INTRODUZIONE ALLA PROBABILITA

Note per il corso di

Laboratorio di Tecniche Computazionali in Biologia

Pierluigi Contucci

Anno Accademico 2006-2007

Università di Bologna

Queste brevi note (in forma ancora non definitiva) sono state scritte come supporto alle lezioni di Matematica per il corso di Laboratorio di Tecniche Computazionali in Biologia degli studenti del secondo anno. La scelta degli argomenti trattati riflette le necessità della pratica scientifica. Essa è fondata infatti su operazioni di misura che producono dei dati sperimentali composti dai valori numerici delle grandezze osservate e dai relativi errori. Si veda in particolare l'intervento di F.S. Collins [1] sulla ricerca in biologia. Oltre alle lezioni frontali in cui i concetti di base del ragionamento di natura probabilistica vengono esposti il corso si varrà di alcune lezioni con illustrazione al calcolatore. La scelta del metodo utilizzato che ci piace chiamare sperimentale *assolve il duplice scopo di svincolare l'insegnamento della matematica dalla linea ipotesi-tesi-dimostrazione e di permettere di avvicinare lo studente alla simulazione col calcolatore. Con questo infatti si eseguono esperimenti interagendo con la macchina attraverso un browser internet (explorer, netscape, etc) senza nessun prerequisito informatico. L'esperimento si effettua letteralmente con un* click *che avvia un applet in java. In un istante per esempio si potrà simulare il lancio di 100 monete ripetuto 100.000 e osservando la statistica delle teste e delle croci avere una idea del significato dei fenomeni dei grandi numeri e della legge Gaussiana. Sull'efficacia del metodo sperimentale in matematica si leggano gli interventi di W. Thurston [2], [3] e Epstein e Levi. [4]*

Saranno molto apprezzati gli studenti che mi segnaleranno eventuali errori o sviste o vorrano darmi suggerimenti per migliorare l'esposizione.

Ringrazio la Prof. Barbara Pecori per l'attenta lettura del manoscritto e per gli utili commenti e suggerimenti.

Pierluigi Contucci, Bologna, Ottobre 2006 (prima stesura Novembre 2002).

email: contucci@dm.unibo.it
homepage: http://www.dm.unibo.it/~contucci

Consideriamo una macchina da scrivere che stampa 3 lettere, a, b e c. Quante parole diverse di 2 lettere si possono scrivere? Avremo aa, ab, ac, ba, bb, bc, ca, cb e cc. Cioè 9 parole. Invece che elencarle e successivamente contarle possiamo contarle direttamente osservando che la prima lettera può essere scelta tra 3 possibilità, la seconda tra 3 possibilità quindi il numero totale sarà di $3 \cdot 3 = 9$. In generale con una macchina da scrivere che stampa n lettere diverse possiamo scrivere

$$n \cdot n \cdots n = n^k \tag{1}$$

parole distinte lunghe k lettere.

Consideriamo ora invece un problema diverso. Invece di una macchina da scrivere abbiamo un sacchetto che contiene 3 lettere a, b e c. Quante parole di 2 lettere possiamo scriverci? In questo caso ciascuna lettera potrà essere utilizzata una sola volta: ab, ac, ba, bc, ca e cb. Cioè 6 possibilità. In generale con un sacchetto che contiene n lettere distinte possiamo scrivere

$$n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) \tag{2}$$

parole diverse lunghe k dato che la prima lettera potrà essere scelta tra n possibilità, la seconda tra $n-1$ etc. Introducendo il simbolo di *fattoriale*

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 \tag{3}$$

che conta il numero di parole lunghe n che si possono scrivere con un sacchetto con n lettere distinte il numero precedente si può esprimere come:

$$n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!} \tag{4}$$

Consideriamo infine un sacchetto che contiene due a e una b. Quante parole diverse possiamo formare usando tutte le tre lettere? Avremo aab, aba, baa. In generale se il sacchetto contiene n lettere di cui k a e $(n-k)$ b possiamo procedere al conteggio in questo modo: dapprima fingiamo che le lettere siano tutte distinte, in tal caso si avrebbero $n!$ parole diverse. Cosi facendo tuttavia abbiamo contato parole uguali come se fossero distinte. Dobbiamo quindi dividere il risultato per i modi in cui si permutano tra loro le lettere coincidenti dipo a cioè $k!$ e di tipo b cioè $(n-k)!$. Avremo quindi che il numero di parole distinte è il coefficiente binomiale:

$$\frac{n!}{k!(n-k)!} \tag{5}$$

Il nome di tale coefficiente deriva dal fatto che esso compare nello sviluppo del binomio di Newton:

$$(a+b)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} a^k b^{n-k} \tag{6}$$

