Lezione n.6: Procedure inferenziali bayesiane

Roma, 20 marzo 2003

Brunero Liseo
Dipartimento di studi geoeconomici, linguistici, statistici e storici
per l'analisi regionale
Università di Roma "La Sapienza"
Rome, Italy
brunero.liseo@uniroma1.it
tel. 06-49766110

Procedure inferenziali classiche

- stima puntuale
- stima per intervallo
- verifica d'ipotesi

Sebbene la trattazione più elegante di questi aspetti preveda l'introduzione della teoria delle decisioni, noi ne faremo a meno, scegliendo le procedure in modo intuitivo.

Stima Puntuale

Abbiamo già visto che il complesso informativo di un'analisi bayesiana è racchiuso dalla distribuzione finale

$$h(\theta \mid \mathbf{x}) \propto h(\theta)L(\theta; \mathbf{x}).$$

La scelta di un indicatore puntuale di questa distribuzione comporta scelte individuali: perlopiù utilizzeremo

- moda
- mediana
- media (aritmetica, ma anche geometrica o armonica...)

Esempi

Abbiamo già visto diversi casi

1. Dati gaussiani con media μ incognita
2. Dati dicotomici
3. Dati esponenziali

Nel primo caso, partendo da una a priori $\mu \sim N(\mu_0, \tau^2)$, arrivavamo a

$$\mu \mid \mathbf{x} \sim N(\mu; \mu^*, \tau^{2*})$$

dove

$$\mu^* = \frac{\frac{\mu_0}{\tau^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} = \frac{\mu_0\sigma^2 + \bar{x}n\tau^2}{\sigma^2 + n\tau^2}, \quad \tau^{*2} = \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}. \quad (1)$$

In questo caso media, moda e mediana coincidono.

Dati Uniformi

$X_1, X_2, \cdots, X_n$ n v.a. i.i.d. $| \, \theta$ con distribuzione $X_j \sim U(0, \theta)$.
La verosimiglianza associata alle n osservazioni è

$$L(\theta; \mathbf{x}) \propto \frac{1}{\theta^n} I_{(x_{(n)}, +\infty)}(\theta),$$

dove $x_{(n)} = \max\{x_1, x_2, \cdots, x_n\}$.

Un'inferenza basata su $L(\theta)$ conduce a $\hat{\theta} = x_{(n)}$, la cui distribuzione, in virtù di risultati generali relativi alle statistiche d'ordine è data da

$$f_{\hat{\theta}}(t \; \theta) = \frac{n x^{(n-1)}}{\theta^n} I_{(0,\theta)}(t),$$

ovvero $Y = \hat{\theta}/\theta \sim Beta(n, 1)$. Dalle proprietà della Beta si ricava allora che

$$E(\hat{\theta}; \theta) = \frac{n\theta}{n+1},$$

ovvero $\hat{\theta}$ è uno stimatore solo asintoticamente corretto.

Basta però porre $T_{ND}(\mathbf{x}) = \frac{(n+1)\hat{\theta}}{n}$

Un'analisi bayesiana convenzionale del problema parte dalla scelta di una famiglia di distribuzioni a priori.

$$\pi(\theta) \propto \frac{1}{\theta^{\alpha}}, \qquad \alpha > 0.$$

La distribuzione a posteriori è allora

$$\pi(\theta \mid \mathbf{x}) \propto \frac{1}{\theta^{n+\alpha}} I_{(x_n, +\infty}(\theta),$$

Essa è integrabile quando $\alpha + n > 1$.
La costante di normalizzazione vale

$$\int_{x_{(n)}}^{\infty} \theta^{-(\alpha+n)} d\theta = \frac{1}{x_{(n)}^{\alpha+n-1}(n+\alpha-1)}.$$

La stima di Bayes per θ, ovvero la media a posteriori per un valore di α fissato, risulta perciò

$$E_\alpha(\theta \mid \mathbf{x}) = x_{(n)}^{n+\alpha-1}(n+\alpha-1)\int_{x_{(n)}}^{\infty} \theta^{-(n+\alpha-1} d\theta = \frac{n+\alpha-1}{n+\alpha-2} x_{(n)}.$$

Si vede facilmente che lo SMV si ottiene per $\alpha \to \infty$ mentre i casi $\alpha = 2$ e $\alpha = 3$ corrispondono, rispettivamente allo stimatore non distorto T_{ND} e allo stimatore che minimizza l'errore quadratico medio T_{c^*}.

