

**Appunti di Calcolo delle probabilità**

**Riccardo Lo Iacono**

Dipartimento di Matematica & Informatica  
Università degli studi di Palermo  
Sicilia  
a.a. 2023-2024

## Indice.

<b>1</b>	<b>Introduzione alla probabilità: storia e concetti base</b>	<b>2</b>
1.1	Partizione dell'evento certo . . . . .	2
<b>2</b>	<b>Impostazioni della probabilità</b>	<b>3</b>
2.1	Impostazione assiomatica . . . . .	3
2.2	Impostazione frequentista . . . . .	3
2.3	Impostazione soggettiva . . . . .	3
<b>3</b>	<b>Eventi e probabilità condizionate</b>	<b>5</b>
3.1	Teorema delle probabilità composte . . . . .	5
3.2	Formula di disintegrazione . . . . .	5
3.3	Teorema di Bayes . . . . .	6
<b>4</b>	<b>Numeri aleatori</b>	<b>7</b>
4.1	Numeri aleatori semplici . . . . .	7
4.2	Numeri aleatori discreti . . . . .	9
<b>5</b>	<b>Distribuzioni assolutamente continue</b>	<b>12</b>
5.1	Previsione e varianza . . . . .	12
5.2	Distribuzione uniforme . . . . .	13
5.3	Distribuzione esponenziale . . . . .	13
5.4	Distribuzione Normale . . . . .	13
5.5	Distribuzione Gamma . . . . .	14
<b>6</b>	<b>Vettori aleatori</b>	<b>15</b>
6.1	Vettori aleatori discreti . . . . .	15
6.2	Covarianza . . . . .	16
6.3	Matrice di varianze e covarianze . . . . .	16
6.4	Funzione caratteristica . . . . .	17
6.5	Convergenza . . . . .	17

## – 1 – Introduzione alla probabilità: storia e concetti base.

Storicamente il primo ad interessarsi a quella che sarebbe poi divenuto lo studio della probabilità fu il cavaliere *De Mere*. Questi accanito scommettitore, suppose che vi fosse un legame tra la frequenza con cui uscisse un certo risultato e la sua probabilità.

**Esempio:** Come esempio della logica di De Mere si consideri quanto segue: siano

$A =$  “Lanciando un dado a sei facce 4 volte, esce almeno un 6”

$B =$  “Lanciando due dadi a sei facce 24 volte, esce almeno una coppia (6,6)”

ci si chiede:  $\Pr(A) = \Pr(B)$ ?

Secondo De Mere si avrebbe

$$\Pr(A) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3}$$

$$\Pr(B) = \underbrace{\frac{1}{36} + \frac{1}{36} + \cdots + \frac{1}{36}}_{\times 36} = \frac{2}{3} \implies \Pr(A) = \Pr(B)$$

Si osserva però che una valutazione più “corretta” è ottenuta sottraendo a tutti casi possibili, quelli a sfavore di  $A$  e  $B$  rispettivamente, da cui

$$\Pr(A) = \frac{6^4 - 5^4}{6^4} \approx 0.518$$

$$\Pr(B) = \frac{36^{24} - 35^{24}}{36^{24}} \approx 0.491$$

Concludendo dunque che  $\Pr(A) > \Pr(B)$ .

Introducendo un po di formalismo, quelli che nell’esempio di sopra sono stati indicati con  $A$  e  $B$  prendono il nome di eventi. Più in generale, fissato un certo spazio campionario  $\Omega$ , dicasi ogni suo sottoinsieme  $E$  evento. Cioè un evento descrive uno dei possibili esiti di un esperimento. Casi particolari di evento sono l’evento certo  $\Omega$  e quello impossibile  $\emptyset$ . In fine dato  $E$  un evento, si definisce  $|E|$  il suo indicatore e si ha

$$|E| = \begin{cases} 1, & \text{se } E \text{ è vero;} \\ 0, & \text{se } E \text{ è falso.} \end{cases}$$

Dove con “vero” e “falso” ci si riferisce al verificarsi o meno di  $E$ .

**Nota.** la definizione di probabilità data da De Mere prende il nome di *criterio classico*.

### – 1.1 – Partizione dell’evento certo.

Siano  $E_1, \dots, E_n$  una famiglia di eventi. Allora se

$$1. \quad \forall i, j, i \neq j \quad E_i \wedge E_j = \emptyset$$

$$2. \quad \bigcup_{i=1}^n (E_i) = \Omega$$

si dirà che  $E_1, \dots, E_n$  formano una partizione di  $\Omega$ .

## – 2 – Impostazioni della probabilità.

Quella dell'impostazione classica, non è l'unica impostazione della probabilità esistente. Di fatti in aggiunta a quella sinora vista, si hanno le impostazioni *assiomatica*, *frequentista* e *soggettiva*, analizzate in maggior dettaglio nelle sezioni a seguire.

### – 2.1 – Impostazione assiomatica.

L'impostazione assiomatica da una definizione rigorosamente matematica di probabilità, a seguito riportata.

