Università degli studi di Milano

Corso di laurea in scienze e tecnologie fisiche

Dipartimento di Fisica Aldo Pontremoli

Problemi di statistica bayesiana

Settore Scientifico Disciplinare FIS/30

Autore

Riccardo Loddo

Anno accademico 2022/23

Indice

1 Confronto tra modelli

Supponiamo di avere un insieme di dati, che indicherò con D. Questo potrebbe essere qualsiasi insieme di misurazioni su qualsiasi dominio. Ad esempio potrebbe essere una serie temporale di punti come la quantità di glucosio nel flusso sanguigno dopo un pasto misurata ad ogni minuto. Abbiamo quindi alcuni modelli proposti che potrebbero adattarsi ai dati e vogliamo sapere qual è il modello migliore. Il modello potrebbe essere un polinomio, un'equazione differenziale o altro. Diciamo che abbiamo due modelli M1 e M2. La premessa alla base dell'inferenza bayesiana è che a qualsiasi cosa può essere assegnata una probabilità, quindi ciò che vogliamo valutare è la probabilità di un modello dati i dati D. Il modello con la probabilità più alta è quindi il modello più probabile. Quindi vogliamo valutare $P(M1|D)$ e confrontarlo con $P(M2|D)$. Un frequentista non potrebbe farlo perché solo variabili casuali possono essere assegnate probabilità e infatti non esiste un modo sistematico per valutare i modelli utilizzando la statistica frequentista. Per valutare in generale un modello possiamo utilizzare il Teorema di Bayes:

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)} \tag{1.1}$$

Questo è generalmente anche il punto in cui gli occhi delle persone iniziano a velarsi. Non riescono davvero a capire perché la semplice riscrittura della probabilità possa essere di qualche utilità. Il punto è che mentre non è chiaro come stimare la probabilità di un modello dati i dati, è solo una questione di calcolo per ottenere la probabilità che i dati nascano dato un modello, ad esempio P(D—M). Tuttavia, c'è un problema con il teorema di Bayes e cioè non è chiaro come calcolare P(D), che è la probabilità di dati marginalizzati su tutti i modelli! La finezza attorno a questo problema è guardare i rapporti di probabilità.

$$\frac{P(M1|D)}{P(M2|D)} = \frac{P(D|M1)}{P(D|M2)}\frac{P(M1)}{P(M2)} \tag{1.2}$$

dove il fattore $\frac{P(D|M1)}{P(D|M2)}$ è chiamato fattore di Bayes e questa è la quantità che vogliamo calcolare. Se le probabilità a priori dei modelli sono uguali, il fattore di Bayes è tutto ciò che serve per determinare il modello migliore. In generale, prendiamo i log del fattore di Bayes e quindi il modello che ha la più alta probabilità di log del modello (cioè $\ln P(M|D)$) è il modello migliore. Questo approccio si generalizza anche al confronto di molti modelli. Li confronti semplicemente a coppie e se usi le probabilità logaritmiche, vince il modello con la probabilità logaritmica più alta di tutti i modelli. La prossima mossa è ciò che alimenta un enorme dibattito tra bayesiani e frequentisti. Supponiamo che ogni modello dipenda da un insieme di parametri indicati con θ. Ogni volta che confrontiamo il modello con i dati, dobbiamo scegliere un'istanza dei parametri. Ad esempio, se si dispone di un modello di equazione differenziale per una serie temporale, selezionare alcuni parametri per specificare il modello, simulare l'equazione differenziale e confrontare i risultati con i dati. È quindi possibile provare altri parametri e confrontare i risultati con i dati. L'argomento bayesiano è che per calcolare P(D—M), devi marginalizzare i parametri da ottenere:

$$P(D|M) = \int P(M|D,\theta)P(\theta|M,D)d\theta \tag{1.3}$$

Che cosa significa questo? Ebbene si dice che la probabilità che i dati provengano da un modello, è la performance "media" di un modello "pesata" dalla probabilità a priori dei parametri. Quindi, ora puoi capire perché questo passaggio è così controverso. Suggerisce che un modello è buono solo quanto la probabilità a priori per i parametri del modello. Il punto di vista bayesiano è che, anche se un modello ha un insieme di parametri che fanno sì che il modello si adatti esattamente ai dati, se non hai idea di dove siano quei parametri, quel modello non è molto buono. Ma c'è ancora di più in questo. Quanto è buono un modello dipende anche da quanto "ampia" la distribuzione a priori viene confrontata con la funzione di verosimiglianza perché la "sovrapposizione" tra le due funzioni nell'integrando dell'integrale in (1.3) determina quanto è grande la probabilità. Un modo semplice per vedere questo supponiamo che ci sia un parametro e il "prior" per il parametro abbia la forma

$$P(\theta|M) = \frac{1}{\sigma}, 0 < \theta < \sigma \tag{1.4}$$

Allora la likelihood per il modello è data da

$$P(D|M) = \frac{1}{\sigma}\int_0^\sigma P(D|M,\theta)d\theta \tag{1.5}$$

Supponiamo ora che la likelihood sia una singola funzione di picco con un valore di massima likelihood di L. Quindi possiamo riscrivere la likelihood del modello come

$$P(D|M) = L(\frac{\delta}{\sigma}), \tag{1.6}$$

dove δ è solo un costante che dà una "misura" della "larghezza" della distribuzione. Vediamo quindi che la probabilità di un modello è data dalla massima likelihood, ponderata dal rapporto tra la larghezza della likelihood dei dati e il precedente $\frac{\delta}{\sigma} \leq 1$. Questo rapporto è chiamato fattore di Occam. Quando è generalizzato a k parametri, il fattore di Occam ha la forma $\left(\frac{\delta}{\sigma}\right)^k$. Quindi, vediamo che all'aumentare del numero di parametri, il fattore di Occam diminuisce e quindi i modelli con più parametri sono sempre penalizzati. Se prendiamo il log della probabilità otteniamo

$$ln(P(D|M)) \sim ln(L) - kln(\frac{\sigma}{\delta}) \tag{1.7}$$

Quindi, per confrontare due modelli, calcoliamo semplicemente la probabilità logaritmica bayesiana del modello e il modello con il valore più alto è più probabile. Se si dispone di più modelli, è sufficiente confrontare tutti i modelli tra loro a coppie e il modello con la più alta probabilità logaritmica bayesiana è il migliore. La cosa bella dell'inferenza bayesiana è che il valore di massima verosimiglianza per i parametri, la distribuzione a posteriori e il fattore di Bayes possono essere stimati in un unico calcolo MCMC.

Esempio

Due persone hanno lasciato tracce del proprio sangue sulla scena di un delitto. Un sospetto, Oliver, viene testato e si scopre che ha sangue di tipo O. I gruppi sanguigni delle due tracce risultano essere di tipo O (tipo comune nella popolazione locale, con frequenza 60%) e di tipo AB (tipo raro, con frequenza 1%). Questi dati forniscono prove a favore dell'affermazione che Oliver fosse una delle due persone presenti al crimine?
Soluzione:
In realtà il problema non è semplice da risolvere. Possiamo immaginare di avere due modelli:

- Modello 1 (S): sulla scena del crimine è presente Oliver ed uno sconosciuto.

- Modello 2 ($\bar{S}$): sulla scena del crimine erano presenti due sconosciuti.