SPAZI DI PROBABILITÀ E MODELLI DI PROVE RIPETUTE

Consideriamo un insieme di indici $i \in I$ e ad essi associamo la collezione di numeri $p_i \in P$ che verifica le proprietà

$$\textit{Assioma } 0) \qquad\qquad 0 \leq p_i \leq 1 \tag{7}$$

$$\textit{Assioma } 1) \qquad \sum_i p_i = 1 \tag{8}$$

Ciascun indice i identifica un *evento elementare* e p_i è la sua *probabilità*. La coppia (I, P) è uno spazio di *probabilità discreto* e le relazioni (7) e (8) sono il primo e secondo assioma.

Esempio 1: la moneta simmetrica ha spazio di probabilità identificato da

$$T \equiv \textit{testa} , \tag{9}$$

$$C \equiv \textit{croce} , \tag{10}$$

quindi

$$I = (T, C) ; \tag{11}$$

e dalle relative probabilit'a

$$p_T = \frac{1}{2} , \tag{12}$$

$$p_C = \frac{1}{2} . \tag{13}$$

$$P \equiv \left(\frac{1}{2}, \frac{1}{2}\right) . \tag{14}$$

Per esercizio si verfichi che i due assiomi sono soddisfatti.

Esempio 2: la moneta asimmetrica (o truccata) ha spazio di probabilità identificato dallo stesso spazio di indici

$$I \equiv (T, C) , \tag{15}$$

ma da un diverso valore delle probabilità degli eventi

$$p_T = p , \tag{16}$$

con

$$0 \leq p \leq 1 , \tag{17}$$

e da

$$p_C = 1 - p , \tag{18}$$

quindi

$$P \equiv (p, 1 - p) . \tag{19}$$

Per esercizio si verfichi che anche in questo caso che generalizza quello simmetrico il primo e secondo assioma sono soddisfatti. Si diano inoltre alcuni esempi numerici di p.

Esempio 3: il dado regolare a sei facce ha spazio di probabilità identificato da

$$I \equiv (1, 2, 3, 4, 5, 6) , \tag{20}$$

e da

$$P \equiv (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}) . \tag{21}$$

Un possibile esempio di dado truccato è quello in cui l'evento 6 da solo ha probabilità un mezzo e gli altri eventi $1, 2, 3, 4, 5$ hanno la probabilità p: in tal caso l'assioma (8) impone

$$5p + \frac{1}{2} = 1 , \tag{22}$$

e quindi $p = \frac{1}{10}$. Si diano altri esempi di spazi di proabilità di un dado truccato.

È di grande importanza considerare il lancio della moneta (o del dado) ripetuto più volte in modo indipendente. Consideriamo dapprima il lancio della moneta simmetrica ripetuto due volte e costruiamo il suo spazio delle probabilità. Avremo che gli eventi elementari saranno

$$I \equiv (TT, TC, CT, CC) , \tag{23}$$

e le probabilità

$$P \quad \equiv \quad (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}) \; . \tag{24}$$

È utile nell' esempio appena dato analizzare alcuni eventi che non siano eventi elementari: l' evento A definito dall' uscita di croce al primo lancio potrà realizzarsi o come CT o come CC e utilizzando il simbolo di unione insiemistica (elementi che stanno o nel primo o nel secondo insieme) scriveremo

$$A \; = \; CT \, \cup \, CC \; ; \tag{25}$$

talvolta si utilizza per esso anche il simbolo

$$CX = CT \, \cup \, CC \; , \tag{26}$$

e la notazione significa che siamo nello spazio di probabilità di due lanci in cui il primo è croce e il secondo non è specificato. Similmente l' evento TXC indica l'uscita, nello spazio dei tre lanci di testa al primo e croce al terzo e vale:

$$TXC = TTC \, \cup \, TCC \; . \tag{27}$$

Gli esempi precedenti illustrano che l'unione di eventi è essa stessa un evento. La stessa proprietà vale per l'intersezione (elementi che stanno sia nel primo che nel secondo insieme). Per esempio se consideriamo

$$TX \, \cap \, XC \; , \tag{28}$$

ci riferiamo all'evento nello spazio dei 2 lanci che ha testa al primo e croce al secondo quindi

$$TX \, \cap \, XC \; = TC \; , \tag{29}$$

In generale non c'è nessuna relazione tra tra le probabilità di certi eventi e quelle della loro intersezione. Tuttavia quando per due eventi A e B vale la

$$P(A \cap B) \;=\; P(A)P(B)\;, \tag{30}$$

i due eventi si dicono *indipendenti*. Due eventi A e B si dicono mutuamente esclusivi (o disgiunti) quando la loro intersezione è l' insieme vuoto:

$$A \cap B = \emptyset, \tag{31}$$

per essi vale

$$\textit{Assioma } 2) \qquad\qquad P(A \cup B) = P(A) + P(B)\;. \tag{32}$$

Più in generale quando la loro intersezione è non vuota l'assioma si generalizza in

