

Appunti Lab

Francesco Sermi

Indice

Capitolo 1	Introduzione	Pagina 2
1.1	L'errore massimo e le sue limitazioni	2
1.2	Precisione vs accuratezza	3
Capitolo 2	Probabilità	Pagina 4
2.1	Definizione operativa della probabilità	6
	Definizione combinatoriale — 6 • Definizione frequentista — 6 • Definizione soggettivista della probabilità — 7	7
2.2	Elementi di calcolo combinatorio	7
2.3	Probabilità condizionata	10
2.4	Teorema di Bayes	11
2.5	Variabili casuali e funzioni di distribuzione	12
2.6	Valore di aspettazione, varianza e momenti di una distribuzione	13
2.7	Momenti di una distribuzione	16
2.8	Funzione cumulativa	17
2.9	Variabili multi-variate	17
Capitolo 3	Variabili campione e propagazione dell'errore statistico	Pagina 22

Capitolo 1

Introduzione

Il concetto di *incertezza* e di *misura* gioca un ruolo centrale nelle scienze sperimentali siccome noi **non** siamo mai in grado di misurare una grandezza fisica con un'accuratezza infinita, dunque il nostro risultato manca di una parte sostanziosa del proprio contenuto se non vi è assegnata una stima dell'incertezza compiuta nella misura effettuata.

Il primo sforzo nella storia delle scienze verso un sistema standard di unità di misure è stato costituito, nella Francia del XVIII secolo, del sistema metrico decimale. Una domanda molto interessante che il lettore si potrebbe chiedere sarebbe la seguente *"Come possiamo definire un'unità di misura?"* La risposta più semplice, ma non banale, è tramite dei campioni inalterabili e riproducibili: si riporta proprio la realizzazione del metro campione e del kilogrammo campione realizzati, sebbene subiscano una contaminazione, ogni anno, pari a $1\mu\text{g}$.

Oggi è stato definito il Sistema internazionale delle misure che si basa sul fissare 7 unità di base da cui è possibile esprimere tutte le altre come combinazione delle altre (si rimanda la lettura delle dispense di Baldini per usufruire dei suoi grafici e/o tabelle).

1.1 L'errore massimo e le sue limitazioni

Si osserva che nella sua formulazione più elementare il concetto di *errore massimo* è legato alla domanda *"Qual è il più piccolo intervallo che contiene con certezza il valore numerico della quantità che sto misurando?"*. In altre parole, il risultato della misura di una generica grandezza fisica x risulta essere pari a:

$$x = \hat{x} \pm \Delta x \quad (1.1)$$

intendo come Δx l'errore massimo, ovvero stiamo dicendo che l'intervallo $[\hat{x} - \Delta x; \hat{x} + \Delta x]$ è il più piccolo intervallo possibile che ci dà la certezza di includere il valore incognito x . Definiamo adesso una serie di termini comodi per esprimere al meglio i concetti successivi:

- x è il valore, incognito (siccome non lo conosciamo mai appieno), della grandezza che vogliamo misurare e che chiameremo *misurando*;
- \hat{x} è la migliore stima di x che possiamo fornire a partire dei dati a nostra disposizione (e che chiameremo *valore centrale* o *migliore stima* della nostra misura);
- Δx è l'incertezza di misura e, in questo **caso**, coincide con l'**errore massimo**

Tuttavia, nel caso di misure ripetute, giungiamo ad un assurdo: infatti, se una misura fluttua, da un punto di vista operativo, non possiamo escludere che una nuova misura della grandezza non fornisca un valore al di fuori dell'intervallo iniziale di incertezza e, ove questo accade, siamo costretti ad allargare tale intervallo, giungendo, dunque, ad un evidente paradosso, siccome acquisire nuove informazioni può solo peggiorare (o lasciare invariato, se otteniamo delle misure che fanno ancora parte di $[x - \Delta x; x + \Delta x]$) il nostro stato di conoscenza, almeno determinando l'incertezza come errore massimo.

Questo è il motivo per cui non utilizziamo quasi mai il concetto di errore massimo, ma utilizzeremo il concetto di errore statistico.

$$x = \hat{x} \pm \sigma_x \quad (1.2)$$

La 1.2 ha un significato diverso da 1.1 prima siccome questa definisce un intervallo che non ci dà la *certezza* ma solo la probabilità, ben definita, di contenere il valore del misurando.

Adesso definiamo un altro modo molto utile per quantificare quanto è grosso, rispetto alla misura, l'errore che noi commettiamo:

Definizione 1.1: Errore percentuale

L'errore relativo è il rapporto tra l'incertezza della misura e il suo valore centrale:

$$e_{\%} = \frac{\sigma_x}{|\hat{x}|} \quad (1.3)$$

1.2 Precisione vs accuratezza

E' molto sottile la distinzione fra la precisione di uno strumento e la sua accuratezza: con il primo termine si indica l'accordo tra il valore misurato dallo strumento e quello effettivo del misurando, mentre con il secondo si intende il grado di consistenza fra i risultati di misure successive della stessa quantità nelle medesime condizioni

Capitolo 2

Probabilità

Grazie al matematico Kolmogorov abbiamo la prima costruzione rigorosa della teoria della probabilità, in una struttura che, sostanzialmente, sopravvive ancora oggi nei manuali moderni.

La struttura di base su cui si fonda la struttura assiomatica della probabilità parte dal definire lo spazio campionario

Definizione 2.1: Spazio campionario Ω

Lo spazio campionario Ω è l'insieme (numerabile) di tutte le possibili realizzazioni elementari di un dato fenomeno e lo spazio degli eventi \mathcal{F} l'insieme di tutti i sottoinsiemi di Ω tale che:

$$\#\mathcal{F} = 2^{\#\Omega}$$

L'idea di Kolmogorov è quella di definire la probabilità direttamente sullo spazio degli eventi-cioè possiamo assegnare una probabilità non solo ad ogni elemento dello spazio campionario, ma anche ad uno qualsiasi dei suoi sottoinsiemi. Definiamo adesso il concetto di probabilità:

Definizione 2.2: Probabilità

Definiamo probabilità una misura P su \mathcal{F} che associ univocamente ad ogni elemento E di \mathcal{F} un numero reale $P(E)$ che soddisfa le seguenti tre proprietà (o **assiomi di Kolmogorov**):

- ① $0 \leq P(E) \leq 1 \forall E \in \mathcal{F}$
- ② $P(\Omega) = 1$
- ③ $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ se $E_1 \cap E_2 = \emptyset$

Possiamo utilizzare il terzo assioma di Kolmogorov all'unione numerabili di eventi *disgiunti* (ovvero eventi per cui la loro intersezione è nulla)

Corollario 2.1 Unione numerabili di eventi disgiunti

$$P\left(\bigcup_i^n E_i\right) = \sum_i^n P(E_i) \text{ se } E_1 = E_2 = \dots = E_n = \emptyset$$

Dimostrazione: Si procede per induzione su n . Per $n = 1$ è banale, siccome:

$$P\left(\bigcup_i^1 E_i\right) = P(E_1) = \sum_i^1 P(E_i) = P(E_i)$$

Adesso mostriamo $n \implies n + 1$:

$$\sum_i^{n+1} P(E_i) = P(E_1) + \dots + P(E_{n+1}) = (\text{ip. induttiva}) P\left(\bigcup_i^n E_i\right) + P(E_{n+1}) = P\left(\sum_i^{n+1} E_i\right)$$

Siccome possiamo vedere nell'ultimo passaggio la somma fra due eventi che sono fra loro disgiunti, ovvero fra l'evento $E_1 \cup E_2 \cdots \cup E_n$ e l'evento E_{n+1} e per ipotesi sappiamo che $E_i \cap E_j = \emptyset \forall i, j$.
La dimostrazione è dunque conclusa. \odot

Corollario 2.2 Probabilità complementare

Dato un evento E e detto \bar{E} il suo complementare in \mathcal{F} , si ha che:

$$P(\bar{E}) = 1 - P(E)$$

dove $P(\bar{E})$ è detta probabilità complementare di E

Dimostrazione: Sapendo che $\bar{E} \cap E = \emptyset$ ma $\bar{E} \cup E = \Omega$:

$$1 = P(\Omega) = P(\bar{E} + E) = P(\bar{E}) + P(E) \implies P(\bar{E}) = 1 - P(E)$$

