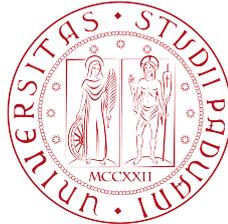


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**Modelli statistici
per l'analisi di processi di punto:
un'applicazione sullo storico derby di Milano**

Relatore Prof. Emanuele Aliverti
Dipartimento di Scienze Statistiche

Laureando Shehata Marco
Matricola 2027129

Anno Accademico 2021/2022

Abstract

La definizione di metodi e modelli statistici applicati al calcio è diventata un tema sempre più di interesse tra gli studiosi. L'obiettivo è quello di estrapolare conoscenza dall'elevata mole di dati disponibili, in tempi sempre più brevi. In questo elaborato viene effettuata l'analisi di processi di punto spaziali, con il fine di individuare le zone di campo in cui vengono osservati con maggiore intensità la partenza e l'arrivo dei passaggi di una squadra. Da questi è possibile definire le tattiche di gioco adottate, informazione utile per l'allenatore prima, durante e dopo la partita. L'applicazione si focalizza su una singola partita: lo storico derby Milan contro Inter, avvenuto il 15 ottobre 2017.

Indice

Introduzione	1
1 Introduzione ai processi di punto	3
1.1 Assunzioni	3
1.2 Notazione	4
1.3 I dati: eventi calcistici	5
1.4 Intensità	6
1.5 Processo Poisson Omogeneo	7
1.5.1 Completa casualità spaziale	7
1.5.2 CSR: verifica di ipotesi	8
1.6 Approccio non parametrico	10
2 Processi di Poisson Disomogenei	13
2.1 Processo Poisson Disomogeneo	13
2.1.1 Processi spaziali parametrici	13
2.1.2 Massima Verosimiglianza	14
2.1.3 Quadratura numerica	16
2.1.4 Applicazione e bontà di adattamento	19
2.2 Analisi dei residui	22
2.2.1 Introduzione	22
2.2.2 Residuo del processo di punto	23
2.3 Correlazione tra punti	24
2.4 Momenti dei processi di punto	25
2.5 Pair correlation function	26
2.5.1 Definizione	26
2.5.2 Stima Empirica	27
2.5.3 Interpretazione di $\rho(\cdot)$	28

3	Processi di Cox	31
3.1	Processi di Cox	31
3.1.1	Eterogeneità spaziale	31
3.1.2	Processi Log-Gaussiani di Cox	32
3.2	Stima di LGCP	34
3.2.1	Discretizzazione	34
3.2.2	Dipendenza markoviana	35
3.2.3	Conditional AutoRegressive	36
3.2.4	Funzione Whittle-Matern	37
3.2.5	Criterio di informazione: DIC	38
3.3	Integrated nested Laplace Approximation	39
3.3.1	Introduzione ad INLA	39
3.3.2	Inferenza bayesiana con INLA	42
3.4	Applicazione ai passaggi rossoneri	43
3.5	Il derby di Milano	44
3.5.1	Due squadre, quattro processi	44
3.5.2	Analisi dei residui in contesto Bayesiano	45
3.5.3	Tattiche individuate	46
4	Limiti e possibili miglioramenti	49
4.1	Punti coincidenti	49
4.2	Processi di punto multitypo	50
4.3	Piastrellatura e SPDE	52
4.4	Spazio e Tempo	53
	Conclusion	57

Introduzione

Chi segue qualsiasi sport con interesse sa bene che uno dei fattori chiave da studiare è quello legato all'analisi dei dati, con la quale risulta possibile estrarre informazioni utili inerenti ai singoli atleti, ai singoli team oppure ai vari incontri disputati. Le varie squadre, tra cui quelle calcistiche, attualmente hanno a disposizione una mole sempre maggiore di dati, considerando sia le sessioni d'allenamento che le partite disputate durante il campionato. L'obiettivo dell'analisi di questi dati complessi e di elevata dimensionalità è quello di guadagnare un vantaggio competitivo contro le squadre avversarie.

L'obiettivo di questo elaborato è quindi presentare modelli che descrivano al meglio il processo associato ai dati oggetto di studio, che permettano di ottenere informazioni utili all'allenatore, e che siano di supporto decisionale in termini di gestione di squadra prima, durante e dopo la partita. Questo tipo di analisi deve tener conto della natura dinamica dello sport, non omettendo così l'aspetto spaziale. Le analisi descrittive non sono uno strumento sufficiente per raggiungere tale scopo. Per questo, negli ultimi anni, è esploso l'interesse nell'applicare metodi e modelli statistici opportuni [10].

Prima di effettuare qualsiasi tipo di analisi dei dati in ambito sportivo, è opportuno possedere nozioni base associate allo sport di interesse. Questo permette non solo di tener conto di fatti stilizzati durante le analisi, ma anche di consultare la letteratura statistica associata all'analisi dei dati. La National Basketball Association, per esempio, utilizza tecniche di analisi molto più sofisticate di quelle adottate dagli altri sport, e può essere vista come punto di riferimento.

Nel seguente elaborato verrà presentato il processo di punto spaziale e la sua applicazione in contesto calcistico. La stima di un opportuno modello è fondamentale per ottenere una corretta rappresentazione grafica delle zone di intensità in cui avvengono eventi di interesse durante i 90 minuti di gioco. Nello specifico, nel capitolo 1 verranno introdotti i processi di punto e i processi Poisson Omogenei; nel capitolo 2 verranno descritti i processi Poisson Disomogenei; nel capitolo

3 si parlerà di Processi di Cox e, in ultima battuta, nel capitolo 4 si discuterà dei limiti dell'analisi e dei possibili miglioramenti.

Capitolo 1

Introduzione ai processi di punto

In questo capitolo viene fornita la definizione di processo di punto, indicando le assunzioni poste al fine di definire una completa teoria statistica. Dopo aver descritto i dati in esame e la notazione adottata durante il prosieguo dell'elaborato, viene dedicata una sezione alla quantità di primario interesse in questo elaborato: l'intensità. In ultima battuta vengono introdotti i processi Poisson Omogenei.

1.1 Assunzioni

Per punti spaziali si intende un insieme di dati sotto forma di coordinate spaziali, rappresentanti la posizione in cui è stato osservato un evento di interesse. Questi dati spaziali sono realizzazioni di processi stocastici

$$Z(u) = \{z(u), u \in R\}$$

con $u = (x, y)$ posizione del punto in $R \subset \mathbb{R}^2$. Mentre $z(u)$ assume valori 0 o 1, indicando se l'evento è stato osservato o meno, la posizione u è realizzazione della variabile aleatoria U .

A differenza delle classiche analisi statistiche, durante l'analisi di punti spaziali, il ricercatore ha come obiettivo quello di definire se:

- i punti sono distribuiti uniformemente all'interno della regione di interesse;
- i punti sono disposti casualmente all'interno di R ;
- sono presenti cluster spaziali.

Assumendo che il numero di eventi osservati nello spazio R sia finito, allora in questo contesto siamo interessati all'analisi di processi di punto spaziale finito.

Si tratta di meccanismi generatori dei dati basati su 3 assunti fondamentali e sufficienti per poter definire una completa teoria statistica:

- sia $\bar{u} = \{u_1, \dots, u_n\}$ l'insieme degli n punti osservati, realizzazione del processo \bar{U} , allora $u_i \in R, \forall i \in \{1, \dots, n\}$;
- per ogni sotto-regione B , il numero di eventi osservati in B , $n(\bar{U} \cap B)$, segue un'opportuna legge probabilistica discreta;
- $u_i \neq u_j \forall i \neq j$: non sono ammessi punti duplicati.

1.2 Notazione

Durante il prosieguo di questo elaborato, si assume di interesse l'analisi degli eventi osservati all'interno della regione finita $R \subset \mathbb{R}^2$. Nel nostro caso la regione R è pari alla dimensione del campo da gioco, ovvero $R = [0, 105] \times [0, 68]$.

La posizione dell' i -esimo punto verrà indicata con le lettere minuscole quali per esempio u_i , o in caso di generalizzazione con la semplice lettera u , tale per cui $u_i = (x_i, y_i)$ oppure $u = (x, y)$, dove x rappresenta l'ascissa e y l'ordinata del punto.

Per sotto-regione (detta anche cella) si intende un sottoinsieme della regione di interesse, e verrà indicata con lettere maiuscole. Le sotto-regioni hanno una forma specifica, selezionata dal ricercatore in base all'analisi da effettuare. Per esempio, la generica sotto-regione $B = [a, b] \times [c, d]$ rappresenta un generico rettangolo o quadrato contenente tutti i punti u tali per cui $a \leq x \leq b$ e $c \leq y \leq d$. Per cerchio, centrato in u e di raggio $r > 0$, si intende il sottoinsieme della regione R formato da tutti i punti $v \in R$ tali per cui:

$$b(u, r) = \{v : \|u - v\| \leq r\}.$$

L'insieme di osservazioni spaziali viene indicato con $\bar{u} = \{u_1, \dots, u_n\}$, dove $n \in \mathbb{N}_0$ indica il numero di eventi osservati in R .

Dati \bar{u} e B sotto-regione di R , $\bar{u} \cap B$ indica il sottoinsieme di punti osservati appartenenti a B . Il numero di punti osservati in B è indicato con $n(\bar{u} \cap B)$.

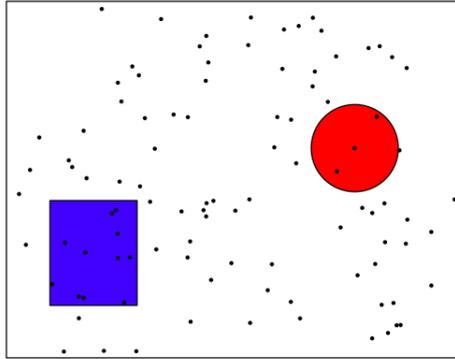


Figura 1.1: Esempio di rappresentazione della regione R , delle sotto-regioni B in blu e $b(u, r)$ in rosso; i punti sono stati generati casualmente

1.3 I dati: eventi calcistici

Per evento osservato si intende qualsiasi tipo di azione svolta dai giocatori in campo durante una partita di calcio, di cui le coordinate spaziali sono note. Esistono pertanto diverse tipologie di eventi, quali i passaggi, i tiri, i duelli tra due giocatori, e altri. Un esempio di dataset a disposizione è riportato in Tabella 1.1. L'unità statistica fa riferimento al singolo evento osservato (ID), di cui si conoscono le coordinate in metri di partenza (PX e PY) e quelle di arrivo della palla (AX e AY). Il tempo in cui viene osservato l'evento è misurato in secondi (Tempo).

I dati raccolti e utilizzati in questo elaborato sono forniti da Wyscout [19], un'azienda leader nel settore del calcio a supporto di più di 50 associazioni calcistiche e di più di 1.000 club professionistici in tutto il mondo. La procedura di raccolta dati è svolta da video-analisti esperti, formati e focalizzati sulla raccolta dati attraverso uno specifico software. Per garantire l'accuratezza della raccolta dei dati, il tagging degli eventi viene effettuato da tre operatori: un operatore per ogni squadra e un operatore che funge da supervisore responsabile dell'output dell'intera partita. Sotto richiesta dei club, per ottenere i dati in tempo reale, viene predisposto un team di quattro operatori. Uno di questi agisce per accelerare la raccolta di eventi complessi che richiedono attributi aggiuntivi e specifici. In aggiunta, si prende a carico la responsabilità di effettuare una rapida revisione. Si hanno a disposizione informazioni diverse, associate a tutte le partite della stagione 2017/2018 dei 5 campionati europei calcistici più importanti: Serie A (Italia), Bundesliga (Germania), LaLiga (Spagna), Premier Ligue (Inghilterra) e Ligue 1 (Francia). In questo elaborato ci si concentrerà sull'analisi dei passaggi

associati ad una singola partita della Serie A: il derby di San Siro, disputatosi il 15 ottobre 2017.