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,. \tag{33}$$

Verifichiamo che i due eventi TX e XC sono indipendenti.

$$P(TX) = P(TT \cup TC) = P(TT) + P(TC) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}\,, \tag{34}$$

dove la prima uguaglianza viene dall'assioma 2 dato che $TT \cap TC = \emptyset$. Analogamente si ha $P(XC) = 1/2$ e quindi poichè $TX \cap XC = TC$ e $P(TC) = 1/4$ la relazione di indipendenza è verificata.

Problema. Dati due eventi indipendenti A e B con $P(A) = 1/2$ e $P(B) = 1/3$ calcoliamo la probabilità della loro unione. Osserviamo anzitutto che senza l'informazione di indipendenza il problema sarebbe indeterminato dato che manca il valore di $P(A \cap B)$. Grazie all'indipendenza tuttavia sappiamo che $P(A \cap B) = P(A)P(B)$ quindi per la (33) abbiamo:

$$P(A \cup B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{2} \cdot \frac{1}{3} = \frac{2}{3} \tag{35}$$

Esercizio. Si dica quali delle seguenti coppie di eventi sono indipendenti: (TX, XT), (CX, CX), (XT, CX), (CT, CC), (TC, XT).

Problema. Consideriamo il lancio di un dado ripetuto 3 volte, levento A in cui esce il 2 al secondo lancio e levento B in cui esce il 6 al primo lancio e il 4 al terzo. Ci proponiamo di vedere quali sono gli eventi elementari che compongono A e quelli che compongono B, di calcolare le relative probabilità e infine di determinare se A e B sono eventi indipendenti. Dalla definizione di A si ha:

$$A = X2X\,, \tag{36}$$

quindi A à formato dall unione dei $6 \cdot 6$ eventi ottenibili inserendo una delle cifre del dado al posto delle X, come per esempio $A_1 = 121$ (da leggere uno-due-uno e non centoventuno!) oppure $A_2 = 122$, etc. Analogamente avremo

$$B = 6X4\,, \tag{37}$$

quindi B sarà formato dai 6 eventi elementari che si ottengono inserendo una delle sei cifre del dado al posto della X, quali $B_1 = 614$, $B_2 = 624$, etc. Per il calcolo di P(A) la regola di somma fornisce:

$$P(A) = P(\cup_{i=1}^{36} A_i) = \sum_{i=1}^{36} P(A_i) = 36P(A_1) = 36\frac{1}{216} = \frac{1}{6} \tag{38}$$

e analogamente per il calcolo di P(B)

$$P(B) = P(\cup_{i=1}^{6} B_i) = \sum_{i=1}^{6} P(B_i) = 6P(B_1) = 6\frac{1}{216} = \frac{1}{36}\,. \tag{39}$$

Infine per stabilire se i due eventi A e B sono indipendenti notiamo anzitutto che l'evento intersezione $A \cap B = 624$ quindi $P(A \cap B) = \frac{1}{6^3} = \frac{1}{216}$ e

$P(A)P(B) = \frac{1}{6}\frac{1}{36} = \frac{1}{216}$, quindi $P(A \cap B) = P(A)P(B)$ cioè gli eventi A e B sono indipendenti.

PROBABILITÀ CONDIZIONATE E FORMULA DI BAYES

È utile nelle applicazioni definire il concetto di probabilità condizionata. La probabilità dell' evento A condizionato dallevento B è definita da:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \, . \tag{40}$$

Si vede che nel caso che due eventi siano indipendenti

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) \tag{41}$$

Sperimentalmente la probabilità condizionata è legata alla frequenza di un evento dato il verificarsi di un altro. Talvolta come nei problemi di diagnostica è utile poter risalire alla probabilità di una causa dato il sintomo quando si conosce la probabilità a priori del sintomo e la probabilità del sintomo data la causa. Chiamiamo con B_k uno dei possibili eventi cause dell'evento sintomo A. Assumeremo che la famiglia dei $\{B_k\}_{k=1}^n$ sia esaustiva cioè contenga tutte le possibilità e che i diversi B_k siano disgiunti: matematicamente cioò si esprime dicendo che

$$\cup_{k=1}^n B_k \; = \; I \; , \tag{42}$$

$$B_k \cap B_l \; = \; \emptyset \qquad \text{per} \qquad k \neq l \; . \tag{43}$$

