Probabilità e Statistica[1]

Isadora Antoniano-Villalobos
isadora.antoniano@unive.it

Laurea in Informatica
(Data science/ Tecnologie e scienze dell'informazione)
Università Ca' Foscari di Venezia

Anno academico 2023/2024

[1]Materiale didattico redatto da: Isadora Antoniano-Villalobos & Federica Giummolè

Distribuzioni congiunte

Introduzione

Fino ad ora ci siamo occupati delle distribuzioni di probabilità di variabili aleatorie prese singolarmente. Invece nella pratica si è spesso interessati a formulare delle affermazioni di tipo probabilistico riguardanti **due o più variabili aleatorie contemporaneamente**.

- Di seguito introduciamo i concetti fondamentali sulle distribuzioni congiunte, considerando il semplice caso di **coppie** di variabili aleatorie (X, Y) ➔ **distribuzioni bivariate**.
- Quello che diremo si può facilmente generalizzare a **vettori** di variabili aleatorie $(X_1, \ldots, X_n)$ di dimensione $n > 1$ ➔ **distribuzioni multivariate**.

Distribuzioni congiunte

funzione di ripartizione congiunta

La **funzione di ripartizione congiunta** della coppia di variabili aleatorie (X, Y) è una funzione $F : \mathbb{R}^2 \to [0, 1]$ definita così:

$$F(x, y) = \mathbb{P}\left[X \leq x, Y \leq y\right], \quad \forall (x, y) \in \mathbb{R}^2.$$

Dalla f.r. congiunta si possono facilmente ottenere le f.r. delle singole v.a. X e Y:

$$\begin{aligned} F_X(x) &= \mathbb{P}\left[X \leq x\right] \\ &= \mathbb{P}\left[X \leq x, Y < \infty\right] \\ &= \mathbb{P}\left[\lim_{y\to\infty}\{X \leq x, Y \leq y\}\right] \\ &= \lim_{y\to\infty} \mathbb{P}\left[X \leq x, Y \leq y\right] \\ &= \lim_{y\to\infty} F(x, y), \end{aligned}$$

Distribuzioni congiunte

e, in modo simile,

$$F_Y(y) = \lim_{x \to \infty} F(x, y).$$

- Si noti quindi che la f.r. congiunta di (X, Y) contiene anche tutta l'informazione riguardante le distribuzioni di probabilità delle singole v.a. X e Y, ovvero le **distribuzioni marginali**.

- Invece il viceversa non è vero: non è possibile ricavare la f.r. congiunta a partire dalle marginali, a meno che non si abbia ulteriore informazione sulla relazione fra le due variabili.

Calcolo di probabilità congiunte

A tutte le domande riguardanti probabilità congiunte di (X, Y) si può rispondere usando la f.r. congiunta $F(x, y)$.

Ad esempio,

$$\mathbb{P}\left[a_1 < X \leq a_2,\, b_1 < Y \leq b_2\right] = F(a_2, b_2) + F(a_1, b_1) - F(a_1, b_2) - F(a_2, b_1).$$

Funzione di probabilità congiunta

Funzione di probabilità congiunta

Se le v.a. X e Y sono entrambe discrete, si definisce la **funzione di probabilità congiunta**

$$p(x,y) = \mathbb{P}\left[X = x, Y = y\right], \quad \forall (x,y) \in \mathbb{R}^2.$$

Le **funzioni di probabilità marginali** si ottengono sommando rispetto a tutti i valori possibili dell'altra variabile:

$$p_X(x) = \sum_{y:p(x,y)>0} p(x,y)$$

e

$$p_Y(y) = \sum_{x:p(x,y)>0} p(x,y).$$

Funzione di probabilità congiunta

La funzione di probabilità congiunta e le marginali si possono rappresentare nella stessa **tabella a due entrate**:

	Y			
X	y_1	y_2	$\dots$	$p_X(x)$
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$\dots$	$p_X(x_1)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$\dots$	$p_X(x_2)$
$\vdots$		$\dots$		$\vdots$
$p_Y(y)$	$p_Y(y_1)$	$p_Y(y_2)$	$\dots$	1

Funzione di probabilità congiunta

Esempio: Un programma consiste di due moduli. Il numero di errori X nel primo modulo e il numero di errori Y nel secondo modulo hanno distribuzione congiunta: $p(0,0) = p(0,1) = p(1,0) = 0.2$, $p(1,1) = p(1,2) = p(1,3) = 0.1$, $p(0,2) = p(0,3) = 0.05$.

