Università degli Studi di Roma Tre

DIPARTIMENTO DI MATEMATICA E FISICA
Corso di Laurea in Fisica

Statistica Bayesiana e sue applicazioni

Laureando:
Lorenzo Carducci

Relatore:
Vittorio Lubicz

Anno Accademico
2015/2016

Alla mia famiglia che ha sempre creduto in me,
e a Cassandra, senza la quale tutto sarebbe meno bello.

Indice

Introduzione

Le espressioni statistica frequentista e statistica bayesiana si riferiscono a due modi diversi di intendere la teoria della probabilità e, di conseguenza, di affrontare l'inferenza statistica. Nella statistica frequentista si assume che il concetto di probabilità sia strettamente legato a quello di frequenza (relativa). Più precisamente, secondo questo approccio si può parlare di probabilità soltanto con riferimento agli esiti aleatori di esperimenti ripetuti nelle stesse condizioni. Nell'approccio bayesiano invece, il concetto di probabilità è semplicemente legato al suo significato intuitivo e all'etimologia dell'aggettivo probabile, ovvero alla plausibilità che eventi dall'esito incerto possano accadere, o che delle proposizioni possano risultare vere. Questo concetto di probabilità è anche detto soggettivo in quanto diverse persone possono avere un diverso stato di informazione e quindi è più che naturale che esprimano diverse valutazioni di probabilità. Come si può facilmente intuire, questo approccio è di più ampia applicazione di quello legato al limite delle frequenze relative poiché non è sottoposto a nessuna restrizione. Nell'approccio bayesiano il concetto di probabilità può essere applicato sia alle osservazioni che alle ipotesi. In particolare, la cosiddetta "inversione di probabilità", che permette di valutare la probabilità condizionate delle varie ipotesi a partire dalle probabilità condizionate delle varie osservazioni, è basata sul teorema di Bayes, da cui il nome all'approccio. Nell'approccio frequentista, invece, è proibito parlare di probabilità di ipotesi, di valori veri, di parametri di un modello o di una popolazione. E' importante sottolineare che l'approccio Bayesiano, sebbene possa sembrare nella maggior parte dei casi poco familiare soprattutto perché nei corsi di studio solitamente si predilige un'impostazione frequentista, è in realtà più esauriente, preciso e generale; può essere applicato a una più vasta gamma di problemi proprio perché generalizza il concetto di probabilità come ”grado di fiducia” di un avvenimento qualsiasi. Ad esempio non avrebbe nessun senso dal punto di vista frequentista chiedersi quali siano le condizioni meteorologiche di un dato giorno perché chiaramente la definizione di probabilità come "casi favorevoli su casi possibili" non è applicabile a questo problema; dal punto di vista dell'approccio Bayesiano invece ci si chiede quale sia il grado di fiducia che si ha nei confronti del verificarsi di una certa condizione meteorologica se si è in possesso di certe informazioni. E poiché riguardo ad una qualsiasi ipotesi siamo sempre a conoscenza di una serie (più o meno vasta e più o meno esauriente) di informazioni, possiamo sempre quantificare il grado di fiducia dell'ipotesi stessa. Nel corso della tesi verranno illustrati più a fondo i concetti della statistica bayesiana, facendo confronti con la statistica frequentista e utilizzando alcuni esempi per osservare in maniera più concreta il modus operandi dell'approccio bayesiano.

Nell'affrontare l'argomento ho fatto riferimento soprattutto ai testi di Giulio D'Agostini, [1] ,[2], [3] .

Capitolo 1

Il concetto di probabilità: approccio frequentista e approccio bayesiano

1.1 Statistica frequentista e sue definizioni

L'origine della parola "probabile" (dal latino probàbilis, da probàre, provare) sta ad indicare "degno di approvazione", "verosimile", "accettabile", "credibile", "ammissibile in base ad argomenti abbastanza sicuri". Di fronte a diverse affermazioni si può parlare di alcune più probabili e di altre meno probabili, a seconda della loro plausibilità. Il termine probabilità viene quindi usato come misura del grado di plausibilità di una affermazione, ovvero del "verificarsi di un certo evento". Si definisce "evento" una qualsiasi affermazione della quale sia verificabile il contenuto. Se l'evento è ben definito può essere dal punto di vista logico vero o falso. Consideriamo tre esempi:

- Il 31 Dicembre 2016 a Roma pioverà.
- Il prossimo numero estratto sulla ruota di Roma sarà il 53.
- Avendo letto su una bilancia 75,5 Kg il mio peso sarà compreso tra 75,4 e 75,6 Kg se la sensibilità del mio strumento è di 0,1 Kg.

In tutti questi eventi ci troviamo in una condizione di incertezza e ad ognuno di questi possiamo associare un *grado di fiducia* a seconda delle conoscenze che si hanno sull'evento stesso. Approfondiremo più avanti che questo grado di fiducia è strettamente collegato alla conoscenza soggettiva che si ha circa il verificarsi o meno dell'evento: per esempio posso essere a conoscenza della provenienza della mia bilancia e quindi attribuire alla misura un grado di fiducia più o meno elevato di una persona che non sa dove la bilancia è stata comprata. Questo concetto di *probabilità* soggettiva sarà fondamentale nel seguito della trattazione.

Analizziamo ora brevemente l'approccio frequentista alla statistica. Assumiamo che l'esito di un possibile esperimento consista nel verificarsi di uno tra i possibili risultati *elementari*, dove con *elementari* si intende che i risultati:

- non possono essere classificati ulteriormente in base a qualche altra caratteristica che possiedono alcuni e non altri;
- sono tutti i possibili per un particolare esperimento;
- sono a due a due incompatibili, ovvero che il verificarsi dell'uno esclude il verificarsi di tutti gli altri.

L'esempio più naturale in questo senso è il lancio di un dado: se il dado è perfettamente regolare, (domanda lecita può essere: "Possiamo essere sicuri della perfetta regolarità del dado?") abbiamo la stessa probabilità per ogni sua faccia. La stima di probabilità più naturale in questo caso è quella detta "definizione di Laplace":

Il valore della probabilità di un dato evento è pari al rapporto tra il numero dei casi favorevoli all'evento e il numero di casi possibili, se questi sono egualmente probabili:

$$P = \frac{numero\ dei\ casi\ favorevoli}{numero\ dei\ casi\ egualmente\ possibili} \tag{1.1}$$

e quindi ogni faccia avrà ovviamente probabilità $P = 1/6$. Questa definizione di probabilità, come vedremo più approfonditamente, contiene già al suo interno un problema: come si può definire la probabilità utilizzando il concetto di equiprobabilità?

Facciamo un secondo esempio: vogliamo valutare la probabilità di sopravvivenza nell'arco di un anno delle persone di una certa classe di età. Assumiamo che il 1° gennaio di un certo anno la popolazione sia costituita da n individui di una certa età e che il 1° gennaio dell'anno successivo siano ancora in vita n_f di quegli n. Questo è equivalente ad aver compiuto n prove delle quali n_f hanno dato esito favorevole e n-n_f esito sfavorevole. Ciascuna prova è inoltre considerata indipendente dall'altra e la probabilità di esito favorevole è ritenuta costante. In generale se la valutazione di probabilità è corretta, l'esito a cui è assegnata la probabilità più alta è quello che si ritiene possa accadere più facilmente. Supponendo che il processo si ripeta nel futuro con le stesse condizioni con cui era avvenuto nel passato si può pensare di valutare la probabilità futura $P(F)$ stimandola come proporzionale a quella degli anni precedenti. Segue che $P(F) = n_{\mathrm{f}}/n$. Secondo un ragionamento di questo tipo la frequenza relativa di successo nel passato viene adottata a misura della probabilità di successo nel futuro. Estendendo il semplice problema dell'esperimento con due esiti ad esperimenti a molti esiti, come potrebbe essere ad esempio quello di n misure nelle quali il risultato si può presentare in tante modalità diverse, otteniamo:

$$P(E_{\mathrm{i}}) \approx \frac{n(E_{\mathrm{i}})}{n} \tag{1.2}$$

dove E_i è l'evento di cui si vuole stimare la probabilità ed n il numero di prove. In questo modo la probabilità viene valutata come *frequenza relativa* di successo.
È naturale fare queste due osservazioni riguardo a questo metodo di valutazione della probabilità:

- il segno " $\approx$ "anziché il segno " $=$ " segnala che questa valutazione è affetta da un errore più grande quanto meno il numero n delle prove è elevato;

- il metodo presuppone che gli esperimenti (passati, presenti e futuri) possano essere ripetuti nelle stesse condizioni.

Oltre a questi esempi facilmente riportabili al quotidiano si può considerare il problema da un punto di vista più scientifico. Ogni qualvolta si vuole trovare il valore di una certa grandezza fisica o stabilire quale teoria descriva meglio i dati ci si basa su misurazioni soggette ad errori, motivo per cui si devono associare alle conclusioni delle incertezze che tengano conto in maniera coerente di questi errori. Secondo l'approccio frequentista il valore vero di una misura è il valore medio ottenuto dopo infinite misure sotto le stesse condizioni con strumenti che non commettono errori sistematici. A questo valore (ottenuto dopo un numero sufficientemente grande di misure) viene associata un'incertezza.
Nella valutazione statistica sulle incertezze associate alle grandezze spesso può essere utile utilizzare il concetto di "intervallo di confidenza", cioè la probabilità che la grandezza si trovi effettivamente in un certo intervallo stabilito. È di fondamentale importanza però ragionare su che tipo di informazione ci può dare il concetto di intervallo di confidenza se stiamo seguendo un approccio frequentista nella nostra valutazione statistica. Consideriamo n misure indipendenti di una stessa quantità fisica sotto le stesse condizioni (con $n \longrightarrow \infty$) nel caso in cui si suppone che le misure siano affette da errori casuali. Quello che si ottiene dai dati sperimentali e da ragionamenti statistici è la seguente relazione:

$$\mu = \bar{x} \pm \frac{\sigma}{\sqrt{n}} \tag{1.3}$$

con μ valore aspettato, $\bar{x}$ valore medio, σ deviazione standard. Una comune e comoda, ma dal punto di vista frequentista, errata interpretazione di questa espressione e del concetto di intervallo di confidenza è la seguente:

$$P(\bar{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}}) = 68\%. \tag{1.4}$$

In realtà la statistica frequentista può solo dire che:

$$P(\mu - \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + \frac{\sigma}{\sqrt{n}}) = 68\% \tag{1.5}$$

cioè fare una dichiarazione probabilistica sul valor medio. Non si possono infatti accettare in un approccio frequentista stime di probabilità su μ perché essa non è una variabile casuale a cui si può associare una certa probabilità ma un solo valore preciso che si otterrebbe sotto le condizioni di perfetta e infinita ripetibilità, assenza di errori sistematici ed eseguendo un numero $n \to \infty$ di misure. Poiché però quando si effettuano esperimenti di questo tipo l'interesse è quello di aumentare la nostra conoscenza del valore vero μ dell'osservabile, è naturale parlare di intervalli di confidenza del valore vero cioè della probabilità che μ sia compreso all'interno di un certo intervallo, cosa come vedremo consentita in un approccio bayesiano ma ovviamente priva di senso in ambito frequentista, dove non ha senso associare una probabilità al valore vero μ.

