Probabilità

7 giugno 2023

Indice

Capitolo 1

Calcolo combinatorio

1.1 Introduzione

Definizione 1.1.1 (Calcolo combinatorio)**.** Studio di come quantificare raggruppamenti aventi determinate caratteristiche degli elementi di un insieme finito di oggetti.

Osservazione 1. È fondamentale per il calcolo delle probabilità in quanto spesso la probabilità di un evento è calcolabile come il numero di modi in cui detto evento può verificarsi in rapporto al numero di casi possibili.

Definizione 1.1.2 (Principio fondamentale del calcolo combinatorio)**.** Se si realizzano due esperimenti:

- in cui il primo ha m esiti possibili;
- e per ognuno di questi il secondo ha n esiti possibili;
- e l'ordinamento conta per qualificare un esito (ossia sequenze diverse dei singoli esiti dei due esperimenti producono esiti finali distinti):

allora i due esperimenti (considerati congiuntamente) hanno $m \cdot n$ esiti possibili.

Osservazione 2. Generalizzato, con r esperimenti nel quale il primo abbia n_1 esiti possibili, per ciascuno di questi il secondo ne abbia n_2 ... per ogni esito dei primi due $r-1$ l'r-esimo n_r esiti possibili e l'ordinamento conta, allora gli esperimenti hanno in tutto $\prod_{i=1}^{r} n_i$ esiti possibili.

Definizione 1.1.3 (Funzione fattoriale)**.** Il fattoriale di n, indicato con $n!$ è una funzione $f : \mathbb{N} \to \mathbb{N}$ è definito come il prodotto dei primi n numeri interi:

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1 \tag{1.1}$$

Si conviene che $0! = 1$.

Osservazione 3 (Definizione ricorsiva). Dato che $(n-1) \cdot (n-2) \cdot \ldots \cdot 1 = (n-1)!$ il fattoriale può esser definito anche come:

$$n! = \begin{cases} 1 & n \in \mathbb{N}, n = 0 \\ n \cdot (n-1)! & n \in \mathbb{N}, n \neq 0 \end{cases} \tag{1.2}$$

Osservazione 4 (Una semplificazione utile). Se $0 < k < n$, si ha:

$$\frac{n!}{(n-k)!} = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) \tag{1.3}$$

1.2 Casistica principale

Supponendo di voler costruire sottoinsiemi contenenti k elementi scelti tra gli n elementi di un insieme U:

- nel caso in cui l'*ordine* abbia importanza (configurazioni con gli stessi elementi posti in ordine diverso danno origine ad esiti diversi) abbiamo a che fare con:
  - **permutazioni**: disponiamo di $k = n$ slot ed n elementi ($\in U$) da utilizzare per riempirli. Ci interessa sapere in quanti modi si possono ordinare gli n oggetti: ognuno di questi ordinamenti si chiama *permutazione*. Possiamo avere due casi:
    1. permutazioni *semplici*: gli n elementi da ordinare sono unici (ad esempio gli anagrammi della parola “AMORE”);
    2. permutazioni *con ripetizione*: ammettono che un elemento si presenti più volte tra gli n dai quali si può pescare (ad esempio gli anagrammi della parola “PEPPER”).
  - **disposizioni** (che costituiscono una versione generalizzata della permutazioni): gli slot sono in numero $k \leq n$ inferiore (o uguale) rispetto agli elementi n con il quale possiamo riempirli. Di fatto qua si considera che gli n elementi siano tutti distinti/diversi. Abbiamo:
    1. disposizioni *semplici*: i k elementi sono pescati da un insieme di n elementi distinti e una volta che l'elemento è stato scelto esce dal pool degli utilizzabili;
    2. disposizioni *con ripetizione*: ciascun elemento dei n può essere estratto più volte
- se viceversa l'*ordine non ha rilevanza*, ossia sottoinsiemi composti da medesimi elementi posti in ordine differente sono considerati uguali (ad esempio quando si vogliono contare insiemi nell'accezione matematica del termine) si ha a che fare con le **combinazioni**. Le combinazioni semplici sono le più utilizzate e si hanno quando il pool dal quale si pesca è composto da oggetti diversi/distinti tra loro.

1.2.1 Permutazioni

Proposizione 1.2.1 (Permutazioni semplici)**.** *Il numero di permutazioni di n elementi distinti in n slot è:*

$$\boxed{P_n = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1 = n!} \tag{1.4}$$

Dimostrazione. Nella prima posizione possiamo porre n alternative, nella seconda $n-1$ (visto che una è già andata nella prima), e così via; arrivando così all'ultima posizione rimane un solo oggetto possibile degli n iniziali. Pertanto per il principio fondamentale del calcolo combinatorio si conclude. $\square$

Osservazione 5. Nel caso in cui vi siano elementi ripetuti/uguali dai quali pescare (ad esempio se vogliamo permutare le lettere di "PEPPER") vogliamo che il numero di esiti complessivi diminuisca (evitando di contare come differenti due configurazioni con elementi uguali permutati tra loro)

Proposizione 1.2.2 (Permutazioni con ripetizione)**.** *Tra gli n dai quali pescare vi siano $i = 1, 2 \ldots r$ elementi univoci che si possono ripetere, aventi numerosità rispettivamente $k_1, k_2 \ldots k_r$ (ossia si ha $\sum_{i=1}^{r} i \cdot k_i = n$). Le permutazioni uniche (non ripetute) sono:*

$$\boxed{P_n^{k_1,k_2,\ldots,k_r} = \frac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_r!}} \tag{1.5}$$

Dimostrazione. Si parte dal numero di permutazioni degli n oggetti al numeratore Applicando il principio fondamentale del calcolo combinatorio al contrario, si tratta di dividere queste per il numero delle $k_1!$ permutazioni uguali fra loro (dovute al "girare" di uno stesso elemento), poi per le $k_2!$ permutazioni del secondo elemento multiplo, e così via. □

Esempio 1.2.1. Considerando le permutazioni PEPPER ad ogni sequenza univoca (ad esempio REPPEP) corrisponderanno 3!2! sequenze che sono di fatto uguali. Pertanto il numero di permutazioni univoche (con ripetizione) di PEPPER saranno $6!/(3! \cdot 2!)$.

Osservazione 6. La formula delle permutazioni è una generalizzazione e vale in realtà per qualsiasi permutazione, anche senza ripetizioni di elementi. Infatti, se abbiamo elementi univoci, ossia $k_1 = k_2 = \ldots = k_r = 1$, otteniamo esattamente la formula delle permutazioni semplici in quanto:

$$P_n^{k_1,k_2,\ldots,k_r} = \frac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_r!} = \frac{n!}{1! \cdot 1! \cdot \ldots \cdot 1!} = n! \tag{1.6}$$

1.2.2 Disposizioni

Definizione 1.2.1 (Disposizioni semplici)**.** Se il numero degli slot disponibili è inferiore (o uguale) al numero di elementi dai quali si pesca, gli elementi dai quali si pesca sono distinti tra loro e non vengono reinseriti nel pool dove pescare si hanno le disposizioni semplici.
Sono quello che in statistica si chiama *campionamento senza ripetizione.*

Proposizione 1.2.3 (Numero di disposizioni semplici)**.** *Il numero $D_{n,k}$ di disposizioni semplici di $k \leq n$ oggetti estratti da un insieme di n oggetti differenti è:*

$$\boxed{D_{n,k} = n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}} \tag{1.7}$$

Dimostrazione. Il primo componente di una tale sequenza può essere scelto in n modi diversi, il secondo in $(n-1)$ e così via, sino al k-esimo che può essere scelto in $(n-k+1)$ modi diversi. □

Osservazione 7. Le permutazioni semplici (quando $k = n$) sono casi particolari delle disposizioni semplici (quando $k \leq n$):

$$P_n = D_{n,n} = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = n! \tag{1.8}$$

Definizione 1.2.2 (Disposizioni con ripetizione)**.** Le disposizioni con ripetizione sono caratterizzate dal fatto che ciascuno degli n elementi possa essere estratto più volte per riempire i k slot.
Sono quello che in statistica si chiama *campionamento con ripetizione*.

Proposizione 1.2.4 (Numero di disposizioni con ripetizione)**.** *Il numero di disposizioni con ripetizione di n elementi in k slot:*

$$\boxed{D'_{n,k} = \underbrace{n \cdot n \cdot \ldots \cdot n}_{\text{k } volte} = n^k} \tag{1.9}$$

Dimostrazione. Si hanno n possibilità per scegliere il primo componente, n per il secondo, altrettante per il terzo e così via, sino al k-esimo; si conclude per il principio fondamentale del calcolo combinatorio. □

1.2.3 Combinazioni

1.2.3.1 Combinazioni semplici

Osservazione 8. Gli n elementi dai quali si pesca sono univoci: si pescano k elementi, l'ordine/disposizione di questi non è rilevante a qualificare un esito differente. Si hanno le combinazioni semplici che conteggiano il numero di sottoinsiemi di ampiezza definita di un determinato insieme base.

Proposizione 1.2.5 (Combinazioni semplici)**.** *Il numero delle combinazioni semplici di n elementi di lunghezza k, indicato con $C_{n,k}$ è:*

$$\boxed{C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!} = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k}} \tag{1.10}$$

Dimostrazione. Analogamente alle disposizioni semplici sceglieremo k elementi da n: si inizierà avendo n possibilità per il primo, sino a $n-k+1$ per il k-esimo. Tuttavia all'interno dei gruppi così determinati ci saranno combinazioni che sono formate dagli stessi elementi di altre, anche se in ordine inverso. Per non contare tali gruppi più volte (dato che l'ordine non interessa), sempre applicando il principio fondamentale del calcolo combinatorio, occorrerà dividere le disposizioni per il numero di permutazioni dei k elementi estratti ($k!$). □

1.2.3.2 Combinazioni con ripetizione

Osservazione 9. Nelle combinazioni semplici non è ammesso pescare lo stesso elemento più volte. Una volta estratto non rimane negli oggetti estraibili.
Nelle combinazioni con ripetizione invece vogliamo determinare quanti modi vi sono di scegliere k volte da un insieme di n oggetti diversi tra loro, ammettendo che però uno stesso oggetto possa essere pescato più volte.
L'ordine continua a non essere importante (ci interessa sono quante volte ogni oggetto è stato scelto, non l'ordine con cui esso appare).
Le combinazioni con ripetizione contano i *multiset* (insiemi che ammettono ripetizioni) sottoinsieme di un insieme dato.

Proposizione 1.2.6. *Il numero di combinazioni con ripetizione di k oggetti scelti tra n è*

$$\boxed{C^*_{n,k} = \binom{n+k-1}{k}} \tag{1.11}$$

Dimostrazione. Se l'ordine contasse il numero di combinazioni sarebbe n^k, ma questo non è il caso. Per dimostrare la formula risolviamo narrativamente un problema isomorfico (stesso problema con setup differente).
Il problema può essere posto come: porre k palline identiche in n scatole differenti: quello che conta è solamente il numero di palline in ciascuna scatola. Una qualsiasi configurazione può essere rappresentata come una sequenza di `|` per rappresentare i lati di una scatola e `o` per rappresentare le palline in essa. Ad esempio ipotizzando di avere $k = 7$ palline e $n = 4$ scatole, per rappresentare una pallina nella prima scatola, due nella seconda, tre nella terza e una nella quarta:

`|o|oo|ooo|o|`

Per essere valida ciascuna sequenza deve iniziare e finire con `|`: pertanto si tratta solo di contare il modo in cui si possono riarrangiare i termini rimanenti al suo interno (varie configurazioni di scatole). I termini all'interno dei bordi numero $n + k - 1$: di questi k (le palline) ed $((n - k - 1) - k) = n - 1$ anche (i bordi rimanenti utili per formare le n scatole, una volta che due sono stati impiegati per i lati). La soluzione è pertanto

$$\frac{(n+k-1)!}{k! \cdot (n-1)!} = \binom{n+k-1}{k}$$

□

1.3 Coefficiente binomiale e multinomiale

1.3.1 Coefficiente binomiale

1.3.1.1 Definizione

Osservazione 10. Approfondiamo il coefficiente che risulta dal calcolo del numero di combinazioni semplici di k elementi presi da n.

Definizione 1.3.1 (Coefficiente binomiale)**.** Indicato con $\binom{n}{k}$ e pronunciato "n su k" si definisce come

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!} = \frac{n!}{k! \cdot (n-k)!}$$

se $k \leq n$. Se $n < k$ si pone $\binom{n}{k} = 0$.

Osservazione 11. Per quanto riguarda il calcolo a mano, spesso è più utile/veloce la prima definizione, mentre la seconda è più compatta ed utilizzabile nelle parti teoriche.

1.3.1.2 Proprietà

Proposizione 1.3.1. *Si ha che:*

$$\boxed{\binom{n}{k} = \binom{n}{n-k}} \tag{1.12}$$

Dimostrazione.

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

□

Osservazione 12. Una intuizione sul significato di 1.12: per scegliere un comitato di k persone tra n sappiamo che ci sono $\binom{n}{k}$ modi. Un'altro modo di scegliere il comitato è specificare quali $n-k$ non ne faranno parte; specificare chi è nel comitato determina chi non vi è e viceversa. Pertanto i due lati sono uguali dato che sono due modi di contare la stessa cosa.

Osservazione 13. Esempi notevoli/utili della 1.12 sono:

$$\binom{n}{0} = \binom{n}{n} = 1, \qquad \binom{n}{1} = \binom{n}{n-1} = n \tag{1.13}$$

Proposizione 1.3.2.

$$\boxed{\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}} \tag{1.14}$$

Dimostrazione.

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

□

Osservazione 14. Per il significato di 1.14: se ho un insieme di n oggetti $I_n = \{1, \ldots, n\}$ isolando un oggetto (diciamo l'n-esimo) posso dividere i sottoinsiemi di I_n che hanno k oggetti in quelli che non contengono l'n-esimo (che sono $\binom{n-1}{k}$, essendo esattamente i sottoinsiemi di I_{n-1} a p oggetti) ed in quelli che lo contengono, i quali si ottengono aggiungendo n ad un insieme di $p-1$ oggetti di I_{n-1} e quindi sono in numero di $\binom{n-1}{p-1}$[1]; questi due gruppi di sottoinsimi di I_n sono evidentemente disgiunte, quindi l'unione ha la somma come cardinale, e quindi si ha la formula.

[1]Sarebbero $\binom{n-1}{p-1} \cdot 1$ poiché vi è un solo modo di aggiungere l'n-esimo ad un insimeme di $p-1$ elementi già formati (scelti tra $n-1$ elementi disponibili)

Proposizione 1.3.3 (Identità di Vandermonde)**.**

$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-k} \tag{1.15}$$

Dimostrazione. La prova mediante espansione dei termini e forza bruta ce la si può evitare. Una dimostrazione narrativa sul perché l'uguaglianza valga è comunque efficace.
Considerando un gruppo di m uomini ed n donne dal quale un comitato di k persone verrà scelto: ci sono $\binom{m+n}{k}$ per farlo. Se vi sono j uomini nel comitato, allora vi debbono essere $k-j$ donne. Il lato destro dell'uguaglianza smma per il numero j di uomini. □

Proposizione 1.3.4 (Squadra con capitano)**.** *Per $k, n \in N$ con $k \leq n$ si ha*

$$n\binom{n-1}{k-1} = k\binom{n}{k} \tag{1.16}$$

Dimostrazione. Una dimostrazione narativa: consideriamo un gruppo di n persone dal quale una squadra di k verrà scelta; uno di queste sarà capitano. Il numero possibile di team così formati può derivare da (lato sinistro) prima scegliere il capitano tra gli n e poi scegliere i $k-1$ rimanenti tra gli $n-1$ disponibili. Oppure ed equivalentemente scegliendo gli $\binom{n}{k}$ componenti e tra questi sceglierne uno dei k come capitano. □

1.3.1.3 Origine del nome

Osservazione 15. Il coefficiente binomiale prende nome dal fatto che determina i coefficienti dello sviluppo della potenza del binomio $(x+y)^n$

Proposizione 1.3.5 (Teorema binomiale)**.**

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \tag{1.17}$$

Dimostrazione. Per provare il teorema espandiamo il prodotto:

$$(x+y)^n = \underbrace{(x+y)\cdot(x+y)\cdot\ldots\cdot(x+y)}_{n \text{ fattori}}$$

I termini del prodotto $(x+y)^n$ sono ottenuti scegliendo la x o la y da ognuno dei fattori. Vi sono $\binom{n}{k}$ modi per scegliere esattamente k volte x (scegliendo y nei $n-k$ rimanenti): in questi casi si ottiene il termine $x^k y^{n-k}$. Il teorema si ottiene facendo variare il numero k di x scelti e sommando i termini risultati. □

1.3.2 Il coefficiente multinomiale

1.3.2.1 Definizione

Proposizione 1.3.6. *Il numero di modi in cui è possibile distribuire n oggetti distinti in r scatole distinte in modo che queste contengano, nell'ordine, $n_1, n_2, \ldots, n_r$ oggetti ($\sum_{i=1}^{r} n_i = n$) è:*

$$\boxed{\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_r!}} \tag{1.18}$$

Dimostrazione. Vi sono $\binom{n}{n_1}$ possibili scelte per gli oggetti della prima scatola; per ogni tale scelta vi sono $\binom{n-n_1}{n_2}$ scelte per la seconda; per ogni scelta effettuata nelle prime due vi sono $\binom{n-n_1-n_2}{n_3}$ nella terza e così via. Dal principio fondamentale del calcolo combinatorio discende che il risultato cercato è:

$$\binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \ldots \cdot \binom{n-n_1-\ldots-n_{r-1}}{n_r} \tag{1.19}$$

Sviluppando si ha

$$\frac{n!}{(n-n_1)!n_1!} \cdot \frac{(n-n_1)!}{(n-n_1-n_2)!n_2!} \cdot \ldots \cdot \frac{(n-n_1-n_2-\ldots-n_{r-1})!}{0!n_r!}$$

dalla quale, in seguito alle semplificazioni, si ottiene il coefficiente. □

Osservazione 16. Costituisce una generalizzazione del coefficiente binomiale (che si ottiene considerando due scatole).

Osservazione 17. Il coefficiente multinomiale è la formula che viene utilizzato nelle permutazioni con ripetizione (utile ad esempio per il numero di permutazioni di una parola con lettere ripetute).

1.3.2.2 Origine del nome

Osservazione 18. La formula del coefficiente multinomiale determina i coefficienti dello sviluppo di un polinomio di r termini

Proposizione 1.3.7 (Teorema multinomiale)**.**

$$(x_1 + x_2 + \cdots + x_r)^n = \sum_{\substack{(n_1,n_2,\ldots,n_r): \\ n_1+n_2+\cdots+n_r=n}} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n1} \cdot x_2^{n_2} \cdot \ldots \cdot x_r^{n_r}$$

Dimostrazione. Analoga al caso binomiale. □

Esempio 1.3.1. Nello sviluppo del cubo di un trinomio potremmo procedere manualmente:

$$(a+b+c)^3 = a^3 + b^3 + c^3 + 3a^2b + 3a^2c + 3b^2a + 3b^2c + 3c^2a + 3c^2b + 6abc$$

o calcolare più velocemente, ad esempio che:

- il termine $a^2b^0c^1$ presenta come coefficiente:

$$\binom{3}{2,0,1} = \frac{3!}{2! \cdot 0! \cdot 1!} = \frac{6}{2 \cdot 1 \cdot 1} = 3$$

- il termine $a^1b^1c^1$ ha invece coefficiente pari a:

$$\binom{3}{1,1,1} = \frac{3!}{1! \cdot 1! \cdot 1!} = \frac{6}{1 \cdot 1 \cdot 1} = 6$$

1.4 L'utilizzo di R

Fattoriale In R il fattoriale è ottenuto mediante la funzione `factorial`; poi nei casi di fattoriali di numeri molto grandi si può utilizzare `lfactorial` che calcola il logaritmo naturale del fattoriale di un numero, ossia $\log(n!)$. Nell'ultimo esempio di sotto si ha che approssimativamente $1000000! = 10^{5565709}$

```
factorial(4)

## [1] 24

factorial(1000000)

## [1] Inf

lfactorial(1000000)

## [1] 12815518

# convertiamo in logaritmo base 10
lfactorial(1000000)/log(10)

## [1] 5565709
```

Coefficiente binomiale Il coefficiente binomiale è implementato in R, mediante `choose`, mentre `lchoose` ne calcola il logaritmo:

```
choose(4,2)

## [1] 6

choose(1e06, 500)

## [1] Inf

lchoose(1e06, 500)

## [1] 4296.3
```

Permutazioni Per determinare tutte le permutazioni di un vettore si può utilizzare `permn` dal pacchetto `combinat`:

```
combinat::permn(1:3)

## [[1]]
## [1] 1 2 3
##
## [[2]]
## [1] 1 3 2
##
```

```
## [[3]]
## [1] 3 1 2
##
## [[4]]
## [1] 3 2 1
##
## [[5]]
## [1] 2 3 1
##
## [[6]]
## [1] 2 1 3
```

Combinazioni Per la determinazione delle combinazioni degli elementi da un vettore `x` presi tot (nell'esempio 2) alla volta si usa `combn`:

```
combn(letters[1:4], 2)

##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,] "a"  "a"  "a"  "b"  "b"  "c"
## [2,] "b"  "c"  "d"  "c"  "d"  "d"
```

1.5 Calcolo combinatorio e funzioni

Il calcolo combinatorio può essere applicato per contare le funzioni aventi determinate caratteristiche tra due insiemi finiti. Vediamo innanzitutto un criterio utile per contare e poi alcune applicazioni al conteggio delle funzioni.

1.5.1 Principio dell'overcounting

Sia $f : X \to Y$ suriettiva; si ha allora

$$\mathrm{Card}\,(X) = \sum_{y \in Y} \mathrm{Card}\,(f^{\leftarrow}(\{y\})) \tag{1.20}$$

In particolare se tutte le fibre $f^{\leftarrow}(\{y\})$ hanno una stessa cardinalità, ossia $\mathrm{Card}\,(f^{\leftarrow}(\{y\})) = \alpha$, si ha:

$$\mathrm{Card}\,(X) = \alpha\,\mathrm{Card}\,(Y) \tag{1.21}$$

essendo X una unione disgiunta delle fibre $f^{\leftarrow}(\{y\})$ al variare di $y \in Y$.
Anche detto principio del pastore, questo torna utile quando conosciamo la cardinalità di uno dei due insiemi (ad esempio pecore) e desideriamo ricavare quella dell'altro (numero di zampe).

1.5.2 Funzioni (disposizioni con ripetizione)

Si indica con X^{I_p} l'insieme di tutte le funzioni $f : I_p \to X$ con $\mathrm{Card}\,(I_p) = p$ e $\mathrm{Card}\,(X) = m$. Il numero di tutte le funzioni possibili tra i due insiemi è

$$\mathrm{Card}\left(X^{I_p}\right) = \underbrace{m \cdot m \cdot \ldots \cdot m}_{p \text{ volte}} = m^p \tag{1.22}$$

e corrisponde alle disposizioni con ripetizione, a p a p degli m oggetti di X.

1.5.3 Funzioni iniettive (disposizioni semplici)

Siamo interessati a quantificare la cardinalità del sottoinsieme delle funzioni inettive $\Lambda(n,p) \subset I_n^{I_p}$ del tipo $f : I_p \to I_n$. Si ha che

$$\mathrm{Card}\,(\Lambda(n,p)) = n \cdot (n-1) \cdot \ldots \cdot (n-(p-1)) \tag{1.23}$$

vedendo che all'ultimo elemento di I_p ho già fatto $p-1$ collegamenti, quindi me ne rimangono possibili $n-(p-1)$.

1.5.4 Permutazioni di un insieme (permutazioni semplici)

In particolare se $p = n$si hanno le biiezioni di un insieme I_n in se stesso, ossia le permutazioni dell'insieme, che sono in numero $n!$

1.5.5 Funzioni caratteristiche (coefficiente binomiale)

Il calcolo del numero di sottoinsiemi a p elementi di un insieme di n oggetti equivale a quantificare la cardinalità delle funzioni caratteristiche che scelgono p elementi tra un insieme di n (ossia tali che $\sum \chi(I_n) = p$).
Indicando con $C(n,p)$ l'insieme dei sottoinsiemi di I_n che hanno cardinale p si ha una funzione suriettiva:

$$s : \Lambda(n,p) \to C(n,p) \tag{1.24}$$

Il dominio $\Lambda(n,p)$ è un insieme di funzioni mentre il codominio $C(n,p)$ è un insieme di insiemi: la funzione suriettiva è quella che ad ogni funzione iniettiva $f : I_p \to I_n$ (con $f \in \Lambda(n,p)$) associa l'immagine $f(I_p) \in C(n,p)$, sottoinsieme a p oggetti di I_n.
Essendo che due funzioni iniettive facenti parte del dominio $f, g \in \Lambda(n,p)$ hanno la stessa immagine se e solo se differiscono per una permutazione sul proprio dominio, le fibre di s hanno tutte cardinale $p!$ (ossia ciascun insieme di p elementi si presenta in $p!$ ordini possibili), segue dal principio dell'overcounting che $\mathrm{Card}\,(\Lambda(n,p)) = p!\,\mathrm{Card}\,(C(n,p))$, quindi:

$$\mathrm{Card}\,(C(n,p)) = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-(p-1))}{p!} = \binom{n}{p} = \frac{n!}{p!(n-p)!} \tag{1.25}$$

Capitolo 2

Introduzione al calcolo della probabilità

2.1 Concetti primitivi

Definizione 2.1.1 (Esperimento)**.** Attività il cui risultato non può essere previsto con certezza (es lancio di un dado), sebbene l'insieme di tutti i possibili esiti lo sia.

Definizione 2.1.2 (Replica)**.** Viene detta così ciascuna singola ripetizione di un esperimento

Definizione 2.1.3 (Esito (Outcome))**.** È uno dei possibili risultati di una esperimento, ossia uno degli elementi dello spazio campionario.

Definizione 2.1.4 (Spazio campionario)**.** Insieme (non vuoto) di tutti i possibili esiti di un esperimento. Si indica con Ω o altre volte con S.

Esempio 2.1.1. Se l'esperimento è il sesso di un nascituro lo spazio campionario sarà $\Omega = \{m, f\}$. m è un possibile esito.

Esempio 2.1.2. Se l'esperimento è "l'ordine di arrivo in una corsa di sete cavalli" (numerati da 1 a 7) $\Omega = \{$tutte le 7! permutazioni di (1,2,3,4,5,6,7)$\}$.

Esempio 2.1.3. Lo spazio del " lancio di due dadi" sarà $\Omega = \{(1, 1), (2, 1), \ldots, (6, 6)\}$.

Esempio 2.1.4. Il numero di macchine conteggiate durante l'ora di punta in un incrocio ha $\Omega = \{0, 1, 2, \ldots\}$

Esempio 2.1.5. Il tempo di vita di un transistor avrà spazio campionario consistente nell'insieme dei numeri reali non negativi $\Omega = \{x \in \mathbb{R}^+ | x \geq 0\}$.

Definizione 2.1.5 (Spazi campionari finiti)**.** Sono gli spazi campionari che contengono un numero finito di possibili esiti.

Esempio 2.1.6. Degli esempi precedenti lo sono quelli di 2.1.1, 2.1.2 e 2.1.3

Definizione 2.1.6 (Spazi campionari infiniti)**.** Spazi la cui cardinalità è infinita: si suddividono in spazi contabili (in biezione con $\mathbb{N}$) e non contabili (rispettivamente con $\mathbb{R}$).

Esempio 2.1.7. Lo spazio dell'esempio 2.1.4 è infinito contabile, mentre quello di 2.1.5 è infinito non contabile.

Definizione 2.1.7 (Evento)**.** Un qualsiasi sottoinsieme dello spazio campionario (ovvero un insieme di esiti formato a piacere).

Definizione 2.1.8 (Eventi elementari)**.** Eventi rappresentati dai singoli esiti dell'esperimento che costituiscono lo spazio campionario

Esempio 2.1.8. Nella esperimento "lancio di una moneta", $E = \{T\}$ (si ottiene testa) è un evento elementare elementare (è uno dei possibili 2 esiti).

Definizione 2.1.9 (Eventi composti)**.** Eventi non elementari

Esempio 2.1.9. Nell'esperimento lancio di un dado l'evento $E = \{5, 6\}$ (si ottiene un 5 o un 6) è un evento composto

Esempio 2.1.10. L'evento che la somma dei due dadi sia 7 è $E = \{(1, 6), (2, 5), \ldots, (6, 1)\}$, ed è un evento composto

Definizione 2.1.10 (Evento verificato)**.** Se l'esito di un esperimento, una volta eseguito, è contenuto in E allora affermiamo che l'*evento E si è verificato.*

Osservazione 19*.* Dato che ogni sottoinsieme dello spazio campionario può costituire un evento, lo possono costituire anche l'insieme stesso e l'insieme vuoto, configurando due eventi particolari, l'evento certo e quello impossibile.

Definizione 2.1.11 (Evento certo)**.** Si indica (come lo spazio campionario) con Ω. Essendo lo spazio campionario l'insieme di tutti gli esiti di una esperimento, se lo consideriamo come esito esso si verifica sempre, poichè si verifica sempre almeno uno degli esiti dell'esperimento.

Definizione 2.1.12 (Evento impossibile)**.** Si indica con $\emptyset$ e consiste nell'insieme vuoto sottoinsieme di Ω. Verificandosi sempre almeno un esito appartenente allo spazio campionario, l'evento impossibile non si verifica mai.

Definizione 2.1.13 (Insieme degli eventi possibili)**.** L'insieme di tutti gli eventi possibili dato un determinato spazio campionario Ω si indica con $\mathcal{E}$

Osservazione 20 (Insieme di tutti gli eventi possibili)*.* Nel caso (quanto meno) di un Ω discreto, $\mathcal{E}$ coincide con l'insieme delle parti di Ω, ossia $\mathcal{E} = \mathcal{P}(\Omega)$.

Osservazione 21 (Rappresentazioni grafiche)*.* Nelle rappresentazioni grafiche si fanno tipicamente uso dei diagrammi di Venn; lo *spazio campionario* di un esperimento è rappresentato mediante un rettangolo (analogamente all'insieme universo nell'insiemistica), mentre gli *eventi* o mediante cerchi o rettangoli

2.2 Algebra degli eventi

L'algebra degli eventi presenta forti analogie con la logica e l'insiemistica.

2.2.1 Operatori

Vediamo qui quali sono gli operatori utili per creare nuovi eventi combinando dati eventi di partenza.

Definizione 2.2.1 (Unione di eventi)**.** L'unione di due eventi E_1 e E_2, indicato con $E_1 \cup E_2$, è quell'evento che si verifica se si verifica *almeno uno* dei due eventi E_1 e E_2.

Esempio 2.2.1. Se nel lancio di due monete, ad esempio, $E_1 = \{(T,T),(T,C)\}$ ed $E_2 = \{(C,T)\}$, allora

$$E_1 \cup E_2 = \{(T,T),(T,C),(C,T)\} \tag{2.1}$$

che coinciderà con l'evento "si verifica almeno una testa nei due lanci".

Osservazione 22. L'unione si può estendere ad un numero finito o infinito numerabile di eventi; indicando con

$$E_1 \cup E_2 \cup \ldots \cup E_n \cup \ldots = \bigcup_{i=1}^{\infty} E_i \tag{2.2}$$

l'evento che detto evento si verifica se si verifica almeno uno tra gli eventi E_i.

Definizione 2.2.2 (Intersezione)**.** L'intersezione tra due eventi E_1, E_2, indicato con $E_1 \cap E_2$ (o con E_1E_2), è quell'evento che si verifica se e solo se si verificano entrambi E_1 e E_2.

Esempio 2.2.2. Se $E_1 = \{(T,T),(T,C),(C,T)\}$ (ottenere almeno una volta testa) ed $E_2 = \{(C,C),(T,C),(C,T)\}$ (almeno una volta croce), allora

$$E_1E_2 = E_1 \cap E_2 = \{(C,T),(T,C)\}$$

che corrisponde all'evento di trovare esattamente una volta testa e una volta croce.

Osservazione 23. Analogamente al caso dell'unione, se $E_1, E_2, \ldots, E_n$ costituisce una collezione numerabile di eventi, indicheremo con:

$$E_1 \cap E_2 \cap \ldots \cap E_n \cap \ldots = \bigcap_{i=1}^{\infty} E_i \tag{2.3}$$

l' evento si verifica se si verifica tutti gli eventi E_i.

Definizione 2.2.3 (Negazione (o evento complementare))**.** A contrario dei precedenti è un operatore unario: la negazione di un evento E_i, indicata con $\overline{E_i}$ o E_i^c, è l'evento che si verifica quando non si verifica l'evento E_i. Ω.

Esempio 2.2.3. Ad esempio se $E = \{(T,T)\}$ (avere due teste nel lancio di due monete), considerando lo spazio campionario, $\overline{E} = \{(C,T),(T,C),(C,C)\}$ si verifica quando abbiamo almeno una croce.

Osservazione 24. La negazione della negazione coincide con l'evento stesso: $\overline{\overline{E_i}} = E_i$.

Proprietà	Unione	Intersezione
Idempotenza	$A \cup A = A$	$A \cap A = A$
Elemento neutro	$A \cup \emptyset = A$	$A \cap \Omega = A$
Commutativa	$A \cup B = B \cup A$	$A \cap B = B \cap A$
Associativa	$(A \cup B) \cup C = A \cup (B \cup C)$	$(A \cap B) \cap C = A \cap (B \cap C)$
Distributiva	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Tabella 2.1: Proprietà di unione ed intersezione

2.2.2 Proprietà delle operazioni tra eventi

2.2.2.1 Proprietà di unione e intersezione

In tabella 2.1 le principali proprietà di unione ed intersezione.

2.2.2.2 Leggi di De Morgan

Sono rappresentate dalle seguenti equazioni

$$\overline{A \cap B} = \overline{A} \cup \overline{B} \tag{2.4}$$
$$\overline{A \cup B} = \overline{A} \cap \overline{B} \tag{2.5}$$

Nella versione generalizzata a più di due eventi si scrivono rispettivamente come:

$$\overline{\bigcap_i E_i} = \bigcup_i \overline{E_i} \tag{2.6}$$
$$\overline{\bigcup_i E_i} = \bigcap_i \overline{E_i} \tag{2.7}$$

2.2.3 Relazioni tra eventi

Si specificano qui alcune proprietà che gli eventi presentano se considerati reciprocamente.

Definizione 2.2.4 (Inclusione)**.** Un evento E_1 è incluso nell'evento E_2 se tutte le volte che si verifica E_1, certamente si verifica E_2 (scritto $E_1 \subseteq E_2$).

Osservazione 25. L'evento $E_1 = \{1, 2\}$ ("risultato inferiore a 3") è contenuto nell'evento $E_2 = \{1, 2, 3\}$ ("risultato inferiore a 4")

Definizione 2.2.5 (Incompatibilità)**.** Due eventi E_1 e E_2 si dicono *incompatibili* (o *disgiunti*) se non possono verificarsi contemporaneamente, ovvero se $E_1 \cap E_2 = \emptyset$

Esempio 2.2.4. Se $E_1 = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$ (evento: somma dei due dadi =7) ed $E_2 = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$ (evento: somma dei due dadi =6), si dice che i due eventi sono incompatibili (o disgiunti) poichè, come si può verificare $E_1 \cap E_2 = \emptyset$.

