

Università degli Studi di Padova  
Dipartimento di Scienze Statistiche

Corso di Laurea Triennale in  
Statistica per l'Economia e l'Impresa



**Local Gaussian correlation:  
una misura di dipendenza per serie storiche non lineari**

Relatrice: Prof.ssa Luisa Bisaglia  
Dipartimento di Scienze Statistiche

Candidato: Alessandro Boscarini  
Matricola n. 1164766

Anno Accademico 2022/2023

# Indice

<b>Introduzione</b>	<b>2</b>
<b>1 L'indice <math>\rho</math> di Pearson</b>	<b>3</b>
1.1 Definizione . . . . .	3
1.2 Criticità dell'indice $\rho$ di Pearson . . . . .	4
1.2.1 Non-gaussianità . . . . .	4
1.2.2 Robustezza . . . . .	5
1.2.3 Non-linearità . . . . .	5
1.3 Misurare la dipendenza statistica . . . . .	6
<b>2 Local Gaussian correlation</b>	<b>7</b>
2.1 La funzione di densità . . . . .	8
2.2 La funzione parametro $\theta(x)$ . . . . .	9
2.3 Teoria asintotica con $b$ fissato . . . . .	12
2.4 Considerazioni per la pratica . . . . .	15
2.5 Proprietà della <i>local Gaussian correlation</i> . . . . .	16
<b>3 Applicazione della LGC</b>	<b>19</b>
3.1 Distribuzione Normale bivariata . . . . .	20
3.2 Distribuzione t-Student bivariata . . . . .	22
3.3 Modello non lineare $Y = X^2$ . . . . .	24
3.4 Il modello GARCH(1,1) . . . . .	26
<b>Conclusione</b>	<b>28</b>

## Introduzione

Uno dei punti cruciali nelle analisi statistiche è misurare e valutare la correlazione tra le variabili. La conclusione alla quale si arriva determina se effettivamente le variabili siano più o meno correlate e questo dato può influenzare l'intera analisi. L'indice che si occupa di suggerirci se vi è correlazione o meno tra le variabili è l'indice  $\rho$  di Pearson. Un'errata misura di questo indice porterebbe a conclusioni fuorvianti, perciò avere la certezza che il valore che si ottiene sia corretto è un passo fondamentale a fini statistici. E' noto però da ormai un po' di tempo che questo indice presenta delle criticità quando lo si utilizza per certi tipi di modelli. Queste criticità, come in tutti i campi statistici, rappresentano una problematica. E' dunque indispensabile avere un indice alternativo a  $\rho$  e alla sua applicazione analoga nel campo delle serie storiche (l'ACF o funzione di autocorrelazione). A questo proposito nel tempo gli scienziati si sono messi al lavoro per trovare un rimedio. Tra le possibili soluzioni in questo elaborato è stata trattata una misura di dipendenza che calcola l'indice  $\rho$  di Pearson localmente, nota come *local Gaussian correlation* (LGC).

Nel capitolo 1 si darà una visione generale di  $\rho$  andando poi nello specifico ad elencare e valutare le sue criticità. Nel capitolo 2 si analizzerà la teoria e la funzionalità della *local Gaussian correlation*. Nel capitolo 3, infine, verranno simulati alcuni modelli non lineari e, in particolare, verranno simulate serie storiche partendo da un modello GARCH(1,1), così da estendere la teoria e la pratica che sta alla base della LGC al campo delle serie storiche.

# 1 L'indice $\rho$ di Pearson

## 1.1 Definizione

Uno degli indici più usati per la misura della dipendenza statistica è il  $\rho$  di Pearson. E' un indice che permette di calcolare sia la forza che la direzione della relazione lineare tra due variabili quantitative nel caso gaussiano e in alcuni casi di non gaussianità. Una delle caratteristiche che lo contraddistingue è la versatilità, in quanto può essere sfruttato in moltissimi ambiti per calcolare la correlazione tra variabili. E' stato denominato "indice  $\rho$  di Pearson", ma in realtà chi lo inventò fu Francis Galton verso la fine dell'Ottocento. Pearson (1896), partendo dagli studi compiuti in precedenza da Galton, li sviluppò in maniera più meticolosa, soprattutto dal punto di vista matematico.

Nello specifico, date due variabili aleatorie  $X$  e  $Y$ , aventi i momenti secondi finiti, l'indice  $\rho$  di Pearson è definito da:

$$\rho = \frac{E[(X - E(X))(Y - E(Y))]}{\sigma_X \sigma_Y} \quad (1.1)$$

dove il numeratore rappresenta la covarianza fra le due variabili, mentre il denominatore il prodotto tra le deviazioni standard delle variabili  $X$  e  $Y$ .

Dato un insieme di osservazioni  $(x_1, y_1), \dots, (x_n, y_n)$  la versione campionaria di  $\rho$  è data da:

$$r = \hat{\rho} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1.2)$$

dove  $\bar{x} = n^{-1} \sum_{i=1}^n x_i$  e  $\bar{y} = n^{-1} \sum_{i=1}^n y_i$ .

La correlazione assume valori che variano nell'intervallo  $[-1, +1]$ , dove  $-1$  indica una correlazione negativa,  $0$  indica incorrelazione e infine  $+1$  correlazione positiva.

## 1.2 Criticità dell'indice $\rho$ di Pearson

Come detto precedentemente l'indice  $\rho$  di Pearson è uno degli indici più usati nelle analisi statistiche, però presenta anch'esso delle criticità che possono portare a conclusioni fuorvianti.

### 1.2.1 Non-gaussianità

Considerando la distribuzione di una variabile aleatoria  $X$  che segue quella di una normale multivariata definita come  $X \sim N(u, \Sigma_p)$  e sussistendo una connessione significativa tra i concetti di correlazione e gaussianità, è possibile dimostrare che se due o più componenti gaussiane sono incorrelate allora vi è indipendenza tra queste.

