Università degli Studi di Napoli Federico II
Scuola Politecnica e delle Scienze di Base
Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione

Corso di Laurea in Informatica

Appunti del corso 2015/2016 di

Calcolo delle Probabilità e Statistica Matematica

dello studente Luigi L. L. Starace

Indice

Capitolo 1
Introduzione

1.1 Informazioni su questo documento

Questo documento è da considerarsi una ricopiatura con rielaborazioni ed integrazioni degli appunti presi durante le lezioni frontali della Prof. Caputo (corso del 2015/2016). Non ha la pretesa di essere privo di errori o completo, ma può considerarsi un valido strumento per la ripetizione degli argomenti trattati durante le lezioni.

1.2 Versione del documento, segnalazione errori e futuri aggiornamenti

Questo documento è stato compilato in data **25 febbraio 2016**.
La data in grassetto permette di discriminare le possibili varie versioni del documento.
Al momento della compilazione, il documento non contiene errori noti. É possibile segnalare errori e inesattezze ortografiche, concettuali, tipografiche tramite il thread ufficiale sul forum degli studenti:
`http://informatica-unina.com/forum/viewtopic.php?f=130&t=1122`
Eventuali aggiornamenti del documento apportanti correzioni e/o ampliamenti saranno resi disponibili nel thread ufficiale. Consiglio di verificare periodicamente la presenza di release successive.

1.2.1 Collegamenti cliccabili (quasi) ovunque!

Nel caso vi fosse sfuggito, vi segnalo che tutti gli argomenti citati nell'indice, nella lista di teoremi, nella lista di definizioni nonché i riferimenti a capitoli/sezioni/teoremi/definizioni specifici e/o indirizzi web in cui vi imbatterete durante la lettura sono cliccabili e vi porteranno direttamente alla risorsa in questione!

1.3 Informazioni sul corso 2015/2016

Per informazioni dettagliate e programma del corso si consulti il sito della docente: `https://www.docenti.unina.it/LUIGIA.CAPUTO`.

1.3.1 Modalità di accertamento del profitto

Lo studente dovrà sostenere un colloquio orale con la commissione. Durante il colloquio il candidato dovrà rispondere a domande teoriche e svolgere esercizi.

1.3.2 Libri di testo consigliati

[1] Sheldon M. Ross, *Probabilità e statistica per l'ingegneria e le scienze*, Apogeo Editore, seconda edizione 2008.

[2] Antonio Di Crescenzo, Virginia Giorno, Amelia Giuseppina Nobile, Luigi Maria Ricciardi, *Un primo corso in probabilità*, Liguori Editore, 2009.

[3] Rita Giuliano, *Elementi di calcolo delle probabilità e statistica*, edizioni ETS, 2007.

Capitolo 2
Statistica descrittiva

La statistica descrittiva è quel ramo delle scienze statistiche che si occupa delle tecniche e degli strumenti finalizzati a raccogliere, descrivere, rappresentare e sintetizzare i dati.

2.1 Alcuni concetti iniziali

Definizione 2.1.1 (Popolazione). *La popolazione è la totalità dei casi (unità statistiche) su cui è possibile individuare uno o più* ***caratteri*** *che rivestono particolare importanza nel fenomeno in esame.*

Definizione 2.1.2 (Carattere). *Un carattere è un tipo di dati su cui si concentra uno studio statistico, cioè una variabile oggetto di studio. Un carattere si può presentare in varie* ***modalità****.*

Definizione 2.1.3 (Modalità). *Le modalità sono i modi in cui si può presentare un carattere. Se il carattere è una domanda, le modalità sono le possibili risposte.*

I caratteri possono essere classificati in:

Caratteri qualitativi nominali: quei caratteri le cui modalità sono etichette non ordinabili. Ad esempio, il colore degli occhi è un carattere qualitativo nominale.

Caratteri qualitativi ordinali: quei caratteri le cui modalità sono etichette ordinabili. Ad esempio, in un test di valutazione, il livello di soddisfazione che si presenta con le modalità *insoddisfatto, soddisfatto, molto soddisfatto* è un carattere qualitativo ordinale.

Caratteri quantitativi discreti: quei caratteri le cui modalità sono numeri interi determinabili con un conteggio. Ad esempio, il numero di maschi e di femmine presenti al corso di Laurea in Informatica sono caratteri quantitativi discreti.

Caratteri quantitativi continui: quei caratteri le cui modalità sono non numerabili, ma continue. Sono quantità che vanno misurate assumendo un'unità di misura e che possono assumere tutti i valori all'interno di un intervallo. Esempi possibili: altezza, peso, reddito.

Definizione 2.1.4 (Taglia della popolazione)**.** *La cardinalità della popolazione, ovvero il numero di unità da cui è composta, prende il nome di* ***taglia*** *della popolazione.*

2.2 Frequenza assoluta e frequenza relativa

Generalmente i caratteri saranno rappresentati con caratteri latini maiuscoli.
Sia Y il carattere in esame e sia N la taglia della popolazione. Effettuando i rilevamenti, si otterrà un *N-pla* $\underline{y} = (y_1, \cdots, y_N)$, dove y_i è l'*i-esimo* dato. Dal momento che nulla vieta che esistano una o più coppie i, j tali che $y_i = y_j$, le modalità distinte possono essere meno dei dati rilevati. In particolare, chiamiamo $x_1, \cdots, x_k$ $(k \leq N)$ le modalità di Y.

Definizione 2.2.1 (Frequenza assoluta di una modalità)**.** *La frequenza assoluta n_i della i-esima modalità x_i è il numero delle unità statistiche su cui la modalità x_i è stata rilevata.*

Dalla definizione segue banalmente che:

- $n_1 + n_2 + \cdots + n_k = \sum_{j=1}^{k} n_j = N$
- $0 \leq n_i \leq N \quad \forall i \in [1, k]$

Definizione 2.2.2 (Frequenza relativa)**.** *La frequenza relativa f_i della i-esima modalità x_i è il valore del rapporto tra n_i ed N. In simboli:*

$$f_i = \frac{n_i}{N}$$

.

Dalle considerazioni già fatte in 2.2.1 segue banalmente che

$$0 \leq f_i = \frac{n_i}{N} \leq 1 \qquad \forall i \in [1, k]$$

.

Definizione 2.2.3 (Frequenza assoluta cumulata di una modalità)**.** *La frequenza assoluta cumulata della modalità x_i è la somma delle frequenze assolute di quella modalità e di tutte quelle che la precedono (in un ordinamento delle modalità stabilito). In simboli:*

$$F_i = \sum_{j=1}^{i} n_j = F_{i-1} + n_i$$

Dalla definizione segue banalmente che $F_k = N$ e $F_1 = n_1$.

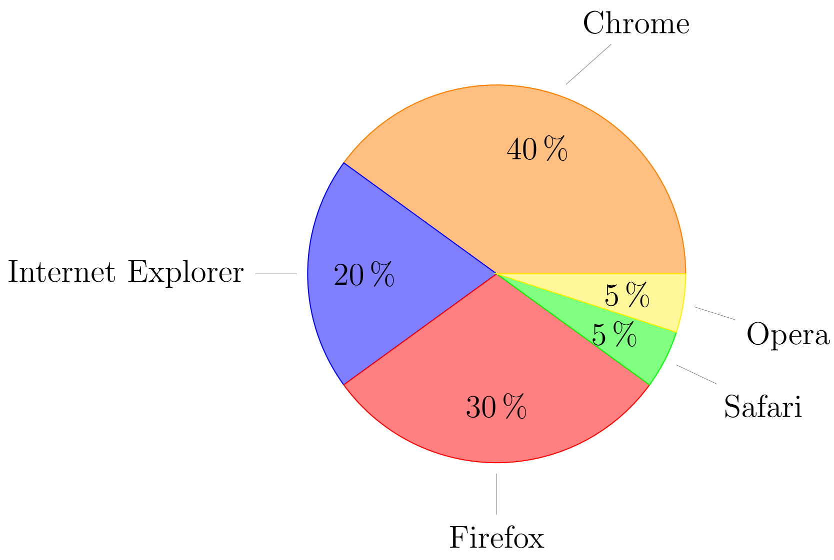


Figura 2.1: Esempio di diagramma circolare

2.3 Rappresentazione dei dati

Supponiamo ancora di studiare un carattere Y su una popolazione di taglia N è che il carattere Y si presenti con le modalità $x_1, \cdots, x_k$ con $(k \leq N)$.

2.3.1 Distribuzione di frequenze

La distribuzione di frequenze è una vista tabellare che permette di rappresentare i dati in modo più leggibile.

	Carattere		
Modalità	n_i	f_i	F_i
x_1	n_1	f_1	$F_1 = n_1$
x_2	n_2	f_2	$F_1 = n_2 + F_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	f_k	$F_k = N$
	tot. N	tot. 1	

Tabella 2.1: Distribuzione di frequenze

Un carattere quantitativo continuo, chiaramente, non potrà essere rappresentato direttamente, ma necessiterà del raggruppamento di modalità in classi (cioè intervalli).

2.3.2 Diagramma circolare

Il diagramma circolare (esempio in 2.1), si presta alla rappresentazione di caratteri qualitativi e quantitativi discreti (o resi discreti da un raggruppamento in classi).

2.3.3 Diagramma a barre verticali

Così come il diagramma circolare, anche i diagrammi a barre si prestano alla rappresentazione di caratteri qualitativi e quantitativi discreti. Di seguito è riportato un esempio.

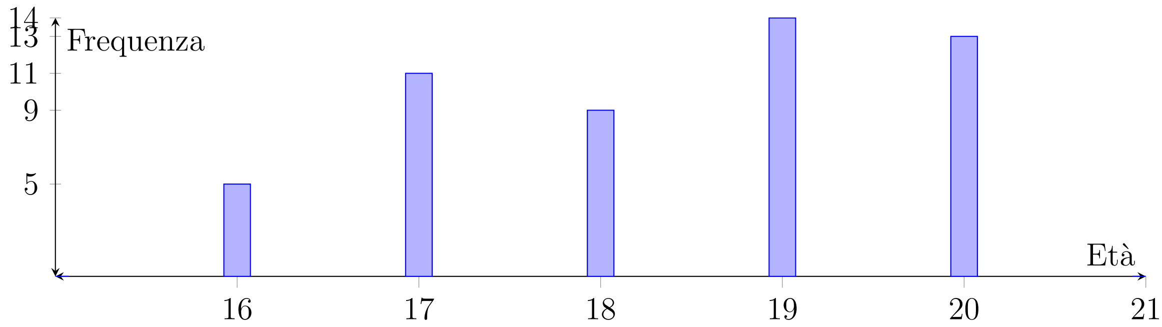


Figura 2.2: Esempio di diagramma a barre verticali

2.3.4 Istogrammi

Un istogramma è una rappresentazione grafica di una distribuzione in classi di un carattere quantitativo continuo. A differenza del diagramma a barre (2.2), le barre non sono separate tra loro proprio in virtù della continuità del carattere rappresentato. L'altezza h_i di ciascuna barra è la **densità di frequenza** di quella classe di modalità, ovvero il rapporto tra la frequenza assoluta di rilevamenti all'interno della classe n_i e l'ampiezza della classe stessa (che indicheremo con b_i). In simboli:

$$h_i = \frac{n_i}{b_i} \Longleftrightarrow n_i = h_i b_i$$

Quindi, in sostanza, la frequenza assoluta della *i-esima* classe è l'area del rettangolo che la rappresenta (e non l'altezza, come avveniva invece nel diagramma a barre).

2.4 Sintesi dei dati: indici statistici

In questa sezione considereremo lo studio di un carattere Y su una popolazione di taglia N. Siano $\underline{y} = (y_1, \cdots, y_n)$ i rilevamenti effettuati e $x_1, \cdots, x_k$ le modalità distinte con cui il carattere Y si presenta.

2.4.1 Indici di posizione

Gli indici di posizione servono ad individurare un elemento in qualche senso "centrale" nella distribuzione.

2.4.2 Media campionaria

Definizione 2.4.1 (Media campionaria)**.** *La media campionaria della distribuzione* $\overline{y} = (y_1, \cdots, y_N)$ *è definita come segue:*

$$\overline{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Teorema 2.4.1 (La media campionaria gode di linearità)**.** *Se* $\forall i \in [1, N]$ $\exists z_i$ *tale che* $y_i = az_i + b$ *con* a, b *costanti , allora* $\overline{y} = a\overline{z} + b$.

Dimostrazione.

$$\begin{aligned}
\underline{y} &= \frac{1}{N}\sum_{i=1}^{N} y_i && \text{(per la definizione di media 2.4.1)}\\
&= \frac{1}{N}\sum_{i=1}^{N}(az_i + b) && \text{(per ipotesi)}\\
&= \frac{1}{N}\sum_{i=1}^{N}(az_i) + \frac{1}{N}\sum_{i=1}^{N} b\\
&= a\frac{1}{N}\sum_{i=1}^{N} z_i + \frac{1}{N}\sum_{i=1}^{N} b\\
&= a\overline{z} + \frac{1}{N}Nb && \text{(si ricordi che } \tfrac{1}{N}\textstyle\sum_{i=1}^{N} z_i = \overline{z}\text{)}\\
&= a\overline{z} + b
\end{aligned}$$

□

Osservazione 2.4.1. *Conoscendo le frequenze assolute della distribuzione si può velocizzare / semplificare il calcolo della media, avendosi:*

$$\overline{y} = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{N}\sum_{i=1}^{k} n_i x_i$$

Osservazione 2.4.2. *L'unità di misura della media campionaria rimane la stessa dei dati. La media campionaria può anche assumere un valore che non corrisponde a nessuna modalità con cui il carattere si presenta.*

2.4.3 Mediana campionaria

Definizione 2.4.2 (Mediana campionaria)**.** *La mediana campionaria è quel dato tra gli* N *disponibili che, previo ordinamento dei dati stessi, è tale che il* 50% *dei dati si trovi prima di esso e il restante* 50% *dopo.*

Si indica con $y_{(1)}$ l'elemento minore tra i valori osservati secondo una qualche relazione d'ordine. In generale, $y_{(k)}$ sarà l'elemento di posto *k-esimo* nell'ordinamento. Di conseguenza, l'ordinamento dei dati rilevati sarà indicabile con $y_{(1)}, y_{(2)}, \cdots, y_{(N)}$.

Dalla definizione 2.4.2 e da quanto appena specificato, si ha quindi che:

- Se N è **dispari**, allora la mediana è un valore tra quelli rilevati e precisamente quello con posizione $\frac{N}{2}$ nell'ordinamento, ovvero $y_{(\frac{N}{2})}$.
- Se N è **pari**, allora si avranno due dati "centrali" $y_{(\frac{N}{2})}$ e $y_{(\frac{N}{2}+1)}$. Tutti i numeri nell'intervallo $\left[y_{\left(\frac{N}{2}\right)},\ y_{\left(\frac{N}{2}+1\right)}\right]$ si trovano nell'**intervallo mediano**. Per convenizione si considera la mediana campionaria come la media dei due valori mediani:

$$Me = \frac{y_{\left(\frac{N}{2}\right)} + y_{\left(\frac{N}{2}+1\right)}}{2}$$

2.4.4 Moda campionaria

Definizione 2.4.3 (Moda campionaria)**.** *La moda campionaria di una distribuzione di dati è il valore che ha frequenza massima. Se la moda è unica, la distribuzione è detta* ***unimodale****. Se vi sono due mode distinte, la distribuzione si dice* ***bimodale****, altrimenti viene definita* ***polimodale / multimodale****. In simboli:*

$$Mo = x_i \text{ per cui } n_i \text{ è tale che } \nexists n_j,\ j \neq i\ :\ n_j > n_i$$

Osservazione 2.4.3. *A differenza della moda campionaria e della mediana campionaria, la moda è* ***sempre*** *un valore rilevato.*

2.4.5 Percentile

Definizione 2.4.4 (Percentile)**.** *Sia* $k \in [0, 100] \cap \mathbb{N}$*, il k-esimo percentile è quel dato della distribuzione ordinata dei dati tale che prima di lui ci sia il* k*% dei dati e dopo il* $(100 - k)$*%.*

Alcuni percentili "famosi":

- **Primo quartile**, corrispondente al **venticinquesimo percentile**.
- **Secondo quartile**, corrispondente al **cinquantesimo percentile**, nonchè alla **mediana campionaria** (2.4.2).
- **Terzo quartile**, corrispondente al **settantacinquesimo percentile**.

2.4.5.1 Box-plot

La rappresentazione dei dati con Box-plot permette di visualizzare il livello di dispersione/concentrazione dei dati. Un box plot, riportato di seguito in figura 2.4.5.1, ha sull'asse le modalità ordinate del carattere. Quindi si rappresenta un rettangolo (la "scatola") tra il primo ed il terzo quartile. La mediana divide questo rettangolo. Infine due segmenti sono tracciati ai lati del rettangolo giungendo rispettivamente fino alla modalità minima e massima.

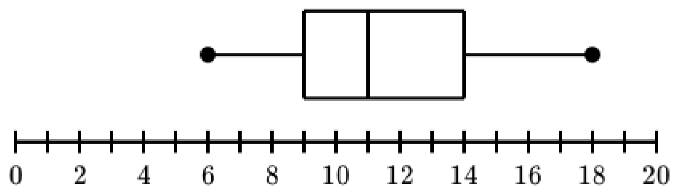


Figura 2.3: Diagramma Box-Plot

Dalla figura 2.4.5.1 si può evincere, ad esempio, che i dati sono maggiormente concentrati tra il primo e il secondo quartile.

2.4.6 Indici di dispersione

Gli indici di dispersione misurano quanto i valori di una distribuzione di discostano da un valore "centrale".

2.4.7 Varianza campionaria

Definizione 2.4.5 (Varianza campionaria). *Dato un insieme di dati* $\underline{y} = (y_1, \cdots, y_N)$ *che assumono valori in* $x_1, \cdots, x_k$ *modalità, con* $k \leq N$*, si definisce varianza campionaria la quantità:*

$$s^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \overline{y})^2$$

Definizione 2.4.6 (Varianza campionaria corretta). *Dato un insieme di dati* $\underline{y} = (y_1, \cdots, y_N)$ *che assumono valori in* $x_1, \cdots, x_k$ *modalità, con* $k \leq N$*, si definisce varianza campionaria corretta la quantità corrispondente alla varianza campionaria semplice moltiplicata per un* ***fattore correttivo*** *corrispondente a* $\frac{N}{N-1}$*. In siboli:*

$$s_c^2 = \left(\frac{N}{N-1} \right) s^2 = \left(\frac{\cancel{N}}{N-1} \right) \frac{1}{\cancel{N}} \sum_{i=1}^{N} (y_i - \overline{y})^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \overline{y})^2$$

D'ora in avanti, faremo sempre riferimento solo alla varianza campionaria corretta, indicandola anche talvolta per semplicità con s^2.

Proposizione 2.4.1 (Forumula alternativa per la varianza). *Sia dato un insieme di dati* $\underline{y} = (y_1, \cdots, y_N)$ *che assumono valori in* $x_1, \cdots, x_k$ *modalità, con* $k \leq N$*. Sia* $\overline{y}$ *la media campionaria. Si ha:*

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i + \overline{y})^2 = \frac{1}{N-1} \left(\sum_{i=1}^{N} y_i^2 + N\overline{y}^2 \right)$$

Dimostrazione.

$$\begin{aligned}
\sum_{i=1}^{N} (y_i - \overline{y})^2 &= \sum_{i=1}^{N} \left(y_i^2 + \overline{y}^2 - 2y_i\overline{y} \right) \\
&= \sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} \overline{y}^2 - \sum_{i=1}^{N} 2y_i\overline{y} \\
&= \sum_{i=1}^{N} y_i^2 + N\overline{y}^2 - 2\overline{y} \sum_{i=1}^{N} y_i \\
&= \sum_{i=1}^{N} y_i^2 + N\overline{y}^2 - 2\overline{y}N\overline{y} \qquad \text{(poichè } \tfrac{1}{n} \textstyle\sum_{i=1}^{N} y_i = \overline{y} \Leftrightarrow \sum_{i=1}^{N} y_i = N\overline{y}\text{)} \\
&= \sum_{i=1}^{N} y_i^2 + N\overline{y}^2 - 2N\overline{y}^2 \\
&= \sum_{i=1}^{N} y_i^2 - N\overline{y}^2
\end{aligned}$$

□

Teorema 2.4.2 (La varianza campionaria (corretta) gode di linearità). *Se* $\forall i \in [1, N] \cap \mathbb{N}, \exists z_i$ *tale che* $y_i = az_i + b$*, con* a, b *costanti, allora*

$$s_y^2 = a^2 s_z^2$$

Dimostrazione.

$$\begin{aligned}
\sum_{i=1}^{N} (y_i - \overline{y})^2 &= \sum_{i=1}^{N} (az_i + b - (a\overline{z} + b))^2 \qquad \text{(usando l'ipotesi e il teorema 2.4.1)} \\
&= \sum_{i=1}^{N} (az_i + \cancel{b} - a\overline{z} - \cancel{b}))^2 \\
&= \sum_{i=1}^{N} (a(z_i - \overline{z})))^2 \\
&= a^2 \sum_{i=1}^{N} (z_i - \overline{z})^2
\end{aligned}$$

Moltiplicando entrambi i membri per $\frac{1}{N-1}$ si ha, applicando la definizione 2.4.6,

$$\left(\frac{1}{N-1}\right) \sum_{i=1}^{N} (y_i - \overline{y})^2 = s_y^2 = \left(\frac{1}{N-1}\right) a^2 \sum_{i=1}^{N} (z_i - \overline{z})^2 = a^2 s_z^2$$

□

Osservazione 2.4.4. *La varianza campionaria non conserva l'unità di misura (distanza quadratica).*

2.4.8 Deviazione standard campionaria

Definizione 2.4.7 (Deviazione standard campionaria). *Dato un insieme di dati* $\underline{y} = (y_1, \cdots, y_N)$ *che assumono valori in* $x_1, \cdots, x_k$ *modalità, con* $k \leq N$*, si definisce deviazione standard campionaria la quantità:*

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (y_i - \overline{y})^2}$$

2.4.9 Coefficiente di correlazione campionaria

Il coefficiente di correlazione campionaria misura il grado di correlazione tra due caratteri.

Definizione 2.4.8 (Coefficiente di correlazione campionaria). *Siano Y e Z i caratteri da studiare, sia N la taglia della popolazione e siano $\underline{y} = (y_1, \cdots, y_N)$, $\underline{z} = (z_1, \cdots, z_N)$ i rilevamenti effettuati. Si definisce coefficiente di correlazione campionaria la quantità:*

$$
\begin{aligned}
r &= \frac{1}{N-1} \frac{\sum_{i=1}^{N}(y_i - \overline{y})(z_i - \overline{z})}{s_y s_z} \\
&= \frac{\sum_{i=1}^{N}(y_i - \overline{y})(z_i - \overline{z})}{(N-1)\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y})^2}\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(z_i - \overline{z})^2}} \\
&= \frac{\sum_{i=1}^{N}(y_i - \overline{y})(z_i - \overline{z})}{(N-1)\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y})^2 \frac{1}{N-1}\sum_{i=1}^{N}(z_i - \overline{z})^2}} \\
&= \frac{\sum_{i=1}^{N}(y_i - \overline{y})(z_i - \overline{z})}{\sqrt{(N-1)^2\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y})^2 \frac{1}{N-1}\sum_{i=1}^{N}(z_i - \overline{z})^2}} \\
&= \frac{\sum_{i=1}^{N}(y_i - \overline{y})(z_i - \overline{z})}{\sqrt{\sum_{i=1}^{N}(y_i - \overline{y})^2 \sum_{i=1}^{N}(z_i - \overline{z})^2}}
\end{aligned}
$$

Osservazione 2.4.5. *Si ha che:*

- $-1 \leq r \leq 1$.
- *se* $r > 0$ *i dati sono correlati positivamente (direttamente proporzionali).*
- *se* $r < 0$ *i dati sono correlati negativamente (inversamente proporzionali).*
- *se* $r = 0$ *i dati non sono correlati.*
- *se* $r \approx 1$ *i dati sono positivamente correlati.*
- *se* $r \approx -1$ *i dati sono negativamente correlati.*
- *se* $r \approx 0$ *i dati non sono correlati.*
- *se* $r \in (-0.5, 0.5)$ *i dati sono scarsamente correlati.*

2.4.10 Indici di forma

Sono indici di confronto di una distribuzione con la distribuzione normale (o distribuzione di Gauss).

2.4.11 Skewness

Definizione 2.4.9 (Skewness)**.** *La skewness è un indice di asimmetria. È definita come la quantità:*

$$Skewness = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i - \overline{y}}{s}\right)^3$$

Osservazione 2.4.6. *Si ha che*

$$\begin{aligned} Skewness &> 0 \Longleftrightarrow \textit{coda a destra} \\ &< 0 \Longleftrightarrow \textit{coda a sinistra} \\ &= 0 \Longleftrightarrow \textit{distribuzione simmetrica} \end{aligned}$$

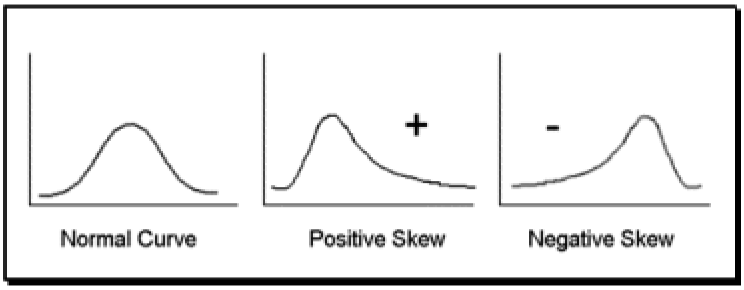


Figura 2.4: Raffigurazione di distribuzioni con diversa skewness

2.4.12 Curtosi

Definizione 2.4.10 (Curtosi)**.** *La curtosi è un indice di allontanamento dalla distribuzione normale, rispetto alla quale si verifica un maggiore appiattimento (**distribuzione platicurtica**) o un maggiore allungamento (**distribuzione leptocurtica**)*

$$Curtosi = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i - \overline{y}}{s}\right)^4 - 3$$

Osservazione 2.4.7. *Si ha che*

$$\begin{aligned} Curtosi &> 0 \iff \text{distribuzione “più appuntita” (leptocurtica)} \\ &< 0 \iff \text{distribuzione “meno appuntita” (platicurtica)} \\ &= 0 \iff \text{distribuzione simmetrica} \end{aligned}$$

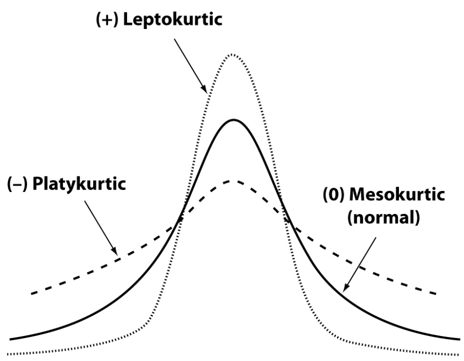


Figura 2.5: Raffigurazione di distribuzioni con diversa curtosi

Capitolo 3

Calcolo delle probabilità

3.1 Esperimenti casuali

Definizione 3.1.1 (Esperimento casuale o aleatorio)**.** *Un esperimento casuale è un esperimento per cui non è possibile prevedere a priori il risultato. Ad ogni esperimento casuale è possibile associare una terna* $(\Omega, \mathcal{F}, \mathbb{P})$.

3.2 Spazio degli esiti possibili

Definizione 3.2.1 (Spazio degli esiti possibili o spazio campione)**.** *Lo spazio degli esiti possibili (o spazio campione)* Ω *dell'esperimento in esame è l'insieme dei possibili esiti dell'esperimento.*

Osservazione 3.2.1.

