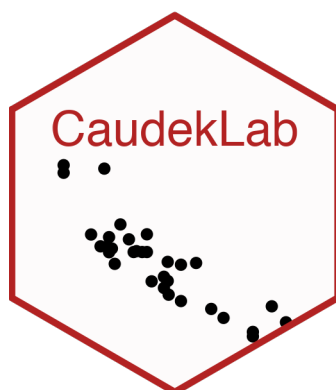


# Psicometria

Corrado Caudek

Questo documento è stato realizzato con:

- $\text{\LaTeX}$  e la classe memoir (<http://www.ctan.org/pkg/memoir>);
- R (<http://www.r-project.org/>) e RStudio (<http://www.rstudio.com/>);
- bookdown (<http://bookdown.org/>) e memoir (<https://ericmarcon.github.io/memoir/>).



Nel blog della mia pagina personale sono forniti alcuni approfondimenti degli argomenti qui trattati. <https://ccaudek.github.io/caudeklab/>

# Indice

<b>Indice</b>	<b>iii</b>
<b>Prefazione</b>	<b>vii</b>
La psicologia e la Data science . . . . .	vii
Come studiare . . . . .	viii
Sviluppare un metodo di studio efficace . . . . .	viii
<b>1 Distribuzioni di v.c. discrete</b>	<b>1</b>
1.1 Una prova Bernoulliana . . . . .	1
1.2 Una sequenza di prove Bernoulliane . . . . .	1
Valore atteso e deviazione standard . . . . .	4
1.3 Distribuzione di Poisson . . . . .	4
1.4 Alcune proprietà della variabile di Poisson . . . . .	6
Considerazioni conclusive . . . . .	6
<b>Bibliografia</b>	<b>7</b>
<b>Elenco delle figure</b>	<b>9</b>



Copyright © 2022.

Data della versione presente: Dicembre 25, 2021.



# Prefazione

*Data Science per psicologi* contiene il materiale delle lezioni dell'insegnamento di *Psicometria B000286* (A.A. 2021/2022) rivolto agli studenti del primo anno del Corso di Laurea in Scienze e Tecniche Psicologiche dell'Università degli Studi di Firenze. *Psicometria* si propone di fornire agli studenti un'introduzione all'analisi dei dati in psicologia. Le conoscenze/competenze che verranno sviluppate in questo insegnamento sono quelle della Data science, ovvero un insieme di conoscenze/competenze che si pongono all'intersezione tra statistica (ovvero, richiedono la capacità di comprendere teoremi statistici) e informatica (ovvero, richiedono la capacità di sapere utilizzare un software).

## La psicologia e la Data science

*It's worth noting, before getting started, that this material is hard. If you find yourself confused at any point, you are normal. Any sense of confusion you feel is just your brain correctly calibrating to the subject matter. Over time, confusion is replaced by comprehension [...]*

— Richard McElreath

Sembra sensato spendere due parole su un tema che è importante per gli studenti: quello indicato dal titolo di questo Capitolo. È ovvio che agli studenti di psicologia la statistica non piace. Se piacesse, forse studierebbero Data science e non psicologia; ma non lo fanno. Di conseguenza, gli studenti di psicologia si chiedono: “perché dobbiamo perdere tanto tempo a studiare queste cose quando in realtà quello che ci interessa è tutt'altro?” Questa è una bella domanda.

C'è una ragione molto semplice che dovrebbe farci capire perché la Data science è così importante per la psicologia. Infatti, a ben pensarci, la psicologia è una disciplina intrinsecamente statistica, se per statistica intendiamo quella disciplina che studia la variazione delle caratteristiche degli individui nella popolazione. La psicologia studia *gli individui* ed è proprio la variabilità inter- e intra-individuale ciò che vogliamo descrivere e, in certi casi, predire. In questo senso, la psicologia è molto diversa dall'ingegneria, per esempio. Le proprietà di un determinato ponte sotto certe condizioni, ad esempio, sono molto simili a quelle di un altro ponte, sotto le medesime condizioni. Quindi, per un ingegnere la statistica è poco importante: le proprietà dei materiali sono unicamente dipendenti dalla loro composizione e restano costanti. Ma lo stesso non può dirsi degli individui: ogni individuo è unico e cambia nel tempo. E le variazioni tra gli individui, e di un individuo nel tempo, sono l'oggetto di studio proprio della psicologia: è dunque chiaro che i problemi che la psicologia si pone sono molto diversi da quelli affrontati, per esempio, dagli ingegneri. Questa è la ragione per cui abbiamo tanto bisogno della Data science in psicologia: perché la Data science ci consente di descrivere la variazione e il cambiamento. E queste sono appunto le caratteristiche di base dei fenomeni psicologici.