Si ritiene opportuno quindi chiedersi se questo forte legame tra il concetto di correlazione e gaussianità valga anche per altre distribuzioni. Per rispondere a questa domanda ci si sofferma sulla distribuzione gaussiana multivariata la quale appartiene ad una classe molto più ampia di distribuzioni di probabilità, ovvero le distribuzioni ellittiche. Per questa tipologia di distribuzioni, infatti, la relazione tra incorrelazione e indipendenza non è sempre vera. A tal proposito si consideri ad esempio la distribuzione t-Student multivariata e, in particolare, quella bivariata la quale appartiene alla classe di distribuzioni ellittiche: è possibile dimostrare che se c'è incorrelazione tra due variabili non vale per forza anche l'indipendenza tra queste, in quanto considerando una generica distribuzione multivariata t-Student con  $v$  gradi di libertà

$$f(x) = \frac{\Gamma(\frac{p+v}{2})}{(\pi v)^{p/2} \Gamma(v/2) |\Sigma|^{1/2}} \times \left( 1 + \frac{(x-u)^T \Sigma^{-1} (x-u)}{v} \right)^{-\frac{p+v}{2}}, \quad (1.3)$$

dove  $\Sigma$  è una matrice definita positiva. Se anche quest'ultima fosse una matrice diagonale, ovvero una matrice quadrata con valori diversi da zero solo sulla diagonale e nulli altrove, non garantirebbe comunque l'incorrelazione tra le variabili.

Si può quindi concludere che è necessario prestare molta attenzione all'utilizzo dell'indice  $\rho$  di Pearson, non solo quando si va a calcolare la dipendenza per le distribuzioni ellittiche, ma in generale per tutte le distribuzioni non gaussiane in quanto l'incorrelazione tra le componenti non implica anche l'indipendenza tra queste.

### 1.2.2 Robustezza

La seconda criticità consiste nella non-robustezza dell'indice in presenza di valori anomali, detti anche *outliers*. Molti studiosi si sono interessati al problema e ad oggi esistono principalmente due metodi che permettono a  $\rho$  di godere di robustezza: il coefficiente di correlazione  $\tau$  di Kendall (1938) e il coefficiente di correlazione basato su ranghi di Spearman (1961).

### 1.2.3 Non-linearità

La non-linearità è l'ultima criticità che viene trattata ma di sicuro quella più problematica, oltre che per  $\rho$  di Pearson anche per gli indici di Spearman e Kendall. Se si considerasse infatti un modello non lineare in cui è presente una dipendenza significativa,  $\rho$  non sarebbe in grado di rilevarla in quanto quest'ultimo è in grado di misurare unicamente la relazione lineare tra due variabili. Applicato al campo delle serie storiche un modello non lineare potrebbe essere il GARCH(1,1) che è definito da:

$$\begin{aligned} X_t &= \epsilon_t h_t^{1/2}, \\ h_t &= \alpha + \beta h_{t-1} + \gamma X_{t-1}^2, \end{aligned} \tag{1.4}$$

dove il processo stocastico  $\{h_t\}$  è indipendente da  $\{\epsilon_t\}$ , ovvero una serie storica di variabili i.i.d. a media zero.

Un altro esempio potrebbe essere banalmente il modello parabolico  $Y = X^2$ . Anche in questo caso è possibile dimostrare che  $\rho$  non è attendibile in quanto simulando un modello come quello appena descritto l'indice  $\rho$  di Pearson risulterebbe vicino a 0, quando in realtà vi è una chiara dipendenza tra  $X$  e  $Y$ .

### 1.3 Misurare la dipendenza statistica

Dopo aver chiarito le criticità di  $\rho$ , alcune delle misure principalmente utilizzate per aggirare queste problematiche sono:

- la copula, Nelsen (2007)
- la *distance covariance*, Székely e Rizzo (2012)
- misure di dipendenza alternative

Tra le varie misure di dipendenza alternative nelle prossime sezioni ci si soffermerà sulla *local Gaussian correlation* ritenendo opportuno rimandare a Tjøstheim *et al.* (2022) per una più completa trattazione delle altre metodologie.

## 2 Local Gaussian correlation

Da quanto appena illustrato, è evidente, quindi, che l'indice  $p$  di Pearson non possa essere utilizzato in tutti i casi. Per far fronte a questo inconveniente si usufruirà di un metodo proposto da Tjøstheim e Hufthammer (2013), ovvero la *local Gaussian correlation*: un approccio per la misura della dipendenza statistica tra due variabili stocastiche  $X$  e  $Y$ . Tale misura gode di un'importante proprietà: se le due variabili sono congiuntamente distribuite come normali la LGC è in grado di separare la dipendenza non lineare da quella positiva a quella negativa attraverso una distribuzione gaussiana bivariata  $\psi_{x,y}(v, u)$ , pur riconducendosi al coefficiente di correlazione di Pearson usualmente adottato.

**Notazione:** Durante tutta la trattazione teorica, per comodità, si utilizzerà  $(x, y)$  anziché  $(x_1, x_2)$ , con  $x = (x_1, x_2)$ . Analogamente,  $(v, u)$  verrà sostituito da  $v = (v_1, v_2)$ . Per quanto riguarda la media e la varianza della funzione  $\psi : u(x) = (u_1(x), u_2(x))$  sarà il vettore delle medie,  $\sigma(x) = (\sigma_1(x), \sigma_2(x))$  rappresenterà il vettore delle deviazioni standard, mentre  $\rho(x)$  l'indice di correlazione. Inoltre, si utilizzerà il termine "famiglia" per indicare un gruppo di elementi.

Il primo passo, quindi, per spiegare in cosa consiste questa nuova misura di dipendenza è quello di definire la funzione di densità di una distribuzione gaussiana bivariata.



