Lezione 09/11/2022

Variabile binomiale

Definizione. *Sia X una variabile binomiale di parametri n e p ($X \sim Bin(n,p)$), allora*

$$E(X) = np \quad e \quad Var(X) = np(1-p).$$

Esercizio 1. *Ad un esame Paolo deve rispondere a 5 quesiti a risposta multipla. Ogni quesito è costituito da 4 risposte, di cui solo una è quella esatta. Il test è superato rispondendo in modo corretto ad almeno 3 domande. Essendo del tutto impreparato Paolo risponde a caso a tutti i quesiti.*

(a) Qual è la probabilità che Paolo superi il test?

(b) Qual è il numero medio di risposte esatte che Paolo può aspettarsi di aver dato?

Svolgimento. *(a) Consideriamo X la variabile a. che rappresenta il numero di quesiti risposti correttamente. X è una variabile binomiale perchè ogni quesito può essere considerato come una prova di Bernoulli, ovvero un esperimento con due esiti possibili: risposto correttamente / non risposto correttamente. X ha parametri 5 (numero di quesiti) e $\frac{1}{4}$ (la probabilità di rispondere correttamente a un quesito è 1 su 4, perchè su 4 risposte possibili solo una è quella corretta). La probabilità che Paolo superi l'esame è $P(X \geq 3) = P(X=3) + P(X=4) + P(X=5) = 8,8\% + 1,5\% + 0,1\% = 10,4\%$.*

(b) $E(X) = np = 5\frac{1}{4} = \frac{5}{4} = 1,25$. Ci aspettiamo che in media Paolo dia 1,25 risposte corrette.

Osservazione $X \sim Bin(1,p)$ è una variabile di Bernoulli.

Teorema (Funzione di ripartizione della variabile binomiale)**.** *Sia $X \sim Bin(n,p)$ dove $X(\Omega) = \{0, \dots n\}$, allora*

$$P(X \leq k) = \sum_{i \leq k} P(X=i) = \sum_{i=0}^{k} \binom{n}{i} (p)^i (1-p)^{n-i}.$$

Esempio. *Supponiamo di lanciare 3 volte una moneta e consideriamo la variabile binomiale* $X \sim Bin(3, \frac{1}{3})$ *che rappresenta il numero di teste uscite nei 3 lanci. La probabilità di ottenere al più 2 teste è*

$$P(X \leq 2) = \binom{3}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^3 + \binom{3}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2 + \binom{3}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^1 + \binom{3}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^0.$$

La distribuzione binomiale viene rappresentata graficamente per mezzo di un istogramma o di un diagramma a barre. La forma della distribuzione dipende dal valore della probabilità di successo p. Nel caso $p = \frac{1}{2}$ (ed anche $1 - p = \frac{1}{2}$), il successo e l'insuccesso sono ugualmente probabili; da questo segue che la probabilità di avere ad esempio 2 successi (e quindi $n-2$ insuccessi) è uguale alla probabilità di avere $n-2$ successi (e quindi 2 insuccessi): l'istogramma della distribuzione è quindi simmetrico (figura 1, pagina seguente). Se invece $p < \dfrac{1}{2}$ e $p > \dfrac{1}{2}$, l'istogramma è asimmetrico; nel primo caso l'asimmetria è positiva, la distribuzione è obliqua a destra (figura 2), nel secondo caso l'asimmetria è negativa, la distribuzione è obliqua a sinistra (figura 3).

Disuguaglianza di Chebyshev

La disuguaglianza di Chebyshev consente di stimare la probabilità di una variabile aleatoria di cui si conoscono la media e la varianza.

Teorema (Disuguaglianza di Chebyshev)**.** *Se X è una v.a. di media μ e varianza σ^2, allora*

$$P(\mu - \varepsilon \leq X \leq \mu + \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

per ogni $\varepsilon > 0$.

