STATISTICA BAYESIANA

LORENZO POMPILI

Sommario. Note molto semplici che esplorano la statistica Bayesiana da zero mentre l'autore la impara.

"Everybody is to some extent a Bayesian especially when using common sense."
I. J. Good, 1985

1. Esercizio di calcolo Bayesiano

Problema. Assumiamo di lanciare una moneta n volte e di ottenere testa x volte. Come facciamo a capire con una certa confidenza se la moneta è biased oppure no?

1.1. **Premessa: modello e dati.** Assumiamo, come è ragionevole che sia, che i lanci della moneta $X_0, X_1, \ldots, X_n$ siano indipendenti e identicamente distribuiti secondo una distribuzione Bernoulli

$$X_i \sim \text{Bernoulli}(\theta), \tag{1}$$

che vuol dire semplicemente che valgono 1 con probabilità θ e 0 altrimenti,

$$\mathbb{P}(X_i = 1) = \theta, \qquad \mathbb{P}(X_i = 0) = 1 - \theta.$$

La variabile X_i vale 1 se il risultato del lancio i-esimo è testa, 0 se è croce. La variabile aleatoria che conta il numero totale di teste in n lanci è la somma delle X_i e, facendo i conti, si dimostra che segue una distribuzione binomiale:

$$X := \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, \theta), \tag{Modello}$$

che è un modo compatto per scrivere con che probabilità X assume ciascun possibile valore k:

$$\mathbb{P}(X = k) = \begin{cases} \binom{n}{k}\theta^k(1-\theta)^{n-k} & \text{se } k \text{ è intero e } 0 \le k \le n, \\ 0 & \text{altrimenti.} \end{cases}$$

La legge di X scritta in (Modello) è il nostro *modello* (in realtà il modello vero è (1), ma è essenzialmente uguale perché non ci servirà mai sapere i risultati dei singoli lanci, solo il totale delle teste). È un modello parametrico, perché dipende dal valore di θ, il quale è sconosciuto.

Oltre al modello, abbiamo dei *dati* osservati. Il nostro unico dato è che è uscito testa x volte:

$$X = x. \tag{Dati}$$

1.2. **Approccio classico.** L'approccio classico è il più banale: si basa solamente su (Modello) e (Dati). In sintesi, si formulano delle ipotesi e le si compara con i dati, assumendo che valga il modello scelto[1]. Ipotesi nulla:

$$H_0: \quad \theta = 1/2.$$

Ipotesi alternativa:

$$H_1: \quad \theta \in [0,1],\ \theta \neq 1/2.$$

[1] Il modello deve essere ragionevole. Se il modello scelto fa schifo, anche ottimi p-value non rifletteranno dei buoni risultati.

Fissiamo la soglia $\alpha = 0.05$. Calcoliamo il p-value associato al risultato $X = x$ con n lanci, che facendo i conti si dimostra essere

$$p_{\text{value}} = \begin{cases} 1 & \text{se } n \text{ è pari e } x = \dfrac{n}{2} \\ 2 \displaystyle\sum_{y=0}^{\min\{x,n-x\}} \binom{n}{y} \theta^y (1-\theta)^{n-y} & \text{altrimenti.} \end{cases}$$

Se $p_{\text{value}} < \alpha$, rigettiamo l'ipotesi nulla. Se $p_{\text{value}} \geq \alpha$, ce lo tiriamo nei denti, perché non possiamo concludere nulla. Fine della storia. Non si capisce neanche tanto bene cosa c'entri H_1.