☞ Per prima cosa rappresentiamo la funzione di probabilità congiunta e le marginali in una tabella:

	Y				
X	0	1	2	3	$p_X(x)$
0	0.20	0.20	0.05	0.05	0.5
1	0.20	0.10	0.10	0.10	0.5
$p_Y(y)$	0.40	0.30	0.15	0.15	1

Funzione di probabilità congiunta

- La probabilità che non ci siano errori nel primo modulo, si calcola sommando tutte le probabilità congiunte della riga corrispondente a $X = 0$:

$$\mathbb{P}\left[X = 0\right] = \sum_{y=0}^{3} p(0, y) = 0.20 + 0.20 + 0.05 + 0.05 = 0.5.$$

- La probabilità che nel primo modulo ci siano meno errori che nel secondo modulo è

$$\begin{aligned}\mathbb{P}\left[X < Y\right] &= 1 - \mathbb{P}\left[X \geq Y\right] \\ &= 1 - \mathbb{P}\left[X = 0, Y = 0\right] - \mathbb{P}\left[X = 1, Y = 0\right] - \mathbb{P}\left[X = 1, Y = 1\right] \\ &= 1 - (0.2 + 0.2 + 0.1) = 0.5.\end{aligned}$$

Densità congiunta

Densità congiunta

Si dice che X e Y sono **congiuntamente continue** se esiste una funzione $f : \mathbb{R}^2 \to \mathbb{R}$ integrabile e tale che:

1. $f(x, y) \geq 0, \; \forall (x, y) \in \mathbb{R}^2$,
2. $\iint_{\mathbb{R}^2} f(x, y) dx dy = 1$.

f è la **funzione di densità congiunta** di (X, Y) e si utilizza per calcolare probabilità riguardanti le due variabili congiuntamente.

Ad esempio,

- se $C \subset \mathbb{R}^2$,

$$\mathbb{P}\left[(X, Y) \in C\right] = \iint_C f(x, y) dx\, dy;$$

Densità congiunta

- se $A, B \subset \mathbb{R}$,

$$\mathbb{P}[X \in A, Y \in B] = \int_B \left(\int_A f(x,y)dx \right) dy.$$

➔ In particolare, dalla densità congiunta si può ricavare la f.r. congiunta:

$$F(x,y) = \mathbb{P}[X \leq x, Y \leq y] = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t)ds\,dt.$$

➔ Viceversa, dalla f.r. congiunta si ottiene facilmente (in tutti i punti in cui F è derivabile)

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}.$$

Densità congiunta

Le **funzioni di densità marginali** si ricavano integrando la densità congiunta rispetto all'altra variabile:

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy$$

e

$$f_Y(y) = \int_{\mathbb{R}} f(x, y) dx.$$

Densità congiunta

Esempio: La densità congiunta di X e Y è

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y}, & x > 0,\ y > 0 \\ 0 & \text{altrove.} \end{cases}$$

- Allora:

$$\begin{aligned} \mathbb{P}\left[X > 1, Y < 1\right] &= \int_0^1 \left(\int_1^{+\infty} 2e^{-x}e^{-2y}dx \right) dy \\ &= \int_0^1 2e^{-2y}(-e^{-x})_1^{+\infty} dy = e^{-1}\int_0^1 2e^{-2y}dy \\ &= e^{-1}(-e^{-2y})_0^1 = e^{-1}(1-e^{-2}), \end{aligned}$$

Densità congiunta

$$\mathbb{P}[X < Y] = \iint_{\{(x,y):x<y\}} 2e^{-x}e^{-2y}dx\,dy$$
$$= \int_0^{+\infty}\left(\int_0^y 2e^{-x}e^{-2y}dx\right)dy = \cdots = 1/3,$$

$$\mathbb{P}[X < 3] = \int_0^3 e^{-x}\left(\int_0^{+\infty} 2e^{-2y}dy\right)dx$$
$$= \int_0^3 e^{-x}dx = 1 - e^{-3}.$$

Variabili aleatorie indipendenti

Il concetto di indipendenza di eventi, già visto in probabilità elementare, si estende anche alle v.a.

Variabili aleatorie indipendenti

Si dice che due v.a. sono **indipendenti** se

$$\mathbb{P}\left[X \in A, Y \in B\right] = \mathbb{P}\left[X \in A\right] \mathbb{P}\left[Y \in B\right], \quad \forall A, B \subset \mathbb{R}.$$

Si può dimostrare che questo è equivalente a dire che:

$$F(x, y) = F_X(x) F_Y(y), \quad \forall (x, y) \in \mathbb{R}^2,$$

ovvero

$$p(x, y) = p_X(x) p_Y(y), \quad \forall (x, y) \in \mathbb{R}^2$$ ➔ caso discreto

$$f(x, y) = f_X(x) f_Y(y), \quad \forall (x, y) \in \mathbb{R}^2$$ ➔ caso continuo

Variabili aleatorie indipendenti

Esercizio:

$$f(x,y) = \begin{cases} 24xy, & x, y \in (0,1),\ x+y \in (0,1) \\ 0 & \text{altrove.} \end{cases}$$

X e Y sono indipendenti?

