A ...

"Without data you're just another person with an opinion"
W. Edwards Deming

Indice

Sommario

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras mattis tincidunt ligula. Duis ante neque, convallis vel vulputate vel, dignissim vel enim. Proin et iaculis libero. Aliquam erat volutpat. Cras ac purus non ante ultricies scelerisque. Donec lobortis lorem imperdiet leo consequat nec iaculis velit adipiscing. Curabitur nec gravida neque. Nunc vel dui vitae ante dapibus sagittis ac non libero. Suspendisse gravida commodo arcu bibendum luctus. Nam placerat pharetra massa, aliquam rutrum arcu fermentum nec. In non ultrices ante. Pellentesque pretium, felis ac mattis condimentum, dui massa ultricies nisl, hendrerit malesuada magna risus eget dolor. Pellentesque lobortis eleifend nibh, sed gravida sem fringilla eget. Proin pretium, arcu in ornare pellentesque, elit ante faucibus sem, at convallis eros ante ut velit. Donec ornare erat non diam tristique vitae congue nulla commodo. Proin fermentum fringilla mattis. Pellentesque ut dolor hendrerit tellus tincidunt egestas at sit amet velit.

Capitolo 1

La distribuzione Normale Asimmetrica multidimensionale

Definita da A. Azzalini e A. Dalla Valle nel 1996 [?] come generalizzazione della corrispettiva classe di distribuzioni univariata proposta da Azzalini nel 1985 [?] la famiglia di distribuzioni Normale Asimmetrica multidimensionale presenta proprietà ideali tra cui:

- inclusione della distribuzione Normale;
- tracciabilità matematica;
- ampio numero di indici per il calcolo di asimmetria e curtosi;
- possibilità di passaggio tra normalità e non normalità attraverso un parametro di regolarizzazione.

Grazie a queste proprietà la classe di distribuzioni Normali Asimmetriche permette un ottimo adattamento ai dati con relativa semplicità nel trattamento dal punto di vista matematico vista l'analogia con la distribuzione normale. Attraverso l'utilizzo del parametro di regolarizzazione è inoltre possibile modificare asimmetria e curtosi della distribuzione in quanto l'introduzione di asimmetria modifica anche le code della distribuzione portando ad una modifica della curtosi della stessa. Risulta necessario quindi ottenere degli indici che permettano di calcolare sia l'asimmetria [?] che della curtosi [?] in modo da permettere una descrizione completa della forma della distribuzione.

1.1 Distribuzione Normale Asimmetrica multidimensionale

La definizione della famiglia di distribuzioni Normale Asimmetrica multivariata nel campo pratico è stata estremamente rilevante in quanto, nel caso multivariato, il numero di distribuzioni capaci di modellare leggere asimmetrie per le distribuzioni marginali è estremamente ridotto; inoltre la famiglia Normale Asimmetrica multivariata presenta una grandissima flessibilità grazie al parametro di regolarizzazione permettendo quindi un'ottima capacità di adattamento a diversi tipi di dati. La generalizzazione multivariata della famiglia Normale Asimmetrica fornita da Azzalini e Dalla Valle permette di ottenere una distribuzione le cui marginali sono a loro volta delle distribuzioni Normali Asimmetriche.

Definizione 1.1. (Azzalini e Dalla Valle, 1996) una variabile continua p-dimensionale Z è detta avere distribuzione Normale Asimmetrica multivariata ($Z \sim SN_p(\overline{\Omega}, \alpha)$) se è continua e ha funzione di densità

$$2\phi_p(z; \overline{\Omega})\Phi(\alpha^\top z) \qquad (z \in \Re)$$

Il parametro di regolarizzazione α è detto in questo caso parametro di forma e permette di regolare asimmetria e curtosi. Con $\alpha = 0$ si ottiene la distribuzione Normale Multivariata con matrice di correlazione $\overline{\Omega}$. Questa definizione assume $\mu = 0$, per questo Azzalini nel 2005 propone una generalizzazione con l'introduzione di un parametro di posizione (ξ di dimensione $p \times 1$) e uno di scala (ω matrice diagonale di dimensione $p \times p$). L'introduzione del parametro di posizione ξ permette di centrare la distribuzione in un valore diverso da 0.

