Introduzione alla statistica bayesiana

Problemi diretti e inversi, Teorema di Bayes, statistica classica, statistica Bayesiana (parte 2)

Docente: Matilde Trevisani

DEAMS

A.A. 2025/2026

(aggiornato: 2025-09-29)

Agenda

Introduzione alla Statistica Bayesiana

- Problemi diretti e inversi
- Il Teorema di Bayes
- Statistica moderna (classica)
- Statistica Bayesiana

- Probabilità soggettiva
- Distribuzioni a priori
- Approcci odierni

Problemi diretti e inversi

Da problemi diretti (processo noto) a $\longrightarrow$

Il calcolo della probabilità, dalle origini all'enunciazione del Teorema di Bayes, si è sviluppato per risolvere problemi ***diretti***: conosco il meccanismo casuale che genera le osservazioni e posso calcolare la probabilità dei vari risultati.

In un pacchetto di *jelly bears* ce ne sono R rossi e G gialli. Se ne estraiamo m (con "reimbussolamento"), qual è la probabilità che x siano rossi?

Dato $\theta = \frac{R}{R+G}$,

$$\Pr(X = x) = \binom{m}{x} \theta^x \, (1-\theta)^{m-x}$$

Supponiamo di non conoscere R, possiamo ancora calcolare la probabilità sopra ipotizzando una distribuzione di probabilità per (la v.a.) R e applicando il teorema delle probabilità totali.

Supponiamo, e.g., che R sia deciso lanciando un dado a n facce (dove $n = R + G$).

$$\Pr(X = x) = \sum_{i=1}^{n} \Pr(R = i \cap X = x) = \sum_{i=1}^{n} \Pr(R = i) \Pr(X = x | R = i)$$

dove $X = 1, \ldots, m$ e $\theta = R/n$.

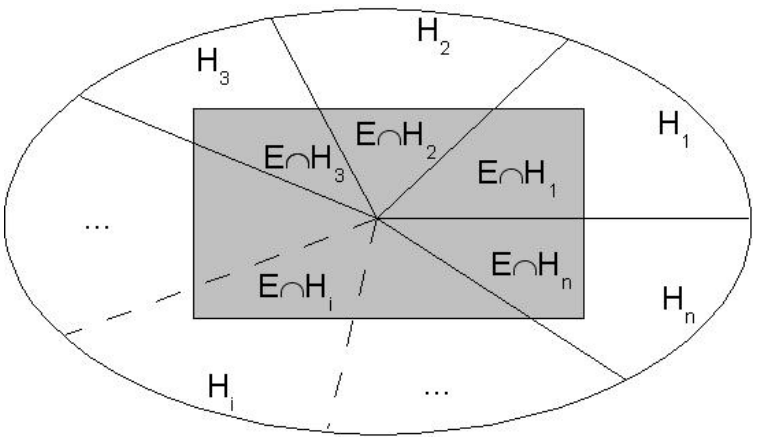
Law of total probability

Let $\{H_i | i = 1, \ldots, n\}$ be a partition of Ω, i.e.,

1. $\bigcup_{i=1}^{n} H_i = \Omega$ (exhaustive),
2. $H_i \cap H_j = \phi$ if $i \neq j$ (pairwise incompatible),

then

$$P(E) = P(E \cap \Omega) = \sum_{i=1}^{n} P(H_i \cap E) = \sum_{i=1}^{n} P(H_i) P(E|H_i)$$



Law of total probability in tabular form

E.g., se $n = 10, m = 5$, allora $X = 0, 1, \ldots, 5, \theta = 0.1, 0.2, \ldots, 0.9, 1$

Probabilità condizionali:

$$\Pr(X = x|R = i) \; o \; \Pr(X = x|R = 10\theta)$$

		Urn composition (θ, proportion of red marbles)									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Sample (x)	0	.5905	.3277	.1681	.0778	.0312	.0102	.0024	.0003	.0000	0
	1	.3280	.4096	.3601	.2592	.1562	.0768	.0284	.0064	.0004	0
	2	.0729	.2048	.3087	.3456	.3125	.2304	.1323	.0512	.0081	0
	3	.0081	.0512	.1323	.2304	.3125	.3456	.3087	.2048	.0729	0
	4	.0005	.0064	.0284	.0768	.1562	.2592	.3601	.4096	.3280	0
	5	.0000	.0003	.0024	.0102	.0312	.0778	.1681	.3277	.5905	1

Probabilità congiunte:

$$\Pr(R = i \cap X = x) = \Pr(R = i)\Pr(X = x|R = i))$$

x	Urn composition (θ, proportion of red marbles)										$P(X = x)$
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	
0	.05905	.03277	.01681	.00778	.00313	.00102	.00024	.00003	.00000	0	.12083
1	.03280	.04096	.03601	.02592	.01562	.00768	.00283	.00064	.00004	0	.16252
2	.00729	.02048	.03087	.03456	.03125	.02304	.01323	.00512	.00081	0	.16665
3	.00081	.00512	.01323	.02304	.03125	.03456	.03087	.02048	.00729	0	.16665
4	.00005	.00064	.00284	.00768	.01562	.02592	.03602	.04096	.03281	0	.16253
5	.00000	.00003	.00024	.00102	.00313	.00778	.01681	.03277	.05905	.1	.22083

che sommandole danno: $\sum_{i=1}^{10} \Pr(R = i \cap X = x) = \Pr(X = x)$

problemi indiretti o inversi (probabilità delle cause)

Nell'ambito esperimento di cui sopra, possiamo anche porre la seguente domanda

Avendo estratto $X = x$ orsetti, qual è la probabilità che nel pacchetto ce ne siano R rossi?

Questo problema è risolto dal **Teorema di Bayes**.

Thomas Bayes (c. 1702-1761)} was a Presbyterian minister. In *Essay Towards Solving a Problem in the Doctrine of Chances* (1763) he considers the inverse probability problem for which he formalizes a solution. His work was published posthumously by his friend Richard Price (1723-1791).

Teorema di Bayes

Formulazione originaria del Teorema di Bayes

Il reverendo Thomas Bayes scrisse un articolo intitolato "Essay Towards Solving a Problem in the Doctrine of Chances" (pubblicato post-mortem nel 1763)

Theorem (PROP. 3)

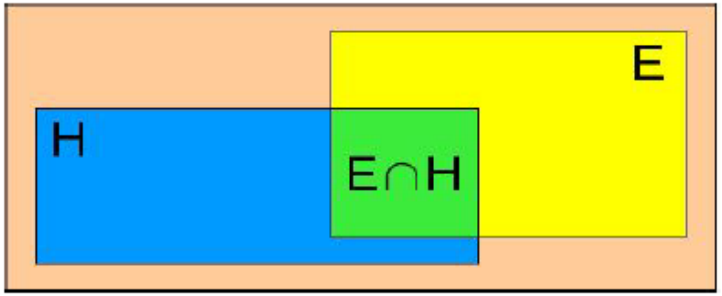
> The probability that two subsequent events will both happen is a ratio compounded of the probability of the 1st, and the probability of the 2d on supposition the 1st happens.