Ω *può essere un insieme finito oppure infinito (discreto o continuo).*

3.3 Eventi

Definizione 3.3.1 (Evento)**.** *Un evento è un sottoinsieme dello spazio degli esiti possibili* Ω.

Consideriamo come esperimento casuale il lancio di un dado a sei facce. Si avrà chiaramente che $\Omega = \{1, 2, 3, 4, 5, 6\}$. Consideriamo ora gli eventi:
$E = \{1, 3, 5\} =$"In seguito al lancio del dado, questo si ferma rivolgendo verso l'alto una faccia con un numero di pallini dispari"
$F = \{2, 4, 6\} =$"In seguito al lancio del dado, questo si ferma rivolgendo verso l'alto una faccia con un numero di pallini pari"

$G = \{3, 6\} =$"In seguito al lancio del dato, questo si ferma rivolgendo verso l'alto una faccia con un numero di pallini multiplo di tre"
Abbiamo che:

- $E \cup F$ è l'insieme degli oggetti che sono in E oppure in F. L'evento $E \cup F$ si verifica se si verifica almeno uno tra E e F.
- $E \cap F$ è l'insieme degli oggetti che sono sia in E che in F. L'evento $E \cap F$ si verifica se si verificano sia E che F.
- Se $E \cap F = \varnothing$, E e G non hanno elementi in comune (si dicono **disgiunti**). L'evento $E \cap F = \varnothing$ è impossibile.
- $E^c = \overline{E}$ è l'insieme degli oggetti che appartengono ad Ω ma non ad E. L'evento $\overline{E}$ si verifica se non si verifica l'evento E.

3.3.1 Alcune proprietà

- **Proprietà commutativa**

$$E \cup (F \cup G) = (E \cup F) \cup G$$
$$E \cap (F \cap G) = (E \cap F) \cap G$$

- **Proprietà distributiva**

$$E \cup (F \cap G) = (E \cup F) \cap (E \cup G)$$
$$E \cap (F \cup G) = (E \cap F) \cup (E \cap G)$$

- **Leggi di *De Morgan***

$$\overline{(E \cup F)} = \overline{E} \cap \overline{F} \text{ oppure } (E \cup F) = \overline{\overline{E} \cap \overline{F}}$$
$$\overline{(E \cap F)} = \overline{E} \cup \overline{F} \text{ oppure } (E \cap F) = \overline{\overline{E} \cup \overline{F}}$$

Definizione 3.3.2 (Unione e intersezione numerabili)**.** *Si definisce l'unione numerabile di una famiglia numerabile di insiemi l'insieme:*

$$\bigcup_{n \in \mathbb{N}} E_n = E_1 \cup E_2 \cup \cdots \cup E_n \cup \cdots$$

Si definisce l'intersezione numerabile di una famiglia numerabile di insiemi l'insieme:

$$\bigcap_{n \in \mathbb{N}} E_n = E_1 \cap E_2 \cap \cdots \cap E_n \cap \cdots$$

3.4 La famiglia di insiemi $\mathcal{F}$

Nella terna citata in 3.1.1, $\mathcal{F}$ è una famiglia di insiemi tale che:

1. $\Omega \in \mathcal{F}$ (ciò assicura che $\mathcal{F}$ sia non vuota.)
2. $F \in \mathcal{F} \Longrightarrow \overline{F} \in \mathcal{F}$
3. Data una famiglia numerabile di insiemi $\{F_n\}_{n\in\mathbb{N}}$, allora $F_n \in \mathcal{F}\ \forall n \in \mathbb{N} \Longrightarrow \bigcup_{n\in\mathbb{N}} F_n \in \mathcal{F}$

La famiglia $\mathcal{F}$ è chiusa rispetto all'intersezione numerabile e all'unione numerabile. La famiglia che soddisfa le tre proprietà sopra è detta anche σ-algebra di Ω.

3.5 Probabilità

3.5.1 Definizione di probabilità legata alla frequenza

Definizione 3.5.1 (Definizione di probabilità legata alla frequenza)**.** *(Legge empirica del caso) Al crescere delle ripetizioni di un esperimento aleatorio (sempre nelle medesime condizioni!) la frequenza relativa di un esito (o di un insieme di esiti, quindi di un evento) si attesta intorno ad un valore costante. Tale valore costante è la* ***probabilità****. Premettendo che il limite utilizzato di seguito non è quello puramente "matematico", in simboli potremmo scrivere:*

$$\mathbb{P}(E) = \lim_{n\to\infty} f_E = \lim_{n\to\infty} \frac{n_E}{n}$$

Dalla definizione 3.5.1 seguono tre proprietà:

1. $0 \leq \mathbb{P}(E) \leq 1$

   Dimostrazione. La dimostrazione segue banalmente dal fatto che $0 \leq f_E \leq 1$ □

2. $\mathbb{P}(\Omega) = 1$

   Dimostrazione.
   $$f_\Omega = \frac{n_\Omega}{n} = \frac{n}{n} = 1 \Longrightarrow \mathbb{P}(\Omega) = 1$$
   □

3. Se E ed F sono eventi disgiunti, allora $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$

Dimostrazione.

$$\begin{aligned} f_{E\cup F} &= \frac{n_{E\cup F}}{n} \\ &= \frac{n_E + n_F}{n} && (E \text{ ed } F \text{ sono disgiunti per ipotesi}) \\ &= \frac{n_E}{n} + \frac{n_F}{n} = \mathbb{P}(E) + \mathbb{P}(F) \end{aligned}$$

□

3.5.2 Definizione classica di probabilità

Definizione 3.5.2 (Definizione classica della probabilità)**.** *La definizione classica di probabilità si può applicare solo quando gli esiti possibili sono equiprobabili. Indicato con E un esito, la probabilità di E è definita come:*

$$\mathbb{P}(E) = \frac{\#\ \textit{casi favorevoli}}{\#\ \textit{casi possibili}} = \frac{|E|}{|\Omega|}$$

Mostriamo che anche dalla definizione 3.5.2 seguono le stesse tre proprietà già dimostrate per 3.5.1:

1. $0 \leq \mathbb{P}(E) \leq 1$

   Dimostrazione. Poichè $|\Omega| > 0$ affinchè l'esperimento abbia senso e $0 \leq |E| \leq |\Omega|$, allora $0 \leq \frac{|E|}{|\Omega|} = \mathbb{P}(E) \leq 1$ □

2. $\mathbb{P}(\Omega) = 1$

   Dimostrazione.
   $$\mathbb{P}(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$$
   □

3. Se E ed F sono eventi disgiunti, allora $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$

   Dimostrazione.
   $$\begin{aligned} \mathbb{P}(E\cup F) &= \frac{|E\cup F|}{|\Omega|} \\ &= \frac{|E| + |F|}{|\Omega|} && (E \text{ ed } F \text{ sono disgiunti per ipotesi}) \\ &= \frac{|E|}{|\Omega|} + \frac{|F|}{|\Omega|} = \mathbb{P}(E) + \mathbb{P}(F) \end{aligned}$$
   □

3.5.3 Definizione soggettiva di probabilità

Definizione 3.5.3 (Definizione soggettiva di probabilità)**.** *Si definisce probabilità di un evento il prezzo (la posta) che un individuo ritiene equo pagare per ricevere 1 se l'evento si verifica e 0 se l'evento non si verifica.*

3.6 Calcolo combinatorio

3.6.1 Principio di enumerazione

Teorema 3.6.1 (Principio di enumerazione)**.** *Se consideriamo la realizzazione di due diversi esperimenti di cui il primo ha n esiti possibili ed il secondo m esiti possibili, allora l'*esperimento totale *ha $n \cdot m$ esiti possibili.*

Dimostrazione. Ogni esito dell'esperimento totale sarà una coppia di esiti $(E_1,\ E_2)$, con E_1 esito del primo esperimento ed E_2 del secondo. È possibile elencare tutti i possibili esiti con una matrice $n \times m$:

$$\begin{pmatrix} (1,1) & (1,2) & \cdots & (1,m) \\ (2,1) & (2,2) & \cdots & (2,m) \\ \vdots & \vdots & \ddots & \vdots \\ (n,1) & (n,2) & \cdots & (n,m) \end{pmatrix}$$

Dal momento che abbiamo enumerato tutti possibili esiti dell'esperimento totale usando una matrice $n \times m$, il numero di detti esiti sarà $n \cdot m$. □

Teorema 3.6.2 (Principio di enumerazione generalizzato)**.** *Supponiamo di avere $r \in \mathbb{N}$ esperimenti. Il primo degli r esperimenti ha n_1 esiti possibili. Per ciascuno di questi n_1 esiti, il secondo esperimento ha n_2 esiti possibili. Per ciascuna possibile combinazione degli esiti dei primi due esperimenti il terzo ha n_3 esiti possibili. Sotto queste ipotesi, vi sono un totale di $n_1 \times n_2 \times \cdots \times n_r = \prod_{i=1}^{r} n_i$ combinazioni degli esiti degli r esperimenti.*

Dimostrazione. Dimostrazione per induzione sul numero di esperimenti r. Se $r = 2$, il caso base è stato già dimostrato in 3.6.1. Sia per ipotesi induttiva il teorema valido su $r - 1$ elementi. Allora avremo che il numero di combinazioni degli esiti di questi $r - 1$ elementi è $E_{r-1} = n_1 \times n_2 \times \cdots \times n_{r-1}$. Considerando le possibili combinazioni degli $r - 1$ esperimenti come un unico esperimento, possiamo applicare nuovamente il passo base, avendosi che $E_r = E_{r-1} \times n_r = n_1 \times n_2 \times \cdots \times n_{r-1} \times n_r$, provando il teorema per ogni $n \in \mathbb{N}$. □

3.6.2 Permutazioni

Definizione 3.6.1 (Permutazione)**.** *Ogni configurazione di n distinti elementi in n posti senza ripetizioni è una permutazione (l'ordine degli elementi è rilevante!). Il numero di permutazioni di n elementi è*

$$P_n = n!$$

.

Intuitivamente, si consideri di voler posizionare n elementi in n slot. Per il primo di questi slot si avranno n possibili scelte, per il secondo $n-1$, e così via fino all'ultimo, la cui scelta sarà vincolata (ultimo elemento rimasto). Da ciò e dal principio di enumerazione generalizzato 3.6.2 segue la formula.

Definizione 3.6.2 (Permutazione con ripetizioni)**.** *Ogni configurazione di n distinti elementi in n posti con possibili ripetizioni è una permutazione con ripetizioni (l'ordine degli elementi è rilevante!). Il numero di permutazioni con ripetizioni di n elementi è*

$$P_n^r = n^n$$

.

Intuitivamente, si consideri di voler posizionare n elementi in n slot. Essendo possibili ripetizioni, per ogni slot si avranno n possibili scelte. Da ciò e dal principio di enumerazione generalizzato 3.6.2 segue la formula.

3.6.3 Disposizioni

Definizione 3.6.3 (Disposizione)**.** *Sia S un insieme con n elementi distinti. k sia il numero di slot da riempire (numerosità del gruppo da scegliere tra gli n elementi). Ciascuna configurazione di n oggetti distinti presi a gruppi di k senza ripetizioni è detta disposizione (l'ordine degli elementi è rilevante!).*
Il numero di disposizioni senza ripetizioni di n oggetti a gruppi di k è

$$D_{n,k} = n \times (n-1) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!}$$

Intuitivamente, si consideri di voler posizionare n elementi in k slot senza ripetizioni. Per riempire il primo slot si avranno n possibili scelte, per il secondo slot $(n-1)$ possibili scelte e così via fino al *k-esimo slot*, per il quale si avranno $(n-k+1)$ possibili scelte. Da ciò e dal principio di enumerazione generalizzato 3.6.2 segue la formula.

Definizione 3.6.4 (Disposizione con ripetizioni)**.** *Sia S un insieme con n elementi distinti. k sia il numero di slot da riempire (numerosità del gruppo da scegliere tra gli n*

elementi). Ciascuna configurazione di n oggetti distinti presi a gruppi di k con possibili ripetizioni è detta disposizione con ripetizioni (l'ordine degli elementi è rilevante!).
Il numero di disposizioni con ripetizioni di n oggetti a gruppi di k è

$$D^r_{n,k} = n \times (n) \times \cdots \times (n) = n^k$$

Intuitivamente, si consideri di voler posizionare n elementi in k slot con possibili ripetizioni. Per riempire ciascuno dei k slot si avranno n possibili scelte. Da ciò e dal principio di enumerazione generalizzato 3.6.2 segue la formula.

3.6.4 Combinazioni

Definizione 3.6.5 (Combinazione semplice)**.** *Siano n e k due numeri interi con $k \leq n$. Si definisce combinazione di n elementi presi a gruppi di k (o alternativamente combinazione di n elementi di classe k) ciascuno dei sottoinsiemi di cardinalità k presi da un insieme di n elementi. I sottoinsiemi sono indipendenti dall'ordine degli elementi! Si ha in particolare che il numero di combinazioni semplici di n elementi presi in gruppi di k è*

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

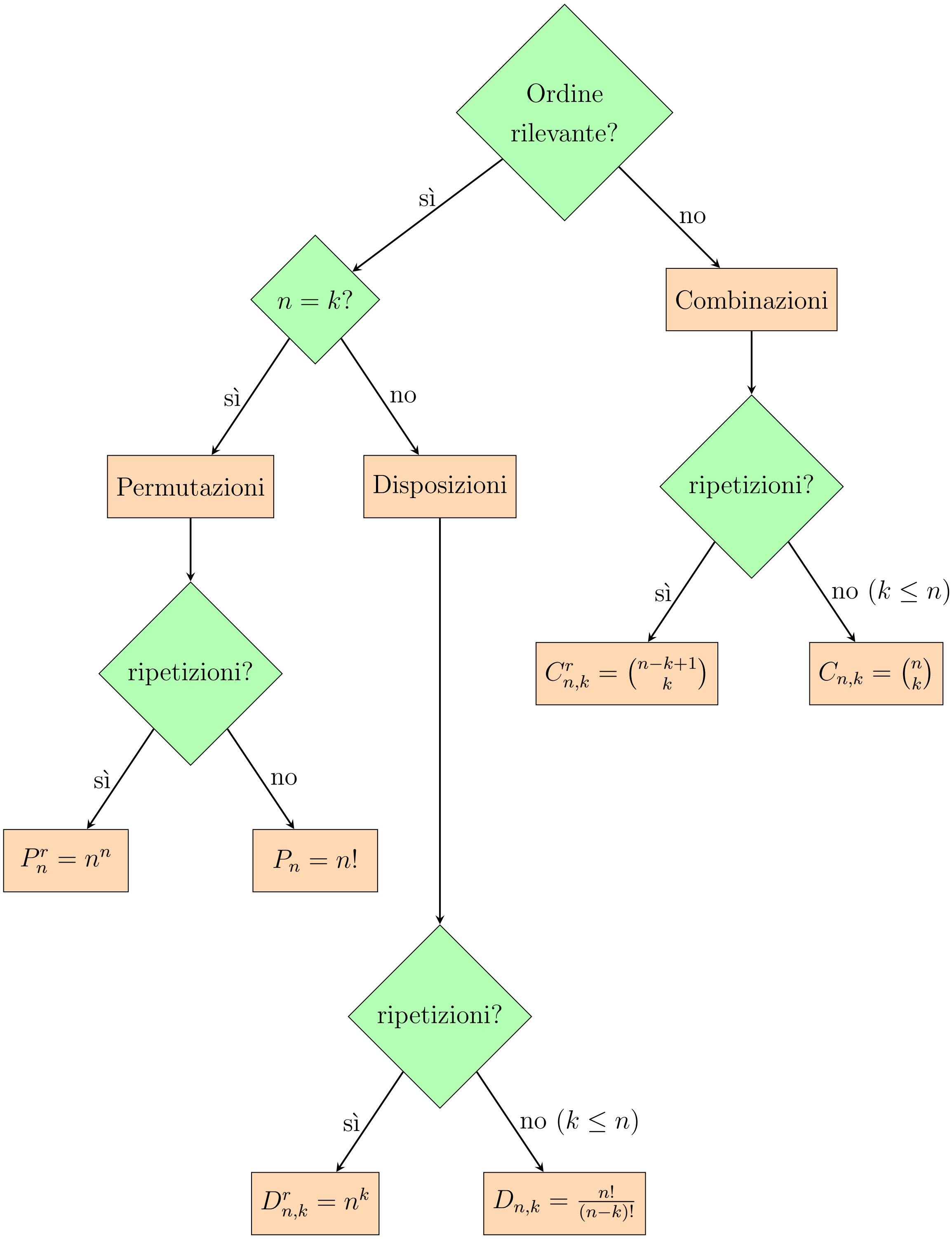
Si consideri che $D_{n,k}$ è il numero di configurazioni di n elementi senza ripetizioni in gruppi di k, tenendo conto dell'ordine degli elementi. In $D_{n,k}$ ciascun sottoinsieme di k elementi è contato tante volte quante sono le permutazioni possibili di k elementi. Dividendo $D_{n,k}$ per tale numero di permutazioni, si ottiene quindi il numero di combinazioni desiderato.

Definizione 3.6.6 (Combinazione con ripetizioni)**.** *Siano n e k due numeri interi con $k \leq n$. Si definisce combinazione con ripetizioni di n elementi presi a gruppi di k (o alternativamente combinazione di n elementi di classe k) ciascuno dei gruppi di cardinalità k presi da un insieme di n elementi, con possibili ripetizione. I gruppi sono indipendenti dall'ordine degli elementi! Si ha*[1] *in particolare che il numero di combinazioni con ripetizioni di n elementi presi in gruppi di k è*

$$C^r_{n,k} = C_{n-k+1,k} = \binom{n-k+1}{k}$$

[1]Per una dimostrazione si consulti `https://it.wikipedia.org/wiki/Combinazione#Combinazioni_con_ripetizione`

3.6.5 Quadro generale sul calcolo combinatorio



3.7 Definizione assiomatica della probabilità

Definizione 3.7.1 (Definizione assiomatica della probabilità). *(Impostazione assiomatica di Kolmogorov) La probabilità è una funzione*

$$\mathbb{P} : \mathcal{F} \to [0, +\infty)$$

che soddisfa le seguenti tre proprietà:

i) $\forall A \in \mathcal{F},\ 0 \leq \mathbb{P}(A) \leq 1$

ii) $\mathbb{P}(\Omega) = 1$

iii) $\{F_n\}_{n\in\mathbb{N}}\ F_n \in \mathcal{F}\ \forall n \in \mathbb{N},\ F_i \cap F_j = \varnothing\ \forall i \neq j \Rightarrow \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} F_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(F_n)$

Osservazione 3.7.1. *Nel terzo assioma (**additività numerabile**), il fatto che la funzione $\mathbb{P}$ possa applicarsi a $\bigcup_{n\in\mathbb{N}} F_n$ è garantito dalla proprietà di chiusura rispetto all'unione numerabile di $\mathcal{F}$, già descritta in 3.4.*

Teorema 3.7.1 (Additività numerabile implica additività finita). *Consideriamo la famiglia finita di insiemi $F_1, F_2, \cdots, F_m,\ F_i \in \mathcal{F}\ \forall i \in [1, m]$, con m finito e tale che $F_i \cap F_j = \varnothing\ \ \forall i \neq j$. Allora $\mathbb{P}(\bigcup_{i=1}^m F_i) = \sum_{i=1}^m \mathbb{P}(F_i)$.*

Dimostrazione. Costruiamo una successione numerabile $\{G_n\}_{n\in\mathbb{N}}$ partendo da quella finita che sia equivalente a quest'ultima.

$$G_1 = F_1,\ G_2 = F_2,\ \cdots,\ G_m = F_m,\ G_{m+1} = \varnothing,\ \cdots,\ G_n = \varnothing, \cdots$$

La famiglia finita di insiemi di partenza è stata resa numerabile aggiungendovi infiniti insiemi vuoti. Ciascuno degli insiemi vuoti aggiunti, chiaramente, sarà disgiunto dai precedenti e ciò implica che alla famiglia numerabile $\{G_n\}_{n\in\mathbb{N}}$ si può applicare l'assioma di additività numerabile. Quindi:

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} G_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(G_n) = \sum_{n=1}^{m} \mathbb{P}(G_n) + \sum_{n=m+1}^{\infty} \mathbb{P}(G_n)$$

Notiamo però che:

i) $\displaystyle\bigcup_{n\in\mathbb{N}} G_n = \bigcup_{n=1}^{m} F_n$ per come abbiamo costruito la famiglia G

ii) $\displaystyle\sum_{n=m+1}^{\infty} \mathbb{P}(G_n) = 0$, essendo tutti i G_n con $n > m$ uguali a $\varnothing$

iii) $\displaystyle\sum_{n=1}^{m} \mathbb{P}(G_n) = \sum_{n=1}^{m} \mathbb{P}(F_n)$ per come abbiamo costruito la famiglia G

Quindi si ha:

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} G_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{m} F_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(G_n) = \sum_{n=1}^{m} \mathbb{P}(G_n) + \sum_{n=m+1}^{\infty} \mathbb{P}(G_n) = \sum_{n=1}^{m} \mathbb{P}(F_n)$$

come volevasi dimostrare.

$\square$

Teorema 3.7.2 (Probabilità dell'evento complementare)**.** *Dato* $A \in \mathcal{F}$*, si ha che* $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$.

Dimostrazione.
Chiaramente si ha che $\Omega = A \cup A^C$, quindi $\mathbb{P}(\Omega) = \mathbb{P}(A \cup A^C) \overset{\textit{Per teo. } 3.7.1}{=} \mathbb{P}(A) + \mathbb{P}(A^C)$.
Ma per l'assioma *ii*) della defīnzione assiomatica 3.7.1 si ha che $\mathbb{P}(\Omega) = 1$, quindi:
$\mathbb{P}(\Omega) = 1 = \mathbb{P}(A) + \mathbb{P}(A^C) \iff \mathbb{P}(A^C) = 1 - \mathbb{P}(A)$, concludendo la dimostrazione. $\square$

Teorema 3.7.3 (Probabilità dell'evento impossibile)**.** *La probabilità dell'evento impossibile* $\mathbb{P}(\varnothing) = 0$.

Dimostrazione. Semplicemente consideriamo che

$$\mathbb{P}(\varnothing) = 1 - \mathbb{P}(\varnothing^C) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0$$

$\square$

Teorema 3.7.4 (Probabilità dell'unione di due eventi qualsiasi)**.** *Dati* $E, F \in \mathcal{F}$ *(*E, F *qualsiasi, non necessariamente disgiunti), allora* $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$.

Dimostrazione. Iniziamo provando ad esprimere $\mathbb{P}(E \cup F)$ come unione disgiunta di due insiemi:

$$\begin{aligned} \mathbb{P}(E \cup F) &= \mathbb{P}(E \uplus (F \cap E^C)) \\ &= \mathbb{P}(E) + \mathbb{P}(F \cap E^C) \qquad (\maltese) \end{aligned}$$

quindi consideriamo che

$$\begin{aligned} \mathbb{P}(F) &= \mathbb{P}\left((F \cap E) \uplus (F \cap E^C)\right) \\ &= \mathbb{P}(F \cap E) + \mathbb{P}(F \cap E^C) \iff \mathbb{P}(F \cap E^C) = \mathbb{P}(F) - \mathbb{P}(F \cap E) \qquad (\star) \end{aligned}$$

Sostituendo il risultato trovato in $(\star)$ in $(\maltese)$, segue la tesi

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F \cap E^C) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$$

$\square$

Teorema 3.7.5 (Probabilità dell'unione di tre eventi qualsiasi)**.** *Dati tre eventi (non necessariamente a due a due disgiunti)* $E, F, G \in \mathcal{F}$ *, si ha che*

$$\mathbb{P}(E \cup F \cup G) = \mathbb{P}(E) + \mathbb{P}(F) + \mathbb{P}(G) - \mathbb{P}(E \cap F) - \mathbb{P}(E \cap G) - \mathbb{P}(F \cap G) + \mathbb{P}(E \cap F \cap G)$$

Dimostrazione. Poichè $\mathcal{F}$ è una σ-algebra, $(E \cup F) \in \mathcal{F}$. Sia quindi $A = (E \cup F)$. Si ha

$$\begin{aligned}
\mathbb{P}(E \cup F \cup G) &= \mathbb{P}(A \cup G) = \mathbb{P}(A) + \mathbb{P}(G) - \mathbb{P}(A \cap G) && \text{(applicando teo. 3.7.4)} \\
&= \mathbb{P}(E \cup F) + \mathbb{P}(G) - \mathbb{P}\left((E \cup F) \cap G\right) && \text{(sostituendo } A = E \cup F\text{)} \\
&\text{Applicando la proprietà distributiva dell'unione sull'intersezione} \\
&= \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F) + \mathbb{P}(G) - \left[\mathbb{P}\left((E \cap G) \cup (F \cap G)\right)\right] \\
&\text{Applicando ancora il teorema 3.7.4} \\
&= \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F) + \mathbb{P}(G) - [\mathbb{P}(E \cap G) + \mathbb{P}(F \cap G) - \mathbb{P}(E \cap F \cap G)] \\
&= \mathbb{P}(E) + \mathbb{P}(F) + \mathbb{P}(G) - \mathbb{P}(E \cap F) - \mathbb{P}(E \cap G) - \mathbb{P}(F \cap G) + \mathbb{P}(E \cap F \cap G)
\end{aligned}$$

□

Teorema 3.7.6 (Formula di inclusione-esclusione generalizzata per n eventi)**.** *Siano* $F_1, F_2, \cdots, F_n \in \mathcal{F}$*. Allora*

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} F_i\right) &= \sum_{i=1}^{n} \mathbb{P}(F_i) - \sum_{1 \le i_1 < i_2 \le n} \mathbb{P}(F_{i_1} \cap F_{i_2}) + \cdots \\
&+ (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \cdots < i_r \le n} \mathbb{P}(F_{i_1} \cap F_{i_2} \cap \cdots \cap F_{i_k}) + \cdots \\
&+ (-1)^{n+1} \mathbb{P}(F_1 \cap F_2 \cap \cdots \cap F_n)
\end{aligned}$$

Teorema 3.7.7 (Proprietà di monotonia)**.** *Siano* $A, B \in \mathcal{F}$ *due eventi tali che* $A \subseteq B$*. Ciò implica che* $\mathbb{P}(A) \le \mathbb{P}(B)$

Dimostrazione. $B = A \cup (B \cap A^C) \overset{\text{teo. 3.7.1}}{\Longrightarrow} \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^C)$
Notiamo che (thanks Postiglione) $B \cap A^C \overset{\text{de Morgan}}{=} (B^C \cup A)^C$. Poichè $\mathcal{F}$ è una σ-algebra (vedi 3.4), allora $B^C \in \mathcal{F}$ e anche l'unione $(B^C \cup A) \in \mathcal{F}$. Se l'unione $(B^C \cup A) \in \mathcal{F}$, allora apperterrà ad $\mathcal{F}$ anche il complemento di quest'ultima. Usando ancora una volta de Morgan abbiamo che $(B^C \cup A)^C = (B \cap A^C)$, dunque possiamo concludere che $B \cap A^C \in \mathcal{F}$.e, di conseguenza, che $0 \le \mathbb{P}(B \cap A^C) \le 1$ per il primo assioma di 3.7.1. Possiamo quindi minorare l'espressione di partenza come segue

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^C) \ge \mathbb{P}(A) + 0 = \mathbb{P}(A)$$

□

3.8 Probabilità condizionata

Definizione 3.8.1 (Probabilità condizionata)**.** *Considerato un esperimento caratterizzato da $(\Omega, \mathcal{F}, \mathbb{P})$, siano $A, B \in \mathcal{F}$ due eventi. Se $\mathbb{P}(B) > 0$, si definisce la probabilità condizionata di A rispetto a B e si indica con $\mathbb{P}(A|B)$ oppure con $\mathbb{P}_B(A)$ la quantità*

$$\mathbb{P}(A|B) = \mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{\#\ \text{casi fav. } A \cap B}{\#\ \text{casi possibili}}}{\frac{\#\ \text{casi fav. } B}{\#\ \text{casi possibili}}} = \frac{\#\ \text{casi fav. } A \cap B}{\#\ \text{casi fav. } B}$$

Teorema 3.8.1 (La probabilità condizionata è una probabilità)**.** *Mostriamo che la probabilità condizionata appena definita è una probabilità, cioè rispetta gli assiomi di Kolmogorov (3.7.1).*

Dimostrazione. Innanzitutto notiamo che $\mathbb{P}_B : \mathcal{F} \to [0, +\infty)$. Ciò segue dalla definizione di $\mathbb{P}_B$ e dal fatto che $\mathbb{P}(F) \geq 0$ per il primo assioma della definizione assiomatica. Infatti

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B) \geq 0}{\mathbb{P}(B) > 0} \Rightarrow \mathbb{P}_B(A) \geq 0$$

Proseguendo, dimostriamo la validità degli assiomi per $\mathbb{P}_B(A)$

i) $\forall A, B \in \mathcal{F},\ \mathbb{P}(B) > 0$ si ha che $0 \leq \mathbb{P}_B(A) \leq 1$
Per definizione si avrà $\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. Inoltre, pochè $(A \cap B) \subseteq B$, si avrà per la proprietà di monotonia 3.7.7 che $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$. Dunque è possibile maggiorare il rapporto espresso prima in questo modo:

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$$

Banalmente dalle proprietà di $\mathbb{P}$ segue che $\mathbb{P}_B(A)$ non può assumere valori < 0.

ii) $\mathbb{P}_B(\Omega) = 1$
Applicando la definizione si ha che

$$\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$$

iii) Data una famiglia numerabile di insiemi $\{F_n\}_{n\in\mathbb{N}},\ F_n \in \mathcal{F}$ mostriamo che

$$\mathbb{P}_B\left(\dot{\bigcup_{n\in\mathbb{N}}} F_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}_B(F_n)$$

Applichiamo la definizione

$$\begin{aligned}
\mathbb{P}_B\left(\dot{\bigcup_{n\in\mathbb{N}}} F_n\right) &= \frac{\mathbb{P}\left(\left(\dot{\bigcup}_{n\in\mathbb{N}} F_n\right)\cap B\right)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}\left(\dot{\bigcup}_{n\in\mathbb{N}} (F_n\cap B)\right)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(F_1\cap B)}{\mathbb{P}(B)} + \frac{\mathbb{P}(F_2\cap B)}{\mathbb{P}(B)} + \cdots + \frac{\mathbb{P}(F_k\cap B)}{\mathbb{P}(B)} + \cdots \\
&= \mathbb{P}_B(F_1) + \mathbb{P}_B(F_2) + \cdots + \mathbb{P}_B(F_k) + \cdots \\
&= \sum_{n\in\mathbb{N}} \mathbb{P}_B(F_n)
\end{aligned}$$

Da *i*), *ii*), *iii*) appena dimostrati segue la tesi. □

3.8.1 Formula della probabilità totale

Definizione 3.8.2 (Formula della probabilità totale)**.** *Invertendo la definizione di* $\mathbb{P}_B(A)$ *possiamo ottenere l'espressione della* $\mathbb{P}(B\cap A)$*:*

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \Rightarrow \mathbb{P}(A\cap B) = \mathbb{P}(A|B)\cdot\mathbb{P}(B)$$

La formula si può generalizzare per n *eventi a probabilità non nulla:*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\cdot\mathbb{P}(A_2|A_1)\cdot\mathbb{P}(A_3|A_1\cap A_2) + \cdots + \mathbb{P}(A_n|A_1\cap A_2\cap\cdots\cap A_{n-1})$$

Esercizio 3.8.1 (Gioco dell'estrazione di palline colorate (formula prob. totale))**.** *In un urna vengono poste 5 palline, di cui 4 bianche ed una nera. Le palline sono tra loro indistinguibili al tatto. Il primo giocatore estrae una pallina dall'urna. Se la pallina estratta è nera, il giocatore vince, altrimenti la pallina estratta viene scartata e il gioco prosegue con il secondo giocatore estrae un'altra pallina.*
Calcoliamo le probabilità di vittoria di ciascun giocatore.