ID	PX	PY	AX	AY	Evento	Squadra	Tempo
1	50.40	34.68	61.95	23.80	Pass.	MILAN	2.53
2	61.95	23.80	86.10	30.60	Pass.	MILAN	4.12
3	86.10	30.60	102.90	38.76	Pass.	MILAN	6.58
4	102.90	38.76	56.70	61.20	Pass.	MILAN	8.63
5	56.70	61.20	43.05	51.68	Pass.	INTER	14.91
6	43.05	51.68	63.00	46.24	Pass.	INTER	18.69
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Tabella 1.1: Esempio di informazioni estratto dal dataset in cui sono raccolti i dati oggetto di studio

1.4 Intensità

La scelta della strategia da adottare per la modellazione e l'interpretazione dei dati oggetto di studio dipende sempre dagli obiettivi di indagine. Il focus principale può tradursi nella modellazione dell'intensità, oppure nell'analisi della dipendenza spaziale tra i vari punti, oppure ancora nella rappresentazione congiunta di intensità e dipendenza. È opportuno soffermarsi nel comprendere affondo il concetto di intensità, quantità utile a trarre conclusioni sul fenomeno oggetto di studio.

Per intensità λ si intende la densità di punti attesa e localizzata per area unitaria. In un contesto calcistico si decide di far riferimento al metro quadro. Questa viene interpretata come tasso di occorrenza, frequenza o incidenza degli eventi registrati. L'unità di misura di λ è pari al reciproco del quadrato dell'area unitaria. L'obiettivo primario dell'elaborato è quello di modellare questa quantità, decidendo se assumerla costante (λ) o variabile nello spazio ($\lambda(u)$), e se le covariate spaziali note la influenzano.

L'intensità è una quantità associata al momento primo. È possibile, quindi, definire dei modelli per l'intensità attraverso processi Poisson disomogenei, che si concentrino sulla rappresentazione di questa quantità, in funzione della latitudine e della longitudine opportunamente trasformate, e di covariate note. Usando approcci parametrici o semi-parametrici si ricava il trend spaziale associato al processo, trascurando però la dipendenza tra i punti, in quanto non considerata e complessa da integrare.

Tuttavia, se la dipendenza risulta essere non trascurabile al fine di ottenere una corretta stima dell'intensità, una possibile alternativa ai processi disomogenei è il

modello di Cox, che si basa sulla formulazione di processi *doppiamente stocastici*: l'intensità viene trattata a sua volta come processo stocastico nello spazio.

Il primo passo consiste nel definire se i punti sono casualmente disposti all'interno di R . La modellazione spaziale del processo avviene solo dopo aver verificato l'assenza di omogeneità, mediante un opportuno test e opportune analisi grafiche.

1.5 Processo Poisson Omogeneo

1.5.1 Completa casualità spaziale

In letteratura si parla di *Complete Spatial Randomness* (CSR) [1] quando il processo generatore dei dati porta ad osservare gli eventi di interesse nella regione R in modo totalmente aleatorio.

Dato B una sotto-regione di R tale che $B \subset R$, si parla di completa casualità spaziale se il processo presenta le seguenti proprietà: omogeneità, se i punti osservati avvengono in modo casuale all'interno di R ; indipendenza, se il numero di eventi osservati in una sotto-regione B è indipendente dal numero di eventi osservati in una sotto-regione A , dati A e B due sotto-regioni disgiunte, tali per cui $A \cap B = \emptyset$.

Con il fine di definire una distribuzione appropriata per il processo appena descritto, si consideri una sotto-regione B contenente n eventi osservati, e si suddivida B in celle disgiunte di piccole dimensioni. Si pensi, ad esempio, allo schermo di un generico PC come rappresentazione di B , e ai pixel dello schermo come la suddivisione di interesse. Si assuma che le sotto-regioni di B contengano al massimo un solo punto osservato. Sotto ipotesi di indipendenza tra sotto-regioni disgiunte, allora ogni sotto-regione di B presenta probabilità non nulla di contenere un singolo punto, seguendo una legge Bernoulliana. Di conseguenza $n(\bar{U} \cap B)$ è pari al numero di successi su un numero elevato di eventi indipendenti, dove ogni evento ha una probabilità non nulla di avvenire. Si conclude che

$$n(\bar{U} \cap B) \sim \text{Poisson}(\lambda|B|),$$

dove la proprietà di omogeneità si riassume in

$$E[n(\bar{U} \cap B)] = \lambda|B| = \text{Var}[n(\bar{U} \cap B)]$$

e dove $|B|$ rappresenta l'area di B . Essendo $\lambda|B| > 0$ allora $\lambda > 0$, ed è chiamato intensità del processo, che risulta essere costante nello spazio.

Riassumendo, si parla di processo di punto Poisson omogeneo (PPO) [1] se sono soddisfatte le seguenti proprietà:

- omogeneità: $E[n(\bar{U} \cap B)] = \lambda|B|, \forall B \subseteq R$;
- indipendenza: dati $A, B \subset R$, con $A \cap B = \emptyset$, allora $n(\bar{U} \cap A) \perp\!\!\!\perp n(\bar{U} \cap B)$;
- distribuzione Poisson: $n(\bar{U} \cap B) \sim Poisson(\lambda|B|)$.

1.5.2 CSR: verifica di ipotesi

Come detto in precedenza, è opportuno verificare che il processo non sia spazio-indipendente, ossia si vuole dimostrare che l'intensità non è costante nello spazio. Pertanto, bisogna stimare l'intensità λ assumendo un modello PPO e, successivamente, verificare mediante opportuni grafici o test statistici l'ipotesi che il processo spaziale risulti omogeneo. Se questa ipotesi viene rifiutata, il passaggio successivo è quello di modellare l'intensità in funzione dello spazio. Per il modello PPO è possibile dimostrare che lo stimatore di massima verosimiglianza, consistente e non distorto per λ , è pari a

$$\hat{\lambda} = \frac{n(\bar{u})}{|R|},$$

ovvero è pari alla densità empirica dei punti, dove $n(\bar{u})$ rappresenta il numero di eventi osservati in tutta la regione di interesse R , e $|R|$ è pari all'area di R . Si consideri come evento di interesse i punti di partenza dei passaggi effettuati dal Milan durante i 90 minuti di gioco, al fine di capire quali sono le maggiori zone di campo da cui partono i passaggi e identificare i giocatori chiave, punti di partenza della tattica di gioco rossonero. La difesa rossonera è stata posizionata nella porzione sinistra del campo. Con i dati attualmente presi in considerazione, $\hat{\lambda} = 448/7140 = 0.0627$. La varianza dello stimatore è pari a

$$V[\hat{\lambda}] = \frac{\lambda}{|R|},$$

di conseguenza

$$\widehat{V[\hat{\lambda}]} = \frac{\hat{\lambda}}{|R|}.$$

Nel caso studio considerato $\widehat{V}[\widehat{\lambda}] = 8.78 \times 10^{-6}$.

Al fine di verificare l'ipotesi di omogeneità, che si traduce nella verifica dell'ipotesi

$$\lambda(u) = \lambda, \quad \forall u \in R,$$

si usano approcci non parametrici per effettuare un'opportuna analisi esplorativa. Un modo semplice per confermare l'assenza di omogeneità è verificare se regioni di uguale area contengono un numero approssimativamente uguale di punti, come accadrebbe nel caso in cui il processo di punto sia PPO.

Si utilizza una rappresentazione grafica, detta *lattice*, come quella raffigurata in Figura 1.2 [4]. Si divide la regione R in sotto-regioni B_1, \dots, B_m t.c. $|B_1| = |B_2| = \dots = |B_m|$. L'assunzione di uguale area delle sotto-regioni è stata posta per pure ragioni esplicative. In generale è possibile assumere che il generico $|B_i| \neq |B_j|$, per $i \neq j$. L'intensità associata alla regione B_j è pari a

$$\lambda_j = n(\bar{u} \cap B_j) / |B_j|.$$

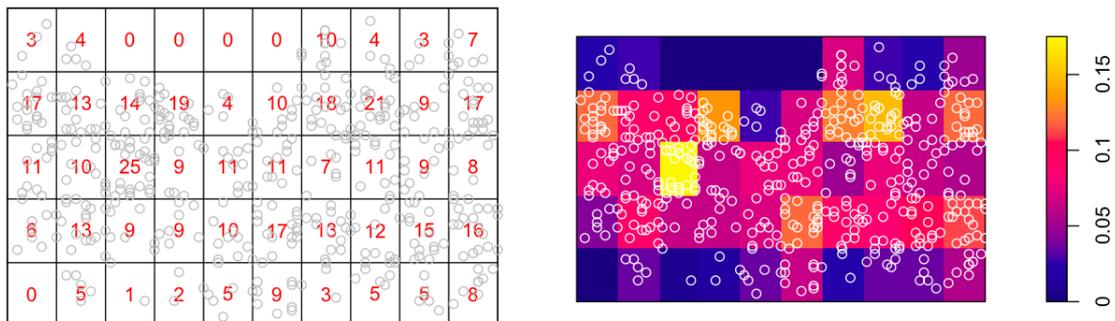


Figura 1.2: Rappresentazione grafica lattice applicata all'esempio della partenza passaggeri rossonera, utilizzando 50 sotto-regioni di area e forma coincidente: a sinistra rappresentazione mediante conteggio, a destra mediante intensità stimata

Dalla Figura 1.2 si evince la possibile assenza di omogeneità. Le fasce, soprattutto la fascia sinistra, presentano un livello di intensità evidentemente minore rispetto alla zona centrale del campo. Addirittura, in alcune regioni B_j il numero di eventi osservati è nullo.

Questa prima indicazione ci suggerisce, quindi, di effettuare un opportuno test per la verifica di ipotesi di omogeneità del processo. Basandoci sempre su una rappresentazione lattice del campo, è possibile utilizzare un test χ^2 bontà di adattamento, dove sotto ipotesi nulla si assume che il processo sia Poisson omogeneo. Siano n_1, \dots, n_m il numero di eventi osservati in B_1, \dots, B_m rispettivamente. Sot-

to ipotesi nulla, n_j è realizzazione di una variabile aleatoria indipendente con distribuzione Poisson di media $\mu_j = \lambda|B_j|$. Ponendo $n = \sum_{j=1}^m n_j$, ed essendo $|R| = \sum_{j=1}^m |B_j|$, allora $\bar{\lambda} = n/|R|$, e il numero di eventi attesi per ogni quadrato $e_j = \bar{\lambda}|B_j| = n|B_j|/|R|$. La statistica test, sotto ipotesi nulla, è definita come

$$X^2 = \sum_{j=1}^m \frac{(n_j - e_j)^2}{e_j} \sim \chi^2(m - 1).$$

Con i dati a nostra disposizione, provando a dividere il campo in 9, 25 e 50 rettangoli di uguale area, si ottiene $\alpha^{oss} \doteq 0$ per ogni partizione adottata. Quindi si rifiuta l'ipotesi nulla di omogeneità.

E' possibile anche stimare la distribuzione sotto ipotesi nulla della distribuzione di X^2 attraverso un approccio bootstrap [6]. Si ottiene un valore di α^{oss} pari a 0.001 per ogni ripartizione selezionata (a 9, 25, 50 rettangoli rispettivamente). Anche in questo caso viene rifiutata l'ipotesi nulla, confermando così l'assenza di omogeneità legata al processo.

La principale criticità del test χ^2 appena adottato è la debolezza dell'informazione estrapolata dalla verifica di ipotesi effettuata. Rifiutare l'ipotesi nulla permette solo di affermare che il processo generatore non è Poisson Omogeneo. Questo implica la presenza di disomogeneità oppure l'esistenza di una qualche forma di dipendenza tra i punti, ma non fornisce indicazioni su quali delle due caratteristiche siano effettivamente presenti nei dati.

Oltre a questo aspetto, risulta problematico definire una opportuna partizione di R . Da un punto di vista pratico, la procedura adottata dallo statistico consiste nel provare varie partizioni e confrontare i risultati ottenuti.

1.6 Approccio non parametrico

Date le lacune del test χ^2 , è necessario introdurre un nuovo strumento utile a comprendere la funzione di intensità del processo. Esistono tre stimatori non parametrici proposti e tendenzialmente adottati al fine di avere una prima idea della forma di $\lambda(u)$. Siano

$$\tilde{\lambda}(u) = \sum_{i=1}^n k(u - x_i) \tag{1.1}$$

$$\tilde{\lambda}_U(u) = \frac{1}{e(u)} \sum_{i=1}^n k(u - x_i) \quad (1.2)$$

$$\tilde{\lambda}_D(u) = \sum_{i=1}^n \frac{1}{e(x_i)} k(u - x_i) \quad (1.3)$$

rispettivamente lo stimatore non corretto, lo stimatore uniformemente corretto e lo stimatore proposto da Diggle [9].