Osservazione 26. Nei diagrammi di Venn, due eventi compatibili sono rappresentati da due aree che non overlappano.

Definizione 2.2.6 (Incompatibilità a coppie di una collezione di eventi)**.** Data una collezione di eventi $E_i \in \mathcal{A}$, $1 \leq i \leq \infty$, diremo che sono incompatibili a due a due se qualunque coppia di eventi distinti non può verificarsi contemporaneamente

$$E_i \cap E_j = \emptyset \quad \forall i \neq j$$

Osservazione 27. Analogamente si può parlare di incompatibilità a tre a tre, a n a n, ecc. Si può dedurre che l'incompatibilità a 2 a 2 implica le incompatibilità di ordine maggiore, ma non viceversa.

Definizione 2.2.7 (Eventi necessari)**.** Due eventi A e B si dicono *necessari* se la loro unione è l'evento certo, cioè se si verifica certamente almeno uno di essi, ovvero se $A \cup B = \Omega$.

Osservazione 28. Analogamente una collezione di eventi $E_i, 1 \leq i \leq \infty$ si dice necessaria se almeno uno di essi certamente si verifica, ovvero se $\bigcup_{i=1}^{\infty} E_i = \Omega$

Definizione 2.2.8 (Partizione di Ω)**.** Una *partizione di* Ω è una collezione di eventi $\{E_i\} \in \Omega$ necessari e incompatibili:

$$\bigcup_{i=1}^{\infty} E_i = \Omega$$
$$E_i \cap E_j = \emptyset \quad i \neq j$$

Osservazione 29. Sui diagrammi di Venn, una partizione di eventi si rappresenta come una collezione di figure che ricoprono Ω senza sovrapporsi.

2.3 Probabilità

Al fine di poter sviluppare la teoria della probabilità si fanno alcune assunzioni sugli eventi e sulle probabilità.

2.3.1 Assunzioni

2.3.1.1 σ-algebra degli eventi

Si ipotizza che l'insieme degli eventi possibili $\mathcal{E}$ costituisca una σ-algebra, o algebra di Boole completa.

Definizione 2.3.1 (Algebra di Boole completa / σ-algebra)**.** Una collezione $\mathcal{E}$ di eventi $E_1, E_2, \ldots, E_n$ costituisce un'algebra di Boole se

$$E_i \in \mathcal{E} \implies \overline{E_i} \in \mathcal{E}$$
$$E_i \in \mathcal{E} \implies \bigcup_{i=1}^{n} E_i \in \mathcal{E}$$

ossia $\mathcal{E}$ è chiusa rispetto alla negazione e all'unione di una quantità numerabile di eventi (se ciò avviene solo per n finito essa si chiama semplicemente algebra di Boole).

2.3.1.2 Funzione di probabilità e assiomi di Kolmogorov

Questi stabiliscono le caratteristiche base di una qualsivoglia funzione di probabilità. Consideriamo un esperimento il cui spazio campionario sia Ω.
Per ogni evento $E_i \in \mathcal{E}$, supponiamo sia definito un numero $\mathrm{P}(E_i)$, generato da una funzione che soddisfa i seguenti tre assiomi.

Assioma 1. La probabilità è una funzione, indicata con P o Pr, che assegna ad ogni evento un numero reale compreso tra 0 e 1:

$$\mathrm{P} : \mathcal{E} \to [0, 1]$$

Assioma 2. L'evento certo ha probabilità 1:

$$\mathrm{P}(\Omega) = 1 \tag{2.8}$$

Assioma 3. La probabilità dell'unione di un'infinità numerabile di eventi incompatibili è la somma (eventualmente di un numero infinito di termini) delle singole probabilità:

$$E_i \cap E_j = \emptyset, \forall i \neq j \implies \mathrm{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(E_i) \tag{2.9}$$

Osservazione 30. Su questi postulati vi è quasi accordo unanime, in quanto reputati naturali e in accordo con il concetto intuitivo di probabilità che ognuno di noi ha.

Osservazione 31. Da notare che i 3 assiomi non dicono quale funzione di probabilità utilizzare: semplicemente limitano il campo delle funzioni utilizzabili.

2.3.2 Primi risultati

Nel seguito consideriamo A e B come generici eventi tali che $A, B \subset \Omega$.

Proposizione 2.3.1.

$$\boxed{\mathrm{P}\left(\overline{A}\right) = 1 - \mathrm{P}(A)} \tag{2.10}$$

Dimostrazione.

$$\begin{aligned} A \cup \overline{A} &= \Omega \\ P\left(A \cup \overline{A}\right) &= P(\Omega) \\ \mathrm{P}(A) + \mathrm{P}\left(\overline{A}\right) &= 1 \end{aligned}$$

□

Esempio 2.3.1. Ad esempio se la probabilità di ottenere testa lanciando una moneta è pari a $\frac{3}{8}$ allora la probabilità di ottenere croce dovrà essere $\frac{5}{8}$.

Proposizione 2.3.2.

$$\boxed{\mathrm{P}(\emptyset) = 0} \tag{2.11}$$

Dimostrazione. Imponendo $A = \Omega$ in 2.10,

$$\begin{aligned} \mathrm{P}\left(\overline{\Omega}\right) &= 1 - \mathrm{P}\left(\Omega\right) \\ \mathrm{P}\left(\emptyset\right) &= 1 - 1 \end{aligned}$$

□

Proposizione 2.3.3.

$$\boxed{A \subseteq B \implies \mathrm{P}\left(A\right) \leq \mathrm{P}\left(B\right)} \tag{2.12}$$

Dimostrazione. Se A è incluso in B, allora può esser scritto mediante l'unione di due eventi incompatibili (da cui applicare il terzo postulato):

$$\begin{aligned} B &= A \cup (B \cap \overline{A}) \\ \mathrm{P}\left(B\right) &= \mathrm{P}\left(A\right) + \mathrm{P}\left(B \cap \overline{A}\right) \end{aligned}$$

pertanto chiaramente $\mathrm{P}\left(B\right) \geq \mathrm{P}\left(A\right)$. □

Proposizione 2.3.4. *La probabilità che si verifichi l'evento A, ma non l'evento B:*

$$\boxed{\mathrm{P}\left(A \cap \overline{B}\right) = \mathrm{P}\left(A\right) - \mathrm{P}\left(A \cap B\right)} \tag{2.13}$$

Dimostrazione. Infatti, vedendo A come l'unione di due parti così definite, e sviluppando:

$$\begin{aligned} A &= (A \cap B) \cup (A \cap \overline{B}) \\ \mathrm{P}\left(A\right) &= \mathrm{P}\left((A \cap B) \cup (A \cap \overline{B})\right) \\ \mathrm{P}\left(A\right) &= \mathrm{P}\left(A \cap B\right) + \mathrm{P}\left(A \cap \overline{B}\right) \end{aligned}$$

e si conclude come in proposizione. □

Proposizione 2.3.5.

$$\boxed{\mathrm{P}\left(A \cup B\right) = \mathrm{P}\left(A\right) + \mathrm{P}\left(B\right) - \mathrm{P}\left(A \cap B\right)} \tag{2.14}$$

Dimostrazione. Scriviamo $A \cup B$ come l'unione tra due eventi incompatibili, applichiamo il terzo postulato e poi la 2.13:

$$\begin{aligned} A \cup B &= A \cup (B \cap \overline{A}) \\ \mathrm{P}\left(A \cup B\right) &= \mathrm{P}\left(A\right) + \mathrm{P}\left(B \cap \overline{A}\right) \\ \mathrm{P}\left(A \cup B\right) &= \mathrm{P}\left(A\right) + \mathrm{P}\left(B\right) - \mathrm{P}\left(A \cap B\right) \end{aligned}$$

□

Proposizione 2.3.6 (Principio di inclusione/esclusione)**.** *Nel caso dell'unione di n eventi, il calcolo della probabilità del relativo evento è determinato* *dal* principio di inclusione/esclusione[1], *che può esser scritto in forma compatta come:*

$$\mathrm{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n}(-1)^{r+1} \sum_{i_1<\ldots<i_r} \mathrm{P}\left(E_{i1} \cap E_{i2} \cap \ldots \cap E_{ir}\right) \tag{2.15}$$

$$= \sum_{i} \mathrm{P}\left(E_i\right) - \sum_{i<j} \mathrm{P}\left(E_i \cap E_j\right) + \sum_{i<j<k} \mathrm{P}\left(E_i \cap E_j \cap E_k\right) - \ldots + (-1)^{n+1}\,\mathrm{P}\left(E_1 \cap \ldots \cap E_n\right) \tag{2.16}$$

Dimostrazione. Rimandata a sezione 5.2.4.3. □

Osservazione 32. La 2.15 stabilisce che la probabilità dell'unione di n eventi è uguale alla somma delle probabilità di questi eventi, meno la somma delle probabilità delle intersezioni degli eventi presi a due a due, più la somma delle probabilità delle intersezioni degli eventi presi a tre a tre e così via.

Esempio 2.3.2. Possiamo sviluppare il caso di tre eventi, E, F, G:

$$\begin{aligned}\mathrm{P}\left(E \cup F \cup G \cup\right) &= \mathrm{P}\left((E \cup F) \cup G\right) \\ &= \mathrm{P}\left(E \cup F\right) + \mathrm{P}\left(G\right) - \mathrm{P}\left((E \cup F) \cap G\right)\end{aligned}$$

Ora $(E \cup F) \cap G$ per la legge distributiva è un evento equivalente a $(E \cap G) \cup (F \cap G)$, quindi dalla precedente equazione otteniamo

$$\begin{aligned}\mathrm{P}\left(E \cup F \cup G\right) &= \mathrm{P}\left(E\right) + \mathrm{P}\left(F\right) - \mathrm{P}\left(E \cap F\right) + \mathrm{P}\left(G\right) - \mathrm{P}\left((E \cap G) \cup (F \cap G)\right) \\ &= \mathrm{P}\left(E\right) + \mathrm{P}\left(F\right) - \mathrm{P}\left(E \cap F\right) + \mathrm{P}\left(G\right) - \mathrm{P}\left(E \cap G\right) - \mathrm{P}\left(F \cap G\right) + \mathrm{P}\left((E \cap G) \cap (F \cap G)\right) \\ &= \mathrm{P}\left(E\right) + \mathrm{P}\left(F\right) + \mathrm{P}\left(G\right) - \mathrm{P}\left(E \cap F\right) - \mathrm{P}\left(E \cap G\right) - \mathrm{P}\left(F \cap G\right) + \mathrm{P}\left(E \cap G \cap F\right)\end{aligned}$$

Proposizione 2.3.7 (Disuguaglianza di Boole)**.** *Una relazione notevole che sussiste tra la probabilità dell'unione e la somma delle singole probabilità è*

$$\mathrm{P}\left(E_1 \cup E_2 \cup \ldots \cup E_n\right) \leq \sum_{i=1}^{n} \mathrm{P}\left(E_i\right) \tag{2.17}$$

Dimostrazione. Rimandata a sezione 5.2.4.3. □

Proposizione 2.3.8 (Diseguaglianza di Bonferroni)**.** *Una relazione notevole per ciò che riguarda invece l'intersezione è data*

$$\mathrm{P}\left(E_1 \cap E_2 \cap \ldots \cap E_n\right) \geq 1 - \sum_{i=1}^{n} \mathrm{P}\left(\overline{E_i}\right) \tag{2.18}$$

Dimostrazione. Rimandata a sezione 5.2.4.3 □

2.4 La misura della probabilità

Nelle parti precedenti sono stati discussi risultati concernenti eventi e probabilità, senza mai pervenire alla misura di queste.In questa sezione si introduce il calcolo delle probabilità effettivo.

[1]Dimostrabile per induzione.

2.4.1 Spazi campionari con esiti (discreti) equiprobabili

In molti esperimenti è ragionevole assumere che tutti gli esiti dello spazio campionario siano equiprobabili (es lancio di un dado). Se lo spazio campionario Ω è un *insieme finito*.

$$\Omega = \{1, 2, \ldots, N\}$$

e i singoli esiti sono equiprobabili

$$\mathrm{P}(1) = \mathrm{P}(2) = \ldots = \mathrm{P}(N) = \mathrm{P}(i)$$

si avrà che (dato che gli eventi sono disgiunti e la loro probabilità somma a 1, assiomi 2 e 3):

$$N\,\mathrm{P}(i) = 1 \tag{2.19}$$

$$\mathrm{P}(i) = \frac{1}{n} \tag{2.20}$$

Conoscendo la probabilità del singolo esito, siamo interessati a determinare la probabilità di un generico evento composto da uno o più esiti. Per l'assioma 3 avremo che la probabilità di un evento E sarà determinata da:

$$\mathrm{P}(E) = \frac{\#\text{ esiti E}}{\#\text{ esiti }\Omega}$$

a parole proporzione degli esiti dello spazio campionario contenuti in E, con una formula nota "casi favorevoli su casi possibili". La quantificazione di casi favorevoli e possibili, infine, passa spesso attraverso il calcolo combinatorio.

Esempio 2.4.1 (Birthday problem)**.** Ci sono k persone in una stanza. Assumendo che siano nate in uno dei 365 giorni dell'anno con probabilità uguale per ciascun giorno (escludiamo anni bisestili) e che i compleanni siano indipendenti (es non vi sono gemelli nella stanza), quale è la probabilità che due o più persone nel gruppo compiano gli anni lo stesso giorno?
La calcoliamo come complemento della probabilità che nessuno faccia compleanno lo stesso giorno: questa è data da casi favorevoli (numero di modi possibili per avere compleanni in date differente) fratto casi possibili (numero di possibili configurazioni di compleanni. Si ha:

$$\mathrm{P}(k\text{ compleanni diversi}) = \frac{365 \cdot \ldots \cdot (365 - k + 1)}{365^k}$$

da cui

$$\mathrm{P}(\text{Almeno due uguali tra } k) = 1 - \frac{365 \cdot \ldots \cdot (365 - k + 1)}{365^k}$$

Eseguendo i conti si nota come si supera la probabilità del 50% già con $k = 23$ persone (ossia in un gruppo di 23 persone c'è poco più del 50% di probabilità di averne due o più che fanno gli anni lo stesso giorno) mentre a $k = 57$ la probabilità è già oltre il 99%.

```
prob_birthday <- function(k){
    # vectorized for several k
    num <- unlist(lapply(k, function(k2) prod(seq(365, 365 - k2 + 1))))
    den <- 365^k
    1 - num/den
}
k <- 1:60
round(prob_birthday(k = k), 4)
```

```
##  [1] 0.0000 0.0027 0.0082 0.0164 0.0271 0.0405 0.0562 0.0743 0.0946 0.1169
## [11] 0.1411 0.1670 0.1944 0.2231 0.2529 0.2836 0.3150 0.3469 0.3791 0.4114
## [21] 0.4437 0.4757 0.5073 0.5383 0.5687 0.5982 0.6269 0.6545 0.6810 0.7063
## [31] 0.7305 0.7533 0.7750 0.7953 0.8144 0.8322 0.8487 0.8641 0.8782 0.8912
## [41] 0.9032 0.9140 0.9239 0.9329 0.9410 0.9483 0.9548 0.9606 0.9658 0.9704
## [51] 0.9744 0.9780 0.9811 0.9839 0.9863 0.9883 0.9901 0.9917 0.9930 0.9941
```

Capitolo 3

Probabilità condizionata e indipendenza

La probabilità condizionata è un concetto importante: talvolta occorre calcolare la probabilità di un evento dopo che si è (o nell'ipotesi che si sia) verificato un altro evento, logicamente o temporalmente collegato al primo. Altresì le probabilità condizionate sono spesso utilizzate per calcolare più facilmente le probabilità richieste.

3.1 Probabilità condizionata e affini

3.1.1 Probabilità condizionata

Osservazione 33. Considerando due eventi A e B, ipotizziamo che si sia verificato A; la conseguenza più evidente sulla probabilità dell'evento B è il fatto che ora lo spazio campione Ω è diventato A. L'evento A assume il ruolo di evento certo, e qualunque evento che ha probabilità maggiore di 0 *deve esser incluso* in A. Se A si realizza, dunque, B si realizza solo se si realizza l'evento $A \cap B$. Pertanto la probabilità condizionata potrà esser definita in maniera analoga alla probabilità normale, come frequenza di casi favorevoli, $\mathrm{P}(A \cap B)$, su possibili $\mathrm{P}(A)$.

Definizione 3.1.1 (Probabilità condizionata)**.** Si definisce probabilità condizionata di A dato B, indicata con $\mathrm{P}(A|B)$, la seguente:

$$\boxed{\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}} \tag{3.1}$$

Osservazione 34. Sia da un punto di vista logico che matematico, deve essere $\mathrm{P}(A) > 0$.

Osservazione 35. $\mathrm{P}(A|B) \neq \mathrm{P}(B|A)$; in particolar modo le due probabilità differiscono per il denominatore

Osservazione 36. I casi estremi della probabilità condizionata si avranno in corrispondenza di:

$$\begin{aligned} A \cap B = \emptyset &\implies \mathrm{P}(A|B) = 0 \\ A \subseteq B &\implies \mathrm{P}(A|B) = 1 \end{aligned}$$

3.1.2 Probabilità dell'intersezione di eventi

Proposizione 3.1.1 (Probabilità dell'intersezione di 2 eventi)**.** *Se* $\mathrm{P}(A) \neq 0$ *si ha che:*

$$\boxed{\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B|A)} \tag{3.2}$$

Data la commutatività dell'intersezione, se $\mathrm{P}(B) \neq 0$*:*

$$\boxed{\mathrm{P}(A \cap B) = \mathrm{P}(B)\,\mathrm{P}(A|B)} \tag{3.3}$$

Dimostrazione. Si tratta di semplici rielaborazioni algebriche della 3.1. □

Proposizione 3.1.2 (Probabilità dell'intersezione di n eventi)**.** *Nel caso in cui* $\mathrm{P}(E_1 \cap E_2 \cap \ldots \cap E_{n-1}) > 0$*, la probabilità dell'intersezione di n eventi:*

$$\mathrm{P}\left(\bigcap_{i=1}^{n} E_i\right) = \mathrm{P}(E_1)\cdot\mathrm{P}(E_2|E_1)\cdot\mathrm{P}(E_3|E_1 \cap E_2)\cdot\ldots\cdot\mathrm{P}(E_n|E_1 \cap E_2 \cap \ldots \cap E_{n-1})$$

Questa è nota anche come regola del prodotto.

Dimostrazione. Per verificarla è possibile sviluppare la regola del prodotto applicando ricorsivamente la 3.2 al secondo membro della regola del prodotto:

$$\mathrm{P}(E_1) \cdot \frac{\mathrm{P}(E_1 \cap E_2)}{\mathrm{P}(E_1)} \cdot \frac{\mathrm{P}(E_1 \cap E_2 \cap E_3)}{\mathrm{P}(E_1 \cap E_2)} \cdot \ldots \cdot \frac{\mathrm{P}(E_1 \cap E_2 \cap \ldots \cap E_n)}{\mathrm{P}(E_1 \cap E_2 \cap \ldots \cap E_{n-1})} \tag{3.4}$$

e dopo le semplificazioni rimane $\mathrm{P}(E_1 \cap E_2 \cap \ldots \cap E_n) = \mathrm{P}\left(\bigcap_{i=1}^{n} E_i\right)$. □

Osservazione 37. Le probabilità che compaiono nei denominatori di 3.4 sono strettamente positive grazie all'ipotesi fatta:

$$\mathrm{P}(E_1 \cap E_2 \cap \ldots \cap E_{n-1}) > 0$$

ossia che l'intersezione di tutti gli eventi ad eccezione di uno non sia vuota. Dall'ipotesi discende che anche la probabilità dell'intersezione di un numero minore di eventi sarà positiva (poiché stiamo "eliminando condizionamenti").

Osservazione 38. Nella pratica possiamo girare gli eventi come torna più comodo al calcolo in quanto, ad esempio:

$$\begin{aligned} \mathrm{P}(E_1 \cap E_2 \cap E_3) &= \mathrm{P}(E_1) \cdot \mathrm{P}(E_2|E_1) \cdot \mathrm{P}(E_3|E_1 \cap E_2) \\ &= \mathrm{P}(E_3) \cdot \mathrm{P}(E_2|E_3) \cdot \mathrm{P}(E_1|E_3 \cap E_2) \end{aligned}$$

3.1.3 Odds ratio (rapporto a favore) di un evento

Definizione 3.1.2 (Odds ratio (rapporto a favore))**.** L'odds ratio di un evento A è definito come

$$\mathrm{OR}(A) = \frac{\mathrm{P}(A)}{\mathrm{P}(\overline{A})} = \frac{\mathrm{P}(A)}{1 - \mathrm{P}(A)} \tag{3.5}$$

ed esprime quanto è più probabile che l'evento si realizzi rispetto al fatto che non si realizzi.

Osservazione 39. Per convertire da odds ratio a probabilità, come si può verificare sostituendo, si ha:

$$\mathrm{P}(A) = \frac{\mathrm{OR}(A)}{1 + \mathrm{OR}(A)} \tag{3.6}$$

Osservazione 40. Può essere di interesse la modifica della probabilità che una ipotesi H sia vera $\mathrm{P}(H)$ quando si dispone di informazioni su una prova E; le probabilità condizionate dato E che H sia vera o meno

$$\mathrm{P}(H|E) = \frac{\mathrm{P}(H \cap E)}{\mathrm{P}(E)} = \frac{\mathrm{P}(H)\,\mathrm{P}(E|H)}{\mathrm{P}(E)}$$
$$\mathrm{P}(\overline{H}|E) = \frac{\mathrm{P}(\overline{H} \cap E)}{\mathrm{P}(E)} = \frac{\mathrm{P}(\overline{H})\,\mathrm{P}(E|\overline{H})}{\mathrm{P}(E)}$$

Definizione 3.1.3 (Odds ratio condizionato)**.** L'odds ratio dell'ipotesi H non è più $\frac{\mathrm{P}(H)}{\mathrm{P}(\overline{H})}$, ma a seguito delle conoscenze su (o nell'ipotesi di) E è dato da:

$$\frac{\mathrm{P}(H|E)}{\mathrm{P}(\overline{H}|E)} = \frac{\frac{\mathrm{P}(H)\,\mathrm{P}(E|H)}{\mathrm{P}(E)}}{\frac{\mathrm{P}(\overline{H})\,\mathrm{P}(E|\overline{H})}{\mathrm{P}(E)}} = \frac{\mathrm{P}(H)}{\mathrm{P}(\overline{H})} \cdot \frac{\mathrm{P}(E|H)}{\mathrm{P}(E|\overline{H})} \tag{3.7}$$

Osservazione 41. A seguito dell'introduzione di una prova l'originale rapporto a favore $\frac{\mathrm{P}(H)}{\mathrm{P}(\overline{H})}$ viene moltiplicato per un secondo termine che ne determina l'eventuale variazione: il rapporto a favore finale (e quindi la probabilità di H) aumenta se E è più probabile quando H è vera che quando H è falsa (secondo termine del prodotto) e diminuisce in caso contrario.

3.2 Il teorema di Bayes

3.2.1 Teorema delle probabilità totali

Osservazione 42 (Versione semplice). Se E ed C sono due eventi, possiamo scomporre E nell'unione di due eventi disgiunti come segue:

$$E = (E \cap C) \cup (E \cap \overline{C})$$

Trattandosi di eventi disgiunti, se vogliamo calcolare $\mathrm{P}(E)$ possiamo scrivere (applicando il terzo assioma di Kolmogorov):

$$\begin{aligned}
\boxed{\mathrm{P}(E)} &= \mathrm{P}\left((E \cap C) \cup (E \cap \overline{C})\right) && (3.8)\\
&= \mathrm{P}(E \cap C) + \mathrm{P}\left(E \cap \overline{C}\right)\\
&= \boxed{\mathrm{P}(C)\,\mathrm{P}(E|C) + \mathrm{P}\left(\overline{C}\right)\mathrm{P}\left(E|\overline{C}\right)} && (3.9)\\
&= \mathrm{P}(C)\,\mathrm{P}(E|C) + (1 - P(C))\,\mathrm{P}\left(E|\overline{C}\right)
\end{aligned}$$

La 3.9 afferma che $\mathrm{P}(E)$ può essere riscritta come media ponderata della probabilità di verificarsi di E dato il verificarsi di C o meno.

Osservazione 43 (Condizionamento per la risoluzione dei problemi). È una formula utile in quanto in alcuni casi è difficile determinare direttamente la probabilità di E; in tali casi si possono impiegare (se disponibili) le informazioni riguardo alle sue relazioni con C.
Condizionare è uno strumento potente per risolvere problemi: quando incontriamo un problema che sarebbe semplice se solo sapessimo se C è accaduto o meno, possiamo condizionare su C e $\overline{C}$, considerare le ipotesi separatamente e combinare mediante la legge della probabilità totale.
Prassi comuni sono condizionare su quello che vorremmo sapere (ipotizzando che sia vera una cosa o l'altra) oppure negli esperimenti a step successivi possiamo condizionare sull'esito del primo step.

Osservazione 44. Vediamo ora la versione generalizzata ad una partizione di n (> 2) elementi di quanto fatto sinora.

Teorema 3.2.1 (Teorema delle probabilità totali). *Sia $C_1, C_2, \ldots, C_n$ una partizione di Ω; la probabilità di un generico evento E può essere scritta come:*

$$\boxed{\mathrm{P}(E) = \sum_{i=1}^{n} \mathrm{P}(C_i)\,\mathrm{P}(E|C_i)} \tag{3.10}$$

Dimostrazione. Se $C_1, C_2, \ldots, C_n$ sono una partizione di Ω sono a due a due disgiunti ed esauriscono Ω. Pertanto considerando l'evento E, lo possiamo spezzare analogamente a quanto fatto in precedenza (utilizzando non più due eventi ma una molteplicità di eventi che esauriscono comunque Ω):

$$E = \Omega \cap E = \left(\bigcup_{i=1}^{n} C_i\right) \cap E = (C_1 \cap E) \cup (C_2 \cap E) \cup \ldots \cup (C_n \cap E)$$

Essendo gli $(C_i \cap A)$ eventi incompatibili, la probabilità dell'evento E può esser scritta come somma:

$$\mathrm{P}(E) = \sum_{i=1}^{n} \mathrm{P}(C_i \cap E) = \sum_{i=1}^{n} \mathrm{P}(C_i)\,\mathrm{P}(E|C_i) \tag{3.11}$$

in cui, nell'ultimo passaggio, abbiamo utilizzato la 3.2. □

3.2.2 La formula di Bayes

Osservazione 45. La formula di Bayes e la legge delle probabilità totali permettono il calcolo delle probabilità condizionate in un ampio ventaglio di casi. La formula di Bayes mette in relazione $\mathrm{P}(B|A)$ e $\mathrm{P}(A|B)$

Teorema 3.2.2 (Formula di Bayes)**.**

$$\boxed{\mathrm{P}(A|B) = \frac{\mathrm{P}(A) \cdot \mathrm{P}(B|A)}{\mathrm{P}(B)}} \tag{3.12}$$

Dimostrazione. Si può derivare equivalentemente ponendo ad uguaglianza le equazioni di proposizione 3.1.1, oppure sostituendo nella definizione 3.1 al numeratore la 3.2 □

Osservazione 46 (Impiego in ambito decisionale). Quando si svolge un test per verificare una ipotesi, viene utilizzato come segue (dove sfruttiamo il teorema delle probabilità totali al denominatore). Siano H è "la mia ipotesi è vera", T "test positivo"; allora:

$$\mathrm{P}(H|T) = \frac{\mathrm{P}(H) \cdot \mathrm{P}(T|H)}{\mathrm{P}(T)} = \frac{\mathrm{P}(H) \cdot \mathrm{P}(T|H)}{\mathrm{P}(H) \cdot \mathrm{P}(T|H) + \mathrm{P}\left(\overline{H}\right) \cdot \mathrm{P}\left(T|\overline{H}\right)}$$

In questo caso $\mathrm{P}(H|T)$ si dice *probabilità a posteriori*, $\mathrm{P}(H)$ *probabilità a priori* e $\mathrm{P}(T|H)$ *verosimiglianza.*

Osservazione 47 (VPP e VPN). In contesto diagnostico, ad esempio, se D è "essere malato" e T è "essere positivo al test", $\mathrm{P}(D|T)$ (e l'applicazione del teorema di Bayes) ci restituisce la stima del valore predittivo positivo. Viceversa $\mathrm{P}\left(\overline{D}|\overline{T}\right)$ del valore predittivo negativo.

Osservazione 48. Il seguente risultato è semplicemente l'inserimento del teorema delle probabilità totali al denominatore della formula di Bayes e torna spesso utile quando non si conosce direttamente la probabilità al denominatore del teorema di Bayes.

Corollario 3.2.3. *Siano C_i una partizione di Ω ed E un generico evento; la probabilità condizionata del verificarsi del j-esimo evento C_j, per $1 \leq j \leq n$, dato il verificarsi di E è data da:*

$$\boxed{\mathrm{P}(C_j|E) = \frac{\mathrm{P}(C_j)\,\mathrm{P}(E|C_j)}{\sum_{i=1}^{n} \mathrm{P}(C_i)\,\mathrm{P}(E|C_i)}}$$

La $\mathrm{P}(C_j|E)$ è detta probabilità a posteriori; le $\mathrm{P}(C_i)$ sono dette probabilità a priori; le $\mathrm{P}(E|C_i)$ verosimiglianze.

Osservazione 49 (Una interpretazione del teorema). Supponiamo che E sia un evento che possa realizzarsi in conseguenza di n cause C_i, una delle quali certamente agisce ed ognuna delle quali ha probabilità $\mathrm{P}(C_i)$ di verificarsi.
Supponiamo anche che $\mathrm{P}(E|C_j)$ sia la probabilità che E si verifichi qualora si sia verificata la causa j-esima C_j. Allora la formula di Bayes esprime la probabilità a posteriori $\mathrm{P}(C_j|E)$, cioè la probabilità che avendo osservato l'evento E, esso sia stato generato dalla causa C_j, in funzione della probabilità a priori $\mathrm{P}(C_j)$ e delle verosimiglianze $\mathrm{P}(E|C_j)$.

Osservazione 50. Questa interpretazione rende utile la formula poiché di fronte ad un risultato osservato (l'evento E), che può esser stato generato da più cause (gli eventi C_i), è possibile calcolare la probabilità che abbia agito ciascuna delle differenti n cause C_i (e quindi concludere investigativamente per la causa che rende massima la probabilità a posteriori).

Osservazione 51. Il confronto tra più cause dipende esclusivamente dal numeratore, essendo il denominatore una costante di normalizzazione (il denominatore fa sì che la somma delle probabilità a posteriori sia 1, ma assume un valore comune per tutte le cause considerate).
Per questo si scrive anche che:

$$\mathrm{P}\left(C_j|E\right) \propto \mathrm{P}\left(C_j\right)\mathrm{P}\left(E|C_j\right) \tag{3.13}$$

ossia la probabilità a posteriori della causa j-esima è proporzionale alla probabilità a priori, moltiplicata per la verosimiglianza.

Osservazione 52. Il problema, nelle applicazioni pratiche, è che non sempre si riesce a pervenire a valutazioni incontrovertibili delle probabilità a priori.

Esempio 3.2.1 (Moneta bilanciata)**.** Abbiamo una moneta bilanciata e una sbilanciata che cade su testa con probabilità $3/4$. Si sceglie una moneta a caso e la si lancia tre volte; restituisce testa tutte e tre le volte. Quale è la probabilità che la moneta scelta sia quella bilanciata?
Se H è l'evento "testa tre volte" e B è l'evento "scelta la moneta bilanciata"; siamo interessati alla probabilità $\mathrm{P}\left(B|H\right)$. Ci risulta tuttavia più semplice trovare $\mathrm{P}\left(H|B\right)$ e $\mathrm{P}\left(H|\overline{B}\right)$ dato che aiuta sapere quale moneta consideriamo per calcolare la probabilità di tre teste. Questo suggerisce l'utilizzo del teorema di Bayes e della legge delle probabilità totali. Si ha

$$\begin{aligned}
\mathrm{P}\left(B|H\right) &= \frac{\mathrm{P}\left(B\right)\cdot\mathrm{P}\left(H|B\right)}{\mathrm{P}\left(B\right)\cdot\mathrm{P}\left(H|B\right)+\mathrm{P}\left(\overline{B}\right)\cdot\mathrm{P}\left(H|\overline{B}\right)} \\
&= \frac{(1/2)\cdot(1/2)^3}{(1/2)\cdot(1/2)^3+(1/2)\cdot(3/4)^3} \\
&\approx 0.23
\end{aligned}$$

Esempio 3.2.2 (Test di una malattia rara)**.** Un paziente è testato per una malattia che colpisce l'1% della popolazione. Sia D l'evento che "il paziente ha la malattia" e T il test è positivo (ossia suggerisce che il paziente abbia la malattia). Il paziente sottoposto al test risulta effettivamente positivo. Supponendo che il test sia accurato al 95%, ossia che $\mathrm{P}\left(T|D\right) = 0.95$ (la sensitività) ma anche che $\mathrm{P}\left(\overline{T}|\overline{D}\right) = 0.95$ (la specificità), qual è la probabilità che il paziente abbia effettivamente la malattia data la positività del test?
Applicando la formula di Bayes:

$$\begin{aligned}
\mathrm{P}\left(D|T\right) &= \frac{\mathrm{P}\left(D\right)\mathrm{P}\left(T|D\right)}{\mathrm{P}\left(T\right)} \\
&= \frac{0.01\cdot 0.95}{\mathrm{P}\left(T\right)}
\end{aligned}$$

$\mathrm{P}(T)$ non è così facile da ottenere (necessiterebbe di provare il test su tutta la popolazione), ma il teorema delle probabilità totali ci viene in soccorso:

$$\begin{aligned}\mathrm{P}(D|T) &= \frac{0.01 \cdot 0.95}{\mathrm{P}(D)\,\mathrm{P}(T|D) + \mathrm{P}\left(\overline{D}\right)\mathrm{P}\left(T|\overline{D}\right)} \\ &= \frac{0.01 \cdot 0.95}{0.01 \cdot 0.95 + 0.99 \cdot 0.05} \\ &\approx 0.16\end{aligned}$$

Pertanto vi è il 16% di probabilità che il paziente sia malato, anche se il test è positivo e lo strumento è affidabile: il fatto è che la malattia è estremamente rara e potrebbe essere un falso positivo, ossia un errore del test applicato (nella maggioranza dei casi) ad individui negativi.
Un altro modo conveniente era utilizzare 3.1.3 per il calcolo dell'odds ratio (e poi la 3.6 per passare a probabilità), evitando di dover utilizzare il teorema delle probabilità totali:

$$\frac{\mathrm{P}(D|T)}{\mathrm{P}\left(\overline{D}|T\right)} = \frac{\mathrm{P}(D)\,\mathrm{P}(T|D)}{\mathrm{P}\left(\overline{D}\right)\mathrm{P}\left(T|\overline{D}\right)} = \frac{0.01}{0.99} \cdot \frac{0.95}{0.05} \approx 0.19$$

da cui applicando la 3.6 si ha:

$$\mathrm{P}(D|T) \approx 0.19/(1 + 0.19) \approx 0.16$$

3.3 Probabilità condizionata, sviluppi

3.3.1 È una probabilità

Osservazione 53. Quando condizioniamo su un evento F, aggiorniamo la nostra idea per essere coerente con questta conoscenza, ponendoci in un universo dove sappiamo che F è accaduto.
Entro questo nuovo universo, tuttavia, le leggi della probabilità funzionano come in precedenza dato che le probabilità condizionate sono probabilità a tutti gli effetti.

