Un problema di stima

Luca La Rocca

15 dicembre 2014

Alla Sagra della Salsiccia un tale vestito da Brighella, circondato da enormi barattoli colmi di fagioli bianchi e rossi, offre ai passanti 155 euro per partecipare al seguente gioco: Brighella indicherà uno dei barattoli e il passante ne estrarrà casualmente 9 fagioli, conterà quanti di questi sono rossi e proverà a indovinare la percentuale di fagioli rossi nel barattolo; quindi il passante pagherà, in euro, il quadrato della differenza tra la sua risposta e la vera percentuale.

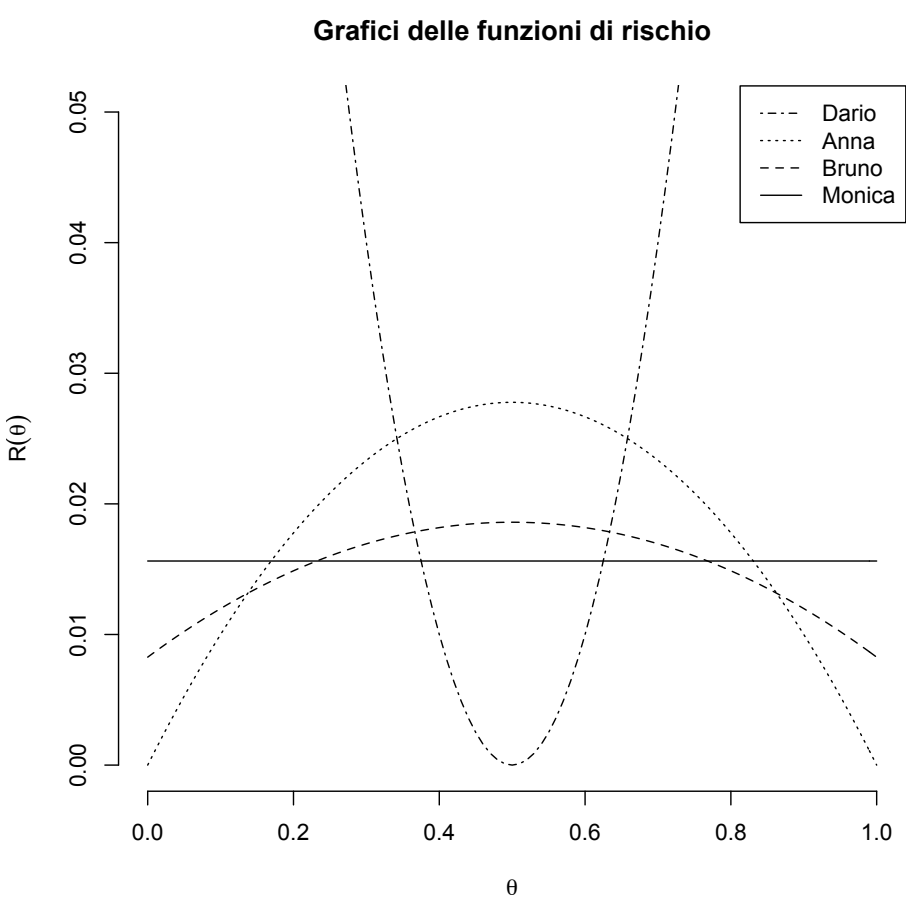
Quattro amici valutano l'eventualità di giocare. Dario dice che può essere sicuro di non pagare più di 2500 euro. Anna si aspetta di fare meglio, ma teme comunque di rimetterci oltre cento euro. Bruno dice che lui, personalmente, si aspetta di guadagnarci qualche euro. Monica osserva che Brighella non è proprio uno stimato concittadino e pertanto è più ragionevole aspettarsi di perdere un euro.

Avendo accettato di giocare, i quattro amici contano 3 fagioli rossi. Un tale vestito da Balanzone compare dal nulla, afferma che bisogna evitare distorsioni e pertanto Anna ha in mente l'unica strategia giusta. Inoltre, dice, possiamo solo confidare che non più di due fagioli su tre siano rossi. Dopo di che sparisce così come era apparso.

Riferimenti

Baldi P. (1998). Calcolo delle Probabilità e Statistica (seconda edizione). McGraw-Hill, Milano.