Sono sicuro che, leggendo queste righe, a molti studenti sarà venuta in mente la seguente domanda: perché non chiediamo a qualche esperto di fare il “lavoro sporco” (ovvero le analisi statistiche) per noi, mentre noi (gli psicologi) ci occupiamo solo di ciò che ci interessa, ovvero dei problemi psicologici slegati dai dettagli “tecnici” della Data

science? La risposta a questa domanda è che non è possibile progettare uno studio psicologico sensato senza avere almeno una comprensione rudimentale della Data science. Le tematiche della Data science non possono essere ignorate né dai ricercatori in psicologia né da coloro che svolgono la professione di psicologo al di fuori dell'Università. Infatti, anche i professionisti al di fuori dall'università non possono fare a meno di leggere la letteratura psicologica più recente: il continuo aggiornamento delle conoscenze è infatti richiesto dalla deontologia della professione. Ma per potere fare questo è necessario conoscere un bel po' di Data science! Basta aprire a caso una rivista specialistica di psicologia per rendersi conto di quanto ciò sia vero: gli articoli che riportano i risultati delle ricerche psicologiche sono zeppi di analisi statistiche e di modelli formali. E la comprensione della letteratura psicologica rappresenta un requisito minimo nel bagaglio professionale dello psicologo.

Le considerazioni precedenti cercano di chiarire il seguente punto: la Data science non è qualcosa da studiare a malincuore, in un singolo insegnamento universitario, per poi poterla tranquillamente dimenticare. Nel bene e nel male, gli psicologi usano gli strumenti della Data science in tantissimi ambiti della loro attività professionale: in particolare quando costruiscono, somministrano e interpretano i test psicometrici. È dunque chiaro che possedere delle solide basi di Data science è un tassello imprescindibile del bagaglio professionale dello psicologo. In questo insegnamento verranno trattati i temi base della Data science e verrà adottato un punto di vista bayesiano, che corrisponde all'approccio più recente e sempre più diffuso in psicologia.

## Come studiare

*I know quite certainly that I myself have no special talent. Curiosity, obsession and dogged endurance, combined with self-criticism, have brought me to my ideas.*

— Albert Einstein

Il giusto metodo di studio per prepararsi all'esame di Psicometria è quello di seguire attivamente le lezioni, assimilare i concetti via via che essi vengono presentati e verificare in autonomia le procedure presentate a lezione. Incoraggio gli studenti a farmi domande per chiarire ciò che non è stato capito appieno. Incoraggio gli studenti a utilizzare i forum attivi su Moodle e, soprattutto, a svolgere gli esercizi proposti su Moodle. I problemi forniti su Moodle rappresentano il livello di difficoltà richiesto per superare l'esame e consentono allo studente di comprendere se le competenze sviluppate fino a quel punto sono sufficienti rispetto alle richieste dell'esame.

La prima fase dello studio, che è sicuramente individuale, è quella in cui è necessario acquisire le conoscenze teoriche relative ai problemi che saranno presentati all'esame. La seconda fase di studio, che può essere facilitata da scambi con altri e da incontri di gruppo, porta ad acquisire la capacità di applicare le conoscenze: è necessario capire come usare un software (R) per applicare i concetti statistici alla specifica situazione del problema che si vuole risolvere. Le due fasi non sono però separate: il saper fare molto spesso ci aiuta a capire meglio.

## Sviluppare un metodo di studio efficace

*Memorization is not learning.*

— Richard Phillips Feynman

Avendo insegnato molte volte in passato un corso introduttivo di analisi dei dati ho notato nel corso degli anni che gli studenti con l'atteggiamento mentale che descriverò qui sotto generalmente ottengono ottimi risultati. Alcuni studenti sviluppano naturalmente questo approccio allo studio, ma altri hanno bisogno di fare uno sforzo per maturarlo.



