MODULO: Variabili aleatorie

Francesco Bologna   Enrico Rogora

CASIO – Università di Roma

10 - 14 Luglio 2017 - Avellino

Incertezza

Condizione abituale: incertezza.

- Domani pioverà?
- Sta piovendo a Nairobi?
- In che anno morirà l'attuale segretario dei DS?
- In che anno è morto Napoleone Bonaparte?
- Cosa succede se premo il tasto π sulla calcolatrice?
- Cosa succede se premo il tasto `ALEA` sulla calcolatrice?

Oggetti aleatori

Aleatorio significa *incerto* (per Te) ma di per sè *ben determinato*.

- Eventi
- Numeri
- Funzioni o processi
- Enti

Caso limite dell'incertezza, evento/numero/processo/ente *pleonastico*.

Eventi aleatori come particolari numeri aleatori

Un evento è un fatto preciso, qualificato da una proposizione logica, che può essere vera o falsa. Si dice aleatorio se non abbiamo necessariamente conoscenza completa del suo verificarsi.
Esempi: Domani pioverà? Sta piovendo a Nairobi? È piovuto ieri a Nantukett? Premendo il tasto Alea della calcolatrice, il numero che esce è maggiore di 0.5
Ad ogni evento aleatorio si può associare il corrispondente *numero aleatorio caratteristico* che vale 1 se l'evento si verifica e zero altrimenti.
Esempio: realizzare programmino con la calcolatrice per calcolare il numero caratteristico dell'ultimo evento proposto; analogo con una tabella di numeri casuali.

Variabile aleatoria

Una variabile aleatoria è un numero aleatorio, con la specificazione della probabilità di ogni valore. Se i possibili valori sono in numero finito, una variabile aleatoria è completamente specificato da una tabella, come la seguente

a_1	a_2	$\dots$	a_k
p_1	p_2	$\dots$	p_k

Valore atteso

Il *valore atteso* della variabile aleatoria discreta

$$X = \begin{array}{|c|c|c|c|} \hline a_1 & a_2 & \dots & a_k \\ \hline p_1 & p_2 & \dots & p_k \\ \hline \end{array}$$

è per definizione

$$E(X) = p_1 x_1 + \dots p_k x_k.$$

Distribuzione congiunta di variabili aleatorie

La *distribuzione congiunta* di due variabili aleatorie

$$X = \begin{array}{|c|c|c|c|}\hline a_1 & a_2 & \dots & a_k \\ \hline p_1 & p_2 & \dots & p_k \\ \hline \end{array} \qquad Y = \begin{array}{|c|c|c|c|}\hline b_1 & b_2 & \dots & b_h \\ \hline q_1 & q_2 & \dots & q_h \\ \hline \end{array}$$

assegna ad ogni coppia a_i, b_j la probabilità p_{ij} che la variabile aleatoria X valga a_i e la variabile aleatoria Y valga b_j. In simboli, $p_{ij} = P(X = a_i, Y = b_j)$.
Le probabilità p_{ij} non sono determinate, in generale, dalle p_i e q_j ma sono soggette alle condizioni

$$\sum_{i=1}^{h} p_{ij} = q_j \qquad \sum_{j=1}^{k} p_{ij} = p_i.$$

Variabili aleatorie indipendenti

Due variabili aleatorie si dicono *indipendenti* se le probabilità $p_{ij} = P(X = i, Y = j)$ che X valga a_i e Y valga b_j verificano la condizione

$$p_{ij} = p_i q_j.$$

dove $p_i = P(X = a_i)$ e $q_j = P(Y = b_j)$

Esempio: lancio ripetuto di una moneta I

Consideriamo per esempio le variabili aleatorie X_i che descrivono l'esito dell'i-esimo lancio nel lancio ripetuto di una stessa moneta. Se indichiamo con 0 il risultato "esce croce" e con 1 "esce testa", le variabili aleatorie X_i sono determinate dalle seguenti tabelle (identiche, le variabili si dicono quindi *identicamente distribuite*)

$$X_i = \begin{array}{|c|c|} \hline 0 & 1 \\ \hline (1-p) & p \\ \hline \end{array}$$