Proposizione 3.3.1. *La probabilità condizionata è una valida funzione di probabilità a tutti gli effetti in quanto rispetta gli assiomi di Kolmogorov. Si ha:*

$$\begin{aligned}&0 \leq \mathrm{P}(E|F) \leq 1 \\ &\mathrm{P}(\Omega|F) = 1 \\ &\mathrm{P}\left(\bigcup_{i=1}^{\infty} E_i|F\right) = \sum_{i=1}^{\infty} \mathrm{P}(E_i|F) \quad \text{se } E_i \cap E_j = \emptyset, \forall i \neq j\end{aligned}$$

Dimostrazione. Per la prima dobbiamo mostrare che:

$$0 \leq \frac{\mathrm{P}(E \cap F)}{\mathrm{P}(F)} \leq 1$$

La prima disuguaglianza è ovvia, mentre la seconda discende dal fatto che $(E \cap F) \subseteq F$, da cui $\mathrm{P}(E \cap F) \leq \mathrm{P}(F)$.

La seconda segue dalla:

$$\mathrm{P}(\Omega|F) = \frac{\mathrm{P}(\Omega \cap F)}{\mathrm{P}(F)} = \frac{\mathrm{P}(F)}{\mathrm{P}(F)} = 1$$

Per la terza

$$\begin{aligned}
\mathrm{P}\left(\bigcup_{i=1}^{\infty} E_i | F\right) &= \frac{\mathrm{P}\left(\left(\bigcup_{i=1}^{\infty} E_i\right) \cap F\right)}{\mathrm{P}(F)} && \text{applicata la def. di } \mathrm{P}(A|B)\text{; per la prop. distributiva, poi} \ldots \\
&= \frac{\mathrm{P}\left(\bigcup_{i=1}^{\infty} (E_i \cap F)\right)}{\mathrm{P}(F)} && \ldots \text{ma dato che si tratta di unione eventi disgiunti} \\
&= \frac{\sum_{i=1}^{\infty} \mathrm{P}(E_i \cap F)}{\mathrm{P}(F)} && \text{e portando il denominatore sotto sommatoria} \\
&= \sum_{i=1}^{\infty} \mathrm{P}(E_i|F)
\end{aligned}$$

□

Osservazione 54 (Notazione). A volte si vuole esprimere compattamente la probabilità condizionata di un evento E condizionata al verificarsi di un altro evento F. Per farlo definiamo

$$\widetilde{\mathrm{P}}(E) = \mathrm{P}(E|F)$$

Osservazione 55. Pertanto si ha che ogni probabilità condizionata è una probabilità. Allo stesso modo *tutte le probabilità possono essere pensate come probabilità condizionate.* Vi è sempre qualche informazione di fondo sulla quale condizioniamo anche se non esplicitata. Quando scriviamo pertanto $\mathrm{P}(A)$ stiamo pensando a $\mathrm{P}(A|K)$ con K background knowledge.

3.3.2 Risultati

Osservazione 56. Il fatto che, in seguito a 3.3.1, la probabilità condizionata sia una funzione di probabilità a tutti gli effetti, fa sì che tutti i risultati sviluppati in precedenza (per la probabilità non condizionata) valgano anche per la probabilità condizionata.
Possiamo aggiornare tutti i risultati visti in precedenza aggiungendo F a destra della barra di condizionamento. Ne mostriamo alcuni.

Lemma 3.3.2.

$$\widetilde{\mathrm{P}}\left(\overline{A}\right) = 1 - \widetilde{\mathrm{P}}(A) \tag{3.14}$$

Dimostrazione. Infatti

$$\begin{aligned}
1 - \widetilde{\mathrm{P}}(A) &= 1 - \mathrm{P}(A|F) = 1 - \frac{\mathrm{P}(A \cap F)}{\mathrm{P}(F)} = \frac{\mathrm{P}(F) - \mathrm{P}(A \cap F)}{\mathrm{P}(F)} = \frac{\mathrm{P}\left(\overline{A} \cap F\right)}{\mathrm{P}(F)} \\
&= \mathrm{P}\left(\overline{A}|F\right) = \widetilde{\mathrm{P}}\left(\overline{A}\right)
\end{aligned}$$

□

Lemma 3.3.3 (Probabilità dell'unione e principio di inclusione/esclusione). *Si ha*

$$\widetilde{\mathrm{P}}(A \cup B) = \widetilde{\mathrm{P}}(A) + \widetilde{\mathrm{P}}(B) - \widetilde{\mathrm{P}}(A \cap B)$$

o equivalentemente

$$\mathrm{P}(A \cup B|F) = \mathrm{P}(A|F) + \mathrm{P}(B|F) - \mathrm{P}(A \cap B|F)$$

Lemma 3.3.4 (Condizionamento ulteriore). *La probabilità condizionata $A|B$ dove B è un nuovo condizionamento e F è già presente/sottointeso si sviluppa come*

$$\widetilde{\mathrm{P}}(A|B) = \frac{\widetilde{\mathrm{P}}(A \cap B)}{\widetilde{\mathrm{P}}(B)} = \frac{\mathrm{P}(A \cap B|F)}{\mathrm{P}(B|F)} = \frac{\dfrac{\mathrm{P}(A \cap B \cap F)}{\mathrm{P}(F)}}{\dfrac{\mathrm{P}(B \cap F)}{\mathrm{P}(F)}} = \mathrm{P}(A|B \cap F)$$

Lemma 3.3.5 (Regola di Bayes con condizionamento ulteriore). *A patto che* $\mathrm{P}(A \cap F) > 0$ *e* $\mathrm{P}(B \cap F) > 0$ *si ha*

$$\widetilde{\mathrm{P}}(A|B) = \frac{\widetilde{\mathrm{P}}(A) \cdot \widetilde{\mathrm{P}}(B|A)}{\widetilde{\mathrm{P}}(B)} = \frac{\mathrm{P}(A|F) \cdot \mathrm{P}(B|A \cap F)}{\mathrm{P}(B|F)}$$

Lemma 3.3.6 (Odds ratio con condizionamento ulteriore). *Si ha:*

$$\frac{\widetilde{\mathrm{P}}(A|B)}{\widetilde{\mathrm{P}}(\overline{A}|B)} = \frac{\mathrm{P}(A|B \cap F)}{\mathrm{P}(\overline{A}|B \cap F)} = \frac{\mathrm{P}(A|F) \cdot \mathrm{P}(B|A \cap F)}{\mathrm{P}(\overline{A}|F) \cdot \mathrm{P}(B|\overline{A} \cap F)} \tag{3.15}$$

Lemma 3.3.7 (Teorema delle probabilità totali 1). *La probabilità condizionata dell'evento E può essere spezzata come somma delle probabilità di eventi incompatibili, analogamente a quanto fatto in 3.9*

$$\widetilde{\mathrm{P}}(E) = \widetilde{\mathrm{P}}(C)\,\widetilde{\mathrm{P}}(E|C) + \widetilde{\mathrm{P}}(\overline{C})\,\widetilde{\mathrm{P}}(E|\overline{C})$$

ossia, equivalentemente

$$\mathrm{P}(E|F) = \mathrm{P}(C|F)\,\mathrm{P}(E|C \cap F) + \mathrm{P}(\overline{C}|F)\,\mathrm{P}(E|\overline{C} \cap F)$$

Lemma 3.3.8 (Teorema delle probabilità totali (versione generica)). *Se $C_1, \dots, C_n$ è una partizione di Ω e nell'ipotesi che* $\mathrm{P}(C_i \cap F) > 0$ *per ogni i, allora analogamente a 3.10 si ha*

$$\widetilde{\mathrm{P}}(E) = \mathrm{P}(E|F) = \sum_{i=1}^{n} \mathrm{P}(C_i|F) \cdot \mathrm{P}(E|C_i \cap F)$$

Esempio 3.3.1 (Moneta bilanciata 2). Riprendendo l'esempio 3.2.1, supponiamo di aver visto la moneta uscire testa tre volte. Se la rilanciamo quale è la probabilità che esca testa una volta ancora?
Sia H l'evento testa tre volte, e T esce testa anche la quarta volta. Siamo interessati a $\mathrm{P}(T|H)$; la legge delle probabilità totali ci permette di scriverla come media ponderata dei condizionamenti su B (scelta la moneta bilanciata)

$$\begin{aligned}\mathrm{P}(T|H) &= \mathrm{P}(B|H)\,\mathrm{P}(T|B \cap H) + \mathrm{P}(\overline{B}|H)\,\mathrm{P}(T|\overline{B} \cap H) \\ &= 0.23 \cdot \frac{1}{2} + (1 - 0.23) \cdot \frac{3}{4} \\ &\approx 0.69\end{aligned}$$

con $\mathrm{P}(B|H) = 0.23$ come derivato in esempio 3.2.1.

3.3.3 Condizionare su più eventi

Spesso si vuole condizionare su più eventi/informazioni, ora abbiamo vari modi per farlo. Ipotizzando di essere interessati a $\mathrm{P}(A|B \cap C)$, ossia di voler condizionare a sia B che C:

- possiamo utilizzare la definizione di probabilità condizionata
$$\mathrm{P}(A|B \cap C) = \frac{\mathrm{P}(A \cap B \cap C)}{\mathrm{P}(B \cap C)}$$
- possiamo utilizzare la regola di Bayes condizionando ulteriormente su C (questo è l'approccio naturale se pensiamo che ogni evento nel nostro problema sia condizionato su C)
$$\mathrm{P}(A|B \cap C) = \frac{\mathrm{P}(A|C) \cdot \mathrm{P}(B|A \cap C)}{\mathrm{P}(B|C)}$$
- viceversa utilizzare la regola di Bayes condizionando ulteriormente su B (questo è l'approccio naturale se pensiamo che ogni evento nel nostro problema sia condizionato su B)
$$\mathrm{P}(A|B \cap C) = \frac{\mathrm{P}(A|B) \cdot \mathrm{P}(C|A \cap B)}{\mathrm{P}(C|B)}$$

3.4 Eventi indipendenti

Osservazione 57. Abbiamo visto diversi esempi dove il condizionamento su un evento cambia la probabilità di un altro evento. Viceversa la situazione dove gli eventi non forniscono informazioni uno dell'altro è chiamata indipendenza.

3.4.1 Indipendenza di 2 eventi

Definizione 3.4.1 (Eventi indipendenti)**.** Due eventi A, B sono indipendenti se e solo se:

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B) \tag{3.16}$$

Osservazione 58. Intuitivamente sono indipendenti se il loro verificarsi congiunto avviene quando si realizzano separatamente (senza influenze reciproche) i due eventi componenti.

Osservazione 59. L'indipendenza è una relazione simmetrica: se A è indipendente da B, allora B è indipendente da A.

Definizione 3.4.2 (Eventi dipendenti)**.** Due eventi che non sono indipendenti si dicono dipendenti.

Proposizione 3.4.1 (Probabilità condizionata di eventi indipendenti)**.** *Se A e B sono due eventi indipendenti*

$$\boxed{\mathrm{P}(A|B) = \mathrm{P}(A)} \tag{3.17}$$

e specularmente

$$\mathrm{P}(B|A) = \mathrm{P}(B)$$

Dimostrazione. A titolo esemplificativo per la prima:

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A)\,\mathrm{P}(B)}{\mathrm{P}(B)} = \mathrm{P}(A)$$

□

Osservazione 60 (Indipendenza e incompatibilità). I concetti di indipendenza e incompatibilità sono molto differenti tra loro:

- l'incompatibilità è una *relazione tra eventi*, ed è rappresentata sui diagrammi di Venn come (ad esempio) due sottoinsiemi di Ω che non hanno aree di sovrapposizione);
- l'indipendenza è una *relazione tra probabilità di eventi*; poiché sui diagrammi di Venn si rappresentano gli eventi, e non le probabilità, l'indipendenza non è rappresentabile graficamente.

In generale, indipendenza e incompatibilità non hanno alcun legame tra loro, tranne nel caso seguente.

Proposizione 3.4.2. *Se due eventi A, B incompatibili hanno probabilità positive, allora non possono esser indipendenti, e viceversa.*

Dimostrazione. Se A e B sono incompatibili deve esser:

$$\mathrm{P}(A \cap B) = \mathrm{P}(\emptyset) = 0$$

Ma se fossero anche indipendenti dovrebbe esser che:

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B) > 0$$

dato che per ipotesi $\mathrm{P}(A), \mathrm{P}(B) > 0$; ma questo contraddirrebbe il risultato precedente. Viceversa se la probabilità dell'intersezione è positiva (indipendenti) non può essere nulla (incompatibili). □

Proposizione 3.4.3 (Indipendenza e negazioni)**.** *Siano E e F due eventi indipendenti; allora E e $\overline{F}$ sono indipendenti. Allo stesso modo avviene per $\overline{E}$ e F, nonché per $\overline{E}$ e $\overline{F}$.*

Dimostrazione. Dimostriamo la prima. Supponiamo che E ed F siano indipendenti; si ha che

$$\mathrm{P}(E \cap F) = \mathrm{P}(E)\,\mathrm{P}(F)$$

vogliamo dimostrare che

$$\mathrm{P}\left(E \cap \overline{F}\right) = \mathrm{P}(E)\,\mathrm{P}\left(\overline{F}\right)$$

Dato che possiamo spezzare $E = (E \cap F) \cup (E \cap \overline{F})$ e che $(E \cap F)$ è disgiunto da $(E \cap \overline{F})$ la probabilità di E può essere riscritta come una somma:

$$\mathrm{P}(E) = \mathrm{P}(E \cap F) + \mathrm{P}\left(E \cap \overline{F}\right)$$

da cui

$$\begin{aligned}
\mathrm{P}\left(E \cap \overline{F}\right) &= \mathrm{P}(E) - \mathrm{P}(E \cap F) \\
&= \mathrm{P}(E) - \mathrm{P}(E)\,\mathrm{P}(F) \\
&= \mathrm{P}(E)\,[1 - \mathrm{P}(F)] \\
&= \mathrm{P}(E)\,\mathrm{P}\left(\overline{F}\right)
\end{aligned}$$

Per $\overline{E}$ e F, nonché per $\overline{E}$ e $\overline{F}$ basta invertire i ruoli (evento è il non evento dell'evento considerato in precedenza) dei casi considerati. □

3.4.2 Indipendenza di 3+ eventi

Osservazione 61. Per definire l'indipendenza di tre eventi E, F, G sembra opportuno non limitarsi alla richiesta di indipendenza di tutte le $\binom{3}{2}$ coppie di eventi, ma richiedere anche l'indipendenza dei tre eventi considerati congiuntamente.

Definizione 3.4.3 (Indipendenza di 3 eventi)**.** Tre eventi E, F, G si dicono mutuamente indipendenti se:

$$\begin{aligned}
\mathrm{P}(E \cap F) &= \mathrm{P}(E)\,\mathrm{P}(F) \\
\mathrm{P}(E \cap G) &= \mathrm{P}(E)\,\mathrm{P}(G) \\
\mathrm{P}(F \cap G) &= \mathrm{P}(F)\,\mathrm{P}(G) \\
\mathrm{P}(E \cap F \cap G) &= \mathrm{P}(E)\,\mathrm{P}(F)\,\mathrm{P}(G)
\end{aligned}$$

Definizione 3.4.4 (Indipendenza a coppie)**.** Tre eventi E, F, G si dicono indipendenti a coppie se valgono le prime tre condizioni di cui sopra.

Osservazione 62. L'indipendenza a coppie non è sufficiente per avere indipendenza.

Esempio 3.4.1. Lanciando due monete, se A è "prima moneta testa", B "seconda moneta testa" e C i due lanci danno lo stesso risultato. Si ha $\mathrm{P}(A) = \mathrm{P}(B) = \mathrm{P}(C) = 1/2$. A, B, C sono indipendenti a coppie in quanto $\mathrm{P}(A|B) = \mathrm{P}(A|C) = \mathrm{P}(B|C) = 1/4$; tuttavia non sono indipendenti perché $\mathrm{P}(A \cap B \cap C) = 1/4$ mentre $\mathrm{P}(A) = \mathrm{P}(B) = \mathrm{P}(C) = 1/8$.
Il punto è che sapere cosa è sucesso sia con A che con B determina/ci da informazione completa su C.

Osservazione 63. Se E, F, G sono indipendenti, allora E è indipendente da ogni evento che si formi a partire da F e G (mediante unioni, negazioni, intersezioni).

Esempio 3.4.2. E è indipendente da $F \cup G$ essendo

$$\begin{aligned}
\mathrm{P}(E \cap (F \cup G)) &= \mathrm{P}((E \cap F) + (E \cap G)) \\
&= \mathrm{P}(E \cap F) + \mathrm{P}(E \cap G) - \mathrm{P}(E \cap F \cap G) \\
&= \mathrm{P}(E)\,\mathrm{P}(F) + \mathrm{P}(E)\,\mathrm{P}(G) - \mathrm{P}(E)\,\mathrm{P}(F \cap G) \\
&= \mathrm{P}(E)\,[\mathrm{P}(F) + \mathrm{P}(G) - \mathrm{P}(F \cap G)] \\
&= \mathrm{P}(E)\,\mathrm{P}(F \cup G)
\end{aligned}$$

Definizione 3.4.5 (Indipendenza di m eventi)**.** n eventi $A_1, \ldots, A_n \subset \Omega$ si dicono mutuamente indipendenti se per ogni sottogruppo di m eventi, $1 < m \leq n$, si ha:

$$\mathrm{P}\left(\bigcap_{i=1}^{m} A_i\right) = \prod_{i=1}^{m} \mathrm{P}(A_i) \tag{3.18}$$

Osservazione 64. In generale, l'indipendenza ad n ad n implica quella ad $n-1$ ad $n-1$, ma non viceversa (ad esempio l'indipendenza a due a due non implica quella a tre a tre).

Definizione 3.4.6 (Indipendenza di ∞ eventi)**.** Diciamo che infiniti eventi sono indipendenti se ogni sottoinsieme finito di essi è formato da eventi indipendenti.

3.4.3 Indipendenza condizionata, aggiornamento delle stime

Definizione 3.4.7 (Indipendenza condizionata)**.** Gli eventi A e B sono indipendenti condizionatamente dato l'evento F se

$$\mathrm{P}(A \cap B|F) = \mathrm{P}(A|F) \cdot \mathrm{P}(B|F) \tag{3.19}$$

Osservazione 65. Attenzione, due eventi:

- possono essere indipendenti condizionatamente (dato F), ma non indipendenti;
- possono essere indipendenti, ma non indipendenti condizionatamente (dato F);
- possono essere indipendenti condizionatamente dato F ma non dato $\overline{F}$.

Lo vediamo nei seguenti esempi.

Esempio 3.4.3 (Eventi indipendenti condizionatamente ma non indipendenti)**.** Tornando al setup di esempio 3.2.1, sia F "ho scelto la moneta bilanciata", A_1 "primo lancio da testa" e A_2 "secondo lancio da testa". Condizionatamente a F, A_1 e A_2 sono indipendenti; ma A_1 e A_2 non sono indipendenti da soli perché A_1 fornisce informazioni su A_2.

Esempio 3.4.4 (Eventi indipendenti ma non condizionatamente)**.** Siano Alice e Bob sono le uniche due persone che mi telefonano; ogni giorno decidono indipendentemente se farlo e sia A "mi chiama Alice", B "mi chiama Bob". Questi sono eventi indipendenti. Ma supponendo che R "il telefono squilla", condizionatamente a questo A e B non sono più indipendenti, perché se non è Alice deve essere Bob, ossia

$$\mathrm{P}(B|R) < 1 = \mathrm{P}\left(B|\overline{A} \cap R\right)$$

per cui B e $\overline{A}$ non sono condizionalmente indipendenti dato R (e allo stesso modo A e B)

Esempio 3.4.5. Supponendo che vi siano solo due tipi di classi: classi buone dove se si lavora tanto si prendono buoni voti e classi cattive dove il professore assegna voti a caso. Sia G "classe è buona", W "si lavora tanto" e A "si prende un bel voto". Allora W, A sono indipendenti condizionatamente a $\overline{G}$, ma non lo sono dato G.

Esempio 3.4.6 (Aggiornamento delle stime (e indipendenza condizionale))**.** Riprendendo l'esempio 3.2.2 sul test della malattia rara, ipotizziamo che il paziente decida di intraprendere un secondo test; questo è indipendente dal primo test effettuato (condizionatamente allo stato di malattia) e ha la stessa sensibilità e specificità. Il paziente risulta positivo per la seconda volta. Come si aggiorna la sua probabilità di essere effettivamente malato?
Siamo interessati a $\widetilde{\mathrm{P}}\left(D|T_2\right)$, condizionata a T_1, dove D è essere malato, T_1 è essere risultati positivi al primo test e T_2 al secondo. Utilizziamo la forma per l'odds ratio per ricondurci in secondo luogo alla probabilità; si ha

$$\begin{aligned}
\frac{\widetilde{\mathrm{P}}\left(D|T_2\right)}{\widetilde{\mathrm{P}}\left(\overline{D}|T_2\right)} &= \frac{\mathrm{P}\left(D|T_1 \cap T_2\right)}{\mathrm{P}\left(\overline{D}|T_1 \cap T_2\right)} = \frac{\mathrm{P}\left(D\right) \cdot \mathrm{P}\left(T_1 \cap T_2|D\right)}{\mathrm{P}\left(\overline{D}\right) \cdot \mathrm{P}\left(T_1 \cap T_2|\overline{D}\right)} \\
&= \frac{\mathrm{P}\left(D\right) \cdot \mathrm{P}\left(T_1|D\right) \cdot \mathrm{P}\left(T_2|D\right)}{\mathrm{P}\left(\overline{D}\right) \cdot \mathrm{P}\left(T_1|\overline{D}\right) \cdot \mathrm{P}\left(T_2|\overline{D}\right)} = \boxed{\frac{\mathrm{P}\left(D|T_1\right)}{\mathrm{P}\left(\overline{D}|T_1\right)} \cdot \frac{\mathrm{P}\left(T_2|D\right)}{\mathrm{P}\left(T_2|\overline{D}\right)}} \\
&= 0.19 \cdot \frac{0.95}{0.05} \approx 3.646
\end{aligned}$$

Di particolare interesse è la seconda riga dove, in contesto di indipendenza condizionata, si vede che aggiorniamo i risultati cui eravamo giunti in precedenza mediante le informazioni sul nuovo test. Passiamo alla probabilità seguendo la consueta formula

$$\mathrm{P}\left(D|T_1 \cap T_2\right) = \frac{3.646}{1 + 3.646} = 0.78$$

La probabilità di essere malati in seguito ad un secondo test positivo (indipendente condizionalmente) aumenta molto, da 0.16 a 0.78.

Esempio 3.4.7 (Calcolo diretto della probabilità)**.** Volendo invece calcolare direttamente la probabilità in un colpo solo si applica Bayes e torna comodo il teorema delle probabilità totali condizionando su D:

$$\begin{aligned}
\mathrm{P}\left(D|T_1 \cap T_2\right) &= \frac{\mathrm{P}\left(D\right) \cdot \mathrm{P}\left(T_1 \cap T_2|D\right)}{\mathrm{P}\left(T_1 \cap T_2\right)} \\
&= \frac{\mathrm{P}\left(D\right) \cdot \mathrm{P}\left(T_1 \cap T_2|D\right)}{\mathrm{P}\left(D\right) \cdot \mathrm{P}\left(T_1 \cap T_2|D\right) + \mathrm{P}\left(\overline{D}\right) \cdot \mathrm{P}\left(T_1 \cap T_2|\overline{D}\right)} \\
&= \frac{\mathrm{P}\left(D\right) \cdot \mathrm{P}\left(T_1|D\right) \cdot \mathrm{P}\left(T_2|D\right)}{\mathrm{P}\left(D\right) \cdot \mathrm{P}\left(T_1|D\right) \cdot \mathrm{P}\left(T_2|D\right) + \mathrm{P}\left(\overline{D}\right) \cdot \mathrm{P}\left(T_1|\overline{D}\right) \cdot \mathrm{P}\left(T_2|\overline{D}\right)} \\
&= \frac{0.01 \cdot 0.95 \cdot 0.95}{0.01 \cdot 0.95 \cdot 0.95 + 0.99 \cdot 0.05 \cdot 0.05} = 0.78
\end{aligned}$$

Soffermandoci un attimo sulla equazione prima del calcolo dell'ultima riga, se dividiamo algebricamente per $\mathrm{P}\left(T_1\right)$ sia numeratore che denominatore si ottiene:

$$\begin{aligned}
\mathrm{P}\left(D|T_1 \cap T_2\right) =& \frac{\mathrm{P}\left(D|T_1\right) \cdot \mathrm{P}\left(T_2|D\right)}{\mathrm{P}\left(D|T_1\right) \cdot \mathrm{P}\left(T_2|D\right) + \mathrm{P}\left(\overline{D}|T_1\right) \cdot \mathrm{P}\left(T_2|\overline{D}\right)} \\
&= \frac{0.16 \cdot 0.95}{0.16 \cdot 0.95 + 0.84 \cdot 0.05} \approx 0.78
\end{aligned}$$

che equivale ad un normale teorema di Bayes dove al posto delle probabilità a priori secca $\mathrm{P}\left(D\right)$ che avevamo utilizzato in esempio 3.2.2, abbiamo sostituito i risultati disponibili alla fine del primo test, ossia $\mathrm{P}\left(D|T_1\right) = 0.16$ e

$\mathrm{P}\left(\overline{D}|T_1\right) = 1 - 0.16 = 0.84$; come si nota l'unica cosa che cambia nella formula (anche perché T_1 e T_2 performano allo stesso modo), sono tali parti, evidenziate in rosso. Aggiorniamo dunque i risultati al termine del primo test con le informazioni del secondo test, per arrivare alla probabilità a posteriori $\mathrm{P}\left(D|T_1 \cap T_2\right)$. Seguendo questa impostazione, è facile generalizzare ad n test applicando ripetutamente il teorema.

Capitolo 4

Variabili casuali

4.1 Introduzione

4.1.1 Variabile casuale e suo supporto

Definizione 4.1.1 (Variabile casuale)**.** Dato un esperimento con spazio campionario Ω una variabile casuale X è una funzione $X : \Omega \to \mathbb{R}$ che assegna ad ogni esito dell'esperimento $s \in \Omega$ un numero $X(s)$.

Esempio 4.1.1. L'esperimento "Lancio di due monete" ha $\Omega = \{TT, TC, CT, CC\}$. Su questo esperimento possiamo definire (tra le altre) la variabile casuale X "numero di teste", come segue:

$$X(TT) = 2;\ X(TC) = 1;\ X(CT) = 1;\ X(CC) = 0;$$

Osservazione 66. Nel prosieguo abbreviamo "variabil casuale" con vc e "variabili casuali" con vcs.

Osservazione 67. Pertanto una variabile casuale X assegna un valore numerico $X(s)$ a ciascun esito possibile s dell'esperimento: la randomicità deriva dal fatto di avere a che fare con un esperimento casuale, mentre il mapping effettuato dalla vc è deterministico.

Definizione 4.1.2 (Supporto di una vc)**.** Indicato con R_X è l'immagine $X(\Omega)$ della variabile casuale X o, in altre parole l'insieme dei suoi esiti possibili. È indicato in maniera estensiva come $R_X = \{x_1, x_2, \dots\}$

Esempio 4.1.2. Nell'esempio 4.1.1, $R_X = \{0, 1, 2\}$.

Definizione 4.1.3 (Vc indicatrice)**.** Una variabile I che assegna il valore 1 al valore l'evento E si verifica e il valore 0 a l'evento $\overline{E}$ si verifica è detta vc indicatrice dell'evento E.

4.1.2 Variabili casuali discrete e continue

Definizione 4.1.4 (vc discreta)**.** La cui cardinalità del supporto è finita o infinita numerabile (in biezione con $\mathbb{N}$.)

Esempio 4.1.3. Il numero di teste nel lancio di due monete è discreta in quanto $\text{Card}\,(R_X) = |\{0, 1, 2\}| = 3$.

Definizione 4.1.5 (vc continua)**.** La cui cardinalità del supporto è infinito non numerabile (in biezione con $\mathbb{R}$).

Esempio 4.1.4. Il tempo in minuti T impiegato ad una lampadina per rompersi è una vc continua in quanto $R_T = \{t \in \mathbb{R} : t > 0\}$ (ipotizzando di avere un cronometro dalla precisione infinita, che non si limiti ai secondi).

4.2 Funzioni su variabili casuali

Osservazione 68. Vogliamo poter descrivere il funzionamento della variabile casuale utilizzando il linguaggio della probabilità: ad esempio qual è la probabilità che nel lancio di due monete (esempio 4.1.1) il numero di teste X sia 1? E almeno 1? Qual è la probabilità che la lampadina si rompa entro i primi 10 minuti di utilizzo.

4.2.1 Variabili discrete (PMF, CDF)

Definizione 4.2.1 (PMF di vc discreta (Funzione di massa di probabilità))**.** Data una vc discreta $X : \Omega \to \mathbb{R}$ la PMF è una funzione $p : \mathbb{R} \to \mathbb{R}$ che associa ad ogni possibile esito della variabile casuale la probabilità con cui esso si verifichi

$$p_X(x) = \mathrm{P}\left(X = x\right) = \begin{cases} \mathrm{P}\left(X(s) = x\right) & \text{se } x \in X(\Omega) \\ 0 & \text{se } x \in \mathbb{R} \setminus X(\Omega) \end{cases} \tag{4.1}$$

Esempio 4.2.1. Per esempio 4.1.1 si ha

$$\begin{aligned} p_X(X = 0) &= 1/4 \\ p_X(X = 1) &= 1/2 \\ p_X(X = 2) &= 1/4 \end{aligned}$$

e $p_X(x) = 0$ per $x \notin \{0, 1, 2\}$.

Proposizione 4.2.1 (PMF valida)**.** *Sia X una vc discreta con supporto $X(\Omega) = \{x_1, x_2, \ldots\}$. Una corretta PMF p_X deve soddisfare i seguenti criteri:*

$$p_X(x) \geq 0, \quad \forall x \in \mathbb{R} \tag{4.2}$$

$$\sum_{x \in \mathbb{R}} p_X(x) = 1 \tag{4.3}$$

Dimostrazione. Il primo criterio deve esser valido dato che la probabilità è non negativa. Il secondo deve essere valido dato che gli eventi $X = x_1, X = x_2, \ldots$ sono disgiunti e X dovrà assumere pur qualche valore:

$$\begin{aligned} \sum_{x \in \mathbb{R}} p_X(x) = \sum_{x \in X(\Omega)} p_X(x) = \sum_j \mathrm{P}\left(X = x_j\right) = \mathrm{P}\left(\bigcup_j \{X = x_j\}\right) \\ = \mathrm{P}\left(X = x_1 \text{ or } X = x_2 \ldots\right) = 1 \end{aligned}$$

□

Definizione 4.2.2 (CDF di vc discreta (Funzione di ripartizione))**.** Data una vc X è definita come:

$$F_X(x) = \mathrm{P}\left(X \leq x\right) = \sum_{x_j \in X(\Omega): x_j \leq x} p_X(x_j) \tag{4.4}$$

Osservazione 69 (Forma della funzione). Se X è una vc discreta, $F_X(x)$ è una funzione a gradini e cresce un numero finito o numerabile di volte, in corrispondenza dei valori $x_1, x_2, \ldots$ assunti dalla vc discreta: il gradino è esattamente pari alla probabilità $p_X(x_1), p_X(x_2), \ldots$ con le quali la vc discreta assume quei valori.

Proposizione 4.2.2 (CDF valida)**.** *Sia X una vc discreta con supporto $X(\Omega) = \{x_1, x_2, \ldots\}$. Una corretta CDF F_X deve soddisfare i seguenti criteri:*

$$x_1 \leq x_2 \implies F_X(x_1) \leq F_X(x_2) \tag{4.5}$$

$$\lim_{x \to x_j^+} F_X(x) = F_X(x_j) \quad \textit{(continuità da destra)} \tag{4.6}$$

$$\lim_{x \to -\infty} F_X(x) = 0, \qquad \lim_{x \to +\infty} F_X(x) = 1 \tag{4.7}$$

Dimostrazione. La prima è giustificata dal fatto che dato che, dato che l'evento $\{X \leq x_1\}$ si verifica sempre quando si verifica $\{X \leq x_2\}$ allora $\mathrm{P}\left(X \leq x_1\right) \leq \mathrm{P}\left(X \leq x_2\right)$.
La continuità da destra deriva dall'aver definito $F_X(x_0)$ come $\mathrm{P}\left(X \leq x_0\right)$ (coerentemente con la letteratura internazionale prevalente); altri autori definiscono $F_X(x_0) = \mathrm{P}\left(X < x_0\right)$, il che implica la continuità da sinistra.
Per la terza, dato che $F_X(x_{min}) = 0$ con $x_{min} = min(x_1, x_2, \ldots)$ e $-\infty < x_{min}$ allora per la prima proprietà si ha che $F(-\infty) \leq 0$, ma non potendo una probabilità esser negativa, sarà nulla, dunque si conclude che $\lim_{x \to -\infty} F_X(x) = 0$.
Altresì sfruttando sempre il fatto che $\{X = x_j\}$ sono eventi indipendenti

$$\lim_{x \to +\infty} F_X(x) = \sum_{x_j \in X(\Omega)} p_X(x_j) = 1$$

□

Esempio 4.2.2. Dato l'esperimento lancio di due dati, l'evento X somma degli esiti ha PMF e CMF riportate in figura 4.1. Ad esempio $\mathrm{P}\left(X = 2\right) = \mathrm{P}\left(\{1, 1\}\right) = (\frac{1}{6})^2 = 1/36 \approx 0.02778$. I "salti" nella CDF sono di entità pari alla PMF

4.2.2 Variabili continue (PDF, CDF)

Osservazione 70. Nel caso continuo l'equivalente della funzione massa di probabilità e la funzione di densità, e il calcolo della probabilità di un evento (ossia che l'esito della variabile casuale appartenga ad un intervallo A) si ha mediante integrazione

Definizione 4.2.3 (PDF di una vc continua (Funzione di densità))**.** Sia X una vc continua. Si definisce funzione di densità, la funzione $f : \mathbb{R} \to \mathbb{R}$, $f_X(x)$ tale che, considerato l'esito $X \in A \subseteq \mathbb{R}$, sia:

$$\mathrm{P}\left(X \in A\right) = \int_{x \in A} f_X(x)\, dx \tag{4.8}$$

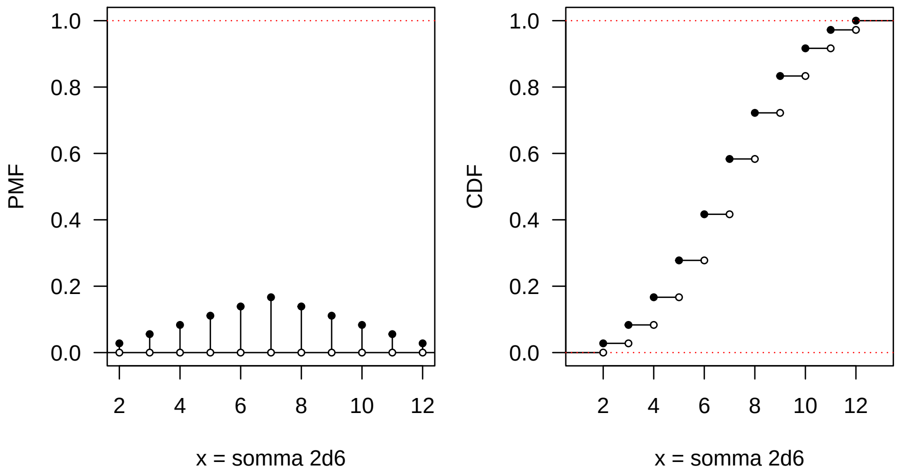


Figura 4.1: Somma del lancio di due d6

Ad esempio siano $a, b \in \mathbb{R}$, con $a < b$:

$$\mathrm{P}\left(X \in [a, b]\right) = \int_a^b f_X(x)\ dx \tag{4.9}$$

Proposizione 4.2.3 (Corretta PDF)**.** *Una corretta PDF è caratterizzata da:*

$$f_X(x) \geq 0 \tag{4.10}$$

$$\int_{-\infty}^{\infty} f_X(t)\ dt = 1 \tag{4.11}$$

Dimostrazione. Il primo criterio è necessario perché la probabilità è non negativa: se $f_X(x_0)$ fosse negativa, allora potremmo integrare su un piccolo intorno di x_0 e ottenere una probabilità negativa.
Il secondo criterio è necessario dato che la X, variabile quantitativa, deve avere un esito che sta in $\mathbb{R}$. □

Osservazione 71. A differenza del caso discreto dove una probabilità mostrata dalla PMF non può superare 1, nel caso continuo una densità può essere maggiore di 1 fino a tanto che l'integrazione su tutto $\mathbb{R}$ rimane comunque 1.