Fornisco qui sotto una breve descrizione del “metodo di studio” che, nella mia esperienza, è il più efficace per affrontare le richieste di questo insegnamento (Burger & Starbird, 2012).

- Dedicate un tempo sufficiente al materiale di base, apparentemente facile; assicuratevi di averlo capito bene. Cercate le lacune nella vostra comprensione. Leggere presentazioni diverse dello stesso materiale (in libri o articoli diversi) può fornire nuove intuizioni.
- Gli errori che facciamo sono i nostri migliori maestri. Istitivamente cerchiamo di dimenticare subito i nostri errori. Ma il miglior modo di imparare è apprendere dagli errori che commettiamo. In questo senso, una soluzione corretta è meno utile di una soluzione sbagliata. Quando commettiamo un errore questo ci fornisce un’informazione importante: ci fa capire qual è il materiale di studio sul quale dobbiamo ritornare e che dobbiamo capire meglio.
- C’è ovviamente un aspetto “psicologico” nello studio. Quando un esercizio o problema ci sembra incomprensibile, la cosa migliore da fare è dire: “mi arrendo”, “non ho idea di cosa fare!”. Questo ci rilassa: ci siamo già arresi, quindi non abbiamo niente da perdere, non dobbiamo più preoccuparci. Ma non dobbiamo fermarci qui. Le cose “migliori” che faccio (se ci sono) le faccio quando non ho voglia di lavorare. Alle volte, quando c’è qualcosa che non so fare e non ho idea di come affrontare, mi dico: “oggi non ho proprio voglia di fare fatica”, non ho voglia di mettermi nello stato mentale per cui “in 10 minuti devo risolvere il problema perché dopo devo fare altre cose”. Però ho voglia di *divertirmi* con quel problema e allora mi dedico a qualche aspetto “marginale” del problema, che so come affrontare, oppure considero l’aspetto più difficile del problema, quello che non so come risolvere, ma invece di cercare di risolverlo, guardo come altre persone hanno affrontato problemi simili, oppure lo stesso problema in un altro contesto. Non mi pongo l’obiettivo “risolvi il problema in 10 minuti”, ma invece quello di farmi un’idea “generale” del problema, o quello di capire un caso più specifico e più semplice del problema. Senza nessuna pressione. Infatti, in quel momento ho deciso di non lavorare (ovvero, di non fare fatica). Va benissimo se “parto per la tangente”, ovvero se mi metto a leggere del materiale che sembra avere poco a che fare con il problema centrale (le nostre intuizioni e la nostra curiosità solitamente ci indirizzano sulla strada giusta). Quando faccio così, molto spesso trovo la soluzione del problema che mi ero posto e, paradossalmente, la trovo in un tempo minore di quello che, in precedenza, avevo dedicato a “lavorare” al problema. Allora perché non faccio sempre così? C’è ovviamente l’aspetto dei “10 minuti” che non è sempre facile da dimenticare. Sotto pressione, possiamo solo agire in maniera automatica, ovvero possiamo solo applicare qualcosa che già sappiamo fare. Ma se dobbiamo imparare qualcosa di nuovo, la pressione è un impedimento.
- È utile farsi da soli delle domande sugli argomenti trattati, senza limitarsi a cercare di risolvere gli esercizi che vengono assegnati. Quando studio qualcosa mi viene in mente: “se questo è vero, allora deve succedere quest’altra cosa”. Allora verifico se questo è vero, di solito con una simulazione. Se i risultati della simulazione sono quelli che mi aspetto, allora vuol dire che ho capito. Se i risultati sono diversi da quelli che mi aspettavo, allora mi rendo conto di non avere capito e ritorno indietro a studiare con più attenzione la teoria che pensavo di avere capito – e ovviamente mi rendo conto che c’era un aspetto che avevo frainteso. Questo tipo di verifica è qualcosa che dobbiamo fare da soli, in prima persona: nessun altro può fare questo al posto nostro.

