



Università di Pavia

Richiami di Probabilità e Statistica

Eduardo Rossi



Un *insieme* è una collezione di oggetti. $x \in A$ indica che l'oggetto x è un elemento dell'insieme A .

$A \subseteq B$ indica che l'insieme A è un *sottoinsieme* dell'insieme B , con tutti gli elementi di A che sono elementi anche di B . L'insieme vuoto, quello che non ha elementi, è indicato con \emptyset .

L'appartenenza ad un insieme può essere indicato con l'indicizzazione degli elementi, come in $A = \{x_i, i \in I\}$ dove I è l'insieme degli indici. I può essere finito o infinito.

Se $I = \mathbb{N}$ (i numeri naturali $1, 2, 3, \dots$) la numerosità degli elementi è detta *infinitamente numerabile*. Alcuni insiemi, come \mathbb{R} sono invece *infinitamente non numerabili*, cioè i loro elementi non possono esaustivamente numerati con gli elementi di \mathbb{N} .



UNIONE ED INTERSEZIONE

Unione: L'*unione* di A e B , $A \cup B$, è l'insieme degli elementi sia in A che in B .

Intersezione: $A \cap B$, è l'insieme degli elementi in entrambi gli insiemi, A e B .

Differenza: L'insieme degli elementi in A ma non in B , $A - B$ o $A \setminus B$. Indicando con X l'insieme universale, cioè l'insieme che contiene tutti gli elementi, allora *l'insieme complemento* di A , $A^c = X - A$.



Collezione $\mathcal{C} = \{A_i, i \in I\}$ è un insieme i cui elementi sono a loro volta insiemi.

L'unione e l'intersezione di questi insiemi, $\bigcup_{i \in I} A_i$ e $\bigcap_{i \in I} A_i$, godono della legge associativa:

$$\left(\bigcup_{i \in I} A_i \right) \cap B = \bigcup_{i \in I} (A_i \cap B)$$

$$\left(\bigcap_{i \in I} A_i \right) \cup B = \bigcap_{i \in I} (A_i \cup B)$$



Leggi di De Morgan per i complementi:

$$\left(\bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c$$

$$\left(\bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c$$



Algebra: Un'algebra di X è una collezione contenente X che ha le seguenti proprietà:

1. $X \in \mathcal{F}$
2. se $A \in \mathcal{F}$ allora $A^c \in \mathcal{F}$
3. se $A_1 \in \mathcal{F}$ e $A_2 \in \mathcal{F}$ allora $A_1 \cup A_2 \in \mathcal{F}$

σ – Algebra



σ -Algebra: Una σ -Algebra di X è un'algebra \mathcal{F} di sottoinsiemi di X che obbedisce alla regola: Se A_1, A_2, \dots è una collezione infinitamente numerabile di insiemi in \mathcal{F} , allora

$$\bigcup_i A_i \in \mathcal{F}$$



Spazio campionario: Lo spazio campionario Ω dell'esperimento è l'insieme di tutti i distinti, possibili esiti, detti eventi elementari (o stati di natura).

Esempio: Lancio della moneta. Lo spazio campionario in questo caso è $\{T, C\}$, dove $T = \text{Testa}$, $C = \text{Croce}$.

Evento: E' un sottoinsieme degli stati di natura.



Probabilità: Una misura di probabilità è una funzione reale di sottoinsiemi dello spazio campionario Ω che soddisfa certi assiomi: se A indica un certo sottoinsieme di Ω ($A \in \mathcal{F}$), allora

$$0 \leq Pr(A) \leq 1 \quad \forall A \in \mathcal{F}$$

$$Pr(\Omega) = 1$$

Se A_1 and A_2 sono insiemi disgiunti ($A_1 \cap A_2 = \emptyset$), allora

$$Pr(A_1 \cup A_2) = Pr(A_1) + Pr(A_2)$$

Se A_1, A_2, A_3, \dots è una collezione infinitamente numerabile di insiemi disgiunti di \mathcal{F} , tale che $A_i \cap A_j = \emptyset$ when $i \neq j$, then

$$Pr\left(\bigcup_i A_i\right) = \sum_i Pr(A_i)$$



Il modello probabilistico di un esperimento casuale ha tre elementi:

- Lo spazio campionario Ω .
- Una σ -algebra di eventi \mathcal{F} (sottoinsiemi di Ω).
- Una misura di probabilità, Pr , cioè una funzione che restituisce un numero reale, nell'intervallo $[0, 1]$, ad ogni elemento di \mathcal{F} .



La tripletta di oggetti $(\Omega, \mathcal{F}, Pr)$ definisce uno spazio di probabilità.
Per esempio, nel caso del lancio della moneta il modello probabilistico è descritto da:

$$\begin{aligned}\Omega &= \{T, C\} \\ \mathcal{F} &= \{\Omega, \{T\}, \{C\}, \emptyset\} \\ Pr(T) &= \frac{1}{2}\end{aligned}$$

Si può dedurre $Pr(T) = 1 - Pr(C) = \frac{1}{2}$.



Variabile casuale: E' una funzione reale $Y(s), s \in S$ ($Y : \Omega \rightarrow \mathbb{R}$) tale che $\{s \in S | Y(s) \leq y\}$ è un evento per ogni $y \in \mathbb{R}$.