In realtà come detto sopra ci basta fare il rapporto di questi due modelli per vedere effettivamente quale dei due è più probabile. Anzitutto assumo che il prior sia uguale per i due modelli, infatti abbiamo gli stessi dati (con dati si intende che il crimine è lo stesso e quindi i modelli si riferiscono allo stesso crimine). Dobbiamo calcolare:

$$\frac{P(D|S)}{P(D|\bar{S})} \tag{1.8}$$

La probabilità $P(D|S)$ ovvero che sulla scena del crimine ci sia Oliver ed un tizio con sangue AB è solamente la probabilità che il tizio abbia sangue AB. Infatti $P(D|S)$ esprime la probabilità che Oliver sia sulla scena del crimine (dato di fatto) e che un tizio con sangue AB sia sulla scena del crimine (dato di fatto con probabilità 1%). Quindi $P(D|S) = 1\%$. Mentre $P(D|\bar{S})$ è la probabilità che due tizi a caso siano sulla scena del crimine quindi: $P(D|\bar{S}) = 60\% \cdot 1\%$ ma in questo modo sto considerando solo che il primo tizio abbia sangue 0 e il secondo tizio abbia sangue AB ma può essere anche il viceversa, quindi $P(D|\bar{S}) = 2 \cdot 60\% \cdot 1\%$. In definitiva:

$$\frac{P(D|S)}{P(D|\bar{S})} = \frac{0.01}{2 \cdot 0.60 \cdot 0.01} = 0.83 \tag{1.9}$$

Ovvero un rapporto tra le likelihood minore di 1. Ciò implica che il secondo modello è quello più probabile, ovvero $P(D|\bar{S})$. Per calcolare la probabilità che Oliver sia effettivamente sulla scena del crimine:

$$\begin{aligned} P(S|D) &= \frac{P(D|S)P(S)}{P(D)} = \frac{P(D|S)P(S)}{P(D|S)P(S) + P(D|\bar{S}))P(\bar{S}))} = \\ &= \frac{P(D|S)}{P(D|S) + P(D|\bar{S})))} = \frac{0.01}{0.01 + 2 \cdot 0.60 \cdot 0.01} = 45.45\% \end{aligned} \tag{1.10}$$

Se si considera ora il caso in cui al posto di Oliver ci sia Alberto che ha il gruppo sanguigno di tipo AB, allora:

- Modello 3 ($S^{'}$): Alberto ed uno sconosciuto erano sulla scena del crimine.

Allora si ha che:

$$\frac{P(D|S^{'})}{P(\bar{S})} = \frac{0.6}{2 \cdot 0.6 \cdot 0.01} = 50 \tag{1.11}$$

In questo caso il rapporto è maggiore di 1 quindi il modello più favorevole è $P(D|S^{'})$ ovvero è molto probabile che Alberto sia uno dei due assassini. Infatti la probabilità che Alberto sia uno dei due assassini è:

$$\begin{aligned} P(S^{'}|D) &= \frac{P(D|S^{'})P(S^{'})}{P(D)} = \frac{P(D|S^{'})P(S^{'})}{P(D|S^{'})P(S^{'}) + P(D|\bar{S}))P(\bar{S}))} = \\ &= \frac{P(D|S^{'})}{P(D|S^{'}) + P(D|\bar{S})))} = \frac{0.6}{0.6 + 2 \cdot 0.60 \cdot 0.01} = 98\% \end{aligned} \tag{1.12}$$

Vediamo ora il caso generale. Sia n_0 le macchie di sangue di individui di tipo O e n_{AB} le macchie di sangue di tipo AB. Si ha un totale di N individui e le persone sconosciute provengono da una vasta popolazione con frazioni p_0 e p_{AB} (potrebbero esserci anche altri gruppi sanguigni). Il compito è valutare il rapporto di verosimiglianza per le due ipotesi:

- S: ”il sospetto di tipo O e gli altri sconosciuti N-1 hanno lasciato N macchie”.

- $\bar{S}$: N persone hanno lasciato N macchie".

Allora si ha che:

$$P(n_0, n_{AB}|S) = \frac{(N-1)!}{(n_0-1)!n_{AB}!} p_0^{n_0-1} p_{AB}^{n_{AB}} \tag{1.13}$$

mentre:

$$P(n_0, n_{AB}|\bar{S}) = \frac{N!}{n_0!n_{AB}!} p_0^{n_0} p_{AB}^{n_{AB}} \tag{1.14}$$

Da cui il rapporto:

$$\frac{P(n_0, n_{AB}|S)}{P(n_0, n_{AB}|\bar{S})} = \frac{n_0}{Np_0} \tag{1.15}$$

2 MCMC

Supponiamo di avere dati $D(t_i)$, che consistono in alcune misurazioni (scalari o vettoriali) in punti temporali discreti t_i e un modello proposto, che produce una serie temporale $y(t|\theta)$, dove θ rappresenta un insieme di parametri liberi che cambia la funzione. Questo modello potrebbe essere un sistema di equazioni differenziali o solo una funzione proposta come un polinomio. L'obiettivo è trovare l'insieme di parametri che meglio si adatta ai dati e valutare quanto sia buono il modello. Ora, il modo corretto per farlo è utilizzare l'inferenza bayesiana e il confronto del modello, che può essere calcolato utilizzando MCMC. Tuttavia, l'MCMC può essere utilizzato anche solo per ottenere i parametri nel senso di trovare l'adattamento migliore secondo alcuni criteri.

3 Catene di Markov

Il Campionamento funziona generando una serie di punti che sono distribuiti secondo una distribuzione d'interesse (nel nostro caso la distribuzione a posteriori). Supponiamo di avere una serie di punti distribuiti secondo P(x) che conosciamo. P(x) vive nello spazi n-dimensionale mentre g(x) vive in uno spazio m dimensionale.

4 Metodo di Laplace

L'idea alla base dell'approssimazione di Laplace è semplice, ovvero è una tecnica usata per approssimare integrali. Assumiamo che una densità di probabilità non normalizzata $P^*(x)$, la cui costante di normalizzazione è descritta in formula (4.1), abbia un picco in un punto x_0 (Figura (1)):

$$Z_P \equiv \int P^*(x) dx \tag{4.1}$$

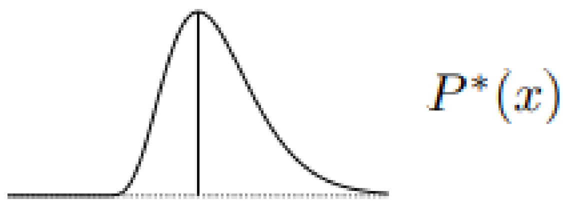

Grafico 1: Distribuzione non normalizzata $P^*(x)$.

Consideriamo ora lo sviluppo di Taylor:

$$ln(P^*(x)) \sim ln(P^*(x_0)) - \frac{c}{2}(x - x_0)^2 + ..., \tag{4.2}$$

dove:

$$c = -\frac{\partial^2}{\partial x^2} ln(P^*(x))\Big|_{x=x_0} \tag{4.3}$$

La $P^*(x)$ allora è diventata:

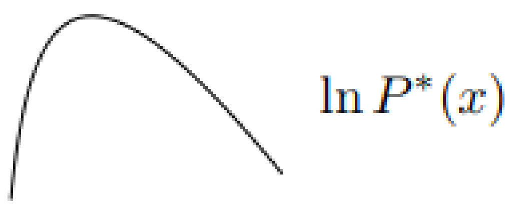

Grafico 2: Distribuzione di Taylor di $P^*(x)$, ovvero $ln(P^*(x))$.