Corollary (PROP. 3)

> Hence if of two subsequent events the probability of the 1st be a/N , and the probability of both together be P/N , then the probability of the 2d on supposition the 1st happens is P/a.

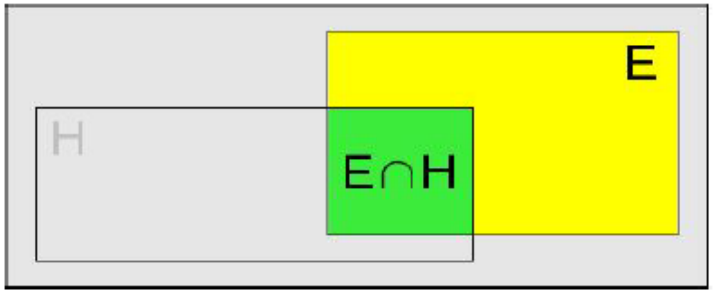
Teorema di Bayes

Siano E e H due eventi, con $\mathrm{P}(E) \neq 0$,



allora

$$\mathrm{P}(H|E) = \frac{\mathrm{P}(H \cap E)}{\mathrm{P}(E)} = \frac{\mathrm{P}(H)\mathrm{P}(E|H)}{\mathrm{P}(E)}$$



Bayes' Theorem for the *jelly bears* example

Avendo estratto $X = x$ orsetti, qual è la probabilità che nel pacchetto ce ne siano R rossi?

La risposta ci viene dal Teorema di Bayes:

$$\mathrm{P}(R = \theta n | X = x) = \frac{\mathrm{P}(R = \theta n \cap X = x)}{\mathrm{P}(X = x)} = \frac{\mathrm{P}(R = \theta n)\mathrm{P}(X = x | R = \theta n)}{\mathrm{P}(X = x)}$$

Supponiamo che $X = 2$, si considerino le probabilità congiunte

$$\Pr(R = 10\theta \cap X = 2)$$

x	Urn composition (θ, share of red marbles)										$P(X = x)$
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	
0	.05905	.03277	.01681	.00778	.00313	.00102	.00024	.00003	.00000	0	.12083
1	.03280	.04096	.03601	.02592	.01562	.00768	.00283	.00064	.00004	0	.16252
2	.00729	.02048	.03087	.03456	.03125	.02304	.01323	.00512	.00081	0	.16665
3	.00081	.00512	.01323	.02304	.03125	.03456	.03087	.02048	.00729	0	.16665
4	.00005	.00064	.00284	.00768	.01562	.02592	.03602	.04096	.03281	0	.16253
5	.00000	.00003	.00024	.00102	.00313	.00778	.01681	.03277	.05905	.1	.22083

da cui

$$\Pr(R = 10\theta \mid X = 2) = \frac{\Pr(R = 10\theta \cap X = 2)}{\Pr(X = 2)}$$

θ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$P(R = 10\theta \mid X = 2)$	.0437	.1229	.1852	.2074	.1875	.1383	.0794	.0307	.0049	0

Cosa c'è di così strano?

Ciò che abbiamo appena ottenuto, la probabilità di ogni possibile composizione dell'urna, è incontrovertibile e standard.

A statistical problem in the search for *causes*

Ora rendiamo questo problema più interessante:

In un pacchetto di n *jelly bears* ce ne sono R rossi ($R \geq 1$). Ne estraiamo m (con "reimbussolamento") e osserviamo x rossi, cosa possiamo dire riguardo a R ?

Nella presentazione precedente del problema R si era supposto generato da un meccanismo casuale e ciò rendeva la soluzione standard (non controversa).

Ora, R ($R = R(\theta)$) è semplicemente *non noto*.

Ora il problema è presentato come un **problema statistico** generale in cui si sa che l'osservazione è generata da un meccanismo casuale di cui non si conoscono tutte le *caratteristiche*.

- Come interpretiamo la probabilità di una probabilità θ (delle cause), $P(\theta|X = x)$?
- Può rappresentare le nostre opinioni su θ ?
- Per alcuni Sì, per altri No.

Due tipi di incertezza

Distinguiamo due tipi di incertezza:

- **aleatoria** in cui l'incertezza è dovuta alla casualità (*randomness*)
    - non siamo in grado di ottenere osservazioni che potrebbero ridurre questa incertezza
- **epistemica** in cui l'incertezza è dovuta alla mancanza di conoscenza
    - Siamo in grado di ottenere osservazioni che possono ridurre questa incertezza
    - due osservatori possono avere un'incertezza epistemica diversa

Aggiornamento dell'incertezza (probabilità)

Probabilità di rosso $\frac{\mathrm{R}}{\mathrm{R+G}} = \theta$

- $p(y = \text{rosso}|\theta) = \theta$ incertezza **aleatoria**
- $p(\theta)$ incertezza **epistemica**

L'estrazione ripetuta di orsetti aggiorna la nostra incertezza sulla proporzione

- $p(\theta|\mathrm{y} = \text{rosso}, \text{giallo}, \text{rosso}, \text{rosso}, \ldots) = ?$
- secondo la **regola di Bayes** $p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$

(*Anticipazione*) Modello *vs* Verosimiglianza (*Likelihood*), a priori *vs* a posteriori

- regola di Bayes $p(\theta|y) \propto p(y|\theta)p(\theta)$
- Modello: $p(y|\theta)$ come funzione di y dato un θ fissato descrive l'incertezza aleatoria
- Verosimiglianza: $p(y|\theta)$ ($= L(\theta|y)$) come funzione di θ dato un y fissato **fornisce informazione sull'incertezza epistemica**, ma non è una distribuzione di probabilità
- *Bayes rule* combina la verosimiglianza con l'incertezza a priori $p(\theta)$ e li transforma nella incertezza a posteriori aggiornata.

Bayes: inference for a probability

This is stated and more or less solved in Bayes essay as follows.

> *Given* the number of times on which an unknown event has happened and failed:
>
> *Required* the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.

If nothing is known of an event but that it has happened p times and failed q in $p+q$ or n trials, and from hence I judge that the probability of it's happening in a single trial lies between $\frac{p}{n}+z$ and $\frac{p}{n}-z$ my chance to be right is *greater* than $\frac{\sqrt{\mathrm{K}pq}\times h}{2\sqrt{\mathrm{K}pq}+hn^{\frac{1}{2}}+hn^{-\frac{1}{2}}} \times 2\mathrm{H}-\frac{\sqrt{2}}{\sqrt{\mathrm{K}}}\times\frac{n+1}{n+2}\times\frac{1}{mz}\times\sqrt{1-\frac{2m^2z^2}{n}|^{\frac{n}{2}+1}}$ and less than $\frac{\sqrt{\mathrm{K}pq}\times h}{2\sqrt{\mathrm{K}pq}-hn^{\frac{1}{2}}-hn^{-\frac{1}{2}}}$ multiplied by the 3 terms $2\mathrm{H}-\frac{\sqrt{2}}{\sqrt{\mathrm{K}}}\times\frac{n+1}{n+2}\times\frac{1}{mz}\times\sqrt{1-\frac{2m^2z^2}{n}|^{\frac{n}{2}+1}}+\frac{\sqrt{2}}{\sqrt{\mathrm{K}}}\times\frac{n}{n+2}\times\frac{n+1}{n+4}\times\frac{1}{m^3z^3}\times\sqrt{1-\frac{2m^2z^2}{n}|^{\frac{n}{2}+2}}$ where m^2, K, h and H stand for the quantities already explained.