- Non aspettatevi di capire tutto la prima volta che incontrate un argomento nuovo.<sup>1</sup> È utile farsi una nota mentalmente delle lacune nella vostra comprensione e tornare su di esse in seguito per cercare di colmarle. L'atteggiamento naturale, quando non capiamo i dettagli di qualcosa, è quello di pensare: "non importa, ho capito in maniera approssimativa questo punto, non devo preoccuparmi del resto". Ma in realtà non è vero: se la nostra comprensione è superficiale, quando il problema verrà presentato in una nuova forma, non riusciremo a risolverlo. Per cui i dubbi che ci vengono quando studiamo qualcosa sono il nostro alleato più prezioso: ci dicono esattamente quali sono gli aspetti che dobbiamo approfondire per potere migliorare la nostra preparazione.
- È utile sviluppare una visione d'insieme degli argomenti trattati, capire l'obiettivo generale che si vuole raggiungere e avere chiaro il contributo che i vari pezzi di informazione forniscono al raggiungimento di tale obiettivo. Questa organizzazione mentale del materiale di studio facilita la comprensione. È estremamente utile creare degli schemi di ciò che si sta studiando. Non aspettate che sia io a fornirvi un riepilogo di ciò che dovete imparare: sviluppate da soli tali schemi e tali riassunti.
- Tutti noi dobbiamo imparare l'arte di trovare le informazioni, non solo nel caso di questo insegnamento. Quando vi trovate di fronte a qualcosa che non capite, o ottenete un oscuro messaggio di errore da un software, ricordatevi: "Google is your friend".

Corrado Caudek

---

<sup>1</sup>Ricordatevi inoltre che gli individui tendono a sottostimare la propria capacità di apprendere (Horn & Loewenstein, 2021).

# Capitolo 1

## Distribuzioni di v.c. discrete

In questo Capitolo verranno esaminate le principali distribuzioni di probabilità delle variabili casuali discrete. Un esperimento casuale che può dare luogo a solo due possibili esiti (successo, insuccesso) è modellabile con una variabile casuale di Bernoulli. Una sequenza di prove di Bernoulli costituisce un processo Bernoulliano. Il numero di successi dopo  $n$  prove di Bernoulli corrisponde ad una variabile casuale che segue la legge binomiale. La distribuzione binomiale risulta da un insieme di prove di Bernoulli solo se il numero totale  $n$  è fisso per disegno. Se il numero di prove è esso stesso una variabile casuale, allora il numero di successi nella corrispondente sequenza di prove bernoulliane segue al distribuzione di Poisson.

### 1.1 Una prova Bernoulliana

Se un esperimento casuale ha solo due esiti possibili, allora le repliche indipendenti di questo esperimento sono chiamate “prove Bernoulliane” (il lancio di una moneta è il tipico esempio).

**Definizione 1.1.** Viene detta variabile di Bernoulli una variabile casuale discreta  $Y = \{0, 1\}$  con la seguente distribuzione di probabilità:

$$P(Y | \theta) = \begin{cases} \theta & \text{se } Y = 1, \\ 1 - \theta & \text{se } Y = 0, \end{cases}$$

con  $0 \leq \theta \leq 1$ . Convenzionalmente l'evento  $\{Y = 1\}$  con probabilità  $\theta$  viene chiamato “successo” mentre l'evento  $\{Y = 0\}$  con probabilità  $1 - \theta$  viene chiamato “insuccesso”.

Applicando l'operatore di valore atteso e di varianza, otteniamo

$$\mathbb{E}(Y) = 0 \cdot Pr(Y = 0) + 1 \cdot Pr(Y = 1) = \theta, \quad (1.1)$$

$$\text{Var}(Y) = (0 - \theta)^2 \cdot Pr(Y = 0) + (1 - \theta)^2 \cdot Pr(Y = 1) = \theta(1 - \theta). \quad (1.2)$$

Scriviamo  $Y \sim \mathcal{B}(\theta)$  per indicare che la variabile casuale  $Y$  ha una distribuzione Bernoulliana di parametro  $\theta$ .

**Esercizio 1.1.** Nel caso del lancio di una moneta equilibrata la variabile casuale di Bernoulli assume i valori 0 e 1. La distribuzione di massa di probabilità è pari a  $\frac{1}{2}$  in corrispondenza di entrambi i valori. La funzione di distribuzione vale  $\frac{1}{2}$  per  $Y = 0$  e 1 per  $Y = 1$ .

