

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea in Matematica

**APPROCCIO BAYESIANO
NELL'ANALISI
DEI DATI ARCHEOLOGICI**

Tesi di Laurea in
Calcolo delle Probabilità e Statistica

Relatore:
Chiar.mo Prof.
Massimo Campanino

Presentata da:
Monia Tomassini

I Sessione
Anno Accademico 2011/2012

*a tutti coloro che
nel loro lavoro mettono
passione e dedizione*

Introduzione

L'approccio bayesiano all'inferenza sta acquisendo un ruolo sempre più importante nella letteratura statistica: è infatti in continuo aumento il numero di elaborazioni di dati in campo medico-sanitario, economico-finanziario, socio-politico e forse ancor di più nelle scienze sperimentali, dove si fa utilizzo più o meno esplicito di tecniche riconducibili al paradigma bayesiano dell'inferenza. Le ragioni di questa improvvisa accelerazione e della diffusione dei metodi bayesiani nella statistica applicata, iniziata più o meno negli anni '90 del secolo scorso, sono molteplici. Innanzitutto bisogna osservare che la logica bayesiana è coerente e dotata di solide basi logiche. Questo fa sì che l'impostazione bayesiana dell'inferenza statistica formalizzi in modo semplice e diretto il ragionamento induttivo di una persona che, in base alle informazioni disponibili su un certo insieme di fenomeni, vuole calcolare la probabilità di eventi futuri. Inoltre l'enorme sviluppo di nuove metodologie computazionali rende possibile analizzare modelli statistici estremamente complessi. Questa tesi vuole mostrare come, con l'utilizzo della probabilità e del teorema di Bayes è possibile risolvere problemi di incertezza di qualsiasi genere. In particolare osserveremo che l'inferenza statistica viene applicata anche in campi come l'archeologia. Vedremo quindi come l'analisi dei reperti archeologici sfrutti il teorema di Bayes per poter combinare insieme le informazioni a priori e i dati e utilizzare la probabilità a posteriori per valutare la veridicità di certe ipotesi. Tra i vari metodi di datazione ci occuperemo della datazione al radiocarbonio considerato come una delle tecniche con il più alto numero di applicazioni del metodo bayesiano. Vedremo quindi un esempio concreto di approccio bayesiano nell'interpretazione dei dati archeologici e come gli archeologi sfruttino i metodi probabilistici per stimare il margine di errore del risultato ottenuto e per introdurre un intervallo di confidenza cioè un intervallo di valori entro

cui il valore vero si trova con probabilità sufficientemente grande. Lo scopo di questa tesi è sottolineare i vantaggi dell'approccio bayesiano e dell'inferenza statistica, illustrare come un problema reale può essere affrontato attraverso la modellizzazione in termini matematici ed infine mostrare alcune caratteristiche essenziali dell'approccio bayesiano nel campo dell'analisi dei dati.

Indice

Introduzione	i
1 Inferenza bayesiana	1
1.1 Teorema di Bayes e sue applicazioni	3
1.1.1 Esempio 1	4
1.1.2 Esempio 2	6
1.2 Test di significatività	10
1.3 Induzione sulla media della normale	11
1.3.1 Esempio	12
2 Approccio bayesiano alla statistica archeologica	15
2.1 Statistica e archeologia	17
2.2 Metodo di datazione al radiocarbonio	17
2.3 Calibrazione	20
2.3.1 Esempio	21
2.4 Margine di errore	23
2.4.1 Sindone di Torino	24
3 Conclusioni	29
Bibliografia	31

Elenco delle figure

1.1	Densità a priori e a posteriori di Θ	8
1.2	Densità a priori e a posteriori di Θ e verosimiglianza	10
2.1	Legge di decadimento del radiocarbonio	20
2.2	Curva di calibrazione e datazione dei reperti di Similaun	23

Capitolo 1

Inferenza bayesiana

L'inferenza statistica è il procedimento che permette di ricavare le caratteristiche di una popolazione dall'osservazione di una parte di essa, detta campione, selezionata solitamente mediante un esperimento casuale (aleatorio). Questo avviene molto spesso nelle indagini socio-economiche o ambientali dove lo scopo è proprio quello di valutare degli indici di sintesi di una popolazione e nelle indagini statistiche che, basandosi su ipotesi (fissate alla luce di esperienze passate) inerenti ai tassi di mortalità, ai salari, ai tassi di interesse e ai dividendi, si prefiggono di predire risultati futuri. L'inferenza bayesiana è un approccio all'inferenza statistica in cui le probabilità sono interpretate come livelli di fiducia nel verificarsi di un dato evento. Il fondamento di questo approccio è dato dal teorema di Bayes la cui applicazione permette di tenere conto di opinioni e conoscenze eventualmente esistenti a priori sul fenomeno oggetto di studio. Esso connette distribuzione a priori, verosimiglianza e distribuzione a posteriori in un'unica formula e fornisce un metodo per modificare il livello di fiducia in una data ipotesi alla luce di una nuova informazione. Prima di introdurre il teorema spieghiamo il significato dei seguenti concetti che ci seguiranno per tutta l'argomentazione di questa tesi:

probabilità a priori: è il modo di rappresentare l'informazione a priori riguardo ai valori dei parametri prima di osservare i dati. È indipendente dall'esperienza attuale e può derivare dalle esperienze passate.

verosimiglianza: è lo strumento che collega i dati e i parametri e introduce i dati alle analisi.

probabilità a posteriori: è il risultato della combinazione dei dati e dell'informazione a priori e può essere letta come la probabilità che il parametro sconosciuto assuma un certo valore alla luce dei dati e dell'informazione a priori. In termini probabilistici è la probabilità condizionata che il parametro sconosciuto assuma un certo valore date le informazioni a priori.

Teorema (di Bayes). *La densità di probabilità a posteriori è proporzionale al prodotto della verosimiglianza per la densità di probabilità a priori. Se Θ è un numero aleatorio e se X_1, X_2, \dots sono numeri aleatori stocasticamente indipendenti subordinatamente alla conoscenza di $(\Theta = \theta)$ si ha che:*

$$\pi(\theta|x_1, \dots, x_n) = K\pi_0(\theta)p(x_1, \dots, x_n|\theta) \quad (1.1)$$

dove

- K è la costante di normalizzazione
- $\pi_0(\theta)$ è la densità di probabilità a priori di Θ
- $p(x_1, \dots, x_n|\theta)$ è la verosimiglianza ed è equivalente alla densità di probabilità subordinata di variabili casuali che formano il campione e ha la forma

$$p(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) \quad (1.2)$$

Pertanto, dato un campione aleatorio cioè un insieme di individui scelti a caso della popolazione, è possibile ottenere valutazioni sulla distribuzione delle grandezze nella popolazione a partire dal risultato di un campionamento. Se si dà a Θ una densità di probabilità a priori $\pi_0(\theta)$ e se si vuole vedere come cambia la distribuzione di Θ dopo aver effettuato n esperimenti, la densità a posteriori di Θ è data dalla formula di Bayes.

1.1 Teorema di Bayes e sue applicazioni

Lo scopo primario dell'inferenza statistica è quello di acquisire ulteriore conoscenza su quantità incerte, sulle quali spesso si hanno informazioni parziali. Gli esempi che seguono possono contribuire a chiarire questi concetti.

Esempio 1 [*Sondaggio pre-elettorale*]

Per un determinato seggio elettorale, si vuole stimare la percentuale di elettori che, nelle prossime elezioni politiche, voterà per le due coalizioni di centro-destra e di centro-sinistra.

Esempio 2 [*Numerosità di persone affette da una malattia*]

Si vuole stimare la prevalenza di una malattia non contagiosa in un'antica società usando informazioni derivanti da un grandissimo cimitero.

