

# Appunti di probabilità e statistica per Laboratorio I

Francesco Sermi

# Indice

<b>Capitolo 1</b>	<b>Probabilità</b>	<b>Pagina 2</b>
1.1	Definizione operativa della probabilità Definizione combinatoriale — 4 • Definizione frequentista — 4 • Definizione soggettivista della probabilità — 5	4
1.2	Elementi di calcolo combinatorio	5
1.3	Probabilità condizionata	8
1.4	Teorema di Bayes	9
1.5	Variabili casuali e funzioni di distribuzione	10
1.6	Valore di aspettazione, varianza e momenti di una distribuzione	11
1.7	Momenti di una distribuzione	14
1.8	Funzione cumulativa	15
1.9	Variabili multi-variate	15
<b>Capitolo 2</b>	<b>Variabili campione e propagazione dell'errore statistico</b>	<b>Pagina 20</b>
2.1	Campionamenti singoli e ripetuti Campionamenti singoli — 20 • Interludio: somma di variabili aleatorie — 21 • Misure ripetute — 22	20
2.2	Covarianza e correlazione campione	24
2.3	Media e varianza di una funzione a variabili casuali	25
2.4	Propagazione dell'errore statistico	26
<b>Capitolo 3</b>	<b>Distribuzioni uni-variate di uso comune</b>	<b>Pagina 27</b>
3.1	La distribuzione binomiale Normalizzazione, media e varianza — 27	27
3.2	La distribuzione di Poisson La poissoniana come limite della distribuzione binomiale — 30 • Normalizzazione, media e varianza — 31 • Somma di variabili poissoniane — 33	29
3.3	La distribuzione uniforme Normalizzazione, media, varianza e asimmetria — 33	33
3.4	La distribuzione esponenziale Normalizzazione, media, varianza e asimmetria — 35 • Funzione cumulativa — 37 • Assenza di memoria — 37	34
3.5	Distribuzione di Gauss Normalizzazione, media e varianza — 38	37
3.6	Il teorema centrale del limite	40
<b>Capitolo 4</b>	<b>Funzioni di probabilità di funzioni qualunque</b>	<b>Pagina 41</b>
4.1	Variabili discrete	41

# Capitolo 1

## 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 1.1: Spazio campionario $\Omega$

Lo spazio campionario  $\Omega$  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  $\Omega$  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 1.2: 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 1.1 Unione numerabili di eventi disgiunti

$P(\bigcup_i^n E_i) = \sum_i^n P(E_i)$  se gli eventi sono tutti disgiunti fra loro (formalmente,  $\forall 1 \leq i < j \leq n, E_i \cap E_j = \emptyset$ )

**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.  $\oplus$

### Corollario 1.2 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  $\bar{E} \cup E = \Omega$ :

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

$\oplus$

### Corollario 1.3 Probabilità dell'insieme nullo

$$P(\emptyset) = 0$$

**Dimostrazione:**

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

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

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

La dimostrazione è dunque conclusa.  $\oplus$

### Corollario 1.4 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  $E_1 \cap (E_2 \setminus E_1) = \emptyset$ :

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

$\oplus$

### Teorema 1.1 Addizione delle probabilità

Dati due eventi  $E_1$  ed  $E_2$ , si ha che:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) \quad (1.1)$$

**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:

$$P(E_2) = P((E_2 \cap E_1) \cup (E_2 \cap \bar{E}_1)) = P(E_2 \cap E_1) + P(E_2 \cap \bar{E}_1) \quad (1.2)$$

d'altra parte abbiamo che

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

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

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

e combinando le due relazioni ottenute

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

dunque

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

⊗

## 1.1 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.

### 1.1.1 Definizione combinatoriale

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

#### Definizione 1.3: 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

$$P(E) = \frac{n}{N} \quad (1.3)$$

#### 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.

### 1.1.2 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 1.4: Definizione frequentista della probabilità

La probabilità di un evento  $E$  si definisce come

$$P(E) = \lim_{N \rightarrow +\infty} \frac{n}{N} \quad (1.4)$$

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

$$\forall \epsilon, \delta > 0 \exists \tilde{N} > 0 : \forall N, N > \tilde{N} \implies P\left(\left|\frac{n}{N} - P(E)\right| \geq \delta\right) \leq \epsilon \quad (1.5)$$

#### 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  $\epsilon$ : 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  $\delta$ .

#### Esempio 1.1 (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)

### 1.1.3 Definizione soggettivista della probabilità

#### Definizione 1.5: 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*<sup>1</sup>

## 1.2 Elementi di calcolo combinatorio

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

Introduciamo il fattoriale di un numero nella seguente maniera

<sup>1</sup>Il 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

**Definizione 1.6: Fattoriale**

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

$$n! = \prod_{k=1}^n k = n(n-1) \cdots 1 \quad (1.6)$$

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 1.7: 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

$$D_{n,k} = \frac{n!}{(n-k)!} \quad (1.7)$$

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

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \quad (1.8)$$

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

**Definizione 1.8: Coefficiente binomiale  $n$  su  $k$** 

Il coefficiente binomiale  $n$  su  $k$  è definito come

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (1.9)$$

**Corollario 1.5**

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

**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 1.2.1** 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 1.2** Teorema binomiale

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

$$(x_1 + x_2)^n = \sum_{k=0}^n \binom{n}{k} x_1^{n-k} x_2^k \quad (1.10)$$

**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  $n \implies n+1$ :

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

Da qua soffermiamoci sul primo termine: si osserva che

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

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:

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

### 1.3 Probabilità condizionata

Tutto ciò che è stato mostrato nella sezione precedente può essere comodo per calcolare la probabilità di determinati eventi<sup>2</sup> e consiglio caldamente di guardare gli esempi proposti da Baldini nel suo libro, su cui non mi soffermerò. Diamo delle definizioni

#### Definizione 1.9: 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

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \quad (1.11)$$

#### 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 1.10: 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

$$P(E_1|E_2) = P(E_1) \text{ e } P(E_2|E_1) = P(E_2) \quad (1.12)$$

e per la definizione di probabilità condizionata possiamo riscrivere che

$$P(E_1 \cap E_2) = P(E_1)P(E_2) \quad (1.13)$$

<sup>2</sup>usufruendo 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

## 1.4 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 1.3 Teorema di Bayes

La probabilità condizionata di  $E_1$  dato  $E_2$  è pari a

$$P(E_1|E_2) = \frac{P(E_2|E_1)P(E_1)}{P(E_2)} \quad (1.14)$$

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

$$P(E_1|E_2) = \frac{P(E_2|E_1)P(E_1)}{P(E_1 \cap E_2) + P(\bar{E}_1 \cap E_2)} \quad (1.15)$$

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 1.6 Caratterizzazione di un evento tramite le partizioni

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

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

<sup>a</sup>ricordiamo 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

$$P(A_1|E_2) = \frac{P(E_2|A_1)P(A_1)}{\sum_i P(E_2|A_i)P(A_i)} \quad (1.16)$$

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".

## 1.5 Variabili casuali e funzioni di distribuzione

Iniziamo questa sezione introducendo il concetto di variabile casuale

### Definizione 1.11: 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 1.2

- 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 <sup>3</sup> nella seguente maniera

### Definizione 1.12: 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 (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  $[x_0, x_0 + dx]$

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

<sup>3</sup>in 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

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 :

$$p(x_0) = \lim_{dx \rightarrow 0} \frac{P(x_0, dx)}{dx} \quad (1.17)$$

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 della 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

$$P(x_1, x_2) = \int_{x_1}^{x_2} p(x)dx \quad (1.18)$$

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

$$\int_{-\infty}^{+\infty} p(x)dx = 1 \quad (1.19)$$

#### 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  $x$

## 1.6 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 1.13: Valore di aspettazione di $f(x)$

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

$$E[f(x)] = \begin{cases} \sum_k f(x_k)P(x_k) & \text{nel caso di variabili discrete} \\ \int_{-\infty}^{+\infty} f(x)p(x)dx & \text{nel caso di variabili continue} \end{cases} \quad (1.20)$$

#### 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

$$E[c_1f(x) + c_2g(x)] = \int_{-\infty}^{+\infty} (c_1f(x) + c_2g(x))dx = c_1 \int_{-\infty}^{+\infty} f(x)dx + c_2 \int_{-\infty}^{+\infty} g(x)dx = c_1E[f(x)] + c_2E[g(x)]$$

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

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

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

**Definizione 1.14: Valore medio di una variabile casuale  $x$** 

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

$$\mu = E[x] = \begin{cases} \sum_k x_k P(x_k) \\ \int_{-\infty}^{+\infty} x p(x) dx \end{cases} \quad (1.21)$$

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 1.15: 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 1.16: 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:

$$\sigma^2 = E[(x - \mu)^2] = \begin{cases} \sum_{k=-\infty}^{+\infty} (x_k - \mu)^2 P(x_k) \\ \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx \end{cases} \quad (1.22)$$

Definiamo dunque la variazione di una distribuzione come:

**Definizione 1.17: Varianza di una distribuzione**

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

$$\sigma^2 = E[(x - \mu)^2] \quad (1.23)$$

e definiamo la **deviazione standard**  $\sigma$  come la radice quadrata della varianza

$$\sigma = \sqrt{\sigma^2} \quad (1.24)$$

**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

$$\sigma^2 = E[x^2] - \mu^2 \quad (1.25)$$

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 1.18: FWHM e HWHM**

La quantità

$$\text{FWHM} = |x_b - x_a| \quad (1.26)$$

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

$$\text{HWHM} = \frac{|x_b - x_a|}{2} \quad (1.27)$$

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

$$\text{HWHM} = c\sigma \quad (1.28)$$

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 1.4** Disuguaglianza di Chebyshev

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

$$P(|x - \mu| \geq c\sigma) \leq \frac{1}{c^2} \quad (1.29)$$

**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

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

☺

## 1.7 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 1.19: 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

$$\mathcal{M}_n(x_0) = E[(x - x_0)^n] = \begin{cases} \sum_k (x_k - x_0)^n P(x_k) \\ \int_{-\infty}^{+\infty} (x - x_0)^n p(x) dx \end{cases} \quad (1.30)$$

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

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

ed i momenti centrali  $\mu_n$ , ossia i momenti di ordine generico attorno al valore medio  $\mu$  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  $\mu_3$

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

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

$$\mu_3 = E[x^3] - 3\mu(\sigma^2 + \mu^2) + 3\mu^3 - \mu^3 = E[x^3] - 3\sigma^2\mu - \mu^3 \quad (1.31)$$

A questo punto, definiamo il *coefficiente di asimmetria*  $\gamma_1$

### Definizione 1.20: Coefficiente di asimmetria $\gamma_1$

Il coefficiente di asimmetria si definisce come il rapporto fra il momento algebrico di ordine  $n = 3$  ( $\mu_3$ ) e la deviazione standard al cubi ( $\sigma^3$ )

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \quad (1.32)$$

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  $\gamma_1 > 0 \implies$  la coda a destra è più lunga.