Definizione 4.2.4 (CDF di vc continua (Funzione di ripartizione))**.** Data una vc continua X, è la funzione $F : \mathbb{R} \to \mathbb{R}$ che restituisce la probabilità che la variabile casuale assuma un valore $\leq x$:

$$F_X(x) = \mathrm{P}\left(X \leq x\right) = \int_{-\infty}^{x} f_X(t)\ dt \tag{4.12}$$

Proposizione 4.2.4 (Corretta CDF)**.** *Affinché una generica funzione possa costituire una CDF di una variabile casuale, deve rispettare:*

$$x_1 \leq x_2 \implies F_X(x_1) \leq F_X(x_2) \tag{4.13}$$

$$\lim_{x \to x_0^+} F_X(x) = F_X(x_0) \quad \textit{(continuità da destra)} \tag{4.14}$$

$$\lim_{x \to -\infty} F_X(x) = 0 \qquad \lim_{x \to +\infty} F_X(x) = 1 \tag{4.15}$$

Osservazione 72 (Calcolo di una probabilità mediante CDF)*.* Se conosciamo la CDF possiamo calcolare la probabilità di un intervallo $a \leq X \leq b$, $a, b \in \mathbb{R}$ come segue:

$$\mathrm{P}(a \leq X \leq b) = \mathrm{P}(X \leq b) - \mathrm{P}(X \leq a) = F_X(b) - F_X(a)$$

Osservazione 73 (Probabilità di un singolo valore)*.* A differenza delle variabili discrete, nel caso continuo si ha che:

$$\mathrm{P}(X = a) = \int_a^a f_X(x)\, dx = F_X(a) - F_X(a) = 0$$

Intuitivamente, se vi sono infiniti esiti possibili la probabilità del singolo è nulla.

Osservazione 74 (Irrilevanza degli estremi di integrazione)*.* Siano $a, b \in \mathbb{R}$, $a < b$:

$$\mathrm{P}(X \in [a, b]) = \mathrm{P}(X \in (a, b]) = \mathrm{P}(X \in [a, b)) = \mathrm{P}(X \in (a, b)) = \int_a^b f_X(x)\, dx$$

Esempio 4.2.3 (Vc logistica)**.** La variabile casuale logistica, rappresentata in figura 4.2, e caratterizzata da:

$$F(x) = \frac{e^x}{1 + e^x}; \quad f(x) = \frac{e^x}{(1 + e^x)^2}$$

4.3 Altre funzioni utili per vc

4.3.1 Funzione indicatrice del supporto

Osservazione 75*.* Nel seguito servirà essere compatti/sicuri sul fatto che, al di fuori del supporto R_X della vc X, la probabilità/densità sia nulla. Per farlo si moltiplicherà la PMF/PDF per la funzione indicatrice applicata al supporto della variabile casuale.

Definizione 4.3.1 (Funzione indicatrice del supporto di una vc)**.** Definita come:

$$\mathbb{1}_{R_X}(x) = \begin{cases} 1 & \text{se } x \in R_X \\ 0 & \text{se } x \notin R_X \end{cases}$$

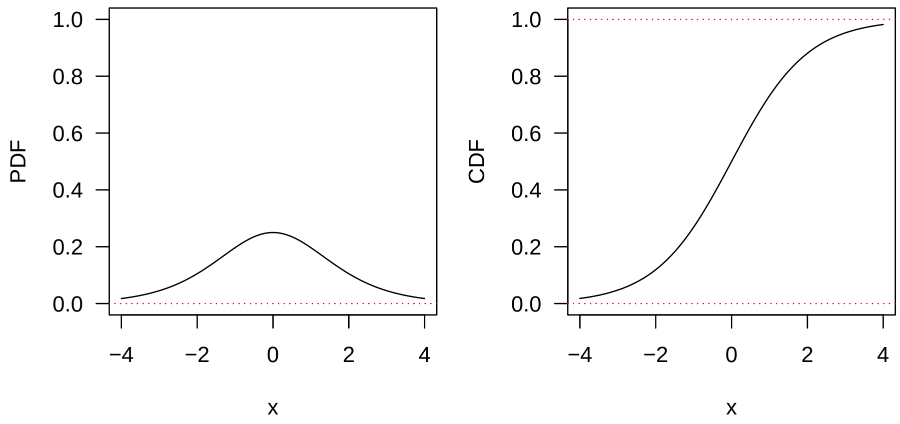


Figura 4.2: Vc logistica

4.3.2 Funzione di sopravvivenza e di azzardo

Osservazione 76. Se la vc T possiede supporto non negativo (ad esempio denota la durata di un fenomeno), allora è utile definire le seguenti funzioni (sopravvivenza per tutte, azzardo solo per variabili continue).

Definizione 4.3.2 (Funzione di sopravvivenza)**.** Per una vc T tale che $\mathrm{P}(T \geq 0) = 1$, la funzione di sopravvivenza è il complemento a 1 della funzione di ripartizione

$$S(t) = \mathrm{P}(T > t) = 1 - \mathrm{P}(T \leq t) = 1 - F_T(t) \tag{4.16}$$

Osservazione 77. Per le sole variabili positive continue T, la funzione di azzardo (o rischio), indicata con $H(x)$ e definita come la probabilità che la vc T duri un infinitesimo dopo t, essendo durata sino al tempo t.

Definizione 4.3.3 (Funzione di azzardo (o rischio))**.** Per una vc continua T tale che $\mathrm{P}(T \geq 0) = 1$, la funzione di azzardo è:

$$H(t) = \frac{f_T(t)}{1 - F_T(t)} \overset{(1)}{=} -\frac{d}{dt} \log(1 - F_T(t)) = -\frac{d}{dt} \log(S(t)) \tag{4.17}$$

Osservazione 78. In (1) l'uguaglianza è giustificata dal nota come la funzione prima coincida, cambiata di segno, con la derivata logaritmica (ossia fare la derivata del logaritmo) di $1 - F_T(t)$.

Osservazione 79. Dall'ultima relazione si ha il legame tra azzardo, sopravvivenza, ripartizione e densità. Integrando entrambi i membri per t tra $-\infty$ a x

$$\begin{aligned} H(t) &= -\frac{d}{dt} \log(S(t)) \\ \int_{-\infty}^{x} H(t)\, dt &= \int_{-\infty}^{x} -\frac{d}{dt} \log(S(t)) \\ \int_{-\infty}^{x} H(t)\, dt &= -\log(S(t)) \end{aligned}$$

Da cui:

$$\log(S(t)) = -\int_{-\infty}^{x} H(t)\,dt$$
$$S(t) = \exp\left(-\int_{-\infty}^{x} H(t)\,dt\right) \tag{4.18}$$

Mentre per quanto riguarda $F_T(t)$ e $f_T(t)$, si ha:

$$F_T(t) = 1 - \exp\left(-\int_{-\infty}^{x} H(t)\,dt\right) \tag{4.19}$$
$$f_T(t) = H(t) \cdot \exp\left(-\int_{-\infty}^{x} H(t)\,dt\right) \tag{4.20}$$

(nell'estremo inferiore di integrazione si poteva anche scrivere 0 al posto di $-\infty$).

4.4 Funzioni di variabili casuali (trasformazioni)

Definizione 4.4.1 (Funzione di una vc)**.** Dato un esperimento con spazio campionario Ω, una vc X sullo stesso, e una funzione $g : \mathbb{R} \to \mathbb{R}$, la funzione $g(X)$ è la variabile casuale che mappa $s \to g(X(s))$, $\forall s \in \Omega$.

Osservazione 80 (Cambio nel supporto)*.* Invece del supporto $X(s_1), X(s_2), \ldots$ la nuova variabile casuale $g(X)$ avrà supporto $g(X(s_1)), g(X(s_2)), \ldots$. Di fatto il supporto si ottiene mediante la composizione $s \circ X$ delle funzioni X ed s.

Osservazione 81 (Cambio nelle funzioni di riferimento)*.* Prendiamo il caso di una variabile discreta a titolo di esempio. Data una vc discreta X con una PMF conosciuta, come possiamo trovare la PMF di $Y = g(X)$? Il discorso varia a seconda che g sia iniettiva o meno.
Se g è iniettiva, ossia $X(s_1) \neq X(s_2) \implies g(X(s_1)) \neq g(X(s_2))$, allora la PMF di Y è la stessa di X:

$$\mathrm{P}(Y = g(x)) = \mathrm{P}(g(X) = g(x)) = \mathrm{P}(X = x)$$

Se viceversa la funzione non è iniettiva e vi potrebbero essere casi in cui $X(s_1) \neq X(s_2)$ ma $\implies g(X(s_1)) = g(X(s_2))$ allora dobbiamo sommare le probabilità dei diversi esiti di x che risultano in uno stesso y. Il risultato è più generale (va bene anche per il caso precedente) e lo esprimiamo a seguire.

Proposizione 4.4.1 (PMF di $g(X)$)**.** *Sia X una variabile casuale discreta e $g : \mathbb{R} \to \mathbb{R}$. Allora il supporto di $g(X)$ è l'insieme di y tali che $g(x) = y$ per almeno un $x \in R_X$ e la PMF di $g(X)$ è*

$$\mathrm{P}(g(X) = y) = \sum_{x: g(x) = y} \mathrm{P}(X = x), \qquad \forall y \in R_{g(X)} \tag{4.21}$$

Esempio 4.4.1. In tabella 4.1 un esempio con X, $Y = 2X$ ($g(x) = 2 \cdot x$ iniettiva) e $Z = X^2$ ($g(x) = x^2$ non iniettiva).

Osservazione 82. È un errore comune applicare g alla PMF (che potrebbe portare a probabilità oltre l'unità): g va applicata al dominio della PMF.

X	$\mathrm{P}(X=x)$	$Y=2X$	$\mathrm{P}(Y=y)$	$Z=X^2$	$\mathrm{P}(Z=z)$
-1	0.33	-2	0.33	1	0.66
0	0.33	0	0.33	0	0.33
1	0.33	2	0.33		

Tabella 4.1: Funzioni di una variabile casuale discreta: un esempio

Definizione 4.4.2 (Trasformazioni di locazione-scala per vc continue)**.** Sia X una vc continua e sia $Y=\sigma X+\mu$ con $\sigma,\mu\in\mathbb{R}$. Si dice che Y è stata ottenuta mediante una trasformazione di posizione e scala a partire da X.

Osservazione 83. Qui σ controlla la scala (se positivo allarga Y rispetto a X) mentre μ la locazione (se positivo sposta la distribuzione Y a destra di X). Per tornare a X si standardizza Y, ossia $X=\dfrac{Y-\mu}{\sigma}$.

Proposizione 4.4.2. *Y presenta la stessa (famiglia di) distribuzione di X*

Dimostrazione. È stata ottenuta mediante una trasformazione lineare iniettiva. □

Osservazione 84. La trasformazione non si aplica a famiglie di distribuzioni discrete: o quanto meno non vale più la precedente proposizione dato che si cambia il supporto (es trasformare linearmente una binomiale non restituisce più una binomiale, funzione definita su $0,1,\ldots$).

4.5 Indipendenza di vc

4.5.1 Indipendenza di 2+ vc, variabili iid

Definizione 4.5.1 (Indipendenza di 2 vc)**.** Due vc X,Y si dicono indipendenti se

$$\mathrm{P}(X\le x\cap Y\le y)=\mathrm{P}(X\le x)\cdot\mathrm{P}(Y\le y),\qquad \forall x,y\in\mathbb{R} \tag{4.22}$$

Osservazione 85. Nel caso discreto la 4.22 è equivalente a

$$\mathrm{P}(X=x\cap Y=y)=\mathrm{P}(X=x)\cdot\mathrm{P}(Y=y),\qquad \forall x,y\in\mathbb{R}$$

Esempio 4.5.1. Nel lancio di due dati indipendenti, se X è l'esito del primo e Y del secondo, allora le vc somma somma e differenza $X+Y$, $X-Y$ degli esiti non sono indipendenti. Infatti si ha che:

$$\mathrm{P}(X+Y=12\cap X-Y=1)=0\neq\mathrm{P}(X+Y=12)\,\mathrm{P}(X-Y=1)=\frac{1}{6}\cdot\frac{5}{6}$$

Quindi dato che abbiamo trovato un controesempio alla definizione $X+Y$ e $X-Y$ sono dipendenti. Questo ha anche un senso: sapere che la somma di due dati è 12 ci dice che la loro differenza deve essere 0, pertanto le due vc forniscono informazioni una dell'altra.

Proposizione 4.5.1. *Se X e Y sono indipendenti, allora qualsiasi funzione di X è indipendente da qualsiasi funzione di Y*

Dimostrazione. Non dimostrata ma dovrebbe essere avere senso dall'intuizione sulla informazione □

Definizione 4.5.2 (Indipendenza di molteplici vc)**.** Le variabili casuali $X_1, \ldots, X_n$ (in numero finito) sono indipendenti se:

$$\mathrm{P}\left(X_1 \leq x_1 \cap \ldots \cap X_n \leq x_n\right) = \mathrm{P}\left(X_1 \leq x_1\right) \cdot \ldots \cdot \mathrm{P}\left(X_n \leq x_n\right), \qquad \forall x_1, \ldots, x_n \in \mathbb{R} \tag{4.23}$$

Viceversa un *numero infinito* di variabili casuali si ha indipendenza se ogni sottoinsieme finito di vc è indipendente.

Proposizione 4.5.2. *Se $X_1, \ldots, X_n$ sono indipendenti, allora sono indipendenti anche a coppie. O, generalizzando, è indipendente qualsiasi sottoinsieme di eventi.*

Dimostrazione. Se $X_1, \ldots, X_n$ sono indipendenti si ha (considerando a titolo di esempio la coppia X_1, X_2) che

$$\mathrm{P}\left(X_1 \leq x_1 \cap X_2 \leq x_2\right) = \mathrm{P}\left(X_1 \leq x_1\right) \cdot \mathrm{P}\left(X_2 \leq x_2\right)$$

Per vedere perché sia così basta far tendere a $+\infty$ gli $x_3, \ldots, x_n$ in maniera tale che a sinistra dell'uguale, nella definizione 4.23, entro parentesi si abbiano eventi certi e a destra dell'uguale si moltiplichi per 1. □

Osservazione 86. Il viceversa non vale, ossia l'indipendenza a coppie non implica l'indipendenza compelssiva.

Definizione 4.5.3 (Variabili casuali i.i.d.)**.** Si chiamano i.i.d. variabili casuali che sono *indipendenti* e hanno *identica distribuzione* (stessa CDF).

4.5.2 Indipendenza condizionata

Definizione 4.5.4 (Indipendenza condizionata)**.** Le vc X e Y sono indipendenti condizionatamente a Z se $\forall x, y \in \mathbb{R}$ ed ogni $z \in R_Z$ si ha:

$$\mathrm{P}\left(X \leq x \cap Y \leq y | Z = z\right) = \mathrm{P}\left(X \leq x | Z = z\right) \cdot \mathrm{P}\left(Y \leq y | Z = z\right) \tag{4.24}$$

Osservazione 87. Per vc discrete, una definizione equivalente è 4.24

$$\mathrm{P}\left(X = x \cap Y = y | Z = z\right) = \mathrm{P}\left(X = x | Z = z\right) \cdot \mathrm{P}\left(Y = y | Z = z\right) \tag{4.25}$$

Definizione 4.5.5 (PMF condizionata)**.** Date due vc discrete X e Z la funzione $\mathrm{P}\left(X = x | Z = z\right)$, quando considerata come funzione di x per un dato z fissato è detta PMF condizionata di X dato che $Z = z$

Proposizione 4.5.3. *L'indipendenza di vc non implica l'indipendenza condizionata e viceversa.*

Dimostrazione. Lo si può mostrare mediante controesempi, vedi Blitzstein pag 121. □

4.6 Momenti

Osservazione 88. La funzione di ripartizione costituisce l'elemento unificante della teoria delle vcs discrete e continue. La conoscenza di F_X equivale alla conoscenza dell'intera struttura probabilistica della vc.
Tuttavia per confrontare vc differenti, si rende spesso necessario utilizzare uno/più indicatori sintetici, detti momenti.

Definizione 4.6.1 (Momenti di una vc)**.** Indicatore sintetico legato ad una funzione di probabilità/densità attraverso:

$$\sum_{i=1}^{\infty} g(x_i) \cdot p_X(x_i) \qquad \text{se } X \text{ è discreta}$$

$$\int_{-\infty}^{+\infty} g(x) \cdot f_X(x)\, dx \qquad \text{se } X \text{ è continua}$$

Differenti funzioni g, applicate ai valori della vc X definiscono momenti differenti (momenti r-esimi, momenti r-esimi rispetto al valore medio, momenti r-esimi standardizzati).

Osservazione 89. Qui e nel seguito supponiamo che gli integrali siano finiti oppure che le serie convergano; infatti si dice che i momenti esistono se e solo se sono finiti. Non tutte le vc possiedono momenti.

Definizione 4.6.2 (Momenti caratteristici)**.** I momenti utili per la descrizione di una vc sono valore atteso, varianza, asimmetria e curtosi: sono detti momenti caratteristici.

Proposizione 4.6.1. *Se X e Y sono vc con la stessa distribuzione (PMF/PDF/CDF) allora hanno uguali momenti.*

Dimostrazione. Questo deriva dal fatto che nella definizione dei momenti si fa utilizzo solamente della distribuzione, che se coincide condurrà a risultati uguali per le due distribuzioni. □

4.6.1 Valore atteso

Definizione 4.6.3 (Momenti r-esimi di X)**.** Se g corrisponde alla potenza r-esima della vc X abbiamo i momenti r-esimi di X:

$$\mu_r = \mathbb{E}\left(X^r\right) = \begin{cases} \displaystyle\sum_i {x_i}^r \cdot p_X(x_i) & \text{se } X \text{ è discreta} \\ \displaystyle\int_{-\infty}^{+\infty} x^r \cdot f_X(x)\, dx & \text{se } X \text{ è continua} \end{cases} \tag{4.26}$$

Definizione 4.6.4 (Valore atteso)**.** Detto così il momento primo di una vc X, indicato con $\mathbb{E}(X)$ o con μ: fornisce una media pesata (per le rispettive probabilità) di tutti i possibili valori di X

$$\mathbb{E}(X) = \begin{cases} \displaystyle\sum_i x_i \cdot p_X(x_i) & \text{se } X \text{ è discreta} \\ \displaystyle\int_{-\infty}^{+\infty} x \cdot f_X(x)\, dx & \text{se } X \text{ è continua} \end{cases} \tag{4.27}$$

Esempio 4.6.1 (Valore atteso del lancio di un dado)**.** Se X rappresenta l'esito del lancio di un dado equilibrato si ha $p_X(1) = \ldots = p_X(6) = 1/6$ e dunque:

$$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

Proposizione 4.6.2 (Proprietà del valore atteso)**.**

$$\mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b \tag{4.28}$$
$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) \tag{4.29}$$
$$\mathbb{E}(g(X)) = \sum_i g(x_i) \cdot p_X(x_i) \tag{4.30}$$
$$\min(X) \leq \mathbb{E}(X) \leq \max(X) \tag{4.31}$$
$$\mathbb{E}(X - \mu) = 0 \tag{4.32}$$
$$\textit{minimizza il valore } \mathbb{E}\left((X - \mu)^2\right) \tag{4.33}$$

Osservazione 90. Congiuntamente alle 4.28 e 4.29 ci si riferisce come linearità del valore atteso, che torna spesso comodo per il calcolo soprattutto se si riesce a scrivere una vc come somma di due o più vc. La linearità è un mero fatto algebrico e di bello c'è che, ad esempio per 4.29, non è necessaria l'indipendenza tra X e Y affinché valga.

Dimostrazione. Mostriamo con riferimento alle variabili discrete. Per la 4.28

$$\begin{aligned}
\mathbb{E}(aX + b) &= \sum_i (ax_i + b) \cdot \mathrm{P}(aX + b = ax_i + b) = \sum_i (ax_i + b) \cdot \mathrm{P}(X = x_i) \\
&= \sum_i ax_i \cdot \mathrm{P}(X = x) + \sum_i b \cdot \mathrm{P}(X = x) \\
&= a \sum_i x_i \cdot \mathrm{P}(X = x) + b \underbrace{\sum_i \mathrm{P}(X = x)}_{1} \\
&= a\,\mathbb{E}(X) + b
\end{aligned}$$

Per 4.29 facendo un passo indietro, possiamo scrivere un generico valore atteso facendo riferimento all'evento $s \in \Omega$ e applicando la funzione X ad esso, al fine di ottenere x_i:

$$\mathbb{E}(X) = \sum_i x_i \cdot \mathrm{P}(X = x_i) = \sum_s X(s) \cdot \mathrm{P}(\{s\})$$

Da questa possiamo generalizzare alla somma di due funzioni:

$$\begin{aligned}
\mathbb{E}(X + Y) &= \sum_s (X + Y)(s) \cdot \mathrm{P}(\{s\}) = \sum_s (X(s) + Y(s)) \cdot \mathrm{P}(\{s\}) \\
&= \sum_s X(s) \cdot \mathrm{P}(\{s\}) + \sum_s Y(s) \cdot \mathrm{P}(\{s\}) \\
&= \mathbb{E}(X) + \mathbb{E}(Y)
\end{aligned}$$

Per il valore atteso della trasformazione g, 4.30, sfruttiamo la stessa tecnica facendo un passo indietro (rispetto all'applicazione della funzione X agli eventi dello spazio campionario): sia $s \in \Omega$ un evento dello spazio campionario e X la vc considerata. Come detto possiamo scrivere il valore atteso $\mathbb{E}(X)$ come prodotto del risultato di X per la probabilità che si verifichi quell'evento:

$$\mathbb{E}(X) = \sum_s X(s)\,\mathrm{P}(\{s\})$$

L'applicazione della trasformazione g porta il valore atteso $\mathbb{E}(g(X))$:

$$\begin{aligned}
\mathbb{E}(g(X)) &= \sum_s g(X(s)) \cdot \mathrm{P}(\{s\}) \\
&\overset{(1)}{=} \sum_i \sum_{s:X(s)=x_i} g(X(s))\,\mathrm{P}(\{s\}) \\
&= \sum_i g(x_i) \sum_{s:X(s)=x_i} \mathrm{P}(\{s\}) \\
&= \sum_i g(x_i) \cdot \mathrm{P}(X = x_i) \\
&= \sum_i g(x_i) \cdot p_X(x_i)
\end{aligned}$$

dove in (1) semplicemente raggruppiamo per i diversi s che attraverso X forniscono lo stesso x_i.
La 4.31 è ovvia essendo $\mathbb{E}(X)$ una media pesata da probabilità dei valori assunti da X; l'uguaglianza vale in caso di variabili degeneri.
La 4.32 è una applicazione della linearità

$$\mathbb{E}(X-\mu) = \mathbb{E}(X) - \mu = \mathbb{E}(X) - \mathbb{E}(X) = 0$$

□

Proposizione 4.6.3 (Monotonicità del valore atteso)**.** *Se X e Y sono vc tali che $X \geq Y$ con probabilità 1, allora $\mathbb{E}(X) \geq \mathbb{E}(Y)$, e nello specifico $\mathbb{E}(X) = \mathbb{E}(Y)$ se e solo se $X = Y$ con probabilità 1.*

Dimostrazione. Anche questo risultato vale sia per vc discrete che continue; lo proviamo con riferimento alle prime. Sia $Z = X - Y$; si ha che Z è non negativa dunque $\mathbb{E}(Z) \geq 0$ dato che definita come la somma di termini non negativi. Sfruttando la linearità si ha che

$$\mathbb{E}(Z) = \mathbb{E}(X-Y) = \mathbb{E}(X) - \mathbb{E}(Y) \geq 0 \iff \mathbb{E}(X) \geq \mathbb{E}(Y)$$

Se $\mathbb{E}(X) = \mathbb{E}(Y)$ allora per la linearità si ha che $\mathbb{E}(Z) = 0$, ma dato che $Z \geq 0$, allora deve essere che $Z = 0$ ossia $\mathrm{P}(Z = 0) = 1$ visto che anche solo un termine positivo della sommatoria del valore atteso darebbe un valore atteso positivo. Pertanto $\mathrm{P}(Z = 0) = 1 \iff \mathrm{P}(X = Y) = 1$. □

Esempio 4.6.2 (Valore atteso di trasformazione)**.** Supponiamo che X sia una vc che assuma i valori $-1, 0, 1$ con probabilità pari a $\mathrm{P}(x = -1) = 0.2$, $\mathrm{P}(x = 0) = 0.5$, $\mathrm{P}(x = 1) = 0.3$. Calcoliamo $\mathbb{E}(X^2)$ applicando prima la trasformazione e poi moltiplicando per la probabilità:

$$\mathbb{E}(X^2) = (-1)^2(0.2) + 0^2 \cdot (0.5) + 1^2(0.3) = 0.5$$

Proposizione 4.6.4 (Valore atteso di funzioni non lineari di vc)**.** *In generale non vale* $\mathbb{E}(g(X)) = g\,\mathbb{E}(X)$ *per una qualsiasi funzione* g.

Esempio 4.6.3. Sia X il lancio di un dado: calcoliamo $\exp(\mathbb{E}(X))$ e $\mathbb{E}(\exp X)$; ricordando che $\mathbb{E}(X) = 7/2$ si ha

$$
\begin{aligned}
g(\mathbb{E}(X)) &= \exp(7/2) \approx 33.12 \\
\mathbb{E}(g(X)) &= e^1 \cdot \frac{1}{6} + \ldots + e^6 \cdot \frac{1}{6} \approx 106.1
\end{aligned}
$$

Considerando invece una trasformazione lineare $g(x) = 2x + 1$ i due risultati coincidono, come in mostrato 4.28. Si ha:

$$
\begin{aligned}
g(\mathbb{E}(X)) &= 2 \cdot \frac{7}{2} + 1 = 8 \\
\mathbb{E}(g(X)) &= 3\frac{1}{6} + 5\frac{1}{6} + 7\frac{1}{6} + 9\frac{1}{6} + 11\frac{1}{6} + 13\frac{1}{6} = 8
\end{aligned}
$$

4.6.2 Varianza

Definizione 4.6.5 (Momenti r-esimi di X ripetto al valore medio)**.** Si hanno se $g = (x-\mu)^r$:

$$
\overline{\mu_r} = \mathbb{E}((X-\mu)^r) = \begin{cases} \displaystyle\sum_i (x_i - \mu)^r \cdot p_X(x_i) & \text{se } X \text{ è discreta} \\ \displaystyle\int_{-\infty}^{+\infty} (x-\mu)^r \cdot f_X(x)\, dx & \text{se } X \text{ è continua} \end{cases} \tag{4.34}
$$

Osservazione 91. Poiché $\overline{\mu_0} = 1, \overline{\mu_1} = 0$, l'interesse per tali momenti nasce a partire dal momento secondo ($r = 2$).

Definizione 4.6.6 (Varianza)**.** Definita come

$$
\overline{\mu_2} = \operatorname{Var}(X) = \sigma^2 = \mathbb{E}\left((X-\mu)^2\right) \tag{4.35}
$$

misura la dispersione di una vc attorno al suo valore medio.

Proposizione 4.6.5 (Formula rapida per il calcolo)**.**

$$
\operatorname{Var}(X) = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2 \tag{4.36}
$$

Dimostrazione. Ponendo $\mathbb{E}(X) = \mu$ per comodità tipografica:

$$
\begin{aligned}
\operatorname{Var}(X) = \mathbb{E}\left((X-\mu)^2\right) &= \sum_i (x_i - \mu)^2 \cdot p_X(x_i) = \sum_i (x_i^2 - 2\mu x_i + \mu^2) \cdot p_X(x_i) \\
&= \sum_i x_i^2 \cdot p_X(x_i) - 2\mu \sum_i x_i \cdot p_X(x_i) + \mu^2 \sum_i p_X(x_i) \\
&= \mathbb{E}\left(X^2\right) - 2\mu^2 + \mu^2 = \mathbb{E}\left(X^2\right) - \mu^2 \\
&= \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2
\end{aligned}
$$

Alternativamente/compattamente si poteva espandere $(X-\mu)^2$ e utilizzare la linearità del valore atteso:

$$
\begin{aligned}
\operatorname{Var}(X) &= \mathbb{E}\left((X-\mu)^2\right) = \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right) = \mathbb{E}\left(X^2\right) - 2\mu\,\mathbb{E}(X) + \mu^2 \\
&= \mathbb{E}\left(X^2\right) - \mu^2 = \mathbb{E}\left(X^2\right) - [\mathbb{E}(X)]^2
\end{aligned}
$$

□

Esempio 4.6.4. Per $\mathrm{Var}(X)$ dove X rappresenta l'esito del lancio di un dado, in precedenza si è calcolato che $\mathbb{E}(X) = 7/2$; inoltre si ha che

$$\mathbb{E}\left(X^2\right) = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right) = \left(\frac{1}{6}\right)(91)$$

Per cui

$$\mathrm{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

Proposizione 4.6.6 (Proprietà della varianza)**.** *Se* $a, b, c \in \mathbb{R}$*:*

$$\mathrm{Var}(X) \geq 0 \tag{4.37}$$

$$\mathrm{Var}(X) = 0 \iff \mathrm{P}(X = c) = 1 \tag{4.38}$$

$$\mathrm{Var}(aX \pm b) = a^2\,\mathrm{Var}(X) \tag{4.39}$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y), \quad \textit{se } X, Y \textit{ indipendenti} \tag{4.40}$$

Dimostrazione. Per la 4.37, la varianza è il valore atteso della vc nonnegativa $(X - \mu)^2$, motivo per cui è non negativa.
Per 4.38 se $\mathrm{P}(X = c) = 1$ per qualche costante c allora $\mathbb{E}(X) = c$ e $\mathbb{E}\left(X^2\right) = c^2$, pertanto $\mathrm{Var}(X) = 0$; viceversa se $\mathrm{Var}(X) = 0$ allora $\mathbb{E}\left((X - \mu)^2\right) = 0$ che mostra che $(X - \mu)^2 = 0$ ha probabilità 1, che a sua volta mostra che X è uguale alla sua media con probabilità 1.
Per la 4.39 ponendo sempre $\mathbb{E}(X) = \mu$ e per la linearità del valore atteso si ha:

$$\begin{aligned}
\mathrm{Var}(aX + b) &= \mathbb{E}\left((aX + b - (a\mu + b))^2\right) \\
&= \mathbb{E}\left((aX + b - a\mu - b)^2\right) \\
&= \mathbb{E}\left((aX - a\mu)^2\right) \\
&= \mathbb{E}\left(a^2(X - \mu)^2\right) \\
&= a^2\,\mathbb{E}\left((X - \mu)^2\right) \\
&= a^2\,\mathrm{Var}(X)
\end{aligned}$$

La 4.40 verrà dimostrata e generalizzata in seguito. □

Osservazione 92 (La varianza non è lineare)*.* A differenza del valore atteso la varianza non è lineare; la costante a risulta al quadrato, b non è considerata e la varianza della somma di vc potrebbe non coincidere con la somma delle varianze.

Definizione 4.6.7 (Deviazione standard)**.** La radice quandrata di $\mathrm{Var}(X)$ è detta deviazione standard e la si indica con σ_X o σ:

$$\sigma = \sigma_X = \sqrt{\mathrm{Var}(X)} \tag{4.41}$$

4.6.3 Asimmetria e curtosi

Definizione 4.6.8 (Vc standardizzata)**.** Se X ha valore atteso $\mathbb{E}(X) = \mu$ e varianza $\mathrm{Var}(X) = \sigma^2 \in (0, +\infty)$ si definisce la vc standardizzata Z come:

$$Z = \frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}} = \frac{X - \mu}{\sigma} \tag{4.42}$$

Questa trasformazione rende la variabile casuale indipendente dall'unità di misura.

Definizione 4.6.9 (Momenti r-esimi standardizzati di X)**.** Se $g = \left(\frac{x-\mu}{\sigma}\right)^r$ si hanno:

$$\overline{\overline{\mu_r}} = \mathbb{E}\left(\left(\frac{X-\mu}{\sigma}\right)^r\right) = \begin{cases} \sum_i \left(\frac{x_i-\mu}{\sigma}\right)^r \cdot p_X(x_i) & \text{se } X \text{ è discreta} \\ \int_{-\infty}^{+\infty} \left(\frac{x-\mu}{\sigma}\right)^r \cdot f_X(x)\, dx & \text{se } X \text{ è continua} \end{cases} \tag{4.43}$$

Osservazione 93. Per qualsiasi vc $\overline{\overline{\mu_0}} = 1$, $\overline{\overline{\mu_1}} = 0$, $\overline{\overline{\mu_2}} = 1$; pertanto i momenti standardizzati di interesse sono quelli di ordine 3 e 4.

4.6.3.1 Asimmetria/Skewness

Definizione 4.6.10 (Vc simmetrica)**.** La vc X è simmetrica rispetto a μ se la variabile casuale $X - \mu$ ha la stessa distribuzione di $\mu - X$.

