

CONSIDERAZIONI SUL TEST Z PER GLI INDICI DI MORAN

Giuliano Visini

Riassunto - In questa nota l'autore considera gli indici di Moran, ne esamina i primi due momenti e analizza i comportamenti dei relativi tests z dandone un'efficace rappresentazione parametrica.

Summary - Moran index is the objet of this work, the author takes into examination the first two moments and analyzes the line of conduct of the relative z tests giving of them a clarifying parametrical representation.

1. INTRODUZIONE

L'analisi spaziale si scontra di continuo tra tre esigenze contrastanti che rispondono a tre angolazioni diverse, eppur complementari, per affrontare lo stesso problema: una modalità territoriale, articolata per unità di rilevamento statistico, di quale processo soggiacente è la realizzazione? Come valutarne la eventuale casualità? Come rappresentarne la variabilità nello spazio, in modo da poterne rilevare in maniera corretta, le relazioni di vicinato tra elementi contigui statistico-territoriali?

La letteratura offre una pluralità di risposte parziali, ed anche questo contributo si muove ed intende procedere in uno specifico sentiero disciplinare, ma nella piena consapevolezza di una struttura concettuale di fondo pluridisciplinare, contesto non solo scientifico, ma anche culturale.

Il tema centrale da esplicitare, in seguito nei termini propri del linguaggio formale, è quello delle tendenze centrali in un territorio privo di gerarchizzazioni aprioristiche dei luoghi: ogni unità di rilevamento è considerato "luogo centrale", dal quale procedere all'esplorazione dello spazio circostante al fine di individuare l'incidenza della "posizione geografica" nelle realizzazioni di un dato processo.

Il fatto che esso si debba necessariamente assumere come in una logica probabilistica - come tale negazione dell'autocorrelazione spaziale - costringe da un lato, a considerare le relazioni di contiguità, e, da un altro, le condizioni che giustificano una distribuzione normale delle stesse in relazione a variabili aleatorie di modalità qualitative.

2. GLI INDICI DI MORAN: MEDIA E VARIANZA

Posto $N_n = \{1, 2, \dots, n\}$ ed $n \geq 2$, sia $\{a_i \mid i \in N_n\}$ una suddivisione territoriale finita di un territorio e sia il vettore aleatorio

$$\underline{K} = (K_1, K_2, \dots, K_n)$$

costituito da una n-pla di variabili aleatorie (v.a.) di modalità qualitative (ad es. variabili policrome) ciascuna osservabile sulla corrispondente unità territoriale $a_i (i \in N_n)$.

Se le distribuzioni delle v.a. $K_i (i \in N_n)$ sono tra loro mutuamente indipendenti e identicamente distribuite e se indichiamo con:

- a) $K = \{B, R, V, \dots, N\}$,
(ove ad es.: B=Bianco, R=rosso, V=verde, ..., N=nero), lo spazio campionario di ogni singola v.a. K_i ;
- b) p_r la probabilità della modalità r-esima della v.a. K_i osservabile sull'unità territoriale i-esima;
- c) K^n , l'universo dei campioni relativo alla situazione osservabile sulle unità territoriali $a_i (i \in N_n)$;
- il modello relativo resta definito da:

$$\{\underline{K}, K^n, k \geq 2, \prod_{i=1}^n p_{ir}, \sum_{r=1}^k p_{ir} = 1\}.$$

Introdotta poi un criterio di contiguità spaziale tra le unità territoriali, resta definito un operatore di associazione L,

$$L: \{a_i \mid i \in N_n\} \rightarrow \{a_j \mid i \in N_n, j \neq i\},$$

che esprime una corrispondenza generalmente multivoca e non ovunque definita, tale che:

$$L(a_i) \subseteq \{a_j \mid j \in N_n, j \neq i\}$$

nel caso valga l'implicazione:

$$a_j \in L(a_i) \implies a_i \in L(a_j) \quad (j \in N_n),$$

l'operatore di associazione lo diremo simmetrico.

La rappresentazione di L nell'insieme prodotto cartesiano $\{a_i \mid i \in N_n\} \times \{a_j \mid j \in N_n\}$ come $\{a_p, a_q \mid a_i \in L(a_p), p \in N_n\}$ induce sulla matrice aleatoria

$$\underline{K} \cdot \underline{K}' = (K_{ij})_{i \in N_n, j \in N_n}$$

una limitazione ristretta alle sole variabili aleatorie K_{pq} i cui indici p e q siano rispettivamente quelli corrispondenti alle unità territoriali poste nella relazione $a_i \in L(a_p)$ per ogni $p \in N_n$. Per tale limitazione si considera assegnata la distribuzione di $\underline{K} \cdot \underline{K}'$.

In tal modo per ciascun operatore di associazione L è possibile considerare la tavola aleatoria di variabili aleatorie

	B	R	V	.	.	N
B	N_{BB}	N_{BR}	N_{BV}	.	.	N_{BN}
R	N_{RB}	N_{RR}	N_{RV}	.	.	N_{RN}
V	N_{VB}	N_{VR}	N_{VV}	.	.	N_{VN}
.
.
N	N_{NB}	N_{NR}	N_{NV}	.	.	N_{NN}

(Tav. 2.1)

le cui determinazioni sono costituite dalle possibili frequenze assolate associate alle limitazioni predette dei risultati individuati dallo spazio

$$IK^2 = \{BB, BR, BV, \dots, BN, \dots, NB, NR, NV, \dots, NN\}$$

e le cui distribuzioni di probabilità si ottengono per "riduzione" da quelle di $\underline{K} \cdot \underline{K}'$.

Gli elementi della tav. 2.1, noti come indici di contiguità di Moran, costituiscono una misura dell'influenza che modalità disposte su zone contigue

esercitano vicendevolmente.

Ipotizzata $p_r = p_i (i \in N_n)$, con probabilità a priori che si abbia a realizzare l'evento r -esimo su una unità territoriale, di tali indici sono noti i momenti.

Infatti, considerate le variabili di comodo:

$$S_0 = \sum_{i=1}^n L_i = \sum_{j=1}^n L_j$$

$$S_1 = \left(\sum_{i=1}^n \sum_{j=1}^n (\delta_{ij} + \delta_{ji})^2 \right) / 2$$

$$S_2 = \left(\sum_{i=1}^n (L_i + L_i)^2 \right) / 2$$

ove

$$\delta_{ij} = 1 \text{ se } a_j \in L(a_i)$$

$$\delta_{ij} = 0 \text{ se } a_j \notin L(a_i)$$

$$L_i = \sum_{j=1}^n \delta_{ij}$$

$$L_j = \sum_{i=1}^n \delta_{ij}$$

le medie e le varianze restano individuate dalle:

I) nel caso in cui due unità territoriali contigue abbiano medesimo colore, ad es. RR:

$$E(N_{RR}) = (S_0 p_R^2) / 2$$

$$\text{Var}(N_{RR}) = (S_1 p_R^2 + (S_2 - 2S_1) p_R^3 + (S_1 - S_2) p_R^4) / 4$$

II) nel caso in cui due unità territoriali contigue abbiano diversità assegnata di colore ad es. RN:

$$E(N_{RN}) = S_0 p_R p_N$$

$$\text{Var}(N_{RN}) = (2S_1 p_R p_N + (S_2 - 2S_1) p_R p_N (p_R + p_N) + 4(S_1 - S_2) p_R^2 p_N^2) / 4$$

III) nel caso in cui due unità territoriali contigue abbiano generici colori diversi:

$$E(N_{BR\dots N}) = S_0 \sum_{r=1}^{k-1} \sum_{s=r+1}^k p_r p_s$$

$$\text{Var}(N_{BR\dots N}) = S_0 \sum_{r=1}^{k-1} \sum_{s=r+1}^k p_r p_s + (2S_1 - 5S_2) \sum_{r=1}^{k-2} \sum_{s=r+1}^{k-1} \sum_{t=s+1}^k p_r p_s p_t +$$

$$+4(S_1 - S_2) \left(\sum_{r=1}^{k-1} \sum_{s=r+1}^k p_r p_s - 2 \sum_{r=1}^{k-3} \sum_{s=r+1}^{k-2} \sum_{t=s+1}^{k-1} \sum_{u=t+1}^k p_r p_s p_t p_u \right) / 4$$