Rappresentate graficamente il dominio di (X, Y) e provate a rispondere.

Variabili aleatorie indipendenti

Esempio: Torniamo all'esempio del numero di errori nei due moduli di un programma:

	Y				
X	0	1	2	3	$p_X(x)$
0	0.20	0.20	0.05	0.05	0.5
1	0.20	0.10	0.10	0.10	0.5
$p_Y(y)$	0.40	0.30	0.15	0.15	1

Sono indipendenti le due v.a. X e Y?

☞ No, perchè, ad esempio,

$$p(1,1) = 0.10 \neq p_X(1)p_Y(1) = 0.50 \cdot 0.30.$$

Variabili aleatorie indipendenti

Esempio: Consideriamo il caso in cui X e Y sono v.a. indipendenti con funzioni di probabilità marginali:

x	0	1	2	3
$p_X(x)$	0.5	0.3	0.1	0.1
$p_Y(x)$	0.7	0.2	0.1	0

Se sono note le distribuzioni marginali di X e Y e inoltre si sa che le due variabili sono indipendenti, allora è possibile ricostruire la distribuzione congiunta di (X, Y).

Variabili aleatorie indipendenti

- La funzione di probabilità congiunta si può calcolare sfruttando l'indipendenza, in questo modo:

$$p(0,0) = \mathbb{P}[X=0, Y=0] = \mathbb{P}[X=0] \cdot \mathbb{P}[Y=0] = 0.5 \cdot 0.7 = 0.35,$$

$$p(0,1) = \mathbb{P}[X=0, Y=1] = \mathbb{P}[X=0] \cdot \mathbb{P}[Y=1] = 0.5 \cdot 0.2 = 0.1,$$

e così via.
La tabella delle probabilità congiunte è dunque:

	Y				
X	0	1	2	3	$p_X(x)$
0	0.35	0.10	0.05	0	0.5
1	0.21	0.06	0.03	0	0.3
2	0.07	0.02	0.01	0	0.1
3	0.07	0.02	0.01	0	0.1
$p_Y(y)$	0.7	0.2	0.1	0	1

Variabili aleatorie indipendenti

- Con la funzione di probabilità congiunta possiamo calcolare tutte le probabilità riguardanti X e Y congiuntamente.
  Possiamo anche calcolare, ad esempio, la distribuzione della v.a. $X+Y$. Infatti:

$$\mathbb{P}[X+Y=0] = \mathbb{P}[X=0, Y=0] = 0.35,$$

$$\begin{aligned} \mathbb{P}[X+Y=1] &= \mathbb{P}[X=0, Y=1] + \mathbb{P}[X=1, Y=0] \\ &= 0.21 + 0.10 = 0.31 \end{aligned}$$

e così via.

Variabili aleatorie indipendenti

- Completate i calcoli e controllate che

$$p_{X+Y}(w) = \begin{cases} 0.35 & \text{if } w = 0 \\ 0.31 & \text{if } w = 1 \\ 0.18 & \text{if } w = 2 \\ 0.12 & \text{if } w = 3 \\ 0.03 & \text{if } w = 4 \\ 0.01 & \text{if } w = 5. \end{cases}$$

Variabili aleatorie indipendenti

Esempio: Un programma è formato da due blocchi che vengono processati sequenzialmente durante la compilazione. Ogni blocco viene processato in un tempo medio di 5 minuti e si assume che i tempi di compilazione dei due blocchi, X e Y, siano indipendenti e con distribuzione esponenziale.

