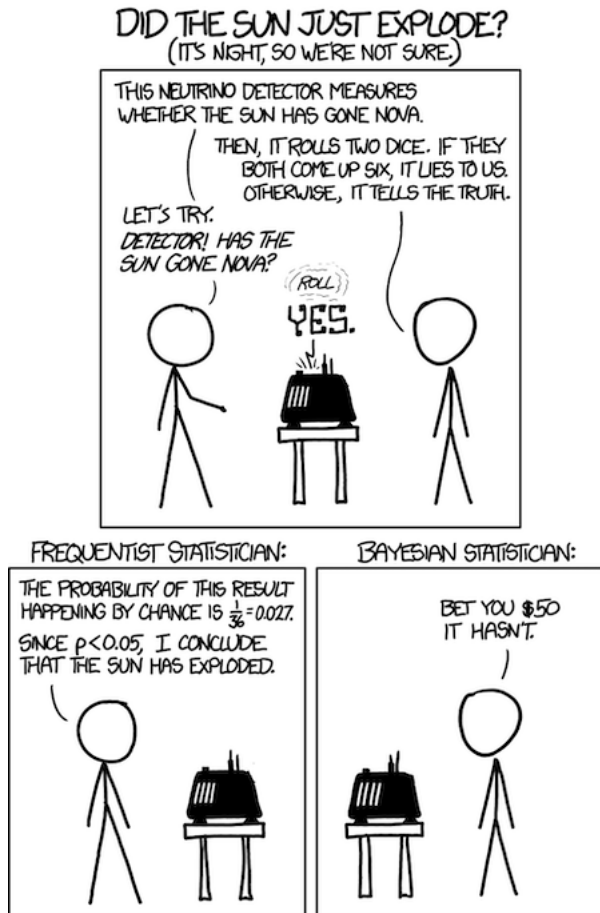


Corrado Caudek

Data Science per psicologi



Psicometria – AA 2021/2022





Indice

Elenco delle figure	vii
Elenco delle tabelle	ix
Prefazione	xi
I Il calcolo delle probabilità	1
1 La logica dell'incerto	3
1.1 Che cos'è la probabilità?	3
1.2 Variabili casuali e probabilità di un evento	6
1.2.1 Eventi e probabilità	6
1.2.2 Spazio campione e risultati possibili	6
1.3 Variabili casuali	7
1.4 Usare la simulazione per stimare le probabilità	8
1.5 La legge dei grandi numeri	11
1.6 Variabili casuali multiple	14
1.7 Funzione di massa di probabilità	16
2 Sintesi a posteriori	21
2.1 Stima puntuale	21
2.2 Intervallo di credibilità	22
2.2.1 Intervallo di credibilità a code uguali	23
2.2.2 Intervallo di credibilità a densità a posteriori più alta	23
2.2.3 Interpretazione	24
2.3 Un esempio concreto	24
2.3.1 Stime puntuali della distribuzione a posteriori	25
2.3.2 Intervallo di credibilità	26
2.3.3 Probabilità della distribuzione a posteriori	28
2.4 La funzione di perdita attesa	32



Elenco delle figure

1.1	Stima della probabilità di successo in funzione del numero di lanci di una moneta.	12
1.2	Stima della probabilità di successo in funzione del numero di lanci di una moneta – scala logaritmica.	13
1.3	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.	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

Il calcolo delle probabilità



1

La logica dell'incerto

In questa parte della dispensa verrà introdotta la teoria delle probabilità. Prima di entrare nei dettagli, cerchiamo di capire perché la probabilità sia cruciale per la ricerca scientifica.

La teoria delle probabilità è cruciale per la scienza perché la ricerca procede mediante l'inferenza induttiva. Non siamo mai completamente sicuri della verità di una proposizione (ipotesi, teoria): al valore di verità di una proposizione possiamo solo assegnare un giudizio probabilistico. L'approccio bayesiano è una scuola di pensiero che usa la probabilità per quantificare il grado di fiducia che può essere attribuito ad una proposizione. L'inferenza statistica bayesiana è un tipo di inferenza induttiva che ha lo scopo di quantificare la fiducia che si ha nell'ipotesi H dopo il verificarsi del dato d'evidenza E . Per quantificare un tale grado di fiducia l'inferenza statistica bayesiana utilizza la teoria delle probabilità. Una comprensione dell'inferenza statistica bayesiana richiede dunque, preliminarmente, la conoscenza della teoria delle probabilità.

1.1 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 ab-

biamo del mondo, non al mondo in sé. Di conseguenza è detta, in contrapposizione alla precedente definizione, “soggettiva”.

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

Nell'interpretazione frequentista, la probabilità $P(E)$ rappresenta la frequenza relativa a lungo termine di un grande numero di ripetizioni di un esperimento casuale sotto le medesime condizioni. Viene stressata qui l'idea che ciò di cui parliamo è qualcosa che emerge nel momento in cui è possibile ripetere l'esperimento casuale tante volte sotto le medesime condizioni – sono invece esclusi gli eventi unici e irripetibili.

L'interpretazione bayesiana della probabilità fa invece ricorso ad una concezione più ampia, non legata al solo evento in sé ma che include anche il soggetto assegnante la funzione di probabilità. In pratica l'assegnazione di probabilità bayesiana viene effettuata dal decisore, in base alle proprie conoscenze a priori integrate con tutto il generico bagaglio culturale personale. In questo modo, la probabilità non sarà obbligatoriamente la stessa per tutti i soggetti, ma varierà a seconda delle informazioni a disposizione, dell'esperienza personale e soprattutto del punto di vista proprio di ogni decisore ed è dunque assimilabile al “grado di fiducia” – in inglese *degree of belief* – di un dato soggetto, in un dato istante e con un dato insieme d'informazioni, circa l'accadere dell'evento E . “[N]essuna 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” (de Finetti, 1931).

L'impostazione bayesiana, sviluppata da Ramsey e de Finetti, riconduce l'assegnazione di probabilità allo scommettere sul verificarsi di un evento:

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 .