## 2.1 La funzione di densità

La funzione di densità di una distribuzione gaussiana bivariata è definita da:

$$\begin{aligned}
 & \psi(v, u_1(x), u_2(x), \sigma_1^2(x), \sigma_2^2(x), \rho(x)) \\
 &= \frac{1}{2\pi\sigma_1(x)\sigma_2(x)\sqrt{1-\rho^2(x)}} \\
 & \times \exp \left[ -\frac{1}{2} \frac{1}{1-\rho^2(x)} \left( \frac{(v_1 - u_1(x))^2}{\sigma_1^2(x)} \right. \right. \\
 & \quad \left. \left. - 2\rho(x) \frac{(v_1 - u_1(x))(v_2 - u_2(x))}{\sigma_1(x)\sigma_2(x)} \right. \right. \\
 & \quad \left. \left. + \frac{(v_2 - u_2(x))^2}{\sigma_2^2(x)} \right) \right].
 \end{aligned} \tag{2.1}$$

Quando ci si sposta dal punto  $x$  al punto  $x'$  si ottiene una nuova densità  $\psi(v, u_1(x'), u_2(x'), \sigma_1^2(x'), \sigma_2^2(x'), \rho(x'))$ , che approssima  $f$  (generica funzione di densità bivariata) in un intorno di  $x'$ . In questo modo  $f$  è rappresentata da una famiglia di funzioni di densità bivariate al variare di  $x$ . In sostanza, quindi, in ogni specifico intorno di un punto, le proprietà di dipendenza locale sono descritte dalla matrice di covarianza/correlazione locale e, poichè la funzione di densità è gaussiana, si ottiene una caratterizzazione completa della relazione di dipendenza in quel determinato intorno.

Tutto questo, però, solo a livello teorico. Infatti, nella pratica c'è bisogno

di introdurre un parametro  $\theta(x) \doteq \{u_1(x), u_2(x), \sigma_1(x), \sigma_2(x), \rho(x)\}$  che rappresenti la popolazione, andandolo successivamente anche a stimare. Questo problema è stato trattato in maniera esaustiva da Hjort, Jones e Loader (1996) che hanno deciso di approssimare  $f(x)$  con una generica famiglia di funzioni di densità, tra cui anche la gaussiana.

## 2.2 La funzione parametro $\theta(x)$

Prima di proseguire si ritiene opportuno definire con più precisione  $\theta(x)$ . Il procedimento per giungere a ciò verrà suddiviso in due parti: il primo, andando a definire la funzione di verosimiglianza locale in un intorno definito da delle bande di confidenza  $b = (b_1, b_2)$  nella direzione di  $x = (x_1, x_2)$  e il secondo, dove si calolerà il limite per  $b \rightarrow 0$ .

Il primo passaggio, quindi, consiste nel definire la funzione di verosimiglianza locale, ma prima si ritiene necessario fare alcune precisazioni.

Hjort e Jones (1996) considerano una famiglia parametrica  $F_\theta$ , ma invece di cercare un singolo valore del parametro  $\theta_0$ , per il quale  $f_X(x) = f(x; \theta_0)$  affermano piuttosto che diversi membri di  $F_\theta$  possano approssimare  $f_X$  localmente in diverse parti del suo dominio. Come viene riportato da Tjøstheim e Hufthammer (2013), ci sarà bisogno di costruire un'approssimazione gaussiana che approssimi  $f(x)$  in un intorno di  $x$  in modo tale che la (2.1) sia valida in quel punto. Nel caso in cui  $X \sim N(u, \sigma^2)$  questo lo si ottiene banalmente utilizzando una gaussiana con  $u(x) = x$  e  $\sigma^2(x) = \sigma^2$  per tutti gli  $x$ .

- **La funzione di verosimiglianza locale**

Fatte queste premesse, la funzione di verosimiglianza locale è definita da:

$$L_n(\theta, x) = \operatorname{argmax}_{\theta \in \Theta} \frac{1}{nb} \sum_{i=1}^n K\left(\frac{X_i - x}{b}\right) \log f(X_i; \theta) - \int \frac{1}{b} K\left(\frac{y - x}{b}\right) f(y; \theta) dy, \quad (2.3)$$

in cui  $K$  rappresenta una funzione di densità simmetrica mentre il secondo termine nella funzione di verosimiglianza locale garantisce che la densità stimata da  $\hat{f}_X(x) = f(x; \hat{\theta}(x))$  converga correttamente alla funzione di densità vera  $f_X(x)$  quando la numerosità del campione  $n$  tende all'infinito e l'ampiezza della banda di confidenza  $b$  diminuisce fino allo zero.

D'altra parte nella pratica si cerca di stimare una funzione parametro  $\theta_0(x)$  per la quale  $f_X(x) = f(x; \theta_0(x))$  andando a massimizzare la funzione di verosimiglianza locale in ogni punto  $x$ .

Da questo procedimento si ottiene, quindi, la stima della funzione parametro che è data da:

$$\hat{\theta}(x) = \operatorname{argmax}_{\theta \in \Theta} L_n(\theta, x). \quad (2.2)$$

- **Il limite per  $b \rightarrow 0$**

In questa prima fase si considera il valore limite  $\theta(x) = \lim_{b \rightarrow 0} \theta_b(x)$  e si lascia  $b \rightarrow 0$ . Per quanto riguarda la stima di  $\theta(x) = \theta_b(x)$  la si ottiene

massimizzando la funzione di verosimiglianza locale ipotizzando che la banda di confidenza  $b$  abbia larghezza finita. A questo punto si riscrive la funzione in questo modo:

$$\begin{aligned}
L(X_1, \dots, X_n, \theta(x)) &= \\
&= n^{-1} \sum_{i=1}^n K_b(X_i - x) \ln \psi(X_i, \theta(x)) \\
&\quad - \int K_b(v - x) \psi(v, \theta(x)) dv.
\end{aligned} \tag{2.4}$$

L'ultimo termine è essenziale in quanto implica che  $\psi(x, \theta_b(x))$  non può allontanarsi molto da  $f(x)$ , dato che  $b \rightarrow 0$ .