☞ Per l'indipendenza, la densità congiunta di X e Y è

$$f(x,y) = f_X(x) f_Y(y) = \frac{1}{5} e^{-x/5} I_{(0,+\infty)}(x) \cdot \frac{1}{5} e^{-y/5} I_{(0,+\infty)}(y).$$

Variabili aleatorie indipendenti

Allora, la probabilità che l'intero programma venga compilato in meno di 12 minuti è:

$$\begin{aligned}\mathbb{P}[X+Y<12] &= \iint_{\{(x,y):x+y<12\}} f(x,y)\,dx\,dy \\ &= \int_0^{12}\left(\int_0^{12-x}\frac{1}{5}e^{-y/5}\,dy\right)\frac{1}{5}e^{-x/5}\,dx \\ &= \int_0^{12}\left(1-e^{-(12-x)/5}\right)\frac{1}{5}e^{-x/5}\,dx \\ &= \left(1-e^{-12/5}\right)-\frac{12}{5}e^{-12/5} = 0.69156.\end{aligned}$$

☞ Qual è la probabilità che il primo blocco del programma venga compilato in un tempo inferiore rispetto al secondo blocco?

Distribuzioni condizionate: variabili discrete

Funzioni di probabilità condizionata

Siano X e Y due v.a. discrete con funzione di probabilità congiunta $p(x, y)$.
La **funzione di probabilità condizionata** di X dato $Y = y$ è

$$p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)}, \quad \forall y \text{ t.c. } p_Y(y) > 0.$$

Allo stesso modo, la **funzione di probabilità condizionata** di Y dato $X = x$ è

$$p_{Y|X}(y|x) = \frac{p(x, y)}{p_X(x)}, \quad \forall x \text{ t.c. } p_X(x) > 0.$$

Distribuzioni condizionate: variabili discrete

- In effetti, ad esempio, se $p_Y(y) > 0$,

$$\begin{aligned} p_{X|Y}(x|y) &= \mathbb{P}[X = x | Y = y] \\ &= \frac{\mathbb{P}[X = x, Y = y]}{\mathbb{P}[Y = y]} = \frac{p(x,y)}{p_Y(y)}, \end{aligned}$$

- Se X e Y sono indipendenti, allora

$$p_{X|Y}(x|y) = p_X(x) \text{ e } p_{Y|X}(y|x) = p_Y(y).$$

Esempio: Siano X e Y due v.a. discrete tali che: $p(0,0) = 0.4$, $p(0,1) = 0.2$, $p(1,0) = 0.1$, $p(1,1) = 0.3$. Allora,

	Y		
X	0	1	$p_X(x)$
0	0.4	0.2	0.6
1	0.1	0.3	0.4
$p_Y(y)$	0.5	0.5	1

Distribuzioni condizionate: variabili discrete

- La funzione di probabilità di X dato $Y = 1$ si calcola utilizzando le informazioni contenute nella colonna corrispondente a $Y = 1$:

$$p_{X|Y}(0|1) = \frac{p(0,1)}{p_Y(1)} = \frac{0.2}{0.5} = 0.4$$

e

$$p_{X|Y}(1|1) = \frac{p(1,1)}{p_Y(1)} = \frac{0.3}{0.5} = 0.6,$$

in modo che

$$X|Y = 1 \sim p_{X|Y}(x|1) = \begin{cases} 0.4 & \text{if } x = 0 \\ 0.6 & \text{if } x = 1 \end{cases}$$

Essendo questa distribuzione diversa dalla marginale di X, concludiamo che le due variabili non sono indipendenti.

Distribuzioni condizionate: variabili continue

Funzioni di densità condizionata

Siano X e Y v.a. continue con funzione di densità congiunta $f(x, y)$.
La **funzione di densità condizionata** di X dato $Y = y$ è

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}, \quad \forall y \text{ t.c. } f_Y(y) > 0.$$

Allo stesso modo, la **funzione di densità condizionata** di Y dato $X = x$ è

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, \quad \forall x \text{ t.c. } f_X(x) > 0.$$

Allora,

$$\mathbb{P}[X \in A | Y = y] = \int_A f_{X|Y}(x|y)\, dx, \quad \mathbb{P}[Y \in B | X = x] = \int_B f_{Y|X}(y|x)\, dy.$$

Distribuzioni condizionate: variabili continue

- Se X e Y sono indipendenti, allora

$$f_{X|Y}(x|y) = f_X(x) \text{ e } f_{Y|X}(y|x) = f_Y(y).$$

Esempio: Sia $f(x,y) = \frac{15}{2}x(2-x-y)I_{(0,1)\times(0,1)}(x,y)$. Allora, dalla definizione,

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad \forall y \in (0,1).$$

- Calcoliamo prima la densità marginale di Y:

$$f_Y(y) = \int_0^1 \frac{15}{2}(2x - x^2 - xy)I_{(0,1)}(y)\,dx = \cdots = \frac{15}{2}\left(\frac{2}{3} - \frac{y}{2}\right)I_{(0,1)}(y).$$

Infine

$$f_{X|Y}(x|y) = \frac{x(2-x-y)I_{(0,1)}(x)}{\left(\frac{2}{3} - \frac{y}{2}\right)}, \quad \forall y \in (0,1).$$

Valore atteso di trasformazioni di (X, Y)

Come già abbiamo visto nel caso univariato, esiste un risultato che permette di calcolare il valore atteso di una trasformazione $g(X, Y)$ senza doverne calcolare la distribuzione.