Steinhaus, H. (1957). The problem of estimation. The Annals of Mathematical Statistics 28(3), 633–648.


Grafici delle funzioni di rischio
Dario
Anna
Bruno
Monica
R(θ)
0.00
0.01
0.02
0.03
0.04
0.05
0.0
0.2
0.4
0.6
0.8
1.0
θ

Confronto di stimatori per una proporzione

Il numero di fagioli estratto è un numero aleatorio con distribuzione approssimativamente binomiale: $Y \sim \text{Bin}(n, \theta)$, dove θ è la proporzione di fagioli rossi nel barattolo ed $n = 9$ il numero totale di fagioli estratti; l'approssimazione è valida nella misura in cui nel barattolo ci sono "molti" fagioli e ne vengono estratti "pochi". In alternativa, due scenari applicativi: controllo di qualità, dove θ è la proporzione di pezzi difettosi; misura dell'efficacia di un trattamento, dove θ è la proporzione di casi in cui il trattamento è efficace. Sorvoliamo su effetto placebo e rarità dei difetti, nel caso in cui questi due aspetti siano rilevanti.

Se conoscessimo θ, potremmo calcolare

$$P_\theta\{Y \in A\} = \sum_{y \in A} f(y|\theta), \quad A \subseteq \mathcal{Y},$$

dove $\mathcal{Y} = \{0, 1, \ldots, n\}$ ed

$$f(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}, \quad y \in \mathcal{Y};$$

tuttavia sappiamo solo che $\theta \in \Theta =]0, 1[$. Ricordiamo che $E_\theta Y = n\theta$ e $V_\theta Y = n\theta(1-\theta)$, come si verifica scrivendo $Y = X_1 + \cdots + X_n$ con $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Ber}(\theta)$ e sfruttando la linearità del valore atteso.

Se osserviamo y, per un certo valore di n, cosa possiamo dire su θ? Questo, in estrema sintesi e generalità, il problema dell'inferenza su una proporzione. Uno *stimatore puntuale* di θ basato su y è definito da una funzione $t : \mathcal{Y} \to \Theta$ che al dato y associ la stima $\hat{\theta}$; lo stimatore definito da $t(\cdot)$ sarà tanto più "buono" quanto più fornirà stime "vicine" al parametro θ. Si può formalizzare la valutazione della "bontà" di una stima puntuale $\hat{\theta}$ mediante una funzione di perdita $L : \Theta \times \Theta \to \Re$ che alla coppia $(\theta, \hat{\theta})$ associ la perdita $L(\theta, \hat{\theta})$; per esempio Brighella suggerisce di usare la *funzione di perdita quadratica*

$$L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2, \quad \theta \in]0, 1[, \quad \hat{\theta} \in]0, 1[,$$

da moltiplicare per 10000 in quanto l'importo da pagare sarà pari al quadrato della differenza tra le percentuali (non tra le proporzioni).

Osserviamo innanzi tutto che non possiamo trovare uno stimatore ottimo uniformemente rispetto a θ: lo stimatore definito da $t_0(y) \equiv \theta_0$ è ottimo per $\theta = \theta_0$, visto che $L(\theta_0, t_0(y)) \equiv 0$, ma per $\theta = \theta_1 \neq \theta_0$ realizzerà una perdita $L(\theta_1, t_0(y)) \equiv (\theta_0 - \theta_1)^2 > 0 \equiv L(\theta_1, t_1(y))$, dove $t_1(y) \equiv \theta_1$ definisce l'analogo stimatore ottimo per $\theta = \theta_1$. Questa osservazione prescinde da come consideriamo la variabilità di y.

Dario è molto cauto e cerca il suo stimatore come

$$t_D(\cdot) = \operatorname*{argmin}_{t(\cdot)} \sup_{\theta \in \Theta, y \in \mathcal{Y}} L(\theta, t(y)),$$

trovando $t_D(y) \equiv 1/2$. Infatti, se per un qualche $y \in \mathcal{Y}$ si ha $t(y) > 1/2$, o $t(y) < 1/2$, allora $\sup_{\theta \in \Theta, y \in \mathcal{Y}} L(\theta, t(y)) \geq L(0, t(y)) = t(y)^2 > 1/4$, o $\sup_{\theta \in \Theta, y \in \mathcal{Y}} L(\theta, t(y)) \geq L(1, t(y)) = (t(y) - 1)^2 > 1/4$, mentre nel caso di $t_D(\cdot)$ abbiamo $\sup_{\theta \in \Theta, y \in \mathcal{Y}} L(\theta, t_D(y)) = L(0, 1/2) = 1/4$. Dario, usando $t_D(\cdot)$, al massimo pagherà $10000/4 = 2500$ euro; non farà però alcun uso dell'informazione che y porta su θ tramite $f(y|\theta)$. Forse, in cuor suo, Dario pensa che Brighella possa scegliere θ sulla base di y.