La grandezza di nostro interesse è

$$P(B_k|A) \; = \; \frac{P(B_k \cap A)}{P(A)} \, , \tag{44}$$

in particolare vorremmo conoscere quale di esse per i diversi k è la più grande, individuare cioè la causa più probabile per un dato sintomo. Osserviamo anzitutto che i diversi eventi $B_k \cap A$ sono disgiunti e che per l'ipotesi di esaustività risulta

$$A = \cup_{l=1}^{n}(B_l \cap A) \tag{45}$$

e quindi

$$P(A) = P(\cup_{l=1}^{n}(B_l \cap A)) = \sum_{l=1}^{n} P(B_l \cap A)\,, \tag{46}$$

inoltre per definizione di probabilità condizionata

$$P(B_l \cap A) = P(A|B_l)P(B_l)\,, \tag{47}$$

da cui

$$P(A) = \sum_{l=1}^{n} P(A|B_l)P(B_l)\,, \tag{48}$$

che sostituita nella (44) fornisce la formula di Bayes:

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{l=1}^{n} P(A|B_l)P(B_l)}\,. \tag{49}$$

Sommario:

- Spazio di probabilità:

$$I \equiv (1, 2, 3 \cdots, n)$$

$$P \equiv (p_1, p_2, p_3 \cdots p_n)$$

$$\begin{aligned} \mathcal{A}_0 &: \quad 0 \le p_i \le 1 \\ \mathcal{A}_1 &: \quad \sum_i p_i = 1 \\ \mathcal{A}_2 &: P(A \cup B) = P(A) + P(B)\,, \quad A \cap B = \emptyset \end{aligned}$$

- Probabilità condizionata:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{50}$$

- Eventi indipendenti:

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \quad \text{oppure} \\ P(A|B) &= P(A) \end{aligned}$$

- Formula di Bayes:

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{l=1}^{n} P(A|B_l)P(B_l)}\,.$$

Una funzione che associa un numero a ciascun evento elementare di uno spazio di probabilità è chiamata *variabile aleatoria*. Per esempio la funzione che vale 1 in corrispondenza della testa della moneta e -1 in corrispondenza della croce:

$$\begin{aligned} I &= (T, C) \\ P &= (p, 1-p) \\ f &= (1, -1) \, . \end{aligned} \tag{51}$$

In generale avremo

$$\begin{aligned} I &= (1, 2, \cdots n) \\ P &= (p_1, p_2, \cdots p_n) \\ f &= (f_1, f_2, \cdots f_n) \, . \end{aligned}$$

Si definisce *valor medio* della funzione f la quantità

$$E(f) = \sum_{i=1}^{n} f_i p_i \, . \tag{52}$$

Notiamo che il valore medio di una combinazione lineare di due funzioni è uguale alla combinazione lineare dei valori medi:

$$E(\alpha f + \beta g) = \sum_{i=1}^{n} (\alpha f_i + \beta g_i) p_i = \alpha \sum_{i=1}^{n} f_i p_i + \beta \sum_{i=1}^{n} g_i p_i = \alpha E(f) + \beta E(g) \, . \tag{53}$$

Inoltre se una variabile aleatoria è non negativa anche la sua media è non negativa:

$$f_i \geq 0 \qquad \textit{implica} \qquad E(f) \geq 0 \, .$$

Infine si osserva che la variabile aleatoria costante unitaria ha valor medio unitario:

$$f_i \;=\; 1 \qquad\quad \textit{implica} \qquad E(f) = 1\ .$$

Si definisce varianza di una variabile aleatoria la quantità

$$\begin{aligned} V(f) &= E([f - E(f)]^2) \;=\; E\left(f^2 - 2fE(f) + E(f)^2\right) \\ &= E(f^2) - E(f)^2 \;=\; \sum_{i=1}^{n} f_i^2 p_i - (\sum_{i=1}^{n} f_i p_i)^2\,, \end{aligned} \tag{54}$$

dove abbiamo usato le proprietà della media. Notiamo che per definizione si ha sempre $V(f) \geq 0$ e quindi $E(f^2) \;\geq\; E(f)^2$. Vale inoltre $V(\alpha f) = \alpha^2 V(f)$.