Osservazione 94 (Intuizione significato). $X - \mu$ sposta la densità/probabilità, così com'è, centrandola sullo 0. Intuitivamente $-X$ ha l'effetto di ottenere la densità probabilità simmetrica/specchiata rispetto a $x = 0$; infine $-X + \mu$ specchia la densità/probabilità rispetto a 0 e poi la ricentra su 0. Pertanto se $X-\mu$ e $-X+\mu$ coincidono, è perché la distribuzione di partenza X è simmetrica rispetto al centro.

Osservazione 95 (Cosa è μ). Il numero μ nella definizione deve essere $\mathbb{E}(X)$ se esiste. Infatti si ha che

$$\mathbb{E}(X) - \mu \overset{(1)}{=} \mathbb{E}(X-\mu) \overset{(2)}{=} \mathbb{E}(\mu - X) \overset{(3)}{=} \mu - \mathbb{E}(X)$$

con (1) e (3) per le proprietà del valore atteso, (2) per l'ipotesi di uguaglianza delle distribuzioni (che si manifesta nell'uguaglianza dei valori attesi). Considerando il primo e ultimo termine dell'uguaglianza, allora:

$$\begin{aligned} \mathbb{E}(X) - \mu = \mu - \mathbb{E}(X) &\iff \mathbb{E}(X) - \mu = -(\mathbb{E}(X) - \mu) \iff \mathbb{E}(X) - \mu = 0 \\ &\iff \mathbb{E}(X) = \mu \end{aligned}$$

Tra l'altro μ, se esiste, è anche la mediana di X (non lo si dimostra, Blitzstein pag 249).

Proposizione 4.6.7 (Simmetria di una vc continua (PDF))**.** *Sia X una vc continua con PDF f. Allora è simmetrica su μ se e solo se $f(x) = f(2\mu - x)$.*

Osservazione 96. La definizione è meramente quella di una funzione simmetrica rispetto a $x = \mu$ (vedi calcolo).

Dimostrazione. Sia F la CDF di X; dimostriamo la doppia implicazione.
Se la simmetria vale ($X - \mu = \mu - X$) abbiamo:

$$\begin{aligned} F(x) &= \mathrm{P}(X - \mu \le x - \mu) \overset{(1)}{=} \mathrm{P}(\mu - X \le x - \mu) \overset{(2)}{=} \mathrm{P}(X \ge 2\mu - x) \\ &= 1 - F(2\mu - x) \end{aligned}$$

dove in (1) abbiamo sfruttato la simmetria ($X - \mu = \mu - X$) e in (2) abbiamo elaborato algebricamente. Facendo la derivata dei membri estremi dell'equazione si ottiene $f(x) = f(2\mu - x)$.
Viceversa supponendo che $f(x) = f(2\mu - x)$ valga $forallx$, vogliamo dimostrare che $\mathrm{P}(X - \mu \leq t) = \mathrm{P}(\mu - X \leq t)$, ossia vi è simmetria e le cumulate CDF coincidono. Si ha

$$\begin{aligned}\mathrm{P}(X - \mu \leq t) = \mathrm{P}(X \leq \mu + t) = \int_{-\infty}^{\mu+t} f(x)\,dx &\overset{(1)}{=} \int_{-\infty}^{\mu+t} f(2\mu - x)\,dx \\ &\overset{(2)}{=} \int_{\mu-t}^{\infty} f(w)\,dw = \mathrm{P}(\mu - X \leq t)\end{aligned}$$

dove in abbiamo sfruttato che $f(x) = f(2\mu - x)$, mentre in (2) deve avvenire qualche trick di integrazione (integra $f(-x)$ ad indici invertiti e moltiplicati direi). □

Definizione 4.6.11 (Asimmetria/Skewness)**.** È il momento standardizzato di ordine 3:

$$\mathrm{Asym}(X) = \overline{\overline{\mu_3}} = \mathbb{E}\left(\left(\frac{X - \mu}{\sigma}\right)^3\right) \tag{4.44}$$

Osservazione 97. Una skewness positiva indica una coda destra più lunga della sinistra, mentre una asimmetria negativa indica una coda a destra più lunga.

4.6.3.2 Curtosi

Definizione 4.6.12 (Curtosi)**.** È il momento standardizzato di ordine 4:

$$\mathrm{Kurt}(X) = \overline{\overline{\mu_4}} = \mathbb{E}\left(\left(\frac{X - \mu}{\sigma}\right)^4\right) \tag{4.45}$$

Osservazione 98. Alcuni definiscono la curtosi, centrandola sul valore assunto dalla Normale (3), come:

$$\mathrm{Kurt}(X) = \mathbb{E}\left(\left(\frac{X - \mu}{\sigma}\right)^4\right) - 3 \tag{4.46}$$

In questo modo una normale ha curtosi 0 mentre le rimanenti hanno un valore positivo o negativo a seconda che diano più o meno peso alle code rispetto a questa (eccesso di curtosi).

Osservazione 99. Una distribuzione con eccesso di curtosi (4.46) negativo (detta *platicurtica*) tende ad avere un profilo più piatto della normale e una minore importanza delle code. Produce outlier in misura minore o meno estremi rispetto alla normale. Un esempio è l'uniforme.
Viceversa una distribuzione con eccesso di curtosi positivo è detta *leptocurtica* (ad esempio distribuzione T di Student, logistica, Laplace): ha code che si avvicinano allo zero più lentamente rispetto una gaussiana, per cui produce più outlier della stessa.
In fig 4.3 alcune distribuzioni (con media 0 e varianza 1) e relativa curtosi.

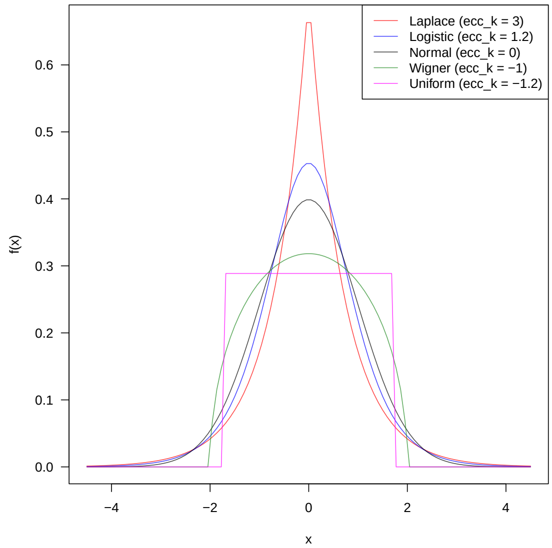


Figura 4.3: PDF per alcune distribuzioni con media 0 e variaanza 1 e relative curtosi

4.7 Modelli probabilistici e utilizzo di R

Osservazione 100. Nei capitoli successivi approfondiamo i modelli probabilistici principali, ossia le diverse famiglie di distribuzioni più note/utilizzate.

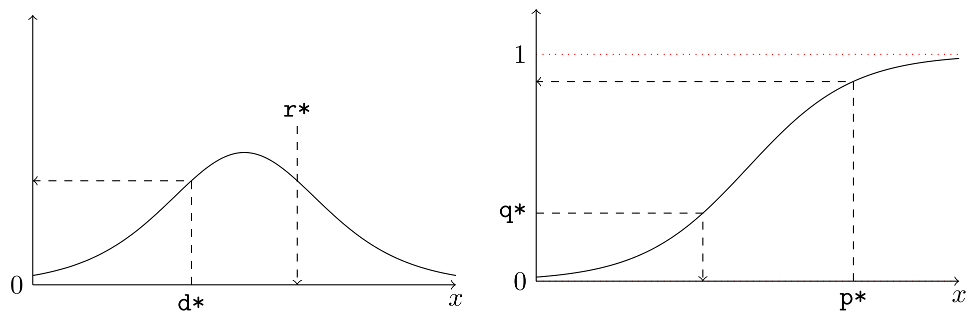
Definizione 4.7.1 (Famiglia parametrica di variabili casuali)**.** Insieme di funzioni di ripartizione $F(x; \boldsymbol{\Theta})$ aventi la medesima forma funzionale, ma che si differenziano per uno o più parametri, indicati con $\boldsymbol{\Theta}$.

Definizione 4.7.2 (Spazio parametrico)**.** Indicato con $\boldsymbol{\Theta}$, consiste nell'insieme dei valori possibili dei parametri che caratterizzano una distribuzione di probabilità.

Osservazione 101. In tabella 4.2 i prefissi delle funzioni e i suffissi delle famiglie parametriche principali in R. In figura 4.4 un utile promemoria di input (dove parte la freccia, cosa fornire alla funzione) e output (dove arriva la freccia, cosa restituisce la funzione) delle quattro funzioni. Ad esempio `dnorm` restituisce la funzione di densità della normale.

Funzione	Prefisso	Famiglia	Suffisso	Famiglia	Suffisso
Densità/Probabilità	`d`	Bernoulli	`binom`	Uniforme cont.	`unif`
Ripartizione	`p`	Binomiale	`binom`	Esponenziale	`exp`
Quantile	`q`	Geometrica	`geom`	Normale	`norm`
Generazione pseudoc.	`r`	Binomiale neg.	`nbinom`	Gamma	`gamma`
		Ipergeometrica	`hyper`	Chi-quadrato	`chisq`
		Poisson	`pois`	Beta	`beta`
		Uniforme disc.	$*$	T di Student	`t`
				F	`f`
				Logistica	`logis`
				Lognormale	`lnorm`
				Weibull	`weibull`
				Pareto (pac. `VGAM`)	`pareto`

Tabella 4.2: Funzioni per famiglie parametriche in R



(a) PMF/PDF (b) CDF

Figura 4.4: Funzioni in R

Capitolo 5

Variabili casuali discrete

5.1 Degenere

Definizione 5.1.1 (Vc degenere)**.** X è una variabile degenere sul valore c, e si scrive $X \sim \mathrm{D}(c)$ se $\mathrm{P}(X = c) = 1$.

Proposizione 5.1.1 (Momenti)**.**

$$\begin{aligned} &\mathbb{E}(X) = c \\ &\mathrm{Var}(X) = 0 \end{aligned}$$

Dimostrazione.

$$\begin{aligned} &\mathbb{E}(X) = c \cdot 1 = c \\ &\mathrm{Var}(X) = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2 = c^2 \cdot 1 - c^2 = 0 \end{aligned}$$

□

Osservazione 102. Come si è visto tra le proprietà della varianza, la degenere è l'unica vc ad avere varianza nulla.

5.2 Bernoulli

5.2.1 Definizione

Osservazione 103. Viene utilizzata quando si ha a che fare con un esperimento il cui esito possibile è dicotomico (es $X = 1$ successo, $X = 0$ insuccesso).

Definizione 5.2.1 (vc di Bernoulli)**.** Una vc X ha distribuzione di Bernoulli con parametro $0 \leq p \leq 1$, e si scrive $X \sim \mathrm{Bern}(p)$, se $\mathrm{P}(X = 1) = p$ e $\mathrm{P}(X = 0) = 1 - p$.

Definizione 5.2.2 (Prova bernoulliana)**.** Singola esecuzione dell'esperimento.

Osservazione 104. Se $p = 0 \vee p = 1$ si ottiene un caso di *distribuzione degenere.*

Osservazione 105. Qualsiasi evento ha una vc di Bernoulli associata, la variabile indicatrice. Se $p = \mathrm{P}(A)$, $I_A \sim \mathrm{Bern}(p)$

5.2.2 Funzioni

Osservazione 106 (Supporto e spazio parametrico).

$$R_X = \{0, 1\}$$
$$\mathbf{\Theta} = \{p \in \mathbb{R} : 0 \leq p \leq 1\}$$

Definizione 5.2.3 (Funzione di massa di probabilità)**.**

$$p_X(x) = \mathrm{P}(X = x) = p^x \cdot (1-p)^{1-x} \cdot \mathbb{1}_{R_X}(x) \tag{5.1}$$

Definizione 5.2.4 (Funzione di ripartizione)**.**

$$F_X(x) = \mathrm{P}(X \leq x) = \begin{cases} 0 & \text{se } x < 0 \\ 1-p & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases} \tag{5.2}$$

5.2.3 Momenti

Proposizione 5.2.1 (Momenti caratteristici)**.**

$$\mathbb{E}(X) = p \tag{5.3}$$
$$\mathrm{Var}(X) = p(1-p) \tag{5.4}$$
$$\mathrm{Asym}(X) = \frac{1-2p}{\sqrt{p(1-p))}} \tag{5.5}$$
$$\mathrm{Kurt}(X) = \frac{3p^2 - 3p + 1}{p(1-p)} \tag{5.6}$$

Dimostrazione. Per il valore atteso

$$\mathbb{E}(X) = 1 \cdot p + 0 \cdot (1-p) = p$$

Per la varianza, dato che $X^2 = X$ e dunque $\mathbb{E}(X^2) = \mathbb{E}(X)$ si ha:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p - p^2 = p(1-p)$$

□

Osservazione 107. In particolare il valore atteso coincide con la probabilità di successo e la varianza è sempre compresa nell'intervallo $[0; 0.25]$, raggiungendo il massimo per $p = 1/2$.

5.2.4 Vc indicatrici

5.2.4.1 Definizione e proprietà

Definizione 5.2.5 (Vc indicatrice dell'evento A)**.** Definita come

$$I_A = I(A) = \begin{cases} 1 & \text{se l'evento } A \text{ si verifica} \\ 0 & \text{altrimenti} \end{cases}$$

Osservazione 108. Pertanto se $p = \mathrm{P}(A)$ si ha che $I_A \sim \mathrm{Bern}(p)$

Proposizione 5.2.2 (Proprietà della vc indicatrice).

$$(I_A)^n = I_A, \quad \forall n \in \mathbb{N} : n > 0 \tag{5.7}$$
$$I_{\overline{A}} = 1 - I_A \tag{5.8}$$
$$I_{A \cap B} = I_A \cdot I_B \tag{5.9}$$
$$I_{A \cup B} = I_A + I_B - I_A \cdot I_B \tag{5.10}$$

Dimostrazione. La 5.7 vale dato che $0^n = 0$ e $1^n = 1$ per qualsiasi intero positivo n. La 5.8 vale dato che $1 - I_A$ è 1 se A non accade e 0 se accade. Per la 5.9, $I_A \cdot I_B$ è 1 solo se sia I_A che I_B sono 1 e 0 altrimenti. Per la 5.10,

$$\begin{aligned} I_{A \cup B} &\overset{(1)}{=} 1 - I_{\overline{A} \cap \overline{B}} = 1 - I_{\overline{A}} \cdot I_{\overline{B}} = 1 - (1 - I_A)(1 - I_B) \\ &= I_A + I_B - I_A I_B \end{aligned}$$

dove in (1) abbiamo sfruttato De Morgan. □

5.2.4.2 Legame tra probabilità e valore atteso

Osservazione 109. Le variabili indicatrici forniscono un collegamento tra probabilità e valore atteso

Proposizione 5.2.3 (FUndamental bridge). *Vi è corrispondenza biunivoca tra eventi e vc indicatrici: la probabilità di un evento A è il valore atteso della sua vc indicatrice I_A, ossia:*

$$\mathrm{P}(A) = \mathbb{E}(I_A) \tag{5.11}$$

Dimostrazione. Per qualsiasi evento A abbiamo una vc indicatrice I_A. Questa è una corrispondenza biunivoca dato che A determina univocamente I_A e viceversa (per andare da I_A a A possiamo utilizzare il fatto che $A = \{s \in S : I_A(s) = 1\}$). Dato che $I_A \sim \mathrm{Bern}(p)$ con $p = \mathrm{P}(A)$, si ha

$$\mathbb{E}(I_A) = \mathbb{E}(\mathrm{Bern}(p)) = p = \mathrm{P}(A)$$

□

Osservazione 110. Il precedente risultato connette eventi e la loro vc indicatrice e permette di esprimere qualsiasi probabilità come un valore atteso. Alcuni esempi seguono nella prossima sezione.

Osservazione 111. Altresì le variabili indicatrici sono utili in molti problemi su valori attesi. Possiamo spesso esprimere una vc discreta complecata la cui distribuzione non conosciamo come somma di variabili indicatrici che sono semplici. Il fundamental bridge ci permette di trovare il valore atteso delle variabili indicatrici, poi per linearità otteniamo il valore atteso della vc dalla quale siamo partiti.

5.2.4.3 Applicazioni: probabilità

Proposizione 5.2.4 (Disuguaglianza di Boole). *Se $E_1, \ldots, E_n$ sono eventi si ha che:*

$$\mathrm{P}(E_1 \cup \ldots \cup E_n) \leq \mathrm{P}(E_1) + \ldots + \mathrm{P}(E_n) \tag{5.12}$$

Dimostrazione. Siano $E_1, \dots, E_n$ gli eventi considerati. Notiamo che

$$I_{E_1 \cup \ldots \cup E_n} \leq I_{E_1} + \ldots + I_{E_n}$$

dato che il lato sinistro è 1 se si verificano tutti gli eventi mentre il destro è 1 se ne si verifica anche solo uno. Prendendo il valore atteso di entrambi i lati si ha:

$$\begin{aligned}
&\mathbb{E}\left(I_{E_1 \cup \ldots \cup E_n}\right) \leq \mathbb{E}\left(I_{E_1} + \ldots + I_{E_n}\right) \quad \text{applicando linearità si arriva a} \ldots \\
&\mathbb{E}\left(I_{E_1 \cup \ldots \cup E_n}\right) \leq \mathbb{E}\left(I_{E_1}\right) + \ldots + \mathbb{E}\left(I_{E_n}\right) \quad \text{applicando la 5.11 a} \ldots \\
&\mathrm{P}\left(E_1 \cup \ldots \cup E_n\right) \leq \mathrm{P}\left(E_1\right) + \ldots + \mathrm{P}\left(E_n\right)
\end{aligned}$$

□

Proposizione 5.2.5 (Diseguaglianza di Bonferroni)**.** *Se $E_1, \dots, E_n$ sono eventi si ha che:*

$$\mathrm{P}\left(E_1 \cap \ldots \cap E_n\right) \geq 1 - \sum_{i=1}^{n} \mathrm{P}\left(\overline{E_i}\right) \tag{5.13}$$

Dimostrazione. Analogamente a quanto fatto per la disuguaglianza di Boole si ha che, per De Morgan

$$I_{E_1 \cap \ldots \cap E_n} = 1 - I_{\overline{E_1} \cap \ldots \cap \overline{E_n}}$$

Passando ai valori attesi

$$\begin{aligned}
&\mathbb{E}\left(I_{E_1 \cap \ldots \cap E_n}\right) = \mathbb{E}\left(1 - I_{\overline{E_1} \cup \ldots \cup \overline{E_n}}\right) \quad \text{per linearità} \ldots \\
&\mathbb{E}\left(I_{E_1 \cap \ldots \cap E_n}\right) = 1 - \mathbb{E}\left(I_{\overline{E_1} \cup \ldots \cup \overline{E_n}}\right) \quad \text{passando alle probabilità} \ldots \\
&\mathrm{P}\left(E_1 \cap \ldots \cap E_n\right) = 1 - \mathrm{P}\left(\overline{E_1} \cup \ldots \cup \overline{E_n}\right)
\end{aligned}$$

Infine applicando la 5.12 si ha

$$\mathrm{P}\left(E_1 \cap \ldots \cap E_n\right) = 1 - \mathrm{P}\left(\overline{E_1} \cup \ldots \cup \overline{E_n}\right) \geq 1 - \mathrm{P}\left(\overline{E_1}\right) - \ldots - \mathrm{P}\left(\overline{E_n}\right)$$

□

Proposizione 5.2.6 (Principio di inclusione/esclusione)**.** *Nel caso di due eventi*

$$\mathrm{P}\left(A \cup B\right) = \mathrm{P}\left(A\right) + \mathrm{P}\left(B\right) - \mathrm{P}\left(A \cap B\right) \tag{5.14}$$

Nel caso generale

$$\mathrm{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1 < \ldots < i_r} \mathrm{P}\left(E_{i1} \cap E_{i2} \cap \ldots \cap E_{ir}\right) \tag{5.15}$$

$$= \sum_{i} \mathrm{P}\left(E_i\right) - \sum_{i<j} \mathrm{P}\left(E_i \cap E_j\right) + \sum_{i<j<k} \mathrm{P}\left(E_i \cap E_j \cap E_k\right) - \ldots + (-1)^{n+1} \mathrm{P}\left(E_1 \cap \ldots \cap E_n\right) \tag{5.16}$$

Dimostrazione. Per la 5.14 basta prendere il valore atteso di entrambi i membri di 5.10. Per la 5.15 in possiamo utilizzare le proprietà delle vc indicatrici come segue

$$\begin{aligned}
1 - I_{A_1 \cup \ldots \cup A_n} &= I_{\overline{A_1} \cap \ldots \cap \overline{A_n}} \\
&= I_{\overline{A_1}} \cdot \ldots \cdot I_{\overline{A_n}} \\
&= (1 - I_{A_1}) \cdot \ldots \cdot (1 - I_{A_n}) \\
&\overset{(1)}{=} 1 - \sum_i I_{A_i} + \sum_{i<j} I_{A_i} I_{A_j} - \ldots + (-1)^n I_{A_1} \cdot \ldots \cdot I_{A_n}
\end{aligned}$$

dove in (1) l'1 significa selezionare tutti gli 1 nei fattori, il $\sum_i I_{A_i}$ si ottiene selezionando tutti gli 1 a meno di un fattore a turno che ha sempre il - davanti, $\sum_{i<j} I_{A_i} I_{A_j}$ si ottiene selezionando tutti gli 1 ad eccezione di due fattori. Prendendo i valori attesi di ambo i membri si ha

$$\begin{aligned}
\mathbb{E}\left(1 - I_{E_1 \cup \ldots \cup E_n}\right) &= \mathbb{E}\left(1 - \sum_i I_{E_i} + \sum_{i<j} I_{E_i} I_{E_j} - \ldots + (-1)^n I_{E_1} \cdot \ldots \cdot I_{E_n}\right) \\
1 - \mathbb{E}\left(I_{E_1 \cup \ldots \cup E_n}\right) &\overset{(1)}{=} 1 - \mathbb{E}\left(\sum_i I_{E_i} - \sum_{i<j} I_{E_i} I_{E_j} + \ldots + (-1)^{n+1} I_{E_1} \cdot \ldots \cdot I_{E_n}\right) \\
\mathbb{E}\left(I_{E_1 \cup \ldots \cup E_n}\right) &= \mathbb{E}\left(\sum_i I_{E_i}\right) - \mathbb{E}\left(\sum_{i<j} I_{E_i} I_{E_j}\right) + \ldots + \mathbb{E}\left((-1)^{n+1} I_{E_1} \cdot \ldots \cdot I_{E_n}\right) \\
\mathrm{P}\left(\bigcup_{i=1}^n E_i\right) &= \sum_i \mathrm{P}\left(E_i\right) - \sum_{i<j} \mathrm{P}\left(E_i \cap E_j\right) + \ldots + (-1)^{n+1}\, \mathrm{P}\left(E_1 \cap \ldots \cap E_n\right)
\end{aligned}$$

dove in (1) abbiamo raccolto un meno al secondo membro entro parentesi. □

5.2.4.4 Applicazioni: calcolo valori attesi

Esempio 5.2.1 (Matching carte)**.** Abbiamo un mazzo di n carte numerate da 1 a n ben mischiato. Una carta è un match se la sua posizione nell'ordine del mazzo matcha con il suo numero. Sia X il numero totale di match nel mazzo: qual è il valore atteso di X?
Se scriviamo $X = I_1 + \ldots + I_n$ con

$$I_i = \begin{cases} 1 & \text{se l'}i\text{-esima carta matcha col proprio numero} \\ 0 & \text{altrimenti} \end{cases}$$

Si ha che, non condizionando a nulla e pensando ad un singolo shuffle/match

$$\mathbb{E}\left(I_i\right) = \frac{1}{n}$$

pertanto per linearità

$$\mathbb{E}\left(X\right) = \mathbb{E}\left(I_1\right) + \ldots + \mathbb{E}\left(I_n\right) = n \cdot \frac{1}{n} = 1$$

Quindi il numero di match medi è 1, indipendentemente da n. Anche se I_i sono dipendenti in maniera complicata, la linearità del valore atteso vale sempre.

Fisso ...	con reinserimento	senza reinserimento
n trial	binomiale	ipergeometrica
n successi	binomiale negativa	ipergeometrica negativa

Tabella 5.1

Esempio 5.2.2 (Valore atteso di Ipergeometrica Negativa)**.** Un'urna contiene w palline bianche e b palline nere che sono estratte senza reinserimento. Il numero di palline nere estratte prima di pescare la prima bianca ha una distribuzione Ipergeometrica negativa (in tab 5.1 una sintesi dei casi). Trovare il valore atteso. Trovarlo dalla definizione della variabile è complicato, ma possiamo esprimere la variabile come somma di indicatrici. Etichettiamo le palline nere con $1, 2, \ldots b$ e sia I_i l'indicatrice che la pallina nera i è stata estratta prima di qualsiasi bianca. Si ha che

$$\mathrm{P}\left(I_i = 1\right) = \frac{1}{w+1}$$

dato nel listare l'ordine in cui la pallina nera i e le altre bianche son pescate (ignorando le altre) tutti gli ordine sono equiprobabili. Pertanto per linearità

$$\mathbb{E}\left(\sum_{i=1}^{b} I_i\right) = \sum_{i=1}^{b} \mathbb{E}\left(I_i\right) = \frac{b}{w+1}$$

La risposta ha n senso dato che aumenta con b, diminuisce con w ed è corretta nei casi estremi $b = 0$ (nessuna pallina nera sarà estratta) e $w = 0$ (tutte le palline nere saranno esaurite prima di pescare una non esistente bianca).

5.3 Binomiale

5.3.1 Definizione

Osservazione 112. Viene utilizzata quando si vuole sapere qual'è la probabilità che si osservi numero x di successi in una sequenza $n \geq x$ di prove Bernoulliane indipendenti aventi la stessa probabilità p.

Definizione 5.3.1 (vc binomiale)**.** Eseguiamo n prove bernoulliane indipendenti, aventi comune probabilità di successo p. Sia X la somma dei successi ottenuti: allora X si distribuisce come una vc binomiale di parametri n e p, e si scrive $X \sim \mathrm{Bin}\left(n, p\right)$.

Osservazione 113. Se $n = 1$ la distribuzione Binomiale coincide con quella di Bernoulli, ossia $\mathrm{Bin}\left(1, p\right) = \mathrm{Bern}\left(p\right)$

Proposizione 5.3.1. *La binomiale può essere generata sommando bernoulliane iid; se X_i, $i = 1, \ldots, n$ sono vc bernoulliane iid $X_i \sim \mathrm{Bern}\left(p\right)$ allora la loro somma $X = \sum_{i=1}^{n} X_i \sim \mathrm{Bin}\left(n, p\right)$*

Dimostrazione. Sia $X_i = 1$ se l'i-esimo trial ha successo o 0 in caso contrario. Se pensiamo di avere una persona per ciascun trial, chiediamo di alzare la mano se si ha successo e contiamo le mani alzate (che equivale a sommare X_i) otteniamo il numero totale di successi in n trial che è X. $\square$

5.3.2 Funzioni

Osservazione 114 (Supporto e spazio parametrico).

$$\begin{aligned} R_X &= \{0, 1, \dots, n\} \\ \mathbf{\Theta} &= \{n \in \mathbb{N} \setminus \{0\}, \; p \in \mathbb{R} : 0 \leq p \leq 1\} \end{aligned}$$

Definizione 5.3.2 (Funzione di massa di probabilità)**.**

$$p_X(x) = \mathrm{P}\left(X = x\right) = \binom{n}{x} \cdot p^x (1-p)^{n-x} \cdot \mathbb{1}_{R_X}(x) \tag{5.17}$$

con: x è il numero di successi, n è il numero di esperimenti, p probabilità di successo in ogni esperimento.

Osservazione 115. Nella 5.17 la prima parte (il coefficiente binomiale) serve per quantificare il numero di casi in cui si verificano il numero di successi desiderati; questa viene moltiplicata per la seconda che costituisce la probabilità di un tale esito (determinato come probabilità di eventi indipendenti di successo/insuccesso).

Definizione 5.3.3 (Funzione di ripartizione)**.**

$$F_X(x) = \mathrm{P}\left(X \leq x\right) = \sum_{k=0}^{x} \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

Validità PMF. Si ha che

$$\sum_{x=0}^{n} p_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \overset{(1)}{=} \left(p + (1-p)\right)^n = 1$$

dove in (1) si è sfruttata la proprietà del coefficiente binomiale:

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

□

5.3.3 Momenti

Proposizione 5.3.2 (Momenti caratteristici)**.**

$$\mathbb{E}\left(X\right) = np \tag{5.18}$$

$$\mathrm{Var}\left(X\right) = np(1-p) \tag{5.19}$$

$$\mathrm{Asym}\left(X\right) = \frac{1-2p}{\sqrt{np(1-p))}} \tag{5.20}$$

$$\mathrm{Kurt}\left(X\right) = 3 + \frac{1-6p+6p^2}{np(1-p)} \tag{5.21}$$

Dimostrazione. Per il valore atteso, sfruttando il fatto che $X \sim \mathrm{Bin}\left(n, p\right)$ sia descrivibile come la somma di n vc X_i $\mathrm{Bern}\left(p\right)$, sfruttando la linearità del valore atteso, il risultato è la somma di n valori attesi uguali:

$$\mathbb{E}\left(X\right) = \mathbb{E}\left(\sum X_i\right) = \sum_{i=1}^{n} \mathbb{E}\left(X_i\right) = n\,\mathbb{E}\left(X\right) = np$$

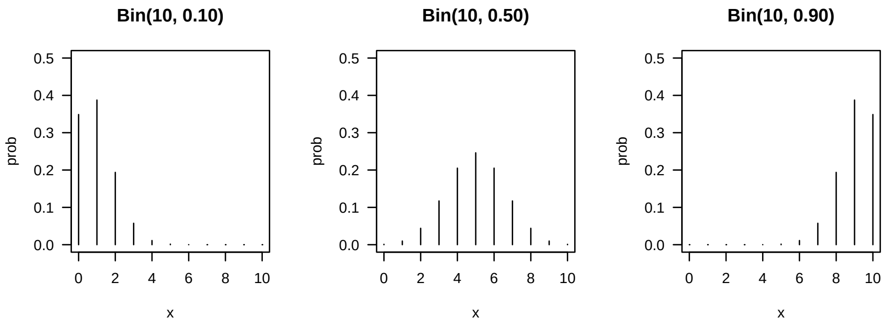


Figura 5.1: Forma distribuzione Bin (n, p)

Alternativamente potevamo sviluppare l'algebra:

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{(n-x)} = \sum_{x=0}^{n} x \cdot \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)} \\ &= \sum_{x=0}^{n} x \cdot \frac{n(n-1)!}{x(x-1)!\,[(n-1)-(x-1)]!} p p^{x-1} (1-p)^{[(n-1)-(x-1)]}\end{aligned}$$

Ora dato che per $x = 0$ il termine entro sommatoria è nullo possiamo portare avanti di uno l'indice inferiore della stessa:

$$\mathbb{E}(X) = \sum_{x=1}^{n} x \cdot \frac{n(n-1)!}{x(x-1)!\,[(n-1)-(x-1)]!} p p^{x-1} (1-p)^{[(n-1)-(x-1)]}$$

ponendo $y = x - 1$ si giunge

$$\begin{aligned}\mathbb{E}(X) &= np \sum_{y=0}^{n-1} \underbrace{\frac{(n-1)!}{y!\,[(n-1)-y]!} p^y (1-p)^{[(n-1)-y]}}_{\text{Bin}\,(n-1,p)} \\ &\overset{(1)}{=} np\end{aligned}$$

con (1) dato che la sommatoria è $= 1$. □

Dimostrazione. Sfruttando sempre il fatto che $X \sim \text{Bin}(n, p)$ sia descrivibile come la somma di n vc iid X_i Bern (p),

$$\text{Var}(X) = \text{Var}\left(\sum_i X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) = n\,\text{Var}(X) = n \cdot p(1-p) \qquad (5.22)$$

□

5.3.4 Forma della distribuzione

Proposizione 5.3.3 (Forma della distribuzione)**.** *La distribuzione è simmetrica se $p = 0.5$, è asimmetrica positiva (coda a destra) se $p < 0.5$, asimmetrica negativa (a sinistra) se $p > 0.5$. (Figura 5.1)*

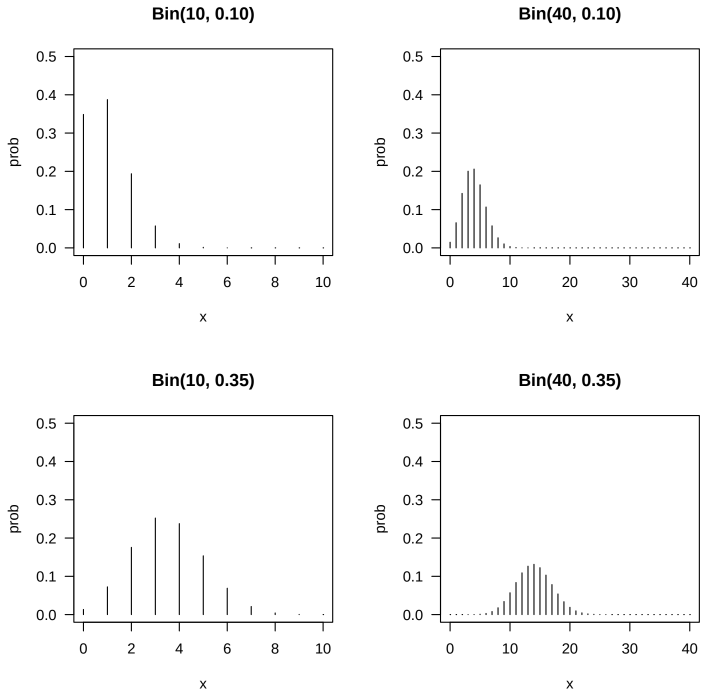


Figura 5.2: Convergenza alla normale della binomiale

Dimostrazione. Per $p = 0.5$ è simmetrica in quanto $p = 1 - p = \frac{1}{2}$ e

$$p_X(x) = \binom{n}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{n-x} = p_X(n-x) = \binom{n}{n-x}\left(\frac{1}{2}\right)^{n-x}\left(\frac{1}{2}\right)^x \tag{5.23}$$

per le proprietà del coefficiente binomiale. E dato che $p_X(x) = p_X(n-x)$, $\forall x \in R_X$, allora la distribuzione è simmetrica attorno al centro del supporto. $\square$

Proposizione 5.3.4. *In una binomiale di parametri n, p, la funzione di densità (per x che varia da 0 a n) è inizialmente strettamente crescente e successivamente strettamente decrescente. Si raggiunge il massimo in corrispondenza del più grande intero $x \leq (n+1)p$*

Dimostrazione. Consideriamo il rapporto $\mathrm{P}(X = x)/\mathrm{P}(X = x-1)$ e determiniamo per quali valori di x esso risulti maggiore (funzione crescente) o minore (decrescente) di 1:

$$\frac{\mathrm{P}(X = x)}{\mathrm{P}(X = x-1)} = \frac{\dfrac{n!}{(n-x)!x!}p^x(1-p)^{n-x}}{\dfrac{n!}{(n-x+1)!(x-1)!}p^{x-1}(1-p)^{n-x+1}} = \frac{(n-x+1)p}{x(1-p)}$$

Quindi tale rapporto ≥ 1 se e solo se:

$$(n-x+1)p \geq x(1-p)$$
$$np - xp + p \geq x - xp$$

ossia $x \leq (n+1)p$ □

Osservazione 116 (Convergenza alla normale)*.* La distribuzione converge verso la Normale (diviene simmetrica e la curtosi tende a 3) al crescere di $n \to \infty$; la convergenza è tanto più veloce per quanto più p è prossimo a 0.5. (figura 5.2)

5.3.5 Variabili derivate

Proposizione 5.3.5 (Vc numero di insuccessi)**.** *Sia* $X \operatorname{Bin}(n,p)$*. Allora* $n - X \sim \operatorname{Bin}(n, 1-p)$

Dimostrazione. Ad intuito basta invertire i ruoli di successo e insuccesso (si inverte anche la probabilità). Volendo tuttavia verificare, sia $Y = n - X$, la PMF è:

$$\begin{aligned} \mathrm{P}(Y = x) &\overset{(1)}{=} \mathrm{P}(X = n - x) = \binom{n}{n-x} p^{n-x}(1-p)^x \\ &\overset{(1)}{=} \binom{n}{x}(1-p)^x p^{n-x} = \operatorname{Bin}(n, 1-p) \end{aligned}$$

dove in (1) diciamo che in n estrazioni la probabilità di avere x fallimenti è uguale alla probabilità di avere $n - x$ successi , mentre in (2) abbiamo sfruttato la proprietà del coefficiente binomiale. □

Osservazione 117*.* Un fatto importante della binomiale è che la somma di binomiali indipendenti aventi la stessa probabilità di successo è un'altra binomiale

Proposizione 5.3.6 (Somma di binomiali)**.** *Se* $X \sim \operatorname{Bin}(n,p)$*,* $Y \sim \operatorname{Bin}(m,p)$ *e* X *è indipendente da* Y*, allora* $X + Y \sim \operatorname{Bin}(n+m, p)$

Dimostrazione. Un modo semplice è rappresentare X e Y come le somma di $X = X_1 + \ldots + X_n$ e $Y = Y_1 + \ldots + Y_n$ con $X_i, Y_i \sim \operatorname{Bern}(p)$ iid. Allora $X+Y$ è la somma di $n+m$ $\operatorname{Bern}(p)$ iid, pertanto la distribuzione è $\operatorname{Bin}(n+m,p)$ per teorema 5.3.1.
Viceversa, mediante la legge delle probabilità totali, possiamo trovare la PMF di $X+Y$ condizionando su X (oppure ugualmente su Y) e sommando:

$$\begin{aligned} \mathrm{P}(X+Y=k) &= \sum_{j=0}^{k} \mathrm{P}(X+Y=k|X=j) \cdot \mathrm{P}(X=j) \\ &= \sum_{j=0}^{k} \mathrm{P}(Y=k-j|X=j) \cdot \mathrm{P}(X=j) \\ &\overset{(1)}{=} \sum_{j=0}^{k} \mathrm{P}(Y=k-j) \cdot \mathrm{P}(X=j) \\ &= \sum_{j=0}^{k} \binom{m}{k-j} p^{k-j}(1-p)^{m-k+j} \cdot \binom{n}{j} p^j (1-p)^{n-j} \\ &= p^k(1-p)^{n+m-k} \sum_{j=0}^{k} \binom{m}{k-j}\binom{n}{j} \\ &\overset{(2)}{=} \binom{n+m}{k} p^k (1-p)^{n+m-k} = \operatorname{Bin}(n+m,p) \end{aligned}$$

dove in (1) abbiamo sfruttato l'indipendenza tra X e Y e in (2) l'identità di Vandermonde (eq 1.15). □

5.4 Ipergeometrica

5.4.1 Definizione

Osservazione 118. La variabile ipergeometrica descrive l'estrazione *senza reinserimento* di palline dicotomiche da un urna. A differenza della binomiale dove la probabilità di successo p non cambiava da una sottoprova Bernoulliana all'altra, qui il non reinserimento fa si che la probabilità di successo vari ad ogni prova.