Anna non vuole rinunciare a usare il dato y e trova naturale definire il suo stimatore per analogia, cioè stimare la frazione di fagioli rossi nel barattolo con l'analoga frazione nella manciata di fagioli estratti:

$$t_A(y) = \frac{y}{n}, \quad y \in \mathcal{Y}.$$

Così facendo Anna può trovarsi a pagare sino a 10000 euro, perché $\sup_{\theta \in \Theta, y \in \mathcal{Y}} L(\theta, t_A(y)) = L(0, 1) = 1$, ma ritiene di non dovere basare la sua decisione su uno scenario tanto sfortunto; ritiene più opportuno tenere conto della variabilità di y mediando la perdita rispetto alla sua distribuzione. Anna valuterà allora $t(\cdot)$ con la sua *funzione di rischio*

$$R_T(\theta) = E_\theta L(\theta, T) = E_\theta (T - \theta)^2, \quad \theta \in \Theta,$$

dove $T = t(Y)$ è lo stimatore puntuale definito da $t(\cdot)$ visto come numero aleatorio prima dell'esperimento che genera il dato y. Troverà

$$R_A(\theta) = E_\theta \left(\frac{Y}{n} - \theta \right)^2 = V_\theta \left(\frac{Y}{n} \right) = \frac{\theta(1-\theta)}{n}, \quad \theta \in]0, 1[,$$

per $T_A = t_A(Y) = Y/n$, dove $R_A(\theta)$ abbrevia $R_{T_A}(\theta)$.

Per $T_D = t_D(Y) \equiv 1/2$ troveremo invece

$$R_D(\theta) = E_\theta \left(\frac{1}{2} - \theta\right)^2 = \left(\theta - \frac{1}{2}\right)^2, \quad \theta \in]0,1[,$$

dove in questo caso il valore atteso è pleonastico perché T_D è deterministico. Per valori di θ in un intorno di $1/2$ lo stimatore T_D è preferibile a T_A, ma il diametro di tale intorno tende a zero, in modo monotono, per $n \to \infty$; solo per $\theta = 1/2$ preferiamo T_D a T_A qualunque sia n. Al di fuori del suddetto intorno, lo stimatore T_A è preferibile a T_D.

Il *rischio massimo* di uno stimatore T del parametro θ è dato da

$$r_M(T) = \sup_{\theta \in \Theta} R_T(\theta);$$

esso esprime, con riferimento alla Sagra della Salsiccia, la perdita attesa nel caso in cui Brighella possa scegliere θ (il barattolo) in base a T (la nostra strategia). Troveremo $r_M(T_D) = 1/4$, per ogni n, mentre $r_M(T_A) = 1/(4n)$ tenderà a zero, in modo monotono, per $n \to \infty$. Per $n = 9$ abbiamo $r_M(T_A) = 1/36 \approx 0.0278$ e quindi Anna rischia di pagare 278 euro, perdendone così $278 - 155 = 123$.

Bruno considera θ uniformemente distribuito su $]0,1[$: $\theta \sim \mathcal{U}(0,1)$. Questo corrisponde a supporre che Brighella scelga "a caso" la composizione del barattolo e permette a Bruno di mediare la funzione di rischio rispetto alla distribuzione di θ, ottenendo $r_B(T) = \int_\Theta R_T(\theta) d\theta$; in generale il *rischio bayesiano* determinato dalla densità iniziale $p(\theta)$, $\theta \in \Theta$, è dato da $r_p(T) = \int_\Theta R_T(\theta) p(\theta) d\theta$ e il caso particolare di Bruno si ottiene prendendo $p(\theta) \equiv 1$.