La funzione $k(\cdot)$ è detta funzione nucleo, tale per cui $k(u) \geq 0 \forall u$, e tale per cui $\int_{\mathbb{R}^2} k(u) du = 1$. Il nucleo coincide con funzioni di densità simmetriche attorno all'origine e la scelta più comune è quella di utilizzare una distribuzione gaussiana isotropica a media nulla, ovvero dato $u = (x, y)$ allora

$$k(u; \sigma) = k(x, y; \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right).$$

La quantità σ è un parametro di regolazione, detto larghezza di banda, in quanto è proporzionale all'ammontare di lisciamiento imposto sull'intensità stimata. La scelta del parametro di regolazione si basa sul classico compromesso varianza-distorsione. A tal fine viene stimato l'errore quadratico medio mediante convalida incrociata, seguendo la procedura proposta da Diggle [9].

La quantità

$$e(u) = \int_R k(u - v) dv$$

viene applicata come correzione, in quanto lo stimatore (1.1) risulta distorto per valori di u prossimi ai confini della regione R , effetto noto in letteratura come effetto bordo. Questa elevata distorsione si presenta soprattutto in contesti in cui gli eventi avvenuti al di fuori di R influenzano l'avvenire degli eventi osservati all'interno della regione. Essendo assente il loro contributo nella stima, all'avvicinarsi del confine la distorsione aumenta. Gli stimatori (1.2) e (1.3) sono stati introdotti col fine di superare questo limite. Tuttavia, in contesto calcistico non si osservano eventi al di fuori del campo. Vengono presentate in Figura 1.3 le stime ottenute, da cui si evince chiaramente che i tre stimatori restituiscono valori pressoché identici, come ci si poteva aspettare.

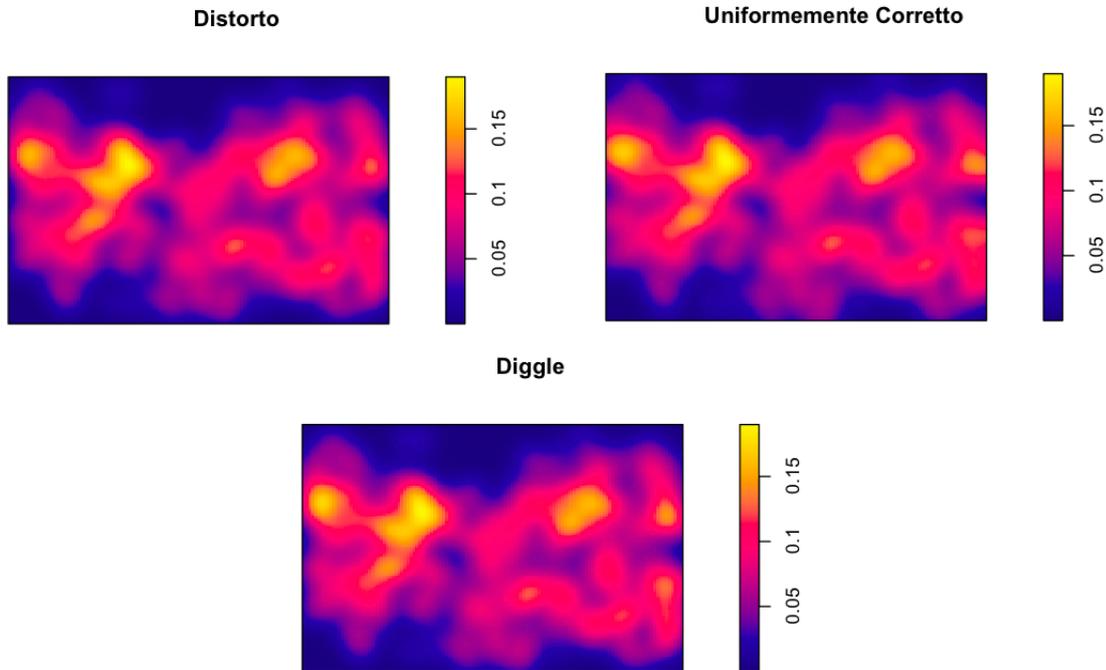


Figura 1.3: Stimatori kernel dell'intensità in funzione dello spazio, applicata ai passaggi in partenza rossoneri

Come si evince dalla Figura 1.3 la disposizione dei punti nello spazio non è completamente casuale: sono ben visibili due zone di alta intensità vicino alle porte e la presenza del trend spaziale osservato nella fascia centrale del campo. Possiamo concludere che esistono zone di campo specifiche da cui partono con maggior frequenza il numero di passaggi.

Capitolo 2

Processi di Poisson Disomogenei

In questo capitolo si introducono i Processi di Poisson Disomogenei, modelli utili a rappresentare il fenomeno oggetto di studio. Vengono descritte caratteristiche, procedura di stima e metodi di analisi della bontà di adattamento di modelli, utili a sottolineare le lacune di queste rappresentazioni inadeguate del fenomeno oggetto di studio.

2.1 Processo Poisson Disomogeneo

2.1.1 Processi spaziali parametrici

Approcci non parametrici come quelli proposti nella sezione 1.6 tendono a sovrastimare la presenza di cluster di punti all'interno di R . È necessario definire modelli più stabili attraverso l'imposizione di assunzioni parametriche. Se il nuovo modello si adatterà bene ai dati, restituirà una rappresentazione migliore della funzione di intensità.

Un processo di Poisson omogeneo si basa sugli assunti descritti nella sezione 1.5.1. Il processo Poisson disomogeneo (PPD) [1] può essere visto come una modifica del processo omogeneo, in quanto:

- $E[n(\bar{U} \cap B)] = \int_B \lambda(u) du$, dove $\lambda(u)$ è detta funzione di intensità;
- gli $n = n(\bar{U} \cap B)$ punti in B sono indipendenti e identicamente distribuiti e seguono distribuzione

$$f(u) = \frac{\lambda(u)}{\int_B \lambda(u) du}.$$

Sotto questi assunti, possiamo definire in qualsiasi modo la funzione $\lambda(u)$, a patto che i suoi valori siano non-negativi e la funzione sia integrabile, ovvero

$\int_B \lambda(u) du < +\infty \forall B \subseteq R$. Esistono scelte comunemente adottate in base al contesto in cui ci si trova.

Il modello omogeneo è un caso particolare dei modelli parametrici disomogenei, in cui $\lambda(u)$ è funzione costante e pari a λ , $\forall u \in R$.

Se risulta sensato dividere R in B_1, \dots, B_m , allora è possibile assumere che $\lambda_j(u) = \beta_j$, $u \in B_j$, ponendo costante l'intensità nella sotto-regione B_j . Questo tipo di modello viene applicato tendenzialmente in contesti geologici, dove R è un paese di interesse e B_1, \dots, B_m è una opportuna partizione nota di R . L'obiettivo, per esempio, potrebbe essere quello di definire le regioni maggiormente a rischio sismico.

Possiamo assumere anche che $\lambda(u)$ sia proporzionale ad una baseline nota di riferimento, ovvero $\lambda(u) = \theta b(u)$, con $b(u)$ noto e θ unico parametro da stimare. In un contesto epidemiologico, se la baseline $b(u)$ rappresenta la densità della popolazione allora stiamo assumendo che ogni membro della popolazione abbia probabilità pari a θ di contrarre una malattia rara. Questo tipo di modello è un modello a rischio costante.

L'opzione più sensata è quella di considerare un modello di regressione loglineare, in quanto racchiude come casi particolari le possibili scelte precedentemente descritte e in quanto tiene conto dell'effetto di covariate spaziali. In questo caso

$$\log(\lambda(u)) = b(u) + \theta^T Z(u), \quad (2.1)$$

dove $b(u)$ è una funzione nota, $Z(u) = (Z_1(u), \dots, Z_p(u))^T$ rappresenta le p variabili esplicative spazio-dipendenti e $\theta = (\theta_1, \dots, \theta_p)^T$ il vettore di parametri da stimare. È possibile assumere $b(u) = \alpha$ costante nello spazio ed ignoto, imponendo la presenza dell'intercetta. La scelta di definire il logaritmo dell'intensità, ispirata dal legame canonico del modello lineare generalizzato con risposta Poisson, è opportuna per evitare l'introduzione di vincoli durante la stima dei parametri, in quanto $\lambda(u) > 0 \forall u \in R$.

2.1.2 Massima Verosimiglianza

Per un processo disomogeneo, $\lambda(u)$ dipende da parametri ignoti θ che vanno opportunamente stimati. Avendo assunto che il numero di eventi osservati in una specifica sotto-regione di R segua distribuzione Poisson, è possibile fare inferenza su θ via massima verosimiglianza.

Per semplicità esplicativa, partiamo considerando $\lambda(u) = \lambda$ costante, ovvero

considerando un processo omogeneo. Dalla verosimiglianza

$$L(\lambda; \bar{u}) = L(\lambda) = \lambda^{n(\bar{u})} e^{(1-\lambda)|R|} \propto \lambda^{n(\bar{u})} e^{-\lambda|R|}$$

è possibile ricavare la funzione di log-verosimiglianza

$$l(\lambda) = \log L(\lambda) = n(\bar{u}) \log(\lambda) - \lambda|R|.$$

La funzione punteggio è pari a

$$l_*(\lambda) = \frac{d}{d\lambda} l(\lambda) = \frac{n(\bar{u})}{\lambda} - |R|.$$

Ponendo $l_*(\lambda) = 0$ si ricava lo stimatore di massima verosimiglianza

$$\hat{\lambda} = n(\bar{u})/|R|.$$

Per un processo disomogeneo, la funzione di verosimiglianza è definita in funzione di θ e verrà indicata con $\lambda(u; \theta)$ la funzione di intensità dato θ .

Sia $\bar{u} = \{u_1, \dots, u_n\}$ l'insieme di punti osservati e $n = n(\bar{u})$ il numero di punti osservati in $|R|$, allora la verosimiglianza è pari a

$$L(\theta) = L(\theta; \bar{u}) = \prod_{i=1}^n \lambda(u_i; \theta) \exp\left(\int_R -\lambda(u; \theta) du\right)$$

e la logverosimiglianza risulta

$$l(\theta) \propto \sum_{i=1}^n \log(\lambda(u_i; \theta)) - \int_R \lambda(u; \theta) du.$$

Data l'ampia flessibilità, la maggior parte dei processi di Poisson adottati nella pratica si basano sull'utilizzo dei modelli loglineari definiti in equazione (2.1). Sotto questa ipotesi

$$l(\theta) = \sum_{i=1}^n b(u_i) + \theta^T \sum_{i=1}^n Z(u_i) - \int_R \exp(b(u) + \theta^T Z(u)) du \quad (2.2)$$

è funzione concava e differenziabile rispetto a θ , anche nel caso in cui $b(\cdot)$ e $Z_j(\cdot)$ siano funzioni non continue. Se la matrice

$$M = \int_R Z(u) Z(u)^T du$$

è definita positiva, allora il modello è identificabile. Se $\sum_{i=1}^n Z_j(u_i) \neq 0, \forall j$, esiste la stima di massima verosimiglianza, è unica e la stima di θ si ottiene massimizzando la logverosimiglianza, ovvero risolvendo in θ l'equazione

$$l_*(\theta) = 0.$$

La funzione score è un vettore formato dagli elementi $l_*(\theta) = (l_*^{(1)}\theta, \dots, l_*^{(p)}\theta)^T$, dove p è il numero di covariate spaziali associate al modello, e tali per cui

$$l_*^{(j)}(\theta) = \sum_{i=1}^n Z_j(u_i) - \int_R Z_j(u) \lambda(u; \theta) du = 0.$$

Essendo il modello regolare, e ponendo $\hat{\theta}$ come stimatore di massima verosimiglianza di θ , allora valgono le usuali proprietà degli stimatori di massima verosimiglianza:

- $\hat{\theta} \xrightarrow{P} \theta$, e di conseguenza è stimatore consistente;
- $\hat{\theta}$ si distribuisce approssimativamente come una normale p -dimensionale di media θ e varianza pari al reciproco dell'informazione attesa di Fisher $I^{-1}(\theta)$, dove

$$I(\theta) = E [l_*(\theta; X) l_*(\theta; X)^T] = E \left[-\frac{d}{d\theta} l_*(\theta; X) \right];$$

- è possibile effettuare test per verifiche di ipotesi utilizzando gli strumenti noti dalla teoria della verosimiglianza, quali, per esempio, il test log-rapporto di verosimiglianza e il test alla Wald.