Per il lancio della moneta, è comune assegnare valore 1 a {testa} e 0 a {croce}. Questa assegnazione è arbitraria. Le variabili casuali sono indicate con le lettere maiuscole Y , particolari numeri casuali (o realizzazioni) sono invece indicati con le lettere minuscole y .



FUNZIONE DI DISTRIBUZIONE CUMULATA

La funzione di distribuzione cumulata (f.d.c.) è la probabilità :

$$F_Y(y) = Pr(Y \leq y)$$

La f.d.c. è una funzione non decrescente del suo argomento: se $y_1 < y_2$, allora

$$\begin{aligned} Pr(Y \leq y_1) &\leq Pr(Y \leq y_1) + Pr(y_1 < Y \leq y_2) \\ &= Pr(Y \leq y_2) \end{aligned}$$

La f.d.c. è continua da destra:

$$\begin{aligned} \lim_{\epsilon \rightarrow 0^+} Pr(Y \leq y_1) &\equiv Pr(Y \leq y_1 + 0) \\ &= Pr(Y \leq y_1) + \lim_{\epsilon \rightarrow 0^+} Pr(y_1 < Y \leq y_1 + \epsilon) \\ &= Pr(Y \leq y_1) \end{aligned}$$



FUNZIONE DI DISTRIBUZIONE CUMULATA - VARIABILI DISCRETE

Variabili stocastiche discrete hanno f.d.c. a scalini. L'immagine del loro spazio campionario può essere ridotto ad un insieme numerabile

$$\mathbb{S}_Y = \{\dots, y_{-1}, y_0, y_1, \dots\}$$

$$Pr\{\mathbb{S}_Y\} = \sum_{t=-\infty}^{\infty} Pr\{Y = y_i\} = 1$$

Da cui

$$F_Y(y) = \sum_{i=-\infty}^{\infty} \mathbf{1}\{y_i \leq y\} Pr\{y = y_i\}$$

dove

$$\mathbf{1}\{y_i \leq y\} = \begin{cases} 1 & y_i \leq y \\ 0 & y_i > y \end{cases}$$



questo significa che vi è un salto in corrispondenza di ogni elemento $y_i \in \mathbb{S}_Y$ con ordinata

$$Pr\{Y = y_i\} = Pr\{Y \leq y_i\} - Pr\{Y \leq y_i - 0\}$$



Gli elementi di S_Y con probabilità strettamente positiva sono chiamati punti di massa o atomi della distribuzione. L'insieme dei punti di massa è detto *supporto*.

Un caso speciale di variabile discreta è una costante. Il supporto di una costante ha un solo elemento.

La variabile casuale Y ha una *distribuzione degenera* se $Pr\{Y = y_1\} = 1$, tale per cui Y è una costante uguale a y_1 con probabilità uguale a 1.



Variabili casuali continue possiedono f.d.c. continue.

$$Pr\{Y \in (y_1, y_2]\} = F_Y(y_2) - F_Y(y_1) > 0$$

dove $y_1 < y_2$. La probabilità di un singolo y_1 valore è zero. Per la continuità di $F_Y(\cdot)$:

$$\lim_{\epsilon \rightarrow 0} F_Y(y_1 + \epsilon) \equiv F_Y(y_1 \pm 0) = F_Y(y_1)$$

$$\begin{aligned} Pr\{Y = y_1\} &= \lim_{\epsilon \rightarrow 0^+} Pr\{|Y - y_1| \leq \epsilon\} \\ &= F_Y(y_1 + 0) - F_Y(y_1 - 0) \\ &= 0 \end{aligned}$$



Il supporto della distribuzione di una variabile continua è l'unione degli intervalli sui quali $F_Y(\cdot)$ è strettamente continua. E' possibile che sottoinsiemi del supporto di una v.c. continua abbiano probabilità zero. Il sottoinsieme $\{y_1\}$ ne è un esempio. Ogni sottoinsieme numerabile $\{y_1, y_2, \dots\}$ ha probabilità zero.



La funzione di probabilità della variabile discreta Y con supporto $S_Y = \{\dots, y_0, y_1, y_2, \dots\}$ è

$$f_Y(y) = \begin{cases} Pr\{Y = y\} & y \in S_Y \\ 0 & y \notin S_Y \end{cases}$$

$\forall y \in \mathbb{R}$.

La funzione di probabilità è non nulla nei punti in cui la f.d.c. cresce.



FUNZIONE DI DENSITÀ DI PROBABILITÀ

La funzione di densità di probabilità (f.d.p.) della variabile continua Y è la funzione $f_Y(\cdot)$ che soddisfa

$$F_Y(y) = \int_{-\infty}^y f_Y(x) dx$$

$\forall y \in \mathbb{R}$.