Possiamo approssimare ora che $P^*(x)$ da una gaussiana non normalizzata:

$$Q^*(x) \equiv P^*(x_0) exp\left[-\frac{c}{2}(x - x_0)^2\right] \tag{4.4}$$

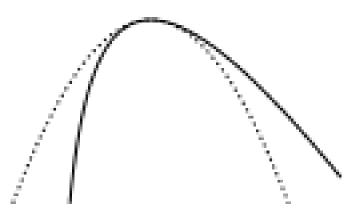

Grafico 3: Distribuzione di $ln(P^*(x))$ e $Q^*(x)$.

e dunque di conseguenza la normalizzazione Z_P la approssimiamo ad una normalizzazione della gaussiana:

$$Z_Q = P^*(x_0)\sqrt{\frac{2\pi}{c}} \tag{4.5}$$

da cui finalmente possiamo rappresentare la Gaussiana $Q^*(x)$ che approssima la $P^*(x)$:

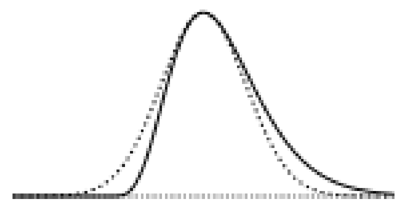

Grafico 4: Distribuzione di $P^*(x)$ e $Q^*(x)$.

Ovviamente il tutto si può generalizzare in k dimensioni. L'approssimazione di Laplace è "base-dipendente": se x viene trasformato in una funzione non lineare u(x) e la densità viene trasformata in P(u) = P(x)—dx/du— allora in generale le costanti di normalizzazione approssimative ZQ saranno diverse. Questo può essere visto come un difetto poiché il vero valore ZP è indipendente dalla base perché possiamo cercare una scelta di base in cui l'approssimazione di Laplace è più accurata.

5 Rasoio di Occam

Il rasoio di Occam è un principio metodologico che indica di scegliere, tra più soluzioni possibili di un problema, quella più semplice. Se diverse spiegazioni sono compatibili con una serie di osservazioni, il rasoio di Occam ci consiglia di acquistare il più semplice. Questo principio è spesso sostenuto per uno dei due motivi: il primo è estetico ("È più probabile che una teoria con una bellezza matematica sia corretta di una brutta che si adatta ad alcuni dati sperimentali". Il secondo motivo è il successo stesso di tale metodologia. Il rasoio di Occam rappresenta il fatto che i modelli più complessi sono quelli meno probabilità.

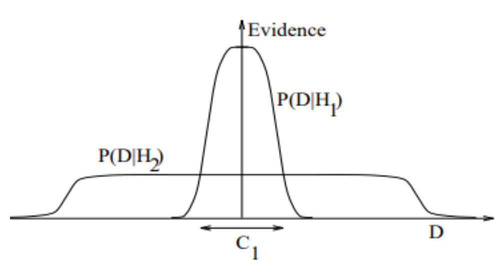

Grafico 5: Rasoio di Occam.

L'asse orizzontale rappresenta lo spazio dei possibili set di dati. Il teorema di Bayes premia i modelli in proporzione a quanto hanno previsto i dati che si sono verificati. Questa previsione sono quantificate da una distribuzione di probabilità normalizzata sui dati. Supponiamo un modello semplice H1 in un intervallo limitato di previsioni. Supponiamo ora un modello più potente H2 che è in grado di prevedere una maggiore varietà di insiemi dati. Ciò significa che H2 non prevede i set di dati nella regione di H1 con la stessa forza che lo fa H1. Se il set di dati cade nella regione di H1 il modello meno potente (quindi H1) sarà quello più probabille.

6 Probabilità marginale e probabilità condizionata

- **Probabilità marginale**: la probabilità congiunta è definita come la probabilità che due eventi X ed Y avvengano *congiuntamente* ovvero avvengano nello stesso momento, che avvengano "insieme". Per esempio possiamo supporre di lanciare due dadi contemporaneamente e ci chiediamo qual'è la probabilità che esca $\{2,6\}$. In questo caso il lancio dei due dadi è un **evento indipendente** ovvero il primo lancio non influenza il secondo lancio (almeno che i dadi non si scontrino ma trascuriamo tale situazione) per cui la probabilità che escano $\{2,6\}$ è il prodotto delle singole probabilità: $P(x,y) = P_x(x)P_y(y) = \frac{1}{36}$. Tuttavia non è detto che sia sempre così ovvero le variabili non sono sempre indipendenti. Infatti supponiamo di dividere la popolazione delle persone a seconda del colore dei capelli e del colore degli occhi. La probabilità che presa una persona a caso abbia occhi celesti e sia biondo non è la probabilità che abbiamo gli occhi celesti per la probabilità che sia biondo. Il colore degli occhi e dei capelli sono altamente correlati tra loro. In generale quindi si definisce la probabilità marginale per l'evento y:

$$P(y) = \sum_x P(x,y) \tag{6.1}$$

  la probabilità dell'evento y, considerando che è correlata con l'evento x. Per esempio: qual'è la probabilità che una persona abbia gli occhi azzurri? Sarà la probabilità che la persona abbia gli occhi azzurri congiunta al fatto che ha i capelli marroni, più la probabilità che la persona abbia gli occhi azzurri congiunta al fatto che ha i capelli biondi..e così via. Tale probabilità presne il nome di **probabilità marginale**.

- **Probabilità condizionata**: la probabilità condizionata è la probabilità che accada (si verifichi) l'evento x dato il fatto che si è verificato (conosciamo) l'evento y. Essa è definita come:

$$P(x|y) = \frac{P(x,y)}{P(y)} = \frac{P(x,y)}{\sum_x P(x,y)} = \frac{P(x \cap y)}{P(y)} \tag{6.2}$$

  dove $P(x \cap y)$ è la probabilità che avvenga x ed y ed è pari a:

$$P(x \cap y) = P(x) + P(y) - P(x \cup y)$$

  con $P(x \cup y)$ la probabilità che avvenga x o y. Attenzione che la p(y) non è 1. Ovvero la P(y) a denominatore esprime la probabilità che avvenga y (senza sapere che è avvenuto). Infatti $P(x|y)$ è la probabilità di x DATO y, ma P(y) è la probabilità della sola y. Un altra cosa fondamentale da osservare è che $P(x|y) \neq P(y|x)$

- **Regola del prodotto**: avendo dato la definizione di probabilità condizionata si può definire la chain rule ovvero la regola della catena:

$$P(x,y) = P(x|y)P(y) = P(y|x)P(x) \tag{6.3}$$

- **Regola della somma**: dalla definizione di probabilità marginale si può definire:

$$P(y) = \sum_x P(y|x)P(x) \tag{6.4}$$

Problema 1

Charlie ha due fratelli, Alf e Bob.

a) Qual'è la probabilità che che Charlie sia più vecchio di Bob?

b) Charlie dice di essere più vecchia di Alf: qual'è la probabilità che Charlie sia più vecchia anche di Bob? **Soluzione:**