Consideriamo gli eventi $V_i =$ L'*i-esimo* giocatore vince, estraendo la pallina nera. Agli eventi possiamo applicare la definizione classica di probabilità.

$$
\begin{aligned}
\mathbb{P}(V_1) &= \frac{1}{5} \\
\mathbb{P}(V_2) &= \mathbb{P}(V_2 \cap \Omega) = \mathbb{P}\left[V_2 \cap (V_1 \cup V_1^C)\right] \\
&= \mathbb{P}\left[(V_2 \cap V_1) \cup (V_2 \cap V_1^C)\right] \\
&= \mathbb{P}(V_2 \cap V_1) + \mathbb{P}(V_2 \cap V_1^C) \\
&\text{Notando che } (V_2 \cap V_1) \text{ è un evento a probabilità nulla} \\
&= \mathbb{P}(V_2 \cap V_1^C) = \mathbb{P}(V_2 | V_1^C) \cdot \mathbb{P}(V_1^C) \qquad \text{(applicando 3.8.2)} \\
&= \frac{1}{4} \cdot \frac{4}{5} = \frac{1}{5} \\
\mathbb{P}(V_3) &= \mathbb{P}(V_3 \cap \Omega) = \mathbb{P}\left[V_3 \cap \big((V_1 \cap V_2) \cup (V_1 \cap V_2^C) \cup (V_1^C \cap V_2) \cup (V_1^C \cap V_2^C)\big)\right] \\
&= \mathbb{P}(V_3 \cap (V_1^C \cap V_2^C)) \\
&= \mathbb{P}(V_1^C) \cdot \mathbb{P}(V_2^C | V_1^C) \cdot \mathbb{P}(V_3 | V_1^C \cap V_2^C) \\
&= \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{5} \\
\mathbb{P}(V_4) &= \mathbb{P}(V_4 \cap \Omega) = \\
&= \mathbb{P}(V_4 \cap ((V_1 \cap V_2) \cup (V_2 \cap V_3) \cup (V_1 \cap V_3) \cup (V_1^C \cap V_2) \cup (V_2^C \cap V_3) \cup (V_1^C \cap V_3) \cup \\
&\quad (V_1 \cap V_2^C) \cup (V_2 \cap V_3^C) \cup (V_1 \cap V_3^C) \cup (V_1^C \cap V_2) \cup (V_2^C \cap V_3) \cup (V_1^C \cap V_3) \cup \\
&\quad (V_1^C \cap V_2^C) \cup (V_2^C \cap V_3^C) \cup (V_1^C \cap V_3^C))) \\
&= \mathbb{P}(V_4 \cap V_1^C \cap V_2^C \cap V_3^C) \\
&= \mathbb{P}(V_1^C) \cdot \mathbb{P}(V_2^C | V_1^C) \cdot \mathbb{P}(V_3^C | V_1^C \cap V_2^C) \cdot \mathbb{P}(V_4 | V_1^C \cap V_2^C \cap V_3^C) \\
&= \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{5}
\end{aligned}
$$

Per il quinto giocatore possiamo ragionare analogamente ai primi quattro oppure giungere allo stesso risultato considerando che $\mathbb{P}(V_5^C) = \mathbb{P}(V_1 \cup V_2 \cup V_3 \cup V_4) = \frac{4}{5}$, quindi si avrà $\mathbb{P}(V_5) = 1 - \frac{4}{5} = \frac{1}{5}$

3.9 Formula di fattorizzazione (o delle alternative)

Definizione 3.9.1 (Sistema completo di alternative per Ω)**.** *La famiglia finita di insiemi* $F_1, F_2, \cdots, F_n$ *con* $F_i \in \mathcal{F}\ \forall i \in [1, n]$ *è un sistema completo di alternative per* Ω *se e solo se:*

i) $\mathbb{P}(F_i) > 0\ \forall i \in [1, n]$

ii) $F_i \cap F_j = \varnothing\ \forall i \neq j$

iii) $\dot{\bigcup}_{i=1}^{n} F_i = \Omega$

Teorema 3.9.1 (Formula di fattorizzazione (o delle alternative))**.** *Si consideri la terna* $(\Omega, \mathcal{F}, \mathbb{P})$ *Sia* $A \in \mathcal{F}$ *un evento qualsiasi e sia* $\{F_i\}_{i=1\cdots n}$ *un sistema completo di alternative per* Ω*. Si ha che*

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|F_i) \cdot \mathbb{P}(F_i)$$

Dimostrazione. Dal momento che $A = A \cap \Omega$, possiamo scrivere

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) \\
&= \mathbb{P}\left(A \cap \left(\dot{\bigcup_{i=1}^{n}} F_i\right)\right) && \text{perchè } \{F\}_{i=1\cdot n} \text{ è un sist. completo di alternative} \\
&= \mathbb{P}\left(\dot{\bigcup_{i=1}^{n}}(A \cap F_i)\right) && \text{applico distributività} \\
&= \sum_{i=1}^{n} \mathbb{P}(A \cap F_1) && \text{per additività finita} \\
&= \sum_{i=1}^{n} \mathbb{P}(A|F_i) \cdot \mathbb{P}(F_i) && \text{usando la formula delle prob. totali 3.8.2}
\end{aligned}$$

□

Esercizio 3.9.1 (Estrazione da urne diverse (formula di fattorizzazione))**.** *Si considerino le tre urne di seguito raffigurate. Una pallina sarà estratta casualmente da una delle tre urne. Le singole urne sono equipreferibili per l'estrattore. Qual è la probabilità che la pallina estratta sia bianca?*

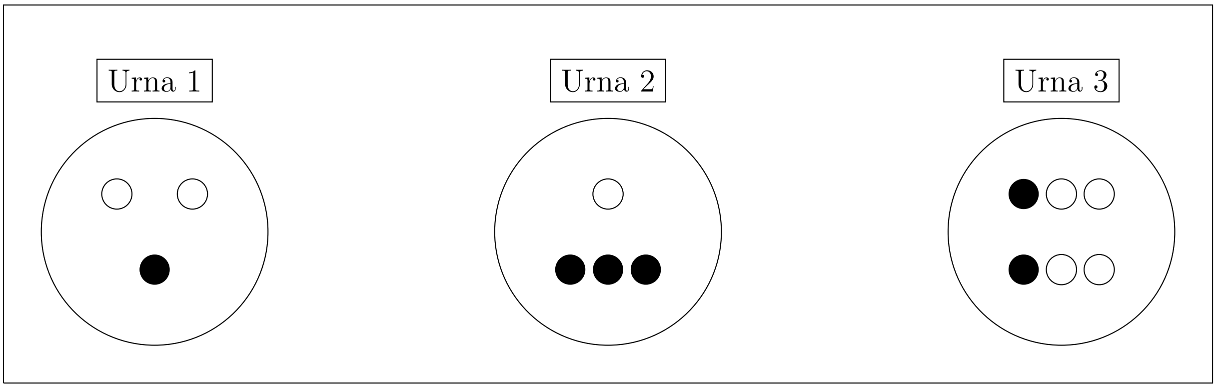


Figura 3.1: Rappresentazione delle urne

Iniziamo identificando gli eventi:

1. $A =$"La pallina estratta è bianca"
2. $F_i =$ "La pallina è estratta dalla i-esima urna (i=1,2,3)"

Verifichiamo che F_1, F_2, F_3 è un set completo di alternative per Ω:

i) $\forall i \ \mathbb{P}(F_i) = \frac{1}{3} > 0$ dal momento che le urne sono equipreferibili (prob. classica).

ii) $F_i \cap F_j = \varnothing \ \forall i \neq j$ perchè non posso estrarre una pallina contemporaneamente da due urne diverse.

iii) $\dot\bigcup_{i=1}^{3} F_i = \Omega$ dal momento che devo estrarre necessariamente la pallina da una delle tre urne.

Dal momento che F_1, F_2, F_3 è un set completo di alternative possiamo applicare la formula di fattorizzazione 3.9.1:

$$\begin{aligned}
\mathbb{P}(A) &= \sum_{i=1}^{3} \mathbb{P}(A|F_i) \cdot \mathbb{P}(F_i) \\
&= \mathbb{P}(A|F_1) \cdot \mathbb{P}(F_1) + \mathbb{P}(A|F_2) \cdot \mathbb{P}(F_2) + \mathbb{P}(A|F_3) \cdot \mathbb{P}(F_3) \\
&= \frac{2}{3} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{3} + \frac{4}{6} \cdot \frac{1}{3} = \frac{4}{9} + \frac{1}{12} = \frac{16+3}{36} = \frac{19}{36}
\end{aligned}$$

Esercizio 3.9.2 (Indagine (formula di fattorizzazione))**.** *Ad un certo punto di un'indagine, l'investigatore è convinto al 60% della colpevolezza del sospettato principale. Dalle indagini emerge che il colpevole ha una certa caratteristica (es. calvo, capelli castani, pelle nera). Anche il sospettato principale ha quella caratteristica. Detta caratteristica ha un'incidenza del 20% sulla popolazione. Qual è la probabilità che il sospettato sia colpevole secondo l'investigatore, dato che ha la caratteristica in questione?*

Come di consueto assegnamo dei nomi agli eventi:

- $C =$ "Il sospetto è colpevole"
- $A =$ "Il sospetto possiede la caratteristica del colpevole"

Abbiamo, dalla definizione di probabilità condizionata, che

$$\mathbb{P}(C|A) = \frac{\mathbb{P}(C \cap A)}{\mathbb{P}(A)} \qquad (1)$$

Calcoliamo $\mathbb{P}(C \cap A)$:

$$\begin{aligned}\mathbb{P}(C \cap A) &= \mathbb{P}(A|C)\mathbb{P}(C) \\ &= (\text{Prob. che il sospetto abbia caratt. dato che è colpevole}) \cdot (\text{Prob. che sia colpevole}) \\ &= 1 \cdot 0.6 = 0.6\end{aligned}$$

Quindi troviamo $\mathbb{P}(A)$, con C e C^C alternative

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A|C)\mathbb{P}(C) + \mathbb{P}(A|C^C)\mathbb{P}(C^C) \\ &= 1 \cdot 0.6 + 0.2 \cdot 0.4 = 0.68\end{aligned}$$

Quindi possiamo sostituire nella formula (1) ed ottenere

$$\mathbb{P}(C|A) = \frac{\mathbb{P}(C \cap A)}{\mathbb{P}(A)} = \frac{0.6}{0.68} \approx 0.882$$

3.10 Teorema di Bayes

Teorema 3.10.1 (Teorema di Bayes (teorema della probabilità delle cause))**.** *Il teorema di Bayes (noto anche come teorema della probabilità delle cause) afferma che, dati un evento $E \in \mathcal{F}$ con $\mathbb{P}(E) > 0$ e un sistema completo di alternative $\{F_i\}_{i=1\cdots n}$, la probabilità che si verifichi l'i-esima alternativa dato che si è verificato E è uguale a:*

$$\mathbb{P}(F_i|E) = \frac{\mathbb{P}(E|F_i) \cdot \mathbb{P}(F_i)}{\sum_{j=1}^{n} \mathbb{P}(E|F_j)\mathbb{P}(F_j)}$$

Dimostrazione.

$$\begin{aligned}
\mathbb{P}(F_i|E) &= \frac{\mathbb{P}(F_i \cap E)}{\mathbb{P}(E)} \\
&= \frac{\mathbb{P}(F_i \cap E)}{\mathbb{P}(E)} \cdot \frac{\mathbb{P}(F_i)}{\mathbb{P}(F_i)} \\
&= \frac{\mathbb{P}(F_i \cap E)}{\mathbb{P}(F_i)} \cdot \frac{\mathbb{P}(F_i)}{\mathbb{P}(E)} \\
&= \mathbb{P}(E|F_i) \cdot \frac{\mathbb{P}(F_i)}{\mathbb{P}(E)} \\
&= \frac{\mathbb{P}(E|F_i) \cdot \mathbb{P}(F_i)}{\mathbb{P}(E)} \\
&= \frac{\mathbb{P}(E|F_i) \cdot \mathbb{P}(F_i)}{\sum_{j=1}^{n} \mathbb{P}(E|F_j) \cdot \mathbb{P}(F_j)} && \text{(ho applicato 3.9.1 a } \mathbb{P}(E))
\end{aligned}$$

□

Focalizzandoci sugli elementi in gioco nella formula:

- E è l'effetto.
- F_i sono le cause che possono generare l'effetto.
- $\mathbb{P}(F_i)$ sono **probabilità a priori**.
- $\mathbb{P}(F_i|E)$ sono **probabilità a posteriori** (si è già verificato l'effetto).
- $\mathbb{P}(E|F_i)$ sono **verosimiglianze** (quanto è verosimile che, avutosi l'effetto, questo sia dovuto a F_i).

Esercizio 3.10.1 (Il dilemma di Monty Hall (teorema di Bayes)). *Siete concorrenti al gioco finale dello show "Let's make a deal" condotto da Monty Hall. Vi trovate davanti 3 porte. Dietro una delle porte è celata una macchina, mentre dietro le altre due si trovano due capre. Scegliete una porta. Il conduttore quindi apre una porta tra le rimanenti che contiene una capra. Vi viene offerta la possibilità di cambiare porta. Vi conviene accettare l'offerta?*

Assegnamo dei nomi agli eventi:

- M_i = "La porta i-esima cela una macchina"
- C_i = "Il conduttore apre la porta i-esima che cela una capra"

Consideriamo senza ledere la generalità dell'analisi (dal momento che posso numerare le porte a partire da quella scelta e che gli altri casi sono simmetrici), che il concorrente scelga inizialmente la porta 1 e che il presentatore apra sempre la porta 3 contenente la capra. Calcoliamo la probabilità di vincere scegliendo di cambiare:

$$\begin{aligned}
\mathbb{P}(M_2|C_3) &= \frac{\mathbb{P}(C_3|M_2)\cdot\mathbb{P}(M_2)}{\sum_{j=1}^{3}\mathbb{P}(C_3|M_j)\mathbb{P}(M_j)} \\
&= \frac{\mathbb{P}(C_3|M_2)\cdot\mathbb{P}(M_2)}{\mathbb{P}(C_3|M_1)\mathbb{P}(M_1)+\mathbb{P}(C_3|M_2)\mathbb{P}(M_2)+\mathbb{P}(C_3|M_3)\mathbb{P}(M_3)} \\
&= \frac{1\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3}+1\cdot\frac{1}{3}+0\cdot\frac{1}{3}} = \frac{2}{3}
\end{aligned}$$

Dunque la strategia più conveniente è quella di cambiare porta accettando l'offerta di Monty.

3.11 Eventi indipendenti

Definizione 3.11.1 (Eventi indipendenti)**.** *Si consideri un esperimento individuato dalla terna* $(\Omega, \mathcal{F}, \mathbb{P})$*. Gli eventi* $A, B \in \mathcal{F}$ *sono indipendenti se e solo se*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Teorema 3.11.1 (Implicazioni dell'indipendenza di due eventi)**.** *Se due eventi* $A, B \in \mathcal{F}$ *sono indipendenti, allora sono indipendenti anche*

i) A^C, B

ii) A, B^C

iii) A^C, B^C

Dimostrazione. Dimostriamo il punto $i)$ mostrando che $\mathbb{P}(A^C \cap B) = \mathbb{P}(A^C) \cdot \mathbb{P}(B)$.

$$\begin{aligned}
\mathbb{P}(B) &= \mathbb{P}(B \cap \Omega) = \mathbb{P}\left(B \cap (A \cupdot A^C)\right) \\
&= \mathbb{P}\left((B \cap A) \cupdot (B \cap A^C)\right) \\
&= \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^C)
\end{aligned}$$

Ricavando $\mathbb{P}(B \cap A^C)$ da quanto ottenuto sopra

$$\begin{aligned}
\mathbb{P}(B \cap A^C) &= \mathbb{P}(B) - \mathbb{P}(B \cap A) \\
&= \mathbb{P}(B) - \mathbb{P}(B) \cdot \mathbb{P}(A) && \text{(usando l'ipotesi che siano } A, B \text{ indipendenti)} \\
&= \mathbb{P}(B)\,(1 - \mathbb{P}(A)) && \text{(raccogliendo } \mathbb{P}(B)\text{)} \\
&= \mathbb{P}(B) \cdot \mathbb{P}(A^C)
\end{aligned}$$

Il punto $ii)$ segue direttamente dal punto $i)$ precedentemente mostrato scambiando i nomi degli insiemi iniziali. Il punto $iii)$ segue applicando due volte i punti $i)$ e $ii)$:

$$A, B \text{ indip.} \overset{i)}{\Longrightarrow} A^C, B \text{ indip.} \overset{ii)}{\Longrightarrow} A^C, B^C \text{ indip.}$$

$\square$

Teorema 3.11.2 (Relazione tra indipendenza e probabilità condizionata)**.** *Si considerino un esperimento caratterizzato da* $(\Omega, \mathcal{F}, \mathbb{P})$ *e gli eventi* $A, B \in \mathcal{F}$ *con* $\mathbb{P}(A) > 0,\ \mathbb{P}(B) > 0$*. Le seguenti sono equivalenti:*

a) $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

b) $\mathbb{P}(A|B) = \mathbb{P}(A)$

c) $\mathbb{P}(B|A) = \mathbb{P}(B)$

Dimostrazione. Daremo di questo teorema una dimostrazione ciclica, ovvero mostreremo che $a \Rightarrow b$, $b \Rightarrow c$, $c \Rightarrow a$.

$a \Rightarrow b$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \overset{a)}{=} \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

$b \Rightarrow c$

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B) \cdot \mathbb{P}(B)}{\mathbb{P}(A)} \overset{b)}{=} \frac{\mathbb{P}(A) \cdot \mathbb{P}(B)}{\mathbb{P}A} = \mathbb{P}(B)$$

$c \Rightarrow a$

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A) \overset{c)}{=} \mathbb{P}(B)\mathbb{P}(A)$$

$\square$

Definizione 3.11.2 (Eventi collettivamente indipendenti (definizione generalizzata))**.** *Una quantità finita* $[F_1, F_2, \cdots, F_n](n > 2)$, $F_i \in \mathcal{F}$ *di eventi si dice collettivamente indipendente se, comunque presi* $i_1, i_2, \cdots, i_k$ *con* $k \leq n$ *si ha*

$$\mathbb{P}(F_{i_1} \cap F_{i_2} \cap \cdots \cap F_{i_k} = \prod_{r=1}^{k} F_{i_r})$$

Teorema 3.11.3 (Proprietà dell'unione du eventi indipendenti)**.** *Siano* $A, B \in \mathcal{F}$ *eventi indipendenti. Si ha che*

$$\mathbb{P}(A \cup B) = 1 - \mathbb{P}(A^C) \cdot \mathbb{P}(B^C)$$

Dimostrazione. Notiamo innanzitutto che per il teorema 3.11.1 si ha che
A, B indipendenti $\Rightarrow A^C, B^C$ indipendenti

$$\mathbb{P}(A \cup B) \overset{\text{de Morgan}}{=} \mathbb{P}((A^C \cap B^C)^C) = 1 - \mathbb{P}(A^C \cap B^C) = 1 - \mathbb{P}(A^C) \cdot \mathbb{P}(B^C)$$

$\square$

Esercizio 3.11.1 (Lancio di dadi (eventi indipendenti))**.** *Si considerino i due seguenti esperimenti:*

1. *Lancio di un dado equo ripetuto 4 volte.*

2. *Lancio di una coppia di dadi equi ripetuto 24 volte.*

Si calcoli per il primo esperimento la probabilità che si ottenga 6 almeno in un lancio e per il secondo la probabilità che si ottenga almeno una coppia di 6 in ventiquattro lanci.

Cominciamo dal primo esperimento. Sia

- A = "ottengo almeno un 6"
- C_i = "al lancio i-esimo non ottengo 6"

Si ha che

$$\begin{aligned}
\mathbb{P}(A) &= 1 - \mathbb{P}(A^C) = 1 - \mathbb{P}(C_1 \cap C_2 \cap C_3 \cap C_4) \\
&= \prod_{i=1}^{4} \mathbb{P}(C_i) = 1 - \left(\frac{5}{6}\right)^4 \approx 0.52
\end{aligned}$$

Per quanto riguarda il secondo, siano

- B = "ottengo almeno una coppia di 6"
- C_i = "al lancio i-esimo non ottengo una coppia di 6"

Si ha che

$$\begin{aligned}
\mathbb{P}(B) &= 1 - \mathbb{P}(B^C) = 1 - \mathbb{P}(C_1 \cap C_2 \cap \cdots \cap C_{24}) \\
&= 1 - \prod_{i=1}^{24} \mathbb{P}(C_i) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0.49
\end{aligned}$$

3.12 Variabili aleatorie

Definizione 3.12.1 (Variabile aleatoria)**.** *Dato uno spazio di probabilità* $(\Omega, \mathcal{F}, \mathbb{P})$*, una variabile aleatoria (chiamata anche variabile casuale o variabile stocastica)* X *è una funzione*

$$X : \Omega \to \mathbb{R}$$

tale che $\forall x \in \mathbb{R},\ \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$*.*

Definizione 3.12.2 (Spettro di una variabile aleatoria)**.** *Lo spettro (o supporto) di una variabile aleatoria* X *è l'insieme* S_X *di valori (o stati) che questa può assumere.*

Esercizio 3.12.1 (Somma delle facce nel lancio di due dadi (var. aleatorie))**.** *Si consideri il lancio di due dadi equi. Denotiamo con* X *la variabile aleatoria definita come la somma dei pallini neri sulle facce dei due dadi. Qual è la probabilità che sia* $X = 3$*?*

Gli $\omega \in \Omega$ tali che $X(\omega) = 3$ sono $(1, 2), (2, 1)$. Quindi

$$\mathbb{P}(X = 3) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = 3\}) = \mathbb{P}(\{(1, 2), (2, 1)\}) = \frac{2}{36}$$

Si noti che, avendo assunto i dadi equi, è stato possibile applicare la definizione classica di probabilità 3.5.2 $\left(\frac{\#casi\ favorevoli}{\#casi\ possibili}\right)$.

Osservazione 3.12.1. *Si consideri quanto segue*

$$\mathbb{P}(\Omega) = 1 = \mathbb{P}\left(\dot{\bigcup_{i \in S_X}} \{X = i\}\right) = \sum_{i \in S_X} \mathbb{P}(X = i)$$

Proseguiamo calcolando le probabililtà per tutti i valori che X può assumere, ovvero per i valori in $S_X = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

$$
\begin{aligned}
\mathbb{P}(X=2) &= \mathbb{P}\{(1,1)\} = \frac{1}{36} \\
\mathbb{P}(X=3) &= \mathbb{P}\{(1,2),(2,1)\} = \frac{2}{36} \\
\mathbb{P}(X=4) &= \mathbb{P}\{(1,3),(3,1),(2,2)\} = \frac{3}{36} \\
\mathbb{P}(X=5) &= \mathbb{P}\{(1,4),(4,1),(2,3),(3,2)\} = \frac{4}{36} \\
\mathbb{P}(X=6) &= \mathbb{P}\{(1,5),(5,1),(2,4),(4,2),(3,3)\} = \frac{5}{36} \\
\mathbb{P}(X=7) &= \mathbb{P}\{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)\} = \frac{6}{36} \\
\mathbb{P}(X=8) &= \mathbb{P}\{(2,6),(6,2),(3,5),(5,3),(4,4)\} = \frac{5}{36} \\
\mathbb{P}(X=9) &= \mathbb{P}\{(3,6),(6,3),(4,5),(5,4)\} = \frac{4}{36} \\
\mathbb{P}(X=10) &= \mathbb{P}\{(4,6),(6,4),(5,5)\} = \frac{3}{36} \\
\mathbb{P}(X=11) &= \mathbb{P}\{(5,6),(6,5)\} = \frac{2}{36} \\
\mathbb{P}(X=12) &= \mathbb{P}\{(6,6)\} = \frac{1}{36}
\end{aligned}
$$

Notiamo che $\sum_{i=2}^{12} \mathbb{P}(X=i) = 1 = \mathbb{P}(\Omega)$, come da 3.12.1.

