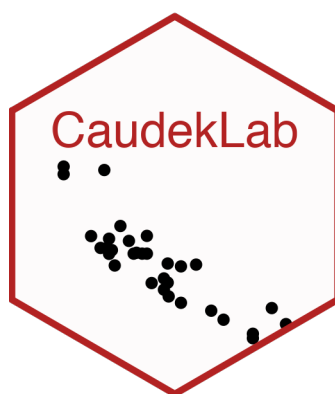


# Psicometria

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>v</b>
La psicologia e la Data Science . . . . .	v
Come studiare . . . . .	vi
Sviluppare un metodo di studio efficace . . . . .	vii
<b>Nozioni di base</b>	<b>3</b>
<b>1 Il calcolo delle probabilità</b>	<b>3</b>
1.1 La probabilità come la logica della scienza . . . . .	3
1.2 Che cos'è la probabilità? . . . . .	5
1.3 Variabili casuali e probabilità di un evento . . . . .	5
Variabili casuali . . . . .	5
Eventi e probabilità . . . . .	6
1.4 Spazio campionario e risultati possibili . . . . .	6
1.5 Usare la simulazione per stimare le probabilità . . . . .	7
1.6 La legge dei grandi numeri . . . . .	9
1.7 Variabili casuali multiple . . . . .	11
1.8 Funzione di massa di probabilità . . . . .	13
Considerazioni conclusive . . . . .	15
<b>Inferenza statistica bayesiana</b>	<b>19</b>
<b>2 Il problema inverso</b>	<b>19</b>
2.1 Inferenza bayesiana come un problema inverso . . . . .	19
Funzioni di probabilità . . . . .	19
La regola di Bayes . . . . .	20
2.2 Inferenza bayesiana . . . . .	21
Notazione . . . . .	22
Il problema inverso . . . . .	22
Cos'è un parametro del modello? . . . . .	22
La distribuzione a priori sui parametri . . . . .	23
Scelta della distribuzione a priori . . . . .	24
2.3 Verosimiglianza marginale . . . . .	25
2.4 La distribuzione a posteriori . . . . .	25
Considerazioni conclusive . . . . .	25

<b>Bibliografia</b>	<b>27</b>
<b>Elenco delle figure</b>	<b>29</b>

# 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.

L'insegnamento di 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 le 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 Psicometria molte volte in passato 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.

- 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. Istintivamente 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. 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

Febbraio 2022



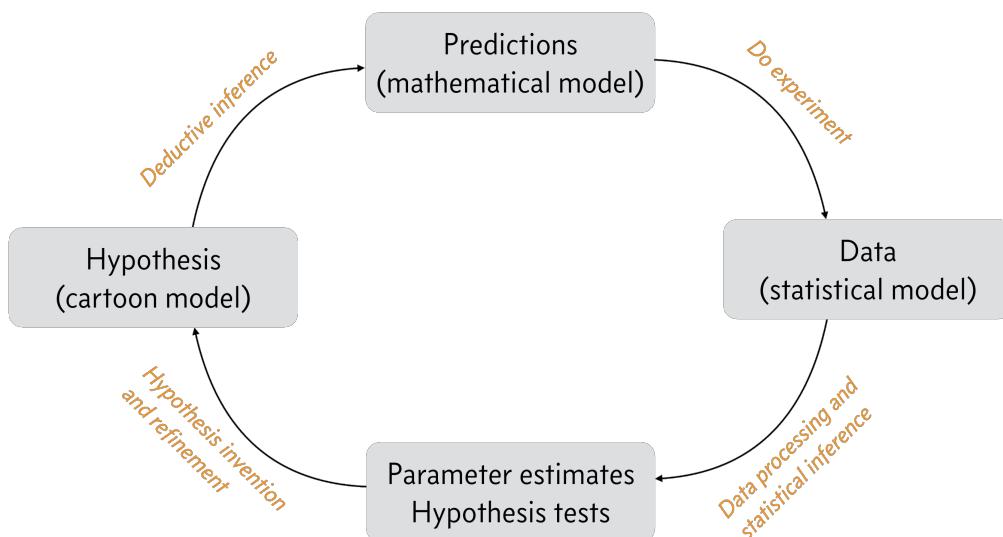
# Nozioni di base



# Il calcolo delle probabilità

## 1.1 La probabilità come la logica della scienza

La figura 1.1 fornisce una rappresentazione schematica del processo dell'indagine scientifica. Possiamo pensare al progresso scientifico come alla ripetizione di questo ciclo mentre i fenomeni naturali (e, ovviamente psicologici) vengono esplorati e i ricercatori imparano sempre di più sul loro funzionamento. Le caselle della figura riportano le varie fasi del processo di ingagine scientifica e lungo le frecce sono riportati i compiti che conducono i ricercatori da una fase alla successiva.



**Figura 1.1:** Rappresentazione schematica del processo scientifico (figura adattata dalla Fig. 1.1 di P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge, 2005).

Consideriamo i compiti e le fasi dell'indagine scientifica. Iniziamo in basso a

sinistra.

- *Invenzione e perfezionamento delle ipotesi* In questa fase del processo scientifico, i ricercatori pensano ai fenomeni naturali, a ciò che è presente nella letteratura scientifica, anche ai risultati dei loro esperimenti, e formulano ipotesi o teorie che possono valutare con degli esperimenti. Questo passaggio richiede innovazione e creatività.
- *L'inferenza deduttiva* procede in maniera deterministica dai fatti verso le conclusioni. Ad esempio, se dico che tutti gli uomini sono mortali e che Socrate è un uomo, posso concludere deduttivamente che Socrate è mortale. Quando i ricercatori progettano i loro esperimenti in base alle loro ipotesi, usano la logica deduttiva per dire: "Se A è vero, allora B è vero", dove A è l'ipotesi considerata e B è un'osservazione sperimentale.
- *Esecuzione degli esperimenti.* Questa fase richiede molte risorse, di tempo e denaro. Richiede anche innovazione e creatività. Nello specifico, i ricercatori devono pensare attentamente a come costruire il loro esperimento per verificare l'ipotesi di interesse. Il risultato degli esperimenti sono i dati.
- *L'inferenza induttiva*, invece, procede dalle osservazioni ai fatti. Se pensiamo ai fatti come a ciò che governa o genera le osservazioni, allora l'induzione è una sorta di inferenza inversa. Supponiamo di avere osservato B. Questo rende A vero? Non necessariamente. Ma può rendere A più plausibile. Questo è un sillogismo debole. Ad esempio, si consideri la seguente coppia ipotesi/osservazioni.
  - A = L'iniezione di acque reflue dopo la fratturazione idraulica, nota come fracking, può portare a una maggiore frequenza di terremoti.
  - B = La frequenza dei terremoti in Oklahoma è aumentata di 100 volte dal 2010, quando il fracking è diventato una pratica comune.
  - Poiché B è stato osservato, A è più plausibile. A non è necessariamente vero, ma decisamente più plausibile.
- *L'inferenza statistica* è un tipo di inferenza induttiva che è specificamente formulata come un problema inverso. L'inferenza statistica è quell'insieme di procedure che hanno lo scopo di quantificare quanto più plausibile è A dopo aver osservato B. Per svolgere l'inferenza statistica, è necessario quantificare la plausibilità. Lo strumento che ci consente di fare questo è la teoria della probabilità.