Bayes solution was not actually very clear, the one from Laplace was better.

Laplace e la probabilità (epistemica) che nasca una femmina

Laplace (1749-1827) è stato il primo a formulare un problema statistico e a risolverlo con la statistica bayesiana.

La domanda che si pose era se la probabilità che nasca una femmina (θ) fosse o no inferiore a 0.5.

Il problema è del tutto analogo a quello dell'estrazione dall'*urna* sopra, se non per il fatto che esiste un continuum di possibili composizioni dell'urna.

Egli osservò che a Parigi, dal 1745 al 1770, le nascite erano state 493,472 di cui 241,945 femmine da cui derivò che

$$P(\theta \geq 0.5 \,|\, \text{dati}) \approx 1.15 \times 10^{-42}$$

ricavando la "certezza morale" che $\theta < 0.5$.

Pierre-Simon Laplace (1749-1827) in *Essai philosophique sur les probabilites (1814)* gives a systematic treatment to the approach which we call Bayesian today.

Laplace e l'estensione del teorema di Bayes

Laplace estese il teorema di Bayes a n possibili cause, H_i, $i = 1, \ldots, n$, di un evento E.

Dati:

- $H_1, \ldots, H_n$, un insieme di ipotesi
- $\mathrm{P}(H_i)$, $i = 1, \ldots, n$, probabilità a priori, $\sum_i \mathrm{P}(H_i) = 1$
- $\mathrm{P}(E|H_i)$, $i = 1, \ldots, n$, verosimiglianza di E quando H_i è vera

$$\mathrm{P}(H_i|E) = \frac{\mathrm{P}(E|H_i)\mathrm{P}(H_i)}{\sum_{i=1}^{n} \mathrm{P}(E|H_i)\mathrm{P}(H_i)}$$

La probabilità a posteriori di H_i dato E è proporzionale al prodotto della probabilità a priori di H_i e della verosimiglianza di E quando H_i è vera.

Corollario: confronto tra due ipotesi

Si supponga che si confrontino due particolari ipotesi, H_i e H_j, il rapporto a posteriori è dato da:

$$\frac{\mathrm{P}(H_i|E)}{\mathrm{P}(H_j|E)} = \frac{\mathrm{P}(H_i)}{\mathrm{P}(H_j)} \times \frac{\mathrm{P}(E|H_i)}{\mathrm{P}(E|H_j)}$$

i.e., il rapporto delle probabilità a priori (*prior odds*) per il rapporto delle verosimiglianze (*likelihood ratio*).

Esempio: Diagnosi medica

Obiettivo: valutare se una persona sottoposta a test - risultato positivo - sia malata

- Un paziente può essere affetto da una data malattia (essere nello stato H_1) o non esserlo (essere nello stato H_2)
- $P(H_1)$ è la prevalenza della malattia nella popolazione alla quale si suppone che il paziente appartenga (probabilità **a priori** di H_1, $P(H_2)$ = $1 - P(H_1)$)
- Il paziente viene sottoposto a test da cui si ricava l'**informazione** se sia verosimilmente malato, test positivo (E) o test negativo ($\bar{E}$):
- $P(E|H_1)$, $P(\bar{E}|H_2)$ sono i tassi di *true posititive* e *true negative*, anche detti sensibilità e specificità del test clinico

> La situazione ottimale è
>
> $$\begin{aligned} P(E|H_1) = 1 \quad &\rightarrow \quad P(\bar{E}|H_1) = 0 \quad \text{false negative} \\ P(\bar{E}|H_2) = 1 \quad &\rightarrow \quad P(E|H_2) = 0 \quad \text{false positive} \end{aligned}$$

- Il teorema di Bayes ci permette di capire combinando le caratteristiche del test con la prevalenza della malattia quale sia il **potere discriminante** del test diagnostico.

Covid-19 e tamponi naso-faringei

Studi recenti suggeriscono che i tamponi naso-faringei usati per la diagnosi di COVID-19 abbiano sensibilità $P(E|H_1)$ = 0.777 e specificità $P(\bar{E}|H_2)$= 0.988 e che la prevalenza di COVID-19 in Italia (in fase pandemica) sia circa pari a $P(H_1)$ = 0.13.
Applichiamo il teorema di Bayes per calcolare la probabilità di essere malato in seguito a risultato positivo di un test diagnostico:

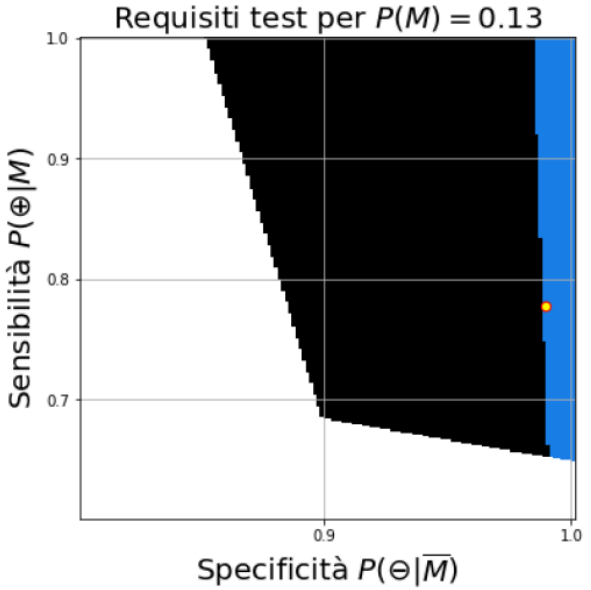
$$\begin{aligned} P(H_1|E) &= \frac{\mathrm{P}(E|H_1)\mathrm{P}(H_1)}{\mathrm{P}(E|H_1)\mathrm{P}(H_1) + \mathrm{P}(E|H_2)\mathrm{P}(H_2)} \\ &= \frac{\text{sensibilità} \cdot \text{prevalenza}}{\text{sensibilità} \cdot \text{prevalenza} + (\text{1-specificità}) \cdot (\text{1-prevalenza})} = \text{‘} \\ &= \frac{0.777 \cdot 0.13}{0.777 \cdot 0.13 + (1 - 0.988) \cdot (1 - 0.13)} = 0.9063257 \end{aligned}$$

la probabilità di non essere malato in seguito a risultato negativo:

$$\begin{aligned} P(H_2|\bar{E}) &= \frac{\text{specificità} \cdot \text{1-prevalenza}}{\text{specificità} \cdot \text{1-prevalenza} + (\text{1-sensibilità}) \cdot (\text{prevalenza})} = \\ &= \frac{0.988 \cdot (1 - 0.13)}{0.988 \cdot (1 - 0.13) + (1 - 0.777) \cdot 0.13} = 0.9673738 \end{aligned}$$ ‘

Test ideale

Data la prevalenza di una malattia, M, le caratteristiche ideali del test per una diagnosi il più corretta possibile sono mostrate nel grafico.