**Definizione:** dato  $(\Omega, A)$  uno spazio di misura: ossia  $\Omega$  è uno spazio con un'algebra (o una  $\sigma$ -algebra) e  $A$  una sua sottofamiglia di eventi; una funzione  $\Pr : A \rightarrow [0, 1]$ , dicasi probabilità se:

1.  $\forall E \in A, \Pr(E) \geq 0$ ;
2.  $\Pr(\Omega) = 1$ ;
3.  $\Pr(E_1 \vee E_2) = \Pr(E_1) + \Pr(E_2), \forall E_1, E_2 \in A : E_1 E_2 = \emptyset$

Nel caso in cui  $A$  sia infinito, il punto 3 diventa

$$\Pr\left(\bigvee_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \Pr(E_i)$$

### – 2.2 – Impostazione frequentista.

Come evidente dall'esempio di De Mere, vi è una naturale tendenza ad associare la probabilità di un evento alla frequenza con cui lo stesso avviene. Tale concezione di probabilità è detta *impostazione frequentista*.

**Definizione:** considerata una successione di prove indipendenti, ripetute sotto le stesse condizioni, posto  $E$  un evento e  $f_N$  la frequenza di successo entro le prime  $N$  prove, si pone

$$\Pr(E) = \lim_{N \rightarrow \infty} f_N$$

### – 2.3 – Impostazione soggettiva.

Le impostazioni sinora viste, sono generalizzabili nella *teoria soggettiva*. Con tale impostazione si ha una divisione rigorosa tra gli aspetti del certo e del probabile. Più in generale si ha quanto segue.

**Definizione:** sia  $E$  un evento, la probabilità  $\Pr(E) = p$  rappresenta il grado di fiducia che un individuo ha del verificarsi di  $E$ .

#### – 2.3.1 – Criterio della scommessa.

Da quanto detto precedentemente, sia supposto  $\Pr(E) = p$  per un qualche evento  $E$ , si può intendere  $p$  come la quota che, in un ipotetica scommessa, un individuo è disposto a scommettere. Cioè

$$\text{si paga } p \text{ e si riceve } \begin{cases} 1, & \text{se } E \text{ vero} \\ 0, & \text{altrimenti} \end{cases}$$

più in generale,  $\forall S \in \mathbb{R}, S \neq 0$

$$\text{si paga } pS \text{ e si riceve } \begin{cases} S, & \text{se si verifica } S, \\ 0, & \text{altrimenti} \end{cases}$$

## – 2.3.2 – Guadagno e criterio di coerenza.

**Definizione:** sia  $\Pr(E) = p$  la probabilità che un individuo attribuisce al verificarsi di  $E$ . Dicesi la seguente quantità

$$\mathcal{G} = S|E| - pS = S(|E| - p)$$

guadagno.

In generale, affinché una scommessa sia sensata, questa deve essere coerente, cioè: l'assegnazione di  $p$  non deve portare ad una perdita certa, e ciò si verifica se e solo se

$$\min \{\mathcal{G}\} \max \{\mathcal{G}\} \leq 0, \forall S \in \mathbb{R}$$

### – 3 – Eventi e probabilit  condizionate.

**Definizione:** siano  $E$  e  $H, H \neq 0$  eventi. Dicesi evento condizionato il seguente ente logico.

$$(E | H) = \begin{cases} \text{vero, se } E \text{ e } H \text{ veri,} \\ \text{falso, se } E^C \text{ e } H \text{ veri,} \\ \text{indeterminato, se } H^C \text{ vero.} \end{cases}$$

Banalmente se  $H = \Omega$ , si ha  $(E | H) = (E | \Omega) = E$ . Vale inoltre

$$(E | H) = ((E \wedge \Omega) | H) = [(E \wedge H) \vee (E \wedge H^C)] | H = ((E \wedge H) | H)$$

Applicando dunque il criterio di scommessa si ha che

$$\Pr(E | H) = \begin{cases} S, & \text{se } EH \text{ vero,} \\ 0, & \text{se } E^C H \text{ vero} \\ pS, & \text{se } H \text{ falso.} \end{cases}$$

da cui, considerando il guadagno si ha

$$\mathcal{G} = S |EH| - pS |H| = S |H| (|E| - p)$$

posto dunque  $G = \{S(1 - p), -pS\}$ , si deve verificare

$$\begin{aligned} \min \{G\} \max \{G\} &\leq 0, \forall S \neq 0 \\ \implies S^2(1 - p)(-p) &\leq 0 \iff p \in [0, 1] \end{aligned}$$

affinche la condizione di coerenza sia rispettata.

#### – 3.1 – Teorema delle probabilit  composte.

**Teorema 3.1.** *Dati due eventi  $E$  e  $H, H \neq 0$ , siano  $(\Pr(H), \Pr(E | H), \Pr(EH))$  valutazioni di probabilit  sugli eventi  $\{H, E \wedge H, (E | H)\}$ , allora se*

$$\Pr(EH) = \frac{\Pr(E | H)}{\Pr(H)}$$

*l'assegnazione risulta essere coerente.*

**Corollario 3.1.1.** *Se  $\Pr(H) \neq 0$ , allora*

$$\Pr(E | H) = \frac{\Pr(EH)}{\Pr(H)}$$

#### – 3.2 – Formula di disintegrazione.

Il *Teorema 3.1* pu  essere generalizzato a  $n$  eventi, come segue. Siano  $\{H_1, \dots, H_n\}$  una partizione di  $\Omega$  e sia  $E$  un evento. Allora

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

---

**– 3.3 – Teorema di Bayes.**

Dalla formula di disintegrazione segue il seguente risultato, noto come *Teorema di Bayes*.

**Teorema di Bayes.** Siano  $E, H$  due eventi. Allora

$$\Pr(H | E) = \frac{\Pr(H) \Pr(E | H)}{\Pr(E)}$$

Questi, nella sua versione pi  generale stabilisce quanto segue.