Assumiamo che $F_Y(y)$ sia differenziabile ovunque, così da avere

$$f_Y(y) = \frac{dF_Y(y)}{dy}$$

La f.d.p. è nulla in regioni nelle quali la f.d.c. è costante (non crescente), proprio come per la f.p. per le v.c. discrete.

$$\int_{-\infty}^{\infty} f_Y(x) dx = 1$$



Entrambe, la f.p. per v.c. discrete e la f.d.p. per v.c. continue, possono essere ricavate dalle f.d.c. e viceversa: possiamo ottenere le f.p. e le f.d.p. dalle f.d.c.

$$F_Y(y) = \begin{cases} \sum_{i:y_i \leq y} f_Y(y_i) & Y \text{ discreta} \\ \int_{-\infty}^y f_Y(x) dx & Y \text{ continua} \end{cases}$$

$$F_Y(y) = \begin{cases} \sum_{i=-\infty}^{\infty} \mathbf{1}\{y_i \leq y\} f_Y(y_i) & Y \text{ discreta} \\ \int_{-\infty}^{\infty} \mathbf{1}\{x \leq y\} f_Y(x) dx & Y \text{ continua} \end{cases}$$



$$E[g(Y)] = \begin{cases} \sum_{i=-\infty}^{\infty} g(y_i) f_Y(y_i) & Y \text{ discreta} \\ \int_{-\infty}^{\infty} g(x) f_Y(x) dx & Y \text{ continua} \end{cases}$$

Le probabilità possono essere sempre scritte come valori attesi:

$$Pr\{Y \in A\} = E[\mathbf{1}\{Y \in A\}]$$



Linearità del valore atteso

$$E[g(Y_1) + g(Y_2)] = E[g(Y_1)] + E[g(Y_2)]$$

Momenti: Sono i valori attesi

$$m_r \equiv E[Y^r] \quad r = 0, 1, 2, 3, \dots$$

Media

$$\mu \equiv m_1 = E[Y]$$

Momenti centrali:

$$\mu_r \equiv E[(Y - \mu)^r]$$

Varianza

$$Var[Y] \equiv \sigma^2 \equiv E[(Y - \mu)^2]$$



PROPRIETÀ DELLA VARIANZA

$$E[Y^2] = \text{Var}[Y] + (E[Y])^2$$

Sia $\alpha \in \mathbb{R}$

$$\text{Var}[\alpha Y] = \alpha^2 \text{Var}[Y]$$



VARIABILI CASUALI STANDARDIZZATE

La varianza misura la dispersione della funzione di probabilità, o della f.d.p., intorno al momento primo (la media).

Variabili casuali standardizzate:

$$W = \frac{Y - \mu}{\sigma}$$

$$E[W] = 0$$

$$Var[W] = 1$$

Il momento terzo standardizzato è il momento terzo di W ed è chiamato *skewness*

$$\gamma_1 \equiv E \left[\left(\frac{Y - \mu}{\sigma} \right)^3 \right]$$

La skewness è sensibile all'asimmetria di una f.d.p.: se la f.d.p. di Y è simmetrica intorno a μ allora $\gamma_1 = 0$.



Il momento quarto standardizzato è

$$\gamma_2 \equiv E \left[\left(\frac{Y - \mu}{\sigma} \right)^4 \right] - 3$$

detto anche coefficiente di curtosi.



FUNZIONE GENERATRICE DEI MOMENTI

La funzione generatrice dei momenti della variabile casuale Y , indicata con $M_Y(\cdot)$, è

$$M_Y(t) \equiv E[e^{tY}]$$

Se la r -esima derivata di $M_Y(t)$ esiste a $t = 0$ allora

$$\left. \frac{d^r M_Y(t)}{dt^r} \right|_{t=0} = E \left[\left. \frac{d^r e^{tY}}{dt^r} \right|_{t=0} \right] = E[Y^r] \equiv \mu_r$$

Non è detto che la funzione esista sempre per $t \neq 0$. Questo significa che né la sequenza dei momenti μ_r né la funzione generatrice dei momenti caratterizzano le distribuzioni.



TRASFORMAZIONE DI VARIABILE

Sia Y una v.c. continua e $g(\cdot)$ sia una funzione reale biiettiva differenziabile su \mathbb{S}_Y . Allora la f.d.p. della v.c.

$$Z = g(Y)$$

è

$$f_Z(z) = \begin{cases} \left| \frac{dh(z)}{dz} \right| f_Y[h(z)] & \text{se } z \in \mathbb{S}_Y \\ 0 & \text{se } z \notin \mathbb{S}_Y \end{cases}$$

dove $h(z) \equiv g^{-1}(\cdot)$ è la funzione inversa di $g(\cdot)$ e \mathbb{S}_Y è l'immagine di \mathbb{S}_Y sotto $g(\cdot)$.



Il termine di derivata può essere trovato attraverso la differenziazione implicita:

$$\begin{aligned}z &= g[h(z)] \\1 &= g'[h(z)] \frac{dh(z)}{dz} \\ \frac{dh(z)}{dz} &= \frac{1}{g'[h(z)]}\end{aligned}$$

La moltiplicazione dell'originale f.d.p. f_Y per il valore assoluto di $\frac{dh(z)}{dz}$ riflette il cambiamento di unità di misura da Y a quelle di Z .



Esempio: $Y \sim N(\mu, \sigma^2)$. Qual'è la distribuzione di $Z = \frac{Y-\mu}{\sigma}$?

$$h(z) = \sigma z + \mu$$

$$\frac{dh(z)}{dz} = \frac{dy}{dz} = \sigma$$

$$\begin{aligned} f_Z(z) &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{[(\sigma z + \mu) - \mu]^2}{2\sigma^2} \right\} |\sigma| \\ &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{z^2}{2} \right\} \end{aligned}$$

quindi $Z \sim N(0, 1)$.