Definizione 5.4.1 (Distribuzione ipergeometrica)**.** Supponiamo di dover estrarre un campione di n palline senza reinserimento da un'urna che contiene w palline bianche (successo) e b nere. Il numero X di palline bianche (successi) tra le estratte si distribuisce come una ipergeometrica con parametri w, b ed n e si scrive $X \sim \mathrm{HGeom}\,(w, b, n)$.

5.4.2 Funzioni

Osservazione 119 (Supporto e spazio parametrico).

$$\begin{aligned} R_X &= \{0, 1, \dots, n\} \\ \boldsymbol{\Theta} &= \{w, b \in \mathbb{N} : w + b \geq 1;\ n \in \{0, \dots, w + b\}\} \end{aligned}$$

Definizione 5.4.2 (Funzione di massa di probabilità)**.**

$$p_X(x) = \mathrm{P}\,(X = x) = \frac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}} \cdot \mathbb{1}_{R_X}(x) \tag{5.24}$$

Osservazione 120 (Interpretazione). Al denominatore sono quantificati il numero di modi con cui posso estrarre n palline qualsiasi dall'urna. Di queste estrazioni, al numeratore sono quantificati il numero di modi in cui nelle n palline estratte ci sono x bianche (successi); ossia devo averne x bianche scelte tra b, e $n - x$ nere scelte tra b.

Validità PMF. Facendo la somma del numeratore si ha:

$$\sum_{x=0}^{n} \binom{w}{x}\binom{b}{n-x} \overset{(1)}{=} \binom{w+b}{n}$$

con (1) per l'identità di Vandermonde (eq 1.15), per cui la PMF somma a 1. □

Osservazione 121. In R per la PMF si usa `dhyper(x, m, n, k)` dove `x` è il supporto (ossia il numero di palline bianche estratte), `m` il numero di palline bianche nell'urna, `n` il numero di palline nere e k il numero di estrazioni.

5.4.3 Momenti

Proposizione 5.4.1 (Momenti caratteristici)**.**

$$\mathbb{E}\left(X\right) = n\frac{w}{w+b} \tag{5.25}$$

$$\mathrm{Var}\left(X\right) = np(1-p)\left(\frac{w+b-n}{w+b-1}\right), \qquad \textit{con } p = \frac{w}{w+b} \tag{5.26}$$

Dimostrazione. Per il valore atteso, come nel caso binomiale possiamo scrivere X come somma di Bernoulliane $I_i \sim \mathrm{Bern}\left(p\right)$ con $p = w/(w+b)$.

$$X = I_1 + \ldots + I_n$$

A differenza della binomiale le I_i non sono indipendenti, tuttavia la linearità del valore atteso non lo richiede, quindi

$$\mathbb{E}\left(X\right) = \mathbb{E}\left(I_1 + \ldots + I_n\right) = \mathbb{E}\left(I_1\right) + \ldots + \mathbb{E}\left(I_n\right) = np = n\frac{w}{w+b}$$

□

Dimostrazione. Per la varianza invece essendo variabili non indipendenti non possiamo sommare le varianze direttamente. Vedremo in seguito la dimostrazione della formula riportata. □

5.4.4 Struttura essenziale ed esperimenti assimilabili

Osservazione 122. L'idea dell'Ipergeometrica è classificare una popolazione utilizzando due set di tag consecutivi (entrambi dicotomici successo/insuccesso) e ottenere il numero degli elementi caratterizzati dal successo in entrambi i tag. Nell'esempio delle palline il primo tag è il colore della pallina (bianco = successo), mentre il secondo è estrazione (estratta = sucesso).
Problemi aventi la stessa struttura presenteranno medesima distribuzione.

Esempio 5.4.1. Il numero A di assi estratti (sono 4 in un mazzo di 52 carte) in una mano di poker (5 carte estratte) si distribuirà come $A \sim \mathrm{HGeom}\left(4, 48, 5\right)$.

Osservazione 123. La struttura essenziale ci permette di dimostrare facilmente l'uguaglianza di due ipergeometriche dove l'ordine dei set di tag viene invertito

Proposizione 5.4.2. $\mathrm{HGeom}\left(w, b, n\right)$ *e* $\mathrm{HGeom}\left(n, w+b-n, w\right)$ *sono identiche.*

Dimostrazione. Sia $X \sim \mathrm{HGeom}\left(w, b, n\right)$ è il numero di palline bianche tra le estratte campione; sia $Y \sim \mathrm{HGeom}\left(n, w+b-n, w\right)$ il numero di palline estratte tra le bianche (pensando ad estratto/non estratto come il primo tag e al colore come secondo. Entrambe X, Y contano il numero di bianche estratte pertanto avranno la stessa distribuzione.
Alternativamente possiamo controllare algebricamente che

$$\mathrm{P}\left(X = x\right) = \frac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}} = \frac{\frac{w!}{k!(w-k)!}\frac{b!}{(n-k)!(b-n+k)!}}{\frac{(w+b)!}{n!(w+b-n)!}} = \frac{w!b!n!(w+b-n)!}{k!(w-k)!(n-k)!(b-n+k)!}$$

$$\mathrm{P}\left(Y = y\right) = \frac{\binom{n}{y}\binom{w+b-n}{w-y}}{\binom{w+b}{w}} = \frac{\frac{n!}{k!(n-k)!}\frac{(w+b-n)!}{(w-k)!(b-n+k)!}}{\frac{(w+b)!}{w!b!}} = \frac{w!b!n!(w+b-n)!}{k!(w-k)!(n-k)!(b-n+k)!}$$

e dunque $\mathrm{P}\left(X = x\right) = \mathrm{P}\left(Y = y\right)$. □

5.4.5 Connessioni con la binomiale

Osservazione 124. Binomiale ed ipergeometrica sono connesse: possiamo ottenere la binomiale calcolando un limite sull'ipergeometrica, oppure ottenere una ipergeometrica condizionando una binomiale.

5.4.5.1 Dall'ipergeometrica alla binomiale

Proposizione 5.4.3. *Se $X \sim \mathrm{HGeom}(w, b, n)$ e $w + b \to \infty$ ma $p = w/(w + b)$ rimane fisso, allora la PMF di X converge a $\mathrm{Bin}(n, p)$.*

Dimostrazione. Sviluppiamo algebricamente per essere comodi prima di applicare il limite:

$$\mathrm{P}(X = x) = \frac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}} \overset{(1)}{=} \binom{n}{x} \frac{\binom{w+b-n}{w-x}}{\binom{w+b}{w}}$$

dove in (1) abbiamo sfruttato che $\mathrm{HGeom}(w, b, n) = \mathrm{HGeom}(n, w + b - n, w)$ come nella dimostrazione di 5.4.2. Ora sviluppiamo il rapporto al secondo fattore ricordando che $\binom{n}{d} = \frac{n!}{d!(n-d)!}$; si ha:

$$\begin{aligned}
\frac{\binom{w+b-n}{w-x}}{\binom{w+b}{w}} &= \frac{(w+b-n)!}{(w-x)!(w+b-n-w+x)!} : \frac{(w+b)!}{w!(w+b-w)!} \\
&= \frac{(w+b-n)!}{(w-x)!(b-n+x)!} \cdot \frac{w!b!}{(w+b)!} \\
&= \frac{w!}{(w-x)!} \frac{b!}{(b-n+x)!} \frac{(w+b-n)!}{(w+b)!} \\
&= \frac{w \cdot \ldots \cdot (w-x+1)(w-x)!}{(w-x)!} \frac{b \cdot \ldots \cdot (b-n+x+1)(b-n+x)!}{(b-n+x)!} \frac{(w+b-n)!}{(w+b) \cdot \ldots \cdot (w+b-n+1)(w+b-n)!} \\
&= \frac{w \cdot \ldots \cdot (w-x+1)}{1} \frac{b \cdot \ldots \cdot (b-n+x+1)}{1} \frac{1}{(w+b) \cdot \ldots \cdot (w+b-n+1)}
\end{aligned}$$

ora al numeratore del primo rapporto abbiamo $w - (w - x + 1) + 1 = x$ fattori, al numeratore del secondo ne abbiamo $b - (b - n + x + 1) + 1 = n - x$ elementi. Pertanto complessivamente al numeratore abbiamo n fattori. Al denominatore invece abbiamo $(w + b) - (w + b - n + 1) + 1 = n$ fattori anche qui. Pertanto possiamo dividere per $(w + b)$, applicandolo n volte sia al numeratore che al denominatore, ottenendo

$$\frac{\binom{w+b-n}{w-x}}{\binom{w+b}{w}} = \frac{\frac{w}{w+b} \cdot \ldots \cdot \left(\frac{w}{w+b} - \frac{x-1}{w+b}\right) \cdot \left(\frac{b}{w+b}\right) \cdot \ldots \cdot \left(\frac{b}{w+b} - \frac{n-x-1}{w+b}\right)}{1 \cdot \ldots \cdot \left(1 - \frac{n-1}{w+b}\right)}$$

ora sostituendo $p = \frac{w}{w+b}$, $1 - p = \frac{b}{w+b}$ e al denominatore $w + b = N$ dove utile si ha:

$$\frac{\binom{w+b-n}{w-x}}{\binom{w+b}{w}} = \frac{p \cdot \ldots \cdot \left(p - \frac{x-1}{N}\right) \cdot (1-p) \cdot \ldots \cdot \left(1 - p - \frac{n-x-1}{N}\right)}{\left(1 - \frac{1}{N}\right) \ldots \left(1 - \frac{n-1}{N}\right)}$$

Ora tornando da dove siamo partiti abbiamo:

$$\mathrm{P}(X=x)=\binom{n}{x}\frac{p\cdot\ldots\cdot\left(p-\frac{x-1}{N}\right)\cdot(1-p)\cdot\ldots\cdot\left(1-p-\frac{n-x-1}{N}\right)}{\left(1-\frac{1}{N}\right)\ldots\left(1-\frac{n-1}{N}\right)}$$

Infine per $N\to+\infty$ il denominatore va a 1 mentre il numeratore va a $p^x(1-p)^{n-x}$ pertanto

$$\mathrm{P}(X=x)\to\binom{n}{x}p^x(1-p)^{n-x}$$

che è la $\mathrm{Bin}(n,p)$.
Intuitivamente data un'urna con w palline bianche e b nere, la binomiale sorge dall'estrarre n palline con replacement, mentre l'ipergeometrica senza. Se il numero di palline nell'urna sale notevolmente rispetto al numero di palline estratte, il campionamento con ripetizione e senza diventano essenzialmente equivalenti. (l'estrazione di una pallina non cambia la probabilità delle prossime estrazioni perché data la grande numerosità nell'urna non modifica praticamente la probabilità di successo) □

Osservazione 125. In termini pratici il teorema ci dice che se $N=w+b$ è grande rispetto a n possiamo approssimare la PMF di $\mathrm{HGeom}(w,b,n)$ con $\mathrm{Bin}(n,w/(w+b))$.

5.4.5.2 Dalla binomiale all'ipergeometrica

Proposizione 5.4.4. *Se $X\sim\mathrm{Bin}(n,p)$, $Y\sim\mathrm{Bin}(m,p)$ e X è indipendente da Y, allora la distribuzione condizionata di X dato che $X+Y=r$ è* $\mathrm{HGeom}(n,m,r)$

Osservazione 126. Dimostriamo attraverso un esempio (distribuzione del test esatto di Fisher).

Dimostrazione. Un ricercatore vuole studiare se la prevalenza di una data malattia sia uguale o meno tra maschi e femmine. Raccoglie un campione di n donne ed m uomini e testa la malattia. Sia $X\sim\mathrm{Bin}(n,p_1)$ il numero di donne con la malattia nel campione e $Y\sim\mathrm{Bin}(n,p_2)$ il numero di uomini. Qui p_1 e p_2 sono sconosciuti.
Supponiamo che siano osservate $X+Y=r$ persone malate. Siamo interessati a testare se $p_1=p_2=p$ (la cd ipotesi nulla); il test di Fisher si fonda sul condizionare sui totali di riga e colonna (quindi n,m,r sono considerati fissi) e verificare se il valore osservato X (numero di donne malate) sia estremo (dato che il tot malati è r) sotto ipotesi nulla. Assumendo l'ipotesi nulla vera troviamo la PMF condizionale di X dato che $X+Y=r$.
La tabella 2×2 di riferimento è la 5.2. Costruiamo PMF condizionata attraverso la regola di Bayes:

$$\mathrm{P}(X=x|X+Y=r)=\frac{\mathrm{P}(X+Y=r|X=x)\,\mathrm{P}(X=x)}{\mathrm{P}(X+Y=r)}=\frac{\mathrm{P}(Y=r-x|X=x)\,\mathrm{P}(X=x)}{\mathrm{P}(X+Y=r)}$$
$$\overset{(1)}{=}\frac{\mathrm{P}(Y=r-x)\,\mathrm{P}(X=x)}{\mathrm{P}(X+Y=r)}$$

dove in (1) abbiamo sfruttato l'indipendenza di X e Y. Assumendo per buona l'ipotesi nulla e impostando $p_1=p_2=p$ si hanno le vc indipendenti $X\sim$

	Donne	Uomini	Tot
Malato	x	$r-x$	r
Sano	$n-x$	$m-r+x$	$n+m-r$
Tot	n	m	$n+m$

Tabella 5.2

$\mathrm{Bin}(n,p)$ e $Y \sim \mathrm{Bin}(m,p)$, per cui $X+Y \sim \mathrm{Bin}(n+m,p)$ (per il risultato 5.3.6). Pertanto sostituendo le formule per esteso si ha

$$\begin{aligned}\mathrm{P}(X=x|X+Y=r) &= \frac{\binom{m}{r-x}p^{r-x}(1-p)^{m-r+x}\cdot\binom{n}{x}p^x(1-p)^{n-x}}{\binom{n+m}{r}p^r(1-p)^{n+m-r}} \\ &= \frac{\binom{n}{x}\binom{m}{r-x}}{\binom{n+m}{r}} = \mathrm{HGeom}(n,m,r)\end{aligned}$$

Intuitivamente questo avviene perché condizionatamente ad avere $X+Y=r$ malati (primo tag), X è il numero di donne (secondo tag) tra quelli. □

5.5 Geometrica

5.5.1 Definizione

Osservazione 127. Supponiamo di ripetere in maniera indipendente diverse prove bernoulliane, ciascuna avente p probabilità di successo, sino a che che si verifica il primo successo. Sia X il numero di *fallimenti* necessari per ottenere il primo successo; X si distribuisce come una variabile geometrica con parametro p e si scrive $X \sim \mathrm{Geom}(p)$.

Esempio 5.5.1. Il numero di croci sino alla prima testa si distribuisce come $\mathrm{Geom}(1/2)$.

5.5.2 Funzioni

Osservazione 128 (Supporto e spazio parametrico).

$$\begin{aligned}R_X &= \{x \in \mathbb{N}\} \\ \boldsymbol{\Theta} &= \{p \in (0,1)\}\end{aligned}$$

Definizione 5.5.1 (Funzione di massa di probabilità)**.**

$$p_X(x) = \mathrm{P}(X=x) = (1-p)^x p \cdot \mathbb{1}_{R_X}(x) \tag{5.27}$$

Validità PMF. Si ha che

$$\sum_{x=0}^{\infty}(1-p)^x p = p\sum_{x=0}^{\infty}(1-p)^x \overset{(1)}{=} p\cdot\frac{1}{p} = 1$$

con l'uguaglianza (1) dovuta alla serie geometrica. □

Osservazione 129. Come il teorema binomiale mostra che la PMF binomiale sia valida, la serie geometrica mostra che la PMF Geometrica sia valida.

Osservazione 130 (Interpretazione). La probabilità di avere x fallimenti consecutivi seguiti da un successo è data dalla probabilità di x fallimenti per la probabilità di un successo.

Definizione 5.5.2 (Funzione di ripartizione)**.** Si ha

$$F_X(x) = \mathrm{P}\left(X \leq x\right) = 1 - (1-p)^{x+1} \tag{5.28}$$

Derivazione della CDF. Si ha

$$F_X(x) = \mathrm{P}\left(X \leq x\right) = 1 - \mathrm{P}\left(X > x\right) = 1 - \sum_{k=x+1}^{\infty} (1-p)^k p$$

Espandendo la sommatoria:

$$\begin{aligned}
\sum_{k=x+1}^{\infty} (1-p)^k p &= (1-p)^{x+1} \cdot p + (1-p)^{x+2} \cdot p + \ldots + (1-p)^{\infty} \cdot p \\
&= p(1-p)^x \left[(1-p) + (1-p)^2 + \ldots + (1-p)^{\infty}\right] \\
&= p(1-p)^x \left[\sum_{i=1}^{\infty}(1-p)^i\right] \\
&= p(1-p)^x \left[\sum_{i=0}^{\infty}(1-p)^i - 1\right] \\
&= p(1-p)^x \left(\frac{1}{p} - 1\right) = p(1-p)^x \frac{1-p}{p} \\
&= (1-p)^{x+1}
\end{aligned}$$

Pertanto:

$$F_X(x) = 1 - (1-p)^{x+1}$$

□

5.5.3 Momenti

Proposizione 5.5.1 (Momenti caratteristici)**.**

$$\mathbb{E}\left(X\right) = \frac{1-p}{p}$$
$$\mathrm{Var}\left(X\right) = \frac{1-p}{p^2}$$

Dimostrazione. Per il valore atteso abbiamo

$$\mathbb{E}\left(X\right) = \sum_{x=0}^{\infty} x \cdot (1-p)^x p$$

Non può essere ricondotta a serie geometrica direttamente per la presenza entro sommatoria di x come primo fattore. Ma notiamo che il termine entro sommatoria assomiglia a $x(1-p)^{x-1}$ ossia la derivata di $(1-p)^x$ rispetto a $1-p$, quindi partiamo da li:

$$\sum_{x=0}^{\infty}(1-p)^x = \frac{1}{p}$$

Questa serie converge dato che $0 < p < 1$. Derivando entrambi i membri rispetto a p.

$$\sum_{x=0}^{\infty} x(1-p)^{x-1} \cdot (-1) = -\frac{1}{p^2}$$

$$\sum_{x=0}^{\infty} x(1-p)^{x-1} = \frac{1}{p^2}$$

e se moltiplichiamo entrambi i lati per $p(1-p)$ otteniamo la somma dalla quale siamo partiti

$$p(1-p)\sum_{x=0}^{\infty} x(1-p)^{x-1} = \frac{1}{p^2}p(1-p)$$

$$\sum_{x=0}^{\infty} xp(1-p)^x = \frac{1-p}{p}$$

□

Dimostrazione. Per la varianza dobbiamo calcolare $\mathbb{E}\left(X^2\right)$:

$$\mathbb{E}\left(X^2\right) = \sum_{x=0}^{\infty} x^2 \cdot \mathrm{P}\left(X = x\right) = \sum_{x=0}^{\infty} x^2 \cdot (1-p)^x \cdot p \overset{(1)}{=} \sum_{x=1}^{\infty} x^2 \cdot (1-p)^x \cdot p$$

con (1) dato dal fatto che se $x = 0$ il termine entro sommatoria è nullo e si può portare avanti l'indice della stessa. Anche qui cerchiamo di sfruttare la serie geometrica per arrivare ad una espressione compatta equivalente all'ultimo termine di sopra. La serie è

$$\sum_{x=0}^{\infty}(1-p)^x = \frac{1}{p}$$

Derivando rispetto a p entrambi i membri, come visto in precedenza si ha:

$$\sum_{x=0}^{\infty} x \cdot (1-p)^{x-1} = \frac{1}{p^2}$$

Possiamo portare avanti di 1 l'indice di sommatoria dato che se $x = 0$ è nullo il termine dentro

$$\sum_{x=1}^{\infty} x \cdot (1-p)^{x-1} = \frac{1}{p^2}$$

Ora, derivando ancora si andrebbe a $x(x-1)$ entro sommatoria, invece di x^2 desiderato, pertanto moltiplichiamo per $(1-p)$ entrambi i membri giungendo a:

$$\sum_{x=1}^{\infty} x \cdot (1-p)^x = \frac{1-p}{p^2}$$

Derivando ambo i membri nuovamente rispetto a p si va a

$$\sum_{x=1}^{\infty} x^2 \cdot (1-p)^{x-1} \cdot (-1) = \frac{(-1) \cdot p^2 - 2p \cdot (1-p)}{p^4}$$
$$\sum_{x=1}^{\infty} x^2 \cdot (1-p)^{x-1} = (-1)\frac{p^2 - 2p}{p^4}$$
$$\sum_{x=1}^{\infty} x^2 \cdot (1-p)^{x-1} = \frac{2-p}{p^3}$$

Moltiplicando entrambi i membri per $(1-p) \cdot p$ si arriva al punto dove eravamo rimasti con $\mathbb{E}\left(X^2\right)$

$$\sum_{x=1}^{\infty} x^2 \cdot (1-p)^x \cdot p = \frac{2-p}{p^3} \cdot (1-p) \cdot p = \frac{(2-p)(1-p)}{p^2}$$

Per cui

$$\mathbb{E}\left(X\right) = \sum_{x=1}^{\infty} x^2 \cdot (1-p)^x \cdot p = \frac{(2-p)(1-p)}{p^2}$$

e dunque:

$$\begin{aligned} \operatorname{Var}\left(X\right) &= \mathbb{E}\left(X^2\right) - \left(\mathbb{E}\left(X\right)\right)^2 = \frac{(2-p)(1-p)}{p^2} - \frac{(1-p)^2}{p^2} \\ &= \frac{(1-p)(2-p-1+p))}{p^2} = \frac{1-p}{p^2} \end{aligned}$$

□

5.5.4 Forma della distribuzione

Osservazione 131 (Forma della distribuzione). Tutte le geometriche hanno forma simile: la funzione è decrescente, con probabilità più alte associate ai valori più piccoli di x. Ha asimmetria positiva che aumenta al crescere di p (più p è alto più velocemente la PMF discende verso 0). Ha una notevole curtosi (figura 5.3)

5.5.5 Assenza di memoria

Osservazione 132. Una proprietà peculiare della geometrica è di esser l'unica vc discreta senza memoria (a parte la sua riformulazione).

Proposizione 5.5.2 (Assenza di memoria)**.**

$$\mathrm{P}\left(X > t+s | X > t\right) = \mathrm{P}\left(X > s\right) \tag{5.29}$$

Dimostrazione. Si ha:

$$\begin{aligned} \mathrm{P}\left(X > t+s | X > t\right) &= \frac{\mathrm{P}\left(X > t+s\right)}{\mathrm{P}\left(X > t\right)} = \frac{1 - F_X(t+s)}{1 - F_X(t)} = \frac{1 - 1 + (1-p)^{t+s+1}}{1 - 1 + (1-p)^{t+1}} \\ &= (1-p)^s = 1 - F_X(s) = \mathrm{P}\left(X > s\right) \end{aligned}$$

□

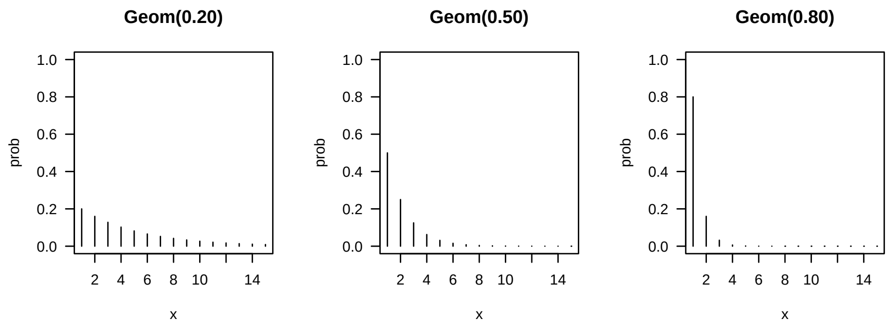


Figura 5.3: Forma distribuzione Geom (p)

5.5.6 Definizione alternativa (first success distribution)

Osservazione 133. Altri definiscono X come il numero di *prove* necessarie per ottenere il primo successo (incluso quest'ultimo). Qui la chiamiamo FS distribution e la indichiamo con $X \sim \mathrm{FS}\,(p)$

Osservazione 134. Se $Y \sim \mathrm{FS}\,(p)$ allora $Y-1 \sim \mathrm{Geom}\,(p)$ e possiamo convertire tra le PMF di Y e $Y-1$ scrivendo

$$\mathrm{P}\,(Y = k) = \mathrm{P}\,(Y - 1 = k - 1)$$

Viceversa se $X \sim \mathrm{Geom}\,(p)$ allora $X + 1 \sim \mathrm{FS}\,(p)$

Osservazione 135 (Supporto e spazio parametrico).

$$R_X = \{x \in \mathbb{N} \setminus \{0\}\}$$
$$\boldsymbol{\Theta} = \{p \in (0, 1)\}$$

Definizione 5.5.3 (Funzione di massa di probabilità)**.**

$$p_X(x) = \mathrm{P}\,(X = x) = (1-p)^{x-1} p \cdot \mathbb{1}_{R_X}(x) \tag{5.30}$$

Osservazione 136 (Interpretazione). La probabilità di avere il primo successo all'n-esima estrazione e data dalla probabilità di $n-1$ fallimenti per la probabilità di un successo.

Definizione 5.5.4 (Funzione di ripartizione)**.**

$$\begin{aligned} F_X(x) = \mathrm{P}\,(X \leq x) &= \sum_{k=1}^{x} \mathrm{P}\,(X = k) = \sum_{k=1}^{x} (1-p)^{k-1} p \\ &= 1 - (1-p)^x \end{aligned} \tag{5.31}$$

Proposizione 5.5.3 (Momenti caratteristici)**.**

$$\begin{aligned}
\mathbb{E}(X) &= \frac{1}{p} \\
\mathrm{Var}(X) &= \frac{1-p}{p^2} \\
\mathrm{Asym}(X) &= \frac{2-p}{\sqrt{1-p}} \\
\mathrm{Kurt}(X) &= 9 + \frac{p^2}{1-p}
\end{aligned}$$

Dimostrazione. Sia $Y = X + 1 \sim \mathrm{FS}(p)$ con $X \sim \mathrm{Geom}(p)$. Allora sfruttando le conoscenze sulla geometrica e le proprietà di valore atteso e varianza

$$\begin{aligned}
\mathbb{E}(Y) &= \mathbb{E}(X+1) = \mathbb{E}(X) + 1 = \frac{1-p}{p} + 1 = \frac{1}{p} \\
\mathrm{Var}(Y) &= \mathrm{Var}(X+1) = \mathrm{Var}(X) = \frac{1-p}{p^2}
\end{aligned}$$

□

Proposizione 5.5.4 (Assenza di memoria)**.** *Analogamente a quanto avviene per la geometrica* $\mathrm{P}(X > t+s | X > t) = \mathrm{P}(X > s)$.

Dimostrazione. Si ha:

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

ovvero il ritardo accertato di un evento in t sottoprove indipendenti non modifica la probabilità che esso si verifichi entro ulteriori s sottoprove. □

5.6 Binomiale Negativa

Osservazione 137. Generalizza la distribuzione Geometrica: invece di aspettare il primo successo conta i fallimenti prima di ottenere il k-esimo successo.

5.6.1 Definizione

Definizione 5.6.1. In una sequenza di prove Bernoulliane indipendenti con probabilità di successo p, se X è il numero di fallimenti prima del k-esimo successo, allora X ha una distribuzione binomiale negativa con parametri k e p e si scrive $X \sim \mathrm{Bn}(k, p)$

Osservazione 138. Anche a livello di notazione, nei parametri, si nota subito la differenza con la binomiale: questa fissa il numero di trial mentre la binomiale negativa fissa il numero di successi.

5.6.2 Funzioni

Osservazione 139 (Supporto e spazio parametrico).

$$R_X = \mathbb{N}$$
$$\mathbf{\Theta} = \{k \in \mathbb{N} : k \geq 1, p \in \mathbb{R} : 0 \leq p \leq 1\}$$

Definizione 5.6.2 (Funzione di massa di probabilità)**.**

$$p_X(x) = \mathrm{P}\left(X = x\right) = \binom{x+k-1}{k-1} p^k (1-p)^x \cdot \mathbb{1}_{R_X}(x) \tag{5.32}$$

Osservazione 140 (Interpretazione). Ci sono $\binom{x+k-1}{k-1}$ sequenze possibili di x fallimenti e $k-1$ successi. Ciascuna di esse ha probabilità $p^{k-1}(1-p)^x$. Si termina con un success, quindi moltiplicando per p.

Osservazione 141. Come una binomiale può essere rappresentata da una somma di Bernoulli iid, una binomiale negativa può essere rappresentata come somma di Geometriche iid, come mostrato dal seguente teorema.

Proposizione 5.6.1. *Sia $X \sim \mathrm{Bn}\left(k, p\right)$ il numero di fallimenti prima del k-esimo successo in una sequenza di probe bernoulliane indipendenti con probabilità di successo p. Allora possiamo scrivere $X = X_1 + \ldots + X_k$ dove gli X_i sono iid e $X_i \sim \mathrm{Geom}\left(p\right)$.*

Dimostrazione. Sia X_1 il numero di fallimenti prima del primo successo, X_2 il numero di fallimenti tra il primo successo e il secondo e, in generale, X_i il numero di fallimenti tra $(i-1)$-esimo succeso e l'i-esimo.
Allora $X_1 \sim \mathrm{Geom}\left(p\right)$ per la definizione della geometrica, $X_2 \sim \mathrm{Geom}\left(p\right)$ e così via. Inoltre le X_i sono indipendenti dato che le prove bernoulliane sono indipendenti l'un l'altra. Sommando gli X_i si ottiene il totale di fallimenti prima del k-esimo successo, che è X. □

5.6.3 Momenti

Proposizione 5.6.2 (Momenti caratteristici)**.**

$$\mathbb{E}\left(X\right) = k\frac{1-p}{p} \tag{5.33}$$

$$\mathrm{Var}\left(X\right) = k\frac{1-p}{p^2} \tag{5.34}$$

Dimostrazione. Per il valore atteso sfruttiamo che X è scrivibile come somma di k vc Geometriche X_i. Il valore atteso è la somma dei valori attesi delle geometriche:

$$\mathbb{E}\left(X\right) = \mathbb{E}\left(X_1 + \ldots + X_k\right) = \mathbb{E}\left(X_1\right) + \ldots \mathbb{E}\left(X_k\right) = k\frac{1-p}{p}$$

Per la varianza avviene lo stesso, dato che le variabili sono indipendenti:

$$\mathrm{Var}\left(X\right) = \mathrm{Var}\left(X_1 + \ldots + X_k\right) = \mathrm{Var}\left(X_1\right) + \ldots \mathrm{Var}\left(X_k\right) = k\frac{1-p}{p^2}$$

□

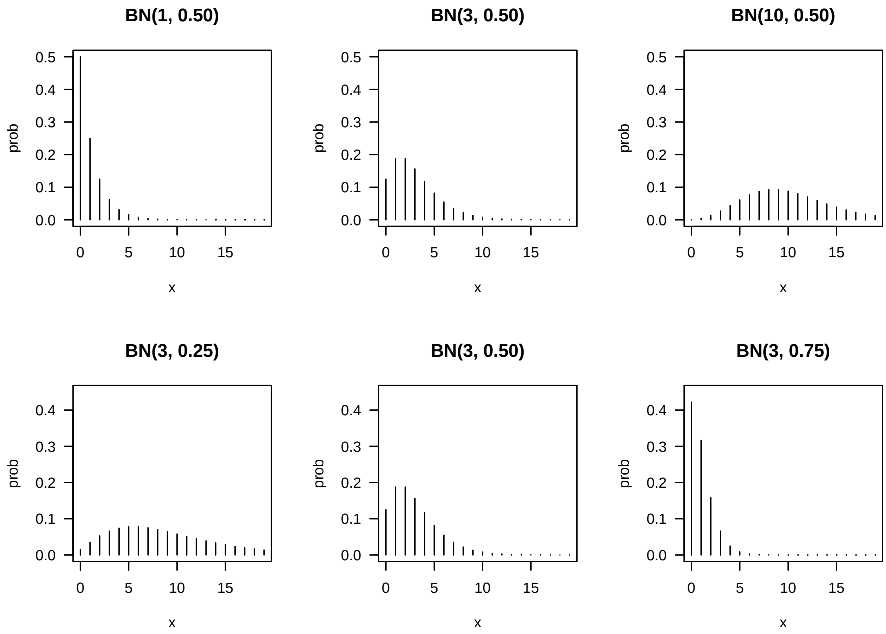


Figura 5.4: Distribuzione binomiale negativa

5.6.4 Forma della distribuzione

Osservazione 142 (Forma della distribuzione). Si nota che così al crescere di k, la distribuzione diviene più simmetrica e la curtosi tende a 3 indicando convergenza alla normalità. All'aumentare di p assume asimmetria positiva. (figura 5.4)

5.6.5 Definizione alternativa

5.6.5.1 Definizione

Definizione 5.6.3 (Distribuzione binomiale negativa)**.** Il numero di prove indipendenti X (ciascuna con probabilità p di essere successo) necessarie per avere $k \geq 1$ successi si distribuisce come una binomiale negativa di parametri k e p, ossia $X \sim \mathrm{Bn}\,(k, p)$.