1.1.1 Esempio 1

Immaginiamo di essere interessati alla percentuale di voti che otterrà lo schieramento di centro-sinistra. Dalla popolazione di votanti in quel seggio elettorale si estrae un campione casuale di n elettori (il valore n è considerato qui fissato in anticipo uguale a dieci) e a ciascuno di loro si chiede per chi voterà alle prossime elezioni; a ciascun componente del campione si associa una variabile aleatoria Y_j che vale 1 se l'elettore vota per il centro-sinistra e vale 0 altrimenti (se vota altri partiti, oppure scheda bianca o nulla). Il valore incognito da stimare è un valore $\Theta \in [0, 1]$. Condizionatamente al valore θ assunto da Θ , le n osservazioni Y_1, \dots, Y_n vengono supposte indipendenti e tutte con la stessa distribuzione di probabilità, ovvero, $\forall j = 1, \dots, n$ si ha:

$$\mathbf{P}(Y_j = 1|\theta) = 1 - \mathbf{P}(Y_j = 0|\theta) = \theta \quad (1.3)$$

Supponiamo che l'esperimento si concluda con $k = 7$ successi (valori di Y_j uguali a 1) e $n - k = 3$ insuccessi (valori di Y_j uguali a 0) allora la funzione di verosimiglianza associata a tale esperimento è

$$p(x_1, \dots, x_n|\theta) = \theta^7(1 - \theta)^3 \quad (1.4)$$

La funzione di verosimiglianza “pesa” i diversi valori possibili che θ può assumere sulla base dei dati rilevati. Prima di proseguire nella trattazione matematico-statistica di questo esempio è bene considerare altri esempi, molto diversi tra loro, che però presentano, una volta formalizzati, una struttura identica a quest'esempio. Consideriamo le tre seguenti situazioni:

1. Tizio sostiene di essere in grado di riconoscere se un brano musicale sia stato scritto da Mozart oppure da Beethoven dopo appena quattro note. Gli sottoponiamo allora gli incipit di dieci brani scelti a caso dal repertorio dei due autori e verifichiamo le sue capacità.

2. La signora Bianchi sostiene che bevendo una tazza di tè al latte, è in grado di stabilire se sia stato versato prima il latte oppure il tè: anche in questo caso sottoponiamo la signora a un test di dieci prove.
3. Il signor Rossi sostiene di possedere capacità soprannaturali e di essere in grado di prevedere il risultato di un lancio di una moneta regolare; allo stesso modo, effettuiamo dieci prove sperimentali.

In tutti i casi si hanno n variabili aleatorie che assumono il valore 1 con probabilità incognita Θ , che danno luogo a un vettore di dati osservati (stringa di valori 0 e 1). Assumiamo che in tutti gli esperimenti si osservino $k = 7$ successi su $n = 10$ prove. Ne segue che le funzioni di verosimiglianza associate ai tre esperimenti saranno del tutto identiche, così come le stime puntuali del parametro incognito Θ dal punto di vista classico (non bayesiano). Valuteremo pari a 0.7 sia la probabilità dell'esperto di musica di riconoscere un brano sia la capacità del presunto sensitivo di prevedere il futuro. C'è qualcosa di insoddisfacente in questa coincidenza di conclusioni: per quanto l'esperimento statistico abbia fornito lo stesso risultato, è abbastanza ragionevole avere un diverso grado di fiducia sulle affermazioni dei tre personaggi in questione. Ad esempio, un appassionato di musica reputa possibile e ragionevole che un vero esperto possa riconoscere un brano dopo quattro note mentre potrebbe non aver nessuna fiducia su chi sostiene di avere doti di preveggenza, e non saranno certo 10 prove (troppo poche) a fargli cambiare idea. Ciò che non è stato considerato finora sono le informazioni a priori che avevamo su Θ nei diversi esperimenti. Nell'1.1.1 Θ rappresentava la percentuale di votanti per la lista di centro-sinistra in un dato collegio elettorale, e certamente, sulla base delle informazioni relative alle elezioni precedenti, abbiamo idea dell'ordine di grandezza della percentuale di voti che ci possiamo attendere per quella lista. Allo stesso modo molte persone darebbero più fiducia all'esperto musicale che non al preveggenente. Come inserire tali informazioni nell'analisi statistica? Lo si può fare attraverso il teorema di Bayes: quello che serve è una formalizzazione di tutto ciò che sappiamo su Θ in termini di una distribuzione di probabilità iniziale. Matematicamente si tratta di scegliere una densità di probabilità π_0 per Θ . Supponiamo che le nostre informazioni sul collegio in questione siano molto scarse e assumiamo per Θ una distribuzione iniziale uniforme nell'intervallo

$[0, 1]$ ovvero

$$\pi_0(\theta) = 1 \quad \text{per } 0 \leq \theta \leq 1 \quad (1.5)$$

Combinando la distribuzione a priori con la funzione di verosimiglianza si ottiene la densità di probabilità a posteriori

$$\pi(\theta|y) = K\pi_0(\theta)\pi(y|\theta) = K\theta^7(1-\theta)^3 \quad (1.6)$$

che è una beta di parametri 8 e 4. La costante di normalizzazione è dunque data da:

$$K = \frac{\Gamma(8+4)}{\Gamma(8)\Gamma(4)} = \frac{11!}{7!3!} = 1320 \quad (1.7)$$

Quindi la densità di probabilità a posteriori di θ è

$$\pi(\theta|y) = 1320 \cdot \theta^7(1-\theta)^3 \quad \text{per } 0 \leq \theta \leq 1 \quad (1.8)$$

La densità di probabilità a posteriori non è altro che la versione normalizzata della funzione di verosimiglianza: questo avviene perché la distribuzione a priori uniforme non ha aggiunto informazioni ulteriori rispetto a quanto già fornito dalla funzione di verosimiglianza. Invece di utilizzare la distribuzione uniforme π_0 , in presenza di specifiche informazioni su Θ (ad esempio, se il collegio elettorale è in Emilia-Romagna abbiamo ragione di credere che la percentuale di voti della lista del centro-sinistra si attesterà intorno a valori superiori al 40-50 per cento), potremmo decidere di utilizzare un'altra distribuzione iniziale, ad esempio una beta di parametri $\alpha = \beta = 5$. Con calcoli del tutto simili ai precedenti si arriva alla nuova distribuzione a posteriori

$$\pi(\theta|y) = 604656 \cdot \theta^{7+5-1}(1-\theta)^{3+5-1} \quad \text{per } 0 \leq \theta \leq 1 \quad (1.9)$$

$$= 604656 \cdot \theta^{11}(1-\theta)^7 \quad \text{per } 0 \leq \theta \leq 1 \quad (1.10)$$

che rappresenta ancora una distribuzione di tipo Beta con parametri modificati in $\alpha = 12$ e $\beta = 8$.