Problema. calcolare media e varianza della variabile aleatoria definita dalle (51) nello spazio della moneta asimmetrica:

$$E(f) = f_T p_T + f_C p_C = p - (1-p) = 2p - 1\,, \tag{55}$$

$$E(f^2) = f_T^2 p_T + f_C^2 p_C = p + (1-p) = 1\,, \tag{56}$$

$$V(f) = E(f^2) - E(f)^2 = 1 - (2p-1)^2 = 1 - 4p^2 + 4p - 1 = 4p(1-p)\ . \tag{57}$$

Problema. Consideriamo un dado asimmetrico in cui la probabilità del 6 è di un mezzo mentre tutte le altre facce hanno uguale probabilità 1/10 (vedi la 22). Vogliamo calcolare nel lancio del dado ripetuto tre volte la probabilità degli eventi $X2X$ e $6X4$ e della loro intersezione 624; infine vogliamo calcolare media della variabile aleatoria $f_i = i$ e confrontarla col caso del dado simmetrico. Dal calcolo diretto per somma degli eventi elementari si ha

$$P(X2X) = \frac{1}{10}\,, \tag{58}$$

$$P(6X4) = \frac{1}{20}\,, \tag{59}$$

$$P(624) = \frac{1}{200}\,. \tag{60}$$

Notiamo quindi che lindipendenza tra gli eventi A e B vale anche nel caso del dado asimmetrico. Infine abbiamo:

$$E(f) = \sum_{i=1}^{6} ip_i \;=\; \frac{1}{10}(1+2+3+4+5) + \frac{1}{2}6 = 4\,, \tag{61}$$

mentre per il dado simmetrico vale

$$E(f) = \sum_{i=1}^{6} ip_i \;=\; \frac{1}{6}(1+2+3+4+5+6) = 3.5 \tag{62}$$

CORRELAZIONI

Due variabili aleatorie tali che

$$P(f = f_i, g = g_j) \;=\; p(f)_i p(g)_j \tag{63}$$

si dicono indipendenti. Per esse valgono alcune notevoli proprietà:

$$E(fg) \;=\; \sum_{i,j} f_i g_j p(f)_i p(g)_j \;=\; \left(\sum_i f_i p(f)_i\right)\left(\sum_j g_j p(f)_j\right) \;=\; E(f)E(g)\;, \tag{64}$$

inoltre

$$V(f+g) \;=\; E([f+g-E(f)-E(g)]^2) \;=\; V(f)+V(g) \tag{65}$$

In generale però la quantità $P(f = f_i, g = g_j)$ non può essere ricondotta alle distribuzioni delle singole variabili aleatorie e quindi la coppia (f, g)

sarà descritta dalle distribuzioni delle singole funzioni e da quella congiunta $p_{i,j}$. In tal caso si introduce la *covarianza*: posto $E(f) = \mu_f$, $E(g) = \mu_g$, $V(f) = \sigma_f^2$ e $V(g) = \sigma_g^2$ essa è definita da

$$\mathrm{Cov}(f,g) \;=\; E\left[(f-\mu_f)(g-\mu_g)\right] \;=\; E\left[fg\right] - E[f]E[g]\;. \tag{66}$$

Chiaramente

$$\mathrm{Cov}(f,f) \;=\; V(f)\,, \tag{67}$$

inoltre se f e g sono indipendenti

$$\mathrm{Cov}(f,g) \;=\; 0\;. \tag{68}$$

È utile anche introdurre la *correlazione* che misura quanto due variabili aleatorie siano correlate

$$\rho(f,g) \;=\; \frac{\mathrm{Cov}(f,g)}{\sqrt{\mathrm{Cov}(f)\mathrm{Cov}(g)}}\;. \tag{69}$$

Quando $f = g$ si ha $\rho = 1$. Quando $f = -g$ si ha $\rho = -1$. Quando f è indipendente da g si ha $\rho = 0$. Poichè per la disuguaglianza di Schwartz

$$\mathrm{Cov}(f,g) \leq \sqrt{V(f)V(g)} \tag{70}$$

si ha sempre

$$-1 \;\leq\; \rho(f,g) \;\leq\; 1\;. \tag{71}$$

DISTRIBUZIONE BINOMIALE

Vogliamo ora concentrarci sul caso della moneta asimmetrica e considerarne il lancio un numero arbitrario di volte n. In particolare vogliamo stabilire qual è la probabilità che in n lanci si abbiano k teste. Il problema si riconduce ad uno di tipo combinatorio del contare quante parole diverse possiamo formare con k lettere di tipo T e $n-k$ lettere di tipo C. Per cominciare facciamo il caso con $k=2$ e $n=3$. Le possibilità sono TTC, TCT, CTT ed ognuna di esse ha (per l'indipendenza dei lanci) probabilità:

$$p^2(1-p)\,. \tag{72}$$

In generale il numero di parole sarà dato dal fattore binomiale

$$\frac{n!}{k!(n-k)!}\,, \tag{73}$$

e la relativa probabilità di ciascuno degli eventi

$$p^k(1-p)^{(n-k)}\,. \tag{74}$$

Pertanto utilizzando la proprietà (32) otteniamo

$$P_n(k) = \frac{n!}{k!(n-k)!}p^k(1-p)^{(n-k)}\,. \tag{75}$$

La formula sopra scritta definisce la distribuzione binomiale della variabile aleatoria f = numero di teste in n lanci. Notiamo infatti che la somma delle probabilità totale risulta

$$\sum_{k=0}^{n} \frac{n!}{k!(n-k)!}p^k(1-p)^{(n-k)} = (p+1-p)^n = 1, \tag{76}$$

dove si è utilizzata la formula del binomio di Newton.

Calcoliamo media e varianza di f.

$$E(f) \;=\; \sum_{k=0}^{n} k\frac{n!}{k!(n-k)!}p^k(1-p)^{(n-k)} \tag{77}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{(n-k)} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{(n-k)}
\end{aligned}
$$

dove si è osservato che il primo termine della somma non contribuisce essendo zero e si è semplificato un fattore k dai rimanenti. Ora poniamo $l = k - 1$ e otteniamo

$$
\begin{aligned}
E(f) &= \sum_{l=0}^{n-1} \frac{n!}{l!(n-l-1)!} p^{l+1} (1-p)^{(n-l-1)} \qquad (78) \\
&= np \sum_{l=0}^{n-1} \frac{(n-1)!}{l!(n-1-l)!} p^l (1-p)^{(n-1-l)} \\
&= np \sum_{l=0}^{n-1} P_{n-1}(l) = np\,.
\end{aligned}
$$

Per il calcolo della varianza si procede in modo analogo

$$
\begin{aligned}
E(f^2) &= \sum_{k=0}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{(n-k)} \qquad (79) \\
&= \sum_{k=1}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{(n-k)} \\
&= \sum_{k=1}^{n} k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{(n-k)}
\end{aligned}
$$

Ora poniamo $l = k - 1$ e otteniamo

$$
\begin{aligned}
E(f^2) &= \sum_{l=0}^{n-1} (l+1) \frac{n!}{l!(n-1-l)!} p^{l+1} (1-p)^{(n-1-l)} \qquad (80) \\
&= np \sum_{l=0}^{n-1} l P_{n-1}(l) + np \sum_{l=0}^{n-1} P_{n-1}(l)
\end{aligned}
$$

$$= \quad np(n-1)p + np = n^2p^2 - np^2 + np\,.$$

da cui

$$V(f) = np(1-p)\ . \tag{81}$$

DISTRIBUZIONE DI POISSON

In talune applicazioni quando n è molto grande (popolazione numerosa) e p molto piccolo (evento raro) è conveniente approssimare la distribuzione binomiale con la distribuzione di Poisson. Se per esempio ci interessa conoscere qual'è la probabilità che in una popolazione di un milione di individui ci siano in un giorno cinque decessi causati da una malattia la cui incidenza statistica è di 10^{-6} la distribuzione binomiale ci dice che essa vale

$$\frac{10^6!}{5!(10^6-5)!}10^{-30}(1-10^{-6})^{(10^6-5)} \tag{82}$$

un numero di non agevole computo! Si può tuttavia dimostrare che defínto $np = \lambda$ tale probabilità è ben approssimata dalla distribuzione di Poisson. Quest'ultima è definita su uno spazio di probabilità con infiniti eventi elementari indicizzati da un numero intero k. La probabilità dell'evento k è data da:

$$p_\lambda(k) \;=\; \frac{\lambda^k e^{-\lambda}}{k!}\,. \tag{83}$$

Osserviamo anzitutto che valendo per la funzione esponenziale la formula $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^\lambda$ otteniamo che il primo assioma è verificato:

$$\sum_{k=0}^{\infty} p_\lambda(k) \;=\; e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \;=\; e^{-\lambda}e^{\lambda} \;=\; 1\,. \tag{84}$$

Calcoliamo media e varianza per la distribuzione di Poisson.

$$E(f) \,=\, \sum_{k=0}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} \,=\, \sum_{k=1}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} \,=\, \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} \tag{85}$$

e ponendo $l = k - 1$

$$E(f) \,=\, \lambda \sum_{l=0}^{\infty} \frac{\lambda^l e^{-\lambda}}{l!} \,=\, \lambda\,. \tag{86}$$