L'inferenza statistica è l'aspetto del processo dell'indagine scientifica che è l'oggetto centrale di questo insegnamento. Il risultato dell'inferenza statistica è la conoscenza di quanto siano plausibili le ipotesi e le stime dei parametri sotto le ipotesi considerate. Ma l'inferenza statistica richiede una teoria della probabilità, laddove la teoria della probabilità può essere vista come una generalizzazione della logica. A causa di questa connessione con la logica e del suo ruolo cruciale nella scienza, E. T. Jaynes afferma infatti che la probabilità è la "logica della scienza". È dunque necessario esaminare preliminarmente alcune nozioni di base della teoria della probabilità.

## 1.2 Che cos'è la probabilità?

La definizione della probabilità è un problema estremamente dibattuto ed aperto. Sono state fornite due possibili soluzioni al problema di definire il concetto di probabilità.

- (a) La natura della probabilità è “ontologica” (ovvero, basata sulla metafisica): la probabilità è una proprietà della realtà, del mondo, di come sono le cose, indipendentemente dalla nostra esperienza. È una visione che qualcuno chiama “oggettiva”.
- (b) La natura della probabilità è “epistemica” (ovvero, basata sulla conoscenza): la probabilità si riferisce alla conoscenza che abbiamo del mondo, non al mondo in sé. Di conseguenza è detta, in contrapposizione alla precedente definizione, “soggettiva”.

In termini epistemici, la probabilità fornisce una misura della nostra incertezza sul verificarsi di un fenomeno, alla luce delle informazioni disponibili. Potremmo dire che c'è una “scala” naturale che ha per estremi il vero (1: evento certo) da una parte ed il falso (0: evento impossibile) dall'altra. La probabilità è la quantificazione di questa scala: quantifica lo stato della nostra incertezza rispetto al contenuto di verità di una proposizione (ovvero, quantifica la plausibilità di una proposizione).

- Nell'interpretazione frequentista della probabilità, la probabilità  $P(A)$  rappresenta la frequenza relativa a lungo termine nel caso di un grande numero di ripetizioni di un esperimento casuale sotto le medesime condizioni. L'evento  $A$  deve essere una proposizione relativa alle variabili casuali<sup>1</sup>.
- Nell'interpretazione bayesiana della probabilità  $P(A)$  rappresenta direttamente il grado di credenza, o plausibilità, a proposito di  $A$ , dove  $A$  può essere qualsiasi proposizione logica.

In questo insegnamento utilizzeremo l'interpretazione bayesiana della probabilità.

## 1.3 Variabili casuali e probabilità di un evento

Esaminiamo qui di seguito alcuni concetti di base della teoria della probabilità.

### Variabili casuali

Sia  $Y$  il risultato del lancio di moneta equilibrata. Non di un generico lancio di una moneta, ma un'istanza specifica del lancio di una specifica moneta in

---

<sup>1</sup>Viene stressata qui l'idea che ciò di cui parliamo è qualcosa che emerge nel momento in cui è possibile ripetere l'esperimento casuale tante volte sotto le medesime condizioni. Le variabili casuali, infatti, forniscono una quantificazione dei risultati che si ottengono ripetendo tante volte l'esperimento casuale sotto le medesime condizioni.

un dato momento. Definita in questo modo,  $Y$  è una *variabile casuale*, ovvero una variabile che assume valori diversi con probabilità diverse. Se la moneta è equilibrata, c'è una probabilità del 50% che il lancio della moneta dia come risultato “testa” e una probabilità del 50% che dia come risultato “croce”.

Per facilitare la trattazione, le variabili casuali assumono solo valori numerici. Per lo specifico lancio della moneta in questione, diciamo, ad esempio, che la variabile casuale  $Y$  assume il valore 1 se esce testa e il valore 0 se esce croce.

## Eventi e probabilità

Nella teoria della probabilità il risultato “testa” nel lancio di una moneta è chiamato *evento*.<sup>2</sup> Ad esempio,  $Y = 1$  denota l'evento in cui il lancio di una moneta produce come risultato testa.

Il funzionale  $Pr[\cdot]$  definisce la probabilità di un evento. Ad esempio, per il lancio di una moneta equilibrata, la probabilità dell'evento “il risultato del lancio della moneta è testa” è scritta come

$$Pr[Y = 1] = 0.5.$$

Se la moneta è equilibrata dobbiamo anche avere  $Pr[Y = 0] = 0.5$ . I due eventi  $Y = 1$  e  $Y = 0$  sono *mutuamente esclusivi* nel senso che non possono entrambi verificarsi contemporaneamente. Nella notazione probabilistica,

$$Pr[Y = 1 \text{ e } Y = 0] = 0.$$

Gli eventi  $Y = 1$  e  $Y = 0$  si dicono *esaustivi*, nel senso che almeno uno di essi deve verificarsi e nessun altro tipo di evento è possibile. Nella notazione probabilistica,

$$Pr[Y = 1 \text{ o } Y = 0] = 1.$$

Il connettivo logico “e” specifica eventi *congiunti*, ovvero eventi che possono verificarsi contemporaneamente (eventi *compatibili*) e per i quali, perciò, la probabilità della loro congiunzione è  $Pr(A \text{ e } B) > 0$ . Il connettivo logico “o” specifica eventi *disgiunti*, ovvero eventi che non possono verificarsi contemporaneamente (eventi *incompatibili*) e per i quali, perciò, la probabilità della loro congiunzione è  $Pr(A \text{ e } B) = 0$ .

## 1.4 Spazio campionario e risultati possibili

Anche se il lancio di una moneta produce sempre uno specifico risultato nel mondo reale, noi possiamo anche immaginare i possibili risultati alternativi che si sarebbero potuti osservare. Quindi, anche se in uno specifico lancio la moneta dà testa ( $Y = 1$ ), possiamo immaginare la possibilità che il lancio possa avere prodotto croce ( $Y = 0$ ). Tale ragionamento controfattuale è la chiave per comprendere la teoria della probabilità e l'inferenza statistica.

---

<sup>2</sup>Per un ripasso delle nozioni di base della teoria degli insiemi, si veda l'Appendice ??.



I risultati possibili che si possono osservare come conseguenza del lancio di una moneta determinano i valori possibili che la variabile casuale può assumere. L'insieme di tutti i risultati possibili è chiamato *spazio campionario*. Lo spazio campionario può essere concettualizzato come un'urna contenente una pallina per ogni possibile risultato del lancio della moneta. Su ogni pallina è scritto il valore della variabile casuale. Uno specifico lancio di una moneta – ovvero, l'osservazione di uno specifico valore di una variabile casuale – è chiamato *esperimento casuale*.