La domanda a) è molto semplice in quanto prende in considerazione solamente due dei tre fratelli. Non sappiamo nulla sul terzo, semplicemente ci chiede la probabilità che su tre fratelli uno sia più vecchio di un altro (il terzo fratello può anche non esistere). Dalla definizione di probabilità classica sappiamo che:

$$P = \frac{\text{n° casi favorevoli}}{\text{n° casi possibili}}$$

per cui nel nostro caso si ha semplicemente che il numero di casi favorevoli è 1 ovvero Charlie è più vecchia di Bob mentre i casi possibili sono che Charli sia più vecchia di Bob oppure che Bob sia più vecchia di Charlie quindi in totale due. Ne segue che la probabilità che Charlie sia più vecchio di Bob è:

$$\frac{\text{n° casi favorevoli}}{\text{n° casi possibili}} = \frac{1}{2}$$

Per la domanda b) invece le cose sono diverse: ci chiede infatti qual'è la probabilità che Charlie sia più vecchia di Bob sapendo che Charlie è più vecchia di Alf. In questo caso dobbiamo quindi utilizzare la probabilità condizionata:

$$P(C > B|C > A) = \frac{P(C > B, C > A)}{P(C > A)}$$

La probabilità $P(C > B, C > A)$ che Charlie sia più grande di Bob e di Alf (ovvero che sia più grandi di tutti e due) è ovviamente $\frac{1}{3}$ ovvero i casi possibili sono che Charlie sia più grande di entrambi mentre i casi possibili sono che Charli sia più grande di Bob, o di Alf o di entrambi quindi in totale 3. mentre $P(C > A)$ è la probabilità che Charlie sia piu vecchio di A. Questa probabilità è uguale alla richiesta a) ovvero la probabilità che un fratello sia piu grande dell'altro (non mi interessa sulla di Bob). Quindi in totale:

$$P(C > B|C > A) = \frac{P(C > B, C > A)}{P(C > A)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Problema 2

Gli abitanti di un isola dicono il vero per un terzo delle volte (mentono per i due terzi delle volte). Un'abitante esprime una dichiarazione. Chiediamo ad un altro abitante dell'isola se l'affermazione che ha fatto il primo abitante sia vera. La risposta che ci viene data è "si". Qual'è la probabilità che l'affermazione fatta dal primo abitante sia vera?

Soluzione:

In questo esempio si capisce molto chiaramente che le due affermazioni fatti dai due abitante dell'isola sono fortemente correlate. Partiamo dal fatto che la risposta che ci viene data è "si". Questa NON è vera per $\frac{1}{3}$ delle volte (o falsa per $\frac{2}{3}$) per il seguente motivo. Il primo abitante fa una dichiarazione: essa può essere vera per $\frac{1}{3}$ oppure po essere falsa per $\frac{2}{3}$. Supponiamo che abbia realmente detto il vero (quindi siamo nella frazione $\frac{1}{3}$ delle volte). Allora il secondo abitante, che sappiamo dice si, non sta mentendo neanche lui. Ovvero se il primo tizio dice la verità allora la dice anche il secondo perché dice ”si”. Se il primo tizio dice il falso allora anche il secondo dice il falso perché dicendo ”si” ci sta mentendo (gli abbiamo chiesto se fosse vera la dichiarazione). Quindi:

$$P(\text{Dichiarazione vera}|\text{Si}) = \frac{P(\text{Dichiarazione vera}, \text{Si})}{P(Si)}$$

dove applicando la chain rule a numeratore e la sum rule a denominatore:

$$P(\text{Dichiarazione vera}, \text{Si}) = P(\text{Si}|\text{Dichiarazione vera})P(\text{Dichiarazione vera}) = \frac{1}{3}\frac{1}{3} = \frac{1}{9}$$

$$P(\text{Si}) = P(\text{Si}|\text{D. vera})P(\text{D. vera}) + P(\text{Si}|\text{D.falsa})P(\text{D. falsa}) = \frac{1}{3}\frac{1}{3} + \frac{2}{3}\frac{2}{3} = \frac{5}{9}$$

da cui la probabilità che il primo tizio dica il vero è:

$$P(\text{Dichiarazione vera}|\text{Si}) = \frac{1}{5}$$

Problema 3

In un sacchetto ci sono 100 monete. Una moneta sulle 100 è truccata ovvero ha testa da tutte e due le parti. Una moneta viene presa casualmente dal sacchetto.

a) Qual’è la probabilità di ottenere due teste?

b) Qual’è la probabilità di ottenere due croci?

Soluzione:

In questo caso l’esercizio è più facile del solito. Infatti basta la regola della somma. Ci chiede infatti di calcolare la probabilità che presa una moneta a caso, e fatta lanciare, esca due volte testa. Se le monete fossero tutte normali, allora la probabilità di pescare una moneta è 1 (sono tutte uguali) e la probabilità di ottenere due volte testa sarebbe $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Nel nostro caso invece la probabilità di estrarre una moneta normale sono i casi favorevoli (quindi 99) rispetto ai casi totale (100). Una volta pescata la moneta normale essa ha probabilità $\frac{1}{4}$ di far uscire due volte testa. Se si pesca invece la moneta truccata (e la probabilità che si pesca quella truccata è $\frac{1}{100}$), allora avremmo probabilità 1 che esca testa. In formule:

$$P(\text{Ottenere testa due volte}) = \frac{99}{100} \cdot \left(\frac{1}{2}\right)^2 + \frac{1}{100} \cdot 1 = 0.2575$$

mentre la probabilità di ottenere due volte croce invece è:

$$P(\text{Ottenere croce due volte}) = \frac{99}{100} \cdot \left(\frac{1}{2}\right)^2 + \frac{1}{100} \cdot 0 = 0.2475$$

Osserviamo che $\left(\frac{1}{2}\right)^2$ indica la probabilità che esca due volte testa (o due volte croce) sapendo che è uscita la moneta normale.

6.1 Problema 4

Il signor Rossi è risultato positivo al tampone rapido del covid. Il test ha un affidabilità del 95% ovvero si ha la probabilità del 5% di avere un falso positivo (il tampone è positivo ma il signor Rossi non ha il covid) ed il 5% di avere un falso negativo (tampone negativo ma il signor Rossi ha il covid). Sappiamo

inoltre che 1% della popolazione dove vive il signor Rossi ha il covid. Qual'è la probabilità che il signor Rossi abbia il covid?
Soluzione:

Definiamo con p il fatto che Rossi sia positivo, con m il fatto che è malato, con n il fatto che Rossi è negativo e con s il fatto che sia sano. Allora:

- $P(p|m) = 95\%$, $P(p|s) = 5\%$, $P(n|m) = 5\%$, $P(n|s) = 95\%$, $P(m) = 1\%$, $P(s) = 99\%$

A noi interessa $P(m|p)$. Usando la probabilità condizionata possiamo ricavare che:

$$P(m|p) = \frac{P(p,m)}{P(p)}$$

La probabilità che rossi sia positivo unita al fatto che il test sia positivo è: $P(p,m) = P(p|m)P(m) = 0.95 \cdot 0.01 = 0.0095$, dove è stata applicata la regola della catena. Mentre $P(p)$ è la probabilità che il test sia positivo ovvero: $P(p) = \sum_x P(p|x)P(x)$ dove x gioca il ruolo di malato o sano: $x = \{m, s\}$. Quindi $P(p) = \sum_x P(p|x)P(x) = P(p|m)P(m) + P(p|s)P(s) = 0.95 \cdot 0.01 + 0.05 \cdot 0.99 = 0.059$. In definitiva si ha che:

$$P(m|p) = \frac{0.0095}{0.059} = 0.161 \sim 16\%$$

Ovvero anche se il test è risultato positivo, la probabilità che il signor Rossi sia realmente covidizzato è solo del 16%.