1.3. **Formule di base.** Prima di passare all'approccio Bayesiano, diamo le due formule principali che ci servono. Il *teorema di Bayes* afferma che dati due eventi A e B vale sempre la formula

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \tag{Bayes}$$

Questa è una conseguenza diretta della definizione di probabilità condizionata:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Inoltre, dato un evento A e un insieme di eventi disgiunti $B_1, ..., B_k$ tali che esauriscano tutti i casi possibili, $\mathbb{P}(B_1) + \cdots + \mathbb{P}(B_k) = 1$, possiamo sempre spezzare la probabilità di A nei vari casi possibili, il che si può scrivere in due modi diversi usando la definizione di probabilità condizionata:

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap B_1) + \cdots + \mathbb{P}(A \cap B_k) \\ &= \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \cdots + \mathbb{P}(A|B_k)\mathbb{P}(B_k). \end{aligned} \tag{Split}$$

1.4. **Approccio Bayesiano.** Nell'approccio Bayesiano non ci si accontenta più solo di (Modello) e (Dati). I "meta-parametri" da due diventano tre, perché si aggiunge il *prior*, una distribuzione di probabilità a priori sui parametri possibili del modello. Questa distribuzione la chiamiamo D_{prior} e rappresenta il nostro guess iniziale dei valori possibili dei parametri del modello, e della loro probabilità.

Esempio. Assumiamo di aver investigato un giro di contraffazione di monete così grande che girano tante monete false quante quelle vere. Abbiamo appena scoperto un'informazione cruciale per il caso, della quale nessun altro è al corrente: quelle false si riconoscono immediatamente perché sono sbilanciate, e la probabilità che esca testa è 1/4. Se passeggiando troviamo una moneta per terra, possiamo già sospettare che la moneta sia contraffatta con probabilità 1/2 con un parametro di bias 1/4, mentre invece con probabilità 1/2 la moneta è bilanciata, $\theta = 1/2$.

Il nostro modello si amplia in questo modo: il parametro θ è il risultato di una variabile aleatoria Θ che vale 1/4 e 1/2 con probabilità 1/2:

$$\theta = \Theta,$$
$$\mathbb{P}(\Theta = k) = \begin{cases} 1/2 & \text{se } k = 1/4 \text{ o } 1/2 \\ 0 & \text{altrimenti.} \end{cases}$$

Detta in modo sintetico, la distribuzione D_{prior} della variabile Θ è la distribuzione uniforme sull'insieme dei valori $\{1/4, 1/2\}$:

$$D_{prior} = \text{Unif}(\left\{\frac{1}{4}, \frac{1}{2}\right\}) \tag{Prior}$$

Questo che stiamo facendo non deve sorprendere: la probabilità serve a modellare l'incertezza, oltre che la casualità, quindi è naturale dire che il nostro parametro sconosciuto è rappresentato da una variabile aleatoria.

Vediamo come procedere. Noi ora facciamo l'esperimento, otteniamo che il numero di deste è x, cioè raccogliamo i (Dati), e calcoliamo una nuova distribuzione di probabilità per θ condizionata all'esito dell'esperimento $X = x$, cioè una probabilità *a posteriori*:

$$\Theta \,|\, \{X = x\} \quad \sim \quad D_{post},$$

dove D_{post} è la distribuzione che ci interessa calcolare. Per farlo, basta applicare il teorema di Bayes. Per semplificare, usiamo le seguenti notazioni:

$$\begin{aligned} \mathbb{P}(X = x) &=: p(x), \\ \mathbb{P}(X = x \,|\, \Theta = \theta) &=: p(x|\theta), \\ \mathbb{P}(\Theta = \theta) &=: p(\theta), \\ \mathbb{P}(\Theta = \theta \,|\, X = x) &=: p(\theta|x). \end{aligned}$$

Applicando (Bayes) con $A = \{\Theta = \theta\}$, $B = \{X = x\}$, otteniamo

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}. \tag{2}$$

La quantità a sinistra dell'equazione è quella che vogliamo calcolare, perché descrivere la probabilità che Θ dia come risultato un certo numero θ condizionata al dato $X = x$. In particolare, ci interessa $p(1/2|x)$ per sapere con che probabilità la moneta è bilanciata.