Possiamo scrivere

$$\begin{aligned} r_p(T) &= \int_\Theta p(\theta) \sum_{y \in \mathcal{Y}} L(\theta, t(y)) f(y|\theta) d\theta \\ &= \sum_{y \in \mathcal{Y}} f(y) \int_\Theta L(\theta, t(y)) p(\theta|y) d\theta, \end{aligned}$$

dove $f(y) = \int_\Theta f(y|\theta) p(\theta) d\theta$ è la probabilità marginale di osservare $\{Y = y\}$, mentre $p(\theta|y) = f(y|\theta)p(\theta)/f(y)$ è la densità finale di θ osservando $\{Y = y\}$.

Nel caso specifico

$$\begin{aligned} r_B(T) &= \int_0^1 \sum_{y=0}^{n} \{t(y)-\theta\}^2 \binom{n}{y} \theta^y (1-\theta)^{n-y} d\theta \\ &= \sum_{y=0}^{n} \int_0^1 \{t(y)-\theta\}^2 \frac{n!}{y!(n-y)!} \theta^y (1-\theta)^{n-y} d\theta \\ &= \sum_{y=0}^{n} \frac{1}{n+1} \int_0^1 \{t(y)-\theta\}^2 \frac{(n+1)!}{y!(n-y)!} \theta^y (1-\theta)^{n-y} d\theta \\ &= \sum_{y=0}^{n} f(y) \int_0^1 \{t(y)-\theta\}^2 p(\theta|y) d\theta, \end{aligned}$$

dove la garanzia di avere correttamente individuato $p(\theta|y)$ ed $f(y) \equiv 1/(n+1)$ deriva dal fatto che $\int_0^1 p(\theta|y)d\theta = 1$; infatti

$$\int_0^1 \theta^y (1-\theta)^{n-y} d\theta = \frac{y!(n-y)!}{(y+n-y+1)!}$$

come avremo modo di verificare nel seguito.

Per definire il suo stimatore in modo da minimizzare il rischio bayesiano,

$$t_B(\cdot) = \operatorname*{argmin}_{t(\cdot)} r_B(t(Y)),$$

Bruno cerca, in corrispondenza del dato y, una stima $\hat{\theta}$ che minimizzi il *rischio a posteriori*

$$r(\tilde{\theta}|y) = \int_0^1 (\tilde{\theta}-\theta)^2 p(\theta|y) d\theta, \quad \tilde{\theta} \in]0,1[.$$

Poiché $r(\tilde{\theta}|y) = \tilde{\theta}^2 - 2\tilde{\theta}\int_0^1 \theta p(\theta|y)d\theta + \int_0^1 \theta^2 p(\theta|y)d\theta$, la stima di Bruno sarà $\hat{\theta} = \int_0^1 \theta p(\theta|y)d\theta$ e il suo *stimatore bayesiano* sarà definito da $t_B(y) = \int_0^1 \theta p(\theta|y)d\theta$, $y \in \mathcal{Y}$.

In concreto, per $y = 0, 1, \ldots, n$, troverà

$$\begin{aligned} t_B(y) &= \int_0^1 \frac{(n+1)!}{y!(n-y)!} \theta^{y+1} (1-\theta)^{n-y} d\theta \\ &= \frac{(n+1)!}{y!(n-y)!} \frac{(y+1)!(n-y)!}{(y+1+n-y+1)!} \\ &= \frac{y+1}{n+2}. \end{aligned}$$

Per comprendere meglio il risultato ottenuto, possiamo scrivere

$$\begin{aligned} t_B(y) &= \frac{n}{n+2}\frac{y}{n} + \frac{2}{n+2}\frac{1}{2} \\ &= \frac{n}{n+2}\frac{y}{n} + \frac{2}{n+2}\int_0^1 \theta d\theta, \\ &= \frac{n}{n+2}t_A(y) + \frac{2}{n+2}t_D(y), \end{aligned}$$

dove si vede che Bruno miscela, in pratica, l'idea di Anna con quella di Dario, usando quest'ultima come informazione iniziale (per definire il valore atteso iniziale di θ).

