

Appunti di teoria per l'esame di Laboratorio di Meccanica e Termodinamica

Alberto Zaghini

a.a. 2022-2023

Indice

1	Grandezze fisiche, dimensioni, cifre significative	4
1.0.1	Analisi dimensionale	5
1.1	Misura	5
1.1.1	Sistema internazionale	5
1.1.2	Cifre significative	6
2	Incertezze	8
2.1	Misure dirette	8
2.1.1	Incertezza assoluta	9
2.1.2	Errore (di misura)	9
2.2	Discrepanza	9
2.3	Incertezze sistematiche	9
2.4	Incertezze statistiche	10
2.5	Istogrammi	10
2.6	Parametri (caratteristici)	11
2.6.1	Indici di posizione	11
2.6.2	Indici di dispersione	11
2.7	Campioni di dati	12
3	Propagazione delle incertezze	14
3.1	Leggi di Propagazione	14
3.1.1	Note importanti	16
4	Calcolo delle probabilità	17
4.1	Glossario probabilistico	17
4.2	Definizioni di probabilità	18
4.3	Probabilità di eventi complessi	18
4.4	Teorema di Bayes	19
5	PDF	20
5.1	Proprietà della PDF	20
5.2	Valore atteso (o di aspettazione)	21
5.3	Distribuzione normale (Gaussiana)	21
5.3.1	Proprietà	21
5.3.2	Normalizzazione	22
5.3.3	Media	22
5.3.4	Varianza	22
5.4	Gaussiana standard	23
5.5	Random Walk	23
5.6	Stima dei parametri	25
5.7	Cumulative Distribution Function	25
5.8	Distribuzioni di variabili casuali	26
5.9	Distribuzione uniforme	26
5.10	Distribuzione esponenziale	27
5.11	Distribuzione di Cauchy-Lorentz	28

6	Distribuzioni multivariate	30
6.0.1	Scatter plot	30
6.0.2	Lego plot	30
6.0.3	Superficie limite	30
6.0.4	PDF condizionale	31
6.1	Bivariate	31
6.2	Covarianza e correlazione	33
6.3	Somma in quadratura	33
6.4	Incertezze in misure dipendenti	34
6.5	Combinazione lineare	34
6.6	Correlazione lineare	35
6.7	Media pesata di misure indipendenti	35
6.8	Regressione lineare	36
6.8.1	Linearizzazione	40
6.9	Rigetto di dati	40
7	DAQ	
	Sistemi di acquisizione dati	42
7.1	Campionamento	43
8	Campioni ridotti e intervalli di confidenza	44
8.1	Confidence level	46
8.2	t di Student	46
8.2.1	Proprietà della distribuzione di Student	46
8.3	Test di ipotesi	47
8.3.1	Confronto misura con valore accettato	47
8.3.2	Confronto di due misure	48
8.4	Rigetto per campioni ridotti	48
8.5	Test di ipotesi: nulla vs alternativa	48
8.5.1	Test parametrici	49
9	Funzioni di probabilità di variabili discrete	51
9.1	Distribuzione uniforme	51
9.2	Calcolo combinatorio	52
9.3	Distribuzione binomiale	52
9.3.1	Proprietà	53
9.4	Efficienza di conteggio	53
9.5	Distribuzione geometrica	54
9.5.1	Proprietà	54
9.5.2	Assenza di memoria	55
9.6	Distribuzione di Poisson	56
9.6.1	Proprietà della poissoniana	57
9.6.2	Singola misura di conteggio	58
9.7	Eliminazione del rumore di fondo	58
10	Chi quadro	59
10.1	Confronto dati - distribuzione	59
10.1.1	Stabilire i bin	60
10.1.2	DoF	60
10.1.3	Chi ridotto	60
10.2	Distribuzione Chi quadro	60
10.2.1	Proprietà di $\Phi(\chi^2)$	61
10.3	Coppie vs relazione funzionale	61
10.4	Soglie di probabilità	61
10.5	Minimo Chi quadro	62
11	Teorema del limite centrale	63

12 PRNG e Monte Carlo	64
12.1 Altre distribuzioni	65
12.2 Metodo del rigetto	65
12.2.1 Aumentare l'efficienza	65
12.3 Metodo della Trasformazione inversa	66
12.4 Inversa per funzioni notevoli	66
12.4.1 Esponenziale	66
12.4.2 Gaussiana	67

Capitolo 1

Grandezze fisiche, dimensioni, cifre significative

Scopo dell'attività scientifica = **comprensione dei fenomeni naturali**. Alla base di ciò si hanno **osservazioni** (serie di) → riconoscimento aspetti ricorrenti (caratteristici).

Caratteristiche Osservabili (analisi preliminare *qualitativa*) → **misurabili** (analisi *quantitativa*) = **ESPERIMENTI**.

Metodo scientifico galileiano

1. Osservazione acritica
2. Riduzione delle osservazioni → selezione delle informazioni (*)
3. Formulazione delle leggi
4. verifica sperimentale :
$$\begin{cases} NO & \rightarrow 1. \\ SI' & \rightarrow \text{Legge convalidata} \end{cases}$$

La verifica sperimentale avviene in condizioni *privilegiate* (*).

1,2 introduzione di nuove quantità caratterizzanti → migliore descrizione dei fenomeni

3 modello interpretativo

4 rafforzamento / falsificazione

Leggi = relazioni funzionali tra quantità **misurabili** → **Grandezze fisiche** = proprietà quantificabili
Ogni grandezza ∈ **classe** = **dimensione** ([1] = adimensionali, **come gli argomenti di tutte le funzioni trascendenti**) -> ogni classe contiene tutte le grandezze **omogenee** che sole sono confrontabili e sommabili/sottraibili.

Per tutte le grandezze fisiche possibile stabilire tramite un parametro d'ordine ($> = <$) relazione d'ordine transitiva. Grandezze per cui è definita anche regola di composizione (somma, differenza) sono **additive**, non additive in caso contrario.

Stessa classe = stesso significato ma differente procedura:

- Misura diretta
- Misura **indiretta** (derivazione attraverso procedura matematica)

Angolo piano

$$\alpha = \frac{\widehat{AB}}{r} \quad 0 \leq \alpha \leq 2\pi \text{ in radianti (rad)}$$

Angolo solido

$$\Omega = \frac{A}{r^2} \quad 0 \leq \Omega \leq 4\pi \text{ in steradiani (sr)}$$

1.0.1 Analisi dimensionale

$$G = cA^\alpha B^\beta C^\gamma \implies [G] = [A]^\alpha [B]^\beta [C]^\gamma$$

Dimensione fisica di una grandezza = funzione che determina fattore di conversione della stessa nel passaggio ad un differente sistema di u.m.

1.1 Misura

Grandezza fisica \rightarrow processo di **misurazione (o misura)** \rightarrow assegnazione di un **valore numerico** (misura).
3 elementi:

1. Materiale / oggetto (o sistema di oggetti)
2. Grandezza caratterizzante materiale/sistema
3. Procedura utilizzata per misura

1,2,3 \rightarrow **precisa descrizione**.

Valore numerico permette confronto con **campione di riferimento (= unità di misura)** = grandezza omogenea \rightarrow rapporto numerico quantitativo.

1.1.1 Sistema internazionale

Sistema di unità di misura = insieme di

- **Definizioni** delle unità fondamentali / di base misurate (solitamente) in modo **diretto**
- **Regole** per la definizione di grandezze **derivate**: il coefficiente moltiplicativo nell'espressione come funzione di grandezze fondamentali è **sempre 1** (**coerenza** sistema di udm)

Definizioni devono essere quanto più

1. Indipendenti da tempo e luogo (stabili)
2. Precise
3. Riproducibili (legato a 1.)

\Rightarrow **Aggiornamento**.

7 grandezze fondamentali definite sulla base di **7 costanti fisiche** fondamentali. La loro scelta garantisce

- Indipendenza reciproca
- Completezza (sufficienti per ottenere tutte le derivate)

I multipli ed i sottomultipli devono essere **decimali**.

Bureau International des Poids et Mesures stabilisce unità e costanti e presiede all'aggiornamento delle definizioni e dei campioni di riferimento

Unità e costanti

Grandezza	udm	Costante
tempo	secondo s	frequenza di transizione iperfine Cesio
lunghezza	metro m	velocità della luce nel vuoto
massa	chilogrammo kg	costante di Planck
temperatura termodinamica	kelvin K	carica elementare
q.tà di sostanza	mole mol	numero di Avogadro
intensità luminosa	candela cd	efficacia luminosa
intensità di corrente	ampère A	carica elementare

Convenzioni d'uso

- nomi comuni: **iniziale minuscola**
- invarianti al plurale
- se
 - **non** accompagnate da numero: scritte per esteso (*Il metro*)
 - accompagnate: simbolo (maiuscola) **non in corsivo, no seguito da punto, dopo il valore**

Esponenti delle gr derivate $\in \mathbb{Z}$

Multipli e sottomultipli solo un prefisso! (per kg si fa riferimento a g)

10^{18}	exa (E)	10^{-1}	deci (d)
10^{15}	peta (P)	10^{-2}	centi (c)
10^{12}	tera (T)	10^{-3}	milli (m)
10^9	giga (G)	10^{-6}	micro μ
10^6	mega (M)	10^{-9}	nano (n)
10^3	kilo (k)	10^{-12}	pico (p)
10^2	etto (h)	10^{-15}	femto (f)
10	deca (da)	10^{-18}	atto (a)

Fattore di conversione C tra due u.m. u_1, u_2

$$1 \cdot u_1 = C \cdot u_2$$

1.1.2 Cifre significative

= tutte le cifre di un numero a partire dalla prima **diversa da 0**, lette da sx a dx.

Per evitare ambiguità con **zeri non significativi**, si utilizza **notazione scientifica** (gli zeri in mezzo od in fondo **sono significativi**, in quanto le c.s. descrivono la precisione di una misura)

$$A \times 10^n \quad 1 \leq A < 10 \quad n \in \mathbb{Z}$$

ove A non termina con 0.

Operazioni e CS

Regola generale:

numero di cifre significative nel risultato di operazioni condotte su due o più misure di grandezze fisiche = numero di cifre significative della misura meno accurata

Arrotondamento

- prima cifra eliminata ≥ 5 : si aumenta ultima c.s. di 1
- se < 5 (**escluso**) : ultima c.s. invariata

Addizione e sottrazione ultima c.s. = ultima ottenuta da somma o differenza c.s. delle due misure iniziali

Prodotto e quoziente numero di c.s. = minimo numero di c.s. tra le misure iniziali

Numeri esatti \rightarrow considerati come n. con *infinite* c.s.

Ordine di grandezza = prima cifra a sx $\neq 0$ (dopo arrotondamento) moltiplicata per opportuna potenza di 10

Utilizzate in confronto grandezze omogenee.

Capitolo 2

Incertezze

ad ogni misura è associata un'incertezza

Non è **mai** eliminabile del tutto (per quanto riducibile) x due ragioni:

1. Sensibilità strumentale limitata (soglia di risoluzione, sotto cui impossibile distinguere grandezze)
2. Inevitabilità errori nell'effettuazione dell'operazione di misura

Processo di misura = confronto grandezze con udm \rightarrow determina **Intervallo di valori**:

$$n_0 u + \frac{n_1}{10} u + \dots + \frac{n_k}{10^k} u < G < n_0 u + \frac{n_1}{10} u + \dots + \frac{(n_k + 1)}{10^k} u$$

- **Limite di riproducibilità** di scala
- **Soglia di riproducibilità** : condizioni del sistema o dell'ambiente che pregiudicano riproducibilità operazione di misura

Perché ridurre l'incertezza?

- Evidenziare fenomeni precedentemente ignorati / nascosti
- Permettere confronto tra misure omogenee: valutazione compatibilità reciproca e con grandezze di riferimento

2.1 Misure dirette

incertezza = risoluzione strumento = più piccola variazione di grandezza che str. riesce ad apprezzare

misura = localizzazione punto su scala graduata / display digitale

- Strumenti analogici: risoluzione = metà minima distanza tra tacche
- Str. digitali: ris = mezza unità digit meno significativo

Ogni strumento è dotato di **data sheet** che specifica

1. Range (portata)
2. Risoluzione
3. Condizioni ambientali adatte all'utilizzo

Per digitali: risoluzione diminuisce aumentando numero di bit in uscita.

2.1.1 Incertezza assoluta

Per una grandezza x è Δx (chiaramente le due sono grandezze omogenee)
L'esito di una misura è espresso secondo

$$x = (x_{best} \pm \Delta x) \text{ u.m.}$$

inc si rappresenta graficamente tramite **barra di errore**; si suppone il *valor vero* della grandezza da misurare ricada nell'intervallo di misura così definito.

Regole

1. L'incertezza va arrotondata a **una sola** c.s.
2. L'ultima c.s. della x_{best} deve essere **dello stesso ordine di grandezza** dell'ultima c.s. di Δx (ovvero stessa posizione decimale se espresse in notazione scientifica con stessa potenza di 10)

Incerezze **strumentali** o **di lettura** = sinonimi; entrambe si riferiscono a incerezze **massime**.

Precisione = **incertezza relativa** = rapporto inc assoluta / valore ottimale

$$\frac{\Delta x}{x_b}$$

1. Può essere espressa in percentuale
2. É adimensionale (rapporto omogenee) → permette confronto tra misure di grandezze non omogenee

2.1.2 Errore (di misura)

Non è sinonimo di incertezza!

Corrisponde a **differenza tra valore reale grandezza (che si ipotizza esista) e valore best trovato**

$$E = |X - x_b|$$

Non si misura (inconoscibile)! L'incertezza ne è la **miglior stima**

2.2 Discrepanza

= differenza tra due valori misurati della stessa grandezza (omogenei → confrontabili). Se entrambi espressi come

$$x_i = (x_b^i \pm \Delta x_i)$$

allora

$$discr = |x_b^2 - x_b^1|$$

Può essere

- **Significativa** se **non** \exists valori compatibili con entrambi gli intervalli di misura (non si sovrappongono)

$$[x_1 \pm \Delta x_1] \cap [x_2 \pm \Delta x_2] = \emptyset$$

- **Non s.** se si ha almeno un punto di sovrapposizione

Data una grandezza con valore accettato X noto e misure x_i , se la discrepanza $|X - x_i|$ è

- Significativa: la misura x_i è **incompatibile** con X
- Non significativa: la misura x_i è **compatibile** con X

2.3 Incerezze sistematiche

= dovute a fattori non controllati (ma controllabili) **insiti nell'apparato di misura o dovuti ad operazioni errate di misura**. **Non** includono sbagli occasionali (e.g. letture errate)

Portano a **sottostima** / **sovrastima sistematica** (ovvero sempre nel medesimo verso e - circa - della stessa entità).

		Accuratezza	
		Alta	Bassa
Precisione	Alta		
	Bassa		

Possibili sorgenti

- Calibrazione errata (*offset*) di uno strumento o riproduzione errata di u.m. nella scala di uno strumento
- Condizioni ambientali differenti da quelle prescritte per la procedura di misura (che possono dare e.g. effetti termici non considerati)
- Presenza di fattori che influiscono sulla grandezza stessa che si va a misurare (e.g. mancato isolamento termico o presenza di fondo radioattivo)

Come rilevarli? si ripete misura in condizioni sperimentali differenti, cercando di ottenere sistematico trascurabile rispetto a risoluzione e quindi incertezza. Generalmente **non** si hanno **regole fisse** (non sono computabili e riducibili come incertezze massime).

"La stima della incertezza statistica è un calcolo, la stima della incertezza sistematica è un'arte"

Se si riescono ad ottenere sistematici trascurabili rispetto a incertezza (strumentale o statistica) della misura \Rightarrow misura è **accurata**

2.4 Incertezze statistiche

Misure ripetute \rightarrow valori/risultati differenti \rightarrow vari casi:

- Grandezza riferita a individui in stessa **popolazione** (classe) \rightarrow varia tra ind
- Grandezza **intrinsecamente stocastica**
- Lettura della scala dello strumento effettuata **oltre limite riproducibilità**

si ha

Errore stocastico (o casuale)

= deviazione della misura dal valore vero dovuta a molte piccole perturbazioni incontrollabili nel processo di misura

A differenza del sistematico, ha *stessa probabilità di aumentare o diminuire il valore di ciascuna misura* \longleftrightarrow la media degli errori casuali tende statisticamente a 0

Dunque **aumentando il numero di misure** se ne può ridurre l'entità

Serie di misure \Rightarrow **Campione di dati** estratto da **popolazione** (individui). Importante sia **rappresentativo** = no *bias*

Studio statistico insieme di dati \rightarrow miglior stima valor vero (x_b) e errore (Δx)

Errori casuali = ee. **statistici** = oggetto di studio **Teoria degli errori**

Errore statistico ridotto = misura **precisa** (non necessariamente accurata!) **Lo studio della Teoria degli Errori è possibile se si presuppone che i sistematici siano trascurabili rispetto a errori statistici o strumentali**

2.5 Istogrammi

Campione di dati \rightarrow tabella **dato / occorrenze** o **dato / frequenze**

Si definiscono **Bin** = intervalli di suddivisione del dominio sulle ascisse della grandezza misurata (variabile)

Indicato con k indice del bin

$$\sum n_k = N \quad \sum f_k = \sum \frac{n_k}{N} = 1$$

Istogramma in occorrenze → istogramma in frequenze (permette di confrontare campioni con N diversi!)
 → Istogramma in **frequenza cumulativa**

$$F_k = \sum_{i=1}^k f_i$$

che va a saturare per $k = k_{max}$

$$F_{k_{max}} = 1$$

Si definisce **densità di frequenza**

$$\Phi_k = \frac{f_k}{\Delta x}$$

ove Δx indica l'ampiezza del bin sulle ascisse. Chiaramente

$$\sum \Phi_k \Delta x = \sum f_k = 1$$

corrisponde alla somma delle aree nel grafico in densità, che è quindi **normalizzato** \implies indipendente dall'ampiezza degli intervalli e dal numero totale di dati del campione!

É possibile confrontare il grafico in Φ (e quindi effettuare fit) con **Probability Density Function (PDF)**

2.6 Parametri (caratteristici)

Definiscono (sintetizzano) proprietà generali di un campione. Due tipologie principali

2.6.1 Indici di posizione

1. **Centro dell'intervallo** $= \frac{x_{min} + x_{max}}{2}$
2. **Moda** x_k : n_k massimo
3. **Mediana** x_k : $N/2$ elementi del campione hanno valori $< x_k$
4. **Media aritmetica** $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$

2.6.2 Indici di dispersione

1. **Semi-dispersione massima** $= \frac{x_{max} - x_{min}}{2}$
 usata per stimare incertezza sulla media di **poche** misure (solo se maggiore della risoluzione dello strumento)

2. **Scarto medio** definito lo scarto i -esimo come $d_i = x_i - \bar{x}$ si ha

$$\sum d_i = 0 \quad \bar{d} = 0$$

per definizione della media. Si ha allora

3. **Scarto quadratico medio** = deviazione standard

$$\sigma_x = S_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N d_i^2} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

4. **Varianza** $\sigma_x^2 = S_x^2$

Gradi di libertà

Il termine $N - 1$ al denominatore è giustificabile:

1. A livello pratico col fatto che inserendo N con una sola misura si avrebbe $S_x = 0$, che non è corretto.
 Invece si ha $S_x = \frac{0}{0}$ che non ha senso (come non ha senso calcolare la dispersione con una sola misura)

2. A livello statistico considerando il numero di **Gradi di libertà**

Gradi di libertà (di una variabile aleatoria o di una statistica)

Numero minimo di dati sufficienti a valutare la quantità di informazione contenuta nella statistica.

Per ottenere gli scarti quadratici è prima necessario calcolare la media: si è utilizzata un'informazione contenuta nei dati e si perde dunque un dof (un dato è vincolato).

2.7 Campioni di dati

Dato un campione statistico di N misure ripetute x_i si assume che

1. La miglior stima del valor vero X sia la media aritmetica \bar{x}
2. La miglior stima dell'incertezza **SULLA SINGOLA MISURA** del campione sia la deviazione standard S_x
3. La miglior stima dell'incertezza da associare **ALLA MEDIA** sia la **deviazione standard della media**

$$S_{\bar{x}} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^N (x_i - \bar{x})^2} = \frac{S_x}{\sqrt{N}}$$

La dimostrazione è in fondo al capitolo.

Il risultato della misura della grandezza è quindi da esprimersi come:

$$x = (\bar{x} \pm S_{\bar{x}}) \text{ u.m.}$$

La varianza campionaria si può alternativamente esprimere come **la differenza tra la media dei quadrati ed il quadrato della media**

$$S_x^2 = \frac{N}{N-1} (\overline{x^2} - \bar{x}^2)$$

Dimostrazione.

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{N}{N-1} \cdot \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{N}{N-1} \left(\frac{\sum x_i^2}{N} + \frac{\sum \bar{x}^2}{N} - 2 \frac{\sum x_i \bar{x}}{N} \right)$$

Portando fuori la media (termine non indicizzato) nell'ultima sommatoria

$$S_x^2 = \frac{N}{N-1} \left(\frac{\sum x_i^2}{N} + \frac{N}{N} \bar{x}^2 - 2 \bar{x} \frac{\sum x_i}{N} \right) = \frac{N}{N-1} \left(\overline{x^2} + \frac{\sum \bar{x}^2}{N} - 2 \bar{x}^2 \right) = \frac{N}{N-1} (\overline{x^2} - \bar{x}^2)$$

■

Dimostrazione $S_{\bar{x}}$

Si fa ricorso a regole espone nel capitolo seguente.