Benissimo, ora noi abbiamo i nostri (Dati) (o meglio, il nostro dato) x, abbiamo un (Modello) che ci dà l'espressione esplicita di $p(x|\theta)$:

$$p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{1-x},$$

e abbiamo un (Prior) che ci dà l'espressione di $p(\theta)$:

$$p(\theta) = \begin{cases} 1 & \theta \in [0,1] \\ 0 & \text{altrimenti.} \end{cases}$$

Noi vogliamo calcolare $p(\theta|x)$, che ci darà la densità di probabilità di Θ condizionata all'evento $X = x$. Per farlo, manca sapere $p(x)$. Come si fa? Al liceo abbiamo imparato che con un'equazione si può trovare una sola incognita, ma noi ne abbiamo due: $p(\theta|x)$ e $p(x)$. Come si fa? Fortunatamente abbiamo l'equazione (Split), che ci dice che la probabilità $\mathbb{P}(X = x)$ può essere spezzata nella somma delle probabilità di tutti i casi possibili $\{X = x\} \cap \{\Theta = \theta\}$ per tutti i possibili valori di θ. In formule,

$$p(x) = p(x|1/4)p(1/4) + p(x|1/2)p(1/2).$$

Non ci interessa calcolarlo, ma basta sapere che possiamo ricavarcelo da $p(x|\theta)$ e $p(\theta)$, quindi dal (Modello) e dal (Prior). Quindi ora abbiamo anche $p(x)$. Mettendo tutto insieme,

$$p(1/2 \,|\, x) = \frac{p(x|1/2)p(1/2)}{p(x|1/4)p(1/4) + p(x|1/2)p(1/2)}.$$

Per $n = 10$, $x = 2$, la probabilità che la moneta sia bilanciata dopo aver effettuato l'esperimento è circa

$$p(1/2 \,|\, 2) = 0.135$$

Se in 10 lanci hai fatto due teste, $\theta = 1/4$ diventa molto più probabile di $\theta = 1/2$ se inizialmente le ritenevi ugualmente probabili.

Attenzione. Se invece il tuo prior favoriva molto il caso $\theta = 1/2$, perché il sospetto iniziale che la moneta fosse bilanciata era alto, la probabilità a posteriori sarebbe stata più alta. Per esempio, una moneta trovata a caso per strada è molto improbabile che sia biased, quindi avremo un prior diverso, per esempio,

$$\mathbb{P}(\Theta = 1/2) = 0.99, \qquad \mathbb{P}(\Theta \neq 1/2) = 0.01,$$

e il calcolo finale darà come risultato

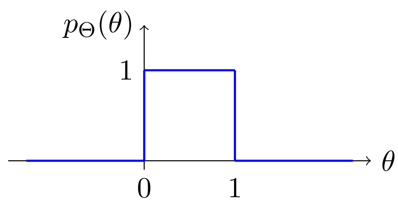
$$p(1/2 \,|\, 2) = 0.9392.$$

Il solo fatto di ottenere 2 teste su 10 non rende l'ipotesi della moneta bilanciata impossibile, se l'alternativa era estremamente improbabile in partenza.

1.5. **Approccio Bayesiano - Caso continuo.** Altro esempio, possiamo pensare di assegnare una probabilità uniforme sul parametro $\theta \in [0,1]$ prima di fare l'esperimento e contare il numero delle teste:

$$\theta = \Theta \quad \sim \quad D_{prior} = \mathrm{Unif}([0,1]). \tag{Prior$_2$}$$

L'equazione sopra vuol dire: "Il parametro θ del nostro modello, che è un numero che noi non conosciamo, è il risultato di una variabile aleatoria Θ, la quale ha distribuzione D_{prior}. E nel nostro caso, questa distribuzione D_{prior} è quella uniforme su $[0,1]$". La probabilità uniforme sull'intervallo $[0,1]$ è rappresentata tramite una *funzione di densità* p_Θ che vale 1 nell'intervallo $[0,1]$ e 0 altrimenti:



La scelta della distribuzione uniforme non è scontata e dovrebbe essere motivata (perché non si dà preferenza a un valore di θ rispetto a un altro?).

