

Modelli probabilistici – variabili casuali

Le variabili casuali costituiscono il legame tra il calcolo della probabilità e gli strumenti di statistica descrittiva visti fino ad ora.

Idea: pensiamo al ripetersi di un esperimento casuale in cui viene rilevata la distribuzione di un fenomeno statistico

Possiamo considerare le frequenze relative osservate sui differenti valori come la probabilità con cui questi valori possono presentarsi

L'insieme dei valori assumibili dal fenomeno statistico e le probabilità associate costituiscono un *modello del fenomeno*

Le variabili casuali sono proprio tali modelli: studieremo le variabili casuali discrete (descrivono principalmente fenomeni di conteggio) e le variabili casuali continue (descrivono fenomeni che possiamo misurare)

Variabili casuali discrete

Dato un esperimento casuale i cui possibili esiti sono in numero finito (o al più numerabile), una variabile casuale è un risultato numerico attribuito a ogni possibile esito di tale esperimento

Def: La variabile casuale è una funzione che associa ad ogni evento elementare un valore numerico

$$X : \Omega \rightarrow \mathbb{R}$$

$$\omega \mapsto k$$

Esempio: abbiamo il seguente esperimento. Lanciamo due volte una moneta regolare. Le probabilità associate agli eventi elementari in un lancio sono $P(T) = P(C) = \frac{1}{2}$.

Gli eventi elementari in due lanci sono

$$\Omega = \{TT, TC, CT, CC\}$$

Sia X il numero di teste ottenuto nei due lanci, la variabile casuale assume i valori riportati in tabella

$\omega \in \Omega$	TT	TC	CT	CC
$P(\omega)$	$1/4$	$1/4$	$1/4$	$1/4$
$X(\omega) = x$	2	1	1	0

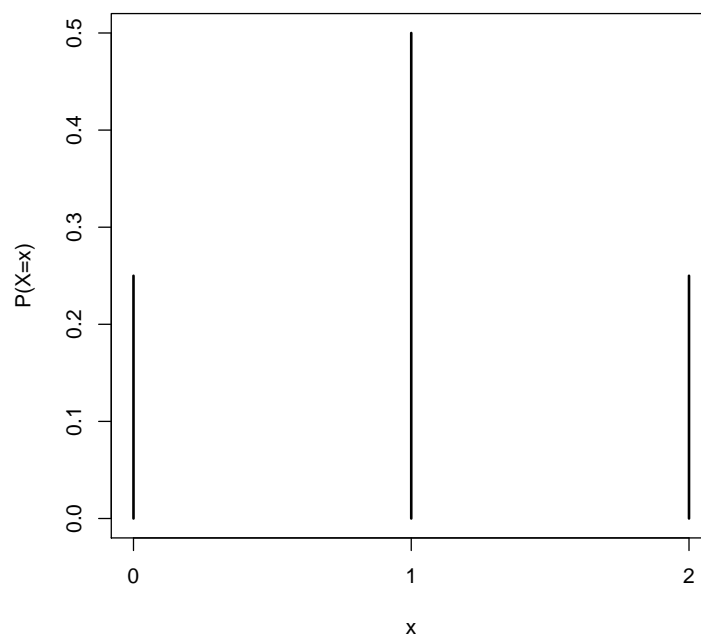
X rappresenta il risultato numerico dell'esperimento

Possiamo sintetizzare la tabella raccogliendo i valori distinti assunti da X e le probabilità con cui assume ogni valore

Abbiamo la tabella della **distribuzione di probabilità** di X

$X = x$	0	1	2
$P(X = x)$	1/4	1/2	1/4

La rappresentazione grafica della distribuzione di probabilità è analoga a quella della distribuzione di frequenza relativa per i caratteri discreti



Riassumendo:

- Una variabile casuale discreta è un modello matematico X
- Assume i valori x_1, x_2, \dots, x_k
- Con rispettive probabilità $P(X = x_i), i = 1, \dots, k$.

