Lezione 16/11/2022

Variabile di Poisson

Un'importante distribuzione di probabilità discreta, che ha una vasta gamma di applicazioni in aree diverse, è la cosiddetta distribuzione di Poisson. Essa può venire ricavata come approssimazione della distribuzione binomiale, secondo il procedimento che ora spieghiamo.

Consideriamo una variabile aleatoria binomiale X di parametri n e p; sappiamo che:

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

Indichiamo inoltre con λ il valore medio di X, cioè poniamo $\lambda = np$.

Esprimiamo la probabilità che sia $X = k$ in funzione di λ e di n.

Poiché $\lambda = np$, ne segue che $p = \frac{\lambda}{n}$, dunque

$$P(X=k)=\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}. \tag{1}$$

Supponiamo ora che n sia molto grande (e di conseguenza che p sia molto piccolo dal momento che $p = \lambda n$); in tal caso il valore della [1] può essere approssimato dal suo limite per n che tende a più infinito:

$$\lim_{n\to+\infty}\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}=\lim_{n\to+\infty}\frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{k!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k}=$$

$$\lim_{n\to\infty}\frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{n^k}\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k}=(*)$$

Osserviamo che

- $\lim_{n\to+\infty}\frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{n^k}=1$ (al numeratore, eseguendo i prodotti si ottiene un polinomio il cui termine di grado massimo è n^k);
- $\lim_{n\to+\infty}\frac{\lambda^k}{k!}=\frac{\lambda^k}{k!}$;
- $\lim_{n\to+\infty}\left(1-\frac{\lambda}{n}\right)^n=e^{-\lambda}$ (limite notevole);

- $\lim_{n\to+\infty}\left(1-\dfrac{\lambda}{n}\right)^{-k} = 1$ ($\dfrac{\lambda}{n}$ tende a 0 per n che tende a $+\infty$, dunque tutto il limite è 1).

  $= (*) = 1 \cdot \dfrac{\lambda^k}{k!} \cdot e^{-\lambda} \cdot 1 = e^{-\lambda}\dfrac{\lambda^k}{k!}$.

Riassumendo: se X è una variabile aleatoria binomiale tale che n è "grande" e p è "piccolo", vale la seguente approssimazione:

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!} \text{ dove } \lambda = np, \tag{2}$$

La formula [2] definisce la distribuzione di Poisson.

Definizione. *Una variabile aleatoria discreta X è detta di Poisson di parametro λ, con $\lambda > 0$, se la sua distribuzione di probabilità è assegnata da:*

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!} \qquad \textit{con } k = 0, 1, 2, \ldots$$

Circa la media e la varianza di una distribuzione di Poisson, si potrebbe dimostrare il seguente teorema.

Teorema. *Sia X una variabile aleatoria di Poisson di parametro λ; allora sia la media sia la varianza di X sono uguali a λ:*

$$E(X) = \lambda \quad \textit{e} \quad V(X) = \lambda.$$

La distribuzione di Poisson nella modellizzazione

La distribuzione di Poisson non è utile soltanto per calcolare approssimativamente probabilità relative a variabili aleatorie binomiali in cui n è grande e p è piccolo, ma riveste un ruolo di particolare importanza ai fini della modellizzazione di fenomeni aleatori. Essa è infatti il modello adatto a interpretare le situazioni descritte da una variabile aleatoria binomiale di cui conosciamo il valore medio, ma non i valori esatti di n e p, purché sia lecito supporre n "grande" e p "piccolo".

