

Corrado Caudek

Data Science per psicologi



Psicometria – AA 2021/2022





Indice

| | |
|---|------------|
| Elenco delle figure | vii |
| Elenco delle tabelle | ix |
| Prefazione | xi |
| I Nozioni preliminari | 1 |
| 1 Concetti chiave | 3 |
| 1.1 Popolazioni e campioni | 3 |
| 1.2 Variabili e costanti | 4 |
| 1.2.1 Variabili casuali | 5 |
| 1.2.2 Variabili indipendenti e variabili dipendenti . . . | 5 |
| 1.2.3 La matrice dei dati | 6 |
| 1.3 Parametri e modelli | 7 |
| 1.4 Effetto | 8 |
| 1.5 Stima e inferenza | 9 |
| 1.6 Metodi e procedure della psicologia | 9 |
| 2 Il calcolo delle probabilità | 3 |
| 2.1 La probabilità come la logica della scienza | 3 |
| 2.2 Che cos'è la probabilità? | 5 |
| 2.3 Variabili casuali e probabilità di un evento | 7 |
| 2.3.1 Variabili casuali | 7 |
| 2.3.2 Eventi e probabilità | 8 |
| 2.4 Spazio campionario e risultati possibili | 9 |
| 2.5 Usare la simulazione per stimare le probabilità | 9 |
| 2.6 La legge dei grandi numeri | 12 |
| 2.7 Variabili casuali multiple | 15 |
| 2.8 Funzione di massa di probabilità | 17 |
| 3 Inferenza bayesiana | 3 |

| | | |
|-----------------------------|--|----------|
| 3.1 | Modellizzazione bayesiana | 3 |
| 3.2 | Inferenza bayesiana come un problema inverso | 4 |
| 3.2.1 | Notazione | 5 |
| 3.2.2 | Funzioni di probabilità | 5 |
| 3.3 | La regola di Bayes | 6 |
| 3.3.1 | Un esempio di aggiornamento bayesiano | 8 |
| 3.4 | Modello probabilistico | 9 |
| 3.5 | Distribuzioni a priori | 10 |
| 3.5.1 | Tipologie di distribuzioni a priori | 11 |
| 3.5.2 | Selezione della distribuzione a priori | 12 |
| 3.5.3 | La distribuzione a priori per i dati di Zetsche et al. (2019) | 13 |
| 3.6 | Verosimiglianza | 14 |
| 3.6.1 | La stima di massima verosimiglianza | 15 |
| 3.6.2 | La log-verosimiglianza | 16 |
| 3.7 | La verosimiglianza marginale | 19 |
| 3.8 | Distribuzione a posteriori | 20 |
| 3.9 | Distribuzione predittiva a priori | 20 |
| 3.10 | Distribuzione predittiva a posteriori | 21 |
| Appendice | | 1 |
| A Simbologia di base | | 1 |

Elenco delle figure

| | | |
|-----|--|----|
| 2.1 | Rappresentazione schematica del processo scientifico (figura adattata dalla Fig. 1.1 di P. Gregory, Bayesian Logical Data Analysis for the Physical Sciences, Cambridge, 2005). | 4 |
| 2.2 | Stima della probabilità di successo in funzione del numero di lanci di una moneta. | 13 |
| 2.3 | Stima della probabilità di successo in funzione del numero di lanci di una moneta – scala logaritmica. | 14 |
| 2.4 | 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. | 19 |
| 3.1 | Esempi di distribuzioni a priori per il parametro θ_c nel Modello Binomiale. | 12 |
| 3.2 | Funzione di verosimiglianza nel caso di 23 successi in 30 prove. | 18 |



Elenco delle tabelle



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

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 pon-

te, 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

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

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.

- 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 possia-

mo 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.¹ È utile farsi una nota mentalmente delle lacune nella vostra comprensione e tornare su di esse in seguito per carcarle 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

¹Ricordatevi inoltre che gli individui tendono a sottostimare la propria capacità di apprendere ([Horn and Loewenstein, 2021](#)).

un software, ricordatevi: “Google is your friend”!

Corrado Caudek
Marzo 2022

Parte I

Nozioni preliminari



1

Concetti chiave

La *data science* si pone all'intersezione tra statistica e informatica. La statistica è un insieme di metodi utilizzati per estrarre informazioni dai dati; l'informatica implementa tali procedure in un software. In questo Capitolo vengono introdotti i concetti fondamentali.

1.1 Popolazioni e campioni

Popolazione. L'analisi dei dati inizia con l'individuazione delle unità portatrici di informazioni circa il fenomeno di interesse. Si dice popolazione (o universo) l'insieme Ω delle entità capaci di fornire informazioni sul fenomeno oggetto dell'indagine statistica. Possiamo scrivere $\Omega = \{\omega_i\}_{i=1,\dots,n} = \{\omega_1, \omega_2, \dots, \omega_n\}$, oppure $\Omega = \{\omega_1, \omega_2, \dots\}$ nel caso di popolazioni finite o infinite, rispettivamente.

L'obiettivo principale della ricerca psicologica è conoscere gli esiti psicologici e i loro fattori trainanti nella popolazione. Questo è l'obiettivo delle sperimentazioni psicologiche e della maggior parte degli studi osservazionali in psicologia. È quindi necessario essere molto chiari sulla popolazione a cui si applicano i risultati della ricerca. La popolazione può essere ben definita, ad esempio, tutte le persone che si trovavano nella città di Hiroshima al momento dei bombardamenti atomici e sono sopravvissute al primo anno, o può essere ipotetica, ad esempio, tutte le persone depresse che hanno subito o saranno sottoporsi ad un intervento di psicoterapia. Il ricercatore deve sempre essere in grado di determinare se un soggetto appartiene alla popolazione oggetto di interesse.

Una *sottopopolazione* è una popolazione in sé e per sé che soddisfa proprietà ben definite. Negli esempi precedenti, potremmo essere interessati alla sottopopolazione di uomini di età inferiore ai 20 anni o di pazienti depressi sottoposti ad uno specifico intervento psicologico. Molte questio-

ni scientifiche riguardano le differenze tra sottopopolazioni; ad esempio, confrontando i gruppi con o senza psicoterapia per determinare se il trattamento è vantaggioso. I modelli di regressione, introdotti nel Capitolo ?? riguardano le sottopopolazioni, in quanto stimano il risultato medio per diversi gruppi (sottopopolazioni) definiti dalle covariate.

Campione. Gli elementi ω_i dell'insieme Ω sono detti *unità statistiche*. Un sottoinsieme della popolazione, ovvero un insieme di elementi ω_i , viene chiamato *campione*. Ciascuna unità statistica ω_i (abbreviata con u.s.) è portatrice dell'informazione che verrà rilevata mediante un'operazione di misurazione.

Un campione è dunque un sottoinsieme della popolazione utilizzato per conoscere tale popolazione. A differenza di una sottopopolazione definita in base a chiari criteri, un campione viene generalmente selezionato tramite un procedura casuale. Il *campionamento casuale* consente allo scienziato di trarre conclusioni sulla popolazione e, soprattutto, di quantificare l'incertezza sui risultati. I campioni di un sondaggio sono esempi di campioni casuali, ma molti studi osservazionali non sono campionati casualmente. Possono essere *campioni di convenienza*, come coorti di studenti in un unico istituto, che consistono di tutti gli studenti sottoposti ad un certo intervento psicologico in quell'istituto. Indipendentemente da come vengono ottenuti i campioni, il loro uso al fine di conoscere una popolazione target significa che i problemi di rappresentatività sono inevitabili e devono essere affrontati.

1.2 Variabili e costanti

Definiamo *variabile statistica* la proprietà (o grandezza) che è oggetto di studio nell'analisi dei dati. Una variabile è una proprietà di un fenomeno che può essere espressa in più valori sia numerici sia categoriali. Il termine “variabile” si contrappone al termine “costante” che descrive una proprietà invariante di tutte le unità statistiche.

Si dice *modalità* ciascuna delle varianti con cui una variabile statistica può presentarsi. Definiamo *insieme delle modalità* di una variabile statistica l'insieme M di tutte le possibili espressioni con cui la variabile può manifestarsi. Le modalità osservate e facenti parte del campione si chiamano *dati* (si veda la Tabella 1.1).

Esempio 1.1. Supponiamo che il fenomeno studiato sia l'intelligenza. In uno studio, la popolazione potrebbe corrispondere all'insieme di tutti gli italiani adulti. La variabile considerata potrebbe essere il punteggio del test standardizzato WAIS-IV. Le modalità di tale variabile potrebbero essere 112, 92, 121, Tale variabile è di tipo quantitativo discreto.

Esempio 1.2. Supponiamo che il fenomeno studiato sia il compito Stroop. La popolazione potrebbe corrispondere all'insieme dei bambini dai 6 agli 8 anni. La variabile considerata potrebbe essere il reciproco dei tempi di reazione in secondi. Le modalità di tale variabile potrebbero essere $1/2.35$, $1/1.49$, $1/2.93$, La variabile è di tipo quantitativo continuo.

