

# Binomiale Negativa e Campionamento Inverso

Teoria delle Decisioni, Aprile 2001 (S Modica)

## 1 Motivazione

La motivazione qui è statistica. In un campionamento da una sequenza di esperimenti vero-falso (0-1, successo-insuccesso...) in genere ci si accontenta di un numero sufficientemente alto di: o osservazioni, o successi osservati. Domanda: questa scelta può essere rilevante ai fini della stima del parametro/probabilità incognita, o no? La risposta bayesiana è no, quella 'classica' sì.

Il tutto è ben spiegato nei libri di statistica (per noi Piccinato *Metodi per le Decisioni Statistiche*, dove abbiamo in mente l'Esempio 4.5 - 1); qui diamo un po' di background sulla binomiale negativa e mostriamo che è effettivamente una distribuzione di probabilità (con somma delle masse uguale ad 1).

## 2 Sequenze di Esperimenti Bernoulliani

Ricordiamo il contesto della binomiale: come spazio di base vero-falso prendiamo  $\Omega = \{0, 1\}$ , con probabilità  $P(\{1\}) = \theta$ . L'esperimento ripetuto indipendentemente  $n$  volte ha come spazio dei risultati  $\Omega^n \ni (x_1, \dots, x_n)$ ,  $x_i = 0, 1$  e probabilità prodotto  $P^n \equiv P \times \dots \times P$  definita da  $P^n(x_1, \dots, x_n) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$ . La variabile  $S$  'numero di successi in  $n$  prove' su  $\Omega^n$  è definita da  $S(x_1, \dots, x_n) = \sum x_i$  ed ha come sappiamo massa di probabilità

$$f_S(s) \equiv P^n[S = s] = \binom{n}{s} \theta^s (1 - \theta)^{n-s}, \quad s = 0, 1, \dots, n. \quad (1)$$

Lo stesso esperimento ripetuto indefinitamente ha come spazio dei risultati quello delle sequenze di zeri e uni, cioè  $\Omega^\infty \ni (x_n) \equiv (x_1, x_2, x_n, \dots)$ ,  $x_n = 0, 1$  per ogni  $n \geq 1$ ; e su questo spazio si può ancora definire una probabilità prodotto  $P^\infty$  che soddisfa la proprietà che intuitivamente ci aspettiamo sulle sequenze finite (ma per farlo ci vuole lavoro), cioè

$$P^\infty\{(x_1, x_2, \dots) \mid x_i = \bar{x}_i, i = 1, 2, \dots, n\} = \theta^{\sum \bar{x}_i} (1 - \theta)^{n - \sum \bar{x}_i}. \quad (2)$$

## 3 Tempi di Attesa

Quante ripetizioni mi devo aspettare prima di vedere realizzati  $s$  successi? Per esaminare questa spontanea domanda si *deve* pensare in termini di sequenze infinite di esperimenti. Perché con una sequenza di  $n$  ripetizioni puoi solo rispondere con un numero da 0 ad  $n$ , e questo non è naturale. Per rispondere in modo soddisfacente dobbiamo avere a disposizione qualunque numero, cioè una sequenza infinita di esperimenti.

La variabile  $S_n$  'numero di successi in  $n$  prove' resta ancora definita su  $\Omega^\infty$ ; per scriverne la definizione conviene usare le variabili 'coordinate'  $X_n: \Omega^\infty \rightarrow \mathbb{R}$ :

$$X_n(x_1, x_2, \dots) = x_n; \quad S_n = \sum_{i=1}^n X_i, \quad n = 1, 2, \dots$$

E dalla (2) è immediato che la distribuzione di  $S_n$  è ancora  $P^\infty[S_n = s] = P^n[S = s]$ ,  $s = 0, 1, \dots, n$ , ultimo membro preso dalla (1).

Più interessante la variabile  $Z$  'tempo di attesa per il primo successo', definita da

$$Z(x_1, x_2, \dots) = \min\{k \mid x_k = 1\}. \quad (3)$$

Nota che se  $(x_n)$  è la sequenza di tutti zeri l'insieme  $\{k \mid x_k = 1\}$  è vuoto, e porremo  $Z(0, 0, \dots) = \infty$ . D'altra parte, per  $\theta = 0, 1$  lo studio dei temi di attesa è inutile: se  $\theta = 0$   $P^\infty[Z = \infty] = 1$ , e se  $\theta = 1$   $P^\infty[Z = 1] = 1$ ; quindi supporremo d'ora in poi  $0 < \theta < 1$ . A questo punto la sequenza  $(0, 0, \dots)$  ha probabilità zero; e in questo contesto si può porre per convenzione  $0 \cdot \infty = 0$ ; conclusione, potremo ignorare il valore  $Z = \infty$  nel calcolo dei valori attesi.