Valore atteso di $g(X, Y)$

$$\mathbb{E}\left[g(X, Y)\right] = \begin{cases} \sum_y \sum_x g(x, y) p(x, y) & \text{caso discreto} \\ \iint_{\mathbb{R}^2} g(x, y) f(x, y)\, dx\, dy & \text{caso continuo} \end{cases}$$

Valore atteso di somma e prodotto

Due importanti conseguenze del precedente risultato sono:

1. $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
   In generale:
$$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i].$$
2. Se X e Y sono indipendenti, allora
$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X]\,\mathbb{E}[Y].$$

➔ **Esercizio:** Provate a dimostrarle.

Valore atteso di somma e prodotto

Esempio: Tornando all'esempio del numero di errori nei due moduli di un programma,

	Y				
X	0	1	2	3	$p_X(x)$
0	0.20	0.20	0.05	0.05	0.5
1	0.20	0.10	0.10	0.10	0.5
$p_Y(y)$	0.40	0.30	0.15	0.15	1

➔ È facile verificare che $\mathbb{E}[X] = 0.5$ e $\mathbb{E}[Y] = 1.05$.
Perciò il numero atteso di errori nell'intero programma è

$$\mathbb{E}[X+Y] = 1.55.$$

La covarianza

La covarianza

La **covarianza** fra due v.a. X e Y è il valore atteso di una particolare trasformazione $g(x, y) = (x - \mathbb{E}[X])(y - \mathbb{E}[Y])$:

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Sviluppando il prodotto dentro il valore atteso e semplificando alcuni termini, otteniamo una formula pratica per il calcolo della covarianza:

$$\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

➔ Se $\text{Cov}[X, Y] = 0$ diciamo che X e Y sono **incorrelate**.

Indipendenza e covarianza

Se X e Y sono indipendenti allora Cov $[X, Y] = 0$.

Il viceversa non è vero!

Esempio: Consideriamo

$$X = \begin{cases} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{cases} \quad \text{e} \quad Y = \begin{cases} 0 & X \neq 0 \\ 1 & X = 0 \end{cases}$$

Ovviamente, le due variabili non sono indipendenti, ma si dimostra facilmente che Cov $[X, Y] = 0$ e perciò che X e Y sono incorrelate.

Proprietà della covarianza

1. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
2. $\text{Cov}[X, X] = \text{Var}[X]$
3. $\text{Cov}[aX, Y] = a\text{Cov}[X, Y]$
4. $\text{Cov}[X, a] = 0$
5. $\text{Cov}\left[\sum_i X_i, \sum_j Y_j\right] = \sum_i \sum_j \text{Cov}[X_i, Y_j]$

Varianza di una somma

Dalle proprietà della covarianza, discende un risultato molto utile nelle applicazioni statistiche:

$$\mathrm{Var}\left[\sum_i X_i\right] = \mathrm{Cov}\left[\sum_i X_i, \sum_j X_j\right] = \sum_i \sum_j \mathrm{Cov}\left[X_i, X_j\right]$$
$$= \sum_i \mathrm{Var}\left[X_i\right] + \sum_i \sum_{j \neq i} \mathrm{Cov}\left[X_i, X_j\right]$$
$$= \sum_i \mathrm{Var}\left[X_i\right] + 2 \sum_i \sum_{j > i} \mathrm{Cov}\left[X_i, X_j\right].$$

☞ In particolare, se le X_i sono **a due a due indipendenti**, allora $\mathrm{Cov}\left[X_i, X_j\right] = 0$ per $i \neq j$ e

$$\mathrm{Var}\left[\sum_i X_i\right] = \sum_i \mathrm{Var}\left[X_i\right].$$

La correlazione

La correlazione

La **correlazione** fra due v.a. X e Y è definita come:

$$\text{Cor}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X] \cdot \text{Var}[Y]}},$$

sempre che $\text{Var}[X]$ e $\text{Var}[Y]$ siano diverse da 0.

Si ha che

$$-1 \leq \text{Cor}[X, Y] \leq 1.$$

Inoltre, se $Y = a + bX$ allora $\text{Cor}[X, Y] = \pm 1$, a seconda del segno di b.