Esempio 1.3. Supponiamo che il fenomeno studiato sia il disturbo di personalità. La popolazione potrebbe corrispondere all'insieme dei detenuti nelle carceri italiane. La variabile considerata potrebbe essere l'assessment del disturbo di personalità tramite interviste cliniche strutturate. Le modalità di tale variabile potrebbero essere i Cluster A, Cluster B, Cluster C descritti dal DSM-V. Tale variabile è di tipo qualitativo.

1.2.1 Variabili casuali

Il termine *variabile* usato nella statistica è equivalente al termine *variabile casuale* usato nella teoria delle probabilità. Lo studio dei risultati degli interventi psicologici è lo studio delle variabili casuali che misurano questi risultati. Una variabile casuale cattura una caratteristica specifica degli individui nella popolazione e i suoi valori variano tipicamente tra gli individui. Ogni variabile casuale può assumere in teoria una gamma di valori sebbene, in pratica, osserviamo un valore specifico per ogni individuo. Quando faremo riferimento alle variabili casuali considerate in termini generali useremo lettere maiuscole come X e Y ; quando faremo riferimento ai valori che una variabile casuale assume in determinate circostanze useremo lettere minuscole come x e y .

1.2.2 Variabili indipendenti e variabili dipendenti

Un primo compito fondamentale in qualsiasi analisi dei dati è l'identificazione delle variabili dipendenti (Y) e delle variabili indipendenti (X). Le variabili dipendenti sono anche chiamate variabili di esito o di risposta e le variabili indipendenti sono anche chiamate predittori o covariate. Ad esempio, nell'analisi di regressione, che esamineremo in seguito, la

domanda centrale è quella di capire come Y cambia al variare di X . Più precisamente, la domanda che viene posta è: se il valore della variabile indipendente X cambia, qual è la conseguenza per la variabile dipendente Y ? In parole povere, le variabili indipendenti e dipendenti sono analoghe a “cause” ed “effetti”, laddove le virgolette usate qui sottolineano che questa è solo un’analogia e che la determinazione delle cause può avvenire soltanto mediante l’utilizzo di un appropriato disegno sperimentale e di un’adeguata analisi statistica.

Se una variabile è una variabile indipendente o dipendente dipende dalla domanda di ricerca. A volte può essere difficile decidere quale variabile è dipendente e quale è indipendente, in particolare quando siamo specificamente interessati ai rapporti di causa/effetto. Ad esempio, supponiamo di indagare l’associazione tra esercizio fisico e insonnia. Vi sono evidenze che l’esercizio fisico (fatto al momento giusto della giornata) può ridurre l’insonnia. Ma l’insonnia può anche ridurre la capacità di una persona di fare esercizio fisico. In questo caso, dunque, non è facile capire quale sia la causa e quale l’effetto, quale sia la variabile dipendente e quale la variabile indipendente. La possibilità di identificare il ruolo delle variabili (dipendente/indipendente) dipende dalla nostra comprensione del fenomeno in esame.

Esempio 1.4. Uno psicologo convoca 120 studenti universitari per un test di memoria. Prima di iniziare l’esperimento, a metà dei soggetti viene detto che si tratta di un compito particolarmente difficile; agli altri soggetti non viene data alcuna indicazione. Lo psicologo misura il punteggio nella prova di memoria di ciascun soggetto.

In questo esperimento, la variabile indipendente è l’informazione sulla difficoltà della prova. La variabile indipendente viene manipolata dallo sperimentatore assegnando i soggetti (di solito in maniera causale) o alla condizione (modalità) “informazione assegnata” o “informazione non data”. La variabile dipendente è ciò che viene misurato nell’esperimento, ovvero il punteggio nella prova di memoria di ciascun soggetto.

1.2.3 La matrice dei dati

Le realizzazioni delle variabili esaminate in una rilevazione statistica vengono organizzate in una *matrice dei dati*. Le colonne della matrice dei dati contengono gli insiemi dei dati individuali di ciascuna variabile statistica considerata. Ogni riga della matrice contiene tutte le informazioni

relative alla stessa unità statistica. Una generica matrice dei dati ha l'aspetto seguente:

$$D_{m,n} = \begin{pmatrix} \omega_1 & a_1 & b_1 & \cdots & x_1 & y_1 \\ \omega_2 & a_2 & b_2 & \cdots & x_2 & y_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \omega_n & a_n & b_n & \cdots & x_n & y_n \end{pmatrix}$$

dove, nel caso presente, la prima colonna contiene il nome delle unità statistiche, la seconda e la terza colonna si riferiscono a due mutabili statistiche (variabili categoriali; A e B) e ne presentano le modalità osservate nel campione mentre le ultime due colonne si riferiscono a due variabili statistiche (X e Y) e ne presentano le modalità osservate nel campione. Generalmente, tra le unità statistiche ω_i non esiste un ordine progressivo; l'indice attribuito alle unità statistiche nella matrice dei dati si riferisce semplicemente alla riga che esse occupano.

1.3 Parametri e modelli

Ogni variabile casuale ha una *distribuzione* che descrive la probabilità che la variabile assuma qualsiasi valore in un dato intervallo.¹ Senza ulteriori specificazioni, una distribuzione può fare riferimento a un'intera famiglia di distribuzioni. I parametri, tipicamente indicati con lettere greche come μ e α , ci permettono di specificare di quale membro della famiglia stiamo parlando. Quindi, si può parlare di una variabile casuale con una distribuzione Normale, ma se viene specificata la media $\mu = 100$ e la varianza $\sigma^2 = 15$, viene individuata una specifica distribuzione Normale – nell'esempio, la distribuzione del quoziente di intelligenza.

I metodi statistici parametrici specificano la famiglia delle distribuzioni e quindi utilizzano i dati per individuare, stimando i parametri, una specifica distribuzione all'interno della famiglia di distribuzioni ipotizzata. Se f è la PDF di una variabile casuale Y , l'interesse può concentrarsi

¹In questo e nei successivi Paragrafi di questo Capitolo introduco gli obiettivi della *data science* utilizzando una serie di concetti che saranno chiariti solo in seguito. Questa breve panoramica risulterà dunque solo in parte comprensibile ad una prima lettura e serve solo per definire la *big picture* dei temi trattati in questo insegnamento. Il significato dei termini qui utilizzati sarà chiarito nei Capitoli successivi.

sulla sua media e varianza. Nell'analisi di regressione, ad esempio, cerchiamo di spiegare come i parametri di f dipendano dalle covariate X . Nella regressione lineare classica, assumiamo che Y abbia una distribuzione normale con media $\mu = \mathbb{E}(Y)$, e stimiamo come $\mathbb{E}(Y)$ dipenda da X . Poiché molti esiti psicologici non seguono una distribuzione normale, verranno introdotte distribuzioni più appropriate per questi risultati. I metodi non parametrici, invece, non specificano una famiglia di distribuzioni per f . In queste dispense faremo riferimento a metodi non parametrici quando discuteremo della statistica descrittiva.

Il termine *modello* è onnipresente in statistica e nella *data science*. Il modello statistico include le ipotesi e le specifiche matematiche relative alla distribuzione della variabile casuale di interesse. Il modello dipende dai dati e dalla domanda di ricerca, ma raramente è unico; nella maggior parte dei casi, esiste più di un modello che potrebbe ragionevolmente usato per affrontare la stessa domanda di ricerca e avendo a disposizione i dati osservati. Nella previsione delle aspettative future dei pazienti depressi che discuteremo in seguito (Zetsche et al., 2019), ad esempio, la specifica del modello include l'insieme delle covariate candidate, l'espressione matematica che collega i predittori con le aspettative future e qualsiasi ipotesi sulla distribuzione della variabile dipendente. La domanda di cosa costituisca un buon modello è una domanda su cui torneremo ripetutamente in questo insegnamento.

1.4 Effetto

L'*effetto* è una qualche misura dei dati. Dipende dal tipo di dati e dal tipo di test statistico che si vuole utilizzare. Ad esempio, se viene lanciata una moneta 100 volte e esce testa 66 volte, l'effetto sarà 66/100. Diventa poi possibile confrontare l'effetto ottenuto con l'effetto nullo che ci si aspetterebbe da una moneta bilanciata (50/100), o con qualsiasi altro effetto che può essere scelto. La *dimensione dell'effetto* si riferisce alla differenza tra l'effetto misurato nei dati e l'effetto nullo (di solito un valore che ci si aspetta di ottenere in base al caso soltanto).

1.5 Stima e inferenza

La stima è il processo mediante il quale il campione viene utilizzato per conoscere le proprietà di interesse della popolazione. La media campionaria è una stima naturale della media della popolazione e la mediana campionaria è una stima naturale della mediana della popolazione. Quando parliamo di stimare una proprietà della popolazione (a volte indicata come parametro della popolazione) o di stimare la distribuzione di una variabile casuale, stiamo parlando dell'utilizzo dei dati osservati per conoscere le proprietà di interesse della popolazione. L'inferenza statistica è il processo mediante il quale le stime campionarie vengono utilizzate per rispondere a domande di ricerca e per valutare specifiche ipotesi relative alla popolazione. Discuteremo le procedure bayesiane dell'inferenza nell'ultima parte di queste dispense.