Bruno, con lo stimatore $T_B = t_B(Y)$, ottiene la funzione di rischio

$$\begin{aligned} R_B(\theta) &= E_\theta\left(\frac{Y+1}{n+2} - \theta\right)^2 \\ &= \frac{1}{(n+2)^2}E_\theta(Y+1-n\theta-2\theta)^2 \\ &= \frac{1}{(n+2)^2}E_\theta\{(Y-n\theta)^2 - 2(Y-n\theta)(1-2\theta) + (1-2\theta)^2\} \\ &= \frac{1}{(n+2)^2}\{n\theta(1-\theta) + (1-2\theta)^2\} \\ &= \frac{1}{(n+2)^2}\{n\theta(1-\theta) + 1 + 4\theta^2 - 4\theta\} \\ &= \frac{(n-4)\theta(1-\theta)+1}{(n+2)^2} \end{aligned}$$

e dunque, essendo $n > 4$, un rischio massimo

$$r_M(T_B) = \frac{(n-4)/4+1}{(n+2)^2} = \frac{n}{4(n+2)^2},$$

pari a 0.0186 per $n = 9$. Bruno rischia quindi di perdere $186-155 = 31$ euro: circa 90 euro meno di Anna. Pur essendo la strategia di Bruno ottimizzata rispetto a una scelta casuale del barattolo, essa si rivela migliore di quella di Anna anche nell'ipotesi che Brighella scelga il barattolo con furbizia (rispetto alla nostra strategia). Si noti che per $n \leq 4$ il rischio massimo varrebbe $1/(n+2)^2$, essendo in questo caso convesso il grafico della funzione di rischio. Nel caso $n = 4$, per dirla tutta, il grafico sarebbe costante, ma su questo torneremo a breve.

Quello che Bruno si aspetta di pagare, dal suo personale punto di vista, è dato dal rischio bayesiano del suo stimatore:

$$\begin{aligned} r_B(T_B) &= \int_0^1 R_B(\theta)d\theta \\ &= \frac{1}{(n+2)^2} + \frac{n-4}{(n+2)^2}\int_0^1 \theta(1-\theta)d\theta \\ &= \frac{1}{(n+2)^2} + \frac{n-4}{(n+2)^2}\left\{\frac{1}{2} - \frac{1}{3}\right\} \\ &= \frac{6+(n-4)}{6(n+2)^2} \\ &= \frac{1}{6(n+2)}. \end{aligned}$$

Per $n = 9$ troviamo $r_B(T_B) = 0.0152$ e quindi Bruno si aspetta di guadagnare dal gioco $155 - 152 = 3$ euro.

Monica osserva che, per $n = 4$, la funzione di rischio dello stimatore bayesiano è identicamente pari a $1/(n+2)^2 = 1/36$; quindi $R_B(\theta) \equiv r_M(T_B)$ ed $r_B(T_B) = \int_0^1 r_M(T_B)d\theta = r_M(T_B)$. Ne segue che, per $n = 4$, lo stimatore definito da $t_B(\cdot)$ minimizza il rischio massimo, oltre a quello bayesiano: se $T = t(Y)$ è un qualsiasi altro stimatore, avremo

$$r_B(T) = \int_0^1 R_T(\theta)d\theta \leq \int_0^1 r_M(T)d\theta = r_M(T),$$

quindi $r_M(T) \geq r_B(T) \geq r_B(T_B) = r_M(T_B)$. In generale, uno stimatore bayesiano che "equalizzi" la funzione di rischio è anche uno *stimatore minimax*: minimizza il rischio massimo ed è lo stimatore "giusto" per giocare contro un soggetto che si adoperi per batterci.

Nel caso $n = 9$ a Monica serve una distribuzione iniziale per θ che dia uno stimatore bayesiano con funzione di rischio "equalizzata". L'espressione di $f(y|\theta)$ le suggerisce di prendere

$$p(\theta) = \frac{1}{B(a,b)}\theta^{a+1}(1-\theta)^{b+1}, \quad \theta \in]0,1[,$$

dove $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, $a > 0$, $b > 0$, è la *funzione beta*, definita in termini della *funzione gamma*

