Legge dei grandi numeri e Teorema Centrale Limite

25 maggio 2017

Discutiamo due teoremi fondamentali per raccordare probabilitá e statistica: la legge dei grandi numeri e il Teorema Centrale Limite (TCL, anche detto Teorema del limite centrale).

La legge dei grandi numeri ci dice che la media aritmetica di un campione di n variabili aleatorie (VA) X_i, indipendenti e identicamente distribuite, ovvero $\frac{1}{n}\sum_{i=1}^{n} X_i$ per n crescente tende o converge al valore atteso teorico μ (cfr. Figura 1).

Ma come si distribuiscono per n crescente le deviazioni / fluttuazioni di $\frac{1}{n}\sum_{i=1}^{n} X_i$ rispetto a μ?

Il TCL asserisce che tali fluttuazioni obbediscono ad una legge Gaussiana , nonostante le distribuzioni delle singole variabili possano essere del tutto generiche (in forma semplice: la somma di un numero sufficientemente grande di variabili é gaussiana). George Polya introdusse per primo l' espressione "Teorema Centrale Limite" ritenendo il TCL un teorema centrale per il calcolo delle probabilitá.

Per ottenere questi due risultati fondamentali, discutiamo prima alcuni aspetti (valore atteso, varianza e funzioni generatrici dei momenti) relativi alla somma di VA indipendenti.

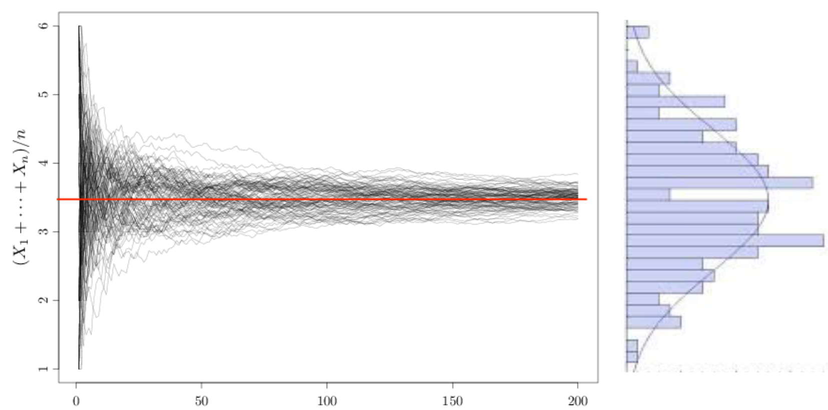

Figura 1: La figura mostra le traiettorie ottenute con diverse realizzazioni della VA $\frac{1}{n}\sum_{i=1}^{n} X_i$ per n crescente: tutte convergono al valore atteso teorico μ (linea rossa). La distribuzione delle deviazioni $|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu|$ segue una legge Gaussiana

1 *Distribuzioni congiunte e marginali*

Consideriamo due variabili aleatorie discrete X e Y. Siamo interessati all'evento congiunto

$$\{X = x\} \cap \{Y = y\}$$

e alla probabilitá di tale evento o *probabilitá congiunta*

$$P_{XY}(\{X = x\} \cap \{Y = y\}) = P_{XY}(X = x, Y = y) = p_{XY}(x, y)$$

Nel caso discreto la congiunta si puo' scrivere in forma tabellare. Un esempio di p. congiunta di due VA che possono assumere valori

$$X = \{0,1,2\}, Y = \{0,1\}$$

é riportato in Figura 2.

y \ x	0	1	2	Py(y)	
0	0.1	0.4	0.1	0.6	$\sum_{x=\{0,1,2\}} p(x, y=0)$
1	0.2	0.2	0	1.4	$\sum_{x=\{0,1,2\}} p(x, y=1)$
Px(x)	0.3	0.6	0.1	1	
	$\sum_{y=\{0,1\}} p(x=0, y)$	$\sum_{y=\{0,1\}} p(x=1, y)$	$\sum_{y=\{0,1\}} p(x=2, y)$		

Figura 2: La distribuzione congiunta $P_{XY}(\{X = x\} \cap \{Y = y\})$ scritta esplicitamente in forma tabellare. Nell'ultima colonna a destra e nell'ultima riga in basso (i margini) sono riportati i valori ottenuti sommando su ciascuna colonna fissata la riga i, $P_Y(Y = i) = \sum_x p_{xy}(x, y = i)$, e su ciascuna riga fissata la colonna j, $P_X(X = j) = \sum_y p_{xy}(x = j, y)$. P_X e P_Y rappresentano le distribuzioni di probabilitá marginali.

Dalla probabilitá congiunta possiamo inferire la probabilitá di qualsiasi evento di interesse che riguarda le VA X, Y. Ad esempio,

$P_{XY}(X + Y > 1) = P_{XY}(1,1) + P_{XY}(2,1) + P_{XY}(2,0) = 0.2 + 0 + 0.1 = 0.3$

Per distribuzioni discrete é utile talvolta semplificare la notazione:

$$P_{XY}(\{X = i\} \cap \{Y = j\}) = p_{XY}(i,j) = p_{i,j}$$

Notiamo che la congiunta é normalizzata, $\sum_{i,j} p_{i,j} = \sum_i \sum_j p_{i,j} = 1$.

