Esempio

Se abbiamo una coppia di variabili casuali discrete, è possibile definire la funzione di massa di probabilità congiunta. Abbiamo visto che è possibile ricavare le funzioni di massa marginale.

coppia di v.c. discrete

(X, Y) $\quad X \in \{x_1, x_2, \dots x_m\}$

$Y \in \{y_1, y_2, \dots y_n\}$

$p(a,b) = P(X=a, Y=b)$

$$p_X(a) = P(X=a, Y \le +\infty) = \sum_{j=1}^{n} P(X=a, Y=y_j)$$

$$p_Y(b) = P(X \le +\infty, Y=b) = \sum_{k=1}^{m} P(X=x_k, Y=b)$$

Perché bisogna fare la somma della funzione di massa per tutti valori che la variabile che non ci interessa assume? Partiamo da un punto di vista diverso e facciamo prima il calcolo delle funzioni marginali.

es. 1 Estrazione senza reimmissione di 2 palline da una scatola che ne contiene 3 numerate (1, 2, 3)

X = somma dei numeri estratti

Y = primo numero estratto

L'esempio ci permette di scrivere tutte le coppie che possono uscire nella prima e seconda estrazione senza reimmissione.

$(1,2) \quad (1,3)$

$(2,1) \quad (2,3)$

$(3,1) \quad (3,2)$

$X =$ 'somma dei numeri estratti'

$X \in \{3, 4, 5\}$

$$p_X(3) = P((1,2) \cup (2,1)) =$$
$$= P((1,2)) + P((2,1)) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$$

$$p_X(4) = P((1,3) \cup (3,1)) = P((1,3)) + P((3,1)) = \frac{2}{6}$$

$$p_X(5) = P((2,3) \cup (3,2)) = P((2,3)) + P((3,2)) = \frac{2}{6}$$

$Y \in \{1, 2, 3\}$

$$p_Y(1) = P((1,2) \cup (1,3)) = P((1,2)) + P((1,3)) = \frac{2}{6}$$

$$p_Y(2) = P((2,1) \cup (2,3)) = P((2,1)) + P((2,3)) = \frac{2}{6}$$

$$p_Y(3) = P((3,1) \cup (3,2)) = P((3,1)) + P((3,2)) = \frac{2}{6}$$

Adesso costruiamo la funzione di massa di probabilità congiunta.

$$p(4,3) = P(X=4, Y=3) = P((3,1)) = \frac{1}{6}$$

$$p(3,3) = P(X=3, Y=3) = 0$$

$$p(5,3) = P(X=5, Y=3) = P((3,2)) = \frac{1}{6}$$

$$p(3,1) = P(X=3, Y=1) = P((1,2)) = \frac{1}{6}$$

$$p(3,2) = P(X=3, Y=2) = P((2,1)) = \frac{1}{6}$$

$$p(3,3) = P(X=3, Y=3) = 0$$
$$p(5,1) = P(X=5, Y=1) = 0$$
$$p(5,2) = P(X=5, Y=2) = P((2,3)) = \frac{1}{6}$$
$$p(4,1) = P(X=4, Y=1) = P((1,3)) = \frac{1}{6}$$
$$p(4,2) = P(X=4, Y=2) = 0$$

Ora scriviamo la tabella.

Y \ X	3	4	5	
1	$\frac{1}{6}$	$\frac{1}{6}$	0	
2	$\frac{1}{6}$	0	$\frac{1}{6}$	
3	0	$\frac{1}{6}$	$\frac{1}{6}$	

Calcolo la funzione marginale di Y.

$$p_Y(1) = P(Y=1) = P((1,2) \cup (1,3))$$
$$p_Y(2) = P(Y=2) = P((2,1) \cup (2,3))$$
$$p_Y(3) = P(Y=3) = P((3,1) \cup (3,2))$$

Y \ X	3	4	5	
1	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{2}{6}$
2	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{2}{6}$
3	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$

La somma è un modo per dire che quando mi riferisco a una sola variabile conto più casi che sono raggruppati nella funzione di massa congiunta in più valori diversi.