Requisiti test per covid-19

L'area **nera** indica i **requisiti necessari**

$$P(M|T) > .5 \quad \text{e} \quad P(M|\bar{T}) < .05$$

per un test con $P(M) = .13$ (**azzurra** i **requisiti ottimali** per $P(M|T) > .9$).

Il punto **giallo** indica i parametri di sensibilità e specificità dei test RT-PCR SARS-CoV-2 RNA test per COVID-19: sensibilità $P(T|M) = 0.777$ e specificità di $P(\bar{T}|M) = 0.988$.

Un problema statistico

Riprendendo l'esempio dell'*urna*, il punto cruciale è che $R(\theta)$ non è casuale nel senso di essere generato attraverso un esperimento casuale (incertezza aleatoria). Piuttosto, $R(\theta)$ è **non noto** (incertezza epistemica).

Come dobbiamo quindi interpretare la probabilità che attribuiamo a θ: $P(\theta| = x)$?

Può rappresentare le nostra opinione sul valore di θ?

Secondo alcuni potrebbe, secondo altri è un'interpretazione priva di senso.

Bayesian approach put aside

- Nei primissimi esempi, l'inferenza statistica - di Thomas Bayes in termini generali e Pierre-Simon Laplace su dati reali - era basata sul paradigma bayesiano.
- Quindi, fine XIX e inizio XX secolo, l'approccio bayesiano fu messo da parte per ragioni filosofiche e tecniche
  - L'idea che la probabilità possa essere utilizzata per modellare l'ignoranza / le opinioni era considerata ridicola.
  - Inoltre, per ottenere $P(\theta|X = x)$ dobbiamo iniziare da $P(\theta)$, un'opinione a priori su θ (che viene prima delle osservazioni), il che comporta l'introduzione di un elemento di soggettività nell'analisi, che allora era, ancora una volta, ritenuto non scientifico.
  - Infine, c'erano problemi di calcolo: anche per problemi relativamente semplici, l'approccio bayesiano comporta solitamente calcoli intrattabili.

Modern (classical) statistics

New questions, new answers

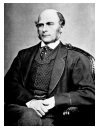
Tra il XIX e il XX-esimo secolo nascono nuovi ambiti di applicazione delle tecniche statistiche come:

-la genetica e l'ereditarietà;
-il controllo di qualità

One has a, possibly incomplete, theory, i.e., a $Model$, on how some outcome is generated and wants to use $Data$ to confirm/clarify the said model.

Vengono sviluppati nuovi approcci in cui il parametro θ è un numero fisso.

William Gosset (1876-1937) Working for Guiness, he developed the *Student-t* distribution to evaluate quality of barley.



sir Francis Galton (1822-1911) Founded the Eugenics Record Office in London, later the Galton labooratory. Develops *linear regression*.

Karl Pearson (1857-1936) Introduces the concept of correlation and of goodness of fit.

Verosimiglianza e il campionamento ripetuto

- L'inferenza è basata sulla **verosimiglianza**: si confronta $P(Data|Model)$ (nell'esempio dell'urna $L(\theta) \propto P(X = x|R = n\theta)$) per i differenti modelli
  (Nella statistica bayesiana confrontiamo $P(Model|Data)$),

- la performance è valutata secondo il principio del **campionamento ripetuto**: le procedure sono valutate in base a ripetizioni fittizie dell'esperimento (*Mean Square Error*, *coverage probability*, *significance level*, etc.).

sir Ronald Fisher (1890-1932) introduces, among other things, the concepts of *likelihood*, *analysis of variance*, *experimental design*. Also, he originates the ideas of *sufficiency*, *ancillarity*, and *information*. His main works: *Statistical Methods for Research Workers* (1925), *The design of experiments* (1935), *Contributions to mathematical statistics (1950)*, *Statistical methods and statistical inference* (1956).

Egon Pearson (1895-1980) with Jerzy Neyman develops the *theory of hypotheses testing*.

Verosimiglianza

La verosimiglianza sintetizza l'informazione su θ proveniente da $X = x$

$$L(\theta) \propto P(X = x | R = 10\theta)$$

x	Urn composition (θ, share of red marbles)									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	.5905	.3277	.1681	.0778	.0312	.0102	.0024	.0003	.0000	0
1	.3280	.4096	.3601	.2592	.1562	.0768	.0284	.0064	.0004	0
2	.0729	.2048	.3087	.3456	.3125	.2304	.1323	.0512	.0081	0
3	.0081	.0512	.1323	.2304	.3125	.3456	.3087	.2048	.0729	0
4	.0005	.0064	.0284	.0768	.1562	.2592	.3601	.4096	.3280	0
5	.0000	.0003	.0024	.0102	.0312	.0778	.1681	.3277	.5905	1

Dalla verosimiglianza possiamo ottenere:

- lo stimatore di massima verosimiglianza:
  - $\hat{\theta} = X/5 = 0.6$
- p-values:
  - il p-value per l'ipotesi $\theta \leq 0.2$ è 0.0579 (*come si calcola?*)

Principio del campionamento ripetuto

Secondo il principio del campionamento ripetuto, valutiamo le procedure in base a come si comporterebbero a lungo termine con nuovi set di dati

Utilizzando il principio del campionamento ripetuto possiamo valutare le prestazioni di

- stimatori $\Rightarrow$ Errore quadratico medio (*Mean Square Error*)
- intervalli di confidenza $\Rightarrow$ probabilità di copertura (*coverage probability*)

Il *test d' ipotesi di Neyman-Pearson* ha il legame più evidente con campionamento ripetuto:

- il *livello di significatività* (α) è la frequenza relativa con cui ci aspettiamo di rifiutare un'ipotesi nulla se dovessimo eseguire il test su un certo numero di campioni provenienti da una popolazione per la quale l'ipotesi nulla è vera.
- la *potenza* è ...

Approccio classico

Per quanto riguarda la probabilità che nasca una femmina, Laplace osservò che a Parigi, dal 1745 al 1770, le nascite erano state 493,472 di cui 241,945 femminili $\rightarrow$

- stima di ML $\hat{\theta} = \frac{241{,}945}{493{,}472} = 0.4903$
- 95% IC $[0.4889, 0.4917]$
- *p-value* per l'ipotesi $H_0 : \theta \geq 0.5$ è ≈ 0

- La stima migliore per θ è 0.4903
- abbiamo ottenuto un intervallo $[0, 4889, 0, 4917]$ come una realizzazione di un intervallo casuale che ha probabilità del 95% di coprire il vero valore di θ
- se $H_0 : \theta \geq 0.5$ fosse vera, la probabilità di osservare una statistica estrema come quella osservata sarebbe ≈ 0

Ciò che ci dice non è di immediata traduzione . . . Con questo intendo dire che dobbiamo fare un ulteriore passo avanti per tradurlo in informazioni su θ. (vedi slide sull'inferenza frequentista)

Classical approach or approaches?

Note that within the classical approach different views can be distinguished, this is particularly evident in hypotheses testing.

A **Fisherian approach** is to view the likelihood as central as a measure of evidence brought by the data. As such, a p-value is a measure of evidence against a given hypotheses.

The **Neyman-Pearson** view is behavioural, they devise a decision rule which controls the probability of error (not the overall one, but at least the conditional ones).