Capitolo 2

Probabilità come grado di fiducia

2.1 Problemi dell'approccio frequentista e definizione della probabilità come grado di fiducia

Un'analisi più attenta mostra che le definizioni di probabilità frequentiste presentano qualche lacuna di costruzione. Vediamo quali:

- Nella definizione come "rapporto tra casi favorevoli e casi possibili" è chiaro che occorre aggiungere la fondamentale condizione che i casi possibili siano tutti *equiprobabili* tra di loro. Ciò significa che stiamo usando il concetto di *probabilità* per definire la *probabilità* stessa.
- Nella definizione come "rapporto tra numero di volte che un evento si è verificato su un numero totale di prove", bisogna aggiungere la condizione che il numero di prove sia infinito. Inoltre con questa definizione si utilizza la *frequenza relativa* con cui un evento si è verificato nel passato per valutare la *probabilità* che esso possa verificarsi nuovamente nel futuro e per poterlo fare si deve assumere che il fenomeno sia avvenuto nel passato e avverrà nel futuro con la stessa *probabilità*; ma nessuno può assicurare questo e comunque anche in questo caso si sta riutilizzando il concetto di *probabilità* per definire la *probabilità stessa.*
- Parlare di esperimenti privi di errori sistematici è un'approssimazione che può essere più o meno lecita. Oltre al fatto che non possono esistere strumenti completamente immuni da errori sistematici, anche le condizioni di ripetibilità possono dipendere in misura più o meno grande da una vasta quantità di fattori come condizioni ambientali o interferenze dello sperimentatore.

Per generalizzare il concetto di probabilità dobbiamo quindi trovare una definizione alternativa, più generale e soddisfacente. Si è quindi condotti a pensare alla probabilità come al grado di fiducia che soggettivamente si associa ad un evento:

La probabilità è la misura del grado di fiducia che un evento si verifichi.

Sebbene questa sembri una definizione vaga, essa è molto attinente al concetto primitivo di probabilità ed ha una serie di importanti vantaggi:

- È *naturale*, *generale* e può essere applicata ad ogni evento pensabile indipendentemente dalla vastità e qualità dei possibili eventi o senza ripetere esperimenti sotto le stesse condizioni di probabilità.
- Ci permette, per quanto riguarda le misure, di parlare di *probabilità del valore vero* di una quantità fisica, mentre nell'approccio frequentista è possibile solo considerare il valore vero come una costante.
- Rende possibile elaborare una teoria *generale* dell'incertezza che tiene conto di ogni errore statistico o sistematico, indipendentemente dalla loro distribuzione.

2.2 Probabilità come scommessa

Si può pensare il concetto di probabilità come grado di fiducia in questo modo: più alto è il grado di fiducia che un evento accada, più alta è la somma in denaro A che si è disposti a scommettere sull'avvenimento dell'evento E per ricevere come vincita la somma in denaro S se l'evento si verifica. La scommessa deve essere accettabile in tutte e due le direzioni. È chiaro che la somma A che un giocatore assennato scommetterebbe aumenta con il grado di fiducia ed è proporzionale alla potenziale vincita S. Dunque se qualcuno ritiene che la probabilità di un evento E sia $P(E)$, costui sarà pronto a scommettere $A = P(E) \cdot S$ per vincere S se l'evento E si verifica e per perdere A se non si verifica.

Attraverso un'impostazione di questo tipo è possibile derivare in maniera naturale le proprietà fondamentali della probabilità come la proprietà di positività e la proprietà di certezza.
Vediamole brevemente:

- $0 \leq P(E) \leq 1$ (proprietà di positività) ;

Partendo dalla relazione:

$$P(E) = \frac{A}{S}; \tag{2.1}$$

poiché ovviamente $A, S > 0$ ma allo stesso tempo $A < S$ (nessuno accetterebbe di scommettere una cifra A più grande di una potenziale vincita S), è chiaro che $0 \leq P(E) \leq 1$, cioè la proprietà di positività.

- $P(\Omega) = 1$, $P(\emptyset) = 0$ con Ω evento certo, $\emptyset$ evento impossibile (proprietà di certezza) ;

Questa proprietà si può intuitivamente derivare dal fatto che nessuno sarebbe disposto a scommettere su un evento impossibile. Quindi nel caso $E = \emptyset$ si ha che $A = 0$, per cui:

$$P(\emptyset) = \frac{A}{S} = 0. \tag{2.2}$$

D'altro canto nel caso limite in cui l'evento E è l'evento certo, cioè $E = \Omega$ saremmo disposti a scommettere $A = S$ e si otterrebbe:

$$P(\Omega) = \frac{A}{S} = 1 \tag{2.3}$$

2.3 Probabilità soggettiva e condizionata

Come si accennava prima, e come si può osservare nell'approccio alla probabilità come scommessa, la definizione di probabilità come grado di fiducia che si ha circa il verificarsi o meno di un certo evento, è strettamente collegata alle valutazioni personali di chi sta effettuando la valutazione. Nell'approccio Bayesiano la probabilità è soggettiva e non è una grandezza intrinseca dell'oggetto. Ovviamente lo stato di informazione di chi valuta la probabilità cambia in maniera importante la quantificazione della probabilità come numero e, soprattutto, rende soggettiva questa stima. L'ideale di oggettività secondo questo approccio si raggiunge nel momento in cui tutti i hanno lo stesso grado di informazione.
Inoltre un altro aspetto fondamentale da considerare è che non ha senso parlare di "probabilità assoluta", ma si può parlare solo di "probabilità condizionata" ad un certo numero di informazioni. La indichiamo come $P(E|I)$ "probabilità di E dato I". Infatti quando si parla semplicemente di $P(E)$ senza aggiungere altro si fa riferimento ad una serie di circostanze convenzionali. Per esempio se si dice che la probabilità di ciascuna faccia di un dado è $1/6$ si sta assumendo come condizionante che il dado e il lancio siano perfettamente regolari. La probabilità soggettiva è basata sull'idea che la probabilità quantifichi il grado di fiducia che un evento avvenga, essendo a conoscenza di certe informazioni a priori. La statistica Bayesiana si basa sull'approccio soggettivo alla probabilità e in questo contesto gioca un ruolo fondamentale il teorema di Bayes che permette di aggiornare la probabilità alla luce di nuove informazioni.

2.4 Regole fondamentali della probabilità

Per proseguire occorre fare un rapido accenno ad alcune proprietà fondamentali del calcolo delle probabilità:

1. $P(\bigcup_{i=1}^{n} E_{\mathrm{i}}) = \sum_{i=1}^{n} P(E_{\mathrm{i}})$, se $E_{\mathrm{i}} \cap E_{\mathrm{j}} = \emptyset, i \neq j$ (proprietà di unione) ;
2. $P(B) = P(\bigcup_{i=1}^{n}(B \cap E_{\mathrm{i}})) = \sum_{i=1}^{n} P(B \cap E_{\mathrm{i}})$;
3. $P(E \cap H) = P(H) \cdot P(E|H) = P(E) \cdot P(H|E)$ (formula della probabilità composta).

Per quanto riguarda la proprietà 2 si ha che un qualsiasi evento B può essere scritto in generale come:

$$B = B \cap \Omega = B \cap (A \cup \bar{A}), \tag{2.4}$$

dove $\bar{A}$ è l'evento contrario ad A e quindi A ed $\bar{A}$ partizionano lo spazio dei risultati Ω. Da questa relazione si ottiene:

$$P(B) = P(B \cap A) + P(B \cap \bar{A}), \tag{2.5}$$

che generalizzata ad una qualsiasi partizione di Ω è proprio la proprietà 2.

Consideriamo infine la proprietà 3. È utile osservare la differenza tra $E \cap H$ ed $E|H$: il primo è l'intersezione tra i due eventi riferita allo spazio campionario Ω mentre il secondo è ancora l'intersezione dei due ma riferita allo spazio campionario ridotto costituito da H. Si può quindi capire che varrà la relazione $P(E|H) \geq P(E \cap H)$ in quanto l'evento E è contemplato all'interno di una classe di ipotesi più ristretta. Possiamo quindi scrivere:

$$P(E|H) = \frac{P(E \cap H)}{P(H)} \; con \; P(H) \neq 0, \tag{2.6}$$

che è proprio la formula 3 invertita. La proprietà 3 sarà di fondamentale importanza nel corso della trattazione, perché permetterà di ricavare una delle basi fondamentali della statistica Bayesiana: il *Teorema di Bayes*.

2.5 Un'altra chiave di lettura della probabilità condizionata

Un modo interessante di interpretare il concetto di probabilità condizionata è quello di pensare al condizionante come causa dell'evento e all'evento come effetto:

$$P(E|H) = P(\; Effetto \mid Causa \;). \tag{2.7}$$

La causa è da intendersi in maniera più generico come "teoria" che può provocare i vari effetti:

$$P(\; Fenomenologia \mid teoria \cap condizioni\ al\ contorno \;), \tag{2.8}$$

ovvero:

$$P(\; Osservazioni \mid grandezza \cap fattori\ di\ influenza \;), \tag{2.9}$$

Considerando la probabilità condizionata da questo punto di vista risulta evidente l'importanza che avrebbe invertire la probabilità per ottenere $P(H|E)$ a partire da $P(E|H)$, perché questo vuol dire valutare la probabilità delle cause. A tal proposito introduciamo il teorema di Bayes.