La media stessa di un campione stocastico è una variabile casuale, in quanto il suo valore dipende dal particolare campione considerato.

Se si disponesse di differenti insiemi dello stesso numero n di misure e se ne calcolasse la rispettiva media, si otterrebbe un **campione di medie**. Poiché ogni media è ottenuta secondo la formula vista da n misure che

si assumono indipendenti, è possibile calcolare la varianza del campione di appartenenza tramite la somma in quadratura

$$\sigma^2(\bar{x}) = \sum_i \left| \frac{\partial \bar{x}}{\partial x_i} \right|^2 \sigma^2(x_i) = n \left(\frac{1}{n^2} \right) \sigma^2(x) = \frac{\sigma^2(x)}{n}$$

da cui, prendendo la radice per ottenere la SD

$$\sigma(\bar{x}) = \frac{\sigma(x)}{\sqrt{n}}$$

Capitolo 3

Propagazione delle incertezze

Misure indirette → procedura comporta due passi (con due corrispondenti nella stima dell'incertezza x misura finale):

1. Misure dirette \Rightarrow Incertezze massime
2. Calcolo \Rightarrow propagazione in stima incertezza finale da associare

3.1 Leggi di Propagazione

Somma

$$a = x + y \quad x = x_b \pm \Delta x \quad y = y_b \pm \Delta y$$

allora

$$\begin{cases} a_{max} = (x_b + y_b) + (\Delta x + \Delta y) \\ a_{min} = x_b + y_b - \Delta x - \Delta y = (x_b + y_b) - (\Delta x + \Delta y) \end{cases} \Rightarrow a_b = x_b + y_b \quad \Delta a = \Delta x + \Delta y$$

Per più misure generalmente

$$a = \sum x_i \Rightarrow a = \left(\sum x_b^i \pm \sum \Delta x_i \right)$$

Differenza

$$a = x - y \quad x = x_b \pm \Delta x \quad y = y_b \pm \Delta y$$

allora

$$\begin{cases} a_{max} = (x_b - y_b) + (\Delta x + \Delta y) \\ a_{min} = x_b - y_b - \Delta x - \Delta y = (x_b - y_b) - (\Delta x + \Delta y) \end{cases} \Rightarrow a_b = x_b - y_b \quad \Delta a = \Delta x + \Delta y$$

Per più misure generalmente

$$a = \sum \pm x_i \Rightarrow a = \left(\sum \pm x_b^i \pm \sum \Delta x_i \right)$$

Generale per somme e diff (provvisoria)

Incetzza sul valore di una grandezza misurata indirettamente tramite somma o differenza di grandezze misurate direttamente è **la somma delle incertezze**

Prodotto

$$a = xy \quad x = x_b \pm \Delta x \quad y = y_b \pm \Delta y$$

in termini di incertezza relativa

$$x = x_b \left(1 \pm \frac{\Delta x}{|x_b|} \right) \quad y = y_b \left(1 \pm \frac{\Delta y}{|y_b|} \right)$$

La miglior stima per a è

$$a = x_b \cdot y_b$$

Il massimo valore si ottiene per

$$a_{max} = x_b \left(1 + \frac{\Delta x}{|x_b|} \right) \cdot y_b \left(1 + \frac{\Delta y}{|y_b|} \right) = x_b y_b \left(1 + \frac{\Delta x}{|x_b|} + \frac{\Delta y}{|y_b|} + \frac{\Delta x}{|x_b|} \frac{\Delta y}{|y_b|} \right)$$

supponendo che le incertezze siano contenute, l'ultimo termine tra parentesi è trascurabile (II ordine):

$$a_{max} = x_b y_b \left[1 + \left(\frac{\Delta x}{|x_b|} + \frac{\Delta y}{|y_b|} \right) \right] = a_b + a_b \left(\frac{\Delta x}{|x_b|} + \frac{\Delta y}{|y_b|} \right)$$

Procedendo analogamente per il valore minimo si ottiene:

$$a_{min} = x_b y_b \left(1 - \frac{\Delta x}{|x_b|} - \frac{\Delta y}{|y_b|} \right) = a_b - a_b \left(\frac{\Delta x}{|x_b|} + \frac{\Delta y}{|y_b|} \right)$$

Segue quindi che

$$\Delta a = \pm a_b \left(\frac{\Delta x}{|x_b|} + \frac{\Delta y}{|y_b|} \right) \Rightarrow \frac{\Delta a}{|a|} = \frac{\Delta x}{|x|} + \frac{\Delta y}{|y|}$$

Quoziente

$$a = \frac{x}{y} \quad x = x_b \pm \Delta x \quad y = y_b \pm \Delta y$$

Si riconduce ad un procedimento analogo operando la linearizzazione per incertezze ridotte:

$$\begin{aligned} a_{max} &= \frac{x_b \left(1 + \frac{\Delta x}{|x_b|} \right)}{y_b \left(1 - \frac{\Delta y}{|y_b|} \right)} \approx \frac{x_b}{y_b} \left(1 + \frac{\Delta x}{|x_b|} \right) \left(1 + \frac{\Delta y}{|y_b|} \right) \approx \\ &\approx a_b \left[1 + \left(\frac{\Delta x}{|x_b|} + \frac{\Delta y}{|y_b|} \right) \right] \end{aligned}$$

per le consuete considerazioni sui termini trascurabili.

Per il minimo valore si ha invece

$$\begin{aligned} a_{min} &= \frac{x_b \left(1 - \frac{\Delta x}{|x_b|} \right)}{y_b \left(1 + \frac{\Delta y}{|y_b|} \right)} \approx \frac{x_b}{y_b} \left(1 - \frac{\Delta x}{|x_b|} \right) \left(1 - \frac{\Delta y}{|y_b|} \right) \approx \\ &\approx a_b \left[1 - \left(\frac{\Delta x}{|x_b|} + \frac{\Delta y}{|y_b|} \right) \right] \end{aligned}$$

Segue quindi, come per il prodotto:

$$\frac{\Delta a}{|a|} = \frac{\Delta x}{|x|} + \frac{\Delta y}{|y|}$$

Generale per prodotti e quozienti (provvisoria)

L'incertezza relativa su una grandezza ottenuta tramite prodotti e/o quozienti di misure dirette è pari alla **somma delle incertezze relative sulle stesse**

Prodotto per numero intero data

$$q = Bx$$

se B non ha incertezza allora

$$\Delta q = B\Delta x$$

Potenze data

$$q = x^n$$

dall'incertezza sul prodotto si ottiene

$$\frac{\Delta q}{|q_b|} = n \frac{\Delta x}{|x_b|}$$

Generale per prodotti e quozienti con vari esponenti (provv)

Se

$$q = c \cdot x^\alpha \cdot y^\beta \cdot z^\gamma \dots$$

ove c costante allora

$$\frac{\Delta q}{|q|} = |\alpha| \frac{\Delta x}{|x|} + |\beta| \frac{\Delta y}{|y|} + |\gamma| \frac{\Delta z}{|z|}$$

É definito anche metodo **passo-passo**, anche se con ciò si indica più in generale l'applicazione della presente regola assieme a quella vista in precedenza per somme e differenze, di modo da poter trattare anche polinomi.

Funzioni di una variabile se q dipende da **una sola** variabile misurata direttamente secondo

$$q(x) = f(x)$$

allora assumendo l'incertezza sufficientemente ridotta è possibile rettificare in un intorno e ottenere

$$\Delta q = q_{max} - q_b = q(x_b + \Delta x) - q(x_b) = \frac{(x_b + \Delta x) - q(x_b)}{\Delta x} \cdot \Delta x = \left| \frac{dq}{dx} \right|_{x_b} \Delta x$$

Funzioni di più variabili si generalizza tramite derivate parziali

$$q = q(x, y, \dots) \implies \Delta q = \left| \frac{\partial q}{\partial x} \right|_{x_b, y_b, \dots} \Delta x + \left| \frac{\partial q}{\partial y} \right|_{x_b, y_b, \dots} \Delta y + \dots$$

3.1.1 Note importanti

Quando q è espressa da un monomio o da un polinomio ma **una grandezza si presenta più volte**, ad esempio sia al denominatore che al numeratore, è corretto utilizzare **solo la regola delle derivate parziali** e **non** il passo-passo, al fine di evitare una sovrastima dell'errore finale.

Capitolo 4

Calcolo delle probabilità

La **teoria delle probabilità** studia le proprietà dei **fenomeni aleatori**, ovvero di fenomeni

- ripetibili
- che però si manifestano in differenti modalità

tali ff non sono dunque prevedibili **singularmente**: è necessario studiare **serie/campioni** per determinare le **frequenze** dei vari *outcomes* e dunque le rispettive probabilità.

4.1 Glossario probabilistico

L'applicazione della teoria delle probabilità richiede la definizione di contenitori e caratteristiche comuni, ovvero degli oggetti fondamentali della teoria.

- **Insieme (o spazio) dei risultati, o spazio campionario** S = insieme possibili modalità in cui si manifesta fenomeno. Può avere # finita o infinita.
- **Evento semplice o elementare** = singolo elemento di S
- **Evento casuale** $A \subseteq S$ = unione di nessuno, uno o più eventi elementari. In particolare
 - $\emptyset \subseteq S$ è l'evento **impossibile**
 - $S \subseteq S$ è l'evento **certo**
- **Spazio degli eventi** $E = \mathcal{P}(S)$ = insieme degli eventi
- **Evento complementare** di un evento casuale A è \bar{A} , ovvero l'evento corrispondente alla sua mancata realizzazione. Un evento ed il suo complementare
 1. Sono disgiunti, ovvero si escludono mutualmente
 2. Saturano lo spazio campionario con la propria unione (segue dalla definizione)

$$A \cup \bar{A} = S$$

- **Evento complesso** = un evento casuale formato dalla combinazione di **più** eventi elementari. Si ottiene (e/o definisce) secondo:
 1. **Unione (o somma logica) di due eventi** = l'evento verificato se si verifica *almeno uno* dei due
$$A \cup B$$
 2. **Intersezione (o prodotto logico) di due eventi** = l'evento verificato se si verificano *entrambi*
$$A \cap B$$
 3. **Differenza di due eventi** = evento verificato se si verifica A ma non B (o viceversa, non vale chiaramente la commutatività!)
$$A \setminus B \quad o \quad A - B$$

É possibile costruire quindi, dati due eventi, la tabella di verità per gli eventi complessi che se ne ottengono.

4.2 Definizioni di probabilità

Definizione classica o combinatoria probabilità di un evento casuale = rapporto tra casi favorevoli e casi possibili **purché tutti i casi possibili siano equiprobabili**

$$P(A) = \frac{n_{fav}}{N_{tot}}$$

Chiaramente la presupposizione della nozione di equiprobabilità rende la definizione ricorsiva è dunque insoddisfacente

Definizione empirica o frequentista si estende il concetto di frequenza relativa per $N_{prove} \rightarrow +\infty$: la probabilità di un evento casuale è infatti

$$P(A) = \lim_{N \rightarrow \infty} f(A) = \lim_{N \rightarrow \infty} \frac{n}{N}$$

Risulta anch'essa insoddisfacente in quanto presuppone la convergenza del limite; tuttavia utile per valutare efficienza contatori.

Definizione assiomatica Si postula l'esistenza dello spazio campionario e degli eventi, definiti secondo quanto esposto prima. Si definisce probabilità di un evento casuale $A \subseteq S$ un numero, indicato con $P(A)$ associato **univocamente** ad A e soddisfacente le seguenti proprietà:

1. $P(A) \geq 0 \forall A$
2. $P(S) = 1$
3. $P(\bigcup_i A_i) = \sum_i P(A_i)$ per qualsiasi scelta di $A_i \in E$ t.c. $A_i \cap A_j = \emptyset \forall i \neq j$

La definizione assiomatica è stata stabilita con la fondazione della Teoria Assiomatica della Probabilità da parte di Kolmogorov.

È logicamente consistente ma nei suoi principi non stabilisce alcun criterio per l'assegnazione dei valori di $P(A)$

4.3 Probabilità di eventi complessi

Se su N prove l'evento A è realizzato n volte, allora per la frequenza relativa del complementare si ha

$$f(\overline{A}) = \frac{1-n}{N} = 1 - f(A)$$

e dunque, come si ottiene anche applicando gli assiomi della teoria di Kolmogorov

$$P(\overline{A}) = 1 - P(A)$$

Dati due eventi A, B allora

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$$

ove l'uguaglianza vale chiaramente nel caso di eventi disgiunti.

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

ove $P(A|B)$ indica la **probabilità condizionale**, ovvero che si verifichi A dato il verificarsi di B . Nel caso di eventi **indipendenti**, ovvero che non influiscono sulla probabilità reciproca si ha

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

nel caso di N eventi indipendenti, ovvero tali per cui

$$P(A_i|A_j) = P(A_i) \quad \forall i \neq j$$

segue

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j) \quad \forall i \neq j$$

da cui

$$P\left(\bigcap_i A_i\right) = \prod_i P(A_i)$$

come vale, ad esempio, per serie ripetute di misure (campione sperimentale).

4.4 Teorema di Bayes

Teorema fondamentale della probabilità (ed in particolare dell'interpretazione soggettivista, che non è qui trattata).

Teorema di Bayes

Siano dati N eventi H_i t.c.

1. siano mutualmente escludenti, ovvero

$$P(H_i \cap H_j) = P(\emptyset) = 0 \quad \forall i \neq j$$

2. saturino lo spazio campionario, ovvero

$$P\left(\bigcup_i H_i\right) = P(S) = 1$$

dato allora il verificarsi di un evento E , indicata con $P(H_i)$ la probabilità **a priori** del verificarsi dell'evento H_i (ovvero prima del verificarsi di E), la probabilità **a posteriori** del medesimo evento H_i è data secondo

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{P(E)}$$

ove la probabilità totale a priori di E è data secondo

$$P(E) = \sum_i P(E|H_i)P(H_i)$$

Esso può essere applicato in modo *ricorsivo*, ovvero per determinare la probabilità H_i di una specifica causa per un effetto osservato E . Se E indica l'esito di un test, si può valutare quindi l'efficacia dello stesso, indicate con H_i le probabilità, ad esempio, della presenza di specifici patogeni etc.

Capitolo 5

PDF

Si consideri nuovamente un istogramma in densità di frequenza, ottenuto da una serie di misure dirette. Si supponga le fluttuazioni casuali siano **maggiori della risoluzione dello strumento**. Aumentando idealmente in modo arbitrario il numero di misure e la precisione, ovvero passando al limite

$$N \rightarrow \infty \quad \Delta x \rightarrow dx$$

l'istogramma in Φ tende ad una **curva continua**. Essa è definita **distribuzione limite**. La funzione che descrive tale curva è una **funzione densità di probabilità, o *Probability Density Function* (PDF)** (il nome segue considerando la formulazione frequentista vista in precedenza).

Si tratta di una costruzione **puramente teorica**, che tuttavia in casi in cui è possibile assumere valido con buona approssimazione il limite permette una efficace trattazione matematica del campione, di cui si possono ricavare i parametri per via analitica.

Il suo dominio è di consueto $] -\infty, +\infty[$

Poiché definita come densità di probabilità, dimensionalmente si ha

$$[f] = [x]^{-1}$$

5.1 Proprietà della PDF

Segue dalla definizione che la frazione di misure che ricade fra un valore x ed uno distante dx , corrispondente quindi alla probabilità che una misura casuale estratta dalla medesima popolazione cada in tale intervallo infinitesimo, è data secondo

$$P(x \leq x' \leq x + dx) = f(x) dx$$

E quindi tra due valori a e b , integrando:

$$P(a \leq x \leq b) = \int_a^b f(x) dx$$

si osserva che la probabilità di un *singolo* valore è quindi pari a 0.

Chiaramente la PDF restituisce valori di probabilità meno efficaci nelle code, in quanto assume valori non nulli per qualsiasi x reale e soddisfa (come esposto di seguito) il vincolo di normalizzazione.

Si dimostrano le seguenti proprietà:

Normalizzazione

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

Per definizione dell'istogramma in densità

$$\sum_{i=1}^N \Phi_k \Delta x = \sum_{i=1}^N 1$$

dunque passando al limite la somma diviene continua e si ha

$$\lim_{\substack{N \rightarrow +\infty \\ \Delta x \rightarrow dx}} \sum_{i=1}^N \Phi_k \Delta x = \int_{-\infty}^{+\infty} \Phi(x) dx = 1$$

Media

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

Varianza

$$\sigma_x^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 dx$$

ove μ indica il **valore centrale**.

5.2 Valore atteso (o di aspettazione)

Si descrive lo strumento matematico tramite cui si ricava, ad esempio, l'espressione generale per i parametri di una PDF.

Sia X variabile aleatoria continua, i cui valori sono distribuiti secondo $f(x)$. Sia quindi $g(X)$ una generica funzione. Si definisce **valore atteso di g** l'espressione

$$\mathbb{E}[g(X)] = \int g(x) f(x) dx$$

Nel caso di X discreta, detta p_i la probabilità di ogni outcome x_i si ha

$$\mathbb{E}[g(X)] = \sum g(x_i) p_i$$

La media \bar{x} quindi altro non è che il valore di aspettazione di $g(x) = x$, e la varianza di $g(x) = (x - \mu)^2$

Si osserva che $\mathbb{E}[x]$ dipende dalla forma della PDF piuttosto che da x .

Si definisce più generalmente l' **n -esimo momento algebrico** di una PDF come $\mathbb{E}[x^n]$ (la media è il primo) e l' **n -esimo momento centrale** come $\mathbb{E}[(x - \mathbb{E}[x])^n]$. Il secondo momento centrale, ovvero la varianza, si denota anche con $\mathbb{V}[x]$.

L'operatore definito è **lineare**:

$$\mathbb{E}[a \cdot g(X) + b \cdot h(X)] = a \cdot \mathbb{E}[g(X)] + b \cdot \mathbb{E}[h(X)]$$

e chiaramente per costanti vale

$$\mathbb{E}[c] = c$$

per la normalizzazione. Quindi

$$\mathbb{E}[\mathbb{E}[x]] = \mathbb{E}[x]$$

5.3 Distribuzione normale (Gaussiana)

La distribuzione limite per una misura soggetta ad un grande numero di piccole cause non valutabili che possono influenzare il risultato sia verso valori maggiori o minori del valor medio, ovvero **sole** incertezze di tipo stocastico, è la **distribuzione normale o di Gauss**:

Gaussiana

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

5.3.1 Proprietà

- Definita su tutta la retta
- Simmetrica rispetto a $x = \mu$, che indica l'ascissa del massimo assoluto della curva: tale primo parametro determina quindi la posizione del grosso della curva su x
- Ha una forma a campana con punti di flesso in

$$x_f = \mu \pm \sigma$$

σ determina quindi la forma della curva: la sua altezza e la sua larghezza (legate dal vincolo di normalizzazione). Si dimostrerà coincide con la SD campionaria.

- maggiore $\sigma \implies$ picco più alto, curva più stretta
- minore $\sigma \implies$ picco più basso, curva più larga

5.3.2 Normalizzazione

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

Dimostrazione. Si pone

$$G(x) = N \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

e si determina N tale da dare la normalizzazione.

Si esegue il cambio di variabile $y = x - \mu$:

$$\int_{-\infty}^{+\infty} G(x) dx = N \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2\sigma^2}} dy$$

E quindi un secondo $z = \frac{y}{\sigma}$

$$N \int_{-\infty}^{+\infty} e^{-z^2/2} \sigma dz = N \sigma \int_{-\infty}^{+\infty} e^{-z^2/2} dz$$

Ora, quello che si è isolato è un integrale notevole (funzione errore) che dà

$$N \sigma \int_{-\infty}^{+\infty} e^{-z^2/2} dz = N \sigma \sqrt{\frac{\pi}{1/2}} = N \sigma \sqrt{2\pi}$$

Imposta la condizione

$$N \sigma \sqrt{2\pi} = 1 \implies N = \frac{1}{\sigma \sqrt{2\pi}}$$

■

5.3.3 Media

$$\bar{x} = \int_{-\infty}^{+\infty} x G(x) dx = \mu$$

Dimostrazione. Si porta fuori dall'integrale il termine di normalizzazione:

$$\int_{-\infty}^{+\infty} x G(x) dx = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Si effettua il cambio di variabile $y = x - \mu$ e si distribuisce per linearità

$$\frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} (y + \mu) \cdot e^{-\frac{y^2}{2\sigma^2}} dy = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} y \cdot e^{-\frac{y^2}{2\sigma^2}} dy + \frac{\mu}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2\sigma^2}} dy$$

Si osserva ora che la bisettrice del primo terzo quadrante, ovvero $f(y) = y$ è dispari; poiché la gaussiana centrata nell'origine è pari la prima integranda è una funzione dispari: **il suo integrale su tutta la retta è nullo in quanto aree simmetriche si annullano** (ovvero la primitiva di una funzione dispari è pari). Si ha dunque

$$\frac{\mu}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2\sigma^2}} dy = \frac{\sigma \sqrt{2\pi}}{\sigma \sqrt{2\pi}} \mu = \mu$$

per la normalizzazione.