\odot

Corollario 2.3 Probabilità dell'insieme nullo

$$P(\emptyset) = 0$$

Dimostrazione:

$$P(\Omega) = P(\Omega \cup \emptyset)$$

ma siccome $\Omega \cap \emptyset = \emptyset$ allora

$$P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = P(\Omega) \implies 1 + P(\emptyset) = 1 \implies P(\emptyset) = 0$$

La dimostrazione è dunque conclusa. \odot

Corollario 2.4 Limitatezza della probabilità di un sottoinsieme

$$\text{Se } E_1 \subset E_2 \implies P(E_1) \leq P(E_2)$$

Dimostrazione: Se $E_1 \subset E_2 \implies E_2 = E_1 + (E_2 \setminus E_1)$ ma siccome $E_1 \cap (E_2 \setminus E_1) = \emptyset$:

$$P(E_2) = P(E_1) + P(E_2 \setminus E_1) \implies P(E_1) \leq P(E_2)$$

\odot

Teorema 2.1 Addizione delle probabilità

Dati due eventi E_1 ed E_2 , si ha che:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) \quad (2.1)$$

Dimostrazione: Possiamo scrivere gli insiemi E_1 ed E_2 come unione di eventi disgiunti, infatti:

$$E_2 = E_2 \cap \Omega = E_2 \cap (E_1 \cup \bar{E}_1) = (E_2 \cap E_1) \cup (E_2 \cap \bar{E}_1)$$

ma siccome $(E_2 \cap E_1) \cap (E_2 \cap \bar{E}_1) = \emptyset$ (altrimenti si giungerebbe ad un assurdo, visto che se $E_2 \cap E_1$ avesse degli elementi in comune con $E_2 \cap \bar{E}_1$ implicherebbe, visto che l'intersezione fra insiemi è un'operazione che gode di proprietà associativa e commutativa, che $E_1 \cap \bar{E}_1 \cap E_2 \cap E_2 \neq \emptyset$ ma $E_1 \cap \bar{E}_1 = \emptyset$, dunque assurdo). Tornando alla dimostrazione:

$$P(E_2) = P((E_2 \cap E_1) \cup (E_2 \cap \bar{E}_1)) = P(E_2 \cap E_1) + P(E_2 \cap \bar{E}_1) \quad (2.2)$$

d'altra parte abbiamo che

$$E_1 \cup E_2 = (E_1 \cup E_2) \cap \Omega = (E_1 \cup E_2) \cap (E_1 \cup \bar{E}_1) = (E_1 \cap E_1) \cup (E_1 \cap \bar{E}_1) \cup (E_2 \cap E_1) \cup (E_2 \cap \bar{E}_1) \\ = E_1 \cup (E_2 \cap E_1) \cup (E_2 \cap \bar{E}_1) \text{ ma si osserva che il termine } E_1 \cup (E_2 \cap E_1) \text{ è ridondante, siccome } E_1 \cup (E_2 \cap E_1) = E_1, \\ \text{dunque}$$

$$E_1 \cup E_2 = E_1 \cup (E_2 \cap \bar{E}_1) \implies P(E_1) = P(E_1) + P(E_2 \cap \bar{E}_1)$$

e combinando le due relazioni ottenute

$$\begin{cases} P(E_2) = P(E_2 \cap E_1) + P(E_2 \cap \bar{E}_1) \\ P(E_1 \cup E_2) = P(E_1) + P(E_2 \cap \bar{E}_1) \end{cases}$$

dunque

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

☺

2.1 Definizione operativa della probabilità

Il lettore più attento si sarà però accorto di come la definizione che abbiamo dato di probabilità non ci dà alcun modo con cui calcolare la probabilità, ma piuttosto una serie di assiomi da cui possiamo ricavare una serie di proprietà utili di cui gode la probabilità.

2.1.1 Definizione combinatoriale

Nella sua definizione combinatoriale, definiamo la probabilità di un evento E con quanto segue

Definizione 2.3: Definizione combinatoriale della probabilità

La probabilità di un evento E coincide con il rapporto tra il numero di casi favorevoli n e il numero di casi possibili N , a condizione che questi siano tutti equiprobabili

$$P(E) = \frac{n}{N} \quad (2.3)$$

Osservazione:-

Si osserva che questa definizione di probabilità rispetta i tre assiomi di Kolmogorov: il primo assioma discende dalla ovvia condizione per cui $0 \leq n \leq N$ ed il secondo fatto deriva dal fatto che se $n = N \implies P(E) = 1$. Per quanto riguarda la terza condizione, si osserva che se E_1 e E_2 sono due eventi disgiunti con rispettivamente n_1 e n_2 casi favorevoli, allora:

$$P(E_1 \cup E_2) = \frac{n_1 + n_2}{N} = \frac{n_1}{N} + \frac{n_2}{N} = P(E_1) + P(E_2)$$

dunque anche il terzo assioma è rispettato

Tuttavia questa definizione operativa di probabilità possiede un grande problema: nella definizione è compiuto infatti un ragionamento circolare, siccome richiediamo l'equiprobabilità dei casi nella definizione stessa di probabilità.

2.1.2 Definizione frequentista

Quando è possibile ripetere un esperimento in condizioni controllate, è possibile definire la probabilità di un evento E come il limite della frequenza relativa all'evento stesso quando il numero di ripetizioni N dell'esperimento tende all'infinito. Possiamo quindi pensare di effettuare un esperimento un numero N arbitrariamente grande di volte, contare le n volte in cui è avvenuto l'evento E e definire la probabilità $P(E)$ dell'evento come il limite del rapporto $\frac{n}{N}$

Definizione 2.4: Definizione frequentista della probabilità

La probabilità di un evento E si definisce come

$$P(E) = \lim_{N \rightarrow +\infty} \frac{n}{N} \quad (2.4)$$

dove il limite va inteso nei termini della convergenza statistica, ovvero

$$\forall \epsilon, \delta > 0 \exists \tilde{N} > 0 : \forall N, N > \tilde{N} \implies P\left(\left|\frac{n}{N} - P(E)\right| \geq \delta\right) \leq \epsilon \quad (2.5)$$

Osservazione:-

Il senso di questo limite, che va inteso come limite in senso statistico piuttosto che nel senso usuale dell'analisi matematica, è il fatto che non è possibile garantire a priori l'esistenza di un numero N di ripetizioni del nostro esperimento che mi permetta di affermare con certezza che la differenza tra la frequenza registrata $\frac{n}{N} - P(E)$ sia minore di una certa quantità ϵ : infatti, se effettuiamo due diverse serie di N ripetizioni dell'esperimento otterrò frequenze relative $\frac{n}{N}$ diverse. Quello che possiamo però dire è il fatto che se N è abbastanza grande allora posso rendere piccola a piacere la probabilità che $\frac{n}{N}$ si discosti da $P(E)$ di un valore prefissato δ .

Esempio 2.1 (Lancio di un dado)

Se lanciamo N volte un dado equo a sei facce e registriamo (al crescere di N) il numero n di volte in cui esce, ad esempio, il numero 3, per N molto grande il rapporto $\frac{n}{N}$ tenderà a $P(3) = \frac{1}{6}$ (nel senso della convergenza statistica)

2.1.3 Definizione soggettivista della probabilità

Definizione 2.5: Definizione soggettivista

La probabilità di un evento E si identifica con la misura del grado di fiducia che un individuo attribuisce al verificarsi di E , sulla base dell'informazione a sua disposizione

Osservazione:-

Il termine "soggettivo" si riferisce al fatto che persone diverse, sulla base di differenti informazioni, assoceranno, in generale, una probabilità diversa allo stesso evento e, proprio per questo fatto, si dice che questa definizione è *soggettiva*.

Alla fine, sebbene questa definizione lasci inizialmente sbigottiti siccome si perde quell'oggettività che, in un certo senso, assicuravano le altre due definizioni, questa definizione rispetto di fatto come noi operiamo nella vita di tutti i giorni. All'interno della scuola soggettivista vi sono diversi approcci distinti per derivare le regole fondamentali della probabilità in un modo logicamente consistente (sebbene, con questo *approccio*, gli assiomi non sono tali, ma regole che si ricavano da un principio più formale): il più popolare di questi approcci è il principio della **scommessa coerente**, il quale afferma che *una volta assegnata la probabilità ad un evento dovremo essere disposti ad accettare scommesse sul verificarsi dell'evento stesso con un rapporto tra puntata e vincita determinato dalla probabilità stessa*¹

2.2 Elementi di calcolo combinatorio

La combinatoria è quella branca della matematica che si occupa del *contare*, dunque è strettamente connessa alla probabilità.