Utilizzando la notazione,

$$u_j(\cdot, \theta) \doteq \frac{\partial}{\partial \theta_j} \ln \psi(\cdot, \theta), \tag{2.5}$$

per la legge dei grandi numeri, assumendo che  $E\{K_b(X_i - x) \ln \psi(X_i, \theta_b(x))\} < \infty$ , si può dire che:

$$\begin{aligned}
\frac{\partial L}{\partial \theta_j} &= n^{-1} \sum_{i=1}^n K_b(X_i - x) u_j(X_i, \theta_b(x)) \\
&\quad - \int K_b(v - x) u_j(v, \theta_b(x)) \psi(v, \theta_b(x)) dv \\
&\quad \rightarrow \int K_b(v - x) u_j(v, \theta_b(x)) \\
&\quad \quad \times [f(v) - \psi(v, \theta_b(x))] dv.
\end{aligned} \tag{2.6}$$

Ponendo uguale a zero l'espressione della prima riga di (2.6) si ottiene la

stima di massima verosimiglianza locale  $\hat{\theta}_b(x) = \hat{\theta}(x)$ .

Inoltre, la funzione di verosimiglianza locale dell'equazione (2.4) può essere sfruttata per dati dipendenti, in quanto, massimizzandola, si otterrà la stessa espressione di  $\hat{\theta}_b(x)$ , ma con i termini delle variabili  $X_t$ .

### 2.3 Teoria asintotica con $b$ fissato

Si può ora discutere le stime di verosimiglianza locale e dimostrarne la consistenza e la normalità asintotica. Si definisce la stima di verosimiglianza locale  $\hat{\theta}_b(x)$  come una soluzione di  $\frac{\partial L}{\partial \theta_j} = 0$ .

Fino ad ora si è utilizzato  $b \rightarrow 0$ , ma con  $b$  fissato la questione è differente, infatti Tjøstheim e Hufthammer (2013) hanno formulato un teorema partendo da una proposizione suggerita da Hjort e Jones (1996).

**Proposizione** (Hjort e Jones (1996)). *Sotto le ipotesi di esistenza e unicità di  $\theta_b(x)$  e considerando  $\hat{\theta}_{n,b}(x)$  come la stima di verosimiglianza locale di  $\theta_b(x)$ , si ha che:*

$$(nb_1b_2)^{1/2}[\hat{\theta}_{n,b}(x) - \theta_b(x)] \xrightarrow{d} N(0, J_b^{-1}M_b(J_b^{-1})^T), \quad (2.7)$$

dove

$$\begin{aligned} J_b &\doteq \int K_b(v-x)u(v, \theta_b(x))u^T(v, \theta_b(x))\psi(v, \theta_b(x))dv \\ &\quad - \int K_b(v-x)\nabla u(v, \theta_b(x))[f(v) - \psi(v, \theta_b(x))]dv, \end{aligned} \quad (2.8)$$

con  $u(x, \theta) = \nabla \log \psi(x, \theta)$  che rappresenta il vettore colonna delle funzioni punteggio, e dove

$$\begin{aligned}
 M_b &\doteq \lim_{n \rightarrow \infty} \text{Var}((nb_1b_2)^{1/2}V_n(\theta_b)) \\
 &= b_1b_2 \int K_b^2(v-x)u(v, \theta_b(x))u^T(v, \theta_b(x))f(v)dv \\
 &\quad - b_1b_2 \int K_b(v-x)u(v, \theta_b(x))f(v)dv \\
 &\quad \times \int K_b(v-x)u^T(v, \theta_b(x))f(v)dv
 \end{aligned} \tag{2.9}$$

con  $V_n(\theta) = \frac{\partial L(\theta)}{\partial \theta}$ .

**Teorema** (Tjøstheim e Hufthammer (2013)). Sia  $\{X_t\}$  una serie storica bivariata stazionaria con  $X_t$  che ha densità bivariata  $f$ . Assumiamo che (i)  $\{X_t\}$  sia geometricamente ergodica, (ii)  $f$  abbia un supporto su tutto  $\mathbb{R}^2$  e sia tale che  $E[K_b(X_t-x)\partial/\partial\theta_j \log \psi(X_t, \theta)]^\gamma < \infty$ ,  $j = 1, \dots, 5$  per qualche  $\gamma > 2$ , e (iii) esiste una banda di confidenza  $b_0$  tale che esiste un unico minimizzatore  $\theta_b(x)$  di:

$$q = \int K_b(v-x)[\psi(v, \theta(x)) - \log \psi(v, \theta(x))]f(v)dv, \tag{2.10}$$

per ogni  $0 < b < b_0$ . Allora per ogni  $\epsilon > 0$  è presente un evento  $A$  dipendente da  $x$  con  $P(A^c) < \epsilon$  tale che esiste una sequenza  $\theta_{n,b}(x)$  che minimizza  $Q$  in:

$$\begin{aligned}
 Q_n(X_1, \dots, X_n, \theta_b(x)) &= - \sum K_b(X_t - x) \log \psi(X_t, \theta(x)) \\
 &\quad + n \int K_b(v-x)\psi(v, \theta_b(x))dv,
 \end{aligned} \tag{2.11}$$

allo stesso modo  $\hat{\theta}_{n,b}(x)$  converge quasi certamente a  $\theta_b(x)$  su  $A$ . Inoltre, il risultato del limite centrale della Proposizione continua a valere, con l'integrale nella prima parte della (2.9) come termine principale per  $M_b$ .