■

5.3.4 Varianza

$$s_x^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 G(x) dx = \sigma^2$$

Dimostrazione. Si porta fuori il termine di normalizzazione e si esegue il consueto primo cambio di variabile $y = x - \mu$

$$s_x^2 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 \cdot e^{-\frac{y^2}{2\sigma^2}} dy$$

e quindi il secondo $z = y/\sigma$

$$s_x^2 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma^2 z^2 e^{-z^2/2} \sigma dz = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-z^2/2} dz$$

Si integra ora per parti:

$$s_x^2 = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z \cdot z e^{-z^2/2} dz = \frac{\sigma^2}{\sqrt{2\pi}} \left[-z e^{-z^2/2} \right]_{-\infty}^{+\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} dz$$

Il primo termine si annulla ad entrambi gli estremi in quanto l'esponente del fattore esponenziale va sempre a $-\infty$ ed il suo reciproco è un infinito di ordine superiore alla funzione lineare.

$$s_x^2 = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} dz = \frac{\sigma^2}{\sqrt{2\pi}} (\sqrt{2\pi}) = \sigma^2$$

■

5.4 Gaussian standard

Data una pdf gaussiana, fissati i valori di μ e σ la probabilità che una misura ricada in un intorno di μ di raggio $k\sigma$ è data da

$$P(x \in I_{k\sigma}(\mu)) = \int_{\mu-k\sigma}^{\mu+k\sigma} G(x) dx$$

Per intervalli peculiari:

- $k = 1 \rightarrow P = 68.3\%$
- $k = 2 \rightarrow P = 95.4\%$
- $k = 3 \rightarrow P = 99.7\%$

Per le probabilità di un **intervallo generico**, è possibile rifarsi alle tabelle degli integrali della **Gaussiana standard**, ovvero la distribuzione normale con

$$\mu = 0 \quad \sigma = 1$$

nella **variabile normale standard** z (adimensionale). Per ricondurre un valore x di una variabile normale distribuita secondo pdf con parametri noti al corrispondente di z :

$$z = \frac{x - \mu}{\sigma}$$

(ovvero z = quante σ dista x da μ). La corrispondenza tra valori di x e z è chiaramente biunivoca, secondo la funzione lineare

$$x = \mu + \sigma z$$

5.5 Random Walk

Si giustifica la forma della distribuzione normale come distribuzione limite di una misura soggetta a sole incertezze casuali.

Dimostrazione. Si abbia una grandezza x la cui misura è soggetta a sole incertezze stocastiche e sia X il suo valor vero. Sia n il numero di effetti casuali (fluttuazioni) che agiscono sulla singola misura e a l'entità dello spostamento provocato da ognuno. Indicati con n_+ e n_- , rispettivamente, il numero di effetti che ne incrementano e diminuiscono il valore, per una generica misura si può scrivere

$$x = X + n_+ a - (n - n_+) a = X + (2n_+ - n) a$$

in quanto chiaramente $n_+ + n_- = n$

Definito ora $s \equiv 2n_+ - n$ si ha

$$x = X + sa$$

e dunque una corrispondenza **biunivoca** tra x , s e n_+ in quanto x è funzione lineare di s lo è di n_+ (n e a sono fissi) e quindi ogni misura determina univocamente il valore di s e quindi di n_+ che la dà, e viceversa. Segue che per un qualsiasi outcome

$$P(x) = P(n_+) = P(s)$$

Ora, la proprietà che n_+ passi degli n siano verso valori maggiori è data dalla **distribuzione binomiale** secondo

$$P(n_+) = \binom{n}{n_+} \left(\frac{1}{2}\right)^n$$

in quanto passi verso valori maggiori o minori sono assunti equiprobabili.

EsPLICITANDO E RIESPRIMENDO

$$P(n_+) = \frac{n!}{n_+!(n-n_+)!} \frac{1}{2^n} = \frac{n!}{\left(\frac{n+s}{2}\right)!\left(\frac{n-s}{2}\right)!} \frac{1}{2^n} \propto \frac{n!}{\left(\frac{n+s}{2}\right)!\left(\frac{n-s}{2}\right)!}$$

Ora, prendendo il logaritmo ed applicando l'**approssimazione di Stirling** per $n \rightarrow \infty$

$$\begin{aligned} \ln P(s) &\propto \ln n! - \ln \left(\left(\frac{n+s}{2} \right)! \left(\frac{n-s}{2} \right)! \right) \approx n \ln n - n - \left[\frac{n+s}{2} \ln \frac{n+s}{2} - \frac{n+s}{2} \right] - \left[\frac{n-s}{2} \ln \frac{n-s}{2} - \frac{n-s}{2} \right] = \\ &= n \ln n - n + \frac{n+s+n-s}{2} - \frac{n+s}{2} \ln \frac{n+s}{2} - \frac{n-s}{2} \ln \frac{n-s}{2} = n \ln n - \frac{n+s}{2} \ln \frac{n+s}{2} - \frac{n-s}{2} \ln \frac{n-s}{2} = \\ &= n \ln n - \frac{n}{2} \ln \frac{n+s}{2} - \frac{s}{2} \ln \frac{n+s}{2} - \frac{n}{2} \ln \frac{n-s}{2} + \frac{s}{2} \ln \frac{n-s}{2} = \\ &= n \ln n - \frac{n}{2} \left[\ln(n+s) + \ln(n-s) - 2 \ln 2 \right] - \frac{s}{2} \left[\ln(n+s) - \ln(n-s) \right] = \\ &= n \ln n - \frac{n}{2} \left[\ln \left(n \left(1 + \frac{s}{n} \right) \right) + \ln \left(n \left(1 - \frac{s}{n} \right) \right) - 2 \ln 2 \right] - \frac{s}{2} \left[\ln \left(n \left(1 + \frac{s}{n} \right) \right) - \ln \left(n \left(1 - \frac{s}{n} \right) \right) \right] = \\ &= n \ln n - \frac{n}{2} \ln n - \frac{n}{2} \ln \left(1 + \frac{s}{n} \right) + n \ln 2 - \frac{n}{2} \ln n - \frac{n}{2} \ln \left(1 - \frac{s}{n} \right) - \frac{s}{2} \ln n - \frac{s}{2} \ln \left(1 + \frac{s}{n} \right) + \frac{s}{2} \ln n + \frac{s}{2} \ln \left(1 - \frac{s}{n} \right) = \\ &= -\frac{n}{2} \left[\ln \left(1 + \frac{s}{n} \right) + \ln \left(1 - \frac{s}{n} \right) \right] + n \ln 2 - \frac{s}{2} \left[\ln \left(1 + \frac{s}{n} \right) - \ln \left(1 - \frac{s}{n} \right) \right] \end{aligned}$$

ora, per n grande risulta $\frac{|s|}{n} \ll 1$ e quindi, sviluppando in serie di Taylor al II ordine rispetto al rapporto:

$$\ln \left(1 + \frac{s}{n} \right) \approx \frac{s}{n} - \frac{1}{2} \left(\frac{s}{n} \right)^2 \quad \ln \left(1 - \frac{s}{n} \right) \approx -\frac{s}{n} - \frac{1}{2} \left(\frac{s}{n} \right)^2$$

sostituendo

$$\begin{aligned} &-\frac{n}{2} \left[\frac{s}{n} - \frac{1}{2} \left(\frac{s}{n} \right)^2 - \frac{s}{n} - \frac{1}{2} \left(\frac{s}{n} \right)^2 \right] + n \ln 2 - \frac{s}{2} \left[\frac{s}{n} - \frac{1}{2} \left(\frac{s}{n} \right)^2 + \frac{s}{n} + \frac{1}{2} \left(\frac{s}{n} \right)^2 \right] = \\ &= \frac{n}{2} \left(\frac{s}{n} \right)^2 + n \ln 2 - \frac{s}{2} \left(2 \frac{s}{n} \right) = \frac{s^2}{2n} + n \ln 2 - \frac{s^2}{n} = \ln(2^n) - \frac{s^2}{2n} \end{aligned}$$

dunque esponenziando

$$e^{\ln P(s)} = P(s) \propto 2^n e^{-s^2/2n}$$

sostituendo ora $s = \frac{x-X}{a}$

$$P(s) \propto e^{-\frac{(x-X)^2}{2na^2}}$$

Interpretando ora X come μ e $a\sqrt{n}$ come σ , si è verificata l'ipotesi, in quanto secondo quanto visto in precedenza

$$P(x) = P(s)$$

■

5.6 Stima dei parametri

Dato un campione di dati (x_i) estratto da una **popolazione gaussiana**, è possibile stimare i parametri della pdf limite tramite il metodo della **Maximum Likelihood**:

Le migliori stime dei parametri della DL sono quelle che rendono **massima la probabilità di uscita del campione effettivamente ottenuto**

ovvero i parametri che definiscono la curva che meglio si adatta all'istogramma in densità del campione.

Si ottiene che

$$\hat{\mu} = \bar{x} \quad \hat{\sigma} = s_x$$

Dimostrazione. Poiché ogni dato del campione è estratto da popolazione gaussiana:

$$P(x_i; \sigma, \mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

considerati i singoli esiti delle misure come indipendenti, la **probabilità congiunta** è quindi data da

$$P(x; \sigma, \mu) = \prod_{i=1}^N P(x_i; \sigma, \mu)$$

ovvero

$$P(x; \sigma, \mu) = \frac{1}{(\sigma\sqrt{2\pi})^N} e^{-\frac{\sum_{i=1}^N (x_i - \mu)^2}{2\sigma^2}} \propto \frac{1}{\sigma^N} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$$

Si definisce ora la **Funzione di verosimiglianza (o di likelihood)**

$$\mathcal{L} = \ln[P(x; \sigma, \mu)]$$

Poiché il logaritmo è monotono crescente, la massimizzazione di P è equivalente a quella di \mathcal{L} . Per proprietà dei logaritmi:

$$\mathcal{L} = -N \ln \sigma - \frac{\sum_i (x_i - \mu)^2}{2\sigma^2}$$

Ora, il valore di μ che massimizza L è ottenuto annullando la derivata rispetto al parametro:

$$\frac{\partial \mathcal{L}}{\partial \mu} = 0 = -\frac{-2 \sum_i (x_i - \mu)}{2\sigma^2} \implies \sum_{i=1}^N (x_i - \mu) = 0 \implies \sum_{i=1}^N -N\mu = 0$$

da cui

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}$$

Analogamente per σ si annulla la derivata parziale:

$$\frac{\partial \mathcal{L}}{\partial \sigma} = 0 = -\frac{N}{\sigma} - \frac{\sum_i (x_i - \mu)^2}{2} \left(-\frac{2}{\sigma^3} \right) \implies -N + \frac{\sum_i (x_i - \mu)^2}{\sigma^2} = 0 \implies \sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

da cui

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

tenendo conto del numero di dof (si è già determinato con i dati l'altro parametro). ■

5.7 Cumulative Distribution Function

Definita come l'integrale della pdf:

$$\int_a^b f(x) dx = F(b) - F(a)$$

Il valore in un punto dà la probabilità che la variabile abbia valori non superiori:

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

- È strettamente crescente (per il Teorema Fondamentale ha derivata sempre positiva)
- Presenta un flesso obliquo in μ , in cui la slope della tangente è determinata dal valore della pdf nel massimo
- Tende a 0 per x tendente a $-\infty$ (o generalmente all'estremo inferiore del dominio)
- Tende a 1 per x tendente a $+\infty$ (o generalmente all'estremo superiore del dominio)

5.8 Distribuzioni di variabili casuali

Concetto di pdf = generalizzabile ed applicabile a tutti i fenomeni **a carattere stocastico**.

Nota etimologica

Dal greco $\sigma\tau\omicron\chi\alpha\sigma\tau\iota\kappa\omega\omega$ = congetturale

Si parla generalmente di **variabile casuale** quando la variabilità non è necessariamente dovuta alla procedura di misura, ma è (anche) **intrinseca, connaturata al fenomeno**.

Per ogni PDF definita per tali fenomeni valgono alcune **proprietà e definizioni comuni**:

- La frazione di misure in un intervallo infinitesimo $[x, x + dx]$ è sempre data da $\Phi(x) dx$
- Per due valori a, b :

$$P(x \in [a, b]) = \int_a^b \Phi(x) dx$$

- Vigge sempre la normalizzazione
- Media e varianza sono sempre ottenute secondo:

$$\bar{x} = \mathbb{E}[x] \quad s_x^2 = \mathbb{E}[(x - \bar{x})^2]$$

Vale inoltre sempre

$$s_x^2 = \overline{x^2} - \bar{x}^2$$

Dimostrazione.

$$\begin{aligned} s_x^2 &= \int_{-\infty}^{+\infty} (x - \bar{x})^2 \Phi(x) dx = \int_{-\infty}^{+\infty} x^2 \Phi(x) dx + \bar{x}^2 \int_{-\infty}^{+\infty} \Phi(x) dx - 2\bar{x} \int_{-\infty}^{+\infty} x \Phi(x) dx = \\ &= \overline{x^2} + \bar{x}^2 - 2\bar{x} \cdot \bar{x} = \overline{x^2} - \bar{x}^2 \end{aligned}$$

■

5.9 Distribuzione uniforme

$$f(x) = \begin{cases} 0 & \text{se } x < a \vee x > b \\ \frac{1}{b-a} & \text{se } a \leq x \leq b \end{cases}$$

Per media e deviazione standard si dimostra che

$$\bar{x} = \frac{a+b}{2} \quad s_x = \frac{b-a}{2\sqrt{3}}$$

Dimostrazione. Per entrambe si nota che è possibile ridurre l'intervallo di integrazione ad $[a, b]$ in quanto al di fuori la pdf si annulla.

Per la media

$$\bar{x} = \mathbb{E}[x] = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}$$

Per la SD si calcola la varianza, applicando l'espressione dimostrata

$$\begin{aligned} s_x^2 &= \overline{x^2} - \bar{x}^2 = \int_a^b \frac{x^2}{b-a} dx - \frac{(a+b)^2}{4} = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \\ &= \frac{4b^2 + 4a^2 + 4ab - (3a^2 + 3b^2 + 6ab)}{12} = \frac{a^2 + b^2 - 2ab}{12} = \frac{(b-a)^2}{12} \end{aligned}$$

Da cui

$$s_x = \sqrt{s_x^2} = \frac{b-a}{\sqrt{12}} = \frac{b-a}{2\sqrt{3}}$$

■

Per la SD si osserva che essa fornisce, a partire dall'incertezza massima di una misura (data come metà dell'intervallo), quella che è definita **incertezza standard di risoluzione**

$$s_x = \frac{b-a}{2} \frac{1}{\sqrt{3}} = \frac{\text{incertezza massima}}{\sqrt{3}}$$

ovvero lo scarto quadratico medio (SD) di una **misura diretta ripetibile** dovuto alla **risoluzione** dello strumento viene calcolato come SD della distribuzione uniforme (di modo da essere trattabile come incertezza di tipo statistico!).

5.10 Distribuzione esponenziale

$$\Phi(t) = ce^{-\frac{t}{\tau}}$$

ove $\hat{\tau}$ = vita media.

Definita su $[0, +\infty[$

Si dimostra che

$$c = \frac{1}{\tau} \quad \bar{t} = \tau \quad s_t^2 = \tau^2$$

Dimostrazione. Si impone la normalizzazione

$$\int_0^{+\infty} ce^{-\frac{t}{\tau}} dt = 1 = [-c\tau e^{-\frac{t}{\tau}}]_0^{+\infty}$$

Ora, si osserva che il limite della primitiva per $t \rightarrow +\infty$ è 0. Dunque

$$[-c\tau e^{-\frac{t}{\tau}}]_0^{+\infty} = c\tau = 1 \implies c = \frac{1}{\tau}$$

Per la media

$$\bar{t} = \mathbb{E}[t] = \frac{1}{\tau} \int_0^{+\infty} t e^{-\frac{t}{\tau}} dt$$

integrando per parti

$$\bar{t} = \frac{1}{\tau} [-t\tau e^{-\frac{t}{\tau}}]_0^\infty + \frac{1}{\tau} \int_0^\infty \tau e^{-\frac{t}{\tau}} dt$$

Si osserva quindi che il primo termine ottenuto va a 0 per t tendente a $+\infty$, in quanto l'esponenziale è infinito di ordine superiore alla funzione lineare. In 0 invece si annulla a causa del termine di risonanza t :

$$\bar{t} = \frac{1}{\tau} [0 - 0] + [-\tau e^{-\frac{t}{\tau}}]_0^\infty = -\tau(0 - 1) = \tau$$

Per la varianza

$$s_t^2 = \mathbb{E}[t^2] - \mathbb{E}[t]^2 = \frac{1}{\tau} \int_0^{+\infty} t^2 e^{-\frac{t}{\tau}} dt - \tau^2$$

ora, integrando per parti

$$\int_0^{+\infty} e^{-\frac{t}{\tau}} dt = \underbrace{\left[-\tau t^2 e^{-\frac{t}{\tau}} \right]_0^{+\infty}}_0 + 2\tau \int_0^{+\infty} e^{-\frac{t}{\tau}} dt$$

Integrando quindi per parti il secondo termine

$$2\tau \int_0^{+\infty} e^{-\frac{t}{\tau}} dt = \underbrace{\left[-2\tau^2 t e^{-\frac{t}{\tau}} \right]_0^{+\infty}}_0 + 2\tau^2 \int_0^{+\infty} e^{-\frac{t}{\tau}} dt = 2\tau^3$$

per la normalizzazione. Allora

$$s_t^2 = \frac{2\tau^3}{\tau} - \tau^2 = \tau^2$$

■

È possibile verificare $\bar{t} = \tau$ anche applicando il metodo della ML:

Dimostrazione. Sia dato un campione t_1, \dots, t_N da una medesima popolazione esponenziale. Allora calcolando la probabilità congiunta:

$$P(\bigcap_i t_i) = \prod_i \Phi(t_i) = \frac{1}{\tau^N} e^{-\frac{\sum_i t_i}{\tau}}$$

Introducendo ora la funzione di Likelihood, ovvero prendendo il logaritmo:

$$\mathcal{L} = \ln P = -N \ln \tau - \frac{\sum_i t_i}{\tau}$$

annullandone la derivata rispetto al parametro:

$$\frac{\partial \mathcal{L}}{\partial \tau} = -\frac{N}{\tau} + \frac{\sum_i t_i}{\tau^2} = 0 \implies \hat{\tau} = \frac{\sum_i t_i}{N} = \bar{t}$$

■

5.11 Distribuzione di Cauchy-Lorentz

$$f(x) = \frac{1}{\pi} \frac{s}{x^2 + s^2}$$

corrisponde alla distribuzione di probabilità della posizione di incidenza (calcolata rispetto all'intersezione con la perpendicolare) di una radiazione isotropa su $(-\frac{\pi}{2}, \frac{\pi}{2})$. Il parametro s è la distanza tra la sorgente (puntiforme) e la superficie irradiata.

È utilizzata in fisica delle particelle, nella forma definita **distribuzione di Breit-Wigner**, per descrivere **fenomeni di risonanza**, ovvero picchi di energia in corrispondenza di specifici valori negli urti tra particelle. È più bassa al centro è più alta ai lati di una distribuzione normale. È particolarmente **patologica**, in quanto non sono definite né la media né la varianza: infatti gli integrali per calcolarle divergono!

Inoltre

- Il rapporto di due variabili gaussiane è una variabile distribuita secondo una pdf di Cauchy-Lorentz
- Somma di variabili di Cauchy-Lorentz è ancora una variabile di Cauchy-Lorentz (il che non contraddice tuttavia il teorema del limite centrale)

Si espone la derivazione

Dimostrazione. Si ha in primo luogo

$$x = s \tan \theta$$

differenziando

$$dx = \frac{dx}{d\theta} d\theta = \frac{s}{\cos^2 \theta} d\theta = s \cdot \frac{1}{s^2/(x^2 + s^2)} = \frac{x^2 + s^2}{s} d\theta$$

Ora, la probabilità per intorno infinitesimo di un valore di x e del corrispondente θ (a s parametro fissato) deve essere uguale:

$$\phi(\theta) \, d\theta = f(x) \, dx$$

Ma essendo θ distribuita secondo un'uniforme:

$$\frac{1}{\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)} d\theta = f(x) \, dx \implies f(x) = \frac{1}{\pi} \frac{d\theta}{dx} = \frac{1}{\pi} \frac{s}{x^2 + s^2}$$

■

Se l'intercetta sull'asse orizzontale ha ascissa $x_0 \neq 0$, si può applicare l'espressione generale per la distribuzione:

$$f(x) = \frac{1}{\pi} \frac{s}{(x - x_0)^2 + s^2}$$

Capitolo 6

Distribuzioni multivariate

Vi sono alcuni fenomeni che richiedono, per la propria descrizione, l'utilizzo di **più variabili casuali**, di cui si studia la distribuzione complessiva. Se $n = 2$ si parla quindi di **distribuzioni bivariate**, per $n > 1$ generalmente di **multivariate**.

