PROBABILITÀ – SCHEDA N. 6
LE VARIABILI ALEATORIE DI BERNOULLI E BINOMIALE

In molte situazioni si è interessati a verificare se una determinata caratteristica si presenta oppure no (l'efficacia di un vaccino, il manifestarsi di una malattia, la difettosità di un pezzo, ...). Ciò corrisponde ad un esperimento con solo due possibili esiti (detto anche **dicotomico**), che può essere modellato con una variabile aleatoria Y che assume valore 1 (successo) con probabilità p e valore 0 (insuccesso) con probabilità 1-p, con 0<p<1.

$$Y = \begin{cases} 1 & \text{con probabilità} \quad p \\ 0 & \text{con probabilità} \quad 1-p \end{cases}$$

Una variabile aleatoria di questo tipo è chiamata di **Bernoulli di parametro p.**

Il valore atteso e la varianza di Y sono

$$E(Y) = 0 \times (1-p) + 1 \times p = p$$
$$Var(Y) = E(Y^2) - (E(Y))^2 = (0^2 \times (1-p) + 1^2 \times p) - p^2 = p\,(1-p)$$

È modellabile con una variabile aleatoria di Bernoulli, ad esempio, il lancio di una moneta, non necessariamente equilibrata (*p* è la probabilità che esca ad esempio *T*).

Ora consideriamo la ripetizione di *n* esperimenti *indipendenti* ciascuno dei quali è modellabile con una variabile aleatoria di Bernoulli con probabilità di successo *p*. Per avere il numero di successi in questi n esperimenti basta sommare i valori delle n variabili aleatorie di Bernoulli, consideriamo quindi la variabile aleatoria

$$X = \sum_{i=1}^{n} Y_i \qquad \text{con } Y_i \text{ v.a. indipendenti di Bernoulli di parametro } p$$

I possibili valori assunti da *X* sono 0, 1, ... , n. Vogliamo conoscere le probabilità con cui tali valori sono assunti, cioè vogliamo costruire la densità di probabilità di *X*.

ESEMPIO 1. Riprendiamo l'esempio della moneta; i singoli lanci sono indipendenti fra di loro. Siamo interessati al numero di uscite di testa in 10 lanci; quindi n = 10. Qual è la probabilità di ottenere 6 uscite di testa?
Consideriamo prima una prefissata sequenza di teste e croci; ad esempio, qual è la probabilità di ottenere la sequenza

T T C T T C C T T C

Essendo i lanci indipendenti, la probabilità cercata è data da

$$p\ p\ (1-p)\ p\ p\ (1-p)\ (1-p)\ p\ p\ (1-p) = p^6\ (1-p)^4.$$

Si capisce subito che la probabilità di avere 6 teste (e quindi 4 croci) anche in posizioni diverse risulta la stessa. Allora la probabilità di avere *una determinata* sequenza di k teste e n-k croci in n lanci sarà

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

Se ora volessimo calcolare la probabilità di *tutte* le possibili sequenze di 10 lanci in cui ci siano 6 teste dobbiamo stabilire *quante* sono le sequenze che presentano 6 teste e 4 croci. Il problema è equivalente a contare tutti i modi in cui è possibile scegliere di mettere le 6 teste nei 10 lanci. Senza entrare nei dettagli del calcolo diciamo che questo numero è $\dfrac{10!}{6!(10-6)!}$ che si indica con $\dbinom{10}{6}$ e si chiama *coefficiente binomiale.*

Allora la probabilità di ottenere 6 teste lanciando la moneta 10 volte sarà:

$$\binom{10}{6} p^6\ (1-p)^4 \quad .$$

In generale, il **coefficiente binomiale** vale: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Un modo più veloce per calcolarlo è:

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \quad .$$

Parliamo di **esperimento binomiale** quando consideriamo un esperimento casuale in cui

1. interessa esclusivamente il successo (codificato con 1) o l'insuccesso (codificato con 0)
2. il successo ha probabilità $\boldsymbol{p}$
3. si effettuano $\boldsymbol{n}$ prove indipendenti.

Se X è la variabile che indica il numero di successi in n prove di un esperimento binomiale, avremo che la probabilità che X assuma il valore k è:

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

La variabile casuale X così definita è detta **variabile casuale binomiale di parametri n** (numero di prove) e **p** (probabilità di successo in una prova) e si indica con

$$X \sim B(n,p)$$

Di seguito riportiamo alcuni grafici delle funzioni di densità per n = 20 e valori diversi di p.