Stima per intervallo

L'operazione di stima di un parametro incognito mediante un singolo valore $\tilde{\theta}$ è destinata a fornire un risultato sbagliato con pratica certezza. Meglio allora sintetizzare il risultato attraverso la produzione di un insieme di valori $C \subset \Theta$, che, con una certa probabilità a posteriori, contiene il vero valore di θ. Partendo dalla distribuzione a posteriori $\pi(\theta \mid \mathbf{x})$, il modo più ragionevole di costruire un insieme di valori di θ di livello $1-\alpha$, ovvero che contenga il valore vero di θ con probabilità non inferiore a $1-\alpha$, è quello di determinare un insieme *HPD*, acronimo inglese che sta per *Highest Posterior Density*; si tratta in pratica di inserire nell'insieme C tutti i valori di θ la cui densità a posteriori risulta più elevata, fino a raggiungere una probabilità complessiva non inferiore al livello prescelto $1-\alpha$.

Il caso normale-normale

Qui si ottiene una distribuzione finale di tipo $N(\mu^*, \tau^{2*})$, e quindi unimodale e simmetrica. Qualunque sia il livello di credibilità $1-\alpha$, l'insieme HPD risulterà un intervallo simmetrico intorno alla media finale μ^*:

$$(\mu^* - h, \mu^* + h),$$

dove h è determinato dal vincolo

$$\int_{\mu^*-h}^{\mu^*+h} h(\theta \mid \mathbf{x}) = 1 - \alpha$$

Per simmetria, si trova che $h = \tau^* z_{1-\frac{\alpha}{2}}$.

Nel caso si utilizzi una distribuzione iniziale non informativa ($h^J(\mu) \propto 1$), risulta $\mu^* = \bar{x}$ e $\tau^{2*} = \sigma_0^2/n$ e l'insieme bayesiano coincide numericamente con l'intervallo di confidenza ottenuto attraverso l'approccio frequentista.

La determinazione dell'insieme HPD non è sempre possibile per via analitica.
Può accadere ad esempio che la distribuzione finale risulti multimodale.
Se la distribuzione finale è monotona decrescente, l'insieme HPD di livello $1-\alpha$ si ottiene determinando il quantile di ordine $1-\alpha/2$ della distribuzione finale.
Nel caso più generale in cui la distribuzione finale $\pi(\theta \mid \mathbf{x})$ è unimodale, occorre ricorrere a metodi numerici. In grandi linee la procedura consiste in tre passi

1. Per k fissato ($k < \max(\pi(\theta \mid \mathbf{x}))$, determina le due radici $\theta_1(k)$ e $\theta_2(k)$ dell'equazione $\pi(\theta \mid \mathbf{x}) = k$;

2. Calcola $C(k) = \mathrm{P}(\theta \in (\theta_1(k), \theta_2(k)))$

3. Determina il valore $\tilde{k}$ che risolve l'equazione $C(k) = 1 - \alpha$;

l'intervallo bayesiano HPD sarà allora

$$\left(\theta_1(\tilde{k}), \theta_2(\tilde{k})\right).$$

Esempio Per un lotto di componenti elettroniche occorre determinare il tempo medio di vita. Si può assumere che ciascuna componente abbia una durata di vita esponenziale di parametro θ (e dunque media $1/\theta$). Da informazioni passate possiamo affermare che $\theta \sim Gamma(20, 2)$. Vengono osservate cinque lampadine e il vettore delle osservazioni è $\mathbf{x} = (15, 12, 14, 10, 12)$, per una media campionaria $\bar{x} = 12.8$.

La distribuzione finale per θ è ancora di tipo Gamma$(\tilde{\lambda}, \tilde{\delta})$, dove

$$\tilde{\delta} = 2 + 5 \times 12.8 = 66, \qquad \tilde{\lambda} = 20 + 5 = 7$$

.