Le variabili aleatorie X_i sono tra loro *indipendenti*. Questa ipotesi permette di calcolare esplicitamente la probabilità di una qualsiasi successione di esiti possibili.
Per esempio la probabilità
$P(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 1, X_7 = 1)$, o,

Esempio: lancio ripetuto di una moneta II

più brevemente, la probabilità della sequenza $(0, 0, 1, 0, 1, 1, 1)$, ovvero di osservare croce al primo lancio, croce al secondo lancio, testa al terzo lancio, croce al quarto lancio, testa al quinto al sesto e al settimo lancio lancio è, nell'ipotesi di indipendenza,

$$(1-p)\cdot(1-p)\cdot p\cdot(1-p)\cdot p\cdot p\cdot p = p^4(1-p)^3,$$

ovvero p elevato al numero delle teste moltiplicato per $(1-p)$ elevato al numero delle croci.
Quest'ultima descrizione si generalizza facilmente. Se codifichiamo una successione di esiti possibili nel lancio ripetuto di una moneta con una *successione binaria*

$$\beta = (\beta_1, \dots, \beta_n)$$

Esempio: lancio ripetuto di una moneta III

dove ogni β_i è un numero che vale 0 o 1, allora la probabilità di osservare β vale

$$p(\beta) = p^s(1-p)^{n-s}$$

dove s è il numero delle teste (che coincide con $\sum_i \beta_i$, n è la lunghezza della stringa binaria e $n-s$ è il numero delle croci.

Somma di variabili aleatorie

La somma di due variabili aleatorie X e Y

$$X = \begin{array}{|c|c|c|c|} \hline a_1 & a_2 & \dots & a_k \\ \hline p_1 & p_2 & \dots & p_k \\ \hline \end{array} \qquad Y = \begin{array}{|c|c|c|c|} \hline b_1 & b_2 & \dots & b_h \\ \hline q_1 & q_2 & \dots & q_h \\ \hline \end{array}$$

è la variabile aleatoria

$$X + Y = \begin{array}{|c|c|c|c|} \hline c_1 & c_2 & \dots & c_m \\ \hline \pi_1 & \pi_2 & \dots & \pi_k \\ \hline \end{array}$$

dove c_h sono i possibili valori distinti delle somme $a_i + b_j$ (in generale lo stesso valore c_h può essere ottenuto per scelte diverse degli indici (i, j)) e la probabilità π_h di osservare c_h si ottiene sommando le probabilità p_{ij} su tutte le coppie di indici (i, j) tali che $a_i + b_j = c_h$. In simboli, $\pi_h = \sum_{(i,j)|a_i+b_j=c_h} p_{ij}$.

Esempio

Siano

$$X_1 = X_2 = \begin{array}{|c|c|} \hline 0 & 1 \\ \hline (1-p) & p \\ \hline \end{array}$$

Allora

$$X_1 + X_2 = \begin{array}{|c|c|c|} \hline 0 & 1 & 2 \\ \hline p_{00} & p_{10} + p_{01} & p_{11} \\ \hline \end{array}$$

Se le variabili sono indipendenti, $p_{00} = (1-p)^2$, $p_{10} = p_{01} = p(1-p)$, $p_{11} = p^2$, e quindi

$$X_1 + X_2 = \begin{array}{|c|c|c|} \hline 0 & 1 & 2 \\ \hline (1-p)^2 & 2p(1-p) & p^2 \\ \hline \end{array}$$

Esempio

Siano

$$X_1 = X_2 = X_3 = \begin{array}{|c|c|} \hline 0 & 1 \\ \hline (1-p) & p \\ \hline \end{array}$$

Allora

$$X_1+X_2+X_3 = \begin{array}{|c|c|c|c|} \hline 0 & 1 & 2 & 3 \\ \hline p_{000} & p_{100}+p_{010}+p_{001} & p_{110}+p_{101}+p_{011} & p_{111} \\ \hline \end{array}$$