Esercizio 3.12.2 (Acquisto dispositivi elettronici (var. aleatorie))**.** *Un cliente acquista due dispositivi elettronici e ognuno di questi può essere accettabile o difettoso. La probabililtà che siano entrambi accettabili è di* 0.49*, quella che siano entrambi difettosi vale* 0.09 *mentre la probabilità che solo uno dei due sia difettoso è* 0.21*. Consideriamo come variabile aleatoria X la funzione che conta il numero di dispositivi accettabili acquistati e calcoliamo la probabilità per ciascun valore che X assume.*

Innanzitutto identifichiamo lo spazio degli esiti $\Omega = \{(a,a),(a,d),(d,a),(d,d)\}$ (ho indicato con d il prodotto difettoso e con a quello accettabile). Dalla traccia sappiamo che $\mathbb{P}(\{(a,a)\}) = 0.49$, $\mathbb{P}(\{(a,d)\}) = \mathbb{P}(\{(d,a)\}) = 0.21$ e che $\mathbb{P}(\{(d,d)\}) = 0.09$. Lo spettro di X è $S_X = \{0,1,2\}$. Calcoliamo le probabililtà:

$$
\begin{aligned}
&\mathbb{P}(X=0) = \mathbb{P}(\omega \in \Omega : X(\omega) = 0) = \mathbb{P}(\{(d,d)\}) = 0.09 \\
&\mathbb{P}(X=1) = \mathbb{P}(\omega \in \Omega : X(\omega) = 1) = \mathbb{P}(\{(a,d)\} \cup \{(d,a)\}) = \mathbb{P}(\{(a,d)\}) + \mathbb{P}(\{(d,a)\}) = 0.42 \\
&\mathbb{P}(X=2) = \mathbb{P}(\omega \in \Omega : X(\omega) = 2) = \mathbb{P}(\{(a,a)\}) = 0.49
\end{aligned}
$$

Notiamo ancora che$\sum_{i=0}^{2} \mathbb{P}(X=i) = 1 = \mathbb{P}(\Omega)$, come da 3.12.1.

Definizione 3.12.3 (Variabile aleatoria indicatrice di un evento)**.** *Si consideri uno spazio di probabilità* $(\Omega, \mathcal{F}, \mathbb{P})$ *e un evento* $A \in \mathcal{F}$*. La variabile aleatoria indicatrice di* A *è definita come segue:*

$$I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

3.13 Funzione di distribuzione di una variabile aleatoria

Definizione 3.13.1 (Funzione di distribuzione)**.** *Si consideri lo spazio di probabilità* $(\Omega, \mathcal{F}, \mathbb{P})$*. Ad ogni* X *variabile aleatoria è associata una funzione di distribuzione* F_X *definita in questo modo:*

$$F_X : \mathbb{R} \to [0,1] \qquad F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})$$

Teorema 3.13.1 (Proprietà della funzione di distribuzione)**.** *Una funzione di distribuzione* F_X *gode delle seguenti proprietà:*

i) F_X *è una funzione monotòna non decrescente, ovvero*

$$\forall x_1, x_2 \in \mathbb{R},\ x_1 < x_2 \Longrightarrow F_X(x_1) \leq F_X(x_2)$$

ii) F_X *è una funzione continua a destra, ovvero*

$$\lim_{\varepsilon \to 0^+} F_X(x + \varepsilon) = F_X(x)$$

iii) F_X *ha un certo comportamento agli estremi, ovvero*

$$\lim_{x \to +\infty} F_X(x) = 1 \qquad \lim_{x \to -\infty} F_X(x) = 0$$

Dimostrazione. Incominciamo dimostrando la proprietà i (F_X **non decrescente**). Siano $x_1, x_2 \in \mathbb{R}$ con $x_1 < x_2$. Si ha che:

$$\begin{aligned} F_X(x_2) &= \mathbb{P}(X \leq x_2) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x_2\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x_1\} \cup \{\omega \in \Omega : x_1 < X(\omega) \leq x_2\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x_1\}) + \mathbb{P}(\{\omega \in \Omega : x_1 < X(\omega) \leq x_2\}) \\ &= F_X(x_1) + \mathbb{P}(\{\omega \in \Omega : x_1 < X(\omega) \leq x_2\}) \\ &\geq F_X(x_1) + 0 \quad (\text{poichè } 0 \leq \mathbb{P}(\{\omega \in \Omega : x_1 < X(\omega) \leq x_2\}) \leq 1) \\ &\geq F_X(x_1) \end{aligned}$$

concludendo quindi che F_X è monotòna non decrescente.
Dimostriamo la proprietà *ii* (F_X **continua a destra**). Vogliamo dimostrare che

$$\forall x, \lim_{\varepsilon \to 0^+} F_X(x+\varepsilon) = F_X(x)$$

Dal momento che F_X è monotona non decrescente, è possibile dimostrare la proprietà dimostrando equivalentemente che

$$\forall x, \lim_{n \to +\infty} F_X\left(x + \frac{1}{n}\right) = F_X(x)$$

Infatti, la monotonìa di F_X dimostrata precedentemente garantisce che, nell'avvicinarsi per "piccoli salti" ad x da destra come nella seconda espressione (anzichè per continuità come nella prima) non si rischia di "perdere" nulla, dal momento che il comportamento nell'intervallo saltato sarà comunue monotòno decrescente. Proseguendo si ha:

$$\begin{aligned}
F_X\left(x + \frac{1}{n}\right) &= \mathbb{P}\left(X \le x + \frac{1}{n}\right) = \mathbb{P}\left(\left\{\omega \in \Omega : X(\omega) \le x + \frac{1}{n}\right\}\right) \\
&= \mathbb{P}\left(\left\{\omega \in \Omega : X(\omega) \le x\right\} \uplus \left\{\omega \in \Omega : x < X(\omega) \le x + \frac{1}{n}\right\}\right) \\
&= \mathbb{P}\left(\left\{\omega \in \Omega : X(\omega) \le x\right\}\right) + \mathbb{P}\left(\left\{\omega \in \Omega : x < X(\omega) \le x + \frac{1}{n}\right\}\right) \\
&= F_X(x) + \mathbb{P}\left(x < X \le x + \frac{1}{n}\right)
\end{aligned}$$

Sia la successione di insiemi $A_n = \left\{\omega \in \Omega : x < X(\omega) \le x + \frac{1}{n}\right\}$. Sicuramente si avrà $\forall n$ che $A_n \supseteq A_{n+1}$, dal momento che l'intervallo tra x ed $x + \frac{1}{n}$ si riduce al crescere di n. Dunque A_n è monotòna descrescente. Quando $n \to +\infty$, $A_n = \varnothing$, dal momento che non esistono ω tali che $x < X(\omega) \le x + 0$. Poichè quindi A_n è monotòna

$$\lim_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\lim_{n \to \infty} A_n\right) = \mathbb{P}(\varnothing) = 0$$

Usando quindi quanto ricavato finora:

$$\begin{aligned}
\lim_{n \to \infty} F_X\left(x + \frac{1}{n}\right) &= \lim_{n \to \infty} \left(F_X(x) + \mathbb{P}(A_n)\right) \\
&= \lim_{n \to \infty} F_X(x) + \lim_{n \to \infty} \mathbb{P}(A_n) \\
&= F_X(x) + 0 = F_X(x)
\end{aligned}$$

provando la continuità a destra.

Dimostriamo la proprietà *iii* (**comportamento agli estremi**). In particolare vogliamo mostrare che:

$$\lim_{x\to+\infty} F_X(x) = 1 \qquad\qquad \lim_{x\to-\infty} F_X(x) = 0$$

Dimostriamo il comportamento all'estremo superiore. Sia $B_n = \{\omega \in \Omega : X(\omega) \leq n\}$. Si ha $\forall n$ che $B_{n+1} \supseteq B_n$, quindi B_n successione crescente. Per $n \to +\infty$, inoltre, $B_n = \Omega$, dal momento che ogni ω sarà tale che $X(\omega) \leq +\infty$. Analogamente, per $n \to -\infty$, $B_n = \varnothing$. Poichè B_n monotòna, è possibile scrivere

$$\begin{aligned}
\lim_{n\to+\infty} F_X(n) &= \lim_{n\to+\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\lim_{n\to+\infty} B_n\right) = \mathbb{P}(\Omega) = 1 \\
\lim_{n\to-\infty} F_X(n) &= \lim_{n\to-\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\lim_{n\to-\infty} B_n\right) = \mathbb{P}(\varnothing) = 0
\end{aligned}$$

Dalla dimostrazione delle tre proprietà segue la verità del teorema. □

3.13.1 Esprimere la probabililtà di intervalli usando la funzione di distribuzione

Definizione 3.13.2 (Funzione di ripartizione)**.** *Siano* $(\Omega, \mathcal{F}, \mathbb{P})$ *uno spazio di probabilità,* X *una variabile aleatoria e* F_X *la funzione di distribuzione di* X*.*

$$F_X(x^-) = \lim_{\varepsilon\to0^+} F_X(x-\varepsilon) = \lim_{\varepsilon\to0^+} \mathbb{P}(X \leq x-\varepsilon) = \mathbb{P}(X < x)$$

N.B. Negli esempi che seguono si considerino $x, x_1, x_2 \in \mathbb{R}$, con $x_1 < x_2$.

- $\mathbb{P}(x_1 < X \leq x_2) = \mathbb{P}\left(x \in]x_1, x_2]\right) = \mathbb{P}(\{\omega \in \Omega : x_1 < X(\omega) \leq x_2\})$

$$\begin{aligned}
\mathbb{P}(X \leq x_2) &= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \leq x_2\}\right) \\
&= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \leq x_1\} \,\dot\cup\, \{\omega \in \Omega : x_1 < X(\omega) \leq x_2\}\right) \\
&= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \leq x_1\}\right) + \mathbb{P}\left(\{\omega \in \Omega : x_1 < X(\omega) \leq x_2\}\right) \\
&= \mathbb{P}\left(X \leq x_1\right) + \mathbb{P}\left(x_1 < X \leq x_2\right) \\
&\Updownarrow \\
\mathbb{P}\left(x_1 < X \leq x_2\right) &= \mathbb{P}(X \leq x_2) - \mathbb{P}(X \leq x_1) \\
&= F_X(x_2) - F_X(x_1)
\end{aligned}$$

- $\mathbb{P}(x_1 \le X \le x_2) = \mathbb{P}\left(x \in [x_1, x_2]\right) = \mathbb{P}(\{\omega \in \Omega : x_1 \le X(\omega) \le x_2\})$

$$\begin{aligned}
\mathbb{P}(X \le x_2) &= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \le x_2\}\right) \\
&= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) < x_1\} \cup \{\omega \in \Omega : x_1 \le X(\omega) \le x_2\}\right) \\
&= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) < x_1\}\right) + \mathbb{P}\left(\{\omega \in \Omega : x_1 \le X(\omega) \le x_2\}\right) \\
&= \mathbb{P}\left(X < x_1\right) + \mathbb{P}\left(x_1 \le X \le x_2\right) \\
&\Updownarrow \\
\mathbb{P}\left(x_1 \le X \le x_2\right) &= \mathbb{P}(X \le x_2) - \mathbb{P}(X < x_1) \\
&= F_X(x_2) - F_X(x_1^-)
\end{aligned}$$

- $\mathbb{P}(x_1 < X < x_2) = \mathbb{P}\left(x \in]x_1, x_2[\right) = \mathbb{P}(\{\omega \in \Omega : x_1 < X(\omega) < x_2\})$

$$\begin{aligned}
\mathbb{P}(X < x_2) &= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) < x_2\}\right) \\
&= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \le x_1\} \cup \{\omega \in \Omega : x_1 < X(\omega) < x_2\}\right) \\
&= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \le x_1\}\right) + \mathbb{P}\left(\{\omega \in \Omega : x_1 < X(\omega) < x_2\}\right) \\
&= \mathbb{P}\left(X \le x_1\right) + \mathbb{P}\left(x_1 < X < x_2\right) \\
&\Updownarrow \\
\mathbb{P}\left(x_1 < X < x_2\right) &= \mathbb{P}(X < x_2) - \mathbb{P}(X \le x_1) \\
&= F_X(x_2^-) - F_X(x_1)
\end{aligned}$$

- $\mathbb{P}(x_1 \le X < x_2) = \mathbb{P}\left(x \in [x_1, x_2[\right) = \mathbb{P}(\{\omega \in \Omega : x_1 \le X(\omega) < x_2\})$

$$\begin{aligned}
\mathbb{P}(X < x_2) &= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) < x_2\}\right) \\
&= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) < x_1\} \cup \{\omega \in \Omega : x_1 \le X(\omega) < x_2\}\right) \\
&= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) < x_1\}\right) + \mathbb{P}\left(\{\omega \in \Omega : x_1 \le X(\omega) < x_2\}\right) \\
&= \mathbb{P}\left(X < x_1\right) + \mathbb{P}\left(x_1 \le X < x_2\right) \\
&\Updownarrow \\
\mathbb{P}\left(x_1 \le X < x_2\right) &= \mathbb{P}(X < x_2) - \mathbb{P}(X < x_1) \\
&= F_X(x_2^-) - F_X(x_1^-)
\end{aligned}$$

- $\mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$

$$\begin{aligned}
\mathbb{P}(X \leq x) &= \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\}) \\
&= \mathbb{P}(\{\omega \in \Omega : X(\omega) < x\} \cup \{\omega \in \Omega : X(\omega) = x\}) \\
&= \mathbb{P}(\{\omega \in \Omega : X(\omega) < x\}) + \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) \\
&= \mathbb{P}(X < x) + \mathbb{P}(X = x) \\
&\Updownarrow \\
\mathbb{P}(X = x) &= \mathbb{P}(X \leq x) - \mathbb{P}(X < x) \\
&= F_X(x) - F_X(x^-)
\end{aligned}$$

Osservazione 3.13.1. *Se F_X è continua anche a sinistra (cioè è continua in senso generale, visto che è continua a destra per definizione), allora si ha che*

$$\mathbb{P}(X = x) = F_X(x) - F_X(x^-) = F_X(x) - F_X(x) = 0$$

Esercizio 3.13.1 (Calcolo prob. di intervallo (var. aleatorie))**.** *Si considerino* $(\Omega, \mathcal{F}, \mathbb{P})$ *e X variabile aleatoria che ammette la funzione di distribuzione*

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-x^2}, & x > 0 \end{cases}$$

Verificare che F_X è una funzione di distribuzione, quindi calcolare la probabilità che X assuma valori maggiori di 1.

Innanzitutto osserviamo che F_X è definita in tutto $\mathbb{R}$. Verifichiamo quindi che sia $F_X : \mathbb{R} \to [0, 1]$.
Se $x \leq 0$, allora $F_X = 0$, quindi nell'intervallo voluto. Analizziamo il comportamento di F_X se $x > 0$. Consideriamo $1 - e^{-x^2} = 1 - \frac{1}{e^{x^2}}$. Si ha che e^{x^2} è sempre maggiore di 1 (ricordando che siamo nel caso in cui $x > 0$), quindi $0 \leq \frac{1}{e^{x^2}} < 1$. Di conseguenza $0 < 1 - \frac{1}{e^{x^2}} \leq 1$.
Verifichiamo che F_X rispetti quindi le tre proprietà della definizione 3.13.1:

i) F_X monotòna non decrescente. Siano $x_1, x_2 \in \mathbb{R}$, con $x_1 < x_2$.
Se $x_1 \leq 0$, per quanto detto anche precedentemente (dimostrazione del fatto che F_X ha valori in $[0, 1]$), la funzione non può decrescere.
Se $0 < x_1 < x_2$ si ha:

$$\begin{aligned}
x_1 < x_2 \Leftrightarrow x_1^2 < x_2^2 \Leftrightarrow -x_1^2 > -x_2^2 \Leftrightarrow e^{-x_1^2} > e^{-x_2^2} \Leftrightarrow -e^{-x_1^2} < -e^{-x_2^2} \\
\Leftrightarrow 1 - e^{-x_1^2} < 1 - e^{-x_2^2} \Leftrightarrow F_X(x_1) < F_X(x_2)
\end{aligned}$$

ii) continua a destra.

$$\lim_{\varepsilon\to 0^+} F_X(x+\varepsilon) = \lim_{\varepsilon\to 0^+} 1 - e^{-(x-\varepsilon)^2} = 1 - e^{-x^2} = F_X(x)$$

iii) comportamento agli estremi.

$$\lim_{x\to+\infty} F_X(x) = \lim_{x\to+\infty} 1 - e^{-x^2} = \lim_{x\to+\infty} 1 - \frac{1}{e^{x^2}} = 1$$
$$\lim_{x\to-\infty} F_X(x) = 0$$

Abbiamo quindi verificato che F_X è una funzione di distribuzione. Calcoliamo la probabililtà che sia $X > 1$

$$\mathbb{P}(X > 1) = \mathbb{P}((X \le 1)^C) = 1 - \mathbb{P}(X \le 1) = 1 - F_X(1) = 1 - \left(1 - \frac{1}{e^{1^2}}\right) = \frac{1}{e}$$

3.14 Variabili aleatorie discrete

Definizione 3.14.1 (Variabile aleatoria discreta). *Una v.a. X definita su una terna di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ si dice discreta se il suo spetto S_X è discreto, cioè costituito da una quantità finita o alpiù numerabile di valori. In tal caso, anche la funzione di distribuzione F_X associata ad X si dirà discreta.*

Definizione 3.14.2 (Funzione di massa di probabilità). *Se X v.a. discreta allora esiste una funzione*

$$p_X : \mathbb{R} \to [0,1]$$
$$p_X(x) = \mathbb{P}(X = x), \forall x \in \mathbb{R}$$

Tale funzione p_X è detta funzione di massa di probabililtà (talvolta è indicata più brevemente con massa di probabililtà, funzione di massa o distribuzione).

Teorema 3.14.1 (Condizione di normalizzazione per la massa di probabilità). *Si consideri lo spazio $(\Omega, \mathcal{F}, \mathbb{P})$ e la v.a. X su esso definita. Sia $S_X = \{x_1, x_2, \cdots\}$. Allora si ha che*

$$\sum_{i\in\mathbb{N}} p_X(x_i) = 1$$

Dimostrazione.

$$\sum_{i\in\mathbb{N}} p_X(x_i) = \sum_{i\in\mathbb{N}} \mathbb{P}(X = x_i) = \mathbb{P}\left(\dot{\bigcup_{i\in\mathbb{N}}}(X = x_i)\right) = \mathbb{P}(X \in S_X) = \mathbb{P}(\Omega) = 1$$

□

3.14.1 Funzione di massa di probabilità e funzione di distribuzione

Si consideri lo spazio $(\Omega, \mathcal{F}, \mathbb{P})$ e la v.a. X su esso definita. Sia $S_X = \{x_1, x_2, \cdots\}$. Si ha che

$$F_X(x) = \sum_{i:x_i \leq x} p_X(x_i)$$

Al contrario, se nota la funzione di distribuzione, è possibile calcolare la massa di distribuzione in un punto con

$$p_X(x) = \mathbb{P}(X = x) = F_X(x) - F_X(x^-)$$

Esercizio 3.14.1 (Massa di prob. e funzione di distrubuzione)**.** *Si consideri nello spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *la v.a.* X. $S_X = \{1, 2, 3\}$. *Si ha inoltre che* $p_X(1) = \frac{1}{2}$, $p_X(2) = \frac{1}{3}$. *Definire e rappresentare graficamente la funzione di distribuzione di* X.

Innanzitutto, dal risultato 3.14.1 possiamo ricavare

$$p_X(1) + p_X(2) + p_X(3) = 1 \Longrightarrow p_X(3) = 1 - \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

Quindi possiamo definire la funzione di distribuzione di X:

$$F_X(x) = \begin{cases} 0 & , \ x \leq 0 \\ \frac{1}{2} & , \ 1 < x \leq 2 \\ \frac{1}{2} + \frac{1}{3} = \frac{5}{6} & , \ 2 < x \leq 3 \\ \frac{1}{2} + \frac{1}{3} + \frac{1}{6} = 1 & , \ x \geq 3 \end{cases}$$

3.15 Variabili aleatorie assolutamente continue

Definizione 3.15.1 (Variabile aleatoria assolutamente continua)**.** *Dato uno spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *la v.a.* X *su esso definita si dice assolutamente continua se esiste una funzione non negativa*

$$f_X : \mathbb{R} \to \mathbb{R}^+ \ \textit{tale che} \ F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

Osservazione 3.15.1. *In ogni punto* x *di continuità di* f_X *si ha che*

$$f_X(x) = \frac{d}{dx} F_X(x)$$

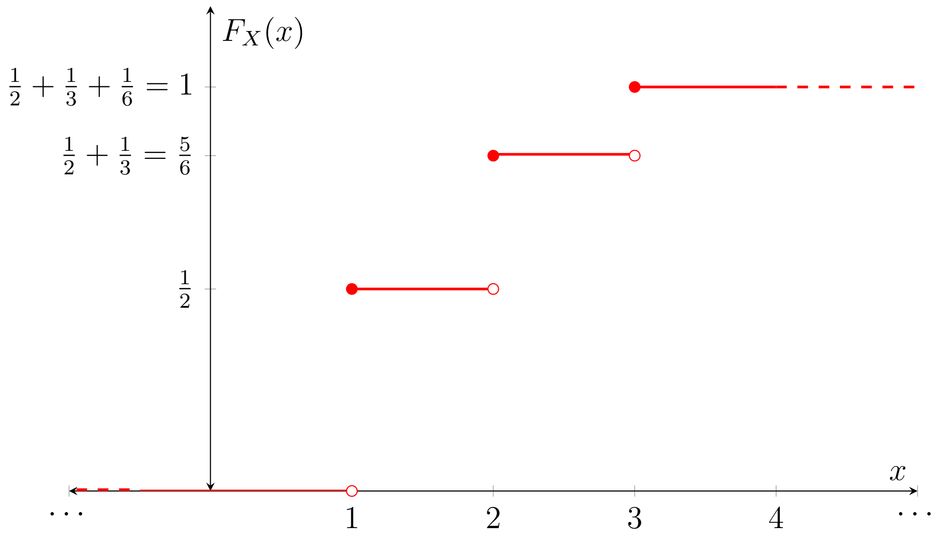


Figura 3.2: Rappresentazione di F_X dell'esercizio 3.14.1

Definizione 3.15.2 (Funzione di densità di probabilità)**.** *La funzione f_X della definizione precedente è detta funzione di densità di probabililtà della variabile aleatoria X.*

Teorema 3.15.1 (Condizione di normalizzazione per la densità di probabilità)**.** *Si consideri lo spazio $(\Omega, \mathcal{F}, \mathbb{P})$ e la v.a. X su esso definita. Si ha che*

$$\int_{-\infty}^{+\infty} f_X(t)dt = 1$$

Dimostrazione. Ricordando le proprietà (in particolare la *iii*) della funzione di distribuzione (3.13.1) si ha che

$$\lim_{x\to\infty} F_X(x) = 1 \Leftrightarrow \lim_{x\to\infty} \int_{-\infty}^{x} f_X(t)dt = 1 \Leftrightarrow \int_{-\infty}^{+\infty} f_X(t)dt = 1$$

□

Teorema 3.15.2 (Teorema di Cimino)**.** *Sia X una v.a. assolutamente continua definita sullo spazio $(\Omega, \mathcal{F}, \mathbb{P})$. Sia f_X la funzione di densità di probabilità di X. L'area sotto la curva f_X è uguale a 1.*

Dimostrazione. La dimostrazione segue istantaneamente dal teorema 3.15.1 e dalla definizione di integrale. □

Questo importante risultato deve la sua scoperta allo studente Maurizio Cimino, che giunse alla conclusione durante una lezione di calcolo delle probabililtà e statistica matematica.

Osservazione 3.15.2. *Sia X una v.a. assolutamente continua definita su $(\Omega, \mathcal{F}, \mathbb{P})$. Si ha che*

$$\mathbb{P}(X = x) = F_X(x) - F_X(x^-) = \int_{-\infty}^{x} f_X(t)dt - \int_{-\infty}^{x} f_X(t)dt = 0$$

Osservazione 3.15.3. *Sia X una v.a. assolutamente continua definita su $(\Omega, \mathcal{F}, \mathbb{P})$. Siano $a, b \in \mathbb{R},\ a < b$. Si ha che*

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(t)dt - \int_{-\infty}^{a} f_X(t)dt = \int_{a}^{b} f_X(t)dt$$

Possiamo inoltre confermare il risultato dell'osservazione precedente con

$$\mathbb{P}(X = a) = \mathbb{P}(a \leq X \leq a) = \int_{a}^{a} f_X(t)dt = 0$$

Esercizio 3.15.1 (V.v.a.a. assolutamente continue: densità di probabililtà e funzione di distribuzione)**.** *Sia data una v.a. X definita su $(\Omega, \mathcal{F}, \mathbb{P})$. Sia la densità di probabilità di X la seguente*

$$f_X(x) = \begin{cases} C(4x - 2x^2) & ,\ 0 < x < 2 \\ 0 & ,\ \textit{altrimenti} \end{cases}$$

Quanto deve valere C affinchè f_X sia una funzione di densità? Quanto vale $\mathbb{P}(X > 1)$? Definire la funzione di distribuzione di X.

Affinchè sia una funzione di densità di probabilità, f_X deve essere non negativa e soddisfare la condizione di normalizzazione (3.15.1).

$$\begin{aligned} \int_{-\infty}^{+\infty} f_X(t)dt = 1 &\Leftrightarrow \int_{-\infty}^{0} f_X(t)dt + \int_{0}^{2} f_X(t)dt + \int_{2}^{+\infty} f_X(t)dt = 1 \\ &\Leftrightarrow \int_{0}^{2} f_X(t)dt = 1 \\ &\Leftrightarrow \int_{0}^{2} C(4t - 2t^2)dt = 1 \\ &\Leftrightarrow 4C \int_{0}^{2} tdt - 2C \int_{0}^{2} t^2)dt = 1 \\ &\Leftrightarrow 4C \cdot \left.\frac{t^2}{2}\right|_0^2 - 2C \cdot \left.\frac{t^3}{3}\right|_0^2 = 1 \\ &\Leftrightarrow 4C \cdot 2 - 2C \cdot \frac{8}{3} = 1 \\ &\Leftrightarrow 8C - \frac{16}{3}C = 1 \\ &\Leftrightarrow \frac{8}{3}C = 1 \\ &\Leftrightarrow C = \frac{3}{8} \end{aligned}$$

Si verifica inoltre semplicemente che per $C = \frac{3}{8}$ la funzione non assume valori negativi. Calcoliamo $\mathbb{P}(X > 1)$.

$$\begin{aligned}
\mathbb{P}(X > 1) &= 1 - \mathbb{P}(X \leq 1) \\
&= 1 - F_X(1) \\
&= 1 - \int_{-\infty}^{1} f_X(t)dt \\
&= 1 - \left(\int_{-\infty}^{0} f_X(t)dt + \int_{0}^{1} f_X(t)dt \right) \\
&= 1 - 0 - \int_{0}^{1} \frac{3}{8} \left(4t - 2t^2\right) dt \\
&= 1 - \frac{3}{8} \left(4 \int_{0}^{1} tdt - 2 \int_{0}^{1} t^2 dt \right) \\
&= 1 - \frac{3}{2} \cdot \left.\frac{t^2}{2}\right|_0^1 + \frac{3}{4} \cdot \left.\frac{t^3}{3}\right|_0^1 \\
&= 1 - \frac{3}{2} \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{1}{3} \\
&= 1 - \frac{3}{4} + \frac{1}{4} \\
&= \frac{1}{2}
\end{aligned}$$

Definiamo ora F_X. Dalla definizione di funzione di densità e dalla traccia si ha:

$$F_X(x) = \begin{cases} \int_0^x \frac{3}{8} \left(4t - 2t^2\right) dt = \frac{3}{2} \cdot \left.\frac{t^2}{2}\right|_0^x - \frac{3}{4} \cdot \left.\frac{t^3}{3}\right|_0^x = \frac{3x^2}{4} - \frac{x^3}{4} & , \ 0 < x < 2 \\ 0 & , \ \text{altrimenti} \end{cases}$$

3.16 Variabili aleatorie studiate congiuntamente

Solitamente non si studia una sola variabile aleatoria, ma si è interessati al comportamento di 2 (o più) variabili aleatorie (d'ora in poi abbreviato in v.v.a.a.) definite sullo stesso spazio di probabilità. Di seguito sono trattati i casi con due v.v.a.a. per semplicità. I casi generici sono comunque analoghi.