**Teorema di Bayes (generalizzato).** Siano  $\{H_1, \dots, H_n\}$  un partizione di  $\Omega$ , sia  $E$  un evento. Si ha allora

$$\Pr(H_i | E) = \frac{\Pr(H_i) \Pr(E | H_i)}{\Pr(E | H_1) \Pr(H_1) + \dots + \Pr(E | H_n) \Pr(H_n)}$$

**– 3.3.1 – Indipendenza stocastica.**

**Definizione:** dati due eventi  $E$  e  $H$ ,  $H \neq \emptyset$ , si dir  che questi sono stocasticamente indipendenti se

$$\Pr(E | H) = \Pr(E)$$

da cui

$$\Pr(EH) = \Pr(E) \Pr(H)$$

## – 4 – Numeri aleatori.

Dato  $\Omega$  uno spazio di probabilità, un numero aleatorio può essere inteso come una funzione a variabili reali definita su  $\Omega$  stesso. Sia dunque  $X$  un numero aleatorio, e sia  $x \in \mathbb{R}$  un valore assumibile da  $X$ . Allora posto considerato l'evento  $(X = x)$ , a questi si associa la probabilità  $\Pr(X = x)$ .

In genere, si distinguono due “classi” di numeri aleatori:

- **discreti:** se il numero di valori assumibili dal numero aleatorio è finito o al più numerabile;
- **continuo:** se non discreto.

**Osservazione.** banalmente, si deve verificare che

$$\sum_{i=1}^n \Pr(X = x_i) = \sum_{i=1}^n p_i = 1$$

affinche l'assegnazione di probabilità risulti coerente.

### – 4.1 – Numeri aleatori semplici.

Siano  $E_1, \dots, E_n$  eventi e siano  $\alpha_1, \dots, \alpha_n$  numeri reali. Si dice che

$$X = \alpha_1 |E_1| + \dots + \alpha_n |E_n|$$

è un numero aleatorio semplice.

Per come è costruito, il dominio di  $X$  non è definito ma lo è il codominio, codominio che può essere stabilito considerando i costituenti dell'evento. Nello specifico, se tutti gli eventi sono indipendenti, allora  $X$  può assumere  $2^n$  valori.

#### – 4.1.1 – Distribuzione ipergeometrica.

Si assuma di possedere un'urna con  $N$  palline, di cui  $pN$  bianche e  $qN$  nere sono note, con  $qN + pN = N$ . E da questa effettuare  $n$  estrazioni con restituzione. Sia posto  $E_i$  l'evento “l'i-esima pallina estratta è bianca” e sia

$$X = \sum_{i=1}^n |E_i|.$$

Si ha che

$$\Pr(E_i) = \frac{\binom{N-1}{pN-1}}{\binom{N}{pN}} = p$$

Cioè gli eventi sono equiprobabili.

Sia ora considerato il generico costituente di  $X$ , si ha

$$\Pr(E_{i_1} E_{i_2} \dots E_{i_h} E_{i_{h+1}}^C \dots E_{i_n}^C) = \dots = \frac{\binom{N-n}{pN-h}}{\binom{N}{pN}}$$

da cui dunque

$$(X = h) = \bigvee_{\{i_1, \dots, i_h\} \subseteq \{1, \dots, n\}} E_{i_1} E_{i_2} \dots E_{i_h} E_{i_{h+1}}^C \dots E_{i_n}^C$$

implicando che

$$\Pr(X = h) = \frac{\binom{N-n}{pN-h} \binom{n}{h}}{\binom{N}{pN}}$$

poiché esistono  $\binom{n}{h}$  modi per ottenere  $(X = h)$ .



#### – 4.1.2 – Distribuzione binomiale.

Siano  $E_1, \dots, E_n$  eventi indipendenti ed equiprobabili, con probabilità  $p$ . Sia  $q = 1 - p$  e sia  $X$  un numero aleatorio che rappresenta il numero di successi alle  $n$  prove. Cioè

$$X = |E_1| + \dots + |E_n|$$

Banalmente  $X$  assume valori in  $\{0, \dots, n\}$ . Per stabilire la distribuzione di  $X$  è necessario calcolare  $\Pr(X = h), \forall h \in \{0, \dots, n\}$ . Risulta dunque più conveniente considerare il numero di costituenti a favore ad esso. Segue che

$$\begin{aligned}\Pr(X = 0) &= \Pr(E_1^C E_2^C \dots E_n^C) = \Pr(E_1^C) \Pr(E_2^C) \dots \Pr(E_n^C) = (1 - p)^n \\ \Pr(X = 1) &= \Pr(E_1 E_2^C \dots E_n^C) = \Pr(E_1) \Pr(E_2^C) \dots + \Pr(E_n^C) = p(1 - p)^{n-1}\end{aligned}$$

Si dimostra banalmente, che in generale vale

$$\Pr(X = h) = \binom{n}{h} p^h q^{n-h}$$

Dicasi che  $X$  ha distribuzione binomiale. In simboli  $X \sim B_{n,p}$

#### – 4.1.3 – Mistura di binomiali.

Si assuma di possedere un'urna con  $N$  palline, di cui  $r$  bianche, con  $r$  non noto. E da questa effettuare  $n$  estrazioni con restituzione. Siano ora  $\{H_1, \dots, H_n\}$  una partizione dell'evento certo, con

$$H_r = \text{“Ci sono } r \text{ palline bianche”}$$

Sia  $E_i = \text{“l'i-esima pallina estratta è bianca”}$ , per  $i \in \{1, \dots, n\}$ . Si verifica che

$$\Pr(E_i | H_r) = \frac{r}{N}, \quad \forall i \in \{1, \dots, n\}$$

Da ciò

$$\begin{aligned}\Pr(E_i) &= \sum_{r=0}^N \Pr(E_i | H_r) \Pr(H_r) \\ &= \sum_{r=0}^N \Pr(H_r r/N)\end{aligned}$$

Da cui risulta  $\Pr(E_i)$  è equiprobabile, per ogni  $i$ .