5.6.5.2 Funzioni

Osservazione 143 (Supporto e spazio parametrico).

$$R_X = \{k, k+1, \ldots\}$$
$$\mathbf{\Theta} = \{k \in \mathbb{N} \setminus \{0\}\,, p \in \mathbb{R} : 0 \leq p \leq 1\}$$

Definizione 5.6.4 (Funzione di massa di probabilità)**.**

$$p_X(x) = \mathrm{P}\,(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k} \cdot \mathbb{1}_{R_X}(x) \tag{5.35}$$

Osservazione 144 (Interpretazione). La formula deriva dalla considerazione che per ottenere il k-esimo successo nella n-esima prova, ci dovranno essere $k-1$ successi nelle prime $n-1$ prove, la cui probabilità

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

è moltiplicata per la probabilità di un successo nella n-esima, ossia p.

5.6.5.3 Momenti

Proposizione 5.6.3 (Momenti caratteristici)**.**

$$\begin{aligned}
\mathbb{E}(X) &= \frac{k}{p} \\
\operatorname{Var}(X) &= \frac{k(1-p)}{p^2} \\
\operatorname{Asym}(X) &= \frac{2-p}{\sqrt{k(1-p)}} \\
\operatorname{Kurt}(X) &= 3 + \frac{6}{k} + \frac{p^2}{k(1-p)}
\end{aligned}$$

5.7 Poisson

5.7.1 Definizione

Osservazione 145. È una vc utilizzabile per modellare conteggi (motivo per cui il supporto è $\mathbb{N}$); sull'origine definizione ragioniamo in seguito. Per ora ci accontentiamo di definire la Poisson come la distribuzione caratterizzata dalle funzioni presentate in seguito: se la vc X è distribuita come una Poisson con parametro λ scriveremo $X \sim \operatorname{Pois}(\lambda)$.

Osservazione 146. Un risultato che ci servirà per questa distribuzione è il seguente

Proposizione 5.7.1 (Sviluppo di Maclaurin della funzione esponenziale)**.**

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{5.36}$$

Dimostrazione. Si ha:

$$e^x = e^0 + \frac{e^0}{1!}(x-0) + \frac{e^0}{2!}(x-0)^2 + \ldots \frac{e^0}{m!}(x-0)^m + \ldots = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

□

5.7.2 Funzioni

Osservazione 147 (Supporto e spazio parametrico).

$$\begin{aligned}
R_X &= \mathbb{N} \\
\boldsymbol{\Theta} &= \{\lambda \in \mathbb{R} : \lambda > 0\}
\end{aligned}$$

Definizione 5.7.1 (Funzione di massa di probabilità)**.**

$$p_X(x) = \mathrm{P}\left(X = x\right) = \frac{e^{(-\lambda)} \cdot \lambda^x}{x!} \cdot \mathbb{1}_{R_X}(x) \tag{5.37}$$

Validità PMF. Si ha:

$$\sum_{x=0}^{\infty} p_X(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \stackrel{(1)}{=} e^{-\lambda} e^{\lambda} = 1$$

dove in (1) abbiamo sfruttato la 5.36 con le dovute sostituzioni di lettere. □

5.7.3 Momenti

Proposizione 5.7.2 (Momenti caratteristici)**.**

$$\mathbb{E}\left(X\right) = \lambda \tag{5.38}$$

$$\mathrm{Var}\left(X\right) = \lambda \tag{5.39}$$

$$\mathrm{Asym}\left(X\right) = \frac{1}{\sqrt{\lambda}} \tag{5.40}$$

$$\mathrm{Kurt}\left(X\right) = 3 + \frac{1}{\lambda} \tag{5.41}$$

Dimostrazione. Per il valore atteso

$$\begin{aligned} \mathbb{E}\left(X\right) &= \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^x}{x!} \stackrel{(1)}{=} e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &\stackrel{(2)}{=} \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = \lambda e^{-\lambda} e^{\lambda} = \lambda \end{aligned}$$

dove in (1) abbiamo anche portato avanti di 1 la sommatoria dato che il primo termine è nullo e in (2) abbiamo sostituito $y = x - 1$ e sfruttato 5.36. □

Dimostrazione. Per la varianza troviamo innanzitutto $\mathbb{E}\left(X^2\right)$:

$$\mathbb{E}\left(X^2\right) = \sum_{x=0}^{\infty} x^2 \cdot \mathrm{P}\left(X = x\right) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!}$$

Ora prendiamo la serie dell'esponenziale e la deriviamo rispetto a λ ad entrambi i membri (x costante)

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \stackrel{(1)}{=} \sum_{x=0}^{\infty} x \frac{\lambda^{x-1}}{x!} \stackrel{(2)}{=} \sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{x!}$$

dove in (1) abbiamo effettuato la derivazione (il primo membro rimane invariato), in (2) abbiamo portato avanti l'indice di sommatoria perché il primo termine è nullo. Ora moltiplicando per λ entrambi i lati si ottiene

$$\lambda e^{\lambda} = \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!}$$

Effettuando gli stessi passaiggi, nell'ordine derivare entrambi i membri rispetto a λ e moltiplicandoli per λ si prosegue come

$$\sum_{x=1}^{\infty} x^2 \frac{\lambda^{x-1}}{x!} = e^{\lambda} + \lambda e^{\lambda} = e^{\lambda}(1+\lambda)$$
$$\sum_{x=1}^{\infty} x^2 \frac{\lambda^{x}}{x!} = e^{\lambda}\lambda(1+\lambda)$$

E infine riprendendo da dove eravamo arrivati con la main quest

$$\mathbb{E}\left(X^2\right) = e^{-\lambda}\sum_{x=0}^{\infty} x^2 \frac{\lambda^{x}}{x!} = e^{-\lambda}e^{\lambda}\lambda(1+\lambda) = \lambda(1+\lambda)$$

per cui

$$\mathrm{Var}\left(X\right) = \mathbb{E}\left(X^2\right) - \left(\mathbb{E}\left(X\right)\right)^2 = \lambda(1+\lambda) - \lambda^2 = \lambda$$

□

Dimostrazione. Dimostrazione alternativa per la varianza:

$$\begin{aligned}
\mathrm{Var}\left(X\right) &= \mathbb{E}\left(X^2\right) - \left[\mathbb{E}\left(X\right)\right]^2 \\
&= \left(\sum_{x=0}^{\infty} x^2 \cdot \frac{e^{-\lambda}\lambda^x}{x!}\right) - \lambda^2 \\
&= \left(\sum_{x=0}^{\infty} (x^2 + x - x) \cdot \frac{e^{-\lambda}\lambda^x}{x!}\right) - \lambda^2 \\
&= \left(\sum_{x=0}^{\infty} (x(x-1) + x) \cdot \frac{e^{-\lambda}\lambda^x}{x!}\right) - \lambda^2 \\
&= \left(\sum_{x=0}^{\infty} (x(x-1))\frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!}\right) - \lambda^2 \\
&= \left(\sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^2\lambda^{x-2}}{x(x-1)(x-2)!} + \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda\lambda^{x-1}}{x(x-1)!}\right) - \lambda^2 \\
&= \left(\sum_{x=0}^{\infty} \frac{\lambda^{x-2}}{(x-2)!}e^{-\lambda}\lambda^2 + \sum_{x=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}e^{-\lambda}\lambda\right) - \lambda^2 \\
&\overset{(1)}{=} \left(\sum_{z=0}^{\infty} \frac{\lambda^{z}}{z!}e^{-\lambda}\lambda^2 + \sum_{y=0}^{\infty} \frac{\lambda^{y}}{y!}e^{-\lambda}\lambda\right) - \lambda^2 \\
&= (e^{\lambda}e^{-\lambda}\lambda^2 + e^{\lambda}e^{-\lambda}\lambda) - \lambda^2 \\
&= (\lambda^2 + \lambda) - \lambda^2 \\
&= \lambda
\end{aligned}$$

dove in (1) abbiamo posto $y = x-1$, $z = x-2$ per sfruttare 5.36 nel seguito. □

5.7.4 Forma della distribuzione

Osservazione 148 (Forma della distribuzione). Quindi valore medio e varianza della vc di Poisson coincidono con il parametro λ; la distribuzione ha picco

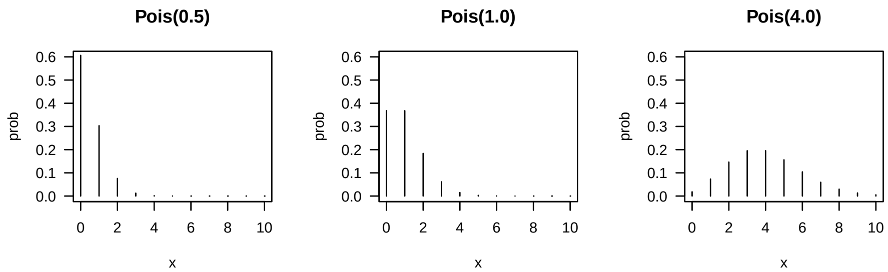


Figura 5.5: Distribuzione Poisson

intorno a λ. Al crescere di questo, la distribuzione diventa più simmetrica e la curtosi tende a 3 (convergendo ad una Normale). Se $\lambda < 1$ la distribuzione ha un andamento decrescente, mentre se > 1 è prima crescente e poi decrescente. (figura 5.5)

5.7.5 Origine e approssimazione

Osservazione 149. È utilizzata per modellare il numero di eventi registrati in un ambito circoscritto (temporale o spaziale), in cui vi è un largo numero di prove indipendenti (o quasi) caratterizzate ciascuna da una bassa probabilità di successo (per questa è chiamata legge degli eventi rari)

Proposizione 5.7.3 (Paradigma di Poisson)**.** *Siano $E_1, \ldots, E_n$ eventi con $p_i = \mathrm{P}(E_i)$, dove n è largo, p_i sono piccoli e gli E_i sono vc indipendenti o debolmente dipendenti. Sia*

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

la somma di quanti eventi E_i siano accaduti. Allora X è abbastanza bene distribuita come una $\mathrm{Pois}(\lambda)$ *con $\lambda = \sum_i p_i$.*

Dimostrazione. La prova dell'approssimazione di sopra è complessa, richiede definire la dipendenza debole e buona approssimazione; è omessa qui. □

Osservazione 150 (Ruolo di λ). Il parametro λ è interpretato come *rate di occorrenza*: ad esempio $\lambda = 2$ mail di spam per giorno.

Osservazione 151. Nell'esempio sopra il numero di eventi X non è esattamente distribuito come Poisson perché una variabile di Poisson non ha limite superire, mentre $I_{E_1} + \ldots + I_{E_n}$ somma al più a n. Ma la distribuzione di Poisson da spesso una buona approssimazione e le condizioni per il verificarsi della situazione di sopra sono abbastanza flessibili: infatti i p_i non devono essere uguali e le prove non devono essere strettamente indipendenti. Questo fa sì che il modello di Poisson sia spesso un buon punto di partenza per dati che assumono valore intero non negativo (chiamati conteggi)
È comunque possibile quantificare l'errore commesso.

Proposizione 5.7.4 (Errore di approssimazione)**.** *Se E_i sono indipendenti e sia $N \sim$* $\mathrm{Pois}(\lambda)$*, allora l'errore di approssimazione che si fa nell'utilizzare la* *poisson per stimare la probabilità di un dato set di interi non negativi $I \subset \mathbb{N}$, è dato dalla seguente:*

$$\mathrm{P}(X \in I) - \mathrm{P}(N \in I) \leq \min\left(1, \frac{1}{\lambda}\right) \sum_{i=1}^{n} p_i^2 \tag{5.42}$$

Dimostrazione. Anche questa è per ora complessa (necessita di una tecnica chiamata metodo di Stein). □

Osservazione 152. La 5.42 fornisce un limite superiore dell'errore commesso nell'utilizzare unaapprossimazione di Poisson: non solo per l'intera distribuzione (se $I = \mathbb{N}$) ma per qualsiasi suo sottoinsieme. Altresì precisa quanto i p_i dovrebbero essere piccoli: vogliamo che $\sum_{i=1}^{n} p_i^2$ sia molto piccolo, o quanto meno lo sia rispetto a λ.

5.7.6 Legami con la binomiale

Osservazione 153. La relazione tra Poisson e Binomiale è simile a quella intercorrente tra Binomiale e Ipergeometrica: possiamo andare dalla Poisson alla binomiale condizionando, e viceversa dalla Binomiale alla Poisson prendendo un limite. Prima un risultato strumentale.

Proposizione 5.7.5 (Somma di Poisson indipendenti)**.** *Siano* $X \sim \mathrm{Pois}(\lambda_1)$ *e* $Y \sim \mathrm{Pois}(\lambda_2)$ *vc indipendenti. Allora* $X + Y \sim \mathrm{Pois}(\lambda_1 + \lambda_2)$

Dimostrazione. Per ottenere la PMF di $X+Y$ condizioniamo su X e utilizziamo il teorema delle probabilità totali

$$\begin{aligned}
\mathrm{P}(X + Y = k) &= \sum_{j=0}^{k} \mathrm{P}(X + Y = k | X = j) \cdot \mathrm{P}(X = j) \\
&= \sum_{j=0}^{k} \mathrm{P}(Y = k - x | X = j) \cdot \mathrm{P}(X = j) \\
&\overset{(1)}{=} \sum_{j=0}^{k} \mathrm{P}(Y = k - x) \cdot \mathrm{P}(X = j) \\
&= \sum_{j=0}^{k} \frac{e^{-\lambda_2}\lambda_2^{k-j}}{(k-j)!} \frac{e^{-\lambda_1}\lambda_1^{j}}{(j)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{j=0}^{k} \binom{k}{j} \lambda_1^j \lambda_2^{k-j} \\
&\overset{(2)}{=} \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^k)}{k!} = \mathrm{Pois}(\lambda_1 + \lambda_2)
\end{aligned}$$

con (1) data l'indipendenza e in (2) si è utilizzato il teorema binomiale $(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}$ □

Osservazione 154. A intuito se vi sono due tipi di eventi che accadono ai rate λ_1 e λ_2 indipendentemente, allora il rate complessivo di eventi è $\lambda_1 + \lambda_2$.

5.7.6.1 Dalla Poisson alla binomiale

Proposizione 5.7.6. *Se* $X \sim \mathrm{Pois}(\lambda_1)$ *e* $Y \sim \mathrm{Pois}(\lambda_2)$ *sono indipendenti, allora la ditribuzione condizionata di* X *dato che* $XY = n$ *è* $\mathrm{Bin}(n, \lambda_1/(\lambda_1\lambda_2))$.

Dimostrazione. Utilizziamo la regola di Bayes per calcolare la PMF condizionata $\mathrm{P}(X = x|X + Y = n)$:

$$
\begin{aligned}
\mathrm{P}(X = x|X + Y = n) &= \frac{\mathrm{P}(X + Y = n|X = x) \cdot \mathrm{P}(X = x)}{\mathrm{P}(X + Y = n)} \\
&= \frac{\mathrm{P}(Y = n - x|X = x) \cdot \mathrm{P}(X = x)}{\mathrm{P}(X + Y = n)} \\
&\overset{(1)}{=} \frac{\mathrm{P}(Y = n - x) \cdot \mathrm{P}(X = x)}{\mathrm{P}(X + Y = n)}
\end{aligned}
$$

con (1) per indipendenza delle due. Ora sostituendo le PMF di X, Y e $X + Y$; questa al denominatore è distribuita come $\mathrm{Pois}(\lambda_1 + \lambda_2)$ per proposizione 5.7.5. Si ha:

$$
\begin{aligned}
\mathrm{P}(X = k|X + Y = n) &= \frac{\left(\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\right)\left(\frac{e^{\lambda_1}\lambda_2^{k}}{k!}\right)}{\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}} = \frac{\frac{e^{-(\lambda_1+\lambda_2)}\cdot\lambda_1^k\cdot\lambda_2^{n-k}}{k!(n-k)!}}{\frac{e^{-(\lambda_1+\lambda_2)}\cdot(\lambda_1+\lambda_2)^n}{n!}} \\
&= \frac{e^{-(\lambda_1+\lambda_2)} \cdot \lambda_1^k \cdot \lambda_2^{n-k}}{k!(n-k)!} \cdot \frac{n!}{e^{-(\lambda_1+\lambda_2)} \cdot (\lambda_1 + \lambda_2)^n} \\
&= \frac{n!}{k!(n-k)!} \cdot \frac{\lambda_1^k \cdot \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n} \\
&= \binom{n}{k} \left(\frac{\lambda_1^k}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2^k}{\lambda_1 + \lambda_2}\right)^{n-k} \\
&= \mathrm{Bin}\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)
\end{aligned}
$$

□

5.7.6.2 Dalla binomiale alla Poisson

Osservazione 155. Viceversa se prendiamo il limite della $\mathrm{Bin}(n, p)$ per $n \to \infty$ e $p \to 0$ con np fisso arriviamo alla Poisson.

Proposizione 5.7.7 (Approssimazione Poissoniana della binomiale)**.** *Se* $X \sim \mathrm{Bin}(n, p)$ *e facciamo tendere* $n \to \infty$, $p \to 0$ *ma* $\lambda = np$ *rimane fisso, allora la PMF di* X *converge a* $\mathrm{Pois}(\lambda)$.
La stessa conclusione si ha se $n \to \infty$, $p \to 0$ *ed* np converge *ad una costante* λ.

Osservazione 156. Questo è un *caso speciale* del paradigma di Poisson dove E_i sono indipendenti e hanno la stessa probabilità, quindi $\sum_{i=1}^{n} I_{E_i}$ ha distribuzione binomiale. In questo caso speciale possiamo dimostrare che l'approssimazione di Poisson ha senso limitandoci a prendere il limite della Binomiale.

Dimostrazione. Effettueremo la dimostrazione per $\lambda = np$ fisso (considerando $p = \lambda/n$), mostrando che la PMF $\mathrm{Bin}(n,p)$ converge alla $\mathrm{Pois}(\lambda)$. Per $0 \le x \le n$:

$$\begin{aligned}
\mathrm{P}(X = x) &= \binom{n}{x} p^x (1-p)^{n-x} \\
&= \frac{n(n-1)\cdot \ldots \cdot (n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\
&= \frac{\lambda^x}{x!} \frac{n(n-1)\cdot \ldots \cdot (n-x+1)}{n^x} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}
\end{aligned}$$

Per $n \to \infty$ con k fisso

$$\begin{aligned}
&\frac{\overbrace{n(n-1)\cdot \ldots \cdot (n-x+1)}^{x \text{ termini}}}{n^x} \overset{(1)}{=} \frac{n \cdot n\left(1 - \frac{1}{n}\right) \cdot \ldots \cdot n\left(1 - \frac{k-1}{n}\right)}{n^x} \to 1 \\
&\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda} \\
&\left(1 - \frac{\lambda}{n}\right)^{-k} = \left[\left(1 - \frac{\lambda}{n}\right)^n\right]^{-\frac{k}{n}} \to e^{-\frac{k}{n}} = 1
\end{aligned}$$

dove in (1) abbiamo raccolto un n a partire dal secondo fattore, lasciando fuori parentesi k n che si moltiplicano. Pertanto

$$\mathrm{P}(X = x) \to \frac{e^{-\lambda}\lambda^x}{x!} = \mathrm{Pois}(\lambda)$$

□

Osservazione 157. Il precedente risultato implica che se n è grande, p piccolo e np moderato, possiamo approssimare $\mathrm{Bin}(n,p)$ con $\mathrm{Pois}(np)$; come visto in precedenza l'errore nell'approssimare $\mathrm{P}(X \in I)$ con $\mathrm{P}(N \in I)$ per $X \sim \mathrm{Bin}(n,p)$ e $N \sim \mathrm{Pois}(np)$ è al massimo $\min(p, np^2)$.

Esempio 5.7.1. Il proprietario di un sito vuole studiare la distribuzione del numero di visitatori. Ogni giorno un millione di persone in maniera indipendente decide se visitare il sito o meno, con probabilità $p = 2 \times 10^{-1}$. Fornire una approssimazione della probabilità di avere almeno tre visitatori al giorno.
Se $X \sim \mathrm{Bin}(n,p)$ è il numero di visitatori con $n = 10^6$, fare i calcoli con la binomiale va incontro a difficoltà computazionali ed errori numerici del pc (dato che n è largo e p molto basso). Ma data la situazione con n largo p basso e $np = 2$ moderato, $\mathrm{Pois}(2)$ è una buona approssimazione. Questo porta a

$$\mathrm{P}(X \ge 3) = 1 - \mathrm{P}(X < 3) \approx 1 - e^{-2} - e^{-2} \cdot 2 - e^{-2} \cdot \frac{2^2}{2!} = 1 - 5e^{-2} \approx 0.3233$$

che è una approssimazione molto accurata.

5.7.7 Processo di Poisson

Definizione 5.7.2 (Processo di Poisson)**.** È una insieme di prove E_i che si possono verificare ciascuna in un dato arco temporale $[0, T]$. Le prove sono svolte nelle medesime condizioni e soddisfano di assiomi:

- il verificarsi di E nell'intervallo (t_1, t_2) è indipendente dal verificarsi di E nell'intervallo (t_3, t_4) (se gli intervalli non si sovrappongono);

- la probabilità del verificarsi di E in un intervallo infinitesimo $(t_0, t_0 + dt)$ è proporzionale ad un parametro $\lambda > 0$ che caratterizza la prova;

- la probabilità che due eventi si verifichino nello stesso intervallo di tempo è un infinitesimo di ordine superiore rispetto alla probabilità che se ne verifichi soltanto uno.

5.8 Uniforme discreta

5.8.1 Definizione

Osservazione 158. La prova che genera la vc Uniforme discreta si può assimilare all'estrazione di una pallina da un'urna che contiene n palline identiche numerate da 1 a n. Viene in genere utilizzata quanto tutti i risultati dell'esperimento sono equiprobabili

Definizione 5.8.1 (Uniforme discreta)**.** Il numero X della pallina estratta dall'urna contenente n palline numerate (da 1 a n) si distribuisce come Uniforme discreta $X \sim \mathrm{DUnif}\,(n)$.

5.8.2 Funzioni

Osservazione 159 (Supporto e spazio parametrico).

$$\begin{aligned} R_X &= \{1, \dots, n\} \\ \mathbf{\Theta} &= \{n \in \mathbb{N} \setminus \{0\}\} \end{aligned}$$

Proposizione 5.8.1 (Funzione di massa di probabilità)**.**

$$p_X(x) = \mathrm{P}\,(X = x) = \frac{1}{n} \cdot \mathbb{1}_{R_X}(x) \tag{5.43}$$

Definizione 5.8.2 (Funzione di ripartizione)**.**

$$F_X(x) = \mathrm{P}\,(X \leq x) = \begin{cases} 0 & \text{se } x < 1 \\ \frac{k}{n} & \text{se } k \leq x < k+1, (k = 1, 2, \dots, n-1) \\ 1 & \text{se } x \geq n \end{cases} \tag{5.44}$$

Osservazione 160. La funzione di ripartizione è nulla in $(\infty; 1)$ ed è una funzione a gradini di altezza costante pari a $1/n$, in corrispondenza di ogni valore intero $1 \leq x \leq n$ e vale 1 in $[n; +\infty)$.

5.8.3 Momenti

Proposizione 5.8.2 (Momenti caratteristici)**.**

$$\mathbb{E}(X) = \frac{n+1}{2} \tag{5.45}$$

$$\mathrm{Var}(X) = \frac{n^2-1}{12} \tag{5.46}$$

$$\mathrm{Asym}(X) = 0 \tag{5.47}$$

$$\mathrm{Kurt}(X) = 1.8 \tag{5.48}$$

Dimostrazione.

$$\mathbb{E}(X) = \sum_{x=1}^{n} x\frac{1}{n} = \frac{1}{n}(1+2+\ldots+n) = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}$$

□

Dimostrazione.

$$\begin{aligned}
\mathrm{Var}(X) = \mathbb{E}\left(X^2\right) - [\mathbb{E}(x)]^2 &= \left(\sum_{x=1}^{n} x^2\frac{1}{n}\right) - \left(\frac{n+1}{2}\right)^2 \\
&= \left(\frac{1}{n}(1^2+2^2+\ldots+n^2)\right) - \left(\frac{n+1}{2}\right)^2 \\
&= \left(\frac{1}{n}\cdot\frac{n(n+1)(2n+1)}{6}\right) - \left(\frac{n^2+1+2n}{4}\right) \\
&= \left(\frac{(n+1)(2n+1)}{6}\right) - \left(\frac{n^2+1+2n}{4}\right) \\
&= \frac{2(2n^2+2n+n+1) - 3(n^2+1+2n)}{12} \\
&= \frac{4n^2+4n+2n+2-3n^2-3-6n}{12} \\
&= \frac{n^2-1}{12}
\end{aligned}$$

□

5.8.4 Variabili discrete con supporto finito in R

Osservazione 161. Per quanto riguarda la simulazione di queste (tra le quali l'uniforme discreta) si fa utilizzo della funzione `sample` alla quale, oltre a specificare l'urna `x`, il numero `size` di estrazioni desiderate, l'estrazione con reinserimento (`replace`) o meno, si possono specificare le probabilità `prob` di ciascun elemento nell'urna.

```
## DUnif(100)
sample(x = 1:100, size = 10, replace = TRUE)

##  [1] 43 96  6 12 62 29 33 46 47  7
```

```r
## Urna discreta custom
sample(x = 1:3, prob = c(0.4, 0.4, 0.2), size = 10, replace = TRUE)

##  [1] 2 2 2 2 3 3 2 2 3 1
```

Capitolo 6

Variabili casuali continue

6.1 Logistica

6.1.1 Origine/definizione

Osservazione 162. Viene utilizzata per modelli di crescita di grandezze nel tempo, dove la crescita segue le fasi di crescita esponenziale, saturazione e arresto. Un buon modello per rappresentare fenomeni di questo tipo è rappresentato dalla funzione di ripartizione logistica.

Osservazione 163. Deriva il nome dall'avere la funzione di ripartizione che soddisfa l'equazione logistica: $F'(x) = \frac{1}{s}F(x)(1 - F(x))$.

Osservazione 164. E' matematicamente semplice e ci permette di focalizzarci su aspetti non numerici; è altresì importante nella regressione logistica.

6.1.2 Funzioni

Definizione 6.1.1 (Funzione di ripartizione)**.** Ha CDF

$$F_X(x) = \mathrm{P}\left(X \leq x\right) = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}}, \qquad x \in \mathbb{R} \tag{6.1}$$

Osservazione 165. Si trovano entrambe le definizioni e si passa dall'una all'altra moltiplicando a numeratore e denominatore per e^x

Definizione 6.1.2 (Funzione di densità)**.** Derivando entrambe le espressioni si hanno, equivalentemente:

$$f_x(x) = \frac{e^x}{(1 + e^x)^2} = \frac{e^{-x}}{(1 + e^{-x})^2} \tag{6.2}$$

6.1.3 Versione generale

Osservazione 166 (Supporto e spazio parametrico)*.*

$$\begin{aligned} R_X &= \mathbb{R} \\ \mathbf{\Theta} &= \{\mu \in \mathbb{R}, s \in \mathbb{R} : s > 0\} \end{aligned}$$

Definizione 6.1.3 (Funzione di ripartizione)**.** La funzione di densità di una vc $X \sim \text{Logis}\,(\mu, \sigma)$ è

$$F_X(x) = \frac{e^{\frac{x-\mu}{\sigma}}}{\left(1 + e^{\frac{x-\mu}{\sigma}}\right)} \cdot \mathbb{1}_{R_X}(x) \tag{6.3}$$

Definizione 6.1.4 (Funzione di densità)**.** La funzione di densità di una vc $X \sim \text{Logis}\,(\mu, \sigma)$ è

$$f_X(x) = \frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1 + e^{\frac{x-\mu}{\sigma}}\right)^2} \cdot \mathbb{1}_{R_X}(x) \tag{6.4}$$

Proposizione 6.1.1 (Momenti caratteristici)**.**

$$\mathbb{E}\,(X) = \mu$$
$$\text{Var}\,(X) = \frac{\pi^2}{3}\sigma^2$$

TODO: perché la varianza non è σ^2 applicando le regole su trasf lineari?

Mia dimostrazione, controllare. Sia $Z \sim \text{Logis}\,(0, 1)$ e sia $X = \sigma Z + \mu$, con σ parametro di scala e μ di posizione. Allora si ha che

$$Z = \frac{X - \mu}{\sigma} \sim \text{Logis}\,(0, 1)$$

Per cui possiamo scrivere che

$$F_X(x) = \frac{e^{\frac{x-\mu}{\sigma}}}{1 + e^{\frac{x-\mu}{\sigma}}}$$

Derivando per ottenere $f_X(x)$ si ha

$$f_X(x) = \frac{\left(e^{\frac{x-\mu}{\sigma}} \cdot \frac{1}{\sigma}\right)\left(1 + e^{\frac{x-\mu}{\sigma}}\right) - \left(e^{\frac{x-\mu}{\sigma}} \cdot \frac{1}{\sigma}\right)\left(e^{\frac{x-\mu}{\sigma}}\right)}{\left(1 + e^{\frac{x-\mu}{\sigma}}\right)^2} = \frac{\left(e^{\frac{x-\mu}{\sigma}} \cdot \frac{1}{\sigma}\right)\left(1 + e^{\frac{x-\mu}{\sigma}} - e^{\frac{x-\mu}{\sigma}}\right)}{\left(1 + e^{\frac{x-\mu}{\sigma}}\right)^2}$$
$$= \frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1 + e^{\frac{x-\mu}{\sigma}}\right)^2}$$

$\square$

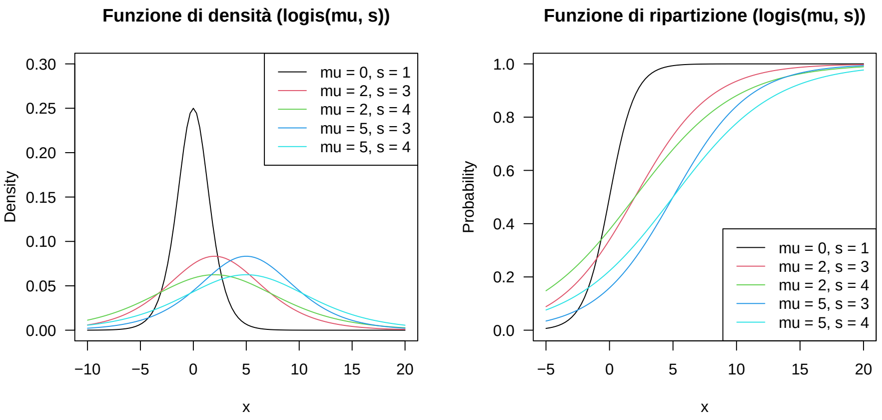


Figura 6.1: Distribuzione logistica

Capitolo 7

Old shit

7.1 Uniforme continua

Osservazione 167. È una vc continua X definita sul supporto (a, b), con $a < b$ ed ed esiti aventi la medesima densità, indicata con $X \sim \text{Unif}\,(a, b)$

Osservazione 168. Una formulazione usuale per tale modello probabilistico è la uniforma continua sull'intervallo con $a = 0, b = 1$.

Osservazione 169 (Supporto e spazio parametrico).

$$\begin{aligned} R_X &= [a, b] \\ \boldsymbol{\Theta} &= \{a, b \in \mathbb{R}, a < b\} \end{aligned}$$

Definizione 7.1.1 (Funzione di densità)**.** In figura 7.1

$$f_X(x) = \frac{1}{b-a} \cdot \mathbb{1}_{R_X}(x) \tag{7.1}$$

Proposizione 7.1.1. *L'area è 1.*

Dimostrazione.

$$(b-a) \cdot \frac{1}{(b-a)} = 1$$

□

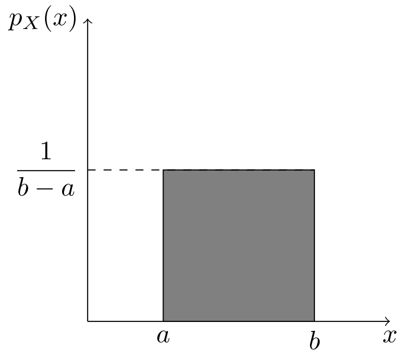


Figura 7.1: Uniforme continua

Definizione 7.1.2 (Funzione di ripartizione)**.**

$$F_X(x) = \begin{cases} 0 & \text{per } x \leq a \\ \dfrac{x-a}{b-a} & \text{se } a < x < b \\ 1 & \text{per } x \geq b \end{cases} \tag{7.2}$$

Proposizione 7.1.2 (Momenti caratteristici)**.**

$$\mathbb{E}\left(X\right) = \frac{a+b}{2} \tag{7.3}$$

$$\mathrm{Var}\left(X\right) = \frac{(b-a)^2}{12} \tag{7.4}$$

$$\mathrm{Asym}\left(X\right) = 0 \tag{7.5}$$

$$\mathrm{Kurt}\left(X\right) = 1.8 \tag{7.6}$$

Dimostrazione.

$$\begin{aligned} \mathbb{E}\left(X\right) &= \int_a^b x\frac{1}{b-a}\, dx = \left[\frac{x^2}{2(b-a)}\right]_a^b \\ &= \left(\frac{b^2}{2(b-a)} + c\right) - \left(\frac{a^2}{2(b-a)} + c\right) \\ &= \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2} \end{aligned}$$

□

Dimostrazione.

$$\begin{aligned} \mathrm{Var}\left(X\right) &= \left(\int_a^b x^2\frac{1}{b-a}\, dx\right) - \left(\frac{a+b}{2}\right)^2 \\ &= \left[\frac{x^3}{3(b-a)}\right]_a^b - \left(\frac{a+b}{2}\right)^2 \\ &= \left(\frac{b^3}{3(b-a)} + c\right) - \left(\frac{a^3}{3(b-a)} + c\right) - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{b^3-a^3}{3(b-a)} - \frac{(a+b)^2}{4} \\ &= \frac{(b-a)(a^2+b^2+ab)}{3(b-a)} - \frac{(a+b)^2}{4} \\ &= \frac{a^2+b^2+ab}{3} - \frac{a^2+b^2+2ab}{4} \\ &= \frac{4a^2+4b^2+4ab-3a^2-3b^2-6ab}{12} \\ &= \frac{a^2+b^2-2ab}{12} = \frac{(a-b)^2}{12} = \frac{(b-a)^2}{12} \end{aligned}$$

□

Osservazione 170. Si tratta di una variabile simmetrica e platicurtica (ovvero con una distribuzione molto piatta).