Teorema 1.2. *(Azzalini e Capitanio, 1999) Sia Z una variabile Normale Asimmetrica p-dimensionale, si ottiene che $Y = \xi + \omega Z$ con $\xi = (\xi_1, \ldots, \xi_p)^\top$ e $\omega = diag(\omega_1, \ldots, \omega_p)$ ha funzione di densità*

$$2\phi_p((y - \xi); \Omega)\Phi(\alpha^\top \omega^{-1}(y - \xi)) \qquad (z \in \Re)$$

con $\phi_p((y-\xi); \Omega)$ densità di una normale con media ξ e matrice di varianze e covarianze $\Omega = \omega\overline{\Omega}\omega$, per la variabile Y si indicherà la distribuzione come $Y \sim SN_p(\xi, \Omega, \alpha)$

1.1.1 I primi quattro momenti della distribuzione Normale Asimmetrica multidimensionale

Azzalini e Dalla Valle (1996) hanno calcolato la funzione generatrice dei momenti nel caso in cui $Y \sim SN_p(\Omega, \alpha)$

Teorema 1.3. *(Azzalini e Dalla Valle, 1996) Sia Z una variabile Normale Asimmetrica p-dimensionale con $Z \sim SN_p(\Omega, \alpha)$, allora la funzione generatrice dei momenti per Z è pari a:*

$$\begin{aligned} M_4 &= 2 \int_{\Re^p} \exp(t^\top z) \phi_p(z; \Omega) \Phi(\alpha^\top z) dz \\ &= 2 \exp(\tfrac{1}{2} t^\top \Omega t) \Phi(\delta^\top t) \qquad (t \in \Re^p) \end{aligned}$$

con

$$\delta = \frac{\Omega\alpha}{(1 + \alpha^\top \Omega \alpha)^{\frac{1}{2}}}$$

Grazie a questo teorema Genton nel 2001 calcolata i primi quattro momenti per una distribuzione Normale Asimmetrica p-variata

Teorema 1.4. *(Genton et al., 2001): Sia Z un vettore casuale con distribuzione $Z \sim SN_p(\alpha, \Omega)$, allora:*

$$\begin{aligned} M_1 &= \sqrt{\tfrac{2}{\pi}} \delta \\ M_2 &= \Omega \\ M_3 &= \sqrt{\tfrac{2}{\pi}} \left[\delta \otimes \Omega + vec\Omega\delta^\top + (I_p \otimes \delta) \Omega - (I_p \otimes \delta) \left(\delta \otimes \delta^\top \right) \right] \\ M_4 &= (I_{p^2} + U_{p,p})(\Omega \otimes \Omega) + vec(\Omega) vec(\Omega^\top) \end{aligned}$$

Generalizzando i momenti per gli una distribuzione con parametro di posizione $\xi \neq 0$ Genton (2001) calcola i primi quattro momenti non centrati per la distribuzione $Y \sim SN_p(\xi, \Omega, \alpha)$

Teorema 1.5. *(Azzalini, 2005): Sia Y un vettore casuale con distribuzione $Y \sim SN_p(\xi, \Omega, \alpha)$, allora la funzione generatrice dei momenti per Y è:*

$$M(t) = 2 \exp(\xi^\top t + \frac{1}{2} t^\top \Omega t) \Phi(\delta^\top t) \qquad (t \in \Re^p)$$

Teorema 1.6. *(Genton et al., 2001): Sia Y un vettore casuale con distribuzione $Y \sim SN_p(\xi, \Omega, \alpha)$, allora i primi quattro momenti di Y sono:*

