Statistica inferenziale – Esercizi #4

Dario Comanducci, 2 giugno 2024

Esercizio 1

Conditional probability: approximately 1/125 of all births are fraternal twins and 1/300 of births are identical twins. Elvis Presley had a twin brother (who died at birth). What is the probability that Elvis was an identical twin? (You may approximate the probability of a boy or girl birth as 1/2.)

Svolgimento

Una coppia di gemelli (evento G) può essere omozigote (evento O, gemelli identici dello stesso sesso) oppure eterozigote (evento E, in qual caso possono essere dello stesso sesso o meno); per cui, sapendo che una persona ha un gemello:

$$\mathbb{P}[\mathrm{O}|G] = \frac{\mathbb{P}[G|O]\mathbb{P}[O]}{\mathbb{P}[G]} = \frac{\mathbb{P}[G|O]\mathbb{P}[O]}{\mathbb{P}[O] + \mathbb{P}[E]}$$

in quanto O ed E sono eventi disgiunti ($\mathbb{P}[G] = \mathbb{P}[O \cup E] = \mathbb{P}[O] + \mathbb{P}[E]$). Inoltre

- $\mathbb{P}[O] = 1/300$
- $\mathbb{P}[E] = 1/125$
- $O \subset G \Rightarrow \mathbb{P}[G|O] = 1$

per cui la probabilità di essere omozigote in una coppia di gemelli è

$$\mathbb{P}[\mathrm{O}|G] = \frac{1 \cdot 1/300}{1/300 + 1/125} = \frac{1/300}{(125 + 300)/(300 \cdot 125)} = \frac{125}{425} = \frac{5}{17}$$

Quindi nel caso di una coppia di gemelli, la probabilità di essere identici (evento I) è

$$\mathbb{P}[I] = \mathbb{P}[O|G] = \frac{5}{17}$$

mentre la probabilità di non esserlo vale

$$\mathbb{P}[\neg I] = 1 - \mathbb{P}[I] = \mathbb{P}[E|G] = \frac{12}{17}$$

Consideriamo adesso le probabilità di essere maschio o femmina a seconda di gemelli identici o meno; in mancanza di osservazioni in merito possiamo ritenere tali eventi come indipendenti, per cui

- $\mathbb{P}[M|I] = \mathbb{P}[M]$

- $\mathbb{P}[F|I] = 1 - \mathbb{P}[M|I] = 1 - \mathbb{P}[M] = \mathbb{P}[F]$
- $\mathbb{P}[M|\neg I] = \mathbb{P}[M]$
- $\mathbb{P}[F|\neg I] = \mathbb{P}[F]$

Inoltre, $\mathbb{P}[M] = \mathbb{P}[F] = 1/2$. Costruendo quindi l'albero dei possibili casi abbiamo

- nel caso di gemelli identici ($\mathbb{P}[I] = 5/17$):
  - $\mathbb{P}[M|I] = \mathbb{P}[M] = 1/2$
  - $\mathbb{P}[F|I] = \mathbb{P}[F] = 1/2$
- nel caso di gemelli non identici ($\mathbb{P}[\neg I] = 12/17$)
  - $\mathbb{P}[M|\neg I] = \mathbb{P}[M] = 1/2$
  - $\mathbb{P}[F|\neg I] = \mathbb{P}[F] = 1/2$

Ne segue che la probabilità che il fratello gemello (maschio) di Elvis fosse identico a Elvis è

$$\mathbb{P}[I|M] = \frac{\mathbb{P}[M|I]\,\mathbb{P}[I]}{\mathbb{P}[M]} = \mathbb{P}[I] = \mathbb{P}[O|G] = \frac{5}{17} \approx 0.294$$

a causa dell'assunzione di indipendenza tra sesso e gemelli identici/non identici (per cui $\mathbb{P}[M|I] = \mathbb{P}[M]$)

Esercizio 2

Suppose you have a Beta(4, 4) prior distribution on the probability θ that a coin will yield a head. Then the coin is independently spun ten times. Heads appear 3 times. Calculate the posterior density of θ and give the Bayes estimate $\bar{\theta}$.

Svolgimento

Ponendo la probabilità a priori $\pi(\theta) \sim \mathrm{Beta}(\alpha, \beta)$ con $\alpha = \beta = 4$ abbiamo

$$\pi(\theta) \;\propto\; \theta^{\alpha-1}(1-\theta)^{\beta-1} = \theta^3(1-\theta)^3$$

Se inoltre per $n = 10$ lanci otteniamo $s = 3$ successi,

$$L(\theta) = \theta^s(1-\theta)^{n-s} = \theta^3(1-\theta)^7$$

ottenendo infine come densità a posteriori

$$p(\theta|\mathbf{Y}) \;\propto\; \theta^{s+\alpha-1}(1-\theta)^{f+\beta-1} = \theta^6(1-\theta)^{10}$$

dove abbiamo effettuato uno scambio di variabile $f = n - s = 7$, così da indicare con f il numero di fallimenti.