Secondo De Finetti, le valutazioni di probabilità degli eventi devono rispondere ai principi di equità e coerenza. Una scommessa risponde al principio di *equità* se il ruolo di banco e giocatore sono scambiabili in ogni momento del gioco e sempre alle stesse condizioni. Una scommessa risponde al principio di *coerenza* se non vi sono combinazioni di scommesse che consentano (sia al banco che al giocatore) di realizzare perdite o vincite certe.

L'approccio definettiano dell'impostazione della scommessa si basa dunque sulle assunzioni di razionalità e coerenza del decisore, al quale è fatto esplicito divieto di effettuare scommesse a perdita o guadagno certo. Il decisore, proponendo la scommessa, deve essere disposto a scambiare il posto dello scommettitore con quello del banco.

Il metodo della scommessa, oltre che una definizione, fornisce un mezzo operativo di assegnazione della probabilità. Sulla base di questa definizione operativa, che si può ritenere ragionevolmente soddisfatta dal comportamento di un qualunque individuo che agisca in modo razionale in condizioni di incertezza, possono essere agevolmente dimostrate tutte le proprietà classiche della probabilità: essa non può assumere valori negativi, né può essere superiore all'unità; se E è un evento certo, la sua probabilità è 1; se invece E è un evento impossibile, la sua probabilità è 0.

I problemi posti dall'approccio definettiano riguardano l'arbitrarietà dell'assegnazione soggettiva di probabilità la quale sembra negare la validità dell'intero costrutto teorico. In risposta a tale critica, i bayesiani sostengono che gli approcci oggettivisti alla probabilità nascondono scelte arbitrarie preliminari e sono basate su assunzioni implausibili. È molto più onesto esplicitare subito tutte le scelte arbitrarie effettuate nel corso dell'analisi in modo da controllarne coerenza e razionalità.

1.2 Variabili casuali e probabilità di un evento

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

1.2.1 Eventi e probabilità

Nella teoria delle probabilità il risultato “testa” nel lancio di una moneta è chiamato *evento*.¹ Ad esempio, $Y = 1$ denota l'evento in cui il lancio di una moneta produce come risultato testa. Il funzionale $P(\cdot)$ 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 $P(Y = 1) = 0.5$.

Se la moneta è equilibrata dobbiamo anche avere $P(Y = 0) = 0.5$. I due eventi $Y = 1$ e $Y = 0$ sono *mutuamente esclusivi* nel senso che non possono entrambi verificarsi contemporaneamente: $P(Y = 1 \text{ e } Y = 0) = 0$. 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, $P(Y = 1 \text{ o } Y = 0) = 1$.

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

1.2.2 Spazio campione e risultati possibili

Anche se il lancio di una moneta produce sempre uno specifico risultato nel mondo reale, 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.

¹Per un ripasso delle nozioni di base della teoria degli insiemi, si veda l'Appendice ??.

I risultati possibili che si possono osservare come conseguenza del lancio di una moneta determinano i valori possibili che la variabile casuale può assumere. L'insieme Ω di tutti i risultati possibili è chiamato *spazio campione* (*sample space*). Lo spazio campione 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 campione: in questo caso, l'insieme dei risultati $\{1, 3, 5\}$. Se esce 3, per esempio, diciamo che si è verificato l'evento “dispari” (ma l'evento “dispari” si sarebbe anche verificato anche se fosse uscito 1 o 5).



1.3 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 i cui valori non possono essere previsti con esattezza. Se la moneta è equilibrata, c'è una probabilità del 50% che il lancio della moneta dia come risultato “testa” e una probabilità del 50% che dia come risultato “croce”. Per facilitare la trattazione, le variabili casuali assumono solo valori numerici. Per lo specifico lancio della moneta in questione, diciamo, ad esempio, che la variabile casuale Y assume il valore 1 se esce testa e il valore 0 se esce croce.

Una variabile casuale può essere *discreta* o *continua*. Una variabile casuale discreta può assumere un numero finito di valori x_1, \dots, x_n , in corrispondenza degli eventi E_1, \dots, E_n che si verificano con le rispettive probabilità p_1, \dots, p_n . Un esempio è il punteggio totale di un test psicometrico costituito da item su scala Likert. Invece un esempio di una variabile casuale continua è la distanza tra due punti, che può assumere infiniti valori all'interno di un certo intervallo. L'insieme S dei valori che la variabile casuale può assumere è detto *spazio dei valori* o *spazio*

degli stati. La caratteristica fondamentale di una variabile casuale è data dall'insieme delle probabilità dei suoi valori, detta *distribuzione di probabilità*. Nel seguito useremo la notazione $P(\cdot)$ per fare riferimento alle distribuzioni di probabilità delle variabili casuali discrete e $p(\cdot)$ per fare riferimento alla densità di probabilità delle variabili casuali continue. In questo contesto, l'insieme dei valori che la variabile casuale può assumere è detto *supporto* della sua distribuzione di probabilità. Il supporto di una variabile casuale può essere finito (come nel caso di una variabile casuale uniforme di supporto $[a, b]$) o infinito (nel caso di una variabile causale gaussiana il cui supporto coincide con la retta reale).

1.4 Usare la simulazione per stimare le probabilità

I metodi basati sulla simulazione consentono di stimare le probabilità degli eventi in un modo diretto, se siamo in grado di generare molteplici e casuali realizzazioni 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, la stessa operazione si può realizzare 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 $P(Y = 1)$ è data dalla frequenza relativa del numero di volte in cui abbiamo osservato l'evento di interesse ($Y = 1$):

```
M <- 100
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] = 0.53
```

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(100)
}
#> estimated Pr[Y = 1] = 0.44
#> estimated Pr[Y = 1] = 0.52
#> estimated Pr[Y = 1] = 0.46
#> estimated Pr[Y = 1] = 0.57
#> estimated Pr[Y = 1] = 0.47
#> estimated Pr[Y = 1] = 0.46
#> estimated Pr[Y = 1] = 0.48
#> estimated Pr[Y = 1] = 0.49
```