The above is a very simplistic summary, however it is true that the two approaches are incompatible and there have been harsh debates between the proponents.

Interpretazione dei risultati nell'approccio Bayesiano

> A primary motivation for Bayesian thinking is that it facilitates a ***common sense interpretation*** of statistical conclusions.
>
> -- *Gelman*

Al contrario della stima d'intervallo o della verifica d'ipotesi, la statistica Bayesiana ci dice quello che vogliamo sapere, la statistica classica no, ed è probabile che molti utenti interpreterebbero erroneamente i risultati della statistica classica alla maniera bayesiana.

> Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a ***probability distribution on the parameters of the model and on unobserved quantities*** such as predictions for new observations.
>
> -- *Gelman*

Approccio classico vs approccio bayesiano all'inferenza

Nell'INFERENZA CLASSICA

- il parametro è una costante.
- la conclusione non è derivata all'interno delle regole del calcolo delle probabilità (queste sono usate in effetti, ma la conclusione non ne è una diretta conseguenza)
- la verosimiglianza e la distribuzione di probabilità del campione vengono utilizzate

Framework per estrarre evidenza dai dati.

Nell'INFERENZA BAYESIANA

- il parametro è una v.a.
- il ragionamento e la conclusione sono una conseguenza immediata delle regole del calcolo delle probabilità (del Teorema di Bayes in particolare);
- la verosimiglianza e la distribuzione a priori vengono utilizzate

Framework per aggiornare le informazioni.

Statistica Bayesiana

Il teorema di Bayes a fondamento dell'inferenza Bayesiana

> Il teorema di Bayes costituisce la chiave di volta e il concetto informatore di ogni attività costruttiva del pensiero.
>
> -- *Bruno de Finetti, 1970*

il teorema di Bayes è lo strumento par excellence capace di aggiornare le probabilità delle ipotesi H_j alla luce di nuovi fatti.

L'esempio che segue evidenzia l'errore che si commette quando si trascurano le informazioni disponibili, nel nostro caso le probabilità a priori, e si traggono inferenze (e conclusioni) usando solo le verosimiglianze.

Esempio: Diagnosi medica
(Gigerenzer, 2002)

- Consider women aged 40-50 with no family history of cancer and no symptoms of cancer
  - The proportion that have breast cancer is .008
- Conduct mammogram screening
  - If a woman has breast cancer, the probability of a positive mammogram is .90
  - If a woman does not have breast cancer, the probability of a positive mammogram is .07
  - A woman undergoes the mammogram, the result is positive
- What should we infer?
  - What's the probability the woman has breast cancer?

See natural frequencies →

Maximum Likelihood

Logica della Massima Verosimiglianza (*Maximum Likelihood*)

- A general approach to parameter estimation
- The use of a model implies that the data may be sufficiently characterized by the features of the model, including the unknown parameters
- Parameters govern the data in the sense that the data ***depend*** on the parameters
    - Given values of the parameters we can calculate the (conditional) probability of the data
    - Mammogram (data) depends on breast cancer status (parameter)
    - When conventional statistical approaches discuss a “model” they usually refer to this dependence structure
- Maximum likelihood (ML) estimation asks: *“What are the values of the parameters that make the data most probable?”*

Maximum Likelihood

- Specify the conditional probability of the data as $p(x|\theta)$
  - E.g., $\theta = (\mu, \sigma^2)$, $p(x|\mu, \sigma^2) = N(\mu, \sigma^2)$
  - This describes the structure as function of x
- When taken as a function of θ, this is referred to as the ***likelihood*** $L(\theta|x) = p(x|\theta)$
- ML estimation then maximizes this function w.r.t. θ, using the known values of x

Diagnosi medica
Conditional Probability Distribution

- If a woman has breast cancer, the probability of a positive mammogram is .90

  $p(Mam = +|BC = Y) = .90\,,\ p(Mam = -|BC = Y) = .10$

- If a woman does not have breast cancer, the probability of a positive mammogram is .07