Il q -esimo quantile ($0 \leq q \leq 1$) della f.d.c. $F_Y(\cdot)$ è l'insieme

$$\{y | q = F_Y(y)\}$$

Se F_Y è una funzione 1-a-1, allora $F_Y^{-1}(q)$ è l'unico q -esimo quantile.

Quando si prende una trasformazione monotona di Y , $h(Y)$,
l'immagine del q -esimo quantile di Y è il q -esimo quantile di $h(Y)$.

Questa proprietà non è rispettata dai momenti:

$$E[h(Y)] \neq h[E(Y)]$$

con la sola eccezione di $h(\cdot)$ trasformazione lineare.



Se $h(\cdot)$ è una funzione convessa e $E[Y]$ esiste, allora

$$h(E[Y]) \leq E[h(y)]$$

Se $h(\cdot)$ è strettamente convessa in ogni punto di $(S)_Y$, allora la disuguaglianza è stretta a meno che Y sia uguale ad una costante con probabilità 1.

Esempio:

$$Var[Y] = E[Y^2] - (E[Y])^2 \geq 0$$

perchè la funzione quadratica è convessa.



DISEGUAGLIANZA DI CHEBYCHEV

Per ogni variabile casuale Y con momenti secondi finiti,

$$Pr\{|Y - b| > a\} \leq \frac{E[(Y - b)^2]}{a^2}$$

per ogni b e per ogni $a > 0$.

Se $b = E(Y)$ e $a = k\sigma$ allora:

$$Pr\{|Y - E(Y)| > k\sigma\} \leq \frac{E[(Y - E(Y))^2]}{a^2} = \frac{\sigma^2}{k^2\sigma^2}$$

questo significa, che indipendentemente dalla distribuzione della v.c. Y , la probabilità che Y assuma valori distanti dalla media più di k deviazioni standard è al più $\frac{1}{k^2}$.



Eventi come risultati di un esperimento comune. A_1 e A_2 si verificano quando elementi comuni dello spazio campionario sono realizzati:

$$Pr\{A_1 \text{ e } A_2\} = Pr\{A_1 \cap A_2\}$$

La funzione di distribuzione cumulata di un insieme finito di variabili casuali $\{Y_1, \dots, Y_N\}$

$$F_Y(y_1, \dots, y_N) = Pr\{Y_1 \leq y_1, \dots, Y_N \leq y_N\}$$



- Y continua, funzione di densità congiunta

$$f_Y(y) = \frac{\partial^N F_Y(y)}{\partial y_1 \cdots \partial y_N}$$

- Y discreta, funzione di massa di probabilità

$$f_Y(y) = Pr\{Y_1 = y_1, \dots, Y_N = y_n\}$$



FUNZIONE DI DISTRIBUZIONE CUMULATA MARGINALE

Per un sottoinsieme di v.c. $Z = \{Y_1, \dots, Y_k\}$, $k < N$

$$F_Z(z) = Pr\{Y_1 \leq y_1, \dots, Y_K \leq y_k\} = F_Y(y_1, \dots, y_K, \infty, \dots, \infty)$$

Funzione di densità marginale (v.c. continue)

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y_1, \dots, y_N) dy_{k+1} \dots dy_N$$

$$f_Z(z) = \frac{\partial^k F_Z(z)}{\partial y_1 \dots \partial y_k}$$



Funzione di distribuzione cumulata

$$F_{X,Y}(x, y) = Pr(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(t, s) ds dt$$

Funzione di densità marginale

$$f_X(x) = \int_y f(x, t) dt$$



La probabilità condizionale di A_1 dato A_2 è pari a:

$$Pr\{A_1|A_2\} = \frac{Pr\{A_1 \cap A_2\}}{Pr\{A_2\}}$$

$$\begin{aligned} Pr\{A_1|A_2\} + Pr\{A_1^c|A_2\} &= \frac{Pr\{A_1 \cap A_2\} + Pr\{A_1^c \cap A_2\}}{Pr\{A_2\}} \\ &= \frac{Pr\{(A_1 \cap A_2) \cup (A_1^c \cap A_2)\}}{Pr\{A_2\}} \\ &= 1 \end{aligned}$$

$Pr\{A_2\}$ probabilità marginale.



TEOREMA DI BAYES

Sia $\{A_i : i = 1, 2, \dots\}$ una collezione numerabile di eventi disgiunti tali che $\bigcup_i A_i = S$. Allora

$$Pr\{A_i|B\} = \frac{Pr\{B|A_i\}Pr\{A_i\}}{\sum_j Pr\{B|A_j\}Pr\{A_j\}} = \frac{Pr\{B|A_i\}Pr\{A_i\}}{Pr\{B\}}$$

$\forall B \subseteq S$, tale che $Pr\{B\} > 0$.



