

Data Science per psicologi

Corrado Caudek

2021-09-13

Indice

| | |
|--|-----------|
| Indice | 1 |
| 1 Distribuzioni a priori coniugate | 3 |
| 1.1 Il denominatore bayesiano | 3 |
| 1.2 Le distribuzioni coniugate | 4 |
| 1.3 Il modello Beta-Binomiale | 4 |
| 1.4 Principali distribuzioni coniugate | 11 |
| Considerazioni conclusive | 11 |
| Bibliografia | 13 |

Distribuzioni a priori coniugate

1.1 Il denominatore bayesiano

In generale, in un problema bayesiano i dati y provengono da una densità $p(y \mid \theta)$ e al parametro θ viene assegnata una densità a priori $p(\theta)$. La scelta della distribuzione a priori per i parametri è l'aspetto cruciale dell'impostazione bayesiana in quanto la distribuzione a priori è lo strumento attraverso cui le informazioni extra-sperimentali vengono inserite nel procedimento induttivo. La scelta della distribuzione a priori che meglio rappresenta le credenze iniziali implica che agenti diversi possono specificare sia le stesse distribuzioni di probabilità ma con diversi valori dei parametri, sia diverse distribuzioni di probabilità. Da ciò deriva una componente soggettiva. La scelta della distribuzione a priori ha inoltre importanti conseguenze di tipo computazionale. A meno di non utilizzare particolari forme analitiche, in generale risulta impossibile ottenere espressioni analitiche in forma esplicita per le distribuzioni a posteriori. Ciò dipende dal denominatore del rapporto di Bayes.

Dopo che un campione $Y = y$ è stato osservato, la funzione di verosimiglianza è uguale a $\mathcal{L}(\theta) = p(y \mid \theta)$ e la densità a posteriori è

$$p(\theta \mid y) = \frac{p(\theta)p(y \mid \theta)}{\int p(\theta)p(y \mid \theta)d\theta}.$$

Si noti che al denominatore troviamo un integrale. Tale integrale può essere risolto per via analitica solo se vengono usate le cosiddette *distribuzioni a priori coniugate per la verosimiglianza*.

Inoltre, i sommari della distribuzione a posteriori sono espressi come rapporto di integrali. Ad esempio, la media a posteriori di θ è data da

$$\mathbb{E}(\theta \mid y) = \frac{\int \theta \pi(\theta)p(y \mid \theta)d\theta}{\int \pi(\theta)p(y \mid \theta)d\theta}.$$

Il calcolo del valore atteso a posteriori richiede dunque la valutazione di due integrali, ciascuno dei quali non esprimibile in forma chiusa.

Per questa ragione, la strada principale che viene seguita nella modellistica bayesiana è quella che porta a determinare la distribuzione a posteriori non per via analitica, ma bensì mediante metodi numerici. La simulazione fornisce dunque la strategia generale del calcolo bayesiano. A questo fine vengono usati i metodi di campionamento detti Monte-Carlo Markov-Chain (MCMC). Tali metodi costituiscono una potente e praticabile alternativa per la costruzione della distribuzione a posteriori per modelli complessi e consentono di decidere quali distribuzioni a priori e quali distribuzioni di verosimiglianza usare sulla base di considerazioni teoriche soltanto, senza dovere preoccuparsi di altri vincoli. Dato che è basata su metodi computazionalmente intensivi, la stima numerica della funzione a posteriori può essere svolta soltanto mediante software. In anni recenti i metodi bayesiani di analisi dei dati sono diventati sempre più popolari proprio perché la potenza di calcolo necessaria per svolgere tali calcoli è ora alla portata di tutti. Questo non era vero solo pochi decenni fa.

1.2 Le distribuzioni coniugate

Una distribuzione di probabilità a priori si dice coniugata al modello utilizzato se appartiene alla stessa classe di distribuzioni della distribuzione a posteriori; in tali circostanze la distribuzione a priori e la distribuzione a posteriori differiscono solo per il valore dei parametri.