1.1.2 Esempio 2

Supponiamo di voler stimare la prevalenza di una malattia non contagiosa basandoci su dati derivanti da un cimitero. L'identificazione della malattia mediante gli scheletri

è difficile e necessita di analisi scientifiche costose quindi non è conveniente analizzare tutti gli scheletri del cimitero ma solo un campione. Supponiamo che cinque scheletri vengono selezionati casualmente per un'analisi scientifica e di questi, si scopre che per tre la causa di morte sia la malattia. Come possiamo leggere questo risultato sull'intera popolazione? Una stima schietta potrebbe essere che $3/5$ o equivalentemente il 60% della popolazione soffre di questa malattia. Questo si basa esclusivamente sui dati e quindi non tiene conto di eventuali altri elementi di prova. Sia Θ il tasso di incidenza sconosciuta nell'intera popolazione al tempo di studio. Sebbene la malattia non sia contagiosa, ci potrebbe essere qualche fattore genetico coinvolto che spieghi che gli individui che hanno stretti rapporti di sangue possono essere più sensibili di quelli che non ne hanno. Supponiamo che il campione sia rappresentativo e che quindi siamo in grado di garantire per quanto possibile che gli scheletri possano essere considerati intercambiabili. Pertanto la probabilità che un particolare individuo abbia la malattia è θ . Così la probabilità che non soffra di tale malattia è $1 - \theta$. Nell'ipotesi di indipendenza dei campioni, la verosimiglianza che 3 su 5 siano stati affetti dalla malattia e conseguentemente 2 su 5 no è

$$\theta \times \theta \times \theta \times (1 - \theta) \times (1 - \theta) \implies \theta^3(1 - \theta)^2 \quad (1.11)$$

La domanda da porci è: prima di conoscere i dati, cosa sappiamo riguardo il tasso di incidenza della malattia in quel periodo? Una risposta potrebbe essere che non ci sono informazioni disponibili. Nel qual caso si pensa a una densità a priori di Θ che dà la stessa probabilità a intervalli della stessa lunghezza. In altre parole prendiamo come funzione di densità di probabilità a priori di Θ , una distribuzione uniforme tra 0 e 1. Quindi

$$p(\theta) = 1 \quad \text{per } 0 \leq \theta \leq 1 \quad (1.12)$$

La distribuzione a posteriori riguardo il tasso di incidenza è espressa, per il teorema di Bayes, nei seguenti termini:

$$p(\theta|x) = Kp(\theta)p(x|\theta) = K\theta^3(1 - \theta)^2 \quad (1.13)$$

che è una beta di parametri 4 e 3. La costante di normalizzazione è data da:

$$K = \frac{\Gamma(4 + 3)}{\Gamma(4)\Gamma(3)} = \frac{6!}{3!2!} = 60 \quad (1.14)$$

Quindi la distribuzione a posteriori di Θ è

$$p(\theta|x) = 60 \cdot \theta^3(1 - \theta)^2 \quad \text{per } 0 \leq \theta \leq 1 \quad (1.15)$$

La figura 1.1 mostra il grafico della distribuzione a priori e a posteriori di Θ . Osserviamo che quella a priori è una linea orizzontale a indicare che la densità è costante, in contrasto con quella a posteriori che ha una moda distinta in $\theta = 0,6$.

Introduciamo una piccola variante in questo problema per vedere come cambiano i

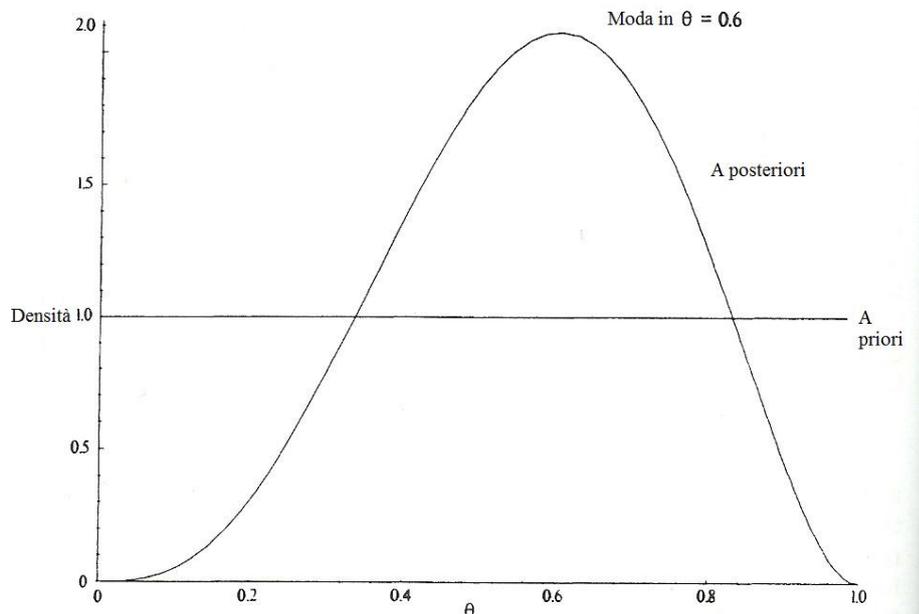


Figura 1.1: Densità a priori e a posteriori di Θ

risultati e come cambia il grafico della distribuzione a priori e a posteriori. Supponiamo quindi che vengano analizzati più scheletri. Assumiamo, per esempio, che invece di 5 siano 25 e di questi 15 hanno riscontrato la malattia. Supponiamo inoltre di avere una distribuzione a priori differente. L'analisi di scheletri provenienti da altri luoghi nella stessa regione e dello stesso periodo suggerisce che il tasso di incidenza giace probabilmente in un range che va dal 20% al 80% con una probabilità dello 0,8 così una possibile

distribuzione a priori per Θ è data da:

$$p(\theta) = 30 \cdot \theta^2(1 - \theta)^2 \quad \text{per } 0 \leq \theta \leq 1 \quad (1.16)$$

Infatti effettuando i calcoli si ha:

$$\begin{aligned} 30 \cdot \int_{0.2}^{0.8} \theta^2(1 - \theta)^2 d\theta &= 30 \cdot \int_{0.2}^{0.8} \theta^2(1 + \theta^2 - 2\theta) d\theta \\ &= 30 \cdot \int_{0.2}^{0.8} (\theta^2 - 2\theta^3 + \theta^4) d\theta \\ &= \left[\frac{\theta^3}{3} - \frac{2\theta^4}{4} + \frac{\theta^5}{5} \right]_{0.2}^{0.8} \simeq 0.8 \end{aligned}$$

Pertanto la distribuzione a posteriori questa volta sarà:

$$p(\theta|x) = Kp(\theta)p(x|\theta) = K\theta^3(1 - \theta)^2\theta^2(1 - \theta)^2 = K\theta^5(1 - \theta)^4 \quad \text{per } 0 \leq \theta \leq 1 \quad (1.17)$$

che è una beta di parametri 6 e 5. La costante di normalizzazione è dunque data da:

$$K = \frac{\Gamma(6 + 5)}{\Gamma(5)\Gamma(6)} = \frac{10!}{4!5!} = 1260 \quad (1.18)$$

Quindi la distribuzione a posteriori di Θ è

$$p(\theta|x) = 1260 \cdot \theta^5(1 - \theta)^4 \quad \text{per } 0 \leq \theta \leq 1 \quad (1.19)$$

Il grafico della distribuzione a priori, della verosimiglianza e della distribuzione a posteriori è mostrato nella figura 1.2. La moda della distribuzione a posteriori è $\theta = 0,56$.