La difficoltà principale nell'estendere il teorema alla situazione attuale è che le matrici  $J_b$  e  $M_b$  non sono più definite positive fintanto che  $b \rightarrow 0$ . Infatti, con  $J_b$  e  $M_b$  definite dagli integrali (2.8) e (2.9), ma con il parametro  $\theta_b(x)$  sostituito da  $\theta(x) = \theta$ , si ha  $J_b \rightarrow J \doteq u(x, \theta)u^T(x, \theta)\psi(x, \theta)$  e  $M_b \rightarrow M \doteq u(x, \theta)u^T(x, \theta)f(x) \int K^2(v)dv$  (e l'ultimo termine in (2.9) tende a diminuire dal momento che  $b \rightarrow 0$ ).

Le matrici  $J$  ed  $M$  si ottengono mantenendo solo il termine del primo ordine degli sviluppi di Taylor. D'altra parte, però, per essere sicuri di ottenere la positività è necessario includere termini di ordine superiore per cui gli sviluppi di Taylor per il termine  $b$  (generico) risultano quindi:

$$I_M = b_1 b_2 \int K_b^2(v - x)u(v, \theta(x))u^T(v, \theta(x))f(v)dv,$$

che corrispondono ad  $M_b$  in (2.9).

## 2.4 Considerazioni per la pratica

Dalle argomentazioni sopra sviluppate si può notare che è difficile ottenere delle espressioni utili di (2.8) e (2.9) da utilizzare poi nella pratica. Come soluzione in letteratura sono stati proposti due metodi alternativi. Il primo, che vale solo quando l'insieme di variabili  $\{X_i\}$  di osservazioni i.i.d. (indipendenti ed identicamente distribuite), è il bootstrap, che è stato usato da Berentsen *et al.* (2012) e Berentsen e Tjøstheim (2012) per valutare la significatività dei test di bontà di adattamento e di indipendenza delle copule sul quale, però, non ci si soffermerà.

La seconda alternativa, definita da Tjøstheim e Hufthammer (2013), è quella di approssimare  $M_b$  in (2.9) come segue:

$$\begin{aligned} \hat{M}_b = b_1 b_2 & \left[ \frac{1}{n} \sum_{i=1}^n K_b^2(X_i - x) u(X_i, \theta_b(x)) u^T(X_i, \theta_b(x)) \right. \\ & - \frac{1}{n} \sum_{i=1}^n K_b(X_i - x) u(X_i, \theta_b(x)) \\ & \left. \times \frac{1}{n} \sum_{i=1}^n K_b(X_i - x) u^T(X_i, \theta_b(x)) \right], \end{aligned} \tag{2.12}$$

dove  $\hat{M}_b \rightarrow M_b$  quasi certamente (per un fissato  $b$ ) per la legge forte dei grandi numeri.

Nella pratica si dovrà sostituire  $\theta_b(x)$  con la stima  $\hat{\theta}_{n,b}(x)$ .



## 2.5 Proprietà della *local Gaussian correlation*

Come descritto da Rényi (1959), alla *local Gaussian correlation* si possono attribuire le seguenti proprietà:

1.  $\rho(x)$  è una misura locale, ma è definita per qualsiasi coppia di variabili  $X_1$  e  $X_2$  sotto opportune condizioni di regolarità.
2. Poiché  $\rho(x)$  è una misura locale, si ha generalmente  $\rho(x_1, x_2) \neq \rho(x_2, x_1)$  ma  $\rho_{X_1, X_2}(x_1, x_2) = \rho_{X_2, X_1}(x_2, x_1)$ .
3. Dalle definizioni (1.1) e (1.2) segue che  $-1 \leq \rho(x) \leq 1$  e  $-1 \leq \hat{\rho}(x) \leq 1$ . La capacità di misurare in maniera qualitativa e quantitativa sia la dipendenza negativa che quella positiva rappresenta uno dei principali vantaggi della *local Gaussian correlation*.
4. L'indipendenza di  $X_1$  e  $X_2$  implica che  $\rho(x_1, x_2) \equiv 0$  ma questa condizione non è sufficiente per l'indipendenza. Si deve richiedere in aggiunta  $u_i(x_1, x_2) \equiv u_i(x_i)$  e  $\sigma_i(x_1, x_2) \equiv \sigma_i(x_i)$  per  $i = 1, 2$ , Tjøstheim e Hufthammer (2013). Probabilmente,  $\rho_Z(z)$  è una misura migliore per la dipendenza tra  $X_1$  e  $X_2$ . (Si noti che anche se  $Z_i \sim N(0, 1)$  non si ha necessariamente  $u_i(z) = 0$  e  $\sigma_i(z) = 1$  in un'approssimazione gaussiana locale bivariata. Per distribuzioni in cui questo è vero,  $X_1$  e  $X_2$  sono indipendenti se e solo se  $\rho(z) \equiv 0$ ).
5. Se  $X_1 = f(X_2)$  o  $X_2 = g(X_1)$ , allora il valore limite  $\rho(x_1, x_2)$  man mano che l'intorno si restringe al punto  $(x_1, x_2)$  è uguale a 1 o -1, a seconda che  $f$  o  $g$  abbiano una pendenza positiva o negativa su  $x_2$  o  $x_1$ . Lo stesso vale per una relazione di curva chiusa  $g(X_1) + f(X_2) = c$  per una costante  $c$ , allora  $\rho(x)$  è uguale a 1 o -1 lungo la curva, a seconda che la tangente al punto  $x$  abbia pendenza positiva o negativa (Tjøstheim e Hufthammer 2013).

- 
6. Come il  $\rho$  di Pearson, l'LGC  $\rho(x)$  dipende dalle marginali, ma sulla scala  $Z$ , in quanto  $\rho_Z(z)$  è indipendente dalle marginali.
7. Nel caso gaussiano si ha che  $\rho(x) \equiv \rho$  per costruzione.