Definizione 1.1. Sia data la funzione di verosimiglianza $p(\theta \mid y)$ e sia \mathcal{D} una classe di densità a priori per θ . La famiglia di distribuzioni a priori \mathcal{D} è coniugata al modello se avviene che

$$p(\theta \mid y) \in \mathcal{D}.$$

Pertanto ciò che cambia nella distribuzione a posteriori non è la forma della distribuzione ma solo i parametri.

L'uso di distribuzioni a priori coniugate consente di determinare la distribuzione a posteriori per via analitica. Per quanto “semplice” in termini formali, questo approccio, però, limita di molto le possibili scelte del ricercatore. Nel senso che non è sempre sensato, dal punto di vista teorico, utilizzare distribuzioni a priori coniugate per la verosimiglianza per i parametri di interesse. Detto in altre parole: è possibile ottenere la distribuzione posteriore per via analitica solo per alcune specifiche combinazioni di distribuzione a priori e verosimiglianza, ma questo limita considerevolmente la flessibilità della modellizzazione.

1.3 Il modello Beta-Binomiale

Obiettivo di questo Capitolo è mostrare come sia possibile calcolare in maniera esatta la distribuzione a posteriori mediante l'uso di una distribuzione a priori coniugata. Esamineremo qui il caso più semplice, ovvero quello Beta-Binomiale.

Per fare un esempio concreto, verranno qui considerati 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 risultino distorte negativamente. Nell'esempio considereremo 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 cinque giorni successivi. Lo studio prende in esame diverse emozioni, ma qui ci concentreremo unicamente sulla tristezza.

Utilizzando i dati forniti dagli autori, è stata calcolata 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.

Nel seguito, indicherò con θ la probabilità che le aspettative del partecipante siano distorte negativamente (ovvero, la probabilità dell'evento codificato con “1”). Il problema dell'esercizio è quello di ottenere la stima a posteriori di θ alla luce dei dati corrispondenti a 23 “successi” in 30 prove.¹

Nel caso presente, i dati (23 “successi” in 30 prove) possono essere considerati come realizzazioni indipendenti di una variabile casuale Bernoulliana. In tali circostanze, esiste una famiglia di distribuzioni che, qualora venga scelta per la distribuzione a priori, fa sì che la distribuzione a posteriori abbia la stessa forma funzionale della distribuzione a priori. Questo consente una soluzione analitica dell'integrale che compare nel denominatore bayesiano. Nel caso del modello binomiale, la famiglia di distribuzioni che ha questa proprietà è la distribuzione Beta.

¹Si 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 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 $p(\theta | y)$. Ovviamente, $p(\theta | y)$ assume forme molto diverse nei quattro casi descritti sopra.

1.3.1 Parametri della distribuzione Beta

Usando la distribuzione Beta è possibile esprimere diverse credenze iniziali rispetto a θ . Scegliere $a = 4$ e $b = 4$ per la distribuzione a priori Beta significa dire che, prima di avere esaminato i dati, una “aspettativa futura distorta negativamente” è tanto probabile quanto una “aspettativa futura non distorta negativamente”; tuttavia, ad una tale proposizione è associato un livello di certezza piuttosto basso. Possiamo esprimere la nostra incertezza calcolando, con un grado di fiducia del 95%, la regione nella quale, in base a tale ipotesi a priori, si trova il valore del parametro. Per ottenere tale intervallo di credibilità a priori, usiamo la funzione `qbeta()` di R. In `qbeta()` i parametri a e b sono chiamati `shape1` e `shape2`:

```
qbeta(c(0.025, 0.975), shape1 = 4, shape2 = 4)
#> [1] 0.1840516 0.8159484
```