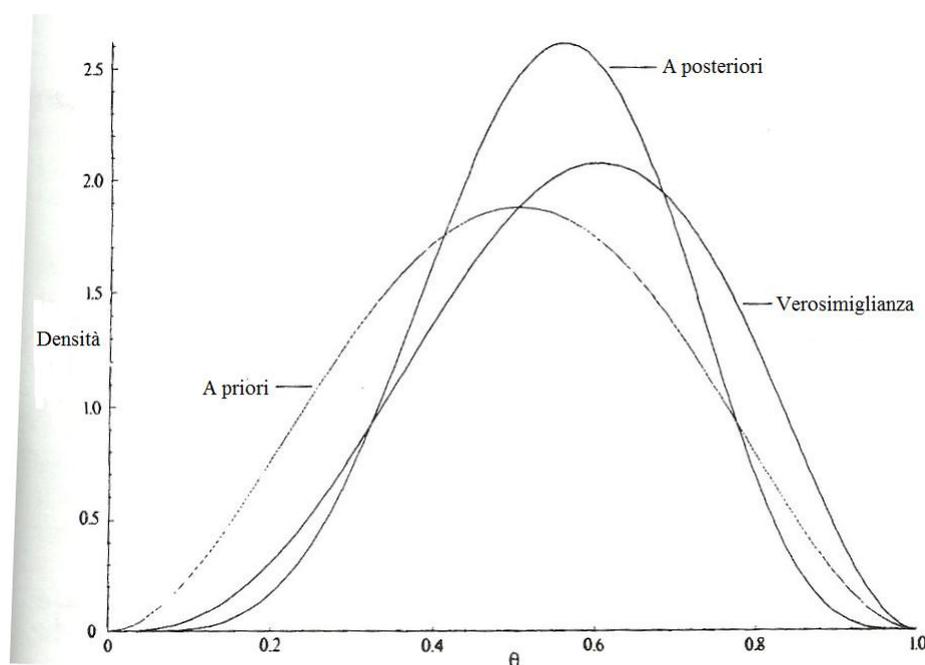


Figura 1.2: Densità a priori e a posteriori di Θ e verosimiglianza

1.2 Test di significatività

Molto spesso la stima puntuale di un parametro non fornisce sufficienti garanzie di precisione, ed è più ragionevole fornire un insieme di valori che, alla luce del campione osservato, possano essere considerati una ragionevole stima del parametro. Pertanto un modo per descrivere in maniera sintetica la distribuzione a posteriori è introdurre intervalli di fiducia (o di confidenza) o nel caso multidimensionale regioni di fiducia cioè regioni tali che la probabilità che un certo vettore appartenga alla regione sia per esempio del 95%. L'intervallo di confidenza è un intervallo di valori entro il quale si ritiene sia compreso il parametro in esame con un certo grado di confidenza. Fare un test di significatività a livello α significa trovare un intervallo che contenga il valore vero di θ con probabilità non inferiore $1 - \alpha$. I livelli di significatività (α) più usati sono 0,05 e 0,01 a cui corrispondono i coefficienti di fiducia 0,95 e 0,99, spesso espressi in termini percentuali (95% e 99%). È chiaro che vi è arbitrarietà nella scelta di un tale intervallo o regione. Ma nelle situazioni concrete la scelta può essere fatta seguendo criteri di simme-

tria se la densità a posteriori è simmetrica rispetto ad un punto, scegliendo la lunghezza dell'intervallo il più piccola possibile o, nel caso multidimensionale, minimizzando il volume nello spazio dei parametri.

Il test statistico è uno strumento utile perché consente di decidere circa l'attendibilità dell'ipotesi. In particolare il test è una procedura che, sulla base di dati campionari e con un certo grado di probabilità, consente di decidere se è ragionevole respingere l'ipotesi oppure se non esistono elementi sufficienti per respingerla. Pertanto fatta una determinata congettura o un'ipotesi sul parametro Θ , questo test mi permette di decidere se essa sia accettabile o meno. Riportiamo ora un esempio concreto di test di verifica dell'ipotesi ma prima introduciamo il concetto di induzione sulla media della normale che ci permetterà di risolvere e capire a pieno l'esempio.

1.3 Induzione sulla media della normale

Sia $(X_i)_{i \in N}$ una successione di numeri aleatori stocasticamente indipendenti subordinatamente alla conoscenza di un parametro Θ con densità subordinata

$$f(x_i|\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}} \quad (1.20)$$

con $\sigma \in \mathfrak{R}^+$ conosciuta.

Se la densità di probabilità a priori $\pi_o(\theta)$ è una gaussiana $N(\mu_0, \sigma_0^2)$, dopo aver ottenuto i risultati dei primi n esperimenti la distribuzione a posteriori è allora:

$$\begin{aligned} \pi_n(\theta|x_1, \dots, x_n) &= K \pi_0(\theta) f(x_1, \dots, x_n|\theta) \\ &= K \pi_0(\theta) \prod_{i=1}^n f(x_i|\theta) \\ &= K e^{-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}} e^{-\sum_{i=1}^n \frac{(x_i-\theta)^2}{2\sigma^2}} \\ &= K e^{-\frac{1}{2}(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2})\theta^2 - 2(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2})\theta} \\ &= K e^{-\frac{1}{2} \frac{(\theta-\mu_n)^2}{\sigma_n^2}} \end{aligned}$$

dove

$$\mu_n = \frac{n\bar{x}\sigma^{-2} + \mu_0\sigma_0^{-2}}{n\sigma^{-2} + \sigma_0^{-2}} \quad (1.21)$$

$$\sigma_n^{-2} = n\sigma^{-2} + \sigma_0^{-2} \quad (1.22)$$

con

- n : numero di misurazioni fatte
- μ_0 : media della distribuzione a priori
- σ_0^{-2} : inverso della varianza della distribuzione a priori
- σ^{-2} : inverso della varianza della densità subordinata in 1.20

Si noti che all'aumentare del numero delle osservazioni ovvero per n abbastanza grande si ha che:

$$\begin{aligned} \mu_n - \bar{x} &\xrightarrow{n \rightarrow \infty} 0 \\ \sigma_n^{-2} &\xrightarrow{n \rightarrow \infty} \infty \implies \sigma_n^2 \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

cioè la distribuzione diventa approssimativamente una normale con media pari alla media campionaria e varianza pari a $\frac{\sigma^2}{n}$. Questo può essere tradotto affermando che più osservazioni faccio più dò peso alle osservazioni e alla varianza. Pertanto la stima è migliore quanto è maggiore la numerosità campionaria.

1.3.1 Esempio

Nella preparazione di un materiale isolante, le misure di conduttività sono fatte usando uno strumento con deviazione standard conosciuta. La distribuzione a priori è una normale $N(16, \frac{1}{4})$ e la densità subordinata è

$$f(x_i|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2}} \quad (1.23)$$

quindi nella notazione usata nella sezione 1.3 si ha $\mu_0 = 16$, $\sigma_0^2 = \frac{1}{4}$ e $\sigma^{-2} = 1$. Vengono fatte dieci letture e si ottengono i seguenti valori:

16,11 17,37 16,35 15,16 18,82 18,12 15,82 16,34 16,64 15,01

La media è $\bar{x} = 16,57$ pertanto la distribuzione a posteriori data dalla formula di Bayes è ancora una normale di parametri $N(\mu_n, \sigma_n^2)$ dove μ_n , e σ_n^2 sono dati rispettivamente dalla 1.21 e dalla 1.22. Quindi nel nostro caso avremo:

$$\mu_{10} = \frac{10 \cdot 16,57 + 4 \cdot 16}{10 + 4} = 16,41$$

$$\sigma_{10}^{-2} = 10 + 4 = 14 \implies \sigma_{10} = \frac{1}{\sqrt{14}} = 0,27$$

Pertanto la distribuzione a posteriori è una normale di parametri $N(\mu_{10}, \sigma_{10}^2)$. Supponiamo che la conduttività non debba superare 17 altrimenti la proprietà di isolamento non è soddisfatta. Allora la domanda a cui dobbiamo rispondere è: la conduttività sta sotto 17? Quindi dobbiamo studiare la distribuzione a posteriori e la sua posizione rispetto al valore 17. Un intervallo di confidenza del 95% della media è:

$$\bar{x} \pm \lambda_\alpha \frac{\sigma}{\sqrt{n}} = 16,57 \pm 1,96 \cdot \frac{1}{\sqrt{10}} = 16,57 \pm 0,61$$

cioè l'intervallo (15,96 ; 17,19). In conclusione possiamo affermare che la conducibilità giace molto probabilmente in questo intervallo. Ma questo intervallo contiene il valore 17 così è possibile che questo materiale non soddisfi le caratteristiche richieste. In altre parole il risultato è insignificante a livello del 5%. Usando un intervallo di confidenza dell' 80% otteniamo l'intervallo (16,17 ; 16,97) che non include il valore 17. Questo mi dice che la proprietà di isolamento è soddisfatta quindi il test di significatività a livello del 20% dell'ipotesi che la conduttività non superi 17 ossia che il materiale sia isolante non mi permette di respingere l'ipotesi.