Infine, se l'operatore di associazione spaziale è simmetrico, introdotte le variabili di comodo

$$A = \sum_{i=1}^n L_i / 2 = \sum_{j=1}^n L_j / 2$$

$$D = \sum_{i=1}^n \binom{L_i}{2} = \sum_{j=1}^n \binom{L_j}{2}$$

legate alle precedenti dalle:

$$A = S_0 / 2 = S_1 / 4$$

$$D = (S_2 - 2S_1) / 8 = (S_2 - 4S_0) / 8$$

le medie e le varianze si trasformano nelle:

$$I') \quad E(N_{RR}) = Ap_R^2$$

$$\text{Var}(N_{RR}) = Ap_R^2 + 2Dp_R^3 - (A + 2D)p_R^4$$

$$II') \quad E(N_{RN}) = 2Ap_R p_N$$

$$\text{Var}(N_{RN}) = Ap_R p_N + 2Dp_R p_N (p_R + p_N) - 4(A + 2D)p_R^2 p_N^2$$

$$III') \quad E(N_{BR...N}) = 2A \sum_{r=1}^{k-1} \sum_{s=r+1}^k p_r p_s$$

$$\text{Var}(N_{BR...N}) = 2(A + D) \sum_{r=1}^{k-1} \sum_{s=r+1}^k p_r p_s - 2(4A + 5D) \sum_{r=1}^{k-3} \sum_{s=r+1}^{k-2} \sum_{t=s+1}^{k-1} p_r p_s p_t +$$

$$- 4(A + 2D) \left(\sum_{r=1}^{k-1} \sum_{s=r+1}^k p_r^2 p_s^2 - 2 \sum_{r=1}^{k-3} \sum_{s=r+1}^{k-2} \sum_{t=s+1}^{k-1} \sum_{u=t+1}^k p_r p_s p_t p_u \right)$$

3. LA VARIANZA NELLE SUE COMPONENTI

Come è possibile osservare i valori medi e le varianze degli indici dipendono, oltre che dalle probabilità p_i , che si realizzi l'evento r -esimo sull'unità territoriale i -esima, anche dalle variabili A e D , a loro volta, dipendenti dalla numerosità n delle unità territoriali.

In particolare, i valori medi sono influenzati solamente dalla variabile A , mentre le varianze lo sono generalmente da entrambi.

Ora i valori assumibili da A sono propri anche di altre possibili ipotesi di operatori di associazione, mentre i valori assumibili da D diversificano talune di queste ipotesi, che, essendo scelta del ricercatore, sono soggettive.

Nell'esempio, che si riporta, riferito al caso in cui $n=5$, accanto alle colonne dei valori assunti da A e D è posta la colonna F delle numerosità delle possibili ipotesi di operatori di associazione simmetrici corrispondenti.

A	D	F	A	D	F	A	D	F
1	0	10	5	5	12	7	13	30
2	0	15	5	6	120	7	14	60
2	1	30	5	7	60	7	15	30
3	1	30	5	8	60	8	18	15
3	2	60	6	9	70	8	19	30
3	3	30	6	10	75	9	24	10
4	3	70	6	11	60	10	30	1
4	4	75	6	12	5			
4	5	60						
4	6	5						

D'altra parte le varianze espresse dalle I') sono monotone crescenti rispetto a D mentre quelle espresse dalle II') sono monotone non decrescenti rispetto a D, come anche quelle espresse dalle III'), almeno nei casi per i quali è $k=3, 4$.

È allora di interesse ricercare una (o più d'una) relazione che consenta di stabilire i valori assumibili da D fissato A, essendo D funzione multivoca di A.

Il sistema 3.1, a variabili intere, formalizza l'interesse della ricerca e, come tipica formulazione di un problema di programmazione a variabili intere ed a soluzione ottima intera massimizzante o minimizzante, consente di determinare la serie marginale dei valori massimi indicata D_{\max} o quella dei valori minimi indicata D_{\min} di D.

$$(3.1) \begin{cases} \sum_{i=0}^{n-1} \left(\begin{matrix} ix_i + \sum_{j=i+1}^{n-1} x_j \\ 2 \end{matrix} \right) - D = 0 & ((\max D) \text{ o } (\min D)) \\ \sum_{i=0}^{n-1} ix_i = A \\ 1 \leq A \leq \binom{n}{2} \\ 0 \leq x_i \leq 1 \end{cases}$$

La serie marginale dei valori $D_{\min} = D_{\min}(A)$, soluzione ottima minimizzante

il sistema (3.1), è definibile attraverso la:

$$D_{\min}(A) = (2A \pmod n) (\text{int}(2A/n) + 1)^2 + (n - 2A \pmod n) (\text{int}(2A/n))^2 / 2 - A.$$

Per la serie marginale dei valori massimi $D_{\max} = D_{\max}(A)$ si individuano immediatamente le corrispondenze:

$$\begin{aligned} 1 \leq A \leq n-1 & \quad D_{\max}(A) = \binom{A}{2} \\ A = n & \quad D_{\max}(A) = \binom{n-1}{2} + 2 \\ A = 2n-3 & \quad D_{\max}(A) = n(n-2) \\ A = h + \binom{n-1}{2} \quad (0 \leq h \leq n-1) & \quad D_{\max}(A) = \binom{h}{2} + (n-h-1) \binom{n-2}{2} + h \binom{n-1}{2} \end{aligned}$$

Tuttavia, per quest'ultima si constatano delle regolarità con possibilità di algoritmare negli intervalli di tipo I e J sotto specificati:

$$\begin{aligned} I_1 &= [1, n-1] \\ I_{r_1} &= \left[1 + \sum_{i=1}^{r_1-1} (n-i), \sum_{i=1}^{r_1} (n-i) \right] \quad (r_1 \in N_{n-1}) \\ I_{n-1} &= \left[\binom{n}{2}, \binom{n}{2} \right] \\ J_1 &= \left[\binom{n-1}{2} + 1, \binom{n}{2} \right] \\ J_{r_2} &= \left[\binom{n-r_2}{2} + 1, \binom{n-r_2+1}{2} \right] \quad (r_2 \in N_{n-1}) \\ J_{n-1} &= [1, 1]. \end{aligned}$$

Infatti, se si considera il valore assunto da A, individuati gli indici r_1 e r_2 dei rispettivi intervalli di appartenenza di tipo I e J e gli indici h_1 e h_2 di rispettiva posizione all'interno di questi intervalli per cui:

$$A = h_1 + \sum_{i=1}^{r_1-1} (n-i) = h_2 + \binom{n-r_2}{2}$$

si ha in corrispondenza:

$$\begin{aligned} D_{\max}(A) = \max & \left[2(r_1-1)h_1 + \binom{h_1}{2} + \sum_{s=1}^{r_1-1} \left(2(s-1)(n-s) + \binom{n-s}{2} \right); \right. \\ & \left. \binom{n-r_2}{2} h_2 + \binom{h_2}{2} + (n-r_2-h_2) \binom{n-r_2-1}{2} \right] \end{aligned}$$

ove $h_1 \in N_{n-r_1}$ e $h_2 \in N_{n-r_2}$.

Supposto poi che la probabilità p_r che si realizzi l'evento r -esimo su una unità territoriale sia costante per ogni evento di K , ossia $p_r=1/k$, le relazioni precedenti I', II' e III' si riducono alle:

$$\begin{aligned} \text{I''}) \quad & E(N_{RR}) = A / k^2 \\ & \text{Var}(N_{RR}) = (Ak^2 + 2Dk - (A + 2D)) / k^4 \\ \text{II''}) \quad & E(N_{RN}) = 2A / k^2 \\ & \text{Var}(N_{RN}) = 2(Ak^2 + 2Dk - 2(A + 2D)) / k^4 \\ \text{III''}) \quad & E(N_{BR...N}) = A(k-1) / k \\ & \text{Var}(N_{BR...N}) = A(k-1) / k^2. \end{aligned}$$

Come è osservabile, nella III'' la varianza è indipendente da D ed ancora nel particolare caso di $k=2$ perde di significato la III'' e le I'' e le II'' diventano:

$$\begin{aligned} \text{I''')} \quad & E(N_{RR}) = A / 4 \\ & \text{Var}(N_{RR}) = (3A + 2D) / 16 \\ \text{II''')} \quad & E(N_{RN}) = A / 2 \\ & \text{Var}(N_{RN}) = A / 4 \end{aligned}$$

4. LE DISTRIBUZIONI CAMPIONARIE DEGLI INDICI DI MORAN

In condizioni di indipendenza delle osservazioni spaziali e di equiprobabilità delle modalità rilevabili, l'ipotesi statistica che si vuole provare attraverso il test z è la presenza o meno di autocorrelazione spaziale, ossia se l'osservazione di una modalità di una variabile qualitativa in una unità territoriale influenzi quelle in unità contigue e viceversa ne resti da queste influenzata.

4.1 CASO DELLA VARIABILE $N_{BR...N}$

Essendo $K=\{B, R, \dots, N\}$ e $|K|=k$, indicata con r la genera modalità sia nelle ipotesi formulate $p_r=1/k$ ($r \in K$). Con riferimento alla variabile casuale $N_{BR...N}$, indicata con x una sua determinazione, è

$$z_{BR...N} = z_k = (xk - A(k-1))\sqrt{A(k-1)}$$

Questa è funzione reale delle variabili intere x , A , k , ed in quanto definita per

$$\left\{ \begin{array}{l} k \geq 2 \\ 1 \leq A \leq \binom{n}{2} \\ 0 \leq x \leq \binom{n}{2} - \sum_{h=1}^{n-k} \text{int}((n-h)/k) \\ x \leq A \end{array} \right.$$

lo è anche di n con l'aggiunta della dipendenza delle determinazioni x da n,k e dagli operatori di associazione L.