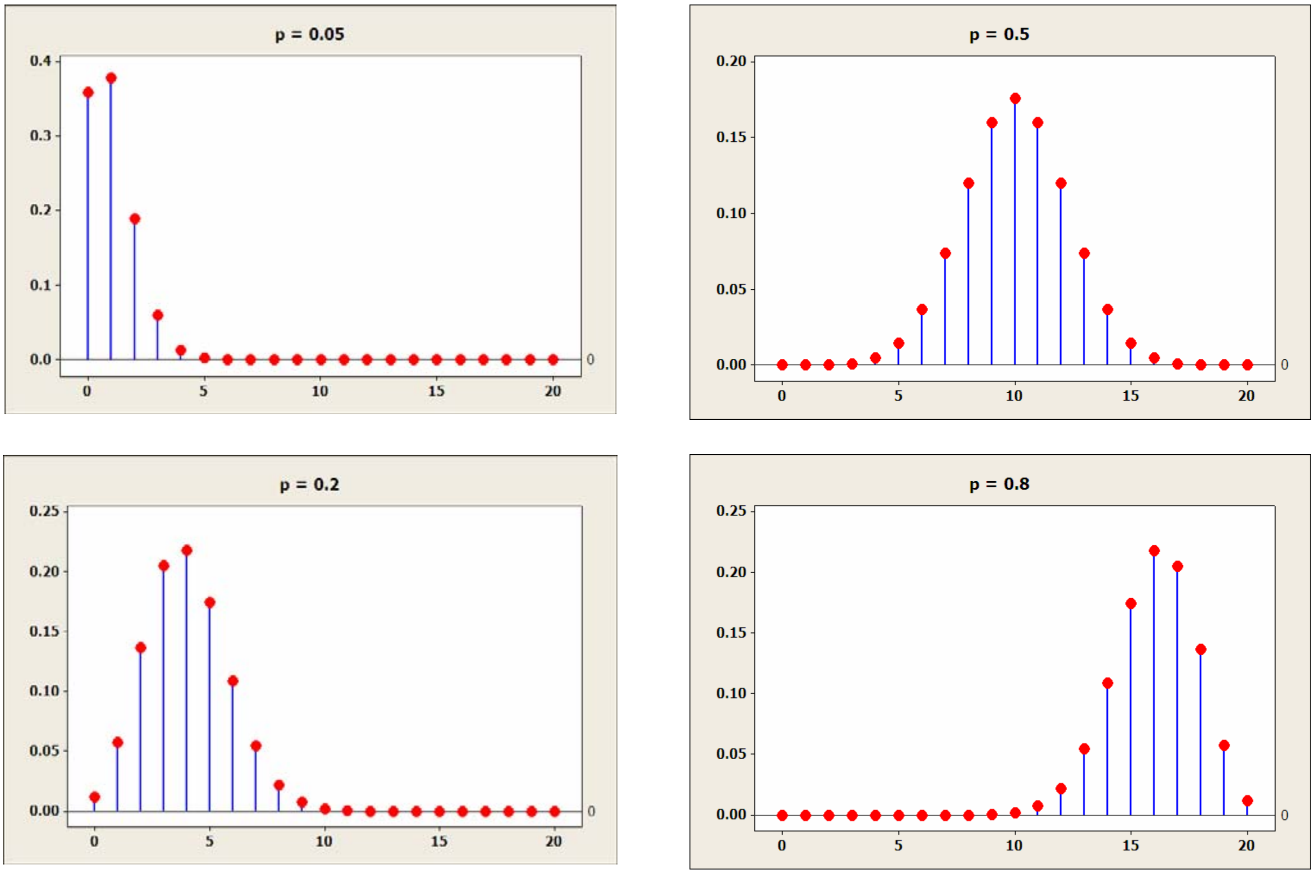

Figura 1

Si può osservare che

- il grafico è simmetrico quando p=0.5;
- il grafico è più concentrato a sinistra quando p assume valori bassi (<0.5);
- i grafici con p=0.2 e 0.8 sono simmetrici l'un l'altro rispetto alla retta verticale passante per n/2;
- ogni grafico ha solo un massimo che è vicino al valore atteso della variabile aleatoria.

La variabile aleatoria di Bernoulli che rappresenta l'esito di una sola prova è quindi una binomiale di parametri 1 e p. Come abbiamo già detto, siccome X rappresenta la somma dei successi in n prove, X può essere interpretata come somma di n variabili aleatorie binomiali B(1,p) indipendenti:

$$X = \sum_{i=1}^{n} Y_i \qquad \text{con } Y_i \text{ v.a. indipendenti di Bernoulli di parametro } p$$

Utilizzando le proprietà del valore atteso e della varianza della somma di variabili aleatorie indipendenti, abbiamo:

$$E(X) = E\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} p = np$$

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \mathrm{Var}(Y_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Ricordiamo che il valore atteso della somma è sempre uguale alla somma dei valori attesi, mentre la varianza della somma è uguale alla somma delle varianze solo se le variabili sono indipendenti.