Il seguente codice in R fornisce la soluzione al nostro problema:

```
function(k)
{la= 25
de=66
popo_function(t) {dgamma(t,de,rate=la)-k}
t1_uniroot(popo, c(0, 2.64), tol = 0.0001)
t2_uniroot(popo, c(2.64, 10), tol = 0.0001)
a1_t1$root
a2_t2$root
#trova radici t1 e t2 di equazione popo=k) }
valore_integrate(popo, a1,a2)
list("estr. inf"= a1, "estr. sup"= a2, "prob."=valore)
}
```

L'intervallo di credibilità di livello 0.95 per θ così calcolato vale

$$(1.80, 3.60).$$

Va notato che, essendo la distribuzione finale asimmetrica, l'intervallo non risulta centrato né sulla media a posteriori (pari in questo caso a $\tilde{\delta}/\tilde{\lambda} = 2.64$) né sulla moda.

Per questo esempio specifico è possibile una soluzione alternativa che non conduce ad un intervallo HPD ma ad una sua buona approssimazione: è noto che, se una v.a. $Z \sim Gamma(c, d)$, allora la sua trasformazione $Y = 2cX$ ha distribuzione χ^2_{2d}: nel nostro caso questo implica che $Y = 50\,\theta \sim \chi^2_{132}$. E' sufficiente allora determinare i quantili di ordine 0.025 e 0.975 (rispettivamente 102.08 e 165.69) di tale distribuzione per ottenere un intervallo di credibilità di livello 0.95 costruito eliminando le code della distribuzione finale. Ne segue che $\mathrm{P}(102.08 \leq 50\,\theta \leq 165.69) = 0.95$, cosicché

$$(2.042, 3.313)$$

è un intervallo di credibilità (non HPD!) di livello 0.95 per θ.

La verifica di ipotesi

Nel contesto classico, si confrontano due alternative:

$$H_0: \ \theta \in \Theta_0, \quad vs. \quad H_1: \ \theta \in \Theta_1$$

dove $\Theta_0 \cup \Theta_1 = \Theta$ e, ovviamente,

$$\Theta_0 \cap \Theta_1 = \emptyset$$

.

Secondo l'impostazione classica, i dati osservati vengono utilizzati per verificare la loro *compatibilità* con l'ipotesi nulla attraverso il calcolo del valore-*p*, ovvero la probabilità di, essendo vera l'ipotesi nulla, osservare un campione che fornisca un risultato "più lontano" dall'ipotesi nulla rispetto a quello osservato.

In ambito bayesiano, il modo più naturale di quantificare, a posteriori, il peso dell'ipotesi nulla H_0 è quello di calcolarne la probabilità

$$P(H_0 \mid \mathbf{x})$$

definita da

$$P(H_0 \mid \mathbf{x}) = P(\Theta_0 \mid \mathbf{x}) = \int_{\Theta_0} \pi(\theta \mid \mathbf{x}) d\theta \tag{2}$$

La formula (**??**) assume che la variabile aleatoria Θ sia dotata di densità rispetto alla misura di Lebesgue; ovvie modifiche si applicano nel caso in cui Θ sia una v.a. discreta.

Il caso di due ipotesi semplici

Il ruolo delle informazioni a priori e della verosimiglianza nella verifica di ipotesi bayesiana è chiaramente espresso nel caso artificialmente semplice di due ipotesi puntuali

$$H_0 : \theta = \theta_0, \quad vs. \quad H_1 : \theta = \theta_1$$

.

In questo caso semplici calcoli mettono in luce la natura della (**??**); siano infatti

$$\pi_0 = P(H_0) \qquad \text{e} \qquad \pi_1 = P(H_1) = 1 - \pi_0;$$

Il peso relativo a posteriori delle due ipotesi è dato dal rapporto

$$\frac{P(H_0 \mid \mathbf{x})}{P(H_1 \mid \mathbf{x})} = \frac{\pi(\theta_0 \mid \mathbf{x})}{\pi(\theta_1 \mid \mathbf{x})} = \frac{\pi_0}{1 - \pi_0} \frac{L(\theta_0; \mathbf{x})}{L(\theta_1; \mathbf{x})},$$

Tale rapporto è il prodotto di due quantità: la prima, $\pi_0/(1-\pi_0)$, rappresenta il peso relativo delle due ipotesi prima di osservare i dati

la seconda viene in genere denotata con

$$B = \frac{L(\theta_0; \mathbf{x})}{L(\theta_1; \mathbf{x})}$$

e si chiama fattore di Bayes. Esso rappresenta il fattore moltiplicativo che trasforma il rapporto di probabilità a priori in quello a posteriori: in questo senso, B è una misura dell'evidenza sperimentale a favore dell'ipotesi nulla.