Definizione 3.16.1 (Funzione di distribuzione congiunta di due v.v.a.a). *Date due v.v.a.a. X, Y definite su uno spazio $(\Omega, \mathcal{F}, \mathbb{P})$ comune, si definisce la funzione di distribuzione congiunta di X, Y come*

$$F_{X,Y} : \mathbb{R}^2 \to [0,1]$$
$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x; Y \leq y) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\} \cap \{\omega \in \Omega : Y(\omega) \leq y\})$$

Teorema 3.16.1 (Ricavare le funzioni di distribuzioni marginali dalla congiunta). *Date due v.v.a.a. X, Y definite su uno spazio $(\Omega, \mathcal{F}, \mathbb{P})$ e nota la funzione di distribuzione congiunta $F_{X,Y}$ si ha che*

$$F_X(x) = \lim_{y \to +\infty} F_{X,Y}(x,y)$$
$$F_Y(y) = \lim_{x \to +\infty} F_{X,Y}(x,y)$$

Dimostrazione.

$$\begin{aligned} F_X(x) = \mathbb{P}(X \leq x) &= \lim_{y \to +\infty} \mathbb{P}(X \leq x; Y < y) \qquad \text{(prende tutti i valori di Y)} \\ &= \lim_{y \to +\infty} F_{X,Y}(x,y) \end{aligned}$$

□

3.16.1 Caso discreto

Definizione 3.16.2 (Funzione di massa di probabilità congiunta)**.** *Si considerino* X, Y *v.v.a.a. discrete definite sullo spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *. Si avranno* $S_X = \{x_1, x_2, \cdots, x_i, \cdots\}$ *ed* $S_Y = \{y_1, y_2, \cdots, y_j, \cdots\}$*. Si definisce la funzione di massa di probabilità congiunta come*

$$p_{X,Y} : \mathbb{R}^2 \to [0, 1]$$
$$p_{X,Y}(x, y) = \mathbb{P}(X = x; Y = y)$$

Teorema 3.16.2 (Ricavare le funzioni di massa di probabililtà marginali dalla congiunta)**.** *Date due v.v.a.a.* X, Y *definite su uno spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *e nota la funzione di massa di probabilità congiunta* $p_{X,Y}$ *si ha che*

$$p_X(x) = \sum_j p_{X,Y}(x, y_i)$$
$$p_Y(y) = \sum_i F_{X,Y}(x_i, y)$$

Dimostrazione.

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}\left(\dot{\bigcup_j}\{X = x; Y = y_j\}\right) = \sum_j \mathbb{P}(X = x; Y = y_j) = \sum_j p_{X,Y}(x, y_j)$$

□

Esercizio 3.16.1 (Massa di probabilità di v.v.a.a.). *In una certa popolazione il 15% delle coppie non ha figli, il 20% ha un figlio, il 35% ha due figli e il 30% ne ha tre. Indipendentemente dagli altri eventuali fratelli/sorelle, ogni bambino ha pari possibilità di essere maschio o femmina. Si denoti con X il numero di figli femmina e con Y il numero di maschi di una data famiglia. Costruire una tabella a doppia entrata con le masse di probabilità congiunte di X, Y.*

Lo spettro di X così come quello di Y sarà uguale a $S_X = S_Y = \{0, 1, 2, 3\}$.

$$\begin{aligned}
p_{X,Y}(0,0) &= \mathbb{P}(X = 0; Y = 0) = 0.15 \\
p_{X,Y}(1,0) &= p_{X,Y}(0,1) = \mathbb{P}(X = 1; Y = 0) \\
&= \mathbb{P}(\text{Ha un figlio ed è femmina}) \\
&= \mathbb{P}(\text{Dato che ha un figlio, questo è femmina}) \cdot \mathbb{P}(\text{Ha un figlio}) \\
&= 0.5 \cdot 0.2 = 0.1 \\
p_{X,Y}(2,0) &= p_{X,Y}(0,2) = \mathbb{P}(X = 2; Y = 0) \\
&= \mathbb{P}(\text{Ha due figli e sono entrambi femmina}) \\
&= \mathbb{P}(\text{Dato che ha due figlii, sono entrambi femmina}) \cdot \mathbb{P}(\text{Ha due figli}) \\
&= 0.25 \cdot 0.35 = 0.0875 \\
p_{X,Y}(3,0) &= p_{X,Y}(0,3) = \mathbb{P}(X = 3; Y = 0) \\
&= \mathbb{P}(\text{Ha tre figli e sono tutti femmina}) \\
&= \mathbb{P}(\text{Dato che ha tre figli, sono tutti femmina}) \cdot \mathbb{P}(\text{Ha tre figli}) \\
&= 0.125 \cdot 0.3 = 0.0375 \\
p_{X,Y}(1,1) &= \mathbb{P}(X = 1; Y = 1) \\
&= \mathbb{P}(\text{Ha due figli e sono un maschio e una femmina}) \\
&= \mathbb{P}(\text{Dato che ha due figli, sono un maschio e una femmina}) \cdot \mathbb{P}(\text{Ha due figli}) \\
&= 0.5 \cdot 0.35 = 0.175 \\
p_{X,Y}(2,1) &= p_{X,Y}(1,2) = \mathbb{P}(X = 2; Y = 1) \\
&= \mathbb{P}(\text{Ha tre figli e sono un maschio e due femmina}) \\
&= \mathbb{P}(\text{Dato che ha tre figli, sono un maschio e due femmine}) \cdot \mathbb{P}(\text{Ha tre figli}) \\
&= \frac{3}{8} \cdot 0.3 = 0.1125 \\
p_{X,Y}(2,2) &= p_{X,Y}(3,1) = p_{X,Y}(1,3) = p_{X,Y}(2,3) = p_{X,Y}(3,2) = p_{X,Y}(3,3) = 0
\end{aligned}$$

Costruiamo quindi la tabella a doppia entrata:

		X				
		0	1	2	3	
	0	0.15	0.1	0.0875	0.0375	0.375
Y	1	0.1	0.175	0.1125	0	0.3875
	2	0.0875	0.1125	0	0	0.2
	3	0.0375	0	0	0	0.375
		0.375	0.3875	0.2	0.375	1

Si noti come, nella tabella precedente, la somma dei valori delle colonne corrisponda a $p_X(x)$ mentra la somma dei valori di ciascuna riga corrisponde a $p_Y(y)$. Si può verificare la condizione di normalizzazione notando che la somma delle somme delle righe (così come la somma della somma delle colonne) ammonta sempre ad 1.

3.16.2 Caso assolutamente continuo

Definizione 3.16.3 (V.v.a.a. distribuite congiuntamente e continue). *Siano dati* $(\Omega, \mathcal{F}, \mathbb{P})$ *e* X, Y *v.v.a.a. definite su esso.* X, Y *si dicono distribuite congiuntamente e continue se esiste una funzione non negativa*

$$f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$$
$$F_{X,Y}(x, y) = \iint_{(x,y):X \le x; Y \le y} f_X(s, t)\ ds\ dt$$

Definizione 3.16.4 (Funzione di densità congiunta). *La funzione* $f_{X,Y}$ *della definizione precedente è detta funzione di densità congiunta delle variabili aleatorie* X, Y*.*

Teorema 3.16.3 (Ricavare la densità congiunta dalla funzione di distribuzione congiunta). *Siano dati* $(\Omega, \mathcal{F}, \mathbb{P})$ *e* X, Y *v.v.a.a. definite su esso.* X, Y *si dicono distribuite congiuntamente e continue. Si ha che*

$$f_{X,Y} = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$$

Teorema 3.16.4 (Ricavare le densità marginali dalla congiunta). *Date due v.v.a.a.* X, Y *definite su uno spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *e nota la funzione di massa di probabilità congiunta* $p_{X,Y}$

si ha che

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$
$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)dx = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx$$

3.16.3 Variabili aleatorie indipendenti

Definizione 3.16.5 (Variabili aleatorie indipendenti)**.** *Sia dato uno spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *e due v.v.a.a.* X, Y *su esso definite.* X, Y *sono indipendenti se ogni evento relativo a* X *è indipendente da ogni evento relativo a* Y*.*
Siano I *e* J *intervalli di* $\mathbb{R}$*. Se* X, Y *sono indipendenti, la loro funzione di distribuzione congiunta si fattorizza nel prodotto delle distribuzioni singole.*

$$\begin{aligned} X, Y \text{ indipendenti } &\Leftrightarrow \mathbb{P}(X \in I; Y \in J) = \mathbb{P}(X \in I) \cdot \mathbb{P}(Y \in J) \\ &\Leftrightarrow F_{X,Y}(x,y) = \mathbb{P}(X \le x; Y \le y) = \mathbb{P}(X \le x) \cdot \mathbb{P}(Y \le y) = F_X(x) \cdot F_Y(y) \end{aligned}$$

Se X, Y *sono v.v.a.a. discrete, si ha che*

$$X, Y \text{ indipendenti } \Leftrightarrow p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$$

Se X, Y *sono v.v.a.a. assolutamente continue, si ha che*

$$X, Y \text{ indipendenti } \Leftrightarrow f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

Osservazione 3.16.1 (Calcolo della funzione di distribuzione congiunta dalla densità congiunta nel caso di v.v.a.a. indipendenti e assolutamente continue)**.** *In queste particolari condizioni si ha che:*

$$\begin{aligned} F_{X,Y}(x,y) &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)\ ds\ dt \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_X(s) f_Y(t)\ ds\ dt \\ &= \int_{-\infty}^{x} f_X(s)\ ds \cdot \int_{-\infty}^{y} f_Y(t)\ dt \end{aligned}$$

Esercizio 3.16.2 (Variabili aleatorie indipendenti)**.** *Siano X, Y v.v.a.a. indipendenti definite su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Si ha che*

$$f_X(t) = f_Y(t) = \begin{cases} e^{-t} & , \; t > 0 \\ 0 & , \; altrove \end{cases}$$

Calcolare la densità della variabile aleatoria $Z = \frac{X}{Y}$.

$$\begin{aligned}
F_Z(z) &= \mathbb{P}(Z \le z) \\
&= \mathbb{P}(\frac{X}{Y} \le z) \\
&= \iint\limits_{(x,y):\frac{x}{y}\le z} f_{X,Y}(s,t)\ ds\ dt \\
&= \iint\limits_{(x,y):\frac{x}{y}\le z} f_X(s)\ ds\ f_Y(t)\ dt \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{zy} f_X(s)\ ds\ f_Y(t)\ dt \quad (\frac{x}{y} \le z \Leftrightarrow x \le yz) \\
&= \int_0^{+\infty} e^{-t} \boxed{\int_0^{zy} e^{-s}\ ds}\ dt \quad \big(\text{divido gli integrali e semplifico perchè } f_X(t) = F_Y(t) = 0 \text{ quando } t \le 0\big) \\
&= \int_0^{+\infty} e^{-t} \left[-e^{-s} \Big|_0^{yz} \right] dt \quad \text{(ho svolto l'integrale evidenziato)} \\
&= \int_0^{+\infty} e^{-t} \left[1 - e^{-yz} \right] dt \\
&= 1 - \frac{1}{z+1}
\end{aligned}$$

Per individuare quindi f_Z è necessario derivare F_Z:

$$\frac{d}{dz} F_Z(z) = \frac{d}{dz}\left(\frac{1}{z+1}\right) = -\frac{1}{(z+1)^2}$$

3.17 Distribuzione condizionale di due v.v.a.a.

Definizione 3.17.1 (Funzione di massa di probabilità condizionata)**.** *Siano X, Y v.v.a.a. discrete definite su $(\Omega, \mathcal{F}, \mathbb{P})$. p_X, p_Y, $p_{X,Y}$ sono le funzioni di massa di probabilità rispettivamente di X, Y e della distribuzione congiunta di X e Y.*

$$p_{X|Y}(x|y) =\Leftrightarrow \mathbb{P}(\{X = x\}|\{Y = y\}) = \frac{\mathbb{P}(\{X = x\} \cap \{Y = y\})}{\mathbb{P}(\{Y = y\})} = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

Esercizio 3.17.1 (Funzione di massa di probabilità condizionata). *Si consideri l'esercizio 3.16.1 svolto precedentemente. Se si aggiunge l'informazione che la famiglia scelta ha esattamente una figlia, qual è la funzione di massa di probabilità condizionata del numero di figli maschi?*

Sappiamo che si verifica l'evento $\{X = 1\}$. Vogliamo calcolare $p_{Y|X}(y, 1)$. Conosciamo $S_Y = \{0, 1, 2, 3\}$.

$$\begin{aligned}
p_{Y|X}(0,1) &= \frac{p_{Y,X}(0,1)}{p_X(1)} = \frac{0.1}{0.3875} = 0.2580 \\
p_{Y|X}(1,1) &= \frac{p_{Y,X}(1,1)}{p_X(1)} = \frac{0.175}{0.3875} \approx 0.45 \\
p_{Y|X}(2,1) &= \frac{p_{Y,X}(2,1)}{p_X(1)} = \frac{0.1125}{0.3875} \approx 0.29 \\
p_{Y|X}(3,1) &= \frac{p_{Y,X}(3,1)}{p_X(1)} = \frac{0.0}{0.3875} = 0
\end{aligned}$$

Esercizio 3.17.2 (Funzione di massa di probabilità condizionata). *Si consideri l'esercizio 3.16.1 svolto precedentemente.Si aggiunge l'informazione che la famiglia scelta ha esattamente una figlia, come nell'esercizio precedente. Qual è la probabilità che una famiglia abbia alpiù un figlio maschio?*

Vogliamo calcolare $\mathbb{P}(Y \leq 1 | X = 1)$.

$$\begin{aligned}
\mathbb{P}(Y \leq 1 | X = 1) &= F_{Y|X}(1, \{X = 1\}) \\
&= \sum_{i=0}^{1} p_{Y|X}(i|1) \\
&= p_{Y|X}(0,1) + p_{Y|X}(1,1) \\
&\approx 0.2580 + 0.45 \approx 0.708
\end{aligned}$$

Definizione 3.17.2 (Funzione di densità condizionata). *Siano X, Y v.v.a.a. assolutamente continue definite su $(\Omega, \mathcal{F}, \mathbb{P})$. f_X, f_Y, $f_{X,Y}$ sono le funzioni di densità di probabilità rispettivamente di X, Y e della distribuzione congiunta di X e Y. Si dice densità condizionata di $X|Y$ la funzione*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \qquad (f_Y(y) > 0\ \forall y)$$

Esercizio 3.17.3 (Funzione di densità condizionata). *Siano X, Y v.v.a.a. definite su uno spazio $(\Omega, \mathcal{F}, \mathbb{P})$. È data la funzione di densità congiunta*

$$f_{X,Y}(x,y) = \begin{cases} \frac{12}{5}(2-x-y) & \text{, se } 0 < x < 1 \wedge 0 < y < 1 \\ 0 & \text{, altrimenti} \end{cases}$$

Calcolare la densità condizionata di X rispetto ad Y

Per la definizione 3.17.2 si ha

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{\frac{12}{5}x(2-x-y)}{f_Y(y)} & \text{, se } 0 < x < 1 \wedge 0 < y < 1 \\ 0 & \text{, altrimenti} \end{cases} \qquad (\star)$$

Occorre quindi calcolare $f_Y(y)$ dalla densità congiunta:

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} f_{X,Y}(x,y)\ dx \\ &= \int_{-\infty}^{0} 0\ dx + \int_0^1 \frac{12}{x}(2-x-y)\ dx + \int_1^{+\infty} 0\ dx \\ &= \int_0^1 \frac{12}{5}x(2-x-y)\ dx \\ &= \int_0^1 \frac{24}{5}x\ dx - \int_0^1 \frac{12}{5}x^2\ dx - \int_0^1 \frac{12}{5}xy\ dx \\ &= \frac{24}{5}\cdot\frac{x^2}{2}\Big|_0^1 - \frac{12}{5}\cdot\frac{x^3}{3}\Big|_0^1 - \frac{12}{5}y\cdot\frac{x^2}{2}\Big|_0^1 \\ &= \frac{24}{5}\cdot\frac{1}{2} - \frac{12}{5}\cdot\frac{1}{3} - \frac{12}{5}y\cdot\frac{1}{2} \\ &= \frac{12}{5} - \frac{4}{5} - \frac{6}{5}y \\ &= \frac{8}{5} - \frac{6}{5}y \end{aligned}$$

Sostituendo in $(\star)$, quindi, ricaviamo che

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{\frac{12}{5}x(2-x-y)}{\frac{8}{5}-\frac{6}{5}y} & \text{, se } 0 < x < 1 \wedge 0 < y < 1 \\ 0 & \text{, altrimenti} \end{cases}$$

3.18 Valore atteso

Definizione 3.18.1 (Valore atteso di una v.a. discreta)**.** *Sia X una v.a. discreta definita su $(\Omega, \mathcal{F}, \mathbb{P})$. Si avrà $S_X = \{x_1, x_2, \cdots\}$. Si definisce valore atteso di X (anche detto media, valore medio, aspettazione o expectation) il numero*

$$E[X] = \sum_i (x_i \cdot p_X(x_i))$$

Esercizio 3.18.1 (Valore atteso di una v.a. discreta: lancio dado)**.** *Sia $(\Omega, \mathcal{F}, \mathbb{P})$ la terna costruita sul lancio di un dado equo. Sia X la v.a. (dicreta) che conta il numero di pallini disegnati sulla faccia del dado rivolta verso l'alto dopo un lancio. Calcolare il valore atteso di X.*

$$E[X] = \sum_{i=1}^{6} (ip_X(i)) = \frac{1}{6} \cdot (1+2+3+4+5+6) = \frac{7}{2} = 3.5$$

Si noti come il valore atteso conservi l'unità di misura, ma non assuma necessariamente valori nello spettro della variabile aleatoria.

Esercizio 3.18.2 (Valore atteso di una v.a. discreta: funzione rilevatrice)**.** *Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità. Sia $A \in \mathcal{F}$. Si calcoli il valore atteso della variabile aleatoria indicatrice di A (definita in 3.12.3).*

$$I_A = \begin{cases} 1 & , \text{ se } \omega \in A \\ 0 & , \text{ se } \omega \notin A \end{cases} \qquad\qquad S_{I_A} = \{0, 1\}$$

Applicando la definizione di ha

$$E[I_A] = 0 \cdot p_{I_A}(0) + 1 \cdot p_{I_A}(1)$$

$$\begin{aligned} p_{I_A}(1) &= \mathbb{P}(I_A = 1) = \mathbb{P}(\{\omega \in \Omega : I_A(\omega) = 1\}) \\ &= \mathbb{P}(\{\omega \in \Omega : \omega \in A\}) = \mathbb{P}(A) \end{aligned}$$

Analogamente si avrà

$$P_{I_A}(0) = \mathbb{P}(A^C)$$

Sostituendo quindi nella formula della definizione si ha

$$E[I_A] = 0 \cdot \mathbb{P}(A^C) + 1 \cdot \mathbb{P}(A) = \mathbb{P}(A)$$

Ottenendo quindi che il valore atteso della variabile aleatoria indicatrice di un evento coincide con la probabilità dell'evento stesso.

Definizione 3.18.2 (Valore atteso di una v.a. assolutamente continua)**.** *Sia X una v.a. assolutamente continua definita su $(\Omega, \mathcal{F}, \mathbb{P})$. f_X sia la sua densità di probabilità. Si definisce valore atteso di X (se esiste) il numero*

$$E[X] = \int_{\mathbb{R}} x \cdot f_X(x)\ dx$$

Esercizio 3.18.3 (Valore atteso di una v.a. ass. cont.: tempo di attesa)**.** *Si è in attesa di una comunicazione che deve arrivare dopo le 17. Sia $X =$"il numero di ore di attesa dopo le 17 per ricevere la comunicazione". X è una v.a. assolutamente continua. È data la densità di X:*

$$f_X(x) = \begin{cases} \frac{1}{1.5} & , \textit{ se } 0 < x < 1.5 \\ 0 & , \textit{ altrove} \end{cases}$$

Calcolare il valore atteso del tempo che trascorre tra le 17 e l'arrivo della comunicazione.

Verifichiamo innanzitutto che f_X sia una densità:

✓ f_X assume valori in $\mathbb{R}^+$ (lo si verifica immediatamente dalla definizione di f_X.

✓ f_X soddisfa la condizione di normalizzazione avendosi $\int_{\mathbb{R}} f_X(x)\ dx = 1$.

Calcoliamo il valore atteso:

$$\begin{aligned} E[X] &= \int_{\mathbb{R}} x \cdot f_X(x)\ dx \\ &= \int_{-\infty}^{0} x \cdot 0\ dx + \int_{0}^{1.5} x \cdot \frac{1}{1.5}\ dx + \int_{1.5}^{+\infty} x \cdot 0\ dx \\ &= \frac{1}{1.5} \cdot \int_{0}^{1.5} x\ dx \\ &= \frac{1}{1.5} \cdot \left. \frac{x^2}{2} \right|_{0}^{1.5} \\ &= \frac{1}{1.5} \cdot \frac{(1.5)^2}{2} \\ &= \frac{1.5}{2} = 0.75 \end{aligned}$$

3.19 Trasformazioni di variabili aleatorie

Sia X una variabile aleatoria su $(\Omega, \mathcal{F}, \mathbb{P})$. Se consideriamo una funzione $g : \mathbb{R} \to \mathbb{R}$, in quali condizioni è possibile affermare che $Y = g(X)$ è a sua volta una v.a.?

Ricordando la definizione di v.a. (3.12.1), possiamo affermare che, preso uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, la $X : \Omega \to \mathbb{R}$ su esso definita è una v.a. se, per ogni aperto $\mathbb{B} =]-\infty, x]$ di $\mathbb{R}$ [2] si ha che

$$\{\omega \in \Omega : X(\omega) \in \mathbb{B}\} \in \mathcal{F}$$

Dunque,

$$\begin{aligned} Y = g(X) \text{ è una v.a. } &\iff \{\omega \in \Omega : Y(\omega) \in]-\infty, y]\} \in \mathcal{F} \qquad && \forall y \\ &\iff \{\omega \in \Omega : g(X(\omega)) \in]-\infty, y]\} \in \mathcal{F} \qquad && \forall y \end{aligned}$$

In generale è possibile affermare che se g è monotòna o alpiù monotòna a tratti (cioè se g è invertibile), allora $Y = g(X)$ è ancora una variabile aleatoria.

Teorema 3.19.1 (Teorema delle trasformazioni monotòne). *Seguono una serie di risultati (espressione delle funzioni di distribuzione, di massa, di densità di $Y = g(X)$ in funzione delle omologhe di X). In tutti i casi assumeremo X v.a. e $g : \mathbb{R} \to \mathbb{R}$ funzione monotòna tale che $Y = g(X)$ sia una v.a.*

- ***Funzioni di distribuzione.***
  Caso di g non decrescente.

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

  Caso di g non crescente.

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)^-)$$

- ***Funzioni di massa di probabilità.*** *Assumiamo X v.a. discreta.*

$$p_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(g(X) = y) = \mathbb{P}(X = g^{-1}(y)) = p_X(g^{-1}(y))$$

- ***Funzioni di densità di probabilità.*** *Assumiamo X v.a. assolutamente continua.*

[2] chi lo desidera può approfondire qui: `https://it.wikipedia.org/wiki/Algebra_di_Borel`

Sia F_X la sua funzione di distribuzione.

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(g^{-1}(y)) = \frac{d}{dy}F_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y)$$

Esercizio 3.19.1 (Massa di probabilità della v.a. trasformata)**.** *Sia X v.a. discreta. $S_X = \{0, 1, 2\}$. $p_X(0) = 0.2$, $p_X(1) = 0.5$, $p_X(2) = 0.3$.*
$Y = X^2$. Calcolare p_Y.

$S_Y = \{0, 1, 4\}$.

$$\begin{aligned}
p_Y(0) &= \mathbb{P}(Y = 0) = \mathbb{P}(X^2) = 0 = \mathbb{P}(X = 0) = p_X(0) = 0.2 \\
p_Y(1) &= \mathbb{P}(Y = 1) = \mathbb{P}(X^2) = 1 = \mathbb{P}(\{X = 1\} \cup \{X = -1\}) = p_X(1) + 0 = 0.5 \\
p_Y(4) &= \mathbb{P}(Y = 4) = \mathbb{P}(X^2) = 4 = \mathbb{P}(\{X = 2\} \cup \{X = -2\}) = p_X(2) + 0 = 0.3
\end{aligned}$$

Esercizio 3.19.2 (Densità di probabilità della v.a. trasformata)**.** *Sia X v.a. assolutamente continua corrispondente al tempo in ore necessario a localizzare un guasto in un impianto elettrico. È data la funzione di densità:*

$$f_X(x) = \begin{cases} 1 & , 0 < x < 1 \\ 0 & , \textit{altrimenti} \end{cases}$$

Se il danno economico provocato da un'interruzione di x ore è di x^3, calcolare la funzione di distribuzione e la densità della v.a. Y che rappresenta il danno economico.

Nota f_X, calcoliamo la funzione di distribuzione F_X.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\ dt = \begin{cases} \int_{-\infty}^{x} 0\ dt = 0 & , x \leq 0 \\ \int_{-\infty}^{x} 1\ dt = x & , 0 < x < 1 \\ \int_{-\infty}^{x} 0\ dt = 1 & , x \geq 1 \end{cases}$$

Quindi possiamo esprimere la funzione di distribuzione di Y:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^3 \leq y) = \mathbb{P}(X \leq \sqrt[3]{y}) = F_X(\sqrt[3]{y})$$

Infine, possiamo calcolare anche la densità di Y:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} F_X(\sqrt[3]{y}) \\ &= f_X(\sqrt[3]{y}) \cdot \frac{d}{dy} \sqrt[3]{y} = \begin{cases} \frac{1}{3\sqrt[3]{y^2}} & , \ 0 < x < 1 \\ 0 & , \text{ altrimenti} \end{cases} \end{aligned}$$

3.20 Proprietà del valore atteso

Di seguito sono mostrate alcune proprietà del valore atteso.