Esercizio 2. *Il numero di clienti che visitano un concessionario di auto al sabato mattina è una variabile aleatoria con valor medio $\mu = 18$ e deviazione standard $\sigma = 2.5$. Con quale probabilità si può asserire che il numero di clienti è compreso tra 8 e 28?*

Svolgimento. *Se X è la variabile aleatoria che rappresenta il numero di clienti, allora dobbiamo calcolare*

$$P(8 \leq X \leq 28) = P(18 - 10 \leq 18 + 10).$$

Utilizziamo la disuguaglianza di Chebyshev dove $\varepsilon = 10$ e $\sigma = 2.5$:

$$P(8 \leq X \leq 28) \geq 1 - \frac{(2.5)^2}{(10)^2} = 0.9375.$$

La probabilità che tra 8 e 28 clienti visitino il concessionario è almeno 0.9375.

Dal precedente esercizio si osserva che la disuguaglianza di Chebyshev si usa quando si verificano le seguenti condizioni:

- Si conoscono il valor medio μ e la varianza σ^2 (o la deviazione standard σ) di una v.a. X;
- si vuole stimare la probabilità che X assuma valori in un intervallo di centro μ.

Per utilizzare la disuguaglianza di Chebychev bisogna calcolare ε che è la distanza tra μ e gli estremi dell'intervallo rispetto cui vogliamo stimare la probabilità
(nell'esempio $\varepsilon = |28 - 18| = |8 - 18| = 10$).

La disuguaglianza di Chebyshev esprime matematicamente il signifato della varianza (e quindi della deviazione standard).

- Se σ^2 è piccola, allora c'è un'alta probabilità di ottenere valori di X vicini al valor medio;
- se σ^2 è grande, allora c'è una maggiore probabilità di ottenere valori lontani dal valor medio.

Variabile Ipergeometrica

Esempio. *Consideriamo un'urna con 8 palline: 6 bianche e 3 nere. Estraiamo 4 palline senza rimpiazzo, vogliamo calcolare la probabilità di ottenere 3 palline bianche.*

Svolgimento. *Ogni estrazione è un esperimento di Bernoulli e quindi è descritto da una variabile di Bernoulli, dove il successo è l'evento* esce una pallina bianca. *Le prove di Bernoulli non sono indipendenti tra loro e X rappresenta il numero di successi nelle 4 prove.*

$$P(X = 3) = \frac{num.\ di\ casi\ favorevoli}{num.\ di\ casi\ possibili}.$$

- ***numero di casi possibili.*** *Chiamiamo $B_1, B_2, B_3, B_4, B_5, B_6$ le sei palline bianche e N_1, N_2, N_3 le tre palline nere dell'urna. Tutti i casi possibili sono gli insiemi composti da 4 palline prese dalle 9 che si trovano nell'urna. Esempi: $B_1B_2N_1N_3$, $B_1B_2B_6N_3$, $B_1B_2B_4B_3$, ... I casi possibili sono tutte le combinazioni semplici di 9 oggetti e di classe 4, quindi il loro numero è $\binom{9}{4}$.*
- ***numero di casi favorevoli.*** *Tutti i casi favorevoli sono le coppie (x, y) dove x è un insieme di 3 palline dell'insieme $\{B_1, B_2, B_3, B_4, B_5, B_6\}$ e y è una pallina dell'insieme $\{N_1, N_2, N_3\}$. Dunque, l'insieme dei casi favorevoli si rappresenta come $A \times B$ dove*
  - $A = \{$*gruppi di 3 palline di* $\{B_1, \ldots, B_6\}\}$ *e*
  - $B = \{$*gruppi di 1 pallina di* $\{N_1, N_2, N_3\}\}$.

  A è l'insieme delle combinazioni semplici di 6 oggetti e di classe 3, metre B è l'insiem delle combinazioni semplici di 3 oggetti e di classe 1. Quindi, $\#A = \binom{6}{3}$ e $\#B = \binom{3}{1}$.

  Per il principio di moltiolicazione, il $num.\ di\ casi\ favorevoli$ è uguale a $\#(A \times B) = \#A \times \#B = \binom{6}{3} \cdot \binom{3}{1}$.