Il caso trattato in particolare è quello di grandezze derivate misurate **indirettamente** da più grandezze misurate **direttamente** e soggette a errori casuali, ognuna con le proprie proprietà stocastiche, i propri parametri/indici.

Ci si limita inoltre alle bivariate, indicando con x e y le due variabili.

6.0.1 Scatter plot

Il primo passo è la realizzazione di un **grafico di dispersione**, in cui ogni punto corrisponde ad una coppia di valori misurati (x_i, y_i) , che permette di visualizzare l'intersezione tra le rette date da \bar{x} e \bar{y} , definite come medie aritmetiche dei rispettivi campioni.

6.0.2 Lego plot

Si procede quindi a realizzare l'istogramma in tre dimensioni in densità di frequenza $\Phi(x, y)$. Si osserva che è necessario $\Delta x, \Delta y \gg 0$ per il binnaggio.

La densità per un bin con coppia di valori medi (x_i, y_j) è quindi

$$\Phi_{i,j} = \frac{n_{ij}}{N \Delta x \Delta y}$$

e per ogni j si ha chiaramente

$$P(y_j) = \sum_{i=1}^N P(x_i, y_j) = \sum_{i=1}^N \Phi_{ij} \Delta x \Delta y$$

e vale la condizione di normalizzazione

$$\sum_{i=1}^N \sum_{j=1}^N P(x_i, y_j) = \sum_{i=1}^N \sum_{j=1}^N \Phi_{ij} \Delta x \Delta y = 1$$

6.0.3 Superficie limite

Passando quindi al limite

$$N \rightarrow \infty \quad \Delta x, \Delta y \rightarrow 0$$

il lego-plot in densità tende ad una **superficie limite** (continua), descritta da una **pdf bidimensionale**.

$$\Phi_{i,j} = \frac{n_{ij}}{N \Delta x \Delta y} \xrightarrow[\Delta x, \Delta y \rightarrow 0]{N \rightarrow \infty} \Phi(x, y)$$

Vale chiaramente la condizione di normalizzazione in forma integrale:

$$\int \int_{-\infty}^{+\infty} \Phi(x, y) \, dx \, dy = 1$$

e la probabilità per il prodotto cartesiano di due intervalli vale

$$P(a < x < b \wedge c < y < d) = \int_a^b \int_c^d \Phi(x, y) \, dx \, dy$$

e quindi per valori di una sola variabile (ovvero la cosiddetta **PDF marginale**):

$$\Phi_x(\tilde{x}) = \int_{-\infty}^{+\infty} \Phi(\tilde{x}, y) \, dy \quad \Phi_y(\tilde{y}) = \int_{-\infty}^{+\infty} \Phi(x, \tilde{y}) \, dx$$

si osserva tuttavia che, fatto salvo il caso di variabili indipendenti, è impossibile di converso ottenere la Φ note le Φ_x e Φ_y .

In tal caso si avrebbe invece

$$\Phi(x, y) = \Phi_x(x)\Phi_y(y)$$

6.0.4 PDF condizionale

La PDF condizionale per y dato x si ottiene secondo

$$h_y(y|x) = \frac{f(x, y)}{f_x(x)} = \frac{f(x, y)}{\int f(x, y') \, dy'}$$

Si tratta di una funzione della sola y , con x parametro costante. Di fatto il denominatore rappresenta un termine di rinormalizzazione. Ovviamente per x dato y

$$h_x(x|y) = \frac{f(x, y)}{f_y(y)} = \frac{f(x, y)}{\int f(x', y) \, dx'}$$

Si ha quindi il **teorema di Bayes per variabili continue**

$$h_x(x|y) = \frac{h_y(y|x)f_x(x)}{f_y(y)}$$

da cui le espressioni alternative per le marginali

$$f_x(x) = \int_{-\infty}^{\infty} h_x(x|y)f_y(y) \, dy \quad f_y(y) = \int_{-\infty}^{\infty} h_y(y|x)f_x(x) \, dx$$

6.1 Bivariate

Si considera quindi la grandezza misurata in modo indiretto tramite una generica dipendenza funzionale

$$q = q(x, y)$$

si dimostra allora che

$$\bar{q} = q(\bar{x}, \bar{y}) \quad e \quad \sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 + 2\left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right)\sigma_{xy}$$

ove la **covarianza** è definita secondo:

$$\sigma_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

Si dimostra prima per la media

Dimostrazione. Sia dato un campione indicizzato secondo

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

Si indichi quindi $q_i = q(x_i, y_i)$. Sviluppando in serie di Taylor con punto base $q_0 = q(\bar{x}, \bar{y})$ al primo ordine si ha:

$$q_i \approx q(\bar{x}, \bar{y}) + \left(\frac{\partial q}{\partial x}\right)_{\substack{x=\bar{x} \\ y=\bar{y}}} (x_i - \bar{x}) + \left(\frac{\partial q}{\partial y}\right)_{\substack{x=\bar{x} \\ y=\bar{y}}} (y_i - \bar{y})$$

quindi calcolando la media

$$\bar{q} = \frac{\sum_i q_i}{N} = \frac{1}{N} \left[Nq(\bar{x}, \bar{y}) \left(\frac{\partial q}{\partial x} \right)_{x=\bar{x}, y=\bar{y}} \sum_{i=1}^N (x_i - \bar{x}) + \left(\frac{\partial q}{\partial y} \right)_{x=\bar{x}, y=\bar{y}} \sum_{i=1}^N (y_i - \bar{y}) \right]$$

distribuendo ora le sommatorie:

$$\begin{aligned} \bar{q} &= q(\bar{x}, \bar{y}) + \left(\frac{\partial q}{\partial x} \right)_{x=\bar{x}, y=\bar{y}} \left(\frac{\sum_i x_i}{N} - \frac{N}{N} \bar{x} \right) + \left(\frac{\partial q}{\partial y} \right)_{x=\bar{x}, y=\bar{y}} \left(\frac{\sum_i y_i}{N} - \frac{N}{N} \bar{y} \right) = \\ &= q(\bar{x}, \bar{y}) + \left(\frac{\partial q}{\partial x} \right)_{x=\bar{x}, y=\bar{y}} (\bar{x} - \bar{x}) + \left(\frac{\partial q}{\partial y} \right)_{x=\bar{x}, y=\bar{y}} (\bar{y} - \bar{y}) = q(\bar{x}, \bar{y}) \end{aligned}$$

■

Per la varianza

Dimostrazione. Si applica la definizione, considerando i dof (diminuzione dovuta al calcolo della media)

$$s_q^2 = \frac{\sum_i (q_i - \bar{q})^2}{N - 1}$$

Si applica lo sviluppo effettuato in precedenza:

$$q_i - \bar{q} = q_i - q(\bar{x}, \bar{y}) = \left(\frac{\partial q}{\partial x} \right)_{x=\bar{x}, y=\bar{y}} (x_i - \bar{x}) + \left(\frac{\partial q}{\partial y} \right)_{x=\bar{x}, y=\bar{y}} (y_i - \bar{y})$$

inserendo nella formula

$$\begin{aligned} s_q^2 &= \frac{1}{N - 1} \sum_{i=1}^N \left[\left(\frac{\partial q}{\partial x} \right)_{x=\bar{x}, y=\bar{y}} (x_i - \bar{x}) + \left(\frac{\partial q}{\partial y} \right)_{x=\bar{x}, y=\bar{y}} (y_i - \bar{y}) \right]^2 = \\ &= \frac{1}{N - 1} \left(\frac{\partial q}{\partial x} \right)_{x=\bar{x}, y=\bar{y}}^2 \sum_{i=1}^N (x_i - \bar{x}) + \frac{1}{N - 1} \left(\frac{\partial q}{\partial y} \right)_{x=\bar{x}, y=\bar{y}}^2 \sum_{i=1}^N (y_i - \bar{y}) + \frac{2}{N - 1} \left(\frac{\partial q}{\partial x} \right)_{x=\bar{x}, y=\bar{y}} \left(\frac{\partial q}{\partial y} \right)_{x=\bar{x}, y=\bar{y}} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) = \\ &= \left(\frac{\partial q}{\partial x} \right)_{x=\bar{x}, y=\bar{y}}^2 \underbrace{\frac{\sum_i (x_i - \bar{x})}{N - 1}}_{\sigma_x^2} + \left(\frac{\partial q}{\partial y} \right)_{x=\bar{x}, y=\bar{y}}^2 \underbrace{\frac{\sum_i (y_i - \bar{y})}{N - 1}}_{\sigma_y^2} + 2 \left(\frac{\partial q}{\partial x} \right)_{x=\bar{x}, y=\bar{y}} \left(\frac{\partial q}{\partial y} \right)_{x=\bar{x}, y=\bar{y}} \underbrace{\frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{N - 1}}_{\sigma_{xy}} \end{aligned}$$

Si osserva che al denominatore della covarianza è presente $N - 1$, a differenza dell' N della definizione. Ciò segue dal fatto che vi sono $N - 1$ dof nel calcolo di entrambe le varianze campionarie. La contraddizione è evitata in quanto la covarianza è utile/efficace, e viene quindi valutata, per campioni numerosi: per essi vale chiaramente $N \approx N - 1$. (Ciò non è invece necessariamente vero per le varianze). ■

Si dimostra ora che

$$\sigma_{xy} = \overline{xy} - \bar{x} \cdot \bar{y}$$

Dimostrazione. Dalla definizione

$$\begin{aligned} \sigma_{xy} &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) = \frac{\sum_i (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \cdot \bar{y})}{N} = \overline{xy} - \bar{y} \frac{\sum_i x_i}{N} - \bar{x} \frac{\sum_i y_i}{N} + \frac{N}{N} \bar{x} \cdot \bar{y} = \\ &= \overline{xy} - 2\bar{x} \cdot \bar{y} + \bar{x} \cdot \bar{y} = \overline{xy} - \bar{x} \cdot \bar{y} \end{aligned}$$

■

Suddividendo lo scatter plot in quattro quadranti tramite le rette $x = \bar{x}$ e $y = \bar{y}$, si osserva:

I $\sigma_{xy} < 0$	II $\sigma_{xy} > 0$
III $\sigma_{xy} > 0$	IV $\sigma_{xy} < 0$

Se gli scarti in x e quelli in y sono **indipendenti**, poiché si può assumere che i valori delle due variabili siano equamente distribuiti intorno ai rispettivi valori medi, fissato un generico valore di una variabile vi saranno coppie dell'altra che danno scarti simmetrici e quindi la sommatoria che definisce la covarianza sarà complessivamente nulla.

6.2 Covarianza e correlazione

Se gli scarti **non** sono indipendenti, ovvero si ha un'influenza di quelli su x su quelli su y o viceversa (o entrambi) si ha invece una covarianza **non nulla**: ad esempio in caso di correlazione simil-lineare, una covarianza **positiva** se il coeff angolare del fit $y(x)$ è maggiore di 0, viceversa negativa per coeff ang negativi (generalmente $\sigma_{xy} < 0 \implies$ **anticorrelazione**).

Covarianza nulla \nRightarrow indipendenti!

Si tratta di una condizione **necessaria** ma **non sufficiente**: anche una dipendenza quadratica $y = x^2$ dà covarianza nulla in quanto funzione pari!

6.3 Somma in quadratura

Se le misure delle due variabili sono indipendenti, la covarianza si annulla e quindi la varianza di q si riduce a:

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2$$

tale formula è definita **somma in quadratura**. Si ribadiscono le condizioni di applicabilità:

1. Misure dirette soggette a incertezze di tipo casuale
2. Misure indipendenti

La trattazione effettuata finora può essere estesa al caso delle **multivariate** introducendo la **matrice di covarianza**

Matrice di covarianza

Per n variabili casuali x_n e due loro funzioni a, b si costruisce una matrice simmetrica V_{ij} ove gli ingressi diagonali sono le varianze:

$$\sigma_a^2 = V_{aa} = \mathbb{E}[(a - \bar{a})^2] \quad \text{ove } \bar{a} = \mathbb{E}[a(\mathbf{x})] = \int \cdots \int a(\mathbf{x}) \Phi(\mathbf{x}) dx_1 \cdots dx_n$$

e gli altri sono dati dalle covarianze (uguali per permutazione degli indici in quanto simmetrica):

$$V_{ab} = V_{ba} = \mathbb{E}[(a - \bar{a})(b - \bar{b})] = \int \cdots \int a(\mathbf{x})b(\mathbf{x})\Phi(\mathbf{x}) dx_1 \cdots dx_n - \bar{a} \cdots \bar{b}$$

Se tutte le misure sono indipendenti, ovvero la matrice di covarianza è diagonale, si può estendere la formula della somma in quadratura:

$$q = q(\mathbf{x}) \implies \sigma_q^2 = \sum_{i=1}^n \left(\frac{\partial q}{\partial x_i}\right)^2 \sigma_i^2$$

É ora giustificata la dimostrazione della deviazione standard della media.

6.4 Incertezze in misure dipendenti

Se non è possibile una valutazione esatta della covarianza, è comunque possibile stimarla tramite la **disuguaglianza di Schwarz**:

$$|\sigma_{xy}| \leq \sigma_x \sigma_y$$

Dimostrazione. Esprimendo in notazione vettoriale

$$\mathbf{x} = (x_1, \dots, x_N) \quad \mathbf{y} = (y_1, \dots, y_N)$$

e indicando il vettore unitario con $\mathbf{1}$ si ha

$$|\sigma_{xy}| = \frac{1}{N-1} |(\mathbf{x} - \bar{x}\mathbf{1}) \cdot (\mathbf{y} - \bar{y}\mathbf{1})|$$

ove $|\cdot|$ indica la norma euclidea e \cdot il prodotto scalare euclideo. Applicando ora la disuguaglianza di Cauchy-Schwarz, si maggiore la norma del prodotto scalare con il prodotto delle norme:

$$\begin{aligned} |(\mathbf{x} - \bar{x}\mathbf{1}) \cdot (\mathbf{y} - \bar{y}\mathbf{1})| &\leq |(\mathbf{x} - \bar{x}\mathbf{1})| |(\mathbf{y} - \bar{y}\mathbf{1})| \implies |\sigma_{xy}| \leq \frac{1}{N-1} \sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2} = \\ &= \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2} = \sigma_x \sigma_y \end{aligned}$$

■

Si può poi dimostrare che in ogni caso (per misure dipendenti o indipendenti) vale

$$\sigma_q \leq \left| \frac{\partial q}{\partial x} \right| \sigma_x + \left| \frac{\partial q}{\partial y} \right| \sigma_y$$

ovvero la **somma lineare** delle incertezze dà un limite superiore all'errore **per misure correlate** (la quadratura potrebbe infatti dare una sottostima dell'errore a causa del termine di covarianza).

Dimostrazione. Per definizione del valore assoluto

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y} \right)^2 \sigma_y^2 + 2 \left(\frac{\partial q}{\partial x} \right) \left(\frac{\partial q}{\partial y} \right) \sigma_{xy} \leq \left(\frac{\partial q}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y} \right)^2 \sigma_y^2 + 2 \left(\frac{\partial q}{\partial x} \right) \left(\frac{\partial q}{\partial y} \right) |\sigma_{xy}|$$

applicando Schwarz

$$\sigma_q^2 \leq \left(\frac{\partial q}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y} \right)^2 \sigma_y^2 + 2 \left(\frac{\partial q}{\partial x} \right) \left(\frac{\partial q}{\partial y} \right) \sigma_x \sigma_y = \left(\frac{\partial q}{\partial x} \sigma_x + \frac{\partial q}{\partial y} \sigma_y \right)^2$$

Prendendo la radice quadrata si perviene al risultato.

■

6.5 Combinazione lineare

La c.l.

$$q = c_1 \cdot x + c_2 \cdot y$$

di due variabili **gaussiane indipendenti** è **a sua volta una variabile gaussiana** di media

$$\bar{q} = c_1 \cdot \bar{x} + c_2 \cdot \bar{y}$$

e varianza

$$\sigma_q^2 = c_1^2 \cdot \sigma_x^2 + c_2^2 \cdot \sigma_y^2$$

(segue dalla quadratura).

Tale risultato è generalizzabile ad una c.l. di **numero arbitrario** di variabili gaussiane. Si procede alla dimostrazione

Dimostrazione. Per la media

$$q = c_1x + c_2y \implies q_i = c_1x_i + c_2y_i$$

e dunque

$$\bar{q} = \frac{\sum_i q_i}{N} = \frac{\sum_i (c_1x_i + c_2y_i)}{N} = c_1 \frac{\sum_i x_i}{N} + c_2 \frac{\sum_i y_i}{N} = c_1\bar{x} + c_2\bar{y}$$

Per la varianza

$$\begin{aligned} \sigma_q^2 &= \frac{\sum_i (q_i - \bar{q})^2}{N-1} = \frac{\sum_i (q_i^2 + \bar{q}^2 - 2q_i\bar{q})}{N-1} = \frac{\sum_i [(c_1x_i + c_2y_i)^2 + (c_1\bar{x} + c_2\bar{y})^2 - 2(c_1x_i + c_2y_i)(c_1\bar{x} + c_2\bar{y})]}{N-1} = \\ &= \frac{\sum_i c_1^2(x_i^2 + \bar{x}^2 - 2x_i\bar{x})}{N-1} + \frac{\sum_i c_2^2(y_i^2 + \bar{y}^2 - 2y_i\bar{y})}{N-1} + \frac{\sum_i 2c_1c_2(x_iy_i + \bar{x}\bar{y} - x_i\bar{y} - y_i\bar{x})}{N-1} = \\ &= c_1^2 \frac{\sum_i (x_i - \bar{x})^2}{N-1} + c_2^2 \frac{\sum_i (y_i - \bar{y})^2}{N-1} + 2c_1c_2 \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{N-1} = c_1^2\sigma_x^2 + c_2^2\sigma_y^2 + 2c_1c_2\sigma_{xy} \end{aligned}$$

essendo assunte indipendenti, si elide il termine di covarianza:

$$\sigma_q^2 = c_1^2\sigma_x^2 + c_2^2\sigma_y^2$$

■

6.6 Correlazione lineare

Quanto di ha una correlazione dovuta ad una **relazione funzionale**, in particolare di tipo lineare (che non necessariamente rispecchia una relazione fisica tra le grandezze!) si introduce il

Coefficiente di correlazione lineare

$$r \equiv \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

Segue chiaramente da Schwarz, poiché le SD sono positive:

$$|r| = \frac{|\sigma_{xy}|}{\sigma_x\sigma_y} \leq 1 \implies -1 \leq r \leq 1$$

r permette di valutare la **validità dell'ipotesi di dipendenza lineare** tra le due variabili. In funzione di N infatti si può esprimere quantitativamente la **probabilità che, se le due variabili non sono legate da dipendenza lineare, le N misure forniscano un r maggiore o uguale in valore assoluto di quello ottenuto**.

- Se $P(|r'| > |r|) < 5\%$ la correlazione è significativa
- Se $P(|r'| > |r|) < 1\%$ la correlazione è altamente significativa
- Se $P \geq 5\%$ non è significativa

nei primi due casi è molto più probabile vi sia una correlazione. Si osserva che la probabilità diminuisce aumentando N : chiaramente incrementando il numero di misure la correlazione *emerge* in modo maggiormente evidente.

6.7 Media pesata di misure indipendenti

Si abbiano due misure della stessa grandezza X

$$(x_1 \pm \sigma_1) \quad (x_2 \pm \sigma_2)$$

se esse

- sono compatibili **ma**
- hanno incertezze differenti

la miglior stima \hat{X} del valor vero risultante si ottiene combinando le due misure tramite una **media pesata**

$$\hat{X} = x_p = \frac{w_1 x_1 + w_2 x_2}{w_1 + w_2}$$

ove i **pesi** si ottengono secondo

$$w_i = \frac{1}{\sigma_i^2}$$

L'incertezza da associarvi si ottiene secondo

$$\sigma_{x_p} = \frac{1}{\sqrt{w_1 + w_2}}$$

Generalizzando ad un numero arbitrario di misure si ha poi

$$x_p = \frac{\sum_i w_i x_i}{\sum_i w_i} \quad \sigma_{x_p} = \frac{1}{\sqrt{\sum_i w_i}}$$

Dimostrazione. Si supponga $n = 2$. Ogni misura è governata da una distribuzione normale, ed essendo indipendenti la probabilità congiunta è

$$P(x_1, x_2) = P(x_1)P(x_2) \propto \frac{1}{\sigma_1} e^{-\frac{(x_1 - x)^2}{2\sigma_1^2}} \cdot \frac{1}{\sigma_2} e^{-\frac{(x_2 - x)^2}{2\sigma_2^2}} \propto e^{-\frac{(x_1 - x)^2}{2\sigma_1^2} - \frac{(x_2 - x)^2}{2\sigma_2^2}}$$

Si applica ora la ML: per massimizzare la probabilità, essendo l'esponentiale monotona crescente, si annulla la derivata dell'esponente rispetto alla media comune x da determinarsi:

$$\frac{d}{dx} \left(e^{-\frac{(x_1 - x)^2}{2\sigma_1^2} - \frac{(x_2 - x)^2}{2\sigma_2^2}} \right) = 0 \implies 2 \frac{(x_1 - x)}{2\sigma_1^2} + 2 \frac{(x_2 - x)}{2\sigma_2^2} = 0$$

da cui

$$\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2} - \frac{x}{\sigma_1^2} - \frac{x}{\sigma_2^2} = 0 \implies \hat{x} = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

che si riconduce alla formula data sostituendo $w_i = \frac{1}{\sigma_i^2}$ ■

Per l'incertezza

Dimostrazione. Essendo le misure indipendenti, si somma in quadratura:

$$\sigma_x^2 = \left(\frac{w_1}{w_1 + w_2} \right)^2 \sigma_1^2 + \left(\frac{w_2}{w_1 + w_2} \right)^2 \sigma_2^2 = \frac{w_1^2}{(w_1 + w_2)^2} \frac{1}{w_1} + \frac{w_2^2}{(w_1 + w_2)^2} \frac{1}{w_2} = \frac{w_1 + w_2}{(w_1 + w_2)^2} = \frac{1}{w_1 + w_2}$$

da cui si ottiene la SD prendendo la radice. ■

6.8 Regressione lineare

Sia dato un insieme di N misure x_i e corrispondenti $(y_i \pm \sigma_y)$ e si ipotizzi che

1. gli errori sulle x_i siano trascurabili
2. le grandezze y_i siano estratte da popolazioni gaussiane
3. l'errore σ_y sulle y_i sia uniforme (gli errori siano tutti uguali)
4. x e y siano legate da una **relazione lineare** ovvero nella forma

$$y = A + Bx$$

É allora possibile applicare un metodo analitico di interpolazione dei punti sperimentali definito **regressione lineare** o **adattamento dei minimi quadrati per una retta**.