Sia ora posto  $X$  il numero aleatorio di palline estratte alle  $n$  prove. Si ha

$$\Pr(X = h | H_r) = \binom{n}{h} \left(\frac{r}{N}\right)^h \left(\frac{N-h}{N}\right)^{n-h}$$

da cui

$$\Pr(X = h) = \sum_{r=0}^N \binom{n}{h} \left(\frac{r}{N}\right)^h \left(\frac{N-h}{N}\right)^{n-h} \Pr(H_r) \quad (1)$$

Dicasi che  $X$  è una mistura di binomiali.

**Nota.** nel caso si effettuino estrazioni senza restituzione, la (1) diventa

$$\Pr(X = h) = \sum_{r=0}^N \frac{\binom{n}{h} \left(\frac{r}{N}\right)^h \left(\frac{N-h}{N}\right)^{n-h}}{\binom{N}{r}} \Pr(H_r)$$

Si dirà in questo caso che  $X$  è una mistura di iper-geometriche.

### – 4.2 – Numeri aleatori discreti.

Come anticipato, se  $X$  è un numero aleatorio è il numero di valori  $x_i$  che questi può assumere è finito, o al più numerabile, si dice che  $X$  è un numero aleatorio discreto.

Sia  $p_i = \Pr(X = x_i)$ ,  $\forall i \in \mathbb{N}$  la distribuzione di probabilità di  $X$ , sia supposto anche che

$$\sum_{i=1}^{\infty} p_i = 1 \quad \text{e} \quad \sum_{i=0}^{\infty} |x_i| p_i < \infty$$

allora si può definire il *valore atteso* di  $X$ , come

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i p_i = \mu$$

Sia ora assunto che  $\mu$  esista finito, e  $\mathbb{E}[(X - \mu)^2] < \infty$ , si definisce quest'ultima quantità varianza. Cioè

$$\text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \sum_{i=1}^{\infty} (x_i - \mu)^2 p_i$$

In fine, si definisce

$$F(x) = \Pr(X \leq x)$$

*funzione di ripartizione* di  $X$ . Se  $X$  è discreto si ha che

$$F(x) = \Pr(X \leq x) = \sum_{x_i \leq x} p_i, \quad x \in \mathbb{R}$$

#### – 4.2.1 – Distribuzione geometrica.

**Definizione:** dato  $X$  un numero aleatorio discreto, questi si dice avere distribuzione geometrica di parametro  $p \in [0, 1]$  se

$$\Pr(X = n) = (1 - p)^{n-1} p, \forall n \in \mathbb{N}$$

Più in generale, siano  $E_1, \dots, E_n$  eventi equiprobabili e indipendenti. Sia

$X =$  “Numero di prove fino al primo successo”

ossia  $X = \min\{n : |E_n| = 1\}$ . Segue che se  $X$  è discreto si ha

$$(X = n) = \begin{cases} E_1, n = 1 \\ E_1^C, \dots, E_{n-1}^C E_n, \forall n \geq 2 \end{cases}$$

da cui

$$\Pr(X = n) = (1 - p)^{n-1} p, \forall n \in \mathbb{N}$$

Considerando valore atteso e varianza si ha che, poiché  $X \in \mathbb{N}$  e nello specifico  $X = |X > 0| + |X > 1| + \dots$ , segue, posto  $(1 - p) = q$ , che

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \Pr(X > n) = 1 + q + q^2 + \dots = \sum_{i=0}^{\infty} q^i = \frac{1}{1 - q}$$

Per calcolare la varianza, sia considerato  $X^2$ , poiché

$$X^2 = |X > 0| + 3|X > 1| + \dots$$

segue  $\mathbb{E}(X^2) = 1 + 3(1 - p) + \dots = \dots = \frac{2-p}{p^2}$ , da cui

$$\text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \frac{1 - p}{p^2}$$

## – 4.2.2 – Distribuzione di Poisson.

Sia  $X$  un numero aleatorio discreto, questi si dice avere distribuzione di Poisson se

1.  $X \in \mathbb{N}_0 \implies X \in \{0, 1, \dots, n, \dots\}$
2.  $\Pr(X = n) = p_n = \frac{\lambda^n}{n!} e^{-\lambda}, \forall n \in \mathbb{N}_0$

Si dimostra che

$$\mathbb{E}(X) = \text{Var}(X) = \lambda$$

**Nota.** Tale distribuzione permette di approssimare una binomiale.

## – 4.2.2.1 – Proprietà di assenza di memoria.

**Definizione:** sia  $X$  un numero aleatorio discreto positivo. Si dice che  $X$  gode di assenza di memoria se

$$\Pr(X > x_0 + x \mid X > x_0) = \Pr(X > x), \forall x_0, x \in \mathbb{R}_0^+$$

**Osservazione.** Si dimostra che se  $X$  ha distribuzione di Poisson, allora  $X$  gode di assenza di memoria.

## – 4.2.3 – Distribuzione di Pascal e binomiale inversa.

Siano  $E_1, \dots, E_n$  eventi indipendenti ed equiprobabili. Sia  $\Pr(E_i) = p_i$  e sia  $T_k$  il numero di prove fino al  $k$ -esimo successo. Si ha, posto  $(1 - p) = q$ , che

$$\Pr(T_k = n) = \binom{n-1}{k-1} p^k q^{n-k}$$

Dicasi che  $T_k$  ha distribuzione di Pascal di parametri  $p$  e  $k$ .

