6 Distribuzioni di probabilità discrete

Nel mondo reale per descrivere diversi fenomeni si utilizzano delle variabili casuali discrete le quali, per essere caratterizzate, necessitano di poche distribuzioni di probabilità.
Per esempio, per testare l'efficacia di un farmaco, il numero di pazienti che guariscono dopo l'assunzione di tale farmaco può essere descritto da una variabile casuale binomiale; a livello industriale, se si seleziona un campione da un lotto di produzione, il numero di unità difettose presenti nel campione può essere descritto da una variabile casuale ipergeometrica; in un problema di controllo della qualità il numero di campioni richiesti per produrre un falso allarme può essere descritto da una variabile casuale geometrica; il numero di globuli bianchi in un campione di sangue può essere descritto da una variabile casuale poissoniana. Andremo ad analizzare le distribuzioni più comunemente utilizzate distinguendo tra variabile casuali discrete finite ed infinite.

DISTRIBUZIONE UNIFORME

È una distruibuzione che è uniforme su un insieme, cioè che attribuisce la stessa probabilità ad ogni elemento dell'insieme su cui è definita.

$$f(x;n) = \begin{cases} \dfrac{1}{n} & x = 1, 2, ..., n\,, \\ 0 & \text{altrove}\,. \end{cases}$$

Un esempio è fornito dal lancio di un dado equilibrato, in cui ogni faccia ha probabilità $\frac{1}{6}$ di presentarsi ad ogni lancio.

- $E[X] = \sum_{i=1}^{n} x_i f(x_i) = \sum_{i=1}^{n} i \cdot \frac{1}{n} = \sum_{i=1}^{n} \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$
- $\text{var}[X] = E[X^2] - E[X]^2 = \sum_{i=1}^{n} x_i^2 f(x_i) - \left(\frac{n+1}{2}\right)^2 =$
  $\sum_{i=1}^{n} i^2 \cdot \frac{1}{n} - \left(\frac{n+1}{2}\right)^2 = \frac{1}{n} \frac{n(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}$

Questa distribuzione è quella che fornisce la classica definizione di probabilità "casi favorevoli" su "casi possibili".

Esempio. Lotteria (90 biglietti numerati da 1 a 90)

L'organizzatore tiene per sé 6 biglietti il cui numero

$\in (9, 16)$. Qual è la variabile casuale idonea? Qual è la probabilità che l'organizzatore vinca il primo premio?
La variabile casuale è una uniforme di parametro $n = 90$:

- $P[X = x] = \frac{1}{90} \quad x = 1, ..., 90$
- $P[9 < x < 16] = P[9 < X \leq 15] = F(15) - F(9) = \sum_{x_i \leq 15} f(x_i) - \sum_{x_i \leq 9} f(x_i) = \frac{15 - 9}{90} = \frac{1}{15}$

Secondo la definizione classica (gli interi tra 9 e 16 sono 6):

$$p = \frac{f}{n} = \frac{6}{90} = \frac{1}{15}$$

Chiaramente, se i numeri appartengono all'intervallo $[9, 16)$ o all'intervallo $(9, 16]$ o all'intervallo $[9, 16]$ si hanno probabilità diverse:

$$P[9 \leq X < 16] = P[9 \leq X \leq 15] \neq P[9 < X \leq 15]$$

Nel caso discreto considerare o meno l'estremo dell'intervallo fornisce risultati diversi.
$P[9 \leq X \leq 15] = F(15) - F(9) + P[X = 9] =$
$= \frac{15 - 9}{90} + \frac{1}{90} = \frac{7}{90}$

DISTRIBUZIONE BINOMIALE

Spesso un esperimento consiste in una serie di prove ripetute, ognuna con due soli esiti possibili (si/no; vero/falso; successo/insuccesso; 0/1). Ogni prova il cui esito è dicotomico è detta prova di Bernoulli.