Due eventi A_1 e A_2 sono indipendenti se

$$Pr\{A_1 \cap A_2\} = Pr\{A_1\}Pr\{A_2\}$$

Una condizione equivalente per l'indipendenza:

$$Pr\{A_2\} > 0 \quad : \quad Pr\{A_1|A_2\} = Pr\{A_1\}$$



$$F_{ZY}(z, y) = Pr\{Z \leq z, Y \leq y\}$$

Per tutte le variabile casuali, discrete o continue,

$$\begin{aligned} Pr\{Z \leq z | y_a < Y \leq y_b\} &= \frac{Pr\{Z \leq z, y_a < Y \leq y_b\}}{Pr\{y_a < Y \leq y_b\}} \\ &= \frac{F_{ZY}(z, y_b) - F_{ZY}(z, y_a)}{F_Y(y_b) - F_Y(y_a)} \end{aligned}$$

dato che $F_Y(y_b) - F_Y(y_a) > 0$



Due variabili casuali, Z, Y discrete, dove $y_a = y_b = y$ e $Pr\{Y = y\} > 0$, con supporto $S_{ZY}\{(z_i, y_i); i = 1, 2, \dots\}$, la funzione di probabilità condizionale di Z quando $Y = y$ è

$$Pr\{Z = z|Y = y\} = \frac{Pr\{Z = z, Y = y\}}{Pr\{Y = y\}}$$

Due variabili casuali continue, Z, Y , la f.d.c. condizionale

$$\begin{aligned} F_{Z|Y}(z|y) &= Pr\{Z \leq z|Y = y\} \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{F_{ZY}(z, y) - F_{ZY}(z, y - \varepsilon)}{F_Y(y) - F_Y(y - \varepsilon)} \\ &= \frac{\partial F(z, y)/\partial y}{f_Y(y)} \end{aligned}$$

dato $f_Y(y) > 0$.



$$\begin{aligned}f_{Z|Y}(z|y) &= \frac{\partial F_{Z|Y}(z|y)}{\partial z} \\ &= \frac{f_{ZY}(z, y)}{f_Y(y)}\end{aligned}$$

$$f_{ZY}(z, y) = f_{Z|Y}(z|y)f_Y(y)$$

Integrando entrambi i membri

$$\int f_{ZY}(z, y)dz = \int f_{Z|Y}(z|y)f_Y(y)dz$$

$$f_Y(y) = f_{Z|Y}(z|y)f_Y(y)$$



Se Z e Y sono indipendenti, allora

$$f_{Y|Z}(y|z) = f_Y(y)$$

$$f_{Y|Z}(y|z) = f_Z(z)$$



DISTRIBUZIONE DI BERNOULLI

La variabile casuale discreta Y

$$f_Y(y; \theta) = \begin{cases} \theta & \text{se } y = 1 \\ 1 - \theta & \text{se } y = 0 \\ 0 & \text{altrove} \end{cases}$$

$0 \leq \theta \leq 1$, dove il parametro θ

$$\theta = Pr\{Y = 1\}$$



La variabile casuale discreta Y ha funzione di probabilità

$$f_Y(y; \theta) = \begin{cases} \binom{N}{y} \theta^y (1 - \theta)^{N-y} & \text{se } y \in \{0, 1, \dots, N\} \\ 0 & \text{se } y \notin \{0, 1, \dots, N\} \end{cases}$$

dove $0 \leq \theta \leq 1$ e $N \in \mathbb{N}$. Viene spesso caratterizzata come il numero dei successi su N v.c. Bernoulliane indipendenti.



La variabile casuale discreta Y ha funzione di probabilità

$$f_Y(y; \theta) = \begin{cases} \theta(1 - \theta)^{N-y} & \text{se } y \in \mathbb{N} \\ 0 & \text{se } y \notin \mathbb{N} \end{cases}$$

dove $0 \leq \theta \leq 1$. La distribuzione geometrica è la distribuzione del numero di esperimenti bernoulliani che si verificano prima che si realizzi un successo.



DISTRIBUZIONE BINOMIALE NEGATIVA

La variabile casuale discreta Y ha funzione di probabilità

$$f_Y(y; \theta) = \begin{cases} \binom{y-1}{M-1} \theta^M (1-\theta)^{y-M} & \text{se } y \geq M, y \in \mathbb{N} \\ 0 & \text{se } y \notin \{0, 1, \dots, N\} \end{cases}$$

dove $0 \leq \theta \leq 1$. E' una generalizzazione della distribuzione geometrica, sono infatti il numero di esperimenti prima che si verifichi l' M -esimo successo.



DISTRIBUZIONE POISSON

La variabile casuale discreta Y ha funzione di probabilità

$$f_Y(y; \theta) = \begin{cases} \frac{\theta^y e^{-\theta}}{y!} & \text{se } y \in \mathbb{N} \\ 0 & \text{se } y \notin \mathbb{N} \end{cases}$$

dove $0 < \theta$. La distribuzione di Poisson è il limite della distribuzione binomiale quando $N \rightarrow \infty$.



DISTRIBUZIONE UNIFORME

La variabile casuale continua Y ha funzione di densità di probabilità

$$f_Y(y) = \begin{cases} 1 & \text{se } y \in [0, 1] \\ 0 & \text{se } y \notin [0, 1] \end{cases}$$



La variabile casuale continua Y ha funzione di densità di probabilità

$$f_Y(y) = \begin{cases} e^{-y} & \text{se } y \geq 0 \\ 0 & \text{se } y < 0 \end{cases}$$

Si può generare una v.c. esponenziale prendendo il logaritmo di una v.c. uniforme.