Nella tabella riportata in Figura 2 abbiamo anche riportato, a margine, i valori ottenuti sommando su ciascuna colonna fissata la riga i, $P_Y(Y = i) = \sum_x p_{xy}(x, y = i)$, e su ciascuna riga fissata la colonna j, $P_X(X = j) = \sum_y p_{xy}(x = j, y)$.

L'insieme dei valori $P_X(X = j)$ e $P_Y(Y = i)$ definiscono rispettivamente le *probabilitá marginali* di X e Y:

$$P_X(x) = \sum_y p_{XY}(x, y)$$

$$P_Y(y) = \sum_x p_{XY}(x, y)$$

Si noti che P_X e P_Y sono normalizzate, $\sum_x P_X(x) = 1, \sum_y P_Y(y) = 1$.

I concetti introdotti sono ovviamente generalizzabili a VA continue. Se X e Y sono VA (continue o discrete) la finzione di ripartizione congiunta é definita come

$$F_{XY}(x,y) = P_{XY}(X \leq x, Y \leq y)$$

e per essa valgono tutte le proprietá definite per le funzioni di ripartizione di una singola variabile aleatoria.

Nel caso continuo, $F_{XY}(x,y)$ é (assolutamente) continua se esiste la funzione densitá di probabilitá congiunta $f_{XY}(x,y)$ (non negativa e integrabile) tale che

$$F_{XY}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(x,y) dxdy$$

dove $\int_x \int_y f_{XY}(x,y)dxdy = 1$ e

$$f_{XY}(x,y) = \frac{\partial}{\partial y}(\frac{\partial}{\partial x}(F_{XY}(x,y)))\frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

In altri termini la probabilitá congiunta nel caso continuo é definita come

$$prob(x,y \in A) = P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x,y)dxdy$$

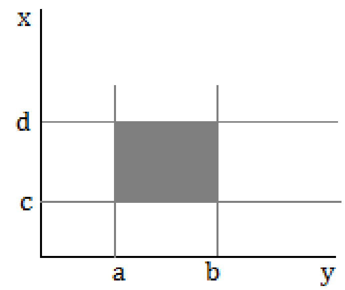

Figura 3: Un evento continuo $A = \{(x,y) \mid a \leq X \leq b, c \leq Y \leq d\}$

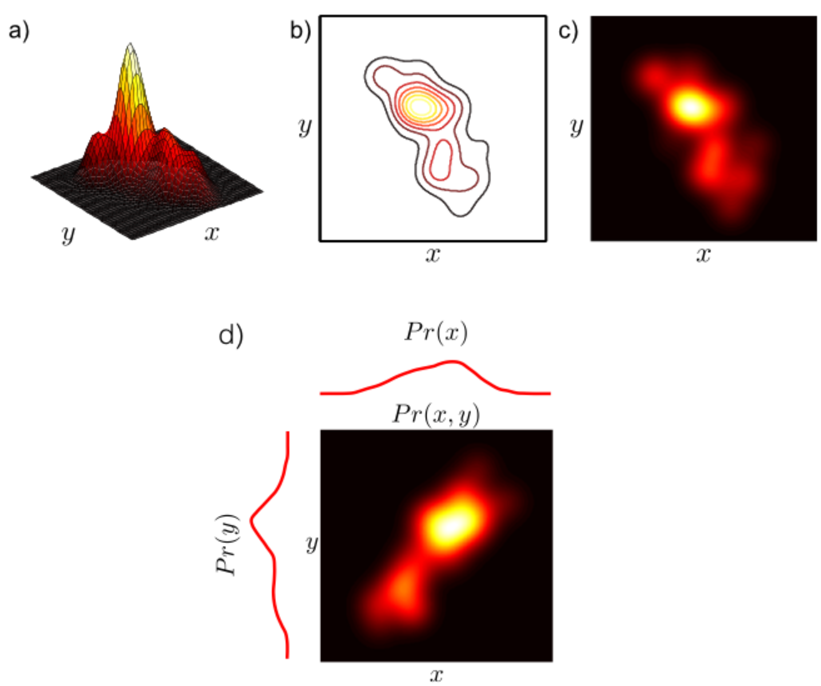

Figura 4: Le figure a), b) e c) sono tre possibili rappresentazioni grafiche di una generica densitá congiunta $prob(x,y)$: a) la rappresenta come un volume sul supporto bidimensionale x,y; b) come una mappa a livelli; c) come una mappa dove il colore é il valore de probabilitá assunto in ciascun punto (x,y) (il nero indica probabilitá nulla). La figura d) mostra le due distribuzioni marginali di x,y rispetto alla congiunta ottenute "accumulando" i valori della congiunta sull'asse x e sull'asse y.

Inoltre:

$$f_X(x) = \int_y f(x,y)dy$$

sará la densitá marginale rispetto a X, e

$$f_Y(y) = \int_x f(x, y)dx$$

la densitá marginale rispetto a Y. Una rappresentazione grafica di congiunte e marginali nel caso continuo é riportato in Figura 4.

