# *Distribuzioni di probabilità e variabili aleatorie discrete*

Variabile aleatoria→ è una variabile che assume valori numerici in corrispondenza ai risultati di un esperimento aleatorio.

Variabile aleatoria discreta→ variabile aleatoria che può assumere al più un insieme numerabile di valori (è discreta anche nel caso in cui il numero di possibili risultati à infinito)

Variabile aleatoria continua→ variabile aleatoria che può assumere un qualunque valore in un intervallo (il punto fondamentale è che ci sia un intervallo)

Vengono considerate variabili aleatorie discrete tutte quelle variabili in cui l'attribuzione di probabilità ai singoli risultati possibili è significativa.

Si indica con X la variabile aleatoria discreta e con x una sua possibile realizzazione, la probabilità che X assuma il valore di x si indica con P(X=x).

La distribuzione di probabilità (o funzione di probabilità) di una variabile aleatoria è la rappresentazione delle probabilità di tutti i possibili valori che può assumere la variabile. Un modo per poter rappresentare la distribuzione di probabilità è il diagramma ad aste.

Proprietà della funzione di probabilità di una variabile aleatoria discreta

- 0<=P(x)<=1 per ogni valore di x (le probabilità non possono essere negative o superare il valore 1)
- La somma delle singole probabilità deve dare 1 (in quanto questi eventi sono mutamente esclusivi e collettivamente esaustivi), questa proprietà ci permette di dire che quando sosteniamo un esperimento aleatorio almeno un risultato si verifica certamente.

La funzione di ripartizione F(x0) esprime la probabilità che X non superi il valore x0, cioè che P(X<=x0).

La probabilità di F(x0) è uguale all'unione degli eventi mutuamente esclusivi X=x per tutti i possibili valori di x minori o uguali a x0. La probabilità dell'unione è quindi la somma delle probabilità individuali dei singoli eventi.

Proprietà della funzione di ripartizione di una variabile aleatoria discreta

- 0<= f(x0) <= 1 per ogni x0
- Se x1 e x2 sono due valori tali che x1<x2, allora f(x1) <= f(x2)

Proprietà delle variabili aleatorie discrete

Nelle variabili aleatorie discrete viene utilizzata una misura di tendenza centrale chiamata "valore atteso" (media) e si calcola facendo:

$$\sum xP(x)$$

La varianza di una variabile aleatoria discreta è la media ponderata dei quadrati di tutti i suoi possibili scarti dalla media. Il peso associato a (X-media) ^2 è la probabilità che la variabile aleatoria assuma il valore x. La varianza può essere perciò vista come il valore medio che sarà assunto dalla funzione (X- media) ^2. La deviazione standard di una variabile aleatoria discreta non è altro che la radice quadrata della varianza.

Distribuzione binomiale

Modello di bernoulli→ considerando un esperimento casuale che può presentare solo due risultati, mutualmente esclusivi e collettivamente esaustivi, che per convenzione vengono chiamati "successo" e "insuccesso". P è probabilità di successo e (1-p) la probabilità di insuccesso. X assume il valore 1 in caso di successo e 0 in caso di insuccesso. L'equazione di probabilità di questa variabile aleatoria è P (0) =(1-p) P (1) =p

La media corrisponde a p, mentre la varianza è p(1-p)

Numero di sequenze di k successi in n prove indipendenti

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

Distribuzione binomiale

considerando un esperimento casuale che può presentare solo due risultati, mutualmente esclusivi e collettivamente esaustivi, che per convenzione vengono chiamati "successo" e "insuccesso". p→ probabilità di successo e n il numero di ripetizioni indipendenti dell'esperimento. La sua funzione di probabilità è chiamata distribuzione binomiale

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

La media è np

La varianza è np(1-p)

Distribuzione di Poisson

Si assuma che un intervallo sia suddiviso in un numero molto grande di sottointervalli, in modo che la probabilità di verificarsi di un evento in ogni sottointervallo sia molto piccola, ne derivano le seguenti ipotesi:

- La probabilità del verificarsi di un evento à costante per tutti i sottointervalli
- La probabilità che un evento si ripeta più di una volta all'interno di un sottointervallo è trascurabile
- Eventi che si verificano in intervalli disgiunti sono indipendenti

<u>Distribuzioni congiunte di due variabili aleatorie discrete</u>

Diversi utilizzi delle statistica riguardano relazioni tra variabili. Le variabili possono avere cambiamenti diversi e quindi sorge il problema di dover analizzare anche le relazioni che le legano.

Distribuzione di probabilità congiunta→ siano X e Y due variabili, la distribuzione di frequenza congiunta esprimerà la probabilità che X assuma un determinato valore x e che Y assuma un determinato valore y in contemporanea.

P (x, y) = p (X=x ∩ Y=y)

<u>Distribuzione di probabilità marginale</u>

Le distribuzioni di probabilità marginali riguardano una delle due variabili, ignorando l'altra.

$$f_X(x) = P(X = x) \qquad f_Y(y) = P(Y = y)$$

Le proprietà sono:

- 0<= P (x, y) <= 1 per ogni coppia di valori x, y
- La somma delle probabilità congiunte P (x, y) su tutte le possibili coppie di valori (x, y) deve valere 1.

Probabilità condizionate

- Le distribuzioni di probabilità condizionate della variabile aleatoria Y esprimono le probabilità di Y condizionatamente ad uno specifico valore x di X:
  $$f_{Y|X}(y \mid x) = P(Y = y \mid X = x)$$
  si calcola così:
  $$f_{Y|X}(y \mid x) = \frac{f_{XY}(x, y)}{f_X(x)}$$
- Esiste anche il caso contrario in cui basta invertire le variabili.

Indipendenza→ 2 variabili si dicono indipendenti se la loro distribuzione di probabilità congiunta è uguale al prodotto delle sue disponibilità di probabilità marginali:

$$f_{XY}(x,y) = f_X(x) f_Y(y)$$

Covarianza→ La covarianzadi due variabili X e Y è il valore atteso del prodotto degli scarti dai rispettivi valori attesi

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

- La covarianza misura la forza della relazione lineare tra due variabili aleatorie
    - Covarianza nulla←→assenza di relazione lineare
- Se due variabili aleatorie sono indipendenti allora non esiste nessun tipo di relazione tra loro→ la loro covarianza è nulla
    - Se la covarianza è nulla non è detto che due variabili siano indipendenti

Correlazione→ la correlazione tra due variabili X, Y è:

$$\rho = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

- Se la correlazione è 0 allora non c'è relazione lineare
- Se la correlazione è >0 allora c'è una relazione lineare positiva
    - Quando una assumerà valori alti allora anche l'altra assumerà valori alti
    - P=1 dipendenza lineare perfetta positiva
- Se la correlazione è <0 c'è una relazione lineare negativa
    - Quando una assumerà valori altri l'altra assumerà valori bassi
    - P=-1 dipendenza lineare perfetta negativa