L'insieme delle $P(X = x_i), i = 1, \dots, k$ si chiama **distribuzione di probabilità** ed è tale per cui

$$\sum_{i=1}^k P(X = x_i) = 1$$

Possiamo scrivere $X = \{x_i, P(X = x_i), i = 1, \dots, k\}$

Quando si definisce una v.c. occorre sempre dare i valori che assume e la distribuzione di probabilità

$$\begin{array}{c|c|c|c|c|c} X = x & x_1 & x_2 & \dots & x_k & \\ \hline P(X = x) & p_1 & p_2 & \dots & p_k & 1 \end{array}$$

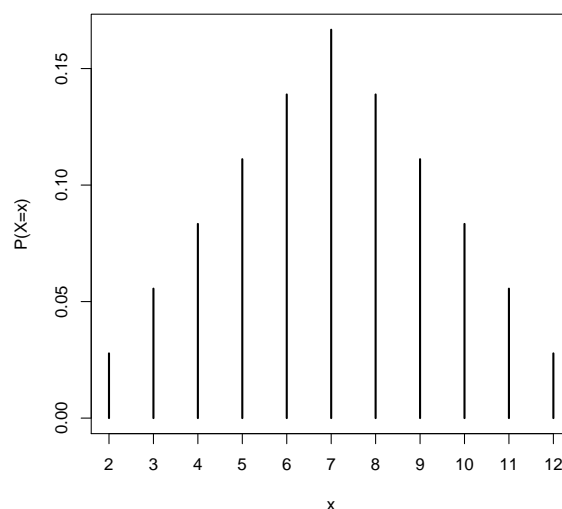
Esempio: consideriamo il lancio di due dadi e la v.c. X somma dei punteggi realizzati con i due dadi

$$\Omega = \{11, 12, 13, 14, 15, 16, \\ 21, 22, 23, 24, 25, 26, \\ 31, 32, 33, 34, 35, 36, \\ 41, 42, 43, 44, 45, 46, \\ 51, 52, 53, 54, 55, 56, \\ 61, 62, 63, 64, 65, 66\}$$

La variabile X può assumere uno dei valori tra 2 e 7 e la distribuzione di probabilità è

$X = x$	2	3	4	5	6	7	8	9	10	11	12	
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

La rappresentazione grafica della distribuzione



Esempio: Popolazione di 5 individui (5 aziende) di cui vogliamo studiare un certa caratteristica (numero di dipendenti). Effettuiamo un campionamento rilevando il dato solo su 3 aziende. Siamo interessati al numero mediano di dipendenti. Trovare la distribuzione di tale numero.

Supponiamo che le aziende abbiano $\{1, 3, 5, 7, 9\}$ dipendenti (a noi non è noto). I possibili campioni che si possono estrarre sono

$$\binom{5}{3} = \frac{5 \cdot 4 \cdot 3!}{3! \cdot 2!} = 10$$

$$\Omega = \{(1, 3, 5), (1, 3, 7), (1, 3, 9), (1, 5, 7), (1, 5, 9), (1, 7, 9), (3, 5, 7), (3, 5, 9), (3, 7, 9), (5, 7, 9)\}$$

Ciascun evento ha probabilità $1/10$ di comparire ($P(\omega) = 1/10$)

Costruiamo la variabile casuale $X = \text{mediana campionaria}$

La mediana vera della popolazione è 5

La v.c. X assumerà valori diversi a seconda del campione estratto.

Vogliamo calcolare la sua distribuzione

Associamo ad ogni elemento ω dello spazio campionario Ω il valore della Media campionaria $X(\omega)$

ω	$X(\omega)$
(1,3,5)	3
(1,3,7)	3
(1,3,9)	3
(1,5,7)	5
(1,5,9)	5
(1,7,9)	7
(3,5,7)	5
(3,5,9)	5
(3,7,9)	7
(5,7,9)	7

Per cui la variabile casuale X risulta avere la seguente distribuzione

x	$P(X = x)$
3	3/10
5	4/10
7	3/10

La frequenza con cui ci aspettiamo che un certo valore x_i della mediana si realizzi è pari alla probabilità dell'evento $P(X = x_i)$.