```
#> estimated Pr[Y = 1] = 0.47  
#> estimated Pr[Y = 1] = 0.62
```

Dato che la moneta è equilibrata, la stima delle probabilità dell'evento $Pr[Y = 1]$ è simile a al valore che ci aspettiamo ($P(Y = 1) = 0.5$), 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(1000)  
}  
#> estimated Pr[Y = 1] = 0.497  
#> estimated Pr[Y = 1] = 0.529  
#> estimated Pr[Y = 1] = 0.493  
#> estimated Pr[Y = 1] = 0.511  
#> estimated Pr[Y = 1] = 0.506  
#> estimated Pr[Y = 1] = 0.52  
#> estimated Pr[Y = 1] = 0.49  
#> estimated Pr[Y = 1] = 0.495  
#> estimated Pr[Y = 1] = 0.489  
#> estimated Pr[Y = 1] = 0.496
```

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] = 0.4885  
#> estimated Pr[Y = 1] = 0.4957  
#> estimated Pr[Y = 1] = 0.4902  
#> estimated Pr[Y = 1] = 0.5032  
#> estimated Pr[Y = 1] = 0.5048  
#> estimated Pr[Y = 1] = 0.4931  
#> estimated Pr[Y = 1] = 0.4965  
#> estimated Pr[Y = 1] = 0.499
```

```
#> estimated Pr[Y = 1] = 0.4979  
#> estimated Pr[Y = 1] = 0.4973
```

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

1.5 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 quello che accade all'aumentare del numero M di ripetizioni del lancio della moneta consiste nel registrare la stima della probabilità dell'evento $P(Y = 1)$ in funzione del numero di ripetizioni dell'esperimento casuale per ogni $m \in 1 : M$. Possiamo ottenere un grafico dell'andamento della stima di $P(Y = 1)$ in funzione di m come:

```
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 <- tibble(  
  n = 1:nrep,  
  estimate  
)  
d %>%  
  ggplot(  
    aes(x = n, y = estimate)
```

```

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

```

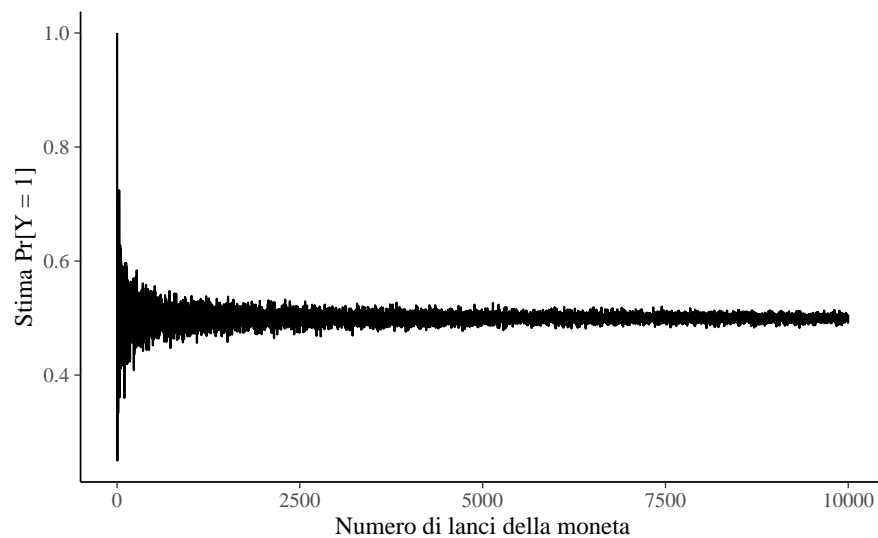


Figura 1.1: Stima della probabilità di successo in funzione del numero di lanci di una moneta.

Dato che il grafico 1.1 espresso su una scala lineare non rivela chiaramente l'andamento della simulazione, imponiamo una scala logaritmica sull'asse delle ascisse (x). Su scala logaritmica, i valori tra 1 e 10 vengono tracciati all'incirca con la stessa ampiezza che si osserva tra i valori 50 e 700, eccetera.

```

d %>%
  ggplot(
    aes(x = n, y = estimate)
  ) +
  geom_line() +

```

```
scale_x_log10(  
  breaks = c(  
    1, 3, 10, 50, 200,  
    700, 2500, 10000  
  )  
) +  
theme(legend.title = element_blank()) +  
labs(  
  x = "Numero di lanci della moneta",  
  y = "Stima Pr[Y = 1]"  
)
```

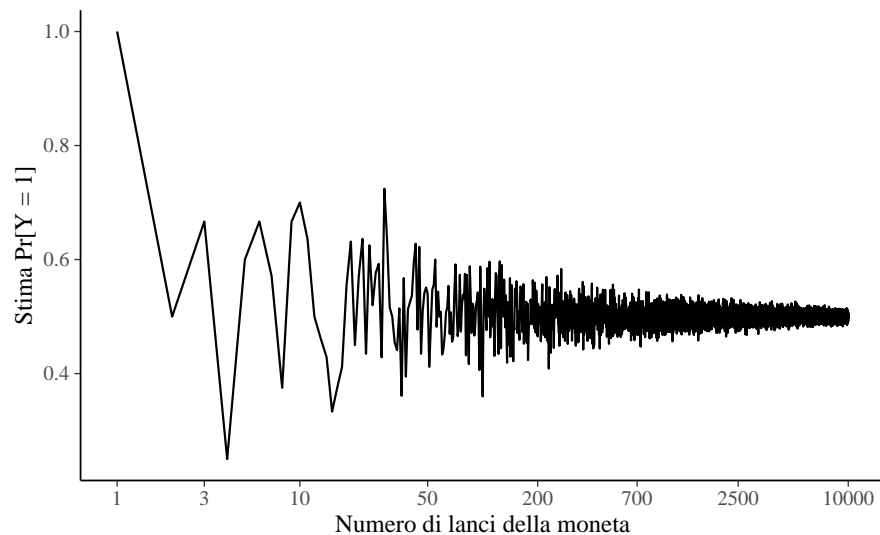


Figura 1.2: Stima della probabilità di successo in funzione del numero di lanci di una moneta – scala logaritmica.

La *legge dei grandi numeri* ci dice che, all'aumentare del numero di ripetizioni dell'esperimento casuale, la media dei risultati ottenuti tende al valore atteso, man mano che vengono eseguite più prove. Nella figura 1.2 vediamo infatti che, all'aumentare del numero M di lanci della moneta, la stima di $P(Y = 1)$ converge al valore 0.5.

1.6 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. I risultati di ciascuno dei tre lanci possono essere rappresentati da una diversa variabile casuale, ad esempio, Y_1, Y_2, Y_3 . 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 $P(Y_n = 1) = 0.5$ e $P(Y_n = 0) = 0.5$.

È possibile 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] 1 0 1
z <- sum(c(y1, y2, y3))
cat("z =", z, "\n")
#> z = 2
```

ovvero,

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

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

oppure, ancora più semplicemente:

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

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

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

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

```
table(z) / M
#> z
#>      0      1      2      3
#> 0.1258 0.3750 0.3748 0.1244
```

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
#> 0.06340 0.24917 0.37360 0.25022 0.06361

```

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

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

1.7 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:

$$\begin{aligned}
 p_Z(0) &= 1/16 & \text{TTTT} \\
 p_Z(1) &= 4/16 & \text{HTTT, THTT, TTHT, TTTH} \\
 p_Z(2) &= 6/16 & \text{HHTT, HTHT, HTTH, THHT, THTH, TTTH} \\
 p_Z(3) &= 4/16 & \text{HHHT, HHHT, HTHH, THHH} \\
 p_Z(4) &= 1/16 & \text{HHHH}
 \end{aligned}$$

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

$$P_Z(z) = P(Z = z).$$

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

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

Una rappresentazione grafica della stima della funzione di massa di probabilità per l'esperimento casuale del lancio di quattro monete equilibrate è fornita nella figura 1.3.

```
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
```

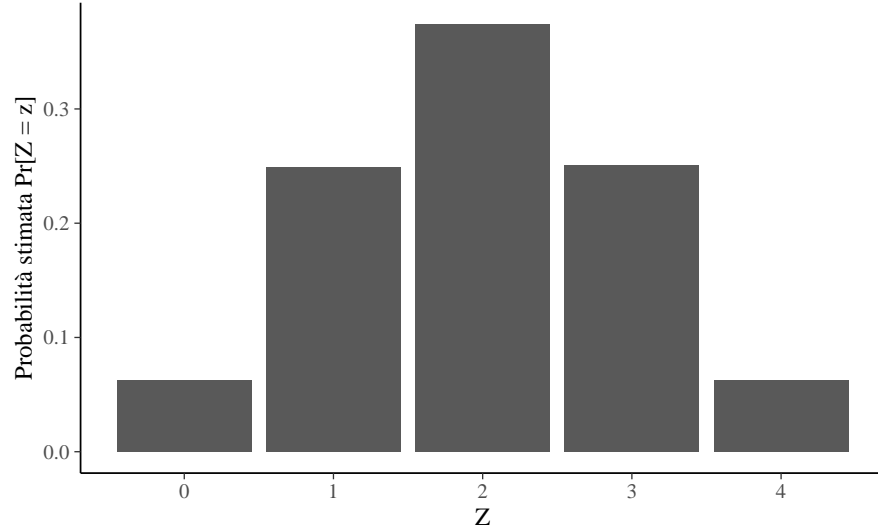


Figura 1.3: Grafico di $M = 100\,000$ simulazioni della funzione di massa di probabilità di una variabile casuale definita come il numero di teste in quattro lanci di una moneta equilibrata.

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 1.1. Nel caso dell'esempio discusso nella Sezione 1.7, la probabilità che la variabile casuale Z sia un numero dispari è

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

Commenti e considerazioni finali

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 campione 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 $F(X) = P(X < x)$, 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.



2

Sintesi a posteriori

La distribuzione a posteriori è un modo per descrivere il nostro grado di incertezza rispetto al parametro incognito (o rispetto ai parametri incogniti) oggetto dell'inferenza. La distribuzione a posteriori contiene tutte le informazioni disponibili sui possibili valori del parametro. Se il parametro esaminato è monodimensionale (o bidimensionale) è possibile fornire un grafico di tutta la distribuzione a posteriori $p(\theta | y)$. Tuttavia, spesso vogliamo anche giungere ad una sintesi numerica della distribuzione a posteriori, soprattutto se il vettore dei parametri ha più di due dimensioni. A a questo proposito è possibile utilizzare le consuete statistiche descrittive, come media, mediana, moda, varianza, deviazione standard e i quantili. In alcuni casi, queste statistiche descrittive sono più facili da presentare e interpretare rispetto alla rappresentazione grafica della distribuzione a posteriori.

La stima puntuale della tendenza centrale della distribuzione a posteriori fornisce informazioni su quello che può essere considerato come il “valore più plausibile” del parametro. L'intervallo di credibilità fornisce invece un'indicazione dell'ampiezza dell'intervallo che contiene una determinata quota della massa della distribuzione a posteriori del parametro.

2.1 Stima puntuale

Per sintetizzare la distribuzione a posteriori in modo da giungere ad una stima puntuale di θ si è soliti scegliere tra moda, mediana o media a seconda del tipo di distribuzione con cui si ha a che fare e della sua forma. Ogni stima puntuale ha una sua interpretazione.

- La media è il valore atteso a posteriori del parametro.
- La moda può essere interpretata come il singolo valore più credibile (“più plausibile”) del parametro, alla luce dei dati, ovvero il valore per

il parametro θ che massimizza la distribuzione a posteriori. Per questa ragione la moda viene detta *massimo a posteriori*, MAP. Il limite della moda quale statistica riassuntiva della distribuzione a posteriori è che, talvolta, tale distribuzione è multimodale e il MAP non è necessariamente il valore “più credibile”.

- La mediana è il valore del parametro tale per cui, su entrambi i lati di essa, giace il 50% della massa di probabilità a posteriori.

La misura di variabilità del parametro è la *varianza a posteriori* la quale, nel caso di una distribuzione a posteriori ottenuta per via numerica, si calcola con la formula della varianza che conosciamo rispetto alla tendenza centrale data dalla media a posteriori. La radice quadrata della varianza a posteriori è la *deviazione standard a posteriori* che descrive l'incertezza a posteriori circa il parametro di interesse nella stessa unità di misura dei dati.