Sia  $Z_k = T_k - k$  il numero di insuccessi prima dei  $k$  successi. Si ha

$$\Pr(Z_k = h) = \Pr(T_k = k + h) = \binom{k+h-1}{k-1} p^k q^h$$

Dicasi che  $Z_k$  ha distribuzione binomiale inversa di parametri  $k$  e  $p$ .

## – 4.2.4 – Disuguaglianza di Markov.

**Teorema 4.1.** Dato un numero aleatorio discreto non-negativo  $X$  con  $\mathbb{E}(X) < \infty$  e un  $\alpha > 0 \in \mathbb{R}$ , si ha

$$\Pr(X \geq \alpha) \leq \frac{\mathbb{E}(X)}{\alpha}$$

*Dimostrazione:* poiché  $X \geq 0$ , segue che  $\Pr(X \geq 0) = 1$ . Posto ora  $\Pr(X \geq x_n) = p_n$ , segue

$$\begin{aligned} \alpha \Pr(X \geq \alpha) &= \alpha \sum_{x_n : x_n \geq \alpha} p_n \\ &\leq \sum_{x_n : x_n \geq \alpha} x_n p_n \\ &\leq \sum_{x_n} x_n p_n = \mathbb{E}(X) \end{aligned}$$

da cui dunque

$$\Pr(X \geq \alpha) \leq \frac{\mathbb{E}(X)}{\alpha}$$

## – 4.2.5 – Disuguaglianza di Chebychev.

**Teorema 4.2.** Dato un numero aleatorio discreto  $X$  con  $\mathbb{E}(X) = \mu < \infty$  e  $\text{Var}(X) < \infty$ , si ha

$$\Pr(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \forall \varepsilon > 0$$

*Dimostrazione:* a partire da Markov (*Teorema 4.1*), si consideri  $Y = (X - \mu)^2$ . Si ha che  $Y \geq 0$  e  $\mathbb{E}(Y) = \sigma^2$ . Da Markov

$$\Pr(Y \geq \alpha) \leq \frac{\mathbb{E}(Y)}{\alpha}$$

posto allora  $\alpha = \varepsilon^2$ , segue

$$\Pr((X - \mu)^2 \geq \varepsilon^2) \leq \frac{\sigma^2}{\varepsilon^2}$$

ma poiché  $(X - \mu)^2 \geq \varepsilon^2 \iff |X - \mu| \geq \varepsilon$ , segue

$$\Pr(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

## – 5 – Distribuzioni assolutamente continue.

Sia  $\Omega$  un evento certo, sia  $A$  una sua partizione, tale che  $\|A\|_2 > \|\mathbb{N}\|_2$ . Si può dimostrare in questi casi che si può attribuire una probabilità solamente ad un sottoinsieme  $B \subset A$  tale che  $\|B\|_2 = \|\mathbb{N}\|_2$ .

In questi casi, se un numero aleatorio  $X$  può assumere valori la cui cardinalità è pari a quella del continuo, si dice che  $X$  è un numero aleatorio continuo. In generale, se  $X$  è un numero aleatorio continuo, si ha

$$\Pr(X = x) = 0, \forall x \in \mathbb{R}$$

Una definizione più rigorosa è la seguente.

**Definizione:** dicasi che un numero aleatorio  $X$  è continuo se

- $\Pr(X = x) = 0, \forall x$ ;
- $\exists f(x) > 0$  integrabile secondo Riemann, tale che  $\forall A \subseteq \mathbb{R}$  misurabile secondo Peano-Jordan, si abbia

$$\Pr(A) = \Pr(X \in A) = \int_A f(x) \, dx$$

**Nota.** Dicasi  $f$  densità di probabilità di  $X$ .

Nel caso di numeri aleatori continui, si dimostra esservi un legame tra funzione di ripartizione e densità di probabilità. Infatti, sia  $F(x)$  la funzione di ripartizione di un qualche numero aleatorio continuo  $X$ , poiché per definizione

$$F(x) = \Pr(X \leq x) = \Pr(X \in ]-\infty, x])$$

segue

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

Si ha cioè che

$$f(x) = F'(x)$$

### – 5.1 – Previsione e varianza.

Sia  $X$  un numero aleatorio continuo, in questo caso estendendo la definizioni data per il caso discreto al caso continuo, il valore atteso e la varianza di  $X$  sono definiti come segue.