## 1.8 Funzione cumulativa

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

### Definizione 1.21: Funzione cumulativa

La **funzione cumulativa** è definita come

$$F(x') = P(x \leq x') \quad (1.33)$$

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

$$F(x) = \begin{cases} \sum_{x_k \leq x} P(x_k) \\ \int_{-\infty}^x p(t) dt \end{cases} \quad (1.34)$$

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*

## 1.9 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

$$p(x_1, x_2) \geq 0 \wedge \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x_1, x_2) dx_1 dx_2 = 1 \quad (1.35)$$

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

$$\iint_A p(x_1, x_2) dx_1 dx_2 \quad (1.36)$$

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

$$E[f(x_1, x_2)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1, x_2) p(x_1, x_2) dx_1 dx_2 \quad (1.37)$$



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 1.22: Probabilità condizionata di due variabili dipendenti

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

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{\int_{-\infty}^{+\infty} p(x_1, x_2) dx_1} \quad (1.38)$$

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  $x_1$ )

$$p_2(x_2) = \int_{-\infty}^{+\infty} p(x_1, x_2) dx_1 \quad (1.39)$$

(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 1.38 è correttamente normalizzata, siccome

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

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 1.23: Indipendenza statistica di due variabili aleatorie**

$$p(x_1, x_2) = p_1(x_1)p_2(x_2) \quad (1.40)$$

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 1.24: Covarianza**

Date due variabili casuali  $x_1$  e  $x_2$  e dette  $\mu_1$  e  $\mu_2$  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

$$\text{Cov}(x_1, x_2) = \sigma_{x_1x_2} = E[(x_1 - \mu_1)(x_2 - \mu_2)] \quad (1.41)$$

Si può dimostrare banalmente che la 1.41 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 1.3 (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 siano 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

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

Per la ② si osserva che

$$\begin{aligned} \text{Cov}(c_1 x_1 + c_3 x_3, x_2) &= E[(c_1 x_1 + c_3 x_3)x_2] - E[c_1 x_1 + c_3 x_3]E[x_2] = E[c_1 x_1 x_2] + E[c_3 x_3 x_2] - E[c_1 x_1]E[x_2] - E[c_3 x_3]E[x_2] \\ &= c_1 E[x_1 x_2] - c_1 E[x_1]E[x_2] + c_3 E[x_3 x_2] - c_3 E[x_3]E[x_2] = c_1 \text{Cov}(x_1, x_2) + c_3 \text{Cov}(x_3, x_2) \end{aligned}$$

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

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

☺

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

$$\text{Cov}(x, x) = \text{Var}(x) \quad (1.42)$$

#### Definizione 1.25: 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

$$\Sigma = \begin{bmatrix} \text{Cov}(x_1, x_1) & \text{Cov}(x_1, x_2) & \cdots & \text{Cov}(x_1, x_n) \\ \text{Cov}(x_2, x_1) & \text{Cov}(x_2, x_2) & \cdots & \text{Cov}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(x_n, x_1) & \text{Cov}(x_n, x_2) & \cdots & \text{Cov}(x_n, x_n) \end{bmatrix} \quad (1.43)$$

#### Definizione 1.26: Correlazione

La correlazione è una versione *risalata* della covarianza

$$\text{Corr}(x_1, x_2) = \rho_{x_1 x_2} = \frac{\text{Cov}(x_1, x_2)}{\sigma_1 \sigma_2} \quad (1.44)$$

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

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

tuttavia si ha che  $\sigma_{x_2} = \sqrt{E[(m x_1 + q - m \mu - q)^2]} = |m| \sigma_1$ , dunque

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

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

**Definizione 1.27: Matrice di correlazione**

$$\mathcal{R} = \begin{bmatrix} \text{Corr}(x_1, x_1) & \text{Corr}(x_1, x_2) & \dots & \text{Corr}(x_1, x_n) \\ \text{Corr}(x_2, x_1) & \text{Corr}(x_2, x_2) & \dots & \text{Corr}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Corr}(x_n, x_1) & \text{Corr}(x_n, x_2) & \dots & \text{Corr}(x_n, x_n) \end{bmatrix} = \begin{bmatrix} 1 & \text{Corr}(x_1, x_2) & \dots & \text{Corr}(x_1, x_n) \\ \text{Corr}(x_2, x_1) & 1 & \dots & \text{Corr}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Corr}(x_n, x_1) & \text{Corr}(x_n, x_2) & \dots & 1 \end{bmatrix} \quad (1.45)$$

## Capitolo 2

# 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 come migliore stima della grandezza il valore centrale della distribuzione e come incertezza la deviazione standard.

Il nuovo errore si presenterà dunque in questa maniera:

$$x = \hat{x} \pm \sigma_x [\text{unità di misura}] \quad (2.1)$$

## 2.1 Campionamenti singoli e ripetuti

### 2.1.1 Campionamenti singoli

Se conosciamo a priori la deviazione standard  $\sigma$  della distribuzione generatrice del misurando (ma anche avendone una stima) che stiamo, per appunto, *misurando* allora una singola misura (ovvero un singolo campionamento) è sufficiente per definire tutte le componenti della 2.1: prenderemo il singolo valore misurato come valore centrale e  $\sigma$  come incertezza associata.

Se conosciamo inoltre la forma della distribuzione possiamo anche determinare il livello di confidenza associato ad una deviazione standard, oppure utilizzare come stima dell'incertezza un multiplo o sottomultiplo della deviazione standard per ottenere un livello di confidenza fissato a priori.

#### Esempio 2.1 (Esempi di campionamenti singoli sapendo la deviazione standard a priori)

- Supponiamo di avere un pesi  $m$  di un oggetto con una bilancia digitale con una risoluzione di 1 g. Se il valore indicato dal display è 58 g, possiamo assumere che, in assenza di errori sistematici, la distribuzione generatrice del misurando sia uniforme tra 57.5 g e 58.5 g e possiamo utilizzare la deviazione standard della funzione uniforme che risulta essere pari a  $\sigma = \sqrt{\text{Var}(x)} = \sqrt{\frac{1}{12}}$  g e il livello

di confidenza (ovvero la probabilità che la variabile disti meno di una deviazione standard) è pari al 58%, dunque

$$m = 58.00 \pm 0.29 \text{ g (58\% CL)}$$

si osserva infatti che la probabilità che si trovi entro una deviazione standard è pari a

$$\int_{58 - \frac{1}{\sqrt{12}}}^{58 + \frac{1}{\sqrt{12}}} \frac{1}{(58.5 - 57.5) \text{ g}} dm \approx 0.577$$

- Il ragionamento che abbiamo fatto prima si applica alla misura di una lunghezza con il metro a nastro e, più in generale a tutti gli strumenti digitali (se decidiamo di non interpolare tra le divisioni)
- In generale gli strumenti si possono *calibrare* tramite misure ripetute di grandezze di riferimento oppure tramite uno strumento di misura più accurato (o usando un metodo indipendente)

### 2.1.2 Interludio: somma di variabili aleatorie

Il problema si pone quando non conosciamo a priori la deviazione standard della nostra misura: come possiamo fare in tal caso? In primis dobbiamo sapere che cosa accade quando sommiamo due o più variabili casuali: in generale non è banale individuare qual è la "forma" della funzione di distribuzione di  $x = \sum_i x_i$ , tuttavia possiamo determinare abbastanza facilmente alcuni parametri che caratterizzano la variabile aleatoria  $x$ , infatti:

$$E[x] = E\left[\sum_i x_i\right] = \sum_i E[x_i] = \sum_i \mu_i$$

La varianza è più complicata (e ci limiteremo quindi al caso, generalizzabile, di due variabili aleatorie)

$$\begin{aligned} \text{Var}(x) &= E[(x - \mu)^2] = E[x^2] - 2\mu E[x] + E[\mu^2] = E[(x_1 + x_2)^2] - 2(\mu_1 + \mu_2)^2 + (\mu_1 + \mu_2)^2 \\ &= E[x_1^2] + E[x_2^2] + 2E[x_1 x_2] - (\mu_1 + \mu_2)^2 = \\ &= E[x_1^2] + E[x_2^2] + 2\text{Cov}(x_1, x_2) - \mu_1^2 - \mu_2^2 - 2\mu_1 \mu_2 = E[x_1^2] - \mu_1^2 + E[x_2^2] - \mu_2^2 + 2\text{Cov}(x_1, x_2) = \\ &= \text{Var}(x_1) + \text{Var}(x_2) + 2\text{Cov}(x_1, x_2) \end{aligned}$$

Nel caso in cui  $x_1$  e  $x_2$  siano variabili aleatorie indipendenti, sappiamo che  $\text{Cov}(x_1, x_2) = 0$  siccome si ha necessariamente che  $E[x_1 x_2] = E[x_1]E[x_2]$ , dunque:

#### Proposizione 2.1 Valore medio e varianza della somma di variabili indipendenti

Sia  $x = \sum_i x_i$  con  $p(x_i)p(x_j) = p(x_i \cap x_j) \forall i \neq j$  ( $x$  è somma di eventi indipendenti) allora

$$\mu = \sum_{i=1} E[x_i] \quad \text{e} \quad \text{Var}(x) = \sqrt{\sum_{i=1} \sigma_i^2}$$

**Dimostrazione:** Si procede per induzione su  $n$ . Per  $n = 1$  si osserva che

$$E[x] = \sum_{i=1} E[x_i] = \mu_1$$

adesso mostriamo che  $n \implies n + 1$ :

$$E[x] = \sum_{i=1}^{n+1} E[x_i] = \sum_{i=1}^n E[x_i] + \mu_{n+1} = \mu_n + \mu_{n+1} = \sum_{i=1}^{n+1} \mu_i$$

Per la varianza si procede alla stessa maniera, sapendo che  $\text{Cov}(x_n, x_{n+1}) = 0$

☺

Comunque si osservi che

$$\sqrt{\sum_{i=1} \sigma_i^2} \leq \sum_{i=1} \sigma_i \quad (2.2)$$

**Dimostrazione:** Si osservi che siccome  $\sigma_i^2 + \sigma_j^2 \leq \sigma_i^2 + \sigma_j^2 + 2\sigma_i\sigma_j = (\sigma_i + \sigma_j)^2$  si deve avere che

$$\sqrt{\sum_i \sigma_i^2} \leq \sqrt{\left(\sum_{i=1} \sigma_i\right)^2} = \sum_{i=1} \sigma_i$$

☺

Abbiamo dunque dimostrato che *date  $n$  variabili casuali indipendenti la **media della somma** è uguale alla **somma delle medie** e la **varianza della somma** è uguale alla **somma delle varianze***

**Osservazione:-**

Un esempio interessante che può essere visto, soprattutto perché utile quando tratteremo la distribuzione di Cauchy, è il cosiddetto *random walk*

### 2.1.3 Misure ripetute

Tuttavia torniamo alla questione originale: se abbiamo una serie di misure  $x_i \wedge i \in I = 1, \dots, n$  indipendenti di una stessa grandezza  $x$  fatte in condizione di ripetitività, quali sono le migliori stime che possiamo dare della media  $\mu$  e della varianza  $\sigma^2$  della distribuzione generatrice? Come stima della media possiamo pensare di prendere la media aritmetica delle misure, chiamata *media campione*:

**Definizione 2.1: Media campione**

$$m = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.3)$$

La media campione ha la proprietà che il suo valore di aspettazione è uguale alla media della distribuzione generatrice:

$$E[m] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \cdot n\mu = \mu$$

e un estimatore che soddisfa questa proprietà si dice *imparziale*.