Processo di Bernoulli

Un processo di Bernoulli ha le seguenti caratteristiche:

1. l'esito è dicotomico
2. la probabilità di successo, indicata con p, è la stessa in ogni prova
3. l'esperimento consiste in prove ripetute
4. le prove sono indipendenti

La variabile casuale X che conta il numero di successi che si registrano in n prove di Bernoulli è chiamata binomiale e la sua funzione di densità è:

$$B(x, n, p) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, ..., n$$

dove $q = 1 - p$ è la probabilità di insuccesso, $p \in [0, 1]$.

Il nome nasce dal fatto che nello sviluppo binomiale di $(p+q)^n$ i termini corrispondono ai valori di $B(x,n,p)$ per $x = 0, ..., n$. Più precisamente:

$$(p+q)^n = \binom{n}{0} p^0 q^n + \binom{n}{1} p^1 q^{n-1} + \cdots + \binom{n}{n} p^n q^0$$
$$= B(0,n,p) + B(1,n,p) + \cdots + B(n,n,p)$$

Nota bene: $\sum_{i=1}^{n} B(i,n,p) = 1$; $(p+q=1)$.

- $E[X] = np$
- $\text{var}[X] = npq$

Proprietà

- $P[X < x] = P[X \leq x-1]$
- $P[X > x] = 1 - P[X \leq x]$
- $P[X \geq x] = 1 - P[X < x] = 1 - P[X \leq x-1]$
- $P[X = x] = P[X \leq x] - P[X < x] =$
  $= P[X \leq x] - P[X \leq x-1]$

Esempio. Lancio di una moneta
Esce testa con probabilità $p \in [0,1]$, esce croce con probabilità $1 - p = q$.

La moneta viene lanciata n volte.
Qual è la probabilità di ottenere x volte testa (ed $n - x$ volte croce)?
Indichiamo con:
$1 =$ "esce testa"
$0 =$ "esce croce"
Lo spazio campione è dato da:

$$\Omega = \{w : w = (w_1, ..., w_n) \text{ con } w_i = 0 \text{ o } 1, i = 1, ..., n\}$$

Per esempio:

$$w = (\underbrace{1, ..., 1}_{x \text{ volte}}, \underbrace{0, ..., 0}_{n-x \text{ volte}})$$

Sia
$A_i = \{w : w_i = 1\} = \{\text{testa nel i-esimo lancio}\}$
$P[A_i] = p$
$\Rightarrow w = A_1 \cap ... \cap A_x \cap \overline{A}_{x+1} \cap ... \cap \overline{A}_n$
Poichè le prove sono indipendenti
$P[w] = P[A_1]...P[A_x] \cdot P[\overline{A}_{x+1}]...P[\overline{A}_n] =$
$= p^x q^{n-x}$
Quante sono le configurazioni possibili?
Tante quante le combinazioni di n oggetti di cui "x" ed "$n - x$" uguali tra loro.

Utilizzando il calcolo combinatorio questo numero è:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

perciò:

$$B(x, n, p) = P[X = x \text{ successi}] = \binom{n}{x} p^x (1-p)^{n-x}$$

La funzione di ripartizione è data da:

$$F(x) = P[X \leq x] = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$$

La distribuzione binomiale viene utilizzata in ambito industriale, in ambito medico, in ambito militare, nelle estrazioni con reinserimento dell'oggetto (per avere l'indipendenza delle prove).

Esempi

1) La probabilità che un componente resista ad una prova d'urto è del 75%. Calcolare la probabilità che esattamente due dei 4 componenti sottoposti alla prova, la superino con successo.

$x = 2, n = 4, p = \frac{3}{4}$.

$B(2, 4, \frac{3}{4}) = \binom{4}{2} \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^2 = \frac{4!}{2!2!} \frac{3^2}{4^4} = \frac{27}{128}$

2) Il 5% dei chip di memoria prodotti da una fabbrica è difettoso. Determinare la probabilità che su 4 chip scelti a caso:
a) uno sia difettoso;
b) nessuno sia difettoso;
c) meno di due siano difettosi.
Infine, se X è la variabile casuale che indica il numero dei chip difettosi su un totale di 400 chip, determinare $E[X]$ e $\mathrm{var}[X]$.