Il lancio di un dado ci fornisce l'esempio di un altro esperimento casuale. Supponiamo di essere interessati all'evento “il lancio del dado produce un numero dispari”. Un *evento* seleziona un sottoinsieme dello spazio campionario: in questo caso, l'insieme dei risultati  $\{1, 3, 5\}$ . Se esce 3, per esempio, diciamo che si è verificato l'evento “dispari” (ma l'evento “dispari” si sarebbe anche verificato anche se fosse uscito 1 o 5).

## 1.5 Usare la simulazione per stimare le probabilità

I metodi basati sulla simulazione ci consentono di stimare le probabilità degli eventi in un modo diretto se siamo in grado di generare realizzazioni molteplici e casuali delle variabili casuali coinvolte nelle definizioni degli eventi. Per simulare il lancio di una moneta equilibrata in R iniziamo a definire un vettore che contiene i possibili risultati del lancio della moneta (ovvero i possibili valori della variabile casuale  $Y$ ):

```
coin <- c(0, 1)
```

L'estrazione casuale di uno di questi due possibili valori (ovvero, la simulazione di uno specifico lancio di una moneta) si realizza con la funzione `sample()`:

```
sample(coin, size = 1)
#> [1] 0
```

In maniera equivalente, lo stesso risultato si ottiene mediante l'istruzione

```
rbinom(1, 1, 0.5)
#> [1] 1
```

Supponiamo di ripetere questo esperimento casuale 100 volte e di registrare i risultati così ottenuti. La stima della probabilità dell'evento  $Pr[Y = 1]$  è data dalla frequenza relativa del numero di volte in cui abbiamo osservato l'evento di interesse ( $Y = 1$ ):

```
M <- 10
y <- rep(NA, M)
for (m in 1:M) {
  y[m] <- rbinom(1, 1, 0.5)
}
estimate <- sum(y) / M
```

```
cat("estimated Pr[Y = 1] =", estimate)
#> estimated Pr[Y = 1] = 0.7
```

Ripetiamo questa procedura 10 volte.

```
flip_coin <- function(M) {
  y <- rep(NA, M)
  for (m in 1:M) {
    y[m] <- rbinom(1, 1, 0.5)
  }
  estimate <- sum(y) / M
  cat("estimated Pr[Y = 1] =", estimate, "\n")
}
```

```
for (i in 1:10) {
  flip_coin(10)
}
#> estimated Pr[Y = 1] = 0.5
#> estimated Pr[Y = 1] = 0.7
#> estimated Pr[Y = 1] = 0.4
#> estimated Pr[Y = 1] = 0.6
#> estimated Pr[Y = 1] = 0.6
#> estimated Pr[Y = 1] = 0.5
#> estimated Pr[Y = 1] = 0.3
#> estimated Pr[Y = 1] = 0.4
#> estimated Pr[Y = 1] = 0.7
#> estimated Pr[Y = 1] = 0.7
```

Dato che la moneta è equilibrata, la stima delle probabilità dell'evento  $Pr[Y = 1]$  è simile a al valore che ci aspettiamo ( $Pr[Y = 1] = 0.5$ ), ma il risultato ottenuto nelle varie simulazioni non è sempre esatto. Proviamo ad aumentare il numero di lanci in ciascuna simulazione:

```
for (i in 1:10) {
  flip_coin(100)
}
#> estimated Pr[Y = 1] = 0.54
#> estimated Pr[Y = 1] = 0.35
#> estimated Pr[Y = 1] = 0.48
#> estimated Pr[Y = 1] = 0.41
#> estimated Pr[Y = 1] = 0.48
#> estimated Pr[Y = 1] = 0.5
#> estimated Pr[Y = 1] = 0.49
#> estimated Pr[Y = 1] = 0.45
#> estimated Pr[Y = 1] = 0.47
#> estimated Pr[Y = 1] = 0.48
```

In questo secondo caso, gli errori tendono ad essere più piccoli della simulazione precedente. Cosa succede se in ciascuna simulazione esaminiamo i risultati di 10,000 lanci della moneta?

```
for (i in 1:10) {  
  flip_coin(1e4)  
}  
#> estimated Pr[Y = 1] = 0.5045  
#> estimated Pr[Y = 1] = 0.5004  
#> estimated Pr[Y = 1] = 0.4965  
#> estimated Pr[Y = 1] = 0.4986  
#> estimated Pr[Y = 1] = 0.5092  
#> estimated Pr[Y = 1] = 0.5033  
#> estimated Pr[Y = 1] = 0.5028  
#> estimated Pr[Y = 1] = 0.5011  
#> estimated Pr[Y = 1] = 0.4974  
#> estimated Pr[Y = 1] = 0.5024
```

Ora le stime ottenute sono molto vicine alla vera probabilità che vogliamo stimare (cioè 0.5, perché la moneta è equilibrata). I risultati delle simulazioni precedenti pongono dunque il problema di determinare quale sia il numero di lanci di cui abbiamo bisogno per assicurarci che le stime siano accurate (ovvero, vicine al valore corretto della probabilità)

## 1.6 La legge dei grandi numeri

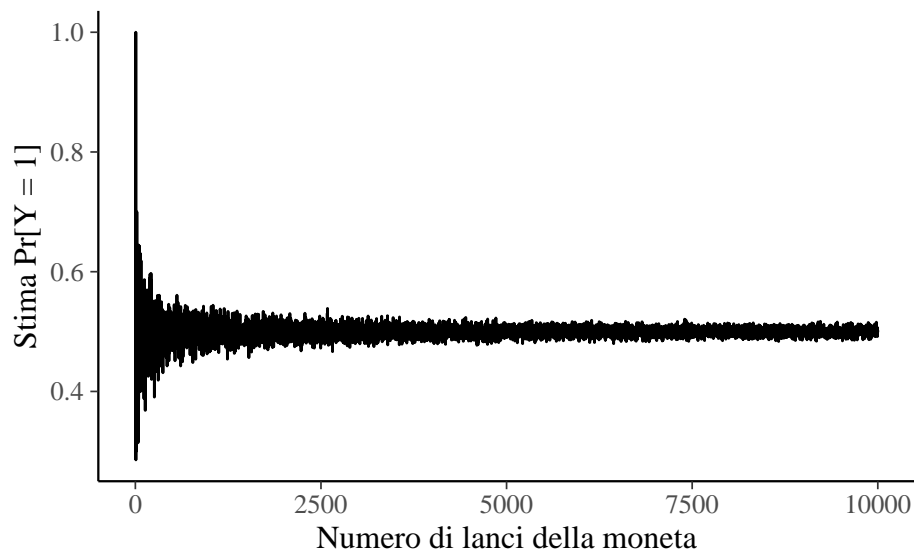
La visualizzazione mediante grafici contribuisce alla comprensione dei concetti della statistica e della teoria della probabilità. Un modo per descrivere ciò che accade all'aumentare del numero  $M$  di ripetizioni del lancio della moneta consiste nel registrare la stima della probabilità dell'evento  $Pr[Y = 1]$  in funzione del numero di ripetizioni dell'esperimento casuale per ogni  $m \in 1 : M$ . Un grafico dell'andamento della stima di  $Pr[Y = 1]$  in funzione di  $m$  si ottiene nel modo seguente.

```
nrep <- 1e4  
estimate <- rep(NA, nrep)  
  
flip_coin <- function(m) {  
  y <- rbinom(m, 1, 0.5)  
  phat <- sum(y) / m  
  phat  
}  
  
for (i in 1:nrep) {  
  estimate[i] <- flip_coin(i)  
}  
  
d <- data.frame(  