7 Teorema di Bayes è calcolo combinatorio

Il punto di vista bayesiano si basa sul fatto che noi abbiamo un certo pensiero...crediamo in qualche modo che sia così. Per esempio la condanna di un presunto assassino...o il fatto che una firma sia falsa o meno...Le regole di probabilità assicurano che se due persone fanno le stesse ipotesi e ricevono gli stessi dati, trarranno conclusioni identiche. Questo uso più generale della probabilità per quantificare credenze è noto come il punto di vista bayesiano. I sostenitori di un approccio bayesiano alla modellazione dei dati e al riconoscimento dei modelli non vedono questa soggettività come un difetto, poiché a loro avviso, non è possibile fare inferenze senza fare ipotesi. Supponiamo che D siano i dati e θ sia il modello fisico che si vuole seguire. Conosciamo la distribuzione dei dati D rispetto a θ ($P(D|\theta)$) ma vorremmo calcolare la distribuzione di θ dato i dati ($P(\theta|D)$). Ovvero vorremmo trovare qual'è la probabilità che il modello fisico che stiamo utilizzando rispecchia effettivamente i dati. Il Teorema di Bayes è definito come:

$$P(\theta|D) = \frac{P(D|\theta) \cdot P(\theta)}{P(D)} \tag{7.1}$$

dove:

- $P(D|\theta)$ è definita come la **likelihood** ovvero una descrizione del problema fisico associata ai risultati

- $P(\theta)$ è la nostra convinzione sui valori dei parametri prima di eseguire qualsiasi misurazione. Prende il nome di Prior.

- $P(\theta|D)$ è la nostra convinzione aggiornata sui valori dei parametri. Prende il nome di Posterior.

Ricordo inoltre che in generale dato un ”esperimento qualsiasi” (per esempio lancio di un dado) in cui un certo evento abbia una probabilità p di realizzarsi, se si ripete l’esperimento N volte la probabilità che tale evento si verifichi n volte è:

$$P = \binom{N}{n} p^n (1-p)^{N-n} \tag{7.2}$$

7.1 Problema 5

Un urna contiene K palline di cui B sono nere e W = K-B sono bianche. Fred pesca una pallina a caso dall’urna e poi la rimette nell’urna da dove l’ha pescata. Questa operazione Fred la esegue N volte. Qual’è la distribuzione di probabilità per le palline nere? Ovvero, qual’è la distribuzione di probabilità che su N estrazioni escano n palline nere?

Soluzione:

Chiaramente è un esperimento ripetuto N volte. La probabilità che esca una pallina nera è $\frac{B}{K} \equiv f$. Allora dalla formula (7) si ha che la distribuzione di probabilità per le palline nere:

$$P(n|N,f) = \binom{N}{n} f^n (1-f)^{N-n}$$

7.2 Problema 6

Ci sono 11 urne ed ogni urna contiene 10 palline. L’urna u-esima contiene u palline nere e 10-u palline bianche. Fred seleziona un urna a caso e la passa a Bill. Bill pesca una pallina e poi la rimette dentro (operazione fatta 10 volte). Se sono uscite 3 volte le palline nere e 7 volte quelle bianche,

a) Qual’è la probabilità che l’urna presa da fred sia la terza?

b) Qual’è la probabilità che la prossima pallina pescata sia nera?

Soluzione:

Siamo in un esperimento ripetuto. Anzitutto capiamo il problema: la prima urna (urna 0) contiene 0 pallina nera e 10 bianche. L’urna 2 contiene 2 palline nere e 8 bianche e così via. L’urna 10 conterrà 10 palline nere e 0 bianche. Ovviamente la probabilità di estrarre la terza urna è $\frac{1}{11}$. L’obbiettivo è trovare la probabilità che l’urna sia la terza sapendo che sono uscite 3 palline nere e 7 bianche e sapendo che l’urna è scelta a caso. In formule:

$$P(n=3|3,7,\text{caso}) = \frac{P(n=3,3,7,\text{caso})}{P(3,7,\text{caso})}$$

La probabilità che dall’urna tre si estraggano 3 palline nere e 7 bianche la conosciamo in quanto la probabilità che esca una pallina nera è $\frac{3}{10} \equiv f$ mentre che esca la pallina bianca è $\frac{7}{10} \equiv g$. Il fatto che escano 3 palline nere e il resto non sono nere è uguale al fatto che escano 7 palline bianche ed il resto non sono bianche, quindi abbiamo un informazione che non ci interessa. Ovvero:

$$\binom{10}{3}\left(\frac{3}{10}\right)^3\left(1-\frac{3}{10}\right)^{10-3} = \binom{10}{7}\left(\frac{7}{10}\right)^7\left(1-\frac{7}{10}\right)^{10-7} = 0.0022235661$$

Si ha quindi che:

$$P(n=3,3,\text{caso}) = P(n=3,3,7\text{caso}) = P(n=3,7,\text{caso}) = \frac{1}{11}\binom{10}{3}\left(\frac{3}{10}\right)^3\left(1-\frac{3}{10}\right)^{10-3}$$

Manca il termine $P(3,7,\text{caso})$ non considerando l'urna 3. Ovvero la probabilità che, per ogni urna, accada che escono 3 palline nere e 7 bianche.

$$P(3,7,\text{caso}) = P(3,\text{caso}) = P(7,\text{caso}) = \sum_x \frac{1}{11}\binom{10}{3}\left(\frac{x}{10}\right)^3\left(1-\frac{x}{10}\right)^{10-3} = \sum_x \frac{1}{11}\binom{10}{7}\left(\frac{x}{10}\right)^7\left(1-\frac{x}{10}\right)^{10-7}$$

Mettendo le informazioni all'interno della $P(n=3|3,7,\text{caso})$ si ottiene:

$$P(n=3|3,7,\text{caso}) = \frac{\frac{1}{11}\binom{10}{3}\left(\frac{3}{10}\right)^3\left(1-\frac{3}{10}\right)^{10-3}}{\sum_x \frac{1}{11}\binom{10}{3}\left(\frac{x}{10}\right)^3\left(1-\frac{x}{10}\right)^{10-3}} = \frac{\binom{10}{7}\left(\frac{7}{10}\right)^7\left(1-\frac{7}{10}\right)^{10-7}}{\sum_x \frac{1}{11}\binom{10}{7}\left(\frac{x}{10}\right)^7\left(1-\frac{x}{10}\right)^{10-7}}$$

La probabilità che la prossima pallina sia nera sapendo che ne sono uscite già 3 è, dalla regola della somma:

$$P(nera|N=10,B=3,W=7,caso) = \sum_x P(x|B=3,N=10)p(nera|x,)$$

ovvero fissata l'urna la probabilità è fissata.

7.3 Problema 7

Bill lancia una moneta piegata N volte ottenendo una sequenza di teste e croci. Assumendo che la moneta abbia probabilità f di uscire testa

a) Qual'è la distribuzione di probabilità per f se sono uscite n teste in N lanci?

b) Qual'è la probabilità che all'N+1 esimo lancio esca testa?