Le migliori stime dei parametri di regressione (intercetta all'origine e coefficiente angolare) sono date da

$$A = \frac{\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i}{\Delta} \quad B = \frac{N \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{\Delta}$$

ove si ha

$$\Delta = N \sum_i x_i^2 - \left(\sum_i x_i \right)^2$$

e le rispettive incertezze sono

$$\sigma_A = \sigma_y \sqrt{\frac{\sum_i x_i^2}{\Delta}} \quad \sigma_B = \sigma_y \sqrt{\frac{N}{\Delta}}$$

Nel caso invece le incertezze sulle y fossero ignote, possono essere stimate tramite i parametri secondo

$$\sigma_y = \sqrt{\frac{\sum (y_i - A - Bx_i)^2}{N - 2}}$$

ove il denominatore è determinato dal numero di dof

Dimostrazione delle stime

Dimostrazione. Si suppone che ogni misura y_i segua una distribuzione normale centrata sul valor vero con incertezza (SD) corrispondente a σ_y .

Allora la probabilità di osservare singolarmente ogni y_i è data da

$$P(y_i) \propto \frac{1}{\sigma_y} e^{-\frac{(y_i - y_{i,vero})^2}{2\sigma_y^2}}$$

Supposto che il valor vero corrisponda a quello dato dalla relazione lineare, si ha

$$P(y_i) \propto \frac{1}{\sigma_y} e^{-\frac{(y_i - A - Bx_i)^2}{2\sigma_y^2}}$$

Considerando ora la probabilità congiunta dell'intero campione, assunte le misure indipendenti

$$P(y_1, \dots, y_N) = \prod_i P(y_i) \propto \frac{1}{\sigma_y^N} e^{-\frac{\sum_i (y_i - A - Bx_i)^2}{2\sigma_y^2}}$$

Si applica la ML: la probabilità è massimizzata quando lo è l'esponente, ovvero la funzione di likelihood $\mathcal{L} = \ln P$. Dunque si imposta un sistema annullando la derivata rispetto ai due parametri:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial A} = 0 \\ \frac{\partial \mathcal{L}}{\partial B} = 0 \end{cases} \Rightarrow \begin{cases} \sum_i \frac{y_i - A - Bx_i}{\sigma_y^2} = 0 \\ \sum_i \frac{(y_i - A - Bx_i)(x_i)}{\sigma_y^2} = 0 \end{cases} \Rightarrow \begin{cases} \sum_i y_i - NA - B \sum_i x_i = 0 \\ \sum_i x_i y_i - A \sum_i x_i - B \sum_i x_i^2 = 0 \end{cases}$$

riordinando e moltiplicando la prima equazione per $\sum_i x_i^2$ e la seconda per $\sum_i x_i$

$$\begin{cases} \sum_i x_i^2 \cdot [NA + B \sum_i x_i] = \sum_i y_i \cdot \sum_i x_i^2 \\ \sum_i x_i \cdot [A \sum_i x_i + B \sum_i x_i^2] = \sum_i x_i y_i \cdot \sum_i x_i \end{cases}$$

Sottraendo ora la seconda equazione alla prima:

$$AN \sum_i x_i^2 - A \left(\sum_i x_i \right)^2 = \sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i$$

e isolando il parametro:

$$\hat{A} = \frac{\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i}{N \sum_i x_i^2 - (\sum_i x_i)^2} = \frac{\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i}{\Delta}$$

Per la stima di B si riprende il sistema precedente e si moltiplica la prima equazione per $\sum_i x_i$ e la seconda per N , quindi si sottrae questa alla precedente:

$$\begin{cases} \sum_i x_i \left[AN + B \sum_i x_i \right] = \sum_i y_i \cdot \sum_i x_i \\ NA \sum_i x_i + NB \sum_i x_i^2 = N \sum_i x_i y_i \end{cases} \implies B \left(\sum_i x_i \right)^2 - NB \sum_i x_i^2 = \sum_i x_i \sum_i y_i - N \sum_i x_i y_i$$

e quindi

$$B \left[N \sum_i x_i^2 - \left(\sum_i x_i \right)^2 \right] = N \sum_i x_i y_i - \sum_i x_i \sum_i y_i$$

da cui

$$\hat{B} = \frac{N \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{N \sum_i x_i^2 - (\sum_i x_i)^2} = \frac{N \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{\Delta}$$

■

si ha quindi la **retta verità**, che interpola gli intervalli delle singole y_i (di semiampiezza σ_y). I valori determinati corrispondono alla media tra i valori estremi, con incertezza data dalla semidispersione:

$$\hat{A} = \frac{A_{max} + A_{min}}{2} \pm \frac{A_{max} - A_{min}}{2} \quad \hat{B} = \frac{B_{max} + B_{min}}{2} \pm \frac{B_{max} - B_{min}}{2}$$

ove i valori massimi e minimi sono quelli estremi per cui la retta ottenuta intercetta tutti gli intervalli.

Dimostrazione delle incertezze

Dimostrazione. Poiché le singole misure sono indipendenti, si ottiene σ_A sommando in quadratura

$$\begin{aligned} \sigma_A^2 &= \sum_J \left(\frac{\partial A}{\partial y_J} \right)^2 \sigma_y^2 = \sum_J \left(\frac{\sum_i x_i^2 - x_J \sum_i x_i}{\Delta} \right)^2 \sigma_y^2 = \sum_J \frac{(\sum_i x_i^2)^2 + x_J^2 (\sum_i x_i)^2 - 2x_J \sum_i x_i \sum_i x_i^2}{\Delta^2} \sigma_y^2 = \\ &= \left[\frac{N(\sum_i x_i^2)^2 + \sum_J x_J^2 (\sum_i x_i)^2 - 2 \sum_J x_J \sum_i x_i \sum_i x_i^2}{\Delta^2} \right] \sigma_y^2 = \frac{[N \sum_i x_i^2 - (\sum_i x_i)^2] \sum_i x_i^2}{[N \sum_i x_i^2 - (\sum_i x_i)^2] \Delta} \sigma_y^2 = \frac{\sum_i x_i^2}{\Delta} \sigma_y^2 \end{aligned}$$

Per B si procede analogamente:

$$\begin{aligned} \left(\frac{\partial B}{\partial y_J} \right)^2 &= \frac{(Nx_J - \sum_i x_i)^2}{\Delta^2} = \frac{N^2 x_J^2 + (\sum_i x_i)^2 - 2Nx_J \sum_i x_i}{\Delta^2} \implies \\ \implies \sum_J \left(\frac{\partial B}{\partial y_J} \right)^2 &= \frac{N^2 \sum_i x_i^2 + N(\sum_i x_i)^2 - 2N(\sum_i x_i)^2}{\Delta^2} = \\ &= \frac{N\Delta}{\Delta^2} = \frac{N}{\Delta} \implies \sigma_B^2 = \frac{N}{\Delta} \sigma_y^2 \end{aligned}$$

■

Correlazione lineare se la correlazione lineare tra x e y è certa, si ha

$$\bar{y} = A + B\bar{x} \implies (y_i - \bar{y}) = B(x_i - \bar{x})$$

e dunque

$$r = \frac{\sum_i (x_i - \bar{x}) B (x_i - \bar{x})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i B^2 (x_i - \bar{x})^2}} = \frac{B \sum_i (x_i - \bar{x})^2}{|B| \sqrt{(\sum_i (x_i - \bar{x})^2)^2}} = \frac{B}{|B|} = \pm 1$$

Regressione pesata si effettua se gli errori sulle y non sono tutti uguali. Assegnando i pesi secondo

$$w_i = \frac{1}{\sigma_{y_i}^2}$$

si ha

$$A = \frac{\sum_i w_i x_i^2 \sum_i w_i y_i - \sum_i w_i x_i \sum_i w_i x_i y_i}{\Delta} \quad B = \frac{\sum_i w_i \sum_i w_i x_i y_i - \sum_i w_i x_i \sum_i w_i y_i}{\Delta}$$

ove è definito

$$\Delta = \sum_i w_i \sum_i w_i x_i^2 - \left(\sum_i w_i x_i \right)^2$$

per le incertezze

$$\sigma_A = \sqrt{\frac{\sum_i w_i x_i^2}{\Delta}} \quad \sigma_B = \sqrt{\frac{\sum_i w_i}{\Delta}}$$

Dimostrazione. La probabilità congiunta dà

$$P_{campione} = \prod P_i \propto e^{-\frac{w_i}{2}(y_i - A - Bx_i)^2}$$

Ottenendo la funzione di likelihood tramite logaritmo ed imponendo l'annullamento delle derivate parziali rispetto ai parametri:

$$\begin{cases} \frac{\partial \ln P}{\partial A} = \sum_i w_i (y_i - A - Bx_i) = 0 \\ \frac{\partial \ln P}{\partial B} = \sum_i w_i x_i (y_i - A - Bx_i) = 0 \end{cases} \Rightarrow \begin{cases} A \sum_i w_i + B \sum_i w_i x_i = \sum_i w_i y_i \\ A \sum_i w_i x_i + B \sum_i w_i x_i^2 = \sum_i w_i x_i y_i \end{cases}$$

Per determinare A , si moltiplica la prima equazione per $\sum_i w_i x_i^2$ e la seconda per $\sum_i w_i x_i$, procedendo poi a sottrarre la seconda alla prima:

$$A \sum_i w_i \sum_i w_i x_i^2 - A \left(\sum_i w_i x_i \right)^2 + B \cdot 0 = \sum_i w_i y_i \sum_i w_i x_i^2 - \sum_i w_i x_i \sum_i w_i x_i y_i$$

Definendo ora

$$\Delta = \sum_i w_i \sum_i w_i x_i^2 - \left(\sum_i w_i x_i \right)^2$$

si ha

$$A = \frac{\sum_i w_i x_i^2 \sum_i w_i y_i - \sum_i w_i x_i \sum_i w_i x_i y_i}{\Delta}$$

Riprendendo ora il sistema precedente, si moltiplica la prima equazione per $\sum_i w_i x_i$ e la seconda per $\sum_i w_i$; si procede quindi a sottrarre la seconda alla prima:

$$A \cdot 0 + B \left(\sum_i w_i x_i \right)^2 - B \sum_i w_i \sum_i w_i x_i^2 = -B \Delta = \sum_i w_i y_i \sum_i w_i x_i - \sum_i w_i \sum_i w_i x_i y_i$$

e quindi

$$B = \frac{\sum_i w_i \sum_i w_i x_i y_i - \sum_i w_i x_i \sum_i w_i y_i}{\Delta}$$

Per le incertezze, notando che il termine Δ dipende dalle sole x_i e dai pesi w_i :

$$\left(\frac{\partial A}{\partial y_J} \right)^2 = \frac{1}{\Delta^2} \left(w_J \sum_i w_i x_i^2 - w_J x_J \sum_i w_i x_i \right)^2 = \frac{1}{\Delta} \left(w_J^2 \left(\sum_i w_i x_i^2 \right)^2 + w_J^2 x_J^2 \left(\sum_i w_i x_i \right)^2 - 2 w_J^2 x_J \sum_i w_i x_i^2 \sum_i w_i x_i \right)$$

Osservando ora che $w_J = \frac{1}{\sigma_J^2} \Rightarrow \sigma_J^2 = \frac{1}{w_J}$ si ha per la somma in quadratura:

$$\sigma_A^2 = \sum_J \left(\frac{\partial A}{\partial y_J} \right)^2 \sigma_J^2 = \frac{1}{\Delta^2} \left[\sum_J w_J \left(\sum_i w_i x_i^2 \right)^2 + \left(\sum_i w_i x_i \right)^2 \sum_J w_J x_J^2 - 2 \left(\sum_i w_i x_i^2 \right) \left(\sum_i w_i x_i \right) \left(\sum_J w_J x_J \right) \right] =$$

$$= \frac{1}{\Delta^2} \left[\sum_i w_i \left(\sum_i w_i x_i^2 \right)^2 - \left(\sum_i w_i x_i \right)^2 \sum_i w_i x_i^2 \right] = \frac{\Delta \sum_i w_i x_i^2}{\Delta^2} = \frac{\sum_i w_i x_i^2}{\Delta}$$

Per l'incertezza su B

$$\left(\frac{\partial B}{\partial y_J} \right)^2 = \frac{1}{\Delta^2} \left(w_J x_J \sum_i w_i - w_J \sum_i w_i x_i \right)^2 = \frac{1}{\Delta^2} \left(w_J^2 x_J^2 \left(\sum_i w_i \right)^2 + w_J^2 \left(\sum_i w_i x_i \right)^2 - 2 w_J^2 x_J \sum_i w_i \sum_i w_i x_i \right)$$

Sommando in quadratura con la sostituzione $\sigma_J^2 = \frac{1}{w_J}$ (si opera implicitamente opportuno cambio di indici)

$$\sigma_B^2 = \sum_J \left(\frac{\partial B}{\partial y_J} \right)^2 \sigma_J^2 = \frac{1}{\Delta^2} \left[\sum_i w_i x_i^2 \left(\sum_i w_i \right)^2 - \sum_i w_i \left(\sum_i w_i x_i \right)^2 \right] = \frac{\Delta \sum_i w_i}{\Delta^2} = \frac{\sum_i w_i}{\Delta}$$

■

Per incertezze non nulle su x si calcola l'errore equivalente secondo

$$\sigma_{y,eq} = \frac{dy}{dx} \sigma_x = B \sigma_x$$

e quindi si somma in quadratura:

$$\sigma_{y,tot} = \sqrt{\sigma_y^2 + B^2 \sigma_x^2}$$

Nel caso di retta per l'origine ovvero $A = 0$ si ha

$$\hat{B} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \quad \sigma_B = \frac{\sigma_y}{\sqrt{\sum_i x_i^2}}$$

6.8.1 Linearizzazione

È possibile trattare analogamente funzioni non lineari **linearizzandole** e quindi applicando la regressione. Ad esempio per dipendenza esponenziale

$$y(x) = y_0 e^{-\mu x}$$

si prende il logaritmo, riducendo a funzione lineare

$$z(x) = \ln y(x) = \underbrace{\ln y_0}_A - \underbrace{\mu x}_{+Bx}$$

il tutto adattando opportunamente le incertezze (minimi quadrati pesati). Vedasi in seguito per la trattazione di altri tipi di dipendenze funzionali.

6.9 Rigetto di dati

Se una misura in un campione **gaussiano** si discosta in modo evidente dalle altre, è possibile applicare un criterio per valutare la possibilità di rigettarla. Esso afferma che:

Se il numero di misure improbabili almeno quanto la misura sospetta è minore di 0.5 (come numero **assoluto**), allora questa può essere rigettata

Le misure parimenti o più improbabili sono quelle la cui probabilità di verificarsi è \leq quella della sospetta. Per calcolare la probabilità si riporta a variabile standard:

$$z_s = \frac{x_{sospetto} - \bar{x}}{\sigma_x}$$

ove σ_x è la SD **campionaria** e si effettua test **a due code** (dunque con val. assoluto: è noto che la curva è simmetrica, e il criterio si focalizza sulla distanza dal valor medio, non il valore esatto della misura!):

$$P(|z| > |z_s|) = 2 \int_{|z_s|}^{\infty} G(z) dz$$

si calcola dunque la *probabilità di fare peggio*. Segue dalla normalizzazione e dalla simmetria della curva:

$$P(|z| > |z_s|) = 1 - 2 \int_0^{|z_s|} G(z) \, dz$$

Quindi il numero di eventi attesi è dato secondo:

$$N_{attesi} = N_{tot} \cdot P(|z| > |z_s|)$$

da confrontarsi con 0.5. Chiaramente si tratta di una soglia convenzionale.

In caso di rigetto si procede quindi a ricalcolare i parametri del campione e della corrispondente distribuzione.

Capitolo 7

DAQ Sistemi di acquisizione dati

Sensore = dispositivo capace di generare risposta Y (**segnale di output**, e.g. una ddp) in presenza di grandezza fisica X (**input**)

$Y = F(X)$ è la **curva caratteristica (o di calibrazione)**

$$\text{Sensibilità} = F'(x) = \frac{dF(x)}{dx}$$

La curva caratt non è sempre lineare \implies sensibilità è specifica del **punto di lavoro** (valore di X)

Accuratezza di un sensore = massima differenza fra output effettivo e corretto (misurato con standard) **a parità di input**

$$accuratezza = \max[F(x) - F_{std}(X)]$$

può essere assoluta o in percentuale sulla lettura. Quantifica sistematici.

Catena di misura

Sensore \longrightarrow Amplificatore (1) \rightarrow Digitizzatore (Analog to digital Converter) (2) \longrightarrow Computer

\uparrow Data Acquisition System

Amplificatore moltiplica segnale per f_{amp} opportuno

ADC converte input continuo in binario

Se $f_{amp} < 1$ si ha **attenuazione**. Fattore scelto in modo da sfruttare tutta la **finestra di acquisizione**.

ADC caratterizzato da N_{bit} impiegati x rappresentare segnale.

+ **livelli** \implies rappresentazione segnale digitale + **fedele** all'analogico

Caratterizzato anche da **portata**, ovvero range **in ingresso**

$$[ADC_{min}, ADC_{max}]$$

L'**output range** dell'amplificatore deve rientrare nella portata:

$$\begin{cases} f_{amp} Y_{min} \geq ADC_{min} \\ f_{amp} Y_{max} \leq ADC_{max} \end{cases}$$

possibili valori portata ELVIS:

- Bipolare $[-10, 10]$ V
- Unipolare $[0, 10]$ V

Code width = più piccola variazione rivelabile nel segnale:

$$c.w. = \frac{portata}{f_{amp} \cdot 2^{N_{bit}}}$$

da cui si ottiene la **risoluzione**

$$risoluzione = \frac{code\ width}{sensibilit\grave{a}} = \frac{portata}{f_{amp} \cdot 2^{N_{bit}} \cdot F'(x)}$$

Dimostrazione. L'aggiunta di un bit permette di raddoppiare il numero di sottointervalli discreti di segnale in ingresso nell'ADC, che permettono di 'risolvere' digitalmente differenze minori. La differenza minima nell'output del sensore apprezzabile dalla DAQ, ovvero il code width, è quindi dato da

$$c.w. = \frac{Y_{max} - Y_{min}}{2^{N_{bit}}}$$

Ora, per quanto visto sull'amplificatore (l'altro componente della DAQ) si può sostituire

$$ADC_{max} - ADC_{min} = f_{amp}(Y_{max} - Y_{min}) \implies Y_{max} - Y_{min} = \frac{ADC_{max} - ADC_{min}}{f_{amp}}$$

e quindi ricavare la formula nota.

Per la risoluzione, si osserva che per variazioni minime di input ed output del sensore è possibile linearizzare:

$$\Delta Y \approx F'(X)\Delta X \implies \Delta X = \frac{\Delta Y}{F'(X)}$$

ove la derivata della curva caratteristica corrisponde, come noto, alla sensibilità.

Poiché il ΔY minimo apprezzabile è proprio il code width, e la risoluzione è per definizione la minima variazione nell'**input** del sensore (ovvero nella grandezza da misurare) che può essere apprezzata dall'intera catena di misura, si ha

$$risoluzione = \frac{c.w.}{F'(X)} = \frac{code\ width}{sensibilit\grave{a}}$$

■

7.1 Campionamento

Per la conversione di un segnale:

analogico (continuo) \longrightarrow **campionamento** \longrightarrow **digitale** (discreto)

Serie discreta di campionamenti \rightarrow numero finito di punti. Campionamento avviene con specifica frequenza: **tasso di campionamento**

Necessario sia adeguata a ricostruzione del segnale originario: **Frequenza di Nyquist** garantisce corretta ricostruzione.