DISTRIBUZIONE NORMALE

$$Y \sim N(\mu, \sigma^2)$$

$$f_Y(y; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(y - \mu)^2}{2\sigma^2} \right\}$$

$$X = a + bY$$

$$X \sim N(a + b\mu, b^2\sigma^2)$$



DISTRIBUZIONE NORMALE STANDARD

$$Z \sim N(0, 1)$$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{z^2}{2} \right\}$$

Funzione di distribuzione cumulata per $\phi(z)$

$$\Phi(z) = \int_{-\infty}^z \phi(s) ds$$

$$\Phi(-z) = 1 - \Phi(z)$$

$$\mu_{2r} = \prod_{n=1}^r (2n - 1)$$

$$\mu_{2r-1} = 0 \quad r = 1, 2, \dots$$



DISTRIBUZIONE NORMALE STANDARD

$$X = \mu + \sigma Z$$

Possiamo esprimere la densità di X come:

$$X \sim N(\mu, \sigma^2)$$

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right)$$

$$Pr(a < x < b) = Pr\left(\frac{a - \mu}{\sigma} < \frac{x - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right) = \Phi(b) - \Phi(a)$$

Siano $Y_1 \sim N(\mu_1, \sigma_1^2)$ e $Y_2 \sim N(\mu_2, \sigma_2^2)$ distribuite indipendentemente

$$Y_1 + Y_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$



FUNZIONE GAMMA

La funzione gamma $\Gamma(\cdot)$

$$\Gamma(y) \equiv \int_0^{\infty} x^{y-1} e^{-x} dx$$

La funzione gamma non ha un'espressione algebrica semplice eccetto nei seguenti casi:

$$\Gamma(1/2) = \sqrt{(\pi)}$$

$$\Gamma(1) = 1$$

$$\Gamma(n+1) = n\Gamma(n)$$

$$= n! \quad \text{se } n \in \mathbb{N}$$

$$\frac{\Gamma(z)\Gamma(y)}{\Gamma(z+y)} = \int_0^{\infty} x^{z-1}(1-x)^{y-1} dx$$



DISTRIBUZIONE CHI-QUADRATO

La variabile casuale continua Y ha funzione di densità di probabilità

$$f_Y(y; \nu) = \begin{cases} \frac{y^{(\nu/2)-1} e^{-\frac{y}{2}}}{2^{\nu/2} \Gamma(\frac{\nu}{2})} & \text{se } y > 0 \\ 0 & \text{se } y \leq 0 \end{cases}$$



DISTRIBUZIONE CHI-QUADRATO

Possiamo generare una v.c. X con distribuzione chi-quadrato:

$$Z \sim N(0, 1)$$

$$X = Z^2$$

$$X \sim \chi_1^2$$

$$E[X] = 1$$

$$Var[X] = 2$$



DISTRIBUZIONE CHI-QUADRATO

Siano Z_1, Z_2, \dots, Z_m v.c. i.i.d. $N(0, 1)$, segue che

$$\sum_{i=1}^m Z_i^2 \sim \chi_m^2$$

$$E \left[\sum_i^m Z_i^2 \right] = m$$

$$\text{Var} \left[\sum_i^m Z_i^2 \right] = 2m$$



DISTRIBUZIONE CHI-QUADRATO

$$X_1^2 \sim \chi_{m_1}^2$$

$$X_2^2 \sim \chi_{m_2}^2$$

$$X_1^2 + X_2^2 \sim \chi_{m_1+m_2}^2$$



DISTRIBUZIONE T-STUDENT

La v.c. continua Y ha una distribuzione t-Student con ν gradi di libertà, $Y \sim t_\nu$, con densità:

$$f_Y(y) = \frac{[\Gamma(\nu + 1)/2]}{\nu^{1/2}\Gamma(1/2)\Gamma(\nu/2)} \left(1 + \frac{y^2}{\nu}\right)^{-(\nu+1)/2}$$

Per $\nu > 1$, μ_r è finito solo se $r < \nu$. I momenti di ordine dispari che esistono sono zero.



I momenti finiti di ordine pari sono:

$$\begin{aligned}\mu_r &= \nu^{r/2} \frac{\Gamma[(\nu + 1)/2] \Gamma[(\nu - r)/2]}{\Gamma(1/2) \Gamma(\nu/2)} \\ &= \nu^{r/2} \frac{1 \cdot 2 \cdots (r - 1)}{(\nu - r)(\nu - r + 2) \cdots (\nu - 2)}\end{aligned}$$

con $r = 2, 4, 6, \dots$

$$\begin{aligned}\mu_2 &= \frac{\nu}{\nu - 2} \\ \mu_4 &= \frac{3\nu^2}{(\nu - 4)(\nu - 2)}\end{aligned}$$



DISTRIBUZIONE T-STUDENT

$$Z \sim N(0, 1)$$

$$X = Z^2 \sim \chi_n^2$$

$$\frac{Z}{\sqrt{X/n}} \sim t_n$$

Con l'aumentare dei gradi di libertà la t-Student non è distinguibile da una distribuzione normale.