Se scegliamo $a = 10$ e $b = 10$ supponiamo ancora a priori che una “aspettativa futura distorta negativamente” sia altrettanto probabile di una “aspettativa futura non distorta negativamente”, ma ora la nostra certezza a priori sul valore del parametro è maggiore, come indicato dall’intervallo al 95%:

```
qbeta(c(0.025, 0.975), shape1 = 10, shape2 = 10)
#> [1] 0.2886432 0.7113568
```

Quale distribuzione a priori dobbiamo scegliere? In un problema concreto di analisi dei dati, la scelta della distribuzione a priori dipende da quali sono le credenze a priori vogliamo includere nell’analisi dei dati. Se non abbiamo alcuna informazione a priori, potremmo usare $a = 1$ e $b = 1$, il che produce una distribuzione a priori uniforme. Questa, tuttavia, è una cattiva idea, se pensiamo al problema della riparametrizzazione. In tali circostanze è invece raccomandato usare una distribuzione a priori vagamente informativa, come $\text{Beta}(2, 2)$.

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

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

1.3.2 La specificazione della distribuzione a posteriori

Una volta scelta la distribuzione a priori di tipo Beta, i cui parametri rispecchiano il giudizio a priori di probabilità, la distribuzione a posteriori viene specificata dalla regola di Bayes:

$$\text{distribuzione a posteriori} = \frac{\text{verosimiglianza} \cdot \text{distribuzione a priori}}{\text{verosimiglianza marginale}}.$$

Nel caso presente abbiamo

$$p(\theta \mid n = 30, y = 23) = \frac{\left[\binom{30}{23} \theta^{23} (1 - \theta)^{30-23} \right] \left[\frac{\Gamma(12)}{\Gamma(2)\Gamma(10)} \theta^{2-1} (1 - \theta)^{10-1} \right]}{p(y = 23)},$$

laddove $p(y = 23)$ è $\int_0^1 p(y = 23 \mid n = 30, \theta)$. Questo termine, ovvero la verosimiglianza marginale, è una costante di normalizzazione che fa sì che l'area sottesa alla densità a posteriori sia unitaria.

Raccogliendo tutte le costanti dell'equazione precedente ottiamo

$$p(\theta \mid n = 30, y = 23) = \left[\frac{\binom{30}{23} \frac{\Gamma(12)}{\Gamma(2)\Gamma(10)}}{p(y = 23)} \right] \theta^{23} (1 - \theta)^7 \theta^1 (1 - \theta)^9.$$

Se ignoriamo il termine costante all'interno della parentesi quadra abbiamo:

$$\begin{aligned} p(\theta \mid n = 30, y = 23) &\propto \theta^{23} (1 - \theta)^7 \theta^1 (1 - \theta)^9 \\ &\propto \theta^{24} (1 - \theta)^{16}. \end{aligned}$$

Il termine di destra dell'equazione precedente corrisponde ad una distribuzione Beta di parametri $\alpha = 25$ e $\beta = 17$. Il risultato così ottenuto individua il *kernel* della distribuzione a posteriori, ovvero la distribuzione a posteriori non normalizzata. Infatti, l'area sotto la curva (AUC)

```
postFun <- function(theta) {
  theta^24 * (1 - theta)^16
}
(AUC <- integrate(postFun, lower = 0, upper = 1)$value)
#> [1] 3.880577e-13
```

non è 1. Per ottenere una corretta distribuzione di densità è necessario includere una costante di normalizzazione. In base alla definizione della distribuzione Beta, tale costante è

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{\Gamma(42)}{\Gamma(25)\Gamma(17)}.$$

Questo termine non è altro che l'integrale della distribuzione a posteriori non normalizzata che è stato calcolato in precedenza:

```
postFun <- function(theta) {
  theta^24 * (1 - theta)^16 / AUC
}
integrate(postFun, lower = 0, upper = 1)$value
#> [1] 1
```