Variabili casuali identicamente distribuite

$$(X_1, X_2, \dots X_N) \quad \text{i.i.d.}$$

i → indipendenti

i.d. → identicamente distribuite

$$E[X_k] = \mu \qquad k = 1, \dots N$$
$$Var(X_k) = \sigma^2 \qquad k = 1, \dots N$$

$$\overline{X} = \frac{\sum_{k=1}^{N} X_k}{N} \qquad \text{media campionaria v.c.}$$

$$\left.\begin{array}{l} E[\overline{X}] = \mu \\ Var(\overline{X}) = \dfrac{\sigma^2}{N} \end{array}\right\} \text{vedi lezione precedente}$$

Questo spiega perché in un esperimento facciamo tante misurazioni. Il valor medio teorico rimane lo stesso, ma la varianza (misura di quanto i dati si allontanano dal valor medio) cala e cala 1/N.

Cosa vuol dire la media della media campionaria? La media della media campionaria è il valore teorico che si vuole misurare, immaginando un errore che può essere positivo o negativo. Mu è il valore teorico e sigma quadro misura l'errore al quadrato e quanto si allontana dal valor medio. Mu è quello che con gli esperimenti vogliamo arrivare a stimare. Il fatto che la media campionaria abbia lo stesso valor medio, significa che può essere usata per stimare mu come possono essere usate le singole misure, con il vantaggio che l'errore è stato ridotto (sommando tanti valori insieme si riduce l'errore).

Tutto ciò è alla base della legge dei grandi numeri. Ma per dimostrarla e dimostrare il corollario di Bernoulli che sono alla base della parte sperimentale della definizione frequentista di probabilità, bisogna introdurre dei teoremi molto facili da dimostrare.

Disuguaglianza di Markov

Disuguaglianza di Markov

Data una v.c. $X \geq 0$ con $E[X] = \mu$ e una costante $a > 0$

$$P(X > a) \leq \frac{\mu}{a}$$

Dobbiamo guardare due aspetti di questo enunciato: il significato di come si usa questa disuguaglianza e la dimostrazione.

Mu/a è interessante, da informazioni al problema, solo se è più piccolo di 1. Altrimenti, la probabilità che X sia maggiore di a è minore o uguale di un numero più grande, non da niente di nuovo.

$\boxed{\text{DIM}}$ CASO CONTINUO

$$\mu = E[X] = \int_{-\infty}^{+\infty} x f(x)\, dx$$

Finora è solo una definizione. Ricordiamoci che X non è negativo.

$$\underset{\substack{\downarrow \\ X \geq 0 \\ \Rightarrow f(x) = 0 \text{ se } x < 0}}{=} \int_{0}^{+\infty} x f(x)\, dx$$

$$= \int_0^a \underbrace{x}_{\substack{x \geq 0 \\ \text{in } [0,a]}} \underbrace{f(x)}_{\geq 0}\, dx + \int_a^{+\infty} x f(x)\, dx \geq \int_a^{+\infty} x f(x)\, dx$$

$$= \int_a^{+\infty} \boxed{x} f(x)\, dx \geq \int_a^{+\infty} a\, f(x)\, dx = a \int_a^{+\infty} f(x) dx =$$

$x \geq a$, $f(x) \geq 0$

$$= a\, P(X \geq a)$$

$$\Downarrow$$

$$\mu \geq a\, P(X \geq a)$$

$$\Rightarrow P(X \geq a) \leq \frac{\mu}{a}$$

CASO DISCRETO SIMILE $\square$

Disuguaglianza di Cebycev

DISUGUAGLIANZA DI ČEBYČEV

Data una v.c. X con $E[X] = \mu$ e $Var(X) = \sigma^2$ e data $r > 0$

$$P(|X - E[X]| \geq r) \leq \frac{Var(X)}{r^2}$$

DIM

$$P(|X - E[X]| \geq r) = P(|X - \mu|^2 \geq r^2) =$$

$$= P((X-\mu)^2 \geq r^2) \leq \frac{E[(X-\mu)^2]}{r^2}$$

DISUG. MARKOV

$(X-\mu)^2$ v.c ≥ 0

$$= \frac{Var(X)}{r^2}$$

$$\Downarrow$$

$$P(|X - E[X]| \geq r) \leq \frac{Var(X)}{r^2} \quad \square$$

N.B. Le due disug. precedenti danno delle stime significative delle probabilità solo se

$\frac{\mu}{a} < 1$ per disug. di Markov

$\frac{\sigma^2}{r^2} < 1$ per disug. di Čebyčev

N.B. Le disug. si possono usare anche nel caso in cui di X si conosca solo $E[X]$ e/o $Var(X)$

es. 1

Una catena di produzione produce in media 50 pezzi a settimana.