DISTRIBUZIONE F

La v.c. Y è distribuita come una F con gradi di libertà ν_1 e ν_2 ,
 $Y \sim F_{\nu_1, \nu_2}$ con densità

$$f_Y(y) = \begin{cases} \frac{\nu_1}{\nu_2} \frac{\Gamma[(\nu_1 + \nu_2)/2]}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \left(\frac{\nu_1}{\nu_2} y\right)^{(\nu_1/2)-1} \left(1 + \frac{\nu_1}{\nu_2} y\right)^{-(\nu_1 + \nu_2)/2} & \text{se } y \geq 0 \\ 0 & \text{se } y < 0 \end{cases}$$

$$E[Y] = \frac{\nu_2}{\nu_2 - 2}$$

se $\nu_2 > 2$, altrimenti Y non ha momenti finiti.



DISTRIBUZIONE F

Se X_1 e X_2 sono due variabili χ^2 indipendenti con ν_1 e ν_2 gradi di libertà

$$\frac{X_1/\nu_1}{X_2/\nu_2} \sim F_{(\nu_1, \nu_2)}$$

F_{ν_1, ν_2} converge a $\chi_{\nu_1}^2/\nu_1$ quando $\nu_2 \rightarrow \infty$.



DISTRIBUZIONE CAUCHY

$$Z_1 \sim N(0, 1) \quad Z_2 \sim N(0, 1)$$

$$\frac{Z_1}{Z_2} \sim t_1$$

I momenti di questa distribuzione non esistono. E'importante notare come

$$t_n^2 \sim F_{(1,n)}$$



DISTRIBUZIONE LOG-NORMALE

$$X \sim LN(E(X), Var(X))$$

$$\log(X) \sim N(\mu, \sigma^2)$$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log(x) - \mu)^2}{2}\right\} \quad x > 0$$

$$E[X] = e^{\mu + \sigma^2/2}$$

$$Var[X] = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$



DISTRIBUZIONE LOG-NORMALE

Se $X \sim LN(\theta, \lambda^2)$ allora

$$\ln(X) \sim N(\mu, \sigma^2)$$

dove

$$\mu = \ln \theta^2 - \frac{1}{2} \ln (\theta^2 + \lambda^2)$$

$$\sigma^2 = \ln \left(1 + \frac{\lambda^2}{\theta^2} \right)$$



Sia \mathbf{x} un vettore di n componenti con funzione di distribuzione cumulata multivariata:

$$F(\mathbf{c}) = Pr(\mathbf{X} \leq \mathbf{c}) \quad \mathbf{c} \in \mathbb{R}^n$$

Indipendenza totale:

$$F(c_1, \dots, c_n) = F_{X_1}(c_1) \cdots F_{X_n}(c_n)$$

Indipendenza a coppie:

$$F_{X_i, X_j}(c_i, c_j) = F_{X_i}(c_i)F_{X_j}(c_j) \quad \forall i \neq j$$

non implica indipendenza totale.



$$\mathbf{x} = (x_1, \dots, x_n)$$

$$E[\mathbf{x}] = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_n) \end{bmatrix}$$

$$Var(\mathbf{x}) \equiv E[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))']$$

$$Var(\mathbf{x}) = \begin{bmatrix} E(x_1 - E(x_1))^2 & \dots & E[(x_1 - E(x_1))(x_n - E(x_n))] \\ \vdots & \ddots & \vdots \\ E[(x_n - E(x_n))(x_1 - E(x_1))] & \dots & E(x_n - E(x_n))^2 \end{bmatrix}$$

La matrice di covarianza di \mathbf{X} è simmetrica e definita positiva.



$$Var(\mathbf{X}) = E(\mathbf{X}\mathbf{X}') - E(\mathbf{X})E(\mathbf{X}')$$

Per ogni vettore \mathbf{a} fissato

$$\begin{aligned} Var(\mathbf{a}'\mathbf{x}) &= E[\mathbf{a}'\mathbf{x} - \mathbf{a}'E(\mathbf{x})]^2 \\ &= E[\mathbf{a}'(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))'\mathbf{a}] \\ &= \mathbf{a}'E[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))']\mathbf{a} \\ &= \mathbf{a}'Var(\mathbf{X})\mathbf{a} \end{aligned}$$



DISTRIBUZIONE NORMALE BIVARIATA

Le v.c. X e Y sono distribuite congiuntamente in modo normale:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$$

con densità:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{(\epsilon_X^2 + \epsilon_Y^2 - 2\rho\epsilon_X\epsilon_Y)}{2(1-\rho^2)} \right\}$$

dove

$$\begin{aligned} \epsilon_X &= \frac{X - \mu_X}{\sigma_X} \\ \epsilon_Y &= \frac{Y - \mu_Y}{\sigma_Y} \end{aligned}$$



DISTRIBUZIONE NORMALE BIVARIATA

La densità è definita solo se $\rho \neq 1$ cioè le due variabili non devono essere linearmente correlate.