2.6 Teorema di Bayes

Immaginiamo che un evento E possa essere causato da una serie di ipotesi H_i. Utilizzando la proprietà 3 si ottiene:

$$P(E \cap H_i) = P(H_i|E) \cdot P(E) = P(E|H_i) \cdot P(H_i), \tag{2.10}$$

da cui segue che:

$$P(H_i|E) = \frac{P(E|H_i) \cdot P(H_i)}{P(E)}. \tag{2.11}$$

Se la classe di ipotesi H_i è completa, cioè le H_i sono a due a due incompatibili e la loro unione è l'evento certo Ω, si può scrivere per la proprietà 2 (si veda la figura (2.1)):

$$P(E) = \sum_{i=1}^{n} P(E \cap H_i). \tag{2.12}$$

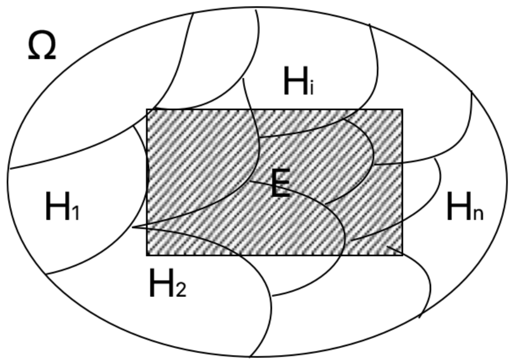


Figura 2.1: Insieme Ω come unione degli insiemi $\mathrm{H_i}$ a due a due disgiunti ed evento E

Sostituendo ora la (2.12) nella (2.11) e applicando la proprietà 3 si ottiene:

$$P(H_i|E) = \frac{P(E|H_i) \cdot P(H_i)}{\sum_{i=1}^{n} P(E|H_i) \cdot P(H_i)}, \tag{2.13}$$

che è proprio la forma finale del teorema di Bayes. Il termine a denominatore della (2.13) è un fattore di normalizzazione tale che:

$$\sum_{i=1}^{n} P(H_i|E) = 1. \tag{2.14}$$

Facciamo ora alcune considerazioni di base sul teorema di Bayes :

- $P(H_i|E) \propto P(H_i) \cdot P(E|H_i)$, esiste cioè un rapporto di proporzionalità diretta tra le due probabilità $P(H_i|E)$ e $P(E|H_i)$.

- Ricordando che le probabilità sono sempre da intendere condizionate alle ipotesi implicite che chiamiamo H_0 possiamo scrivere:

$$P(H_i|E, H_0) \propto P(E|H_i, H_0) \cdot P(H_i|H_0). \tag{2.15}$$

  Questo mostra come la probabilità di una certa ipotesi è aggiornata quando cambia lo stato di informazione. $P(H_i|H_0)$ è la probabilità *iniziale* o a *priori*, ovvero la probabilità di H_i date tutte le informazioni iniziali H_0 con o senza il verificarsi di E. È importante sottolineare che nel condurre una ricerca si ha sempre un grado di informazione iniziale: per quanto le conoscenze basilari su quello che stiamo cercando possano essere ristrette, esse rappresentano sempre un punto di partenza, un grado di informazione importante. $P(H_i|E, H_0)$ è la probabilità *finale* o *a posteriori* cioè la probabilità di H_i riaggiornata alla luce dell'ipotesi che E sia vero. $P(E|H_i, H_0)$ è la *verosimiglianza* ovvero una misura di quanto E sia verosimile alla luce di H_i, e cioè di quanto facilmente H_i possa produrre E. Anche la verosimiglianza è valutata tenendo conto di tutte le ipotesi preliminari H_0.

Il teorema di Bayes può dunque essere schematicamente riletto in questo modo:

$$probabilita' \ a \ posteriori \ \propto \ verosimiglianza \ \times \ probabilita' \ a \ priori \ . \tag{2.16}$$

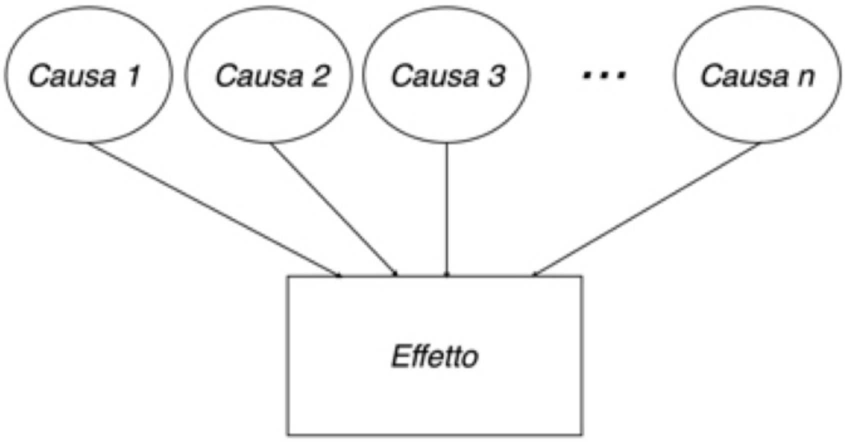


Figura 2.2: Un effetto può essere provocato da diverse cause. Il teorema di Bayes ci permette di calcolare la probabilità di ciascuna causa.

La figura 2.2 rappresenta schematicamente il caso in cui un certo effetto può essere provocato da n possibili cause. È spesso interessante capire quale sia la probabilità che l'evento sia stato causato proprio da una delle possibili n cause e il teorema di Bayes ci permette di farlo. La prima applicazione di uno schema simile è quella di utilizzare il teorema per valutare la probabilità che un effetto sperimentale osservato sia stato provocato da una delle sue possibili cause. Immaginiamo che l'evento E possa essere causato dall'ipotesi H_1 o dall'ipotesi H_2. È utile costruire il rapporto:

$$\frac{P(H_1|E, H_0)}{P(H_2|E, H_0)} = \frac{P(E|H_1, H_0)}{P(E|H_2, H_0)} \cdot \frac{P(H_1|H_0)}{P(H_2|H_0)} \tag{2.17}$$

Dove H_0 sono tutte le ipotesi preliminari. Questo è un test di ipotesi che permette di capire se l'evento E è causato con maggiore probabilità da H_1 o da H_2. Il termine $\frac{P(H_1|H_0)}{P(H_2|H_0)}$ è il rapporto tra le probabilità a priori. Quando non si ha nessuna ragione a priori per scegliere come causa H_1 piuttosto che H_2 vuol dire che il rapporto tra le probabilità a priori vale 1 e il test dipende solo dal fattore di Bayes. Se invece una delle probabilità a priori è molto grande vuol dire che si assegna ad una delle due ipotesi un grado di fiducia molto elevato. Questa dipendenza dalle "prior", cioè le probabilità a priori, viene spesso considerato il punto debole dell'approccio Bayesiano alla statistica; in realtà possiamo renderci conto facilmente che questo non è un difetto (ed inoltre non riguarda solo questo approccio), perché:

- Non si può costruire una teoria dell'incertezza che non sia affetta da pregiudizi. Ad esempio, come abbiamo visto prima parlando dell'equiprobabilità delle facce di un dado, si ha come pregiudizio che il dado sia perfettamente regolare.
- Con il crescere delle informazioni iniziali certe, la dipendenza dai pregiudizi diminuisce.
- Quando si hanno pochissime informazioni iniziali, non è un ragionamento sbagliato quello di basarsi su assunzioni a priori per provare a restringere il campo di ricerca.
- Spesso gli esperimenti e la loro stessa progettazione dipendono dal risultato che ci si aspetta da essi, motivo per cui anche i risultati possono dipendere in maniera più o meno importante dalle assunzioni a priori.

Dedichiamo il prossimo capitolo a due di esempi di applicazione del teorema di Bayes.

Capitolo 3

Applicazioni

3.1 Esperimento di rivelazione

Supponiamo che un rivelatore abbia un'efficienza di identificazione delle particelle μ del 95% e una probabilità di confondere una particella π per un μ del 2%. Se una particella viene identificata come un μ si accende una lampadina (L). Sapendo che le particelle del fascio contengono il 10% di μ e il 90% di π ci chiediamo:

- Quanto vale la probabilità che, se si accende la lampadina, sia passato un μ ?
- Quanto vale la probabilità che, se non si accende la lampadina, sia passato un π ?
- Come cambiano i risultati se si pongono sul fascio due contatori aventi le stesse caratteristiche e funzionanti indipendentemente ?

Per prima cosa schematizziamo le informazioni che abbiamo e le domande a cui vogliamo rispondere:
-$P(L|\mu) = 0,95$ è la probabilità che si accenda la lampadina se passa una vera particella μ. Essa viene stimata attraverso la frequenza relativa con la quale si accende la lampadina quando il rivelatore viene esposto ad un fascio di soli mu, ammettendo che ciò sia possibile.
-$P(\bar{L}|\mu) = 0,05$ è la probabilità complementare a $P(L|\mu)$.
-$P(L|\pi) = 0,02$ è la probabilità che la lampadina si accenda "per errore".
-$P(\bar{L}|\pi) = 0,98$ è la probabilità complementare a $P(L|\pi)$.
-$P_0(\mu) = 0,10$ è la probabilità a priori che una particella che arrivi al rivelatore sia proprio un μ in assenza dell'informazione sull'accensione o meno della lampadina. Questa probabilità include tutte le informazioni che si posseggono sul fascio.
-$P_0(\pi) = 0,90$ probabilità a priori dell'evento π . Ovviamente $P_0(\mu) + P_0(\pi) = 1$
-$P(\mu|L)$ la probabilità che la particella che ha attraversato il rivelatore sia un μ se la la lampadina si è accesa.