- $B > 1 \quad (< 1)$ indica che i dati favoriscono $H_0 \quad (H_1)$

Esempio

Nel 1919, durante un'eclisse solare, l'astronomo Eddington effettuò il seguente esperimento: da due posizioni diverse egli misurò il grado di piegatura della luce emessa in funzione della posizione intorno al sole: effettuò $n_A = 5$ misurazioni dal posto A e $n_B = 7$ misurazioni dal posto B. La teoria di Newton, denotata qui con H_0, prevedrebbe una deflezione di luce di circa 0.875 secondi di arco. Al contrario, la teoria della relatività generale di Einstein, diciamo H_1, conduce qui ad una previsione di 1.75 secondi di arco.

Le osservazioni riportarono un valore medio pari a $\bar{x}_A = 1.98$ secondi, con un errore standard pari a

$$s_A = \sqrt{\frac{\sum_i (x_{iA} - \bar{x}_A)^2}{n_A(n_A - 1)}} = 0.16$$

e $\bar{x}_B = 1.61$ con errore standard pari a $s_B = 0.40$. Possiamo supporre che in entrambi i siti, i dati possano essere considerati avere distribuzione normale con media μ e varianza σ incognite. Sia inoltre, a priori $\pi_0 = \pi_1 = \frac{1}{2}$, ovvero si dà la stessa probabilità alle due teorie.

Consideriamo il primo esperimento, in cui $n_A = 5$; per semplicità assumiamo che s_A possa essere considerato una buona stima puntuale di $\sigma/\sqrt{n_A}$. Allora, la media campionaria $X_A \sim N(\mu, 0.16^2)$ con $\mu = 0.875$ secondo H_0 e $\mu = 1.75$ secondo H_1.

Il fattore di Bayes vale

$$B_A = \frac{\frac{1}{0.16\sqrt{2\pi}} \exp\left\{-(1.98 - 0.875)^2/(2 * 0.16^2)\right\}}{\frac{1}{0.16\sqrt{2\pi}} \exp\left\{-(1.98 - 1.75)^2/(2 * 0.16^2)\right\}} = 1.23 * 10^{-10}$$

In pratica, sono sufficienti 5 osservazioni a fornire evidenza inconfutabile (odds pari a 1 su un milione) in favore di H_1; è però il caso di chiarire bene qual è la conclusione che B suggerisce: il fattore di Bayes non dice che H_1 è vera, ma solo che l'evidenza sperimentale in suo favore, comparativamente con H_0, è un milione di volte superiore...

Consideriamo adesso i dati relativi al secondo punto di osservazione. Calcoli analoghi conducono a

$$B_A = \frac{\frac{1}{0.40\sqrt{2\pi}} \exp\left\{-(1.61-0.875)^2/(2*0.40^2)\right\}}{\frac{1}{0.40\sqrt{2\pi}} \exp\{-(1.61-1.75)^2/(2*0.40^2)\}} = 0.197$$

In questo caso le conclusioni sono molto più incerte e il rapporto di odds è solo di 5 a 1 in favore di H_1.

Ipotesi alternativa composta Consideriamo adesso il caso più generale in cui $H_0 : \theta = \theta_0$ contro un'alternativa composta

$$H_1 : \theta \neq \theta_1$$

; questa situazione corrisponde, da un punto di vista metodologico, al classico test di significatività. Anche in questo caso assumiamo che i dati possano essere considerati realizzazioni di v.a. $X_1, X_2, \cdots, X_n$ con distribuzione normale $N(\theta, \sigma^2)$ con σ noto, per semplicità espositiva.

L'assegnazione delle probabilità a priori richiede qui un minimo di attenzione; poiché H_0 è una ipotesi semplice, la probabilità π_0 risulterà concentrata sul punto θ_0. La probabilità di H_1 deve essere distribuita su tutti i valori di θ diversi da θ_0. Un modo notazionalmente conveniente per descrivere la distribuzione a priori è il seguente

$$\pi_(d\theta) = \pi_0 \delta_{\theta_0}(\theta) + (1 - \pi_0) g(d\theta) I_{(\theta \neq \theta_0}(\theta), \qquad (3)$$

dove $g(\cdot)$ rappresenta la legge di probabilità a priori condizionata ad H_1.