Soluzione:

Questo problema è molto importante in quanto si capisce che cosa vuol dire analizzare un problema con inferenza bayesiana. Ci interessa la distribuzione di probabilità per f sapendo che sono uscite n teste in N lanci, ovvero ci interessa $P(f|n,N)$. Già vediamo l'analogo alla definizione del Teorema di Bayes. Il modello è la nostra f e i dati sono n ed N. Possiamo quindi applicare il teorema:

$$P(f|n,N) = \frac{P(n|f,N)P(f)}{P(n,N)}$$

Cerchiamo di calcolare ogni singola componente: $P(n|f,N)$ è la probabilità che escano n teste sapendo che si sono fatti N lanci e che la distribuzione di probabilità è f. Ma allora è semplicemente:

$$P(n|f,N) = \binom{N}{n} f^n (1-f)^{N-n}$$

Ora bisogna calcolare $P(f)$ ma è qui la potenza del Teorema di Bayes: possiamo infatti decidere di mettere qualsiasi cosa con Prior. Ovvero è un qualcosa che noi pensiamo che sia cosi. Per me la probabilità di f è pari ad 1 (completamente a caso ma per semplicità). Manca la $P(n,N)$ ovvero la probabilità che escano n teste in N lanci non sapendo nulla sulla distribuzione f. Allora se fosse un caso discreto avremmo una sommatoria su tutte le f. In questo caso essendo un caso continuo avremmo l'integrale su tutte le f. In particolare:

$$P(n,N) = \int_0^1 \binom{N}{n} f'^n (1-f')^{N-n} df'$$

Ricordando l'integrale beta il quale è definito come: $\int_0^1 x^n(1-x)^m dx = \frac{n!m!}{(n+m+1)!}$ si ha che:

$$P(n,N) = \binom{N}{n}\frac{n!(N-n)!}{(n+N-n+1)!} = \frac{n!(N-n)!}{(N+1)!}$$

da cui quindi:

$$P(f|n,N) = \frac{\binom{N}{n} f^n(1-f)^{N-n}}{\binom{N}{n}\frac{n!(N-n)!}{(N+1)!}} = \boxed{\frac{f^n(1-f)^{N-n}}{n!(N-n)!}(N+1)!}$$

Per rispondere alla domanda b invece, è del tutto analoga al problema precedente ovvero, la probabilità che esce testa al prossimo lancio sapendo che ne sono già uscite n è, dalla regola della somma:

$$P(n|N+1) = \int P(f|n,N)P(n) = \int_0^1 \frac{f^n(1-f)^{N-n}}{n!(N-n)!} f(N+1)!df = \boxed{\frac{n+1}{N+2}}$$

dove $P(n)$ è la probabilità che esca testa, ovvero è f.

7.4 Problema 8

Due dadi normali sono lanciati:

a) Qual'è la distribuzione di probabilità per la somma dei valori?

b) E per la differenza?

Soluzione:

Sappiamo che la probabilità che esca un determinato numero lanciando un dado è $\frac{1}{6}$. La somma di due dadi può essere un numero che è compreso tra 2 e 12. Tuttavia il numero due ha probabilità $\frac{1}{36}$ poiché deve uscire due volte 1. Analogamente il numero 12. In generale questo problema si risolve con metodi computazionali (in realtà si può fare anche a mano e vedere la frequenza di quante volte esce un numero). Analogamente per la differenza. In questo caso la differenza sarà un numero compreso tra

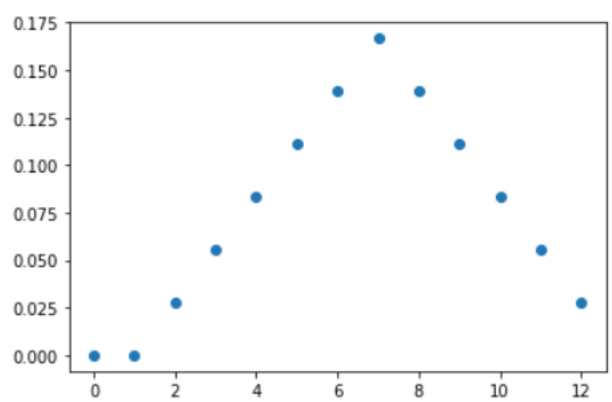

Grafico 6: Distribuzione di probabilità per la somma. Il valore più probabile è 7.

0 e 5.

8 Prosecutor's fallacy

L'errore del pubblico ministero è un errore del ragionamento statistico che coinvolge un test per un evento, come una corrispondenza del DNA. Un risultato positivo nel test può paradossalmente essere più probabile che sia un risultato errato che un evento reale, anche se il test è molto accurato. L'errore prende il nome dal fatto che viene tipicamente utilizzato da un pubblico ministero per esagerare la probabilità della colpevolezza di un imputato criminale. L'errore può essere utilizzato anche per supportare altre affermazioni, inclusa l'innocenza di un imputato.

8.1 Problema 9

Un campione di DNA della scena del crimine viene confrontato con un database di 20.000 uomini. Viene trovata una corrispondenza tra uno dei 20000 uomini ed il campione e quindi l'uomo viene accusato. Al processo viene testimoniato che la probabilità che due profili DNA corrispondano per caso è solo 1 su 10.000. Qual'è la probabilità di trovare un falso positivo in città?
Soluzione: Sembrerebbe che, dato che la probabilità che il DNA coincida per caso è molto piccola ($\frac{1}{10000}$), allora la probabilità che l'uomo sia l'assassino è $\frac{9999}{10000}$ ovvero molto probabile. Tuttavia non è cosi! Il problema ci chiede qual'è la probabilità che, l'uomo sia innocente sapendo però che il test coincide. Ovvero qual'è la probabilità che l'evento sia successo per caso. La probabilità che due profili del DNA corrispondano casualmente dopo un controllo è:

$$\frac{1}{10000} \tag{8.1}$$

Da cui la probabilità che non c'è corrispondenza è:

$$1-\frac{1}{10000} \tag{8.2}$$

La probabilità di non corrispondenza, dopo 20.000 controlli, è di:

$$\left(1-\frac{1}{10000}\right)^{20000} \tag{8.3}$$

e quindi la probabilità di corrispondenza, dopo 20.000 controlli è:

$$1-\left(1-\frac{1}{10000}\right)^{20000} \sim 86\% \tag{8.4}$$

L'idea è che più controlli faccio più aumenta la probabilità di trovare una corrispondenza per caso. Infatti trovare una corrispondenza per caso è $\frac{1}{10000}$, quindi mi aspetto che su 10000 controlli ci sia almeno una persona che è risultata positiva pur non essendo l'assassino. Ciò non vuol dire che la persona è al 99,99% colpevole, anzi!!! Se si fossero fatti solo 10 controlli al posto che 10000 allora la questione sarebbe diversa. Infatti con solo 10 controlli, la probabilità che una impronta coincida per caso è davvero molto bassa. Ma avendo fatto 20000 controlli, la probabilità che una persona coincida non è bassa per niente, anzi è del 86%. Quindi in definitiva la probabilità che l'uomo sia innocente è del 86%.