La questione della varianza è più complicata siccome l'analogo per la varianza sarebbe la seguente

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad (2.4)$$

che si tratta di un estimatore imparziale, siccome si osserva che

$$E[s^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2\right] = \frac{1}{n} E[(x_i - \mu)^2] = \sigma^2$$

ma a priori non conosciamo  $\mu$ . Come migliore stima della varianza potremmo considerare dunque la *varianza campione*

**Definizione 2.2: Varianza campione**

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 \quad (2.5)$$

E quindi ci potremmo chiedere se il valore di aspettazione della varianza campione sia ancora pari a  $\sigma$ . Facendo la derivata rispetto ad una generica stima della media, si osserva che

$$\frac{d}{d\xi} s_n^2 = -\frac{1}{n} \sum_{i=1}^n 2(x_i - \xi) = \frac{2}{n} \left( \sum_{i=1}^n x_i - \sum_{i=1}^n \xi \right) = \frac{2}{n} (nm - n\xi) = 2(m - \xi)$$

dunque la media campionaria è la media che minimizza la stima della varianza campione, dunque questo ci dice che, in media,  $s_n^2$  è una sottostima di  $\sigma^2$ . Calcoliamo il valore di aspettazione di  $s_n^2$ , che risulta essere pari a

$$E[s_n^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - m)^2\right] = \frac{1}{n} \sum_{i=1}^n E[x_i^2 + m^2 - 2mx_i] = \frac{1}{n} \sum_{i=1}^n (E[x_i^2] + E[m^2] - 2E[mx_i])$$

Sapendo che  $E[x_i^2] = \sigma^2 + \mu^2$ , ci resta da calcolare solo  $E[m^2]$  e  $E[mx_i]$ . Si osserva che  $E[m^2]$

$$\begin{aligned} E[m^2] &= E\left[\left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2\right] = \frac{1}{n^2} E\left[\left(\sum_{i,j=1}^n x_i x_j\right)\right] = \frac{1}{n^2} E\left[\left(\sum_{i=1}^n x_i^2 + \sum_{i=1}^n \sum_{j \neq i}^n x_i x_j\right)\right] = \frac{1}{n^2} \left(E\left[\sum_{i=1}^n x_i^2\right] + E\left[\sum_{i=1}^n \sum_{j \neq i}^n x_i x_j\right]\right) = \\ &= \frac{1}{n^2} \left(n(\sigma^2 + \mu^2) + E\left[\sum_{i=1}^n \sum_{j \neq i}^n x_i x_j\right]\right) \end{aligned}$$

Si osserva però che la sommatoria  $E\left[\sum_{i=1}^n \sum_{j \neq i}^n x_i x_j\right]$ , tramite linearità dell'operatore di aspettazione, può essere trasformata come

$$E\left[\sum_{i=1}^n \sum_{j \neq i}^n x_i x_j\right] = \sum_{i=1}^n \sum_{j \neq i}^n E[x_i]E[x_j] = \sum_{i=1}^n (n-1)\mu^2 = n(n-1)\mu^2$$

dunque, abbiamo che

$$E[m^2] = \frac{1}{n^2} [n(\sigma^2 + \mu^2) + n(n-1)\mu^2] = \frac{1}{n^2} [n\sigma^2 + n\mu^2 + n^2\mu - n\mu^2] = \frac{1}{n}\sigma^2 + \mu^2$$

Per quanto riguarda il valore di aspettazione di  $E[mx_i]$  si ha che

$$E[mx_i] = E\left[\frac{1}{n} \sum_{j=1}^n x_j x_i\right] = \frac{1}{n} E\left[x_i^2 + \sum_{j \neq i}^n x_i x_j\right] = \frac{1}{n} \left(E[x_i^2] + \sum_{j \neq i}^n E[x_i]E[x_j]\right) = \frac{1}{n} [\sigma^2 + \mu^2 + (n-1)\mu^2] = \frac{\sigma^2}{n} + \mu^2$$

Rimettendo tutto insieme si osserva che nessuno di questi valori dipende dall'indice  $i$ , dunque

$$E[s_n^2] = \frac{1}{n} \sum_{i=1}^n E[x_i^2] + E[m^2] - 2E[mx_i] = \frac{1}{n} \cdot n \left(\frac{\sigma^2}{n} + \mu^2 - 2\frac{\sigma^2}{n} - 2\mu^2 + \sigma^2 + \mu^2\right) = \frac{n-1}{n}\sigma^2$$

Tuttavia si osserva che  $s_n^2$  non è uno stimatore imparziale, ma solo asintoticamente (e.g. se  $n \rightarrow +\infty$  si ha che  $\frac{n-1}{n} \sim 1$ ) dunque per ovviare a ciò si moltiplica il nostro stimatore  $s_n$  per il fattore correttivo  $\frac{n}{n-1}$  trasformandolo nello stimatore  $s_{n-1}^2$

$$s_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2 \quad (2.6)$$

#### Osservazione:-

Una domanda che può sorgere è quale delle due stimatori sia più giusto: si tratta di una domanda non banale, tuttavia sappiamo che per un numero molto grande di campioni sono asintoticamente imparziali entrambi, mentre per campioni piccoli la domanda non è scontata. Dai grafici riportati nelle dispense del Baldini si osserva che per  $n > 10$  i due stimatori sono compatibili entro il 10% tuttavia per un numero di campioni ancora più basso la domanda è ancora presente.

Si deve osservare che il fattore correttivo fa in modo per campioni piccolo lo stimatore  $s_n^2$  sia più corretto di quello imparziale, oltre al fatto che rende la coda di destra ancora più pronunciata. Un'altra considerazione che si può fare deriva dal fatto che lo stimatore  $s_{n-1}^2$  è uno stimatore imparziale per  $\sigma^2$  ma non per  $\sigma$

Dunque, dopo essersi dilungati anche fin troppo sugli estimatori statistici, ritorniamo alla domanda iniziale: come posso scrivere il risultato di una misura?



Il candidato ideale per il valore centrale sarebbe la media campionaria, mentre per l'incertezza associata non possiamo utilizzare la varianza campione, siccome essa rappresenta le fluttuazioni della singola misura (rispetto al valore medio) ma non quella del valore centrale.

Tuttavia la media campione, essendo somma di variabili aleatorie, sarà anch'essa una variabile casuale, dunque se siamo interessati alle sue fluttuazioni

$$\text{Var}(m) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n x_i\right) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Banalmente, la deviazione standard della media risulta quindi essere

$$\sigma_m = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \quad (2.7)$$

e utilizzando la stima della varianza campione possiamo dire che

$$s_m = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - m)^2} \quad (2.8)$$

dunque la nostra misura possiamo scriverla come

$$x = m \pm s_m$$

## 2.2 Covarianza e correlazione campione

Supponiamo di avere una serie di campionamenti  $x_i$  ed  $y_i$  con  $i \in I = 1, \dots, n$  di due variabili aleatorie  $x, y$ , ovvero sia abbiamo di fatto misurato  $n$  coppie ordinate  $(x_i, y_i)$ . Se indichiamo le medie campionarie, rispettivamente, dalla variabile  $x$  e della variabile  $y$  con

$$m_x = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{e} \quad m_y = \frac{1}{n} \sum_{i=1}^n y_i$$

possiamo stimare la covarianza, proprio come abbiamo fatto con la varianza, tramite la covarianza del campione

### Definizione 2.3: Covarianza campione

$$q_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - m_x)(y_i - m_y) \quad (2.9)$$

### Osservazione:-

Il termine  $n-1$  deriva sempre dalla cosiddetta "correzione di Bessel" ovvero quello che abbiamo fatto prima riguardo allo stimatore statistico della varianza campione

Dunque la stima delle correlazione  $r_{xy}$  campionaria si scrive tramite le stime della varianza campione delle due variabili  $x$  e  $y$ :

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m_x)^2 \quad \text{e} \quad s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - m_y)^2$$

dunque

$$r_{xy} = \frac{q_{xy}}{s_x s_y} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - m_x)(y_i - m_y)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - m_x)^2} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - m_y)^2}} = \frac{\sum_{i=1}^n (x_i - m_x)(y_i - m_y)}{\sqrt{\sum_{i=1}^n (x_i - m_x)^2} \sqrt{\sum_{i=1}^n (y_i - m_y)^2}} \quad (2.10)$$

e talvolta viene chiamato anche come *coefficiente di correlazione lineare* o *indice di correlazione di Pearson*: infatti uno dei metodi più semplici per verificare se sussiste una correlazione tra due variabili di un campione è quello di riportare i dati in un grafico di dispersione. La cosa interessante è il fatto che è possibile, in questa maniera, andare a scrivere la covarianza campione fra due misure come

$$q_{xy} = r_{xy} s_x s_y \quad (2.11)$$

## 2.3 Media e varianza di una funzione a variabili casuali

Se abbiamo una generica funzione  $f(x)$  come possiamo stimare media e varianza? In generale ci aspettiamo che i valori di  $x$  tendano a concentrarsi maggiormente attorno al valore medio, dunque possiamo partire dall'approssimare la funzione tramite sviluppo di Taylor al primo ordine:

$$f(x) \approx f(\mu) + \left. \frac{df}{dx} \right|_{x=\mu} (x - \mu)$$

quindi, si ha che

$$\mu_f = E[f(x)] \approx E \left[ f(\mu) + \left. \frac{df}{dx} \right|_{x=\mu} (x - \mu) \right] = E[f(\mu)] + E \left[ \left. \frac{df}{dx} \right|_{x=\mu} (x - \mu) \right] = f(\mu) + \left. \frac{df}{dx} \right|_{x=\mu} E[x - \mu] = f(\mu)$$

dunque abbiamo che

$$\mu_f \approx f(\mu)$$

La stima della varianza è leggermente più complicata, ma nel caso ad una singola variabile rimane comunque banale siccome

$$\sigma_f^2 = E[(f(x) - f(\mu))^2] \approx E \left[ \left( \left. \frac{df}{dx} \right|_{x=\mu} (x - \mu) \right)^2 \right] = \left( \left. \frac{df}{dx} \right|_{x=\mu} \right)^2 E[(x - \mu)^2] = \left( \left. \frac{df}{dx} \right|_{x=\mu} \right)^2 \sigma_x^2$$

### Definizione 2.4: Media e varianza di una generica funzione $f(x)$

Sia  $f(x)$  una generica funzione e  $x$  variabile aleatoria, il valore medio di  $f(x)$  e la varianza sono rispettivamente

$$\mu_f \approx f(\mu) \qquad \sigma_f^2 \approx \left( \left. \frac{df}{dx} \right|_{x=\mu} \right)^2 \sigma_x^2$$

Nel caso di funzioni che dipendono da un certo numero di variabili casuali  $x_1, x_2, \dots, x_n$  con medie  $\mu_1, \mu_2, \dots, \mu_n$  e varianza  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$  la questione diventa più complicata, siccome ritorna in gioco l'indipendenza statistica fra variabili aleatorie.