Se le variabili sono indipendenti, $p_{000} = (1-p)^3$, $p_{100} = p_{010} = p_{001} = p(1-p)^2$, $p_{110} = p_{101} = p_{011} p^2(1-p)$, $p_{111} = p^3$ e quindi

$$X_1 + X_2 + X_3 = \begin{array}{|c|c|c|c|} \hline 0 & 1 & 2 & 3 \\ \hline (1-p)^3 & 3p(1-p)^2 & 3p^2(1-p) & p^3 \\ \hline \end{array}$$

Diatribuzione binomiale

La distribuzione binomiale, di parametri n, p è la distribuzione della somma di n variabili aleatorie X_i indipendenti e identicamente distribuite

$$X_i = \begin{array}{|c|c|} \hline 0 & 1 \\ \hline (1-p) & p \\ \hline \end{array}$$

Tale distribuzione è quella che assegna la probabilità di osservare k teste in n lanci ripetuti di una moneta in cui la probabilità di osservare testa è p e quella di osservare croce è $(1-p)$. La distribuzione della variabile aleatoria $Y = X_1 + X_2 + \cdots + X_n$ è

$$Y = \begin{array}{|c|c|c|c|c|} \hline 0 & 1 & 2 & \ldots & n \\ \hline (1-p)^n & (1-p)^{n-1}p & (1-p)^{n-2}p^2 & \ldots & p^n \\ \hline \end{array}$$

Valore atteso della somma di variabili aleatorie

Nella distribuzione congiunta di due variabili aleatorie, le probabilità congiunte p_{ij} non sono determinate ma devono soddisfare ai vincoli $\sum_i p_{ij} = b_j$ e $\sum_j p_{ij}$.
Questi vincoli sono sufficienti a determinare il valore di $E(X+Y)$. Si ha infatti

$$E(X+Y) = \sum_{ij} p_{ij}(a_i + b_j) = \sum_{ij} p_{ij} a_i + \sum_{ij} p_{ij} b_j =$$

$$\sum_i \sum_j (a_i p_{ij}) + \sum_j \sum_i (b_j p_{ij}) = \sum_i a_i \sum_j p_{ij} + \sum_j b_j \sum_i p_{ij} =$$

$$\sum_i a_i p_i + \sum_j b_j q_j = E(X) + E(Y).$$

Prodotto di variabili aleatorie

Analogamente alla somma possiamo definire il prodotto di due variabili aleatorie X e Y. Se X e Y sono definite da

$$X = \begin{array}{|c|c|c|c|} \hline a_1 & a_2 & \dots & a_k \\ \hline p_1 & p_2 & \dots & p_k \\ \hline \end{array} \qquad Y = \begin{array}{|c|c|c|c|} \hline b_1 & b_2 & \dots & b_h \\ \hline q_1 & q_2 & \dots & q_h \\ \hline \end{array}$$

il prodotto $X * Y$ è la variabile aleatoria

$$X * Y = \begin{array}{|c|c|c|c|} \hline c_1 & c_2 & \dots & c_m \\ \hline \pi_1 & \pi_2 & \dots & \pi_k \\ \hline \end{array}$$

dove c_h sono i possibili valori distinti dei prodotti $a_i \cdot b_j$ (in generale lo stesso valore c_h può essere ottenuto per scelte diverse degli indici (i, j)) e la probabilità π_h di osservare c_h si ottiene sommando le probabilità p_{ij} su tutte le coppie di indici (i, j) tali che $a_i \cdot b_j = c_h$. In simboli, $\pi_h = \sum_{(i,j)|a_i \cdot b_j = c_h} p_{ij}$.

Valore atteso del prodotto di variabili aleatorie

Il valore atteso del prodotto di due variabili aleatorie, non è in generale il prodotto dei valori aspettati. Lo è se le variabili aleatorie sono *indipendenti*, o più in generale se sono *non correlate*, dove la correlazione di due variabili è per definizione $\mathrm{Corr}(X, Y) = E(X * Y) - E(X) \cdot E(Y)$.

Attenzione

La variabile X^2 che si ottiene quadrando la stessa variabile X (quindi non facendo il prodotto di due variabili indipendenti identicamente distribuite) si ottiene scegliendo come distribuzione congiunta quella definita ponendo

$$p_{ij} = \begin{cases} p_i & \text{se} \quad i = j \\ 0 & \text{se} \quad i \neq j \end{cases}$$