Le procedure bayesiane basate sui metodi MCMC utilizzano un numero finito di campioni dalla distribuzione stazionaria, e una tale caratteristica della simulazione introduce un ulteriore livello di incertezza nella stima del parametro. L'*errore standard della stima* (in inglese *Monte Carlo standard error*, MCSE) misura l'accuratezza della simulazione. La deviazione standard a posteriori e l'errore standard della stima sono due concetti completamente diversi. La deviazione standard a posteriori descrive l'incertezza circa il parametro (l'ampiezza della distribuzione a posteriori) ed è una funzione della dimensione del campione; il MCSE descrive invece l'incertezza nella stima del parametro dovuta alla simulazione MCMC ed è una funzione del numero di iterazioni nella simulazione.

2.2 Intervallo di credibilità

Molto spesso la stima puntuale è accompagnata da una stima intervallare (abbiamo già incontrato questo aspetto nel Capitolo ?? discutendo lo schema beta-binomiale). Nella statistica bayesiana, se il parametro $\theta \in \Theta$ è monodimensionale, si dice *intervallo di credibilità* un intervallo di valori I_α che contiene la proporzione $1 - \alpha$ della massa di probabilità della funzione a posteriori:

$$p(\Theta \in I_\alpha \mid y) = 1 - \alpha. \quad (2.1)$$

L'intervallo di credibilità ha lo scopo di esprimere il nostro grado di incertezza riguardo la stima del parametro. Se il parametro θ è multidimensionale, si parla invece di “regione di credibilità”.

La condizione (2.1) non determina un unico intervallo di credibilità al $(1 - \alpha) \cdot 100\%$. In realtà esiste un numero infinito di tali intervalli. Ciò significa che dobbiamo definire alcune condizioni aggiuntive per la scelta dell'intervallo di credibilità. Esaminiamo due delle condizioni aggiuntive più comuni.

2.2.1 Intervallo di credibilità a code uguali

Un intervallo di credibilità *a code uguali* a livello α è un intervallo

$$I_\alpha = [q_{\alpha/2}, 1 - q_{\alpha/2}],$$

dove q_z è un quantile z della distribuzione a posteriori. Per esempio, l'intervallo di credibilità a code uguali al 95% è un intervallo

$$I_{0.05} = [q_{0.025}, q_{0.975}]$$

che lascia il 2.5% della massa di densità a posteriori in ciascuna coda.

2.2.2 Intervallo di credibilità a densità a posteriori più alta

Nell'intervallo di credibilità a code uguali alcuni valori del parametro che sono inclusi nell'intervallo possono avere una credibilità a posteriori più bassa rispetto a quelli esterni all'intervallo. L'intervallo di credibilità *a densità a posteriori più alta* (in inglese *High Posterior Density Interval*, HPD) è invece costruito in modo tale da assicurare di includere nell'intervallo tutti i valori θ che sono a posteriori maggiormente credibili. Graficamente questo intervallo può essere ricavato tracciando una linea orizzontale sulla rappresentazione della distribuzione a posteriori e regolando l'altezza della linea in modo tale che l'area sottesa alla curva sia pari a $1 - \alpha$. Questo tipo di intervallo è il più stretto possibile, tra tutti i possibili intervalli di credibilità allo stesso livello di fiducia. Se la distribuzione a posteriori è simmetrica unimodale, l'intervallo di credibilità a densità a posteriori più alta corrisponde all'intervallo di credibilità a code uguali.

2.2.3 Interpretazione

L'interpretazione dell'intervallo di credibilità è molto intuitiva: l'intervallo di credibilità è un intervallo di valori all'interno del quale cade il valore del parametro incognito con un particolare livello di probabilità soggettiva. Possiamo dire che, dopo aver visto i dati crediamo, con un determinato livello di probabilità soggettiva, che il valore del parametro (ad esempio, la dimensione dell'effetto di un trattamento) abbia un valore compreso all'interno dell'intervallo che è stato calcolato, laddove per probabilità soggettiva intendiamo “il grado di fiducia che lo sperimentatore ripone nel verificarsi di un evento”. Gli intervalli di credibilità si calcolano con un software.

2.3 Un esempio concreto

Per fare un esempio pratico, consideriamo nuovamente i valori del BDI-II dei 30 soggetti clinici di [Zetsche et al. \(2019\)](#):

```
suppressPackageStartupMessages(library("bayesrules"))

df <- tibble(
  y = c(
    26, 35, 30, 25, 44, 30, 33, 43, 22, 43,
    24, 19, 39, 31, 25, 28, 35, 30, 26, 31,
    41, 36, 26, 35, 33, 28, 27, 34, 27, 22
  )
)
```

Un valore BDI-II ≥ 30 indica la presenza di un livello “grave” di depressione. Nel campione clinico di [Zetsche et al. \(2019\)](#),