Risulta allora

$$\frac{P(H_0 \mid \mathbf{x})}{P(H_1 \mid \mathbf{x}})=\frac{\pi_0 L(\theta_0;\mathbf{x})}{(1-\pi_0)\int_{\theta\neq\theta_0} L(\theta;\mathbf{x})g(d\theta)}$$

mentre il fattore di Bayes è

$$B=\frac{P(H_0 \mid \mathbf{x})}{P(H_1 \mid \mathbf{x})}/\frac{\pi_0}{1-\pi_0}=\frac{L(\theta_0;\mathbf{x})}{\int_{\theta\neq\theta_0} L(\theta;\mathbf{x})g(d\theta)}$$

Per motivi computazionali, spesso si utilizza una distribuzione a priori $g(\cdot)$ di tipo coniugato, e quindi anch'essa normale $N(\xi,\tau^2)$.

In genere, per simmetria, si pone $\xi=\theta_0$, ma si può generalizzare.

Per il calcolo di B è opportuno notare che, per la sufficienza della funzione di verosimiglianza, sia il numeratore che il denominatore sono funzione dei dati solo attraverso la media campionaria $\bar{x}$, la cui densità, condizionatamente ad un θ generico, denotata con $f(\cdot \mid \theta)$ risulta ancora normale con media θ e varianza σ^2/n. Dunque

$$
\begin{aligned}
B &= \frac{f(\bar{x} \mid \theta_0}{\int_{\theta \neq \theta_0} f(\bar{x} \mid \theta) g(d\theta)} \\
&= \frac{\frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left\{-n(\bar{x}-\theta_0)^2/(2\sigma^2)\right\}}{\int_{\theta \neq \theta_1} \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left\{-n(\bar{x}-\theta_0)^2/(2\sigma^2)\right\} \frac{1}{\tau\sqrt{2\pi}} \exp\left\{-(\theta-\theta_0)^2/(2\tau^2)\right\}} d\theta
\end{aligned}
$$

L'integrale al denominatore può essere risolto in modo analitico; è più agevole però ottenerne il valore utilizzando il seguente risultato.

Lemma 1 *Sia X una v.a. con distribuzione, per θ fissato, $N(\theta, \sigma^2)$; sia inoltre $\theta \sim N(\mu, \tau^2)$. Allora è vero che, marginalmente,*

$$X \sim N(\mu, \sigma^2 + \tau^2).$$

Dimostrazione. E' ovvio che $Y = X - \theta \mid \theta \sim N(0, \sigma^2)$, qualunque sia θ. La combinazione lineare $Y + \theta = X$ è allora normale con media pari a $EY + E\theta = 0 + \mu = \mu$ e varianza pari a

$$\mathrm{Var}\,(Y) + \mathrm{Var}\,(\theta) + 2\,\mathrm{Cov}\,(Y, \theta) = \sigma^2 + \tau^2 + 2\,\mathrm{Cov}\,(Y, \theta)\,.$$

Infine,

$$\mathrm{Cov}\,(Y, \theta) = E(Y\theta) = E\left[E\left(Y\theta \mid \theta\right)\right] = E\left[E\left(X - \theta\right)\theta \mid \theta\right] = 0,$$

da cui la tesi. $\diamond$

Il calcolo di B.

$$B = \frac{\frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left\{-n(\bar{x}-\theta_0)^2/(2\sigma^2)\right\}}{\frac{\sqrt{n}}{\sqrt{\sigma^2+n\tau^2}\sqrt{2\pi}} \exp\left\{-n(\bar{x}-\theta_0)^2/(2(\sigma^2+n\tau^2))\right\}}.$$

Semplici elaborazioni conducono a

$$B = \frac{\sqrt{\sigma^2+n\tau^2}}{\sigma} \exp\left\{-\frac{n}{2\sigma^2}\left[(\bar{x}-\theta_0)^2\left(1-\frac{\sigma^2}{\sigma^2+n\tau^2}\right)\right]\right\}$$

Denotando con

$$u = \frac{\sqrt{n}(\bar{x}-\theta_0)}{\sigma}$$

il valore della usuale statistica test, e ponendo $\rho^2 = \tau^2/\sigma^2$, si ottiene

$$B = \sqrt{1+n\rho^2} \exp\left\{-\frac{u^2}{2}\frac{n\rho^2}{1+(n\rho^2)}\right\} \qquad (4)$$

Alcuni commenti

La formula (4) merita qualche approfondimento.