Sotto ipotesi di un processo Poisson loglineare, l'informazione osservata di Fisher è pari a

$$J(\theta) = \int_R Z(u) Z(u)^T \lambda(u; \theta) du,$$

che non dipende da u . Di conseguenza $J(\theta) = I(\theta)$. La matrice J contiene elementi $J(\theta)_{[i,j]} = \int_R Z_i(u) Z_j(u) \lambda(u; \theta) du$.

2.1.3 Quadratura numerica

Nella funzione di logverosimiglianza in equazione (2.2), la componente

$$\int_R \exp(b(u) + \theta^T Z(u)) du$$

non può essere calcolata analiticamente, di conseguenza è opportuno definire una procedura di stima numerica utile a calcolare $l(\theta)$. La strategia adottata in contesti spaziali è quella di definire un'approssimazione della complessa verosimiglianza del modello in equazione (2.1), riconducendola alla più semplice log-verosimiglianza del modello di regressione Poisson per dati di conteggio, in quanto più semplice da trattare e stimare. L'approssimazione si basa sull'utilizzo della quadratura numerica.

Per quadratura numerica si intende un efficiente metodo di calcolo numerico di un integrale di interesse. Data una generica funzione $f(u)$, allora vale

$$\int_R f(u)du \doteq \sum_j w_j f(u_j),$$

ovvero la quantità di interesse si approssima nella somma finita di pesi di quadratura w_j moltiplicati per i punti di quadratura $f(u_j)$. Berman e Turner [5] svilupparono una quadratura numerica opportunamente applicabile per la stima approssimata della verosimiglianza di un processo di Poisson disomogeneo. Si approssima l'integrale

$$\int_R \lambda(u; \theta)du \approx \sum_{j=1}^m \lambda(u_j; \theta)w_j,$$

dove $\lambda(u_j; \theta)$ sono i punti di quadratura, e $w_j > 0$ sono pesi di quadratura tali per cui $\sum_{j=1}^m w_j = |R|$.

Di conseguenza

$$l(\theta) \propto \sum_{i=1}^n \log(\lambda(u_i; \theta)) - \int_R \lambda(u_i; \theta)du \approx \sum_{i=1}^n \log(\lambda(u_i; \theta)) - \sum_{j=1}^m \lambda(u_j; \theta)w_j.$$

Se l'insieme dei punti di quadratura comprendesse anche i punti osservati, allora

$$l(\theta) \propto \sum_{i=1}^n \log(\lambda(u_i; \theta)) - \int_R \lambda(u_i; \theta)du \approx \sum_{j=1}^m (\mathbb{1}_j \log \lambda_j - w_j \lambda_j), \quad (2.3)$$

dove $\lambda_j = \lambda(u_j, \theta)$, e

$$\mathbb{1}_j = \begin{cases} 1 & \text{se } u_j \text{ è un punto osservato;} \\ 0 & \text{se } u_j \text{ non è un punto osservato, detto } \textit{dummy point}. \end{cases}$$

È possibile riformulare l'equazione (2.3) in

$$l(\theta) \approx \sum_{j=1}^m (y_j \log \lambda_j - \lambda_j) w_j \quad (2.4)$$

dove $y_j = \mathbb{1}_j/w_j$. La verosimiglianza approssimata riportata in equazione (2.4) coincide con la verosimiglianza pesata per osservazioni indipendenti, dove $Y_j \sim \text{Poisson}(\lambda_j)$, con $Y_i \perp\!\!\!\perp Y_j \forall i \neq j$ e con pesi w_j . Di conseguenza, è possibile seguire la seguente procedura per la stima dei parametri che regolano il processo di Poisson disomogeneo:

- 1) si generano casualmente *dummy points* e si considera l'insieme formato da questi e i dati osservati;
- 2) si definiscono i pesi di quadratura w_j ;
- 3) vengono calcolati $y_j = \mathbb{1}_j/w_j$;
- 4) viene definito il vettore $z_j = Z(u_j)$ in riferimento alle covariate;
- 5) si richiama la funzione per la stima di un modello lineare generalizzato, specificando un modello Poisson loglineare dove $\log(\lambda_j) = \theta^T z_j$, ponendo y_j come risposta e z_j come vettore di covariate; la stima si ottiene via massima verosimiglianza pesata con pesi w_j , ottenendo così θ .

Usando questo approccio è possibile stimare via massima verosimiglianza θ , ottenere i relativi standard error e il valore della logverosimiglianza calcolata in $\hat{\theta}$. Pertanto è possibile utilizzare procedure note per la selezione dei parametri che governano la dinamica del processo, quali per esempio selezione stepwise via AIC o BIC.

Utilizzando lo stesso approccio è possibile stimare attraverso l'algoritmo di Local Scoring l'intensità $\lambda(u)$. Definendo un opportuno lisciatore, è possibile stimare un modello GAM per dati di conteggio tale per cui

$$\log(\lambda(u)) = b(u) + \sum_{j=1}^p f(Z_j(u))$$

e utilizzabile in un contesto di processo di punto spaziale.

Da un punto di vista computazionale, è conveniente utilizzare una partizione in regioni di uguale dimensioni per definire i pesi w_j . Dividiamo R in sotto-regioni S_j di uguale dimensione, ovvero $|S_j| = |S| < |R| \forall j$. Allora $w_j = |S|/k_j$, dove k_j è

pari al numero di punti di quadratura contenuti in S_j . Un esempio di quadratura è riportato in Figura 2.1.

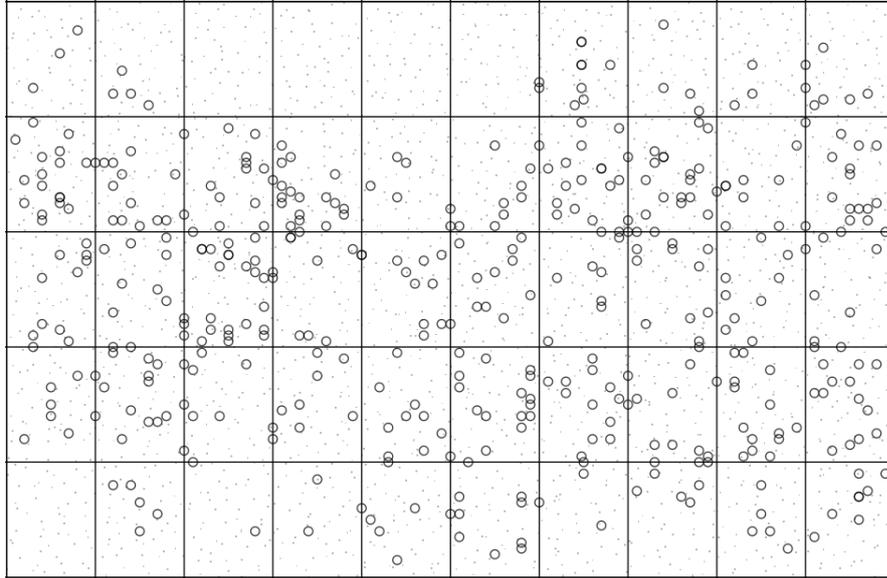


Figura 2.1: Quadratura rettangolare del campo da calcio: i punti pieni sono dummy points generati casualmente, punti vuoti eventi osservati

2.1.4 Applicazione e bontà di adattamento

Considerando ancora i passaggi in partenza del Milan, vengono stimati processi di Poisson disomogenei. Non avendo a disposizione covariate spaziali che possano influire sull'intensità dei passaggi, si stima il processo di Poisson via massima verosimiglianza pesata e mediante l'algoritmo local scoring (da qui in avanti indicati rispettivamente con GLM e GAM), per poi verificare la bontà di adattamento del modello.

Si assuma che i passaggi in partenza rossoneri siano realizzazioni di un processo di Poisson disomogeneo, dove l'intensità è in funzione delle coordinate spaziali. In prima battuta si adotta un modello polinomiale con

$$\begin{aligned} \log(\lambda(u; \theta)) &= \log(\lambda(x, y; \theta)) = \\ &= \theta_0 + \theta_1 x + \theta_2 y + \theta_3 x^2 + \theta_4 xy + \theta_5 y^2 \\ &+ \theta_6 x^3 + \theta_7 x^2 y + \theta_8 xy^2 + \theta_9 y^3. \end{aligned} \quad (2.5)$$

Utilizzando la procedura di stima precedentemente descritta si ottengono stime, standard error, un intervallo di confidenza 0.95 per il singolo parametro e il p -value della statistica test alla Wald per la significatività del singolo coefficiente riportati in Tabella 2.1.

Parametro	Stima	S.E.	Estr. Inf.	Estr. Sup.	p-value
Intercetta	-5.537244	0.677428	-6.864979	-4.209509	0.0000
x	0.045010	0.024317	-0.002650	0.092670	0.0642
y	0.145520	0.044256	0.058781	0.232260	0.0010
x^2	-0.000204	0.000382	-0.000953	0.000545	0.5931
xy	-0.002013	0.000640	-0.003268	-0.000758	0.0017
y^2	-0.000567	0.001149	-0.002818	0.001685	0.6217
x^3	0.000000	0.000002	-0.000004	0.000004	0.9968
x^2y	0.000007	0.000004	-0.000001	0.000014	0.0950
xy^2	0.000017	0.000006	0.000005	0.000029	0.0055
y^3	-0.000022	0.000011	-0.000043	-0.000002	0.0354

Tabella 2.1: Tabella riassuntiva della stima del modello completo

Osservando attentamente i p -value molti parametri risultano statisticamente non significativi. Viene utilizzato un approccio backward per la selezione dei parametri basato sul criterio di informazione

$$BIC(p) = p \log(n) - 2l(\hat{\theta}; p),$$

dove p è il numero di parametri stimati, n è il numero di eventi osservati in R e $l(\hat{\theta}; p)$ è il valore della funzione di logverosimiglianza calcolato nel punto di massimo [17]. Tale procedura mantiene solo 5 variabili all'interno del modello, riportate in Tabella (2.2).

Parametro	Stima	S.E.	Estr. Inf.	Estr. Sup.	p-value
Intercetta	-4.884237	0.368234	-5.605963	-4.162512	0.0000
x	0.020648	0.006015	0.008858	0.032437	0.0006
y	0.110673	0.015902	0.079506	0.141841	0.0000
xy	-0.001169	0.000380	-0.001913	-0.000424	0.0021
xy^2	0.000016	0.000006	0.000004	0.000027	0.0063
y^3	-0.000027	0.000004	-0.000034	-0.000019	0.0000

Tabella 2.2: Tabella riassuntiva della stima del modello ridotto, la selezione dei parametri è avvenuta mediante l'uso di una procedura step backward basata sul criterio di informazione di Schwarz

Per quanto riguarda il modello GAM, sia

$$\log(\lambda(u); \theta) = \theta_0 + f_1(x; \theta) + f_2(y; \theta) + f_3(xy; \theta). \quad (2.6)$$

Il modello in questione presenta vari parametri di regolazione: il tipo di lisciatore adottato, il numero di nodi e il grado dei polinomio. Utilizzando splines cubiche naturali, in cui i nodi interni sono pari ai quartili delle variabili x , y , xy rispettivamente, si ottiene una stima dell'intensità che sembra, a primo impatto, rappresentare meglio la dinamica del processo. In termini di BIC si ottiene 3233.642 per il modello GAM contro 3252.225 per il modello GLM. GAM definisce un trend spaziale più coerente con il processo di interesse, definendo però cluster differenti da quelli ottenuti mediante stimatore basato sul metodo Kernel. In realtà, osservando le intensità stimate dai due modelli precedentemente definiti e confrontandoli con le intensità stimate mediante metodo Kernel e partizione in celle (Figura 2.2), si evince chiaramente che il modello loglineare non si adatta bene ai dati. Questo modello, infatti, cattura bene solo il trend spaziale, inteso come la sezione della regione R in cui si osserva con frequenza maggiore l'evento di interesse. Infatti le partenze dei passaggi dei giocatori rossoneri avvengono principalmente nella fascia centrale del campo, ma raramente nelle fasce esterne. Ciò che i modelli definiti finora non sono in grado di catturare è la presenza di dipendenza tra i punti, colta solo dallo stimatore non parametrico. Questo è causato dalla forma funzionale eccessivamente semplice, e dal tipo di dipendenza imposta.