Capitolo 2

Approccio bayesiano alla statistica archeologica

Per gli archeologi l'informazione a priori è molto importante. Prendiamo per un momento in considerazione lo scavo di un luogo risalente all'età del bronzo appena scoperto in Europa. La persona che coordina il lavoro avrà già intrapreso scavi simili, sarà un esperto dell'epoca al riguardo e avrà già studiato negli stessi settori. Pertanto, quando inizia lo scavo, tutte le caratteristiche e i manufatti recuperati verranno interpretate alla luce delle conoscenze di base di questo periodo storico. Il vantaggio di cui abbiamo parlato dell'approccio bayesiano di incorporare la conoscenza a priori con l'analisi, di combinare la probabilità a priori con i dati e di interpretare la probabilità a posteriori, fa sì che questo metodo sia usato nell'interpretazione dei dati archeologici. Come affermano Bowman e Leese (1995),

“Bayesian methodology is particularly suited to the calibration of groups of dates”

La metodologia bayesiana è particolarmente adatta per la taratura dei gruppi di date. L'approccio bayesiano sull'analisi dei dati è precisamente un meccanismo per gestire l'incertezza e per formalizzare una relazione tra presupposti e conclusioni. Il principio base che governa l'approccio bayesiano è la combinazione in una singola analisi formale della comprensione dei dati di un problema al fine di fare inferenze circa il problema in modo logico e coerente.

2.1 Statistica e archeologia

Il lavoro sul campo e la ricerca hanno sempre prodotto grandi quantità di dati sia qualitativi sia quantitativi, questi ultimi comprendenti conti, misure, pesi e così via. Senza alcuni mezzi di comprimere la loro mole, i dati derivati dalla ricerca archeologica non potrebbero mai essere interpretati o pubblicati. Di conseguenza gli archeologi sono in grado di utilizzare le statistiche riassuntive quali medie, mediane, mode e, per facilitare la lettura dei risultati, di utilizzare grafici come istogrammi, grafici a torta e boxplots. Al fine di adottare l'approccio bayesiano è necessario andare oltre i metodi esplorativi nelle statistiche e utilizzare un approccio basato su modelli. In ogni progetto, i ricercatori hanno bisogno di farsi un'idea del problema in esame indagando, esaminando i dati e osservando come essi vengono distribuiti. Grafici e istogrammi, rivestono una particolare importanza nella costruzione del modello che è molto utile e molto utilizzato dagli archeologi perché permette di semplificare il problema e di interpretare i risultati ottenuti in maniera più chiara e precisa. I passi da seguire sono:

1. individuare il problema e formalizzarlo attraverso una rappresentazione semplificata;
2. formulare la domanda in termini matematici e dare una risposta;
3. interpretare la soluzione e compararla con la realtà;
4. usare i risultati ottenuti ai fini dell'indagine

Grazie alla modellizzazione del problema e alla scelta delle valutazioni a priori, è possibile, attraverso il teorema di Bayes, ottenere l'informazione a posteriori. In altre parole, attraverso l'approccio bayesiano, le conoscenze a priori e le ipotesi possono essere inserite esplicitamente nelle equazioni e quindi si può formalizzare l'interazione tra informazioni precedenti e nuovi dati.

2.2 Metodo di datazione al radiocarbonio

La datazione riveste una particolare importanza per gli archeologi perché senza una scala del tempo non c'è né storia né preistoria. La ricostruzione storica di eventi passati,

naturali o antropici, ha come presupposto l'esigenza di collocare gli eventi medesimi in una scala temporale che possa avere la validità più generale possibile e che tenda verso un riferimento assoluto. Ciò viene generalmente realizzato con l'impiego di modelli deterministici o stocastici che, in ogni caso, devono avere come presupposto una base fenomenologica altamente affidabile. La metodologia della datazione è basata sull'individuazione di archivi temporali, vale a dire di quel complesso di materiali che conservano comunque tracce dei fenomeni passati secondo una ben determinata ascissa temporale. Dunque una caratteristica essenziale della ricerca archeologica consiste nell'essere in grado di datare un sito nel modo più accurato possibile. La datazione al radiocarbonio è uno dei metodi scientifici di datazione più importanti usati dagli archeologi. Esso fu introdotto tra il 1947 ed il 1949 da un gruppo di chimici dell'Università di Chicago, diretti da Willard Libby (che per questo ebbe il premio Nobel nel 1960) ed è il metodo scientifico più conosciuto per determinare l'età dei materiali organici. Si utilizza per reperti di origine organica poiché il carbonio è un elemento che fa parte di tutta la materia vivente e si trova in natura sotto forma dei suoi isotopi: Carbonio-12, Carbonio-13, Carbonio-14. Di questi, solo l'ultimo è radioattivo, mentre gli altri sono stabili. Dopo la formazione, gli atomi di carbonio 14 si combinano rapidamente con l'ossigeno a formare l'anidride carbonica che si mescola con l'atmosfera e si dissolve nell'oceano. Tutti gli organismi viventi che fanno parte del ciclo del carbonio scambiano continuamente carbonio con l'atmosfera attraverso processi di respirazione (animali) o fotosintesi (vegetali), oppure lo assimilano nutrendosi di altri esseri viventi o sostanze organiche. Perciò il nostro corpo, una foglia di albero, oppure un pezzo di mobile di legno, contengono carbonio. Di conseguenza finché un organismo è vivo, il rapporto tra la sua concentrazione di Carbonio-14 e quella degli altri due isotopi di carbonio si mantiene costante e uguale a quella che si riscontra nell'atmosfera. Dopo la morte, però, questi processi terminano e l'organismo non scambia più carbonio con l'esterno per cui il suo Carbonio-14 comincia a diminuire con un ritmo noto a causa del decadimento radioattivo, non venendo più reintegrato dall'esterno. Il metodo si basa dunque sul lento decadimento dell'isotopo radioattivo del carbonio, che costituisce quindi un ottimo "orologio" archeologico. Per effetto del decadimento la concentrazione di Carbonio-14 diminuisce in modo regolare secondo la

formula:

$$c_t = c_0 e^{-\frac{\Delta t}{\tau}}$$

dove

- c_0 è il numero di nuclei di Carbonio-14 contenuti nel reperto al momento della morte dell'organismo
- c_t è il numero di nuclei di Carbonio-14 all'istante t
- Δt è il tempo trascorso dalla morte dell'organismo
- τ è la vita media del Carbonio-14 (pari al tempo di dimezzamento diviso per il logaritmo naturale di 2 ovvero $\frac{5730}{\ln 2} = 8267$ anni)

Misurando dunque la quantità di Carbonio-14 presente in un reperto archeologico organico e conoscendo la curva di decadimento e la quantità iniziale di Carbonio-14 presente nel reperto quando la sua struttura organica era ancora vitale (ovvero un istante prima di morire), si può facilmente stabilire quanti anni sono trascorsi dalla morte dell'organismo applicando la seguente formula:

$$\Delta t = -\tau \ln \frac{c_t}{c_0} \quad (2.1)$$

Per esempio supponiamo di avere un campione con inizialmente un milione di atomi di Carbonio-14 e che quando lo riosserviamo abbia solo duecentomila atomi. In termini matematici si ha $c_0 = 1.000.000$ e $c_t = 200.000$. Sia 8033 il tasso di decadimento misurato sperimentalmente. Allora la stima dell'età del campione è:

$$\Delta t = -8033 \ln \frac{200.000}{1.000.000} = 12.929 \quad (2.2)$$

In altre parole il campione osservato ha all'incirca 13.000 anni.