Uno dei principali vantaggi dell'indice  $\rho$  di Pearson è la sua invarianza lineare di scala, ossia  $\rho_{\alpha_1+\beta_1X_1, \alpha_2+\beta_2X_2} = \rho_{X_1, X_2}$ . Esiste una corrispondente invarianza di scala per l'LGC, ma con la condizione che il punto  $(x_1, x_2)$  venga spostato nel punto  $(\alpha_1 + \beta_1x_1, \alpha_2 + \beta_2x_2)$ . Più in generale, è dimostrato da Tjøstheim e Hufthammer (2013) che per un vettore  $\alpha$ , una matrice  $A$ , una variabile stocastica  $Y = \alpha + AX$  e per un vettore di parametri locali  $\theta(y) = [u(y), \Sigma(y)]$  nel punto  $y = \alpha + Ax$ , si ha  $u(y) = \alpha + \Sigma(x)u(x)$  e  $\Sigma(y) = A\Sigma(x)A^T$ . Ne consegue che abbiamo invarianza di scala nel senso seguente  $\rho_{Y_1, Y_2}(y_1, y_2) = \rho_{\alpha_1+\beta_1X_1, \alpha_2+\beta_2X_2}(\alpha_1 + \beta_1x_1, \alpha_2 + \beta_2x_2) = \rho_{X_1, X_2}(x_1, x_2)$ .

In Tjøstheim e Hufthammer (2013) i risultati della trasformazione per  $\alpha + AX$  sono stati utilizzati per dimostrare una serie di proprietà di simmetria.

Si può distinguere tra 4 diverse tipologie di simmetria e per calcolarle si è assunto che  $u = E(X) = 0$ . Questo è stato fatto perchè altrimenti si rischierebbe di centrare la densità in  $u$  e fare affermazioni sulla simmetria di  $u$ .

(i) Simmetria radiale: Se  $f(x) = f(-x)$ , allora  $\Sigma(-x) = \Sigma(x)$ , da cui  $\rho(-x) = \rho(x)$  e  $u(-x) = -u(x)$ .

(ii) Simmetria di riflessione:  $f(-x_1, x_2) = f(x_1, x_2)$  e/o  $f(x_1, -x_2) =$

$f(x_1, x_2)$  implicano che:

$$\begin{aligned}
 \rho(-x_1, x_2) &= -\rho(x_1, x_2), \rho(x_1, -x_2) & (2.13) \\
 &= -\rho(x_1, x_2), u_1(-x_1, x_2) \\
 &= -u_1(x_1, x_2), u_2(-x_1, x_2) \\
 &= u_2(x_1, x_2), u_1(x_1, -x_2) \\
 &= u_1(x_1, x_2), u_2(x_1, -x_2) \\
 &= -u_2(x_1, x_2).
 \end{aligned}$$

- (iii) Simmetria di scambio: Se  $f(x_1, x_2) = f(x_2, x_1)$ , allora  $\Sigma(x_1, x_2) = \Sigma(x_2, x_1)$  e quindi  $\rho(x_1, x_2) = \rho(x_2, x_1)$ .
- (iv) Simmetria di rotazione: Per una funzione  $\gamma$ ,  $f(x) = \gamma(|x|)$ . Se  $f$  è una densità sferica, allora soddisfa tutti i requisiti di simmetria menzionati sopra. In questo caso si può dimostrare che lungo le rette  $x_1 = x_2$  e  $x_1 = -x_2$ ,  $\rho^2(x)$  assume il suo massimo.

La concreta espressione di tali simmetrie si manifesta compiutamente nei grafici, i quali permettono di concretizzare visivamente le peculiarità di ciascuna. Nella sezione successiva infatti si potrà osservare graficamente alcune di queste simmetrie, in particolare la simmetria radiale e la simmetria di riflessione.

### 3 Applicazione della LGC

In questa sezione verranno esposte alcune simulazioni nelle quali verrà applicata la LGC (*local Gaussian correlation*) così da poter osservare in modo più esplicito l'efficacia e la potenzialità di questa misura di dipendenza locale.

Nello specifico verranno proposti alcuni modelli non lineari per introdurre la sua funzionalità nel caso concreto per poi concludere con l'applicazione in serie storiche non lineari attraverso un modello GARCH(1,1).

Per poter effettuare questa analisi sono state utilizzate le librerie di R: "localgauss", "lg" ed "fGARCH". Le prime due sono molto simili in quanto stimano e permettono di visualizzare la *local Gaussian correlation* sia graficamente che tramite dei test di dipendenza. La terza libreria invece è stata utilizzata per poter simulare serie storiche del tipo GARCH(1,1) attraverso la funzione "garchSim".

### 3.1 Distribuzione Normale bivariata

Il primo caso di modello lineare si è ottenuto simulando una distribuzione gaussiana bivariata di  $n = 1000$  osservazioni con  $\rho = +0.5$ :