```
sum(df$y > 29)
#> [1] 17
```

17 pazienti su 30 manifestano un livello grave di depressione.

Supponiamo di volere stimare la distribuzione a posteriori della probabilità θ di depressione “grave” nei pazienti clinici, così come viene misurata

dal test BDI-II, imponendo su θ una distribuzione a priori $\text{Beta}(8, 2)$.

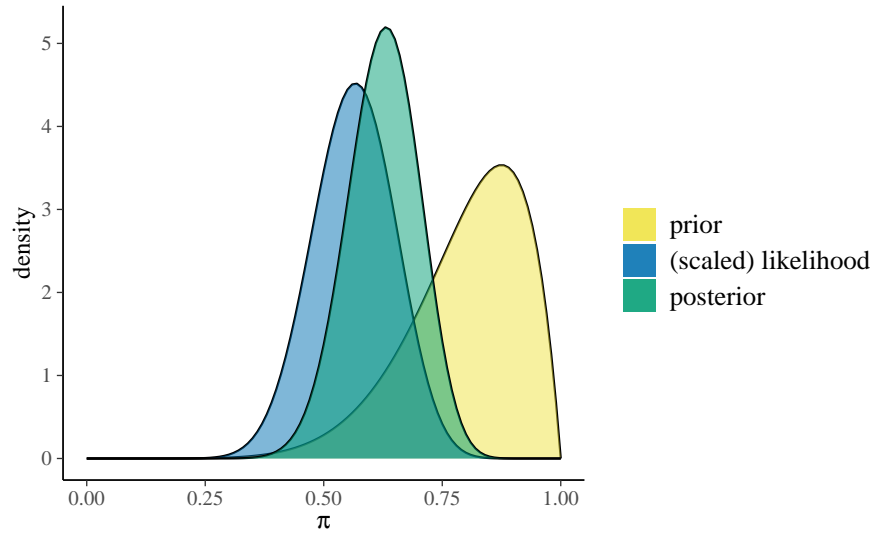
Sappiamo che il modello Beta-Binomiale può essere espresso nella forma seguente:

$$\begin{aligned} Y|\theta &\sim \text{Bin}(30, \theta) \\ \theta &\sim \text{Beta}(8, 2) \end{aligned}$$

con una corrispondente distribuzione a posteriori $\text{Beta}(25, 15)$:

$$f(\theta|y = 17) = \frac{\Gamma(25 + 15)}{\Gamma(25)\Gamma(15)} \theta^{25-1} (1 - \theta)^{15-1} \quad \text{for } \theta \in [0, 1] . \quad (2.2)$$

```
plot_beta_binomial(alpha = 8, beta = 2, y = 17, n = 30)
```



2.3.1 Stime puntuali della distribuzione a posteriori

Una volta trovata l'intera distribuzione a posteriori, quale valore di sintesi è necessario riportare? Questa sembra una domanda innocente, ma in realtà è una domanda a cui è difficile rispondere. La stima bayesiana dei parametri è fornita dall'intera distribuzione a posteriori, che non è

un singolo numero, ma una funzione che mappa ciascun valore del parametro ad un valore di plausibilità. Quindi non è necessario scegliere una stima puntuale. In linea di principio, una stima puntuale non è quasi mai necessaria ed è spesso dannosa in quanto comporta una perdita di informazioni.

Tuttavia talvolta una tale sintesi è richiesta. Diverse risposte sono allora possibili. La media della distribuzione a posteriori per θ è

$$\mathbb{E}(\pi \mid y = 17) = \frac{\alpha}{\alpha + \beta} = \frac{25}{25 + 15} = 0.625.$$

Una stima del massimo della probabilità a posteriori, o brevemente massimo a posteriori, MAP (da *maximum a posteriori probability*), è la moda della distribuzione a posteriori. Nel caso presente, una stima del MAP può essere ottenuta nel modo seguente:

$$\text{Mo}(\pi \mid y = 17) = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{25 - 1}{25 + 15 - 2} = 0.6316.$$

Gli stessi risultati si ottengono usando la chiamata a `bayesrules::summarize_beta_binomial()`:

```
summarize_beta_binomial(alpha = 8, beta = 2, y = 17, n = 30)
#>      model alpha beta mean  mode    var    sd
#> 1  prior      8   2 0.800 0.8750 0.014545 0.12060
#> 2 posterior    25  15 0.625 0.6316 0.005716 0.07561
```

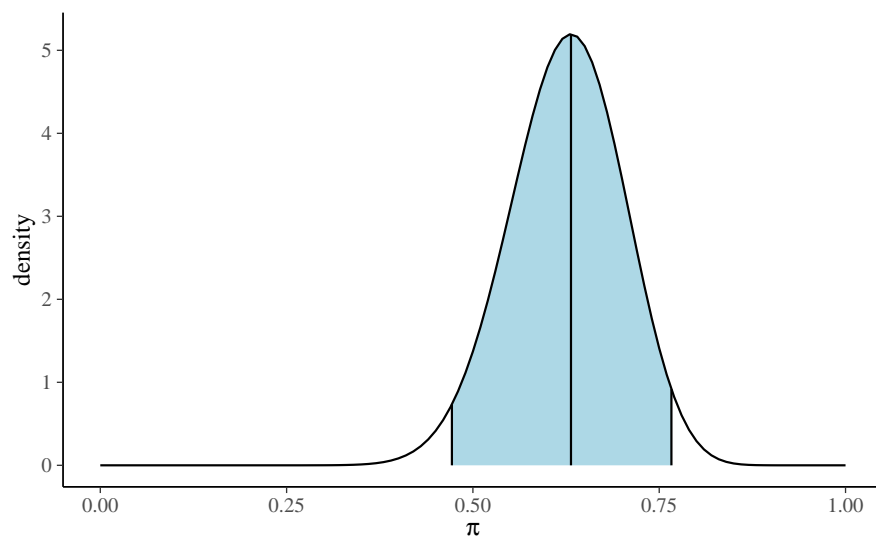
La mediana si ottiene con

```
qbeta(.5, shape1 = 25, shape2 = 15)
#> [1] 0.6271
```

2.3.2 Intervallo di credibilità

È più comune sintetizzare la distribuzione a posteriori mediante l'intervallo di credibilità. Per esempio, l'intervallo di credibilità a code uguali al 95%

```
plot_beta_ci(alpha = 25, beta = 15, ci_level = 0.95)
```



è dato dalla chiamata

```
qbeta(c(0.025, 0.975), 25, 15)
#> [1] 0.4718 0.7664
```

Il calcolo precedente evidenzia l'interpretazione intuitiva dell'intervallo di credibilità. Tale intervallo, infatti, può essere interpretato come la probabilità che θ assuma valori compresi tra 0.472 e 0.766:

$$P(\theta \in (0.472, 0.766) | Y = 17) = \int_{0.472}^{0.766} f(\theta | y = 17) d\theta = 0.95,$$

ovvero

```
postFun <- function(theta) {
  gamma(25 + 15) /
    (gamma(25) * gamma(15)) * theta^24 * (1 - theta)^14
}
integrate(
```