In generale, è possibile radiodattare solo reperti di età fino a 40.000 – 60.000 anni fa. Per campioni più antichi, la concentrazione di Carbonio-14 è troppo bassa per poter essere misurata con sufficiente accuratezza. Perciò se un campione contiene radiocarbonio è buona prova che non ha milioni di anni.

Il grafico 2.1 riportato di seguito mostra come il decadimento di un radioisotopo segua una curva di tipo esponenziale. Si vede chiaramente come all'aumentare degli anni, il numero di atomi rimasti decresca sempre più.

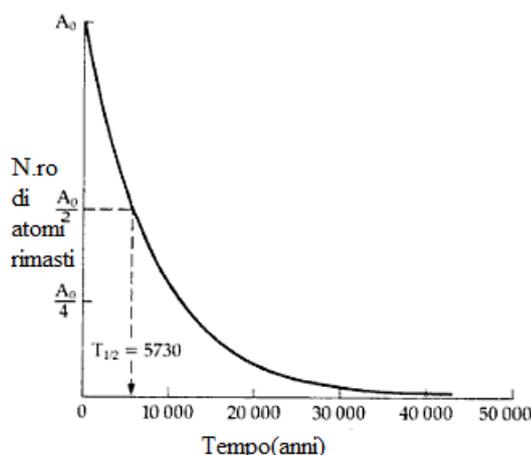


Figura 2.1: Legge di decadimento del radiocarbonio

2.3 Calibrazione

La datazione fornita dai laboratori viene poi sottoposta ad una calibrazione, al fine di ottenere la data reale di calendario, confrontando la datazione convenzionale con quelle ottenute da campioni di età nota. La data calibrata, soprattutto per certi periodi, si discosta notevolmente (anche di parecchi secoli) da quella convenzionale e va considerata come la miglior stima della data “vera”. La datazione radiocarbonica convenzionale (conventional radiocarbon age), non calibrata, è normalmente espressa in anni BP (Before Present) calcolati a ritroso a partire dal 1950, data di riferimento scelta in quanto da quell’anno si è cominciato a datare con il metodo del radiocarbonio. Quindi l’età dei reperti sono sempre riferite al 1950 come all’anno 0. Ad esempio un oggetto datato al 1240 BP vuol dire che è del $1950 - 1240 = 710$ d.C. L’utilità di questa convenzione risiede nel fornire lo stesso valore numerico sia per la data di formazione dei reperti (che hanno avuto origine prima del 1950) sia per la loro età (se riferita al 1950). La datazione calibrata è invece normalmente espressa come data di calendario (calendar age), in anni BC (Before Christ) o AD (Anno Domini), a seconda che si tratti di anni prima o dopo Cristo. La calibrazione si effettua mediante software specializzati che confrontano la datazione radiocarbonica convenzionale con una curva di calibrazione, ottenuta datando reperti di epoca nota. La data calibrata è da considerarsi la miglior stima della data “vera” ed è

quella che si deve prendere in considerazione per trarre conclusioni storiche. Mentre la datazione radiocarbonica convenzionale viene pubblicata con un range di errore espresso in “ \pm anni” con confidenza del 68%, la datazione calibrata viene in genere fornita come intervallo di date di calendario entro il quale la data “vera” ha il 95% probabilità di cadere ovvero con limiti di confidenza del 95%. Le curve di calibrazione purtroppo non hanno un andamento continuo per cui, ad una datazione radiocarbonica convenzionale, possono corrispondere più datazioni di calendario (calibrate). Ad esempio si ha:

Data radiocarbonica convenzionale: 4550 ± 19 BP

Data calibrata: 3370 – 3320 BC (primo range), 3230 – 3100 BC (secondo range)

cioè ad una datazione radiocarbonica convenzionale, corrispondono due datazioni calibrate. In alcuni casi sono le considerazioni storiche a far propendere per un data piuttosto che per un'altra, altrimenti occorre allargare il range per comprendere tutti i possibili periodi. Si può dire quindi che la calibrazione aumenta l'intervallo dei valori e aumenta notevolmente l'accuratezza cioè la vicinanza al valore vero.

2.3.1 Esempio

A titolo di esempio riportiamo la calibrazione riguardante la datazione radiocarbonica della mummia del Similaun. Il Similaun è il corpo mummificato di un uomo che fu rinvenuto nel settembre del 1991, sul fondo coperto di ghiacci, al confine tra l'Austria e l'Italia. Il corpo era in ottimo stato di conservazione e nei suoi pressi furono trovati indumenti (una mantellina d'erba intrecciata, un calzare, resti di pelliccia), attrezzi vari (una borsa, un'ascia, un coltello), armi (arco con frecce) e resti vegetali. In tre serie di misure indipendenti furono trovate le età convenzionali che risultano tra loro consistenti e fanno risalire il decesso a circa 5.300 anni fa.

	Ossa	Tessuti	Tessuti e foglie dal cappello
Età convenzionale	4576 ± 27 BP	4523 ± 27 BP	4546 ± 17 BP
Età calibrata	3350-3300 a.C. 56%	3350-3300 a.C. 56%	3352-3300 a.C. 31%
	3210-3160 a.C. 36%	3210-3160 a.C. 36%	3235-3175 a.C. 36%
	3140-3120 a.C. 8%	3140-3120 a.C. 8%	3166-3108 a.C. 33%

Tabella 2.1: Età convenzionale e corrispondente età calibrata dell'uomo di Similaun

La presenza di tre range è dovuta all'andamento irregolare della curva di calibrazione pertanto la corrispondente età reale non è univoca. L'età può cadere con differenti probabilità in tre distinti intervalli temporali. Nella figura 2.2 riportata di seguito sull'ordinata è riportata l'età in numero di anni BP ricavata con la radiodatazione, sull'ascissa l'anno secondo il calendario ordinario. La curva centrale è la curva di calibrazione. Per convertire l'età radiometrica in quella del calendario ordinario si devono compiere i seguenti passi:

- Tracciare le linee orizzontali passanti per i valori di $4550 - 20$, $4550 + 20$ anni (i valori 4550 e 20 si sono ottenuti rispettivamente mediando i tre valori ottenuti nelle età convenzionali 4576, 4523, 4546 e 27, 27, 17).
- Tracciare le linee verticali che passano per le intersezioni delle linee orizzontali con la curva di calibrazione.
- Tracciare le linee verticali in corrispondenza delle intercette della curva di calibrazione con la retta orizzontale corrispondente a 4540 anni, che individuano i possibili valori degli anni del calendario ordinario.

Come abbiamo spiegato nella sezione 2.3 sulla calibrazione, occorre allargare il range per comprendere tutti i possibili periodi peggiorando così la precisione della misura. Possiamo perciò dire che l'uomo del Similaun è vissuto, con probabilità del 95%, tra il 3370 ed il 3100 a.C. In conclusione possiamo affermare che non è possibile stabilire quale intervallo sia quello reale sulla base delle sole informazioni scientifiche: si deve ricorrere al confronto con altre informazioni, ad esempio storiche, che possono aiutare ad escludere determinati periodi. In termini probabilistici ciò che entra in gioco è la densità a priori.