Variabili Casuali, fenomeni e osservazioni

I fenomeni statistici sono descritti teoricamente dalle variabili casuali

Noi abbiamo visto che un fenomeno statistico osservato ha una distribuzione

Anche la variabile casuale che descrive quel fenomeno ha una distribuzione (di probabilità)

La distinzione fondamentale è che la distribuzione di probabilità ci dice con quali *probabilità osserveremo* certi valori del fenomeno, mentre la distribuzione di frequenza fotografa *quanto è accaduto* all'interno di un particolare campione di dati di ampiezza n

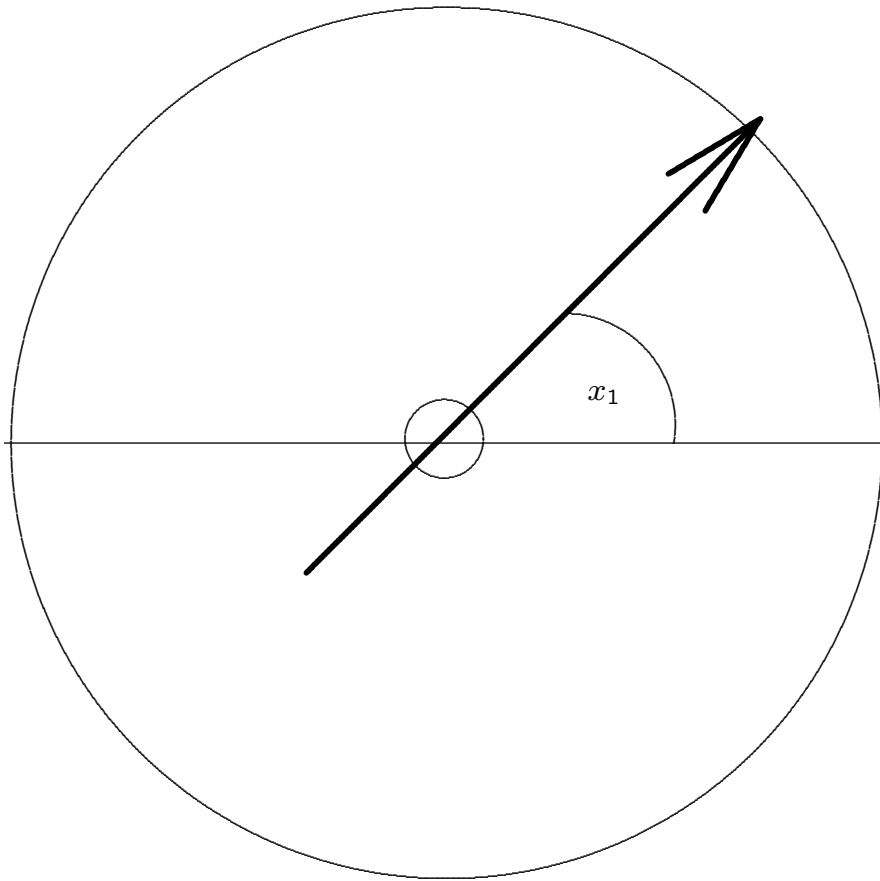
Nota la distribuzione di probabilità di una v.c. possiamo calcolare probabilità del tipo $P(X \leq x)$. Nell'esempio precedente:

$$\begin{aligned} P(X \leq 5) &= P(X = 3) + P(X = 5) \\ &= \frac{3}{10} + \frac{4}{10} = \frac{7}{10} = 0.7 \end{aligned}$$