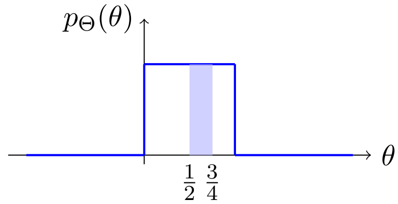
Vediamo come procedere (copio-incollo dall'esempio precedente cambiando solo quello che è necessario). Rifacciamo l'esperimento, otteniamo che il numero di deste è x, cioè raccogliamo i (Dati), e calcoliamo una nuova distribuzione di probabilità per θ condizionata all'esito dell'esperimento $X = x$, cioè una probabilità *a posteriori*:

$$\Theta \,|\, \{X = x\} \quad \sim \quad D_{post},$$

dove D_{post} è la distribuzione che ci interessa calcolare. Per farlo, basta applicare il teorema di Bayes. Per semplificare, diamo i seguenti nomi:

$$\begin{aligned} \mathbb{P}(X = x) &=: p(x), \\ \mathbb{P}(X = x \,|\, \Theta = \theta) &=: p(x|\theta), \\ p_\Theta(\theta) &=: p(\theta), \\ p_\Theta(\theta \,|\, X = x) &=: p(\theta|x). \end{aligned}$$

Qui sopra c'è una parte che meriterebbe un trattamento tecnico e non possiamo approfondire, ma bisogna menzionare velocemente per chiarirci le idee. Siccome la variabile Θ è continua, non discreta, al posto della probabilità $\mathbb{P}(\Theta = \theta)$ usiamo la *funzione di densità di probabilità*, che come detto sopra abbiamo denotato con p_Θ. $p_\Theta(\theta)$ può essere vista come una sorta di probabilità che $\Theta = \theta$ ma "rinormalizzata", poiché essendo la variabile Θ continua e quindi capace di assumere infiniti valori, ogni particolare valore θ è così improbabile che ha probabilità nulla di verificarsi. La probabilità degli eventi che riguardano Θ si calcola a partire da questa funzione p_Θ. Per esempio, la probabilità che $1/2 < \Theta < 3/4$ è l'area sotto la funzione p_Θ nell'intervallo $[1/2, 3/4]$:



che si può scrivere sotto forma di integrale:

$$\mathbb{P}(1/2 < \Theta < 3/4) = \int_{1/2}^{3/4} p_\Theta(\theta)\, d\theta.$$

Nel nostro caso, p_Θ è una funzione costante uguale a 1 nell'intervallo $[0,1]$ e zero al di fuori, così che l'area sotto il suo grafico compresa tra 1/2 e 3/4 è esattamente 1/4. L'area totale sotto il grafico di p_Θ è ovviamente 1, perché Θ assume valori in $[0,1]$ con probabilità 1.

Fatto salvo questa parte, il teorema di (Bayes) applicato al nostro caso ci dà[2], usando le notazioni di cui sopra,

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}. \tag{3}$$

La quantità a sinistra dell'equazione è quella che vogliamo calcolare, perché descriverà la distribuzione di probabilità di Θ condizionata a $X = x$.

Benissimo, ora noi abbiamo i nostri (Dati) (o meglio, il nostro dato) x, abbiamo un (Modello) che ci dà l'espressione esplicita di $p(x|\theta)$ (si veda la definizione di distribuzione binomiale):

$$p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{1-x},$$

e abbiamo un (Prior$_2$) che ci dà l'espressione di $p(\theta)$:

$$p(\theta) = \begin{cases} 1 & \theta \in [0,1] \\ 0 & \text{altrimenti.} \end{cases}$$