Supponiamo ad esempio di guardare il caso di una funzione  $f(x_1, x_2)$  ovvero dipendente da due variabili aleatorie e basta. Innanzitutto, si osserva che attorno al punto medio possiamo approssimare la funzione tramite lo sviluppo di Taylor in due variabili troncato al primo ordine:

$$f(x_1, x_2) \approx f(\mu_1, \mu_2) + \left. \frac{\partial f}{\partial x_1} \right|_{x_1=\mu_1, x_2=\mu_2} (x_1 - \mu_1) + \left. \frac{\partial f}{\partial x_2} \right|_{x_1=\mu_1, x_2=\mu_2} (x_2 - \mu_2)$$

dunque possiamo sviluppare il valore medio al primo ordine nella seguente maniera:

$$\begin{aligned} \mu_f &= E[f(x_1, x_2)] \\ &= E[f(\mu_1, \mu_2)] + \left. \frac{\partial f}{\partial x_1} \right|_{x_1=\mu_1, x_2=\mu_2} E[x_1 - \mu_1] + \left. \frac{\partial f}{\partial x_2} \right|_{x_1=\mu_1, x_2=\mu_2} E[x_2 - \mu_2] \\ &= f(\mu_1, \mu_2) \end{aligned}$$

per la varianza la questione è un po' più complicata siccome:

$$\begin{aligned}
\text{Var}(f(x_1, x_2)) &= E[(f(x_1, x_2) - f(\mu_1, \mu_2))^2] = E \left[ \left( \frac{\partial f}{\partial x_1} \Big|_{x_1=\mu_1, x_2=\mu_2} (x_1 - \mu_1) + \frac{\partial f}{\partial x_2} \Big|_{x_1=\mu_1, x_2=\mu_2} (x_2 - \mu_2) \right)^2 \right] \\
&= \left( \frac{\partial f}{\partial x_1} \Big|_{x_1=\mu_1, x_2=\mu_2} \right)^2 E[(x_1 - \mu_1)^2] - 2 \frac{\partial f}{\partial x_1} \Big|_{x_1=\mu_1, x_2=\mu_2} \frac{\partial f}{\partial x_2} \Big|_{x_1=\mu_1, x_2=\mu_2} E[(x_1 - \mu_1)(x_2 - \mu_2)] + \left( \frac{\partial f}{\partial x_2} \Big|_{x_1=\mu_1, x_2=\mu_2} \right)^2 E[(x_2 - \mu_2)^2] = \\
&= \left( \frac{\partial f}{\partial x_1} \Big|_{x_1=\mu_1, x_2=\mu_2} \right)^2 \text{Var}(x_1) + \left( \frac{\partial f}{\partial x_2} \Big|_{x_1=\mu_1, x_2=\mu_2} \right)^2 \text{Var}(x_2) - 2 \frac{\partial f}{\partial x_1} \Big|_{x_1=\mu_1, x_2=\mu_2} \frac{\partial f}{\partial x_2} \Big|_{x_1=\mu_1, x_2=\mu_2} \text{Cov}(x_1, x_2)
\end{aligned}$$

Naturalmente si semplifica enormemente la formula se  $x_1, x_2$  sono variabili fra loro indipendenti, ergo  $\text{Cov}(x_1, x_2) = 0$ . Si può generalizzare a funzioni dipendenti da  $n$  parametri indipendenti che

$$\sigma_f^2 \approx \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i} \Big|_{x_1=\mu_1, \dots, x_n=\mu_n} \right)^2 \sigma_i^2 \quad (2.12)$$

Questo argomento verrà ripreso più avanti.

## 2.4 Propagazione dell'errore statistico

Nella teoria dei campioni si è visto che l'incertezza di misura ha il significato di stima della deviazione standard della distribuzione generatrice. Siccome sappiamo calcolare, con le formule che abbiamo visto nella sezione precedente media la deviazione standard di una funzione arbitraria di variabili casuali, date le deviazioni standard delle variabili stesse, siamo adesso in grado di propagare gli errori in maniera statisticamente corretta: supponiamo di avere dunque  $n$  grandezze misurare  $x_i = \hat{x}_i \pm \sigma_i$  ed una generica funzione  $f(x_1, \dots, x_n)$  e partiamo dal caso più semplice, ovvero quello in cui le grandezze di partenza sono tutte indipendenti. La formula, nella *forma*, è equivalente alla 2.12, infatti

**Definizione 2.5: Propagazione dell'errore statistico con variabili indipendenti**

$$\sigma_f^2 \approx \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i} \Big|_{x_1=\hat{x}_1, \dots, x_n=\hat{x}_n} \right)^2 \sigma_i^2 \quad (2.13)$$

sebbene, per quanto riguarda la forma, come ho già detto questa scrittura è equivalente con la 2.12, dal punto di vista logico non lo sono siccome nella 2.12 le  $\sigma_i$  rappresentano in generale le nostre migliori stime delle deviazioni standard delle distribuzioni generatrici e non i valori calcolati a partire dalla forma analitica delle funzioni di distribuzione stesse.

Possiamo generalizzare ancora di più: infatti se noi effettuiamo la stima dell'indice di correlazione possiamo dire che

**Definizione 2.6: Propagazione dell'errore statistico con variabili dipendenti**

$$\sigma_f^2 \approx \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_1=\hat{x}_1, \dots, x_n=\hat{x}_n} \frac{\partial f}{\partial x_j} \Big|_{x_1=\hat{x}_1, \dots, x_n=\hat{x}_n} r_{ij} \sigma_i \sigma_j \quad (2.14)$$

## Capitolo 3

# Distribuzioni uni-variate di uso comune

### 3.1 La distribuzione binomiale

Supponiamo di considerare un esperimento che abbia solamente **due esiti possibili distinti**,  $E_1$  ed  $E_2$ , con probabilità pari a  $p$  e  $1 - p$  rispettivamente. La domanda che può sorgere spontanea è la seguente: *qual è la probabilità di ottenere  $k$  volte l'esito  $E_1$  ripetendo l'esperimento  $n$  volte e assumendo che le realizzazioni siano indipendenti?*

Inizialmente potremmo essere tentati di dire  $p^k(1-p)^{n-k}$  ma noi siamo interessati ad una qualunque combinazione degli esiti  $E_1$  ed  $E_2$ , fintanto che avvengano  $k$  occorrenze dell'evento  $E_1$ , pertanto il nostro risultato va moltiplicato per il numero possibile delle combinazioni, pari a  $\binom{n}{k}$  (che coincide con il numero di sottoinsiemi di  $k$  che è possibile formare in un insieme di  $n$  elementi). Possiamo sfruttare il fatto che le probabilità per eventi disgiunti si sommano e possiamo scrivere la probabilità cercata come:

$$\mathcal{B}(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \quad (3.1)$$

#### Osservazione:-

Il punto e virgola nella parentesi degli argomenti di una funzione di distribuzione separa le variabili casuali dai parametri esterni che sono determinati univocamente a priori dal problema

#### 3.1.1 Normalizzazione, media e varianza

Si dimostra facilmente che la condizione di normalizzazione è rispettata.

##### **Teorema 3.1** Condizione di normalizzazione della distribuzione binomiale

La distribuzione binomiale rispetta la condizione di normalizzazione, dunque:

$$\sum_{k=0}^n \mathcal{B}(k; n, p) = 1 \quad (3.2)$$

**Dimostrazione:** si osserva banalmente che la sommatoria di  $k$  fino ad  $n$  rappresenta lo sviluppo della potenza  $n$ -esima del binomio  $p + (1 - p)$ :

$$\sum_{k=0}^n \mathcal{B}(k; n, p) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$$

La tesi è dunque ottenuta



Enunciamo adesso qualche fatto facilmente dimostrabile

**Teorema 3.2** Valore atteso della distribuzione  $\mathcal{B}(k; n, p)$ 

Il valore atteso della distribuzione binomiale è

$$\mu = np \quad (3.3)$$

**Dimostrazione:**

$$\mu = \sum_{k=0}^n k \mathcal{B}(k; n, p) = \sum_{k=0}^n k \cdot \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=0}^n \frac{n(n-1)!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$

Si osserva che possiamo far partire la sommatoria da  $k = 1$  siccome il primo termine non contribuisce alla somma ed effettuiamo un banale cambio di indice: definiamo  $h = k - 1$  e dunque la sommatoria si fermerà a  $m = n - 1$ :

$$\sum_{k=1}^n np \frac{(n-1)!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} = np \sum_{h=0}^m \frac{m!}{h!(m-h)!} p^h (1-p)^{m-h}$$

Si riconosce nella sommatoria di  $\sum_{h=0}^m \frac{m!}{h!(m-h)!} p^h (1-p)^{m-h}$  la condizione di normalizzazione, dunque

$$\mu = np \sum_{h=0}^m \frac{m!}{h!(m-h)!} p^h (1-p)^{m-h} = np$$

☺

**Teorema 3.3** Varianza della distribuzione  $\mathcal{B}(k; n, p)$ 

La varianza della distribuzione binomiale risulta essere

$$\sigma^2 = np(1-p) \quad (3.4)$$

**Dimostrazione:** si utilizza la relazione 1.25 per stimare la varianza:

$$\begin{aligned} E[k^2] &= \sum_{k=0}^n k^2 \mathcal{B}(k; n, p) = \sum_{k=0}^n k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^n k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n np \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} = np \sum_{h=0}^m (h+1) \frac{m!}{h!(m-k)!} p^h (1-p)^{m-k} \\ &= np \sum_{h=0}^m h \frac{m!}{h!(m-k)!} p^h (1-p)^{m-k} + np \sum_{h=0}^m \frac{m!}{h!(m-k)!} p^h (1-p)^{m-k} = np(mp+1) = np(np-p+1) \end{aligned}$$

dunque, si ha che

$$\text{Var}(k) = E[k^2] - \mu^2 = np(np-p+1) - n^2 p^2 = np(1-p)$$

La dimostrazione è dunque conclusa.