➔ Per questa ragione si dice che $\text{Cor}[X, Y]$ misura l'intensità del legame lineare fra X e Y.

La correlazione

Esempio: Tornando all'esempio del numero di errori nei due moduli di un programma.

☞ Abbiamo

$$\begin{aligned}\text{Cov}[X,Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \\ &= 1\cdot 1\cdot 0.1 + 1\cdot 2\cdot 0.1 + 1\cdot 3\cdot 0.1 - 0.5\cdot 1.05 \\ &= 0.6 - 0.525 = 0.075\end{aligned}$$

e, essendo $\text{Var}[X] = 0.5^2$ e $\text{Var}[Y] = 1.0712^2$,

$$\text{Cor}[X,Y] = \frac{0.075}{0.5\cdot 1.0712} = 0.14.$$

La distribuzione normale bivariata

Distribuzione normale bivariata

Siano $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ Cor $[X, Y] = \rho$ e

$$\text{Cov}\,[X, Y] = \sigma_{X,Y} = \rho\sigma_X\sigma_Y.$$

Allora, in termini matriciali, scriviamo

$$\left[\begin{array}{c} X \\ Y \end{array}\right] \sim \mathcal{N}_2\left(\left[\begin{array}{c} \mu_X \\ \mu_Y \end{array}\right], \left[\begin{array}{cc} \sigma_X^2 & \sigma_{X,Y} \\ \sigma_{X,Y} & \sigma_Y^2 \end{array}\right]\right) \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$$

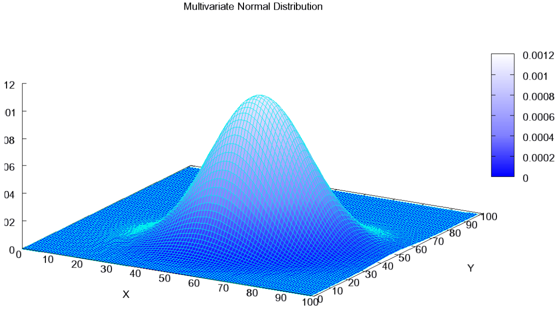
e la densità di $\mathbf{W} = (X, Y)$, posto $(x, y) = \mathbf{w}$, è

$$f(\mathbf{w}) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{w}-\boldsymbol{\mu})}.$$

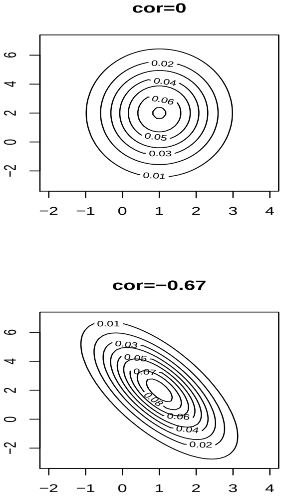
La distribuzione normale bivariata

Svolgendo i calcoli, otteniamo:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$
$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right\}.$$



La distribuzione normale bivariata



Somma di v.a.

Spesso, come abbiamo già visto in alcuno esempi, siamo interessati alla somma di due o più variabili aleatorie.

- Sappiamo già come calcolare il valore atteso e la varianza di una somma di v.a..
- Tuttavia, calcolare l'intera distribuzione di una somma di v.a. può essere un'impresa davvero difficile.

Esempio: somma di Poisson indipendenti

Siano X e Y due v.a. indipendenti con distribuzione di Poisson di media rispettivamente λ e μ.

Somma di v.a.

- Siamo interessati a trovare la distribuzione di $Z = X + Y$.

$$\begin{aligned}
\mathbb{P}[Z = z] &= \sum_{x=0}^{z} \mathbb{P}[X = x, Y = z - x] = \sum_{x=0}^{z} \mathbb{P}[X = x]\,\mathbb{P}[Y = z - x] \\
&= \sum_{x=0}^{z} \frac{\lambda^x}{x!} e^{-\lambda} \frac{\mu^{z-x}}{(z-x)!} e^{-\mu} = \frac{e^{-(\lambda+\mu)}}{z!} \sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \lambda^x \mu^{z-x} \\
&= \frac{e^{-(\lambda+\mu)}}{z!} (\lambda + \mu)^z \,, \qquad z = 0, 1, 2, \ldots
\end{aligned}$$

Allora $Z \sim \mathcal{P}(\lambda + \mu)$.

➔ **Esercizio:** Per le variabili del lucido precedente, si calcoli la distribuzione condizionata di $X|Z = z$.
Si tratta di una distribuzione nota?

Esempio: somma di esponenziali indipendenti

Siano X e Y due v.a. indipendenti con distribuzione Esponenziale di parametro λ.