3.20.1 Formula alternativa per il valore atteso di una v.a. trasformata

Teorema 3.20.1 (Formula alternativa per il valore atteso di una v.a. trasformata (caso discreto))**.** *Sia X una v.a. discreta definita sullo spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *. Sia g una funzione tale che $Y = g(X)$ sia ancora una v.a. discreta. Si ha che*

$$E[Y] = \sum_x g(x) \cdot p_X(x)$$

Dimostrazione. Per definizione di valore atteso si ha

$$E[Y] = \sum_y y \cdot p_Y(y) = \sum_y y \cdot \mathbb{P}(Y = y)$$

Sostituendo l'informazione che $Y = g(X)$, abbiamo

$$E[Y] = E[g(X)] = \sum_x g(x) \cdot p_X(x)$$

□

Esercizio 3.20.1 (Continuando l'es. 3.19.1)**.** *Facendo riferimento all'esercizio 3.19.1 svolto in precedenza, calcolare il valore atteso della Y senza definire p_Y.*

Applichiamo la formula appena mostrata:

$$E[Y] = \sum_x g(x) \cdot p_X(x) = 0^2 \cdot 0.2 + 1^2 \cdot 0.5 + 2^2 \cdot 0.3 = 1.7$$

Sfruttiamo il fatto di aver già calcolato nell'esercizio 3.19.1 la funzione p_Y e verifichiamo il risultato con la formula della definizione di valore atteso:

$$E[Y] = \sum_y y \cdot p_Y(y) = 0 \cdot 0.2 + 1 \cdot 0.5 + 4 \cdot 0.3 = 1.7$$

Teorema 3.20.2 (Formula alternativa per il valore atteso di una v.a. trasformata (caso continuo))**.** *Sia X una v.a. assolutamente continua definita sullo spazio $(\Omega, \mathcal{F}, \mathbb{P})$. Sia g una funzione tale che $Y = g(X)$ sia ancora una v.a. assolutamente continua. Si ha che*

$$E[Y] = \int_{\mathbb{R}} g(x) \cdot f_X(x) \; dx$$

Dimostrazione. Per definizione di valore atteso si ha

$$E[Y] = \int_{\mathbb{R}} y \cdot f_Y(y) \; dy$$

Sostituendo l'informazione che $Y = g(X)$, abbiamo

$$E[Y] = E[g(X)] = \int_{\mathbb{R}} g(x) \cdot f_X(x) \; dx$$

□

Esercizio 3.20.2 (Continuando l'es. 3.19.2)**.** *Facendo riferimento all'esercizio 3.19.2 svolto in precedenza, calcolare il valore atteso della Y senza definire f_Y.*

Applichiamo la formula appena mostrata:

$$E[Y] = \int_0^1 g(x) \cdot 1 \; dx = \int_0^1 x^3 \cdot 1 \; dx = \left.\frac{x^4}{4}\right|_0^1 = \frac{1}{4}$$

Sfruttiamo il fatto di aver già calcolato nell'esercizio 3.19.2 la funzione f_Y e verifichiamo il risultato con la formula della definizione di valore atteso:

$$E[Y] = \int_0^1 y \cdot f_Y(y) = \int_0^1 y \cdot \frac{1}{3} y^{-\frac{2}{3}} \; dy = \frac{1}{3} \int_0^1 y^{\frac{1}{3}} \; dy = \frac{1}{3} \cdot \left.\frac{y^{\frac{4}{3}}}{\frac{4}{3}}\right|_0^1 = \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4}$$

Si noti come il calcolo di f_Y nell'esercizio 3.19.2 non sia stato banale ed abbia richiesto comunque alcuni passaggi. La formula alternativa permette di evitare quei passaggi per il calcolo della densità qualora interessi soltanto il valore atteso.

3.20.2 Valore atteso di una trasformazione lineare di v.a.

Teorema 3.20.3 (Valore atteso di una trasformazione lineare di una v.a.)**.** *Sia X una v.a. definita su $(\Omega, \mathcal{F}, \mathbb{P})$. Si ha che la media di una trasformazione lineare è la trasformazione lineare della media, ovvero:*

$$\forall a, b \in \mathbb{R}, \;\; E(aX + b) = aE[X] + b$$

Dimostrazione. **Caso discreto**. Sia X una v.a. discreta e siano $a, b \in \mathbb{R}$.

$$\begin{aligned}
E[aX + b] &= \sum_x ((ax + b) \cdot p_X(x)) \\
&= \sum_x ((ax \cdot p_X(x)) + (b \cdot p_X(x))) \\
&= \sum_x ax \cdot p_X(x) + \sum_x (b \cdot p_X(x)) \\
&= a \sum_x x \cdot p_X(x) + b \sum_x p_X(x) \\
&= aE[x] + b
\end{aligned}$$

Caso continuo. Sia X una v.a. assolutamente continua e siano $a, b \in \mathbb{R}$.

$$\begin{aligned}
E[aX + b] &= \int_{\mathbb{R}} ((ax + b) \cdot f_X(x) \; dx) \\
&= \int_{\mathbb{R}} ((ax \cdot f_X(x)) + (b \cdot f_X(x))) \\
&= \int_{\mathbb{R}} ax \cdot f_X(x) + \int_{\mathbb{R}} (b \cdot f_X(x)) \\
&= a \int_{\mathbb{R}} x \cdot f_X(x) + b \int_{\mathbb{R}} f_X(x) \\
&= aE[x] + b
\end{aligned}$$

□

Osservazione 3.20.1. *Non abbiamo fatto alcuna assunzione sugli $a, b \in \mathbb{R}$ usati in precedenza, quindi possiamo dedurre da quanto ricavato alcuni casi particolari. Nel caso in cui $b = 0$ si ha*

$$E[aX] = aE[X]$$

Nel caso in cui $a = 0$ si ha

$$E[b] = b$$

Una costante può essere vista come una variabile aleatoria ***degenere*** *che assume sempre il valore della costante. Di conseguenza, anche il valore atteso si attesterà sullo stesso valore.*

3.20.3 Valore atteso di trasformazioni di v.v.a.a. congiuntamente distribuite

Definizione 3.20.1 (Valore atteso di trasformazioni di v.v.a.a. congiuntamente distribuite)**.** *Si consideri lo spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *e le v.v.a.a.* X, Y *su esso definite e congiuntamente distribuite. Se* X, Y *sono discrete, è nota* $p_{X,Y}$ *mentre se continue è nota* $f_{X,Y}$*. Si ha:*

$$E[g(X,Y)] = \begin{cases} \sum\limits_x \sum\limits_y g(x,y) \cdot p_{X,Y}(x,y) & \text{, se } X, Y \text{ discrete} \\ \int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y) \cdot f_{X,Y}(x,y) & \text{, se } X, Y \text{ assolutamente continue} \end{cases}$$

Teorema 3.20.4 (Valore atteso della somma di v.v.a.a.)**.** *Si consideri lo spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *e le v.v.a.a.* X, Y *su esso definite e congiuntamente distribuite. Si ha che il valore atteso della somma di v.v.a.a. è la somma dei singoli valori attesi:*

$$E[X+Y] = E[X] + E[Y]$$

In generale la formula sopra si può estendere a somme di n *v.v.a.a.*

Dimostrazione. **Caso discreto**. È nota $p_{X,Y}$. Sia $g(X,Y) = X + Y$. Applicando 3.20.3:

$$\begin{aligned} E[X+Y] &= \sum_x \sum_y (x+y) \cdot p_{X,Y}(x,y) \\ &= \sum_x \sum_y x \cdot p_{X,Y}(x,y) + \sum_x \sum_y y \cdot p_{X,Y}(x,y) \\ &= \sum_x x \sum_y p_{X,Y}(x,y) + \sum_y y \sum_x p_{X,Y}(x,y) \\ &= \sum_x x \cdot p_X(x) + \sum_y y \cdot p_Y(y) \\ &= E[X] + E[Y] \end{aligned}$$

Caso assolutamente continuo. È nota $f_{X,Y}$. Sia $g(X,Y) = X+Y$. Applicando 3.20.3:

$$\begin{aligned} E[X+Y] &= \int_{\mathbb{R}} \int_{\mathbb{R}} (x+y) \cdot f_{X,Y}(x,y)\ dx\ dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} x \cdot f_{X,Y}(x,y)\ dx\ dy + \int_{\mathbb{R}} \int_{\mathbb{R}} y \cdot f_{X,Y}(x,y)\ dx\ dy \\ &= \int_{\mathbb{R}} x \int_{\mathbb{R}} f_{X,Y}(x,y)\ dy\ dx + \int_{\mathbb{R}} y \int_{\mathbb{R}} f_{X,Y}(x,y)\ dx\ dy \\ &= \int_{\mathbb{R}} x \cdot f_X(x)\ dx + \int_{\mathbb{R}} y \cdot f_Y(y)\ dy \\ &= E[X] + E[Y] \end{aligned}$$

□

Esercizio 3.20.3 (Proprietà del valore atteso: Segretaria fa cadere le lettere)**.** *Una segretaria ha appena finito di scrivere n lettere e compilare le corrispondenti n buste. Prima che possa inserire ciascuna lettera nella propria busta, però, rovescia la pila di lettere e buste. Se inserisce le lettere nelle buste in maniera casuale (ogni lettera può finire in ogni busta con pari probabilità), qual è il numero medio di lettere che capitano nella busta corretta?*

Siano

$$X = \text{“numero di lettere che capitano nella busta corretta”}$$

$$X_i = \begin{cases} 1 \text{ , se la lettera i-esima capita nella busta corretta} \\ 0 \text{ , altrimenti} \end{cases}$$

Quindi

$$X = \sum_{i=1}^{n} X_i$$

Si ha inoltre che

$$p_{X_i}(1) = \mathbb{P}(X_i = 1) = \frac{1}{n} \quad \forall i$$

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n}\sum_{x=0}^{1} x \cdot p_{X_i}(x) = \sum_{i=1}^{n} 1 \cdot \frac{1}{n} = 1$$

3.21 Momenti di ordine n-esimo intorno all'origine

Definizione 3.21.1 (Momento di ordine n-esimo intorno all'origine)**.** *Sia X una v.a. definita su uno spazio $(\Omega, \mathcal{F}, \mathbb{P})$. Si definisce momento di ordine $n \in \mathbb{N}$ intorno all'origine il valore*

$$E[X^n] = \begin{cases} \sum_x x^n p_X(x) & \text{, se } X \text{ discreta} \\ \int_{\mathbb{R}} x^n f_X(x) & \text{, se } X \text{ assolutamente continua} \end{cases}$$

Osservazione 3.21.1. *Si noti che il momento del primo ordine intorno all'origine è il valore atteso.*

Osservazione 3.21.2. *È possibile definire in maniera generale (indipendente dal tipo di v.a.) il valore atteso nel seguente modo:*

$$E[X] = \int_{\mathbb{R}} x \; dF_X(x)$$

Nel caso discreto, l'integrale sopra si può scrivere come la sommatoria che abbiamo usato nelle definizioni finora.

3.22 Momenti di ordine n-esimo intorno alla media

Definizione 3.22.1 (Momenti di ordine n-esimo intorno alla media)**.** *Sia dato uno spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *ed una v.a.* X *su esso definita. Sia* X *dotata di valore atteso* $E[X] = \mu$*. Si definiscono (laddove esistono) i momenti di ordine n intorno alla media come*

$$E\Big[\big[X - E[X]\big]^n\Big] = \begin{cases} \sum_x (x-\mu)^n \cdot p_X(x) & \text{, se } X \text{ discreta} \\ \int_{\mathbb{R}} (x-\mu)^n \cdot f_X(x)\ dx & \text{, se } X \text{ assolutamente continua} \end{cases}$$

3.23 Varianza

Definizione 3.23.1 (Varianza)**.** *La varianza di una v.a.* X *definita sullo spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *è il momento intorno alla media di ordine due:*

$$Var(X) = D^2(X) = E\Big[\big[X - E[X]\big]^2\Big]$$

La varianza indica in media quadratica la distanza della X *dalla sua media.*

Teorema 3.23.1 (La varianza si può esprimere in termini di momenti intorno all'origine)**.** *Sia* X *una v.a. discreta definita su* $(\Omega, \mathcal{F}, \mathbb{P})$ *. Si ha che*

$$Var(X) = E[X^2] - E^2[X]$$

Ovvero la varianza di X *è uguale al momento intorno all'origine di ordine due meno il momento intorno all'origine di ordine uno elevato al quadrato.*

Dimostrazione.

$$\begin{aligned} Var(X) &= E\Big[\big[X - E[X]\big]^2\Big] \\ &= E\Big[X^2 - 2XE[X] + E^2[X]\Big] \\ &= E[X^2] - 2E[X]E[X] + E^2[X] \quad \text{(si ricordi che } E[X] = \mu \text{ è un valore costante)} \\ &= E[X^2] - 2E^2[X] + E^2[X] \\ &= E[X^2] - E^2[X] \end{aligned}$$

□

3.23.1 Varianza di una trasformazione lineare

Teorema 3.23.2 (Varianza di una trasformazione lineare)**.** *Si consideri lo spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *su cui è definita la v.a.* X *dotata di media* $E[X] = \mu$. *Si ha che:*

$$Var(aX + b) = a^2 \cdot Var(X)$$

Ovvero l'operatore varianza è un operatore quadratico.

Dimostrazione.

$$\begin{aligned} Var(aX + b) &= E\Big[\big[(aX + b) - E[aX + b]\big]^2\Big] \\ &= E\Big[\big[aX + b - a\mu - b\big]^2\Big] \\ &= E\Big[\big[aX - a\mu\big]^2\Big] \\ &= E\Big[\big[a(X - \mu)\big]^2\Big] \\ &= E\Big[a^2(X - \mu)^2\Big] \\ &= a^2 \cdot E\Big[(X - \mu)^2\Big] \\ &= a^2 \cdot Var(X) \end{aligned}$$

□

Osservazione 3.23.1. *Se* $b = 0$, $Var(aX) = a^2 Var(X)$.
Se $a = 0$, $Var(b) = 0$.

3.24 Covarianza

Definizione 3.24.1 (Covarianza)**.** *Siano* X, Y *v.v.a.a. definite su* $(\Omega, \mathcal{F}, \mathbb{P})$ *e dotate rispettivamente di valori medi* $\mu_X = E[X]$, $\mu_Y = E[Y]$. *La covarianza di* X *e* Y *è definita come*

$$Cov(X, Y) = E\big[(X - \mu_X)(Y - \mu_Y)\big]$$

La covarianza da misura di quanto siano relate le v.v.a.a.:

$$Cov(X, Y) \begin{cases} = 0 \Rightarrow & X, Y \ \textit{scorrelate} \\ > 0 \Rightarrow & X, Y \ \textit{correlate positivamente} \\ < 0 \Rightarrow & X, Y \ \textit{correlate negativamente} \end{cases}$$

Osservazione 3.24.1. *Si ha che*

$$Cov(X,X) = E\big[(X-\mu_X)^2\big] = Var(X)$$

Osservazione 3.24.2 (Formula alternativa)**.** *Partendo dalla definizione e svolgendo i prodotti si ha che*

$$\begin{aligned}
Cov(X,Y) &= E\big[(X-E[X])(Y-E[Y])\big] \\
&= E\big[XY - XE[Y] - YE[X] + E[X]E[Y]\big] \\
&= E[XY] - E\big[XE[Y]\big] - E\big[YE[X]\big] + E\big[E[X]E[Y]\big] \\
&= E[XY] - E[Y]E[X] - E[X]E[Y] + E[X]E[Y] \\
&= E[XY] - E[X]E[Y]
\end{aligned}$$

Teorema 3.24.1 (Proprietà simmetrica per la covarianza)**.** *Siano X, Y v.v.a.a. definite su $(\Omega, \mathcal{F}, \mathbb{P})$ e dotate rispettivamente di valori medi $\mu_X = E[X]$, $\mu_Y = E[Y]$. Si ha che*

$$Cov(X,Y) = Cov(Y,X)$$

Dimostrazione. Applicando la definizione si ha

$$Cov(X,Y) = E[XY] - E[X]E[Y] = E[YX] - E[Y]E[X] = Cov(Y,X)$$

□

Teorema 3.24.2 (Proprietà distributiva per la covarianza)**.** *Siano X, Y, Z v.v.a.a. definite su $(\Omega, \mathcal{F}, \mathbb{P})$ e dotate di valori medi. Si ha che*

$$\begin{aligned}
Cov(X+Y,Z) &= Cov(X,Z) + Cov(Y,Z) \\
Cov(X,Y+Z) &= Cov(X,Y) + Cov(X,Z)
\end{aligned}$$

Dimostrazione. Di seguito esibisco la dimostrazione soltanto per il primo caso, essendo il secondo totalmente analogo.

$$\begin{aligned}
Cov(X+Y,Z) &= E[(X+Y)Z] - E[X+Y]E[Z] \\
&= E[XZ+YZ] - \big(E[X]+E[Y]\big)E[Z] \\
&= E[XZ] + E[YZ] - E[X]E[Z] - E[Y]E[Z] \\
&= E[XZ] - E[X]E[Z] + E[YZ] - E[Y]E[Z] \\
&= Cov(X,Z) + Cov(Y,Z)
\end{aligned}$$

□

Teorema 3.24.3 (X, Y indipendenti $\Rightarrow E(XY) = E(X)E(Y)$)**.** *Siano X, Y v.v.a.a. discrete e indipendenti, allora si ha che*

$$E[XY] = E[X]E[Y]$$

Caso discreto.

$$E[XY] = E[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot p_{X,Y}(x,y) = \sum_x \sum_y xy \cdot p_{X,Y}(x,y)$$

ma poichè abbiamo assunto X, Y indipendenti, abbiamo che $p_{X,Y}(x,y) = p_X(x)p_Y(y)$, quindi

$$\sum_x \sum_y xy \cdot p_{X,Y}(x,y) = \sum_x \sum_y xy \cdot p_X(x)p_Y(y) = \sum_x xp_X(x) \sum_Y yp_Y(y) = E[X]E[Y]$$

□

Osservazione 3.24.3 (Indipendenza implica scorrelazione, ma non è vero il contrario)**.** *Dall'osservazione 3.24.2 e dal teorema precedente, segue immediatamente che*

$$X, Y \text{ indipendenti } \Rightarrow E[XY] = E[X]E[Y] \Rightarrow E[XY] - E[X]E[Y] = 0 \Rightarrow X, Y \text{ scorrelate}$$

Tale implicazione non è generalmente invertibile!

Teorema 3.24.4 (Varianza di una somma di v.v.a.a. indipendenti)**.** *Siano X, Y, Z v.v.a.a. indipendenti definite su $(\Omega, \mathcal{F}, \mathbb{P})$ e dotate di valori medi. Si ha che*

$$Var(X+Y) = Var(X) + Var(Y)$$

Dimostrazione.

$$\begin{aligned}
Var(X+Y) &= Cov(X+Y, X+Y) && \text{(si veda 3.24.1)}\\
&= Cov(X, X+Y) + Cov(Y, X+Y) && \text{(si veda 3.24.2)}\\
&= Cov(X,X) + Cov(X,Y) + Cov(Y,X) + Cov(Y,Y)\\
&= Var(X) + 2Cov(X,Y) + Var(Y)
\end{aligned}$$

Dall'osservazione 3.24.3 segue che, se X, Y indipendenti, allora

$$Var(X+Y) = Var(X) + Var(Y)$$

□

3.25 Disuguaglianza di Markov

Teorema 3.25.1 (Disuguaglianza di Markov)**.** *Sia X una v.a. non negativa ($X \geq 0$) dotata di valor medio $E[X]$ finito. Allora*

$$\forall a > 0,\ \mathbb{P}(X \geq a) \leq \frac{E[X]}{a},\ \mathbb{P}(X < a) \geq 1 - \frac{E[X]}{a}$$

Dimostrazione. (Caso continuo) Dalla definizione di valore atteso si ha che

$$\begin{aligned} E[X] &= \int_{\mathbb{R}} x f_X(x)\ dx \overset{x\geq 0}{=} \int_0^{+\infty} x f_X(x)\ dx = \\ &= \int_0^a x f_X(x)\ dx + \int_a^{+\infty} x f_X(x)\ dx \end{aligned}$$

dal momento che la densità $f_X(x)$ è per definizione non negativa e che x è per ipotesi non negativa, possiamo concludere che i due integrali siano entrambi quantità non negative. inorando con il secondo addendo l'intera abbiamo

$$\geq \int_a^{+\infty} x f_X(x)\ dx = \int_{x \geq a} x f_X(x)\ dx$$

Minorando la x nell'integrale con a

$$\geq a \int_a^{+\infty} f_X(x)\ dx = a\mathbb{P}(X \geq a)$$

$$E[X] \geq a\mathbb{P}(X \geq a) \Leftrightarrow \mathbb{P}(X \geq a) \leq \frac{E[X]}{a}$$

□

3.26 Disuguaglianza di Chebychev

Teorema 3.26.1 (Disuguaglianza di Chebychev)**.** *Sia X una v.a. definita su $(\Omega, \mathcal{F}, \mathbb{P})$ e dotata di valore atteso $E[X]$ e varianza $Var(X)$ finiti. Allora si ha*

$$\forall r > 0,\ \mathbb{P}\Big(\big|X - E[X]\big| \geq r\Big) \leq \frac{Var(X)}{r^2}$$

Dimostrazione. (Sfruttando la disuguaglianza di Markov già dimostrata in 3.25).
Notiamo innanzitutto che $\mathbb{P}\big(\big|X - E[X]\big| \geq r\big) = \mathbb{P}\big(\big(X - E[X]\big)^2 \geq r^2\big)$. Possiamo considerare $\big(X - E[X]\big)^2$ come una nuova v.a. non negativa ed applicarvi la disuguaglianza di Markov:

$$\mathbb{P}\Big(\big(X - E[X]\big)^2 \geq r^2\Big) \leq \frac{E\Big[\big(X - E[X]\big)^2\Big]}{r^2} = \frac{Var(X)}{r^2}$$

□

Dimostrazione. (Senza usare la disuguaglianza di Markov)

$$\begin{aligned} Var(X) &= E\Big[\big(X - E[X]\big)^2\Big] \\ &= \int_{-\infty}^{+\infty} (x - E[X])^2 f_X(x)\ dx \\ &= \int_{0}^{r^2} (x - E[X])^2 f_X(x)\ dx + \int_{r^2}^{+\infty} (x - E[X])^2 f_X(x)\ dx \end{aligned}$$

Notando come le funzioni integrande siano sempre positive, possiamo applicare considerazioni analoghe a quelle fatte durante la dimostrazione di 3.25 e minorare la somma degli integrali con il secondo addendo e procedere come già visto:

$$\begin{aligned} &\geq \int_{r^2}^{+\infty} (x - E[X])^2 f_X(x)\ dx = \int_{x:(x-E[X])^2 \geq r^2} (x - E[X])^2 f_X(x)\ dx \\ &\geq r^2 \int_{x:(x-E[X])^2 \geq r^2} f_X(x)\ dx = r^2 \mathbb{P}\big((x - E[X])^2 \geq r^2\big) \end{aligned}$$

dal risultato appena ottenuto, invertendo la formula, si ricava la disuguaglianza di Chebychev:

$$Var(X) \geq r^2 \mathbb{P}\big((x - E[X])^2 \geq r^2\big) \Leftrightarrow \mathbb{P}\big((x - E[X])^2 \geq r^2\big) \leq \frac{Var(X)}{r^2}$$

□

3.27 Legge debole dei grandi numeri

Definizione 3.27.1 (Convergenza di una successione di v.v.a.a. in probabilità)**.** *Una successione di v.v.a.a.* $\{X_n\}_{n\in\mathbb{N}}$ *converge in probabilità ad una v.a.* X *(in simboli ciò si indica con* $\{X_n\}_{n\in\mathbb{N}} \xrightarrow{\mathbb{P}} X$*) se e solo se*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|X_n - X\right| < \varepsilon\right) = 1$$

Teorema 3.27.1 (Legge debole dei grandi numeri)**.** *Sia* $\{X_n\}_{n\in\mathbb{N}}$ *una successione di variabili aleatorie indipendenti ed identicamente distribuite. Avendo la stessa funzione di distribuzione, le variabili della successione avranno anche uguale valor medio* $E[X_i] = \mu$ *ed uguale varianza* $Var(X_i) = \sigma^2$*. Si ha*

$$\forall\varepsilon > 0,\ \lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

$$\forall\varepsilon > 0,\ \lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - \mu\right| < \varepsilon\right) = 1$$

Dimostrazione. Consideriamo la successione di variabili aleatorie $\{Y_n\}_{n\in\mathbb{N}} : Y_n = \frac{x_1+x_2+\cdots+x_n}{n}$. Verifichiamo che Y_n sia dotata di media e varianza finite.

$$\begin{aligned} E[Y_n] &= E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\ &= \frac{1}{n} E[X_1 + X_2 + \cdots + X_n] \\ &= \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu \end{aligned}$$

$$\begin{aligned} Var(Y_n) &= Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\ &= \frac{1}{n^2} Var(X_1 + X_2 + \cdots + X_n) \end{aligned}$$

dal momento che le X_i sono per ipotesi indipendenti è possibile scrivere (si riveda 3.24.4)

$$= \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i)$$

poichè abbiamo assunto le X_i identicamente distribuite, inoltre, con $Var(X_i) = \sigma^2 \forall i$ si ha

$$= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Essendo Y_n dotata di media e varianza finite, è possibile applicare ad essa la disuguaglianza di Chebychev (3.26).

$$\forall \varepsilon > 0,\ \mathbb{P}\left(\left|Y_n - \mu\right| \geq \varepsilon\right) \leq \frac{Var(Y_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

Infinte, applicando il limite per n che tende ad infinito ad entrambi i membri della disequazione si ha

$$\forall \varepsilon > 0,\ \lim_{n\to+\infty} \mathbb{P}\left(\left|Y_n - \mu\right| \geq \varepsilon\right) \leq \lim_{n\to+\infty} \frac{\sigma^2}{n\varepsilon^2}$$
$$\lim_{n\to+\infty} \mathbb{P}\left(\left|Y_n - \mu\right| \geq \varepsilon\right) \leq 0$$

Ma $\mathbb{P}\left(\left|Y_n - \mu\right| \geq \varepsilon\right)$ è una probabilità e, pertanto, non può assumere valori minori di zero. Ne segue che, necessariamente,

$$\lim_{n\to+\infty} \mathbb{P}\left(\left|Y_n - \mu\right| \geq \varepsilon\right) = 0$$

come volevasi dimostrare. □

Osservazione 3.27.1. *Spesso come valore di ε è conveniente scegliere la deviazione standard ($D(X) = \sqrt{\sigma^2} = \sigma$). Ad esempio:*

$$\varepsilon = \sigma \qquad \mathbb{P}\left(|X - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\sigma^2} = 1$$
$$\varepsilon = 2\sigma \qquad \mathbb{P}\left(|X - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{4\sigma^2} = \frac{1}{4}$$
$$\varepsilon = 3\sigma \qquad \mathbb{P}\left(|X - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{9\sigma^2} = \frac{1}{9}$$

Capitolo 4

Modelli di variabili aleatorie

Di seguito sono presentate alcune classi di v.v.a.a. che compaiono molto frequentemente in natura o nelle applicazioni.