Le densità marginali sono:

$$f_X(x) = N(\mu_X, \sigma_X^2)$$

$$f_Y(y) = N(\mu_Y, \sigma_Y^2)$$

Le distribuzioni condizionali sono normali:

$$f_{Y|X}(y|x) = N(\alpha + \beta x, \sigma_Y^2(1 - \rho^2))$$

$$\alpha = \mu_Y - \beta\mu_X$$

$$\beta = \frac{\sigma_{XY}}{\sigma_X^2}$$



DISTRIBUZIONE NORMALE BIVARIATA

X e Y sono indipendenti se e solo se $\rho = 0$.



DISTRIBUZIONE NORMALE MULTIVARIATA

Il vettore di v.c. $\mathbf{X} : (n \times 1)$ ha distribuzione congiunta normale:

$$\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

con densità:

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right\}$$

$$\boldsymbol{\Delta} \equiv \text{diag}(\sigma_1, \dots, \sigma_n)$$

La matrice di correlazione:

$$\mathbf{R} = \boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Delta}^{-1}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Delta} \mathbf{R} \boldsymbol{\Delta}$$

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Delta}^{-1} \mathbf{R}^{-1} \boldsymbol{\Delta}^{-1}$$



DISTRIBUZIONE NORMALE MULTIVARIATA

$$|\mathbf{\Sigma}| = |\mathbf{\Delta}|^2 |\mathbf{R}| = (\sigma_1 \cdot \sigma_2 \cdots \sigma_n)^2 |\mathbf{R}|$$

$$(\mathbf{X} - \boldsymbol{\mu})' \mathbf{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = (\mathbf{X} - \boldsymbol{\mu})' \mathbf{\Delta}^{-1} \mathbf{R}^{-1} \mathbf{\Delta}^{-1} (\mathbf{X} - \boldsymbol{\mu})$$

$$\boldsymbol{\epsilon} \equiv \mathbf{\Delta}^{-1} (\mathbf{X} - \boldsymbol{\mu})$$

$$(\mathbf{X} - \boldsymbol{\mu})' \mathbf{\Delta}^{-1} \mathbf{R}^{-1} \mathbf{\Delta}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = \boldsymbol{\epsilon}' \mathbf{R}^{-1} \boldsymbol{\epsilon}$$

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-n/2} (\sigma_1 \cdot \sigma_2 \cdots \sigma_n)^{-1} |\mathbf{R}|^{-1/2} \exp \left\{ -\frac{1}{2} \boldsymbol{\epsilon}' \mathbf{R}^{-1} \boldsymbol{\epsilon} \right\}$$



DISTRIBUZIONE NORMALE MULTIVARIATA

Tre casi:

- $\rho_{ij} = 0$ per $i \neq j$ tutte le variabili non sono correlate, cioè indipendenti ($\mathbf{R} = \mathbf{I}$)

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-n/2} (\sigma_1 \cdot \sigma_2 \cdots \sigma_n)^{-1} \exp \left\{ -\frac{1}{2} \boldsymbol{\epsilon} \boldsymbol{\epsilon}' \right\}$$

- Se $\sigma_i = \sigma$ e $\mu = 0$

$$x_i \sim N(0, \sigma^2)$$

- Se $\sigma_i = 1$, normale multivariata standard

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} \mathbf{x} \mathbf{x}' \right\}$$



DISTRIBUZIONI MARGINALI E CONDIZIONALI

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Se $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$ ha una distribuzione multivariata congiunta le distribuzioni marginali sono

$$\mathbf{X}_1 \sim N(\mu_1, \Sigma_{11})$$

$$\mathbf{X}_2 \sim N(\mu_2, \Sigma_{22})$$

Le distribuzioni condizionali sono:

$$\mathbf{X}_1 | \mathbf{X}_2 \sim N(\mu_{1.2}, \Sigma_{11.2})$$

$$\mathbf{X}_2 | \mathbf{X}_1 \sim N(\mu_{2.1}, \Sigma_{22.1})$$



DISTRIBUZIONI MARGINALI E CONDIZIONALI

$$\mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2)$$

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$



FUNZIONI LINEARI DI UN VETTORE NORMALE

Se $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\mathbf{Ax} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$

Se \mathbf{A} non è di rango pieno allora $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$ è singolare e la densità non esiste.



Consideriamo

$$\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}) \quad (n \times 1)$$

La distribuzione di $\omega = \mathbf{z}'\mathbf{z}$ è la chi-quadrato centrale con n gradi di libertà, indicata con $\chi^2(n)$.

Sia

$$\begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \sim N(\mathbf{0}, \mathbf{I}_{n_1+n_2})$$

$$\mathbf{z}'_1 \mathbf{z}_1 = \omega_1 \sim \chi^2(n_1)$$

$$\mathbf{z}'_2 \mathbf{z}_2 = \omega_2 \sim \chi^2(n_2)$$

$$\omega_1 + \omega_2 = \mathbf{z}'_1 \mathbf{z}_1 + \mathbf{z}'_2 \mathbf{z}_2 \sim \chi^2(n_1 + n_2)$$



Le v.c. ω_1 e ω_2 sono indipendenti se e solo se gli elementi di \mathbf{z}_1 e \mathbf{z}_2 sono indipendenti. Essendo gaussiani la non correlazione è sufficiente per l'indipendenza. La somma di v.c. chi quadrato è una chi quadrato con gradi di libertà pari alla somma dei gradi di libertà di ogni addendo.