8.2 Problema 10

É stato commesso un omicidio. La polizia ha chiamato 10000 persone ed ha confrontato le loro impronte digitali con quella sulla scena del crimine. Un' uomo (tra questi 10000) ha l'importata digitale compatibile con quella sulla scena del crimine (la probabilità che ciò avvenga per caso è di $\frac{1}{20000}$).

a) Qual'è la probabilità di trovare **almeno** un falso positivo in città?

b) Qual'è la probabilità di trovare **esattamente** un solo falso positivo in città?

c) Qual'è la probabilità che l'uomo sia il ladro, se l'impronta coincide?

Soluzione: Con falso positivo si intende la probabilità che l'uomo ha l'impronta che coincide ma non è l'assassino. Mentre con falso negativo si intende che l'impronta non coincide ma l'uomo è il ladro. Analogamente al problema precedente, questa volta è leggermente diverso, ovvero: prima avevamo che la probabilità che il DNA corrispondesse era $\frac{1}{10000}$, ora è $\frac{1}{20000}$ quindi ancora meno probabile. Mi aspetto che se faccio un numero alto di controlli, aumenti la probabilità di trovare un falso positivo. Inoltre mi aspetto che, essendo meno probabile il fatto che le impronte coincidano e che faccio solo 10000 controlli, non mi aspetto assolutamente di ottenere una probabilità maggiore del problema precedente. Infatti, la probabilità che le impronte coincidano per caso, **nel singolo controllo**, è:

$$\frac{1}{20000} \tag{8.5}$$

Quindi la probabilità che le impronte non coincidono (sempre nel singolo controllo) è:

$$1-\frac{1}{2000} \tag{8.6}$$

La probabilità che le impronte non coincidano dopo 10000 controlli è:

$$\left(1-\frac{1}{20000}\right)^{10000} \tag{8.7}$$

Da cui la probabilità che le impronte coincidano dopo 10000 controlli è:

$$1-\left(1-\frac{1}{20000}\right)^{10000} \sim 39,3\% \tag{8.8}$$

Attenzione, questa è la probabilità di trovare **almeno** un falso positivo. La probabilità di trovare esattamente un falso positivo invece è diversa. Ovvero la probabilità di avere un solo falso positivo su 10000 è data dalla probabilità del successo (uno solo in questo caso $\frac{1}{20000}$) (formula (7)). Infatti è semplicemente un evento ripetuto e ci chiede qual'è la probabilità esatta che accade ciò.

$$\binom{10000}{1}\left(\frac{1}{20000}\right)^{1}\left(1-\frac{1}{20000}\right)^{9999} \sim 30\% \tag{8.9}$$

Analogamente la probabilità che ci siano esattamente 4 falsi positivi (mi aspetto sia molto bassa) è di:

$$\binom{10000}{4}\left(\frac{1}{20000}\right)^{4}\left(1-\frac{1}{20000}\right)^{9996} \sim 0.16\% \tag{8.10}$$

La probabilità che l'uomo sia il ladro bisogna applicare il Teorema di Bayes:

$$P(Ass|Coinc)=\frac{P(Coinc|Ass)P(Ass)}{P(Coinc)} \tag{8.11}$$

e sarebbe compito della giuria dire quanto vale $P(Ass)$, in particolare se siamo sicuri che il ladro è tra le persone chiamate dalla polizia è $\frac{1}{10000}$.

9 Cambio di variabili

In alcuni problemi è bene ricordare la distribuzione di Poission:

$$P_\lambda(x) = \frac{\lambda^n}{n!} e^{-\lambda} \tag{9.1}$$

Supponiamo ora di avere il seguente problema: Supponiamo X una variabile casuale distribuita secondo f_X. Definiamo Y anch'essa una variabile casuale (deterministica) in funzione di X attraverso la funzione g: $g(X) = Y$ dove $g : \mathbb{R} \longrightarrow \mathbb{R}$. Allora la distribuzione f_Y può essere calcolata come:

$$f_Y(y) = \int \delta(y - g(x)) f_X(x) dx \tag{9.2}$$

Tuttavia tale integrale vale per variabili in n dimensioni. Si può dimostrare infatti che se g è una funzione monotona crescente o decrescente allora, **in una dimensione, vale che**:

$$f_Y(y) = \frac{f_X(x)}{|g'(x)|} \tag{9.3}$$

9.1 Problema 11

In un box ci sono 1000000 monete ed una di esse ha testa da entrambe le parti. Una moneta viene estratta casualmente e lanciata 15 volte ottenendo 15 teste. Qual'è la probabilità che la moneta sia quella truccata?

Soluzione:

Sicuramente è una azione ripetuta: sappiamo che la probabilità di estrarre la moneta truccata è:$\frac{1}{1000000}$ mentre la probabilità di estrarre una moneta normale è di: $\frac{99999}{1000000}$. Supponiamo quindi che è uscita una moneta normale, allora la probabilità che esca 15 volte testa è:

$$P(n = 15|N = 15, norm) = \frac{99999}{1000000} \left(\frac{1}{2}\right)^{15}.$$

Mentre se la moneta è truccata allora si ha:

$$P(n = 15|N = 15, f) = \frac{1}{1000000} (1)^{15}.$$

La probabilità che la moneta sia quella falsa è:

$$P(f|n = 15, N = 15) = \frac{P(n = 15 \cap N = 15, f)}{P(n = 15|N = 15)} = \frac{P(n = 15 \cap N = 15, f)}{P(n = 15|N = 15, norm) + P(n = 15|N = 15, f)}$$
$$= \frac{\frac{1}{1000000}}{\frac{99999}{1000000}\left(\frac{1}{2}\right)^{15} + \frac{1}{1000000}} = 0.031728.$$

9.2 Problema 12

Un urna A contiene 5 palline: 1 nera, 2 bianche, 1 verde ed 1 rosa. Un urna B contiene 500 palline: 200 nere, 100 bianche, 50 gialle, 40 blu 30 marroni, 25 verdi, 25 argento, 20 oro, 10 rosse. Un'urna è selezionata a caso, e viene pescata una pallina. La pallina è nera. Qual'è la probabilità che l'urna selezionata sia la A?

Soluzione:

Analogamente al problema precedente: ci interessa la probabilità che l'urna estratta sia la A sapendo che è uscita una pallina mera e che l'urna è stata estratta a caso, quindi $P(A|black, caso)$. Essa, secondo la definizione di probabilità condizionata è:

$$P(A|black, caso) = \frac{P(A \cap black, caso)}{P(black, caso)}$$

dove:

$$P(A \cap black, caso) = \frac{1}{2} \cdot \frac{1}{5}$$

$$P(black, caso) = \frac{1}{2} \cdot \frac{1}{5} + \frac{1}{2} \cdot \frac{200}{500}$$

da cui si evince che:

$$P(A|black, caso) = \frac{\frac{1}{10}}{\frac{1}{10} + \frac{1}{5}} = \frac{1}{3}$$

9.3 Problema 13

Delle particelle instabili sono emesse da una sorgente e decadono ad una distanza x seguendo una distribuzione esponenziale con lunghezza λ. Possiamo misurare solo le particelle nella finestra tra x = 1 cm ed x = 10 cm. Come possiamo fare inferenza su λ?