Più in generale si può dimostrare che la somma di due binomiali indipendenti X_1 e X_2 con densità $B(n_1,p)$ e $B(n_2,p)$ rispettivamente è ancora una densità binomiale $B(n_1+n_2,p)$.

ESEMPIO 2. Il rapporto dei sessi nella specie umana alla nascita è di 105 femmine su 100 maschi. Qual è la probabilità che in 6 nascite singole almeno la metà dei neonati siano di sesso femminile?
Dai dati si desume che la probabilità della nascita di una femmina è p=105/205. Se indichiamo con X il numero di neonati di sesso femminile nelle 6 nascite prese in esame, la variabile aleatoria X avrà una densità binomiale B(6, 105/205).
Vogliamo calcolare

$$\begin{aligned} P(X \geq 3) &= P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - \left(\frac{100}{205}\right)^6 - \binom{6}{1}\left(\frac{105}{205}\right)^1\left(\frac{100}{205}\right)^5 - \binom{6}{2}\left(\frac{105}{205}\right)^2\left(\frac{100}{205}\right)^4 = 0.68 \end{aligned}$$

Come varierebbe la probabilità se si considerassero 60 nascite singole? Sarebbe uguale alla precedente? (Rispondi intuitivamente senza fare i calcoli. Vedremo la risposta in una scheda successiva)

ESEMPIO 3. L'incubatrice di un allevamento di polli deve mantenere una temperatura che permetta la schiusa delle uova; per far ciò devono funzionare contemporaneamente almeno 5 e non più di 9 apposite resistenze.
Ciascuna resistenza ha una probabilità di essere funzionante per tutto il periodo necessario alla schiusa pari a 0.85; le resistenze funzionano l'una indipendentemente dall'altra. Supponiamo inoltre che non intervengano altri fattori che incidano sul funzionamento dell'incubatrice.
Si vuole stabilire qual è il numero minimo di resistenze che occorre attivare per avere una probabilità maggiore del 95% che almeno 5 resistenze restino sempre funzionanti.

Indichiamo con X_n una variabile aletoria binomiale di parametri n (con n = 5, 6, 7, 8, 9) e p=0.85 che rappresenta il numero di resistenze funzionanti con n attivate.
Bisogna trovare il più *piccolo n* per cui $P(X_n \geq 5) > 0.95$.
Osserviamo anzitutto che $P(X_n \geq 5) = 1 - P(X_n \leq 4)$; quindi bisogna trovare il più *piccolo n* per cui

$$1 - P(X_n \leq 4) > 0.95 \qquad \text{ovvero} \qquad P(X_n \leq 4) < 0.05$$

Risolviamo il problema calcolando la funzione di distribuzione cumulata di X_n nel valore 4, al variare di n, con un opportuno software statistico, in questo caso Minitab.

	n = 5	n = 6	n = 7	n = 8	n = 9
$F_{X_n}(4)$	0.556295	0.223516	0.0737651	0.0213525	0.00562866

Possiamo quindi concludere che bisogna attivare almeno 8 resistenze.

ESERCIZI

ESERCIZIO 1. Una caratteristica A è presente nel 10% della popolazione 1 e nel 15% della popolazione 2. Con un esperimento di tipo binomiale (con ripetizione) si estraggono 30 individui di cui 10 dalla popolazione 1 e 20 dalla popolazione 2.
a) Qual è la probabilità che un individuo fra i 30 estratti abbia la caratteristica A?
b) Qual è la probabilità che tre o più abbiano la caratteristica A?
c) Qual è il numero medio di individui con la caratteristica A?
d) Qual è la varianza del numero di individui con la caratteristica A?

ESERCIZIO 2. A una lotteria i biglietti in vendita sono 500, metà dei quali assicurano una vincita di una bottiglia di vino; l'altra metà non assicura nessun premio. Pippo acquista 10 biglietti a 6 euro l'uno, contando di vendere a Topolino a 20 euro l'una le bottiglie eventualmente vinte. Valutare la probabilità che Pippo ci rimetta almeno 20 euro.