Per la prima parte dell'esercizio:
$n = 4, \quad p = 0.05$
a) $P[X = 1] = \binom{4}{1}(0.05)^1 \cdot (0.95)^3 = 0.171$
b) $P[X = 0] = \binom{4}{0}(0.05)^0 \cdot (0.95)^4 = 0.814$
c) $P[X < 2] = P[X = 0] + P[X = 1] = 0.985$

Per la seconda parte dell'esercizio:
$n = 400, \quad p = 0.05$
$E[X] = np = 20 \quad \mathrm{var}[X] = npq = 19$

3) Un rivenditore acquista un certo apparecchio elettronico e sa dal produttore che il tasso di prodotti difettosi è del 3%.

a) L'addetto al controllo della qualità seleziona a caso 20 apparecchi da un ordine. Qual è la probabilità che vi sia almeno un apparecchio difettoso?
b) Se il rivenditore riceve 10 ordini al mese e l'addetto alla qualità riceve 20 apparecchi per ogni ordine, qual è la probabilità che vi siano 3 ordini ciascuno contenente almeno un apparecchio difettoso?

a) $X = n^\circ$ apparecchi difettosi.
$n = 20$
$p = 0.03$
$P[X \geq 1] = 1 - P[X < 1] = 1 - P[X = 0] =$
$= 1 - \binom{20}{0}(0.03)^0 \cdot (0.97)^{20} = 1 - (0.97)^{20} = 0.4562$

b) Ogni ordine può contenere almeno un apparecchio difettoso oppure non contenere apparecchi difettosi. Il controllo di ogni ordine è una prova di Bernoulli con probabilità $p = 0.4562$
Gli ordini sono indipendenti
$Y = n^\circ$ di ordini con almeno un apparecchio difettoso.
$Y \sim B(y, n = 10, p = 0.4562)$
$P[Y = 3] = \binom{10}{3}(0.4562)^3 \cdot (0.5438)^7 = 0.1602$

DISTRIBUZIONE IPERGEOMETRICA

Si distingue dalla distribuzione binomiale perchè non viene richiesta l'indipendenza delle prove e si basa sul campionamento senza reinserimento dell'oggetto.
Viene utilizzata nel campionamento in accettazione e nella certificazione di qualità.
Lo scopo è determinare la probabilità di selezionare x successi da k unità indicate come successi ed $n-x$ insuccessi da $N-k$ unità indicate come insuccessi, dato un campione casuale di dimensione n selezionato da N unità.
Il numero x di successi dipende dal numero di successi k presenti nell'insieme N da cui vengono selezionate n unità.

$$f(x, N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, ..., n$$

con $k > n$ e $N - k > n$.

- $E[X] = \frac{nk}{N}$
- $\text{var}[X] = \frac{nk(N-k)(N-n)}{N^2(N-1)}$

Esempio

Un rivenditore acquista chip a lotti di 10. In ogni lotto controlla a caso 3 chip.
Decide di accettare il lotto solo se nessuno dei 3 chip è difettoso.
Sapendo che il 30% dei lotti ha 4 pezzi difettosi su 10 e il 70% dei lotti ha 1 pezzo difettoso su 10, calcolare la percentuale di rifiuto del rivenditore.
Chiamiamo:
A l'evento "il lotto viene accettato"
$R = \overline{A}$ l'evento "il lotto viene rifiutato"
D_i il lotto ha i pezzi difettosi.