¹Il senso è che se diciamo che due eventi sono equiprobabili allora dobbiamo essere pronti ad accettare scommesse 1:1, ovvero che una *scommessa* è coerente se e solo se non determina **a priori** una perdita per il banco o per lo scommettitore, dunque è equivalente anche se i ruoli fossero scambiati

Introduciamo il fattoriale di un numero nella seguente maniera

Definizione 2.6: Fattoriale

Dato un numero $n \in \mathbb{N}$, indichiamo con $n!$

$$n! = \prod_{k=1}^n k = n(n-1) \cdots 1 \quad (2.6)$$

Se ci pensiamo bene, la funzione fattoriale non è altro il numero di *permutazioni*, ovvero il numero di modi in cui si possono disporre n elementi se non possiamo ripeterli e **conta** l'ordine con cui questi elementi vengono disposti: supponiamo infatti di avere 20 oggetti e di volerli disporre all'interno di un cassetto, noi abbiamo ben $20!$ modi possibili per metterli, siccome per il primo "posto" del cassetto possiamo metterci uno dei 20, nel secondo "posto" 19 oggetti e così via; ottenendo ben $20!$ fattoriale di possibilità.

Le permutazioni possono essere viste come un caso particolare delle disposizioni, ovvero i modi con cui è possibile disporre n oggetti in k posti: per esempio, riprendendo l'esempio di prima, se noi volessimo prendere da 20 oggetti 5, presi a caso da questi 20, all'interno di un cassetto in maniera tale che conti l'ordine con cui li mettiamo, si osserva che nel primo *posto* del cassetto abbiamo 20 oggetti disponibili, nel secondo 19, nel terzo 18, nel quarto 17 e nell'ultimo 16; dunque affermare che i modi totali sono $\frac{20!}{15!} = 20 \cdot 19 \cdot 18 \cdot 17 \cdot 16$. Definiamo quindi una disposizione come

Definizione 2.7: Disposizione di n elementi e di ordine k

Definiamo una disposizione di n elementi e di ordine k come il numero di modi con cui è possibile disporre n elementi in k slot, pari a

$$D_{n,k} = \frac{n!}{(n-k)!} \quad (2.7)$$

Tornando alla funzione fattoriale, è comoda l'approssimazione di Stirling (di cui non daremo una dimostrazione) per cui:

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \quad (2.8)$$

Il numero di modi con cui è possibile scegliere k elementi non ordinati è dato dal coefficiente binomiale n su k

Definizione 2.8: Coefficiente binomiale n su k

Il coefficiente binomiale n su k è definito come

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (2.9)$$

Corollario 2.5

$$\binom{n}{k} = \binom{n}{n-k}$$

Dimostrazione: Banalmente, si osserva che

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

☺

Il coefficiente binomiale è strettamente connesso al triangolo di Pascal (o di Newton) e può essere anche caratterizzato attraverso la formula della potenza del binomio (o teorema binomiale). Per farlo però ci servirà il seguente lemma:

Lemna 2.2.1 Proprietà fondamentale del coefficiente binomiale

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$$

Dimostrazione:

$$\begin{aligned} \binom{n}{k+1} + \binom{n}{k} &= \frac{n!}{(k+1)!(n-k-1)!} + \frac{n!}{k!(n-k)!} = \frac{n!}{(k+1)k!(n-k-1)!} + \frac{n!}{k!(n-k)(n-k-1)!} \\ &= \frac{n!}{k!(n-k-1)!} \left(\frac{1}{k+1} + \frac{1}{n-k} \right) = \frac{n!}{k!(n-k-1)!} \left(\frac{n-k+k+1}{(k+1)(n-k)} \right) = \frac{n!}{k!(n-k-1)!} \cdot \frac{(n+1)}{(k+1)(n-k)} \\ &= \frac{(n+1)!}{(k+1)!(n-k)!} = \binom{n+1}{k+1} \end{aligned}$$

⊕

Teorema 2.2 Teorema binomiale

Lo sviluppo della potenza n -esima del binomio $x_1 + x_2$ è pari a

$$(x_1 + x_2)^n = \sum_{k=0}^n \binom{n}{k} x_1^{n-k} x_2^k \quad (2.10)$$

Dimostrazione: si può procedere per induzione. Innanzitutto si osserva che, per $n = 0$, si ha che

$$(x_1 + x_2)^0 = 1 = \sum_{k=0}^0 \binom{0}{k} x_1^{0-k} x_2^k = \binom{0}{0} x_1^0 x_2^0 = 1$$

dunque per $n = 0$ l'ipotesi è verificata. Mostriamo che $n \implies n+1$:

$$(x_1 + x_2)^{n+1} = (x_1 + x_2)^n (x_1 + x_2) = \sum_{k=0}^n \binom{n}{k} x_1^{n-k} x_2^k \cdot (x_1 + x_2) = \sum_{k=0}^n \binom{n}{k} x_1^{n+1-k} x_2^k + \sum_{k=0}^n \binom{n}{k} x_1^{n-k} x_2^{k+1}$$

Da qua soffermiamoci sul primo termine: si osserva che

$$\sum_{k=0}^n \binom{n}{k} x_1^{n+1-k} x_2^k = \binom{n}{0} x_1^{n+1} + \sum_{k=1}^n \binom{n}{k} x_1^{n+1-k} x_2^k = \binom{n}{0} x_1^{n+1} + \sum_{k=0}^{n-1} \binom{n}{k+1} x_1^{n+1-k-1} x_2^{k+1} = x_1^{n+1} + \sum_{k=0}^{n-1} \binom{n}{k+1} x_1^n x_2^{k+1}$$

in cui si è semplicemente effettuato un cambio di variabile di variabile $k' = k+1$ (ma siccome gli indici sono muti abbiamo semplicemente indicato k' sempre con k). Adesso andiamo al secondo termine:

$$\sum_{k=0}^n \binom{n}{k} x_1^{n-k} x_2^{k+1} = \sum_{k=0}^{n-1} \binom{n}{k} x_1^{n-k} x_2^{k+1} + \binom{n}{n} x_1^{n-n} x_2^{n+1} = \sum_{k=0}^{n-1} \binom{n}{k} x_1^{n-k} x_2^{k+1} + x_2^{n+1}$$

Tornando dunque alla relazione iniziale, si deve avere che

$$(x_1 + x_2)^{n+1} = x_1^{n+1} + \sum_{k=0}^{n-1} \binom{n}{k+1} x_1^n x_2^{k+1} + \sum_{k=0}^{n-1} \binom{n}{k} x_1^{n-k} x_2^{k+1} + x_2^{n+1} = x_1^{n+1} + \sum_{k=0}^{n-1} \left[\binom{n}{k} + \binom{n}{k+1} \right] x_1^{n-k} x_2^{k+1} + x_2^{n+1}$$

dunque, siccome per proprietà dei coefficienti binomiali si ha che $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$, allora

$$(x_1 + x_2)^{n+1} = x_1^{n+1} + \sum_{k=0}^{n-1} \binom{n+1}{k+1} x_1^{n-k} x_2^{k+1} + x_2^{n+1} = x_1^{n+1} + \sum_{k=1}^n \binom{n+1}{k} x_1^{n+1-k} x_2^k + x_2^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} x_1^{n+1-k} x_2^k$$

Dunque la tesi è stata dimostrata

⊕

Il coefficiente binomiale è comodo per calcolarsi le potenze del 2 siccome se $x_1 = x_2 = 1$ si ha che

$$2^n = \sum_{k=0}^n \binom{n}{k} 1^{n-k} 1^k = \sum_{k=0}^n \binom{n}{k}$$

Il coefficiente binomiale può anche essere generalizzato supponendo di voler dividere un insieme di n elementi in m sottoinsiemi disgiunti, ciascuno con un numero k_i di elementi, la cui unione costituisca l'insieme di partenza, dunque si deve avere che $\sum_{i=1}^m k_i = n$. Il numero di modi con cui si può effettuare tale suddivisione prende il nome di n su k_1, \dots, k_m e si basa sull'idea che abbiamo n su k_1 modi per scegliere il primo sottoinsieme, $n - k_1$ su k_2 modi per scegliere il secondo sottoinsieme e così via, dunque:

$$\binom{n}{k_1, k_2, \dots, k_m} = \binom{n}{k_1} \binom{n-k_1}{k_2} + \dots + \binom{n-k_1-\dots-k_{m-1}}{k_m}$$