1.6 Metodi e procedure della psicologia

Un modello psicologico di un qualche aspetto del comportamento umano o della mente ha le seguenti proprietà:

1. descrive le caratteristiche del comportamento in questione,
2. formula predizioni sulle caratteristiche future del comportamento,
3. è sostenuto da evidenze empiriche,
4. deve essere falsificabile (ovvero, in linea di principio, deve potere fare delle predizioni su aspetti del fenomeno considerato che non sono ancora noti e che, se venissero indagati, potrebbero portare a rigettare il modello, se si dimostrassero incompatibili con esso).

L'analisi dei dati valuta un modello psicologico utilizzando strumenti statistici.

Questa dispensa è strutturata in maniera tale da rispecchiare la suddivisione tra i temi della misurazione, dell'analisi descrittiva e dell'inferenza. Nel prossimo Capitolo sarà affrontato il tema della misurazione e, nell'ul-

tima parte della dispensa verrà discusso l'argomento più difficile, quello dell'inferenza. Prima di affrontare il secondo tema, l'analisi descrittiva dei dati, sarà necessario introdurre il linguaggio di programmazione statistica R (un'introduzione a R è fornita in Appendice). Inoltre, prima di potere discutere l'inferenza, dovranno essere introdotti i concetti di base della teoria delle probabilità, in quanto l'inferenza non è che l'applicazione della teoria delle probabilità all'analisi dei dati.

Nozioni di base



2

Il calcolo delle probabilità

Una possibile definizione della teoria delle probabilità è la seguente: la teoria delle probabilità ci fornisce gli strumenti per prendere decisioni razionali in condizioni di incertezza, ovvero per formulare le migliori congetture possibili.

2.1 La probabilità come la logica della scienza

La figura 2.1 fornisce una rappresentazione schematica del processo dell'indagine scientifica. Possiamo pensare al progresso scientifico come alla ripetizione di questo ciclo, laddove i fenomeni naturali (e, ovviamente psicologici) vengono esplorati e i ricercatori imparano sempre di più sul loro funzionamento. Le caselle della figura descrivono le varie fasi del processo di indagine scientifica, mentre lungo le frecce sono riportati i compiti che conducono i ricercatori da una fase alla successiva.

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, ai risultati dei loro esperimenti, e formulano ipotesi o teorie che possono essere valutate mediante esperimenti empirici. Questo passaggio richiede innovazione e creatività.
- *L'inferenza deduttiva* procede in maniera deterministica dai fatti alle conclusioni. Ad esempio, se dico che tutti gli uomini sono mortali e che Socrate è un uomo, allora posso concludere deduttivamente che Socrate è mortale. Quando i ricercatori progettano gli esperimenti in base alle teorie, usano la logica deduttiva per dire: "Se A è vero, allora

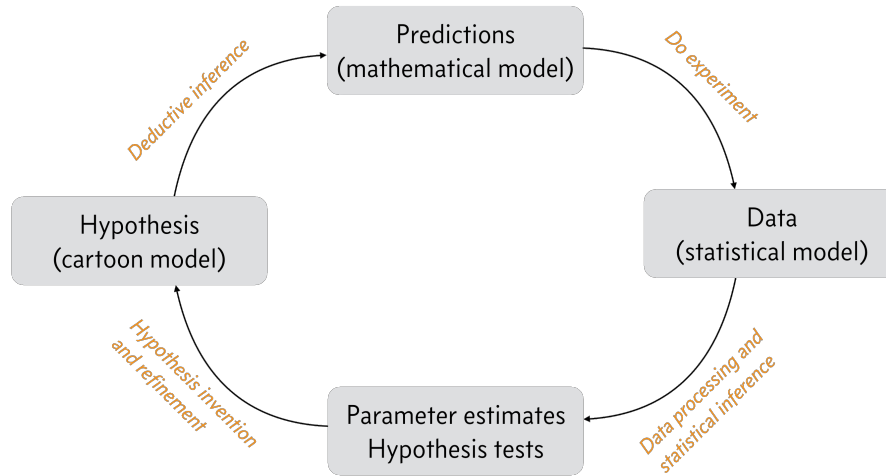


Figura 2.1: Rappresentazione schematica del processo scientifico (figura adattata dalla Fig. 1.1 di P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge, 2005).

B deve essere vero”, dove A è l’ipotesi teorica e B è l’osservazione sperimentale.

- *Esecuzione degli esperimenti.* Questa fase richiede molte risorse (tempo e denaro). Richiede anche innovazione e creatività. Nello specifico, i ricercatori devono pensare attentamente a come costruire l’esperimento necessario per verificare la teoria di interesse. Quale risultato dell’esperimento si ottengono i dati.
- *L’inferenza induttiva* 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 è 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 sia A dopo aver osservato B . Per svolgere l'inferenza statistica è dunque necessario quantificare tale plausibilità. Lo strumento che ci consente di fare questo è la teoria delle probabilità.

L'inferenza statistica è l'aspetto del processo dell'indagine scientifica che costituisce il tema 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 delle probabilità, laddove la teoria delle 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 ha dichiarato che “la probabilità è la logica della scienza”. Per potere trattare i temi di base dell'inferenza statistica è dunque necessario esaminare preliminarmente alcune nozioni della teoria delle probabilità.

2.2 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 di-

sponibili. 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 casuali¹.
- Nell'interpretazione bayesiana della probabilità $P(A)$ rappresenta 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. Possiamo citare De Finetti, ad esempio, il quale ha formulato la seguente definizione “soggettiva” di probabilità la quale risulta applicabile anche ad esperimenti casuali i cui eventi elementari non siano ritenuti ugualmente possibili e che non siano necessariamente ripetibili più volte sotto le stesse condizioni:

Definizione 2.1. La probabilità di un evento E è la quota $p(E)$ che un individuo reputa di dover pagare ad un banco per ricevere “1” ovvero “0” verificandosi o non verificandosi E . Le valutazioni di probabilità degli eventi devono rispondere ai principi di equità e coerenza.

I principi di equità e coerenza sono definiti come segue.

Definizione 2.2. Una scommessa risponde ai principi di *equità* se il ruolo di banco e giocatore sono scambiabili in ogni momento del gioco e sempre alle stesse condizioni; *coerenza* se non vi sono combinazioni di scommesse che consentano (sia al banco che al giocatore) di realizzare perdite o vincite certe.

¹Viene 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.

Secondo [de Finetti \(1931\)](#), “nessuna scienza ci permetterà di dire: il tale fatto accadrà, andrà così e così, perché ciò è conseguenza di tale legge, e tale legge è una verità assoluta, ma tanto meno ci condurrà a concludere scetticamente: la verità assoluta non esiste, e quindi tale fatto può accadere e può non accadere, può andare così e può andare in tutt'altro modo, nulla io ne so. Quel che si potrà dire è questo: io prevedo che il tale fatto avverrà, e avverrà nel tal modo, perché l'esperienza del passato e l'elaborazione scientifica cui il pensiero dell'uomo l'ha sottoposta mi fanno sembrare ragionevole questa previsione.”

In altri termini, de Finetti ritiene che la probabilità debba essere concepita non come una proprietà “oggettiva” dei fenomeni (“la probabilità di un fenomeno ha un valore determinato che dobbiamo solo scoprire”), ma bensì come il “grado di fiducia – in inglese *degree of belief* – di un dato soggetto, in un dato istante e con un dato insieme d'informazioni, riguardo al verificarsi di un evento”. Per denotare sia la probabilità (soggettiva) di un evento sia il concetto di *valore atteso* (che descriveremo in seguito), [de Finetti \(1970\)](#) utilizza il termine “previsione” (e lo stesso simbolo P): “la previsione [...] consiste nel considerare ponderatamente tutte le alternative possibili per ripartire fra di esse nel modo che parrà più appropriato le proprie aspettative, le proprie sensazioni di probabilità.”

2.3 Variabili casuali e probabilità di un evento

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

2.3.1 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 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 nu-

merici. 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.

2.3.2 Eventi e probabilità

Nella teoria delle probabilità il risultato “testa” nel lancio di una moneta è chiamato *evento*.² 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 $Y = 0$ 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 $Pr(A \text{ e } B) = 0$.

²Per un ripasso delle nozioni di base della teoria degli insiemi, si veda l’Appendice ??.

2.4 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 delle probabilità e l'inferenza statistica.

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).