$P[A] = P[A|D_4]P[D_4] + P[A|D_1]P[D_1]$ per il th. delle prob. totali

$P[D_4] = \frac{3}{10}, \quad P[D_1] = \frac{7}{10};$

$$P[A|D_4] = \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} = \frac{1 \cdot \frac{6!}{3!3!}}{\frac{10!}{3!7!}} = \frac{1}{6}$$

$$P[A|D_1] = \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}} = \frac{1 \cdot \frac{9!}{3!6!}}{\frac{10!}{3!7!}} = \frac{7}{10}$$

Perciò:

$P[A] = \frac{1}{6} \cdot \frac{3}{10} + \frac{7}{10} \cdot \frac{7}{10} = \frac{54}{100} = 54\%$

$P[R] = P[\overline{A}] = 1 - P[A] = 46\%$

DISTRIBUZIONE GEOMETRICA

Si consideri un esperimento in cui le proprietà sono le stesse descritte in un esperimento binomiale, ma con l'eccezione che le prove indipendenti vengano ripetute fino al raggiungimento del primo successo.
Perciò l'obiettivo non è calcolare la probabilità di ottenere x successi in n prove, con n noto, bensì la probabilità che si verifichi il primo successo nella x-esima prova.
La variabile casuale geometrica conta il numero x di prove necessarie per ottenere il primo successo.

$$G(x,p) = pq^{x-1} \quad x = 1, 2, 3, \ldots \text{ (non finito)}$$

dove p è la probabilità di successo, $q = 1 - p$ la probabilità di insuccesso.

- $E[X] = \dfrac{1}{p}$
- $\mathrm{var}[X] = \dfrac{q}{p^2} = \dfrac{1-p}{p^2}$

X è detta tempo di attesa (es. il ritardo di un numero nel gioco del lotto).
$E[X] = \frac{1}{p}$ è detta tempo di ritorno, è la media del numero di ripetizioni da tentare per avere un successo.

OSSERVAZIONE

Poichè per studiare l'istante di un primo successo bisogna considerare un numero arbitrariamente grande di prove non può più essere utilizzato lo schema di Bernoulli che descrive un numero prefissato di prove.

Proprietà

La variabile casuale geometrica gode della proprietà "mancanza di memoria".

In uno schema successo-insuccesso supponiamo di non aver ottenuto alcun successo nelle prime k prove.

Qual è la probabilità di dover attendere ancora m prove per avere il primo successo?

È la stessa che si avrebbe se le prime k prove non fossero state eseguite.

Ciò è ovvio perchè in uno schema a prove ripetute indipendenti i risultati delle prime non influenzano le successive.

Infatti

Ω spazio campione

$\{E_i\}$ $i = 1, 2, 3, ...$ successione di eventi indipendenti

$P[E_i] = p$ e $q = 1 - p$
Poichè X è la variabile casuale tempo di attesa del 1° successo.
$X = x$ se $E_1, ..., E_{x-1}$ non si verificano ed E_x si verifica.

$$\begin{aligned} P[X = x] &= P[\overline{E}_1 \cap \overline{E}_2 \cap ... \cap \overline{E}_{x-1} \cap E_x] \\ &= P[\overline{E}_1] \cdot P[\overline{E}_2]...P[\overline{E}_{x-1}] \cdot P[E_x] \\ &= q^{x-1}p = p(1-p)^{x-1} \end{aligned}$$

Esempio. Ritardo di un numero nel gioco del lotto.
$\overline{E}_x$ = "nella x-esima estrazione (a partire da una estrazione "0"), viene estratto un numero su una ruota".
Calcolare la probabilità che il numero abbia accumulato un ritardo di x settimane

$$P[X > x] = 1 - P[X \leq x] = 1 - F(x) =$$
$$1 - \sum_{k=1}^{x} p(1-p)^{k-1} = 1 - \sum_{k=1}^{x} (1-q)q^{k-1} =$$
$$1 - (1 - q + q(1-q) + ... + q^{x-1}(1-q)) =$$
$$1 - (1 - \cancel{q} + \cancel{q} - \cancel{q^2} + \cancel{q^2} - q^3 + ... + q^{x-1} - q^x) =$$
$$1 - (1 - q^x) = q^x = (1-p)^x$$