### 1.2 Una sequenza di prove Bernoulliane

La distribuzione binomiale è rappresentata dall'elenco di tutti i possibili numeri di successi  $Y = \{0, 1, 2, \dots, n\}$  che possono essere osservati in  $n$  prove Bernoulliane indipendenti di probabilità  $\theta$ , a ciascuno dei quali è associata la relativa probabilità. Esempi di

una distribuzione binomiale sono i risultati di una serie di lanci di una stessa moneta o di una serie di estrazioni da un'urna (con reintroduzione). La distribuzione binomiale di parametri  $n$  e  $\theta$  è in realtà una famiglia di distribuzioni: al variare dei parametri  $\theta$  e  $n$  variano le probabilità.

**Definizione 1.2.** La probabilità di ottenere  $y$  successi e  $n - y$  insuccessi in  $n$  prove Bernoulliane è data dalla distribuzione binomiale:

$$\begin{aligned} P(Y = y) &= \binom{n}{y} \theta^y (1 - \theta)^{n-y} \\ &= \frac{n!}{y!(n-y)!} \theta^y (1 - \theta)^{n-y}, \end{aligned} \quad (1.3)$$

dove  $n$  = numero di prove Bernoulliane,  $\theta$  = probabilità di successo in ciascuna prova e  $y$  = numero di successi.

*Dimostrazione.* La (1.3) può essere derivata nel modo seguente. Indichiamo con  $S$  il successo e con  $I$  l'insuccesso di ciascuna prova. Una sequenza di  $n$  prove Bernoulliane darà come esito una sequenza di  $n$  elementi  $S$  e  $I$ . Ad esempio, una sequenza che contiene  $y$  successi è la seguente:

$$\overbrace{SS \dots S}^{y \text{ volte}} \overbrace{II \dots I}^{n-y \text{ volte}}$$

Essendo  $\theta$  la probabilità di  $S$  e  $1 - \theta$  la probabilità di  $I$ , la probabilità di ottenere la specifica sequenza riportata sopra è

$$\overbrace{\theta \dots \theta}^{y \text{ volte}} \overbrace{(1 - \theta)(1 - \theta) \dots (1 - \theta)}^{n-y \text{ volte}} = \theta^y \cdot (1 - \theta)^{n-y}. \quad (1.4)$$

Non siamo però interessati alla probabilità di una *specifica* sequenza di  $S$  e  $I$  ma, bensì, alla probabilità di osservare una *qualsiasi* sequenza di  $y$  successi in  $n$  prove. In altre parole, vogliamo la probabilità dell'unione di tutti gli eventi corrispondenti a  $y$  successi in  $n$  prove.

È immediato notare che una qualsiasi altra sequenza contenente esattamente  $y$  successi avrà sempre come probabilità  $\theta^y \cdot (1 - \theta)^{n-y}$ : il prodotto infatti resta costante anche se cambia l'ordine dei fattori.<sup>1</sup> Per trovare il risultato cercato dobbiamo moltiplicare la (1.4) per il numero di sequenze possibili di  $y$  successi in  $n$  prove.

Il numero di sequenze che contengono esattamente  $y$  successi in  $n$  prove. La risposta è fornita dal coefficiente binomiale<sup>2</sup>:

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}, \quad (1.5)$$

dove il simbolo  $n!$  si legge  $n$  fattoriale ed è uguale al prodotto di  $n$  numeri interi decrescenti a partire da  $n$ . Per definizione  $0! = 1$ .