2.5 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] = 0.5
```

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] = 0.3
#> estimated Pr[Y = 1] = 0.7
#> estimated Pr[Y = 1] = 0.5
#> estimated Pr[Y = 1] = 0.5
#> estimated Pr[Y = 1] = 0.6
#> estimated Pr[Y = 1] = 0.5
#> estimated Pr[Y = 1] = 0.8
#> estimated Pr[Y = 1] = 0.4
#> estimated Pr[Y = 1] = 0.5
```

Dato che la moneta è equilibrata, la stima delle probabilità dell'evento $Pr[Y = 1]$ è simile 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] = 0.44
#> estimated Pr[Y = 1] = 0.53
#> estimated Pr[Y = 1] = 0.43
#> estimated Pr[Y = 1] = 0.58
#> estimated Pr[Y = 1] = 0.5
#> estimated Pr[Y = 1] = 0.41
#> estimated Pr[Y = 1] = 0.51
#> estimated Pr[Y = 1] = 0.49
#> estimated Pr[Y = 1] = 0.5
#> estimated Pr[Y = 1] = 0.57
```

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] = 0.5029  
#> estimated Pr[Y = 1] = 0.4886  
#> estimated Pr[Y = 1] = 0.4956  
#> estimated Pr[Y = 1] = 0.49  
#> estimated Pr[Y = 1] = 0.5032  
#> estimated Pr[Y = 1] = 0.5051  
#> estimated Pr[Y = 1] = 0.4928  
#> estimated Pr[Y = 1] = 0.4968  
#> estimated Pr[Y = 1] = 0.4991  
#> estimated Pr[Y = 1] = 0.4976
```

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)

2.6 La legge dei grandi numeri

La visualizzazione mediante grafici contribuisce alla comprensione dei concetti della statistica e della teoria delle 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]"
  )
)
```

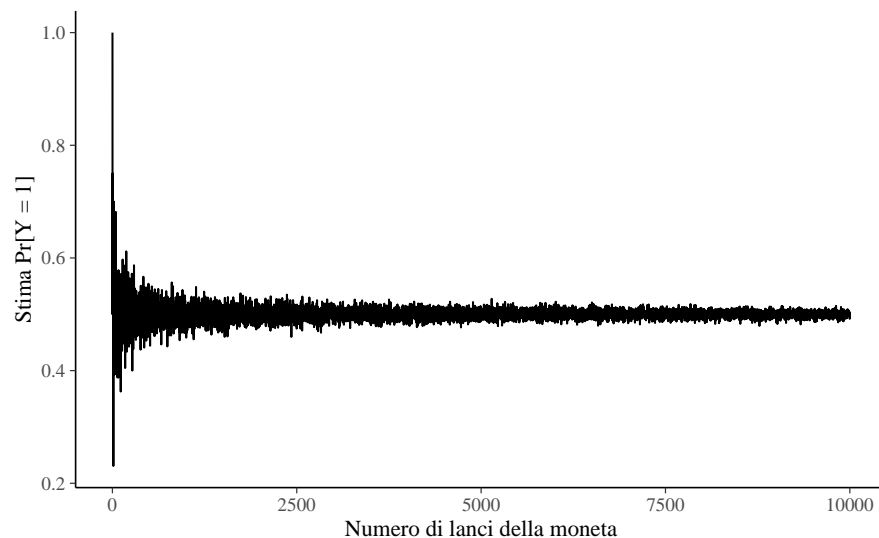


Figura 2.2: Stima della probabilità di successo in funzione del numero di lanci di una moneta.

Dato che il grafico 2.2 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]"
  )
)
```

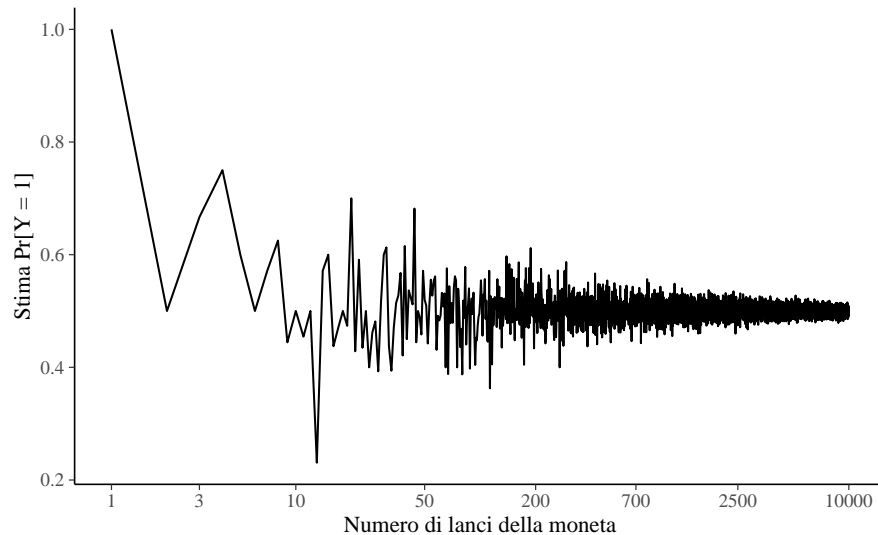


Figura 2.3: Stima della probabilità di successo in funzione del numero di lanci di una moneta – scala logaritmica.

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 2.3 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.

2.7 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 1
z <- sum(c(y1, y2, y3))
cat("z =", z, "\n")
#> z = 1
```

ovvero,

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

oppure, ancora più semplicemente:

```
y <- rbinom(3, 1, 0.5)
y
#> [1] 0 1 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
#> 0.1256 0.3750 0.3749 0.1245
```

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
#> 0.06213 0.25019 0.37400 0.25097 0.06271

```

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

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

2.8 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 |
| $p_Z(1)$ | $=$ | $4/16$ | HTTT, THTT, TTHT, TTTH |
| $p_Z(2)$ | $=$ | $6/16$ | HHTT, HTHT, HTTH, THHT, THTH, TTTH |
| $p_Z(3)$ | $=$ | $4/16$ | HHHT, HHTH, HTHH, THHH |
| $p_Z(4)$ | $=$ | $1/16$ | HHHH |

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 2.4.

```
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

```

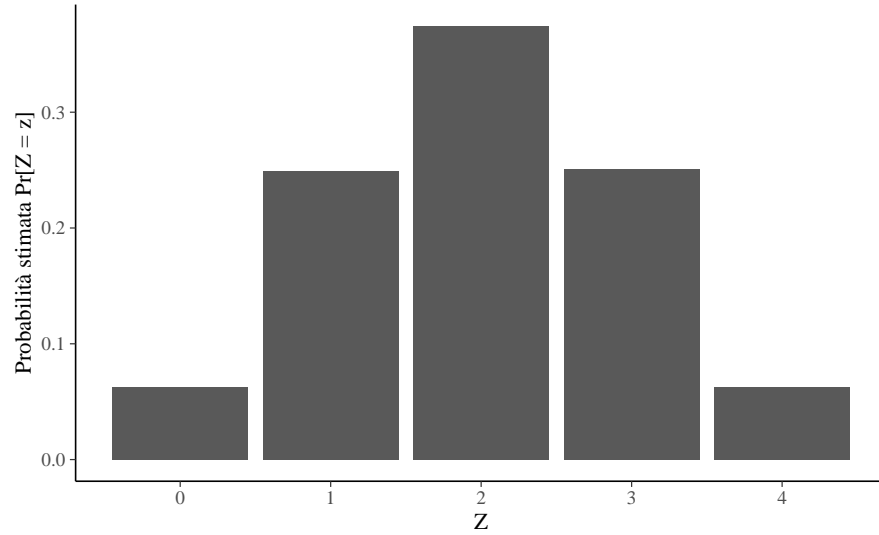


Figura 2.4: Grafico di $M = 100\,000$ simulazioni della funzione di massa di probabilità di una variabile casuale definita come il numero di teste in quattro lanci di una moneta equilibrata.

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 2.1. Nel caso dell'esempio discusso nella Sezione 2.8, la probabilità che la variabile casuale Z sia un numero dispari è $Pr(Z \text{ è un numero dispari}) = P_z(Z = 1) + P_z(Z = 3) = \frac{4}{16} + \frac{4}{16} = \frac{1}{2}$.

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



3

Inferenza bayesiana

La moderna statistica bayesiana viene per lo più eseguita utilizzando un linguaggio di programmazione probabilistico implementato su computer. Ciò ha cambiato radicalmente il modo in cui venivano eseguite le statistiche bayesiane anche fin pochi decenni fa. La complessità dei modelli che possiamo costruire è aumentata e la barriera delle competenze matematiche e computazionali che sono richieste è diminuita. Inoltre, il processo di modellazione iterativa è diventato, sotto molti aspetti, molto più facile da eseguire. Anche se formulare modelli statistici complessi è diventato più facile che mai, la statistica è un campo pieno di sottigliezze che non scompaiono magicamente utilizzando potenti metodi computazionali. Pertanto, avere una buona preparazione sugli aspetti teorici, specialmente quelli rilevanti nella pratica, è estremamente utile per applicare efficacemente i metodi statistici.

3.1 Modellizzazione bayesiana

Seguendo ([Martin et al., 2022](#)), possiamo descrivere il processo della modellazione bayesiana distinguendo 3 passaggi.

1. Dati alcuni dati e alcune ipotesi su come questi dati potrebbero essere stati generati, progettiamo un modello combinando e trasformando variabili casuali.
2. Usiamo il teorema di Bayes per condizionare i nostri modelli ai dati disponibili. Chiamiamo questo processo “inferenza” e come risultato otteniamo una distribuzione a posteriori. Ci auguriamo che i dati riducano l’incertezza per i possibili valori dei parametri, sebbene questo non sia garantito per nessun modello bayesiano.

3. Critichiamo il modello verificando se il modello abbia senso utilizzando criteri diversi, inclusi i dati e la nostra conoscenza del dominio. Poiché generalmente siamo incerti sui modelli stessi, a volte confrontiamo diversi modelli.

Questi 3 passaggi vengono eseguiti in modo iterativo e danno luogo a quello che si chiama un “flusso di lavoro bayesiano” (*bayesian workflow*).