```

postFun,
  lower = 0.4717951,
  upper = 0.7663607
)$value
#> [1] 0.95

```

Possiamo costruire diversi intervalli di credibilità a code equivalenti. Ad esempio, l'intervallo di credibilità compreso tra il 25-esimo e il 75-esimo percentile è

```

qbeta(c(0.25, 0.75), 25, 15)
#> [1] 0.5744 0.6779

```

ovvero, abbiamo una certezza a posteriori del 50% che la probabilità di depressione grave tra i pazienti clinici sia un valore compreso tra 0.57 e 0.68.

Non esiste un livello credibile “corretto”. I ricercatori, utilizzano vari livelli, ad esempio 50%, 80% o 95%, a seconda del contesto dell'analisi. Ciascuno di questi intervalli fornisce un'immagine diversa della nostra comprensione della distribuzione a posteriori del parametro di interesse.

Non è inoltre necessario riportare l'intervallo di credibilità a code uguali. Se la distribuzione a posteriori è fortemente asimmetrica è più sensato riportare l'intervallo di credibilità a densità a posteriori più alta. L'intervallo HPD risulta più semplice da determinare quando la distribuzione a posteriori viene approssimata con il metodo MCMC.

2.3.3 Probabilità della distribuzione a posteriori

Il test di ipotesi è un compito comune dell'analisi della distribuzione a posteriori (si veda anche il Capitolo ??). Supponiamo che si voglia conoscere la probabilità a posteriori che θ sia superiore a 0.5. Per sapere quanto credibile sia l'evento $\theta > 0.5$ possiamo calcolare il seguente integrale:

$$P(\theta > 0.5 \mid y = 17) = \int_{0.5}^1 f(\theta \mid y = 17) d\theta ,$$

dove $f(\cdot)$ è la distribuzione $(25, 15)$:

```
pbeta(0.5, shape1 = 25, shape2 = 15, lower.tail = FALSE)
#> [1] 0.9459
```

il che è equivalente a:

```
postFun <- function(theta) {
  gamma(25 + 15) / (gamma(25) * gamma(15)) * theta^24 * (1 - theta)^14
}
integrate(
  postFun,
  lower = 0.5,
  upper = 1
)$value
#> [1] 0.9459
```

È anche possibile formulare un test di ipotesi contrastando due ipotesi contrapposte. Per esempio, $H_1 : \theta \geq 0.5$ e $H_2 : \theta < 0.5$. Ciò consente di calcolare l'*odds a posteriori* di $\theta > 0.5$:

$$\text{poterior odds} = \frac{H_1 | y = 17}{H_2 | y = 17} \quad (2.3)$$

ovvero

```
posterior_odds <-
  pbeta(0.5, shape1 = 25, shape2 = 15, lower.tail = FALSE) /
  pbeta(0.5, shape1 = 25, shape2 = 15, lower.tail = TRUE)
posterior_odds
#> [1] 17.5
```

L'odds a posteriori rappresenta l'aggiornamento delle nostre credenze dopo avere osservato $y = 17$ in $n = 30$. L'odds a priori di $\theta > 0.5$ era:

```
prior_odds <-
  pbeta(0.5, shape1 = 8, shape2 = 2, lower.tail = FALSE) /
  pbeta(0.5, shape1 = 8, shape2 = 2, lower.tail = TRUE)
prior_odds
#> [1] 50.2
```

Il *fattore di Bayes* (*Bayes Factor*; BF) confronta gli odds a posteriori con gli odds a priori e quindi fornisce informazioni su quanto sia mutata la nostra comprensione relativa a θ dopo avere osservato i nostri dati del campione:

$$\text{BF} = \frac{\text{odds a posteriori}}{\text{odds a priori}}.$$

Nel caso presente abbiamo

```
BF <- posterior_odds / prior_odds
BF
#> [1] 0.3485
```

Quindi, dopo avere osservato i dati, gli odds a posteriori della nostra ipotesi a proposito di θ sono pari a solo il 34% degli odds a priori.

Per fare un altro esempio, consideriamo invece il caso in cui le credenze a priori rivelano una credenza diametralmente opposta rispetto a θ che nel caso considerato in precedenza, ovvero $\text{Beta}(2, 8)$. In questo secondo caso, la distribuzione a posteriori diventa

```
summarize_beta_binomial(alpha = 2, beta = 8, y = 17, n = 30)
#>      model alpha beta  mean  mode    var    sd
#> 1   prior     2    8 0.200 0.1250 0.014545 0.12060
#> 2 posterior   19   21 0.475 0.4737 0.006082 0.07799
```

e il BF è

```
posterior_odds <-
  pbeta(0.5, shape1 = 19, shape2 = 21, lower.tail = FALSE) /
  pbeta(0.5, shape1 = 19, shape2 = 21, lower.tail = TRUE)

prior_odds <-
  pbeta(0.5, shape1 = 2, shape2 = 8, lower.tail = FALSE) /
  pbeta(0.5, shape1 = 2, shape2 = 8, lower.tail = TRUE)

