2 Il modello binomiale

2.1 Il centralino - attività

In questa sezione ci occupiamo di risolvere il problema di conteggio posto nel paragrafo precedente. Ma, prima di affrontarlo nel caso generale, è meglio prendere gradualmente confidenza con il problema ed esaminare il **quesito specifico** proposto al termine del video *Modellizzazione di eventi "rari"*.

Inoltre inizialmente proviamo a modellizzare la situazione mediante una distribuzione che gli studenti già conoscono: la distribuzione binomiale, che preciseremo, per il docente, nel paragrafo successivo al punto b) approfondimenti teorici. Puntiamo così a "**radicare con continuità**" i nuovi saperi su quelli pregressi. Altrimenti il rischio è che gli studenti riescano ad utilizzare gli strumenti matematici introdotti, ma siano in grado di farlo solo localmente e per un periodo limitato di tempo; e poi, a lungo termine è come non li avessero mai visti. Di questo avviso è A. Sfard [Sfa, p.294] secondo la quale "*sembra che la via più sicura per arrivare a nuove esplorazioni sia quella di presentarle come (possibili) miglioramenti di atti familiari*".

> *Modalità di utilizzo dei materiali*
> Gli studenti sono invitati a provare a risolvere autonomamente il problema. In caso di difficoltà, trovano un aiuto nel suggerimento proposto ed eventualmente anche nella soluzione scritta; in ogni caso, è fondamentale che si confrontino con la risoluzione proposta e poi discutano in classe l'attività svolta individualmente. Utilizzeremo anche in seguito questa modalità di lavoro che possiamo sintetizzare con lo schema "quesito-suggerimento-risoluzione".

Osservazione
Nel materiale, l'indicazione principale è di utilizzare il modello binomiale; invece, la scelta di considerare una sequenza di $n = 100$ prove, è solo una tra le altre possibili. Ciò che importa è che n sia sufficientemente grande per poter assumere che in un intervallino di tempo arrivi al più una sola telefonata.

2.2 Il caso generale

a) Aspetti didattici

A questo punto gli studenti dovrebbero riuscire ad apprezzare una prima modellizzazione del problema nel caso generale: essa ripercorre i passi già seguiti nell'esempio e in particolare il ricorso al modello binomiale.

Pertanto uno dei passi da compiere è stabilire il valore da attribuire ai due parametri che caratterizzano la distribuzione. Fissato il generico numero n di prove, per determinare la probabilità p di successo nella singola prova, scegliamo di seguire un approccio frequentista. Ossia poniamo

$$p = \frac{\#\ \textit{realizzazioni evento}}{\#\ \textit{prove}}.$$

La vera scelta, in realtà, risiede nel seguire un approccio intuitivo e non una via formale che, come vedremo, si basa sul significato del valore atteso, ma che è didatticamente meno efficace. Possiamo indicare questo procedimento come una "**reinvenzione guidata**", per usare il termine introdotto da Freudenthal [Fre, p.72]: con la guida del docente gli studenti ricostruiscono gli oggetti e i concetti matematici[1] sottesi, quasi come li stessero inventando essi stessi.

In questo contesto prestiamo attenzione alla **modellizzazione**, in linea con gli obiettivi specifici di apprendimento previsti dalle Indicazioni Nazionali per la classe quinta del secondo ciclo d'istruzione:

"*In relazione con le nuove conoscenze acquisite, anche nell'ambito delle relazioni della matematica con altre discipline, lo studente approfondirà il concetto di modello matematico e svilupperà la capacità di costruirne e analizzarne esempi*".

Per questo investiamo del tempo per **valutare** l'espressione analitica a cui giungiamo, in termini di efficienza e espressività.

Inoltre discutiamo le **ipotesi** da assumere affinché abbia senso utilizzare il modello in esame. Come vedremo, in ultima analisi, ciò comporta **prendere decisioni**. Tale aspetto non è sempre esplicitato nei libri di testo che, anzi, sembrano spesso suggerire l'idea che ci sia un modo unico, assoluto, di schematizzare una situazione.
L'attività di riflessione sulle decisioni che si devono prendere quando si fa matematica a scuola è resa ancora più urgente dagli studi condotti da R. Zan [Zan07, p.127-129] secondo i quali gli studenti percepiscono la matematica come una disciplina nella quale non si debbano prendere decisioni.