Si noti che la funzione di ripartizione marginale $F_X(x)$ é per definizione $P_X(X \leq x)$. Esplicitamente, nel caso continuo:

$$F_X(x) = P_X(X \leq x) = F_{XY}(x, +\infty) = \int_{-\infty}^{x} \left[\int_{-\infty}^{+\infty} f_{X,Y}(u, v)dv \right] du = \int_{-\infty}^{x} f_X(u)du$$

Esempio 1.1 *Sia data la densitá congiunta continua:*

$$f_{XY}(x, y) = \begin{cases} 6e^{-2x-3y} & per\ x > 0, y > 0 \\ 0 & altrimenti \end{cases}$$

- *Calcolare $P_{XY}(1 \leq x \leq 2, 2 \leq y \leq 3)$:*

$$\begin{aligned} P_{XY}(1 \leq x \leq 2, 2 \leq y \leq 3) &= \int_1^2 \int_2^3 6e^{-2x-3y}dxdy \\ &= 6\int_1^2 e^{-2x}dx \int_2^3 e^{-3y}dy = 6\left[\frac{1}{2}e^{-2x}\right]_2^1 \left[\frac{1}{3}e^{-3y}\right]_3^2 \\ &= (e^{-2} - e^{-4})(e^{-6} - e^{-9}) = 0.0003 \end{aligned}$$

- *Calcolare la densitá marginale f_X:*

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{XY}(x, y)dy = \int_0^\infty 6e^{-2-3y}dy = 6e^{-2x}\int_0^\infty e^{-3y}dy = \\ &= 6e^{-2x}\frac{1}{3}[e^{-3y}]_\infty^0 = \frac{6}{3}e^{-2x}(1-0) = 2e^{-2x} \end{aligned}$$

- *Calcolare la CDF congiunta:*

  Ripetendo i calcoli precedenti si ottiene

$$F_{XY}(x, y) = \int_0^x \int_0^y 6e^{-2u-3v}dudv = (1 - e^{-2x})(1 - e^{-3y})$$

1.1 *Probabilitá condizionata*

Nel caso discreto, osservato un valore di $Y = y$, e data la distribuzione congiunta $P_{XY}(X, Y)$ é possibile definire la probabiitá condizionata di $X \mid Y = y$ nello stesso modo in cui avevamo definito la probabilitá condizionata di eventi:

$$P_{X|Y}(X|Y = y) = \frac{P_{XY}(X, Y = y)}{P_Y(Y = y)}$$

dove $P_Y(Y = y) > 0$ ed é ricavabile per marginalizzazione da $P_{XY}(X, Y)$.

Esempio 1.2 *Considerando l'esempio iniziale di Figura 2, possiamo calcolare* $P_{X|Y}(X|Y=1)$

$\rightarrow P_{X|Y}(X=0|Y=1) = \frac{P_{XY}(X=0,Y=1)}{P_Y(Y=1)} = \frac{0.2}{0.4} = 0.5$

$\rightarrow P_{X|Y}(X=1|Y=1) = \frac{P_{XY}(X=1,Y=1)}{P_Y(Y=1)} = \frac{0.2}{0.4} = 0.5$

$\rightarrow P_{X|Y}(X=2|Y=1) = \frac{P_{XY}(X=2,Y=1)}{P_Y(Y=1)} = \frac{0}{0.4} = 0$

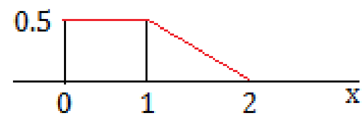

Figura 5: Probabilitá condizionata calcolata nell'Esempio 1.2 (la linea rossa serve solo a guidare l'occhio)

Nel caso continuo la densitá condizionata é definita come:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

.

In termini di CDF:

$$F_{X|Y}(x|y) = \frac{\int_{-\infty}^{x} f_{XY}(u,y)du}{\int_{-\infty}^{-\infty} f_{XY}(u,y)du}$$

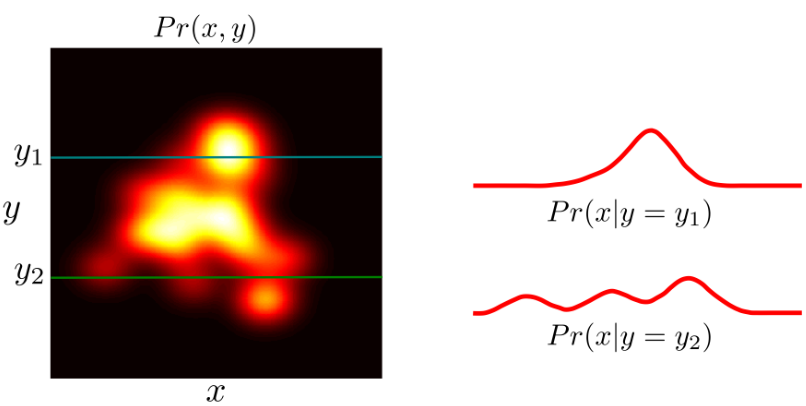

Figura 6: Rappresentazione grafica della densitá condizionata $prob(X \mid Y = y)$: questa é interpretabile come il "profilo" che si ottiene "tagliando" la densitá congiunta in $Y = y$