Sempre come per il coefficiente binomiale, possiamo caratterizzare il coefficiente multinomiale come lo sviluppo per la potenza n di un numero arbitrario di monomi nella seguente maniera:

$$(x_1 + x_2 + \dots + x_m)^n = \sum_{\{(k_1, k_2, \dots, k_m) : \sum_{i=1}^m k_i = n\}} \binom{n}{k_1, \dots, k_m} x_1^{k_1} x_2^{k_2} \dots x_m^{k_m}$$

2.3 Probabilità condizionata

Tutto ciò che è stato mostrato nella sezione precedente può essere comodo per calcolare la probabilità di determinati eventi² e consiglio caldamente di guardare gli esempi proposti da Baldini nel suo libro, su cui non mi soffermerò. Diamo delle definizioni

Definizione 2.9: Probabilità condizionata

Dati due eventi E_1 ed E_2 , con $P(E_2) \neq 0$, definiamo la *probabilità condizionata* $P(E_1|E_2)$ di E_1 dato E_2 (cioè la probabilità che si verifichi E_1 nel caso in cui sappiamo già che si verifica E_2) come

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \quad (2.11)$$

Osservazione:-

La probabilità condizionata può essere vista come una sorta di "misura" dell'intersezione $E_1 \cap E_2$ pesata a E_2 . E' possibile inoltre dimostrare che la probabilità condizionata soddisfa tutti gli assiomi della probabilità siccome ogni evento E può essere visto come una probabilità condizionata alla probabilità dello spazio campionario $P(E|\Omega)$

L'utilità della probabilità condizionata è il fatto che possiamo calcolare la probabilità che due eventi E_1 ed E_2 si verifichino contemporaneamente nella seguente maniera:

$$P(E_1 \cap E_2) = P(E_1)P(E_2|E_1) = P(E_2)P(E_1|E_2)$$

A questo punto diamo la definizione di eventi indipendenti

Definizione 2.10: Eventi indipendenti

Si dice che due eventi E_1 ed E_2 sono **indipendenti** se il fatto che sia verificato E_2 non influenza la probabilità che si verifichi E_1 , ovvero se

$$P(E_1|E_2) = P(E_1) \text{ e } P(E_2|E_1) = P(E_2) \quad (2.12)$$

e per la definizione di probabilità condizionata possiamo riscrivere che

$$P(E_1 \cap E_2) = P(E_1)P(E_2) \quad (2.13)$$

²usufruendo della definizione combinatoriale della probabilità

Osservazione:-

Non bisogna confondere il concetto di con il concetto di *eventi disgiunti* e, dunque, incompatibili: si rimanda all'esempio di Baldini sul libro, in cui si osserva proprio il classico esempio del mazzo da carte in cui viene inserito un jolly. Prima dell'aggiunta del jolly i due eventi $E_1 = \text{"pesco un re"}$ ed $E_2 = \text{"pesco una carta di cuori"}$ erano indipendenti pure avendo un'intersezione non nulla, dopo l'aggiunta del jolly non sono più indipendenti ma non sono nemmeno incompatibili

2.4 Teorema di Bayes

La naturale conseguenza di tutto ciò che abbiamo detto sulla probabilità condizionata è il teorema di Bayes, che lega tra loro la probabilità condizionata $P(E_1|E_2)$ con la probabilità $P(E_2|E_1)$

Teorema 2.3 Teorema di Bayes

La probabilità condizionata di E_1 dato E_2 è pari a

$$P(E_1|E_2) = \frac{P(E_2|E_1)P(E_1)}{P(E_2)} \quad (2.14)$$

dove $P(E_2|E_1)$ è la probabilità condizionata di E_2 dato E_1

Dimostrazione: Dalla definizione di probabilità condizionata $P(E_1|E_2)$ di E_1 dato E_2 sappiamo che

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

con $P(E_2) \neq 0$, ma noi sappiamo pure che $P(E_1 \cap E_2) = P(E_2|E_1)P(E_1)$, dunque si ottiene la tesi:

$$P(E_1|E_2) = \frac{P(E_2|E_1)P(E_1)}{P(E_2)}$$

La dimostrazione è dunque conclusa ⊕

Sfruttando inoltre il fatto che $E_1 \cup \bar{E}_1 = \Omega$ allora $(E_1 \cap E_2) \cup (\bar{E}_1 \cap E_2) = E_2 \cap (E_1 \cup \bar{E}_1) = E_2$, dunque:

$$P(E_2) = P(E_1 \cap E_2) + P(\bar{E}_1 \cap E_2)$$

(si ricorda che $(E_1 \cup E_2) \cup (\bar{E}_1 \cup E_2) = \Omega$ siccome se, per assurdo, $\exists c \in (E_1 \cap E_2) \cap (\bar{E}_1 \cap E_2) \neq \emptyset \implies (c \in E_1 \cap E_2 \wedge c \in \bar{E}_1 \cap E_2)$ ma ciò implica che $(c \in E_1 \wedge c \in \bar{E}_1)$ che è un assurdo; dunque possiamo usare il terzo assioma di Kolmogorov per stimare la probabilità di E_2). Ciò implica che

$$P(E_1|E_2) = \frac{P(E_2|E_1)P(E_1)}{P(E_1 \cap E_2) + P(\bar{E}_1 \cap E_2)} \quad (2.15)$$

Più in generale, possiamo dire che se si dispone di un partizionamento dell'insieme $\{A_i\}$ (tali che $A_i \cap A_j = \emptyset \forall i, j, i \neq j$ e $\bigcup_i A_i = \Omega$) allora possiamo generalizzare, per induzione su i , che $P(E_2) = \sum_i P(E_2|A_i)P(A_i)$, dunque:

Corollario 2.6 Caratterizzazione di un evento tramite le partizioni

Sia Ω lo spazio campionario degli eventi e sia $\{E_i\}$ con $i \in I = [1, \dots, m], m \leq \#\mathcal{F}$ un partizionamento^a dell'insieme $\mathcal{F} = P(\Omega)$, ovvero lo spazio degli eventi, allora, dato un evento A , risulta che

$$A = (A \cap E_1) \cup (A \cap E_2) \dots (A \cap E_m)$$

^aricordiamo che un partizione, per essere tale, deve risultare che $\forall i, j, i \neq j E_i \cap E_j = \emptyset$ e $\bigcup_i E_i = \mathcal{F}$

Dimostrazione: Si ha che $\bigcup_i E_i = \mathcal{F}$ e siccome l'evento $A = A \cap \mathcal{F} = A \cap \bigcup_i E_i = A \cap (E_1 \cup E_2 \dots E_m) = (A \cap E_1) \cup (A \cap E_2) \dots (A \cap E_m) = \bigcup_i (A \cap E_i)$. Dimostriamo che gli $(A \cap E_i)$ sono tutti insiemi disgiunti $\forall i$: infatti se, per assurdo, $\exists c \in \mathcal{F} : \exists \tilde{i}, \tilde{d} \in I | (A \cap E_{\tilde{i}}) \cap (A \cap E_{\tilde{d}}) = c \neq \emptyset \implies (c \in (A \cap E_{\tilde{i}}) \wedge c \in (A \cap E_{\tilde{d}}))$ ma, per la definizione dell'operazione \cap , implica che $c \in E_{\tilde{i}} \wedge c \in E_{\tilde{d}}$ il che è un assurdo siccome gli $\{E_i\}$ sono un partizionamento dell'insieme e, dunque, disgiunti. La dimostrazione è dunque conclusa \ominus

Andiamo adesso a scrivere il teorema di Bayes alla luce del corollario qua sopra

$$P(A_1|E_2) = \frac{P(E_2|A_1)P(A_1)}{\sum_i P(E_2|A_i)P(A_i)} \quad (2.16)$$

Questo teorema, sebbene risulti a prima vista banale, in realtà è molto importante siccome lega la probabilità diretta al problema di probabilità inversa. Si rimanda sempre al libro di Baldini per degli esempi per comprendere meglio il teorema di Bayes, sebbene sia utile osservare una cosa: mentre lo statista, oppure il matematico, studia la probabilità in astratto, ovvero cerca di calcolare la probabilità di un determinato processo "contando"; il fisico, di professione, cerca di inferire sul calcolo della probabilità attraverso delle raccolte dati, dunque tramite queste relazioni possiamo eventualmente ricondurci alla probabilità inversa, tuttavia cambia l'approccio "metodologico".