  $p(Mam = +|BC = N) = .07\,,\ p(Mam = -|BC = N) = .93$

	Mammogram Result	
Breast Cancer	Positive	Negative
Yes	.90	.10
No	.07	.93

Diagnosi medica
Maximum Likelihood

- A woman undergoes the mammogram, the result is positive

$$L(BC = Y|Mam = +) = p(Mam = +|BC = Y) = .90$$

$$L(BC = N|Mam = +) = p(Mam = +|BC = N) = .07$$

	Mammogram Result	
Breast Cancer	Positive	Negative
Yes	.90	
No	.07	.93

- ML: *What is the value of BC that maximizes the likelihood?*

$$\widehat{BC} = Yes$$

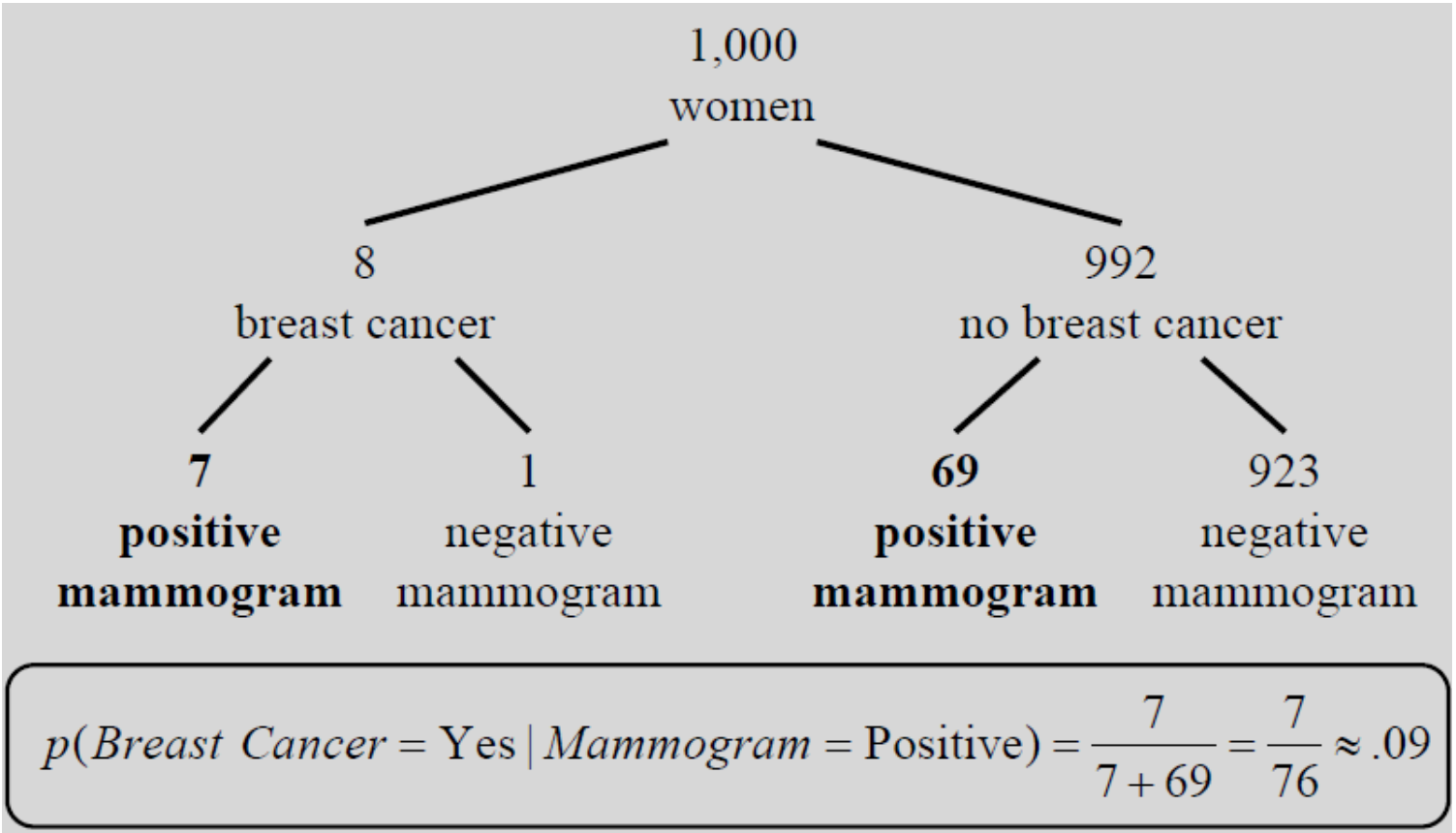
Inferenza Bayesiana

Esempio: Diagnosi medica
Frequenze naturali

← See case description

- Consider 1,000 women aged 40-50 with no family history of cancer and no symptoms of cancer
  - 8 of the 1,000 women have breast cancer
- Conduct mammogram screening
  - Of the 8 women with breast cancer, 7 will have a positive mammogram
  - Of the remaining 992 women w/o breast cancer, 69 will have a positive mammogram
- A woman undergoes the mammogram, the result is positive
- What should we infer?
  - What's the probability the woman has breast cancer?

Esempio: Diagnosi medica
Diagramma ad albero



$$p(Breast\ Cancer = \text{Yes} \mid Mammogram = \text{Positive}) = \frac{7}{7+69} = \frac{7}{76} \approx .09$$

Teorema di Bayes in azione

- The setup
  - Two entities: x (data) and θ (parameter)
  - We know the conditional probabilities $p(x|\theta)$, which tell us what to believe about x if we knew the value of θ
  - When we learn the value of x, what should we believe about θ?
- We combine three things (two of which are important)
  - Conditional probabilities for x given θ, the likelihood
  - Previous probabilities about θ
  - The marginal probability for x

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

Conditional Probability or Likelihood

$p(x|\theta)$:

- Conditional probability of x given θ:
  probability of x given the value of θ;
- likelihood for θ:
  Once value of x is known, and this is viewed as a function of θ it is a likelihood

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

Diagnosi medica
Conditional Probability Distribution

- If a woman has breast cancer, the probability of a positive mammogram is .90

  $p(Mam = +|BC = Y) = .90\,, p(Mam = -|BC = Y) = .10$

- If a woman does not have breast cancer, the probability of a positive mammogram is .07

  $p(Mam = +|BC = N) = .07\,, p(Mam = -|BC = N) = .93$

	Mammogram Result	
Breast Cancer	Positive	Negative
Yes	.90	.10
No	.07	.93

Prior distribution

$p(\theta)$:

- Prior probability distribution for unknown θ:
  Everything we know about θ before we observe value for x

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

Diagnosi medica
Prior Distribution for breast cancer

- The proportion of women aged 40-50 with no family history of cancer and no symptoms that have breast cancer is .008
- Assign a prior distribution

  p(Breast Cancer = Yes) = .008
  p(Breast Cancer = No) = .992

Breast Cancer	
Yes	.008
No	.992

Marginal distribution

$p(x)$:

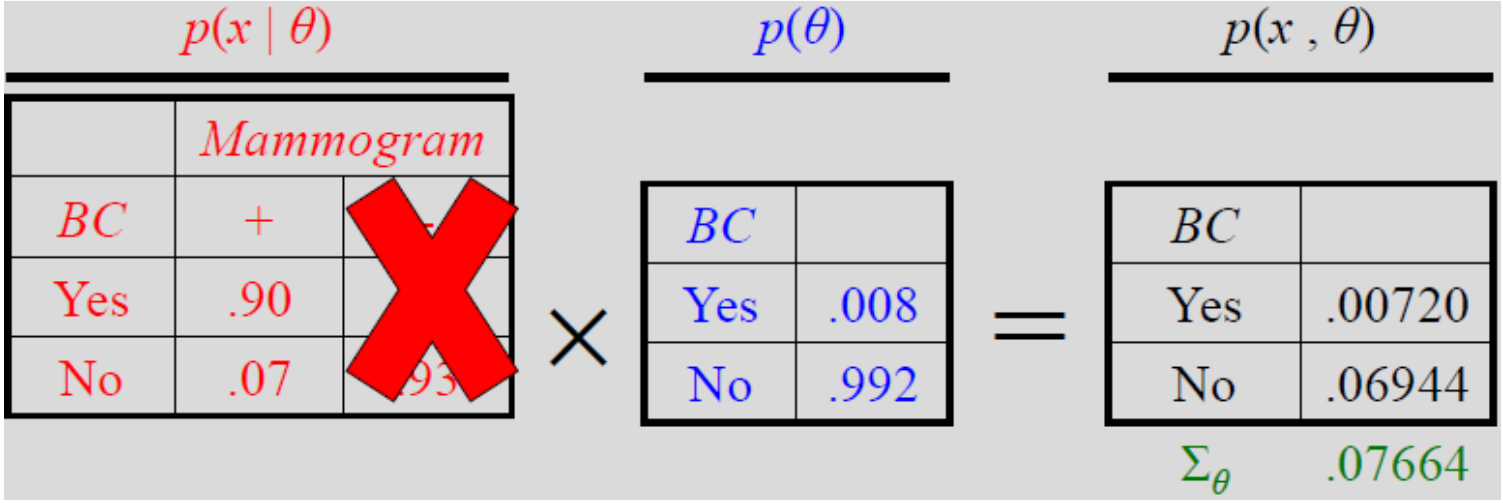
- Marginal probability of x (over θ): $p(x) = \sum_\theta p(x|\theta)p(\theta)$
  - Serves to normalize the distribution
  - Note that $p(x)$ does not vary with θ
    We will eventually discard it

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

Diagnosi medica

Marginal Probability of Mammogram = +

$p(x) = \sum_\theta p(x|\theta)p(\theta)$



Posterior distribution

$p(\theta|x)$:

- Posterior probability distribution for unknown θ given x:
  - Captures what we think about θ now that we have incorporated x

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

Diagnosi medica
Posterior distribution

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \rightarrow \frac{p(Mam = +|BC)p(BC)}{p(Mam = +)}$$

	Mammogram	
BC	+	-
Yes	.90	
No	.07	

×

BC	
Yes	.008
No	.992

=

BC	
Yes	.00720
No	.06944

.07664 .07664

The ***posterior probability*** that the woman has breast cancer

=

BC	
Yes	.09
No	.91

Bayesian Inference as Updating from Probability-Based Reasoning

"Prior to the mammogram, we didn't think the woman had breast cancer.
After the positive result on the mammogram ... we don't think the woman had breast cancer.
So it seems the mammogram is irrelevant, or Bayes doesn't work"

"We are uncertain if the woman has breast cancer. And our language for uncertainty is probabilities.
Before the mammogram, the probability of breast cancer is $p(BC = Y) = .008$.
After the mammogram, the probability of breast cancer is $(BC = Y|M = +) = .09$.
These probabilities are expressions of our beliefs. We still think it's unlikely the patient has breast cancer, but our beliefs have shifted considerably from what they were."