Osservazione. Un modello è uno strumento concettuale che viene utilizzato per risolvere uno specifico problema. In quanto tale, è generalmente più conveniente parlare dell’adeguatezza del modello a un dato problema che di determinare la sua intrinseca correttezza. I modelli esistono esclusivamente come l’ausilio per il raggiungimento di un qualche ulteriore obiettivo. Il problema che i modelli bayesiani cercano di risolvere è quello dell’inferenza¹.

I modelli bayesiani, computazionali o meno, hanno due caratteristiche distintive:

- Le quantità incognite sono descritte utilizzando le distribuzioni di probabilità. Queste quantità incognite sono chiamate parametri.
- Il teorema di Bayes viene utilizzato per aggiornare i valori dei parametri condizionati ai dati. Possiamo anche concepire questo processo come una riallocazione delle probabilità.

3.2 Inferenza bayesiana come un problema inverso

In questo capitolo ci focalizzeremo sul passaggio 2 descritto sopra. Nello specifico, descrivemo in dettaglio il significato dei tre termini a destra del segno di uguale nella formula di Bayes: la distribuzione a priori e la funzione di verosimiglianza al numeratore, e la verosimiglianza marginale al denominatore.

¹In termini colloquiali, l’inferenza può essere descritta come la capacità di giungere a conclusioni basate su evidenze e ragionamenti. L’inferenza bayesiana è una particolare forma di inferenza statistica basata sulla combinazione di distribuzioni di probabilità che ha il fine di ottenere altre distribuzioni di probabilità. Nello specifico, la regola di Bayes ci fornisce un metodo per giungere alla quantificazione della plausibilità di una teoria alla luce dei dati osservati.

3.2.1 Notazione

Per fissare la notazione, nel seguito y rappresenterà i dati e θ rappresenterà i parametri incogniti di un modello statistico. Sia y che θ saranno concepiti come delle variabili casuali.² Con x verranno invece denotate le quantità note, come ad esempio i predittori del modello lineare. Per rappresentare in un modo conciso i modelli probabilistici 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 θ costituisce un campione casuale estratto dalla distribuzione $\text{Beta}(1, 1)$. Allo stesso modo, ad esempio, la verosimiglianza del modello binomiale può essere scritta come $y \sim \text{Bin}(n, \theta)$.

3.2.2 Funzioni di probabilità

Una caratteristica attraente della statistica bayesiana è che la nostra credenza “a posteriori” viene sempre descritta mediante una distribuzione. Questo fatto ci consente di fare affermazioni probabilistiche sui parametri, come ad esempio: “la probabilità che un parametro sia positivo è 0.35”; oppure, “il valore più probabile di θ è 12 e abbiamo probabilità del 50% che θ sia compreso tra 10 e 15”. Inoltre, possiamo pensare alla distribuzione a posteriori come alla logica conseguenza della combinazione di un modello con i dati; quindi, abbiamo la garanzia che le affermazioni probabilistiche associate alla distribuzione a posteriori siano matematicamente coerenti. Dobbiamo solo ricordare che tutte queste belle proprietà matematiche sono valide solo nel mondo platonico delle idee dove esistono oggetti matematici come sfere, distribuzioni gaussiane e catene di Markov. Quando passiamo dalla purezza della matematica al disordine della matematica applicata al mondo reale, dobbiamo sempre tenere a mente che i nostri risultati sono condizionati, non solo dai dati, ma anche dai modelli. Di conseguenza, dati errati e/o modelli errati conducono facilmente a conclusioni prive di senso, anche se matematica-

²Nell’approccio bayesiano 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 inoltre riferimento ad una distribuzione congiunta (di massa o di densità di probabilità) $f(y, \theta)$. Entrambi gli argomenti della funzione y e θ 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 θ è dovuta alla mancata conoscenza del suo valore numerico (*variabilità epistemica*).

mente coerenti. È dunque necessario conservare sempre una sana quota di scetticismo relativamente ai nostri dati, modelli e risultati (Martin et al., 2022).

Avendo detto questo, nell'aggiornamento bayesiano (dai dati ai parametri) vengono utilizzate le seguenti distribuzioni di probabilità (o di massa di probabilità):

- la *distribuzione a priori* $p(\theta)$ — la credenza iniziale (prima di avere osservato i dati $Y = y$) riguardo a θ ;
- la *funzione di verosimiglianza* $p(y \mid \theta)$ — quanto sono compatibili i dati osservati $Y = y$ con i diversi valori possibili di θ ?
- la *verosimiglianza marginale* $p(y)$ — costante di normalizzazione: qual è la probabilità complessiva di osservare i dati $Y = y$? In termini formali:

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

- la *distribuzione a posteriori* $p(\theta \mid y)$ — la nuova credenza relativa alla credibilità di ciascun valore θ dopo avere osservato i dati $Y = y$.

3.3 La regola di Bayes

Assumendo un modello statistico, la formula di Bayes consente di giungere alla distribuzione a posteriori $p(\theta \mid y)$ per il parametro di interesse θ , come indicato dalla seguente catena di equazioni³:

³In realtà, avremmo dovuto scrivere $p(\theta \mid y, \mathcal{M})$, in quanto non condizioniamo la stima di θ solo rispetto ai dati y ma anche ad un modello probabilistico \mathcal{M} che viene assunto quale meccanismo generatore dei dati. Per semplicità di notazione, omettiamo il riferimento a \mathcal{M} .

$$p(\theta | y) = \frac{p(\theta, y)}{p(y)} \quad [\text{def. prob. condizionata}] \quad (3.1)$$

$$= \frac{p(y | \theta) p(\theta)}{p(y)} \quad [\text{legge prob. composta}] \quad (3.2)$$

$$= \frac{p(y | \theta) p(\theta)}{\int_{\Theta} p(y, \theta) d\theta} \quad [\text{legge prob. totale}] \quad (3.3)$$

$$= \frac{p(y | \theta) p(\theta)}{\int_{\Theta} p(y | \theta) p(\theta) d\theta} \quad [\text{legge prob. composta}] \quad (3.4)$$

$$\propto p(y | \theta) p(\theta) \quad (3.5)$$

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 θ (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 distribuzione a posteriori $p(\theta | y)$ contiene tutta l’informazione riguardante il parametro θ e viene utilizzata per produrre indicatori sintetici, per la determinazione di stime puntuali o intervallari, e per la verifica d’ipotesi.

La (3.5) rende evidente che, in ottica bayesiana, la quantità di interesse θ 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 θ , ma invece lo scopo è quello di fornire invece un giudizio di probabilità (o di formulare una “previsione”, nel linguaggio di de Finetti). Prima delle osservazioni, sulla base delle

nostre conoscenze assegnamo a θ una distribuzione a priori di probabilità. Dopo le osservazioni, correggiamo il nostro giudizio e assegniamo a θ una distribuzione a posteriori di probabilità.

3.3.1 Un esempio di aggiornamento bayesiano

Per descrivere l'aggiornamento bayesiano, in questo Capitolo (così come nei successivi) considereremo i dati di [Zetsche et al. \(2019\)](#). Questi ricercatori si sono chiesti se gli individui depressi manifestino delle aspettative accurate circa il loro umore futuro, oppure se tali aspettative siano distorte negativamente. Esamineremo qui i 30 partecipanti dello studio di [Zetsche et al. \(2019\)](#) che hanno riportato la presenza di un episodio di depressione maggiore in atto. All'inizio della settimana di test, a questi pazienti è stato chiesto di valutare l'umore che si aspettavano di esperire nei giorni seguenti della settimana. Mediante una app, i partecipanti dovevano poi valutare il proprio umore in cinque momenti diversi di ciascuno dei cinque giorni successivi. Lo studio considera diverse emozioni, ma qui ci concentriamo solo sulla tristezza.

Sulla base dei dati forniti dagli autori, abbiamo calcolato la media dei giudizi relativi al livello di tristezza raccolti da ciascun partecipante tramite la app. Tale media è stata poi sottratta dall'aspettativa del livello di tristezza fornita all'inizio della settimana. La discrepanza tra aspettative e realtà è stata considerata come un evento dicotomico: valori positivi di tale differenza indicano che le aspettative circa il livello di tristezza erano maggiori del livello di tristezza effettivamente esperito — ciò significa che le aspettative future risultano negativamente distorte (evento codificato con “1”). Viceversa, si ha che le aspettative risultano positivamente distorte se la differenza descritta in precedenza assume un valore negativo (evento codificato con “0”).

Nel campione dei 30 partecipanti clinici di [Zetsche et al. \(2019\)](#), le aspettative future di 23 partecipanti risultano distorte negativamente e quelle di 7 partecipanti risultano distorte positivamente. Chiameremo θ la probabilità dell'evento “le aspettative del partecipante sono distorte negativamente”. Ci poniamo il problema di ottenere una stima a posteriori di θ avendo osservato 23 “successi” in 30 prove.⁴

⁴Si noti un punto importante: dire semplicemente che la stima di θ è uguale a $23/30 = 0.77$ ci porta ad ignorare il livello di incertezza associato a tale stima. Infatti, lo stesso valore (0.77) si può ottenere come $23/30$, o $230/300$, o $2300/3000$, o $23000/30000$, ma l'incertezza di una stima pari a 0.77 è molto diversa nei quattro

3.4 Modello probabilistico

Nel caso dello studio di [Zetsche et al. \(2019\)](#), i dati qui considerati possono essere considerati la manifestazione di una variabile casuale Bernoulliana – 23 “successi” in 30 prove. Se i dati rappresentano una proporzione, allora possiamo adottare un modello probabilistico binomiale quale meccanismo generatore dei dati:

$$y \sim \text{Bin}(n, \theta), \quad (3.6)$$

laddove θ 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 θ . Ma ora abbiamo il problema inverso, ovvero quello di fare inferenza su θ 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: vogliamo dunque ottenere informazioni su θ avendo osservato i dati y .