2.5 Variabili casuali e funzioni di distribuzione

Iniziamo questa sezione introducendo il concetto di variabile casuale

Definizione 2.11: Variabile casuale

Una *variabile casuale* o *variabile aleatoria* è una variabile che rappresenta la realizzazione numerica di un processo causale, per cui il suo valore è soggetto a fluttuazioni casuali e non è noto a priori. Si distinguono in

- discrete, cioè variabili che possono assumere un numero finito o numerabile di valori;
- continue, cioè variabili che possono assumere tutti i valori compresi in un intervallo.

Esempio 2.2

- L'uscita del lancio di un dado a sei facce è una variabile casuale discreta che può assumere esattamente sei valori: 1, 2, 3, 4, 5, 6
- Il tempo necessario per arrivare da casa al luogo di lavoro è un esempio di variabile casuale continua

Occupiamoci inizialmente di una variabile discreta x che può assumere n valori distinti x_1, \dots, x_n e indichiamo con $P(x_k)$ la probabilità che assuma il valore x_k e definiamo la ³ nella seguente maniera

Definizione 2.12: Funzione di distribuzione di x

La *funzione di densità di probabilità* è la funzione che associa ad ogni valore di x_k della variabile la sua probabilità $P(x_k)$

In questo contesto il secondo assioma di Kolmogorov si scrive nella forma di una *condizione di normalizzazione*, che tutte le distribuzioni di distribuzione devono rispettare:

$$\sum_k P(x_k) = 1$$

Nel caso di una variabile continua la definizione di funzione di distribuzione data per una variabile discreta fallisce clamorosamente siccome la probabilità che la variabile aleatoria x assuma un valore *esattamente* definito è zero

³in letteratura, si indica con questa espressione anche la funzione cumulativa di una distribuzione. Nel caso continuo, un termine che useremo sarà funzione di densità di probabilità

(infatti mostreremo che si tratta di un integrale su un dominio di misura nulla). Ha però senso chiedersi qual è la probabilità, dato un punto generico x_0 , che la variabile assuma un valore appartenente all'intervallo $[x_0, x_0 + dx]$

$$P(x_0, dx) = P(x_0 \leq x < x_0 + dx)$$

Se a questo punto noi dividiamo per la larghezza dell'intervallo e consideriamo il limite per $dx \rightarrow 0$ otteniamo una sorta di probabilità specifica o probabilità specifica per unità di intervallo che chiamiamo :

$$p(x_0) = \lim_{dx \rightarrow 0} \frac{P(x_0, dx)}{dx} \quad (2.17)$$

che si tratta di una sorta di rapporto incrementale, per cui possiamo dire che la funzione di densità di probabilità è, in un certo senso, la derivata dalla funzione probabilità, dunque possiamo dire che la probabilità che la variabile aleatoria continua assuma un valore compreso nell'intervallo $[x_0, x_0 + dx]$ risulta essere pari

$$P(x_0, dx) = p(x_0)dx$$

e, dunque, la probabilità che assuma un valore compreso nell'intervallo chiuso e limitato $[x_1, x_2]$ è pari a

$$P(x_1, x_2) = \int_{x_1}^{x_2} p(x)dx \quad (2.18)$$

In questo caso la *condizione di normalizzazione*, nel caso di una variabile aleatoria continua, si scrive come

$$\int_{-\infty}^{+\infty} p(x)dx = 1 \quad (2.19)$$

Osservazione:-

Si osserva che mentre la probabilità che la variabile casuale assuma un valore contenuto in $[x_0, x_0 + dx]$ sia un numero adimensionale, nel caso della funzione di densità di probabilità questa dimensionalmente deve avere dimensione pari all'inverso di x

2.6 Valore di aspettazione, varianza e momenti di una distribuzione

Nel momento in cui andiamo a fare un'indagine statistica o, nel caso nostro, effettuiamo un esperimento è comodo condensare le informazioni contenute nella funzione di distribuzione in una serie di parametri significativi come il valore che assume in media la nostra distribuzione oppure quanto la funzione di distribuzione si discosta in media da questo valore.

Andiamo, proprio per questo, a definire il valore di aspettazione, primo strumento utile per caratterizzare i parametri d'interesse di una distribuzione:

Definizione 2.13: Valore di aspettazione di $f(x)$

Sia data una variabile aleatoria x e una funzione $f(x)$, definiamo il valore di aspettazione come

$$E[f(x)] = \begin{cases} \sum_k f(x_k)P(x_k) & \text{nel caso di variabili discrete} \\ \int_{-\infty}^{+\infty} f(x)p(x)dx & \text{nel caso di variabili continue} \end{cases} \quad (2.20)$$

Osservazione:-

Il valore di aspettazione, alla fine, non è altro che una sorta di media "pesata" (con la probabilità che la variabile casuale assuma il valore x_k) della funzione $f(x)$ calcolata in x_k

Inoltre, il valore di aspettazione è un *operatore lineare*, siccome

$$E[c_1 f(x) + c_2 g(x)] = \int_{-\infty}^{+\infty} (c_1 f(x) + c_2 g(x))dx = c_1 \int_{-\infty}^{+\infty} f(x)dx + c_2 \int_{-\infty}^{+\infty} g(x)dx = c_1 E[f(x)] + c_2 E[g(x)]$$

Inoltre il valore di aspettazione di una costante è pari alla costante stessa, siccome

$$E[c] = \int_{-\infty}^{+\infty} c dx = c \int_{-\infty}^{+\infty} dx = c \text{ per la condizione di normalizzazione}$$

Definiamo a questo punto il valore medio di una variabile casuale x (continua o discreta) come

Definizione 2.14: Valore medio di una variabile casuale x

Il valore medio di una variabile casuale x si definisce come il valore di aspettazione di x

$$\mu = E[x] = \begin{cases} \sum_k x_k P(x_k) \\ \int_{-\infty}^{+\infty} x p(x) dx \end{cases} \quad (2.21)$$

Naturalmente, se c è una costante allora

$$E[cx] = cE[x]$$

per linearità dell'integrale.

Osservazione:-

Nel caso particolare di una variabile casuale e discreta per cui si abbiano n uscite equiprobabili x_k equiprobabili (ovvero $P(x_1) = P(x_2) = \dots = P(x_n) = \frac{1}{n}$) si ha che

$$\mu = \sum_{k=1}^n x_k P(x_k) = \frac{1}{n} \sum_{k=1}^n P(x_k)$$

Tuttavia il valore medio non è l'unica stima possibile di tendenza centrale: possiamo definire anche la :

Definizione 2.15: Mediana

Si definisce **mediana** di una distribuzione quel valore $\mu_{\frac{1}{2}}$ della variabile casuale tale che

$$P(x \leq \mu_{\frac{1}{2}}) = P(x \geq \mu_{\frac{1}{2}})$$

Per una variabile casuale continua la mediana è definita, tramite la condizione di normalizzazione, nella seguente maniera

$$\int_{-\infty}^{\mu_{\frac{1}{2}}} p(x) dx = \int_{\mu_{\frac{1}{2}}}^{+\infty} p(x) dx = \frac{1}{2}$$

Per una variabile discreta non è detto che questo valore esista e sia univocamente determinato (proprio per questo la mediana è rilevante per le distribuzioni continue) e, proprietà degna di nota, è il fatto che se la funzione di distribuzione è simmetrica rispetto al valore medio, allora la media coincide con la mediana. Definiamo adesso la *moda* di una distribuzione

Definizione 2.16: Moda

La **moda** di una variabile casuale x è il valore della variabile casuale (se **esiste** ed è **unico**) in corrispondenza del quale la funzione ha un massimo

Come caratterizziamo quanto si disperde la funzione di distribuzione attorno al valore medio? Si fa definendo una funzione il cui valore di aspettazione definisce in media quanto si disperde rispetto al valore medio, dunque "pesiamo" la dispersione rispetto al valore medio con la probabilità che la variabile aleatoria assume quello specifico valore.