```

```
n = 1:nrep,
estimate
)

d %>%
  ggplot(
    aes(x = n, y = estimate)
  ) +
  geom_line() +
  theme(legend.title = element_blank()) +
  labs(
    x = "Numero di lanci della moneta",
    y = "Stima Pr[Y = 1]"
  )
)
```



**Figura 1.2:** Stima della probabilità di successo in funzione del numero di lanci di una moneta.

Dato che il grafico 1.2 su una scala lineare non rivela chiaramente l'andamento della simulazione, utilizzeremo invece un grafico in cui sull'asse  $x$  è stata imposta una scala logaritmica. Con l'asse  $x$  su scala logaritmica, i valori tra 1 e 10 vengono tracciati all'incirca con la stessa ampiezza come nel caso dei valori tra 50 e 700, eccetera.

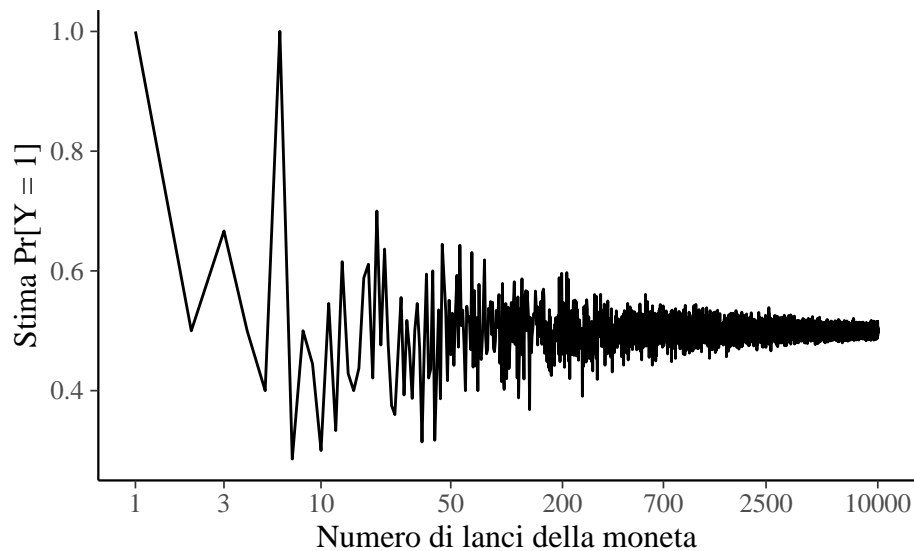
```
d %>%
  ggplot(
    aes(x = n, y = estimate)
  ) +
  geom_line() +
  scale_x_log10(
    breaks = c(
      1, 3, 10, 50, 200,

```

```

700, 2500, 10000
)
) +
theme(legend.title = element_blank()) +
labs(
  x = "Numero di lanci della moneta",
  y = "Stima Pr[Y = 1]"
)

```



**Figura 1.3:** Stima della probabilità di successo in funzione del numero di lanci di una moneta – scala logaritmica.

La *legge dei grandi numeri* ci dice che all'aumentare del numero di ripetizioni dell'esperimento casuale la media dei risultati ottenuti tenderà ad avvicinarsi al valore atteso man mano che verranno eseguite più prove. Nel caso presente, la figura 1.3 mostra appunto che, all'aumentare del numero  $M$  di lanci della moneta, la stima di  $Pr[Y = 1]$  tende a convergere al vero valore di 0.5.

## 1.7 Variabili casuali multiple

Le variabili casuali non esistono isolatamente. Abbiamo iniziato con una singola variabile casuale  $Y$  che rappresenta il risultato di un singolo, specifico lancio di una moneta equilibrata. Ma supponiamo ora di lanciare la moneta tre volte. Ciò suggerisce che possiamo avere le variabili casuali  $Y_1, Y_2, Y_3$  che rappresentano i risultati di ciascuno dei lanci. Possiamo assumere che ogni lancio sia indipendente, ovvero che non dipenda dal risultato degli altri lanci. Ognuna di queste variabili  $Y_n$  per  $n \in 1 : 3$  ha  $Pr[Y_n = 1] = 0.5$  e  $Pr[Y_n = 0] = 0.5$ . Possiamo combinare più variabili casuali usando le operazioni aritmetiche. Se  $Y_1, Y_2, Y_3$  sono variabili casuali che rappresentano tre lanci di una moneta equilibrata (o un lancio di tre monete equilibrate), possiamo definire la somma di tali variabili casuali come

$$Z = Y_1 + Y_2 + Y_3.$$

Possiamo simulare i valori assunti dalla variabile casuale  $Z$  simulando i valori di  $Y_1, Y_2, Y_3$  per poi sommarli.

```
y1 <- rbinom(1, 1, 0.5)
y2 <- rbinom(1, 1, 0.5)
y3 <- rbinom(1, 1, 0.5)
c(y1, y2, y3)
#> [1] 0 0 0
z <- sum(c(y1, y2, y3))
cat("z =", z, "\n")
#> z = 0
```

ovvero,

```
y <- rep(NA, 3)
for (i in 1:3) {
  y[i] <- rbinom(1, 1, 0.5)
}
y
#> [1] 1 1 0
z <- sum(y)
cat("z =", z, "\n")
#> z = 2
```

oppure, ancora più semplicemente:

```
y <- rbinom(3, 1, 0.5)
y
#> [1] 1 0 1
z <- sum(y)
cat("z =", z, "\n")
#> z = 2
```

Possiamo ripetere questa simulazione  $M = 1e5$  volte:

```
M <- 1e5
z <- rep(NA, M)
for (i in 1:M) {
  y <- rbinom(3, 1, 0.5)
  z[i] <- sum(y)
}
```

e calcolare una stima della probabilità che la variabile casuale  $Z$  assuma i valori 0, 1, 2, 3:

```
table(z) / M
#> z
#>      0      1      2      3
#> 0.12608 0.37225 0.37666 0.12501
```

Nel caso di 4 monete equilibrate, avremo:

```
M <- 1e5
z <- rep(NA, M)
for (i in 1:M) {
  y <- rbinom(4, 1, 0.5)
  z[i] <- sum(y)
}
table(z) / M
#> z
#>      0      1      2      3      4
#> 0.06303 0.24984 0.37314 0.25090 0.06309
```

Viene detta *variabile casuale discreta* una variabile casuale le cui modalità possono essere costituite solo da numeri interi:

$$\mathbb{Z} = \dots, -2, -1, 0, 1, 2, \dots$$

## 1.8 Funzione di massa di probabilità

È conveniente avere una funzione che associa ogni possibile valore di una variabile casuale alla sua probabilità. In generale, ciò è possibile se e solo se la variabile casuale è discreta, così com'è stata definita nel Paragrafo precedente.