Vogliamo calcolare $p(\theta|x)$, che ci darà la densità di probabilità di Θ condizionata all'evento $X = x$. Per farlo, manca sapere $p(x)$. Come si fa? Come prima, la probabilità $\mathbb{P}(X = x)$ è uguale alla somma delle probabilità di tutti i casi possibili $\{X = x\} \cap \{\Theta = \theta\}$ per tutti i $\theta \in [0,1]$. In formule,

$$p(x) = \int_0^1 p(x|\theta)p(\theta).$$

Non ci interessa calcolarlo, ma basta sapere che possiamo ricavarcelo da $p(x|\theta)$ e $p(\theta)$, quindi dal (Modello) e dal (Prior$_2$). Quindi ora abbiamo anche $p(x)$. Nel nostro caso quell'integrale vale, incredibilmente,

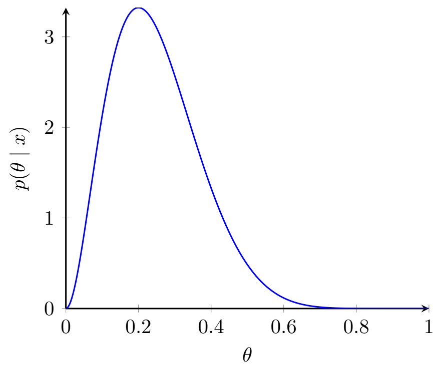
$$p(x) = 1/(n+1).$$

La densità di probabilità a posteriori per il nostro parametro θ, mettendo tutto dentro (3), è quindi

$$p(\theta|x) = (n+1)\binom{n}{x} \theta^x (1-\theta)^{n-x}.$$

Questo è il grafico per $n = 10$ e $x = 2$:

[2]Il fatto che Θ sia continua complica le cose anche per applicare il teorema di Bayes, ma esistono versioni del teorema che si applicano al caso di variabili aleatorie continue.



La densità $p(\theta|x)$ ci dà ““““davvero”””” la (densità di) probabilità dei valori che il parametro del (Modello) θ può assumere. Quindi ora possiamo fare tutti i test che vogliamo, per esempio l'area della curva tra 0 e 0.5 ci dà la probabilità che θ sia minore di 0.5. Se questa è maggiore di 0.95, possiamo scartare l'ipotesi che θ sia più grande o uguale a 0.5. O meglio ancora, possiamo costruire un intervallo di confidenza guardando al 5 e al 95-esimo percentile della distribuzione qui sopra.

Le virgolette qui sopra, va da sé, ci sono perché tutto ciò si basa su quanto è accurato il (Prior_2), oltre che il (Modello) stesso. Quindi sta ancora alla persona che conduce l'esperimento capire. La differenza con l'approccio classico, tuttavia, è chiara: la "probabilità" che la tua ipotesi sia vera non dipende *solamente* dai dati che osservi, ma anche da quanto è buona e verosimile la tua ipotesi in partenza. Più un'ipotesi è inverosimile, più avrai bisogno di dati per essere confidente che sia vera. La statistica Bayesiana tiene in considerazione questa cosa a un livello molto esplicito, quindi più facilmente verificabile: assumendo che i conti siano giusti, mi basta vedere se il tuo prior è ragionevole o meno (e se il modello è fatto bene) per validare lo standard del tuo articolo. Nell'approccio classico queste considerazioni vanno (o andrebbero) fatte lo stesso, ma vengono fatte in modo qualitativo e meno esplicito.

1.6. **Approccio Bayesiano - Reloaded (magagne).** Ora è tutto molto bello quando hai un modello e un prior, ma a volte questo non è possibile. Per esempio, il tuo prior può essere: "Con una certa probabilità una certa variabile Z è normale, altrimenti no". Quell'"altrimenti no" è impossibile da modellare. Si dice che il modello è *incompleto*, perché non è definito in un insieme di casi con una probabilità maggiore di 0. Nella pratica, questo comporta che anche se l'equazione (2) rimane valida, non è più possibile calcolare $p(x)$ come prima usando (Split), perché $p(x \mid \Omega)$ non si può calcolare per l'insieme dei casi Ω dove il modello non è specificato. Servirà per forza di cose stimare $p(x)$. Qui, se si hanno abbastanza dati, se ne usa una parte per approssimare la distribuzione $p(x)$, e il resto di essi per il calcolo della probabilità a posteriori.