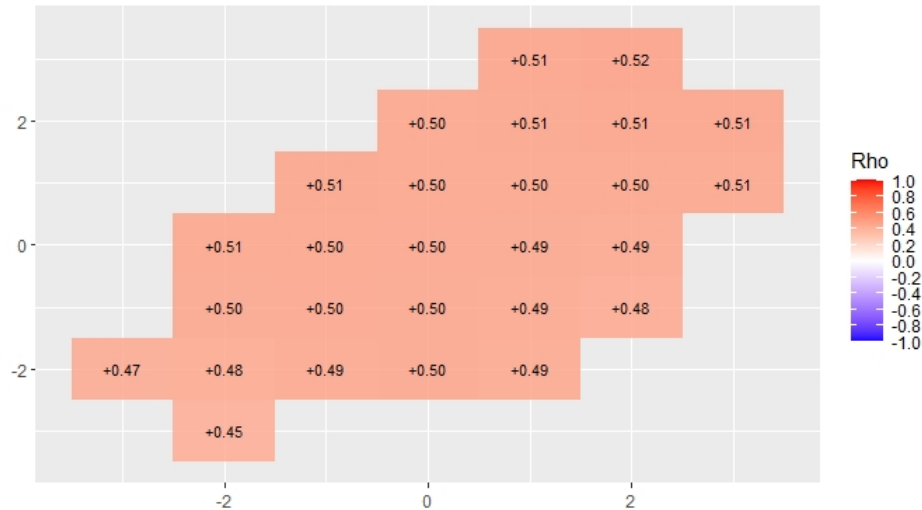


Figura 1: mappa di dipendenza di una gaussiana bivariata,  $\rho = +0.5$

Tale distribuzione è stata rappresentata graficamente tramite una mappa di dipendenza, strumento genericamente utilizzato per la visualizzazione grafica della *local Gaussian correlation*. Si può notare, infatti, come utilizzando  $\rho = +0.5$ , sia evidente la correlazione positiva tra le componenti e i diversi valori dell'indice  $\rho$  di Pearson calcolato localmente.

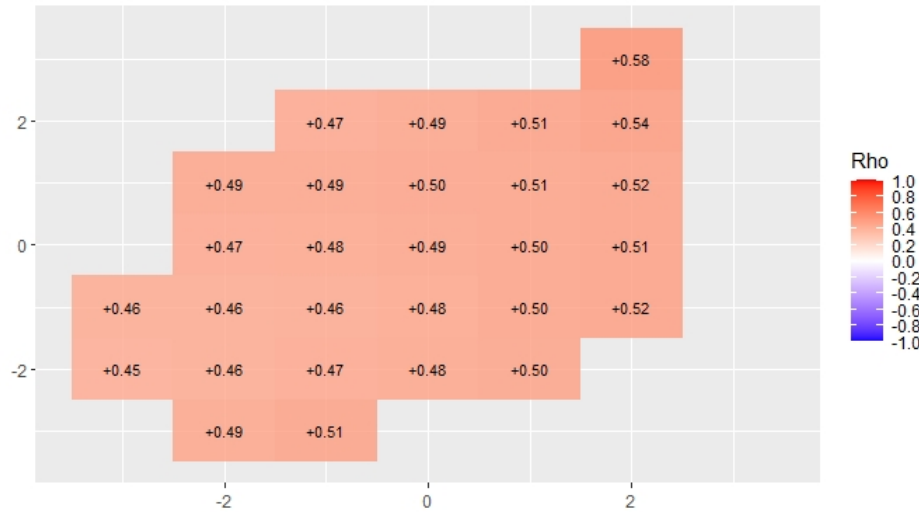


Figura 2: mappa di dipendenza di una gaussiana bivariata,  $\rho = +0.5$ , dati standardizzati

E' possibile osservare che non esiste molta differenza tra i dati originali e quelli standardizzati. Inoltre, i valori ottenuti sono tutti molto vicini al valore  $\rho = +0.5$  utilizzato in input. Si è partiti dal caso più semplice per dimostrare la validità di questa misura di dipendenza e di come effettivamente vengano mostrati i dati graficamente.

### 3.2 Distribuzione t-Student bivariata

Nel secondo caso di modello non lineare si dimostra, partendo da una simulazione di una distribuzione t-Student bivariata con 4 g.d.l. (gradi di libertà) con  $\rho = 0$  e  $n = 1000$ , come si comporta la LGC:



Figura 3: distribuzione t-Student con 4 gradi di libertà

Anche in questo caso si è utilizzato come strumento grafico la mappa di dipendenza, la quale ha evidenziato i valori della LGC di una distribuzione t-Student con 4 gradi di libertà e  $\rho = 0$  (i valori, infatti, sono tutti concentrati vicino allo 0). Come per il modello precedente ora si standardizzeranno i dati per provare a trarre ulteriori conclusioni.



Figura 4: distribuzione t-Student con 4 gradi di libertà, dati standardizzati

Si può notare come non ci sia molta differenza tra dati originali e standardizzati. D'altra parte, tuttavia, è possibile osservare che nonostante l'indice di correlazione globale sia nullo, a livello locale la dipendenza sia, seppur lievemente, maggiore di 0. Inoltre, questo caso è molto interessante perchè è presente una delle proprietà menzionate nella sezione precedente ovvero la simmetria radiale e una simmetria di riflessione dispari; è possibile infatti verificare che ci sia dipendenza positiva nel primo e terzo quadrante e dipendenza negativa nel secondo e quarto quadrante.



### 3.3 Modello non lineare $Y = X^2$

Come terzo modello si è deciso di discutere, come in una delle sezioni precedenti, il caso parabolico:  $Y = X^2 + \epsilon$ .

Per illustrare questo modello è stata utilizzata una funzione della libreria "localgauss", ovvero "localgauss.indtest" che consente di eseguire un test di dipendenza locale passandogli come parametri di input dei punti dove effettuare il test.

Tramite la funzione "localgauss" si è ottenuto il seguente grafico (n=1000):

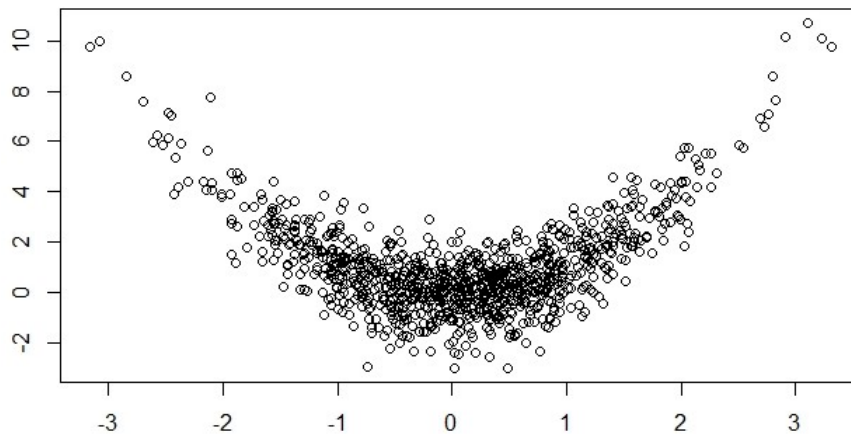


Figura 5: modello di regressione non-lineare  $Y = X^2 + \epsilon$

Eseguendo il test nei punti:  $(-1,1)$ ,  $(0,0)$  e  $(1,1)$  si è ottenuto il seguente output:

```
[1 ] -1 0 1
```

La *local Gaussian correlation* è significativamente negativa in  $(-1,1)$ , significativamente positiva in  $(1,1)$  e nulla in  $(0,0)$  per un livello di significatività del test del 10% (valore di default).