Che funzione possiamo prendere? Una funzione del tipo $f(x) = x - \mu$ non va bene siccome:

$$E[f(x)] = E[x - \mu] = E[x] - E[\mu] = \int_{-\infty}^{+\infty} x p(x) dx - \int_{-\infty}^{+\infty} \mu p(x) dx = \mu - \mu \int_{-\infty}^{+\infty} p(x) dx = \mu - \mu = 0$$

dunque questo valore di aspettazione non ci fornisce niente di utile, siccome le fluttuazioni statistiche attorno al valore medio tendono a compensarsi. Possiamo però pensare di utilizzare le fluttuazioni quadratiche $(x - \mu)^2$, dunque:

$$\sigma^2 = E[(x - \mu)^2] = \begin{cases} \sum (x_k - \mu)^2 P(x_k) \\ \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx \end{cases} \quad (2.22)$$

Definiamo dunque la variazione di una distribuzione come:

Definizione 2.17: Varianza di una distribuzione

Si definisce **varianza** σ^2 di una distribuzione il valore di aspettazione di $(x - \mu)^2$, dunque:

$$\sigma^2 = E[(x - \mu)^2] \quad (2.23)$$

e definiamo la **deviazione standard** σ come la radice quadrata della varianza

$$\sigma = \sqrt{\sigma^2} \quad (2.24)$$

Osservazione:-

La deviazione standard ha le stesse dimensioni fisiche della variabile casuale di partenza, dunque è la deviazione standard a caratterizzare la misura della dispersione attorno alla media cercata.

Osserviamo una proprietà utile della varianza, ovvero che se c è una costante, allora

$$\text{Var}(cx) = E[(cx - c\mu)^2] = E[c^2(x - \mu)^2] = c^2 \text{Var}(x)$$

Dimostriamo una formula equivalente per il calcolo della varianza

$$\sigma^2 = E[(x - \mu)^2] = E[x^2 - 2\mu x + \mu^2] = E[x^2] - 2\mu E[x] + E[\mu^2] = E[x^2] - 2\mu^2 + \mu^2 = E[x^2] - \mu^2$$

dunque

$$\sigma^2 = E[x^2] - \mu^2 \quad (2.25)$$

Un concetto utile che si applica alle funzioni di distribuzione di variabile continua (ma ha senso per lo più se si tratta di una distribuzione unimodale) è quello di semilarghezza a metà altezza, ovvero la distanza fra le ascisse x_a e x_b dei punti intersecati dalla retta orizzontale che interseca l'asse delle ordinate in corrispondenza della metà del valore della moda della distribuzione, ovvero il valore massimo assunto da essa.

Definizione 2.18: FWHM e HWHM

La quantità

$$\text{FWHM} = |x_b - x_a| \quad (2.26)$$

prende il nome di *full width at half maximum*, mentre la quantità

$$\text{HWHM} = \frac{|x_b - x_a|}{2} \quad (2.27)$$

prende il nome di *half width at half maximum*

La seconda quantità, ovvero la HWHM, è una stima abbastanza ragionevole, nella maggior parte delle distribuzioni, della deviazione standard, nel senso che

$$\text{HWHM} = c\sigma \quad (2.28)$$

con c dell'ordine delle unità.

Osservazione:-

In un certo senso possiamo affermare che geometricamente la deviazione standard rappresenta una sorta di *larghezza della distribuzione* che stiamo considerando, anche se la *disuguaglianza di Chebyshev* che andremo a considerare fra poco lo renderà ancora più chiaro

Teorema 2.4 Disuguaglianza di Chebyshev

Sia x una variabile casuale tale che esistano finiti la media μ e la varianza σ^2 ; preso $c \in \mathbb{R}^+$ si ha che

$$P(|x - \mu| \geq c\sigma) \leq \frac{1}{c^2} \quad (2.29)$$

Dimostrazione: senza perdita di generalità consideriamo x come una variabile continua (sebbene la dimostrazione nel caso discreto è analoga).

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx \geq \int_{|x-\mu| \geq c\sigma} (x - \mu)^2 p(x) dx \geq \int_{|x-\mu| \geq c\sigma} c^2 \sigma^2 p(x) dx = c^2 \sigma^2 \int_{|x-\mu| \geq c\sigma} p(x) dx \geq c^2 \sigma^2 P(|x - \mu| \geq c\sigma)$$

ma ciò implica la tesi, siccome

$$c^2 \cancel{\sigma^2} P(|x - \mu| \geq c\sigma) \leq \cancel{\sigma^2} \implies P(|x - \mu| \geq c\sigma) \leq \frac{1}{c^2}$$

☺

2.7 Momenti di una distribuzione

Generalizziamo alcune definizioni che abbiamo dato su alcuni parametri di una distribuzione introducendo il concetto di *momento di ordine n* di una variabile casuale x attorno ad un punto x_0

Definizione 2.19: Momento di ordine n attorno al punto x_0

Il momento di ordine n di una variabile casuale x attorno al punto x_0 si definisce come il valore di aspettazione di $f(x) = (x - x_0)^n$, dunque

$$\mathcal{M}_n(x_0) = E[(x - x_0)^n] = \begin{cases} \sum_k (x_k - x_0)^n P(x_k) \\ \int_{-\infty}^{+\infty} (x - x_0)^n p(x) dx \end{cases} \quad (2.30)$$

Hanno rilevanza particolare i momenti algebrici λ_n , ossia i momenti di ordine generico attorno al punto $x_0 = 0$

$$\lambda_n = \mathcal{M}_n(0)$$

ed i momenti centrali μ_n , ossia i momenti di ordine generico attorno al valore medio μ di x

$$\mu_n = \mathcal{M}_n(\mu)$$

dunque possiamo dire che il valor medio è il momento algebrico di ordine 1 e la varianza invece è il momento centrale di ordine 2. Oltre al valor medio e alla varianza, sono utili i momenti centrali di ordine 3 siccome sono i primi momenti di ordine dispari a non annullarsi e poiché misurano l'eventuale asimmetria della funzione di distribuzione, pesando con il segno le code a destra e a sinistra della media. Troviamo una forma più agevole per calcolare μ_3

$$\mu_3 = E[(x - \mu)^3] = E[x^3 - \mu^3 - 3x^2\mu + 3x\mu^2] \implies E[x^3] - 3\mu E[x^2] + 3\mu^2 E[x] - \mu^3$$

e siccome $\sigma^2 = E[x^2] - \mu^2 \implies E[x^2] = \sigma^2 + \mu^2$ allora

$$\mu_3 = E[x^3] - 3\mu(\sigma^2 + \mu^2) + 3\mu^3 - \mu^3 = E[x^3] - 3\sigma^2\mu - \mu^3$$

A questo punto, definiamo il *coefficiente di asimmetria* γ_1

Definizione 2.20: Coefficiente di asimmetria γ_1

Il coefficiente di asimmetria si definisce come il rapporto fra il momento algebrico di ordine $n = 3$ (μ_3) e la deviazione standard al cubi (σ^3)

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \quad (2.31)$$

e si tratta di una quantità adimensionale che vale zero per le distribuzioni simmetriche rispetto al valore medio e che è diversa da zero se la funzione di distribuzione presenta una coda più lunga dell'altra: nel caso in cui di $\gamma_1 > 0 \implies$ la coda a destra è più lunga.

2.8 Funzione cumulativa

Data una variabile casuale x la *funzione cumulativa* è definita come

Definizione 2.21: Funzione cumulativa

La **funzione cumulativa** è definita come

$$F(x') = P(x \leq x') \quad (2.32)$$

e si indica solitamente con lo stesso nome della funzione di distribuzione siccome ha il suo stesso dominio. In maniera operativa, possiamo affermare che

$$F(x) = \begin{cases} \sum_{x_k \leq x} P(x_k) \\ \int_{-\infty}^x p(t) dt \end{cases} \quad (2.33)$$

La funzione cumulativa è una funzione **monotona crescente** e, per la condizione di normalizzazione, si deve avere che

$$\lim_{x \rightarrow -\infty} F(x) = 0 \qquad \lim_{x \rightarrow +\infty} F(x) = 1$$

Inoltre, siccome è strettamente crescente e continua, è invertibile, dunque esiste uno ed un solo valore di x per cui $F(x) = q$. E' possibile quindi definire un inverso della funzione cumulativa che viene chiamata *funzione di distribuzione inversa*

2.9 Variabili multi-variate

La nostra discussione, fino ad adesso, si è concentrata sulla variabili casuali singolo. Come possiamo caratterizzare un insieme di variabili casuali x_1, \dots, x_n ? In maniera più banale di quanto si creda, si potrebbe pensare di considera la funzione di distribuzione congiunta che, ad ogni punto del polirettangolo $A_1 \cup A_2 \cup \dots \cup A_n$, associa la probabilità che le variabili assumano quel *set* di valori.