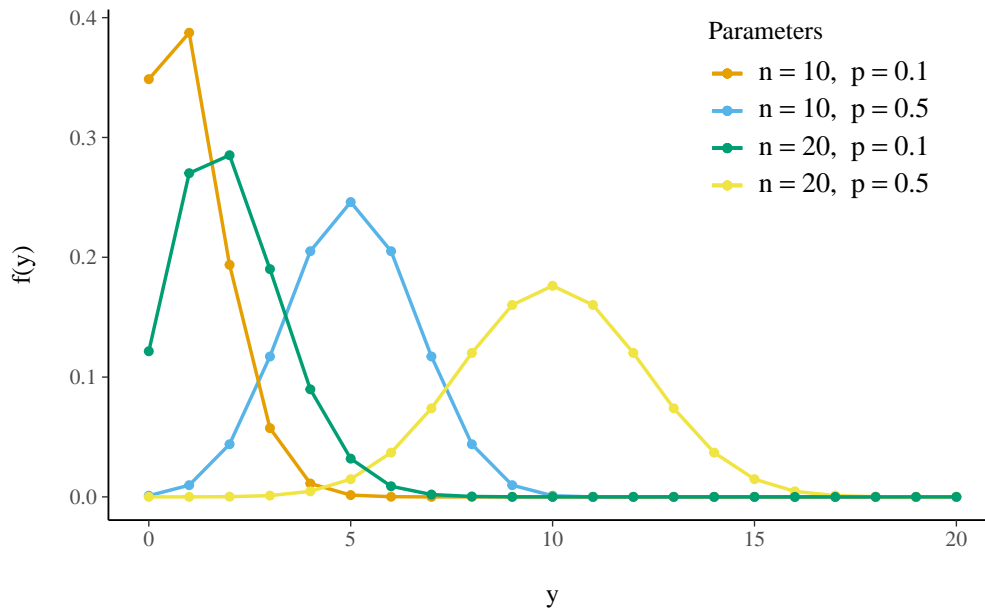
Essendo la probabilità dell'unione di  $K$  elementi incompatibili uguale alla somma delle loro rispettive probabilità, e dato che le sequenze di  $y$  successi in  $n$  prove hanno tutte la stessa probabilità, per trovare la formula della distribuzione binomiale (1.3) è sufficiente moltiplicare la (1.4) per la (1.5).  $\square$

La distribuzione di probabilità di alcune distribuzioni binomiali, per due valori di  $n$  e  $\theta$ , è fornita nella figura 1.1.

---

<sup>1</sup>Viene detta *scambiabilità* la proprietà per cui l'ordine con cui compiamo le osservazioni è irrilevante per l'assegnazione delle probabilità.

<sup>2</sup>La derivazione della formula del coefficiente binomiale è fornita nell'Appendice ??.



**Figura 1.1:** Alcune distribuzioni binomiali. Nella figura, il parametro  $\theta$  è indicato con  $p$ .

**Esercizio 1.2.** Usando la (1.3), si trovi la probabilità di  $y = 2$  successi in  $n = 4$  prove Bernoulliane indipendenti con  $\theta = 0.2$

$$\begin{aligned} P(Y = 2) &= \frac{4!}{2!(4-2)!} 0.2^2 (1-0.2)^{4-2} \\ &= \frac{4 \cdot 3 \cdot 2 \cdot 1}{(2 \cdot 1)(2 \cdot 1)} 0.2^2 0.8^2 = 0.1536. \end{aligned}$$

Ripetendo i calcoli per i valori  $y = 0, \dots, 4$  troviamo la distribuzione binomiale di parametri  $n = 4$  e  $\theta = 0.2$ :

$y$	$P(Y = y)$
0	0.4096
1	0.4096
2	0.1536
3	0.0256
4	0.0016
sum	1.0

Lo stesso risultato si ottiene usando la seguente istruzione R:

```
dbinom(0:4, 4, 0.2)
#> [1] 0.4096 0.4096 0.1536 0.0256 0.0016
```

**Esercizio 1.3.** Lanciando 5 volte una moneta onesta, qual è la probabilità che esca testa almeno tre volte?

In R, la soluzione si trova con

```
dbinom(3, 5, 0.5) + dbinom(4, 5, 0.5) + dbinom(5, 5, 0.5)
#> [1] 0.5
```

Alternativamente, possiamo trovare la probabilità dell'evento complementare a quello definito dalla funzione di ripartizione calcolata mediante `pbinom()`, ovvero

```
1 - pbinom(2, 5, 0.5)
#> [1] 0.5
```

### Valore atteso e deviazione standard

La media (numero atteso di successi in  $n$  prove) e la deviazione standard di una distribuzione binomiale sono molto semplici:

$$\begin{aligned}\mu &= n\theta, \\ \sigma &= \sqrt{n\theta(1-\theta)}.\end{aligned}\tag{1.6}$$