Ad esempio, se consideriamo  $Z = Y_1 + \dots + Y_4$  come il numero di risultati “testa” in 4 lanci della moneta, allora possiamo definire la seguente funzione:

$p_Z(0)$	$=$	$1/16$	TTTT
$p_Z(1)$	$=$	$4/16$	H T T T, T H T T, T T H T, T T T H
$p_Z(2)$	$=$	$6/16$	H H T T, H T H T, H T T H, T H H T, T H T H, T T T H
$p_Z(3)$	$=$	$4/16$	H H H T, H H T H, H T H H, T H H H
$p_Z(4)$	$=$	$1/16$	H H H H

Il lancio di quattro monete può produrre sedici possibili risultati. Dato che i lanci sono indipendenti e le monete sono equilibrate, ogni possibile risultato è ugualmente probabile. Nella tabella in alto, le sequenze dei risultati possibili del lancio delle 4 monete sono riportate nella colonna più a destra. Le probabilità si ottengono dividendo il numero di sequenze che producono lo stesso numero di eventi testa per il numero dei risultati possibili.

La funzione  $p_Z$  è stata costruita per mappare un valore  $u$  per  $Z$  alla probabilità dell'evento  $Z = u$ . Convenzionalmente, queste probabilità sono scritte come

$$p_Z(z) = \Pr[Z = z].$$

La parte a destra dell'uguale si può leggere come: “la probabilità che la variabile casuale  $Z$  assuma il valore  $z$ ”.

Una funzione definita come sopra è detta *funzione di massa di probabilità* della variabile casuale  $Z$ . Ad ogni variabile casuale discreta è associata un'unica funzione di massa di probabilità.

Una rappresentazione grafica della stima della funzione di massa di probabilità per l'esperimento casuale del lancio di quattro monete equilibrate è fornita nella figura 1.4.

```
set.seed(1234)
M <- 1e5
nflips <- 4
u <- rbinom(M, nflips, 0.5)
x <- 0:nflips
y <- rep(NA, nflips + 1)
for (n in 0:nflips) {
  y[n + 1] <- sum(u == n) / M
}
bar_plot <-
  data.frame(Z = x, count = y) %>%
  ggplot(
    aes(x = Z, y = count)
  ) +
  geom_bar(stat = "identity") +
  scale_x_continuous(
    breaks = 0:4,
    labels = c(0, 1, 2, 3, 4)
  ) +
  labs(
    y = "Probabilità stimata Pr[Z = z]"
  )
bar_plot
```

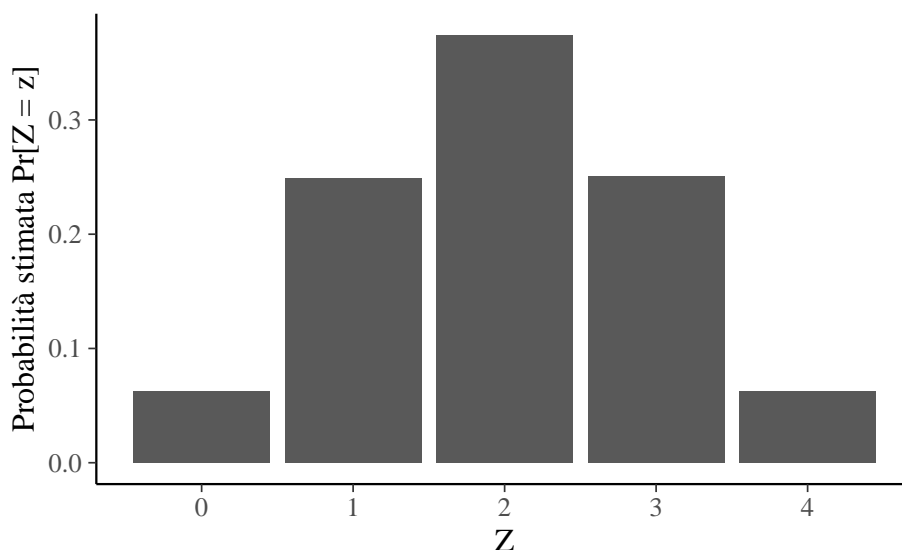
Se  $A$  è un sottoinsieme della variabile casuale  $Z$ , allora denotiamo con  $P_z(A)$  la probabilità assegnata ad  $A$  dalla distribuzione  $P_z$ . Mediante una distribuzione di probabilità  $P_z$  è dunque possibile determinare la probabilità di ciascun sottoinsieme  $A \subset Z$  come

$$P_z(A) = \sum_{z \in A} P_z(Z).$$

**Esempio 1.1.** Per esempio, la probabilità che  $Z$  sia un numero dispari è

$$\Pr(Z \text{ è un numero dispari}) = P_z(Z = 1) + P_z(Z = 3) = \frac{4}{16} + \frac{4}{16} = \frac{1}{2}.$$





**Figura 1.4:** Grafico di  $M = 100\,000$  simulazioni della funzione di massa di probabilità di una variabile casuale definita come il numero di teste in quattro lanci di una moneta equilibrata.

## Considerazioni conclusive

In questo capitolo abbiamo visto come si costruisce lo spazio campionario di un esperimento casuale, quali sono le proprietà di base della probabilità e come si assegnano le probabilità agli eventi definiti sopra uno spazio campionario discreto. Abbiamo anche introdotto le nozioni di “variabile casuale”, ovvero di una variabile che prende i suoi valori casualmente. E abbiamo descritto il modo di specificare la probabilità con cui sono presi i differenti valori, ovvero la funzione di distribuzione probabilistica

$$F(X) = \Pr(X < x),$$

e la funzione di massa di probabilità. Le procedure di analisi dei dati psicologici che discuteremo in seguito faranno un grande uso di questi concetti e della notazione qui introdotta.



# Inferenza statistica bayesiana



## Il problema inverso

### 2.1 Inferenza bayesiana come un problema inverso

L'inferenza bayesiana è formulata nei termini di un problema inverso che segue la regola di Bayes (si veda la Sezione 1.1).<sup>1</sup> Per fissare la notazione, nel seguito  $y$  rappresenterà le variabili osservate, ovvero i dati, e  $\theta$  rappresenterà i parametri incogniti di un modello statistico. Sia  $y$  che  $\theta$  sono concepiti come delle variabili casuali. Con  $x$  verranno invece denotate le quantità note, come i predittori nel modello di regressione.