Consideriamo, per semplicità, due variabili casuali continue x_1 e x_2 descritta dalla densità di probabilità congiunta $p(x_1, x_2)$ tale che

$$p(x_1, x_2) \geq 0 \wedge \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x_1, x_2) dx_1 dx_2 = 1 \quad (2.34)$$

La probabilità che la coppia ordinata $(x_1, x_2) \in A$ è pari a

$$\iint_A p(x_1, x_2) dx_1 dx_2 \quad (2.35)$$

e possiamo sempre definire il valore di aspettazione per una generica funzione $f(x_1, x_2)$ come

$$E[f(x_1, x_2)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1, x_2) p(x_1, x_2) dx_1 dx_2 \quad (2.36)$$

Una domanda che ci potremmo porre è la seguente: come posso capire se due o più variabili x_1, \dots, x_m sono *indipendenti*? Si potrebbe pensare che questa caratteristica si dovrebbe, in qualche maniera, andare a "rintracciare" dalla funzione di densità di probabilità congiunta.

Per fare ciò potremmo pensare di far variare il valore di una variabile aleatoria (come ad esempio x_1 e di fissare quello delle altre variabili. Per esempio, nel caso di due variabili, x_1 e x_2 , allora potremmo pensare di definire una funzione di densità di probabilità *condizionata*, ovvero

$$p(x_1|x_2 \text{ fissato}) = p_1(x_1) \qquad p(x_2|x_1 \text{ fissato})$$

Tuttavia come possiamo scrivere questa probabilità? Si potrebbe pensare che, in maniera abbastanza banale, questa probabilità condizionata si ottiene prendendo la funzione di densità di probabilità e fissando l'altra variabile. Tuttavia, come mostrano molto bene le figure e gli esempi riportati da Baldini (che consiglio di vedere), questa "forma" non è correttamente normalizzata; ma non tutto è da buttare, siccome basta dividere la densità di probabilità per un valore costante ovvero $\int_{-\infty}^{+\infty} p(x_1, x_2) dx_1$:

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{\int_{-\infty}^{+\infty} p(x_1, x_2) dx_1}$$

Definizione 2.22: Probabilità condizionata di due variabili dipendenti

La probabilità condizionata di x_1 relativamente a x_2 (fissato) è definita come

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{\int_{-\infty}^{+\infty} p(x_1, x_2) dx_1} \quad (2.37)$$

Definiamo densità di probabilità marginale il denominatore della precedente formula (siccome dopo l'integrazione sarà esclusivamente in funzione di x_2 e non più di x_1)

$$p_2(x_2) = \int_{-\infty}^{+\infty} p(x_1, x_2) dx_1 \quad (2.38)$$

(per la probabilità condizionata di x_2 relativo a x_1 e la corrispondente probabilità marginale basta sostituire x_1 al posto di x_2 e viceversa nelle precedenti relazioni).

Osservazione:-

Si può dimostrare che la forma 2.37 è correttamente normalizzata, siccome

$$\begin{aligned} p(x_1|x_2 \text{ fissato}) &= \frac{p(x_1, x_2)}{\int_{-\infty}^{+\infty} p(x_1, x_2 \text{ fissato}) dx_1} \implies \int_{-\infty}^{+\infty} p(x_1|x_2) dx_1 = \int_{-\infty}^{+\infty} \frac{p(x_1, x_2)}{\int_{-\infty}^{+\infty} p(x_1, x_2) dx_1} dx_1 \\ &= \frac{1}{\int_{-\infty}^{+\infty} p(x_1, x_2) dx_1} \int_{-\infty}^{+\infty} p(x_1, x_2) dx_1 \implies p(x_1|x_2) = 1 \end{aligned}$$

ed è ragionevole portare fuori l'integrale $p_2(x_2) = \int_{-\infty}^{+\infty} p(x_1, x_2) dx_1$ dall'integrale siccome la densità di probabilità marginale non è più in funzione della variabile x_2 dunque si comporta come una costante rispetto a x_1

In maniera molto simile come per le distribuzioni univariate, diciamo che due variabili aleatorie x_1 e x_2 sono indipendenti se la densità di probabilità congiunta può essere fattorizzata come il prodotto delle due densità di probabilità marginali, ovvero

Definizione 2.23: Indipendenza statistica di due variabili aleatorie

$$p(x_1, x_2) = p_1(x_1)p_2(x_2) \quad (2.39)$$

Dunque, rimettendo insieme con ciò che eravamo partiti, allora possiamo dire che

$$p(x_1|x_2) = \frac{p_1(x_1)p_2(x_2)}{p_2(x_2)} = p_1(x_1) \quad p(x_2|x_1) = \frac{p_1(x_1)p_2(x_2)}{p_1(x_1)} = p_2(x_2)$$

Una proprietà molto comoda delle variabili aleatorie è il fatto che, se due variabili x_1 e x_2 (ci restringiamo al caso in due variabili, ma ciò non è restrittivo) sono **indipendenti**, allora si dimostra, in maniera banale, che il valore di aspettazione del loro prodotto è uguale al prodotto dei valori di aspettazione, infatti:

$$E[x_1x_2] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1x_2p(x_1, x_2)dx_1dx_2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1x_2p_1(x_1)p_2(x_2)dx_1dx_2$$

è possibile applicare il teorema di Fubini, da cui

$$E[x_1x_2] = \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} x_1p_1(x_1)x_2p_2(x_2)dx_1dx_2 = \int_{-\infty}^{+\infty} x_1p_1(x_1)dx_1 \int_{-\infty}^{+\infty} x_2p_2(x_2)dx_2$$

(nell'ultimo passaggio abbiamo portato fuori dei termini che sono solamente in funzione di x_1 e dunque si comportano come una costante rispetto a x_2).

Definizione 2.24: Covarianza

Date due variabili casuali x_1 e x_2 e dette μ_1 e μ_2 il momento algebrico di ordine 1 (ovvero sono il valore medio delle due distribuzioni) definiamo la covarianza $\text{Cov}(x_1, x_2)$ o $\sigma_{x_1x_2}$ come il valore di aspettazione delle relative fluttuazioni attorno al valore medio

$$\text{Cov}(x_1, x_2) = \sigma_{x_1x_2} = E[(x_1 - \mu_1)(x_2 - \mu_2)] \quad (2.40)$$

Si può dimostrare banalmente che la 2.40 può essere scritta come

$$\text{Cov}(x_1, x_2) = E[x_1x_2] - \mu_2E[x_1] - \mu_1E[x_2] + \mu_1\mu_2 = E[x_1x_2] - E[x_1]E[x_2]$$

dunque, si ha che se $E[x_1x_2] = E[x_1]E[x_2] = 0 \iff \text{Cov}(x_1, x_2) = 0$. Dunque si osserva che, se due variabili aleatorie sono indipendenti, allora $\text{Cov}(x_1, x_2) = 0$ ma non è vero il contrario (riporto sotto l'esempio del Baldini).