Detto ciò, uno si può dire: "Vabè, ma anche se il modello non è completo, moralmente è chiaro che ottenere 2 teste su 10 rende l'ipotesi della moneta bilanciata *quantomeno* meno probabile di prima". E' un ragionamento giusto? Vediamo.

1.6.1. *Analisi matematica.* Supponiamo di avere di nuovo un caso discreto come in 1.4, ma stavolta abbiamo $\theta = 1/2$ con probabilità $1/2$, mentre con probabilità $1/2$ la moneta è biased ma il parametro di bias è dato da un fenomeno che non sappiamo modellizzare. Quest'ultimo evento lo chiamiamo Ω, che rappresenta una perdita catastrofica di informazione per cui non è possibile neanche avere una stima a priori sul possibile bias della moneta.

Applichiamo di nuovo, formalmente, la formula di (Bayes) insieme alla formula di decomposizione (Split) e si ottiene:

$$p(1/2 \mid x) = \frac{p(x \mid 1/2)p(1/2)}{p(x \mid 1/2)p(1/2) + \mathbb{P}(x \mid \Omega)\mathbb{P}(\Omega)}.$$

Deve valere per forza $0 \leq \mathbb{P}(x \,|\, \Omega) \leq 1$, che applicato all'equazione di sopra porta a

$$\frac{p(x \,|\, 1/2)p(1/2)}{p(x \,|\, 1/2)p(1/2) + \mathbb{P}(\Omega)} \leq p(1/2 \,|\, x) \leq 1. \tag{4}$$

Quindi non sappiamo quanto vale $p(1/2 \,|\, x)$, ma abbiamo una stima dal basso di questa probabilità. Questo ha senso: se la probabilità che $\theta = 1/2$ è relativamente grande e il valore x ottenuto è verosimile per il caso $\theta = 1/2$, non c'è motivo di ridurre di molto la probabilità di $\theta = 1/2$, a prescindere da quello che succede nel caso non modellizzato. Purtroppo non si può andare oltre questo che abbiamo appena detto, almeno in un modo direttamente matematico. Come già detto sopra, non è possibile dire nulla su $\mathbb{P}(x \,|\, \Omega)$ (a parte che sia compreso tra 0 e 1), quindi la stima (4) è tutto ciò che possiamo dire. In particolare, se lanci una moneta 10 volte ottenendo 2 teste non è mai teoricamente possibile abbassare la probabilità che la moneta sia biased, senza assumere nient'altro sulla tua conoscenza a priori del fenomeno. Ok, ma perché?

1.6.2. *Morale filosofica.* Il motivo per cui quello che abbiamo detto sopra fallisce è presto detto, e si vede benissimo il carico filosofico, quasi magico della formula di Bayes.

Teorema 1.1. *Se vuoi giustificare perché una certa cosa A diventa meno probabile in base ai nuovi dati x che hai ottenuto (cioè $\mathbb{P}(A \,|\, x) < \mathbb{P}(A)$), devi ammettere per forza l'esistenza di una qualche alternativa B che spiega meglio i dati ($p(x \,|\, B) > p(x \,|\, A)$) e che abbia una certa probabilità non nulla di verificarsi ($\mathbb{P}(B) > 0$).*