Nel modello probabilistico che stiamo esaminando, il termine n viene trattato come una costante nota e θ come una *variabile casuale*. Dato che θ è incognito, ma abbiamo a disposizione i dati y , svolgeremo l’inferenza su θ mediante la regola di Bayes per determinare la distribuzione a posteriori $p(\theta \mid y)$.

casi. Quando si traggono conclusioni dai dati è invece necessario quantificare il livello della nostra incertezza relativamente alla stima del parametro di interesse (nel caso presente, θ). Lo strumento ci consente di quantificare tale incertezza è la distribuzione a posteriori $p(\theta \mid y)$. Ovviamente, $p(\theta \mid y)$ assume forme molto diverse nei quattro casi descritti sopra.

Osservazione. Si noti che il modello probabilistico (3.6) 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. 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. 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. 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.

3.5 Distribuzioni a priori

Quando adottiamo un approccio bayesiano, i parametri della distribuzione di riferimento non venono considerati come delle costanti incognite ma bensì vengono trattati come variabili casuali e, di conseguenza, i parametri assumono una particolare distribuzione che nella statistica bayesiana viene definita come “a priori”. I parametri (o il parametro), che possiamo indicare con θ , possono assumere delle distribuzioni a priori differenti; a seconda delle informazioni disponibili bisogna cercare di assegnare una distribuzione di θ in modo tale che venga assegnata una probabilità maggiore a quei valori che si ritengono più plausibili per θ .

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 θ .

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 credenze precedenti (p. es., la maggior parte dei cigni sono bianchi) la nuova osservazione rafforza le mie credenze precedenti: più nuove osservazioni raccolgo (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 credenze precedenti, ciò contribuisce a diminuire la certezza che attribuisco alle mie credenze: tanto maggiori diventano le osservazioni non corrispondenti alle mie credenze (p. es., più cigni neri vedo), tanto più si indeboliscono le mie credenze. Fondamentalmente, tanto più forti sono le mie credenze precedenti, di tante più osservazioni incompatibili (ad esempio, cigni neri) ho bisogno per cambiare idea.

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 vi 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.

3.5.1 Tipologie di distribuzioni 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 informa-*

tive 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 $\frac{\alpha}{\beta} = \frac{5}{2}$.

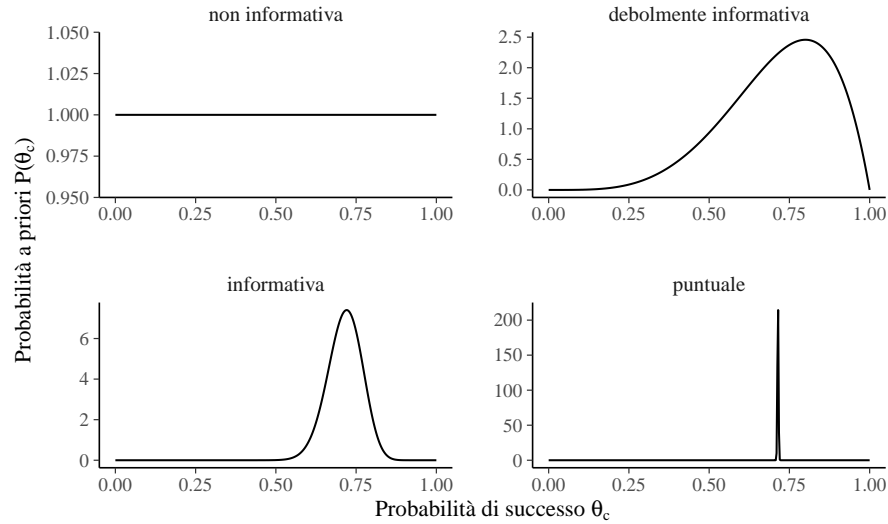


Figura 3.1: Esempi di distribuzioni a priori per il parametro θ_c nel Modello Binomiale.

3.5.2 Selezione 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, [van de 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, [van de Schoot et al. \(2021\)](#) 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 ??.

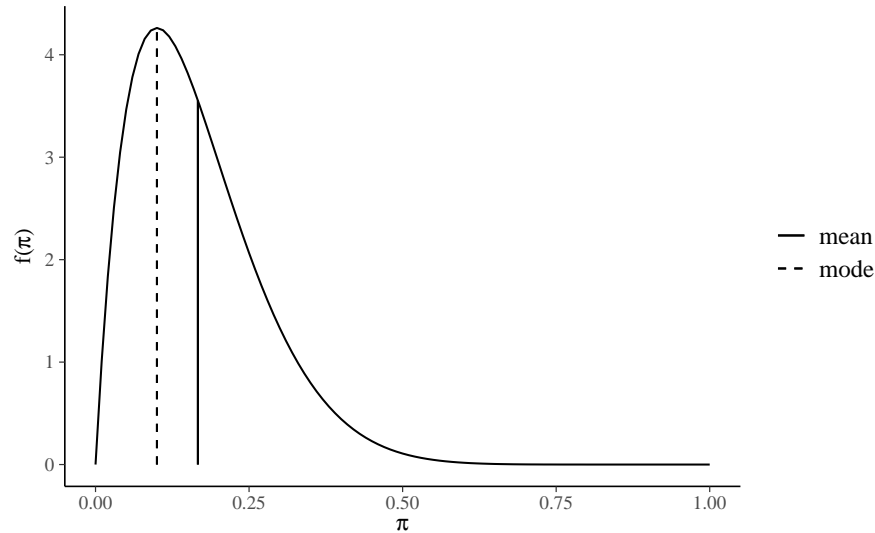
3.5.3 La distribuzione a priori per i dati di [Zetsche et al. \(2019\)](#)

In un problema concreto di analisi dei dati, la scelta della distribuzione a priori dipende dalle credenze a priori che vogliamo includere nell’analisi dei dati. Se non abbiamo alcuna informazione a priori, potremmo pensare di usare una distribuzione a priori uniforme, ovvero una Beta di parametri $\alpha = 1$ e $\beta = 1$. Questa, tuttavia, è una cattiva idea perché il risultato ottenuto non è invariante a seconda della trasformazione della scala dei dati (ad esempio, se esprimiamo l’altezza in cm piuttosto che in m). Il problema della *riparametrizzazione* verrà discusso nel Capitolo ?? **TODO**. È invece raccomandato usare una distribuzione a priori poco informativa, come ad esempio $\text{Beta}(2, 2)$.

Nella presente discussione, per fare un esempio, quale distribuzione a priori useremo una $\text{Beta}(2, 10)$, ovvero:

$$p(\theta) = \frac{\Gamma(12)}{\Gamma(2)\Gamma(10)} \theta^{2-1} (1 - \theta)^{10-1}.$$

```
bayesrules::plot_beta(alpha = 2, beta = 10, mean = TRUE, mode = TRUE)
```



La $\text{Beta}(2, 10)$ esprime la credenza che θ assume valori < 0.5 , con il valore più plausibile pari a circa 0.1. Questo è assolutamente implausibile, nel caso dell'esempio in discussione. Adotteremo una tale distribuzione a priori solo per scopi didattici, per esplorare le conseguenze di tale scelta (molto più sensato sarebbe stato usare $\text{Beta}(2, 2)$).

3.6 Verosimiglianza

Oltre alla distribuzione a priori di θ , nel numeratore della regola di Bayes troviamo la funzione di verosimiglianza. Iniziamo dunque con una definizione.

Definizione 3.1. La *funzione di verosimiglianza* $\mathcal{L}(\theta | y) = f(y | \theta)$, $\theta \in \Theta$, è la funzione di massa o di densità di probabilità dei dati y vista come una funzione del parametro sconosciuto (o dei parametri sconosciuti) θ .

Detto in altre parole, le funzioni di verosimiglianza e di (massa o densità di) probabilità sono formalmente identiche, ma è completamente diversa la loro interpretazione. Nel caso della funzione di massa o di densità di

probabilità la distribuzione del vettore casuale delle osservazioni campionarie y dipende dai valori assunti dal parametro (o dai parametri) θ ; nel caso della la funzione di verosimiglianza la credibilità assegnata a ciascun possibile valore θ viene determinata avendo acquisita l'informazione campionaria y che rappresenta l'elemento condizionante. In altri termini, la funzione di verosimiglianza è lo strumento che consente di rispondere alla seguente domanda: avendo osservato i dati y , quanto risultano (relativamente) credibili i diversi valori del parametro θ ?