Se X e Y sono indipendenti valgono le seguenti

$$P_{XY}(x,y) = P_X(x)P_Y(y)$$

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

In generale (nei casi discreto e continuo):

$$F_{XY}(x,y) = F_X(x)F_Y(y)$$

Piú precisamente si dice che X e Y sono statisticamente o marginalmente indipendenti. Un caso immediato é giá stato visto alla fine dell'Esempio 1.1 dove

$$F_{XY}(x,y) = (1-e^{-2x})(1-e^{-3y})$$

Si puó facilmente verificare che $F_X(x) = (1 - e^{-2x})$ e $F_Y(y) = (1 - e^{-3y})$, pertanto $F_{XY}(x, y) = F_X(x)F_Y(y)$. Nell'esempio, dunque X e Y sono indipendenti, ed é il motivo per cui l'integrazione doppia risulta agevole essendo separabile su x e y.

Riportiamo infine, in Figura 7 un esempio notevole: la distribuzione Normale bivariata che si puó scrivere come

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2\left(-2\rho\frac{(x-\mu_x)(y-\mu_y}{\sigma_x\sigma_y}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

dove $\rho = \frac{cov(X,Y)}{\sigma_x\sigma_y}$ é il coefficiente di correlazione fra le variabili X e Y che vedremo piú avanti. Quando $\rho = 0$ allora X e Y sono indipendenti e la Normale bivariata é semplicemente il prodotto $\mathcal{N}(x \mid \mu_x, \sigma_x)\mathcal{N}(y \mid \mu_y, \sigma_y)$ delle due Normali univariate. Tuttavia, anche nel caso piú generale di $\rho \neq 0$ marginali e condizionate sono sempre distribuzioni Normali.

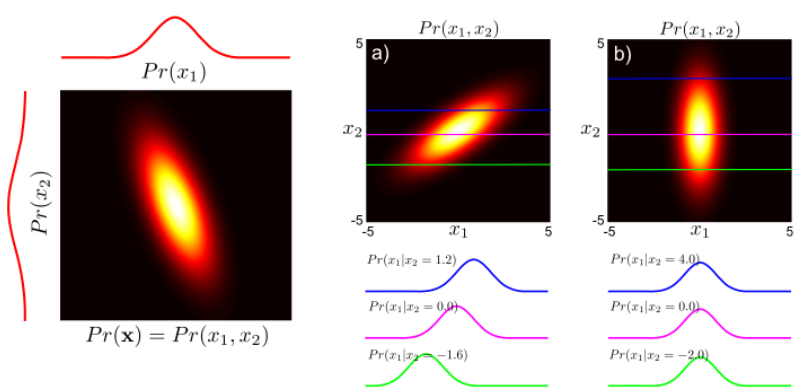

Figura 7: Un caso notevole: per una distribuzione Normale bidimensionale, marginali e condizionate sono ancora distribuzioni Normali (unidimensionali)

2 *Valore atteso e varianza*

Affrontiamo il problema di calcolare valore atteso e varianza per distribuzioni congiunte di VA X e Y. Per brevitá notazionale considereremo il caso continuo, ma i risultati sono estendibili immediatamente al caso discreto.

Data la densitá congiunta $f_{XY}(x, y)$, il suo valore atteso é:

$$\mu_x = E[X] = \int_x \int_y x f_{XY}(x,y)dxdy = \int_x xdx \int_y f_{XY}(x,y)dy = \int_x x f_X(x)dx$$

dove si é usata la definizione di densitá marginale di Y. Analogamente

$$\mu_y = E[Y] = \int_y y f_Y(y) dy$$

In generale, se abbiamo una variabile aleatoria $Z = g(X, Y)$ funzione di X e di Y, di densitá congiunta f_{XY}, avremo che:

$$\mu_z = E[Z] = \int \int g(x, y) f_{XY}(x, y) dx dy$$

Noti i valori di aspettazione μ_x e μ_y, le rispettive varianze sono

$$var(X) = \int_x \int_y (x - \mu_x)^2 f_{XY}(x, y) dx dy = \int_x (x - \mu_x)^2 f_X(x) dx$$

Analogamente:

$$var(Y) = \int_y (y - \mu_y)^2 f_Y(y) dy$$

3 *Valore atteso e varianza della somma di variabili aleatorie*

Date due VA X, Y, consideriamo la VA somma

$$Z = X + Y$$

e calcoliamone il valore atteso rispetto alla densitá congiunta $f(x, y)$:

$$\begin{aligned} E[Z] = E[X+Y] &= \int_{-\infty}^{-\infty} \int_{-\infty}^{-\infty} (x+y) f(x,y) dx dy \\ &= \int_{-\infty}^{-\infty} \int_{-\infty}^{-\infty} x f(x,y) dx dy + \int_{-\infty}^{-\infty} \int_{-\infty}^{-\infty} y f(x,y) dx dy \\ &= \int_{-\infty}^{-\infty} x f_X(x) dx + \int_{-\infty}^{-\infty} y f(y) dy \\ &= E[X] + E[Y] \qquad (1) \end{aligned}$$