La formazione di cluster è dettata dal fatto che i passaggi partono principalmente da giocatori chiave all'interno della formazione del Milan. Esiste quindi una nota dipendenza fra i punti, che i processi disomogenei non sono in grado di cogliere. Oltre all'evidenza grafica, è possibile effettuare il test χ^2 bontà di adattamento per il processo Poisson disomogeneo definito in equazione (2.5) e in equazione (2.6). Dividendo R in m sotto-regioni B_1, \dots, B_m , è possibile stimare il numero atteso di eventi per regione μ_j ponendo:

$$\hat{\mu}_j = \int_{B_j} \hat{\lambda}(u; \hat{\theta}) du.$$

Si dimostra che, sotto ipotesi nulla

$$X^2 = \sum_{j=1}^m \frac{(n_j - \hat{\mu}_j)^2}{\hat{\mu}_j} \sim \chi^2(m - p).$$

Sia per il modello loglineare con selezione delle variabili via procedura step che per il modello GAM si rifiuta l'ipotesi nulla ($\alpha^{oss} \doteq 0$ per entrambi i test, utilizzando diverse partizioni del campo). I modelli ancora non si adattano bene ai dati.

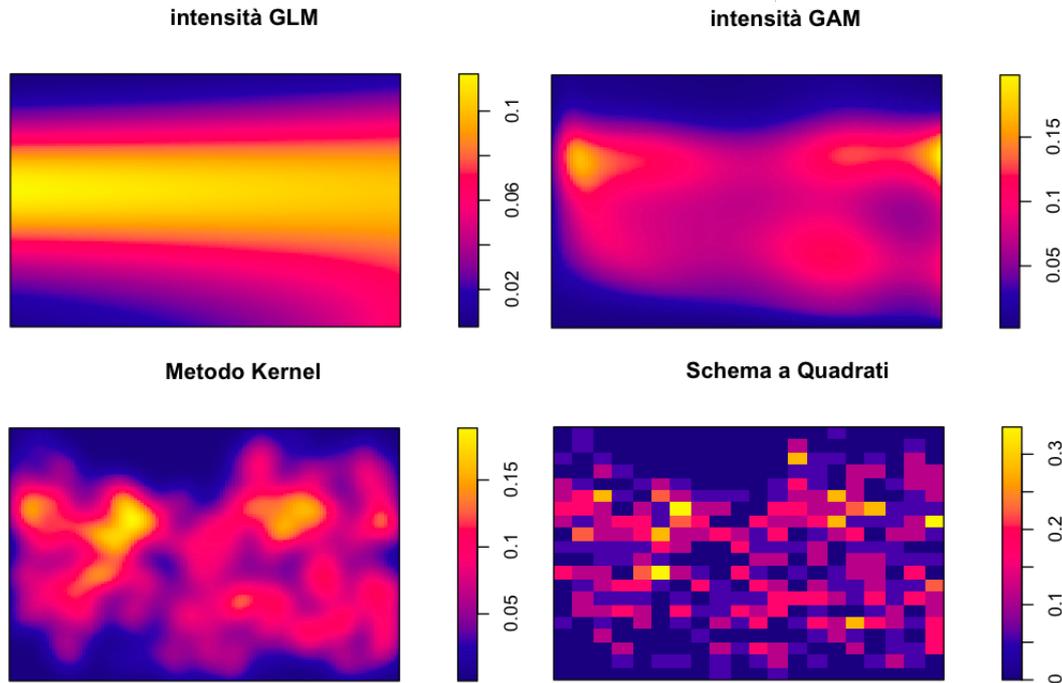


Figura 2.2: Confronto stime intensità: 4 approcci, 2 parametrici (GLM e GAM) e due non parametrici (kernel e schema a quadrati)

2.2 Analisi dei residui

2.2.1 Introduzione

Un altro strumento di diagnosi della bontà di adattamento del modello consiste nell'analisi grafica dei residui [2]. Si ponga

$$r(u) = \lambda(u) / \hat{\lambda}(u; \hat{\theta}),$$

dove $\lambda(u)$ rappresenta la vera intensità del processo e $\hat{\lambda}(u; \hat{\theta})$ la stima dell'intensità del processo. Se le due misure sono in accordo, i valori di $r(u)$ sono prossimi a 1. Non conoscendo la vera intensità, è possibile stimare $r(u)$ ponendo

$$\hat{r}(u) = \tilde{\lambda}_U(u) / \hat{\lambda}(u; \hat{\theta}),$$

ovvero il rapporto della stima dell'intensità non parametrica uniformemente corretta e l'intensità stimata dal modello.

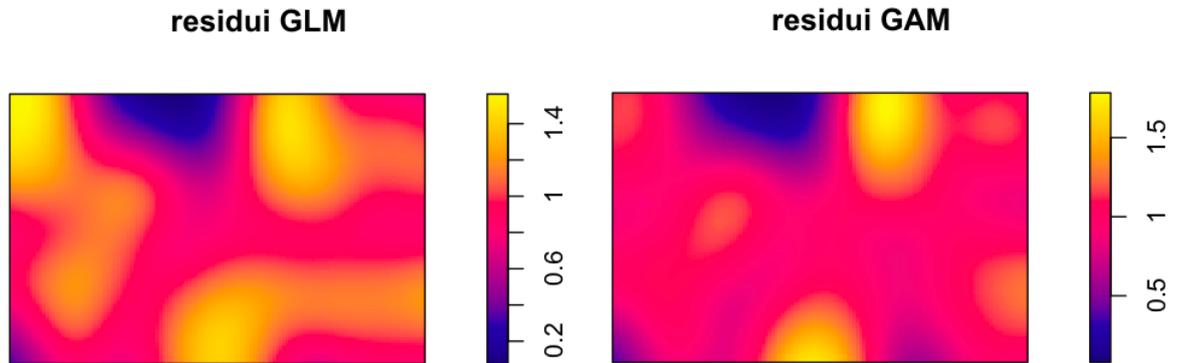


Figura 2.3: Rappresentazione grafica dei residui $\hat{r}(u)$

Se i modelli risultassero correttamente specificati, assumendo che la vera intensità sia pari alla stima dell'intensità non parametrica adottata, il modello restituirebbe grafici diversi da quelli riportati in Figura 2.3. Questo perché i residui dovrebbero avere tutti lo stesso valore (prossimo a 1). Invece, il range osservato per i residui è pari a $[0.07, 1.56]$ per il modello GLM, e $[0.08, 1.78]$ per il modello GAM. Si evince chiaramente una performance migliore in termini dei residui per il modello GAM, in quanto maggiormente presente il colore rosso, che rappresenta un valore del rapporto pari a 1. Tuttavia in molte sezioni del campo il modello tende a sovrastimare l'intensità (regioni in blu) mentre in altre tende a sottostimarle (regioni in giallo).

2.2.2 Residuo del processo di punto

La grave lacuna dei residui appena descritti è l'assunzione che la vera intensità sia pari a quella definita mediante il metodo kernel. Per ovviare al problema, si consideri lo scostamento tra il numero previsto di eventi in una sotto-regione di R e il numero di eventi osservati nella stessa sotto-regione, per poi cercare una opportuna rappresentazione grafica. In generale

$$r(B) = n(\bar{u} \cap B) - \int_B \hat{\lambda}(u; \hat{\theta}) du,$$

dove $r(B)$ è chiamato **residuo del processo di punto**. Si fa notare che $r(B)$ non richiede la definizione specifica dello spazio in sotto-regioni: $r(\cdot)$ è una misura spaziale, che cambia al variare di B , e ci permette di definire gli scostamenti di piccola e larga scala. La funzione $r(\cdot)$ definisce la distribuzione della massa spaziale, dove per massa si intende lo scostamento tra osservato e atteso. In

Figura 2.4 viene riportata l'analisi grafica dei residui al quadrato dei due modelli, dove il campo è stato diviso in celle di area 1 m^2 . E' possibile osservare come i valori dei residui siano tendenzialmente prossimi a zero, tuttavia si nota un trend spaziale in fascia centrale. Anche in questo caso si dubita della bontà di adattamento dei due modelli.

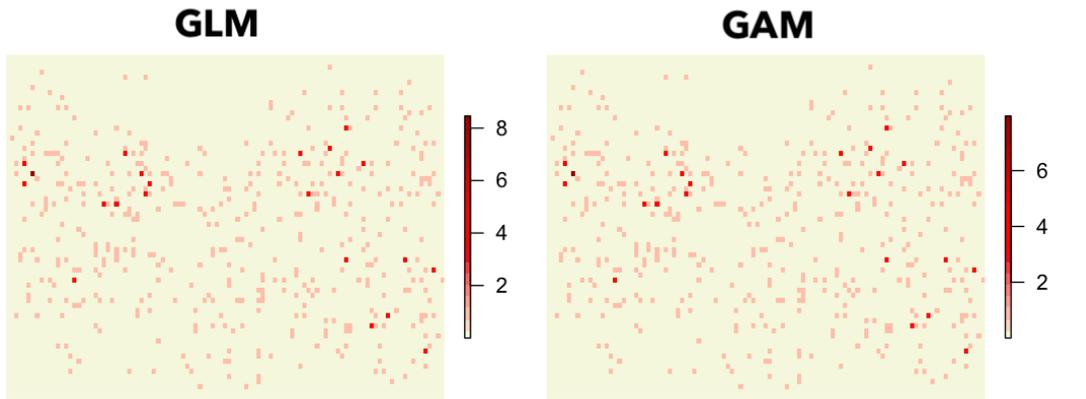


Figura 2.4: Analisi dei residui al quadrato: GLM a sinistra, GAM a destra.

2.3 Correlazione tra punti

In Figura 2.5 vengono rappresentati 3 forme di correlazione tra i punti. Si parla di regolarità quando i punti nello spazio sono distribuiti in modo equi spaziale all'interno di R , di cluster spaziale quando è ben visibile la presenza di un numero rilevante di punti in specifiche sotto-regioni di R e di indipendenza nel caso in cui i punti siano posizionati in modo casuale all'interno della regione di studio. Uno strumento utile per identificare il tipo di dipendenza tra i punti con-

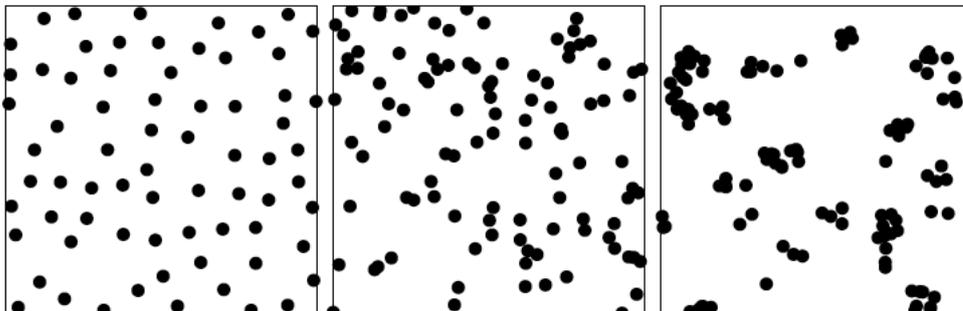


Figura 2.5: Esempio di dipendenza spaziale: in presenza di regolarità (a sinistra), sotto indipendenza (in centro) e in presenza di cluster spaziali (a destra)

siste nell'analisi della correlazione residua. Per correlazione residua si intende la dipendenza tra punti realizzazione di uno stesso processo non spiegata da $\hat{\lambda}(u)$,

intensità stimata e associata ad un generico processo stocastico (per esempio via GLM o GAM). Di norma, se la correlazione residua tra i punti è superiore ad una certa soglia si ha una situazione di cluster non spiegato dal modello, se inferiore si parla di regolarità non spiegata, se pari alla soglia il modello si adatta bene ai dati e tiene conto della dipendenza, se esistente, tra i punti.