Infatti[3], in questo caso, chiamando $C := (A \cup B)^c$ l'insieme dei casi possibili rimanenti ($\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) = 1$), si ha:

$$\begin{aligned}\mathbb{P}(A \,|\, x) &= \frac{\mathbb{P}(x \,|\, A)\mathbb{P}(A)}{\mathbb{P}(x \,|\, A)\mathbb{P}(A) + \mathbb{P}(x \,|\, B)\mathbb{P}(B) + \mathbb{P}(x \,|\, (C)\mathbb{P}(C)} \\ &=: \alpha \mathbb{P}(A),\end{aligned}$$

dove

$$\alpha = \frac{1}{\mathbb{P}(A) + \frac{\mathbb{P}(x \,|\, B)}{\mathbb{P}(x \,|\, A)}\mathbb{P}(B) + \frac{\mathbb{P}(x \,|\, C)}{\mathbb{P}(x \,|\, A)}(1 - \mathbb{P}(A) - \mathbb{P}(B))}.$$

Si vede subito che il coefficiente α è minore di 1 nel caso A e B siano le uniche due alternative ($1 - \mathbb{P}(A) - \mathbb{P}(B) = 0$) perché in quel caso

$$\alpha = \frac{1}{\mathbb{P}(A) + \frac{\mathbb{P}(x \,|\, B)}{\mathbb{P}(x \,|\, A)}\mathbb{P}(B)},$$

e abbiamo detto che $p(x \,|\, B) > p(x \,|\, A)$. Questo ci dice che la probabilità di A è diminuita passando da quella a priori a quella a posteriori[4].

Questo tipo di informazione discussa nel teorema 1.1 va incluso nel prior, anche se parziale, altrimenti non è possibile stimare dall'alto $p(\theta \,|\, x)$. Come chiamiamo questa B...? Eheh... Questa, signori, è esattamente *l'ipotesi alternativa*, e il teorema 1.1 è il motivo per cui serve sempre un'ipotesi alternativa quando si vuole mostrare evidenza contro una certa ipotesi. Questo concetto è noto e discusso nella letteratura della filosofia Bayesiana: un'ipotesi può perdere probabilità solo perché c'è un insieme di ipotesi alternative che spiegano meglio i dati. Quindi, il problema nell'ultimo esempio che abbiamo fatto, è che anche se i dati non sono poi così buoni, non abbiamo un'ipotesi alternativa *verosmile* che spieghi meglio i dati, nel senso che anche se teoricamente non abbiamo escluso che θ

[3]Quella che segue non è una dimostrazione, ma fa capire qual'è il meccanismo.

[4]Nel caso generale in cui A e B non sono le uniche due alternative, allora α potrebbe effettivamente essere maggiore di 1, perché magari i nuovi dati escludono il caso C, rendendo A e B entrambi più probabili. Se C è meno probabile di A ci sono quindi due effetti, B ruba della probabilità ad A e C, mentre A la ruba a C. Il calcolo di α rivela quale dei due effetti è maggiore nel calcolo di $\mathbb{P}(A \,|\, x)$.

potrebbe assumere altri valori più verosimili come $\theta = 1/5$, non abbiamo esplicitato *quanto* questi casi siano probabili in partenza. In un certo senso, fare i conti con la statistica Bayesiana ti forza a fare le cose bene, perché non le basta che tu dica che esiste un'ipotesi alternativa, ma richiede anche che tu dica *perché* e *quanto* quell'ipotesi alternativa è verosimile, costringendoti a motivarla sul serio sulla base di altri dati o di altre considerazioni sul fenomeno in esame.

Nell'esempio 1.4, l'ipotesi nulla era $\theta = 1/2$ e l'ipotesi alternativa era $\theta = 1/4$, e infatti abbiamo dovuto esplicitare quanto l'ipotesi alternativa fosse probabile in partenza. Nell'esempio 1.5, tutti i possibili valori di θ sono ipotesi alternative le une alle altre, e vediamo che passando da D_{prior} a D_{post} (confrontando i grafici di $p(\theta)$ e $p(\theta \,|\, x = 2)$), alcune di queste sono diventate più probabili e altre meno. Volendo, avremmo potuto partire già dall'inizio con un'ipotesi nulla e un'ipotesi alternativa in modo esplicito.