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) \, dx$$

e supposto che

$$\int_{-\infty}^{\infty} |x| f(x) \, dx < \infty$$

segue

$$\text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx$$

### – 5.2 – Distribuzione uniforme.

Sia  $X$  un numero aleatorio continuo, e sia la sua densità di probabilità  $f(x)$  definita come segue

$$f(x) = \begin{cases} k > 0, & \text{se } x \in [a, b] \\ 0, & \text{altrimenti} \end{cases}$$

**Osservazione.** Osservando che

$$\int_{-\infty}^{\infty} f(t) dt = \int_a^b k dx = bk - ak = 1 \implies k = \frac{1}{b-a}$$

Per quanto osservato segue che

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & \text{se } x < a \\ \frac{x-a}{b-a}, & \text{se } x \in [a, b] \\ 1, & \text{se } x > b \end{cases}$$

Segue infine che valore atteso e varianza sono definiti come

$$\begin{aligned} \mathbb{E}[X] &= \int_a^b \frac{x}{b-a} dx = \dots = \frac{b+a}{2} \\ \text{Var}(X) &= \mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (X-\mu)^2 k dx = \frac{(b-a)^2}{12} \end{aligned}$$

### – 5.3 – Distribuzione esponenziale.

Sia  $X$  un numero aleatorio continuo, e si la sua densità di probabilità definita come

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \lambda \in \mathbb{R}^+, x > 0 \\ 0, & x \leq 0 \end{cases}$$

Segue dalla definizione, che

$$F(x) = \begin{cases} \int_0^x \lambda e^{-\lambda x} dx, & \text{se } x > 0 \\ 0, & \text{se } x \leq 0 \end{cases} \implies \begin{cases} 1 - e^{-\lambda x}, & \text{se } x > 0 \\ 0, & \text{se } x \leq 0 \end{cases}$$

Dalla definizione generale di valore atteso e varianza, segue

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x \lambda e^{-\lambda x} dx = \dots = \frac{1}{\lambda} \\ \text{Var}(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 \lambda e^{-\lambda x} dx = \dots = \frac{1}{\lambda^2} \end{aligned}$$

Vale in fine la seguente proprietà.

**Proprietà.** Sia  $X$  un numero aleatorio con distribuzione esponenziale, allora

$$\mathbb{E}(X^n) = \frac{n!}{\lambda^n}$$

### – 5.4 – Distribuzione Normale.

Sia  $X$  un numero aleatorio continuo, e sia la sua funzione di ripartizione definita come

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

Si dimostra che  $\mathbb{E}(X) = \mu$  e  $\text{Var}(X) = \sigma$ . Circa la funzione di ripartizione, questa può solo essere tabulata<sup>2</sup>.

<sup>2</sup>Si osserva banalmente che la funzione  $F(x)$  è riconducibile all'integrale di Gauss, per cui non esiste una primitiva semplice.

## – 5.5 – Distribuzione Gamma.

Sia  $X$  un numero aleatorio continuo, sia la sua distribuzione di probabilità definita come

$$f(x) = \begin{cases} \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

allora il suo valore atteso è definito come

$$\begin{aligned} \mathbb{E}(X) &= \int_0^\infty x \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \, dx \\ &= \dots = \frac{a}{\lambda} \end{aligned}$$

Poiché si dimostra che

$$\mathbb{E}(X^k) = \frac{(a+k-1)(a+k)\cdots a}{\lambda^k}$$

segue che  $\text{Var}(X) = a/\lambda^2$

**Nota.** se  $\lambda = 1/2$  e la densità di probabilità di  $X$  è definita come

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

si dirà che  $X$  ha distribuzione *Chi-Quadro con  $n$  gradi di libertà*.

## – 6 – Vettori aleatori.

Sia  $\Omega$  uno spazio campionario. Al generico  $\omega \in \Omega$  sono associati  $n$  valori  $x_1, \dots, x_n$ ,  $n \geq 2$  che rappresentano i valori assunti da  $n$  numeri aleatori  $X_1, \dots, X_n$ . Quest'ultimi possono essere visti come componenti di un vettore aleatorio  $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ . Come per i singoli numeri aleatori, si distinguono il caso discreto e quello continuo.

### – 6.1 – Vettori aleatori discreti.

**Definizione:** sia  $X = (X_1, \dots, X_n)$  un vettore aleatorio. Questi si dice essere discreto se

$$\exists C \subset \mathbb{R}^n : \forall x \in C, \Pr(X = x) \geq 0 \wedge \Pr(X = y) = 0 \forall y \notin C$$

Sia  $n = 2$ , e siano  $X, Y$  due numeri aleatori. Si indica con

$$C_{(X,Y)} = \{(x_i, y_j) \in C \subset \mathbb{R}^2 : \Pr(X = x_i, Y = y_j) = p_{x_i, y_j} > 0\}$$

Risulta che  $C \subseteq C_X \times C_Y$ .

Sia ora fissato un  $x_i \in C_X$ , osservando che

$$\Omega = \bigvee_{y_j \in C_Y} (Y = y_j)$$

si può decomporre  $(X = x_i)$  come

$$\begin{aligned} (X = x_i) &= (X = x_i) \wedge \Omega \\ &= \dots = \bigvee_{y_j \in C_Y} (X = x_i, Y = y_j) \end{aligned}$$

da cui

$$\Pr(X = x_i, Y = y_j) = \sum_{y_j} p_{x_i, y_j} \quad (2)$$

dicasi la (2), distribuzione marginale di  $X$ . Se

$$\Pr(X = x_i, Y = y_j) = \Pr(Y = y_j \mid X = x_i) \Pr(X = x_i)$$

si parla di distribuzione marginale condizionata.