-$P(\pi|L)$ è la probabilità che la particella che ha attraversato il rivelatore sia un π se la lampadina si è accesa.(Complementare a $P(\mu|L)$).
Applicando il teorema di Bayes si ottiene per le probabilità richieste:

$$P(\mu|L) = \frac{P(L|\mu) \cdot P_0(\mu)}{P(L|\mu) \cdot P_0(\mu) + P(L|\pi) \cdot P_0(\pi)} = 0,84. \tag{3.1}$$

$$P(\pi|L) = 1 - P(\mu|L) = 0,16. \tag{3.2}$$

$$P(\mu|\bar{L}) = \frac{P(\bar{L}|\mu) \cdot P_0(\mu)}{P(\bar{L}|\mu) \cdot P_0(\mu) + P(\bar{L}|\mu) \cdot P_0(\pi)} = 0,0056. \tag{3.3}$$

$$P(\pi|\bar{L}) = 1 - P(\mu|\bar{L}) = 0,994. \tag{3.4}$$

Siamo quindi più sicuri che quando non si accende la luce sia passato un π di quanto non lo siamo quando identifichiamo un μ essendosi accesa la lampadina.
Per rispondere all'ultima domanda, ossia come cambiano le probabilità se si introducono due contatori, possiamo utilizzare la probabilità finale condizionata dall'evento μ al posto di $P_0(\mu)$ e applicare il teorema di Bayes rispetto al secondo condizionamento di L_2:

$$P(\mu|L_1 \cap L_2) = \frac{P(L_2|\mu) \cdot P(\mu|L_1)}{P(L_2|\mu) \cdot P(\mu|L_1) + P(L_2|\pi) \cdot P(\pi|L_1)} = 0,996. \tag{3.5}$$

Questo uso iterativo del teorema di Bayes può essere riassunto dicendo che:

La probabilità iniziale di una inferenza è pari alla probabilità finale dell'inferenza precedente.

Per rispondere a questa domanda avremmo anche potuto calcolare le probabilità condizionate $P(L_1 \cap L_2|\mu)$ e $P(L_1 \cap L_2|\pi)$ e usare il teorema di Bayes per trovare $P(\mu|L_1 \cap L_2)$. Seguendo questa strada il risultato ovviamente sarebbe stato lo stesso ma è fondamentale sottolineare l'importanza dell'uso iterativo del teorema di Bayes. Soprattutto in casi più complessi, in cui entrano in gioco più fattori e l'esperimento è meno schematico, l'aggiornamento delle probabilità in base alle nuove informazioni ci permette di perfezionare la nostra stima di probabilità passo dopo passo senza dover valutare da capo il problema.

3.2 Urne

Consideriamo ora un'altra applicazione del teorema di Bayes. Immaginiamo di avere sei urne H_{j}, con $j = 0, 1, ..., 5$, e cinque palline all'interno di ciascuna urna di cui j sono bianche e $5 - j$ sono nere. Quindi avremo ad esempio nella scatola H_3, tre palline bianche e due palline nere. Immaginiamo che queste urne siano identiche esternamente e che quindi sia impossibile conoscerne il contenuto.

Scegliamo ora a caso una delle urne. È evidente che alla scelta iniziale tutte le urne sono equiprobabili, essendo indistinguibili, per cui la probabilità relativa alla scelta di una qualsiasi delle urne H_{j} sarà $P(H_{\mathrm{j}}) = 1/6$. Dopo aver scelto una delle urne, estraiamo a caso da essa una delle palline.

Ci chiediamo:

- Alla luce del colore della pallina estratta, cosa possiamo dire sulle probabilità che l'urna scelta sia la j-esima? In termini più precisi vorremmo conoscere le probabilità $P(H_{\mathrm{j}}|E_{\mathrm{b}})$ e $P(H_{\mathrm{j}}|E_{\mathrm{n}})$ dove E_{b} è l'evento "estrazione della pallina bianca" e E_{n} è l'evento "estrazione della pallina nera".

- Come si riaggiornano le probabilità se proseguiamo nelle estrazioni delle palline dall'urna scelta reinserendo di volta in volta la pallina? Quanto vale quindi $P(H_{\mathrm{j}}|E_{\mathrm{x}_1} \cap E_{\mathrm{x}_2} \cap E_{\mathrm{x}_3} ... \cap .E_{\mathrm{x}_{\mathrm{k}}})$ dove per gli x_{k} si ha $x_{\mathrm{k}} = \quad b \quad \mathrm{o} \quad n \quad ?$

Per rispondere a queste domande abbiamo implementato una simulazione del problema con il programma di manipolazione algebrica *Mathematica* il cui testo è riportato in appendice A. Un problema di questo tipo può essere facilmente risolto utilizzando il teorema di Bayes in maniera iterata e aggiornando quindi di volta in volta la probabilità associata a ciascuna urna. È chiaro che con il crescere delle estrazioni potremo essere sempre più sicuri circa il contenuto dell'urna scelta perché aggiungeremo sempre più informazioni alla nostra conoscenza del problema, ma è anche altrettanto evidente che non potremo mai raggiungere la certezza assoluta. Per prima cosa si sceglie un'urna. Le probabilità associate a ciascuna urna in questo momento sono tutte pari ad $1/6$ e sono rappresentate graficamente in Figura 3.1(a):

Tabella 3.1: Dati prima dell'inizio del pescaggio

Urne	H_0	H_1	H_2	H_3	H_4	H_5	Estrazioni dall'urna
Probabilità	0.167	0.167	0.167	0.167	0.167	0.167	0

A questo punto la prima pallina estratta nella simulazione risulta essere **bianca**, e le probabilità si aggiornano applicando il teorema di Bayes nel seguente modo:

$$P(H_{\mathrm{j}}|E_{\mathrm{b}}) = \frac{P(E_{\mathrm{b}}|H_{\mathrm{j}}) \cdot P(H_{\mathrm{j}})}{\sum_{j=0}^{5} P(E_{\mathrm{b}}|H_{\mathrm{j}}) \cdot P(H_{\mathrm{j}})}, \tag{3.6}$$

dove $P(E_{\mathrm{b}}|H_{\mathrm{j}})$ è la probabilità di estrarre una pallina bianca dall'urna H_{j}(dati in Tabella 3.2 e grafico in figura 3.1(b)).

La seconda pallina estratta nella simulazione è **nera**. Applicando il meccanismo di aggiornamento delle probabilità utilizziamo questa volta come probabilità a priori delle urne le probabilità finali dell'inferenza precedente cioè le $P(H_{\mathrm{j}}|E_{\mathrm{b}})$. Le nuove probabilità saranno quindi (dati in tabella 3.2 e grafico in Figura 3.1(c)):

$$P(H_{\mathrm{j}}|E_{\mathrm{b}} \cap E_{\mathrm{n}}) = \frac{P(E_{\mathrm{n}}|H_{\mathrm{j}}) \cdot P(H_{\mathrm{j}}|E_{\mathrm{b}})}{\sum_{j=0}^{5} P(E_{\mathrm{n}}|H_{\mathrm{j}}) \cdot P(H_{\mathrm{j}}|E_{\mathrm{b}})}. \tag{3.7}$$

Abbiamo reiterato questo meccanismo per 65 estrazioni facendo compiere al programma 65 cicli.

Tabella 3.2: Riportati il numero di pescaggi, il conteggio di estrazioni di palline bianche e nere (b,n) e le probabilità relative alle urne dopo k pescaggi

Pescaggi (k)	Risultati(b,n)	H_0	H_1	H_2	H_3	H_4	H_5
0	-	0.167	0.167	0.167	0.167	0.167	0.167
1	(1,0)	0	0.067	0.133	0.2	0.267	0.333
2	(1,1)	0	0.2	0.3	0.3	0.2	0
3	(2,1)	0	0.08	0.24	0.36	0.32	0
4	(2,2)	0	0.154	0.347	0.347	0.154	0
5	(2,3)	0	0.246	0.415	0.277	0.061	0
7	(4,3)	0	0.046	0.308	0.463	0.183	0
10	(6,4)	0	0.012	0.244	0.550	0.193	0
15	(9,6)	0	0.002	0.196	0.663	0.138	0
20	(13,7)	0	$5 \cdot 10^{-5}$	0.062	0.706	0.232	0
25	(14,11)	0	$3 \cdot 10^{-4}$	0.224	0.755	0.021	0
30	(18,12)	0	$9 \cdot 10^{-6}$	0.078	0.884	0.038	0
35	(21,14)	0	10^{-6}	0.054	0.922	0.024	0
40	(24,16)	0	$2 \cdot 10^{-7}$	0.037	0.945	0.014	0
45	(28,17)	0	$5 \cdot 10^{-9}$	0.011	0.966	0.023	0
55	(35,20)	0	$2 \cdot 10^{-11}$	0.002	0.975	0.022	0
65	(41,24)	0	$4 \cdot 10^{-13}$	0.001	0.991	0.008	0

Osservando i dati, dopo 65 estrazioni con reinserimento, possiamo essere quasi certi di aver scelto casualmente la scatola H_3, cioè quella contenente tre palline bianche e due nere.
Facciamo ora alcune considerazioni su quanto visto nei grafici e nella tabella:

- Le probabilità di H_0 e H_5 si azzerano a partire dal momento in cui vengono estratte rispettivamente una pallina bianca e una pallina nera. Intuitivamente questo azzeramento è ovvio: una volta pescata una pallina bianca non possiamo aspettarci di avere selezionato la scatola H_0 contenente zero palline bianche. Questo azzeramento è frutto del meccanismo di aggiornamento delle probabilità: dopo aver estratto una pallina bianca infatti la probabilità a priori $P(H_0)$ si annulla per tutte le inferenze successive.