Il concetto di correlazione nei processi di punto è simile al concetto classico di correlazione in statistica. Tuttavia, quando si parla di correlazione tra punti, si intende una misura in grado di descrivere la presenza di un effetto di attrazione o repulsione tra le coppie di punti (u_i, u_j) con $i \neq j$.

2.4 Momenti dei processi di punto

I processi di punto godono di due importanti proprietà, utili alla definizione del concetto di correlazione tra punti. Sia du una regione di area infinitesimale che contiene il punto u e, per alleggerire la notazione, $N(B) = n(\bar{U} \cap B)$. Si parla di proprietà del primo ordine quando

$$\lambda(u) = \lim_{|du| \rightarrow 0} \frac{E[N(du)]}{|du|}.$$

Dati due punti, u e v , per proprietà del secondo ordine si intende

$$\lambda_2(u, v) = \lim_{|du|, |dv| \rightarrow 0} \frac{E[N(du)N(dv)]}{|du||dv|}.$$

La funzione $\lambda_2(\cdot, \cdot)$ è detta intensità prodotto. Per la proprietà del secondo ordine, date due sotto-regioni disgiunte A e B allora

$$E[N(A)N(B)] = \int_A \int_B \lambda_2(u, v) dudv.$$

La quantità $\lambda_c(u|v)$ è detta intensità condizionata, dove

$$\lambda_c(u|v) = \frac{\lambda_2(u, v)}{\lambda(v)}.$$

Il processo di punto è detto stazionario in media se $\lambda(u) = \lambda \forall u \in R$, mentre si parla di stazionarietà nel prodotto se

$$\lambda_2(u, v) = \lambda_2(s),$$

dove s rappresenta una misura della distanza spaziale tra u e v . Si parla di processo stazionario nel prodotto e isotropico se $s = \|u - v\|$. Dati A e B t.c. $A \cap B = \emptyset$

$$\text{Cov}[N(A), N(B)] = \int_A \int_B \lambda_2(u, v) dudv - \int_A \lambda(u) du \int_B \lambda(v) dv.$$

Per i processi Poisson disomogenei

$$\lambda_2(u, v) = \lambda(u)\lambda(v)$$

e, di conseguenza

$$\text{Cov}[N(A), N(B)] = \int_A \int_B \lambda_2(u, v) dudv - \int_A \lambda(u) du \int_B \lambda(v) dv \quad (2.7)$$

$$= \int_A \int_B \lambda(u)\lambda(v) dudv - \int_A \lambda(u) du \int_B \lambda(v) dv \quad (2.8)$$

$$= 0. \quad (2.9)$$

Pertanto la correlazione tra il numero di eventi osservati in A e il numero di eventi osservati in B risulta nulla.

2.5 Pair correlation function

2.5.1 Definizione

Uno strumento più sofisticato per l'analisi della correlazione residua, proposto da Baddeley, Møller and Waagepetersen [3], prende il nome di *pair correlation function* (PCF), dove

$$\rho(s) = \frac{\lambda_2(u, v)}{\lambda(u)\lambda(v)}$$

in funzione della distanza euclidea tra u e v . Mentre $\lambda(u)$ definisce l'intensità spaziale media in funzione della posizione spaziale, $\rho(s)$ può essere interpretata come la variazione spaziale attorno alla media locale.

La funzione $\rho(\cdot)$ è legata ad una seconda quantità, detta funzione $K(\cdot)$. Sia $N_0(s)$ il numero di eventi posizionati a distanza minore o uguale ad s da un evento arbitrario, e si assuma che \bar{U} sia un processo stazionario in media. Allora

$$K(s) = \lambda^{-1} E[N_0(s)].$$

Grazie all'assunzione di assenza di punti coincidenti è possibile definire il legame tra $K(\cdot)$ e $\rho(\cdot)$, in quanto è possibile calcolare $E[N_0(s)]$ come l'integrale dell'intensità condizionata, considerando come sotto-regione un cerchio centrato sull'origine o e di raggio s . Quindi

$$\lambda K(s) = \int_0^{2\pi} \int_0^s \lambda_c(u|o) \partial u \partial \theta.$$

Nel caso in cui il processo risulti stazionario nel prodotto, allora $\lambda_c(u|o) = \lambda_2(u)/\lambda$, ottenendo così

$$\lambda K(s) = 2\pi \lambda^{-1} \int_0^s \lambda_2(u) \partial u. \quad (2.10)$$

Dall'equazione (2.10) si ricava che

$$\lambda_2(s) = \lambda^2 (2\pi s)^{-1} K'(s),$$

dove $K'(s)$ rappresenta la derivata prima di $K(s)$. Il legame tra $\rho(\cdot)$ e $K(\cdot)$ è definito dall'equazione

$$\rho(s) = (2\pi s)^{-1} K'(s).$$

Baddeley, Møller and Waagepetersen [3] dimostrarono che

$$K(s) = 2\pi \int_0^s \rho(s) \partial s \quad (2.11)$$

senza l'imposizione della stazionarietà in media. Per un processo Poisson disomogeneo

$$\rho(s) = \frac{\lambda_2(u, v)}{\lambda(u)\lambda(v)} = \frac{\lambda(u)\lambda(v)}{\lambda(u)\lambda(v)} = 1,$$

e si dimostra che $K(s) = \pi s^2$.

2.5.2 Stima Empirica

Stimare direttamente $\rho(\cdot)$ non è possibile. Conviene quindi conoscere un'opportuna stima di $K(\cdot)$, utilizzata in seguito per la stima di $\rho(\cdot)$. Una stimatore naturale di $E[N_0(s)]$ è

$$\tilde{E}[N_0(s)] = n^{-1} \sum_{i=1}^n \sum_{i \neq j} \mathbb{1}(s_{ij} \leq s)$$

dove, dati due punti u_i e u_j , $s_{ij} = \|u_i - u_j\|$. $\tilde{E}[N_0(s)]$, tuttavia, è stimatore distorto, quindi Ripley [20] propose lo stimatore

$$\hat{E}[N_0(s)] = n^{-1} \sum_{i=1}^n \sum_{i \neq j} w_{ij}^{-1} \mathbb{1}(s_{ij} \leq s),$$

con w_{ij} opportuni pesi di correzione. Sotto stazionarietà in media, allora $\hat{\lambda} = n(\bar{u})/|R|$. Si ottiene quindi

$$\hat{K}(s) = \hat{\lambda}^{-1} \hat{E}[N_0(s)].$$

Tuttavia, non sono di interesse i casi in cui il processo non risulti stazionario in media. Si assuma noto $\lambda(u)$. Si dimostra che

$$\hat{K}_I(s) = |R|^{-1} \sum_{i=1}^n \sum_{j \neq i} w_{ij}^{-1} \frac{\mathbb{1}(s_{ij} \leq s)}{\lambda(u_i)\lambda(u_j)}.$$

Basandoci sull'equazione (2.11), un'opportuno stimatore per $\rho(\cdot)$ è pari a

$$\hat{\rho}(s) = \frac{\hat{K}'_I(s)}{2\pi s}.$$

Essendo $\lambda(u)$ ignoto, questa quantità può essere sostituita con qualsiasi stima di $\hat{\lambda}(u)$. Per ottenere $\hat{\rho}(\cdot)$ è comunque necessario stimare $\hat{K}'_I(\cdot)$. Basta utilizzare un opportuno lisciatore (splines) applicato a $\hat{K}_D(s)$, da cui poi è semplice ricavare la sua derivata prima.

2.5.3 Interpretazione di $\rho(\cdot)$

La funzione $\rho(\cdot)$ è uno strumento utile al fine di comprendere le lacune associate ai processi disomogenei.

Sia $\hat{\lambda}(u)$ la stima della funzione d'intensità associata ad un generico processo disomogeneo. Se $\hat{\rho}(s) > 1$, allora vi è presenza di cluster spaziali che il processo disomogeneo non è in grado di cogliere. Al contrario, se $\hat{\rho}(s) < 1$, allora vi è regolarità residua non considerata dal modello. Se il modello si adatta bene ai dati, allora $\hat{\rho}(s) \doteq 1$.

Riprendendo l'esempio dei passaggi in partenza rossoneri, si stima $\rho(\cdot)$ utilizzando la procedura descritta nella sezione 2.5.2 e ponendo come stima della funzione intensità quella ottenuta e discussa nella sezione 2.1.4. Come si evince dalla Figu-

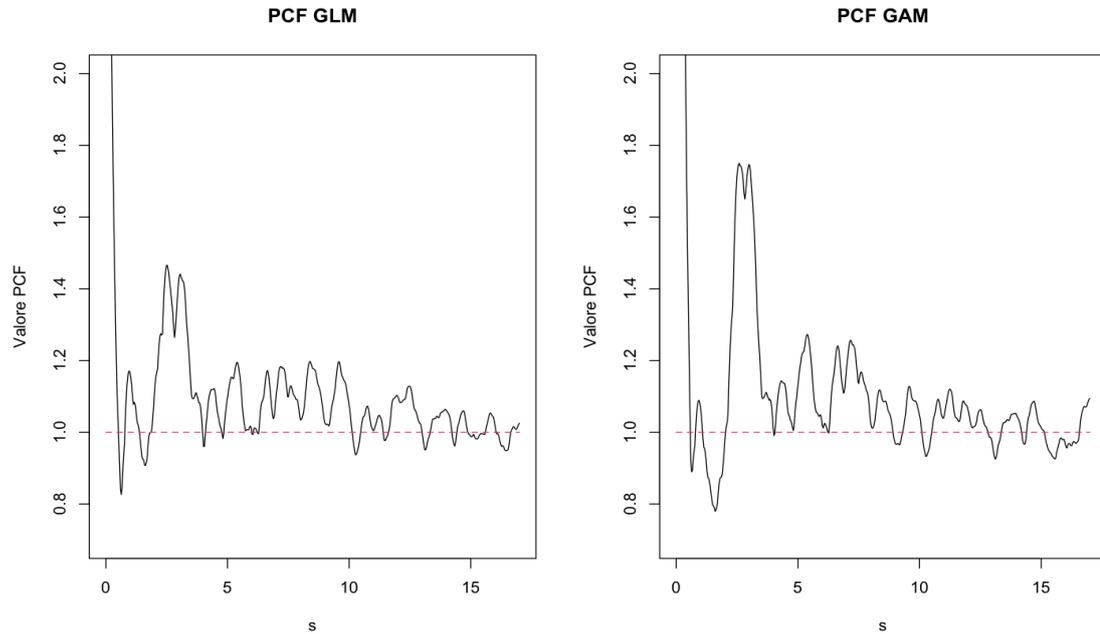


Figura 2.6: Funzione $\rho(\cdot)$ stimata utilizzando i passaggi in partenza rossoneri: a sinistra la stima di $\lambda(\cdot)$ si basa sulla stima GLM, a destra stima GAM

ra 2.6, la funzione $\hat{\rho}(s)$ associata ai processi disomogenei presenta valori maggiori di 1 per $s \in [2, 10]$. Esiste una forma di attrazione tra coppie di punti posizionati a distanza tra 2 e 10 metri, che non viene colta dai processi PPD stimati. Questo è legato a caratteristiche note del fenomeno studiato in quanto, durante il corso di una partita, i passaggi partono dalla stessa rosa di 11 giocatori, per ognuno dei quali è stato assegnato una specifica zona del campo. I giocatori rossoneri rappresentano quindi una dinamica latente non considerata dai processi PPD, che porta ad osservare cluster di punti in sotto-regioni di R .

Questa analisi della correlazione residua conferma la presenza di cluster non identificati da modelli PPD. È necessario considerare processi in grado di cogliere la dinamica spaziale latente (i giocatori) e pertanto si introduce il processo di Cox.

Capitolo 3

Processi di Cox

Si evince chiaramente dalle diagnostiche effettuate mediante test, analisi dei residui e PCF che il processo Poisson disomogeneo non si adatta bene ai dati. Questo accade perché esiste dipendenza tra i punti del processo che non viene considerata dai modelli disomogenei, e non può essere trascurata. In questo capitolo viene presentato il modello di Cox, al fine di stimare la funzione di intensità che consideri la presenza di cluster, e successivamente viene applicato per descrivere le tattiche di gioco adottate durante il derby Milan-Inter.