In termini più generali si può usare la formula

$$P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$$

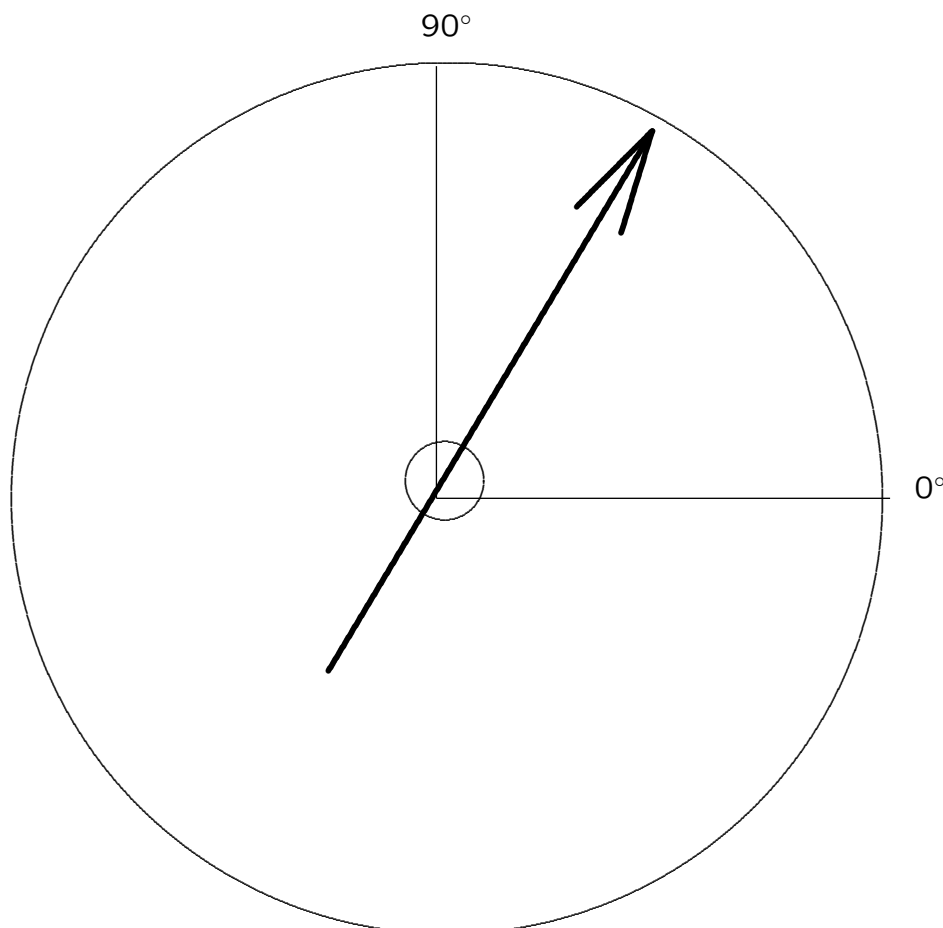
Variabili casuali continue



Sia X la variabile casuale che registra i gradi x_i , assumerà tutti valori dei gradi nell'intervallo $(0, 360)$. Quindi i casi possibili sono $\Omega = (0, 360)$. Se dobbiamo calcolare la probabilità che X assuma un certo valore x_i succede una cosa poco piacevole:

$$P(X = x_i) = \frac{\# \text{ casi favorevoli}}{\# \text{ casi possibili}} = \frac{1}{\infty} = 0$$

Proviamo un'altra strada: con quale probabilità la variabile casuale X , assumerà un valore tra 0 e 90°?



$$P(X \leq 90) = 1/4 = 0.25$$

Questa probabilità può essere calcolata come

$$P(X \leq 90) = \int_0^{90} \frac{1}{360} dx = \frac{90}{360} = 0.25$$

La funzione $f(x) = \frac{1}{360}$, per $0 \leq x < 360$ è la densità di X

Le variabili casuali continue sono definite a partire dall'insieme dei valori assumibili e dalle probabilità del tipo $P(X \leq x)$

Queste probabilità sono espresse attraverso la funzione $f(\cdot)$, detta *densità di probabilità*, come

$$P(X \leq x) = \int_{-\infty}^x f(u) du$$

Questa è un'estensione della formula

$$P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$$

Calcolare l'integrale corrisponde a calcolare l'area sotto la curva $f(\cdot)$ stessa. In generale non dovremo mai calcolare questi integrali perché esistono le *tavole* che contengono i valori di quegli integrali per i vari valori di x