- Siamo interessati a trovare la distribuzione di $Z = X + Y$.
  Sia $t > 0$:

$$\begin{aligned}
\mathbb{P}[Z \leq t] &= \iint_{x+y\leq t} f(x,y)dx\,dy \\
&= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{t-y} f_X(x)dx \right) f_Y(y)dy \\
&= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{t} f_X(z-y)dz \right) f_Y(y)dy \\
&= \int_{-\infty}^{t} \left(\int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)dy \right) dz.
\end{aligned}$$

Somma di v.a.

Allora la densità di Z è:

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) dy \\
&= \int_{-\infty}^{+\infty} \lambda^2 e^{-\lambda(z-y)} e^{-\lambda y} I_{(0,+\infty)}(z-y) I_{(0,+\infty)}(y) dy \\
&= \int_0^z \lambda^2 e^{-\lambda(z-y)} e^{-\lambda y} I_{(0,+\infty)}(z) dy \\
&= \int_0^z \lambda^2 e^{-\lambda z} I_{(0,+\infty)}(z) dy = \lambda^2 z e^{-\lambda z} I_{(0,+\infty)}(z),
\end{aligned}
$$

che coincide con la funzione di densità di una Gamma$(\alpha = 2, \lambda)$.

Somma di v.a.

In generale, valgono i seguenti risultati:

1. La somma di n v.a. Bin $(1, p)$ indipendenti è una binomiale Bin (n, p).
2. La somma di n v.a. Po (λ_i), $i = 1, \ldots n$, indipendenti è una Po $(\sum_i \lambda_i)$.
3. La somma di n v.a. Exp (λ) indipendenti è una Ga (n, λ).
4. La somma di n v.a. N (μ_i, σ_i^2) indipendenti è una Normale N $(\sum_i \mu_i, \sum_i \sigma_i^2)$.

La media campionaria

Media campionaria

Siano $X_1, \ldots, X_n$ n v.a. indipendenti e tutte con la stessa distribuzione (i.i.d.). Scriviamo $\mathbb{E}\left[X_i\right] = \mu$ e $\mathrm{Var}\left[X_i\right] = \sigma^2$.
La **media campionaria** si definisce come

$$\overline{X}_n = \frac{\sum_i X_i}{n}$$

La media campionaria è una v.a. molto importante nelle applicazioni del calcolo delle probabilità in statistica.

- La sua distribuzione dipende dalla distribuzione delle X_i e, in generale, non è facile da calcolare.

La media campionaria

- Si possono però calcolare facilmente la media e la varianza della media campionaria:

$$\mathbb{E}\left[\overline{X}_n\right] = \mathbb{E}\left[\frac{\sum_i X_i}{n}\right] = \frac{\sum_i \mathbb{E}\left[X_i\right]}{n} = \frac{n\mu}{n} = \mu,$$

$$\text{Var}\left[\overline{X}_n\right] = \text{Var}\left[\frac{\sum_i X_i}{n}\right] = \frac{\sum_i \text{Var}\left[X_i\right]}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Due importanti teoremi

I due più importanti teoremi del calcolo delle probabilità parlano proprio della media campionaria e della sua distribuzione. Sono:

- La Legge dei grandi numeri
- Il Teorema del limite centrale.

Per illustrarli, dobbiamo introdurre un risultato preliminare e due concetti di convergenza di successioni di v.a..

La disuguaglianza di Chebyshev

Disuguaglianza di Chebyshev

Sia X una v.a. con valore atteso $\mathbb{E}[X]$ e $\text{Var}[X] < +\infty$. Allora, $\forall \epsilon > 0$,

$$\mathbb{P}[|X - \mathbb{E}[X]| > \epsilon] \leq \frac{\text{Var}[X]}{\epsilon^2}.$$

Esempio: Il numero di richieste giornaliere di collegamento ad un server è una v.a. Y con valore atteso 130 e varianza 50. Qual è la probabilità che in un giorno si colleghino fra i 100 e i 160 clienti?

$$\mathbb{P}[100 \leq Y \leq 160] = \mathbb{P}[|Y - 130| \leq 30] \geq 1 - \frac{50}{30^2} = 0.9444.$$

Convergenza di v.a.

Convergenza di v.a.

Sia $X_1, \ldots, X_n, \ldots$ una sequenza di v.a. con f.r. F_n e X un'ulteriore v.a. con f.r. F.