Chiaramente  $Z = n$  si verifica se i primi  $n - 1$  sono insuccessi e l'ennesimo è un successo, dunque la massa di probabilità di  $Z$  è

$$f_Z(n) \equiv P^\infty[Z = n] = (1 - \theta)^{n-1} \cdot \theta \quad n = 1, 2, \dots$$

È facile vedere che  $\sum_{n=1}^{\infty} f_Z(n) = 1$ .<sup>1</sup> E per calcolare il valore atteso:

$$\begin{aligned} E(Z) &= \sum_{n=1}^{\infty} n f_Z(n) = \theta \sum_{n=1}^{\infty} n (1 - \theta)^{n-1} \\ &= \theta \sum_{k=0}^{\infty} (1 - \theta)^k \sum_{n=0}^{\infty} (1 - \theta)^n = 1/\theta. \quad (4) \end{aligned}$$

Molto semplice: se la probabilità di successo è  $\theta = 1/6$  mi aspetto 6 ripetizioni prima di vederne uno; se  $\theta = 1/10$  me ne aspetto 10. . .

Per passare alla variabile  $W_s$  'tempo di attesa per osservare  $s$  successi' con cui abbiamo aperto la sezione, consideriamo il tempo di attesa  $Z_k$  fra il  $k - 1$ -esimo e il  $k$ -esimo successo. Formalmente possiamo definire recursivamente  $Z_1 = Z$  e (di nuovo a meno di un insieme di probabilità zero)

$$Z_{k+1}(x) = \min\{k > Z_k(x) \mid x_k = 1\} - Z_k(x), \quad x \equiv (x_1, x_2, \dots).$$

Poichè gli esperimenti sono indipendenti le  $Z_k$  hanno distribuzioni uguali, in particolare valore atteso  $1/\theta$ . La  $W_s$  resta definita ponendo

$$W_s = \sum_{k=1}^s Z_k;$$

e da ciò, prima ancora di conoscerne la distribuzione, possiamo ricavare che

$$E(W_s) = s/\theta.$$

<sup>1</sup> $\sum_{n=1}^{\infty} f_Z(n) = \theta \cdot \sum_{n=0}^{\infty} (1 - \theta)^n = \frac{\theta}{1 - (1 - \theta)} = 1$ .

<sup>2</sup>Per vedere la doppia sommatoria scrivi le potenze di  $(1 - \theta)$  affiancate in colonne, la prima lunga 1 elemento ( $= 1$ ), la seconda lunga 2 elementi uguali a  $(1 - \theta)$ , eccetera, e poi somma per righe.

## 4 La Binomiale Negativa

La distribuzione di  $W_s$  si chiama *binomiale negativa*, i cui parametri sono  $\theta$  ed  $s$ . Osservando che  $W_s = n$  se si verificano  $s - 1$  successi nelle prime  $n$  ripetizioni e un successo alla  $n$ -esima si vede che

$$f(n) \equiv P^\infty[W_s = n] = \binom{n-1}{s-1} \theta^s (1-\theta)^{n-s}, \quad n = s, s+1, \dots \quad (5)$$

In questo nota siamo interessati principalmente a ricavare l'uguaglianza  $\sum_{n=s}^\infty f(n) = 1$ , e qui veniamo al dunque, in due modi.

Primo: un argomento probabilistico, in questo corso il più importante, che sta in Ross, *A First Course in Probability*: poichè  $W_s = \sum_{k=1}^s Z_k$ ,  $W_s$  è finito se tutte le  $Z_k$  lo sono; ma sappiamo che  $P^\infty[Z_k < \infty] = 1$  per ogni  $k = 1, \dots, s$ ; dunque  $P^\infty[W_s < \infty] = P^\infty(\cap_{k=1}^s [Z_k < \infty]) = 1$ , come volevamo.<sup>3</sup>

Secondo, una dimostrazione analitica, che si basa su un risultato (dovuto a Cauchy) sulle serie numeriche, in particolare sulle serie prodotto. Per definire una serie prodotto cominciamo dal prodotto finito di due somme  $(a_0 + \dots + a_N)(b_0 + \dots + b_N)$ ; disegniamo una matrice con prima colonna  $a_0 b_0, a_0 b_1, \dots, a_0 b_N$ , seconda  $a_1 b_0, a_1 b_1, \dots, a_1 b_N$  e raggruppiamo prima per righe, poi per diagonali sudovest-norddest; otteniamo:

$$\sum_{n=0}^N a_n \cdot \sum_{n=0}^N b_n = \sum_{n=0}^N a_n \sum_{k=0}^N b_k = \sum_{n=0}^N \sum_{k=0}^n a_k b_{n-k}.$$

Ora ricorda che una serie è definita come limite delle somme parziali. Se proviamo a generalizzare il prodotto di due somme alle serie usando la prima decomposizione ( $\sum_{n=0}^N a_n \sum_{k=0}^N b_k$ ) ci troviamo subito in difficoltà: in ogni somma parziale ogni termine diventerebbe una serie; sicchè si dovrebbero sommare somme infinite, e alla fine fare un limite di limiti —una cosa ingarbugliata. Con la seconda decomposizione invece il problema non si pone: la somma parziale di ordine  $n$  ha  $n + 1$  termini, una somma finita; dunque il processo di limite (per  $N \rightarrow \infty$ ) è quello solito, e abbiamo una serie ben definita, che si chiama *serie prodotto*:

$$\sum_{n=0}^\infty \sum_{k=0}^n a_k b_{n-k}.$$

Il risultato che useremo è il seguente (vedi un libro di Analisi, es. Conti, *Calcolo*):