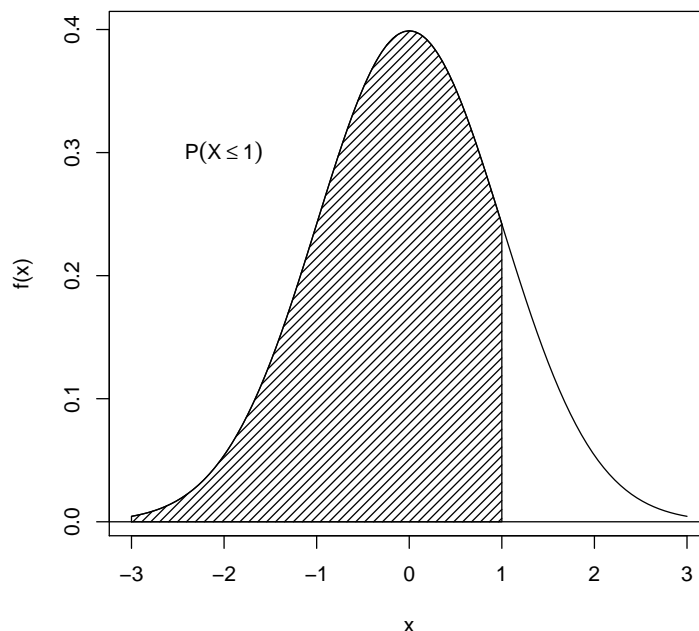
Esempio: Sia X la variabile casuale che assume i valori $(0, 360)$ con funzione di densità $f(x) = \frac{1}{360}$. Abbiamo

$$P(X \leq x) = \int_{-\infty}^x \frac{1}{360} du = \frac{1}{360} x$$

Ad esempio se X ha densità $f(\cdot)$ abbiamo

$$P(X \leq 1) = \int_{-\infty}^1 f(u) du = \text{AREA 1}$$

Il tratteggio rappresenta l'AREA 1

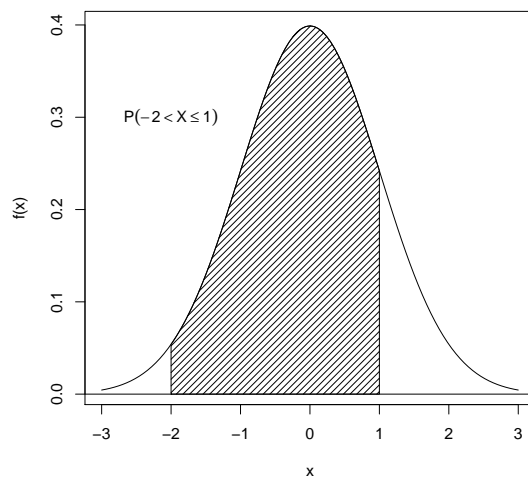


Esempio: continua esempio

$$P(X \leq 90) = \int_0^x \frac{1}{360} du = \frac{1}{360} 90 = \frac{1}{4}$$

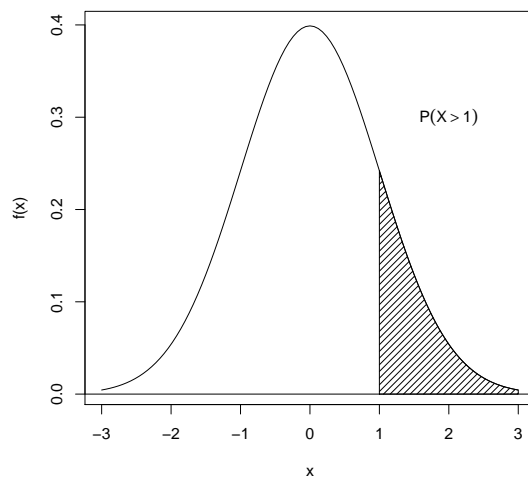
Possiamo anche calcolare:

$$P(-2 < X \leq 1) = \int_{-2}^1 f(u) du = \text{AREA 2}$$



Il tratteggio = AREA 2

$$P(X > 1) = \int_1^{+\infty} f(u) du = \text{AREA 3}$$



Il tratteggio = AREA 3

Riassumendo:

- Una variabile casuale continua è un modello matematico X

- È dotata di una densità $f(x)$ tale che

$$f(x) \geq 0 \quad \text{e} \quad \int_{-\infty}^{\infty} f(u) du = 1$$

- Per calcolare le probabilità con cui la v.c. continua assume valori in un intervallo si utilizza la **funzione di distribuzione**