Proportionality in the Posterior Distribution

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \propto p(x|\theta)p(\theta)$$

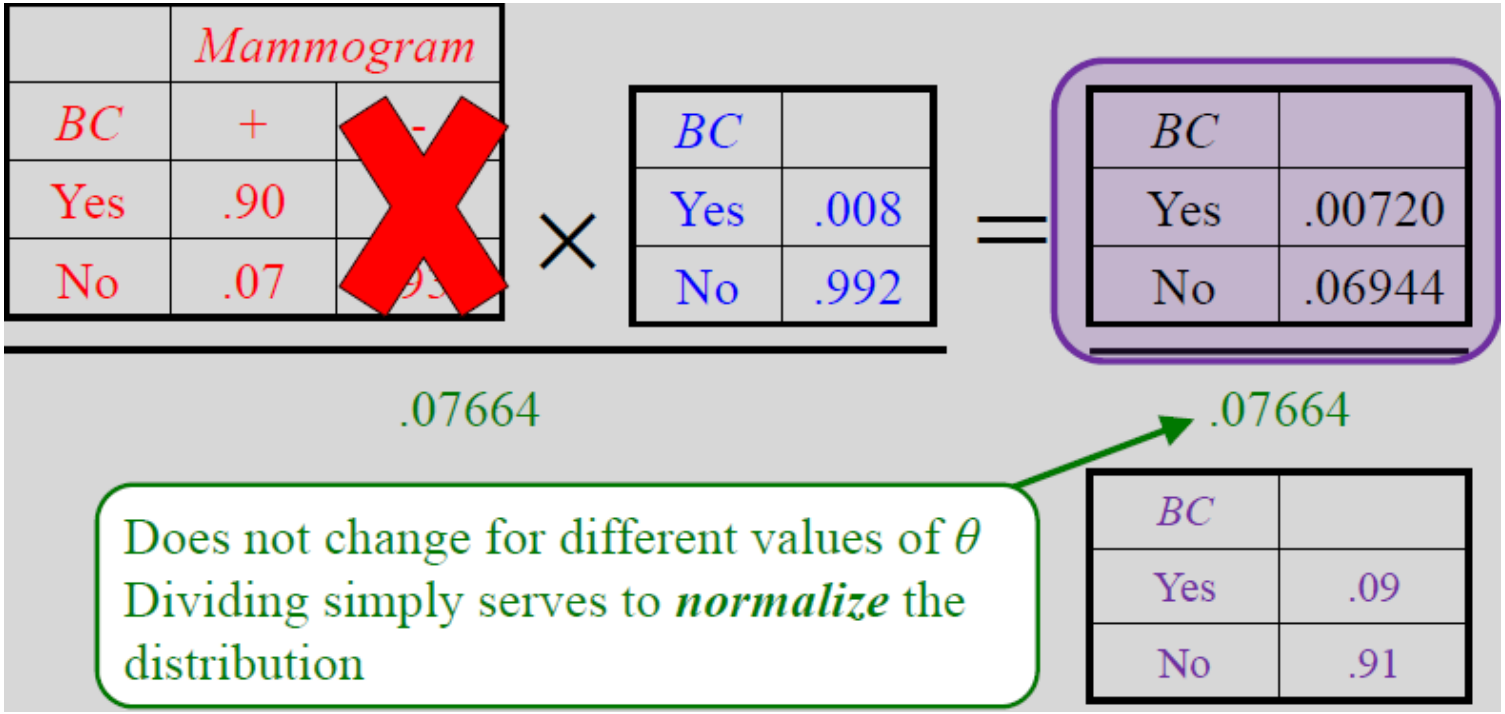
Marginal probability of x (over θ), $p(x)$:

does not change for different values of θ

discarding yields proportionality

Diagnosi medica
Proportionality of Posterior Distribution

$$p(BC|Mam=+) = \frac{p(Mam=+|BC)p(BC)}{p(Mam=+)} \propto p(Mam=+|BC)p(BC)$$



Approccio generale alla modellazione Bayesiana

Approccio generale alla modellazione Bayesiana

Un'analisi ***Bayesianly justifiable*** è una che

> "treats known values as observed values of random variables, treats unknown values as unobserved random variables, and calculates the conditional distribution of unknowns given knowns and model specifications using Bayes'theorem."
>
> -- *Rubin (1984, p. 1152)*

3 Step General Approach to Bayesian Modeling

1. Set up the full probability model: the joint distribution of all entities, including observables (x) and unobservables (θ) in accordance with all that is known about the problem

2. Condition on the observed data (x), calculate the conditional probability distribution for the unobservable entities (θ) of interest given the observed data: the posterior distribution

3. Examine fit, tenability/sensitivity of assumptions, reasonable conclusions?, respecify, summarize results, etc.

Step 1

1. Set up the full probability model: the joint distribution of all entities, including observables (x) and unobservables (θ) in accordance with all that is known about the problem

$$p(x, \theta)$$

Difficult to do as a multivariate system
Joint probability of breast cancer and mammogram results

	Mammogram	
BC	+	-
Yes	0.0072	0.0008
No	0.06944	0.92256

Step 1

1. Set up the ***full probability model*** via a conditional distribution $p(x|\theta)$, and a prior $p(\theta)$

$$p(x, \theta) = p(x|\theta)p(\theta)$$

   - The prior probability of cancer
   - The conditional probability of the mammogram result, given cancer status

BC	
Yes	.008
No	.992

	Mammogram	
BC	+	-
Yes	.90	.10
No	.07	.93

Step 2

1. Condition on the observed data (x), calculate the conditional probability distribution for the unobservable entities (θ) of interest given the observed data: obtain the ***posterior distribution*** as

$$\begin{aligned} p(\theta|x) &= \frac{p(x,\theta)}{p(x)} \\ &= \frac{p(x|\theta)p(\theta)}{p(x)} \\ &\propto p(x|\theta)p(\theta) \end{aligned}$$

Bayes' Theorem for Discrete and Continuous Variables

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} == \frac{p(x|\theta)p(\theta)}{\sum p(x|\theta)p(\theta)} \propto p(x|\theta)p(\theta)$$

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int_\theta p(x|\theta)p(\theta)} \propto p(x|\theta)p(\theta)$$

Bayes' Theorem

Effects a ***reversal*** of the conditional probability, $p(x|\theta) \Rightarrow p(\theta)|x)$

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

From $p(Mam|BC)$ to $p(BC|Mam)$

Confusion of these conditional probabilities is quite common

Confronto tra inferenza frequentista & inferenza bayesiana (sin qui)