*Dimostrazione.* Essendo  $Y$  la somma di  $n$  prove Bernoulliane indipendenti  $Y_i$ , è facile vedere che

$$\mathbb{E}(Y) = \mathbb{E}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathbb{E}(Y_i) = n\theta,\tag{1.7}$$

$$\text{Var}(Y) = \text{Var}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \text{Var}(Y_i) = n\theta(1-\theta).\tag{1.8}$$

□

**Esercizio 1.4.** Si trovino il valore atteso e la varianza del lancio di quattro monete con probabilità di successo pari a  $\theta = 0.2$ .

Il valore atteso è  $\mu = n\theta = 4 \cdot 0.2 = 0.8$ . Ciò significa che, se l'esperimento casuale venisse ripetuto infinite volte, l'esito testa verrebbe osservato un numero medio di volte pari a 0.8. La varianza è  $n\theta(1-\theta) = 4 \cdot (1-0.2) = 0.8$ .<sup>3</sup>

### 1.3 Distribuzione di Poisson

La distribuzione di Poisson è una distribuzione di probabilità discreta che esprime le probabilità per il numero di eventi che si verificano successivamente ed indipendentemente in un dato intervallo di tempo, sapendo che mediamente se ne verifica un numero  $\lambda$ . La distribuzione di Poisson serve dunque per contare il numero di volte in cui un evento ha luogo in un determinato intervallo di tempo. La stessa distribuzione può essere estesa anche per contare gli eventi che hanno luogo in una determinata porzione di spazio.

**Definizione 1.3.** La distribuzione di Poisson può essere intesa come limite della distribuzione binomiale, dove la probabilità di successo  $\theta$  è pari a  $\frac{\lambda}{n}$  con  $n$  che tende a  $\infty$ :

$$\lim_{y \rightarrow \infty} \binom{n}{y} \theta^y (1-\theta)^{n-y} = \frac{\lambda^y}{y!} e^{-\lambda}.\tag{1.9}$$

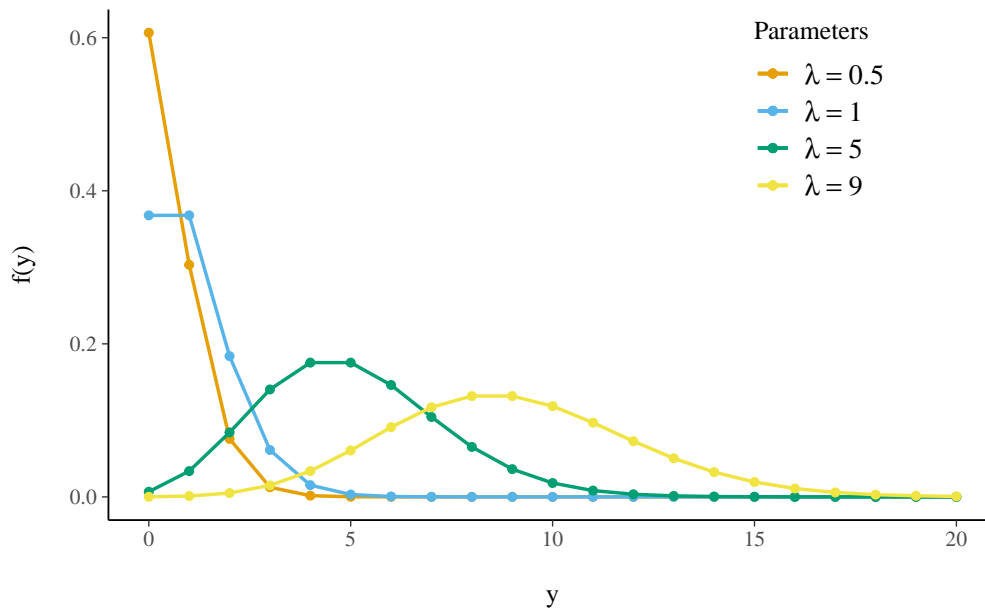
Alcune distribuzioni di Poisson sono riportate nella figura 1.2.

**Esercizio 1.5.** Supponiamo che un evento accada 300 volte all'ora e si vuole determinare la probabilità che in un minuto accadano esattamente 3 eventi.