1. Diciamo che $\{X_n\}$ **converge in probabilità** a X e scriviamo $X_n \xrightarrow{p} X$, se
$$\mathbb{P}\left[|X_n - X| > \epsilon\right] \to 0, \quad \forall \epsilon > 0.$$
2. Diciamo che $\{X_n\}$ **converge in distribuzione** a X e scriviamo $X_n \xrightarrow{d} X$, se in ogni punto x di continuità per F
$$F_n(x) \to F(x).$$
3. Diciamo che $\{X_n\}$ **converge in quasi certamente** o **con probabilità 1** a X e scriviamo $X_n \to X \quad q.c.$, se
$$\mathbb{P}\left[X_n \to X\right] = 1.$$

Le leggi dei grandi numeri (LGN)

Legge debole dei grandi numeri

Sia $X_1, \dots, X_n, \dots$ una sequenza di v.a. i.i.d. con $\mathbb{E}\left[X_i\right] = \mu$ e $\text{Var}\left[X_i\right] = \sigma^2 < +\infty$. Allora

$$\overline{X}_n \xrightarrow{p} \mu.$$

Infatti, per la disuguaglianza di Chebyshev applicata alla media campionaria,

$$\mathbb{P}\left[|\overline{X}_n - \mu| > \epsilon\right] \leq \frac{\text{Var}\left[\overline{X}_n\right]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0.$$

Le leggi dei grandi numeri (LGN)

Legge forte dei grandi numeri

Sia $X_1, \ldots, X_n, \ldots$ una sequenza di v.a. i.i.d. con $\mathbb{E}\left[X_i\right] = \mu$. Allora

$$\overline{X}_n \to \mu \quad q.c..$$

La dimostrazione di questo risultato è più complessa...

Esempio: L'estrazione del Lotto

Immaginiamo di essere interessati all'evento che venga estratto il numero 53 nella ruota di Venezia. Indichiamo con X_i la variabile binaria che vale 1 se all'i-esima estrazione esce il numero 53 e 0 altrimenti. E' chiaro che

$$X_i \sim \mathcal{B}(1, p = 1/90).$$

Le leggi dei grandi numeri (LGN)

Per la legge dei grandi numeri, ci aspettiamo che la frequenza relativa delle volte che si estrae il 53 si avvicini sempre più a p al crescere del numero di estrazioni n. Infatti:

$$\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{p} \mathbb{E}\left[X_i\right] = p = 1/90.$$

Questo però non cambia la probabilità di estrarre 53 ad ogni estrazione, che rimane sempre uguale a $p = 1/90$, anche quando non si è osservato il 53 per molto tempo!

Il teorema del limite centrale (TLC)

Teorema del limite centrale (TLC)

Sia $X_1, \ldots, X_n, \ldots$ una sequenza di v.a. i.i.d. con $\mathbb{E}[X_i] = \mu$ e Var $[X_i] = \sigma^2 < +\infty$. Allora

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1),$$

o, equivalentemente,

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Il teorema del limite centrale (TLC)

Il TLC permette, nella pratica, di approssimare la distribuzione della media campionaria (o di una somma) con la distribuzione normale:

$$\overline{X}_n \dot{\sim} \mathcal{N}(\mu, \sigma^2/n).$$

Ovvero

$$\sum_{i=1}^{n} X_i \dot{\sim} \mathcal{N}(n\mu, n\sigma^2).$$

☞ Si osservi che l'approssimazione normale per la binomiale altro non è che un'applicazione del TLC alla somma di v.a. di Bernoulli (la binomiale).

Esempio: TLC per Poisson

Il numero di studenti iscritti a un corso di laurea segue una distribuzione di Poisson di media 100. Se gli iscritti superano i 120 bisogna sdoppiare i corsi. Qual è la probabilità che ciò accada?

➔ $X \sim \text{Po}\,(100)$, perciò

$$\mathbb{P}\left[X \geq 120\right] = e^{-100} \sum_{i=120}^{+\infty} \frac{100^i}{i!} = \text{1-ppois(119,100)} = 0.0282.$$

Il teorema del limite centrale (TLC)

- Ma X si può vedere come una somma di 100 v.a. Po (1) e la sua distribuzione si può dunque approssimare con quella normale, in virtù del TLC.
  Con la correzione per continuità, otteniamo:

$$
\begin{aligned}
\mathbb{P}\left[X \geq 120\right] = \mathbb{P}\left[X \geq 119.5\right] &= \mathbb{P}\left[\sum_i X_i \geq 119.5\right] \\
&\doteq \mathbb{P}\left[Z \geq \frac{119.5 - 100}{\sqrt{100}}\right] \\
&= 1 - \Phi(1.95) = 0.0256.
\end{aligned}
$$