In generale per n VA $X_1, X_2, \cdots X_n$ vale la seguente:

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] \qquad (2)$$

Date due VA X, Y, **indipendenti**, consideriamo la VA prodotto

$$Z = XY$$

allora vale la seguente:

$$E[Z] = E[XY] = E[X]E[Y] \qquad (3)$$

Infatti:

$$\begin{aligned} E[Z] = E[XY] &= \int_{-\infty}^{-\infty}\int_{-\infty}^{-\infty} xyf(x,y)dxdy \\ &= \int_{-\infty}^{-\infty} xf_X(x)dx \cdot \int_{-\infty}^{-\infty} yf_Y(y)dy \\ &= E[X]E[Y] \end{aligned} \tag{4}$$

dove si é fatto uso dell'ipotesi di indipendenza per fattorizzare la congiunta, $f(x,y) = f_X(x)f_Y(y)$.

Date due VA X, Y, questa volta **indipendenti**, consideriamo di nuovo la VA somma

$$Z = X + Y$$

e calcoliamone la varianza:

$$\begin{aligned} var(Z) = var[X+Y] &= E[(X+Y)^2] - E[(X+Y)]^2 \\ &= E[X^2 + Y^2 + 2XY] - (E[X] + E[Y])^2 \\ &= E[X^2] + E[Y^2] + 2E[XY] - E[X]^2 - E[Y])]^2 - 2E[X]E[Y] \\ &= E[X^2] + E[Y^2] + 2E[X]E[Y] - E[X]^2 - E[Y])]^2 - 2E[X]E[Y] \\ &= var(X) + var(Y) \end{aligned} \tag{5}$$

In generale per n VA $X_1, X_2, \cdots X_n$ **indipendenti** vale la seguente:

$$var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} var(X_i) \tag{6}$$

4 *Legge dei grandi numeri*

Sulla base delle proprietá fin qui dimostrate possiamo ora enunciare e dimostrare la legge dei grandi numeri (in forma debole)

Teorema 4.1 *Sia X una VA tale che*

$$E[X] = \mu \qquad var(X) = \sigma^2$$

*. Siano $X_1, X_2, \cdots X_n$ n VA **indipendenti e identicamente distribuite** (i.i.d.) come X*

$$X_i \sim X, \qquad \forall i,$$

in modo che il loro insieme costituisca un campione aleatorio di taglia n della VA X. Sia

$$S_n = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

la variabile aleatoria somma e

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

la media aritmetica del campione. Allora vale la seguente:

$$\lim_{n\to\infty} P(|\frac{S_n}{n} - \mu| \geq \epsilon) = 0 \qquad \forall \epsilon > 0 \tag{7}$$

ovvero $\frac{S_n}{n}$ *converge in probabilitá al valore atteso teorico* μ*:*

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

Dimostrazione *Il teorema si dimostra applicando la diseguaglianza di Chebychev che, ponendo* $\epsilon = k\ \ var(X)$*, riscriviamo nel seguente modo*

$$\begin{aligned} &P(|X - E[X]| \geq k\ var(X)) \leq \frac{1}{k^2} \\ &\rightarrow P(|X - E[X]| \geq \epsilon) \leq \frac{var(X)}{\epsilon^2} \end{aligned} \tag{8}$$

Usando $\frac{S_n}{n}$ *al posto della generica X*

$$P(|\frac{S_n}{n} - E\left[\frac{S_n}{n}\right]| \geq \epsilon) \leq \frac{var(\frac{S_n}{n})}{\epsilon^2} \tag{9}$$

Utilizzando le proprietá di valore atteso e varianza della somma di n VA iindipendenti , e le ipotesi che esse abbiano la stessa distribuzione (identicamente distribuite), per cui $E[X_i] = \mu,\ var(X_i) = \sigma^2$

$$E\ [\frac{S_n}{n}] = E\ [\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E\ [X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{n}{n}\mu = \mu$$

$$\begin{aligned} var\ (\frac{S_n}{n}) &= var\ (\frac{1}{n}\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} var\ (X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n}{n^2}\sigma^2 \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Sostituendo in Eq.(9) e prendendo il limite per $n \rightarrow \infty$

$$\lim_{n\to\infty} P(|\frac{S_n}{n} - \mu| \geq \epsilon) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0 \tag{10}$$

In conclusione la legge dei grandi numeri ci assicura che la media aritmetica $\frac{S_n}{n}$ converge in probabilitá al valore atteso teorico μ. La $\frac{S_n}{n}$ e' una VA. Qual é la legge di probabilitá che ci descrive la sua distribuzione, ovvero le sue fluttuazioni mentre sta convergendo a μ? Il Teorema Centrale Limite ci dice che tale distribuzione é la Gaussiana, indipendentemente dalle distribuzioni utilizzate per generare le X_i (purché abbiano media e varianza finite!). Per dimostrare questa proprietá ci serve prima definire uno strumento generale e univoco per caratterizzare la distribuzione di una VA.