Per ciascun n, z_k è rappresentabile in uno spazio R¹ da una nuvola di punti non minore di

$$\frac{1}{2} \left(\binom{n}{2}^2 + 3 \binom{n}{2} \right) - \left(\sum_{h=1}^{n-k} \text{int}((n-h)/k) \right) - \sum_{k=1}^{\text{int}(n/2)} \binom{n-2h+1}{2}$$

Essa è limitata inferiormente da

$$z_{k|x=0} = -\sqrt{A(k-1)}$$

limitata superiormente da

$$z_{k|x=A} = \sqrt{A/(k-1)} \text{ per } 1 \leq A \leq \binom{n}{2} - \sum_{h=1}^{n-k} \text{int}((n-h)/k)$$

e maggiorata da

$$z_{k|x=A} = \sqrt{A/(k-1)} \text{ per } \binom{n}{2} - \sum_{h=1}^{n-k} \text{int}((n-h)/k) < A \leq \binom{n}{2}$$

essendo per queste $\lim_{k \rightarrow +\infty} z_{k|x=0} = -\infty$ e $\lim_{k \rightarrow +\infty} z_{k|x=A} = 0$.

Se poi la z_k si pensa definita da variabili reali, essa ha differenziale totale

$$dz_k = \left(k / \sqrt{A(k-1)} \right) dx + \left((A - k(A+x)) / (2A\sqrt{A(k-1)}) \right) dA + \left(((x-A)k - 2x + A) / (2(k-1)\sqrt{A(k-1)}) \right) dk$$

con componenti $\partial z_i / \partial x$ sempre positiva, $\partial z_i / \partial A$ sempre negativa e $\partial z_i / \partial k$ sempre negativa ed ha per differenziale secondo una forma quadratica indefinita.

Impossibilitati a rappresentare graficamente la z_i che nella sua generalità è dipendente da x , A , k e da n , se ne mostra una rappresentazione parametrica.

Considerati come parametri le variabili n e k , le superfici di livello σ_n sono configurabili in una serie di spazi R_{nk}^3 doppiamente indicati e se, inoltre, si ipotizza, in ciascuno di essi, parametro, la variabile A , si ha di tali R_{nk}^3 , una rappresentazione planare con sistema di riferimento xz_i .

Stabilito poi un livello di significatività α , (sia ad es. $\alpha=0.05$), e tracciate nei piani xz_i le rette $z_i = \pm 1.96$ restano, per ciascuna rappresentazione parametrica planare, individuate tre zone una di accettazione caratterizzata dai valori di z_i tali che $-1.96 < z_i < +1.96$ e due di rifiuto l'una per valori di $z_i \geq +1.96$ e l'altra per valori di $z_i \leq -1.96$ per le quali si parla rispettivamente di autocorrelazione spaziale positiva e negativa.

A scopo illustrativo si presentano rappresentazioni parametriche planari di z_i per diversi valori di n e di k (fig. 4.1.n.k) che rendono evidenti alcune interrelazioni significative di tipo matematico-statistico tra le variabili x , A , k , n .

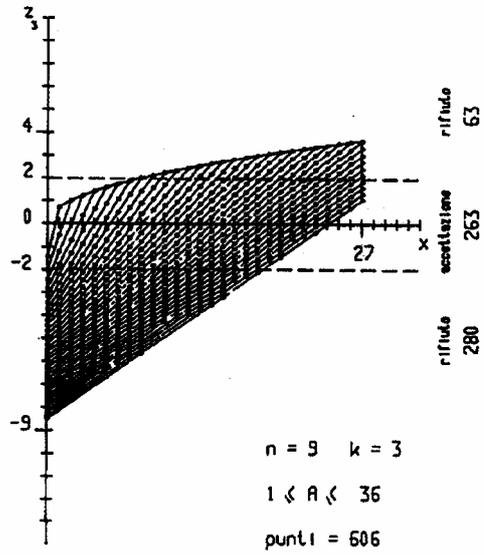
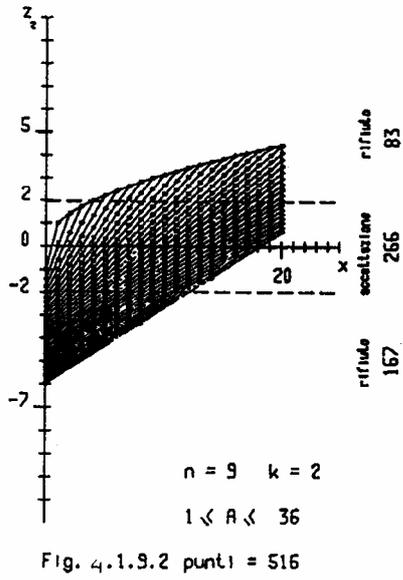