7.2 Esponenziale

Osservazione 171. L'esponenziale è generalmente usata per fenomeni di cui interessa un tempo/durata t (di vita, resistenza, funzionamento).
La derivazione può avvenire se si ipotizza una funzione di rischio/azzardo costante $H(t) = \lambda > 0$. con λ tasso di occorrenza dell'evento (reciproco del numero di eventi per unità di tempo).

Osservazione 172 (Supporto e spazio parametrico).

$$\begin{aligned} R_X &= \{x \in \mathbb{R} : x > 0\} \\ \boldsymbol{\Theta} &= \{\lambda \in \mathbb{R} : \lambda > 0\} \end{aligned}$$

Definizione 7.2.1 (Distribuzione esponenziale)**.** Se $H(t) = \lambda > 0$ la funzione di ripartizione si ricava dalla 4.19 come

$$\begin{aligned} F_X(t) &= 1 - \exp\left(-\int_0^t H(w)\, dw\right) = 1 - \exp\left(-\int_0^t \lambda\, dw\right) \\ &= 1 - \exp\left(-\lambda t\right) \end{aligned}$$

Definizione 7.2.2 (Funzione di ripartizione)**.**

$$F_X(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{per } x \geq 0 \\ 0 & \text{per } x < 0 \end{cases} \tag{7.7}$$

Osservazione 173. La funzione di densità si ottiene derivando dalla 7.7; pertanto una vc continua X si dice vc Esponenziale con parametro $\lambda > 0$, e si scrive $X \sim \text{Exp}\,(\lambda)$ se caratterizzata dalla seguente funzione di densità.

Definizione 7.2.3 (Funzione di densità)**.**

$$f_X(x) = \lambda \exp(-\lambda x) \cdot \mathbb{1}_{R_X}(x) \tag{7.8}$$

Proposizione 7.2.1 (Momenti caratteristici)**.**

$$\mathbb{E}\,(X) = \frac{1}{\lambda} \tag{7.9}$$

$$\text{Var}\,(X) = \frac{1}{\lambda^2} \tag{7.10}$$

$$\text{Asym}\,(X) = 2 \tag{7.11}$$

$$\text{Kurt}\,(X) = 9 \tag{7.12}$$

Osservazione 174 (Forma distribuzione). Tale funzione è decrescente a partire da $x = 0$, in corrispondenza del quale si registra la moda; è asimmetrica positiva e fortemente leptocurtica (a punta), con asimmetria e curtosi costanti al variare di λ. (figura 7.2)

Osservazione 175. La vc Esponenziale presenta una struttura molto semplice ma rigida, per cui non si adatta facilmente a tutte le situazioni reali; infatti, talvolta non è realistico assumere che la funzione di rischio si costante rispetto al tempo. Pertanto si hanno almeno due generalizzazioni: la Weibull e la Gamma.

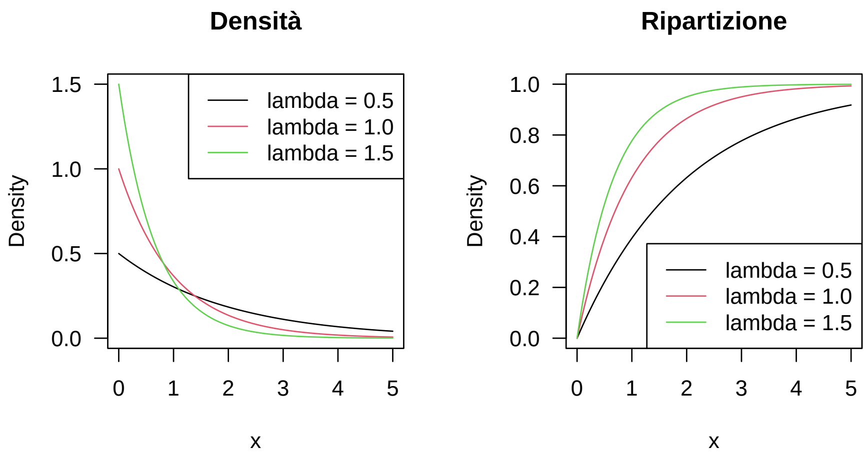


Figura 7.2: Distribuzione esponenziale

7.3 Normale/Gaussiana

Osservazione 176. Viene utilizzata come prima approssimazione per descrivere variabili casuali a valori reali che tendono a concentrarsi attorno a un singolo valor medio.

Osservazione 177. Una vc continua si dice vc Normale con parametri μ e σ^2, e la si indica con $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$ se è definita su tutto l'asse reale e presenta la seguente funzione di densità.

Osservazione 178 (Supporto e spazio parametrico).

$$\begin{aligned} R_X &= \{\mathbb{R}\} \\ \mathbf{\Theta} &= \left\{\mu \in \mathbb{R}; \sigma^2 \in \mathbb{R} : \sigma^2 > 0\right\} \end{aligned}$$

Definizione 7.3.1 (Funzione di densità)**.**

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \cdot \mathbb{1}_{R_X}(x) \tag{7.13}$$

Osservazione 179 (Forma della distribuzione). Ha una forma campanulare e simmetrica rispetto al punto di ascissa $x = \mu$, è crescente in $(-\infty, \mu)$ e decrescente in (μ, ∞). In corrispondenza di μ $f_X(x)$ ha il massimo (perché l'esponente negativo è minimo). Pertanto μ è il valore centrale la moda, mediana e valore medio della vc.
Si dimostra che $f_X(x)$ presenta due flessi in corrispondenza di $x = \mu \pm \sigma$. Ha come asintoto l'asse x
μ è un parametro di posizione mentre σ^2 misura la dispersione attorno a μ. La modifica di μ a parità di σ^2 implica una traslazione della funzione di densità lungo l'asse x; invece, al crescere di σ a parità di μ, i flessi si allontanano da μ e la funzione di den attribuisce maggiore probabilità ai valori lontani dal valore centrale (e viceversa al diminuire di σ^2). (figura 7.3)

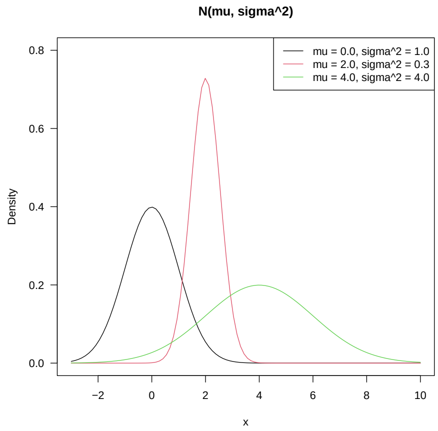


Figura 7.3: Distribuzione normale

Definizione 7.3.2 (Normale standardizzata)**.** Se $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$, la trasformazione lineare $Z = (X - \mu)/\sigma$ definisce la vc Normale standardizzata $Z \sim \mathrm{N}\left(0, 1\right)$

Definizione 7.3.3 (Funzione di densità (Normale standardizzata))**.**

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \cdot \mathbb{1}_{R_X}(x) \tag{7.14}$$

Definizione 7.3.4 (Funzione di ripartizione (Normale standardizzata))**.**

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right) dw \tag{7.15}$$

Osservazione 180. La funzione di ripartizione della vc Z non ammette una formulazione esplicita ed è necessario predisporre delle tavole che per opportuni valori di z forniscano l'integrale con sufficiente accuratezza.

Osservazione 181. Sfruttando la simmetria della funzione di densità, è sufficiente conoscere $\Phi(z)$ per i soli valori di $z > 0$. Infatti $\Phi(0) = 0.5$ ed inoltre:

$$\Phi(-z) = 1 - \Phi(z) \quad \forall z \geq 0 \tag{7.16}$$

Osservazione 182. La conoscenza della funzione di ripartizione della vc $Z \sim \mathrm{N}\left(0, 1\right)$ è sufficiente per calcolare la probabilità di qualsiasi vc $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$ mediante una semplice trasformazione:

$$\begin{aligned} \mathrm{P}\left(x_0 < X \leq x_1\right) &= \mathrm{P}\left(\frac{x_0 - \mu}{\sigma} < \underbrace{\frac{X - \mu}{\sigma}}_{Z} \leq \frac{x_1 - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{x_1 - \mu}{\sigma}\right) - \Phi\left(\frac{x_0 - \mu}{\sigma}\right) \end{aligned}$$

In pratica per calcolare la probabilità ce una vc normale assuma valori in un intervallo basta standardizzare gli estremi dell'intervallo ed utilizzare le tavole di $\Phi(z)$.

Proposizione 7.3.1 (Momenti caratteristici (Normale standardizzata))**.**

$$\mathbb{E}(Z) = 0 \tag{7.17}$$
$$\mathrm{Var}(Z) = 1 \tag{7.18}$$
$$\mathrm{Asym}(Z) = 0 \tag{7.19}$$
$$\mathrm{Kurt}(Z) = 3 \tag{7.20}$$

Proposizione 7.3.2 (Momenti caratteristici (Normale))**.** *Da $X = \mu + \sigma Z$ si ha*

$$\mathbb{E}(X) = \mu \tag{7.21}$$
$$\mathrm{Var}(X) = \sigma^2 \tag{7.22}$$
$$\mathrm{Asym}(X) = 0 \tag{7.23}$$
$$\mathrm{Kurt}(X) = 3 \tag{7.24}$$

Osservazione 183. Nel prosieguo tratteremo della vc Normale standardizzata, per semplicità.

Proposizione 7.3.3. *Se $X_i \sim \mathrm{N}\left(\mu_i, \sigma_i^2\right)$, allora:*

$$\sum_{i=1}^{n} a_i X_i \sim \mathrm{N}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

Osservazione 184. La famiglia delle vc normali è chiusa rispetto ad ogni combinazione lineare: in particolare la combinazione lineare di vc normali e indipendenti è ancora una vc normale che ha per valore medio la combinazione lineare dei valori medi e per varianza la combinazione lineare delle varianze con i quadrati dei coefficienti (proprietà riproduttiva della vc normale).

7.4 Gamma

Osservazione 185. Viene utilizzata quando si deve verificare la lunghezza dell'intervallo di tempo fino all'istante in cui si verifica la n-esima manifestazione di un evento aleatorio di interesse.
Similmente alla Beta è chiamata cosi perché coinvolge l'omonima funzione matematica.

Osservazione 186 (Supporto e spazio parametrico)*.*

$$R_X = \{x \in \mathbb{R} : x > 0\}$$
$$\boldsymbol{\Theta} = \{n, \lambda \in \mathbb{R} : n, \lambda > 0\}$$

Definizione 7.4.1 (Funzione di densità)**.** Una vc continua X si distribuisce come una Gamma con parametri $n > 0, \lambda > 0$, indicata con $X \sim \mathrm{Ga}(n, \lambda)$, se presenta una funzione di densità come la:

$$f_X(x) = \frac{\lambda^n}{\Gamma(n)} \cdot x^{n-1} \exp(-\lambda x) \cdot \mathbb{1}_{R_X}(x) \tag{7.25}$$

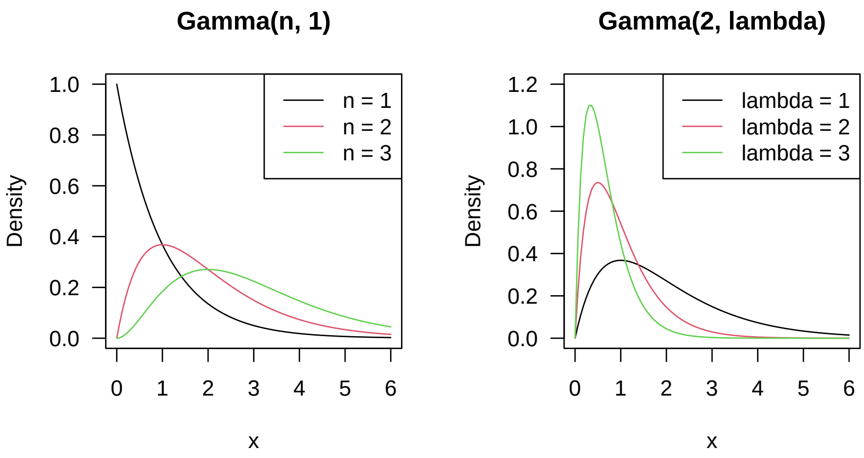


Figura 7.4: Distribuzione gamma

Definizione 7.4.2 (Funzione Gamma)**.** È definita come

$$\Gamma(n) = \int_0^{+\infty} x^{n-1} e^{-x}\, dx \tag{7.26}$$

e presenta le seguenti proprietà: se $n \in \mathbb{R}, n > 1$, $\Gamma(n) = (n-1)\Gamma(n-1)$ (ossia è ricorsiva); se $n \in \mathbb{N} \setminus \{0\}$, $\Gamma(n) = (n-1)!$; ha valore notevole $\Gamma(1/2) = \sqrt{\pi}$.

Osservazione 187 (Funzione di ripartizione). Non si può definire una funzione di ripartizione perché questa dipende dalla funzione Γ (a meno che n sia intero).

Proposizione 7.4.1 (Momenti caratteristici)**.**

$$\mathbb{E}(X) = \frac{n}{\lambda} \tag{7.27}$$

$$\mathrm{Var}(X) = \frac{n}{\lambda^2} \tag{7.28}$$

$$\mathrm{Asym}(X) = \frac{2}{\sqrt{n}} \tag{7.29}$$

$$\mathrm{Kurt}(X) = 3 + \frac{6}{n} \tag{7.30}$$

Osservazione 188 (Forma della distribuzione). λ è un parametro di scala mentre n determina la forma della distribuzione. All'aumentare del parametro λ la distribuzione si concentra sui valori più piccoli. Quando $n \to \infty$ la distribuzione diviene simmetrica e di forma campanulare (curtosi pari a 3). (figura 7.4)

Osservazione 189. Si nota che se $n = 1$, la distribuzione gamma diviene una esponenziale, ovvero $\mathrm{Ga}(1, \lambda) \sim \mathrm{Exp}(\lambda)$; pertanto la gamma è una generalizzazione della esponenziale.

Osservazione 190. Altro caso particolare, se $n = \frac{\nu}{2}$ (con $\nu \in \mathbb{N} \setminus \{0\}$, numero dei gradi di libertà) e $\lambda = \frac{1}{2}$ la distribuzione Gamma coincide con la Chi-quadrato.

Proposizione 7.4.2. *La gamma gode della proprietà riproduttiva nel senso che la somma di gamma indipendenti ancora una gamma:*

$$\sum \mathrm{Ga}(n_i, \lambda) \sim \mathrm{Ga}\left(\sum_i n_i, \lambda\right) \tag{7.31}$$

7.5 Chi-quadrato

Osservazione 191. La somma di ν vc normali standardizzate indipendenti ed elevate al quadrato è una vc continua sul supporto $(0, +\infty)$ che si distribuisce come una vc Chi-quadrato con ν gradi di libertà

$$\sum_{i=1}^{\nu} Z_i^2 \sim \chi_\nu^2 \tag{7.32}$$

Osservazione 192 (Supporto e spazio parametrico).

$$R_X = \{x \in \mathbb{R} : x > 0\}$$
$$\boldsymbol{\Theta} = \{\nu \in \mathbb{N} \setminus \{0\}\}$$

Definizione 7.5.1 (Funzione di densità)**.**

$$f_X(x) = \frac{1}{2^{\left(\frac{\nu}{2}\right)}\Gamma\left(\frac{\nu}{2}\right)} x^{\left(\frac{\nu}{2}-1\right)} e^{\left(-\frac{x}{2}\right)} \cdot \mathbb{1}_{R_X}(x) \tag{7.33}$$

con $x > 0$

Osservazione 193. Anche se ν può esser qualsiasi numero reale positivo, in pratica le applicazioni hanno tipicamente ν intero positivo.

Osservazione 194 (Forma della distribuzione). La vc Chi-quadrato è asimmetrica positiva e, al crescere di $\nu \to \infty$, tende ad assumere una forma sempre più vicina alla Normale. La forma della funzione di densità è monotona decrescente a zero se $\nu \leq 2$; se $\nu > 2$, presenta un picco intermedio in corrispondenza della moda (pari a $\nu - 2$). (figura 7.5)

Proposizione 7.5.1 (Momenti caratteristici)**.**

$$\mathbb{E}(X) = \nu \tag{7.34}$$
$$\mathrm{Var}(X) = 2\nu \tag{7.35}$$
$$\mathrm{Asym}(X) = \sqrt{\frac{8}{\nu}} \tag{7.36}$$
$$\mathrm{Kurt}(X) = 3 + \frac{12}{\nu} \tag{7.37}$$

Proposizione 7.5.2. *Anche la distribuzione Chi-quadrato gode della proprietà riproduttiva:*

$$\sum_{i=1}^{n} \chi_{\nu_i}^2 \sim \chi_{\sum_i \nu_i}^2$$

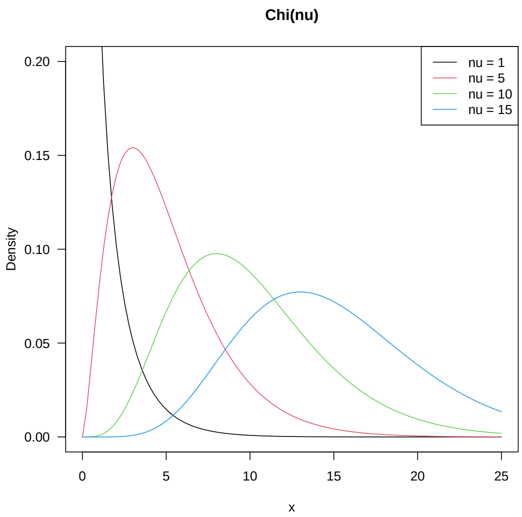


Figura 7.5: Distribuzione χ^2

7.6 Beta

Osservazione 195. Viene utilizzata quando si vogliono definire a priori i valori possibili delle probabilità di successo per variabili Bernoulliane.

Osservazione 196 (Supporto e spazio parametrico).

$$\begin{aligned} R_X &= [0,1] \\ \boldsymbol{\Theta} &= \{\alpha, \beta \in \mathbb{R} : \alpha, \beta > 0\} \end{aligned}$$

Definizione 7.6.1 (Funzione di densità)**.** Una vc continua X si definisce Beta con due parametri $\alpha > 0, \beta > 0$, e la indichiamo con $X \sim \text{Be}(\alpha, \beta)$ se la sua funzione di densità è:

$$f_X(x, \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \cdot \mathbb{1}_{R_X}(x) \tag{7.38}$$

Definizione 7.6.2 (Funzione Beta)**.** Definita come

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} \tag{7.39}$$

Presenta le seguenti proprietà

$$\begin{aligned} B(\alpha, \beta) &= B(\beta, \alpha) \\ B(\alpha, \beta) &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!} \end{aligned}$$

Osservazione 197. Una vc Beta è definita nell'intervallo $[0, 1]$, ma effettuando la trasformazione $Y = X(b-a) + a$, la si può ricondurre all'intervallo $[a, b]$.

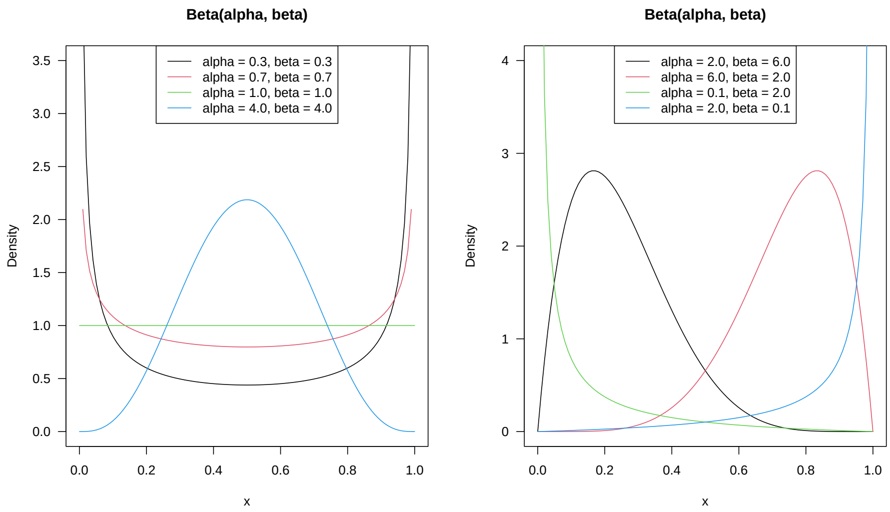


Figura 7.6: Distribuzione beta

Proposizione 7.6.1 (Momenti caratteristici)**.**

$$\mathbb{E}(X) = \frac{\alpha}{\alpha + \beta} \tag{7.40}$$

$$\mathrm{Var}(x) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \tag{7.41}$$

Osservazione 198 (Forma della distribuzione)*.* La forma (figura 7.6) dipende dai parametri α, β:

- se $\alpha = \beta$ la distribuzione è simmetrica rispetto al valore centrale $x = 1/2$; nel caso particolare $\alpha = \beta = 1$, la distribuzione coincide con l'uniforme: $\mathrm{Be}(1, 1) \sim \mathrm{Unif}(0, 1)$;
- altrimenti il segno di $\beta - \alpha$ denota l'asimmetria (es se negativo, perché $\alpha > \beta$, la coda è a sinistra, se positivo la coda a destra); scambiando α con β si inverte l'asse di simmetria.

7.7 T di Student

Osservazione 199*.* Il suo uso è prettamente teorico, in quanto è la risultante di una trasformazione su due variabili, una normale e una chi quadrato.

Osservazione 200 (Supporto e spazio parametrico)*.*

$$R_X = \mathbb{R}$$
$$\boldsymbol{\Theta} = \{g \in \mathbb{N} \setminus \{0\}\}$$

Definizione 7.7.1 (Distribuzione T)**.** Se $Z \sim \mathrm{N}(0, 1)$ ed C è una distribuzione indipendente tale che $C \sim \chi^2_g$ allora si definisce vc di Student la seguente X:

$$X = \frac{Z}{\sqrt{C/g}} \sim \mathrm{T}(g) \tag{7.42}$$

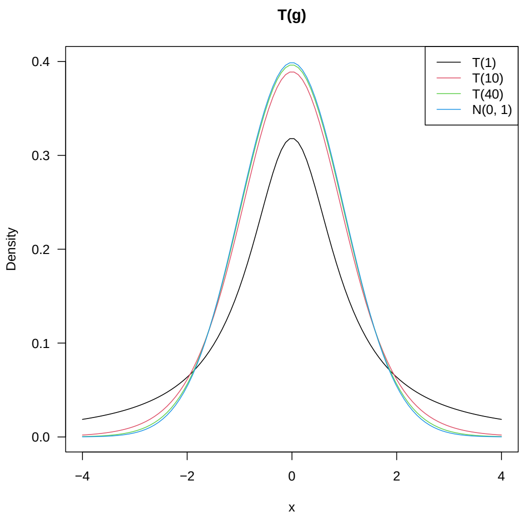


Figura 7.7: Distribuzione t

Definizione 7.7.2 (Funzione di densità)**.**

$$f_X(x) = \frac{\Gamma\left(\frac{g+1}{2}\right)}{\Gamma\left(\frac{g}{2}\right)\sqrt{\pi g}}\left(1+\frac{x^2}{g}\right)^{-\frac{g+1}{2}} \cdot \mathbb{1}_{R_X}(x) \tag{7.43}$$

Proposizione 7.7.1 (Momenti caratteristici)**.**

$$\mathbb{E}(X) = 0 \quad se\ g > 1$$
$$\operatorname{Var}(X) = \frac{g}{g-2} \quad seg > 2$$
$$\operatorname{Kurt}(X) = 3 + \frac{6}{g-4} \quad se\ g > 4$$

Osservazione 201 (Forma della distribuzione). Per $g \to \infty$ si nota la convergenza alla normale standardizzata. Verso $g = 30$, l'approssimazione è già buona; per g via via inferiore permane qualche differenza (code più alte rispetto alla normale, moda e media più basse). (figura 7.7)

7.8 F di Fisher

Osservazione 202. Il suo uso è prettamente teorico, in quanto è risultate di una trasformazione. È la distribuzione che deriva dal rapporto tra due vc Chi quadrato indipendenti tra loro e divise per i rispettivi gradi di libertà.

Osservazione 203. Se $X_1 \sim \chi^2_{g_1}$ e $X_2 \sim \chi^2_{g_2}$, allora

$$X = \frac{X_1/g_1}{X_2/g_2} \sim \mathrm{F}(g_1, g_2) \tag{7.44}$$

ovvero X si distribuisce come una F con g_1 e g_2 gradi di libertà.

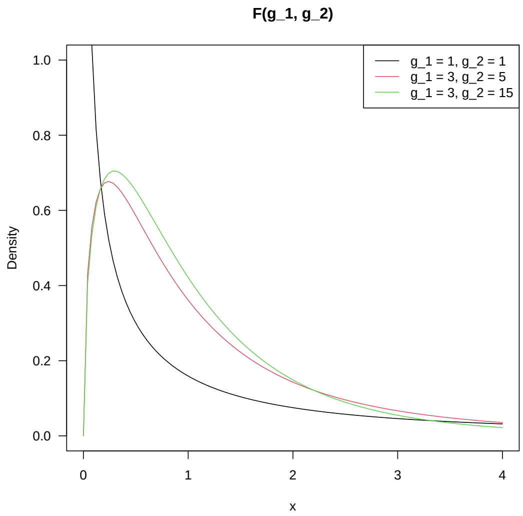


Figura 7.8: Distribuzione F

Osservazione 204 (Supporto e spazio parametrico).

$$R_X = \{x \in \mathbb{R} : x > 0\}$$
$$\boldsymbol{\Theta} = \{g_1, g_2 \in \mathbb{N} \setminus \{0\}\}$$

Definizione 7.8.1 (Funzione di densità)**.**

$$f_X(x) = \frac{\Gamma\left(\frac{g_1+g_2}{2}\right)}{\Gamma\left(\frac{g_1}{2}\right)\Gamma\left(\frac{g_2}{2}\right)} \cdot \left(\frac{g_1}{g_2}\right)^{\frac{g_1}{2}} \cdot \frac{x^{(g_1-2)/2}}{\left(1+\frac{g_1}{g_2}x\right)^{\frac{g_1+g_2}{2}}} \cdot \mathbb{1}_{R_X}(x) \tag{7.45}$$

Osservazione 205 (Funzione di ripartizione). Anche per la F non vi è una forma chiusa della ripartizione e ci si affida alle tavole.

Proposizione 7.8.1 (Momenti caratteristici)**.**

$$\mathbb{E}(X) = \frac{g_2}{g_2 - 2} \quad se\ g_2 > 2$$
$$\text{Var}(X) = \frac{2g_2^2(g_1 + g_2 - 2)}{g_1(g_2-2)^2(g_2-4)} \quad se\ g_2 > 4$$

Osservazione 206 (Forma della distribuzione). Si nota che se $g_1 = g_2 = 1$ la funzione è monotona decrescente, se $g_1, g_2 \neq 1$ la funzione è asimmetrica positiva. (figura 7.8)
La distribuzione converge a quella di una normale solo se contemporaneamente $g_1 \to \infty$ e $g_2 \to \infty$.

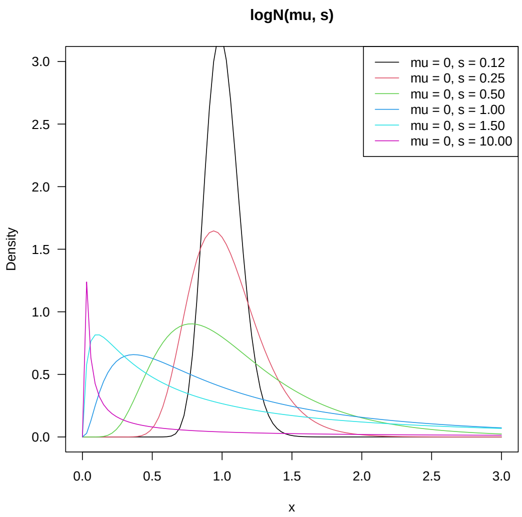


Figura 7.9: Distribuzione lognormale

7.9 Lognormale

Osservazione 207. Viene utilizzata quando la grandezza oggetto di studio è il risultato del prodotto di n fattori indipendenti

Osservazione 208 (Supporto e spazio parametrico).

$$R_X = \{x \in \mathbb{R} : x > 0\}$$
$$\boldsymbol{\Theta} = \left\{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R} : \sigma^2 > 0\right\}$$

Definizione 7.9.1 (Funzione di densità)**.**

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2} \cdot \mathbb{1}_{R_X}(x) \tag{7.46}$$

Proposizione 7.9.1 (Momenti caratteristici)**.**

$$\mathbb{E}\left(X\right) = e^{\mu + \frac{\sigma^2}{2}}$$
$$\mathrm{Var}\left(X\right) = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$$

Osservazione 209. Si ha che se $X \sim \mathrm{logN}\left(mu, \sigma\right)$ allora $\log X \sim \mathrm{N}\left(\mu, \sigma^2\right)$, mentre se $Y \sim \mathrm{N}\left(\mu, \sigma^2\right)$, $e^Y \sim \mathrm{logN}\left(\mu, \sigma^2\right)$

Osservazione 210 (Forma della distribuzione). Con μ fisso all'aumentare di σ l'asimmetria si incrementa (figura 7.9)

7.10 Weibull

Osservazione 211. Viene utilizzata per studiare l'affidabilità dei sistemi di produzione nei processi industriali, in particolare per valutare i tassi di rottura

Osservazione 212. La Weibull presenta la caratteristica di avere una funzione di rischio variabile in funzione di un ulteriore parametro a: se la vc $(X/b)^a \sim \operatorname{Exp}(1)$, allora diremo che la vc continua X, definita sula semiretta positiva è una vc di Weibull con parametri $a > 0, b > 0$.

Osservazione 213 (Supporto e spazio parametrico).

$$R_X = \{x \in \mathbb{R} : x > 0\}$$
$$\mathbf{\Theta} = \{a, b \in \mathbb{R} : a, b > 0\}$$

Osservazione 214 (Forma della funzione). Il parametro a determina la forma (figura 7.10):

- se $a < 1$ il tasso di rottura è decrescente nel tempo, ci sono componenti difettose che si rompono subito e, una volta sostituito, il tasso diminuisce
- se $a = 1$ il tasso di rottura è costante nel tempo: le cause dei difetti sono casuali (e la distribuzione coincide con una esponenziale di parametro $1/b$, ossia $\operatorname{Wei}(1, b) \sim \operatorname{Exp}\left(\frac{1}{b}\right)$)
- se $a > 1$ il tasso di rottura è crescente nel tempo, le cause della rottura dei componenti derivano dall'usura

Definizione 7.10.1 (Funzione di densità)**.**

$$f_X(x) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} e^{\left[-\left(\frac{x}{b}\right)^a\right]} \cdot \mathbb{1}_{R_X}(x) \tag{7.47}$$

Proposizione 7.10.1 (Momenti caratteristici)**.**

$$\mathbb{E}(X) = \frac{\Gamma\left(1 + \frac{1}{b}\right)}{a^{1/b}}$$
$$\operatorname{Var}(X) = \frac{\Gamma\left(1 + \frac{2}{b}\right) - \Gamma^2\left(1 + \frac{1}{b}\right)}{a^{2/b}}$$

7.11 Pareto

Osservazione 215. Viene utilizzata quando si studiano distribuzioni di variabili che hanno un minimo (ad esempio come, con x_m = reddito minimo)

Osservazione 216 (Supporto e spazio parametrico).

$$R_X = (x_m, +\infty)$$
$$\mathbf{\Theta} = \{x_m, k \in \mathbb{R} : x_m, k > 0\}$$

Definizione 7.11.1 (Funzione di densità)**.**

$$f_X(x) = k\frac{x_m^k}{x^{k+1}} \cdot \mathbb{1}_{R_X}(x) \tag{7.48}$$

Proposizione 7.11.1 (Momenti caratteristici)**.**

$$\mathbb{E}(X) = \frac{kx_m}{k-1} \qquad \textit{per } k > 1$$
$$\operatorname{Var}(X) = \left(\frac{x_m}{k-1}\right)^2 \frac{k}{k-2} \qquad \textit{per } k > 2$$

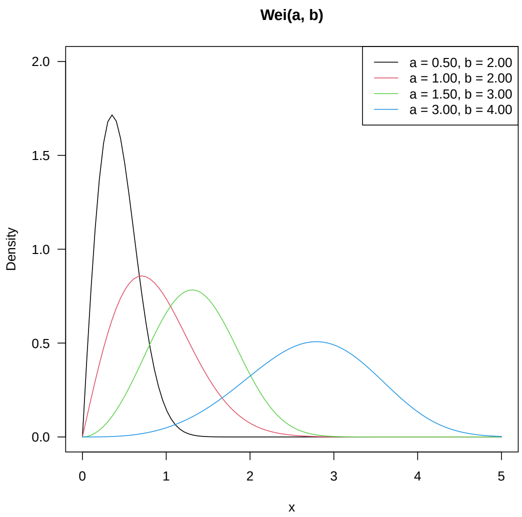


Figura 7.10: Distribuzione Weibull

Osservazione 217 (Forma della distribuzione). Al crescere di k la distribuzione è disuguale, ed è molto probabile trovare valori vicini al limite inferiore x_m, poco proabile trovare valori molto grandi.

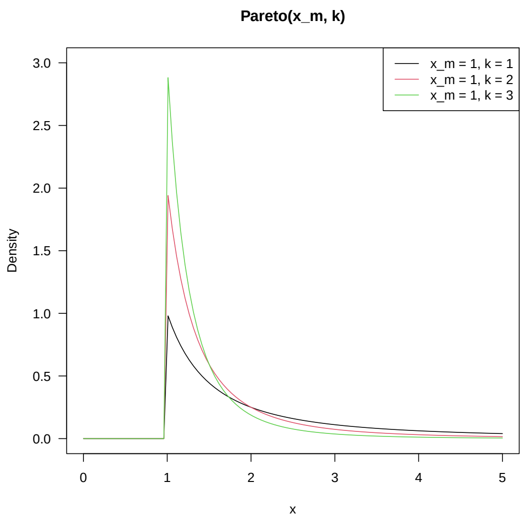


Figura 7.11: Distribuzione di Pareto