Esercizio 1. *Al centralino di un numero verde, in un'ora di punta, arriva un numero medio di 120 telefonate. Qual è la probabilità che il numero effettivo di telefonate arrivate a quel centralino in un'ora di punta sia 110?*

Svolgimento. *Indichiamo con la variabile aleatoria X il numero effettivo di telefonate che arrivano al centralino in un' ora di punta. Supponiamo che il numero n di "utenti potenziali" del centralino sia molto alto, che la* *probabilità p che un singolo utente telefoni al centralino sia molto bassa e che ciascun utente telefoni o meno al centralino indipendentemente dagli altri. Sotto queste ipotesi il numero X di telefonate effettivamente giunte al centralino si può assimilare al numero di successi in un processo diBernoulli di n prove di parametro p. Non conosciamo i valori esatti di n e di p, tuttavia conosciamo il valore medio di X (uguale a 120); inoltre abbiamo supposto che n sia grande e p sia piccolo: dunque siamo nelle condizioni per poter approssimare il modello esatto costituito dalla distribuzione binomiale di parametri incogniti n e p con la distribuzione di Poisson di parametro noto $\lambda = 120$. La distribuzione di Poisson di parametro $\lambda = 120$ è definita da:*

$$P(X = k) = e^{-120}\frac{120^k}{k!}.$$

In particolare: $P(X = 110) = e^{-120}\dfrac{120^{110}}{110!} = 2,5.$

Ragionando come nell'esempio precedente, si può capire perché la distribuzione di Poisson si rivela utile a descrivere, per esempio, i seguenti fenomeni:

- il numero di incidenti che si verificano su un certo tratto autostradale in un datogiorno;
- il numero di errori di stampa che si trovano in una pagina di un libro;
- il numero di auto che passano da un certo casello autostradale tra le 17 e le 18 diun dato giorno;
- il numero di persone di una comunità che superano l'età di 100 anni.

Esercizio 2. *Ad un servizio di guardia medica arrivano in media 3.5 richieste ogni ora di interventi urgenti a domicilio.*

(a) Calcola la probabilità che in una stessa ora arrivino 3, 4, 5 chiamate.

(b) Calcolare la probabilità che in una stessa ora arrivi un numero di chiamate maggiore o uguale di 4

Svolgimento. *Osservo che $X \sim P(3.5)$, allora*

(a) $P(X = 3) = e^{-3.5}\dfrac{(3.5)^3}{3!} = 0.2158$, $P(X = 4) = e^{-3.5}\dfrac{(3.5)^4}{4!} = 0.1888$, $P(X = 5) = e^{-3.5}\dfrac{(3.5)^5}{5!} = 0.1322$;

(b) $P(X \geq 4) = 1 - P(X < 4) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)] = 0.4633$.

Curiosità: La distribuzione di Poisson venne introdotta da S.D. Poisson nell'ambito dei suoi studi circa le applicazioni del calcolo dellaprobabilità alle cause civili e ai processi. Essa fu a lungoignorata, finché non si scoprìche il numero di particelle α emesse da una sostanza radioattiva in un dato intervallo di tempo non ha un valore fisso ma è una variabile aleatoria ben interpretata da una distribuzione di Poisson.

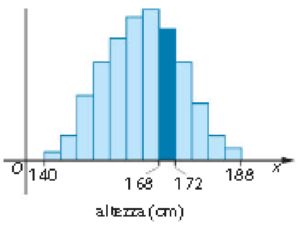
Schema riassuntivo per risolvere i problemi sulle variabili discrete