Fig. 4.1.9.3

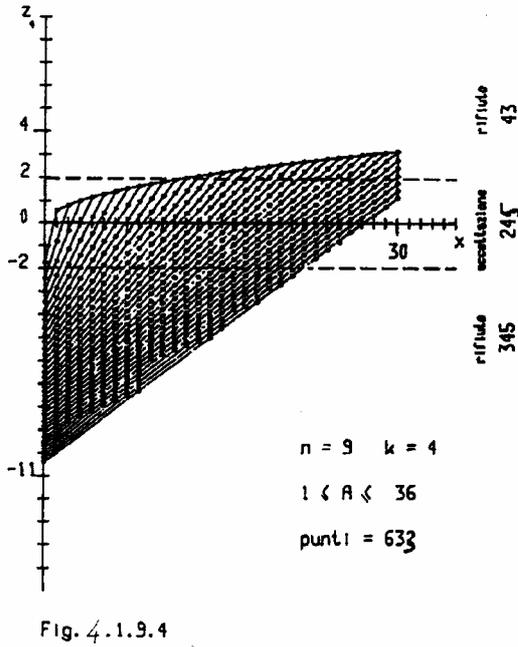


Fig. 4.1.9.4

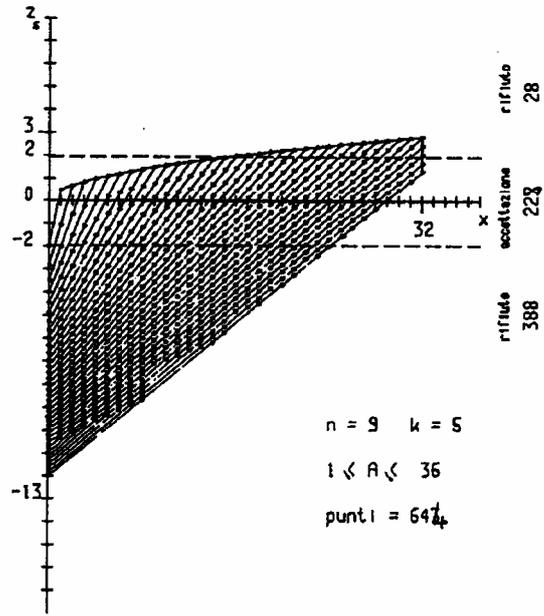


Fig. 4.1.9.5

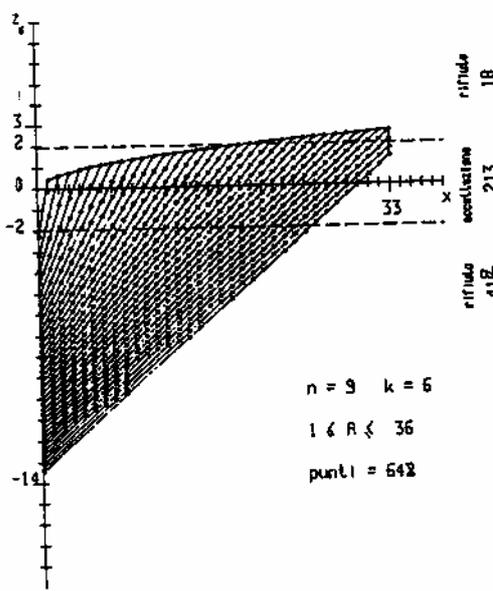


Fig. 4.1.9.6

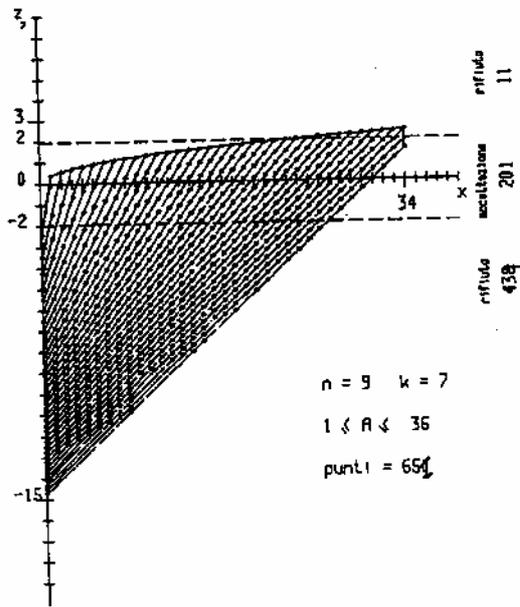


Fig. 4.1.9.7

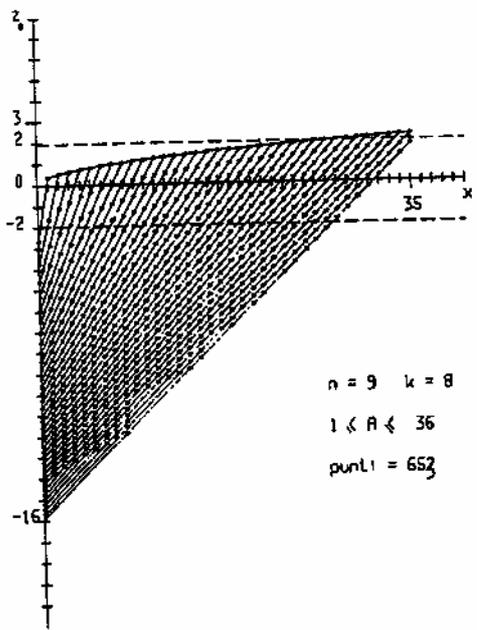


Fig. 4.1.9.8

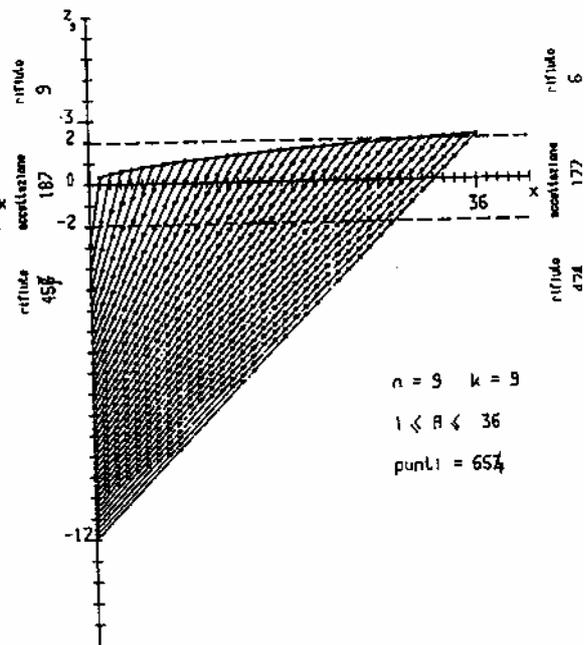


Fig. 4.1.9.9

Si riportano nel seguito alcune considerazioni di tipo matematico:

- nelle rappresentazioni parametriche planari di z_k i punti corrispondenti ad $A=costante$ sono allineati;
- per ogni fissato k , se $n > n'$ i punti della fig. 4.1. $n'.k$ sono anche punti della fig. 4.1. $n.k$;
- la simmetria dei valori è limitata al solo caso $k=2$ e per valori di $A \leq \binom{n}{2} - \text{int}((n-1)/2)^2$;
- per ogni fissato n , se $k > k'$ si ha :

$$0 < z_{k|x=A} < z_{k'|x=A} \text{ per } 0 \leq A \leq \binom{n}{2} - \sum_{h=1}^{n-k'} \text{int}((n-h)/k');$$

- per ogni fissato n , se $k > k'$ si ha :

$$|\{z_{k''} | z_{k''} < 0\}| \geq |\{z_{k'} | z_{k'} < 0\}|;$$

- per ogni fissato k , z_k è dotata di:

$$\text{minimo assoluto } \min z_k = -\sqrt{\binom{n}{2}(k-1)} \text{ e}$$

$$\text{massimo assoluto } \max z_k = \sqrt{\binom{n}{2} - \sum_{h=1}^{n-k} \text{int}((n-h)/k)} / (k-1);$$

- per $k=2$, caso di modalità dicotomiche, è $z_2 = (2x - A) / \sqrt{A}$ rappresentabile in una serie di

$$1/2 \left(\binom{n}{2} - 3 \binom{n}{2} \right) - \left(\sum_{h=1}^{n-2} \frac{\text{int}((n-h)/2)}{2} \right) - \sum_{h=1}^{\text{int}(n/2)} \binom{n-2h+1}{2}$$

$$\text{punti con } \min z_2 = -\sqrt{\binom{n}{2}} \text{ e } \max z_2 = \sqrt{\binom{n}{2} - \text{int}(((n-1)/2)^2)};$$

- per $k=n$, caso in cui le modalità cromatiche siano pari al numero delle unità territoriali, è per z_n :

$$\min z_n = -(n-1)\sqrt{n/2} \text{ e } \max z_n = \sqrt{n/2}.$$

Infine accanto alle considerazioni matematiche le rappresentazioni ne evidenziano alcune statistiche:

- all'aumentare di k , numerosità delle modalità cromatiche, i valori espressi da z_k tendono a collocarsi nella zona di rifiuto caratterizzata dalla presenza di autocorrelazione negativa;
- esiste $k=k^*$, con k^* minimo dipendente dal numero delle unità territoriali n e dal livello di significatività α , tale che l'insieme dei valori z_k appartenenti

- alla zona di rifiuto con la presenza di autocorrelazione positiva è vuoto;
 - esiste $k=k^{**}$, con k^{**} minimo dipendente da n ed α , tale che i valori di z , nella zona di accettazione siano solo $\binom{n}{2}$.

4.2 CASO DELLA VARIABILE N_{rs}

Se ci si limita a tener conto solo di congiunzioni che realizzano coppie prescelte di eventi di tipo rs con $r \neq s$ per la variabile casuale N_{rs} , indicata con x una sua determinazione, è

$$z_{rs,k} = (k^2x - 2A) / \sqrt{2Ak^2 + 4Dk - 4(A + 2D)}.$$

Questa è funzione reale delle variabili intere x, A, D, k ed in quanto definita per

$$\left\{ \begin{array}{l} k \geq 2 \\ 1 \leq A \leq \binom{n}{2} \\ 0 \leq x \leq \binom{n}{2} - \text{int}(((n-1)/2)^2) \\ x \leq A \\ D_{\min}(A) \leq D \leq D_{\max}(A), \end{array} \right.$$

con l'aggiunta della dipendenza delle determinazioni x da n, k e dagli operatori L e di quella dei valori D da n, A e dagli stessi operatori L .

La funzione $z_{rs,k}$ è limitata inferiormente da

$$z_{rs,k|x=0} = -2A / \sqrt{2Ak^2 + 4Dk - 4(A + 2D)},$$

per la quale è $\lim_{k \rightarrow +\infty} z_{rs,k|x=0} = 0$.

È limitata superiormente da

$$z_{rs,k|x=A} = A(k^2 - 2) / \sqrt{2Ak^2 + 4Dk - 4(A + 2D)}$$

per $1 \leq A \leq \binom{n}{2} - \text{int}(((n-1)/2)^2)$ ed è solamente maggiorata da $z_{rs,k|x=A}$ per $\binom{n}{2} - \text{int}(((n-1)/2)^2) < A \leq \binom{n}{2}$, avendosi per questa $\lim_{k \rightarrow +\infty} z_{rs,k|x=A} = +\infty$.

Pensata la $z_{rs,k}$ definita da variabili reali continue, questa ha differenziale totale:

$$\begin{aligned} dz_{rs,k} = & \left(k^2 / \sqrt{2A(k^2 - 2) + 4D(k - 2)} \right) dx + \\ & - \left((k^2x + 2A)(k^2 - 2) - 8D(k - 2) / \sqrt{(2A(k^2 - 2) + 4D(k - 2))^3} \right) dA + \\ & + \left(2(k^2x + 2A)(k - 2) / \sqrt{(2A(k^2 - 2) + 4D(k - 2))} \right) dD + ((2kx(Ak^2 + 3Dk + \\ & - 4A - 8D) + 4A(Ak + D)) / \sqrt{(2A(k^2 - 2) + 4D(k - 2))^3}) dx \end{aligned}$$

con componenti $\partial z_{rs,k} / \partial x$ sempre positiva, $\partial z_{rs,k} / \partial A$ sempre negativa, $\partial z_{rs,k} / \partial D \geq 0$ per $(k^2x - 2A)(k - 2) \geq 0$, $\partial z_{rs,k} / \partial k$ sempre positiva.

