$  non è una variabile aleatoria, per questo motivo riscriviamo l'intervalllo di confidenza in  $\theta_t \in (\theta^* - \sigma, \theta^* + \sigma)$  e la probabilità come  $P[\theta_t \in (\theta^* - \sigma, \theta^* + \sigma)] = 0,68$ , determinando un 68 % di confidenza nell'intercettare  $\theta_t$  ripetendo l'esperimento.



### 6.5.1 Metodo della cintura di confidenza

Definiamo uno stimatore  $\hat{\theta} \equiv \hat{\theta}(x)$ , di un parametro  $\theta$  rispetto ad un campione di N variabili aleatorie IID di una grandezza x, poichè lo stimatore dipende da variabili aleatorie anch'esso è una r.v. di conseguenza seguirà

una distribuzione di probabilità  $f(\hat{\theta} | \theta)$ . Per ciascun campione raccolto  $\hat{\theta}$  definirà una stima  $\theta^*$  del parametro, costruendo la pdf associata. Ripetendo l'esperimento con diversi campionamenti ciascuna distribuzione relativa avrà un valore atteso  $E[\hat{\theta}] = \theta_t^i$  e una varianza  $V[\hat{\theta}] = \sigma_{\theta_t^i}^2$ .

Per ciascun esperimento non conosciamo la forma analitica della pdf oppure non siamo in grado di definirla di conseguenza per stimare l'intervallo di confidenza di un certo valore di  $\theta^*$  stimato usiamo il seguente metodo:

- Per ogni valore di  $\theta_t^i$  definiamo  $pdf_i(\hat{\theta}|\theta_t)$ ;

- Si determinano i punti della  $pdf(\hat{\theta}|\theta_t)$  che definiscono un intervallo di confidenza per il  $\theta_t$  corrispettivo;

- Si tracciano due rette parallele che attraversano ciascuna i punti estremanti di ciascun intervallo di confidenza, definendo una banda nel piano  $(\theta_t, \theta^*)$  che prende il nome di **Confidence Band**.

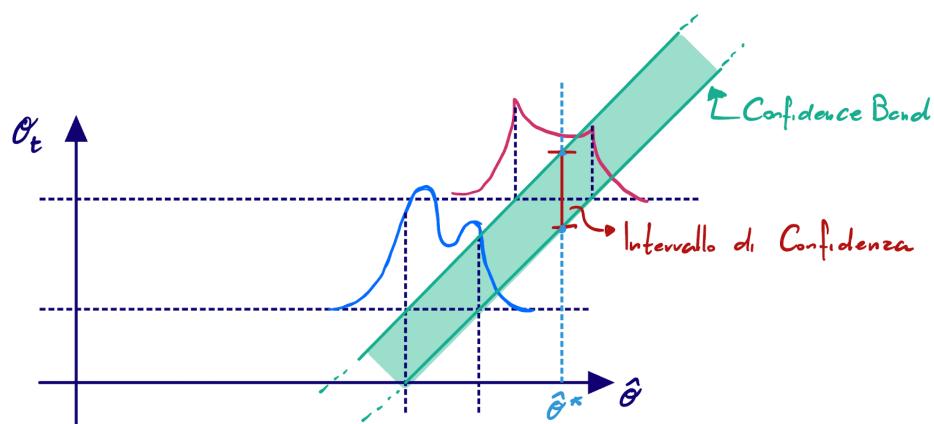
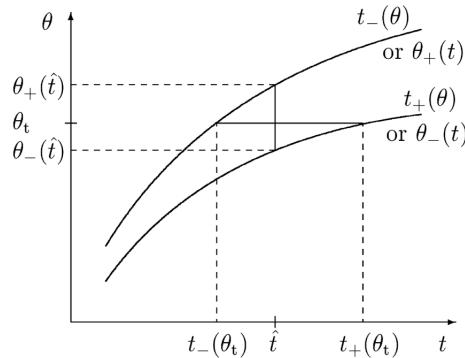


Figure 6.3: Banda di confidenza ed intervallo di confidenza

Quando si effettua una misura si trova  $\theta^*$  e dove tale valore interseca la confidence band, punti d'intersezione determinano l'intervallo di confidenza di  $\theta_t$ . Si dimostra che l'intervallo trovato ha copertura uguale a quello scelto per le singole p.d.f.

## 6.6 Discovery Significance

Consideriamo di rilevare un numero di eventi  $n_0$  in un intervallo di tempo, e che tale numero di eventi può essere distinto in due sotto categorie date da  $n_s$  che è il numero di eventi dovuto a un processo fisico e  $n_b$  numero di eventi che definiscono il "rumore di fondo" (e rappresenta fenomeni non legati al processo fisico osservato) dove  $n_0 = n_s + n_b$ . A priori non abbiamo modo di sapere nel conteggio quanti fenomeni fisici e non compongano  $n_0$ . In compenso conosciamo il numero medio di eventi di entrambi i conteggi  $E[n_s] = \nu_s$  e  $E[n_b] = \nu_b$ .

Costruiamo la nostra null hypothesis  $H_0$  assumendo che del numero complessivo di fenomeni buona parte siano dati dal rumore di fondo; per confutare tale ipotesi utilizzando il p-value è necessario che questo sia più piccolo del valore di soglia  $p_0$ , in fisica delle particelle si sceglie  $p_0 = 3 \times 10^{-7}$  per il segnale che una particella sia stata rilevata. La Poissoniana che descrive la probabilità di  $n_b$  è data da:

$$Poiss(n_b, \nu_b) = \frac{\nu^n}{n!} e^{-\nu}$$

la probabilità di misurare un valore  $n_b$  di quello misurato è data da:

$$\beta = P(n > n_b^0) = \sum_{k=n_0+1}^{\infty} \frac{\nu^k}{k!} e^{-\nu} = 1 - \sum_{k=0}^n \frac{\nu^k}{k!} e^{-\nu} \quad (6.8)$$

Possiamo riscrivere l'equazione (6.7) come:

$$1 - \beta = \sum_{k=0}^n \frac{\nu^k}{k!} e^{-\nu} \approx \int_{2n_b^0}^{\infty} \chi^2(2n+2) d\chi^2 \quad (6.9)$$

approssimabile alla distribuzione del  $\chi^2(2n+2)$  con  $2(n+1)$  g.d.l., ovvero

---

## CHAPTER 6. TEST D'IPOTESI

l'espressione di sinistra della 6.8 coincide con la sua c.d.f. Di conseguenza avremo che il:

$$\text{p-value} = 1 - \beta$$

Per  $\beta \approx 1$  il p-value è molto piccolo e quindi possiamo rigettare la null hypothesis  $H_0$  formulata all'inizio e quindi il segnale osservato è effettivamente un processo fisico. Un valore grande di  $\beta$  ci dice che si ha un alta probabilità che per valori più grandi di  $n_b^0$  i fenomeni osservati siano composti per la maggior parte da rumore di fondo. Mentre per  $1 - \beta$  molto piccolo si ha una bassa probabilità che quanto osservato sia dato da rumore di fondo.