- Come possiamo vedere nella Tabella 3.2 e in maniera visiva nelle Figure 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7 man mano che si procede con le estrazioni la probabilità di una delle urne cresce fino ad arrivare a valori molto prossimi a 1 (in questa particolare simulazione H_3)e le probabilità delle altre urne decrescono fino ad arrivare quasi tutte a zero. Ovviamente non sarà possibile

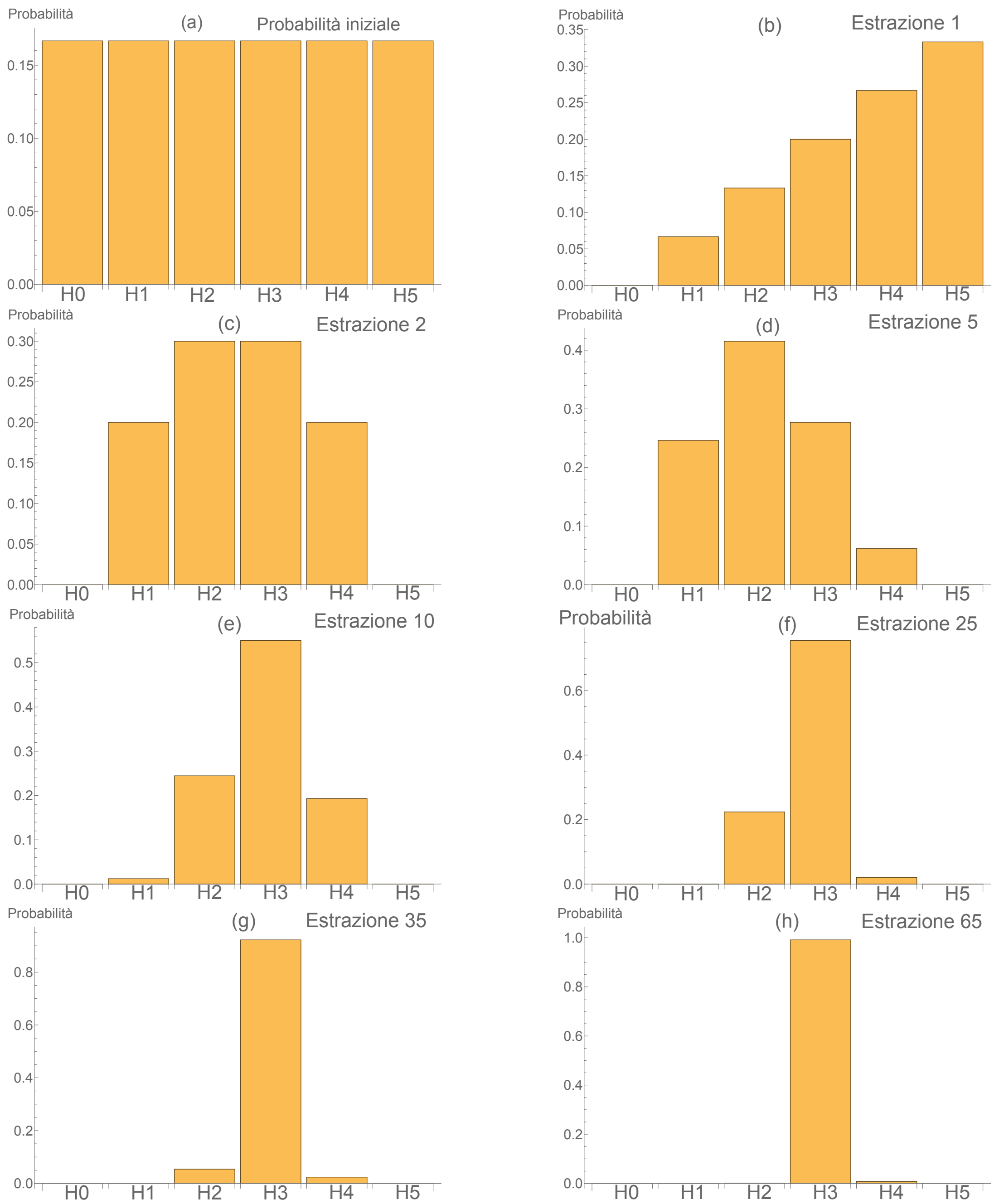


Figura 3.1: In figura sono riportati i grafici delle probabilità di ciascuna urna dopo k estrazioni e i relativi dati si trovano in tabella 3.2

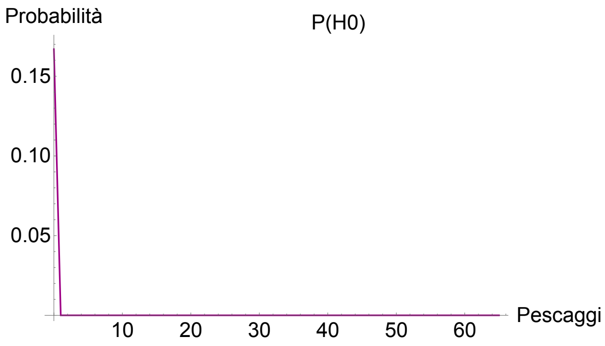


Figura 3.2: Andamento di $P(H_0)$.

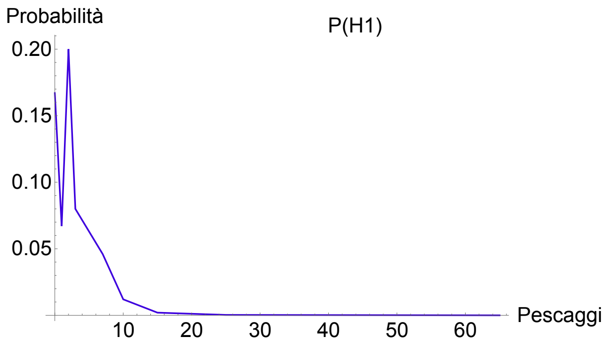


Figura 3.3: Andamento di $P(H_1)$.

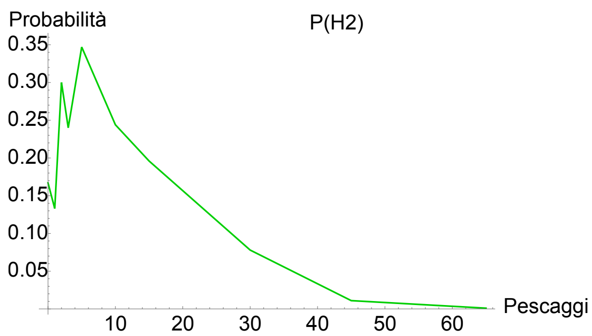


Figura 3.4: Andamento di $P(H_2)$.

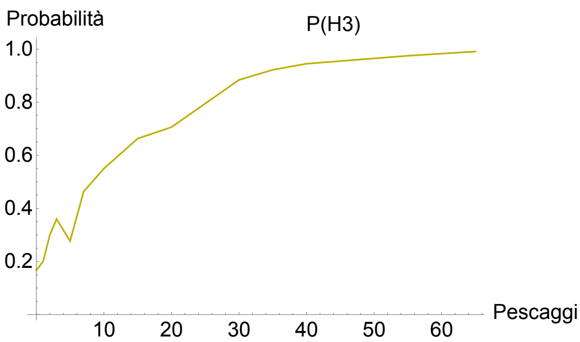


Figura 3.5: Andamento di $P(H_3)$.

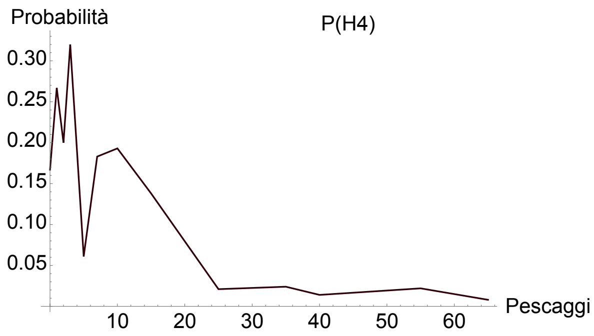


Figura 3.6: Andamento di $P(H_4)$.

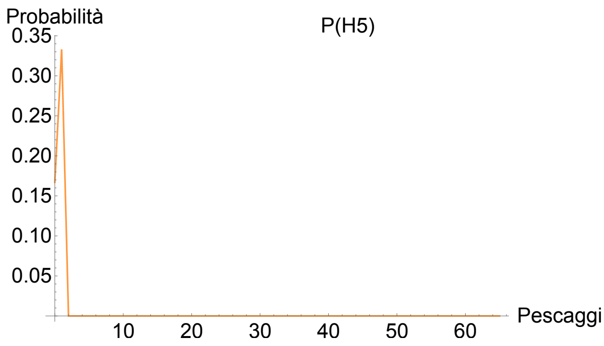


Figura 3.7: Andamento di $P(H_5)$.

raggiungere la certezza *assoluta* ma potremo arrivare ad un grado di fiducia arbitrariamente elevato.

- Se consideriamo le frequenze relative $\frac{palline\ bianche}{palline\ totali} = \frac{41}{65} = 0.631$ e $\frac{palline\ nere}{palline\ totali} = \frac{24}{65} = 0.369$ possiamo osservare che queste frequenze sono rispettivamente vicine a $P(E_{\mathrm{b}}|H_3) = \frac{3}{5} = 0.6$ e $P(E_{\mathrm{n}}|H_3) = \frac{2}{5} = 0.4$ che sono le probabilità associate al pescaggio di una pallina bianca o di una pallina nera dall'urna H_3. Nel limite per di un numero infinito di pescaggi avremo uguaglianza tra queste frequenze.

Capitolo 4

Statistica Bayesiana e funzioni di distribuzione

In questo ultimo capitolo vogliamo estendere il teorema di Bayes al caso in cui le quantità sulle quali si vogliono avere informazioni non sono più probabilità di eventi, ma delle quantità incerte che seguono delle funzioni di distribuzione. Considereremo ad esempio il caso in cui un set di misure è distribuito normalmente e analizzeremo brevemente il meccanismo di aggiornamento delle probabilità che risulta dal confronto di due set di misure provenienti dalla stessa grandezza μ.

4.1 Funzioni di distribuzione

Per proseguire nella trattazione ricordiamo brevemente il concetto di funzione di distribuzione e le sue proprietà. Una distribuzione di probabilità è un modello matematico che collega i valori di una variabile alle probabilità che tali valori possano essere osservati.