Se si considerano inizialmente quali parametri le variabili n,k si hanno ipersuperfici di livello rappresentabili in spazi R_{nk}^4 e se inoltre si considerano parametri A e D si ottengono di $z_{rs,k}$ rappresentazioni planari con sistema di riferimento $xz_{rs,k}$.

Si stabilisca poi anche per $z_{rs,k}$ il livello di significatività α , (nell'es. è $\alpha=0.05$), si traccino nei piani $xz_{rs,k}$ le rette corrispondenti di delimitazione delle regioni di accettazione, in tal modo accanto a delle considerazioni di tipo matematico:

- nelle rappresentazioni parametriche planari di $z_{rs,k}$ i punti corrispondenti ad $A=\text{costante}$ e $D=\text{costante}$ sono allineati;
- per ogni fissato k, se $n'' > n'$ della fig. 4.2. $n'.k_{rs}$ sono anche punti della fig. 4.2. $n''.k_{rs}$;
- nel caso sia $k=2$, la $z_{rs,k} = z_2 = (2x - A) / \sqrt{A}$ è indipendente da D; se ne evidenziano anche alcune di tipo statistico quali:
- all'aumentare di k i valori espressi da $z_{rs,k}$ tendono a collocarsi nella zona di rifiuto caratterizzata dalla presenza di autocorrelazione positiva;
- esiste $k=k^{***}$, con k^{***} minimo dipendente dal numero delle unità territoriali n e dal livello di significatività α , tale che l'insieme dei valori z appartenenti alla zona di accettazione sia minimo.

Tali considerazioni matematico-statistiche sono rese intuitive attraverso le rappresentazioni illustrate (fig. 4.2. $n.k_{rs}$) pur se limitate ai valori marginali degli intervalli di variabilità di D.

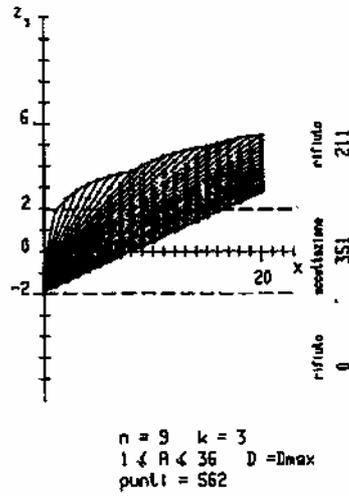
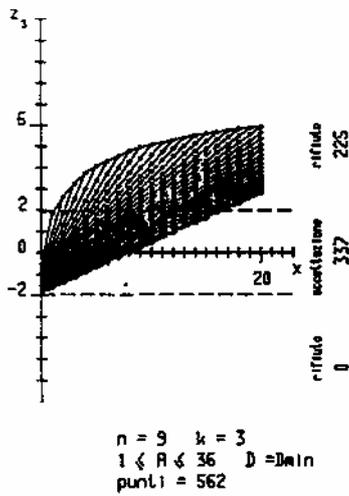


Fig. 4.2.9.3

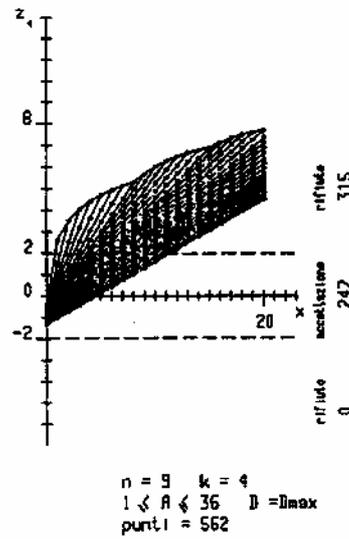
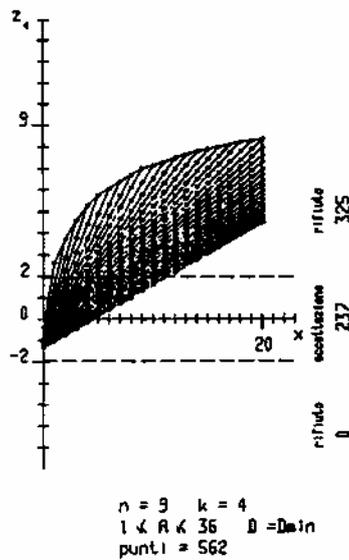
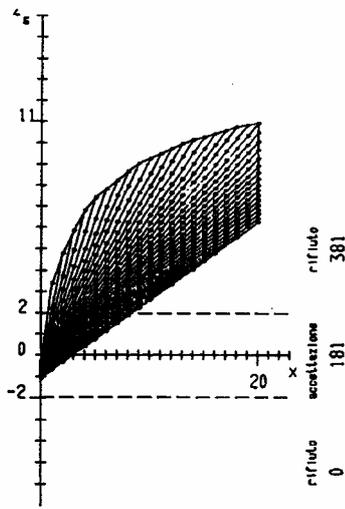
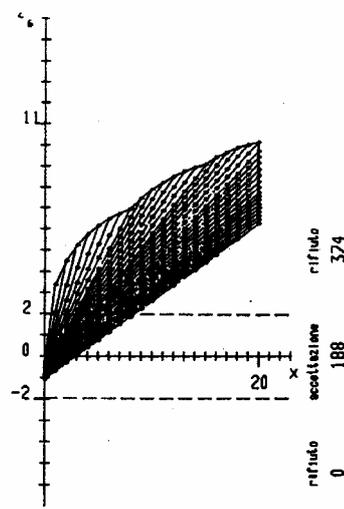