Affinché $p(\theta|\mathbf{Y})$ possa rappresentare una densità di probabilità occorre introdurre una costante di normalizzazione ($p(\theta|\mathbf{Y}) = c\,\theta^6(1-\theta)^{10}$) t.c.

$$\int_0^1 p(\theta|\mathbf{Y})\mathrm{d}\theta = 1$$

ossia (ricordando anche che per $n \in \mathbb{N}_+$ vale $\Gamma(n) = (n-1)!$)

$$c^{-1} = \int_0^1 \theta^6 (1-\theta)^{10} \mathrm{d}\theta = \mathrm{B}(11,7) = \frac{\Gamma(11)\,\Gamma(7)}{\Gamma(11+7)} = \frac{10!\,6!}{17!} = \frac{1}{4\cdot 11\cdot 13\cdot 14\cdot 17}$$
$$= \frac{1}{136136}$$

Infine,

$$\bar{\theta} = \mathbb{E}[\theta|\mathbf{Y}] = \frac{\alpha+s}{\alpha+\beta+n} = \frac{4+3}{4+4+10} = \frac{7}{18} = 0.3\bar{8}$$

Esercizio 3

A popuation of measures is normal with mean θ and standard deviation $\sigma = 40$. Reconsider the example of the two physicist A and B in the notes, where A has a normal prior for θ with mean 900 and standard deviation 20, while B has a normal prior with mean 800 and standard deviation 80.

1. *Assume the the experiment concerns a sample of size $n = 1$ and that the single observation is 850. Find the posteriors for A and B.*
2. *Then, suppose the the sample has size $n = 15$ with a sample mean $\bar{y} = 850$. Find the posteriors for A and B.*
3. *Finally, suppose that $n = 100$ and $\bar{y} = 850$. Compare the posteriors of A and B and the likelihood*

Svolgimento

Fissata la prior $\theta \sim N(\theta_0, \sigma_0^2)$ per $Y \sim N(\theta, \sigma^2 = 40^2)$ dove

- per A, $\theta_0 = 900$ e $\sigma_0^2 = 20^2$
- per B, $\theta_0 = 800$ e $\sigma_0^2 = 80^2$

abbiamo che la probabilità a posteriori per θ obbedisce ad una Normale $N(\bar{\theta}, \bar{\sigma}^2)$ essendo

$$\begin{aligned}
\bar{\theta} &= \frac{w_0\theta_0 + w_n\bar{y}}{w_0 + w_n} \\
\bar{\sigma}^2 &= \frac{1}{w_0 + w_n} \\
w_0 &= 1/\sigma_0^2 \\
w_n &= n/\sigma^2
\end{aligned}$$

La stima per la normale a posteriori, a confronto con i parametri per la Likelihood, è riportata in Tab. 1 secondo il codice nel Listato 1.

Listato 1: Codice R per stima della normale a posteriori

```
rm(list=ls())
```

Tab. 1: Stima per la normale a posteriori (secondo le prior dei due fisici A e B), a confronto con i parametri della likelihood, per la misura Y con media θ e varianza $\sigma^2 = 1600$ date n osservazioni.

	A		B		Likelihood	
	$\theta_0 = 900$, $\sigma_0 = 20$		$\theta_0 = 800$, $\sigma_0 = 40$			
n	$\bar{\theta}$	$\bar{\sigma}^2$	$\bar{\theta}$	$\bar{\sigma}^2$	$\theta_L = \bar{y}$	$\sigma_L^2 = \sigma^2/n$
1	890	320	840	1280	850	1600
15	860.526	84.211	849.180	104.918	850	$106.\bar{6}$
100	851.923	15.385	849.875	15.960	850	16

```
PosteriorParams <- function(theta0, sigma0, n, ybar, sigma)
{
   w0 = 1/sigma0^2
   wn = n/sigma^2
   ws = w0 + wn
   thetaPost = (w0*theta0 + wn*ybar)/ws
   sigma2Post = 1/ws
   return (c(thetaPost, sigma2Post))
}

ybar = 850
sigma = 40

theta0A = 900
sigma0A = 20
(postParsA_1 = PosteriorParams(theta0A, sigma0A, 1, ybar, sigma))
(postParsA_15 = PosteriorParams(theta0A, sigma0A, 15, ybar, sigma))
(postParsA_100 = PosteriorParams(theta0A, sigma0A, 100, ybar, sigma))

theta0B = 800
sigma0B = 80
(postParsB_1 = PosteriorParams(theta0B, sigma0B, 1, ybar, sigma))
(postParsB_15 = PosteriorParams(theta0B, sigma0B, 15, ybar, sigma))
(postParsB_100 = PosteriorParams(theta0B, sigma0B, 100, ybar, sigma))

sigma2Lik1 = sigma^2
sigma2Lik15 = (sigma^2)/15
sigma2Lik100 = (sigma^2)/100
```