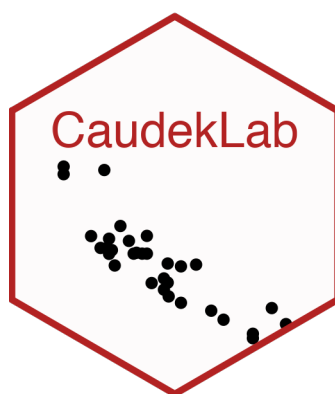


Psicometria

Corrado Caudek

Questo documento è stato realizzato con:

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



Nel blog della mia pagina personale sono forniti alcuni approfondimenti degli argomenti qui trattati.

<https://ccaudek.github.io/caudeklab/>

Indice

Indice	iii
Prefazione	v
La psicologia e la Data Science	v
Come studiare	vi
Sviluppare un metodo di studio efficace	vii
1 Sintesi a posteriori	1
1.1 Stima puntuale	1
1.2 Intervallo di credibilità	2
Intervallo di credibilità a code uguali	2
Intervallo di credibilità a densità a posteriori più alta	3
Interpretazione	3
1.3 Un esempio concreto	3
Stime puntuali della distribuzione a posteriori	4
Intervallo di credibilità	5
Probabilità della distribuzione a posteriori	6
Considerazioni conclusive	9
Bibliografia	11
Elenco delle figure	13

Prefazione

Data Science per psicologi contiene il materiale delle lezioni dell'insegnamento di *Psicometria B000286* (A.A. 2021/2022) rivolto agli studenti del primo anno del Corso di Laurea in Scienze e Tecniche Psicologiche dell'Università degli Studi di Firenze.

L'insegnamento di Psicometria si propone di fornire agli studenti un'introduzione all'analisi dei dati in psicologia. Le conoscenze/competenze che verranno sviluppate in questo insegnamento sono quelle della *Data science*, ovvero le conoscenze/competenze che si pongono all'intersezione tra statistica (ovvero, richiedono la capacità di comprendere teoremi statistici) e informatica (ovvero, richiedono la capacità di sapere utilizzare un software).

La psicologia e la Data Science

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

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

C'è una ragione molto semplice che dovrebbe farci capire perché la Data Science è così importante per la psicologia. Infatti, a ben pensarci, la psicologia è una disciplina intrinsecamente statistica, se per statistica intendiamo quella disciplina che studia la variazione delle caratteristiche degli individui nella popolazione. La psicologia studia *gli individui* ed è proprio la variabilità inter- e intra-individuale ciò che vogliamo descrivere e, in certi casi, predire. In questo senso, la psicologia è molto diversa dall'ingegneria, per esempio. Le proprietà di un determinato ponte sotto certe condizioni, ad esempio, sono molto simili a quelle di un altro ponte, sotto le medesime condizioni. Quindi, per un ingegnere la statistica è poco importante: le proprietà dei materiali sono unicamente dipendenti dalla loro composizione e restano costanti. Ma lo stesso non può dirsi degli individui: ogni individuo è unico e cambia nel tempo. E le variazioni tra gli individui, e di un individuo nel tempo, sono l'oggetto di studio proprio della

psicologia: è dunque chiaro che i problemi che la psicologia si pone sono molto diversi da quelli affrontati, per esempio, dagli ingegneri. Questa è la ragione per cui abbiamo tanto bisogno della *data science* in psicologia: perché la *data science* ci consente di descrivere la variazione e il cambiamento. E queste sono appunto le caratteristiche di base dei fenomeni psicologici.