Il numero medio di eventi in un minuto è pari a

---

<sup>3</sup>L'eguaglianza di  $\mu$  e  $\sigma$  è solo una peculiarità di questo esempio.



**Figura 1.2:** Alcune distribuzioni di Poisson.

```
lambda <- 300 / 60
lambda
#> [1] 5
```

Quindi la probabilità che in un minuto si abbiano 3 eventi è pari a

```
y <- 3
(lambda^y / factorial(y)) * exp(-lambda)
#> [1] 0.14
```

**Esercizio 1.6.** Per i dati dell'esempio precedente, si trovi la probabilità che un evento accada almeno 8 volte in un minuto.

La probabilità cercata è

$$p(y \geq 8) = 1 - p(y \leq 7) = 1 - \sum_{i=0}^7 \frac{\lambda^i}{i!} e^{-\lambda},$$

con  $\lambda = 5$ .

Svolgendo i calcoli in Rotteniamo:

```
1 - ppois(q = 7, lambda = 5)
#> [1] 0.133
ppois(q = 7, lambda = 5, lower.tail = FALSE)
#> [1] 0.133
```

**Esercizio 1.7.** Sapendo che un evento avviene in media 6 volte al minuto, si calcoli (a) la probabilità di osservare un numero di eventi uguale o inferiore a 3 in un minuto, e (b) la probabilità di osservare esattamente 2 eventi in 30 secondi.

(a) In questo caso  $\lambda = 6$  e la probabilità richiesta è

```
pPois(q = 3, lambda = 6, lower.tail = TRUE)
#> [1] 0.151
```

(b) In questo caso  $\lambda = 6/2$  e la probabilità richiesta è

```
dPois(x = 2, lambda = 3)
#> [1] 0.224
```

## 1.4 Alcune proprietà della variabile di Poisson

- Il valore atteso, la moda e la varianza della variabile di Poisson sono uguali a  $\lambda$ .
- La somma  $Y_1 + \dots + Y_n$  di  $n$  variabili casuali indipendenti con distribuzioni di Poisson di parametri  $\lambda_1, \dots, \lambda_n$  segue una distribuzione di Poisson di parametro  $\lambda = \lambda_1 + \dots + \lambda_n$ .
- La differenza di due variabili di Poisson non è una variabile di Poisson. Basti infatti pensare che può assumere valori negativi.

## Considerazioni conclusive

La distribuzione binomiale è una distribuzione di probabilità discreta che descrive il numero di successi in un processo di Bernoulli, ovvero la variabile aleatoria  $Y = Y_1 + \dots + Y_n$  che somma  $n$  variabili casuali indipendenti di uguale distribuzione di Bernoulli  $\mathcal{B}(\theta)$ , ognuna delle quali può fornire due soli risultati: il successo con probabilità  $\theta$  e il fallimento con probabilità  $1 - \theta$ .

La distribuzione binomiale è molto importante per le sue molte applicazioni. Nelle presenti dispense, dedicate all'analisi bayesiana, è soprattutto importante perché costituisce il fondamento del caso più semplice del cosiddetto "aggiornamento bayesiano", ovvero il caso Beta-Binomiale. Il modello Beta-Binomiale ci fornirà infatti un esempio paradigmatico dell'approccio bayesiano all'inferenza e sarà trattato in maniera analitica. È dunque importante che le proprietà della distribuzione binomiale risultino ben chiare.



# Bibliografia

- Burger, E. B. & Starbird, M. (2012). *The 5 elements of effective thinking*. Princeton University Press. (Cit. a p. ix).
- Horn, S. & Loewenstein, G. (2021). Underestimating Learning by Doing. *Available at SSRN 3941441* (cit. a p. x).



## Elenco delle figure

1.1	Alcune distribuzioni binomiali. Nella figura, il parametro $\theta$ è indicato con $p$ .	3
1.2	Alcune distribuzioni di Poisson. . . . .	5

**Abstract** This document contains the material of the lessons of Psicometria B000286 (2021/2022) aimed at students of the first year of the Degree Course in Psychological Sciences and Techniques of the University of Florence, Italy.

**Keywords** Data science, Bayesian statistics.