ESERCIZIO 3. È stata svolta un'indagine per verificare l'utilizzo di un antiparassitario A, non permesso, nella coltivazione di arance. In generale è noto che se viene utilizzato A mediamente marcisce il 5% della produzione di arance; se viene usato l'antiparassitario B, permesso, marcisce il 10%. È noto che il 20% degli agricoltori utilizzano A ed i restanti utilizzano B. Si esaminano 100 arance e si trovano 4 frutti marci.
a) Qual è mediamente la percentuale delle arance che marcisce, essendo stato usato uno dei due antiparassitari?
b) Qual è la probabilità che marciscano 4 arance su 100 se è stato usato l'antiparassitario A?
c) Qual è la probabilità che marciscano 4 arance su 100 se è stato usato l'antiparassitario B?
d) Qual è la probabilità totale che marciscano 4 arance su 100, essendo stato usato uno dei due antiparassitari?
e) Qual è la probabilità che sia stato utilizzato l'antiparassitario A, avendo trovato 4 arance marce su 100 esaminate?

ESERCIZIO 4. Il batterio Cloristridium botulinum, responsabile della malattia mortale detta botulismo, è presente, se non si usa un opportuno trattamento, circa nell'1% di un prodotto alimentare inscatolato. Il trattamento, se effettuato, garantisce la totale assenza del batterio. Un ricercatore ignora se su una partita di cibo inscatolato sia stato effettuato il trattamento. Decide di assumere come equiprobabili i fatti che il trattamento sia stato effettuato oppure no. Mediante un'analisi in grado di rilevare senza errori la presenza o meno del batterio in ciascuna scatola, esamina 100 scatole ed in nessuna trova il batterio.
a) Quanto vale la probabilità del risultato sperimentale riscontrato nell'ipotesi che il trattamento non sia stato effettuato?
b) Preso atto del risultato sperimentale riscontrato, quanto vale la probabilità che il trattamento sia stato effettuato?
c) Si ritiene che, alla luce del risultato del punto precedente e della gravità della malattia, si sia pervenuti a una situazione di sufficiente garanzia dell'assenza del batterio?
In caso contrario che cosa si ritiene opportuno fare.

Esperienze

E1. Esercizi con il calcolatore sulle funzioni di probabilità e di distribuzione cumulata esatte

A Disegnare diversi grafici relativi alla densità B(10,p) facendo variare p. Cosa si osserva per p=0.5? (commentare) Cosa si può dire confrontando i grafici di B(10,0.2) e B(10,0.8)? (commentare)

B Si sa che il 70% di una certa varietà di bulbi di fiori fiorirà. Se si piantano 10 bulbi.

1) Qual è la probabilità che:
   a) esattamente 3 fioriranno
   b) al più 6 fioriranno
   c) almeno 2 fioriranno.

(risolvere l'esercizio usando la funzione di distribuzione cumulata)

2) Quanto deve valere k, affinché la probabilità che fioriscano almeno k fiori piantando 100 bulbi sia maggiore dell'80%?

C In un paese, il 35% degli elettori è a favore del candidato A. In un seggio votano 200 persone. Calcolare la probabilità che in quel seggio voti A non più del 70% degli elettori.
Qual è il numero di voti a favore di A più probabile?

D Data X una variabile aleatoria con densità binomiale B(20,0.5).

a) Determinare k tale che $P(X>k) \leq 0.80$.

b) Utilizzando la simmetria del grafico della densità di X, per quali h, $P(|X\text{-}10| X > k) \leq 0.25$

Sarebbe altrettanto "facile" rispondere ai punti precedenti se X ~ B(20,0.3)?

E Data una variabile aleatoria X con distribuzione B(20,0.2)

1. determinare media e mediana di X;
2. come si può scegliere p, in modo che Y ~ B(20,p) abbia media maggiore a quella di X?
3. come si può scegliere p, in modo che Z ~ B(20,p) abbia mediana maggiore a quella di X?
4. come si possono scegliere n e p, in modo che W ~ B(n,p) abbia stessa media di X ma mediana maggiore a quella di X?

(Confrontare i grafici delle rispettive densità trovate)

F Sia X ~ B(10,0.3) e Y=X+4. Dire, se è possibile, in che relazione stanno fra loro:

a) le medie e le mediane di X e Y;
b) i grafici delle funzioni cumulate di X e Y;
c) i grafici delle densità di X e Y;
d) le varianze di X e Y.

E2. Esperienza. Simulazione di una variabile aleatoria binomiale come somma di variabili aleatorie di Bernoulli

1) Simulare un esperimento binomiale, ad esempio il lancio di una moneta regolare 10 volte (successo: TESTA codificato con 1, insuccesso: CROCE codificato con 0)
   Calcolare la somma dei successi (ovvero la somma dei risultati ottenuti)
2) Ripetere l'esperimento per N=100 volte e determinare
   - la distribuzione dei successi in 10 lanci ottenuta sulle 100 ripetizioni
   - la media dei successi
   - lo scarto quadratico medio dei successi
3) Confrontare i risultati sperimentali ottenuti precedentemente con quelli teorici.