Tale strumento é la **funzione generatrice dei momenti**.

5 Funzione Generatrice dei Momenti (f.g.m)

$$f.g.m. = \varphi_X(u) = E[e^{uX}] \begin{cases} \sum_x e^{ux} \cdot p_X(x) & \text{se } X \text{ discreta,} \\ \int e^{ux} \cdot f_X(x)\, dx & \text{se } X \text{ continua.} \end{cases} \tag{11}$$

Sviluppiamo ora questa funzione utilizzando McLaurin:

$$\begin{aligned} \varphi_X(u) = E[e^{uX}] &= E[1 + u \cdot X + \frac{u^2}{2} \cdot X^2 + \cdots] = \\ &= E[1] + u \cdot E[X] + \frac{u^2}{2} \cdot E[X^2] + \cdots = \\ &= \sum_{i=0}^{+\infty} \frac{E[X^i]}{i!} \cdot u^i \end{aligned} \tag{12}$$

Derivando i volte questa funzione é possibile ricavare il momento i-esimo:

$$E[X^i] = \frac{d^i}{du^i} \varphi_X(u) \bigg|_{u=0} \tag{13}$$

E' possibile verificare subito tale proprietá per media e varianza sfruttando la linearitá del valore atteso

$$\frac{d}{du}\varphi_X(u)|_{u=0} = \frac{d}{du}E[e^{uX}]|_{u=0} = E[\frac{d}{du}e^{uX}|_{u=0}] = E[Xe^{uX}|_{u=0}] = E[X];$$

analogamente

$$\frac{d^2}{du^2}\varphi_X(u)|_{u=0} = \frac{d^2}{du^2}E[e^{uX}]|_{u=0} = E[\frac{d^2}{du^2}e^{uX}|_{u=0}] = E[X^2 e^{uX}|_{u=0}] = E[X^2];$$

5.1 Distribuzione della V.A. somma di due V.A. indipendenti

Che distribuzione ha la VA che si definisce come somma di due V.A. **indipendenti**?

Lemma 5.1 *Siano X, Y V.A. indipendenti e sia $Z = X + Y$. Vale la seguente:*

$$\varphi_Z(u) = \varphi_X(x) \cdot \varphi_Y(y) \tag{14}$$

Dimostrazione *Utilizzando la proprietá di fattorizzazione di $E[]$ per il prodotto di VA indipendenti:*

$$\begin{aligned} \varphi_Z(u) &= E[e^{u(X+Y)}] = \\ &= E[e^{uX} \cdot e^{uY}] = \\ &= E[e^{uX}] \cdot E[e^{uY}] = \\ &= \varphi_X(u) \cdot \varphi_Y(u). \end{aligned} \tag{15}$$

Generalizzando, su n realizzazioni indipendenti $X_1, X_2, \cdots, X_n$ dove ciascuna X_i é campionata dalla stessa distribuzione, e Z é la V.A somma $Z = S_n$

$$S_n = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

$$\varphi_{S_n}(u) = \varphi_{\sum_{i=1}^{n} X_i}(u) = E[e^{u(\sum_{i=1}^{n} X_i)}] = \left[\varphi_{X_i}(u)\right]^n \tag{16}$$

Esempio 5.2 (f.g.m della Binomiale) *La f.g.m. della distribuzione Binomiale si puó calcolare facilmente, ricordando che la Binomiale si ottiene come la somma $X_1, X_2, \cdots, X_n$ dei risultati di n realizzazioni indipendenti $X_1, X_2, \cdots, X_n$ di una V.A. Bernoulliana, (per esempio, n lanci di moneta) cioé dove ogni X_i é distribuita con legge bernoulliana:*

$$X_i \sim Bern(p).$$

Calcoliamo quindi la f.g.m. della Bernoulli

$$\begin{aligned} \varphi_{Bern}(u) = E_{Bern}[e^{uX}] &= \sum_x e^{ux} \cdot p_X(x) = \sum_{x=0}^{1} e^{ux} p^x q^{1-x} \\ &= e^{u0} p^0 q^{1-0} + e^{u1} p^1 q^{1-1} = q + e^u p \end{aligned}$$

La Binomiale conteggia gli eventi di interesse $X_i \sim Bern(p)$ con $i = 1, \cdots, n$ e $X_i = \{0, 1\}$, dove $S_n = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i = k$ ($X_i = 1$ in caso di successo, altrimenti $X_i = 0$)

Facendo uso della f.g.m. della Bernoulli, e applicando l'equazione (16) la f.g.m. della Binomiale é

$$\varphi_{X \sim Bin(n,p)}(u) = \left[\varphi_{X_i \sim Bern(p)}(u)\right]^n = [q + p \cdot e^u]^n \tag{17}$$