Calcolare la probabilità che il numero venga estratto solo tra k settimane sapendo che ha un ritardo di m settimane.

$$P[X = k + m | X > m] = \frac{P[X = m + k, X > m]}{P[X > m]} =$$

$$= \frac{P[X = m + k]}{P[X > m]} = \frac{pq^{\cancel{m}+k-1}}{q^{\cancel{m}}} = pq^{k-1} =$$

$$= p(1-p)^{k-1} = P[X = k]$$

$\Rightarrow$ MANCANZA DI MEMORIA

OSSERVAZIONI

Abbiamo visto che:

$$P[X = x] = p(1-p)^{x-1} \quad x = 1, 2, 3, \ldots$$

rappresenta la probabilità che alla x-esima prova si ottenga il primo successo.

$$P[X > x] = (1-p)^x$$

$$\Rightarrow P[X \geq x] = P[X > x] + P[X = x] =$$

$$(1-p)^x + p(1-p)^{x-1} = (1-p)^x \left(1 + \frac{p}{1-p}\right) =$$

$$= \frac{(1-p)^x}{1-p} = (1-p)^{x-1}$$

$$\Rightarrow P[X \geq x+1] = (1-p)^x$$

Anzichè contare il numero di prove X per ottenere il primo successo possiamo contare il numero Y di insuccessi prima del primo successo.

$$P[Y = y] = p(1-p)^y \quad y = 0, 1, 2, 3, \dots$$

Y è ancora una variabile casuale geometrica.

- $E[Y] = \dfrac{q}{p} = \dfrac{1-p}{p}$
- $\text{var}[Y] = \dfrac{q}{p^2} = \dfrac{1-p}{p^2}$

$P[Y \leq y] = 1 - P[Y > y] = 1 - (1-p)^{y+1}$
$P[Y \geq y] = 1 - P[Y < y] = 1 - P[Y \leq y-1] =$
$= 1 - [1 - (1-p)^y] = (1-p)^y$

Esempi

1) Di un certo processo manifatturiero si conosce che, in media, 1 prodotto su 100 è difettoso.
Qual è la probabilità che il quinto prodotto controllato sia il primo identificato come difettoso?
$p = 0.01 =$ probabilità di avere un difetto
$x = 5$
$G(5, 0.01) = 0.01 \cdot (0.99)^4 = 0.0096$

2) Ad un centralino telefonico tutte le linee sono occupate. Sapendo che $p = 0.05$ è la probabilità di ottenere una linea libera in un momento di massima affluenza di telefonate, calcolare la probabilità che siano necessari 5 tentativi prima di trovare la linea libera.

$p = 0.05 =$ probabilità di avere la linea libera

$x = 5$

$G(5, 0.05) = 0.05 \cdot (0.95)^4 = 0.041$

DISTRIBUZIONE DI POISSON

La distribuzione di probabilità poissoniana si utilizza negli esperimenti in cui è possibile fare un qualche tipo di conteggio, in un certo intervallo di tempo o di spazio. La variabile casuale X che rappresenta il numero di eventi che si vericano in un intervallo di tempo o di spazio è non finita e segue la legge:

$$P(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, ...$$

con $\lambda > 0$.

Il parametro λ rappresenta il numero medio di realizzazioni dell'evento nell'intervallo considerato.

- $E[X] = \lambda$
- $\text{var}[X] = \lambda$

Esempi di esperimenti in cui si utilizza la variabile casuale di Poisson sono:

– il numero di chiamate telefoniche ricevute da un centralino in un intervallo di tempo fissato

– il numero di clienti che entrano in un ufficio in un giorno

– il numero degli incidenti stradali in una regione in una settimana

– il numero di refusi in una pagina di un libro

OSSERVAZIONE
L'evento che si verifica in un certo intervallo è un evento raro.