Modalità di utilizzo dei materiali
Il materiale di questo paragrafo è progettato e realizzato come supporto alla lezione in classe, che prefiguriamo di tipo partecipato. Pertanto non sono indicati per un lavoro unicamente individuale e autonomo. Ciò si può indicare sinteticamente con il termine dispensa.

b) Approfondimenti teorici - la distribuzione binomiale

Nei materiali per lo studente utilizziamo più volte la distribuzione binomiale. La presentiamo agli studenti come una distribuzione di probabilità che nasce da una sequenza di prove svolte nelle stesse condizioni e tra loro indipendenti.
Più formalmente essa si può costruire per passi come di seguito. Riteniamo che allo studente vada presentata una opportuna declinazione di tale schema, che ne espliciti la sostanza senza risultare però troppo tecnica e suggeriamo di riferirsi a [Bon, par.3.1].

Cominciamo col definire la *variabile aleatoria di Bernoulli*. Essa è l'oggetto matematico che rappresenta la singola prova della sequenza.

[1] Ricordiamo che la costruzione di significati è uno dei criteri che abbiamo stabilito per il nostro percorso.

Definizione 2.1. *Si dice variabile aleatoria di* ***Bernoulli*** *di parametro $p \in [0,1]$ la variabile aleatoria che assume valore 0 oppure 1 e ha distribuzione discreta definita da:*

$$P(X=1)=p$$
$$P(X=0)=1-p$$

Convenzionalmente si indica con il termine "successo" l'evento che si realizza con probabilità p e con "insuccesso" l'evento complementare.
Tale variabile aleatoria si denota spesso con $X \sim Be(p)$.

Molte situazioni si possono modellizzare mediante una sequenza di variabili aleatorie di Bernoulli; tale sequenza rappresenta formalmente lo schema di prove ripetute a cui accennavamo all'inizio del paragrafo. Precisamente si definisce nel modo seguente.

Definizione 2.2. *Si definisce "**schema di Bernoulli**", o "schema successo-insuccesso", una sequenza $X_1, X_2, ..., X_n$ di variabili aleatorie di Bernoulli (ossia $X_i \sim Be(p)$ per $i = 1, 2, ..., n$) indipendenti e aventi lo stesso parametro p.*

In questo contesto la questione centrale è determinare la probabilità di avere $k = 0, 1, 2, ...$ successi sulle n prove dello schema. Il problema si esprime rigorosamente mediante una opportuna variabile aleatoria e una distribuzione di probabilità.

Definizione 2.3. *Sia S_n la variabile aleatoria che conta il numero di successi in uno "schema di Bernoulli" costituito da n prove ciascuna con probabilità di successo p. Tale variabile si dice variabile aleatoria **binomiale** e si denota spesso con $S_n \sim B(n,p)$.*
*La distribuzione di probabilità di S_n si dice **distribuzione binomiale** di parametri n e p e si osserva essere la funzione p definita da $p(x) = P(S_n = x)$ per $x = 0, 1, 2, ..., n$.*

Osserviamo che S_n si può esprimere nella forma

$$S_n = X_1 + X_2 + ... + X_n$$

dove $X_i \sim Be(p)$ per $i = 0, 1, 2, ..., n$ e le variabili aleatorie X_i sono indipendenti tra loro.

Proposizione 2.4. *La distribuzione binomiale di parametri n e p ed è data da:*

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \textit{per } k = 0, 1, 2, ...$$

Dimostrazione. Per $k = 0, 1, 2, ...$ vogliamo esprimere la probabilità $P(S_n = k)$ in termini dei parametri della distribuzione .
Ovvero vogliamo calcolare la probabilità di ottenere esattamente k "successi" sulle n prove totali (quindi le restanti $n-k$ prove sono "insuccessi").
Iniziamo osservando che ogni sequenza di k "successi" su n prove si può ottenere in $\binom{n}{k}$ modi, dove il simbolo $\binom{n}{k}$ denota il coefficiente binomiale "n su k" e vale $\dfrac{n!}{k!(n-k)!}$.
Infatti ogni sequenza del tipo richiesto si può interpretare come un sottoinsieme di k elementi, contenuto in un insieme di n elementi, precisamente tale sottoinsieme è costituito dalle posizioni dei k successi nella sequenza delle n posizioni che identificano le prove dello schema di Bernoulli.