*Siano  $\sum a_n, \sum b_n$  due serie convergenti, una assolutamente. Allora la serie prodotto converge al prodotto delle somme; cioè, con  $c_n = \sum_{k=0}^n a_k b_{n-k}$ ,*

$$\sum c_n = \sum a_n \cdot \sum b_n. \quad (6)$$

Tornando al nostro obiettivo di dimostrare che  $\sum_{n=s}^\infty f(n) = 1$  per la  $f(n)$  della (5): scrivi  $n - 1 = n - s + s - 1$ , nota che  $n \geq s$  sse  $n - s \geq 0$  e che nella somma  $\theta^s$  è costante; poni  $1 - \theta = x$ , e riscrivi il risultato voluto come:

$$\forall |x| < 1 \quad \sum_{n=0}^\infty \binom{s+n}{s} x^n = 1/(1-x)^{s+1}, \quad s = 1, 2, \dots \quad (7)$$

(noi abbiamo  $0 < x < 1$ , ma quel che qui importa è il valore assoluto). Lo dimostriamo per induzione su  $s$  (è un esercizio di Conti): Per  $s = 1$  l'asserto è che  $\sum_{n=0}^\infty (n+1)x^n =$

<sup>3</sup>Usiamo il fatto che l'intersezione finita di eventi di probabilità 1 ha probabilità 1. Per due eventi,  $A \cap B = (A^c \cup B^c)^c$ , sicchè  $\mu(A \cap B)^c \leq \mu A^c + \mu B^c = 0$ , cioè  $\mu(A \cap B) = 1$ ; conclusione per induzione.

$1/(1-x)^2$ ; ma sappiamo che  $\sum_{n \geq 0} x^n = 1/(1-x)$ , e che la convergenza è assoluta; dunque possiamo applicare il teorema di Cauchy ottenendo

$$(1-x)^{-2} = \sum_{n \geq 0} x^n \cdot \sum_{n \geq 0} x^n = \sum_{n \geq 0} \sum_{0 \leq k \leq n} x^k x^{n-k} = \sum_{n \geq 0} (n+1)x^n,$$

come volevasi. Resta da verificare che se l'uguaglianza in (7) vale per  $s$  allora vale anche per  $s+1$ . Moltiplicando la serie assolutamente convergente  $\sum_{n \geq 0} x^n$  per la  $\sum_{n \geq 0} \binom{s+n}{s} x^n$ , convergente per ipotesi induttiva otteniamo, usando anche la relazione combinatoriale  $\sum_{0 \leq k \leq n} \binom{s+k}{s} = \binom{s+n+1}{s+1}$ :

$$\begin{aligned} (1-x)^{-(s+2)} &= \sum_{n \geq 0} x^n \cdot \sum_{n \geq 0} \binom{s+n}{s} x^n \\ &= \sum_{n \geq 0} \sum_{0 \leq k \leq n} \binom{s+k}{s} x^n = \sum_{n \geq 0} \binom{s+n+1}{s+1} x^n; \end{aligned} \quad (8)$$

ma questa è la relazione cercata. Fine della dimostrazione analitica (quella probabilistica era forse più elegante).

## 5 Campionamento Inverso

Nel campionamento inverso da sequenze bernoulliane si ripete l'esperimento finchè non si osserva un numero fissato  $s$  di successi (vedi libro di statistica). Dunque lo spazio dei risultati è  $\{(s, n) \mid n \geq s\}$ ; e il numero (aleatorio)  $n$  di prove da fare non è altro che  $W_s$ , sicchè ha una distribuzione binomiale negativa, con parametri  $s$  e la probabilità di successo  $\theta$ .

Vogliamo qui verificare che lo stimatore  $(s-1)/(n-1)$  di  $\theta$  è non distorto; che cioè  $E_\theta(s-1)/(n-1) = \theta$  per ogni  $\theta$  (è un esercizio del Piccinato). Ma abbiamo

$$\begin{aligned} E_\theta \frac{s-1}{n-1} &= \sum_{n \geq s} \frac{s-1}{n-1} f(n) = (s-1) \sum_{n \geq s} \frac{1}{n-1} \binom{n-1}{s-1} \theta^s (1-\theta)^{n-s} \\ &= \theta \sum_{n-1 \geq s-1} \binom{n-2}{s-2} \theta^{s-1} (1-\theta)^{(n-1)-(s-1)} \\ &= \theta \sum_{n \geq s-1} \binom{n-1}{s-2} \theta^{s-1} (1-\theta)^{n-(s-1)} = \theta, \end{aligned}$$

l'ultimo uguale perchè la somma è la massa totale di probabilità –uguale ad uno– della binomiale negativa di parametri  $\theta$  ed  $s-1$ .

Ora che hai chiarito l'aspetto formale della contrapposizione fra principio della verosimiglianza e non distorsione della stima (almeno nel contesto di questo esempio bernoulliano), torna a meditare sull'aspetto concettuale: sei classico o bayesiano?