☺

La distribuzione binomiale è asimmetrica generalmente, dunque ha senso chiedere quanto vale la *skewness*. Ricaviamo un'identità molto utile:

**Lemna 3.1.1** Momenti di ordine superiore per la  $\mathcal{B}(k; n, p)$ 

I momento di ordine superiore rispettano la seguente identità:

$$E[k^{m+1}] = p(1-p) \frac{d}{dp} E[k^m] + np E[k^m] \quad (3.5)$$

**Dimostrazione:**

$$\begin{aligned}
\frac{d}{dp}E[k^m] &= \frac{d}{dp} \sum_{k=0}^n k^m \binom{n}{k} p^k (1-p)^{n-k} = \\
&= \sum_{k=0}^n k^m \binom{n}{k} [kp^{k-1}(1-p)^{n-k} - p^k(n-k)(1-p)^{n-k-1}] \\
&= \sum_{k=0}^n k^{m+1} \binom{n}{k} p^{k-1}(1-p)^{n-k} - \sum_{k=0}^n k^m \binom{n}{k} \frac{n-k}{1-p} p^k (1-p)^{n-k} = \\
&= \sum_{k=0}^n k^{m+1} \binom{n}{k} \frac{1}{p} p^k (1-p)^{n-k} - \sum_{k=0}^n k^m \binom{n}{k} \frac{n}{1-p} p^k (1-p)^{n-k} + \sum_{k=0}^n k^{m+1} \binom{n}{k} p^k (1-p)^{n-k} = \\
&= \frac{1}{p} \sum_{k=0}^n k^{m+1} + \sum_{k=0}^n k^{m+1} \binom{n}{k} p^k (1-p)^{n-k} = \\
&= \left(\frac{1}{p} + \frac{1}{1-p}\right) E[k^{m+1}] - \frac{n}{1-p} E[k^m]
\end{aligned}$$

dunque otteniamo che

$$\frac{d}{dp}E[k^m] + \frac{n}{1-p}E[k^m] = \frac{1}{p(1-p)}E[k^{m+1}] \implies E[k^{m+1}] = p(1-p)\frac{d}{dp}E[k^m] + npE[k^m]$$

☺

Tramite questo semplice lemma, possiamo calcolare i momenti superiori al secondo:

$$E[k^3] = p(1-p)\frac{d}{dp}E[k^2] + npE[k^2] = p(1-p)\frac{d}{dp}[np(np-p+1)] + np[p(n-1)+1] = np(1-p)(1-2p) + 3n^2p^2(1-p) + n^3p^3 \quad (3.6)$$

Dunque, il momento centrale di ordine 3 risulta essere pari a:

$$\mu_3 = E[(k-\mu)^3] = E[k^3] - 3\mu\sigma^2 - \mu^3 = np(1-p)(1-2p) + 3n^2p^2(1-p) + n^3p^3 - 3n^2p^2(1-p) - n^3p^3 = np(1-p)(1-2p) \quad (3.7)$$

dunque si ha che la *skewness*  $\gamma_1$ :

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{1-2p}{\sqrt{np(1-p)}}$$

## 3.2 La distribuzione di Poisson

Formalmente la distribuzione di Poisson può essere ottenuta come limite della binomiale per  $p \rightarrow 0$  e  $n \rightarrow \infty$ , facendo in modo che  $np = \mu$ . Ora, prima di procedere a mostrarlo formalmente, è educativo studiare le caratteristiche di un *processo poissoniano*, prima menzionato, e ricavare in maniera per lo più euristica la forma della distribuzioni. Facciamo le seguenti ipotesi sul processo studiato:

- ① **indipendenza:** gli eventi elementari sono indipendenti, ovvero il verificarsi di un evento ad un determinato istante non influenza la probabilità che un altro evento si verifichi (o non si verifichi) ad un istante successivo;
- ② **stazionarietà:** il numero medio di eventi per unità di tempo è lo stesso in qualsiasi intervallo;
- ③ **non simultaneità:** non si possono verificare due o più eventi nello stesso istante.

Consideriamo un intervallo  $\Delta t$  che dividiamo in  $n$  intervalli più piccoli di lunghezza  $dt = \frac{\Delta t}{n}$ : il numero medio di eventi per unità di tempo  $\lambda$  è il parametro che determina il numero medio  $\mu = \lambda \Delta t$  di eventi nell'intervallo, da cui segue che il numero medio di eventi in uno qualunque degli intervallini è dato esplicitando  $\lambda dt$  nella precedente formula, ricordando che  $\Delta t = n dt$ ,

$$\mu = \lambda n dt \implies \lambda dt = \frac{\mu}{n}.$$

Se chiamiamo  $p$  la probabilità che si verifichi un evento all'interno di un intervallino e ricordando l'ipotesi di non simultaneità, abbiamo che all'interno di un  $dt$  si possono verificare 0 o 1 evento, il che ci consente di dire che la probabilità che si verifichi un solo evento è dato da

$$0 \times (1 - p) + 1 \times p = \frac{\mu}{n} = \lambda dt,$$

dove fra il primo termine e il successivo abbiamo usato la definizione combinatoriale della probabilità, affermando che la probabilità che un evento avvenga in un intervallo è pari a  $\frac{\mu}{n}$ , ovvero il numero medio di eventi nell'intervallo  $\Delta t$  diviso il numero di intervalli. Abbiamo quindi ottenuto che  $\lambda dt$  è esattamente pari alla *probabilità che un evento si verifichi esattamente in uno qualsiasi dei sottointervalli*. Questo ci consente subito di determinare la forma della distribuzione del processo: possiamo infatti chiederci quale sia la probabilità  $P(0; \mu)$  di osservare 0 eventi nell'intervallo  $\Delta t$  quando in media sappiamo che se ne verificano  $\mu$ ? Ragionando come già fatto per un processo a due esiti, abbiamo che la probabilità che non si verifichi in un evento in un intervallino è proprio pari a  $(1 - p)^n$ , pertanto

$$P(0; \mu) = \lim_{n \rightarrow \infty} (1 - p)^n = \lim_{n \rightarrow \infty} (1 - \frac{\mu}{n})^n = e^{-\mu}.$$

La probabilità di osservare un solo evento? Il calcolo è identico, tuttavia bisogna considerare che l'occorrenza dell'evento può avvenire in uno qualunque degli intervalli, quindi è necessario moltiplicare per  $n$  per tenere conto di ciò: ne consegue che

$$P(1; \mu) = \lim_{n \rightarrow \infty} np(1 - p)^{n-1} = \lim_{n \rightarrow \infty} n \frac{\mu}{n} (1 - \frac{\mu}{n})^{n-1} = \lim_{n \rightarrow \infty} \mu (1 - \frac{\mu}{n})^n (1 - \frac{\mu}{n}) = \mu e^{-\mu}.$$

Ma osserviamo proprio che stiamo proprio riderivando la distribuzione binomiale nel caso di  $n$  tendente all'infinito. Questo fissa che la distribuzione poissoniana ha la seguente forma

$$\mathcal{P}(k; \mu) = \frac{\mu^k}{k!} e^{-\mu}. \quad (3.8)$$

### 3.2.1 La poissoniana come limite della distribuzione binomiale

#### **Teorema 3.4** Distribuzione di Poisson come limite della distribuzione binomiale

La distribuzione poissoniana può essere ottenuta come limite della distribuzione binomiale con  $n \rightarrow \infty$ ,  $p \rightarrow 0$  in modo che la media  $\mu = np$  rimanga costante.

$$\lim_{n \rightarrow \infty} \mathcal{B}(k; n, p) = \mathcal{P}(k; \mu) \quad (3.9)$$

**Dimostrazione:** Osserviamo che

$$\lim_{n \rightarrow \infty} \mathcal{B}(k; n, p) = \lim_{n \rightarrow \infty} \binom{n}{k} p^k (1 - p)^{n-k},$$

e, usando l'equazione 3.3, possiamo scrivere che  $p = \frac{\mu}{n}$ , da cui abbiamo che

$$\lim_{n \rightarrow \infty} \mathcal{B}(k; n, p) = \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k}.$$

Osservando che  $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\dots(n-k+1)}{k!}$  abbiamo che

$$\lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k} = \lim_{n \rightarrow \infty} \frac{n(n-1)\dots(n-k+1)}{n^k} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^{n-k}.$$

Dobbiamo adesso procedere con *cautela*: infatti, noi vorremmo dire qualcosa su  $k$  ma non si tratta di un "numero", bensì di una variabile aleatoria. Questo vuol dire che, a priori, può assumere tutti i valori compresi fra 0 e  $n$ , quindi non possiamo, a priori, servirci del suo valore per determinare il comportamento asintotico (per  $n \rightarrow \infty$ ) di questo limite. Possiamo, tuttavia, introdurre una variabile aleatoria ridotta

$$\xi = \frac{k - np}{n}, \quad (3.10)$$

che quantifica quanto si discosta  $k$  dal valore atteso della nostra distribuzione binomiale, confrontandolo con  $n$ . Per il teorema di Chebyshev, sappiamo che è poco "probabile" che la  $k$  si discosti dal valore atteso più una quantità molto grande di  $\sigma = \sqrt{np(1-p)}$ , conseguentemente

$$-\frac{\sqrt{np(1-p)}}{n} \leq \frac{k-np}{n} \leq \frac{\sqrt{np(1-p)}}{n}$$

$$\downarrow \qquad \qquad \qquad \downarrow$$

$$0 \qquad \qquad \qquad 0$$

quindi, per confronto, abbiamo che  $\xi \rightarrow 0$ . Riscrivendo il limite in termini della nuova variabile e raccogliendo  $n$  da ogni fattore presente al numeratore, abbiamo che

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathcal{B}(k; n, p) &= \lim_{n \rightarrow \infty} \frac{n(n-1)\dots(n-k+1)}{n^k} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^{n-k} = \\ &= \lim_{n \rightarrow \infty} \frac{n^k(1 - \frac{1}{n})(1 - \frac{2}{n})\dots(1 - p - \xi + \frac{1}{n})}{n^k} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^{n(1-p-\xi)} = \lim_{n \rightarrow \infty} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^k, \end{aligned}$$

dove si è usato dalla 3.10 che  $\frac{k}{n} = \xi + p$  e il fatto che  $p \rightarrow 0, \xi \rightarrow 0$  per giustificare il fatto che  $(1 - \frac{1}{n})(1 - \frac{2}{n})\dots(1 - p - \xi + \frac{1}{n}) \sim 1$  e  $(1 - \frac{\mu}{n})^{n(1-p-\xi)} \sim (1 - \frac{\mu}{n})^n$ . Ricordando adesso che  $\lim_{n \rightarrow \infty} (1 - \frac{\mu}{n})^n \rightarrow e^{-\mu}$ , possiamo affermare il limite precedente converge a

$$\lim_{n \rightarrow \infty} \mathcal{B}(k; n, p) \rightarrow \frac{\mu^k}{k!} e^{-\mu}. \quad (3.11)$$

☺

#### Osservazione:-

Dall'espressione della poissoniana si vedono due sostanziali differenze fra la distribuzione binomiale e quella di Poisson:

- ① la distribuzione binomiale dipende da due parametri,  $n$  e  $p$ , che contribuiscono entrambi a determinare l'espressione (e, conseguentemente, il valore) del valore medio e la varianza della distribuzione. D'altro canto, la distribuzione di Poisson dipende esclusivamente da un parametro, ovvero la media che coincide pure con la varianza;
- ② nella distribuzione binomiale la variabile  $k \leq n$ , mentre nella distribuzione poissoniana può assumere qualunque valore intero da 0 a  $\infty$ . Questo è un buon indicatore per capire se un processo è *poissoniano* o *binomiale*: il numero delle occorrenze è fissato? A seconda della risposta, è lecito modellizzare un fenomeno con una delle distribuzioni invece che l'altra (poi, chiaramente, la natura non è né perfettamente poissoniana e né perfettamente binomiale).