#### – 6.1.1 – Indipendenza stocastica.

Siano  $X, Y$  due numeri aleatori. Si dirà che questi sono stocasticamente indipendenti se

$$\Pr(X = x_i, Y = y_j) = \Pr(X = x_i) \Pr(Y = y_j)$$

ossia

$$p_{x_i, y_j} = p_{x_i} p_{y_j}, \quad \forall (x_i, y_j) \in C^2$$

Segue banalmente che  $\Pr(X = x_i \mid Y = y_j) = \Pr(X = x_i)$ .



### – 6.2 – Covarianza.

Siano  $X, Y$  due numeri aleatori. La covarianza misura il grado di variabilità reciproca tra i due. Ossia quanto  $X$  varia in funzione di  $Y$  e viceversa. Segue dunque dalla definizione di varianza

$$\begin{aligned}\text{Var}(X + Y) &= \mathbb{E}[(X + Y) - (\mu_X + \mu_Y)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\mathbb{E}[(X - \mu_X) + (Y - \mu_Y)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)\end{aligned}$$

**Osservazione.** dalla definizione di valore atteso, segue

$$\begin{aligned}\text{Cov}(X, Y) &= 2\mathbb{E}[(X - \mu_X) + (Y - \mu_Y)] \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

**Nota.** se  $\text{Cov}(X, Y) = 0$  allora si dirà che  $X, Y$  sono incorrelati.

#### – 6.2.1 – Coefficiente di correlazione.

Siano  $X, Y$  due numeri aleatori. Si definisce

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X\sigma_Y}$$

*coefficiente di correlazione* (o covarianza normalizzata).

### – 6.3 – Matrice di varianze e covarianze.

Siano  $X_1, \dots, X_m, m$  numeri aleatori e siano  $Y_1, \dots, Y_n$  altri  $n$  numeri aleatori. Si considerino le seguenti combinazioni lineari

$$X = \sum_{i=1}^m \alpha_i X_i \quad Y = \sum_{j=1}^n \beta_j Y_j, \quad \alpha_i, \beta_j \in \mathbb{R}$$

calcolando la covarianza dei due si ha

$$\begin{aligned}\text{Cov}(X, Y) &= \text{Cov}\left(\sum_{i=1}^m \alpha_i X_i, \sum_{j=1}^n \beta_j Y_j\right) \\ &= \dots = \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j \sigma_{x_i, y_j}\end{aligned}$$

Nel particolare  $\text{Cov}(X, X) = \text{Var}(X)$ . Posto allora  $\alpha^T = (\alpha_1, \dots, \alpha_m)$  e

$$\Sigma = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,m} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m,1} & \sigma_{m,2} & \cdots & \sigma_{m,m} \end{pmatrix}$$

si ha che

$$\text{Cov}(X, X) = \text{Var}(X) = \alpha^T \Sigma \alpha$$

$$= (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_m) \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,m} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m,1} & \sigma_{m,2} & \cdots & \sigma_{m,m} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} \geq 0, \forall \alpha \in \mathbb{R}^m$$

### – 6.4 – Funzione caratteristica.

La funzione caratteristica è uno strumento teorico utile ad analizzare diversi aspetti dei numeri aleatori. Nello specifico, questa è definita come segue.

**Definizione:** sia  $X$  un numero aleatorio, e sia

$$Y = e^{itX} = \cos(tX) + i \sin(tX)$$

si definisce funzione caratteristica  $\varphi_X$  quanto segue.

$$\begin{cases} \sum_h p_h e^{itx_h}, & \text{se } X \text{ è discreto} \\ \int_{-\infty}^{\infty} e^{itx} dx, & \text{se } X \text{ è continuo} \end{cases}$$

Considerando unicamente il caso continuo, vale lo stesso ragionamento nel caso discreto, valgono le seguenti proprietà:

1.  $\varphi_X(0) = 1$
2.  $|\varphi_X(t)| \leq \varphi_X(0)$ , si ha infatti

$$\begin{aligned} |\varphi_X(t)| &= \left| \int_{-\infty}^{\infty} e^{itx} f(x) dx \right| \leq \int_{-\infty}^{\infty} |e^{itx} f(x)| dx \\ &= \int_{-\infty}^{\infty} |e^{itx}| f(x) dx = 1 \end{aligned}$$

#### – 6.4.1 – Somma di numeri aleatori stocasticamente indipendenti.

Proprietà fondamentale della funzione caratteristica è la seguente: dati  $n$  numeri aleatori  $X_1, \dots, X_n$  stocasticamente indipendenti, posto  $Y = X_1 + \dots + X_n$  si ha che

$$\varphi_Y(t) = \varphi_{X_1}(t) \varphi_{X_2}(t) \cdots \varphi_{X_n}(t)$$

La dimostrazione (che qui non si riporta), segue dalla definizione stessa di valore atteso.

### – 6.5 – Convergenza.

Sia  $\{X_n\}_{n \in \mathbb{N}}$  una successione di numeri aleatori, la *convergenza* cerca di stabilire se tale successione possa convergere, in termini probabilistici, ad un qualche numero aleatorio  $X$  con una qualche distribuzione di probabilità. Cioè, considerato  $B \subset \mathbb{R}$ , (in genere si considera la classe Borel  $\mathcal{B} \subset \mathbb{R}$ ), si deve verificare

$$\lim_{n \rightarrow \infty} \Pr(X_n \in B) = \Pr(X \in B), \forall B \in \mathcal{B}$$

Inoltre, dato il legame tra funzione di ripartizione e numeri aleatori, se  $\{F_n\}_{n \in \mathbb{N}}$  è una successione di funzioni di ripartizione di una successione di numeri aleatori  $\{X_n\}_{n \in \mathbb{N}}$ , allora

$$\lim_{n \rightarrow \infty} \Pr(X_n \in B) = \Pr(X \in B), \forall B \in \mathcal{B} \iff \lim_{n \rightarrow \infty} F_n(x) = F(x), \forall B \in \mathcal{B}, x \in \mathbb{R}$$

con  $F(x)$  funzione di ripartizione del numero aleatori con distribuzione limite.