Sono sicuro che, leggendo queste righe, a molti studenti sarà venuta in mente la seguente domanda: perché non chiediamo a qualche esperto di fare il “lavoro sporco” (ovvero le analisi statistiche) per noi, mentre noi (gli psicologi) ci occupiamo solo di ciò che ci interessa, ovvero dei problemi psicologici slegati dai dettagli “tecnici” della *data science*? La risposta a questa domanda è che non è possibile progettare uno studio psicologico sensato senza avere almeno una comprensione rudimentale della *data science*. Le tematiche della *data science* non possono essere ignorate né dai ricercatori in psicologia né da coloro che svolgono la professione di psicologo al di fuori dell’Università. Infatti, anche i professionisti al di fuori dall’università non possono fare a meno di leggere la letteratura psicologica più recente: il continuo aggiornamento delle conoscenze è infatti richiesto dalla deontologia della professione. Ma per potere fare questo è necessario conoscere un bel po’ di *data science*! Basta aprire a caso una rivista specialistica di psicologia per rendersi conto di quanto ciò sia vero: gli articoli che riportano i risultati delle ricerche psicologiche sono zeppi di analisi statistiche e di modelli formali. E la comprensione della letteratura psicologica rappresenta un requisito minimo nel bagaglio professionale dello psicologo.

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

Come studiare

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

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

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

Sviluppare un metodo di studio efficace

Memorization is not learning. — Richard Phillips Feynman

Avendo insegnato Psicometria molte volte in passato ho notato nel corso degli anni che gli studenti con l'atteggiamento mentale che descriverò qui sotto generalmente ottengono ottimi risultati. Alcuni studenti sviluppano naturalmente questo approccio allo studio, ma altri hanno bisogno di fare uno sforzo per maturarlo. Fornisco qui sotto una breve descrizione del "metodo di studio" che, nella mia esperienza, è il più efficace per affrontare le richieste di questo insegnamento.

- Dedicate un tempo sufficiente al materiale di base, apparentemente facile; assicuratevi di averlo capito bene. Cercate le lacune nella vostra comprensione. Leggere presentazioni diverse dello stesso materiale (in libri o articoli diversi) può fornire nuove intuizioni.
- Gli errori che facciamo sono i nostri migliori maestri. Istintivamente cerchiamo di dimenticare subito i nostri errori. Ma il miglior modo di imparare è apprendere dagli errori che commettiamo. In questo senso, una soluzione corretta è meno utile di una soluzione sbagliata. Quando commettiamo un errore questo ci fornisce un'informazione importante: ci fa capire qual è il materiale di studio sul quale dobbiamo ritornare e che dobbiamo capire meglio.
- C'è ovviamente un aspetto "psicologico" nello studio. Quando un esercizio o problema ci sembra incomprensibile, la cosa migliore da fare è dire: "mi arrendo", "non ho idea di cosa fare!". Questo ci rilassa: ci siamo già arresi, quindi non abbiamo niente da perdere, non dobbiamo più preoccuparci. Ma non dobbiamo fermarci qui. Le cose "migliori" che faccio (se ci sono) le faccio quando non ho voglia di lavorare. Alle volte, quando c'è qualcosa che non so fare e non ho idea di come affrontare, mi dico: "oggi non ho proprio voglia di fare fatica", non ho voglia di mettermi nello stato mentale per cui "in 10 minuti devo risolvere il problema perché dopo devo fare altre cose". Però ho voglia di *divertirmi* con quel problema e allora mi dedico a qualche aspetto "marginale" del problema, che so come affrontare, oppure considero l'aspetto più difficile del problema, quello che non so come risolvere, ma invece di cercare di risolverlo, guardo come altre persone hanno affrontato problemi simili, oppure lo stesso problema in un altro contesto. Non mi pongo l'obiettivo "risolvi il problema in 10 minuti", ma invece

quello di farmi un'idea "generale" del problema, o quello di capire un caso più specifico e più semplice del problema. Senza nessuna pressione. Infatti, in quel momento ho deciso di non lavorare (ovvero, di non fare fatica). Va benissimo se "parto per la tangente", ovvero se mi metto a leggere del materiale che sembra avere poco a che fare con il problema centrale (le nostre intuizioni e la nostra curiosità solitamente ci indirizzano sulla strada giusta). Quando faccio così, molto spesso trovo la soluzione del problema che mi ero posto e, paradossalmente, la trovo in un tempo minore di quello che, in precedenza, avevo dedicato a "lavorare" al problema. Allora perché non faccio sempre così? C'è ovviamente l'aspetto dei "10 minuti" che non è sempre facile da dimenticare. Sotto pressione, possiamo solo agire in maniera automatica, ovvero possiamo solo applicare qualcosa che già sappiamo fare. Ma se dobbiamo imparare qualcosa di nuovo, la pressione è un impedimento.