Nel caso delle distribuzioni discrete abbiamo che la variabile x assume valori numerici interi e la funzione di distribuzione, che indichiamo con $f(x)$, è una funzione pari alla probabilità che la variabile assuma il valore x. Questa funzione gode delle proprietà:

$$0 \leq f(x) \leq 1, \tag{4.1}$$

$$\sum_x f(x) = 1. \tag{4.2}$$

Nel caso discreto definiamo il *valore aspettato* μ e la *varianza* σ^2 rispettivamente come:

$$\mu = \sum_x x \cdot f(x), \tag{4.3}$$

e

$$\sigma^2 = \sum_x (x - \mu)^2 \cdot f(x). \tag{4.4}$$

Nel caso invece in cui la variabile casuale è continua, per prima cosa definiamo la *funzione cumulativa di probabilità*

$$F(x') = P(x \leq x') \tag{4.5}$$

che gode delle proprietà :

$$\lim_{x \to \infty} F(x) = 1, \tag{4.6}$$

$$\lim_{x \to -\infty} F(x) = 0. \tag{4.7}$$

A questo punto introduciamo la funzione di distribuzione di probabilità:

$$f(x) = \frac{d}{dx} F(x), \tag{4.8}$$

tale che:

$$F(x) = \int_{-\infty}^{x} f(x')dx' = P(x \leq x') \tag{4.9}$$

Ovviamente si ha che

$$\int_{-\infty}^{\infty} f(x')dx' = 1. \tag{4.10}$$

Nel caso di variabili casuali continue non ha senso domandarsi quale sia la probabilità che la variabile x assuma un determinato valore, in quanto questa probabilità è certamente nulla. Ci si può domandare solo quale sia la probabilità che la variabile x assuma valori all'interno di un determinato intervallo. Nel caso continuo abbiamo per il valore aspettato μ e la varianza σ^2 le seguenti espessioni:

$$\mu = \int_{-\infty}^{\infty} x \cdot f(x)dx, \tag{4.11}$$

e

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx. \tag{4.12}$$

4.2 Teorema di Bayes su quantità incerte

Fino ad ora abbiamo applicato il teorema di Bayes ad esperimenti di cui conoscevamo già i possibili risultati. In molti casi però, prima di eseguire le misure, ci troviamo in una condizione di incertezza sia sul valore vero della grandezza μ che sul possibile valore osservato x. Quantifichiamo questa incertezza con la funzione di distribuzione $f(x, \mu)$. Per calcolare la probabilità condizionata $f(\mu|x)$ riprendiamo la forma standard del teorema di Bayes (2.13) ed estendiamola al caso continuo:

$$f(\mu|x) = \frac{f(x|\mu) \cdot f_0(\mu)}{\int f(x|\mu) \cdot f_0(\mu)d\mu}, \tag{4.13}$$

dove con $f_0(\mu)$ si intende la probabilità a priori condizionata da tutte le conoscenze preliminari.

Se nell'esecuzione dell'esperimento abbiamo una buona conoscenza delle caratteristiche del nostro apparato, avremo che $f(x|\mu)$ sarà una distribuzione stretta mentre il fattore di imprecisione maggiore proverrà da $f(\mu)$ che sarà una distribuzione molto larga, cosa che tra l'altro ci si può aspettare considerando che è proprio su μ che vogliamo acquisire informazioni. Ovviamente se la probabilità

a priori è molto larga, cioè ci dà un'informazione vaga, significa che a priori riteniamo tutti i valori egualmente probabili.
Le forme assunte da $f(x|\mu)$ e $f(\mu)$ sono ovviamente diverse in base al problema di fronte al quale ci troviamo: possono per esempio essere distribuite normalmente come vedremo nel paragrafo successivo.

4.3 Verosimiglianza Gaussiana

Consideriamo un campione di dati $\mathbf{q}$ di n_1 misure della variabile Q delle quali calcoliamo la media $\bar{q}_{\mathrm{n}_1}$. Assumiamo poi che la deviazione standard σ della variabile Q sia nota o perché n_1 è molto grande e quindi σ può essere stimata dal campione o perché è conosciuta a *priori*. Abbiamo dunque che la verosimiglianza $f(\bar{q}_{\mathrm{n}_1}|\mu, \sigma)$ è una Gaussiana e che la media delle misure $\bar{Q}_{\mathrm{n}_1}$ è distribuita normalmente con parametri μ e $\frac{\sigma}{\sqrt{n_1}}$ cioè $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{n_1}})$. Poniamo per comodità: $x_1 = \bar{q}_{\mathrm{n}_1}$ $\sigma_1 = \frac{\sigma}{\sqrt{n_1}}$.
Applicando l'equazione (4.13) otteniamo:

$$f(\mu|x_1) = \frac{\frac{1}{\sqrt{2\pi}\sigma_1} \cdot exp\left[-\frac{(x_1-\mu)^2}{2\sigma_1^2}\right] \cdot f_0(\mu)}{\int \frac{1}{\sqrt{2\pi}\sigma_1} \cdot exp\left[-\frac{(x_1-\mu)^2}{2\sigma_1^2}\right] \cdot f_0(\mu) d\mu}. \tag{4.14}$$

Supponiamo di avere informazioni a priori quasi nulle su μ e quindi $f_0(\mu)$ uniforme su un intervallo arbitrariamente grande (al limite infinito) che includa x_1. Ovviamente quando siamo molto distanti da x_1, la funzione integranda al denominatore tenderà a zero per le proprietà della Gaussiana. Possiamo allora semplificare l'equazione (4.14) e ottenere:

$$f(\mu|x_1) = \frac{\frac{1}{\sqrt{2\pi}\sigma_1} \cdot exp\left[-\frac{(x_1-\mu)^2}{2\sigma_1^2}\right]}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \cdot exp\left[-\frac{(x_1-\mu)^2}{2\sigma_1^2}\right] d\mu}, \tag{4.15}$$

e poichè l'integrale a denominatore è l'integrale di una funzione di distribuzione di probabilità esso è uguale a 1. Otteniamo:

$$f(\mu|x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \cdot exp\left[-\frac{(x_1-\mu)^2}{2\sigma_1^2}\right], \tag{4.16}$$

per cui:

- il valore vero è distribuito normalmente intorno a x_1;
- la sua migliore stima è il valore aspettato di μ cioè x_1;
- la sua varianza è $\sigma_\mu = \sigma_1$;
- gli intervalli di confidenza nei quale c'è una certa probabilità di trovare il valore vero sono:

Probabilità (%)	intervallo di credibilità
68.3	$x_1 \pm \sigma_1$
90.0	$x_1 \pm 1.65\sigma_1$
95.0	$x_1 \pm 1.96\sigma_1$
99.0	$x_1 \pm 2.58\sigma_1$
99.7	$x_1 \pm 3\sigma_1$

4.4 Combinazione di diversi set di misure

Immaginiamo ora di eseguire un secondo campione di misure della quantità fisica in esame, che assumiamo essere invariata. Cerchiamo di capire come cambia la nostra conoscenza di μ dopo le informazioni acquisite dal nuovo set di dati. Avremo quindi $x_2 = \bar{q}_{\mathrm{n}_2}$ e $\sigma_2 = \frac{\sigma'}{\sqrt{n_2}}$ rispettivamente la nuova media e la nuova deviazione standard dalla media. Applicando nuovamente il teorema di Bayes e assumendo come probabilità a priori la probabilità finale della precedente applicazione otteniamo:

$$f(\mu|x_1,\sigma_1,x_2,\sigma_2,\mathcal{N}) = \frac{\frac{1}{\sqrt{2\pi}\sigma_2}\cdot exp\left[-\frac{(x_2-\mu)^2}{2\sigma_2^2}\right]\cdot f(\mu|x_1,\mathcal{N})}{\int\frac{1}{\sqrt{2\pi}\sigma_2}\cdot exp\left[-\frac{(x_2-\mu)^2}{2\sigma_2^2}\right]\cdot f(\mu|x_1,\mathcal{N})d\mu}. \tag{4.17}$$

Integrando il denominatore dell'eq. (4.17) si ottiene:

$$f(\mu|x_1,\sigma_1,x_2,\sigma_2,\mathcal{N}) = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{A}}}\cdot exp\left[-\frac{(\mu-x_{\mathrm{A}})^2}{2\sigma_{\mathrm{A}}^2}\right], \tag{4.18}$$

dove:

$$x_{\mathrm{A}} = \frac{\frac{x_1}{\sigma_1^2}+\frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}, \tag{4.19}$$

$$\frac{1}{\sigma_{\mathrm{A}}^2} = \frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}, \tag{4.20}$$

la (4.19) è la media ponderata e la (4.20) è la varianza della media ponderata.
Facciamo ora alcune considerazioni:

- Anche in questo caso il meccanismo di aggiornamento del teorema risulta molto comodo e intuitivo. L'aggiunta di un set di dati ci fornisce più informazioni sulla grandezza fisica in esame e quindi una probabilità a posteriori più precisa.
- se $\sigma_1 >> \sigma_2$ (e x_1 non è troppo lontano da x_2, il che significherebbe che la nostra misura è ancora molto vaga visto che i due campioni hanno dato medie molto diverse), allora il risultato finale sarà determinato solo dal secondo set di misure.
- La combinazione dei due campioni di dati richiede una valutazione soggettiva da parte dello sperimentatore sul fatto che le due misure provengano realmente entrambe dallo stesso valore vero μ.

- Scegliendo come probabilità a priori una funzione di distribuzione uniforme tra $-\infty$ e $+\infty$, avremmo che questa non è normalizzabile. Le *priors* non normalizzabili sono chiamate *priors improprie* e possiamo considerarle come degli artifici matematici utili per rappresentare un'incertezza su un grande intervallo. Per esempio se stiamo effettuando esperimenti su valori di massa o lunghezza non ha senso considerare quantità negative. Tuttavia se stiamo effettuando le nostre misurazioni con uno strumento che dà una risposta Gaussiana, il prodotto *verosimiglianza* $\times$ *probabilità a priori* si annulla nei punti in cui la verosimiglianza (distribuita come una Gaussiana), tende a zero, per esempio per valori negativi della massa. Questo mostra che è molto importante saper scegliere le funzioni di distribuzione che si adattano meglio al problema.