4.1 Variabili aleatorie bernoulliane

Si consideri un esperimento il cui esito può essere "successo" oppure "fallimento". Consideriamo gli eventi E e E^C corrispondenti rispettivamente al successo e al fallimento dell'esperimento. La v.a. X individua il successo dell'esperimento:

$$X(\omega) = 1 \Leftrightarrow \omega \in E \Leftrightarrow \text{"successo"}$$

$$X(\omega) = 0 \Leftrightarrow \omega \in E^C \Leftrightarrow \text{"fallimento"}$$

Indicando con p il valore di $\mathbb{P}(E)$, si ha quindi

$$p_X(x) = \begin{cases} p & \text{se } x = 1 \\ 1 - p & \text{se } x = 0 \end{cases}$$

Definizione 4.1.1 (Variabile aleatoria bernoulliana). *Una variabile aleatoria X si dice bernoulliana (o di Bernoulli) se la sua funzione di massa di probabilità è come quella descritta precedentemente. In simboli si indicherà con*

$$X \sim B(1, p)$$

4.1.1 Valore atteso di v.a. bernoulliana

Essendo nota la massa di probabilità, è possibile ricavare il valore atteso in questo modo:

$$E[X] = 1 \cdot p_X(1) + 0 \cdot p_X(0) = 1 \cdot p + 0 \cdot (1 - p) = p$$

4.1.2 Varianza di una v.a. bernoulliana

Conoscendo il valore atteso, possiamo individuare la varianza:

$$Var(X) = E\big[(X - E[X])^2\big] \overset{teo.3.23.1}{=} E\big[E[X^2] - E^2[X]\big] \qquad (\star)$$

Calcoliamo $E[X^2]$:

$$E[X^2] = 1^2 \cdot p_X(1) + 0^2 \cdot p_X(0) = p$$

Quindi, tornando a $(\star)$:

$$Var(X) = E\big[p - p^2\big] = p(1 - p)$$

4.1.3 Funzione di distribuzione di una v.a. bernoulliana

La funzione di distribuzione di una v.a. bernoulliana avrà la seguente forma "a gradini":

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - p & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$$

4.2 Variabili aleatorie binomiali

Definizione 4.2.1 (Variabile aleatoria binomiale)**.** *Si consideri un esperimento che consiste nella ripetizione di $n \in \mathbb{N}$ prove indipendenti tra loro. Sia p la probabilità di successo di ciascuna prova. La v.a. che* ***conta*** *il numero di successi registrati nelle n prove si dice variabile aleatoria (con distribuzione) binomiale. In simboli si indicherà con*

$$X \sim B(n, p)$$

4.2.1 Funzione di massa di probabilità di una v.a. binomiale

Sia X v.a. binomiale. Sia p la probabilità di successo di ogni prova. Calcoliamo la funzione di massa di probabilità:

$$p_X(x) = \mathbb{P}(X = x) = C_{n,x} p^x (1-p)^{n-x} = \binom{n}{x} p^x (1-p)^{n-x}$$

Nel calcolo della massa di probabilità nell'espressione precedente, $p^x(1-p)^{n-x}$ è la probabilità che si verifichino x successi in modo tale che i primi x tentativi siano tutti successi e i restanti siano tutti fallimenti. Chiaramente questo non è l'unico modo in cui possono verificarsi esattamente x successi su n prove. Per l'esattezza, ci sono $C_{n,x}$ modi di scegliere le x prove che dovranno avere successo.
Verifichiamo la condizione di normalizzazione per $p_X(x)$.

$$\begin{aligned}
\sum_{x=0}^{n} p_X(x) &= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \\
&= (p + (1-p))^n && \text{binomio di Newton!} \\
&= 1^n = 1
\end{aligned}$$

Per chiarificare il secondo passaggio della verifica precedente, si ricordi lo sviluppo binomiale di Newton[1]:

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}$$

Congettura 4.2.1 (Congettura di Auriemma-Mazzoccola)**.** *In merito alla verifica della condizione di normalizzazione della funzione di massa di probabilità di una variabile aleatoria binomiale svolta precedentemente, non basterebbe dire che* 1^n *fa* 1*?*

La congettura è stata immediatamente confutata dalla professoressa, che ha notato come siano comunque necessari i passaggi precedentemente svolti per mostrare che $\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = 1^n$.

[1] Ulteriori informazioni qui: `https://it.wikipedia.org/wiki/Teorema_binomiale`

4.2.2 Valore atteso di una variabile aleatoria binomiale

La variabile aleatoria binomiale X descrive la somma di n bernoulliane indipendenti di parametro p.

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n] = np$$

4.2.3 Varianza di una variabile aleatoria binomiale

Usiamo ancora una volta il fatto che X binomiale sia la somma di n bernoulliane indipendenti.

$$Var(X) = Var(X_1) + Var(X_2) + \cdots + Var(X_n) = np(1-p)$$

Nell'espressione precedente, si ricordi che la varianza di una somma di v.v.a.a. è uguale alla somma delle varianze solo se le v.v.a.a. sommate sono indipendenti! Si veda in particolare 3.24.4.

4.3 Variabili aleatorie con distribuzione geometrica

Definizione 4.3.1 (Variabile aleatoria geometrica)**.** *Si consideri un esperimento di prove ripetute. Ciascuna prova ha probabilità di successo p. La variabile aleatoria che conta il numero di ripetizioni necessarie per ottenere il primo successo si dice avere distribuzione geometrica di parametro p. In simboli si indica con:*

$$X \sim G(p)$$

4.3.1 Funzione di massa di probabilità di v.a. con distr. geometrica

Calcoliamo la funzione di massa di probabilità di $X \sim G(p)$.

$$\begin{aligned} p_X(x) &= \mathbb{P}(X = x) \\ &= (1-p)^{x-1}p && (x-1) \text{ fallimenti e un successo} \end{aligned}$$

Verifichiamo la condizione di normalizzazione:

$$\begin{aligned}
\sum_{x=1}^{\infty} p_X(x) &= \sum_{x=1}^{\infty}(1-p)^{x-1}p \\
&= p\sum_{x=1}^{\infty}(1-p)^{x-1} \\
&= p\sum_{s=0}^{\infty}(1-p)^{s} && \text{(posto } s = x-1\text{)} \\
&= p\left(\frac{1}{1-1+p}\right) && \text{(serie geometrica di ragione } < 1\text{)} \\
&= 1
\end{aligned}$$

4.3.2 Funzione di distribuzione di una v.a. con distr. geometrica

Sia $X \sim G(p)$. Si ha

$$\begin{aligned}
F_X(x) &= \sum_{r=1}^{x} p_X(r) \\
&= \sum_{r=1}^{x}(1-p)^{r-1}p \\
&= p\sum_{r=1}^{x}(1-p)^{r-1} \\
&= p\sum_{s=0}^{x-1}(1-p)^{s} && \text{(pongo } s = r-1\text{)} \\
&= p\frac{1-(1-p)^{x-1+1}}{1-(1-p)} && \text{(serie geometrica di ragione } < 1\text{)} \\
&= 1-(1-p)^x
\end{aligned}$$

4.3.3 V.a. con distr. geometrica gode di proprietà di assenza di memoria

Definizione 4.3.2 (Proprietà di assenza di memoria)**.** *Una variabile aleatoria X discreta (risp. continua) gode della proprietà di assenza di memoria se $\forall r, s \in \mathbb{N}$ (risp. $\forall r, s \in \mathbb{R}^+$) si ha che*

$$\mathbb{P}(X > r + s | X > s) = \mathbb{P}(X > r)$$

Teorema 4.3.1 (V.a. con distr. geometrica gode di prop. assenza mem.)**.** *Dimostrazione.* Si consideri $X \sim G(p)$ e due numeri naturali $r, s \in \mathbb{N}$. Si ha

$$\begin{aligned}
\mathbb{P}(X > r + s | X > s) &= \frac{\mathbb{P}(\{X > r + s\} \cap \{X > s\})}{\mathbb{P}(\{X > s\})} \\
&= \frac{\mathbb{P}(X > r + s)}{\mathbb{P}(X > s)} \\
&= \frac{1 - F_X(r + s)}{1 - F_X(s)} \\
&= \frac{1 - (1 - (1 - p)^{r+s})}{1 - (1 - (1 - p)^s)} \\
&= \frac{(1 - p)^{r+s}}{(1 - p)^s} \\
&= (1 - p)^r \\
&= 1 - F_X(r) = \mathbb{P}(X > r)
\end{aligned}$$

□

4.3.4 Valore atteso di v.a. con distr. geometrica

Sia $X \sim G(p)$.

$$
\begin{aligned}
E[X] &= \sum_{n=1}^{\infty} n p_X(n) \\
&= \sum_{n=1}^{\infty} n(1-p)^{n-1} p \\
&= p \sum_{n=1}^{\infty} n(1-p)^{n-1} \\
&= p \sum_{n=1}^{\infty} n q^{n-1} && \text{sia } q = 1-p \\
&= p \sum_{n=1}^{\infty} \frac{d}{dq} q^n \\
&= p \frac{d}{dq} \sum_{n=1}^{\infty} q^n \\
&= p \frac{d}{dq} \frac{1}{1-q} && \text{(serie geometrica di ragione } < 1) \\
&= p \frac{1}{(1-q)^2} \\
&= p \frac{1}{(1-1+p)^2} && \text{(sostituisco nuovamente } q = 1-p) \\
&= p \frac{1}{p^2} = \frac{1}{p}
\end{aligned}
$$

4.4 Variabili aleatorie con distr. di Poisson

Definizione 4.4.1 (Variabile aleatoria con distr. di Poisson)**.** *Una v.a. X con $S_X = N^0$ si dice avere distribuzione di Poisson di parametro λ se la sua massa di probabilità si può scrivere come*

$$p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

In simboli si indica con

$$X \sim Poisson(\lambda)$$

X conta il numero di eventi che si verificano in un dato intervallo di tempo (o di lunghezza). Verifichiamo che la massa di probabilità descritta soddisfi la condizione di normalizzazione:

$$\begin{aligned}\sum_{x=0}^{\infty} p_X(x) &= \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} e^{\lambda} \qquad \text{(notando che } \textstyle\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \text{ è lo sviluppo in serie di } e^{\lambda}) \\ &= 1\end{aligned}$$

4.4.1 Valore atteso di v.a. poissoniana

Sia $X \sim Poisson(\lambda)$. Allora si ha

$$\begin{aligned}E[X] &= \sum_{x=0}^{\infty} x p_X(x) \\ &= \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \\ &= e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= e^{-\lambda} \lambda e^{\lambda} \\ &= \lambda\end{aligned}$$

λ rappresenta il numero di eventi attesi (nell'intervallo).

4.4.2 Varianza di una v.a. poissoniana

Sia $X \sim Poisson(\lambda)$. Allora si ha

$$\begin{aligned}Var(X) &= E\Big[(X - E[X])^2\Big] \\ &= E\Big[X^2 - 2XE[X] + E^2[X]\Big] \\ &= E[X^2] - 2E^2[X] + E^2[X]\Big] \\ &= E[X^2] - E^2[X]\end{aligned}$$

Calcoliamo quindi $E[X^2]$:

$$
\begin{aligned}
E[X^2] &= \sum_{x=0}^{\infty} x^2 p_X(x) \\
&= \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!} e^{-\lambda} \\
&= e^{-\lambda}\lambda \sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{(x-1)!} \\
&= e^{-\lambda}\lambda \sum_{x=1}^{\infty} (x-1+1) \frac{\lambda^{x-1}}{(x-1)!} \\
&= e^{-\lambda}\lambda \left(\sum_{x=1}^{\infty} (x-1) \frac{\lambda^{x-1}}{(x-1)!} + \sum_{x=1}^{\infty} 1 \frac{\lambda^{x-1}}{(x-1)!} \right) \\
&= e^{-\lambda}\lambda \left(\sum_{x=2}^{\infty} \frac{\lambda^{x-1}}{(x-2)!} + e^{\lambda} \right) \\
&= e^{-\lambda}\lambda \sum_{x=2}^{\infty} \frac{\lambda^{x-1}}{(x-2)!} + e^{-\lambda}\lambda e^{\lambda} \\
&= e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda \\
&= e^{-\lambda}\lambda^2 \sum_{s=0}^{\infty} \frac{\lambda^{s}}{(s)!} + \lambda \qquad \text{(posto } s = x-2\text{)} \\
&= \lambda^2 + \lambda
\end{aligned}
$$

Tornando all'espressione della varianza della poissoniana, quindi, si ha:

$$
Var(X) = E\Big[X^2] - E^2[X]\Big] = \lambda^2 - \lambda - \lambda^2 = \lambda
$$

4.4.3 Funzione di distribuzione di v.a. poissoniane

$$
F_X(r) = \begin{cases} 0 & \text{se } r < 0 \\ \sum_{x=0}^{r} \frac{\lambda^x}{x!} e^{-\lambda} & \text{altrimenti} \end{cases}
$$

4.4.4 Legge degli eventi rari

La distribuzione di Poisson può essere ottenuta come limite di una successione di v.v.a.a. binomiali $B(n, p)$ con $\lambda = np$. Per questa convergenza la distribuzione di Poisson è anche nota come Legge degli eventi rari.

Teorema 4.4.1 (Derivare Poisson da successione di binomiali)**.** *Dimostrazione.*
Sia $Y \sim B(n, p)$. Si ha

$$\begin{aligned}
\lim_{\substack{p\to 0\\ np=\lambda}} p_Y(x) &= \lim_{\substack{p\to 0\\ np=\lambda}} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \lim_{\substack{p\to 0\\ np=\lambda}} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \lim_{\substack{n\to \infty\\ p=\frac{\lambda}{n}}} \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1-\left(\frac{\lambda}{n}\right)\right)^{n-x} \\
&= \frac{\lambda^x}{x!} \lim_{\substack{n\to \infty\\ p=\frac{\lambda}{n}}} \frac{n!}{(n-x)!} \left(\frac{1}{n}\right)^x \left(1-\left(\frac{\lambda}{n}\right)\right)^{n-x} \\
&= \frac{\lambda^x}{x!} \lim_{\substack{n\to \infty\\ p=\frac{\lambda}{n}}} \frac{n \times (n-1) \times \cdots \times (n-x+1)}{n^x} \left(1-\left(\frac{\lambda}{n}\right)\right)^{n-x} \\
&= \frac{\lambda^x}{x!} \lim_{\substack{n\to \infty\\ p=\frac{\lambda}{n}}} \frac{n}{n}\frac{n-1}{n} \cdots \frac{n-x+1}{n} \frac{\left(1-\left(\frac{\lambda}{n}\right)\right)^n}{\left(1-\left(\frac{\lambda}{n}\right)\right)^x} \\
&= \frac{\lambda^x}{x!} \frac{e^{-\lambda}}{1}
\end{aligned}$$

□

Nell'ultimo passaggio della dimostrazione, si osservi che i rapporti $\frac{n}{n}$*,* $\frac{n-1}{n}$*, … ,* $\frac{n-x+1}{n}$ *tendono tutti ad 1, così come* $\lim_{n\to\infty} \left(1-\left(\frac{\lambda}{n}\right)\right)^x$*. Infine, per il numeratore dell'ultima frazione si è usato il limite notevole* $\lim_{n\to\infty} \left(1+\left(\frac{\lambda}{n}\right)\right)^x = e^\lambda$*.*

4.5 Distribuzione uniforme

Definizione 4.5.1 (Variabile aleatoria con distr. uniforme)**.** *La v.a. X assolutamente continua ammette distribuzione uniforme in un certo intervallo $[a, b]$ se la sua densità è costante all'interno dell'intervallo e nulla altrove:*

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

In simboli si indica con

$$X \sim U[a, b]$$

Dalla condizione di normalizzazione è possibile ricavare il valore di k essendo noto l'intervallo.

$$\int_a^b f_X(x)dx = 1$$
$$\int_a^b kdx = 1$$
$$kx\Big|_a^b = 1$$
$$k = \frac{1}{b-a}$$

4.5.1 Valore atteso di una v.a. con distr. uniforme

Sia $X \sim U[a,b]$. Si ha:

$$\begin{aligned}
E[X] &= \int_{\mathbb{R}} x f_X(x)dx \\
&= \int_a^b xkdx \\
&= k\int_a^b xdx \\
&= k\frac{x^2}{2}\Big|_a^b \\
&= k\frac{b^2-a^2}{2} \\
&= \frac{1}{b-a}\frac{b^2-a^2}{2} && \text{(sostituendo il valore di } k\text{)} \\
&= \frac{1}{b-a}\frac{(b-a)(b+a)}{2} \\
&= \frac{b+a}{2}
\end{aligned}$$

4.5.2 Varianza di una v.a. con distr. uniforme

Sia $X \sim U[a,b]$. Si ha:

$$Var(X) = E[(X-E[X])^2] = E[X^2] - E^2[X]$$

Calcoliamo $E[X^2]$:

$$\begin{aligned}
E[X^2] &= \int_{\mathbb{R}} x^2 f_X(x)dx \\
&= \int_a^b x^2 \frac{1}{b-a} dx \\
&= \frac{1}{b-a} \int_a^b x^2 dx \\
&= \frac{1}{b-a} \frac{x^3}{3}\Big|_a^b \\
&= \frac{b^3 - a^3}{3(b-a)} \\
&= \frac{(b-a)(a^2 + b^2 + ab)}{3(b-a)} \\
&= \frac{a^2 + b^2 + ab}{3}
\end{aligned}$$

Quindi, sostituendo nell'espressione della varianza:

$$\begin{aligned}
Var(X) = E[(X - E[X])^2] &= E[X^2] - E^X \\
&= \frac{a^2 + b^2 + ab}{3} - \frac{b^2 + a^2 + 2ab}{4} \\
&= \frac{4a^2 + 4b^2 + 4ab - 3b^2 - 3a^2 - 6ab}{12} \\
&= \frac{a^2 + b^2 - 2ab}{12} \\
&= \frac{(b-a)^2}{12}
\end{aligned}$$

4.6 Variabili aleatorie con distr. esponenziale

Una v.a. assolutamente continua si dice avere distribuzione esponenziale di parametro α se la sua densità ha forma

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} & \text{se } x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

In simboli ciò si esprime con

$$X \sim ESP(\alpha)$$

Verifichiamo che la densità specificata soddisfi la condizione di normalizzazione:

$$\begin{aligned}\int_{\mathbb{R}} f_X(x) &= \int_{\mathbb{R}} \alpha e^{-\alpha x}\\ &= -\int_0^{+\infty} -\alpha e^{-\alpha x}\\ &= -e^{-\alpha x}\Big|_0^{+\infty}\\ &= -(e^{-\infty} - e^0) = 1\end{aligned}$$

Per quanto riguarda i passaggi precedenti, si tenga conto che $\int f'(x)e^{f(x)} = e^{f(x)}$.

4.6.1 Valor medio di una v.a. con distr. esponenziale

Sia $X \sim ESP(\alpha)$. Si ha:

$$\begin{aligned}E[X] &= \int_{\mathbb{R}} x f_X(x)dx\\ &= \int_0^{+\infty} x\alpha e^{-\alpha x}dx\\ &= \left(-e^{-\alpha x}x\Big|_0^{\infty}\right) - \int_0^{+\infty} -e^{-\alpha x}\cdot 1dx\\ &= 0 - \frac{1}{\alpha}\int_0^{+\infty} -\alpha e^{-\alpha x}dx\\ &= -\frac{1}{\alpha}e^{-\alpha x}\Big|_0^{+\infty} = \frac{1}{\alpha}\end{aligned}$$

4.6.2 Funzione di distribuzione di v.a. con distr. esponenziale

Calcoliamo la funzione di distribuzione per $X \sim ESP(\alpha)$.

$$\begin{aligned}F_X(x) &= \int_{-\infty}^{x} f_X(x)dx\\ &= \begin{cases}\int_{-\infty}^{x} 0dx = 0 & \text{se } x < 0\\ \int_0^x \alpha e^{-\alpha x}dx & \text{altrimenti}\end{cases}\\ &= \begin{cases}0 & \text{se } x < 0\\ -\int_0^x -\alpha e^{-\alpha x}dx & \text{altrimenti}\end{cases}\\ &= \begin{cases}0 & \text{se } x < 0\\ -e^{-\alpha x}\Big|_0^x & \text{altrimenti}\end{cases}\\ &= \begin{cases}0 & \text{se } x < 0\\ 1 - e^{-\alpha x} & \text{altrimenti}\end{cases}\end{aligned}$$

4.6.3 Varianza di una v.a. con distr. esponenziale

$$Var(X) = \frac{1}{\alpha^2}$$

4.6.4 V.a. esponenziale gode di assenza di memoria

Teorema 4.6.1 (Assenza di memoria per v.v.a.a. esponenziali). *Sia $X \sim ESP(\alpha)$ e siano $r, s \in \mathbb{R}^+$. Allora si ha che $\mathbb{P}(X > r + s|X > r) = \mathbb{P}(X > s)$, cioè vale la proprietà di assenza di memoria 4.3.2.*

Dimostrazione.

$$\begin{aligned}
\mathbb{P}(X > r + s|X > r) &= \frac{\mathbb{P}(X > r + s; X > r)}{\mathbb{P}(X > r)} \\
&= \frac{\mathbb{P}(X > r + s)}{\mathbb{P}(X > r)} \\
&= \frac{1 - F_X(r + s)}{1 - F_X(r)} \\
&= \frac{1 - 1 + e^{-\alpha(r+s)}}{1 - 1 + e^{-\alpha(r)}} \\
&= e^{-\alpha(r+s)+\alpha(r)} \\
&= e^{-\alpha r-\alpha s+\alpha r} \\
&= e^{-\alpha s} = 1 - F_X(s) = \mathbb{P}(X > s)
\end{aligned}$$

□

4.7 Variabili aleatorie con distribuzione normale

Definizione 4.7.1 (Variabili aleatorie normali). *Una v.a. X si dice normale (o gaussiana) si parametri μ e σ^2 se ha funzione di densità data da*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \forall x \in \mathbb{R}$$

In simboli si scrive

$$X \sim N(\mu, \sigma^2)$$

4.7.1 Funzione di distribuzione di v.a. normali

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

Questa funzione non è integrabile!

4.7.2 Operazione di standardizzazione

Si consideri una v.a. X di cui sono noti $F_X(x)$, $E[X]$, $Var(X)$. Allora la v.a. $Y = \frac{X-E[X]}{\sqrt{(Var(X))}} = \frac{X-E[X]}{D[X]}$ è una v.a. con la stessa distribuzione di X ma con media nulla e varianza unitaria. Infatti

$$E[Y] = \frac{1}{D[X]}(E[X - E[X]]) = \frac{1}{D[X]}(E[X] - E[E[X]]) = \frac{1}{D[X]}(E[X] - E[X]) = 0$$

$$Var(Y) = \frac{1}{D^2[X]}Var(X - E[X]) = \frac{1}{D^2[X]}Var(X) = 1$$

4.7.3 Funzione di distribuzione di v.a. normali standard

Se Z è normale standard, si ha

$$F_Z(z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$

Una tabella di valori di questa funzione è riportata in A.1.

Esercizio 4.7.1 (Utilizzare la tabella della distr. normale standard)**.** *Sia* $X \sim N(3, 2)$. *Calcolare* $\mathbb{P}(X > 4)$.

Per potere utilizzare la tabella dei valori in A.1 occorre rendere standard la v.a. X. Consideriamo quindi $Y = \frac{X-3}{\sqrt{2}}$, ottenuta da X eseguendo la standardizzazione. Si ha che $\mathbb{P}(X > 4) = \mathbb{P}(Y > \frac{4-3}{\sqrt{2}})$. Quindi è possibile utilizzare la tabella per calcolare

$$\mathbb{P}(Y > \frac{1}{\sqrt{2}}) = 1 - \mathbb{P}(Y \leq \frac{1}{\sqrt{2}}) = 1 - \mathbb{P}(Y \leq 0.71) = 1 - \Phi(0.71) = 1 - 0.7611 = 0.2389$$

4.7.4 Quantili normali di ordine α

Definizione 4.7.2 (Quantile normale di ordine α)**.** *Sia* $Z \sim N(0, 1)$. *Per ogni* $\alpha \in (0, 1)$ *esiste* z_α *tale che* $\mathbb{P}(Z > z_\alpha) = \alpha$. *Tale* z_α *è detto quantile normale di ordine* α.

Esempio: Sia $\alpha = 0.5$. $\mathbb{P}(Z > z_{0.5}) = 0.5 \Leftrightarrow 1-\Phi(z_{0.5}) = 0.5 \Leftrightarrow \Phi(z_{0.5}) = 0.5 \Leftrightarrow z_{0.5} = 0$.

4.7.5 Legge del 3σ

Sia X v.a. con media $E[X]$ e varianza $Var(X) = \sigma^2$. La disuguaglianza di Chebychev (3.26) ci dice che

$$\forall r > 0,\ \mathbb{P}(|X - E[X]| \geq r) \leq \frac{Var(X)}{r^2}$$

Se si sceglie $r = 3\sigma$ si ha:

$$\forall r > 0,\ \mathbb{P}(|X - E[X]| \geq 3\sigma) \leq \frac{\sigma^2}{9\sigma^2} = \frac{1}{9}$$

Da cui

$$\mathbb{P}(|X - E[X]| < 3\sigma) = 1 - \frac{1}{9} = \frac{8}{9}$$

Nel caso in cui X abbia distibuzione normale standard, si ha che

$$\mathbb{P}(|X| < 3) = \mathbb{P}(-3 < X < 3) = \Phi(3) - \Phi(-3) = \Phi(3) - (1 - \Phi(3)) = 2\Phi(3) - 1 = 0.9974 \geq \frac{8}{9}$$

4.7.6 Teorema del limite centrale

Teorema 4.7.1 (Teorema del limite centrale)**.** *Si consideri la successione di v.v.a.a. $\{X_n\}_{n\in\mathbb{N}}$ definite su $(\Omega, \mathcal{F}, \mathbb{P})$, indipendenti ed identicamente distribuite. Tali v.v.a.a. sono dotate di $E[X_i] = \mu$ e $Var(X_i) = \sigma^2$ finite. Allora si ha, indipendentemente dalla distribuzione originale delle X_i, che*

$$\forall x \in \mathbb{R}, \lim_{n\to\infty} \mathbb{P}\left(\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x)$$

Considerando $S_n = X_1 + X_2 + \dots + X_n$, si ha che $E[S_n] = n\mu$ e che $Var(S_n) = n\sigma^2$, quindi $D[X] = \sqrt{n}\sigma$. Si può dunque riscrivere l'espressione precedente come:

$$\forall x \in \mathbb{R}, \lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - E[S_n]}{D[S_n]} \leq x\right) = \Phi(x)$$

Riconoscendo una standardizzazione della successione delle somme parziali.
Analogamente, si può considerare la v.a. media campionaria $\overline{X_n} = \frac{S_n}{n}$:

$$\forall x \in \mathbb{R}, \lim_{n\to\infty} \mathbb{P}\left(\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) = \Phi(x)$$

Notiamo che $E[\overline{X_n}] = \mu$, $Var(\overline{X_n}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$, $D[\overline{X}] = \frac{\sigma}{\sqrt{n}}$.

Il teorema permette, in presenza di un numero sufficientemente grande di v.a. indipendenti ed identicamente distribuite, di considerare normali standard le successioni normalizzate delle somme parziali o delle medie campionarie.

4.8 Variabili aleatorie con distr. chi-quadro

Definizione 4.8.1 (Distribuzione chi-quadro)**.** *La distribuzione χ^2_k (Chi-quadro) descrive la somma di quadrati di k v.v.a.a. indipendenti aventi distribuzione normale standard. L'ordine k (detto anche grado di libertà) indica il numero di v.v.a.a. sommate. In simboli, χ^2_k descrive la v.a.*

$$X^2 = \sum_{i=1}^{k} x_i^2$$

4.8.1 Funzione di distribuzione di v.a. con distr. chi-quadro

Consideriamo $Z \sim N(0,1)$. $Y = Z^2$ sarà tale che $Y \sim \chi_1^2$.

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(Z^2 \leq y) \\ &= \mathbb{P}(-\sqrt{y} \leq z \leq \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \\ &= \Phi(\sqrt{y}) - (1 - \Phi(\sqrt{y})) \\ &= 2\Phi(\sqrt{y}) - 1 \end{aligned}$$

4.8.2 Varianza e valore atteso di v.a. con distr. chi-quadro

Se $X \sim \chi_\nu^2$, $E[X] = \nu$, $Var(X) = 2\nu$.

Capitolo 5
Statistica inferenziale

La statistica inferenziale è il processo mediante cui si deducono le caratteristiche di un'intera popolazione da quelle rilevate per un particolare sottogruppo (campione) scelto il maniera casuale.
Solitamente in statistica inferenziale si ha a che fare con un **campione casuale** (d'ora in avanti semplicemente c.c.) di taglia n, cioè una n-pla $(X_1, X_2, \cdots, X_n)$ di v.v.a.a. i.i.d. e osservabili. Si avrà $X_i \sim X$, $\forall i$ con X detta **genitrice** del campione. Il campione casuale ha distribuzione
$F = F_{X_1} = F_{X_2} = \cdots = F_{X_n} = F_X(x)$.

Spesso la distribuzione del campione non è completamente nota. In caso di distribuzione nota a meno di uno o più parametri si usano tecniche di **stima parametrica**. La stima parametrica può essere **puntuale**, se produce un singolo valore di stima, oppure **intervallare**, se produce un intervallo di possibili valori. Quando la distribuzione non è nota si usano techiche di **stima non parametrica**.