Dopo aver calcolato la f.g.m. della distribuzione Binomiale, verifichiamo per esercizio come il calcolo del momento di ordine 1 (valor medio) mediante la derivata prima $\frac{d\varphi_{X \sim Bin(n,p)}(u)}{du}|_{u=0}$, corrisponda a $E_{Bin}[X] = np$:

$$\begin{aligned} E_{Bin}[X] = &\; \frac{d}{du} \varphi_{X \sim Bin(n,p)}(u) \Big|_{u=0} \\ = &\; \frac{d}{du} (q + p \cdot e^u)^n \Big|_{u=0} = n \cdot (q + p \cdot e^u)^{n-1} \cdot p \cdot e^u \Big|_{u=0} \\ = &\; n \cdot (q + p)^{n-1} \cdot p = n \cdot 1^{n-1} \cdot p = np. \end{aligned} \tag{18}$$

Esempio 5.3 (f.g.m. della Normale Standard) *Calcoliamo ora la f.g.m. della distribuzione Normale Standard $N(0,1)$che ci servirá per dimostrare il Teorema Centrale Limite:*

$$
\begin{aligned}
\varphi_{X\sim N(0,1)}(u) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{ux} \cdot e^{-\frac{1}{2}x^2} dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{u^2}{2}} \cdot e^{-\frac{1}{2}(x-u)^2} dx = \\
&= e^{\frac{u^2}{2}} \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x-u)^2} dx \\
&= e^{\frac{u^2}{2}}.
\end{aligned}
\tag{19}
$$

6 Teorema Centrale Limite

Come giá anticipato il Teorema Centrale Limite (TCL) ci dice che la distribuzione che descrive le deviazioni (fluttuazioni) delle somma di n variabili aleatorie rispetto al valore atteso teorico é la distribuzione Gaussiana, indipendentemente dalle distribuzioni utilizzate per generare le X_i, purché tali distribuzioni abbiano media e varianza finite.

Le ipotesi del TCL sono le stesse assunte nel caso della legge dei grandi numeri

Teorema 6.1 (TCL) *Siano $X_1, X_2, \cdots, X_n$ V.A. indipendenti e identicamente distribuite (i.i.d), siano $\mu < \infty$ e $\sigma^2 < \infty$ la media e la varianza di ciascuna V.A., sia $S_n = \sum_{i=1}^n X_i$, siano a, b valori reali con $a < b$. Allora vale la seguente:*

$$
\begin{aligned}
&\lim_{n\to+\infty} P(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b) = \\
=&\lim_{n\to+\infty} P(a \leq \frac{\sqrt{n}}{\sigma} \cdot \left(\frac{S_n}{n} - \mu\right) \leq b) = \\
=&\frac{1}{\sqrt{2\pi}} \cdot \int_a^b e^{-\frac{1}{2}x^2} dx = \\
=&\phi(b) - \phi(a).
\end{aligned}
\tag{20}
$$

Dimostrazione *La dimostrazione si riduce a verificare che la funzione generatrice della VA standardizzata*

$$Z_n = \frac{\frac{S_n}{n} - E[\frac{S_n}{n}]}{\sqrt{var(\frac{S_n}{n})}} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{n}}$$

é la f.g.m della Gaussiana standard $\varphi_{N(0,1)}(u) = e^{\frac{u^2}{2}}$.

Riscriviamo la Z_n come

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\frac{X_1-\mu}{\sigma} + \frac{X_2-\mu}{\sigma} + \cdots + \frac{X_n-\mu}{\sigma}}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} t_j$$

dove $t_j = \frac{X_j-\mu}{\sigma}$ *sono standard con distribuzione* $N(0,1)$.
La f.g.m. di Z_n *é, per la proprietá della f.g.m. della somma di VA,*

$$\varphi_{Z_n}(u) = \varphi_{\frac{1}{\sqrt{n}}\sum_{j=1}^{n} t_j}(u) = \varphi_{\sum_{j=1}^{n} t_j}\left(\frac{u}{\sqrt{n}}\right) = \prod_{j=1}^{n}\varphi_{t_j}\left(\frac{u}{\sqrt{n}}\right) = \left[\varphi_{t_j}\left(\frac{u}{\sqrt{n}}\right)\right]^n$$

La f.g.m. delle t_j, *sviluppando in serie l'esponenziale, con* $E[t_j] = 0, E[t_j^2] = 1$ *giacché standard, vale*

$$\varphi_{t_j}\left(\frac{u}{\sqrt{n}}\right) = 1 + \frac{u}{\sqrt{n}}E[t_j] + \frac{u^2}{2n}E[t_j^2] + \cdots = 1 + \frac{u^2}{2n} + o\left(\frac{t_j^2}{n}\right)$$

Pertanto

$$\varphi_{Z_n}(u) = \left[1 + \frac{u^2}{2n} + o\left(\frac{t_j^2}{\sqrt{n}}\right)\right]^n$$

Prendendo il logaritmo a destra e sinistra

$$\log\varphi_{Z_n}(u) = n\log\left[1 + \frac{u^2}{2n} + o\left(\frac{t_j^2}{n}\right)\right]$$