1) Stimare la prob. che nella prossima settimana il numero di pezzi prodotti non sia inferiore a 75

X = 'n° pezzi prodotti la prossima settimana' ≥ 0

$$P(X \geq 75) \underset{\text{DISUG. DI MARKOV}}{\leq} \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

2) Se si suppone che $Var(X) = 25$, cosa si può dire della probabilità

$$P(40 \leq X \leq 60) \, ?$$

Con valor medio e varianza non si può stimare la probabilità (servirebbe la funzione di massa o di ripartizione).

$$P(40 < X < 60) = P(40-50 < X-50 < 60-50) =$$
$$= P(-10 < X-50 < 10)$$
$$= P(|X-50| < 10)$$

Come usare la disuguaglianza di Cebycev in questo caso? Considero il complementare.

$$= 1 - P(|X-50| \geq 10) \geq 1 - \frac{Var(X)}{100} = 1 - \frac{25}{100} = 1 - \frac{1}{4} = \frac{3}{4}$$

$$P(|X-50| \geq 10) \underset{\substack{\downarrow \\ \text{DISUG.} \\ \text{DI ČEBYČEV}}}{\leq} \frac{Var(X)}{100} \Rightarrow -P(|X-50| \geq 10) \geq -\frac{Var(X)}{100}$$

Oss. $P(|X - E[X]| < a) =$
$$= 1 - \underline{P(|X - E[X]| \geq a)} \geq 1 - \frac{Var(X)}{a^2}$$

Questa stima viene impiegata abbastanza di frequente in statistica quando si conosce soltanto il valor medio e la varianza, ma anche nello studio della probabilità non si conosce la funzione densità o di massa. Capiterà negli esercizi.

Legge dei grandi numeri

È uno dei pilastri della teoria della probabilità ed è un collegamento tra probabilità e statistica, ma la dimostrazione è banale. Quindi un grande risultato con una cosa semplice.

Data una successione di v.c. $X_1, X_2, \ldots X_N$ i.i.d. con $E[X_K] = \mu$ e $Var(X_K) = \sigma^2$ $(K = 1, \ldots N)$, la loro media (campionaria) aritmetica converge in probabilità al valor medio μ, ovvero $\forall \varepsilon > 0$

$$P(|\bar{X} - \mu| \geq \varepsilon) \xrightarrow[N \to +\infty]{} 0$$

con $\bar{X} = \dfrac{\sum_{K=1}^{N} X_K}{N}$

Cosa vuol dire? La probabilità che il valore assoluto della media campionaria meno la media teorica sia maggiore o uguale ad eplison, cioè i loro valori differiscano di più di epsilon, va a 0 se considero infinite variabili, infiniti esperimenti.

La probabilità a 0 non vuol dire sempre impossibilità, vuol dire anche che è così raro che non ci si aspetti che capiti.

La convergenza in probabilità (cioè il valore assoluto della media campionaria – mu maggiore o uguale di epsilon) non è la convergenza di analisi. Non sto dimostrando che la variabile campionaria tenda a mu, non si può dimostrare, la media campionaria è una variabile casuale, non possiamo essere in grado di dire che sicuramente la media campionaria fa mu. Dico che la probabilità che la media campionaria non faccia mu quando N va a infinito è 0, perché epsilon si può prendere piccolo a piacere (anche miliardesimi di miliardesimi) e la probabilità resta a 0.

Un concetto nuovo: la convergenza in probabilità. Non siamo nel mondo della certezza, dell'analisi, ma nel mondo dell'incertezza. Non possiamo dimostrare che una cosa tenda ad un'altra nel mondo della probabilità e della statistica.

$\boxed{\text{DIM}}$ $E[\bar{X}] = \mu$ e $Var(\bar{X}) = \dfrac{\sigma^2}{N}$ (per lezione precedente)

$$|\bar{X} - E[\bar{X}]| = |\bar{X} - \mu|$$

$$P(|\bar{X} - \mu| \geq \varepsilon) = P(|\bar{X} - E[\bar{X}]| \geq \varepsilon) \underset{\text{Čebyšev}}{\leq} \frac{Var(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{N\varepsilon^2}$$

$$\Downarrow$$

$$P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{N\varepsilon^2} \xrightarrow[N \to +\infty]{} 0$$

$$\Downarrow$$

$$P(|\bar{X} - \mu| \geq \varepsilon) \leq 0 \text{ se } N \to +\infty$$

MA PROBABILITÀ ≥ 0

$$\Rightarrow P(|\bar{X} - \mu| \geq \varepsilon) \xrightarrow[N \to +\infty]{} 0 \qquad \square$$