Esiste poi un teorema che determina la soglia necessaria per la frequenza di campionamento

Teorema di Nyquist-Shannon

Il segnale analogico può essere ricostruito *completamente* se campionato ad una frequenza

$$f_c \geq 2f_M$$

ove f_M indica la frequenza del componente del segnale a frequenza maggiore.

Campioni ridotti e intervalli di confidenza

Si consideri un campione di N misure $\{x_i\}$ estratte da popolazione **gaussiana**. Per quanto visto le migliori stime dei parametri di $G(x; \mu, \sigma)$ sono

$$\hat{\mu} = \bar{x} \quad \hat{\sigma} = \sigma_x$$

si tratta di **stime puntuali** ottenute per ML, ovvero valori specifici delle statistiche calcolati in corrispondenza dei dati del campione (il procedimento è anche definito **inferenza statistica**).

Se N ridotto sono tuttavia poco accurate. Infatti si ha, per l'incertezza relativa sullo stimatore:

$$\frac{\Delta\sigma}{\sigma} = \frac{1}{\sqrt{2(N-1)}}$$

Si utilizzano quindi **stimatori per intervallo**, ovvero intervalli costruiti *intorno* allo stimatore puntuale (dunque centrati in esso) la cui ampiezza determina un valore noto e fisso della probabilità che il parametro stimato appartenga all'intervallo. Di consueto la probabilità che **non** appartenga è α , dunque quella che vi ricada è $1 - \alpha$.

Dimostrazione: incertezza su σ

Per un generico campione l'incertezza relativa nella stima della varianza vera σ^2 risulta uguale a quella nella sommatoria

$$\mathcal{S} = \sum_{i=1}^N (x_i - \bar{x})^2$$

(il fattore $N - 1$ si elide). Risulta ora che

$$\frac{\Delta\sigma^2}{\sigma^2} \approx \frac{2\sigma\Delta\sigma}{\sigma^2} = 2\frac{\Delta\sigma}{\sigma} \implies \frac{\Delta\sigma}{\sigma} = \frac{1}{2} \frac{\Delta\sigma^2}{\sigma^2}$$

Ora, applicando la formula nota per la SD generica di una grandezza:

$$\Delta\mathcal{S} = \sqrt{\mathcal{S}^2 - \bar{\mathcal{S}}^2}$$

ora, però, si può verificare che la media delle sommatorie dei quadrati di più campioni con medesimo numero di misure N è data secondo:

$$\bar{\mathcal{S}} = (N - 1)\sigma^2$$

ora, per un singolo campione la miglior stima della varianza è quindi data da $\sigma^2 = \frac{\mathcal{S}}{N - 1}$. Sostituendo quindi l'espressione nota si ha:

$$\mathcal{S} = (N - 1)\sigma^2 = (N - 1) \cdot \frac{N}{N - 1} \cdot (\overline{x^2} - \bar{x}^2) = \sum_i x_i^2 - \frac{1}{N} \left(\sum_i x_i \right)^2$$

sviluppando ora il secondo termine

$$\left(\sum_i x_i \right)^2 = \sum_i x_i^2 + \sum_i \sum_{j \neq i} x_i x_j$$

da cui

$$\mathcal{S} = \sum_i x_i^2 - \frac{1}{N} \sum_i x_i^2 - \frac{1}{N} \sum_i \sum_{j \neq i} x_i x_j = \left(1 - \frac{1}{N}\right) \sum_i x_i^2 - \frac{1}{N} \sum_i \sum_{j \neq i} x_i x_j$$

il cui quadrato è dato da

$$\mathcal{S}^2 = \underbrace{\left(\frac{N-1}{N}\right)^2 \sum_i x_i^2 \sum_j x_j^2}_A - 2 \underbrace{\frac{N-1}{N} \sum_i x_i^2 \sum_j \sum_{k \neq j} x_j x_k}_B + \underbrace{\frac{1}{N} \sum_j \sum_{j \neq k} x_j x_k \sum_m \sum_{n \neq m} x_m x_n}_C$$

Si assuma $\bar{x} = 0$ (il caso alternativo vi si riconduce considerando la differenza). Allora per i tre termini:

•

$$A = (N-1)^2 \frac{\sum_i x_i^4 + \sum_i \sum_{j \neq i} x_i^2 x_j^2}{N^2}$$

ora, la media dei termini di medesimo indice, ovvero esponente 4, è \bar{x}^4 ; per gli $N(N-1)$ termini con indice differente si ha invece \bar{x}^2 . Ora, per $\bar{x} = 0$ si ha $\bar{x}^2 = \sigma^2$ (secondo l'espressione alternativa per la varianza).

Integrando si dimostra che $\bar{x}^4 = 3\sigma^4$ (vedasi in fondo al box). La media di A è quindi

$$\bar{A} = \frac{(N-1)^2}{N^2} (3N\sigma^4 + N(N-1)\sigma^4) = \frac{(N-1)^2(N+2)}{N} \sigma^4$$

- ciascun termine nella sommatoria di B contiene una potenza dispari di ogni x_i , ovvero $x_i^1 = x_i$ oppure x_i^3 (in quanto nella seconda sommatoria i valori hanno indici differenti). Essendo x distribuita normalmente intorno alla media 0, risulta chiaramente che la media di potenze dispari sia nulla (contributi opposti si elidono):

$$\bar{B} = 0$$

- la sommatoria in C contiene $N(N-1)$ termini in cui gli indici coincidono a coppie, ovvero $j = m$ e $k = n$. Per ognuno di tali termini la media è quindi $\bar{x}^2 = \sigma^4$. Si ha poi un egual numero di termini ove gli indici coincidenti sono scambiati, ovvero $j = n$ e $k = m$; anche per questi la media è σ^4 . I termini restanti contengono potenze dispari: hanno quindi media nulla. Si ottiene quindi

$$\bar{C} = \frac{2N(N-1)}{N^2} \sigma^4 = \frac{2(N-1)}{N} \sigma^4$$

sommando si ottiene

$$\overline{\mathcal{S}^2} = \frac{(N-1)^2(N+2) + 2(N-1)}{N} \sigma^4 = \frac{N^3 + N - 2N^2 + 2N^2 + 2 - 4N + 2N - 2}{N} \sigma^4 = \frac{N^3 - N}{N} \sigma^4 = (N^2 - 1) \sigma^4$$

da cui

$$\Delta \mathcal{S} = \sqrt{(N^2 - 1) \sigma^4 - (N-1)^2 \sigma^4} = \sqrt{2N - 2} \sigma^2$$

prendendo ora l'incertezza relativa, il termine σ^2 si semplifica:

$$\frac{\Delta \mathcal{S}}{\mathcal{S}} = \frac{\sqrt{2(N-1)}}{N-1} = \frac{\sqrt{2}}{\sqrt{N-1}} = \frac{\Delta \sigma^2}{\sigma^2}$$

da cui

$$\frac{\Delta \sigma}{\sigma} = \frac{1}{2} \frac{\Delta \sigma^2}{\sigma^2} = \frac{1}{\sqrt{2(N-1)}}$$

Si dimostra in ultimo l'espressione per \bar{x}^4 .

$$\mathbb{E}[x^4] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2\sigma^2}} dx$$

Integrando per parti

$$\int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2\sigma^2}} dx = -\sigma^2 \int_{-\infty}^{\infty} x^3 \cdot \left(\frac{-x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}\right) dx = -\underbrace{\sigma^2 x^3 e^{-\frac{x^2}{2\sigma^2}}}_{0} \Big|_{-\infty}^{\infty} + 3\sigma^3 \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2\sigma^2}} dx$$

reintroducendo il termine di normalizzazione:

$$\overline{x^4} = 3\sigma^2 \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx = 3\sigma^2 \mathbb{E}[x^2] = 3\sigma^2 \cdot \sigma^2 = 3\sigma^4$$

8.1 Confidence level

Si considera ad esempio lo stimatore di μ . Si definisce l'**intervallo di confidenza**

$$[\hat{\mu} \pm k\hat{\sigma}]$$

ove k è definito **fattore di copertura**. La probabilità che il **valore vero** del parametro vi ricada

$$P(\hat{\mu} - k\hat{\sigma} < \mu < \hat{\mu} + k\hat{\sigma})$$

è definita **livello di confidenza** o **confidence level**.

8.2 t di Student

Per quantificare il livello di confidenza a seconda del fattore di copertura scelto, si definisce una nuova variabile casuale.

Essa è la **t di Student**:

$$t = \frac{\hat{\mu} - \mu}{\hat{\sigma}}$$

Dunque il confidence level per un dato fattore di copertura k , ovvero un intervallo di confidenza di raggio $k\hat{\sigma}$ è

$$P(|t| < k)$$

si noti il valore assoluto: chiaramente il fattore di copertura è DP e il test si svolge a due code (la distribuzione di Student è simmetrica).

La probabilità si calcola dalla **distribuzione di Student**

$$S_\nu(t) = \frac{c_\nu}{\left(1 + \frac{t^2}{\nu}\right)^{(\nu+1)/2}}$$

che è parametrizzata dal **numero di gradi di libertà**

$$\nu = n - 1$$

ove n è il numero di dati del campione. Viene sottratto un dof in quanto un dato è vincolato in seguito all'ottenimento della stima puntuale di μ . Si ha dunque

$$P(|t| < k) = P(\hat{\mu} - k\hat{\sigma} < \mu < \hat{\mu} + k\hat{\sigma}) = \int_{-k}^k S_\nu(t) dt = 2 \int_0^k S_\nu(t) dt$$

8.2.1 Proprietà della distribuzione di Student

1. È simmetrica rispetto al valor medio 0
2. **Varia a seconda del valore di ν :**
3. La sua varianza è pari a $\frac{\nu}{\nu - 2}$
4. È normalizzata a 1
5. Per $n \rightarrow \infty$ e dunque $\nu \rightarrow \infty$ **tende alla distr. normale standard** $G(z)$, con media 0 e varianza 1

Per n ridotti, invece, risulta più **bassa, schiacciata** della gaussiana: dunque gli integrali sono significativamente differenti, e di conseguenza le probabilità ottenibili. Un basso numero di dati, infatti, comporta una minore fiducia nella stima puntuale e dunque una probabilità meno elevata per $k \rightarrow 0$ e più alta per valori alti.

Se non si richiede eccessiva accuratezza, già per $n \sim 30$ è possibile approssimare la Student con la gaussiana standard, che non ha dipendenza da ν e quindi da n .

Applicazioni di Student

per campioni ridotti (!) si utilizza la distribuzione di St. nel

1. Confronto di una misura **con il valore noto/accettato** di una grandezza
2. Confronto **tra due misure** differenti della medesima grandezza

Entrambi costituiscono casi pertinenti al campo dei **test di ipotesi**, in quanto si valuta l'accettazione o il rigetto di un'ipotesi di compatibilità su basi probabilistiche.

8.3 Test di ipotesi

Il test delle ipotesi consente di **verificare (e quantificare) la compatibilità di un'ipotesi con l'evidenza empirica**.

La determinazione del **livello di accordo o disaccordo** si riconduce a quella di ottenere un **disaccordo maggiore** con i dati dell'ipotesi considerata, ovvero della *probabilità di fare peggio*.

Si effettua quindi un **test a due code**, fissando delle **soglie di probabilità**:

- Se la probabilità di maggiore disaccordo è **alta**, l'ipotesi è accettata: risulta compatibile (in accordo) con i dati e la discrepanza è giustificabile con fluttuazioni casuali
- Se essa è **bassa**, l'ipotesi è rigettata, in quanto la discrepanza può essere imputata (anche) alla sua falsità. Si definiscono, come di consueto, due soglie di significatività:
 1. Discrepanza significativa
 2. D. *altamente* significativa

8.3.1 Confronto misura con valore accettato

Il valore t_{mis} da considerare nell'applicazione della distribuzione di Student si calcola utilizzando come valor vero μ il **valore accettato della grandezza**. Se \bar{x} è la miglior stima dal campione (la misura ottenuta) e $\sigma_{\bar{x}}$ la SDOM del campione (*non campionaria!*):

$$t_{mis} = \frac{|\bar{x} - \mu|}{\sigma_{\bar{x}}}$$

si valuta dunque l'ipotesi di consistenza sul valore di

$$P(|t| > t_{mis}) = 1 - \int_{-t_{mis}}^{t_{mis}} S_{\nu}(t) dt = 1 - 2 \int_0^{t_{mis}} S_{\nu}(t) dt$$

ove

$$\nu = N - 1$$

con N = numero di misure del campione.

SDOM

La scelta della SDOM $\sigma_{\bar{x}}$ è giustificata dal fatto che si sta considerando la popolazione limite *delle medie*, da cui \bar{x} è estratta, che si suppone centrata in μ con SD corrispondente alla SDOM del campione di partenza da cui si è ottenuta la media.

Si utilizza Student per $N < 40$; per $N \geq 40$ in virtù di quanto detto è possibile utilizzare la gaussiana standard $G(z)$ cui si riconduce secondo

$$z_{mis} = \frac{|\bar{x} - \mu|}{\sigma_{\bar{x}}}$$

e quindi

$$P(|z| > z_{mis}) = 1 - 2 \int_0^{z_{mis}} G(z) dz$$

Si valuta quindi, secondo il consueto criterio, la significatività della discrepanza e la conseguente accettazione (o il rigetto) dell'ipotesi di consistenza:

- D. significativa: $P < 0.05$
- D. *altamente* significativa: $P < 0.01$

8.3.2 Confronto di due misure

Date due misure con medie e SDOM

$$\bar{x}_1 \pm \sigma_{\bar{x}_1} \quad \bar{x}_2 \pm \sigma_{\bar{x}_2}$$

si riconduce al caso precedente, in quanto si testa l'**ipotesi di compatibilità della differenza tra le due misure con 0**:

$$t_{mis} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}} \quad \vee \quad z_{mis} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}}$$

(t o z a seconda del numero di dati).

Se $n_1, n_2 \geq 30$, ovvero è possibile utilizzare z , allora la stima della SD della **popolazione delle differenze delle medie** è:

$$\hat{\sigma} = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2}$$

ove $\hat{\sigma}_i$ sono le SD **campionarie dei due campioni da cui sono ottenute le misure**.

Dimostrazione. Si è visto che la c.l. di variabili gaussiane è a sua volta gaussiana; dunque se ottenute da campioni gaussiani le medie \bar{x}_1, \bar{x}_2 sono a loro volta gaussiane e quindi lo è la loro differenza. Essendo assunte indipendenti, si somma in quadratura:

$$\sigma_{\bar{x}_1 - \bar{x}_2}^2 = 1^2 \sigma_{\bar{x}_1}^2 + (-1)^2 \sigma_{\bar{x}_2}^2 = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2$$

■

Il caso di campioni ridotti è trattabile facilmente (e si può dunque ottenere un'espressione per $\hat{\sigma}$ **solo se le SD delle due popolazioni gaussiane da cui sono estratte le misure sono uguali**).

Indicando con $\hat{\sigma}_i$ le SD campionarie (stime della SD della popolazione) si ha

$$t_{mis} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{n_1 - 1}{n_1 + n_2 - 2} \hat{\sigma}_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} \hat{\sigma}_2^2\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

in cui la grandezza

$$\sigma_p^2 = \frac{n_1 - 1}{n_1 + n_2 - 2} \hat{\sigma}_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} \hat{\sigma}_2^2$$

rappresenta una **media pesata** delle due stime (SD campionarie) della uguale SD delle rispettive popolazioni: essa è anche definita **varianza aggregata (o pooled variance)**. Si osserva che non viene assunto che le popolazioni abbiano anche la stessa media, e dunque coincidano: si assume solo l'uguaglianza del parametro di dispersione.

La t_{mis} così ottenuta segue una $S_\nu(t)$ con

$$\nu = n_1 + n_2 - 2$$

8.4 Rigetto per campioni ridotti

É possibile estendere il criterio di Chauvenet riportando a Student tramite

$$t_s = \frac{x_{sospetto} - \bar{x}}{\hat{\sigma}}$$

ove $\hat{\sigma}$ è la SD **campionaria**. Si esegue poi il consueto test a due code

$$\alpha = P(|t| > |t_s|) = 1 - 2 \int_0^{|t_s|} S_\nu(t) dt$$

8.5 Test di ipotesi: nulla vs alternativa

Si intende valutare due ipotesi:

- L'ipotesi nulla H_0
- L'ipotesi alternativa H_1

Si considera lo spazio dei campioni S . Per un qualsiasi campione di misure

$$\{x_1, \dots, x_N\} \in S$$

Le due ipotesi saturano S , definendovi due **regioni**:

- S_0 = regione **di accettazione** (per H_0) = insieme dei campioni compatibili con l'ipotesi nulla
- S_1 = regione **critica** = insieme dei campioni compatibili con l'ipotesi alternativa

Sulla base delle osservazioni, si prende dunque una **decisione** riguardo l'accettazione o il rigetto di una o dell'altra ipotesi. Nel farlo è possibile incorrere in due tipi di errore, con rispettive probabilità:

1. Errore di **tipo I**: rigetto di H_0 quando questa è vera. La corrispondente probabilità è α (falso negativo)
2. Errore di **tipo II**: accettazione di H_0 quando questa è falsa. La corrispondente probabilità è β (falso positivo)

In genere $\beta > \alpha$. Le altre probabilità sono quindi:

- Accettazione di H_0 quando questa è vera: $1 - \alpha$
- Rigetto di H_0 quando questa è falsa: $1 - \beta$

Dunque

$$\alpha = P(\text{campione} \in S_0 | H_1) \quad \beta = P(\text{campione} \in S_1 | H_0)$$

L'errore di tipo I è il **più grave**, e dunque si mira a minimizzarlo.

Si definiscono

Livello di significatività del test = α = probabilità valori oltre **valore critico (cut, decision boundary)**. Si osserva che a seconda che il test sia **unilaterale** o **bilaterale** α può riferirsi ad una o due code

Potenza del test = $(1 - \beta)$ = efficacia del test nel valutare l'ipotesi alternativa (se bassa più arduo definire compatibilità o incompatibilità di H_1 con i dati).

Se le due ipotesi determinano due distribuzioni $g(x|H_0)$ e $g(x|H_1)$ allora **per una coda**:

$$\alpha = \int_{x_{cut}}^{\infty} g(x|H_0) dx \quad \beta = \int_{-\infty}^{x_{cut}} g(x|H_1) dx$$

E dunque la potenza sarà

$$1 - \beta = \int_{x_{cut}}^{+\infty} g(x|H_1) dx$$

8.5.1 Test parametrici

Si considerano ipotesi su di un parametro campionario, H_0 e H_1 definiscono due distribuzioni normalizzate (gaussiane): il test consiste quindi nel valutare la relazione tra le probabilità attese secondo ciascuna ipotesi ed i valori estratti dai campioni. In realtà per i test parametrici è necessario verificare a priori ipotesi sulla distribuzione della variabile. Per test **non parametrici** tali ipotesi non sono necessarie; tuttavia presentano una **potenza minore** e dunque a parità di numerosità campionaria sono meno efficaci nell'evidenziare differenze di minore entità.

Ad esempio il **test di Kolmogorov-Smirnov** è proprio quello applicato per valutare se la distribuzione ha forma Gaussiana.

In genere ipotesi che determinano specificamente una distribuzione $f(x)$ sono dette **semplici**; se è determinata solo la forma e si ha dunque almeno un parametro libero si parla di ipotesi **composte**.

La significatività viene fissata **a priori**.

Dai dati si determina quindi il **p-value**

p-value

Più piccolo valore di significatività α per cui i dati campionari permettono di rifiutare l'ipotesi nulla H_0 , ovvero la probabilità, assunta vera H_0 , di ottenere valori del parametro **meno compatibili** del dato osservato (fare peggio in una o due code, a seconda del tipo di test).

Chiaramente un minore p-value implica una maggiore evidenza **contro** H_0 . In generale il test ha un risultato **significativo** se la statistica test calcolata sui dati cade nella regione di rifiuto, ovvero in termini del p-value:

1. L'ipotesi nulla H_0 è **rigettata** se $p - value < \alpha$. In particolare:
2. Se $p - value < 0.05$ si rifiuta H_0 con un livello di significatività del 5%
3. Se $p - value < 0.01$ si rifiuta H_0 con un livello di significatività dell' 1%

Chiaramente, appurata la natura gaussiana delle distribuzioni, il calcolo delle probabilità avviene riportandosi alla variabile standard. La SD da utilizzare è quella associata al parametro: nel caso di ipotesi su un valor medio è **la SDOM!**

Si osserva che sotto l'assunzione di verità dell'ipotesi nulla, se questa ha la forma $H_0 : \mu = \mu_0$ (per un parametro) allora **la distribuzione seguita dal p-value è uniforme in $[0, 1]$** .

Nel caso di campioni ridotti si utilizza invece la Student: ad esempio per ipotesi sul valore di una grandezza:

$$t_{mis} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

con ipotesi

$$H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0$$

Dalle tabelle di Student, fissato α e $\nu = N - 1$ (con N dati nel campione) si ricava il valore limite a t.c. (se il test è a due code)

$$P(|t| > a) = 1 - 2 \int_0^a S_\nu(t) dt = \alpha$$

se quindi

$$|t_{mis}| > a$$

allora il p-value è inferiore alla significatività e si ha il rigetto dell'ipotesi nulla; viceversa questa è **corroborata** (per esprimersi popperianamente).