Fig. 4.2.9.4

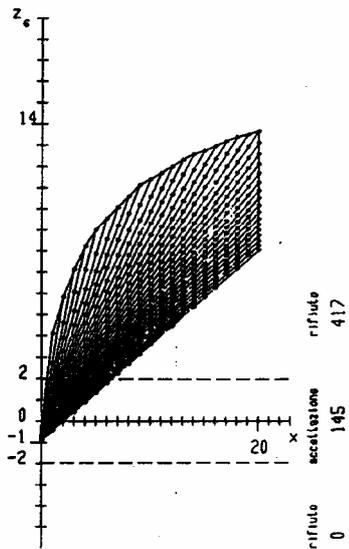


$n = 9 \quad k = 5$
 $1 < A < 36 \quad D = D_{min}$
 punti = 562

Fig. 4.2.9.5

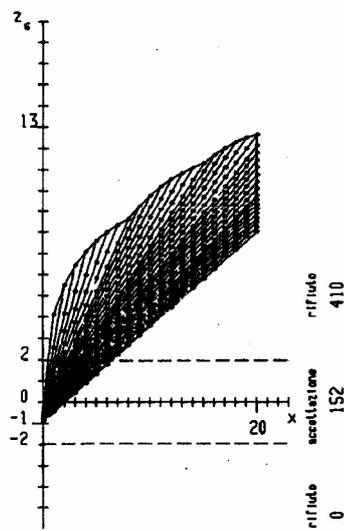


$n = 9 \quad k = 5$
 $1 < A < 36 \quad D = D_{max}$
 punti = 562

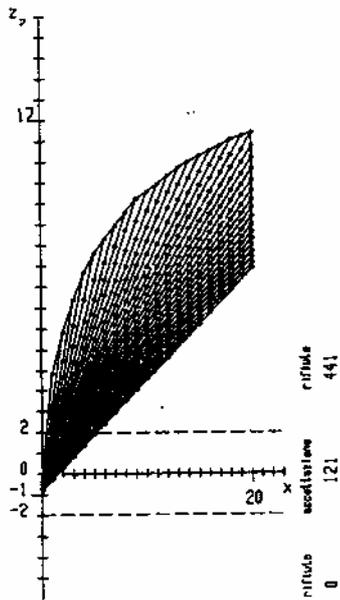


$n = 9 \quad k = 6$
 $1 < A < 36 \quad D = D_{min}$
 punti = 562

Fig. 4.2.9.6

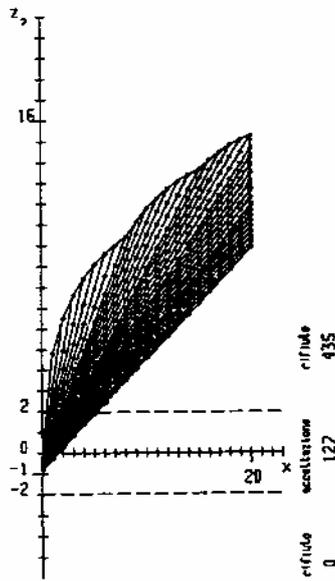


$n = 9 \quad k = 6$
 $1 < A < 36 \quad D = D_{max}$
 punti = 562

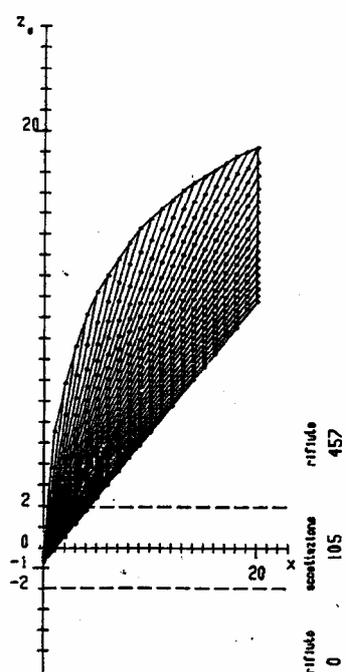


$n = 9 \quad k = 7$
 $1 < R < 36 \quad D = D_{min}$
 punti = 562

Fig. 4.2.9.7

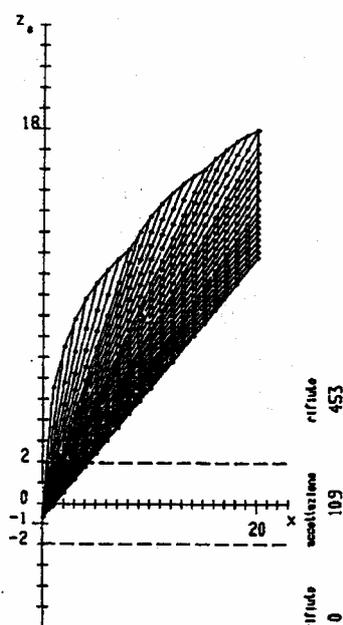


$n = 9 \quad k = 7$
 $1 < R < 36 \quad D = D_{max}$
 punti = 562

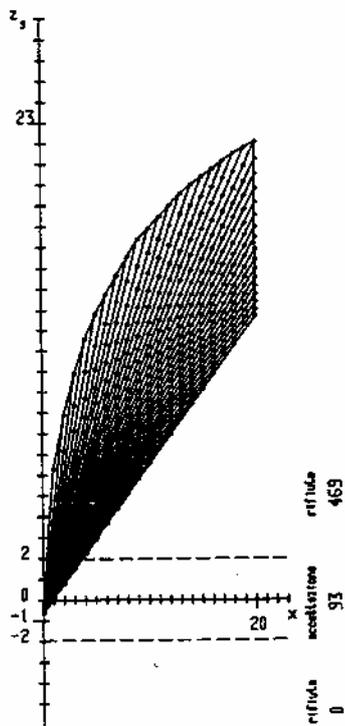


$n = 9 \quad k = 8$
 $1 \leq A \leq 36 \quad D = D_{\min}$
 punti = 562

Fig. 4,2.9.8

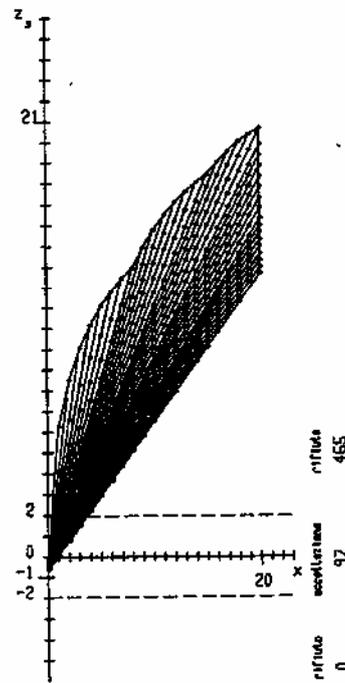


$n = 9 \quad k = 8$
 $1 \leq A \leq 36 \quad D = D_{\max}$
 punti = 562



$n = 9$ $k = 9$
 $1 \leq R \leq 36$ $D = D_{min}$
 punti = 562

Fig. 4.2.9.9



$n = 9$ $k = 9$
 $1 \leq R \leq 36$ $D = D_{max}$
 punti = 562

4.3 CASO DELLA VARIABILE N_{rr}

Infine, se si fa riferimento alla variabile casuale N_r è

$$z_{r,k} = (k^2x - A) / \sqrt{Ak^2 + 2Dk - (A + 2D)}$$

funzione reale delle variabili intere x, A, D, k ed n definita per:

$$\begin{cases} k \geq 2 \\ 0 \leq x \leq A \\ 1 \leq A \leq \binom{n}{2} \\ D_{\min}(A) \leq D \leq D_{\max}(A) \end{cases}$$

con l'attenzione riservata (4.2) per le determinazioni di x e per i valori assumibili dalla variabile D .

È limitata inferiormente da $z_{r,k|x=0}$ e superiormente da $z_{r,k|x=A}$ aventi rispettivamente $\lim_{k \rightarrow \infty} z_{r,k|x=0} = 0$ e $\lim_{k \rightarrow \infty} z_{r,k|x=A} = +\infty$.

Considerata la $z_{r,k}$ definita da variabili reali continue essa ha differenziale totale:

$$\begin{aligned} dz_{r,k} = & \left(k^2 / \sqrt{A(k^2 - 1) + 2D(k - 1)} \right) dx + \\ & - \left(((k + 1)(A + k^2x) + 4D) / \sqrt{4(k - 1)(Ak + A + 2D)^3} \right) dA + \\ & + \left((k - 1)(k^2x - A) / \sqrt{A(k^2 - 1) + 2D(k - 1)} \right) dD + \\ & \left((k^3xA + 3k^2xD + k(A^2 - 2xA - 4xD) + AD) / \sqrt{((k - 1)(Ak + A + 2D))^3} \right) dk \end{aligned}$$

con componenti $\partial z_{r,k} / \partial x$ sempre positiva, $\partial z_{r,k} / \partial A$ sempre negativa, $\partial z_{r,k} / \partial D \geq 0$ per $k^2x - A \geq 0$, $\partial z_{r,k} / \partial x$ sempre positiva.

Si traggono infine dalla $z_{r,k}$ le seguenti osservazioni:

- nelle rappresentazioni parametriche di $z_{r,k}$ i punti corrispondenti ad $A = \text{costante}$ e $D = \text{costante}$ sono allineati e costituiscono una nuvola di al più

$$1/2 \left(\binom{n}{2}^2 + 3 \binom{n}{2} \right) - \sum_{h=2}^{n-2} \binom{n-h}{2} \text{ punti};$$

- per ogni fissato k , se $n > n'$ i punti della fig. 4.3. $n'.k_r$ sono anche punti della fig. 4.3. $n''.k_r$;

- nel caso sia $k=2$, si ha $z_{r,2} = (4x - A)\sqrt{3A + 2D}$ con $\min z_{r,2} = -\sqrt{\binom{n}{2}}/3$

$$\text{e } \max z_{r,2} = 3\binom{n}{2} / \sqrt{3\binom{n}{2} + 2};$$

- nel caso sia $k \geq n$ si ha l'equivalenza logica $z_{r,k} < 0 \iff x=0$.

Tali osservazioni si rendono evidenti nelle rappresentazioni parametriche planari (fig. 4.3. n.k_n) pur se limitate ai valori marginali $D_{\min}(A)$ e $D_{\max}(A)$.

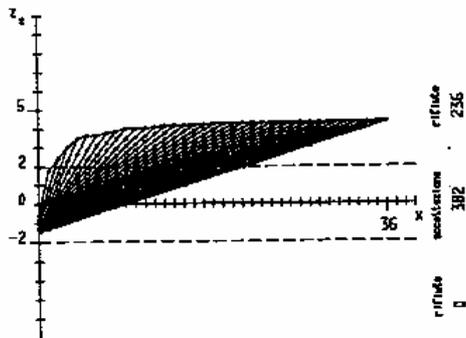
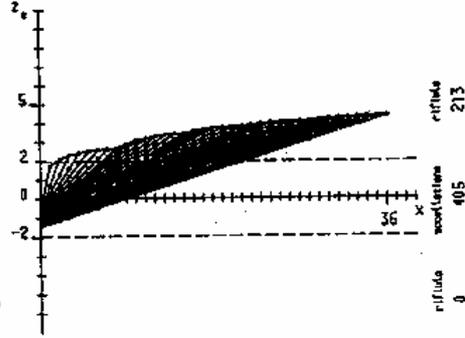