Spesso per indicare la verosimiglianza si scrive $\mathcal{L}(\theta)$ se è chiaro a quali valori y ci si riferisce. La verosimiglianza \mathcal{L} è una curva (in generale, una superficie) nello spazio Θ del parametro (in generale, dei parametri) che riflette la credibilità relativa dei valori θ alla luce dei dati osservati.

Notiamo un punto importante: la funzione $\mathcal{L}(\theta \mid y)$ non è una funzione di densità. Infatti, essa non racchiude un'area unitaria.

In conclusione, la funzione di verosimiglianza descrive in termini relativi il sostegno empirico che $\theta \in \Theta$ riceve da y . Infatti, la funzione di verosimiglianza assume forme diverse al variare di y (lasciamo come esercizio da svolgere la verifica di questa affermazione).

3.6.1 La stima di massima verosimiglianza

La funzione di verosimiglianza rappresenta la “credibilità relativa” dei valori del parametro di interesse. Ma qual è il valore più credibile? Se utilizziamo soltanto la funzione di verosimiglianza, allora la risposta è data dalla stima di massima verosimiglianza.

Definizione 3.2. Un valore di θ che massimizza $\mathcal{L}(\theta \mid y)$ sullo spazio parametrico Θ è detto *stima di massima verosimiglianza* (s.m.v.) di θ ed è indicato con $\hat{\theta}$:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta). \quad (3.7)$$

Il paradigma frequentista utilizza la funzione di verosimiglianza quale unico strumento per giungere alla stima del valore più credibile del parametro sconosciuto θ . Tale stima corrisponde al punto di massimo della funzione di verosimiglianza. In base all'approccio bayesiano, invece, il valore più credibile del parametro sconosciuto θ , anziché alla s.m.v., corrisponde invece alla moda (o media, o mediana) della distribuzione a

posteriori $p(\theta | y)$ che si ottiene combinando la verosimiglianza $p(y | \theta)$ con la distribuzione a priori $p(\theta)$. Per un approfondimento della stima di massima verosimiglianza si veda l'Appendice ??.

3.6.2 La log-verosimiglianza

Dal punto di vista pratico risulta più conveniente utilizzare, al posto della funzione di verosimiglianza, il suo logaritmo naturale, ovvero la funzione di log-verosimiglianza:

$$\ell(\theta) = \log \mathcal{L}(\theta). \quad (3.8)$$

Poiché il logaritmo è una funzione strettamente crescente (usualmente si considera il logaritmo naturale), allora $\mathcal{L}(\theta)$ e $\ell(\theta)$ assumono il massimo (o i punti di massimo) in corrispondenza degli stessi valori di θ :

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \ell(\theta) = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta).$$

Per le proprietà del logaritmo, si ha

$$\ell(\theta) = \log \left(\prod_{i=1}^n f(y | \theta) \right) = \sum_{i=1}^n \log f(y | \theta). \quad (3.9)$$

Si noti che non è necessario lavorare con i logaritmi, ma è fortemente consigliato. Il motivo è che i valori della verosimiglianza, in cui si moltiplicano valori di probabilità molto piccoli, possono diventare estremamente piccoli – qualcosa come 10^{-34} . In tali circostanze, non è sorprendente che i programmi dei computer mostrino problemi di arrotondamento numerico. Le trasformazioni logaritmiche risolvono questo problema.

Osservazione. Seguendo una pratica comune, in questa dispensa spesso useremo la notazione $p(\cdot)$ per rappresentare due quantità differenti, ovvero la funzione di verosimiglianza e la distribuzione a priori. Questo piccolo abuso di notazione riflette il seguente punto di vista: anche se la verosimiglianza non è una funzione di densità di probabilità, noi non vogliamo stressare questo aspetto, ma vogliamo piuttosto pensare alla verosimiglianza e alla distribuzione a priori come a due elementi che sono egualmente necessari per calcolare la distribuzione a posteriori. In altri termini, per così dire, questa notazione assegna lo stesso status epistemologico alle due diverse quantità che si trovano al numeratore della regola di Bayes.

Esercizio 3.1. Per i dati di [Zetsche et al. \(2019\)](#), ovvero 23 “successi” in 30 prove, si trovi e si interpreti la funzione di verosimiglianza.

Per i dati di [Zetsche et al. \(2019\)](#) la funzione di verosimiglianza corrisponde alla funzione binomiale di parametro $\theta \in [0, 1]$ sconosciuto. Abbiamo osservato un “successo” 23 volte in 30 “prove”, dunque, $y = 23$ e $n = 30$. La funzione di verosimiglianza diventa

$$\mathcal{L}(\theta \mid y) = \frac{(23 + 7)!}{23!7!} \theta^{23} + (1 - \theta)^7. \quad (3.10)$$

Per costruire la funzione di verosimiglianza dobbiamo applicare la (3.10) tante volte, cambiando ogni volta il valore θ ma *tenendo sempre costante il valore dei dati*. Per esempio, se poniamo $\theta = 0.1$

$$\mathcal{L}(\theta \mid y) = \frac{(23 + 7)!}{23!7!} 0.1^{23} + (1 - 0.1)^7$$

otteniamo

```
dbinom(23, 30, 0.1)
#> [1] 9.737e-18
```

Se poniamo $\theta = 0.2$

$$\mathcal{L}(\theta \mid y) = \frac{(23 + 7)!}{23!7!} 0.2^{23} + (1 - 0.2)^7$$

otteniamo

```
dbinom(23, 30, 0.2)
#> [1] 3.581e-11
```

e così via. La figura 3.2 — costruita utilizzando 100 valori equispaziati $\theta \in [0, 1]$ — fornisce una rappresentazione grafica della funzione di verosimiglianza.

```
n <- 30
y <- 23
theta <- seq(0, 1, length.out = 100)
```

```

like <- choose(n, y) * theta^y * (1 - theta)^(n - y)
tibble(theta, like) %>%
  ggplot(aes(x = theta, y = like)) +
  geom_line() +
  labs(
    y = expression(L(theta)),
    x = expression("Valori possibili di" ~ theta)
  )

```

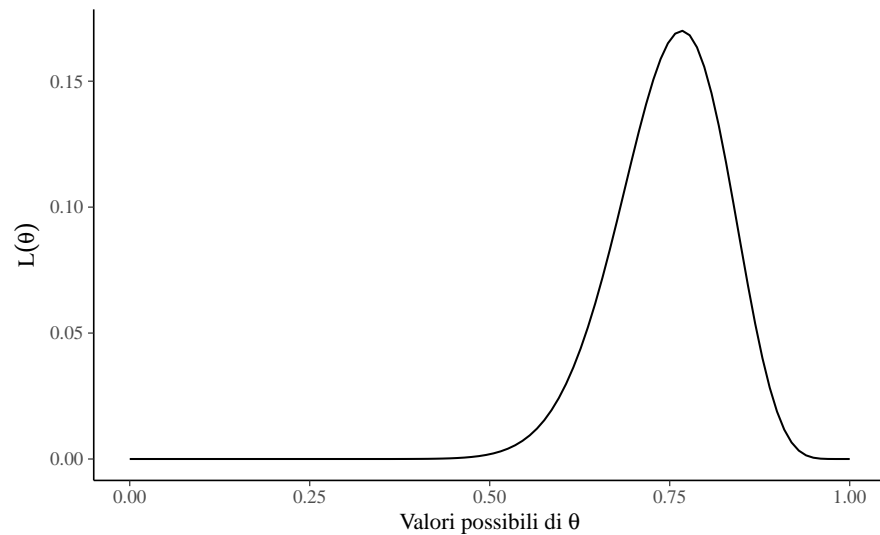


Figura 3.2: Funzione di verosimiglianza nel caso di 23 successi in 30 prove.

Come possiamo interpretare la curva che abbiamo ottenuto? Per alcuni valori θ la funzione di verosimiglianza assume valori piccoli; per altri valori θ la funzione di verosimiglianza assume valori più grandi. Questi ultimi sono i valori di θ “più credibili” e il valore $23/30$ è il valore più credibile di tutti. La funzione di verosimiglianza di θ valuta la compatibilità dei dati osservati $Y = y$ con i diversi possibili valori θ . In termini più formali possiamo dire che la funzione di verosimiglianza ha la seguente interpretazione: sulla base dei dati, $\theta_1 \in \Theta$ è più credibile di $\theta_2 \in \Theta$ come indice del modello probabilistico generatore delle osservazioni se $\mathcal{L}(\theta_1) > \mathcal{L}(\theta_2)$.

3.7 La verosimiglianza marginale

Per il calcolo di $p(\theta | y)$ è necessario dividere il prodotto tra la distribuzione a priori e la verosimiglianza per una costante di normalizzazione. Tale costante di normalizzazione, detta *verosimiglianza marginale*, ha lo scopo di fare in modo che $p(\theta | y)$ abbia area unitaria.

Si noti che il denominatore della regola di Bayes (ovvero la verosimiglianza marginale) è sempre espresso nei termini di un integrale. Tranne in pochi casi particolari, tale integrale non ha una soluzione analitica. Per questa ragione, l'inferenza bayesiana procede calcolando una approssimazione della distribuzione a posteriori mediante metodi numerici.

Esercizio 3.2. Si trovi la verosimiglianza marginale per i dati di [Zetsche et al. \(2019\)](#).