Infatti, l'area sottesa alla funzione

$$p(\theta \mid n = 30, y = 23) = \frac{\Gamma(42)}{\Gamma(25)\Gamma(17)} \theta^{25-1} (1 - \theta)^{17-1} = \text{Beta}(y + a, n - y + b)$$

somma ad 1, ovvero,

```
postFunNor <- function(theta) {
  (factorial(42-1) / (factorial(25-1)*factorial(17-1))) *
  theta^(25-1) * (1 - theta)^(17-1)
}
(AUC2 <- integrate(postFunNor, lower = 0, upper = 1)$value)
#> [1] 1
```

Possiamo quindi riassumere dicendo che siamo partiti da una verosimiglianza $\text{Bin}(n = 30, y = 23 \mid \theta)$ e, moltiplicando la verosimiglianza per la distribuzione a priori $\theta \sim \text{Beta}(2, 10)$, abbiamo ottenuto la distribuzione a posteriori $p(\theta \mid n, y) \sim \text{Beta}(25, 17)$. Questo è un esempio di analisi coniugata: la distribuzione a posteriori del parametro ha la stessa forma funzionale della distribuzione a priori. La presente combinazione di verosimiglianza e distribuzione a priori è chiamata caso coniugato *Beta-Binomiale* ed è descritto dal seguente teorema.

Teorema 1.1. *Sia data la funzione di verosimiglianza $\text{Bin}(n, y \mid \theta)$ e sia $\text{Beta}(\alpha, \beta)$ una distribuzione a priori. In tali circostanze, la distribuzione a posteriori del parametro θ sarà una distribuzione $\text{Beta}(\alpha + y, \beta + n - y)$.*

In particolare, possiamo calcolare qual è il valore atteso a posteriori di θ . Essendo $\mathbb{E}[\text{Beta}(\alpha, \beta)] = \frac{\alpha}{\alpha + \beta}$, il valore atteso a posteriori di θ sarà

$$\mathbb{E}_{\text{post}}[\text{B}(\alpha + y, \beta + n - y)] = \frac{\alpha + y}{\alpha + \beta + n}. \quad (1.1)$$

Esempio 1.1. Per i dati esaminati, usando la 1.1 possiamo ottenere l'intervallo di credibilità a posteriori del 95% per il parametro θ come:

```
qbeta(c(0.025, 0.975), shape1 = 25, shape2 = 17)
#> [1] 0.4450478 0.7368320
```

La media della distribuzione a posteriori è

```
25 / (25 + 17)
#> [1] 0.5952381
```

La moda della distribuzione a posteriori è

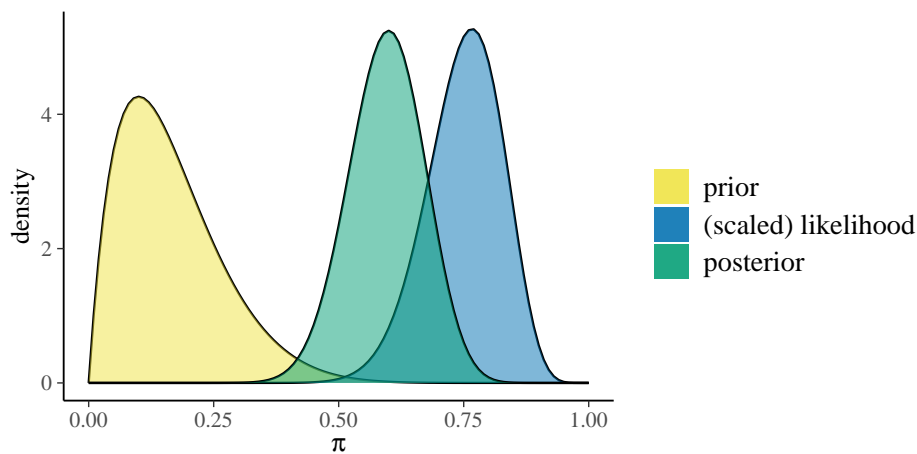

```
(25 - 1) / (25 + 17 - 2)
#> [1] 0.6
```