3.1 Processi di Cox

3.1.1 Eterogeneità spaziale

Per processo di punto di Cox si intende un processo di Poisson la cui intensità è variabile aleatoria (se costante nello spazio) o processo stocastico. In questo contesto, si dice che il processo stocastico $\Lambda(u)$ *guida* il processo \bar{U} generatore dei punti. Qualsiasi rappresentazione adottata per il processo $\Lambda(u)$ deve godere delle seguenti proprietà: i valori di $\Lambda(u)$ devono essere non negativi e ogni realizzazione del processo stocastico $\Lambda(u)$ deve essere integrabile su insiemi limitati (la regione e le sue possibili sotto-regioni).

Come descritto in precedenza, la scelta di adottare il processo di Cox è dettata dal fatto che il processo di Poisson cattura parzialmente l'eterogeneità ambientale del processo di punto. Si consideri il modello più semplice associato al processo di Cox, il Processo Poisson Misto [14]. L'intensità è costante nello spazio ma

variabile aleatoria indicata con Λ . Se Λ fosse noto, allora

$$E[n(\bar{U} \cap B)|\Lambda] = \Lambda|B|$$

e, di conseguenza,

$$E[n(\bar{U} \cap B)] = E(E[n(\bar{U} \cap B)|\Lambda]) = E(\Lambda|B|) = E(\Lambda)|B|.$$

Il processo Poisson misto è un processo di Poisson omogeneo dove

$$n(\bar{U} \cap B) \sim \text{Poisson}(E(\Lambda)|B|),$$

e dove l'intensità $\lambda = E[\Lambda]$. Mentre però il processo omogeneo presenta varianza pari alla media, ovvero $V[n(\bar{U} \cap B)] = E[n(\bar{U} \cap B)] = \lambda$, la varianza di un processo Poisson Misto è pari a

$$\begin{aligned} \text{Var}[n(\bar{U} \cap B)] &= E(\text{Var}[n(\bar{U} \cap B)|\Lambda]) + \text{Var}(E[n(\bar{U} \cap B)|\Lambda]) \\ &= E(\Lambda)|B| + \text{Var}(\Lambda)|B|^2. \end{aligned} \quad (3.1)$$

Il termine $E(\Lambda)|B|$ rappresenta la porzione variabilità associata all'assunzione distributiva Poisson e, a differenza dei processi omogenei, è presente il secondo termine $\text{Var}(\Lambda)|B|^2$ che racchiude la variabilità associata a Λ . Questa quantità si traduce nella variazione non spiegata in media, e di cui si vuole tener conto durante la stima dell'intensità.

3.1.2 Processi Log-Gaussiani di Cox

Il caso più interessante e realistico consiste nel definire il processo stocastico $\{\Lambda(u) : u \in \mathbb{R}^2\}$, ovvero assumendo che $\Lambda(u)$ sia stocastico. $\Lambda(u)$ è detta funzione d'intensità stocastica. Come detto in precedenza, i valori di $\Lambda(u)$ devono essere non negativi e quindi lo stesso Cox propose, ispirato dal legame canonico del modello lineare generalizzato per dati di conteggio, di considerare la trasformata logaritmica del processo stocastico

$$\{G(u) = \log(\Lambda(u)) : u \in \mathbb{R}^2\},$$

assicurandoci così che $\Lambda(u) = \exp[G(u)]$ sia sempre positivo. Questo processo prende il nome di processo Log-Gaussiano di Cox (LGCP) [16].

In molte applicazioni ha senso imporre che $G(u)$ abbia distribuzione gaussiana

per ogni $u \in R$, in quanto risulta semplice definire la correlazione spaziale tra u e v attraverso la matrice di varianza e covarianza di $(G(u), G(v))^T$. Con i dati a disposizione, per ogni set di punti $\bar{u} = \{u_1, \dots, u_n\}$ il corrispondente vettore $(G(u_1), \dots, G(u_n))^T$ presenta a sua volta distribuzione gaussiana n -dimensionale. In generale, una distribuzione gaussiana che gode di queste proprietà è detta *Gaussian Random Field* (GF), la cui matrice di varianza e covarianza descrive la dipendenza spaziale tra le coppie di punti.

Il vantaggio dell'uso del GF è la completa specificazione del processo stocastico associato se si specificano momento primo e secondo, e che gode di proprietà simili a quelle della semplice distribuzione gaussiana.

Un Gaussian Random Field è unicamente identificato e completamente specificato se noti

$$\begin{aligned}\mu(u) &= E[G(u)], \\ C(u, v) &= Cov[G(u), G(v)] = E[G(u)G(v)] - \mu(u)\mu(v),\end{aligned}\tag{3.2}$$

detti rispettivamente funzione media e funzione covarianza. Sia $G_0(u)$ un campo gaussiano a media nulla, tale per cui $\mu(u) = 0 \forall u \in R$ e sia $C_0(u, v)$ la funzione di covarianza associata. Allora, dati $a(u)$ e $b(u)$ due funzioni deterministiche,

$$G(u) = a(u)G_0(u) + b(u)$$

a media $\mu(u) = b(u)$ e funzione di covarianza $C(u, v) = a(u)a(v)C_0(u, v)$. Questo ci permette quindi di costruire modelli per processi logaritmici di Cox disomogenei a partire dalla definizione di $G_0(u)$. Tuttavia la selezione di $b(u)$ e di $C(u, v)$ deve soddisfare alcune condizioni. In particolare $C(u, v)$ deve essere matrice definita positiva.

Si noti che, per un fissato valore di u , se $G(u)$ ha distribuzione gaussiana simmetrica attorno alla sua media, allora l'intensità $\Lambda(u) = \exp[G(u)]$ ha distribuzione log-gaussiana, distribuzione asimmetrica e con la coda destra pesante. Di conseguenza, il processo log-gaussiano di Cox porterà ad una discriminazione nitida tra zone ad alta intensità e zone a bassa intensità.

3.2 Stima di LGCP

3.2.1 Discretizzazione

Il modello LGCP può essere stimato attraverso la discretizzazione del processo [11, 15], basata su una partizione del campo in celle.

Sia R diviso in $n_1 \times n_2 = N$ celle, indicate con $\{S_{ij}\}$, per $i = 1, \dots, n_1$ e $j = 1, \dots, n_2$. Per processi Log-Gaussiani, il numero di eventi attesi nella cella s_{ij} è pari a

$$\Lambda_{ij} = \int_{S_{ij}} \exp[G(u)] du.$$

Come dimostrato da Illian et al. [11] questo integrale può essere approssimato come

$$\Lambda_{ij} \doteq |S_{ij}| \exp(\eta_{ij}).$$

Dato y_{ij} il numero di eventi osservati all'interno della cella S_{ij} , possiamo assumere indipendenza condizionale tra le varie celle e porre

$$y_{ij} | \eta_{ij} \sim \text{Poisson}(|S_{ij}| \exp(\eta_{ij})).$$

Il passaggio fondamentale per l'opportuna definizione di questo processo consiste nello specificare correttamente il predittore lineare η_{ij} . Nella sua forma più generale

$$\eta_{ij} = \beta_0 + \sum_{m=1}^M \beta_m x_{m,ij} + \sum_{k=1}^K f_k(S_{ij}),$$

dove $x_{m,ij}$ rappresenta l' m -esima covariata associata alla cella in posizione ij ed $f_k(S_{ij})$ rappresenta il k -esimo effetto casuale associato alle singole regioni S_{ij} .

Per semplicità, si assuma che lo schema a celle della regione R presenti regioni di uguale area, ponendo $|S_{ij}| = |S|$, $\forall i, j$. Il modello si semplifica e si ottiene

$$\Lambda_{ij} \doteq |S| \exp(\eta_{ij})$$

e, di conseguenza,

$$y_{ij} | \eta_{ij} \sim \text{Poisson}(|S| \exp(\eta_{ij})).$$

Nel seguito dell'elaborato la stima del processo Log-Gaussiano avviene utilizzando un approccio bayesiano. Si considera

$$\eta_{ij} = \beta_0 + f_1(S_{ij}) + f_2(S_{ij}) = \beta_0 + z_{ij} + v_{ij}. \quad (3.3)$$

Si adotta una partizione del campo tale per cui $|S| = 1 \text{ m}^2$ e si assume a priori che z_{ij} abbia distribuzione gaussiana, dove la dipendenza spaziale viene opportunamente strutturata attraverso la matrice di varianza e covarianza del vettore $z = (z_{11}, z_{21}, \dots, z_{n_1 n_2})^T$. In questo elaborato si propone l'utilizzo delle seguenti strutture di covarianza: dipendenza markoviana, dipendenza di natura autoregressiva condizionale e dipendenza isotropica basata sulla funzione di Matern. A priori $v_{ij} \sim N(0, \sigma_v^2)$, dove il logaritmo della precisione $1/\sigma_v^2$ segue distribuzione logGamma con parametro di scala pari a 1 e parametro di forma pari 5^{-5} . Questa componente (v_{ij}) è utile per tenere conto dell'eterogeneità non strutturata e non osservata tra le varie celle. Non è obbligatorio considerare questa componente, anche se Blangiardo et al. [8] consigliano l'utilizzo.

3.2.2 Dipendenza markoviana

La matrice di varianza e covarianza di z , Q^{-1} , descrive la dipendenza spaziale del processo, che si riassume in dipendenza tra celle vicine, e può essere strutturata in diversi modi.

Per semplificare la notazione, assumiamo che ad ognuna delle n celle sia stato assegnato un valore numerico che rappresenta la singola cella, ovvero $\{S_i\}$, con $i = 1, \dots, n$ e con $n = n_1 n_2$ e dove ogni cella ha area pari a $|S|$. Allora

$$y_i | \eta_i \sim \text{Poisson}(|S| \exp(\eta_i))$$

con $\eta_i = \beta_0 + z_i + v_i$.

Si parla di vicini di i , indicato con V_i , l'insieme delle celle direttamente adiacenti a i (detti di primo ordine) e delle celle direttamente adiacenti ai vicini di primo ordine (detti di secondo ordine). Sia z_i il parametro associato all' i -esima area, z_{-i} il vettore di parametri a meno di quello associato all'area i -esima cella, z_{V_i} vettore di parametri associati ai vicini di i .

Allora, assumendo dipendenza di tipo Markoviana, si impone che

$$z_i \perp\!\!\!\perp z_{-i} \mid z_{V_i}$$

La matrice di precisione Q presenterà quindi elementi

$$Q_{ij} = 0 \text{ se } z_i \perp\!\!\!\perp z_j \mid z_{-i,j}$$

ottenendo così una matrice sparsa, che porta ad enormi vantaggi in termini computazionali. Sempre sotto questa specificazione $Q_{ij} \neq 0$ se $j \in \{i, V_i\}$. Di conseguenza, a priori, si assume che $z = (z_1, \dots, z_n)^T$ segua un *Gaussian Markov Random Field* (GMRF) a media nulla e con matrice di precisione pari a $Q = \tau K$, con K detta matrice di costruzione e con elementi

$$K_{ij} = \begin{cases} |V_i| & \text{se } i = j \\ 1 & \text{se } j \in V_i \\ 0 & \text{altrimenti} \end{cases} \quad (3.4)$$

dove $|\cdot|$ indica la cardinalità dell'insieme.

A priori $\log(\tau) \sim \text{logGamma}(1, 5^{-5})$.

3.2.3 Conditional AutoRegressive

È possibile assumere struttura di covarianza tra celle la dipendenza CAR (Conditional AutoRegressive) proposta da Besag (1974). Sia $z = (z_1, \dots, z_n)^T$ il vettore di effetti casuali che definiscono la dipendenza spaziale. Allora

$$z_i | z_{-i} \sim N\left(\mu_i + \sum_{j=1}^n r_{ij}(z_j - \mu_j), s_i^2\right)$$

dove

- μ_i è la media associata all' i -esima area
- $s_i^2 = (\tau|V_i|)^{-1}$ la varianza della stessa area, che dipende dal numero di vicini di i
- r_{ij} indica la vicinanza spaziale ed è calcolata ponendo $r_{ij} = \phi W_{ij}$, $W_{ij} = a_{ij}/V_i$, dove $a_{ij} = 1$ se l'area i e j sono vicini, 0 altrimenti. Il valore di a_{ii} è pari a 0.