Esempio 2.3 (Covarianza nulla, variabili dipendenti)

Consideriamo una variabile casuale continua x con una funzione di distribuzione $p(x)$ simmetrica rispetto a 0, il che implica che tutti i momenti algebrici di ordine dispari sono nulli. Si ha banalmente

$$\text{Cov}(x, x^2) = E[x^3] - E[x]E[x^2] = 0$$

siccome $E[x]$ e $E[x^3]$ sono nulli, tuttavia è ovvio che le due variabili non sia indipendenti statisticamente

Da un punto di vista matematico, possiamo dire che la covarianza è una **forma bilineare simmetrica**, nel senso che gode delle seguenti proprietà:

- ① $\text{Cov}(x_1, x_2) = \text{Cov}(x_2, x_1)$
- ② $\text{Cov}(c_1x_1 + c_2x_2, x_3) = c_1\text{Cov}(x_1, x_3) + c_2\text{Cov}(x_2, x_3)$
- ③ $\text{Cov}(x_1, c_2x_2 + c_3x_3) = c_2\text{Cov}(x_1, x_2) + c_3\text{Cov}(x_1, x_3)$

Dimostrazione: per la ① si osserva banalmente che

$$\text{Cov}(x_1, x_2) = E[x_1x_2] - E[x_1]E[x_2] = E[x_2x_1] - E[x_2]E[x_1] = \text{Cov}(x_2, x_1)$$

Per la ② si osserva che

$$\begin{aligned}\text{Cov}(c_1x_1 + c_3x_3, x_2) &= E[(c_1x_1 + c_3x_3)x_2] - E[c_1x_1 + c_3x_3]E[x_2] = E[c_1x_1x_2] + E[c_3x_3x_2] - E[c_1x_1]E[x_2] - E[c_3x_3]E[x_2] \\ &= c_1E[x_1x_2] - c_1E[x_1]E[x_2] + c_3E[x_3x_2] - c_3E[x_3]E[x_2] = c_1\text{Cov}(x_1, x_2) + c_3\text{Cov}(x_3, x_2)\end{aligned}$$

mentre per la ③ si osserva che, usando le due proprietà dimostrate prima

$$\text{Cov}(x_1, c_2x_2 + c_3x_3) = \text{Cov}(c_2x_2 + c_3x_3, x_1) = c_2\text{Cov}(x_2, x_1) + c_3\text{Cov}(x_3, x_1) = c_2\text{Cov}(x_1, x_2) + c_3\text{Cov}(x_1, x_3)$$

⊕

Una proprietà minore è il fatto che $\text{Cov}(x_1, c) = 0$ con c costante e, inoltre, possiamo definire la *varianza* di una variabile aleatoria come la covarianza della variabile con sé stessa, dunque

$$\text{Cov}(x, x) = \text{Var}(x) \quad (2.41)$$

Definizione 2.25: Matrice di covarianza

La matrice di covarianza è la matrice simmetrica $n \times n$ che organizza le covarianze $\text{Cov}(x_i, x_j)$ di n variabili

$$\Sigma = \begin{bmatrix} \text{Cov}(x_1, x_1) & \text{Cov}(x_1, x_2) & \cdots & \text{Cov}(x_1, x_n) \\ \text{Cov}(x_2, x_1) & \text{Cov}(x_2, x_2) & \cdots & \text{Cov}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(x_n, x_1) & \text{Cov}(x_n, x_2) & \cdots & \text{Cov}(x_n, x_n) \end{bmatrix} \quad (2.42)$$

Definizione 2.26: Correlazione

La correlazione è una versione *risalata* della covarianza

$$\text{Corr}(x_1, x_2) = \rho_{x_1x_2} = \frac{\text{Cov}(x_1, x_2)}{\sigma_1\sigma_2} \quad (2.43)$$

e possiede tutte le "proprietà" della matrice di covarianza.

Osservazione:-

La correlazione misura quanto sono *correlate* fra loro due variabili aleatorie, siccome la correlazione assume tutti i valori compresi fra 1 e -1. Formalmente, si può dimostrare che la correlazione agisce come una sorta di prodotto scalare mentre la norma come una norma, pertanto questa proprietà (ovvero quella che assume i valori compresi in $[-1; 1]$ deriva da Cauchy-Schwartz)

Due variabili x_1 e x_2 sono *scorrelate* se $\text{Corr}(x_1, x_2) = 0$, altrimenti diciamo che se $\text{Corr}(x_1, x_2) > 0 \implies x_1, x_2$ sono *positivamente correlate* altrimenti sono *negativamente correlate*. Assume il valore di 1 (o -1) quando dipendono linearmente una dall'altra: infatti se

$$x_2 = mx_1 + q \implies \text{Corr}(x_1, x_2) = \frac{\text{Cov}(x_1, x_2)}{\sigma_{x_1}\sigma_{x_2}} = \frac{\text{Cov}(x_1, mx_1 + q)}{\sigma_{x_1}\sigma_{mx_1+q}}$$

tuttavia si ha che $\sigma_{x_2} = \sqrt{E[(mx_1 + q - m\mu - q)^2]} = |m|\sigma_1$, dunque

$$\text{Corr}(x_1, x_2) = \frac{\text{Cov}(x_1, mx_1 + q)}{|m|\sigma_{x_1}^2} = \frac{\text{Cov}(x_1, mx_1) + \text{Cov}(x_1, q)}{|m|\sigma_{x_1}^2} = \frac{m\sigma_{x_1}^2}{|m|\sigma_{x_1}^2} = \frac{m}{|m|} = \pm 1$$

Analogamente alla covarianza, si definisce anche una *matrice di correlazione* come

Definizione 2.27: Matrice di correlazione

$$\mathcal{R} = \begin{bmatrix} \text{Corr}(x_1, x_1) & \text{Corr}(x_1, x_2) & \dots & \text{Corr}(x_1, x_n) \\ \text{Corr}(x_2, x_1) & \text{Corr}(x_2, x_2) & \dots & \text{Corr}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Corr}(x_n, x_1) & \text{Corr}(x_n, x_2) & \dots & \text{Corr}(x_n, x_n) \end{bmatrix} = \begin{bmatrix} 1 & \text{Corr}(x_1, x_2) & \dots & \text{Corr}(x_1, x_n) \\ \text{Corr}(x_2, x_1) & 1 & \dots & \text{Corr}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Corr}(x_n, x_1) & \text{Corr}(x_n, x_2) & \dots & 1 \end{bmatrix} \quad (2.44)$$

Capitolo 3

Variabili campione e propagazione dell'errore statistico

Ritorniamo però al motivo per cui abbiamo voluto introdurre degli strumenti statistici: ci siamo accorti che l'errore massimo non è adatto per descrivere le incertezze che si compiono effettuando una misura di una determinata grandezza fisica. Adesso però guardiamo la questione da un punto di vista "nuovo": infatti possiamo immaginare che quando misuriamo una grandezza fisica, in particolare quando il valore della misura fluttua, possiamo pensare che il valore stesso sia una variabile aleatoria con una particolare distribuzione (che a priori **non** è nota) che chiamiamo *distribuzione generatrice*.

In questo schema concettuale fare n misure di una stessa grandezza fisica in condizioni di ripetitività equivale a *campionare* n volte la distribuzione generatrice che caratterizza quel determinato misurando. E' ovvio che non potremo mai conoscere completamente la forma della distribuzione generatrice, ma in maniera intuitiva possiamo pensare che, effettuando sempre più misurazioni, inizieremo ad acquisire progressivamente sempre più informazioni su di essa: se pensiamo di fare un numero molto grande di misure (e.g. $n \rightarrow +\infty$) e riportiamo i risultati di queste misure in un istogramma allora ci aspettiamo che la forma di questo diventi sempre più simile a quello della distribuzione generatrice.

Possiamo quindi intravedere un nuovo modo per operare, ovvero quello di scrivere il risultato di una misura come il valore centrale e di un'incertezza ad esso associato

Indice analitico

approssimazione di Stirling, 8
assiomi di Kolmogorov, 4

coefficiente binomiale, 8
coefficiente di asimmetria, 17
coefficiente multinomiale, 10
condizione di normalizzazione, 12, 13
correlazione, 20
covarianza, 19

densità di probabilità, 13
deviazione standard, 15
disposizione, 8
distribuzione generatrice, 22
disuguaglianza di Chebyshev, 16

errore massimo, 2
errore relativo, 3
eventi indipendenti, 10, 11

fattoriale, 8
funzione cumulativa, 17
funzione di distribuzione, 12
FWHM, 15

FWHM, 15

matrice di correlazione, 21
matrice di covarianza, 20
mediana, 14
moda, 14
momento algebrico, 16, 19
momento centrale, 16
momento di ordine n , 16

probabilità, 4
probabilità complementare, 5
probabilità condizionata, 10, 18

spazio campionario, 4
spazio degli eventi, 4

teorema binomiale, 9
teorema di Bayes, 11

valore di aspettazione, 13
valore medio, 14, 16, 19
variabile aleatoria, 12
variabile casuale, 12
varianza, 15, 16, 20