Fig. 4.3.9.2

$n = 9 \quad k = 2$
 $1 \leq A \leq 36 \quad D = D_{\min}$
 punti = 618



$n = 9 \quad k = 2$
 $1 \leq A \leq 36 \quad D = D_{\max}$
 punti = 618

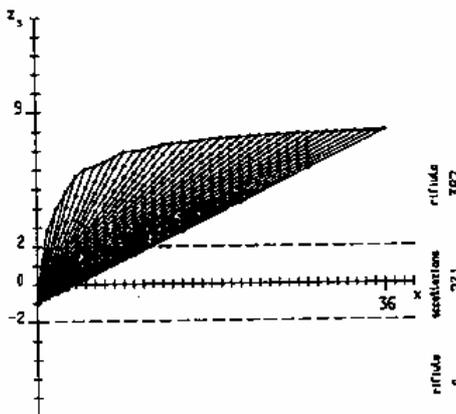
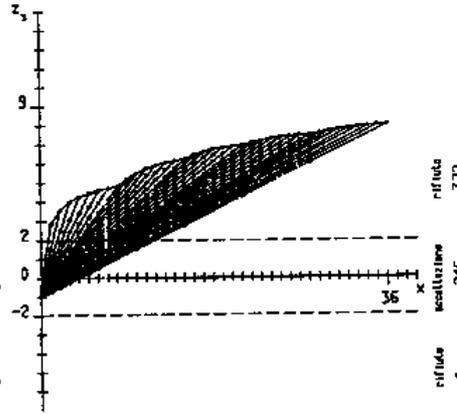


Fig. 4.3.9.3

$n = 9 \quad k = 3$
 $1 \leq A \leq 36 \quad D = D_{\min}$
 punti = 618



$n = 9 \quad k = 3$
 $1 \leq A \leq 36 \quad D = D_{\max}$
 punti = 618

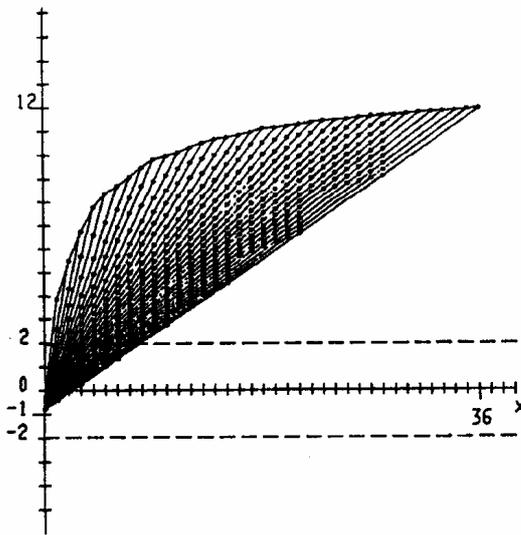
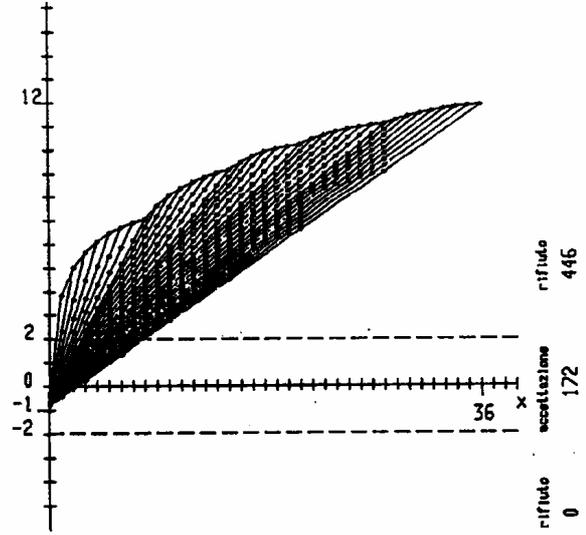


Fig. 4.3.9.4

$n = 9 \quad k = 4$
 $1 \leq A \leq 36 \quad D = D_{min}$
 punti = 618



$n = 9 \quad k = 4$
 $1 \leq A \leq 36 \quad D = D_{max}$
 punti = 618

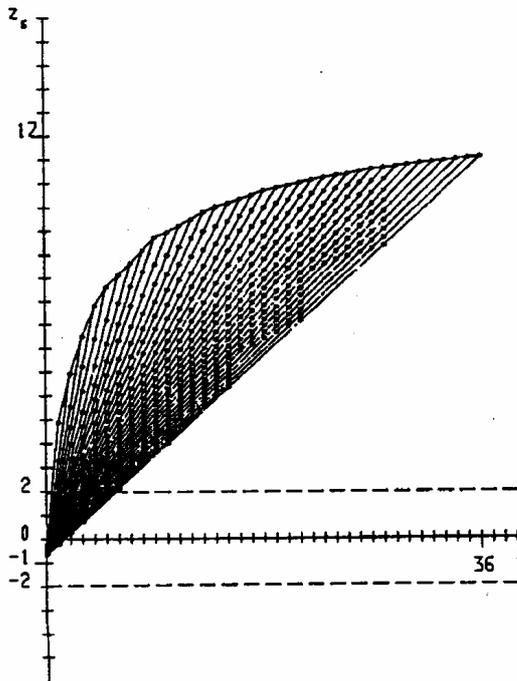
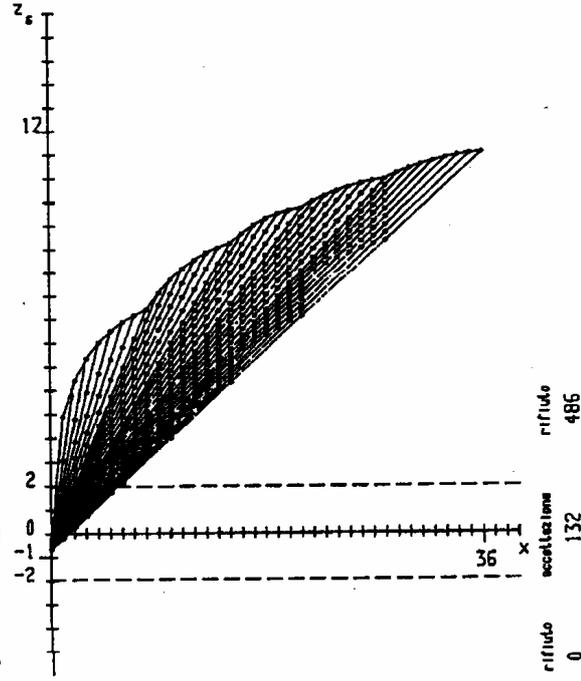