### 3.2.2 Normalizzazione, media e varianza

Siccome la distribuzione di Poisson è ottenuta come limite della binomiale, la media  $np \rightarrow \mu$  e la varianza  $np(1-p) \rightarrow \mu$ , quindi si potrebbe anche non verificare esplicitamente, partendo dalla forma della distribuzione, i loro valori. In ogni caso, non è difficile ricavarli dalla definizione di media e momento: procediamo con ordine, partendo dalla condizione di normalizzazione.

#### Teorema 3.5 Condizione di normalizzazione della distribuzione poissoniana

La distribuzione di Poisson rispetta la condizione di normalizzazione, ovvero

$$\sum_{k=0}^{\infty} \mathcal{P}(k; \mu) = 1.$$

**Dimostrazione:** Procediamo con il calcolo:

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



dove abbiamo riconosciuto la serie esponenziale nell'espressione (si ricorda al lettore che  $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ ). ☺

Passiamo alla media:

### Teorema 3.6 Valore atteso della distribuzione poissoniana

Il valore atteso della distribuzione poissoniana  $\mathcal{P}(k; \mu)$  è pari a  $\mu$

**Dimostrazione:** Sappiamo, per definizione, che la media è data da

$$E[k] = \sum_{k=0}^{\infty} k \mathcal{P}(k; \mu) = e^{-\mu} \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} = e^{-\mu} \sum_{k=1}^{\infty} k \frac{\mu^k}{k!},$$

dove si è osservato che il termine con  $k = 0$  non contribuisce alla somma, per cui abbiamo fatto iniziare la somma da  $k = 1$ . A questo punto, poniamo  $k = h + 1$  dunque

$$E[k] = e^{-\mu} \sum_{h=0}^{\infty} \frac{\mu^{h+1}}{h!} = \mu e^{-\mu} \sum_{h=0}^{\infty} \frac{\mu^h}{h!} = \mu e^{-\mu} e^{\mu} = \mu.$$

☺

Passiamo ai momenti di ordine superiore

### Lemma 3.2.1 Relazione ricorsiva fra i momenti algebrici di ordine superiore per la $\mathcal{P}(k; \mu)$

Abbiamo che i momenti algebrici di ordine superiore della  $\mathcal{P}(k; \mu)$  soddisfano la seguente relazione ricorsiva

$$E[k^{m+1}] = \mu \left( \frac{d}{d\mu} E[k^m] + E[k^m] \right). \quad (3.12)$$

**Dimostrazione:** Partiamo dal calcolare  $\frac{d}{d\mu} E[k^m]$ :

$$\begin{aligned} \frac{d}{d\mu} E[k^m] &= \frac{d}{d\mu} \sum_{k=0}^{\infty} k^m \frac{\mu^k}{k!} e^{-\mu} = \sum_{k=0}^{\infty} \frac{k^m}{k!} \frac{d}{d\mu} [\mu^k e^{-\mu}] = \sum_{k=0}^{\infty} \frac{k^m}{k!} [k \mu^{k-1} e^{-\mu} - \mu^k e^{-\mu}] = \\ &= \sum_{k=0}^{\infty} k^{m+1} \frac{\mu^{k-1}}{k!} e^{-\mu} - \sum_{k=0}^{\infty} k^m \frac{\mu^k}{k!} e^{-\mu} = \sum_{k=0}^{\infty} k^{m+1} \frac{\mu}{\mu} \frac{\mu^{k-1}}{k!} e^{-\mu} - \sum_{k=0}^{\infty} k^m \frac{\mu^k}{k!} e^{-\mu} = \\ &= \frac{1}{\mu} \sum_{k=0}^{\infty} k^{m+1} \frac{\mu^k}{k!} e^{-\mu} - \sum_{k=0}^{\infty} k^m \frac{\mu^k}{k!} e^{-\mu} = \frac{1}{\mu} E[k^{m+1}] - E[k^m]. \end{aligned}$$

Ma allora abbiamo ottenuto la tesi, siccome ciò implica, con un banale riarrangiamento, che

$$E[k^{m+1}] = \mu \left( \frac{d}{d\mu} E[k^m] + E[k^m] \right).$$

☺

Tramite questo lemma, è possibile calcolare la Varianza

### Teorema 3.7 Varianza della distribuzione $\mathcal{P}(k; \mu)$

La varianza della distribuzione di Poisson  $\mathcal{P}(k; \mu)$  è pari a

$$\sigma^2 = \mu$$

**Dimostrazione:** Calcoliamo il momento algebrico di ordine 2 della distribuzione

$$E[k^2] = \mu \left( \frac{d}{d\mu} E[k] + E[k] \right) = \mu \left( \frac{d}{d\mu} (\mu) + \mu \right) = \mu(1 + \mu) = \mu^2 + \mu,$$

a questo punto sappiamo che

$$\sigma^2 = E[k^2] - \mu^2 = \mu^2 + \mu - \mu^2 = \mu.$$

☺

Sempre tramite questo lemma è possibile calcolare il coefficiente di *skewness*: si ha, infatti, che

$$E[k^3] = \mu \left( \frac{d}{d\mu} E[k^2] + E[k^2] \right) = \mu \left( \frac{d}{d\mu} (\mu^2 + \mu) + \mu \right) = \mu(2\mu + 1 + \mu^2 + \mu) = \mu^3 + 3\mu^2 + \mu.$$

e, usando la 1.31, abbiamo che

$$\mu_3 = E[k^3] - 3\mu\sigma^2 - \mu^3 = \mu^3 + 3\mu^2 + \mu - 3\mu^2 - \mu^3 = \mu,$$

da cui possiamo concludere che

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{1}{\sqrt{\mu}}. \quad (3.13)$$

Era possibile ottenerla sempre come limite della distribuzione binomiale per  $p \rightarrow 0$  ricordando che  $\mu = np$ .

### 3.2.3 Somma di variabili poissoniane

Consideriamo due variabili Poissoniane (indipendenti)  $l$  e  $m$  con medie  $\mu_l$  e  $\mu_m$  rispettivamente. Vogliamo adesso capire come si distribuisce la loro somma  $k = m + l$ : formalmente, per ogni valore di  $k$ , dobbiamo sommare per tutte le coppie  $(l, m)$  tali che la loro somma è proprio pari a  $k$ , moltiplicando (essendo indipendenti) la probabilità che esca  $l$  e la probabilità che esca  $m$ . Questo si traduce in

$$\sum_{l=0}^k \mathcal{P}(l; \mu_l) \mathcal{P}(k-l; \mu_m),$$

dove abbiamo usato il fatto che  $k = l + m \implies m = k - l$ , per ricondurci ad una singola sommatoria. Abbiamo allora che

$$P(k) = \sum_{l=0}^k \frac{\mu_l^l}{l!} e^{-\mu_l} \cdot \frac{\mu_m^{k-l}}{(k-l)!} e^{-\mu_m} = e^{-(\mu_l + \mu_m)} \sum_{l=0}^k \frac{\mu_l^l \mu_m^{k-l}}{l!(k-l)!} = e^{-(\mu_l + \mu_m)} \sum_{l=0}^k \frac{k!}{k!} \frac{\mu_l^l \mu_m^{k-l}}{l!(k-l)!} = \frac{e^{-(\mu_l + \mu_m)}}{k!} \sum_{l=0}^k \binom{k}{l} \mu_l^l \mu_m^{k-l},$$

dove nell'ultimo passaggio abbiamo riconosciuto che  $\frac{k!}{l!(k-l)!} = \binom{k}{l}$ . Osserviamo che l'ultima sommatoria coincide con la potenza di un binomio, dunque

$$P(k) = \frac{(\mu_l + \mu_m)^k}{k!} e^{-(\mu_m + \mu_l)}.$$

Abbiamo scoperto che la somma di due variabili poissoniane è ancora una variabile poissoniana.

## 3.3 La distribuzione uniforme

La distribuzione uniforme è l'esempio più semplice di una funzione di distribuzione di variabile casuale continua: la densità di probabilità è costante entro un intervallo finito e nulla fuori

$$u(x; a, b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & x < a; x > b \end{cases}$$

### 3.3.1 Normalizzazione, media, varianza e asimmetria

**Teorema 3.8** Condizione di normalizzazione della distribuzione uniforme

La distribuzione uniforme rispetta la condizione di normalizzazione

$$\int_{-\infty}^{\infty} u(x; a, b) dx = 1$$

**Dimostrazione:** Basta svolgere questo semplice integrale

$$\int_{-\infty}^a u(x; a, b) dx + \int_a^b u(x; a, b) dx + \int_b^{+\infty} u(x; a, b) dx = 0 + \int_a^b u(x; a, b) dx + 0 = \int_a^b \frac{1}{b-a} dx = (b-a) \frac{1}{b-a} = 1.$$

☺

Passiamo alla media

**Teorema 3.9** Valore atteso della distribuzione uniforme

Il valore atteso della distribuzione uniforme è pari a

$$\mu = \frac{b+a}{2}$$

**Dimostrazione:** Per definizione la media è data dal valore di aspettazione della distribuzione:

$$\int_{-\infty}^{\infty} x u(x; a, b) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left. \frac{x^2}{2} \right|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{(b+a)(\cancel{b-a})}{2(\cancel{b-a})} = \frac{b+a}{2}.$$

☺

**Teorema 3.10** Varianza della distribuzione uniforme

La varianza della distribuzione uniforme è pari a

$$\sigma^2 = \frac{(b-a)^2}{12}$$

**Dimostrazione:** Usiamo la relazione 1.25, quindi procediamo a calcolare  $E[x^2]$ :

$$E[x^2] = \int_{-\infty}^{\infty} x^2 u(x; a, b) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{(b-a)} \left. \frac{x^3}{3} \right|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} = \frac{b^2 + ab + a^2}{3},$$

da cui

$$\sigma^2 = E[x^2] - \mu^2 = \frac{b^2 + ab + a^2}{3} - \frac{(b+a)^2}{4} = \frac{4b^2 + 4ab + 4a^2 - 3b^2 - 3a^2 - 6ab}{12} = \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}.$$

☺

Il coefficiente di *skewness* risulta nullo siccome la distribuzione è simmetria.