Analogamente

$$E(f^2) \,=\, \sum_{k=0}^{\infty} k^2\frac{\lambda^k e^{-\lambda}}{k!} \,=\, \lambda \sum_{l=0}^{\infty} (l+1)\frac{\lambda^l e^{-\lambda}}{l!} \,=\, \lambda^2 + \lambda\,, \tag{87}$$

da cui

$$V(f) \,=\, \lambda\,. \tag{88}$$

DISUGUAGLIANZA DI CHEBYSHEV E LEGGE DEI GRANDI NUMERI

In questa paragrafo ci proponiamo di dare un significato preciso all' affermazione che la frequenza empirica tende alla probabilità teorica in un grande numero di lanci della moneta. Partiamo da due semplici disuguaglianze. La prima dice che per due variabili aleatorie tali che $f_i \leq g_i$ si ha

$$\sum_i f_i p_i \leq \sum_i g_i p_i\;. \tag{89}$$

La seconda che per una coppia di eventi in cui uno contiene l'altro $A \subset B$ si ha

$$P(A) \leq P(B) \qquad \textit{equivalentemente} \qquad \sum_{i\in A} p_i \leq \sum_{i\in B} p_i\;, \tag{90}$$

che deriva dall'assioma di positività delle p_i (la seconda somma è più grande perchè si sommano più termini positivi). Ci proponiamo di stimare la quantità

$$P(|f - E(f)| \geq \delta) \ , \tag{91}$$

cioè la probabilità che una funzione scarti dalla propria media per più di una quantità assegnata δ. Per definizione avremo

$$P(|f - E(f)| \geq \delta) = \sum_{i,|f_i - E(f)| \geq \delta} p_i \leq \sum_{i,|f_i - E(f)| \geq \delta} \frac{(f_i - E(f))^2}{\delta^2} p_i \tag{92}$$

dove abbiamo utilizzato la (89). Inoltre per la (90) abbiamo

$$\sum_{i,|f_i - E(f)| \geq \delta} \frac{(f_i - E(f))^2}{\delta^2} p_i \leq \sum_i \frac{(f_i - E(f))^2}{\delta^2} p_i = \frac{1}{\delta^2} V(f) \ . \tag{93}$$

Definendo la *deviazione standard* come $\sigma(f) = \sqrt{V(f)}$ la precedente ha una interpretazione particolarmente efficace quando si sceglie $\delta = l\sigma(f)$. Si ha infatti (disuguaglianza di Chebyshev)

$$P(|f - E(f)| \geq l\sigma(f)) \leq \frac{1}{l^2} \tag{94}$$

cioè la probabilità che una variabile aleatoria scarti dal suo valor medio per pi di 2 sigma è minore di 1/4 , per più di 3 sigma è minore di 1/9 e cosi via.

Una importante applicazione della precedente si ha per la variabile aleatoria binomiale $f_k = k/n$. Anzitutto osserviamo da un calcolo diretto che

$$E(\frac{k}{n}) = p \qquad V(\frac{k}{n}) = \frac{p(1-p)}{n} \ , \tag{95}$$

da cui la (93) fornisce una stima della probabilità che la frequenza $\frac{k}{n}$ delle teste si discosti dalla probabilità p per una quantità piccola

$$P(|\frac{k}{n} - p| \geq \varepsilon) \leq \frac{p(1-p)}{\varepsilon^2 n} \ , \tag{96}$$

quantità questa che tende a zero all' aumentare del numero di lanci. La formula sopra scritta (legge dei grandi numeri) merita alcuni commenti. Essa fa un' affermazione che vale nello spazio di probabilità degli n-lanci e non del singolo lancio! I singoli lanci sono indipendenti e i lanci passati non possono in alcun modo influenzare quelli futuri.

Sommario:

- Media

$$E(f) \;=\; \sum_{i=1}^{n} f_i p_i \,,$$

- Varianza

$$V(f) \;=\; E([f - E(f)]^2) \;=\; E(f^2) - E(f)^2 \;=\; \sum_{i=1}^{n} f_i^2 p_i - (\sum_{i=1}^{n} f_i p_i)^2 \,,$$

- Distribuzione Binomiale

$$P_n(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{(n-k)}$$

- Distribuzione di Poisson

$$P_\lambda(k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{97}$$

- Disuguaglianza di Chebyshev

$$P(|f - E(f)| \geq l\sigma(f)) \leq \frac{1}{l^2} \tag{98}$$

- Legge dei grandi numeri

$$P(|\frac{k}{n} - p| \geq \varepsilon) \leq \frac{p(1-p)}{\varepsilon^2 n} \tag{99}$$