Un'ultima nota vi sono affinità e differenze tra il calcolo del fattore di copertura per un intervallo di confidenza ed un test di ipotesi di compatibilità di un fit.

Infatti per il fit viene **fissato il parametro** e si valuta il p-value in relazione alla significatività; viceversa per il l'intervallo di confidenza viene **fissato il confidence level e quindi α** e si ricava il corrispondente valore di k .

Funzioni di probabilità di variabili discrete

Vi sono fenomeni che richiedono la misura di flussi/densità di **oggetti (discreti)**. Dunque si trattano grandezze che assumono valori **interi**: se sono sempre soggette a sole fluttuazioni di origine stocastica, si tratta di **variabili casuali discrete**.

La trattazione è globalmente affine a quella per le continue, fatto salvo per alcuni accorgimenti, tra cui:

- Le funzioni di probabilità che ne descrivono la p. *non sono delle densità (PDF)*, a causa del carattere discreto.

Proprietà e parametri indicando con k la variabile discreta e k_i i vari valori possibili, le formule espresse per il continuo si traducono nel discreto secondo:

- Normalizzazione: $\sum_i P(k_i) = 1$
- Media $\bar{k} = \sum_i k_i P(k_i)$
- Varianza $\sigma_k^2 = \sum_i (k_i - \bar{k})^2 P(k_i) = \overline{k^2} - \bar{k}^2$

9.1 Distribuzione uniforme

(e.g. dado non truccato con $n = 6$)

$$P(k) = \frac{1}{n} \quad k = 1, \dots, n$$

Normalizzazione

$$\sum_{i=1}^n P(k_i) = n \cdot \frac{1}{n} = 1$$

Media

$$\bar{k} = \sum_{i=1}^n k_i P(k_i) = \frac{1}{n} \sum_{i=1}^n k_i$$

Se $k_i = i$:

$$\bar{k} = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$$

applicando la formula di Gauss.

Varianza

$$\overline{k^2} = \sum_{i=1}^n k_i^2 P(k_i) = \frac{1}{n} \sum_{i=1}^n k_i^2$$

Se $k_i = i$

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6} \implies \overline{k^2} = \frac{(n+1)(2n+1)}{6}$$

e dunque

$$\sigma_k^2 = \overline{k^2} - \bar{k}^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)(n+1)}{4} = (n+1) \frac{4n+2-3n-3}{12} = \frac{n^2-1}{12}$$

da cui

$$\sigma_k = \sqrt{\frac{n^2-1}{12}}$$

9.2 Calcolo combinatorio

Disposizioni

di k oggetti presi da un insieme di n (di n oggetti classe k)

Senza ripetizione solo se $k < n$

$$D_{n,k} = \frac{n!}{(n-k)!}$$

Con ripetizione no vincolo su k

$$D_{n,k}^* = n^k$$

Permutazioni di n oggetti

$$P_n = D_{n,n} = n!$$

Combinazioni

Modi di scegliere k oggetti tra n (ovvero di disporli senza però distinguere solo per l'ordine).

Senza ripetizione se ogni oggetto può venir preso una volta sola \rightarrow coeff binomiale:

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Con ripetizione se ciascun oggetto può essere preso fino a k volte:

$$C_{n,k}^* = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

9.3 Distribuzione binomiale

Si considera un processo casuale di tipo **bernoulliano**, ovvero con due soli possibili esiti mutualmente escludenti:

- 'Successo', con probabilità p
- 'Insuccesso' con probabilità $q = 1 - p$ (definite classicamente, saturano lo spazio degli eventi)

Se si ripete tale processo n volte, la probabilità di ottenere k successi ($0 \leq k \leq n$) è data dalla

Distribuzione di probabilità binomiale

$$P(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

essendo ogni tentativo **indipendente**, la probabilità di ottenere k successi (e dunque $n - k$ insuccessi) in una specifica scelta di tentativi è

$$p^k (1-p)^{n-k}$$

il termine binomiale è quindi aggiunto per considerare tutti gli eventi corrispondenti a differenti scelte di tali k tentativi (in cui chiaramente non conta l'ordine).

La binomiale ha dunque due parametri: n e p , in quanto q è determinato da p .

9.3.1 Proprietà

1. È normalizzata a 1

Dimostrazione.

$$\underbrace{\sum_{i=1}^n \binom{n}{k} p^k (1-p)^{n-k}}_{\text{binomio di Newton}} = (p + (1-p))^n = 1^n = 1$$

■

2. È simmetrica solo se $p = \frac{1}{2}$
3. Per $n \rightarrow \infty$ a p fissa **tende ad una gaussiana** con corrispondenti media e SD.
4. La media è

$$\bar{k} = np$$

Dimostrazione.

$$\bar{k} = \sum_{k=0}^n k P(k) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \frac{n!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}$$

ove l'indice dell'ultima sommatoria parte da 1 in quanto $0 \cdot P(0) = 0$. Quindi effettuando il cambiamento di variabile $m = n - 1$ e $h = k - 1$

$$\bar{k} = np \sum_{h=0}^m \binom{m}{h} p^h (1-p)^{m-h} = np$$

per la normalizzazione.

■

5. La varianza è data da

$$\sigma_k^2 = npq = np(1-p)$$

Dimostrazione.

$$\overline{k^2} = \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^n k \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} (1-p)^{n-k}$$

in cui il termine di indice 0 è eliso per il medesimo motivo esposto in precedenza. Con lo stesso cambio di variabili:

$$\overline{k^2} = np \sum_{h=0}^m (h+1) \binom{m}{h} p^h (1-p)^{m-h} = np \cdot \bar{h} + np \cdot 1 = np(mp+1) = np((n-1)p+1) = n^2 p^2 - np^2 + np$$

Inserendo nella formula per la varianza:

$$\sigma^2 = \overline{k^2} - \bar{k}^2 = n^2 p^2 - np^2 + np - n^2 p^2 = np - np^2 = np(1-p)$$

■

9.4 Efficienza di conteggio

Sia N il numero di particelle che attraversano un rivelatore e n quello di pp. effettivamente rivelate. Si definisce

Efficienza del rivelatore

$$\varepsilon = \frac{n}{N}$$

L'incertezza sull'efficienza, calcolata su più misure, è un errore di tipo **casuale**:

$$\Delta\varepsilon = \sigma_\varepsilon$$

Supponendo di mantenere costante N , la variabilità è data da n e dunque propagando in quadratura:

$$\sigma_\varepsilon^2 = \left(\frac{\partial\varepsilon}{\partial n}\right)^2 \sigma_n^2 = \left(\frac{1}{N}\right)^2 \sigma_n^2 \implies \sigma_\varepsilon = \frac{\sigma_n}{N}$$

che coincide con il risultato ottenuto per propagazione lineare (vale chiaramente sempre nel caso di singola misura con variabilità nel computo). Si osserva che l'incertezza su ε , supposta costante quella su n , è inversamente proporzionale al numero di particelle che attraversano il rivelatore: chiaramente maggiore è il flusso di pp. maggiore è la precisione con cui si può apprezzare l'efficienza dello strumento.

Ora, il conteggio rappresenta un processo bernoulliano: una particella viene rivelata oppure no. Dunque n è una variabile **di tipo binomiale**. I parametri della distribuzione $P(n)$ saranno chiaramente $p = \varepsilon$ (per definizione di efficienza) e $n_{tot} = N$. Quindi

$$\sigma_n = \sqrt{N\varepsilon(1-\varepsilon)}$$

da cui

Errore sull'efficienza (per processo binomiale)

$$\sigma_\varepsilon = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N}}$$

Il contributo di ε dipende chiaramente dalle condizioni ambientali e dall'apparato sperimentale; a parità di tali fattori può essere ridotta l'incertezza incrementando il numero di **particelle primarie** N .

Supponendo ora n come valore dato (costante), si determina il numero "vero" di particelle che hanno attraversato il rivelatore secondo

$$M = \frac{n}{\varepsilon}$$

la variabilità è quindi data da ε :

$$\sigma_M = \left| \frac{\partial M}{\partial \varepsilon} \right| \sigma_\varepsilon = \frac{n}{\varepsilon^2} \sigma_\varepsilon = \frac{M}{\varepsilon} \sigma_\varepsilon$$

9.5 Distribuzione geometrica

Si ha sempre un processo elementare di tipo bernoulliano; tuttavia k indica il numero di tentativi *necessari a ottenere il primo successo* (ovvero il primo tentativo, in ordine, per cui si verifica il successo).

Distribuzione geometrica

$$P(k) = p(1-p)^{k-1}$$

con $k \in [1, \infty[$

Chiaramente essendo $p < 1$ la probabilità decresce all'aumentare di k (sempre più improbabile ottenere tanti insuccessi di fila): è quindi massima per $k = 1$.

$$\lim_{k \rightarrow \infty} P(k) = 0$$

9.5.1 Proprietà

- É normalizzata a 1

Dimostrazione.

$$\sum_{k=1}^{+\infty} P(k) = p \sum_{k=1}^{+\infty} (1-p)^{k-1}$$

Con il cambio di variabile $h = k - 1$ si ottiene la serie geometrica di ragione $1 - p < 1$, che dunque converge

$$\sum_{h=0}^{+\infty} (1-p)^h \rightarrow \frac{1}{1-(1-p)} \implies \sum_{k=0}^{+\infty} P(k) = p \frac{1}{p} = 1$$

■

- La media è data secondo

$$\bar{k} = \frac{1}{p}$$

Dimostrazione.

$$\bar{k} = p \sum_{k=1}^{+\infty} k(1-p)^{k-1} = p \sum_{k=1}^{+\infty} -\frac{d}{dp} [(1-p)^k] = -p \frac{d}{dp} \left[\sum_{h=0}^{+\infty} (1-p)^h \right] = -p \frac{d}{dp} \left(\frac{1}{1-(1-p)} \right) = p \frac{1}{p^2} = \frac{1}{p}$$

(la derivata di costante è nulla).

■

- Per la varianza

$$\sigma_k = \frac{\sqrt{1-p}}{p}$$

Dimostrazione.

$$\overline{k^2} = \sum_{k=1}^{+\infty} k^2 P(k) = p \sum_{k=1}^{\infty} k^2 (1-p)^{k-1}$$

ora, si osserva che

$$k^2 = k^2 - k + k = k(k-1) + k$$

e si distribuisce:

$$\overline{k^2} = p(1-p) \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-2} + \underbrace{\sum_{k=1}^{\infty} kp(1-p)^{k-1}}_{\bar{k}}$$

Per il primo termine si ha invece

$$\sum_{k=1}^{\infty} k(k-1)(1-p)^{k-2} = \sum_{k=1}^{\infty} \frac{d^2}{dp^2} (1-p)^k = \frac{d^2}{dp^2} \sum_{k=0}^{\infty} (1-p)^k = \frac{d^2}{dp^2} \left(\frac{1}{p} \right) = \frac{2}{p^3}$$

e quindi

$$\overline{k^2} = p(1-p) \frac{2}{p^3} + \frac{1}{p} = (1-p) \frac{2}{p^2} + \frac{1}{p} \implies \sigma^2 = \frac{2(1-p) - 1}{p^2} + \frac{1}{p} = \frac{1-2p+p}{p^2} = \frac{1-p}{p^2}$$

■

9.5.2 Assenza di memoria

La distribuzione geometrica è l'**unica** distribuzione discreta caratterizzata da **assenza di memoria**: ovvero descrive fenomeni *la cui evoluzione futura è indipendente dalla storia passata*. Ciò significa che la probabilità di ottenere il successo in $k+m$ tentativi non è influenzata dalla conoscenza del fatto che i primi m tentativi abbiano dato insuccessi.

(Chiaramente ciò non vale per la binomiale: la probabilità di ottenere k successi su n tentativi cambia se è noto quanti se ne sono ottenuti nei primi $m < n$! Basti pensare a $k = n$...)

Si verifica l'assenza di memoria tramite la **probabilità condizionata**, dunque applicando Bayes:

$$P(k+m \mid m \text{ primi insuccessi}) = \frac{P[(k+m) \cap (m \text{ primi insuccessi})]}{P(m \text{ primi insuccessi})}$$

ovvero

$$P = \frac{p(1-p)^{k+m-1}}{(1-p)^m} = p(1-p)^{k-1} = P(k)$$

Nel continuo l'esempio di assenza di memoria è dato dall'analogo della geometrica, ovvero dalla **distribuzione esponenziale**.

Si dimostra che la probabilità una particella decada dopo l'istante t_0 è uguale alla probabilità che decada dopo $t_1 + t_0$ noto che a t_1 non è ancora decaduta.

Per un generico t_x , la probabilità che una particella decada per tempi maggiori (ovvero non sia ancora decaduta) è:

$$P(t > t_x) = \int_{t_x}^{\infty} \frac{1}{\tau} e^{-\frac{t}{\tau}} dt = [-e^{-t/\tau}]_{t_x}^{\infty} = e^{-t_x/\tau}$$

dunque

$$P(t > (t_1 + t_0) \mid \text{non decaduta a } t_1) = \frac{P(\text{decada dopo } t_1 + t_0)}{P(\text{non decaduta a } t_1)} = \frac{e^{-\frac{t_1 + t_0}{\tau}}}{e^{-\frac{t_1}{\tau}}} = e^{-\frac{t_0}{\tau}} = P(\text{decada dopo } t_0)$$

9.6 Distribuzione di Poisson

Si considera un intervallo ΔX di una certa variabile e n oggetti o eventi distribuiti casualmente al suo interno. Considerando ora un sottointervallo Δx , la probabilità di trovarvi uno qualsiasi degli oggetti è data secondo

$$p = \frac{\Delta x}{\Delta X} \ll 1$$

Per n sufficientemente elevato si può dunque definire la **densità media** per unità di X

$$\lambda = \frac{n}{\Delta X}$$

da cui il numero medio di oggetti in Δx secondo:

$$\bar{k} = \underbrace{n \cdot p}_{\text{binom (c'è/non)}} = \lambda \Delta X \frac{\Delta x}{\Delta X} = \lambda \Delta x = \mu$$

dunque è possibile effettuare uno studio senza che siano necessariamente noti n oppure la probabilità di trovare un oggetto nell'intervallo: bastano λ e Δx oppure μ .

La probabilità del verificarsi di k eventi all'interno di un dato sottointervallo Δx , noto che il numero medio per questo è μ e sotto l'assunto di indipendenza tra gli eventi, è data dalla

Distribuzione di Poisson

$$P(k; \mu) = \frac{\mu^k e^{-\mu}}{k!}$$

con $0 \leq k < \infty$

La poissoniana è un **caso limite di una binomiale**, quando il numero totale n è elevato e la probabilità di successo p è piccola (ovvero il rapporto tra il sottointervallo Δx e l'intervallo complessivo della grandezza ΔX è ridotto):

$$P(k; n, p) = \binom{n}{k} p^k q^{n-k} \xrightarrow[p \rightarrow 0]{n \rightarrow \infty} P(k; \mu) = \frac{\mu^k e^{-\mu}}{k!} \quad \text{con } \mu = np$$

Al crescere del parametro di posizione μ si ha quindi che la Poissoniana si sposta verso k maggiori, allargandosi e tendendo a diventare simmetrica: per $\mu \rightarrow \infty$ essa tende quindi a una **gaussiana** di media e varianza

$$\mu_G = \mu \quad \sigma_G = \sqrt{\mu}$$

il livello di approssimazione può allora essere quantificato in percentuale da $(\frac{x}{y} - 1)\%$ ove $x(k)$ è la probabilità calcolata secondo la gaussiana, $y(k)$ secondo la poissoniana.

Dimostrazione: Poissoniana = limite binomiale*Dimostrazione.*

$$P(k; n, p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Si sostituisce $\mu = np \Rightarrow p = \frac{\mu}{n}$

$$\begin{aligned} P(k; n, \mu) &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k} = \\ &= \frac{\mu^k}{k!} \underbrace{\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}}_{\rightarrow 1 \text{ per } n \rightarrow \infty} \underbrace{\left(1 - \frac{\mu}{n}\right)^{-k}}_{\rightarrow 1} \underbrace{\left(1 - \frac{\mu}{n}\right)^n}_{\rightarrow e^{-\mu}} \rightarrow \frac{\mu^k e^{-\mu}}{k!} \end{aligned}$$

■

9.6.1 Proprietà della poissoniana**Normalizzazione**

$$\sum_{k=0}^{\infty} P(k; \mu) = e^{-\mu} \underbrace{\sum_{k=0}^{\infty} \frac{\mu^k}{k!}}_{\text{serie exp}} = e^{-\mu} e^{\mu} = 1$$

Media Si ha

$$\bar{k} = \mu$$

Dimostrazione.

$$\bar{k} = \sum_{k=0}^{\infty} k P(k; \mu) = 0 + \sum_{k=1}^{\infty} \frac{\mu^{k-1} \mu e^{-\mu}}{(k-1)!} = \mu e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!}$$

operando il cambio di variabile $h = k - 1$:

$$\bar{k} = \mu e^{-\mu} \sum_{h=0}^{\infty} \frac{\mu^h}{h!} = \mu e^{-\mu} e^{\mu} = \mu$$

■

Varianza Si ha

$$\sigma_k^2 = \mu$$

Dimostrazione.

$$\overline{k^2} = 0 + \sum_{k=1}^{\infty} k \frac{\mu^k e^{-\mu}}{(k-1)!} = \mu \sum_{k=1}^{\infty} k \frac{\mu^{k-1} e^{-\mu}}{(k-1)!}$$

Operando il cambio di variabile $h = k - 1$

$$\overline{k^2} = \mu \sum_{h=0}^{\infty} (h+1) \frac{\mu^h e^{-\mu}}{h!} = \mu \underbrace{\sum_{h=0}^{\infty} h \frac{\mu^h e^{-\mu}}{h!}}_{\bar{h}=\mu} + \mu \underbrace{\sum_{h=0}^{\infty} \frac{\mu^h e^{-\mu}}{h!}}_1 = \mu^2 + \mu$$

Da cui

$$\sigma^2 = \overline{k^2} - \bar{k}^2 = \mu^2 + \mu - \mu^2 = \mu$$

■

Si perviene al medesimo risultato passando al limite della varianza binomiale:

$$\sigma^2 = np(1-p) = np - np^2 \xrightarrow[n \rightarrow +\infty]{p \rightarrow 0} \mu - 0 \cdot \mu = \mu$$

9.6.2 Singola misura di conteggio

Quando si effettua una singola misura m di una variabile poissoniana:

- La miglior stima di μ è la misura stessa

$$\hat{\mu} = m$$

- L'incertezza da associare è quindi la radice della misura

$$\Delta m = \sqrt{m}$$

- Segue che l'errore relativo è

$$\frac{\Delta m}{m} = \frac{1}{\sqrt{m}}$$

dunque decresce maggiore è il valore di m

Si dimostra quanto affermato per ML

Dimostrazione. Si supponga dato un campione di misure $\{m_1, \dots, m_N\}$ estratte da popolazione poissoniana. La probabilità congiunta, sotto l'assunzione di indipendenza, è

$$P\left(\bigcap_i m_i\right) = \prod_{i=1}^N \frac{\mu^{m_i} e^{-\mu}}{m_i!} = \frac{\mu^{\sum_i m_i} e^{-N\mu}}{\prod_i m_i!}$$

Si ottiene ora la funzione di likelihood prendendo il logaritmo

$$\mathcal{L} = \ln P = \sum_{i=1}^N m_i \ln \mu - N\mu - \sum_{i=1}^N \ln m_i!$$

annullando quindi la derivata rispetto al parametro:

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{\sum_i m_i}{\mu} - N = 0 \implies \hat{\mu} = \frac{\sum_i m_i}{N} = \bar{m}$$

■

9.7 Eliminazione del rumore di fondo

Si considera la misura di un tasso di decadimento.

Si misura in un tempo T_{tot} un numero totale di eventi (segnale + fondo) N_{tot} , ottenendo quindi un tasso complessivo

$$R_{tot} = \frac{N_{tot}}{T_{tot}}$$

si misura poi separatamente, per un tempo T_F , esclusivamente il fondo. Se N_F sono gli eventi osservati, il tasso di solo fondo è

$$R_F = \frac{N_F}{T_F}$$

Dunque la misura del tasso di segnale è data secondo:

$$R_S = R_{tot} - R_F$$

Considerando le incertezze sui tempi trascurabili, la variabilità dipende dal numeratore, ovvero i conteggi:

$$\Delta R_F^2 = \left(\frac{\partial R_F}{\partial N_F}\right)^2 \Delta N_F^2 \implies \Delta R_F = \frac{\partial R_F}{\partial N_F} \Delta N_F = \frac{\Delta N_F}{T_F} \quad \text{analogamente} \quad \Delta R_{tot} = \frac{\Delta N_{tot}}{T_{tot}}$$

(la propagazione lineare coincide con la quadratura). Segue che

$$\Delta R_S = \sqrt{\left(\frac{\partial R_S}{\partial N_{tot}}\right)^2 \Delta N_{tot}^2 + \left(\frac{\partial R_S}{\partial N_F}\right)^2 \Delta N_F^2} = \sqrt{\left(\frac{\Delta N_{tot}}{T_{tot}}\right)^2 + \left(\frac{\Delta N_F}{T_F}\right)^2}$$

Capitolo 10

Chi quadro

Il test del chi quadro permette la verifica della consistenza di un'ipotesi teorica con un insieme di dati sperimentali, fornendo un criterio di valutazione su basi probabilistiche (dunque quantificando il livello di compatibilità).