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx, \quad a > 0.$$

Verifichiamo che $B(a,b)$ è la costante di normalizzazione giusta per fare in modo che $p(\theta)$, $\theta \in]0,1[$, sia una densità:

$$\begin{aligned}
\Gamma(a)\Gamma(b) &= \int_0^\infty x^{a-1}e^{-x}dx \int_0^\infty y^{b-1}e^{-y}dy \\
&= \int_0^\infty \int_0^\infty x^{a-1}y^{b-1}e^{-x-y}dxdy \\
&= \int_0^1 \int_0^\infty u^{a-1}(1-u)^{b-1}z^{a+b-2}e^{-z}zdzdu \\
&= \int_0^\infty z^{a+b-1}e^{-z}dz \int_0^1 u^{a-1}(1-u)^{b-1}du \\
&= \Gamma(a+b)\int_0^1 u^{a-1}(1-u)^{b-1}du,
\end{aligned}$$

dove si è effettuato il cambio di variabili $z = x + y$ e $u = x/(x+y)$, vale a dire $x = zu$ e $y = z(1-u)$, per il quale

$$dxdy = \begin{vmatrix} u & z \\ (1-u) & -z \end{vmatrix} dzdu = |-uz - (1-u)z|dzdu = zdzdu.$$

La funzione gamma può vedersi come un'estensione del fattoriale ai numeri reali positivi, nel senso che integrando per parti si trova

$$\Gamma(a+1) = \int_0^\infty x^a e^{-x}dx = -x^a e^{-x}|_0^\infty + a\int_0^\infty x^{a-1}e^{-x}dx = a\Gamma(a),$$

qualunque sia $a > 0$. Per induzione, osservando che $\Gamma(1) = \int_0^\infty e^{-x} = 1$, se ne ricava $\Gamma(n+1) = n!$ per ogni intero positivo n. Si trova allora

$$\begin{aligned}
\int_0^1 \theta^y(1-\theta)^{n-y}d\theta &= B(y+1, n-y+1) \\
&= \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(y+1+n-y+1)} \\
&= \frac{y!(n-y)!}{(y+n-y+1)!},
\end{aligned}$$

per ogni $y = 0, 1, \ldots, n$, come era stato anticipato.

Se Monica assume per θ la distribuzione iniziale sopra suggerita, detta *distribuzione beta* di parametri a e b, in simboli $\theta \sim \text{Beta}(a,b)$, è immediato verificare che la densità finale di θ sarà proporzionale a $\theta^{a+y}(1-\theta)^{b+n-y}$, $\theta \in]0,1[$, ovvero che $\theta|y \sim \text{Beta}(a+y, b+n-y)$; prendendo $a = b = 1$ si ritroverà il caso particolare di Bruno.

Lo stimatore bayesiano con distribuzione iniziale Beta(a, b) sarà definito da

$$\begin{aligned} t_{a,b}(y) &= \int_0^1 \theta p(\theta|y) d\theta \\ &= \frac{1}{B(a+y, b+n-y)} \int_0^1 \theta^{a+y}(1-\theta)^{b+n-y-1} d\theta \\ &= \frac{\Gamma(a+y+b+n-y)}{\Gamma(a+y)\Gamma(b+n-y)} \frac{\Gamma(a+y+1)\Gamma(b+n-y)}{\Gamma(a+y+1+b+n-y)} \\ &= \frac{\Gamma(a+y+1)}{\Gamma(a+y)} \frac{\Gamma(a+b+n)}{\Gamma(a+b+n+1)} \\ &= \frac{a+y}{a+b+n} \end{aligned}$$

e otterrà la funzione di rischio

$$\begin{aligned} R_{a,b}(\theta) &= E_\theta \left(\frac{Y+a}{n+a+b} - \theta \right)^2 \\ &= \frac{1}{(n+a+b)^2} E_\theta \{Y - n\theta + a - (a+b)\theta\}^2 \\ &= \frac{E_\theta (Y-n\theta)^2 + \{a-(a+b)\theta\}^2}{(n+a+b)^2} \\ &= \frac{a^2 + \{n - 2a(a+b)\}\theta + \{(a+b)^2 - n\}\theta^2}{(n+a+b)^2}, \end{aligned}$$

la quale sarà "equalizzata" se $a+b = \sqrt{n}$ e $a = \sqrt{n}/2$, ovvero quando prenderemo $a = b = \sqrt{n}/2$.