Esempi
1) In un esperimento, il numero medio di particelle radioattive individuate da un contatore in 1 millisecondo è pari a 4.
Qual è la probabilità che 6 particelle vengano individuate in un dato millisecondo?

$x = 6$ e $\lambda = 4$

$$P[6,4] = \frac{e^{-4}4^6}{6!} = 0.1042$$

2) Il numero medio di navi che attraccano in un porto, in un giorno, è pari a 10 e la struttura può gestire 15 navi al giorno. Qual è la probabilità che in un dato giorno le navi non possano attraccare?

$\lambda = 10$

$$P[X > 15] = 1 - P[X \leq 15] = 1 - \sum_{x=0}^{15} \frac{e^{-10}10^x}{x!} =$$

$$1 - \left(e^{-10} + e^{-10} \cdot 10 + \frac{e^{-10} \cdot 10^2}{2!} + \ldots + \frac{e^{-10}10^{15}}{15!} \right)$$

$$= 0.0487$$

3) In un'azienda ogni giorno si assentano in media 1.8 operai. Calcolare la probabilità che in un dato giorno siano assenti 3 operai contemporaneamente.

$x = 3$, $\lambda = 1.8$

$$P[X = 3] = e^{-1.8} \cdot \frac{(1.8)^3}{3!} = 0.16$$

Proprietà
La somma di due variabili casuali di Poisson è una variabile casuale di Poisson

Sia $X_1 \sim P(\lambda_1)$ e $X_2 \sim P(\lambda_2)$ allora

$$Y = X_1 + X_2 \sim P(\lambda = \lambda_1 + \lambda_2)$$

Questa proprietà prende il nome di riproducibilità della distribuzione di Poisson.

Esempio

Il numero di apparecchi difettosi prodotti giornalmente da una fabbrica segue una distribuzione di Poisson di parametro $\lambda = 4$.
Calcolare la probabilità che in due giorni consecutivi non vengano prodotti più di 3 pezzi difettosi.

Il 1° giorno $\quad X_1 \sim P(\lambda_1 = 4)$
Il 2° giorno $\quad X_2 \sim P(\lambda_2 = 4)$
X_1 e X_2 sono indipendenti
$Y = X_1 + X_2 \sim P(\lambda = \lambda_1 + \lambda_2 = 8)$

$$P[Y \leq 3] = \sum_{k=0}^{3} \frac{8^k e^{-8}}{k!} = e^{-8}\left(1 + 8 + \frac{8^2}{2} + \frac{8^3}{3!}\right) =$$

$$= 0.0424$$

APPROSSIMAZIONE DELLA BINOMIALE CON LA POISSON

La variabile casuale di Poisson può essere usata come approssimazione della variabile casuale binomiale quando il numero n di prove è molto elevato e la probabilità di successo p è molto piccola.
Una variabile casuale discreta finita (binomiale) viene approssimata con una variabile casuale discreta infinita (poissoniana)

$$B(x, n, p) \xrightarrow[\substack{n \to \infty,\, p \to 0 \\ np \to \lambda}]{} P(x, \lambda),$$

cioè $E[X_B] = np$ deve essere uguale a $E[X_P] = \lambda$ quando $n \to \infty$. Da un punto di vista empirico l'approssimazione è valida se

$$n \geq 50 \quad \text{e} \quad p \leq 0.1$$

<u>Esempio</u>

Il 3% delle lampadine prodotte da una fabbrica è difettoso. Su un campione di 100 lampadine, calcolare la probabilità che 2 siano difettose.

Con la binomiale

$n = 100, x = 2, p = 0.03$

$P[X = 2] = \binom{100}{2} 0.03^2 \cdot 0.97^{98} \sim 0.225$

Con la poissoniana

$n = 100, x = 2, p = 0.03 \Rightarrow \lambda = np = 3$

$P[X = 2] = \frac{e^{-3} 3^2}{2!} \sim 0.224$