Vi sono in particolare due casi di applicazione:

1. Confronto tra **un campione di dati** di una variabile casuale ed una **distribuzione di probabilità** ipotetica
2. Confronto tra **un insieme di coppie di valori** misurati (x_i, y_i) ed un'ipotetica **relazione funzionale** $y = g(x)$

10.1 Confronto dati - distribuzione

Si suddivide il campione in esame in **bin** e per ogni bin si calcola:

- Il numero di eventi osservati (*observed*) O_i
- La probabilità che una misura ricada nel bin, data secondo

$$P_i = \int_{x_i - \Delta x/2}^{x_i + \Delta x/2} \Phi(x) dx$$

ove x_i è il valore mediano del bin e Δx l'ampiezza (assunto binnaggio uniforme).

- Il numero di eventi attesi (*expected*)

$$E_i = N \cdot P_i$$

ove N è il numero totale **di misure del campione**

Se l'ipotesi di distribuzione seguita $\Phi(x)$ è **corretta**, ovvero Φ descrive correttamente l'andamento della distribuzione limite del campione nell'ipotesi di ripetere le N misure **infinite volte**, allora per ogni bin E_i è **la media degli** O_i , ovvero il valore medio di eventi osservati nel bin ripetendo, appunto, la misura infinite volte.

Si definisce quindi per ogni bin il **chi quadro**

$$\chi_i^2 = \frac{(O_i - E_i)^2}{\sigma_i^2}$$

ove σ_i^2 è la varianza della distribuzione degli O_i con media E_i .

Si calcola quindi il Chi quadro complessivo:

$$\chi^2 = \sum_{i=1}^{N_{bins}} \left(\frac{O_i - E_i(\alpha_J)}{\sigma_i} \right)^2$$

ove gli $\{\alpha_J\}$ sono i parametri della Φ .

Tale funzione è **definita positiva**, in quanto i contributi per ogni bin sono elevati al quadrato per evitare l'annullamento reciproco di scostamenti di segno opposto. I suoi valori seguono una **pdf nota**.

Poiché nell'ipotesi fatta in precedenza O_i risulta una **variabile casuale discreta di conteggio**, la varianza associata è in generale quella della **binomiale**:

$$\sigma_i^2 = NP_i(1 - P_i) = E_i(1 - P_i)$$

nel caso poi di $N_{bins} \rightarrow \infty$ è possibile adottare la varianza di **poissoniana**:

$$\mu = NP_i = E_i \implies \sigma_i^2 = E_i$$

10.1.1 Stabilire i bin

Il binnaggio è da determinarsi avendo premura che siano soddisfatte le seguenti condizioni:

- L'ampiezza sia sempre \geq la risoluzione della variabile (dunque dello strumento in caso di misura diretta)
- Il numero di misure di ciascun bin (O_i) sia $\geq 4-5$
- Il numero di bin sia sufficiente perché $\nu > 0$ (vd. dopo), in quanto altrimenti il test **non avrebbe senso**

In tal modo si evita di creare *false distribuzioni* o viceversa di *celare la forma* di $\Phi(x)$

10.1.2 DoF

La distribuzione del Chi quadro è parametrizzata dal numero di gradi di libertà, definito secondo

$$\nu = N_{bins} - N_{vincoli}$$

ove $N_{vincoli}$ è il numero di termini della somma che definisce il χ^2 **non indipendenti**. Il loro numero si ottiene secondo:

1. Se si stimano i parametri α_J della Φ **dal campione**, ogni parametro corrisponde ad un vincolo
2. Se si calcola dai dati o impone **il numero totale di dati** N_{tot} , si ha un ulteriore vincolo

Di principio, in caso di ipotesi corretta sulla distribuzione, si attende

$$|O_i - E_i| \sim \sigma_i \implies (O_i - E_i)^2 \sim \sigma_i^2 \implies \chi^2 \approx N_{bins}$$

(viceversa per ipotesi errata $\chi^2 \gg N_{bins}$)

In realtà, **non** essendo tutti i termini della somma **indipendenti**, dunque è più corretto attendersi

$$\chi^2 \approx \nu$$

10.1.3 Chi ridotto

Si definisce quindi il **chi quadrato ridotto**

$$\tilde{\chi}^2 = \frac{\chi^2}{\nu}$$

dunque un'ipotesi corretta deve dare

$$\tilde{\chi}^2 \approx 1$$

Chiaramente un tale risultato descrive **esclusivamente la compatibilità tra la PDF ipotizzata e lo specifico campione considerato**, dunque non dà garanzie su eventuali bias nell'ottenimento di questo.

10.2 Distribuzione Chi quadro

Come detto, il χ^2 segue una specifica PDF (è variabile continua) parametrizzata su ν . Essa è espressa secondo:

$$\Phi_\nu(\chi^2) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} (\chi^2)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}} \quad \text{con la funzione Gamma} \quad \Gamma\left(\frac{\nu}{2}\right) = \int_0^\infty x^{\frac{\nu}{2}-1} e^{-x} dx$$

La probabilità è quindi valutata **a una coda**:

$$P(\chi^2 > \chi_0^2) = \int_{\chi_0^2}^\infty \Phi(\chi^2) d\chi^2$$

E analogamente per $\tilde{\chi}^2$ dai valori tabulati per la corrispondente distribuzione.

Si osserva che la $\Phi(\chi^2)$ tende a spostarsi verso destra all'aumentare di ν , tendendo secondo il **teorema del limite centrale** ad una gaussiana per $\nu \rightarrow \infty$.

10.2.1 Proprietà di $\Phi(\chi^2)$

Per la media e la varianza si ha:

$$\mathbb{E}[\chi^2] = \nu \quad \text{Var}[\chi^2] = 2\nu$$

Per il chi ridotto si ha invece

$$\mathbb{E}[\tilde{\chi}^2] = 1 \quad \text{Var}[\tilde{\chi}^2] = \frac{2}{\nu}$$

Normalizzazione della Φ

Per determinare il termine di normalizzazione, si ponga generalmente

$$\Phi(\chi^2) = \eta \cdot (\chi^2)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}}$$

integrando tra 0 e $+\infty$ si ha

$$\begin{aligned} \int_0^{+\infty} \Phi(\chi^2) d\chi^2 &= \eta \int_0^{+\infty} (\chi^2)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}} d\chi^2 = \eta \cdot 2^{\nu/2-1} \int_0^{+\infty} \left(\frac{\chi^2}{2}\right)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}} d\chi^2 = \\ &= \eta \cdot 2^{\nu/2-1} \cdot 2\Gamma\left(\frac{\nu}{2}\right) = \eta \cdot 2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right) \end{aligned}$$

da cui, imposto $\int \Phi = 1$:

$$\eta = \frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})}$$

10.3 Coppie vs relazione funzionale

Il chi è dato secondo

$$\chi^2 = \sum_{i=1}^N \left(\frac{y_i - g(x_i; \alpha_J)}{\sigma_i} \right)^2$$

ove N sono le coppie (x_i, y_i) totali.

Se l'ipotesi di relazione funzionale è corretta, ogni y_i è distribuita con media $g(x_i)$ e varianza σ_i^2 , che corrisponde **all'incertezza sui dati in y** .

Nota

Una sovrastima delle incertezze σ_i può condurre ad una sottostima del chi e quindi una conclusione errata sulla compatibilità con i dati: infatti se si *accetta* una discrepanza eccessivamente ampia, il test non ha particolare validità.

Nel caso di incertezze non uniformi, può essere espresso anche nella forma del **chi quadrato pesato**:

$$\chi^2 = \sum_{i=1}^N w_i (y_i - g(x_i; \alpha_J))^2 \quad \text{ove } w_i = \frac{1}{\sigma_i^2}$$

10.4 Soglie di probabilità

Le soglie sulla probabilità per determinare l'esito del test sono le consuete:

- $P \geq 5\%$ discrepanza non significativa: compatibile
- $P < 5\%$ discrepanza significativa: rigetto ipotesi di compatibilità
- $P < 1\%$ discrepanza altamente significativa: rigetto ipotesi di compatibilità

10.5 Minimo Chi quadro

Si tratta di un metodo alternativo alla ML per la determinazione dei parametri di una relazione funzionale ipotizzata (e.g. regressione lineare).

Lo scopo è determinare i parametri che minimizzano il χ^2 su una data serie di coppie di dati (per cui sono note anche le σ_i o i w_i); in tal modo si ottiene contemporaneamente anche il valore χ^2_{min} da utilizzare per valutare la probabilità e quindi l'ipotesi di compatibilità: se si ha il rigetto, chiaramente era errata l'ipotesi sulla forma della curva.

Nel caso di singolo parametro incognito si calcola il valore del χ^2 per alcuni valori nell'intorno del minimo e si interpola con una parabola convessa, di modo da determinarne il minimo.

In generale la condizione si traduce in:

$$\frac{\partial \chi^2}{\partial \alpha_J} = 0 \quad J = 1, \dots, N_{parametri}$$

Il metodo è applicabile **anche per determinare i parametri di una distribuzione ipotizzata** (caso 1).

Il minimo chi quadro si riconduce al metodo dei minimi quadrati se

- L'incertezza su x è trascurabile
- $\sigma_y \neq 0$ per ogni y_i
- σ_i è un'incertezza di tipo gaussiano

Nel qual caso la minimizzazione del chi quadro e della funzione di likelihood sono equivalenti:

$$\frac{\partial \mathcal{L}}{\partial \alpha_J} = 0 \quad \frac{\partial \chi^2}{\partial \alpha_J} = 0$$

Capitolo 1 1

Teorema del limite centrale

Si tratta di un risultato fondamentale in statistica. Esso enuncia quanto segue

Teorema del limite centrale

La somma di n variabili aleatorie (casuali) continue **indipendenti** distribuite in modo arbitrario (ma identico fra loro) e aventi valori di aspettazione (attesi) comparabili e varianza finita **tende ad una pdf gaussiana per n sufficientemente grande, indipendentemente dalla forma delle distribuzioni di partenza.**

Il risultato vale in realtà più generalmente per una **combinazione lineare** ove l'ordine di grandezza dei vari pesi sia il medesimo.

Se per ogni variabile x_i la media è μ_i e la varianza σ_i^2 , allora per la gaussiana ottenuta

$$\mu = \sum \mu_i \qquad \sigma^2 = \sum \sigma_i^2$$

o nel caso più generale di c.l. di pesi α_i

$$\mu = \sum \alpha_i \mu_i \qquad \sigma^2 = \sum \alpha_i^2 \sigma_i^2$$

Ad esempio per n variabili distribuite secondo $U(0, 1)$

$$\mu = 0.5 \cdot n \qquad \sigma^2 = n \left(\frac{1}{\sqrt{12}} \right)^2 = \frac{n}{12}$$

Capitolo 12

PRNG e Monte Carlo

Il metodo Monte Carlo fu sviluppato negli anni '40 nell'ambito del Progetto Manhattan da vari scienziati, tra cui Enrico Fermi ma soprattutto John Von Neumann, Stanislaw Ulam e Nicholas Metropolis (cui si deve il nome).

Il metodo permette di studiare sistemi per cui procedimenti analitici sono impossibili a livello pratico a causa della complessità computazionale eccessiva. Si basa sulla **generazione di numeri casuali**.

Numero casuale = valore assunto da una variabile aleatoria, per definizione imprevedibile a partire dalle occorrenze passate

Nel concreto è difficile ottenere numeri casuali **in grandi quantità**: per questioni di efficienza si fa quindi ricorso a sequenze prodotte da **generatori di numeri pseudo-casuali (Pseudo-Random Number Generators)**. Alla base dei PRNG vi sono **algoritmi matematici** (dunque perfettamente deterministici). La sequenza può infatti essere descritta matematicamente dall'equazione alle differenze:

$$x_{n+1} = f(x_n)$$

Tuttavia gli algoritmi usati, come indica il nome, hanno proprietà che permettono di ottenere sequenze con comportamento molto simile a numeri veramente casuali.

Ad esempio considerando la **mappa logistica**

$$x_{n+1} = \lambda \cdot x_n(1 - x_n)$$

per piccoli valori del parametro si ha il seguente **comportamento asintotico** con convergenza:

$$n \rightarrow \infty \quad x_\infty = \frac{\lambda - 1}{\lambda}$$

Invece per λ maggiori si ha convergenza verso **un insieme di valori oscillanti** e quindi una transizione ad un comportamento **caotico**.

I PRNG più comuni in genere producono

- Sequenze di **interi** distribuiti in modo **uniforme** in un intervallo $[0, N_{max}]$
- Sequenze di **reali** distribuiti **uniformemente** in $[0, 1]$

Indipendenza e uniformità sono verificate tramite **batterie di test**.

Gli algoritmi sono qualificati in termini di:

- **prestazioni** (e.g. velocità)
- **periodo** = numero di interazioni prima che siano ripercorsi nella sequenza stati già generati.

Una sequenza è inizializzata in base / a partire da **un seme (seed)**, ovvero un valore iniziale. Semi diversi producono sequenze diverse e **indipendenti**.

12.1 Altre distribuzioni

Per generare numeri secondo distribuzioni differenti dall'uniforme, vi sono due tipi di metodi adottabili:

1. **Metodo del rigetto (o *hit or miss*)** → funziona per ogni f ma solo su **dominio finito**; per la sua natura può inoltre risultare inefficiente
2. **Metodo della trasformazione inversa**

12.2 Metodo del rigetto

Si intende generare casualmente una variabile x distribuita secondo una specifica $f(x)$ su di un dominio finito $[a, b]$.

1. Si generano casualmente due valori:

$$\begin{cases} x = U(a, b) \\ y = U(0, \max) \end{cases} \quad \text{ove } \max = \max_{[a, b]} f(x)$$

2. Si definisce una funzione $h(x, y)$ secondo

$$h(x, y) = \begin{cases} 0 & \text{se } y > f(x) \\ 1 & \text{se } y < f(x) \end{cases}$$

3. Si applica il criterio *hit or miss*:

- Se $h(x, y) = 0$ l'estrazione viene **scartata**
- Se $h(x, y) = 1$ l'estrazione viene **accettata**

In tal modo x risulterà distribuita secondo $f(x)$.

Limitazioni Oltre a quella per dominio finito, si ha soprattutto un forte limite sull'efficienza per particolari f .

Infatti se la pdf presenta forti escursioni tra massimi e minimi / picchi particolarmente marcati si ha $A_{\text{rigetto}} \gg A_{\text{sottesa}}$ e dunque gran parte delle coppie generate viene rigettata, con importanti conseguenze sull'efficienza computazionale.

Definendo l'**efficienza** dell'algoritmo di rigetto come la frazione di valori accettati, ovvero il rapporto tra l'area della pdf (normalizzata, dunque 1) e quella complessiva di generazione, ovvero il rettangolo determinato da max e dagli estremi di dominio, chiaramente per pdf del genere si avrà una **bassa efficienza** ($eff \ll 1$)

12.2.1 Aumentare l'efficienza

Una possibile soluzione consiste nel determinare **un'altra pdf** $g(x)$ per cui la generazione di punti random è **semplice e veloce** (ad esempio tramite trasformazione inversa → 2.) e che soddisfi, per un opportuno **fattore di scala** S :

$$g'(x) = Sg(x) \geq f(x)$$

ovvero t.c. la funzione $g'(x)$ **copre** $f(x)$.

A questo punto

1. si generano punti in x secondo la $g'(x)$ e si genera $y \in U(0, 1)$.
2. Se

$$y \cdot g'(x) < f(x)$$

si **accetta** l'estrazione, altrimenti si rigetta.

Di fatto si **ridefinisce** l'area di generazione (ovvero il max precedente) di modo da diminuire il rapporto tra l'area di rigetto e quella di accettazione, e dunque il rate di rigetto:

$$A'_{\text{rigetto}} = A(g'(x)) - A(f(x))$$

si osserva che x risulta sempre distribuita secondo la $f(x)$ desiderata in quanto la probabilità che un valore x sia generato è proporzionale a $g'(x) = Sg(x)$ e la probabilità che sia accettato è proporzionale a $\frac{f(x)}{g'(x)} = \frac{f(x)}{Sg(x)}$.

12.3 Metodo della Trasformazione inversa

Se $f(x)$ è **integrabile** e la corrispondente CDF $F(x)$ è **invertibile*** è più efficace utilizzare un metodo alternativo.

Per la normalizzazione, il valore di

$$F(X) = \int_a^X f(x) dx$$

è una variabile casuale distribuita **uniformemente** tra 0 e 1, *indipendentemente dalla distribuzione di partenza*.

Si definisce quindi la variabile y secondo:

$$y = F^{-1}(U) \quad U \in U(0, 1)$$

che sarà distribuita $f(x)$.

Dimostrazione. Si dimostra che $P(y < X) = F(X)$ e quindi $\phi(y) = f(x)$.

Per definizione di y

$$P(y < X) = P(F^{-1}(U) < X)$$

Si applica quindi F a entrambi i membri della disuguaglianza

$$P(F^{-1}(U) < X) = P(F(F^{-1}(U)) < F(X)) = P(U < F(X))$$

la relazione d'ordine è preservata per la **monotonia di F** .

Poiché U è distribuito uniformemente tra 0 e 1, si ha

$$P(U < A) = \int_0^A \frac{1}{1} dx = A$$

quindi

$$P(U < F(X)) = F(X)$$

e dunque

$$P(y < X) = P(F^{-1}(U) < X) = P(F(F^{-1}(U)) < F(X)) = P(U < F(X)) = F(X)$$

Dunque generando uniformemente una variabile y tra 0 e 1 e invertendo la F si genera la variabile casuale x secondo la $f = F'$. ■

Limitazioni

- Solitamente richiede che la cumulativa sia **descrivibile analiticamente** e **invertibile**.
- É possibile tuttavia alternativamente procedere tramite integrazione numerica; numericamente avviene quindi anche il procedimento di inversione.

12.4 Inversa per funzioni notevoli

12.4.1 Esponenziale

$$f(x) = ke^{-kx} \quad F(X) = \int_0^x f(x) dx = 1 - e^{-kx}$$

Da cui

$$y = F^{-1}(U) = -\frac{\ln(1-U)}{k}$$

definendo un'altra uniforme $U' \equiv 1 - U$

$$y = -\frac{\ln U'}{k}$$

y sarà quindi distribuita secondo una pdf esponenziale di parametro k .

12.4.2 Gaussiana

Per la gaussiana standard

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

si integra in due dimensioni e si passa in polari

$$\int_x \int_y \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy = \int_r \int_\varphi \frac{1}{2\pi} e^{-\frac{r^2}{2}} r dr d\varphi = 1$$

La CDF ha quindi la forma

$$F\left(\frac{R^2}{2}, \Phi\right) = \left(1 - e^{-\frac{R^2}{2}}\right) \frac{1}{2\pi} \Phi$$

ovvero è espressa come prodotto di **due cumulative indipendenti**. La cumulativa per l'angolo Φ è un uniforme: rappresenta infatti solamente un **fattore di scala** (nell'integrale è possibile fattorizzare e integrare separatamente $\int d\varphi = \varphi$).

È quindi necessario generare **due** numeri random (altrimenti in 1D non si ha integrabilità) n_1 e n_2 secondo $U(0, 1)$ e determinare

$$x = \sqrt{-2 \ln(1 - n_1)} \cos(2\pi n_2) \quad y = \sqrt{-2 \ln(1 - n_1)} \sin(2\pi n_2)$$

in quanto

$$x = r \cdot \cos \varphi \quad y = r \cdot \sin \varphi$$

tale metodo di generazione è noto come **Metodo di Box-Muller**.

x e y saranno quindi distribuite **secondo una gaussiana 2D** standard (media 0, varianza 1).

Per ottenere una gaussiana unidimensionale, essendo le due variabili generate **indipendenti** è sufficiente ignorare una delle due, ovvero considerare solo x oppure y .

Per ottenere una gaussiana generica è sufficiente operare la trasformazione

$$x' = \mu + \sigma \cdot x$$

ove μ e σ sono specifici della curva desiderata (analogo sostituendo y ad x per quanto detto).

Piccola dimostrazione di Box-Muller Le due cumulative sono indipendenti. Si considera quindi prima quella per r

Dimostrazione.

$$\begin{aligned} F(r) = \left(1 - e^{-\frac{r^2}{2}}\right) &\implies 1 - F = e^{-\frac{r^2}{2}} \implies -\frac{r^2}{2} = \ln(1 - F) \implies \\ &\implies R = \sqrt{-2(1 - F_r)} \end{aligned}$$

La distribuzione di φ è invece uniforme, dunque

$$F(\varphi) = \frac{1}{2\pi} \varphi \implies \varphi = 2\pi F_\varphi$$

quindi generando uniformemente $n_1 = F(r)$ e $n_2 = F(\varphi)$ e ricordando che

$$x = r \cdot \cos \varphi \quad y = r \cdot \sin \varphi$$

si ottengono le formule precedenti. ■