Monica troverà allora

$$t_M(\cdot) = \operatorname*{argmin}_{t(\cdot)} r_M(t(Y)),$$

per definire lo stimatore minimax $T_M = t_M(Y)$, prendendo

$$t_M(y) = \frac{y + \sqrt{n}/2}{n + \sqrt{n}}, \quad y = 0, 1, \ldots, n;$$

in questo modo otterrà

$$R_M(\theta) \equiv \frac{n/4}{(n+\sqrt{n})^2} = \frac{1}{4(1+\sqrt{n})^2}$$

che per $n = 9$ darà $r_M(T_M) = 1/\{4(1+\sqrt{9})^2\} = 1/64 = 0.0156$. Monica si aspetta dunque di perdere $156-155 = 1$ euro (uniformemente rispetto alla vera proporzione di fagioli rossi nel barattolo).

Per comprendere l'intervento di Balanzone introduciamo la distorsione di uno stimatore T, in inglese *bias*, definita come

$$B_\theta T = E_\theta T - \theta, \quad \theta \in \Theta.$$

Diremo che T è non distorto quando $B_\theta T \equiv 0$, vale a dire $E_\theta T \equiv \theta$. Questo è un requisito ragionevole per uno stimatore e nel caso in esame conduce direttamente allo stimatore $T_A = t_A(Y) = Y/n$: se vi fosse un altro stimatore non distorto $T = t(Y)$, posto $h(\cdot) = t_A(\cdot) - t(\cdot)$, avremmo $E_\theta h(T) \equiv 0$ e questo non è possibile; infatti

$$E_\theta h(Y) = \sum_{y=0}^{n} \binom{n}{y} \theta^y (1-\theta)^{n-y} h(y), \quad \theta \in]0,1[,$$

è un polinomio in θ di grado n che sarà identicamente nullo solo per $h(y) \equiv 0$. Quando, come in questo caso, $E_\theta h(Y) \equiv 0$ implica $h(y) \equiv 0$, si dice che T è una *statistica completa*.

Abbiamo visto che, in termini di rischio massimo, lo stimatore $T_M = t_M(Y)$ fa meglio di T_A: accettando la distorsione

$$B_\theta T_M = \frac{n\theta + \sqrt{n}/2}{n + \sqrt{n}} - \theta = \frac{\sqrt{n}/2 - \theta\sqrt{n}}{n + \sqrt{n}} = \frac{1/2 - \theta}{1 + \sqrt{n}}, \quad \theta \in]0,1[,$$

tendente a zero, in modo monotono, per $n \to \infty$, si riduce il rischo massimo di 122 euro (da 278 a 156). In generale vi è un *trade-off* tra distorsione e varianza nella minimizzazione del rischio quadratico:

$$R_T(\theta) = E_\theta(T - E_\theta T + E_\theta T - \theta)^2 = V_\theta T + (B_\theta T)^2.$$

Imporre $B_\theta T \equiv 0$, per quanto ragionevole, significa rinunciare a questo trade-off. D'altra parte, una volta posto $B_\theta T \equiv 0$ e osservato che il rischio di T coincide con la sua varianza, in diversi casi si riesce a trovare uno stimatore non distorto a varianza uniformemente minima (rispetto a θ); nel caso specifico vi era addirittura un solo candidato.

Balanzone conclude notando che, inevitabilmente, uno stimatore puntuale fornisce il valore esatto del parametro oggetto di stima con probabilità trascurabile. Il rimedio proposto consiste nel sostituire lo stimatore puntuale con uno *stimatore per intervallo*. Vediamo come si sviluppa, concretamente, questo ragionamento.

Supponiamo che n sia "grande" abbastanza da potere invocare il Teorema Limite Centrale per approssimare la distribuzione di $Z = (T_A - \theta)/D_\theta(T_A)$ con la distribuzione normale standard, dove $T_A = n^{-1}Y = n^{-1}\sum_{i=1}^{n} X_i$ è lo stimatore non distorto e

$$D_\theta(T_A) = \sqrt{V_\theta t_A(Y)} = \sqrt{\frac{\theta(1-\theta)}{n}}$$

è la sua deviazione standard (anche detta *errore standard*). Nel caso specifico $n = 9$ è troppo "piccolo" perché questa approssimazione sia valida, ma nelle applicazioni essa permette di trattare un gran numero di casi; per esempio il caso di 1000 intervistati cui si chieda di esprimere un giudizio favorevole o contrario sull'operato del governo in carica.