Andando a sostituire,

$$P(X = 3) = \frac{\binom{6}{3} \cdot \binom{3}{1}}{\binom{9}{4}}.$$

Definizione (Variabile ipergeometrica)**.** *La variabile ipergeometrica conta il numero di successi in uno schema Successo/insuccesso senza rimpiazzo.*

Esempio. *La variabile X dell'esempio precedente è una variabile ipergeometrica (conta il numero di palline bianche in 4 estrazioni senza rimpiazzo (4 prove di Bernoulli non indipendenti)).*

Definizione (Distribuzione di probabilità di una variabile ipergeometrica)**.** *Considera l'esperimento che consiste nell'estrarre m volte e* senza rimpiazzo *una pallina da un'urna di b palline bianche e n palline nere, allora la probabilità di estrarre k palline bianche è data dalla seguente formula:*

$$P(X = k) = \frac{\binom{b}{k} \cdot \binom{n}{m-k}}{\binom{b+n}{m}},$$

dove $X \sim Iper(b, n, m)$

Osserva che

- b : numero di palline bianche nell'urna;
- k : numero di palline bianche in m estrazioni;
- n : numero di palline nere nell'urna;
- $m - k$: numero di palline nere in m estrazioni;
- $b + n$: palline nell'urna;
- m totale delle palline estratte.

Teorema (Funzione di ripartizione della variabile ipergeometrica)**.** *Sia $X \sim Iper(b, n, m)$ dove $X(\Omega) = \{0, \ldots, n\}$, allora*

$$P(X \leq k) = \sum_{i \leq k} P(X = i) = \sum_{i=0}^{k} \frac{\binom{b}{i} \cdot \binom{n}{m-i}}{\binom{b+n}{m}}.$$

Esercizio 3. *Una grande azienda ha 150 impiegati, 12 dei quali hanno un alto tasso di assenteismo. All'avvicinarsi delle feste natalizie, l'azienda decide di assegnare 40 premi agli impiegati estratti a sorte. Calcolare la probabilità che i premi vengano assegnati:*

(a) a nessun impiegato assenteista;

(b) a 3 impiegati assenteisti;

(c) a meno di 5 assenteisti.

Svolgimento.

(a) *Prima di tutto, osserviamo che l'estrazione a sorte di 40 impegati su 150 si può assimilare all'estrazione* senza rimpiazzo *di 40 palline da un'urna di 150 palline, dove ogni pallina ha una tra due possibili etichette:* assenteista *e* non assenteista*. Stiamo parlando di uno schema successo/insuccesso senza rimpiazzo, quindi, se X è la variabile aleatoria che raprresenta il numero di impiegati assenteisti estratti (numero di successi), allora dobbiamo calcolare $P(X=0)$, dove $X \sim Iper(12, 138, 40)$ (osserva che 12 sono gli impiegati assenteisti, 138 quelli non assenteisti e 40 quelli estratti sul totale di 150).*

$$P(X=0) = \frac{\binom{12}{0} \cdot \binom{138}{40-0}}{\binom{150}{40}} = 0.020.$$

(b)

$$P(X=3) = \frac{\binom{12}{3} \cdot \binom{138}{37}}{\binom{150}{40}} = 0.266.$$

(c) $P(X<5) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) = 0.020 + 0.99 + 0.212 + 0.217 = 0.814.$