E' un risultato molto importante in quanto immaginando di dividere il grafico in tre zone si può effettivamente osservare una dipendenza negativa nella prima parte (osservazioni da  $-3$  a  $-1$  dell'asse delle  $x$ ), dipendenza nulla nella zona centrale (all'incirca le osservazioni da  $-1$  a  $+1$  considerando l'asse delle  $x$ ) e infine una dipendenza positiva nell'ultima parte (osservazioni da  $+1$  a  $+3$  sempre guardando l'asse delle  $x$ ). Si può quindi concludere che questa misura di dipendenza non ha solo potenziale a livello grafico, ma permette anche di costruire test in base a ciò che si desidera analizzare.





## Conclusioni

L'estrema utilità nell'adoperare la distribuzione gaussiana si manifesta nelle sue proprietà che altre famiglie di distribuzioni parametriche non hanno. Una di queste è, ad esempio, che per una distribuzione gaussiana multivariata la distribuzione condizionata di variabili aleatorie  $Y_i$  date  $X_i$  per  $i = 1, \dots, n$  è ancora gaussiana.

L'idea che poi ha portato gli studiosi a creare questa nuova misura di dipendenza era fondata sul pensiero che si potesse estendere localmente le proprietà della gaussiana anche alle altre distribuzioni.

Inoltre, uno dei motivi più importanti per il quale è utile introdurre una misura come la *local Gaussian correlation* (LGC) è quello di poter gestire serie storiche, come ad esempio nel campo della finanza i rendimenti finanziari.

Sotto un altro aspetto, dalle simulazioni è possibile concludere che questa nuova misura di dipendenza è a tutti gli effetti uno strumento molto buono nei casi di criticità del  $\rho$  di Pearson e risulta determinante oltre che per analisi statistiche generiche anche nel campo delle serie storiche, tant'è che modelli come quelli analizzati sopra avrebbero per certo portato a conclusioni affrettate e fuorvianti se ci si fosse basati solamente sull'indice  $\rho$ . Il modello GARCH(1,1), infatti, mostra nello specifico come si comporti la LGC nei casi di serie storiche non lineari e alcune delle sue proprietà come la simmetria radiale.

Come si è potuto osservare è possibile sfruttare oltre alla rappresentazione grafica (tramite l'utilizzo di mappe di confidenza) anche i test di dipendenza locale qualora si volesse verificare la correlazione in determinati punti del modello. Questa misura di dipendenza locale gode di molte funzioni, ma per una trattazione completa di tutte le sue funzionalità si consiglia di leggere quanto pubblicato da Berentsen *et al.* (2014) e Otneim (2021).

## Bibliografia

Berentsen G. D.; Kleppe T.; Tjøstheim D. (2014). Introducing localgauss, an R-package for estimating and visualizing local Gaussian correlation. *Journal of Statistic Software*, **56**, 1-18.

Berentsen G.; Støve B.; Tjøstheim D.; Nordbø T. (2012). Recognizing and visualizing copulas: an approach using local Gaussian approximation. *Insurance: Mathematics and Economics*, **57**, 90-103.

Berentsen G.; Tjøstheim D. (2012). Recognizing and visualizing departures from independence in bivariate data using local Gaussian correlation. *Statistics and Computing*, **24**(5), 785-801.

Hjort N. L.; Jones M. C. (1996). Locally parametric nonparametric density estimation. *The Annals of Statistics*, **24**, 1619–1647.

Kendall M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**(1/2), 81-93.

Lacal V.; Tjøstheim D. (2017). Local Gaussian autocorrelation and tests for serial independence. *Journal of Time Series Analysis*, **38**(1), 51–71.

Nelsen R. B. (2007). *An introduction to copulas*. Springer science & business media.

Otneim H. (2021). lg: An R package for Local Gaussian Approximations. *R. J.*, **13**(2), 15.

Pearson K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philos. Trans. R. Soc. Lond.*, **187**, 253–318.

Posedel P. (2005). Properties and estimation of GARCH (1, 1) model. *Advances in Methodology and Statistics*, **2**(2), 243-257.

Rényi A. (1959). On measures of dependence. *Acta mathematica hungarica*, **10**(3-4), 441-451.

Spearman C. (1961). The proof and measurement of association between two things.

Støve B.; Tjøstheim D.; Hufthammer K. O. (2014). Using local Gaussian correlation in a nonlinear re-examination of financial contagion. *Journal of Empirical Finance*, **25**, 62-82.

Székely G. J.; Rizzo M. L. (2012). On the uniqueness of distance covariance. *Statistics & Probability Letters*, **82**(12), 2278-2282.

Tjøstheim D.; Otneim H.; Støve B. (2022). Statistical dependence: Beyond pearson's  $p$ . *Statistical Science*, **37**(1), 90-109.

Tjøstheim D.; Otneim H.; Støve B. (2022). Supplement to “Statistical Dependence: Beyond Pearson's  $p$ .”

Tjøstheim D.; Hufthammer K. O. (2013). Local Gaussian correlation: A new measure of dependence. *Journal of Econometrics*, **172**, 33-48.