Supponiamo che nel numeratore bayesiano la verosimiglianza sia moltiplicata per una distribuzione uniforme, $\text{Beta}(1, 1)$. In questo caso, il prodotto si riduce alla funzione di verosimiglianza. In riferimento ai dati di [Zetsche et al. \(2019\)](#), la costante di normalizzazione per si ottiene semplicemente marginalizzando la funzione di verosimiglianza $p(y = 23, n = 30 | \theta)$ sopra θ , ovvero risolvendo l'integrale:

$$p(y = 23, n = 30) = \int_0^1 \binom{30}{23} \theta^{23} (1 - \theta)^7 d\theta. \quad (3.11)$$

Una soluzione numerica si trova facilmente usando R:

```
like_bin <- function(theta) {
  choose(30, 23) * theta^23 * (1 - theta)^7
}
integrate(like_bin, lower = 0, upper = 1)$value
#> [1] 0.03226
```

La derivazione analitica della costante di normalizzazione qui discussa è fornita nell'Appendice ??.

3.8 Distribuzione a posteriori

La distribuzione a posteriori si trova applicando il teorema di Bayes:

$$\text{probabilità a posteriori} = \frac{\text{probabilità a priori} \cdot \text{verosimiglianza}}{\text{costante di normalizzazione}}$$

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); questo problema verrà discusso nel Capitolo ??

Una volta trovata la distribuzione a posteriori, possiamo usarla per derivare altre quantità di interesse. Questo viene generalmente ottenuto calcolando il valore atteso:

$$J = \int f(\theta) p(\theta | y) \, d\theta$$

Se $f(\cdot)$ è la funzione identità, ad esempio, J risulta essere la media di θ :

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

3.9 Distribuzione predittiva a priori

La distribuzione a posteriori è l'oggetto centrale nella statistica bayesiana, ma non è l'unico. Oltre a fare inferenze sui valori dei parametri, potremmo voler fare inferenze sui dati. Questo può essere fatto calcolando la *distribuzione predittiva a priori*:

$$p(y^*) = \int_{\Theta} p(y^* | \theta) p(\theta) \, d\theta. \quad (3.12)$$

La (3.12) descrive la distribuzione prevista dei dati in base al modello (che include la distribuzione a priori e la verosimiglianza). Questi sono i dati y^* che ci aspettiamo, dato il modello, prima di avere osservato i dati del campione.

Possiamo utilizzare campioni dalla distribuzione predittiva a priori per valutare e calibrare i modelli utilizzando le nostre conoscenze dominio-specifiche. Ad esempio, ci potremmo chiedere: “È sensato che un modello dell’altezza umana preveda che un essere umano sia alto -1.5 metri?”. Già prima di misurare una singola persona, possiamo renderci conto dell’assurdità di questa domanda. Se la distribuzione prevista dei dati consente domande di questo tipo, è chiaro che il modello deve essere riformulato.

Osservazione. Si dice comunemente che l’adozione di una prospettiva probabilistica per la modellazione conduce all’idea che i modelli generano dati. Se i modelli generano dati, possiamo creare modelli adatti per i nostri dati solo pensando a come i dati potrebbero essere stati generati. Inoltre, questa idea non è solo un concetto astratto. Assume una concreta nella forma della distribuzione predittiva a priori. Se la distribuzione predittiva a priori non ha senso, come abbiamo detto sopra, diventa necessario riformulare il modello.

3.10 Distribuzione predittiva a posteriori

Un’altra quantità utile da calcolare è la distribuzione predittiva a posteriori:

$$p(\tilde{y} | y) = \int_{\Theta} p(\tilde{y} | \theta) p(\theta | y) d\theta. \quad (3.13)$$

Questa è la distribuzione dei dati attesi futuri \tilde{y} alla luce della distribuzione a posteriori $p(\theta | y)$, che a sua volta è una conseguenza del modello (distribuzione a priori e verosimiglianza) e dei dati osservati. In altre parole, questi sono i dati che il modello si aspetta dopo aver osservato i dati y . Dalla (3.13) possiamo vedere che le previsioni sui dati attesi futuri sono calcolate integrando (o marginalizzando) sulla distribuzione a posteriori dei parametri. Di conseguenza, le previsioni calcolate in que-

sto modo incorporano l'incertezza relativa alla stima dei parametri del modello.

Considerazioni conclusive

Questo Capitolo ha brevemente passato in rassegna alcuni concetti di base dell'inferenza statistica bayesiana. 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. I concetti importanti che abbiamo appreso in questo Capitolo sono quelli di distribuzione a priori, verosimiglianza, verosimiglianza marginale e distribuzione a posteriori. Questi sono i concetti fondamentali della statistica bayesiana.

A

Simbologia di base

Per una scrittura più sintetica possono essere utilizzati alcuni simboli matematici.

- $\log(x)$: il logaritmo naturale di x .
- L'operatore logico booleano \wedge significa “e” (congiunzione forte) mentre il connettivo di disgiunzione \vee significa “o” (oppure) (congiunzione debole).
- Il quantificatore esistenziale \exists vuol dire “esiste almeno un” e indica l'esistenza di almeno una istanza del concetto/oggetto indicato. Il quantificatore esistenziale di unicità $\exists!$ (“esiste soltanto un”) indica l'esistenza di esattamente una istanza del concetto/oggetto indicato. Il quantificatore esistenziale \nexists nega l'esistenza del concetto/oggetto indicato.
- Il quantificatore universale \forall vuol dire “per ogni.”
- \mathcal{A}, \mathcal{S} : insiemi.
- $x \in A$: x è un elemento dell'insieme A .
- L'implicazione logica “ \Rightarrow ” significa “implica” (se ...allora). $P \Rightarrow Q$ vuol dire che P è condizione sufficiente per la verità di Q e che Q è condizione necessaria per la verità di P .
- L'equivalenza matematica “ \Leftrightarrow ” significa “se e solo se” e indica una condizione necessaria e sufficiente, o corrispondenza biunivoca.
- Il simbolo $|$ si legge “tale che.”
- Il simbolo \triangleq (o $:=$) si legge “uguale per definizione.”
- Il simbolo Δ indica la differenza fra due valori della variabile scritta a destra del simbolo.
- Il simbolo \propto si legge “proporzionale a.”
- Il simbolo \approx si legge “circa.”
- Il simbolo \in della teoria degli insiemi vuol dire “appartiene” e indica l'appartenenza di un elemento ad un insieme. Il simbolo \notin vuol dire “non appartiene.”
- Il simbolo \subseteq si legge “è un sottoinsieme di” (può coincidere con l'insieme stesso). Il simbolo \subset si legge “è un sottoinsieme proprio di.”

- Il simbolo $\#$ indica la cardinalità di un insieme.
- Il simbolo \cap indica l'intersezione di due insiemi. Il simbolo \cup indica l'unione di due insiemi.
- Il simbolo \emptyset indica l'insieme vuoto o evento impossibile.
- In matematica, argmax identifica l'insieme dei punti per i quali una data funzione raggiunge il suo massimo. In altre parole, $\operatorname{argmax}_x f(x)$ è l'insieme dei valori di x per i quali $f(x)$ raggiunge il valore più alto.
- a, c, α, γ : scalari.
- x, y : vettori.
- X, Y : matrici.
- $X \sim p$: la variabile casuale X si distribuisce come p .
- $p(\cdot)$: distribuzione di massa o di densità di probabilità.
- $p(y \mid x)$: la probabilità o densità di y dato x , ovvero $p(y = Y \mid x = X)$.
- $f(x)$: una funzione arbitraria di x .
- $f(X; \theta, \gamma)$: f è una funzione di X con parametri θ, γ . Questa notazione indica che X sono i dati che vengono passati ad un modello di parametri θ, γ .
- $\mathcal{N}(\mu, \sigma^2)$: distribuzione gaussiana di media μ e varianza σ^2 .
- $\text{Beta}(\alpha, \beta)$: distribuzione Beta di parametri α e β .
- $\mathcal{U}(a, b)$: distribuzione uniforme con limite inferiore a e limite superiore b .
- $\text{Cauchy}(\alpha, \beta)$: distribuzione di Cauchy di parametri α (posizione: media) e β (scala: radice quadrata della varianza).
- $\mathcal{B}(p)$: distribuzione di Bernoulli di parametro p (probabilità di successo).
- $\text{Bin}(n, p)$: distribuzione binomiale di parametri n (numero di prove) e p (probabilità di successo).
- $\mathbb{KL}(p \parallel q)$: la divergenza di Kullback-Leibler da p a q .

Bibliografia

- de Finetti, B. (1931). Probabilismo. *Logos*, pages 163–219.
- de Finetti, B. (1970). *Teoria delle probabilità: sintesi introduttiva con appendice critica*. Einaudi.
- Horn, S. and Loewenstein, G. (2021). Underestimating learning by doing. *Available at SSRN 3941441*.
- Martin, O. A., Kumar, R., and Lao, J. (2022). *Bayesian Modeling and Computation in Python*. CRC Press.
- van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., and Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primer*, 1(1):1–26.
- Zetsche, U., Bürkner, P.-C., and Renneberg, B. (2019). Future expectations in clinical depression: Biased or realistic? *Journal of Abnormal Psychology*, 128(7):678–688.