### 3.4 La distribuzione esponenziale

La distribuzione esponenziale è una distribuzione che, spesso, è incontrata per la prima volta quando ci si sofferma a studiare la distribuzione del tempo  $t$  che intercorre fra due eventi di un processo poissoniano. In primis si osserva che  $t$  è una variabile aleatoria: se così non fosse, allora tutti gli eventi sarebbero equispaziati, il che implica che il numero di eventi in un intervallo  $\Delta t$  non fluttuerebbe ma sarebbe determinato a priori e ciò rappresenterebbe un assurdo col fatto che stiamo considerando un processo poissoniano. Ora riprendiamo l'idea del suddividere l'intervallo  $\Delta t$  in tanti intervallini: osserviamo che la media del tempo fra due eventi è allora data dal rapporto

$\frac{\Delta t}{\mu} = \frac{\Delta t}{\lambda \Delta t} = \frac{1}{\lambda}$ , dove si è usato che  $\mu = \lambda \Delta t$ . Tenendo a mente che stiamo lavorando con un processo poissoniano, la probabilità infinitesima che l'evento successivo si verifichi entro un intervallino (infinitesimo) di durata  $dt$  ad una distanza temporale  $t$  è dato dal prodotto della probabilità che non si verifichi un evento all'istante  $t$  con la probabilità che l'evento si verifichi un intervallo  $dt$ , pari, rispettivamente, a  $e^{-\lambda t}$  (si ripensi alla probabilità di osservare 0 eventi in un processo poissoniano) e  $\lambda dt$ . Dunque

$$dP(t, dt) = e^{-\lambda t} \times \lambda dt \implies \frac{dP(t, dt)}{dt} = \lambda e^{-\lambda t}.$$

Ricordando la definizione di densità di probabilità, abbiamo che

$$p(t; \lambda) = \lambda e^{-\lambda t}.$$

Definiamo quindi la distribuzione esponenziale con parametro  $\lambda > 0$  come

$$\varepsilon(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & 0 \leq x \leq +\infty \\ 0 & x < 0. \end{cases}$$

### 3.4.1 Normalizzazione, media, varianza e asimmetria

#### **Teorema 3.11** Condizione di normalizzazione della distribuzione esponenziale

La distribuzione esponenziale rispetta la condizione di normalizzazione

$$\int_{-\infty}^{+\infty} \varepsilon(x; \lambda) dx = 1.$$

**Dimostrazione:** Svolgendo l'integrale abbiamo che

$$\int_{-\infty}^{+\infty} \varepsilon(x; \lambda) dx = \int_0^{+\infty} \lambda e^{-\lambda x} dx = -e^{-\lambda x} \Big|_0^{+\infty} = 1.$$

☺

Per quanto riguarda media e varianza, si ha che

#### **Teorema 3.12** Valore atteso della distribuzione esponenziale

Il valore atteso della distribuzione esponenziale è pari a

$$\mu = \frac{1}{\lambda}.$$

**Dimostrazione:** Per definizione il valore atteso è dato dal valore di aspettazione della distribuzione esponenziale

$$\mu = E[x] = \int_0^{+\infty} x \varepsilon(x; \lambda) dx = \int_0^{+\infty} \lambda x e^{-\lambda x} dx.$$

Facendo il cambio di variabile  $t = \lambda x \implies dt = \lambda dx$  e gli estremi di integrazione rimangono immutati, pertanto

$$\mu = \int_0^{+\infty} x e^{-\lambda x} \lambda dx = \int_0^{+\infty} \frac{t}{\lambda} e^{-t} dt = \frac{1}{\lambda} \int_0^{+\infty} t e^{-t} dt.$$

Procedendo per un'integrazione per parti, prendendo  $e^{-t}$  come funzione da integrare, si ha che

$$\int_0^{+\infty} t e^{-t} dt = -t e^{-t} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-t} dt = 0 + \int_0^{+\infty} e^{-t} dt = -e^{-t} \Big|_0^{+\infty} = 1.$$

Segue allora che

$$\mu = \frac{1}{\lambda} \int_0^{+\infty} t e^{-t} dt = \frac{1}{\lambda}.$$

☺

**Teorema 3.13** Varianza della distribuzione esponenziale

La varianza della distribuzione esponenziale è pari a

$$\sigma^2 = \frac{1}{\lambda^2}.$$

**Dimostrazione:** Usiamo sempre la 1.25, pertanto procediamo a calcolare  $E[x^2]$ :

$$E[x^2] = \int_0^{+\infty} x^2 \varepsilon(x; \lambda) dx = \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx,$$

ed effettuiamo sempre il solito cambio di variabile  $t = \lambda x \implies dt = \lambda dx$ , dunque

$$E[x^2] = \int_0^{+\infty} x^2 e^{-\lambda x} \lambda dx = \frac{1}{\lambda^2} \int_0^{+\infty} t^2 e^{-t} dt.$$

Concentrandoci su quest'ultimo integrale, procediamo effettuando due integrazioni per parti prendendo sempre  $e^{-t}$  come funzione da integrare, pertanto:

$$\int_0^{+\infty} t^2 e^{-t} dt = -t^2 e^{-t} \Big|_0^{+\infty} + 2 \int_0^{+\infty} t e^{-t} dt = 0 + 2(-t e^{-t} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-t} dt) = -2e^{-t} \Big|_0^{+\infty} = 1.$$

Si ottiene la tesi, siccome

$$\sigma^2 = E[x^2] - \mu^2 = \frac{1}{\lambda^2} \int_0^{+\infty} t^2 e^{-t} dt - \mu^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

☺

**Lemma 3.4.1** Momenti algebrici di ordine superiore per la distribuzione  $\varepsilon(x; \lambda)$ 

I momenti algebrici di ordine superiore per la distribuzione  $\varepsilon(x; \lambda)$  soddisfano la seguente relazione ricorsiva

$$E[x^n] = \frac{n}{\lambda} E[x^{n-1}]$$

**Dimostrazione:** Osserviamo che, effettuando un'integrazione per parti nel calcolo del momento algebrico di ordine  $n$  riconoscendo che  $\lambda e^{-\lambda x} = \frac{d}{dx}(-e^{-\lambda x})$ , si ha che

$$\begin{aligned} E[x^n] &= \int_0^{+\infty} x^n \lambda e^{-\lambda x} dx = -x^n e^{-\lambda x} \Big|_0^{+\infty} + n \int_0^{+\infty} x^{n-1} e^{-\lambda x} dx = 0 + n \int_0^{+\infty} x^{n-1} \frac{\lambda}{\lambda} e^{-\lambda x} dx = \\ &= \frac{n}{\lambda} \int_0^{+\infty} x^{n-1} \lambda e^{-\lambda x} dx = \frac{n}{\lambda} E[x^{n-1}]. \end{aligned}$$

☺

Calcoliamo anche qua il coefficiente di *skewness*: si calcola sempre il momento algebrico di ordine 3

$$E[x^3] = \frac{3}{\lambda} E[x^2] = \frac{3}{\lambda^3}$$

e ci calcoliamo  $\mu_3 = E[x^3] - 3\mu\sigma^2 - \mu^3 = \frac{6}{\lambda^3} - 3\frac{1}{\lambda}\frac{1}{\lambda^2} - \frac{1}{\lambda^3} = \frac{2}{\lambda^3}$ . Da cui segue che

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\frac{2}{\lambda^3}}{\frac{1}{\lambda^3}} = 2$$

### 3.4.2 Funzione cumulativa

Possiamo facilmente calcolare la funzione cumulativa di questa distribuzione

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda t} \Big|_0^x = 1 - e^{-\lambda x} \text{ per } x \geq 0.$$

che può essere invertita, esplicitando la  $x$ , ottenendo che

$$F(x) - 1 = -e^{-\lambda x} \implies 1 - F(x) = e^{-\lambda x} \implies F^{-1}(y) = -\frac{\ln 1 - y}{\lambda}$$

### 3.4.3 Assenza di memoria

Una caratteristica molto peculiare della distribuzione esponenziale è costituita dal fatto che

$$P(x \geq x_1 + x_2) = \int_{x_1+x_2}^{+\infty} \varepsilon(x; \lambda) dx = \lambda \int_{x_1+x_2}^{+\infty} e^{-\lambda x} dx = e^{-\lambda(x_1+x_2)} = P(x \geq x_1)P(x \geq x_2)$$

Per definizione di probabilità condizionale, abbiamo che

$$P(x \geq x_1 + x_2) = P(x \geq x_1 + x_2 | x \geq x_1)P(x \geq x_1) = P(x \geq x_2)P(x \geq x_1) \implies P(x \geq x_1 + x_2 | x \geq x_1) = P(x \geq x_2).$$

Una variabile casuale che goda di questa proprietà si dice *senza memoria*.

## 3.5 Distribuzione di Gauss

La distribuzione di Gauss può essere formalmente ottenuta come limite della distribuzione poissoniana per  $\mu \rightarrow +\infty$  oppure direttamente come limite della distribuzione binomiale per  $n \rightarrow +\infty$  senza porre alcuna condizione su  $p$ . La sua importanza deriva certamente dal ruolo che essa gioca nel teorema centrale del limite, ma anche perché sarà il prototipo di distribuzione degli errori delle grandezze su cui vorremo effettuare il bestfit.

#### Teorema 3.14 Distribuzione di Gauss come limite

La distribuzione di Gauss  $\mathcal{N}(x; \mu, \sigma)$  può essere ottenuta come limite della distribuzione poissoniana per  $\mu \rightarrow 0$

**Dimostrazione:** Consideriamo la distribuzione poissoniana  $\mathcal{P}(k; \mu)$  e prendiamone il logaritmo

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

Vorremmo adesso approssimare il fattoriale: ciò viene effettuato considerando l'approssimazione di Stirling ( $k! \approx (\frac{k}{e})^k \sqrt{2\pi k}$ ). Dobbiamo procedere, però, con *cautela*: sempre per le solite ragioni esposte quando è stato effettuato il limite della binomiale, si devono fare i conti col fatto che  $k$  non è un numero, ma una variabile aleatoria. Possiamo introdurre la variabile casuale ridotta

$$\delta = \frac{k - \mu}{\mu},$$

e, grazie al teorema di Chebyshev, sappiamo che le fluttuazioni di  $k$  rispetto al valor medio  $\mu$  non distano più di  $\sigma = \sqrt{\mu}$ : per confronto si ha allora che

$$\begin{array}{ccc} -\frac{\sqrt{\mu}}{\mu} & \leq \delta \leq & \frac{\sqrt{\mu}}{\mu} \\ \downarrow & & \downarrow \\ 0 & & 0 \end{array}$$

quindi  $\delta \rightarrow 0$ . Questo ci consente di scrivere che  $k = \mu(1 + \delta)$  e siamo quindi autorizzati ad utilizzare l'approssimazione di Stirling all'interno della precedente espressione

$$\ln \mathcal{P}(\delta; \mu) \approx \mu(1 + \delta) \ln \mu - \frac{1}{2} \ln(2\pi\mu(1 + \delta)) - \mu(1 + \delta) \ln \mu(1 + \delta) + \mu(1 + \delta) - \mu.$$