- È utile farsi da soli delle domande sugli argomenti trattati, senza limitarsi a cercare di risolvere gli esercizi che vengono assegnati. Quando studio qualcosa mi viene in mente: "se questo è vero, allora deve succedere quest'altra cosa". Allora verifico se questo è vero, di solito con una simulazione. Se i risultati della simulazione sono quelli che mi aspetto, allora vuol dire che ho capito. Se i risultati sono diversi da quelli che mi aspettavo, allora mi rendo conto di non avere capito e ritorno indietro a studiare con più attenzione la teoria che pensavo di avere capito – e ovviamente mi rendo conto che c'era un aspetto che avevo frainteso. Questo tipo di verifica è qualcosa che dobbiamo fare da soli, in prima persona: nessun altro può fare questo al posto nostro.
- Non aspettatevi di capire tutto la prima volta che incontrate un argomento nuovo. È utile farsi una nota mentalmente delle lacune nella vostra comprensione e tornare su di esse in seguito per cercare di colmarle. L'atteggiamento naturale, quando non capiamo i dettagli di qualcosa, è quello di pensare: "non importa, ho capito in maniera approssimativa questo punto, non devo preoccuparmi del resto". Ma in realtà non è vero: se la nostra comprensione è superficiale, quando il problema verrà presentato in una nuova forma, non riusciremo a risolverlo. Per cui i dubbi che ci vengono quando studiamo qualcosa sono il nostro alleato più prezioso: ci dicono esattamente quali sono gli aspetti che dobbiamo approfondire per potere migliorare la nostra preparazione.
- È utile sviluppare una visione d'insieme degli argomenti trattati, capire l'obiettivo generale che si vuole raggiungere e avere chiaro il contributo che i vari pezzi di informazione forniscono al raggiungimento di tale obiettivo. Questa organizzazione mentale del materiale di studio facilita la comprensione. È estremamente utile creare degli schemi di ciò che si sta studiando. Non aspettate che sia io a fornirvi un riepilogo di ciò che dovete imparare: sviluppate da soli tali schemi e tali riassunti.
- Tutti noi dobbiamo imparare l'arte di trovare le informazioni, non solo nel caso di questo insegnamento. Quando vi trovate di fronte a qualcosa

che non capite, o ottenete un oscuro messaggio di errore da un software,
ricordatevi: “Google is your friend”.

Corrado Caudek

Febbraio 2022

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.

1.1 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.

1.2 Intervallo di credibilità

Molto spesso la stima puntuale è accompagnata da una stima intervallare. 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:

$$p(\Theta \in I_\alpha \mid y) = 1 - \alpha. \quad (1.1)$$

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 (1.1) 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.

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.

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.

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.

1.3 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 <- data.frame(
  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. (ibid.),

```
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 Beta(8, 2).

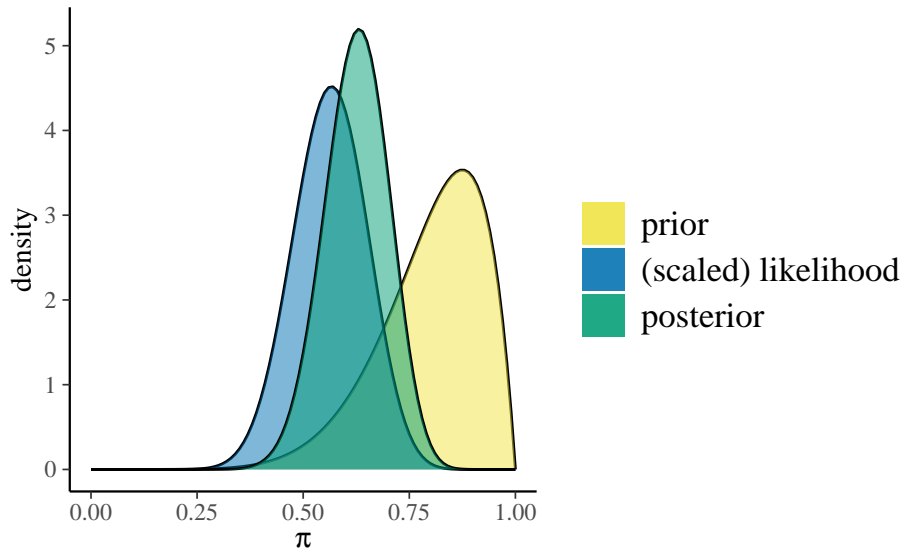
Sappiamo che il modello Beta-Binomiale può essere espresso nella forma seguente:

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

con una corrispondente distribuzione a posteriori Beta(25, 15):

$$f(\theta|y = 17) = \frac{\Gamma(25 + 15)}{\Gamma(25)\Gamma(15)} \theta^{25-1}(1 - \theta)^{15-1} \quad \text{for } \theta \in [0, 1] . \quad (1.2)$$

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



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
#> 1   prior      8    2 0.800 0.8750000 0.014545455
#> 2 posterior    25   15 0.625 0.6315789 0.005716463
#>      sd
#> 1 0.1206045
#> 2 0.0756073
```