- per grandi valori di n, il fattore di Bayes conduce a preferire l'ipotesi nulla H_0, indipendentemente dal valore osservato di u.
- lo stesso accade per grandi valori di τ^2.

Questo è il paradosso di Lindley.

La divergenza a $+\infty$ di B per grandi valori di n può essere imputata al fatto che, per dimensioni campionarie elevate, l'adozione di una ipotesi puntuale cessa di rappresentare un'adeguata approssimazione ad una più ragionevole ipotesi intervallare del tipo $H_0 : \theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon)$.

Il comportamento di B, per grandi valori di τ ci dice invece che, nei problemi di verifica di ipotesi non è possibile, se non in circostanze molto speciali, utilizzare distribuzioni a priori improprie; infatti, al crescere di τ^2 la distribuzione a priori di θ "tende" ad una distribuzione uniforme sull'intera retta reale, ovvero ad una distribuzione impropria.

confronti con l'impostazione classica

Lo stesso problema, affrontato da un punto di vista classico, avrebbe comportato il calcolo della statistica u e del suo relativo valore-p a due code. La tabella mostra il calcolo di B e del corrispondente valore-p per specifici valori della statistica u e per diverse numerosità campionarie. I confronti vengono fatti ponendo $\tau^2 = \sigma^2$: intuitivamente questo significa considerare la distribuzione a priori alla stregua di un'ulteriore osservazione campionaria.

Valore di u	valore-p	$n=1$	$n=10$	$n=50$	$n=100$	$n=1000$
1.64	0.10	0.722	.976	1.91	2.65	8.25
1.96	0.05	0.541	.578	1.08	1.50	4.64
2.56	0.01	0.275	.168	.287	.392	1.19
3.29	0.001	0.09	.024	.035	.047	0.14

Tabella 1: Valori di B per diversi valori di n, nel caso $\rho^2 = 1$, in corrispondenza a valori notevoli del valore-p.

Esempio bernoulliano

Questo esempio mostra chiaramente come un'analisi bayesiana di un problema di verifica di ipotesi possa condurre a conclusioni in completo disaccordo con l'analisi classica.

Esperimento di psicocinesi: per verificare se un certo soggetto è dotato di abilità psicocinetiche, viene utilizzato un generatore di eventi casuali basato su principi della meccanica quantistica: l'esperimento consisteva nel verificare se il soggetto era in grado di influenzare il generatore. Semplificando un pò , supponiamo che il generatore spari delle particelle verso una porta "quantistica": qui le particelle possono proseguire lungo la luce rossa oppure la luce verde. Il soggetto tenta di spingere le particelle lungo la luce rossa; nel caso in cui egli non riesca ad influenzare le particelle il meccanismo aleatorio è tale che ogni particella ha probabilità $1/2$ di proseguire lungo ciascuna dei due fasci di luce.

Vennero effettuate un milione di prove ($n = 1000000$) bernoulliane in cui si osserva $X_i =$ 1(se la particella sceglie la luce rossa) oppure 0 (luce verde).

Sia $\theta = P(X_i = 1)$ e sia $Y = \sum X_i$ il numero di particelle che prende la strada *rossa*. Per quanto visto si ha che $Y \sim Bin(n, \theta)$. Si vuole verificare l'ipotesi nulla

$$H_0 : \qquad \theta = \frac{1}{2} \qquad \text{Il soggetto non infuenza il generatore}$$

contro

$$H_1 : \qquad \theta > \frac{1}{2} \qquad \text{Il soggetto infuenza il generatore}$$

Nell'esempio specifico, il valore osservato per Y è pari a 501550. Data l'enorme numerosità campionaria è lecito utilizzare l'approssimazione normale per cui, sotto l'ipotesi nulla H_0

$$Z = \frac{Y - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \sim N(0,1).$$

Il valore di Z osservato è $z_{oss.} = 3.1$ con conseguente valore p unidirezionale pari a $P(Z \geq 3.1) = 0.000967.$

L'analisi classica conduce quindi ad un chiaro rifiuto dell'ipotesi nulla.