A questo punto possiamo sviluppare in serie attorno a  $\delta = 0$  e, senza riportare i conti espliciti, si ha che

$$\ln \mathcal{P}(k; \mu) \approx \ln \mathcal{P}(0; \mu) + \frac{d \ln \mathcal{P}(0; \mu)}{d\delta} \delta + \frac{1}{2} \frac{d^2 \ln \mathcal{P}(0; \mu)}{d\delta^2} \delta^2 = -\frac{1}{2} \ln 2\pi\mu - \frac{1}{2} \delta - \frac{1}{2} \mu \delta^2 + \frac{1}{4} \delta^2,$$

e osserviamo che il termine  $-\frac{1}{2} \delta$  e  $+\frac{1}{4} \delta^2$  si annullano per  $\delta \rightarrow 0$ , mentre il terzo termine riporta la forma indeterminata  $0 \cdot \infty$ : sappiamo però che  $\delta \sim \frac{1}{\sqrt{\mu}} \implies \delta^2 \sim \frac{1}{\mu}$ , quindi  $\mu \delta^2 \sim 1$ . Riscrivendo tutto nuovamente in termini di  $k$  abbiamo che

$$\ln \mathcal{P}(k; \mu) \approx -\frac{1}{2} \ln 2\pi\mu - \frac{1}{2} \mu \left( \frac{k - \mu}{\mu} \right)^2 = -\frac{1}{2} \ln 2\pi\mu - \frac{1}{2} \frac{(k - \mu)^2}{\mu} \implies \mathcal{P}(k; \mu) \approx \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{1}{2} \frac{(k - \mu)^2}{\mu}}. \quad (3.14)$$

Usando il fatto che  $\sigma^2 = \mu$  possiamo riscrivere la 3.14 come

$$\mathcal{P}(k; \mu) \approx \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{k - \mu}{\sigma} \right)^2}.$$

Possiamo passare al continuo, sostituendo la variabile discreta  $k$  con la variabile continua  $x$ , e lasciare la deviazione standard  $\sigma$  libera di variare indipendentemente dalla media: abbiamo così ottenuto la distribuzione di Gauss! ☺

#### Osservazione:-

Osserviamo una caratteristica importante della gaussiana: la sua simmetria. Avevamo ricavato che la poissoniana ha un coefficiente di *skewness* pari a  $\frac{1}{\sqrt{\mu}}$  (infatti è asimmetrica) ma per  $\mu \rightarrow +\infty$  osserviamo che la *skewness* tende a zero, il che spiega la simmetria.

La distribuzione di Gauss si scrive, quindi, nella forma

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

dove  $\mu$  e  $\sigma^2$  rappresentano proprio la media e la varianza, come vedremo.

### 3.5.1 Normalizzazione, media e varianza

#### Teorema 3.15 Condizione di normalizzazione della distribuzione $\mathcal{N}(x; \mu, \sigma)$

La distribuzione di Gauss  $\mathcal{N}(x; \mu, \sigma)$  rispetta la condizione di normalizzazione

$$\int_{-\infty}^{+\infty} \mathcal{N}(x; \mu, \sigma) dx = 1.$$

**Dimostrazione:** Svolgiamo l'integrale

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} z^2} dz,$$

dove abbiamo fatto il cambio di variabile  $z = \frac{x - \mu}{\sigma} \implies dz = \frac{dx}{\sigma}$ . Purtroppo l'ultimo integrale non ha un'espressione analitica, ovvero non esiste una primitiva dell'integrale esprimibile in forma chiusa in termini di funzioni elementari: si può tuttavia calcolare usando il fatto che, posto  $I = \int_{-\infty}^{+\infty} e^{-\frac{1}{2} x^2} dx$ , allora possiamo considerare la funzione a due variabili  $f(x, y) = e^{-\frac{1}{2}(x^2 + y^2)}$  che integriamo su  $(-\infty, +\infty) \times (-\infty, +\infty)$ . In virtù del teorema di Tonelli, sappiamo che quest'ultimo integrale è pari a

$$\int_{(-\infty, +\infty) \times (-\infty, +\infty)} e^{-\frac{1}{2}(x^2 + y^2)} dx dy = \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y^2)} dy \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = I^2.$$

Quindi, abbiamo ricondotto il nostro integrale ad un integrale a due variabili che possiamo però risolvere passando alle coordinate polari (ricordandosi che l'elemento di superficie  $dx dy$  diventa  $r dr d\theta$ ): si osserva infatti che

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x^2 + y^2)} dx dy = \int_0^{2\pi} \int_0^{+\infty} r e^{-\frac{1}{2}r^2} dr = 2\pi \int_0^{+\infty} r e^{-\frac{1}{2}r^2} dr = -2\pi e^{-\frac{1}{2}r^2} \Big|_0^{+\infty} = 2\pi.$$

Segue che  $I^2 = 2\pi \implies I = \sqrt{2\pi}$ , dunque

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} I = 1.$$

☺

Per svolgere i calcoli successivi è interessante mostrare il seguente lemma

**Lemna 3.5.1** Relazione ricorsiva degli integrali gaussiani

Posto  $I_n = \int_{-\infty}^{+\infty} x^n e^{-\frac{1}{2}x^2} dx$ , abbiamo che

$$I_n = (n-1)I_{n-2}$$

**Dimostrazione:** Consideriamo  $I_n$  e, prendendo "in prestito" una  $x$  dal termine  $x^n$ , possiamo integrare per parti riconoscendo che  $-xe^{-\frac{1}{2}x^2} = \frac{d}{dx}(e^{-\frac{1}{2}x^2})$

$$\begin{aligned} \int_{-\infty}^{+\infty} x^n e^{-\frac{1}{2}x^2} dx &= - \int_{-\infty}^{+\infty} -x^{n-1} x e^{-\frac{1}{2}x^2} dx = - \left( \cancel{x^{n-1} e^{-\frac{1}{2}x^2}} \Big|_{-\infty}^{+\infty} - (n-1) \int_{-\infty}^{+\infty} x^{n-2} e^{-\frac{1}{2}x^2} dx \right) \\ &= (n-1) \int_{-\infty}^{+\infty} x^{n-2} e^{-\frac{1}{2}x^2} dx = (n-1)I_{n-2} \end{aligned}$$

☺

**Osservazione:-**

I termini dispari della successione  $I_n$  sono pari a zero: il motivo deriva dal fatto che  $x^{2k+1}$  è una funzione dispari, dunque  $x^{2k+1}e^{-\frac{1}{2}x^2}$  è una funzione dispari e ciò comporta che l'integrale su un dominio simmetrico è nullo.

Possiamo adesso calcolare più agevolmente media e varianza

**Teorema 3.16** Media della distribuzione  $\mathcal{N}(x; \mu, \sigma)$

La media della distribuzione gaussiana  $\mathcal{N}(x; \mu, \sigma)$  è pari a  $\mu$

**Dimostrazione:** Per definizione la media della distribuzione è data dal valore di aspettazione

$$E[x] = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma z + \mu) e^{-\frac{1}{2}z^2} dz,$$

dove abbiamo operato il cambio di variabile  $z = \frac{x-\mu}{\sigma}$ . Abbiamo allora che

$$E[x] = \frac{1}{\sqrt{2\pi}} (\sigma I_1 + \mu I_0) = \frac{1}{\sqrt{2\pi}} (\mu I_0) = \mu,$$

dove abbiamo usato il fatto che  $I_1 = 0$  e  $I_0 = \sqrt{2\pi}$ .

☺

**Teorema 3.17** Varianza della distribuzione  $\mathcal{N}(x; \mu, \sigma)$

La varianza della distribuzione  $\mathcal{N}(x; \mu, \sigma)$  è pari a  $\sigma^2$

**Dimostrazione:** Per definizione, la varianza è data dal momento centrale di ordine 2:

$$E[(x-\mu)^2] = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma^2 z^2 e^{-\frac{1}{2}z^2} dz,$$

dove abbiamo usato sempre il solito cambio di variabile  $z = \frac{x-\mu}{\sigma}$ . Segue che

$$E[(x-\mu)^2] = \frac{1}{\sqrt{2\pi}} \sigma^2 I_2 = \frac{1}{\sqrt{2\pi}} \sigma^2 (2-1)I_0 = \sigma^2.$$

☺

Il coefficiente di *skewness* è banalmente nullo perché la distribuzione è asimmetrica.



## 3.6 Il teorema centrale del limite

### Teorema 3.18 Teorema centrale del limite

Siano date  $n$  variabili **indipendenti**, ciascuna con media  $\mu_i$  e varianza  $\sigma_i^2$  finite (non necessariamente uguali). Indipendentemente dalla forma delle funzioni di distribuzione delle singole  $x_i$ , la somma  $S = \sum_{i=1}^n x_i$  è asintoticamente normale nel limite di  $n \rightarrow +\infty$ , ovvero la funzione di distribuzione di  $S$  tende ad una distribuzione di Gauss con media e varianza date da

$$E[S] = \sum_{i=1}^n \mu_i \text{ e } \text{Var}(S) = \sum_{i=1}^n \sigma_i^2.$$

## Capitolo 4

# Funzioni di probabilità di funzioni qualunque

In questo capitolo ci si sofferma, limitandoci al caso unidimensionale, sul problema della determinazione della funzione di distribuzione di variabili casuali  $y = f(x)$ , nota la funzione di distribuzione della variabile casuale  $x$ .

### 4.1 Variabili discrete

Nel momento in cui si considera  $f(x)$  dove  $x$  è una variabile aleatoria, la funzione di distribuzione che caratterizza la funzione  $f(x)$ , nel caso in cui  $f$  risulti biunivoca, è, banalmente, pari a

$$p(y) = p(x)$$

tuttavia nel caso di funzioni iniettive, la questione si complica.

# Indice analitico

approssimazione di Stirling, 6, 37

assenza di memoria, 37

assiomi di Kolmogorov, 2

coefficiente binomiale, 6

coefficiente di asimmetria, 14

coefficiente multinomiale, 8

condizione di normalizzazione, 10, 11, 27, 31, 34, 35

correlazione, 18

covarianza, 17

covarianza campione, 24

densità di probabilità, 11

densità di probabilità marginale, 16

deviazione standard, 13

disposizione, 6

distribuzione binomiale, 27, 30

distribuzione di Gauss, 37

distribuzione esponenziale, 34

distribuzione generatrice, 20

distribuzione poissoniana, 29

distribuzione uniforme, 33

disuguaglianza di Chebyshev, 14

eventi indipendenti, 8, 9

fattoriale, 5

funzione cumulativa, 15, 37

funzione di distribuzione, 10

FWHM, 13

FWHM, 13

matrice di correlazione, 18

matrice di covarianza, 18

media campione, 22

mediana, 12

moda, 12

momento algebrico, 14, 17

momento centrale, 14

momento di ordine  $n$ , 14

probabilità, 2

probabilità complementare, 3

probabilità condizionata, 8, 16

processo poissoniano, 29

propagazione dell'errore statistico, 26

skewness, 28, 34

spazio campionario, 2

spazio degli eventi, 2

teorema binomiale, 7

teorema di Bayes, 9

valore atteso, 28, 32, 34, 35

valore di aspettazione, 11

valore medio, 12, 14, 17

variabile aleatoria, 10

variabile casuale, 10

varianza, 13, 14, 18, 28, 32, 34, 36

varianza campione, 22