Conclusioni

Nel corso di questa tesi si è cercato di fornire una panoramica generale dell'approccio Bayesiano alla statistica approfondendo soprattutto concetti di base e inserendo alcuni esempi pratici che mostrano direttamente l'utilità e il procedimento applicativo di questo approccio. Nella prima parte si è parlato brevemente dell' approccio frequentista evidenziando alcune sue problematiche e mettendo in luce come questo risulti spesso poco generale. Successivamente si è introdotto il concetto di probabilità come grado di fiducia nel verificarsi di un dato evento. Dopo aver ripreso alcune proprietà fondamentali della probabilità e dimostrato come queste possano essere semplicemente derivate dal concetto di probabilità come scommessa, si è parlato in modo più specifico del ragionamento Bayesiano illustrando l'idea che la probabilità è sempre soggettiva e condizionata da certe ipotesi di base e pregiudizi diversi a seconda dell'osservatore. Si è poi parlato del teorema di Bayes facendo considerazioni sul suo significato e proponendo due esempi che ne mostrano l'utilizzo; uno di questi esempi ha richiesto lo sviluppo di simulazione con il programma Mathematica e ha mostrato in maniera chiara il meccanismo di aggiornamento delle probabilità, un punto cruciale dell'approccio Bayesiano. Infine si è concluso l'argomento con una generalizzazione del teorema di Bayes analizzando la circostanza in cui si ha a che fare con funzioni di distribuzioni di probabilità, come nel caso di una verosimiglianza Gaussiana, e analizzando il problema del confronto tra due diversi set di misure della stessa osservabile fisica.

Appendice A

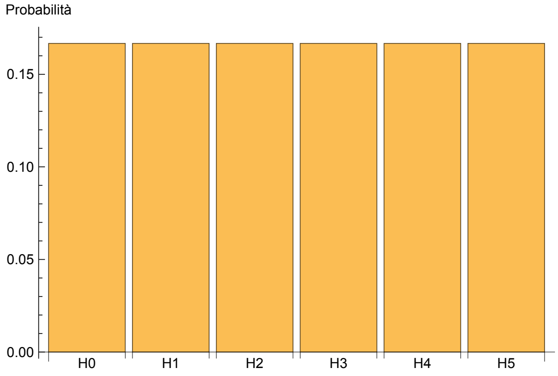
Programma di simulazione delle urne

```
In[368]:= PbiancaSc = {0., 1 / 5., 2/5., 3/5., 4 / 5., 5 / 5.};
          (*Probabilità di pescare una pallina bianca nelle varie urne*)
          PneraSc = {5 / 5., 4 / 5., 3/5., 2/5., 1 / 5., 0.};
          (*Probabilità di pescare una pallina nera nelle varie urne *)

In[370]:= pSc[1] = {1/6., 1/6., 1/6., 1/6., 1/6., 1/6.};
          (*Probabilità a priori di pescare
           una qualunque delle sei urne a disposizione*)

In[371]:= Sc = RandomInteger[{0, 5}];
          Print["Ho scelto casualmente l'urna con ", Sc, " palline bianche"]
          (*Questo comando genera un numero intero casuale tra 0 e 5
           scegliendo così una delle urne con 0,1,2,3,4,5 palline bianche*)
          Pbianca = PbiancaSc[[Sc + 1]];
          Pnera = 1 - Pbianca;
          Print[
           "la probabilità di pescare una pallina bianca dall'urna scelta è " ,  Pbianca]
          Print["la probabilità di pescare una pallina nera dall'urna scelta è = " ,
           Pnera]
          (*Questo comando stampa le probabilità di pescare una pallina
            bianca dall'urna scelta. La probabilità di pescare una
            pallina nera viene di conseguenza calcolato come 1-Pbianca*)

          Ho scelto casualmente l'urna con 5 palline bianche

          la probabilità di pescare una pallina bianca dall'urna scelta è 1.

          la probabilità di pescare una pallina nera dall'urna scelta è = 0.
```

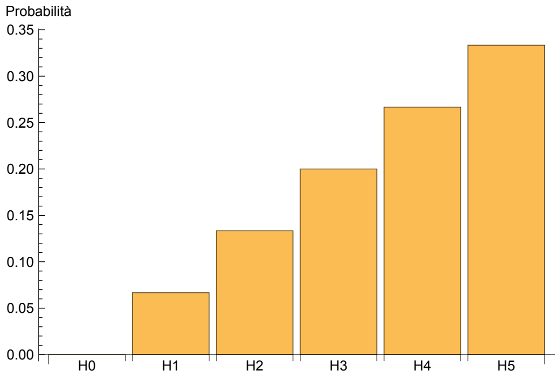
```
For [n = 1, n < 70, n++,
 c = (Sum[PbiancaSc[[i]] * pSc[n][[i]], {i, 1, 6}]);
 s = (Sum[PneraSc[[i]] * pSc[n][[i]], {i, 1, 6}]);
 If[RandomReal[] ≤ Pbianca,
  Pallina = "bianca";
  pSc[n + 1] = PbiancaSc * pSc[n] / c,
  Pallina = "nera";
  pSc[n + 1] = PneraSc * pSc[n] / s];
 If[n < 4 || Mod[n, 5] == 0,
  Print["Pescaggio n. ", n];
  Print["Ho pescato la pallina ", Pallina];
  Print[
   "Le nuove probabilità aggiornate dal pescaggio sono queste: ", pSc[n + 1]];
  Print[BarChart[pSc[n + 1], ChartLabels → {"H0", "H1", "H2", "H3", "H4", "H5"},
    AxesLabel → {None, Probabilità}]], Print[] ]]
ListLinePlot[{Table[{j, pSc[j + 1][[1]]}, {j, 0, 65}]},
 AxesLabel → {"Pescaggi", "Probabilità"},  PlotLabel → "P(H0)",
 PlotStyle → RGBColor[1., 0.36, 0.19], PlotRange → All]
ListLinePlot[{Table[{j, pSc[j + 1][[2]]}, {j, 0, 65}]},
 AxesLabel → {"Pescaggi", "Probabilità"}, PlotLabel → "P(H1)",
 PlotStyle → RGBColor[0.3, 0.19, 1.], PlotRange → All]
ListLinePlot[{Table[{j, pSc[j + 1][[3]]}, {j, 0, 65}]},
 AxesLabel → {"Pescaggi", "Probabilità"}, PlotLabel → "P(H2)",
 PlotStyle → RGBColor[0.5, 0.1, 0.5], PlotRange → All]
ListLinePlot[{Table[{j, pSc[j + 1][[4]]}, {j, 0, 65}]},
 AxesLabel → {"Pescaggi", "Probabilità"}, PlotLabel → "P(H3)",
 PlotStyle → RGBColor[0.3, 1., 0.4], PlotRange → All]
ListLinePlot[{Table[{j, pSc[j + 1][[5]]}, {j, 0, 65}]},
 AxesLabel → {"Pescaggi", "Probabilità"}, PlotLabel → "P(H4)",
 PlotStyle → RGBColor[0.7, 0.8, 0.5], PlotRange → All]
ListLinePlot[{Table[{j, pSc[j + 1][[6]]}, {j, 0, 65}]} ,
 AxesLabel → {"Pescaggi", "Probabilità"}, PlotLabel → "P(H5)",
 PlotStyle → RGBColor[0., 0., 0.], PlotRange → All]
(*Questo comando genera il pescaggio casuale di una pallina bianca
  o nera dall'interno dell'urna scelta al punto precedente. Per
  fare ciò si genera un numero reale compreso tra 0 e 1:
 nel caso in cui questo numero sia minore di Pbianca
  vuol dire che la pallina scelta è bianca,
altrimenti è nera. Dopo di che le probabilità delle urne
 vengono aggiornate  con l'applicazione del teorema di Bayes
 e vengono stampate su un grafico a barre. Dopo aver stampato
 le probabilità iniziali delle urne, utilizzando un ciclo For,
l'operazione di pescaggio viene effettuata per n volte ma
 i risultati delle estrazioni con l'esito del pescaggio,
le probabilità aggiornate e i relativi grafici a barre vengono
 stampati solamente  per gli n selezionati all'interno del ciclo.
```

```
le probabilità delle urne prima del pescaggio sono
 {0.166667, 0.166667, 0.166667, 0.166667, 0.166667, 0.166667}
```

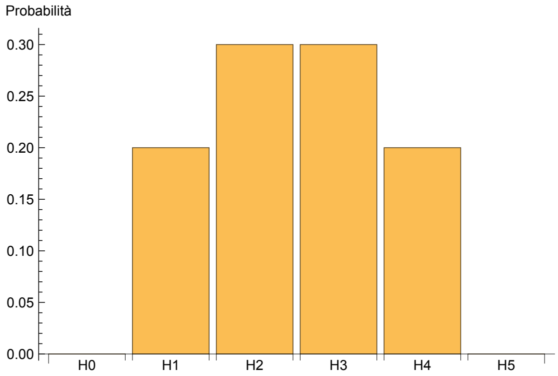
Out[98]=



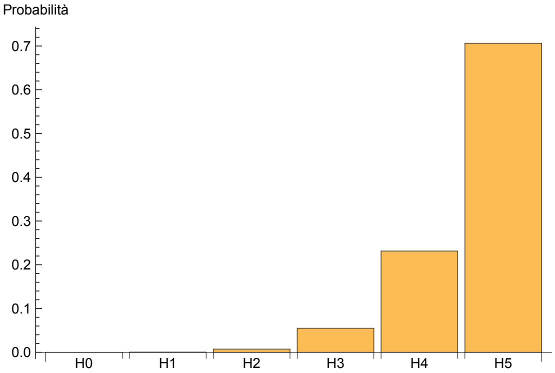

```
Pescaggio n. 1
Ho pescato la pallina bianca
Le nuove probabilità aggiornate dal pescaggio sono queste:
 {0., 0.0666667, 0.133333, 0.2, 0.266667, 0.333333}
```