Poniamo

$$y = \frac{u^2}{2n} + o\left(\frac{t_j^2}{n}\right)$$

e riscriviamo

$$\log\varphi_{Z_n}(u) = ny\frac{\log(1+y)}{y}.$$

di cui vogliamo calcolare il limite:

$$\lim_{n\to\infty}\log\varphi_{Z_n}(u) = \lim_{n\to\infty} ny\frac{\log(1+y)}{y}.$$

Notiamo che per come é stata definita y,

$$\lim_{n\to\infty} y = \lim_{n\to\infty}\left[\frac{u^2}{2n} + o\left(\frac{t_j^2}{n}\right)\right] = 0,$$

e che per $y \to 0$ *(ovvero per* $n \to \infty$*) possiamo applicare il limite notevole*

$$\lim_{y\to 0}\frac{\log(1+y)}{y} = 1.$$

Inoltre,

$$\lim_{n\to\infty} ny = \lim_{n\to\infty} n \cdot \frac{u^2}{2n} + \lim_{n\to\infty} n \cdot o\left(\frac{t_j^2}{n}\right) = \frac{u^2}{2} + 0.$$

Infine

$$\lim_{n\to\infty} \log \varphi_{Z_n}(u) = \frac{u^2}{2},$$

che, ritornando in forma esponenziale, si scrive

$$\lim_{n\to\infty} \varphi_{Z_n}(u) = e^{\frac{u^2}{2}}.$$

Abbiamo cosí dimostrato che la f.g.m. della nostra V.A. Z_n *nel limite* $n \to \infty$ *corrisponde alla f.g.m. di una distribuzione Normale, dunque la* Z_n *é distribuita con legge Normale,* $N(0,1)$. *La tesi del teorema segue immediatamente.*

Il TCL non é solo un fondamentale risultato teorico ma ha anche rilevanza per la pratica sperimentale. Molto spesso una VA puó essere pensata come il risultato finale degli effetti di molte variabili concomitanti che contribuiscono a determinare il valore della variabile considerata. Ad esempio, l?altezza di un individuo é determinata da molti fattori: genetici, alimentari, ambientali. Possiamo descrivere la fluttuazione dell?altezza individuale rispetto al valore medio della popolazione come dovuta ad una somma di contributi dovuti a ciascuna di queste variabili. Il gran numero di fattori che influenzano il valore dell?altezza ci porta ad assumere che l?altezza sia distribuita nella popolazione secondo legge Gaussiana. In generale, il TCL é uno strumento estremamente potente, che consente per esempio di affrontare lo studio della precisione di misure sperimentali.

Vediamo infine un semplice esempio di applicazione pratica del TCL.

Esempio 6.2 Un vostro amico asserisce di aver ottenuto una media di 3.25 punti per lancio su 1000 lanci di un dado non truccato. Qual é la probabilitá che stia mentendo?

Soluzione. Sia X_i *la VA che denota l'esito dell'i-simo lancio di dado.*

Sappiamo che le VA $X_1, X_2, \cdots, X_{1000}$ *sono indipendenti e identicamente distribuite con distribuzione uniforme sull'insieme di interi* $\{1, 2, \cdots, 6\}$ *da cui é facile ricavare il valore atteso e la deviazione standard*

$$E[X_i] = 3.5$$

$$\sigma_{X_i} = \sqrt{E[X_i^2] - E[X_i]^2} \approx 1.7078$$

In breve, il Teorema Centrale Limite ci dice che per un numero n grande ($n \to \infty$) di V.A. indipendenti e identicamente distribuite di valore atteso μ e deviazione standard σ vale

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \tag{21}$$

dove $S_n = \sum_{i=1}^{n} X_i$. Nella fattispecie $n = 1000$ e S_n é il numero di punti totalizzati in 1000 lanci.

Sotto tali ipotesi, la probabilitá che il vostro amico affermi il vero si riduce a calcolare la probabilitá definita nel TCL riscritta come

$$P\left(\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) = \Phi(x) \tag{22}$$

dove $\frac{S_n}{n} = 3.25 = \overline{X}$ é la media campionaria del campione di taglia $n = 1000$ dichiarata dal vostro amico. In buona sostanza il TCL. il teorema garantisce che la variabile ridotta $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ sia una variabile aleatoria la cui funzione di distribuzione tende alla distribuzione normale standard Φ. Pertanto,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{3.25 - 3.5}{1.7078/\sqrt{1000}} = -4.629$$

La probabilitá che il vostro amico dichiari il vero é dunque pari a $\Phi(-4.629)$

Anche senza consultare le tabelle e tenendo presente che per la normale standard $\sigma = 1$ e dunque $z = z\sigma$, significa che stiamo valutando la probabilitá di un valore di media dichiarata che é di 4.629 deviazioni standard al di sotto della media $\mu = 3.5$.

Considerando la legge 3σ, é immediatamente evidente che la probabilitá $\Phi(-4.629)$ é molto bassa e pertanto la probabilitá che che il vostro amico dica il falso molto elevata $(1 - \Phi(-4.629))$.

Se poi, per scrupolo, si consultasse la tabella della normale ridotta, si leggerebbe che $\Phi(-4.629) = 1.84 \times 10^{-7}$