Analisi bayesiana

Sia dunque $\pi = P(H_0)$ la probabilità iniziale dell'ipotesi nulla, che per ora non specifichiamo. Dato H_1 $(\theta \geq 1/2)$ sia $g(\theta)$ la distribuzione iniziale per θ. L'analisi non informativa qui sceglierebbe $\pi = 1/2$ e $g(\theta) \propto 1$ (distribuzione uniforme) oppure la distribuzione iniziale di Jeffreys $g(\theta) \propto (\theta(1-\theta))^{-1/2}$. La probabilità a posteriori di H_0 è così

$$P\left(H_0 \mid y\right) = \frac{\pi P(Y = y \mid \theta = \frac{1}{2})}{\pi P(Y = y \mid \theta = \frac{1}{2}) + (1-\pi)\int_{\theta > \frac{1}{2}} P(Y = y \mid \theta) g(\theta) d_\theta}. \tag{5}$$

Ad esempio, utilizzando $\pi = \frac{1}{2}$ e $g(\theta)$ costante si ottiene $P(H_0 \mid y) = 0.87.$

Un'analisi bayesiana completa prevede il calcolo, oltre che di $P(H_0 \mid y)$, dell'insieme HPD condizionato ad H_1, ovvero quel sottoinsieme di $H_1 : \theta > \frac{1}{2}$ più "probabile" nell'ipotesi che H_1 sia vera. Calcoli già visti conducono all'intervallo HPD di livello 0.95

$$C = (0.50015, 0.5029).$$

L'alta probabilità finale di H_0 potrebbe essere imputata al fatto che, inizialmente avevamo posto $\pi = P(H_0) = \frac{1}{2}$. Per indebolire tale conclusione possiamo seguire due strade

- calcolare la probabilità finale al variare della probabilità iniziale π;
- utilizzare il fattore di Bayes;

Nel primo caso, utilizzando la (5) come funzione di π, e tenendo fissata la $g(\theta)$, si ha

$$P\left(H_0 \mid y\right) = \frac{1}{\left(1 + \frac{b(1-\pi)}{a\pi}\right)},$$

dove, nel nostro caso,

$$a = P(Y = y_{oss} \mid \theta = \frac{1}{2}) = 6.53{\times}10^{-6}, \qquad b = \int_{\theta > \frac{1}{2}} P(Y = y \mid \theta) d_{\theta} = \frac{1}{n+1} = 10$$

Si vede facilmente che

$$P\left(H_0 \mid y\right) > 0.5 \Longleftrightarrow \pi > 0.133$$

.

Nel secondo caso si indeboliscono le assunzioni a priori utilizzando semplicemente le quantità a e b appena definite.

In quest'ottica il fattore di Bayes rappresenta una generalizzazione del rapporto di verosimiglianza dove al denominatore viene considerata una sorta di media della funzione di verosimiglianza, ponderata con la distribuzione iniziale $g(\theta)$; nell'esempio specifico si ottiene, ovviamenente $B_{01} = a/b = 6.53$. Un'analisi più accurata includerebbe il controllo della sensibilità delle risposte al variare della distribuzione $g(\theta)$.

Questo esempio è istruttivo per diversi motivi

1. Il valore p è comunque molto diverso dal fattore di Bayes: anche se i due indicatori misurano effettivamente cose differenti, va sottolineato come l'evidenza fornita contro l'ipotesi nulla dai due metodi sia fortemente contrastante, e la tabella mostra come tale contrasto non dipenda dalla distribuzione $g(\theta)$;

2. in casi come questi ha senso considerare un'ipotesi nulla puntuale come $H_0 = \frac{1}{2}$, poiché tale valore ha un preciso significato nel contesto in esame e corrisponde al fatto che il soggetto in esame non sia in grado di modificare il flusso delle particelle. In altre situazioni, dove non esistono valori parametrici, per così dire, *privilegiati*, tale modellizzazione non è giustificabile e dovremmo ricorrere alla formulazione di due ipotesi composte.

3. abbiamo considerato un esempio in cui l'ipotesi alternativa H_1 è unilaterale (o, come spesso si dice, ad una coda): questo è relativamente importante; le stesse conclusioni si raggiungono in esempi in cui l'ipotesi alternativa è bilaterale.