#### Funzioni di probabilità

L'inferenza bayesiana utilizza le seguenti distribuzioni di probabilità (o densità di probabilità):

- la *distribuzione a priori*  $p(\theta)$  — la credenza iniziale riguardo alla credibilità di ciascun valore  $\theta$ ;
- la *funzione di verosimiglianza*  $p(y | \theta)$  — la credibilità che il ricercatore assegnerebbe ai dati osservati se conoscesse il parametro di interesse  $\theta$ ;
- la *verosimiglianza marginale*  $p(y)$  — quanto sono credibili i dati  $y$  alla luce della nostra credenza a priori relativamente a  $\theta$ . In termini formali:

$$p(y) = \int_{\theta} p(y, \theta) d\theta = \int_{\theta} p(y | \theta) p(\theta) d\theta.$$

---

<sup>1</sup>Nell'approccio bayesiano non si fa riferimento ad un modello probabilistico  $f(y | \theta)$  rappresentativo del fenomeno d'interesse noto a meno del valore assunto dal parametro (o dei parametri) che lo caratterizza. Si fa invece riferimento ad una distribuzione congiunta (di massa o di densità di probabilità)  $f(y, \theta)$ . Entrambi gli argomenti della funzione  $y$  e  $\theta$  hanno natura di variabili casuali, laddove la nostra incertezza relativa a  $y$  è dovuta alla naturale variabilità del fenomeno indagato (*variabilità aleatoria*), mentre la nostra incertezza relativa a  $\theta$  è dovuta alla mancata conoscenza del suo valore numerico (*variabilità epistemica*).

- la *distribuzione a posteriori*  $p(\theta | y)$  — la nuova credenza a posteriori relativamente alla credibilità di ciascun valore  $\theta$  alla luce dei dati  $Y = y$ .

### La regola di Bayes

Nel contesto di un modello statistico, la formula di Bayes permette di giungere alla distribuzione a posteriori  $p(\theta | y)$  per il parametro di interesse  $\theta$ , come indicato dalla seguente catena di equazioni:

$$p(\theta | y) = \frac{p(\theta, y)}{p(y)} \quad [\text{definizione di probabilità condizionata}] \quad (2.1)$$

$$= \frac{p(y | \theta) p(\theta)}{p(y)} \quad [\text{legge della probabilità composta}] \quad (2.2)$$

$$= \frac{p(y | \theta) p(\theta)}{\int_{\Theta} p(y, \theta) d\theta} \quad [\text{legge della probabilità totale}] \quad (2.3)$$

$$= \frac{p(y | \theta) p(\theta)}{\int_{\Theta} p(y | \theta) p(\theta) d\theta} \quad [\text{legge della probabilità composta}] \quad (2.4)$$

$$\propto p(y | \theta) p(\theta) \quad (2.5)$$

La regola di Bayes “inverte” la probabilità della distribuzione a posteriori  $p(\theta | y)$ , esprimendola nei termini della funzione di verosimiglianza  $p(y | \theta)$  e della distribuzione a priori  $p(\theta)$ . L’ultimo passo è importante per la stima della distribuzione a posteriori mediante i metodi Monte Carlo a catena di Markov, in quanto per questi metodi richiedono soltanto che le funzioni di probabilità siano definite a meno di una costante di proporzionalità. In altri termini, per la maggior parte degli scopi dell’inferenza inversa, è sufficiente calcolare la densità a posteriori non normalizzata, ovvero è possibile ignorare il denominatore bayesiano  $p(y)$ . La distribuzione a posteriori non normalizzata, dunque, si riduce al prodotto della verosimiglianza e della distribuzione a priori.

Possiamo dire che la regola di Bayes viene usata per aggiornare le credenze a priori su  $\theta$  (ovvero, la distribuzione a priori) in modo tale da produrre le nuove credenze a posteriori  $p(\theta | y)$  che combinano le informazioni fornite dai dati  $y$  con le credenze precedenti. La distribuzione a posteriori riflette dunque l’aggiornamento delle credenze del ricercatore alla luce dei dati.

La (2.5) rende evidente che, in ottica bayesiana, la quantità di interesse  $\theta$  non è fissata come nell’impostazione frequentista, ma è una variabile casuale la cui distribuzione di probabilità è influenzata sia dalle informazioni a priori sia dai dati a disposizione. In altre parole, nell’approccio bayesiano non esiste un valore vero di  $\theta$ , ma vogliamo fornire invece un giudizio di probabilità. Prima delle osservazioni, sulla base delle nostre conoscenze assegniamo a  $\theta$  una distribuzione a priori di probabilità. Dopo le osservazioni, correggiamo il nostro giudizio e assegniamo a  $\theta$  una distribuzione a posteriori di probabilità. La distribuzione a posteriori  $p(\theta | y)$  contiene tutta l’informazione riguardante il parametro  $\theta$  e viene utilizzata per produrre indicatori sintetici, per la determinazione di stime puntuali o intervallari, e per la verifica d’ipotesi.

## 2.2 Inferenza bayesiana

Un esempio di inferenza bayesiana è quello nel quale i dati sono rappresentati da una proporzione. Per questo tipo di dati possiamo adottare il seguente modello statistico

$$y \sim \text{Bin}(n, \theta), \quad (2.6)$$

laddove  $\theta$  è la probabilità che una prova Bernoulliana assuma il valore 1 e  $n$  corrisponde al numero di prove Bernoulliane. Questo modello assume che le prove Bernoulliane  $y_i$  che costituiscono il campione  $y$  siano tra loro indipendenti e che ciascuna abbia la stessa probabilità  $\theta \in [0, 1]$  di essere un “successo” (valore 1). In altre parole, il modello generatore dei dati avrà una funzione di massa di probabilità

$$p(y \mid \theta) = \text{Bin}(y \mid n, \theta).$$

Nei capitoli precedenti è stato mostrato come, sulla base del modello binomiale, sia possibile assegnare una probabilità a ciascun possibile valore  $y \in \{0, 1, \dots, n\}$  assumendo noto il valore del parametro  $\theta$ . Ma ora abbiamo il problema inverso, ovvero quello di fare inferenza su  $\theta$  alla luce dei dati campionari  $y$ . In altre parole, riteniamo di conoscere il modello probabilistico che ha generato i dati, ma di tale modello non conosciamo i parametri. Nel caso presente, il modello probabilistico è quello binomiale. Noi vogliamo ottenere informazioni sul valore di  $\theta$  conoscendo il numero  $y$  di successi osservati nel campione.

