1.2 Un modello binomiale

Proviamo a modellizzare la situazione proposta mediante la *distribuzione binomiale*. Precisamente interpretiamo la questione nel modo seguente:

- una sequenza[4] di $\mathbf{n} = \mathbf{10.000}$ prove, ciascuna delle quali corrisponde ad una persona;
- ciascuna prova ha due soli esiti possibili: *"la persona vota A"* oppure *"la persona vota B"*; dall'esito del sondaggio stimiamo che la probabilità che l'individuo sia favorevole ad A è $\mathbf{p} = \mathbf{0.4}$. Essa è costante (cioè è la stessa) per ogni prova (individuo).

Affinché abbia senso modellizzare mediante una distribuzione binomiale la v.a. F, assumiamo inoltre che le prove siano indipendenti[5], ossia che il voto della persona non sia influenzato dal voto delle altre.

Secondo il nostro modello la probabilità che il numero di individui F sia, ad esempio, 3900 è[6]

$$P(F = 3900) = \binom{10000}{3900} \cdot 0,4^{3900} \cdot 0,6^{6100}$$

e la probabilità richiesta si può esprimere nella forma

$$P(\mathbf{3900} \leq \mathbf{F} \leq \mathbf{4100}) = P(F = \mathbf{3900}) + P(F = \mathbf{3901}) + ... + P(F = \mathbf{4100})$$

Osservazione. Senza ricorrere a formule ritenute a memoria, il valore di probabilità $P(F = 3900)$ si determina volta per volta in due passi, determinando:

1. la probabilità di una sequenza di 3900 voti a favore di A e 6100 voti a favore di B, quale

$$\underbrace{AAA...A}_{3900 \text{ volte}}\underbrace{BBB...B}_{6100 \text{ volte}}$$

   per la legge della moltiplicazione (eventi indipendenti) la probabilità di tale sequenza è

$$0,4^{3900} \cdot (1 - 0,4)^{6100}$$

2. il numero di tutte le sequenze di 3900 A e 6100 B
   esso è il numero di sottoinsiemi di 3900 *elementi (posizioni per A), contenuti in un insieme di* 10000 *elementi (posizioni possibili nella sequenza); dunque è*

$$\binom{10000}{3900}$$

[4]Tale sequenza si indica spesso come "schema di Bernoulli" o "schema successo-insuccesso".

[5]Non possiamo dire che ciò avvenga realmente. Si tratta solamente di una nostra ipotesi che assumiamo, almeno in prima approssimazione, allo scopo di costruire un modello ragionevolmente semplice.

[6]Ricordiamo che la distribuzione di una variabile aleatoria binomiale S è data da

$$P(S = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

C'è un problema

Il calcolo richiesto è assai articolato per la presenza dei coefficienti binomiali. Basti pensare che, ad esempio, $70! = 1,1979 \cdot 10^{100}$.
Come disse De Moivre nel Settecento, effettuare calcoli analoghi *"non è possibile senza un lavoro immenso, per non dire che è impossibile"*. E' vero che erano altri tempi e all'epoca non si disponeva dei calcolatori di oggi, ma anche noi siamo interessati a trovare un modo più efficiente per risolvere il problema.

C'è un approccio meno dispendioso computazionalmente?