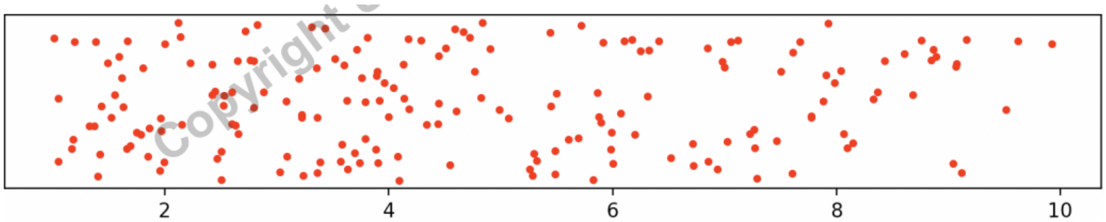

Grafico 7: Distribuzione delle particelle

Soluzione:

Anzitutto ricordo la distribuzione esponenziale:

$$P_\lambda(x) = \lambda e^{-\lambda x} \equiv P(x|\lambda)$$

Bisogna fare inferenza su λ quindi trovare la probabilità di λ dato x. Siccome abbiamo $P(x|\lambda)$ possiamo applicare il Teorema di Bayes e ricavare $P(\lambda|x)$:

$$P(\lambda|x) = \frac{P(x|\lambda)P(\lambda)}{P(x)}$$

In questo caso:

$$P(x|\lambda) = \begin{cases} \frac{\lambda e^{-\lambda x}}{Z(\lambda)} \text{ per } 1 \leq x \leq 10 \\ 0 \text{ altrove} \end{cases}.$$

dove:

$$Z(\lambda) = \int_1^{10} \lambda e^{-\lambda x} dx = ... = e^{-\lambda} - e^{-10\lambda}$$

La $Z(\lambda)$ ha il significato che la likelihood deve essere normalizzata. Da cui:

$$P(\lambda|x) = \frac{\frac{\lambda e^{-\lambda x}}{Z(\lambda)} \cdot P(\lambda)}{\int \tilde{\lambda} e^{-\tilde{\lambda} x} P(\tilde{\lambda}) d\tilde{\lambda}}$$

In generale però, avendo un numero $\{x_n\}$ di punti si ha:

$$P(\lambda|\{x_1, x_2, .., x_n\}) = \frac{P(\{x\}|\lambda)P(\lambda)}{P(\{x\})} \propto \frac{1}{[\lambda Z(\lambda)]^N} exp\left(\frac{-\sum_{N=1}^{N} x_n}{\lambda}\right) P(\lambda)$$

9.4 Problema 14

Gli astronomi misurano il flusso di luminosità F delle stelle e galassie usando una scala logaritmica di magnitudini m. La relazione tra magnitudini e luminosità L = S F, dove S indica la superficie irraggiata dalla stella con luminosità L, mentre F è il flusso.

$$m = -2.5 log_{10} \frac{L}{L_{ref}} \tag{9.4}$$

Supponendo che tutte le stelle e galassie hanno la stessa luminosità assoluta e sono distribuite casualmente nello spazio con densità uniforme (sono distribuite secondo un processo puntuale poissoniano). Quale è la distribuzione di probabilità per m?

Soluzione:

Anzitutto ricordiamo la distribuzione di Poission, la quale:

$$P_\lambda(n) = \frac{\lambda^n}{n!} e^{-\lambda} \tag{9.5}$$

Sia g(L) = m = $-2.5 log_{10} \frac{L}{L_{ref}}$. Sfrutto un cambio di variabili, ovvero quello dettato dalle formule (9.2) ed (9.3). Ovvero noi abbiamo g(L) ma ci interessa f(m). Per farlo calcolo $g^{'}(L)$:

$$g^{'}(L) = -2.5 \frac{L_{ref}}{L} \frac{1}{L_{ref}} \tag{9.6}$$

da cui dalla formula (9.3) si ottiene che:

$$f_Y = \frac{P_\lambda(n)}{|g^{'}(L)|} = \frac{P_\lambda(n) \cdot L}{2.5} \tag{9.7}$$

Ora basta riscrivere L in funzione di m, oovero:

$$m = -2.5 log_{10} \frac{L}{L_{ref}} \longrightarrow e^m = \left(\frac{L}{L_{ref}}\right)^{2.5} \longrightarrow L = e^{-\frac{m}{2.5}} L_{ref} \tag{9.8}$$

Da cui:

$$f_Y = \frac{P_\lambda(n)}{|g^{'}(L)|} = \frac{P_\lambda(n) \cdot L}{2.5} = \frac{P_\lambda(n) \cdot L_{ref}}{2.5 e^{-\frac{m}{2,5}}} \tag{9.9}$$

Un secondo modo (seguire questo perché è di Lombardi) è: Il flusso è definito come:

$$F = \frac{L}{4\pi r^2} \tag{9.10}$$

quindi si ha che:

$$m = -2.5log_{10}\frac{F}{F_{ref}} = -2,5log_{10}\frac{L}{4\pi r^2 F_{ref}} = M_{ref} + 5log_{10}\left(\frac{r}{r_{ref}}\right) \tag{9.11}$$

dove M_{ref} è la magnitudine di riferimento ed è definita come la magnitudine ce una stella avrebbe, se posta a 10 pc ($r_{ref} = 10pc$). Dall'ipotesi di galassie uniformemente distribuite e con pari magnitudine assoluta si ha che:

$$m(r) = M_{ref} + 5log_{10}\left(\frac{r}{r_{ref}}\right) \longrightarrow \frac{m(r)-M}{5} = log_{10}\left(\frac{r}{10pc}\right) \longrightarrow 10^{\frac{m(r)-M}{5}} \cdot 10pc = r \tag{9.12}$$

Ci si aspetta che il numero di stelle in media, entro un raggio r sia:

$$\bar{N}(<r) = n\frac{4}{3}\pi r^3 \tag{9.13}$$

dove n è il numero atteso di stelle. Ma allora:

$$\bar{N}(<m) = n\frac{4}{3}\pi\left(10^{\frac{m(r)-M}{5}}\right)^3 \cdot 10^3 pc = C10^{0.6m} \tag{9.14}$$

9.5 Problema 15

Si lancia una moneta 12 volte e si ottiene 3 volte testa. Cosa si può dire sulla moneta? E una moneta truccata o no? (Comparare due modelli, uno con la moneta non truccata ed uno con probabilità f.)

Soluzione:

Indichiamo con D il numero di volte n in cui si ottiene testa in N = 12 lanci della moneta, ovvero i nostri dati. Assumiamo H_0 il modello per cui la moneta è normale e H_1 il modello per cui la moneta è piegata. Per comparare due modelli è sufficiente fare il rapporto dei posterior $P(H_0|D)$ e $P(H_1|D)$, quindi delle likelihood, assumendo di prendere i prior per i due modelli uguali, avendo lo stesso numero di dati. Quindi la distribuzione di probabilità dato il modello H_0 è

$$P(n|N,H_0) = \binom{N}{n} f^n(1-f)^{N-n} = \binom{N}{n}\frac{1}{2^N} \tag{9.15}$$

dove si considera f = 0.5 perchè indica la probabilità che esca testa se la moneta è normale. La distribuzione di probabilità dato il modello H_1 è

$$P(n|N,H_1) = \binom{N}{n} f^n(1-f)^{N-n} \tag{9.16}$$

dove per la moneta piegata si considera la probabilità f un parametro libero. Infatti per paragonare due modelli non c'è bisogno che il numero di parametri sia uguale.