BF <- posterior_odds / prior_odds
```

BF

#> [1] 30.07

In altre parole, in questo secondo esempio gli odds a posteriori della nostra ipotesi a proposito di θ sono aumentati di 30 volte rispetto agli odds a priori.

In generale, in un test di ipotesi che contrappone un'ipotesi sostantiva H_a ad un'ipotesi nulla H_0 il BF è un rapporto di odds per l'ipotesi sostantiva:

$$\text{Bayes Factor} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{P(H_a | Y)/P(H_0 | Y)}{P(H_a)/P(H_0)}.$$

Essendo un rapporto, il BF deve essere valutato rispetto al valore di 1. Ci sono tre possibilità:

- BF = 1: La credibilità di H_a non è cambiata dopo avere osservato i dati.
- BF > 1: La credibilità di H_a è aumentata dopo avere osservato i dati. Quindi maggiore è BF, più convincente risulta l'evidenza per H_a .
- BF < 1: La credibilità di H_a è diminuita dopo avere osservato i dati.

Non ci sono delle soglie universalmente riconosciute per interpretare il BF. Per esempio, [Lee and Wagenmakers \(2014\)](#) propongono il seguente schema:

BF	Interpretation
> 100	Extreme evidence for H_a
30 - 100	Very strong evidence for H_a
10 - 30	Strong evidence for H_a
3 - 10	Moderate evidence for H_a
1 - 3	Anecdotal evidence for H_a
1	No evidence
1/3 - 1	Anecdotal evidence for H_0
1/10 - 1/3	Moderate evidence for H_0
1/30 - 1/10	Strong evidence for H_0
1/100 - 1/30	Very strong evidence for H_0
< 1/100	Extreme evidence for H_0

Tuttavia, è importante notare che l'opinione maggiormente diffusa nella comunità scientifica sia quella che incoraggia a non trarre conclusioni rigide dai dati utilizzando dei criteri fissati una volta per tutte. Pertanto, non esiste una soglia univoca per il BF che consente di classificare le ipotesi dei ricercatori nelle due categorie “vero” o “falso”. Invece, è più utile adottare una pratica più flessibile capace di tenere in considerazione il contesto e le potenziali implicazioni di ogni singolo test di ipotesi. Inoltre, è stato molte volte ripetuto che la distribuzione a posteriori è molto più informativa di una decisione binaria: la rappresentazione di tutta la distribuzione a posteriori fornisce una misura olistica del nostro livello di incertezza riguardo all'affermazione (il parametro, ovvero l'ipotesi) che viene valutata.

2.4 La funzione di perdita attesa

Un modo razionale per giungere ad una decisione statistica utilizzando l'intera distribuzione a posteriori è quello di usare la *funzione di perdita* (*loss function*). La funzione di perdita è concetto nella teoria delle decisioni statistiche che consente di quantificare il costo derivante dalla decisione di scegliere quale stima del parametro il valore θ_0 quando esso in realtà ha il valore θ .

Per chiarire che cosa si intende per funzione di perdita, esaminiamo qui un semplice esempio nel quale vengono considerati due soli valori di probabilità per l'evento target, anziché l'intera distribuzione a posteriori (il codice è ricavato da [Schmettow, 2021](#)).

Si consideri la scelta di prendere o meno l'ombrello nell'uscire di casa. Le previsioni del tempo sono le seguenti:

```
Risultato <-  
  tibble(  
    risultato = c("piove", "non piove"),  
    prob = c(0.6, 0.4)  
  )  
Risultato  
#> # A tibble: 2 x 2  
#>   risultato prob
```



```
#>   <chr>      <dbl>
#> 1 piove      0.6
#> 2 non piove  0.4
```

Le azioni possibili sono: prendo / non prendo l'ombrello:

```
Azione <-
  tibble(azione = c("prendo l'ombrello", "non prendo l'ombrello"))
Azione
#> # A tibble: 2 x 1
#>   azione
#>   <chr>
#> 1 prendo l'ombrello
#> 2 non prendo l'ombrello
```

Assegniamo un costo massimo (4) alla conseguenza peggiore (“non prendo l'ombrello e piove”) e uno minimo (0) alla conseguenza migliore (“non prendo l'ombrello e non piove”):

```
Costi <-
  expand.grid(
    azione = Azione$azione,
    risultato = Risultato$risultato
  ) %>%
  inner_join(Risultato) %>%
  mutate(costo = c(2, 4, 2, 0))
Costi
#>           azione risultato prob costo
#> 1   prendo l'ombrello   piove  0.6    2
#> 2 non prendo l'ombrello   piove  0.6    4
#> 3   prendo l'ombrello non piove  0.4    2
#> 4 non prendo l'ombrello non piove  0.4    0
```

Calcoliamo ora il costo atteso delle due azioni tenuto conto delle probabilità che si verifichi l'uno o l'altro stato del mondo (ricordiamo che piove/non piove hanno una probabilità rispettivamente del 40% e del 60%), ovvero ponderiamo il costo di ogni azione con la probabilità che si verifichi l'evento corrispondente:

```

Util <-
  Costi %>%
    mutate(costo_condizionato = prob * costo) %>%
    group_by(azione) %>%
    summarise(costo_atteso = sum(costo_condizionato))
Util
#> # A tibble: 2 x 2
#>   azione          costo_atteso
#>   <fct>          <dbl>
#> 1 prendo l'ombrello          2
#> 2 non prendo l'ombrello      2.4

```

La regola di minimizzazione dei costi induce a scegliere l'alternativa che comporta il valore più basso: nel nostro esempio “prendere l'ombrello”.

La stessa logica dell'esempio può essere usata anche nel momento in cui, anziché avere solo due valori per la probabilità dello stato del mondo in questione (per esempio, pioverà / non pioverà), utilizziamo l'intera distribuzione a posteriori (per esempio, quella relativa alla previsione di pioggia). Concludiamo questi brevi accenni relativi alla funzione di perdita con una considerazione di [McElreath \(2020\)](#) il quale nota che, anche se gli statistici e i teorici dei giochi sono da tempo interessati alle funzioni di perdita e alle relazioni che intercorrono tra esse e l'inferenza bayesiana, i ricercatori non le usano quasi mai in modo esplicito.

Commenti e considerazioni finali

Questo capitolo introduce le procedure di base per la manipolazione della distribuzione a posteriori. Lo strumento fondamentale che è stato utilizzato è quello fornito dai campioni di valori del parametro che vengono estratti dalla distribuzione a posteriori. Lavorare con campioni di valori del parametro estratti dalla distribuzione a posteriori trasforma un problema di calcolo integrale in un problema di riepilogo dei dati. Abbiamo visto le procedure maggiormente usate che consentono di utilizzare i campioni a posteriori per produrre indici di sintesi della distribuzione a posteriori: gli intervalli di credibilità e le stime puntuali.

Bibliografia

- de Finetti, B. (1931). Probabilismo. *Logos*, pages 163–219.
- Horn, S. and Loewenstein, G. (2021). Underestimating learning by doing. *Available at SSRN 3941441*.
- Lee, M. D. and Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC Press, Boca Raton, Florida, 2nd edition edition.
- Schmettow, M. (2021). New statistics for design researchers.
- 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.