**Definizione:** una successione di funzioni di ripartizione  $\{F_n\}_{n \in \mathbb{N}}$  convergere a una distribuzione limite se

$$\exists F_X(x) : \lim_{n \rightarrow \infty} F_n(x) = F_X(x), \text{ per ogni punto di } F_X$$

Si scriverà  $F_n \rightarrow F_X$

**Teorema di convergenza in distribuzione e funzione caratteristica.** Sia  $\varphi$  funzione caratteristica di  $F_X$ , allora la successione di funzioni di ripartizione  $\{F_n\}_{n \in \mathbb{N}}$  converge ad  $F_X$  se e solo se la successione  $\{\varphi_n\}_{n \in \mathbb{N}}$  di funzioni caratteristiche corrispondenti tendono a  $\varphi$ . Cioè

$$F_n \rightarrow F_X \iff \varphi_n \rightarrow \varphi$$

**Nota.** Questo risultato, permette la dimostrazione del teorema centrale del limite.

## – 6.5.1 – Teorema centrale del limite.

**Teorema centrale del limite.** Sia  $\{X_n\}_{n \in \mathbb{N}}$  una successione di numeri aleatori indipendenti ed equi-distribuite, con  $\mathbb{E}(X_i) < \infty$ ,  $\text{Var}(X_i) < \infty$ . Sia per ogni  $n \in \mathbb{N}$

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{\sigma^2 n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}$$

la media aritmetica di  $X_1, \dots, X_n$ . Allora, la successione  $\{Z_n\}_{n \in \mathbb{N}}$  converge ad un numero aleatorio  $Z$  con distribuzione normale standard. Ossia

$$\lim_{n \rightarrow \infty} \Pr(Z_n \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \forall z \in \mathbb{R}$$

*Dimostrazione:* sia  $\varphi_n(t) = \mathbb{E}(e^{itZ_n})$ ,  $n \in \mathbb{N}$ , si dimostra banalmente che

$$\varphi_n(t) \rightarrow e^{-\frac{t^2}{2}}$$

Sia ora  $Y_i = (X_i - \mu)/\sigma$ , cioè  $X_i$  standardizzato, si ha

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$$

Considerando la funzione caratteristica di  $Z_n$ , segue

$$\begin{aligned} \varphi_n(t) &= \varphi_{Y_1}(t/\sqrt{n}) \varphi_{Y_2}(t/\sqrt{n}) \cdots \varphi_{Y_n}(t/\sqrt{n}) \\ &= [\varphi_{Y_i}(t/\sqrt{n})]^n \end{aligned}$$

dove l'ultima uguaglianza deriva dal fatto che gli  $Y_i$  sono stocasticamente indipendenti. Dalla relazione tra i momenti e le derivate calcolate in zero, si ha  $\varphi_X^{(k)} = i^k \mathbb{E}(X^n)$ . Da ciò, considerata la serie di MacLaurin arrestata al secondo ordine, si ha

$$\varphi_{Y_1} = \varphi_{Y_1}(0) + \varphi'_{Y_1}(0)t + \varphi''_{Y_1}(0)\frac{t^2}{2!} + o(t^2)$$

da cui

$$\begin{aligned} \lim_{n \rightarrow \infty} \varphi_n(t) &= \lim_{n \rightarrow \infty} \left( 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right)^n \\ &= e^{\lim_{n \rightarrow \infty} n \ln \left( 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right)} \end{aligned}$$

Poiché  $\lim_{x \rightarrow 0} \ln(x+1/x) = 1$ , segue

$$\lim_{n \rightarrow \infty} n \ln \left( 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right) = \cdots = \lim_{n \rightarrow \infty} n \ln \left( 1 - \frac{t^2}{2n} \right)$$

In conclusione, da ciò

$$\lim_{n \rightarrow \infty} \varphi_n(t) = \lim_{n \rightarrow \infty} n \ln \left( 1 - \frac{t^2}{2n} \right) = e^{-\frac{t^2}{2}}$$

**Osservazione.** siano  $E_1, \dots, E_n$  una successione di eventi indipendenti ed equiprobabili, con  $\Pr(E_i) = p$ , e sia  $X_i = |E_i| \forall i$ . Si ha  $\mathbb{E}(X_i) = p$ ,  $\text{Var}(X_i) = p(1-p)$ . Posto, per ogni  $n \in \mathbb{N}$ , che

$$S_n = X_1 + \cdots + X_n = |E_1| + \cdots + |E_n|$$

si ha che  $S_n \sim \text{Bin}(n, p)$ , con  $\mathbb{E}(S_n) = np$  e  $\text{Var}(S_n) = np(1-p)$  per il teorema centrale del limite, segue

$$\Pr \left( \frac{S_n - np}{\sqrt{np(1-p)}} \leq x \right) \rightarrow_{n \rightarrow \infty} \Phi_{0,1}(x)$$

Ossia, per  $n$  grandi, la distribuzione binomiale può essere approssimata da una normale standard.