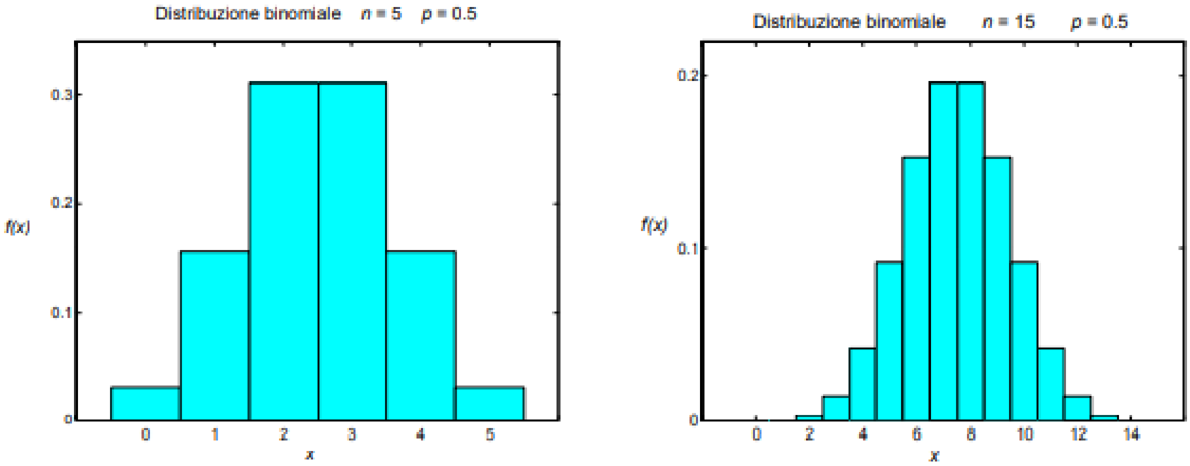

Figura 1

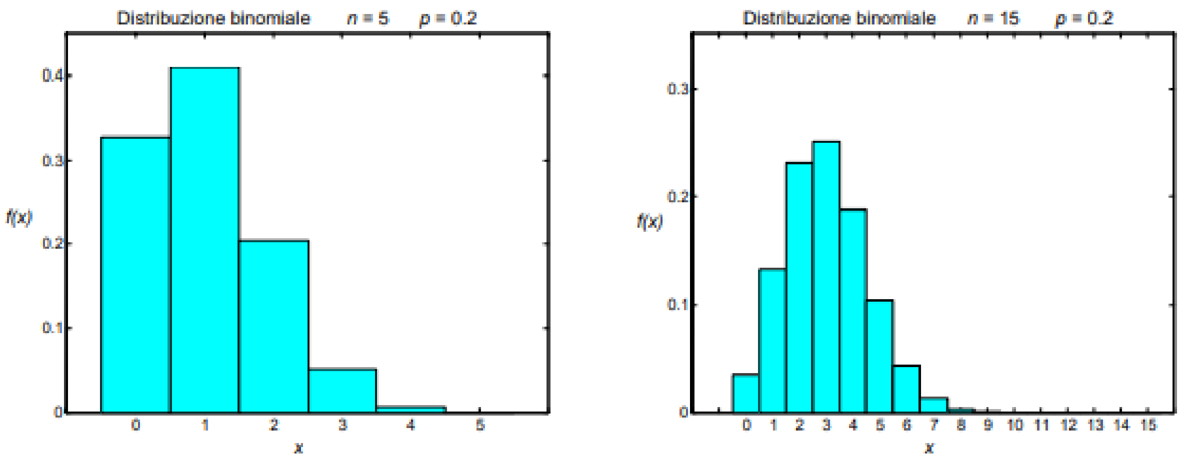

Figura 2

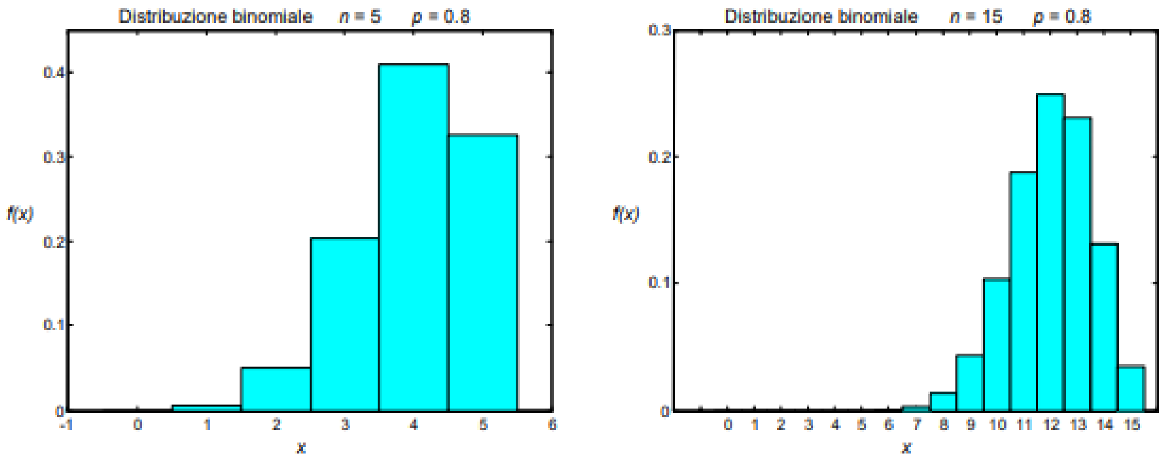

Figura 3