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du$$

- Le probabilità sono calcolate come

$$P(a < X \leq b) = F(b) - F(a)$$

Attenzione: per le variabili casuali continue si ha sempre $P(X = x) = 0$

Valore atteso e varianza

La media di una variabile casuale viene anche chiamata *valore atteso*

Per le v.c. discrete è definito come

$$\mu = E(X) = \sum_{i=1}^k x_i \cdot P(X = x_i)$$

Si noti l'analogia con la media aritmetica

$$\bar{x} = \sum_{i=1}^k x_i \cdot f_i$$

Per le v.c. continue

$$\int_{-\infty}^{\infty} x f(x) dx$$

La varianza per le v.c. discrete è definita

$$\sigma^2 = \text{Var}(X) = \sum_{i=1}^k (x_i - \mu)^2 \cdot P(X = x_i)$$

analogamente al caso della statistica descrittiva

$$\sigma^2 = \sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i$$

Mentre per le v.c. continue

$$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Esempio: calcoliamo il valore atteso e la varianza della variabile casuale Mediana campionaria.

La distribuzione è

x	$P(X = x)$
3	3/10
5	4/10
7	3/10

$$\begin{aligned} \mu = E(X) &= \sum_{i=1}^3 x_i P(X = x_i) \\ &= 3 \cdot \frac{3}{10} + 5 \cdot \frac{4}{10} + 7 \cdot \frac{3}{10} = \frac{9 + 20 + 21}{10} = 5 \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^3 (x_i - \mu)^2 P(X = x_i) \\ &= (3 - 5)^2 \cdot \frac{3}{10} + (5 - 5)^2 \cdot \frac{4}{10} + (7 - 5)^2 \cdot \frac{3}{10} \\ &= \frac{12 + 12}{10} = 2.4 \end{aligned}$$

Esempio: calcoliamo il valore atteso e la varianza della variabile casuale somma dei punteggi realizzati con due dadi

La distribuzione è

$X = x$	2	3	4	5	6	7	8	9	10	11	12	
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

$$\mu = E(X) = \sum_{i=1}^{11} x_i P(X = x_i) = 7$$

La varianza possiamo riscriverla come

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^k (x_i - \mu)^2 P(X = x_i) \\ &= \sum_{i=1}^k x_i^2 P(X = x_i) - \mu^2 \end{aligned}$$

In analogia al caso della statistica descrittiva

$$\sigma^2 = \sum_{i=1}^k x_i^2 f_i - (\bar{x})^2$$

Abbiamo quindi per X

$$\text{Var}(X) = \sum_{i=1}^{11} x_i^2 P(X = x_i) - (7)^2 = 54.83 - 49 = 5.83$$

Esempio: calcoliamo il valore atteso e la varianza della variabile casuale continua X che indica i gradi di una freccia fatta girare su una ruota. Abbiamo visto che i valori possibili sono $\Omega = [0, 360)$, mentre la densità è $f(x) = \frac{1}{360}$. Il valore atteso è

$$\mu = E(X) = \int_0^{360} x \frac{1}{360} du = \frac{1}{2} x^2 \frac{1}{360} \Big|_0^{360} = 180$$

Anche per le v.c. continue possiamo calcolare la varianza in un modo più semplice

$$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

Quindi per il calcolo della varianza calcoliamo prima

$$\int_0^{360} x^2 \frac{1}{360} du = \frac{1}{3} x^3 \frac{1}{360} \Big|_0^{360} = \frac{360^2}{3}$$

La varianza è data da

$$\sigma_X^2 = \frac{360^2}{3} - (180)^2 = \frac{360^2}{12}$$

Proprietà di Media e Varianza

Sia X una v.c. tale che $\mu = E(X)$ e $\sigma_X^2 = \text{Var}(X)$

Sia Y una v.c. definita come $Y = a + bX$, allora

$$E(Y) = E(a + bX) = a + bE(X)$$

e

$$\sigma_Y^2 = \text{Var}(Y) = \text{Var}(a + bX) = b^2 \text{Var}(X) = b^2 \sigma_X^2$$

Se X ed Y sono due v.c. allora

$$E(X + Y) = E(X) + E(Y)$$

Variabili Casuali Indipendenti

Due variabili casuali discrete

$$X = \{x_i, P(X = x_i), i = 1, \dots, h\}$$

$$Y = \{y_j, P(Y = y_j), j = 1, \dots, h\}$$

si dicono indipendenti se e solo se $\forall i, j$

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

Se X ed Y sono anche *indipendenti* si ha che

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$