Fig. 4.3.9.5

$n = 9 \quad k = 5$
 $1 \leq A \leq 36 \quad D = D_{min}$
 punti = 618



$n = 9 \quad k = 5$
 $1 \leq A \leq 36 \quad D = D_{max}$
 punti = 618

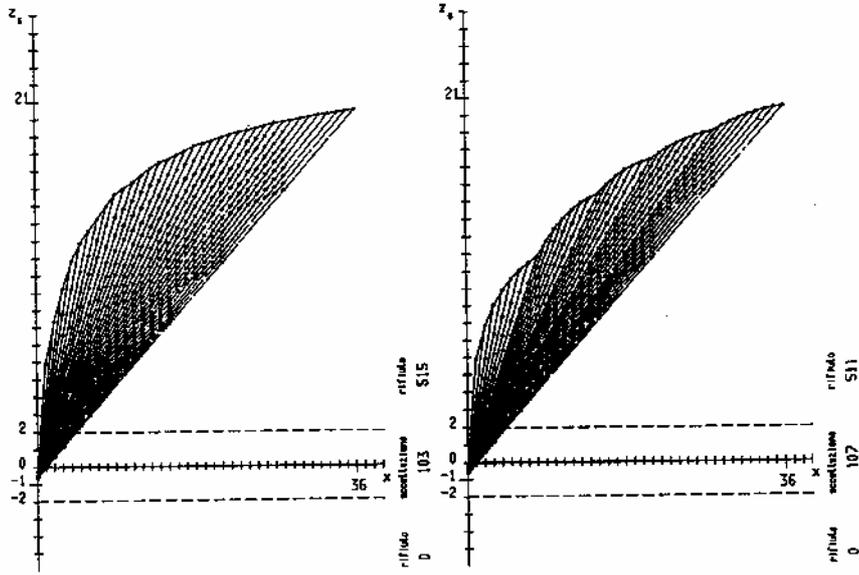


Fig. 4.3.9.6

$n = 9 \quad k = 6$
 $1 \leq R \leq 36 \quad D = \text{Dein}$
 punti = 618

$n = 9 \quad k = 6$
 $1 \leq R \leq 36 \quad D = \text{Deax}$
 punti = 618

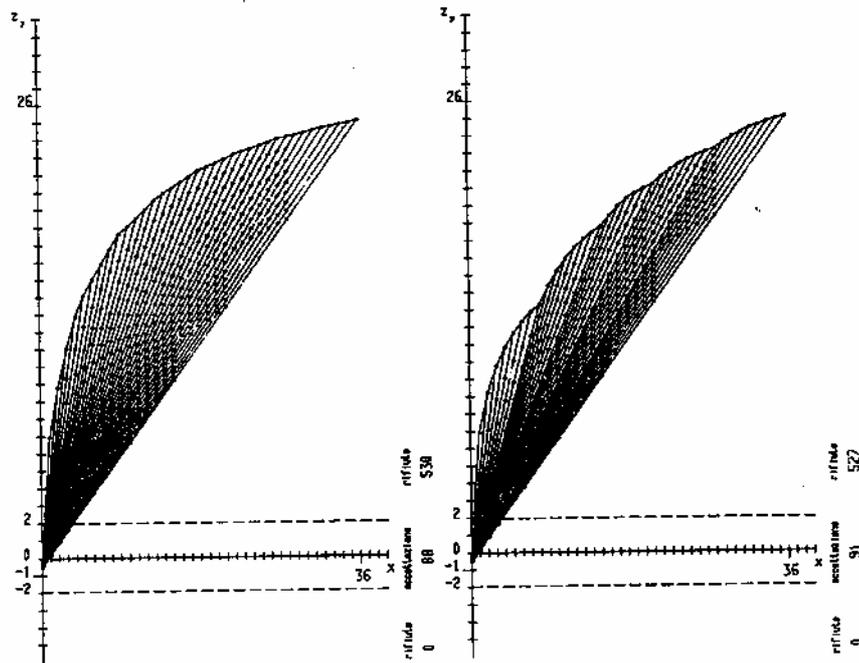


Fig. 4.3.9.7

$n = 9 \quad k = 7$
 $1 \leq R \leq 36 \quad D = \text{Dein}$
 punti = 618

$n = 9 \quad k = 7$
 $1 \leq R \leq 36 \quad D = \text{Deax}$
 punti = 618

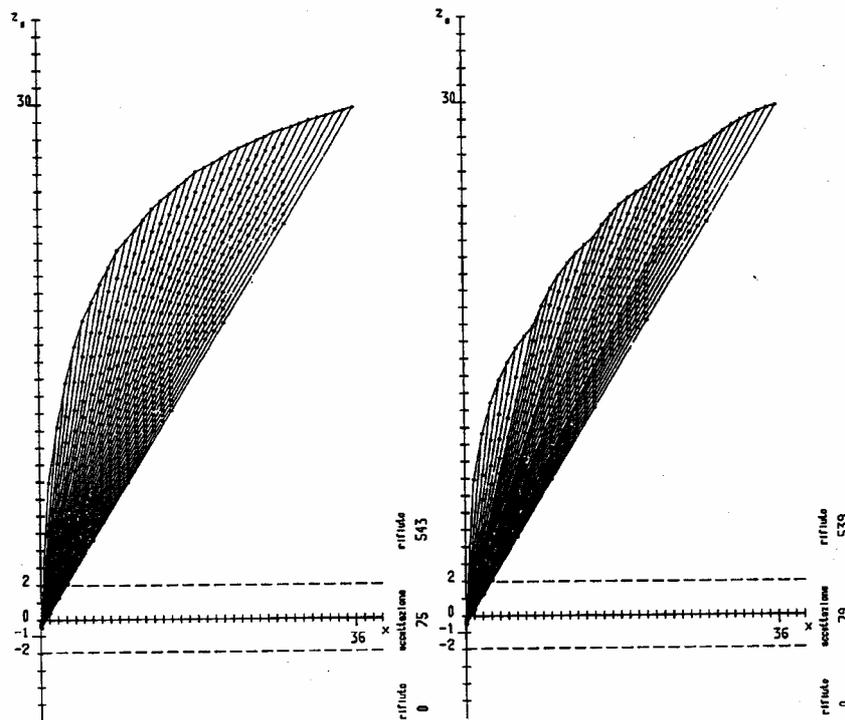


Fig. 4-3.9.8

$n = 9$ $k = 8$
 $1 < A < 36$ $D = D_{min}$
 punti = 618

$n = 9$ $k = 8$
 $1 < A < 36$ $D = D_{max}$
 punti = 618

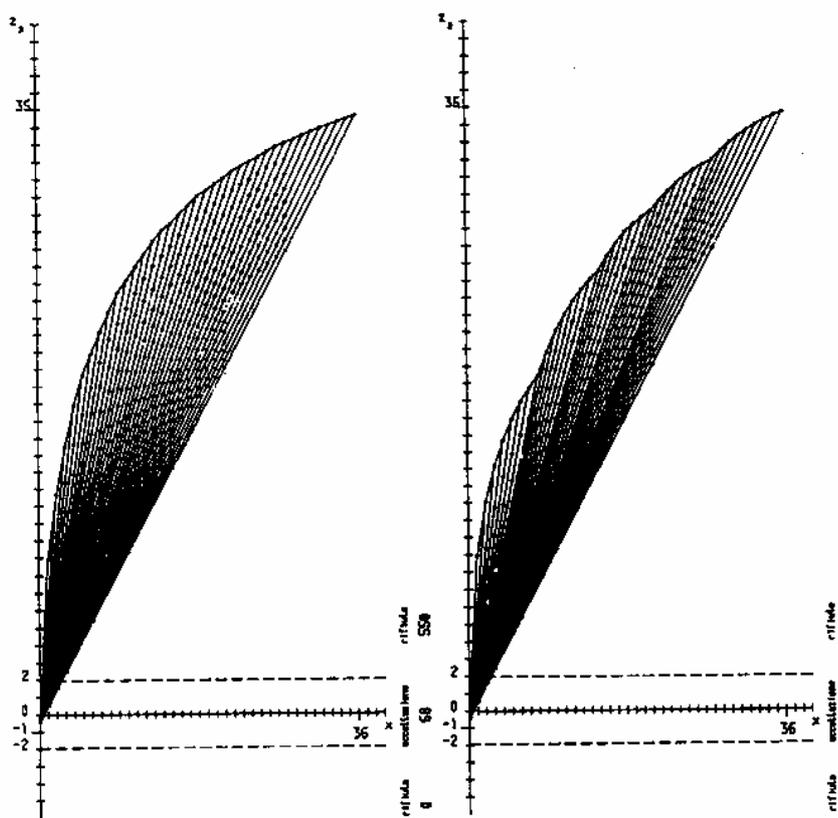


Fig. 4.3.9.9

$n = 9$ $k = 9$
 $1 \leq A \leq 36$ $D = \text{Dmin}$
 punti = 618

$n = 9$ $k = 9$
 $1 \leq A \leq 36$ $D = \text{Dmax}$
 punti = 618

5. CONCLUSIONI

L'itinerario seguito induce a legittimare un'asserzione di fondo ed alcune possibili implicazioni.

La prima si riassume nell'incidenza della numerosità policromica k sulla capacità del test z a svolgere il suo ruolo in tutte le condizioni operative, il che, in altre parole, parla di una necessità di riconsiderare non tanto il campo di applicabilità del test, quanto, piuttosto, la sua formulazione in generale.

Le implicazioni ulteriori, ed in particolare quelle desumibili dalle simulazioni effettuate, secondo cui k possa essere al più dell'ordine di \sqrt{n} , necessitano di ulteriori ricerche finalizzate all'individuazione di un valido criterio per definire la soglia.

In questa strada si cercheranno ulteriori approfondimenti e collegamenti tra sistema informativo e sistema descrittivo dei fatti territoriali.

BIBLIOGRAFIA

1. Anselin L. (1980) "*Estimation methods for spatial autoregressive structures: a study in spatial econometrics*", Ithaca, Cornell N.J. University.
2. Bahrenberg G., Fischer M., Nijdam P. (1984) "*Recent development in spatial data analysis. Methodology measurement models*", Gower Aldershot.
3. Cliff A.D., Ord J.K. (1981a) "*Spatial and temporal analysis: autocorrelation in space and time*", in Wrigley N. e Bennet R.J. (eds), *Quantitative Geography: A British View*, Londra, Routledge e Kegan, 104-110.
4. Cliff A.D., Ord J.K. (1981b) "*The effects of spatial autocorrelation of geographical modelling*", in Craig R.D. e Labowitz M.L. (eds), *Future trends in Geomathematics*, Londra, Pion, 108-137.
5. Gatrell A.C. (1979), "*Autocorrelation in spaces*", *Environment and Planning A* 11, 507-516.
6. Hubert L.J., Golledge R.G. e Costanzo C.N. (1981) "*Generalized procedures for evaluating spatial autocorrelation*", *Geographical Analysis* 13, 224-233.
7. Martin R.L. (1974) "*On autocorrelation, bias and the use of first spatial differences in regression analysis*", *Area* 6, 185,194.
8. Moran P. (1950), "*Notes on continuous stochastic phenomena*", *Biometrika* 37, 17-23.
9. Sclocco T. e Visini G. (1982), "*Sui primi due momenti della statistica joint-count di Moran*", *Atti VI Congresso A.M.A.S.E.S.*, 621-642.
10. Wrigley N. (1985), "*Categorical data analysis for geographer and environmental scientists*", Longman, New York.