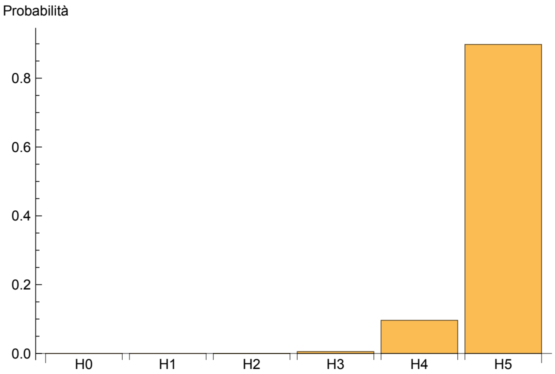

```
Pescaggio n. 2
Ho pescato la pallina nera
Le nuove probabilità aggiornate dal pescaggio sono queste: {0., 0.2, 0.3, 0.3, 0.2, 0.}
```



```
Pescaggio n. 5
Ho pescato la pallina bianca
Le nuove probabilità aggiornate dal pescaggio sono queste:
 {0., 0.000225989, 0.00723164, 0.0549153, 0.231412, 0.706215}
```




```
Pescaggio n. 10
Ho pescato la pallina bianca
Le nuove probabilità aggiornate dal pescaggio sono queste:
```
$\{0., 9.19602 \times 10^{-8}, 0.0000941672, 0.00543016, 0.0964272, 0.898048\}$

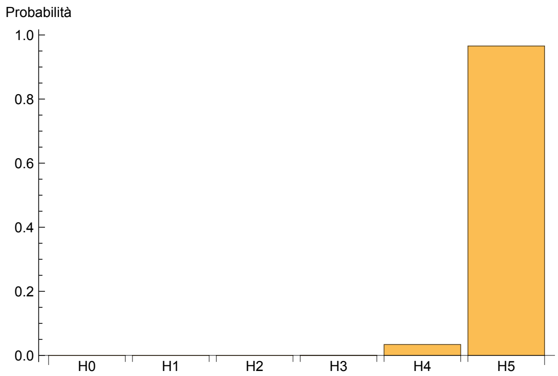


Pescaggio n. 15

Ho pescato la pallina bianca

Le nuove probabilità aggiornate dal pescaggio sono queste:

$\{0., 3.16399 \times 10^{-11}, 1.03677 \times 10^{-6}, 0.000453997, 0.033973, 0.965572\}$

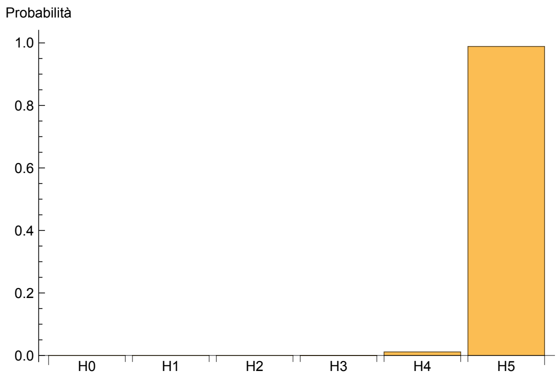


Pescaggio n. 20

Ho pescato la pallina bianca

Le nuove probabilità aggiornate dal pescaggio sono queste:

$\{0., 1.03659 \times 10^{-14}, 1.08694 \times 10^{-8}, 0.0000361436, 0.0113974, 0.988566\}$

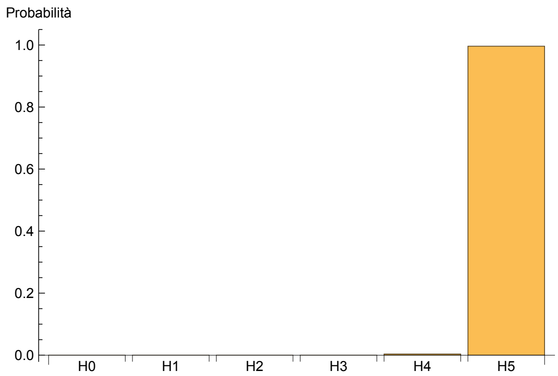


Pescaggio n. 25

Ho pescato la pallina bianca

Le nuove probabilità aggiornate dal pescaggio sono queste:

$\{0., 3.3428 \times 10^{-18}, 1.12166 \times 10^{-10}, 2.83232 \times 10^{-6}, 0.00376366, 0.996234\}$

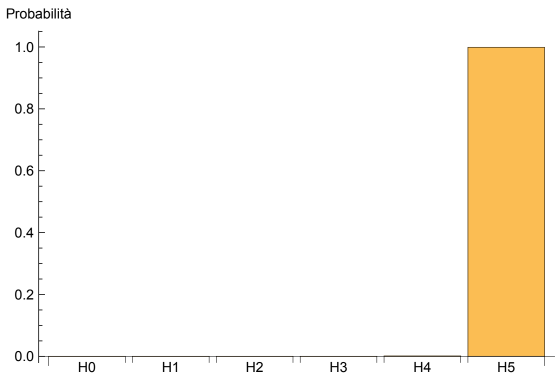


Pescaggio n. 30

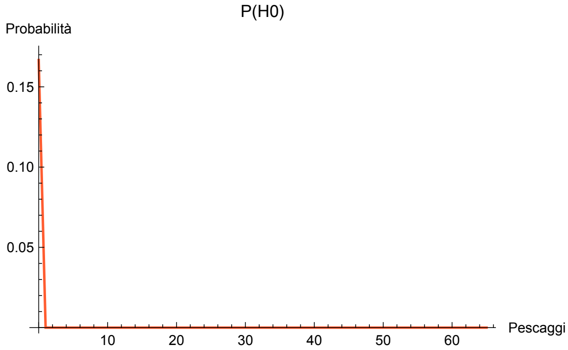
Ho pescato la pallina bianca

Le nuove probabilità aggiornate dal pescaggio sono queste:

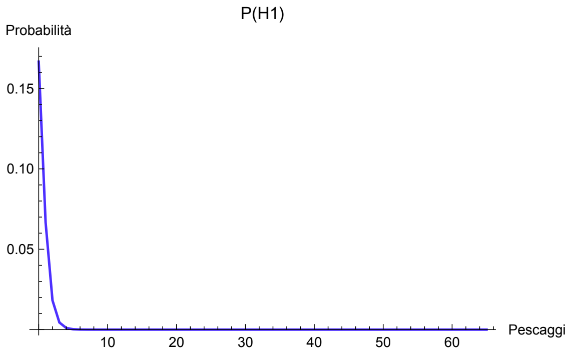
$\{0., 1.07241 \times 10^{-21}, 1.1515 \times 10^{-12}, 2.20801 \times 10^{-7}, 0.00123641, 0.998763\}$



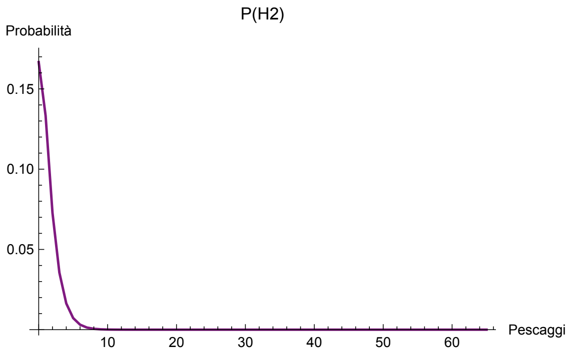
Out[380]=



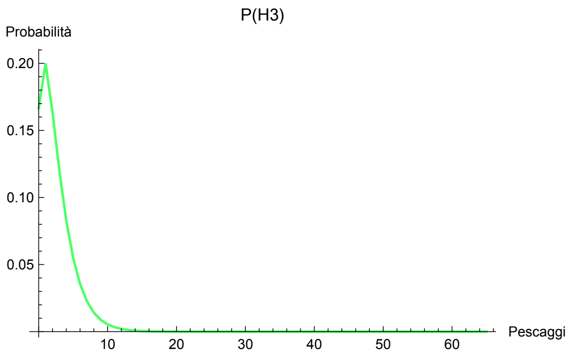
Out[381]=



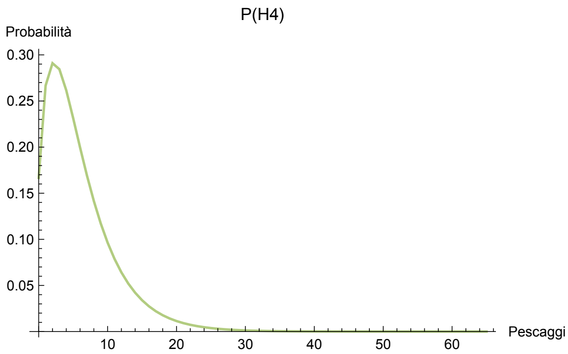
Out[382]=



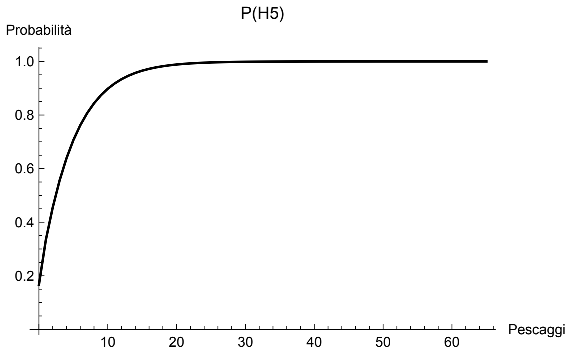
Out[383]=



Out[384]=



Out[385]=



In[386]:=

Bibliografia

[1] Giulio D'Agostini, *Bayesian Reasoning in Data Analysis: a Critical Introduction*. World Scientific Publishing Co. Pte. Ltd.,2003.

[2] Giulio D'Agostini, *Probabilità e incertezza di misura*, Dipartimento di Fisica, Università "La Sapienza", Roma 2001.<https://www.roma1.infn.it/ dagos/PRO/PRO.html>

[3] Giulio D'Agostini, *Teaching statistics in the physics curriculum: Unifying and clarifying role of subjective probability*, pp 4-8, Dipartimento di Fisica dell'Università "La Sapienza" and Istituto Nazionale di Fisica Nucleare (INFN), 1999. <https://www.roma1.infn.it/ dagos/9908014.pdf>.