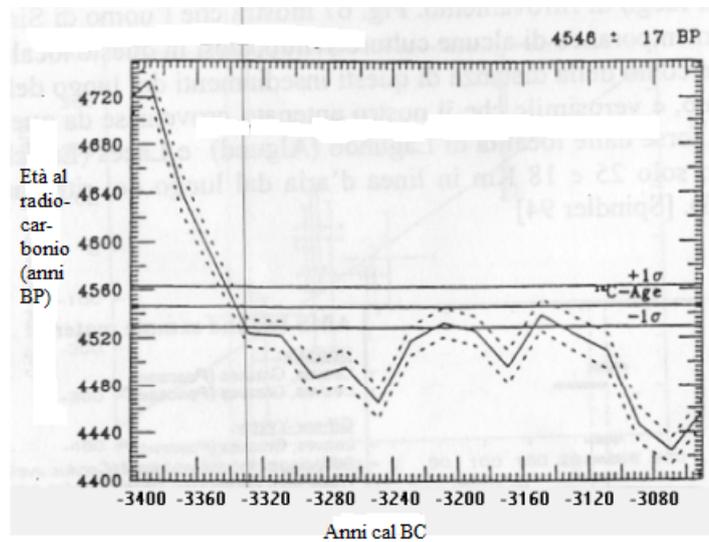


Figura 2.2: Curva di calibrazione e datazione dei reperti di Similaun

2.4 Margine di errore

Gli errori sperimentali, intrinseci in ogni esperimento, sono di solito valutati dalla ripetizione dell'esperimento. Quando una misura viene ripetuta, si suppone che la distribuzione dei risultati sia descritta da una gaussiana. Assumendo questo vale che, l'errore di un sigma ($\pm\sigma$) significa che vi è una probabilità del 68,3% che il valore vero si trovi dentro il risultato sperimentale, una probabilità del 95,4% all'interno di $\pm 2\sigma$ e 99,7% all'interno di $\pm 3\sigma$. Ogni laboratorio include un errore contributivo dal conteggio quindi ogni valore ottenuto sarà della forma $x \pm \sigma$ dove x è l'età del radiocarbonio (che dovrà poi essere convertita nel calendario convenzionale) e σ riporta l'errore stimato dal laboratorio. La legge di decadimento radioattivo dice che la metà degli atomi decade in 5730 anni. Ma in un piccolo intervallo di tempo, il numero che attualmente decade non è esattamente predicibile. Inoltre il tasso di creazione di Carbonio-14 nell'atmosfera non è rimasto sempre costante. Per esempio, con la rivoluzione industriale, e con la combustione di fossili che rilasciano anidride carbonica nella biosfera, la proporzione diminuì. Quindi secondo la datazione carbonica, le piante e animali che morirono in quel periodo sembrano più vecchie. Al contrario, i test di bombe nucleari effettuati agli

inizi degli anni sessanta hanno aumentato la concentrazione di Carbonio-14. Perciò le piante e animali morte in quel periodo, quando sottoposte alla datazione, apparirebbero più giovani di quanto non siano. Infine, un aumento dell'intensità del vento solare o del campo magnetico della Terra porterebbe a un rallentamento del tasso di creazione del Carbonio-14, rendendo la datazione più incerta. Queste variazioni dell'abbondanza del Carbonio-14 forniscono errori sistematici nelle datazioni con il radiocarbonio. Nella datazione al radiocarbonio occorre dunque valutare questi effetti per calibrare attentamente la misura. La tabella riporta una stima della correzione da apportare alle misure di datazione, che tiene conto di tutti questi effetti.

Età	Correzione
0 ÷ 2000 d.C	0
1000 a.C	100 anni
2000 a.C	300 anni
3000 a.C	600 anni
4000 a.C	800 anni

2.4.1 Sindone di Torino

Il caso mediatico più celebre di applicazione del metodo del radiocarbonio è stato la radiodatazione della Sindone di Torino eseguito in tre laboratori da un'equipe internazionale nel 1988, il cui risultato data la Sindone al periodo compreso tra il 1260 e il 1390 con un intervallo di confidenza di almeno il 95% e una approssimazione di 10 anni in più o in meno. La concordanza di alcune caratteristiche delle immagini con la narrazione evangelica fanno pensare che sia il sudario originale della morte di Gesù anche se, una bolla papale del 1390, ne affermava la non autenticità. Sotto la supervisione del British Museum di Londra sono stati incaricati della datazione tre noti laboratori specializzati (Zurigo, Oxford e Tucson). Per la datazione, ad ogni laboratorio fu consegnato un cam-

pione della sindone ed altri tre campioni di lino (campioni di controllo) di età nota (due medievali ed uno di epoca romana).

- Campione 1: lino della sindone di Torino;
- Campione 2: lino trovato in una tomba della Nubia del XII secolo d.C.;
- Campione 3: un pezzo del bendaggio di una mummia egiziana del II secolo d.C.;
- Campione 4: lino della veste di S. Luigi d'Angiò risalente al 1290-1310 d.C.

Vediamo di ripercorrere i passi fatti dagli studiosi per accertare l'autenticità di questo lenzuolo funerario. In base a quanto detto finora, l'idea è quella di costruire un modello che esemplifichi il nostro problema. A tal fine, individuamo le quattro componenti principali per l'analisi bayesiana:

informazione a priori: prima di inviare il campione di fili di lino della Sindone per l'analisi al radiocarbonio l'archeologo ha già alcune informazioni al riguardo per esempio che la Sacra Sindone è stata esposta in Francia nel 1350, portata a Torino nel 1578 e dal 1694 si è tenuta lì in un santuario appositamente progettato. Vi è qualche dubbio sul fatto che il lenzuolo sia lo stesso che quello ritrovato nel 1350. Alla luce di questa incertezza, il modo più ragionevole di procedere è quello di sostenere che niente sia conosciuto con esattezza, a priori, riguardo la datazione. Cioè consideriamo come funzione di densità di probabilità a priori $p(\theta)$ una distribuzione uniforme discreta.

i dati: prendiamo in considerazione una determinazione del radiocarbonio fatta dal laboratorio al radiocarbonio di Oxford che stima un'età pari a 750 con una deviazione standard pari a 30. Usando la stessa notazione riportata nella sezione 2.4, denotiamo quindi $x = 750$ e $\sigma = 30$.

verosimiglianza: data la particolare età al radiocarbonio, la verosimiglianza riunisce la curva di calibrazione del radiocarbonio e il modello statistico per l'incertezza.

informazione a posteriori: dopo aver ricevuto i risultati al laboratorio, l'informazione a posteriori di Θ è espressa in termini della funzione di densità di probabilità a posteriori, $p(\theta|x)$.

Basandosi su questa schematizzazione i risultati ottenuti sono riportati in tabella.