$$
\begin{aligned}
M_1 &= \xi + \sqrt{\tfrac{2}{\pi}}\delta \\
M_2 &= \Omega + \xi\xi^\top + \sqrt{\tfrac{2}{\pi}}\left(\xi\delta^\top + \delta\xi^t op\right) \\
M_3 &= \Omega\otimes\xi + \xi\otimes\Omega + vec(\Omega)\otimes\xi^\top + \xi\otimes\xi^\top\otimes\xi + \sqrt{\tfrac{2}{\pi}}[\delta\otimes\Omega + vec(\Omega)\delta^\top \\
&\quad + (I_p\otimes\delta)\,\Omega - \delta\otimes\delta^\top\otimes\delta + \delta\otimes\xi^\top\otimes\xi + \xi\otimes\delta^\top\otimes\xi + \xi\otimes\xi^\top\otimes\delta] \\
M_4 &= \Omega\otimes\xi\otimes\xi^\top + \xi\otimes\Omega\otimes\xi^\top + vec(\Omega)\otimes\xi^\top\otimes\xi^\top + \xi\otimes\xi^\top\otimes\xi\otimes\xi^\top + \Omega\otimes\Omega \\
&\quad + vec(\Omega)vec(\Omega)^\top + U_{p,p}(\Omega\otimes\Omega) + \xi^\top\otimes\Omega\otimes\xi + \xi\otimes\xi\otimes vec(\Omega)^\top + \xi\otimes\xi^\top\otimes\Omega \\
&\quad + \sqrt{\tfrac{2}{\pi}}[\delta\otimes\Omega\otimes\xi^\top + vec(\Omega)\otimes\delta^\top\otimes\xi^\top + ((I_p\otimes\delta)\Omega)\otimes\xi^\top + \delta\otimes\xi^\top\otimes\xi\otimes\xi^\top \\
&\quad + \xi\otimes\delta^\top\otimes\xi\otimes\xi^\top + \xi\otimes\xi^\top\otimes\delta\otimes\xi^\top + \delta^\top\otimes\Omega\otimes\xi + \delta\otimes vec(\Omega)^\top\otimes\xi \\
&\quad + (\Omega(I_p\otimes\delta^\top))\otimes\xi + \xi^\top\otimes\delta\otimes\Omega + \xi^\top\otimes(vec(\Omega)\delta^\top) + \xi^\top\otimes((I_p\otimes\delta)\Omega) \\
&\quad + \xi\otimes\delta^\top\otimes\Omega + \xi\otimes\delta\otimes vec(\Omega)^\top + \xi\otimes(\Omega(I_p\otimes\delta^\top)) + \xi\otimes\xi^\top\otimes\delta\otimes\delta^\top \\
&\quad - \delta\otimes\delta^\top\otimes\delta\otimes\xi^\top - \delta^\top\otimes\delta\otimes\delta^\top\otimes\xi - \xi^\top\otimes\delta\otimes\delta^\top\otimes\delta - \xi\otimes\delta^\top\otimes\delta\otimes\delta^\top]
\end{aligned}
$$

1.1.2 Le proprietà statistiche della distribuzione Normale Asimmetrica multidimensionale

Capitolo 2

Indici di Curtosi per la Normale Asimmetrica multidimensionale

Nella descrizione descrizione delle caratteristiche di una distribuzione, oltre agli indici di posizione, variabilità globale e asimmetria è utile definire anche un indice di curtosi. Etimologicamente la parola deriva dal greco $\kappa\upsilon\rho\tau\acute{o}\varsigma$ che significa "curvo, arcuato". In generale la curtosi è un indice di forma della curva definito usualmente come rapporto tra la lontananza delle osservazioni dall'indice di posizione rispetto alla sua distanza media, permettendo di definire la pesantezza delle code di una distribuzione. L'indice di curtosi calcola quindi l'allontanamento dalla normalità distributiva a parità di media e varianza, verificando un maggior appiattimento (distribuzione platicurtica) o un maggior appuntimento (distribuzione leptocurtica) della distribuzione. È importante tenere conto però del fatto che oltre a dipendere dall'andamento delle code della distribuzione l'indice di curtosi dipende anche dal comportamento della stessa nella sua parte centrale, un ispessimento delle code della distribuzione porterà ad un minor numero di osservazioni nella parte centrale della distribuzione e viceversa code meno spesse portano ad un maggior numero di osservazioni nella parte centrale in quanto l'integrale della densità deve valere sempre 1. La complessità di questo fenomeno ha portato negli anni alla definizione di vari indici per il calcolo della curtosi.

Nel caso multidimensionale si ha un aumento della complessità di questo indice poiché è intrinsecamente connesso con le altre caratteristiche della distribuzione come la pesantezza delle code, la variabilità e l'asimmetria, inoltre la presenza di più dimensioni connesse tra loro complica l'interpretazione di questo indice

2.1 L'indice di Mardia

2.2 L'indice di Malkovich-Afifi

2.3 Il nuovo indice direzionale

2.4 L'indice di Srivastava

2.5 L'indice di Mori-Rohatgi-Székeley

2.6 L'indice di Kollo

Capitolo 3

Confronto tra le misure di Curtosi

3.1 Performance delle misure di Curtosi

Capitolo 4

Conclusioni

Appendices

Operatori Matematici

The contents...

Codice R

The contents...