La (2.6) è un *modello statistico*. Tale modello non spiega perché, in ciascuna realizzazione,  $Y$  assuma un particolare valore. Questo modello deve piuttosto essere inteso come un costrutto matematico che ha lo scopo di riflettere alcune proprietà del processo corrispondente ad una sequenza di prove Bernoulliane. In questo senso, è simile al modello di Isaac Newton dei moti planetari che utilizza equazioni differenziali. Le equazioni non sono i pianeti, ma solo descrizioni di come si muovono i pianeti in risposta alle forze gravitazionali. Modelli come quello di Newton ci permettono di prevedere alcuni fenomeni, come il moto dei pianeti, ad esempio. Ma in generale i modelli sono solo delle approssimazioni del fenomeno che vogliono descrivere. Anche il modello di Newton, che produce previsioni estremamente accurate di ciò che possiamo osservare a occhio nudo a proposito del moto dei corpi celesti, è solo un'approssimazione dei modelli del moto e dei fenomeni gravitazionali che, in seguito, sono stati introdotti da Albert Einstein. E anche tali modelli successivi sono, a loro volta, solo un caso speciale della più generale teoria della relatività. In altre parole, modelli sempre migliori vengono proposti, laddove ogni successivo modello è migliore di quello precedente in quanto ne migliora le capacità di previsione, è più generale, o è più elegante.

Una parte del lavoro della ricerca in tutte le scienze consiste nel verificare le assunzioni dei modelli e, se necessario, nel migliorare i modelli dei fenomeni considerati. Un modello viene giudicato in relazione al suo obiettivo. Se l'obiettivo del modello molto semplice che stiamo discutendo è quello di prevedere la proporzione di casi nei quali  $y_i = 1$ ,  $i = 1, \dots, n$ , allora un modello con un solo parametro come quello che abbiamo introdotto sopra può essere sufficiente. Ma l'evento  $y_i = 1$  (supponiamo: superare l'esame di Psicometria, oppure risultare positivi al COVID-19) dipende da molti fattori e se vogliamo rendere conto di una tale

complessità, un modello come quello che stiamo discutendo qui certamente non sarà sufficiente.

Per concludere, un modello è un costrutto matematico il cui scopo è quello di rappresentare un qualche aspetto della realtà. Il valore di un tale strumento dipende dalla sua capacità di ottenere lo scopo per cui è stato costruito.

### Notazione

Per rappresentare in un modo conciso i modelli statistici viene usata una notazione particolare. Ad esempio, invece di scrivere

$$p(\theta) = \text{Beta}(1, 1),$$

scriviamo:

$$\theta \sim \text{Beta}(1, 1).$$

Il simbolo “ $\sim$ ” viene spesso letto “è distribuito come”. Possiamo anche pensare che significhi che  $\theta$  costituisce un campione casuale estratto dalla distribuzione  $\text{Beta}(1, 1)$ . Allo stesso modo, per l’esempio presente, la verosimiglianza può essere scritta come:

$$y \sim \text{Bin}(n, \theta).$$

### Il problema inverso

Nel modello statistico che stiamo esaminando, il termine  $n$  viene trattato come una costante nota e  $\theta$  come una *variabile casuale*. Il parametro  $\theta$  del modello rappresenta la probabilità che ciascuna prova Bernoulliana sia un “successo”. Dato che  $\theta$  è incognito, ma abbiamo a disposizione un campione di dati, l’inferenza su  $\theta$  può essere svolta, mediante la regola di Bayes, costruendo la distribuzione a posteriori  $p(\theta | y)$ . Una volta ottenuta la distribuzione a posteriori possiamo riassumerla, ad esempio, riportando l’intervallo centrale al 95% della distribuzione di densità, ovvero

$$\Pr \left[ 0.025 \leq \theta \leq 0.975 \mid Y = y \right].$$

Se vogliamo sapere, per esempio, se la probabilità di  $y_i = 1$  sia maggiore di 0.5, possiamo calcolare la probabilità dell’evento

$$\Pr \left[ \theta > \frac{1}{2} \mid Y = y \right].$$

### Cos’è un parametro del modello?

Il parametro di un modello è un valore che influenza la credibilità dei dati. Ad esempio, il singolo parametro  $\theta$  del modello binomiale determina la forma della funzione di verosimiglianza binomiale. Ricordiamo che, per il modello binomiale, la funzione di verosimiglianza è:

$$p(y | \theta, n) = \text{Bin}(y, n, \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}.$$



### La distribuzione a priori sui parametri

Quando adottiamo un approccio bayesiano, i parametri non più delle costanti incognite ma delle variabili casuali governate da una propria legge di distribuzione delle probabilità (probabilità a priori). La distribuzione a priori sui valori dei parametri  $p(\theta)$  è parte integrante del modello statistico. Ciò implica che due modelli bayesiani possono condividere la stessa funzione di verosimiglianza, ma tuttavia devono essere considerati come modelli diversi se specificano diverse distribuzioni a priori. Ciò significa che, quando diciamo “Modello binomiale”, intendiamo in realtà un’intera classe di modelli, ovvero tutti i possibili modelli che hanno la stessa verosimiglianza ma diverse distribuzioni a priori su  $\theta$ .

Nell’analisi dei dati bayesiana, la distribuzione a priori  $p(\theta)$  codifica le credenze del ricercatore a proposito dei valori dei parametri, prima di avere osservato i dati. Idealmente, le credenze a priori che supportano la specificazione di una distribuzione a priori dovrebbero essere supportate da una qualche motivazione, come ad esempio i risultati di ricerche precedenti, o altre motivazioni giustificabili.

- Quando una nuova osservazione (p. es., vedo un cigno bianco) corrisponde alle mie convinzioni precedenti (p. es., la maggior parte dei cigni sono bianchi) sto rafforzando le mie convinzioni precedenti: più nuove osservazioni (p. es., più cigni bianchi vedo), più forti diventano le mie credenze precedenti.
- Tuttavia, quando una nuova osservazione (p. es., vedo un cigno nero) non corrisponde alle mie convinzioni precedenti, ciò contribuisce a diminuire la certezza delle mie convinzioni precedenti: più nuove osservazioni non corrispondenti raccolgo (p. es., più cigni neri vedo), più deboli diventano le mie convinzioni precedenti. Fondamentalmente, più forti sono le mie credenze precedenti, più osservazioni non corrispondenti (ad esempio, cigni neri) devo raccogliere per cambiarle.