Laboratorio	Campione n.1	Campione n.2	Campione n.3	Campione n.4
Oxford	750 ± 30	940 ± 30	1980 ± 35	755 ± 30
Zurigo	676 ± 24	941 ± 23	1940 ± 30	685 ± 34
Tucson	646 ± 31	927 ± 32	1995 ± 46	722 ± 43

Tabella 2.2: Età convenzionali del radiocarbonio

Ad ogni età convenzionale relativa ad uno stesso campione, ottenuta dai tre laboratori, corrispondono intervalli di età calibrate per effetto dell'andamento irregolare della curva di calibrazione. Combinando insieme questi tre risultati si ottiene una media pesata pari a 689 ± 16 . Applicando la necessaria calibrazione, questa data viene convertita negli intervalli 1262 – 1384 con livello di confidenza 95%. Arrotondando ai 10 anni più vicini otteniamo l'intervallo 1260 – 1390. Supponiamo ora che il laboratorio di Zurigo effettui un'altra misura e che stimi che l'età del radiocarbonio sia 733 con una deviazione standard di 61. Invece di confrontare il risultato ottenuto con gli altri tre campioni di età nota, possiamo usare l'informazione a posteriori prodotta dal primo laboratorio (Oxford) come informazione a priori per analizzare il campione nel secondo laboratorio (Zurigo). In questo modo, dalle due determinazioni al radiocarbonio, il materiale organico è altamente probabile che sia morto tra l'800 e il 660 il che vuol dire, sulla scala del calendario convenzionale, tra il 1150 e il 1290 (ottenute facendo rispettivamente: $1950 - 800$ e $1950 - 600$). La verifica dell'ipotesi consiste nel rispondere alla seguente domanda: dato il risultato al radiocarbonio, qual'è la probabilità che la Sindone di Torino sia autentica?

Sia Θ la data del calendario sconosciuta e sia x la determinazione al radiocarbonio. L'ipotesi è che la Sindone sia datata al tempo della morte di Cristo ovvero tra il 20 d.C. e l'80 d.C.. Quindi Θ dovrebbe stare tra 1930 e il 1870 (ottenute facendo rispettivamente: $1950 - 20$ e $1950 - 80$). Vogliamo dunque calcolare la probabilità che la

Sindone sia datata intorno al tempo della morte di Cristo osservati i dati ovvero:

$$\mathbf{P}(1870 < \Theta < 1930|x) \quad (2.3)$$

e per il teorema di Bayes si ha:

$$\mathbf{P}(1870 < \Theta < 1930|x) = K\mathbf{P}(x|1870 < \Theta < 1930)\mathbf{P}(1870 < \Theta < 1930) \quad (2.4)$$

dove

- $\mathbf{P}(1870 < \Theta < 1930)$ è la probabilità a priori che la Sindone risalga al tempo della morte di Cristo
- K è la costante di normalizzazione

Evitando di entrare nelle polemiche sorte intorno a questa datazione, la Sindone di Torino risulterebbe, con probabilità del 95%, un falso medievale prodotto su di una tela datata tra il 1260 ed il 1390. Equivalentemente possiamo dire che il test di significatività a livello 0,5 dell'ipotesi che la Sindone di Torino sia il famoso sudario utilizzato per avvolgere il volto di Cristo ovvero che sia databile tra il 20 d.C. e l'80 d.C. ci permette di rifiutare l'ipotesi. A conferma di questa tesi sono state fatte alcune supposizioni per spiegare la differenza tra l'età stimata (600 anni) e quella supposta (2.000 anni):

- La radiodatazione della Sindone ha rilevato un'età apparente non congrua con quella reale, determinata dall'inquinamento da Carbonio moderno che potrebbe essere stato prodotto nella fase di estrazione del Carbonio originale dal reperto.
- La presenza di inquinanti di epoche più recenti avrebbe fornito una datazione più giovane rispetto a quella ipotizzata. Ad esempio, le variazioni dell'abbondanza del Carbonio-14 causate da fattori come quelli di cui si è parlato nella sezione 2.4 o l'inquinamento provocato dall'atmosfera surriscaldata dell'incendio della Sainte-Chapelle du Saint-Suaire nella quale era conservata del 1532 d.C.. L'ipotesi è stata avanzata da alcuni ricercatori russi, ma sperimentalmente non si sono ottenute conferme a questa teoria.
- La produzione di Carbonio recente da parte di contaminanti superficiali di natura biologica, ad esempio batteri, muffe e funghi, non interamente eliminati dai trattamenti di pulizia chimica, farebbe apparire la Sindone più giovane.

- La datazione è stata eseguita su una piccolissima porzione della Sindone che potrebbe non avere le caratteristiche medie di tutto il lenzuolo.

Se si tenesse conto di queste perplessità e delle considerazioni storiche e quindi delle informazioni a priori si potrebbe ipotizzare che la datazione non sia corretta, cioè che il reperto sia molto più antico di quanto stimato.

Capitolo 3

Conclusioni

Questa tesi ha voluto approfondire la comprensione della statistica bayesiana e ha voluto sottolineare i numerosi vantaggi che l'inferenza bayesiana offre nella risoluzione ai problemi di diversa natura. Grazie all'introduzione del concetto di probabilità è possibile quantificare l'incertezza delle ipotesi. L'uso dell'approccio bayesiano permette di tenere conto dei valori dei dati mancanti, non semplicemente imputandoli o ignorandoli, ma tenendone conto nel processo di modellazione. Il teorema di Bayes rappresenta lo strumento matematico attraverso il quale un individuo aggiorna le proprie conoscenze sulla base del risultato sperimentale. Così è possibile incorporare le conoscenze a priori pertinenti fornendo un modo logico e coerente di aggiornare le informazioni che si hanno prima di osservare i dati a quelle ottenute dopo aver preso in considerazione i dati. Nella mia trattazione ho voluto mostrare il processo attraverso il quale gli archeologi analizzano i loro dati e come questi possono essere resi rigorosi ed efficaci grazie all'uso dei modelli statistici. Possiamo sottolineare i seguenti vantaggi.

Osservazione 1. *La filosofia bayesiana permette di misurare l'incertezza circa i valori dei parametri esprimendola in termini di probabilità e permette di valutare la veridicità e la qualità di certe ipotesi.*

Osservazione 2. *La caratteristica dell'approccio bayesiano di riuscire a combinare due o più informazioni rende possibile la combinazione delle informazioni provenienti da fonti diverse: nell'esempio del capitolo precedente il laboratorio di Oxford e quello di Zurigo forniscono un risultato più preciso rispetto al primo trovato.*

Osservazione 3. *L'approccio bayesiano all'analisi dei dati fornisce un quadro completo all'interno del quale le informazioni e la conoscenza a priori possono essere incluse esplicitamente nell'inferenza statistica. Gli archeologi si avvalgono quasi sempre di ragionamenti euristici e delle loro conoscenze, sia nel fare inferenza sia nella pianificazione di progetti sul campo e di laboratorio. Pertanto questo approccio risulta molto utile per i loro obiettivi.*

Osservazione 4. *Grazie alla capacità di contare le emissioni radioattive con accuratezza è possibile ottenere l'esatta età radiocarbonica di un campione. Con l'utilizzo della probabilità è possibile stabilire il margine di errore di tali rilevamenti.*

Bibliografia

- [1] J.M.Bernardo e A.F.M.Smith. *Bayesian Theory*. Wiley, Chichester,1994
- [2] S.Bowman. *Interpreting the past: radiocarbon Dating*. British Museum Publications, London,1990
- [3] D.V.Lindley. *Introduction to probability and Statistics from Bayesian Viewpoint*. University Press, Cambridge, 1995
- [4] C.E.Buck, W.G.Cavanagh e C.D.Litton. *Bayesian Approach to Interpreting Archaeological Data*. Wiley, Chichester,1996
- [5] G. Bonani et al. *Fractionation, precision and accuracy in ^{14}C and ^{13}C measurement*. Nucl. Instr. And Meth. Phys.1987
- [6] O'Hagan. *Kendall's advanced Theory for statisticians: Bayesian Inference*. Edward Arnold, London, 1994

Ringraziamenti

Desidero ringraziare infinitamente il Professor Campanino per la sua disponibilità, competenza e soprattutto per l'interesse e la passione che è riuscito a suscitare in me in questa materia. Infine vorrei ringraziare i miei genitori per l'opportunità che mi hanno dato di proseguire gli studi facendomi vivere questa bellissima esperienza che mi ha fatto crescere intellettualmente e personalmente.