DISTRIBUZIONI CONTINUE

In una distribuzione di probabilità continua gli eventi non sono numerabili (non possono essere messi in corrispondenza biunivoca con gli interi). Un esempio è la distribuzione di probabilità uniforme nellintervallo $[0, 1]$. Si dice che una variabile aleatoria ξ ha distribuzione uniforme in tale intervallo se per ogni $0 \leq a < b \leq 1$ si ha

$$P(a \leq \xi \leq b) = \int_a^b dx = b - a \; . \tag{100}$$

Più in generale si dice che una variabile aleatoria ζ ha *densità di probabilità* $p(x)$ nellintervallo $[0, 1]$ se vale la condizione di positività

$$p(x) \geq 0 \; , \tag{101}$$

e la condizione di normalizzazione

$$\int_0^1 p(x) dx = 1 \; . \tag{102}$$

Esercizi:

- Si dica quali tra le seguenti funzioni sono distribuzioni di probabilità nell'intervallo $[0, 1]$: $2x$, $sen(x)$, $1 - x$.
- Si diano esempi che violano la positività, che violino la normalizzazione o che violino entrambe.
- Si calcoli il valore della costante c che rende le seguenti funzioni delle distribuzioni di probabilità in $[0, 1]$: cx^3, $c(x^2 - 1)$

Notiamo che la distribuzione uniforme è un caso particolare del precedente definito da $p(x) = 1.$.

Per le distribuzioni continue valgono le stesse definizioni di media e varianza date per quelle discrete in cui si utitlizza l'integrale al posto delle somme. Data la funzione f dell'intervallo $[0,1]$ (variabile aleatoria) si definisce

$$E(f) = \int_0^1 f(x)p(x)dx \tag{103}$$

$$E(f^2) = \int_0^1 f(x)^2 p(x)dx \ , \tag{104}$$

$$V(f) = E([\zeta^2 - E(\zeta)]^2) \tag{105}$$

Ovviamente le definizioni date per variabili nell' intervallo unitario si estendono a qualsiasi intervallo finito, e vedremo anche alla retta reale.

Esercizi:

- La variabile aleatoria $f(x) = x$ ha densità di probabilità $p(x) = 1 - x/2$ nellintervallo $[0,2]$. Si verifichi che la $p(x)$ data è una funzione densità di probabilità e si calcoli la media e la varianza di f.

- La variabile aleatoria $f(x) = x$ ha densità di probabilità cx^4 nellintervallo [0,1] Si calcoli il valore di c e la media e la varianza di f.

- La variabile aleatoria $f(x) = 3x$ ha distribuzione di probabilità costante nellintervallo $[0,3]$. Si calcoli il valore della costante, la media e la varianza di f.

La distribuzione di Gauss è definita (nellintervallo reale) dalla densità di probabilità

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \ . \tag{106}$$

Per tale distribuzione valgono le seguenti formule

$$\int_{-\infty}^{\infty} \quad \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = 1 \tag{107}$$

$$\int_{-\infty}^{\infty} \quad \frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = 0 \tag{108}$$

$$\int_{-\infty}^{\infty} \quad \frac{x^2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = 1 \ , \tag{109}$$

da cui si può dedurre che per la Gaussiana di media μ e varianza σ^2 di equazione

$$p_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{110}$$

valgono le

$$\int_{-\infty}^{\infty} \quad p_{\mu,\sigma}(x) = 1 \tag{111}$$

$$\int_{-\infty}^{\infty} \quad x p_{\mu,\sigma}(x) = \mu \tag{112}$$

$$\int_{-\infty}^{\infty} \quad (x-\mu)^2 p_{\mu,\sigma}(x) = \sigma^2 \ . \tag{113}$$

È interessante vedere come si trasforma la curva $p_{\mu,\sigma}(x)$ al varioare di μ e σ. Quando μ aumenta la curva si sposta in modo rigido verso destra, quando diminuisce verso sinistra. Al variare di σ invece cambia la *larghezza* della curva: per σ piccoli la curva è molto alta e scende a zero in modo molto rapido. Per σ grandi la curva invece è sparpagliata.

Bibliografia

[1] Mathematics and the Genome. http://mathforum.org/mam/02/

[2] William P. Thurston: "On Proof and Progress in Mathematics" Bulletin of the American Mathematical Society 30 (2), 161-177, (1994).

[3] William P. Thurston: "Mathematical Education." Notices of the American Mathematical Society 37, no. 7, September (1990): 844-850.

[4] D.Epstein and S. Levy: "Experimentation and Proof in Mathematics" Notices of the American Mathematical Society 42, no. 6, June (1995): 670-674