La deviazione standard della distribuzione a priori è

```
sqrt((25 * 17) / ((25 + 17)^2 * (25 + 17 + 1)))
#> [1] 0.0748533
```

Esempio 1.2. La funzione R `plot_beta_binomial()` del pacchetto `bayesrules` descrive in maniera grafica l'aggiornamento bayesiano Beta-Binomiale:

```
bayesrules::plot_beta_binomial(
  alpha = 2, beta = 10, y = 23, n = 30
)
```



Un sommario delle distribuzioni a priori e a posteriori si ottiene usando la funzione `summarize_beta_binomial()`:

```
bayesrules::summarize_beta_binomial(
  alpha = 2, beta = 10, y = 23, n = 30
)
#>      model alpha beta      mean mode      var      sd
#> 1   prior     2   10 0.166667  0.1 0.010683761 0.1033623
#> 2 posterior    25   17 0.5952381 0.6 0.005603016 0.0748533
```

Esempio 1.3. Consideriamo un altro esempio discusso da [Johnson et al. \(2022\)](#). Nel 1963, Stanley Milgram presentò una ricerca sulla propensione delle persone a obbedire agli ordini di figure di autorità, anche quando tali ordini possono danneggiare altre persone ([Milgram, 1963](#)). Nell'articolo, Milgram descrive lo studio come

consist[ing] of ordering a naive subject to administer electric shock to a victim. A simulated shock generator is used, with 30 clearly marked voltage levels that range from 15 to 450 volts. The instrument bears verbal designations that range from Slight Shock to Danger: Severe Shock. The responses of the victim, who is a trained confederate of the experimenter, are standardized. The orders to administer shocks are given to the naive subject in the context of a ‘learning experiment’ ostensibly set up to study the effects of punishment on memory. As the experiment proceeds the naive subject is commanded to administer increasingly more intense shocks to the victim, even to the point of reaching the level marked Danger: Severe Shock.

All’insaputa del partecipante, gli shock elettrici erano falsi e l’attore stava solo fingendo di provare il dolore dello shock.

Johnson et al. (2022) fanno inferenza sui risultati dello studio di Milgram mediante il modello Beta-Binomiale. Il parametro di interesse è θ , la probabilità che una persona obbedisca all’autorità (in questo caso, somministrando lo shock più severo), anche se ciò significa recare danno ad altri. Johnson et al. (2022) ipotizzano che, prima di raccogliere dati, le credenze di Milgram relative a θ possano essere rappresentate mediante una Beta(1, 10). Sia $y = 26$ il numero di soggetti che, sui 40 partecipanti allo studio, aveva accettato di infliggere lo shock più severo. Assumendo che ogni partecipante si comporti indipendentemente dagli altri, possiamo modellare la dipendenza di y da θ usando la distribuzione binomiale. Giungiamo dunque al seguente modello bayesiano Beta-Binomiale:

$$Y|\theta \sim \text{Bin}(n = 40, \theta)$$

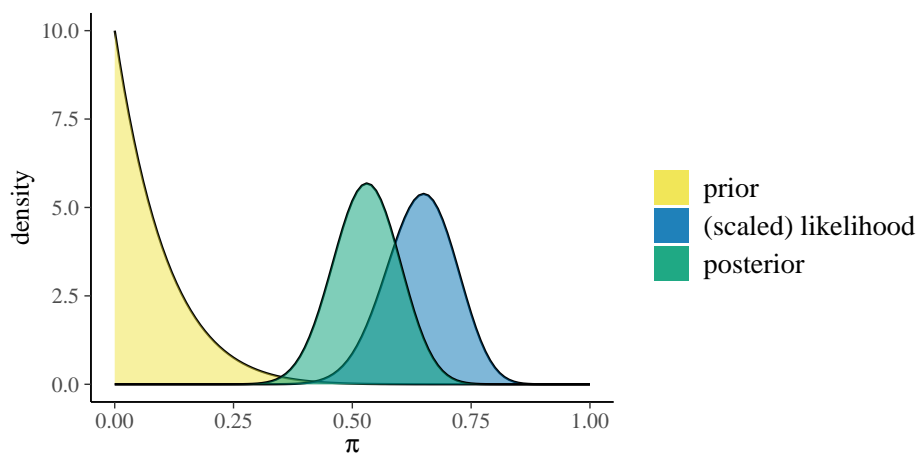
$$\theta \sim \text{Beta}(1, 10) .$$

Usando le funzioni di bayesrules possiamo facilmente calcolare i parametri e le proprietà della distribuzione a posteriori:

```
bayesrules::summarize_beta_binomial(
  alpha = 1, beta = 10, y = 26, n = 40
)
#>      model alpha beta      mean      mode      var
#> 1   prior      1   10 0.09090909 0.0000000 0.006887052
#> 2 posterior    27   24 0.52941176 0.5306122 0.004791057
#>      sd
#> 1 0.08298827
#> 2 0.06921746
```

Il processo di aggiornamento bayesiano è descritto dalla figura seguente:

```
bayesrules::plot_beta_binomial(
  alpha = 1, beta = 10, y = 26, n = 40
)
```



1.4 Principali distribuzioni coniugate

Esistono molte altre combinazioni simili di verosimiglianza e distribuzione a priori le quali producono una distribuzione a posteriori che ha la stessa densità della distribuzione a priori. Sono elencate qui sotto le più note coniugazioni tra modelli statistici e distribuzioni a priori.

- Per il modello Normale-Normale $\mathcal{N}(\mu, \sigma_0^2)$, la distribuzione iniziale è $\mathcal{N}(\mu_0, \tau^2)$ e la distribuzione finale è $\mathcal{N}\left(\frac{\mu_0\sigma^2 + \bar{y}n\tau^2}{\sigma^2 + n\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}\right)$.
- Per il modello Poisson-gamma $\text{Po}(\theta)$, la distribuzione iniziale è $\Gamma(\lambda, \delta)$ e la distribuzione finale è $\Gamma(\lambda + n\bar{y}, \delta + n)$.
- Per il modello esponenziale $\text{Exp}(\theta)$, la distribuzione iniziale è $\Gamma(\lambda, \delta)$ e la distribuzione finale è $\Gamma(\lambda + n, \delta + n\bar{y})$.
- Per il modello uniforme-Pareto $\text{U}(0, \theta)$, la distribuzione iniziale è $\text{Pa}(\alpha, \varepsilon)$ e la distribuzione finale è $\text{Pa}(\alpha + n, \max(y_{(n)}, \varepsilon))$.

Considerazioni conclusive

Lo scopo di questa discussione è stato quello di mostrare come sia possibile combinare le nostre conoscenze a priori (esprese nei termini di una densità di probabilità) con le evidenze fornite dai dati (esprese nei termini della funzione

di verosimiglianza), così da determinare, mediante il teorema di Bayes, una distribuzione a posteriori, la quale condensa l'incertezza che si ha sul parametro θ . Per illustrare tale problema, abbiamo considerato una situazione nella quale θ corrisponde alla probabilità di successo in una sequenza di prove Bernoulliane. Abbiamo visto come, in queste circostanze, è ragionevole esprimere le nostre credenze a priori mediante la densità Beta, con opportuni parametri. L'inferenza rispetto ad una proporzione rappresenta un caso particolare, ovvero un caso nel quale la distribuzione a priori è Beta e la verosimiglianza è Binomiale. In tali circostanze, la distribuzione a posteriori sarà una distribuzione Beta. Per questa ragione, in questo caso specifico, i parametri della distribuzione a posteriori possono essere determinati analiticamente.

Bibliografia

- Johnson, A. A., Ott, M., and Dogucu, M. (2022). *Bayes Rules! An Introduction to Bayesian Modeling with R*. CRC Press.
- Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4):371–378.
- 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.