Inferenza frequentista vs. inferenza bayesiana

Characteristic	Frequentist	Bayesian
Status of Data	Random	Random until obs.
Contribution of Data	Likelihood	Likelihood
Status of Parameters		
Prior		
Solution		

- Data viewed as random, conditionally distributed given parameters
  - Mammogram is conditionally distributed given BC

Inferenza frequentista vs. inferenza bayesiana

Characteristic	Frequentist	Bayesian
Status of Data	Random	Random until obs.
Contribution of Data	Likelihood	Likelihood
Status of Parameters	Fixed (mostly)	
Prior		
Solution		

- Frequentist inference: parameters are ***fixed*** and unknown
  - Frequentist perspective defines probabilities as long-run freq.
  - We don't know if the woman has breast cancer, but she either does or she doesn't, hardly something that invokes long-run frequencies

Inferenza frequentista vs. inferenza bayesiana

Characteristic	Frequentist	Bayesian
Status of Data	Random	Random until obs.
Contribution of Data	Likelihood	Likelihood
Status of Parameters	Fixed (mostly)	
Prior	No	
Solution		

- Random entities are assigned distributions
- Parameters aren't random $\Rightarrow$ not assigned distributions

Inferenza frequentista vs. inferenza bayesiana

Characteristic	Frequentist	Bayesian
Status of Data	Random	Random until obs.
Contribution of Data	Likelihood	Likelihood
Status of Parameters	Fixed (mostly)	
Prior	No	
Solution	Point (e.g., ML)	

- ML estimation attempts to find the value of the parameters that make the data seem most likely

Diagnosi medica
Maximum Likelihood

- A woman undergoes the mammogram, the result is positive

$$L(BC = Y|Mam = +) = p(Mam = +|BC = Y) = .90$$

$$L(BC = N|Mam = +) = p(Mam = +|BC = N) = .07$$

	Mammogram Result	
Breast Cancer	Positive	Negative
Yes	.90	
No	.07	.93

- ML: *What is the value of BC that maximizes this?* $\widehat{BC} = Yes$

Inferenza frequentista vs. inferenza bayesiana

Characteristic	Frequentist	Bayesian
Status of Data	Random	Random until obs.
Contribution of Data	Likelihood	Likelihood
Status of Parameters	Fixed (mostly)	Random (mostly)
Prior	No	(even if not philosophically)
Solution	Point (e.g., ML)	

- In a Bayesian analysis all unknown entities treated as random and assigned a distribution
  - Initially the data, but once observed, they are conditioned on
  - The parameters as well

Inferenza frequentista vs. inferenza bayesiana

Characteristic	Frequentist	Bayesian
Status of Data	Random	Random until obs.
Contribution of Data	Likelihood	Likelihood
Status of Parameters	Fixed (mostly)	Random
Prior	No	Yes
Solution	Point (e.g., ML)	

- All unknown entities are assigned distributions
  - Don’t know if the patient has breast cancer? Assign a distribution!

Inferenza frequentista vs. inferenza bayesiana

Characteristic	Frequentist	Bayesian
Status of Data	Random	Random until obs.
Contribution of Data	Likelihood	Likelihood
Status of Parameters	Fixed (mostly)	Random
Prior	No	Yes
Solution	Point (e.g., ML)	Distribution

- The “solution/answer” in a Bayesian analysis is a ***distribution***
- Not trying to find a point, trying to find a distribution
  - Not trying to find the estimate of the parameter
  - Trying to find the distribution of a parameter
  - Can summarize if desired in the usual ways

ML vs. Bayesian Estimation

- ML seeks to ***maximize*** $L(\theta|x) = p(x|\theta)$
- Bayes seeks to ***obtain*** $p(\theta|x) \propto p(x|\theta)p(\theta)$
- Finding a point vs. finding a distribution
- Other difference is the presence of the prior $p(\theta)$

Ragionare come un Bayesiano ?

Ragioniamo come un Bayesiano?

- Active debate in psychology, neuroscience
- Some examples of natural occurrences in the research base and everyday life ...
- Breast cancer example drawn from work of Gigerenzer; 24 physicians were explicitly asked for the probability that the woman has breast cancer:
  - 8 said < 90%, 8 said = 10%, 8 said between 50% and 80%
  - Most not thinking about the prior, some not thinking of the data
  - Only 2 gave correct reasoning, others seemed to try to do so, but came up with the wrong answer
  - With natural frequencies, majority close to correct
- Research program by Kahneman & Tversky
  - euristics e cognitive biases
  - theory of prospect (1979)
  - basis for behavioral economics
  - Nobel Prize in Economics (2002) to Kahneman

Dovremmo ragionare come un Bayesiano?

- Characterization as “optimal” by psychology, neuroscience
- Should we include prior probability, condition on data, obtain posterior probabilities?
  - Should we include the fact that only the proportion in the population that have breast cancer is .008?
- If so, why?
  - Better descriptions of problems, models, and inference
  - Probability-based reasoning
  - Managing uncertainty
  - Synthesizing information over time
  - Has been shown to be effective in many disciplines
  - The answer to all your questions…
    http://en.wikipedia.org/wiki/Monty_Hall_problem
    https://mathcenter.oxford.emory.edu/site/math117/bayesTheorem/

Pensieri finali & riassunto

General Approach to Bayesian Modeling

A ***Bayesianly justifiable*** analysis is one that

> "treats known values as observed values of random variables, treats unknown values as unobserved random variables, and calculates the conditional distribution of unknowns given knowns and model specifications using Bayes' theorem."
>
> -- *Rubin (1984, p. 1152)*

3 Step General Approach to Bayesian Modeling

1. Set up the full probability model: the joint distribution of all entities, including observables (x) and unobservables (θ) in accordance with all that is known about the problem

$$p(x, \theta) \propto p(x|\theta)p(\theta)$$

Common (shortcut) approach to specifying models for folks new to Bayesian analysis, but familiar with other approaches

A. Set up the model for the conditional probability of the data as you are used to: $p(x|\theta)$
B. List out all the unknown parameter(s) (θ)
C. Specify prior distribution(s) for the parameter(s) $p(\theta)$, reflecting what is believed about the situation

3 Step General Approach to Bayesian Modeling

1. Set up the full probability model: the joint distribution of all entities, including observables (x) and unobservables (θ) in accordance with all that is known about the problem

$$p(x, \theta) \propto p(x|\theta)p(\theta)$$

2. Condition on the observed data (x), calculate the conditional probability distribution for the unobservable entities (θ) of interest given the observed data: the posterior distribution `

$$p(\theta|x) = \frac{p(x, \theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} \propto p(x|\theta)p(\theta)$$

3. Examine fit, tenability/sensitivity of assumptions, reasonable conclusions?, respecify, summarize results, etc.

Una riflessione sulla probabilità

> Probability is not really about numbers; it is about the structure of reasoning
>
> -- *Glenn Shafer, quoted in Pearl, 1988, p. 77*

Riassunto

- Frequentist inference via ML
  - Overview of conceptions
  - Likelihood = conditional probability
- Bayesian inference
  - Prior, likelihood, marginal, posterior
  - Proportionality of posterior
  - Bayes as updater in probability-based reasoning
- Contrasting frequentist and Bayesian inference
- It’s all about reasoning