Proseguiamo osservando che le variabili aleatorie X_i sono indipendenti tra loro e la probabilità di "successo" per ciascuna X_i è p mentre quella di "insuccesso" è $(1-p)$. Pertanto, per la legge del prodotto, si conclude che vale:

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{per } k = 0, 1, 2, ...$$

□

Le variabili aleatorie si possono descrivere sinteticamente mediante opportuni indici; tra questi vi sono il valore atteso e la varianza. Ne faremo uso nel percorso didattico che proponiamo, pertanto ne ricordiamo prima le definizioni per una generica variabile aleatoria discreta, ossia una variabile aleatoria che assume un insieme finito o numerabile di valori.

Definizione 2.5. *Sia X una variabile aleatoria discreta e sia $p(x)$ la sua distribuzione discreta di probabilità (ossia p è la funzione definita da $p(x) = P(X = x)$ per ogni valore x che può assumere la variabile X).*
*Si dice **valore atteso** (o media) di X il numero*

$$E[X] = \sum_x x \cdot p(x)$$

dove la somma è estesa a tutti i valori di X.
Se la variabile aleatoria è numerabile, tale somma è una serie e si richiede che essa sia assolutamente convergente, ossia che

$$\sum_x |x| p(x) < +\infty$$

Definizione 2.6. *Indicato con μ il valore atteso della variabile aleatoria discreta X, si dice **varianza** di X il numero*

$$V[X] = E[(X-\mu)^2]$$

Osservazioni

- per definizione di valore atteso si ha

$$V[X] = \sum_x (x-\mu)^2 p(x)$$

- inoltre, si può dimostrare che vale la seguente uguaglianza, utile nel calcolo

$$V[X] = E[X^2] - \mu^2.$$

Proposizione 2.7. *Sia $S_n \sim B(n,p)$, allora il valore atteso di S_n è np e la sua varianza è $np(1-p)$.*

Dimostrazione. Poiché S_n si può esprimere nella forma $S_n = X_1 + X_2 + ... + X_n$, dove $X_i \sim Be(p)$ per $i = 1, 2, ..., n$ iniziamo col determinare il valore atteso e la varianza delle variabili "elementari" X_i.

- $E[X_i] = 1 \cdot p + 0 \cdot (1-p) = p$

- $V[X_i] = E[{X_i}^2] - (E[X_i])^2 = 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 = p(1-p)$

A questo punto per ricondursi agli indici della variabile aleatoria S_n, basta ricorrere alla linearità del valore atteso e della varianza. E precisamente alle due proprietà che seguono: se X, Y sono due variabili aleatorie e $a \in \mathbb{R}$ vale

$$E[aX + Y] = E[aX] + E[Y] = aE[X] + E[Y]$$

Inoltre se X e Y sono due variabili aleatorie indipendenti che ammettono valore atteso, allora si ha:

$$V[X + Y] = V[X] + V[Y]$$

In realtà tali uguaglianze, formulate nel caso di due variabili aleatorie, si estendono alla somma di n variabili aleatorie.
Così:

- $E[S_n] = E[X_1 + X_2 + ... + X_n] = E[X_1] + E[X_2] + ... + E[X_n] = np$
- e osservando che le X_i sono indipendenti si ha

$$V[S_n] = \sum_{i=1}^{n} V[X_i] = np(1-p).$$

□

Osservazione
Alla luce di queste considerazioni formali possiamo dedurre in modo rigoroso il valore di p ricavato nel paragrafo 3.2.2 utilizzando un approccio frequentista.
Sia S_n la variabile aleatoria binomiale in esame, allora vale $E[S_n] = np$ per la proposizione 2.2.7. D'altra parte, per il significato di λ come numero medio di successi nell'intervallo $[0, t]$, vale

$$\lambda = E[S_n] = np$$

pertanto si conclude che deve essere

$$p = \frac{\lambda}{n}.$$