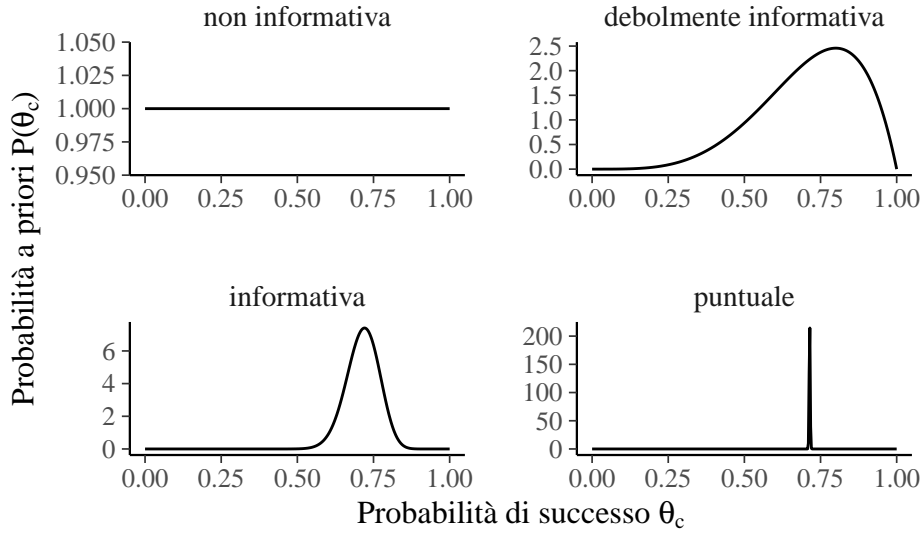
Pertanto, da una prospettiva bayesiana, l’incertezza intorno ai parametri di un modello *dopo* aver visto i dati (ovvero le distribuzioni a posteriori) deve includere anche le credenze precedenti. Se questo modo di ragionare sembra molto intuitivo, non è una coincidenza: vi sono infatti diverse teorie psicologiche che prendono l’aggiornamento bayesiano come modello di funzionamento di diversi processi cognitivi.

Oltre alla motivazione che giustifica una distribuzione a priori, possiamo distinguere tra diverse distribuzioni a priori in base a quanto fortemente impegnano il ricercatore a ritenere come plausibile un particolare intervallo di valori dei parametri. Il caso più estremo è quello che rivela una totale assenza di conoscenze a priori, il che conduce alle *distribuzioni a priori non informative*, ovvero quelle che assegnano lo stesso livello di credibilità a tutti i valori dei parametri. Le distribuzioni a priori informative, d’altra parte, possono essere *debolmente informative* o *fortemente informative*, a seconda della forza della credenza che esprimono. Il caso più estremo di credenza a priori è quello che riassume il punto di vista del ricercatore nei termini di un *unico valore* del parametro, il che assegna tutta la probabilità (massa o densità) su di un singolo valore di un parametro.

Poiché questa non è più una distribuzione di probabilità, sebbene ne soddisfi la definizione, in questo caso si parla di una *distribuzione a priori degenerata*.

La figura seguente mostra esempi di distribuzioni a priori non informative, debolmente o fortemente informative, così come una distribuzione a priori espressa nei termini di un valore puntuale per il modello Binomiale. Le distribuzioni a priori illustrate di seguito sono le seguenti:

- *non informativa* :  $\theta_c \sim \text{Beta}(1, 1)$ ;
- *debolmente informativa* :  $\theta_c \sim \text{Beta}(5, 2)$ ;
- *fortemente informativa* :  $\theta_c \sim \text{Beta}(50, 20)$ ;
- *valore puntuale* :  $\theta_c \sim \text{Beta}(\alpha, \beta)$  con  $\alpha, \beta \rightarrow \infty$  e  $\frac{\alpha}{\beta} = \frac{5}{2}$ .



**Figura 2.1:** Esempi di distribuzioni a priori per il parametro  $\theta_c$  nel Modello Binomiale.

### Scelta della distribuzione a priori

La selezione delle distribuzioni a priori è stata spesso vista come una delle scelte più importanti che un ricercatore fa quando implementa un modello bayesiano in quanto può avere un impatto sostanziale sui risultati finali. La soggettività delle distribuzioni a priori è evidenziata dai critici come un potenziale svantaggio dei metodi bayesiani. A questa critica, Schoot et al. (2021) rispondono dicendo che, al di là della scelta delle distribuzioni a priori, ci sono molti elementi del processo di inferenza statistica che sono soggettivi, ovvero la scelta del modello statistico e le ipotesi sulla distribuzione degli errori. In secondo luogo, Schoot et al. (ibid.) notano come le distribuzioni a priori svolgono due importanti ruoli statistici: quello della “regolarizzazione della stima”, ovvero, il processo che porta ad indebolire l’influenza indebita di osservazioni estreme, e quello del miglioramento dell’efficienza della stima, ovvero, la facilitazione dei processi di calcolo numerico di stima della distribuzione a posteriori. L’effetto della distribuzione a priori sulla distribuzione a posteriori verrà discusso nel Capitolo ??.

## 2.3 Verosimiglianza marginale

Al denominatore della regola di Bayes abbiamo la verosimiglianza marginale  $p(y)$ . Tale denominatore è espresso nei termini di un integrale che, tranne in pochi casi particolari, non ha una soluzione analitica.

## 2.4 La distribuzione a posteriori

Ci sono due metodi principali per calcolare la distribuzione a posteriori  $p(\theta | y)$ :

- una precisa derivazione matematica formulata nei termini della distribuzione a priori coniugata alla distribuzione a posteriori (si veda il Capitolo ??); tale procedura però ha un'applicabilità molto limitata;
- un metodo approssimato, molto facile da utilizzare in pratica, che dipende da metodi Monte Carlo basati su Catena di Markov (MCMC).

Una volta calcolata la distribuzione a posteriori dobbiamo riassumerla in qualche modo. Questo problema verrà discusso nel Capitolo ??.

## Considerazioni conclusive

In base all'approccio bayesiano, invece di dire che il parametro di interesse di un modello statistico ha un valore vero ma sconosciuto, diciamo che, prima di eseguire l'esperimento, è possibile assegnare una distribuzione di probabilità, che chiamano stato di credenza, a quello che è il vero valore del parametro. Questa distribuzione a priori può essere nota (per esempio, sappiamo che la distribuzione dei punteggi del QI è normale con media 100 e deviazione standard 15) o può essere del tutto arbitraria. L'inferenza bayesiana procede poi nel modo seguente: si raccolgono alcuni dati e si calcola la probabilità dei possibili valori del parametro alla luce dei dati osservati e delle credenze a priori. Questa nuova distribuzione di probabilità è chiamata “distribuzione a posteriori” e riassume l'incertezza dell'inferenza.



## Bibliografia

Schoot, R. van de, S. Depaoli, R. King, B. Kramer, K. Märtens, M. G. Tadesse, M. Vannucci, A. Gelman, D. Veen, J. Willemsen e C. Yau (2021). «Bayesian statistics and modelling». In: *Nature Reviews Methods Primer* 1.1, pp. 1–26 (cit. a p. 24).



## Elenco delle figure

1.1	Rappresentazione schematica del processo scientifico (figura adattata dalla Fig. 1.1 di P. Gregory, Bayesian Logical Data Analysis for the Physical Sciences, Cambridge, 2005). . . . .	3
1.2	Stima della probabilità di successo in funzione del numero di lanci di una moneta. . . . .	10
1.3	Stima della probabilità di successo in funzione del numero di lanci di una moneta – scala logaritmica. . . . .	11
1.4	Grafico di $M = 100\,000$ simulazioni della funzione di massa di probabilità di una variabile casuale definita come il numero di teste in quattro lanci di una moneta equilibrata. . . . .	15
2.1	Esempi di distribuzioni a priori per il parametro $\theta_c$ nel Modello Binomiale. . . . .	24

**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.