Nel limite in cui vale l'approssimazione normale, troviamo che T_A fornisce il valore esatto di θ con probabilità $P_\theta\{T_A = \theta\} = 0$. Per fare meglio, in questo senso, possiamo sostituire T_A con una coppia (L, U) di statistiche, $L = \ell(Y)$, $U = u(Y)$, che soddisfi la proprietà

$$P_\theta\{L \leq \theta \leq U\} = 1 - \alpha,$$

dove $\alpha \in]0, 1[$ è un valore "piccolo"; tipicamente $\alpha = 0.05$. Otteniamo così un intervallo aleatorio $[L, U]$ che contiene θ con "buona" probabilità; lo chiameremo *intervallo di confidenza* per θ al livello $1 - \alpha$. Tipicamente avremo $1 - \alpha = 0.95$ e parleremo di intervallo al 95%.

Per costruire (L, U) cerchiamo innanzi tutto z_α con la proprietà

$$P_\theta\{-z_\alpha \leq Z \leq z_\alpha\} = 1 - \alpha,$$

dove Z segue la distribuzione normale standard; per $\alpha = 0.05$ troviamo $z_\alpha \approx 2$. Successivamente scriviamo l'espressione di Z nell'equazione sopra esposta e manipoliamone le disuguaglianze, ottenendo

$$P_\theta\{T_A - z_\alpha D_\theta(T_A) \leq \theta \leq T_A + z_\alpha D_\theta(T_A)\} = 1 - \alpha.$$

Questo non è un intervallo di confidenza, perché $D_\theta(T_A)$, a differenza di T_A, non è una statistica; infatti $D_\theta(T_A)$ dipende dal parametro θ. Possiamo però maggiorare $D_\theta(T_A)$ con $1/\sqrt{4n}$ (errore standard massimo) e ottenere un intervallo di confidenza di livello *almeno* pari a $1-\alpha$ ponendo $L = T_A - z_\alpha/(2\sqrt{n})$ e $U = T_A + z_\alpha/(2\sqrt{n})$.

Nel caso tipico $\alpha = 0.05$, esplicitando l'espressione di $T_A = t_A(Y)$, troviamo

$$P_\theta \left\{ \frac{Y}{n} - \frac{1}{\sqrt{n}} \leq \theta \leq \frac{Y}{n} + \frac{1}{\sqrt{n}} \right\} \geq 0.95,$$

di modo che il *margine di errore* sulla proporzione osservata è $1/\sqrt{n}$. Per un un sondaggio dicotomico con 1000 intervistati il margine d'errore sarà pari a $1/\sqrt{1000} = 0.032 = 3.2\%$. Per il gioco di Brighella, essendo $n = 9$, il margine d'errore varrà $1/3$.

Adesso, per finire, supponiamo di avere estratto $y = 3$ fagioli rossi. Viene naturale scrivere

$$0 = \frac{1}{3} - \frac{1}{3} \leq \theta \leq \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

al 95%, come suggerito da Balanzone, confidando di non avere estratto proprio uno di quei campioni (meno del 5%) nei quali $\theta \notin [L, U]$.

Notiamo però che per l'*intervallo di confidenza stimato* $[0, 2/3]$ non vale alcuna affermazione probabilistica, non essendovi alcuna aleatorietà; l'intervallo $[\ell(y), u(y)]$ è deterministico (una volta osservato y). Di conseguenza è deterministica l'appartenenza o meno di θ all'intervallo $[0, 2/3]$, anche se noi non sappiamo in quale dei due casi siamo. Poiché però $\theta \in [L, U]$ in almeno il 95% dei campioni, confidiamo al livello 95% che $\theta \in [0, 2/3]$.

Un'affermazione probabilistica su θ è invece possibile sulla base della sua densità finale $p(\theta|y)$, $\theta \in \Theta$, una volta che si sia scelta una densità iniziale $p(\theta)$, $\theta \in \Theta$. Per esempio, scegliendo la distribuzione uniforme, come Bruno, si trova $\theta|y \sim \mathrm{B}(y+1, n-y+1)$ e un *intervallo credibile* al 90% per θ, avendo osservato $\{Y = y\}$, si otterrà prendendo il quinto e il novantacinquesimo percentile della distribuzione trovata.