Variabile	Esperimento	Significato della Variabile	Notazione
Binomiale	Estrazione con rimpiazzo di 5 palline da un'urna con 4 palline bianche e 3 rosse (*Schema successo/insuccesso con rimpiazzo: successione di n prove di Bernoulli identiche e indipendenti di parametro p*)	X = numero di palline bianche in 5 estrazioni *X = numero di successi in n prove*	$X \sim Bin(5, \frac{4}{7})$ $X \sim Bin(n, p)$
Ipergeometrica	Estrazione senza rimpiazzo di 5 palline da un'urna con 4 palline bianche e 3 nere (*Schema successo/insuccesso senza rimpiazzo: estrazione di n palline senza rimpiazzo da un'urna con b palline bianche e r palline rosse*)	X = numero di palline bianche in 5 estrazioni *X = numero di successi in n estrazioni*	$X \sim Iper(4, 3, 7)$ $X \sim Iper(b, r, n)$
Geometrica	Estrazione con rimpiazzo di 5 palline da un'urna con 4 palline bianche e 3 rosse (*Schema successo/insuccesso con rimpiazzo: successione di n prove di Bernoulli identiche e indipendenti di parametro p*)	X = numero necessario di estrazioni da effettuare per ottenere una pallina bianca *X = numero necessario di prove da effettuare per ottenere un successo*	$X \sim Geo(\frac{4}{7})$ $X \sim Geo(p)$
Poisson	Estrazione con rimpiazzo di palline da un'urna con palline bianche e rosse, dove la probabilità di estrarre una pallina bianca è molto bassa, il numero complessivo di estrazioni effettuate è molto alto e si sa che in media vengono estratte 50 palline bianche (*Schema successo/insuccesso con rimpiazzo: successione di n prove di Bernoulli identiche e indipendenti di parametro p, dove n e p sono sconosciuti, ma si conosce il valor medio $\lambda = np$*)	X = numero di successi *X = numero di successi*	$X \sim P(50)$ $X \sim P(\lambda)$

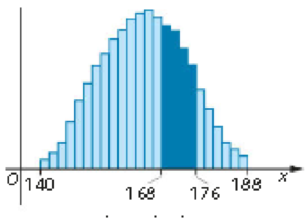
Variabili Continue

Esistono moltissimi fenomeni per la cui descrizione le variabili aleatorie discrete nonsono adatte. Per esempio, è necessaria una variabile aleatoria continua (cioè una variabile aleatoria che può assumere tutti i valori reali di un dato intervallo) per descrivere la misura di grandezze fisiche quali la temperatura, la velocità ecc. In questo contesto, a causa degli inevitabili errori di misura, non è utile chiedersi quanto vale la probabilità che la misura di una data grandezza assuma un valore prefissato, ma piuttosto qual è la probabilità che l'errore commesso non superi una certa soglia: si è interessati cioè a stabilire la probabilità che la misura di una grandezza assuma valori in un determinato intervallo. Per esempio, consideriamo la variabile aleatoria X che rappresenta l'altezza di una persona scelta a caso in un dato insieme di individui; come possiamo interpretare geometricamente la probabilità che X assuma valori in un determinato intervallo?

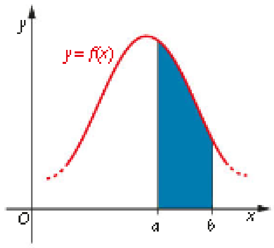
- L'istogramma rappresenta la distribuzione di probabilità delle altezze, dove si sono considerate classi di ampiezza costante uguale a 4 cm. Scegliendo come unità di misura sull'asse x l'ampiezza diciascuna classe, la probabilità che X assuma valori per esempio nell'intervallo [168 cm, 172 cm], è rappresentata dall'area del rettangolo in blu scuro.



- Riducendo l'ampiezza delle classi,l'istogramma ha un contorno più regolare; la probabilità che la variabilealeatoria X assuma per esempio valori nell'intervallo [168 cm, 176 cm] è uguale all'area della regione in blu scuro (sempre assumendo come unità di misura sull'asse x l'ampiezza di ciascuna classe).



- Immaginando di ridurre sempre di più l'ampiezza delle classi, il contorno dell'istogramma tende a diventare una linea continua, il grafico di una funzione reale divariabile reale, e la probabilità che X assuma valori in un datointervallo [a, b] si può interpretare come l'area del trapezoide limitato dal grafico della funzione e dall'asse x nell'intervallo [a, b].



Poiché l'area del trapezoide nella precedente figura è data dall'integrale della funzione f nell'intervallo [a, b], siamo indotti a ritenere che, nel continuo, la probabilità che una variabile aleatoria X assuma valori in un dato intervallo [a,b] possa essere definita come integrale di un'opportuna funzione f, detta densità di X, sull'intervallo [a,b]. Quindi,

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$