5.1 Statistiche

Definizione 5.1.1 (Statistica)**.** *Una statistica è una funzione delle v.v.a.a. che costituiscono il campione casuale.*

$$T = t(X_1, X_2, \cdots, X_n)$$

5.1.1 Statistica: media campionaria

La media campionaria $\overline{X}$ è una statistica definita come

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

5.1.2 Statistica: varianza campionaria

La varianza campionaria S^2 è una statistica definita come

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

5.2 Stimatori

Definizione 5.2.1 (Stimatore)**.** *Sia* $(X_1, X_2, \cdots, X_n)$ *un c.c. di taglia* n*. Sia la distribuzione del c.c. nota a meno di un certo parametro* ϑ*. Uno stimatore per* ϑ *è una statistica i cui valori possono essere usati per stimare* ϑ*.*

Ogni n-pla di valori rilevati $(x_1, x_2, \cdots, x_n)$ è una **realizzazione** del c.c. $(X_1, X_2, \cdots, X_n)$. Se $T = t(X_1, X_2, \cdots, X_n)$ è lo stimatore per il parametro ϑ, allora $t(x_1, x_2, \cdots, x_n)$ è il valore stimato per il parametro in corrispondenza della specifica realizzazione.

5.2.1 Correttezza di stimatori

Definizione 5.2.2 (Correttezza di stimatori)**.** *Si consideri il campione casuale* $(X_1, X_2, \cdots, X_n)$ *la cui distribuzione è nota a meno di un* ϑ *incognito, che assume valori in* $\Theta = S_\vartheta$*. Sia* $\psi(\vartheta)$ *una qualsiasi funzione di* ϑ*. Lo stimatore* T *si dice* ***corrretto*** *per* $\psi(\vartheta)$ *se* $E[T] = \psi(\vartheta),\ \forall\vartheta \in \Theta$*.*

5.2.2 Stimatore per la media di una popolazione

Esercizio 5.2.1 (Media campionaria stimatore corretto per la media)**.** *Si consideri il c.c.* $(X_1, X_2, \cdots, X_n)$ *la cui distribuzione è nota a meno della media* $E[X] = \mu$*. Verificare che la media campionaria* $\overline{X}$ *è uno stimatore corretto per* μ*.*

$$E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}n\mu = \mu \quad \square$$

5.2.3 Stimatore per la varianza una popolazione

Esercizio 5.2.2 (Stimatore corretto per la varianza). *Si consideri il c.c.* $(X_1, X_2, \cdots, X_n)$ *la cui distribuzione è nota a meno della varianza* $Var[X] = \sigma^2$. *Verificare se*

$$T = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

è uno stimatore corretto per σ^2.

$$\begin{aligned}
E[T] &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + \overline{X}^2)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}(E[X_i^2] - 2E[X_i]\overline{X} + \overline{X}^2) \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}(E[X_i^2]) - 2E[\overline{X}\sum_{i=1}^{n}(X_i)] + \sum_{i=1}^{n}E[\overline{X}^2]\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}(E[X_i^2]) - 2E[n\overline{X}^2] + \sum_{i=1}^{n}E[\overline{X}^2]\right] && (\star) \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}(E[X_i^2]) - 2nE[\overline{X}^2] + nE[\overline{X}^2]\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}(E[X_i^2]) - nE[\overline{X}^2]\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right] && (\star\star) \\
&= \frac{1}{n}\left[n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2\right] \\
&= \frac{1}{n}\left[(n-1)\sigma^2\right] \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}$$

Dunque T non è uno stimatore corretto per σ^2. Per quanto riguarda il passaggio $(\star)$, si noti che $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \Leftrightarrow \sum_{i=1}^{n} X_i = n\overline{X}$. Per il passaggio $(\star\star)$ si è invertita la formula $Var(X) = E[X^2] - E^2[X]$. Per la sostituzione di $E[\overline{X^2}]$ si è utilizzata la stessa formula ma per la v.a. varianza campionaria (è necessario calcolare $Var(\overline{X})$). Lo svolgimento dei (semplici) calcoli è lasciato come esercizio al volenteroso lettore.

Da quanto ottenuto nello svolgimento dell'esercizio è possibile individuare uno stimatore corretto per la varianza. Infatti si ha:

$$
\begin{aligned}
E[T] = \frac{n-1}{n}\sigma^2 &\Leftrightarrow \frac{n}{n-1}E[T] = \sigma^2 \\
&\Leftrightarrow E\left[\frac{n}{n-1}T\right] = \sigma^2 \\
&\Leftrightarrow E\left[\frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = \sigma^2 \\
&\Leftrightarrow E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = \sigma^2
\end{aligned}
$$

Dunque $T_C = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ è uno stimatore corretto per la varianza.

5.3 Classificazione degli stimatori

5.3.1 Rischio quadratico

Definizione 5.3.1 (Rischio quadratico di uno stimatore)**.** *Si consideri il c.c.* $(X_1, X_2, \cdots, X_n)$ *la cui distribuzione è nota a meno di un parametro* ϑ*. Sia* T *uno stimatore per* $\psi(\vartheta)$*. Si definisce il rischio quadratico dello stimatore* T *come*

$$R_T(\vartheta) = E\Big[(T - \psi(\vartheta))^2\Big], \quad \forall \vartheta \in \Theta$$

Osservazione 5.3.1 (Caso particolare: rischio quadratico di stimatore corretto)**.** *Nel caso particolare in cui* T *è corretto, si ha che*

$$R_T(\vartheta) = E\Big[(T - \psi(\vartheta))^2\Big] = E\Big[(T - E[T])^2\Big] = Var(T)$$

Definizione 5.3.2 (Stimatore preferibile)**.** *Siano* S, T *due stimatori per* $\psi(\vartheta)$*.* S *è preferibile a* T *se e solo se* $R_S(\vartheta) \leq R_T(\vartheta)$*,* $\forall \vartheta \in \Theta$*.*

Definizione 5.3.3 (Stimatore strettamente preferibile)**.** *Siano* S, T *due stimatori per* $\psi(\vartheta)$*.* S *è strettamente preferibile a* T *se e solo se* S *è preferibile a* T *ed* $\exists \vartheta_0 : R_S(\vartheta_0) < R_T(\vartheta_0)$*.*

Definizione 5.3.4 (Stimatore ammissibile)**.** *Uno stimatore* S *si dice ammissibile se non esiste alcuno stimatore ad esso strettamente preferibile.*

5.4 Correttezza asintotica

Lo stimatore, in linea di massima, dipende dalla taglia del c.c. in uso. Sia T_n uno stimatore per $\psi(\vartheta)$ con dimensione del c.c. uguale ad n.

Definizione 5.4.1 (Correttezza asintotica)**.** *Lo stimatore T_n di $\psi(\vartheta)$ è asintoticamente corretto se*

$$\lim_{n\to\infty} E[T_n] = \psi(\vartheta), \quad \forall\vartheta \in \Theta$$

Esercizio 5.4.1 (Varianza campionaria asintoticamente corretta)**.** *In 5.2.2 si è mostrato che lo stimatore per la varianza*

$$T = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

non è corretto. Verificare che lo stimatore T è asintoticamente corretto.

$$\lim_{n\to\infty} E[T] = \lim_{n\to\infty} \frac{n}{n-1}\sigma^2 = \sigma^2 \quad \square$$

5.5 Consistenza (Coherence)

Definizione 5.5.1 (Consistenza in senso debole)**.** *Uno stimatore T_n si dice consistente in senso debole (debolmente consistente) per $\psi(\vartheta)$ se converge in probabilità a $\psi(\vartheta)$.*

$$T_n \xrightarrow{\mathbb{P}} \psi(\vartheta) \Leftrightarrow \forall\varepsilon > 0,\ \lim_{n\to\infty} \mathbb{P}(|T_n - \psi(\vartheta)| < \varepsilon) = 1$$

Definizione 5.5.2 (Consistenza in senso forte)**.** *Uno stimatore T_n si dice consistente in senso forte (fortemente consistente) per $\psi(\vartheta)$ se T_n converge a $\psi(\vartheta)$ quasi certamente.*

$$\mathbb{P}\left(\lim_{n\to\infty} T_n = \psi(\vartheta)\right) = 1$$

5.5.1 Condizione sufficiente per la consistenza debole

Si ha che

$$T_n \text{ è debolmente consistente} \Leftarrow \begin{cases} T_n \text{ è asintoticamente corretto} \\ \lim_{n\to\infty} Var(T_n) = 0 \end{cases}$$

Esercizio 5.5.1 (Stimatore media campionaria debolmente consistente)**.** *Si consideri il c.c. $(X_1, X_2, \cdots, X_n)$. Siano X la genitrice della popolazione e $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Abbiamo già dimostrato in 5.2.1 che $\overline{X}$ è uno stimatore corretto per $E[X] = \mu$. Verificare che $\overline{X}$ è debolmente consistente.*

Utilizziamo la condizione sufficiente. Come già detto, $\overline{X}$ è corretto. Verifichiamo che $\lim_{n\to\infty} Var(\overline{X}) = 0$.

$$\begin{aligned}
\lim_{n\to\infty} Var(\overline{X}) &= \lim_{n\to\infty} Var(\frac{1}{n}\sum_{x=1}^{n} X_i) \\
&= \lim_{n\to\infty} \frac{1}{n^2} Var(\sum_{x=1}^{n} X_i) \\
&= \lim_{n\to\infty} \frac{1}{n^2} \sum_{x=1}^{n} Var(X_i) \\
&= \lim_{n\to\infty} \frac{1}{n^2} n\sigma^2 \\
&= \lim_{n\to\infty} \frac{\sigma^2}{n} = 0
\end{aligned}$$

Possiamo quindi concludere che per 5.5.1 $\overline{X}$ è debolmente consistente.

Osservazione 5.5.1. *Si ha che*

$$\begin{aligned}
\overline{X} \textit{ debolm. consist.} &\Leftrightarrow \forall \varepsilon > 0, \lim_{n\to\infty} \mathbb{P}(|\overline{X} - \mu| < \varepsilon) = 1 \\
&\Leftrightarrow \forall \varepsilon > 0, \lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| < \varepsilon \right) = 1
\end{aligned}$$

Si confronti quanto osservato con la legge debole dei grandi numeri! 3.27.1

5.6 Legge forte dei grandi numeri

Teorema 5.6.1 (Legge forte dei grandi numeri)**.** *Si consideri la successione di variabili aleatorie $\{X_n\}_{n\in\mathbb{N}}$ i.i.d. Sia $\mu = E[X_i]$, $\forall i \in \mathbb{N}$. Sia $S_n = \sum_{i=1}^{n} X_i$ la successione delle somme parziali. Si ha che*

$$\frac{S_n}{n} \xrightarrow{q.c.} \mu$$

ovvero che $\frac{S_n}{n}$ converge quasi certamente a μ. In altre parole ancora, ripendendo la definizione di convergenza quasi certa, si ha che

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu \right) = 1$$

5.7 Costruzione di stimatori

5.7.1 Metodo dei momenti

Definizione 5.7.1 (Momento teorico di ordine r intorno all'origine)**.** *Si definisce momento teorico di ordine r intorno all'origine di una v.a. X*

$$\mu_r = E[X^r] = \begin{cases} \sum_x x^r p_X(x) & \textit{se } X \textit{ discreta} \\ \int_{\mathbb{R}} x^r f_X(x)dx & \textit{se } X \textit{ assolutamente continua} \end{cases}$$

Definizione 5.7.2 (Momento empirico di ordine r intorno all'origine)**.** *Si definisce momento empirico di ordine r intorno all'origine di una v.a. X su un c.c. $(X_1, X_2, \cdots, X_n)$ di ordine n*

$$\hat{\mu}_r = \overline{X}^{(r)} = \frac{1}{n}\sum_{x=1}^{r} X_i^r$$

Osservazione 5.7.1. *Il momento empirico di ordine* 1 *coincide con la media campionaria.*

Teorema 5.7.1 (I momenti empirici sono stimatori debolmente consistenti per i momenti teorici corrispondenti)**.**

Dimostrazione. Per la legge debole dei grandi numeri (3.27.1) si ha che

$$\forall \varepsilon > 0, \lim_{n\to\infty} \mathbb{P}(|\hat{\mu}_r - \mu_r| < \varepsilon) = 1 \Leftrightarrow \hat{\mu}_r \xrightarrow{\mathbb{P}} \mu_r$$

Si riconosca in $\hat{\mu}_r$ la successione delle medie campionarie e in μ_r il valor medio che compaiono nell'enunciato della legge debole dei grandi numeri. $\square$

Il metodo dei momenti sfrutta il risultato dimostrato in 5.7.1 per ottenere stimatori dei parametri incogniti.
Sia $(X_1, X_2, \cdots, X_n)$ un c.c. con genitrice X. La distribuzione di X è nota a meno dei parametri $\vartheta_1, \vartheta_2, \cdots, \vartheta_q$. Supponiamo che X ammetta q momenti intorno all'origine. Si mettono a sistema le espressioni di dati momenti in funzione dei parametri incogniti:

$$\begin{cases} \mu_1 & = g_1(\vartheta_1, \vartheta_2, \cdots, \vartheta_q) \\ \mu_2 & = g_2(\vartheta_1, \vartheta_2, \cdots, \vartheta_q) \\ & \vdots \\ \mu_q & = g_q(\vartheta_1, \vartheta_2, \cdots, \vartheta_q) \end{cases}$$

Si noti che se X ammette un numero di momenti intorno all'origine minore del numero dei parametri incogniti il sistema rimane indeterminato. Se il numero di momenti di X è non inferiore al numero di parametri incogniti allora è possibile procedere.

Si ricavano espressioni per i parametri incogniti in funzione dei momenti:

$$\begin{cases} \vartheta_1 & = d_1(\mu_1, \mu_2, \cdots, \mu_q) \\ \vartheta_2 & = d_2(\mu_1, \mu_2, \cdots, \mu_q) \\ & \vdots \\ \vartheta_q & = d_q(\mu_1, \mu_2, \cdots, \mu_q) \end{cases}$$

A questo punto è possibile sostituire i momenti empirici ai momenti teorici e individuare gli stimatori per i vari parametri:

$$\begin{cases} \hat{\Theta}_1 & = d_1(\hat{\mu_1}, \hat{\mu_2}, \cdots, \hat{\mu_q}) \\ \hat{\Theta}_2 & = d_2(\hat{\mu_1}, \hat{\mu_2}, \cdots, \hat{\mu_q}) \\ & \vdots \\ \hat{\Theta}_q & = d_q(\hat{\mu_1}, \hat{\mu_2}, \cdots, \hat{\mu_q}) \end{cases}$$

Esercizio 5.7.1 (Metodo dei momenti)**.** *Si consideri il campione casuale* $(X_1, X_2, \cdots, X_n)$ *con distribuzione poissoniana nota a meno del parametro* λ *(*$X_i \sim X \sim Poisson(\lambda)$*). Individuare uno stimatore per* λ *usando il metodo dei momenti.*

Calcoliamo il momento teorico del prim'ordine:

$$\mu_1 = E[X^1] = \lambda \qquad \text{(si veda 4.4 per dimostrazione)}$$

Per definizione, il momento empirico di prim'ordine è

$$\hat{\mu_1} = \frac{1}{n}\sum_{i=1}^{n} x_i^1$$

Sostituiamo il momento empirico al momento teorico

$$\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i^1$$

λ così ottenuto stima il valore del parametro in corrispondenza della realizzazione corrente del campione casuale. É possibile individuare uno stimatore per il parametro tornando alle v.v.a.a. del c.c. come segue:

$$\hat{\Lambda} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

5.8 Metodo della massima verosimiglianza

Si consideri il c.c. $X_1, X_2, \cdots, X_n$ di genitrice X con funzione di massa di probabilità/densità nota a meno di un parametro (o vettore di parametri) ϑ. È possibile costruire la **funzione di verosimiglianza** che rappresenta la probabilità, in funzione di ϑ, che si osservi una particolare realizzazione.

Definizione 5.8.1 (Funzione di verosimiglianza)**.** *Sia $(X_1, X_2, \cdots, X_n)$ c.c. discreto (il caso assolutamente continuo è completamente analogo) con distribuzione nota a meno del parametro (o vettore di parametri) ϑ. É possibile individuare la funzione di massa di probabilità congiunta*

$$\begin{aligned} p_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n; \vartheta) &= \mathbb{P}^{\vartheta}(X_1 = x_1; X_2 = x_2; \cdots; X_n = x_n) \\ &= \prod_{i=1}^{n} p^{\vartheta}_{X_i}(x_i, \vartheta) && \text{(per indip. } X_i\ \forall i\text{)} \\ &= \prod_{i=1}^{n} p^{\vartheta}_{X}(x_i, \vartheta) && (X \text{ genitrice}) \end{aligned}$$

Si definisce quindi la funzione di verosimiglianza (likelyhood) come

$$L(x_1, x_2, \cdots, x_n; \vartheta) = \prod_{i=1}^{n} p^{\vartheta}_{X}(x_i, \vartheta)$$

Teorema 5.8.1 (Postulato zero della statistica)**.** *In assenza di specifiche indicazioni, tra tutte le possibili stime dei parametri incogniti va effettuata la scelta di quella che rende massima la probabilità di occorrenza (o la densità di probabilità, nei casi continui) della realizzazione osservata.*

In corrispondenza di una particolare realizzazione $(x_1, x_2, \cdots, x_n)$ del c.c. si ha che la funzione di verosimiglianza è funzione soltanto di ϑ. Lo **stimatore di massima verosimiglianza** si ottiene scegliendo il ϑ che rende massima la probabilità che si verifichi la realizzazione osservata, ovvero quello che massimizza la funzione di verosimiglianza.

Esercizio 5.8.1 (Stima di massima verosimiglianza su pop. binomiale). *Si consideri il c.c.* $(X_1, X_2, \cdots, X_n)$ *con* $X_i \sim X \sim B(m,p)$ *con* p *parametro incognito. Si individui lo stimatore di massima verosimiglianza per il parametro* p *in corrispondenza della realizzazione del c.c.* $(x_1, x_2, \cdots, x_n)$.

Passo 1: individuazione della funzione di verosimiglianza

$$\begin{aligned}
L(p, x_1, x_2, \cdots, x_n) &= \prod_{i=1}^{n} p_X(x_i; p) \\
&= \prod_{i=1}^{n} \left(\binom{m}{x_i} p^{x_i} (1-p)^{m-x_i} \right) \\
&= \prod_{i=1}^{n} \binom{m}{x_i} \prod_{i=1}^{n} p^{x_i} \prod_{i=1}^{n} (1-p)^{m-x_i} \\
&= \left(\prod_{i=1}^{n} \binom{m}{x_i} \right) p^{\sum_{i=1}^{n} x_i} (p-1)^{\sum_{i=1}^{n}(m-x_i)} \\
&= \left(\prod_{i=1}^{n} \binom{m}{x_i} \right) p^{\sum_{i=1}^{n} x_i} (p-1)^{nm-\sum_{i=1}^{n}(x_i)}
\end{aligned}$$

Abbiamo quindi individuato la funzione di verosimiglianza. Occorre trovare il valore di p che massimizza L. Per comodità, riscriviamo la funzione di verosimiglianza ponendo il valore costante $\prod_{i=1}^{n} \binom{m}{x_i} = A$

$$L(p, x_1, x_2, \cdots, x_n) = Ap^{\sum_{i=1}^{n} x_i} (p-1)^{nm-\sum_{i=1}^{n}(x_i)}$$

Passo 2: applicare il logaritmo naturale alla funzione

Dal momento che il logaritmo naturale è una trasformazione monotòna crescente, la sua applicazione non muta le caratteristiche di crescenza/decrescenza della funzione che ci interessano per individuare il massimo. In compenso, applicare il logaritmo naturale solitamente permette di ottenere una funzione più "semplice" da trattare analiticamente (i prodotti diventano somme). Dunque si avrà

$$\begin{aligned}
\ln(L(x_1, x_2, \cdots, x_n)) &= \ln\left(Ap^{\sum_{i=1}^{n} x_i} (p-1)^{nm-\sum_{i=1}^{n}(x_i)} \right) \\
&= \ln(A) + \ln\left(p^{\sum_{i=1}^{n} x_i} \right) + \ln\left((p-1)^{nm-\sum_{i=1}^{n}(x_i)} \right) \\
&= \ln(A) + \left(\sum_{i=1}^{n} x_i \right) \ln(p) + \left(nm - \sum_{i=1}^{n} (x_i) \right) \ln(1-p)
\end{aligned}$$

Passo 3: derivare la funzione log-verosimiglianza

Proseguiamo nel processo per trovare il massimo calcolando la derivata della funzione ottenuta al passo precedente applicando il logaritmo:

$$\begin{aligned}
&\frac{d}{dp}\left(\ln(A)+\left(\sum_{i=1}^{n}x_i\right)\ln(p)+\left(nm-\sum_{i=1}^{n}(x_i)\right)\ln(1-p)\right)=\\
=&\frac{d}{dp}A+\frac{d}{dp}\left(\left(\sum_{i=1}^{n}x_i\right)\ln(p)\right)+\frac{d}{dp}\left(\left(nm-\sum_{i=1}^{n}(x_i)\right)\ln(1-p)\right)=\\
=&0+\left(\sum_{i=1}^{n}x_i\right)\frac{1}{p}+\left(nm-\sum_{i=1}^{n}(x_i)\right)\frac{-1}{p-1}=\\
=&\frac{\sum_{i=1}^{n}x_i}{p}-\frac{nm-\sum_{i=1}^{n}(x_i)}{p-1}
\end{aligned}$$

Passo 4: individuare (se esiste) il valore $\hat{p}$ che annulla la derivata

$$\begin{aligned}
\frac{\sum_{i=1}^{n}x_i}{\hat{p}}-\frac{nm-\sum_{i=1}^{n}(x_i)}{\hat{p}-1}&=0\\
\frac{(1-\hat{p})\sum_{i=1}^{n}x_i-(nm-\sum_{i=1}^{n}(x_i))\hat{p}}{\hat{p}(1-\hat{p})}&=0\\
(1-\hat{p})\sum_{i=1}^{n}x_i-\left(nm-\sum_{i=1}^{n}(x_i)\right)\hat{p}&=0\\
\sum_{i=1}^{n}(x_i)-\hat{p}\sum_{i=1}^{n}(x_i)-nm\hat{p}+\hat{p}\sum_{i=1}^{n}(x_i)&=0\\
\sum_{i=1}^{n}(x_i)-nm\hat{p}&=0\\
\hat{p}&=\frac{\sum_{i=1}^{n}(x_i)}{nm}
\end{aligned}$$

Passo 5: studiare il segno della derivata seconda Dal momento che la derivata prima non si annulla sono nel punto massimo ma anche in eventuali minimi/flessi è necessario verificare anche che la derivata seconda assuma valori negativi in $\hat{p}$

$$\frac{d^2}{dp^2}\ln\left(L\hat{p},x_1,x_2,\cdots,x_n)\right)<0$$

Abbiamo quindi individuato la **stima di massima verosimiglianza** $\hat{p}=\frac{\sum_{i=1}^{n}x_i}{nm}$ per il parametro p.

Come ottenere uno stimatore? Per ottenere uno stimatore è necessario sostituire in $\hat{p}$ i valori osservati in una realizzazione del campione casuale con le variabili aleatorie corrispondenti. Si ha dunque lo stimatore

$$\hat{P}=\frac{\sum_{i=1}^{n}X_i}{mn}=\frac{\overline{X}}{m}$$

Esercizio 5.8.2 (Stimatore di massima verosimiglianza per v.a. esponenziale)**.** *Si consideri il c.c.* $(X_1, X_2, \cdots, X_n)$ *tale che* $X_i \sim X \sim ESP(\lambda)$ *con* λ *parametro incognito. Costuire lo stimatore di massima verosimiglianza per* λ*.*

Ricordiamo che la densità di una v.a. $X \sim ESP(\lambda)$ è

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{, se } x \geq 0 \\ 0 & \text{, altrimenti} \end{cases}$$

Passo 1: funzione di verosimiglianza

$$\begin{aligned} L(\lambda, x_1, x_2, \cdots, x_n) &= f_{X_1,X_2,\cdots,X_n}(\lambda, x_1, x_2, \cdots, x_n) \\ &= \prod_{i=1}^{n} f_{X_i}(\lambda, x_i) && \text{(per indip.)} \\ &= \prod_{i=1}^{n} f_X(\lambda, x_i) && (X \text{ genitrice}) \end{aligned}$$

Assumiamo $x_i \geq 0 \; \forall i$ per non ricadere nel caso banale.

$$\begin{aligned} &= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \\ &= \lambda^n e^{\sum_{i=1}^{n}(-\lambda x_i)} \\ &= \lambda^n e^{-\lambda \sum_{i=1}^{n}(x_i)} \end{aligned}$$

Passo 2: passaggio alla funzione log-verosimiglianza

$$\begin{aligned} \ln(L(\lambda, x_1, x_2, \cdots, x_n)) &= \ln\left(\lambda^n e^{-\lambda \sum_{i=1}^{n}(x_i)}\right) \\ &= \ln(\lambda^n) + \ln\left(e^{-\lambda \sum_{i=1}^{n}(x_i)}\right) \\ &= n \ln(\lambda) - \lambda \sum_{i=1}^{n}(x_i) \end{aligned}$$

Passo 3: passaggio alla derivata

$$\begin{aligned}\frac{d}{d\lambda}\ln(L(\lambda, x_1, x_2, \cdots, x_n)) &= \frac{d}{d\lambda}\left(n\ln(\lambda) - \lambda\sum_{i=1}^{n}(x_i)\right)\\ &= n\frac{1}{\lambda} - \sum_{i=1}^{n}(x_i)\end{aligned}$$

Passo 4: annullamento derivata

$$\begin{aligned}\frac{d}{d\lambda}\ln(L(\lambda, x_1, x_2, \cdots, x_n)) &= 0\\ n\frac{1}{\lambda} - \sum_{i=1}^{n}(x_i) &= 0\\ \frac{n}{\lambda} &= \sum_{i=1}^{n}(x_i)\\ \lambda &= \frac{n}{\sum_{i=1}^{n}(x_i)}\end{aligned}$$

Otteniamo quindi lo stimatore per il parametro λ di un c.c. con distribuzione esponeziale

$$\hat{\Lambda} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}}$$

5.9 Cenni di stima intervallare

Finora sono stati mostrati metodi di stima puntuale che "provano ad indovinare" esattamente il valore del parametro. Se si individuasse un intervallo di valori per il parametro si potrebbe essere relativamente più confidenti che questo includa il valore reale.
Si dice **intervallo di confidenza** un intervallo casuale che comprende il parametro incognito ϑ con una certa probabilità:

$$\mathbb{P}(T_1 \leq \vartheta \leq T_2) = 1 - \alpha$$

L'intervallo $[T_1, T_2]$ è detto **intervallo di fiducia di livello** α, con α che rappresenta quindi la possibilità di "sbagliare". Analizziamo i passi per ottenere una stima intervallare:

Passo 1: individuare statistica cardine

È necessario innanzitutto individuare una statistica T (detta **statistica cardine** o **quantità pivotale**) che sia funzione soltanto del parametro da stimare.

Passo 2

Verificare che la distribuzione della statistica cardine non dipenda dal parametro da stimare.

Passo 3

$\forall\alpha \in (0,1)$ si individuano α_1, α_2 tali che $\mathbb{P}(\alpha_1 \leq T \leq \alpha_2) = 1-\alpha$.

Passo 4

Attraverso opportune manipolazioni, giungere ad un'espressione del tipo

$$\mathbb{P}(T_1 \leq \vartheta \leq T_2) = 1-\alpha$$

Lista di definizioni

Lista di teoremi

Lista di esercizi

Appendice A
Tabelle e integrazioni

A.1 Tabella della distribuzione normale standard

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998