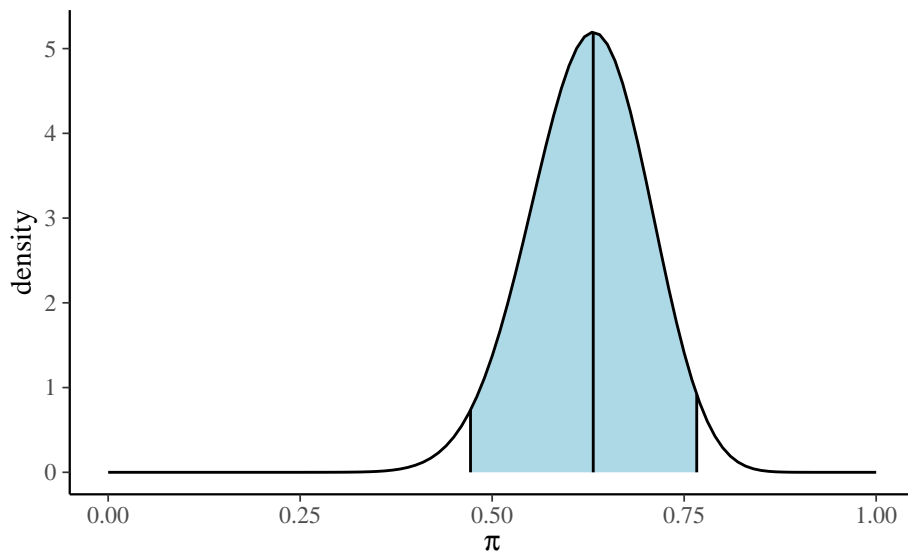
La mediana si ottiene con

```
qbeta(.5, shape1 = 25, shape2 = 15)
#> [1] 0.6271031
```

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] 0.4717951 0.7663607
```

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] 0.5743878 0.6778673
```

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.

Probabilità della distribuzione a posteriori

Il test di ipotesi è un compito comune dell'analisi della distribuzione a posteriori. 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] 0.9459355
```

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] 0.9459355
```

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 $\theta > 0.5$:

$$\text{poterior odds} = \frac{H_1 | y = 17}{H_2 | y = 17} \quad (1.3)$$

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] 17.49642
```

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] 0.3485343
```

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
#> 1   prior      2    8 0.200 0.1250000 0.014545455
#> 2 posterior    19   21 0.475 0.4736842 0.006082317
#>      sd
#> 1 0.12060454
#> 2 0.07798921
```

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] 30.07239
```

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:

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

Considerazioni conclusive

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

- Lee, M. D. e E.-J. Wagenmakers (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press (cit. a p. 9).
- Zetsche, U., P.-C. Bürkner e B. Renneberg (2019). «Future expectations in clinical depression: Biased or realistic?» In: *Journal of Abnormal Psychology* 128.7, pp. 678–688 (cit. a p. 3).

Elenco delle figure

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

Keywords Data science, Bayesian statistics.