Consideriamo W come la matrice degli elementi W_{ij} e sia $S = \text{diag}(s_1, \dots, s_n)$. Per assicurarsi che la matrice di covarianza associata agli effetti casuali $C = (I - \phi W)^{-1} S^2$ sia opportuna, questa deve essere definita positiva. Cressie (1993) dimostrò come C risulti definita positiva se $\phi \in [1/\min(h_i), 1/\max(h_i)]$ dove h_i rappresenta il generico autovalore di W . La specificazione propria quindi della

struttura di correlazione CAR assume

$$z \sim N(\mu, (I - \phi W)^{-1} S^2)$$

con μ vettore delle medie e I matrice identità. Il vettore μ viene posto pari al vettore nullo durante il prosieguo di questo elaborato. La correlazione tra la generica area i e j è pari a

$$\frac{\sqrt{V_i}}{V_j} \frac{(I - \phi W)_{ij}^{-1}}{\sqrt{(I - \phi W)_{ii}^{-1} (I - \phi W)_{jj}^{-1}}}.$$

In termini computazionali può risultare complessa la stima del parametro ϕ a posteriori, di conseguenza in letteratura Besag et al. [7] proposero di porre $\phi = 1$. Questo però pone problemi di natura algebrica, in quanto la matrice di covarianza non risulterebbe mai definita positiva. Per ovviare il problema è sufficiente imporre il vincolo

$$\sum_{i=1}^n z_i = 0.$$

Sotto queste condizioni, allora assumiamo che

$$z_i | z_{-i} \sim N\left(\frac{1}{V_i} \sum_{j=1}^n a_{ij} u_j, s_i^2\right).$$

A priori, $\log(\tau) \sim \log\text{Gamma}(1, 5^{-5})$.

3.2.4 Funzione Whittle-Matern

È possibile assumere che il processo sia isotropico in covarianza, ovvero la covarianza dipenda non tanto dall'esatta posizione dei punti u e v , bensì dalla distanza euclidea dei due punti. Di conseguenza, dato $d = \|u - v\|$, si ha che

$$\text{Cov}(G(u), G(v)) = C(\|u - v\|) = C(d).$$

In aggiunta, sia

$$C(r) = \sigma^2 L(d)$$

con $\sigma^2 \in (0, +\infty)$ parametro di varianza e funzione $L(\cdot)$ tale per cui $L(0) = 1$. La funzione $L(\cdot)$ è detta funzione di correlazione. Nel contesto dei GF, la funzione di correlazione più utilizzata è la funzione di Whittle-Matern. Questa però contiene

al suo interno un'ulteriore parametro ν , detto parametro di forma. La funzione di Matern è pari a

$$L(d) = 2^{1-\nu}\Gamma^{-1}(\nu)(\kappa d)^\nu K_\nu(\kappa d),$$

dove $K_\nu(\cdot)$ è la funzione Bessel modificata del secondo tipo e di ordine ν .

Nel caso di studio, è di interesse la dipendenza tra le varie celle. Si assume $z \sim N(0, C)$ dove C contiene gli elementi

$$[C]_{ij} = Cov(z_i, z_j) = C(d) = \sigma^2 L(d)$$

che dipende dalla distanza d in m^2 e dove:

- $\nu = 1$;
- $\sigma^2 = 1/\tau$ e $\log(\tau) \sim \log\text{Gamma}(1, 5^{-5})$;
- $r = \sqrt{8\nu}/\kappa$ e $\log(r) \sim \log\text{Gamma}(1, 0.01)$;
- d rappresenta la distanza euclidea tra i centri delle celle.

3.2.5 Criterio di informazione: DIC

Data la varietà di possibili strutture di correlazioni adottabili, è necessario definire un criterio di selezione.

Sia $D(\theta)$ la devianza del modello definita come

$$D(\theta) = -2 \log[p(y|\theta)]$$

dove $p(y|\theta)$ rappresenta la verosimiglianza del processo. Tuttavia, in un contesto bayesiano, essendo θ quantità aleatoria, anche $D(\theta)$ è variabile aleatoria, diversamente da quando lo si utilizza in un contesto frequentista. Utilizzando un opportuno indice di sintesi (per esempio la media) allora possiamo ottenere una stima della devianza a posteriori, pari a

$$\bar{D} = E_{\theta|y}[D(\theta)]$$

che si dimostra essere stimatore robusto della devianza a posteriori. Tuttavia, all'aumentare della complessità del modello, la devianza diminuisce. Per questo, similmente a quando si utilizzano criteri quali AIC e BIC, è opportuno introdurre una penalizzazione associata alla complessità del modello.

Spiegelhalter et al. [22] propongono l'utilizzo di DIC (Deviance Information Criterion). Data la natura bayesiana associata al modello, proposero di stimare la complessità del modello definendo una misura concettualmente simile al numero di parametri stimati in contesti frequentisti. Si parla quindi di numero di parametri effettivi pari a

$$p^* = E_{\theta|y}[D(\theta)] - D(E_{\theta|y}(\theta)) = \bar{D} - D(\bar{\theta}).$$

La complessità è pari alla differenza tra il valore atteso della devianza e la devianza calcolata con una opportuna stima dei parametri a posteriori (in questo caso la media a posteriori).

Di conseguenza

$$DIC = \bar{D} + p^* = D(\bar{\theta}) + 2p^*$$

sottolineando così l'analogia con il criterio di informazione di Akaike. Di conseguenza, analogamente all'AIC, si selezionerà come miglior modello il modello con minor DIC.

3.3 Integrated nested Laplace Approximation

3.3.1 Introduzione ad INLA

Nella letteratura associata alla statistica computazionale e alla modellazione per dati spaziali sono stati definiti algoritmi diversi dal classico MCMC, utili nel caso in esame a stimare la distribuzione a posteriori della dinamica latente $\Lambda(u)$. Questo perché, durante l'analisi dei processi di punto, tendenzialmente ci si ritrova davanti al classico problema di efficienza computazionale associato alla grande mole di dati di cui si dispone. Algoritmi quali i classici MCMC, utili a definire i momenti a posteriori delle quantità di interesse, in queste situazioni risultano altamente inefficienti dal punto di vista computazionale.

Rue et al. [21] proposero l'utilizzo dell'algoritmo INLA ("Integrated Nested Laplace approximations Algorithm"), un algoritmo deterministico per effettuare inferenza Bayesiana. A differenza dell'integrazione Monte Carlo, INLA si basa sull'utilizzo dell'approssimazione di Laplace.

Si assuma di essere interessati ad approssimare il generico integrale

$$\int f(x)dx = \int \exp[\log(f(x))]dx,$$

dove $f(x)$ rappresenta la funzione di densità della generica variabile aleatoria X , con supporto S_X . Possiamo approssimare il valore $\log(f(x))$ utilizzando l'espansione in serie di Taylor dove

$$\begin{aligned} \log(f(x)) \doteq & \log(f(x_0)) + (x - x_0) \left[\frac{d \log(f(x))}{dx} \right]_{x=x_0} \\ & + \frac{(x - x_0)^2}{2} \left[\frac{d^2 \log(f(x))}{d^2 x} \right]_{x=x_0} \end{aligned} \quad (3.5)$$

Ponendo $x_0 = x^*$, dove $x^* = \arg \max_{x \in S_X} \log(f(x))$, allora

$$\left[\frac{d \log(f(x))}{dx} \right]_{x=x^*} = 0.$$

L'approssimazione si semplifica in

$$\log(f(x)) \doteq \log(f(x^*)) + \frac{(x - x^*)^2}{2} \left[\frac{d^2 \log(f(x))}{d^2 x} \right]_{x=x^*}$$

Di conseguenza l'integrale di interesse è approssimativamente pari a

$$\begin{aligned} \int f(x) dx & \doteq \int \exp \left[\log(f(x^*)) + \frac{(x - x^*)^2}{2} \left[\frac{d^2 \log(f(x))}{d^2 x} \right]_{x=x^*} \right] dx \\ & = \exp[\log(f(x^*))] \int \exp \left[\frac{(x - x^*)^2}{2} \left[\frac{d^2 \log(f(x))}{d^2 x} \right]_{x=x^*} \right] dx \end{aligned} \quad (3.6)$$

Ponendo

$$\sigma^{2*} = -1 / \left[\frac{d^2 \log(f(x))}{d^2 x} \right]_{x=x^*}$$

otteniamo

$$\int f(x) dx \doteq \exp(\log(f(x^*))) \int \exp \left[\left(-\frac{(x - x^*)^2}{2\sigma^{2*}} \right) \right] dx.$$

La parte integranda dell'approssimazione è pari al Kernel gaussiano a media x^* e varianza σ^{2*} . Più precisamente

$$\int_a^b f(x) dx \doteq f(x^*) \sqrt{2\pi\sigma^{2*}} (\Phi(b) - \Phi(a)) \quad (3.7)$$

dove $\Phi(\cdot)$ rappresenta la funzione di ripartizione di una gaussiana a media x^* e varianza σ^{2*} . L'approssimazione in equazione 3.7 è detta approssimazione di Laplace.

INLA è principalmente utilizzato per la stima di modelli di regressione bayesiani in cui a priori vengono specificate distribuzioni gaussiane con diversa struttura della matrice di varianza e covarianza, come nel caso dei processi LGCP [21].

Sia $y = (y_1, \dots, y_n)$ il campione osservato. Il generico approccio di modellazione adottato con INLA consiste nella specificazione di una opportuna distribuzione dei dati, caratterizzato da un parametro $\mu_i = E[Y_i]$ dove

$$g(\mu_i) = \eta_i = \beta_0 + \sum_{m=1}^M \beta_m x_m + \sum_{j=1}^J f_j(z_{ij})$$

rappresenta il predittore lineare con funzione di legame $g(\cdot)$. Il termine β_0 rappresenta l'intercetta, $\beta = (\beta_1, \dots, \beta_M)^T$ quantifica l'effetto lineare delle covariate $x = (x_1, \dots, x_M)^T$ sul predittore lineare e $f = \{f_1(\cdot), \dots, f_J(\cdot)\}$ rappresenta un set di funzioni associate alle covariate $z = (z_1, \dots, z_J)^T$. Quest'ultime possono rappresentare varie quantità: effetti liscati o non lineari sulla risposta, trend temporali o effetti stagionali, effetti casuali in riferimento al tempo e allo spazio. In generale, sia $\theta = \{\beta_0, \beta, f\}$ l'insieme delle quantità di interesse che governano la dinamica latente associata al modello, e sia $\psi = (\psi_1, \dots, \psi_k)^T$ il vettore di iperparametri. Assumendo indipendenza condizionale tra le varie osservazioni allora la verosimiglianza di y è pari a

$$p(y|\theta, \psi) = \prod_{i=1}^n p(y_i|\theta_i, \psi)$$

dove ogni y_i è associato agli elementi θ_i della variabile latente θ .

Si parla di *Latent Gaussian Model* perché si assume che tutti gli effetti casuali seguano distribuzione congiunta $N(0, Q^{-1}(\psi))$.

Assumendo a sua volta che gli iperparametri ψ siano a loro volta variabili aleatorie con densità pari a $p(\psi)$ allora la distribuzione congiunta

$$\begin{aligned} p(\theta, \psi|y) &\propto p(\psi)p(\theta|\psi)p(y|\theta, \psi) \\ &\propto p(\psi)p(\theta|\psi) \prod_{i=1}^n p(y_i|\theta_i, \psi) \\ &\propto p(\psi)|Q(\psi)|^{1/2} \exp\left(-\frac{1}{2}\theta^T Q(\psi)\theta\right) \prod_{i=1}^n \exp(\log p(y_i|\theta_i, \psi)) \\ &\propto p(\psi)|Q(\psi)|^{1/2} \exp\left(-\frac{1}{2}\theta^T Q(\psi)\theta + \sum_{i=1}^n \log(p(y_i|\theta_i, \psi))\right). \end{aligned} \quad (3.8)$$

3.3.2 Inferenza bayesiana con INLA

L'uso di INLA consiste nel definire i parametri a posteriori che governano la dinamica di interesse, attraverso la stima delle distribuzioni marginali a posteriori usando l'approssimazione di Laplace.

Si vuole quindi ottenere

$$p(\theta_i|y) = \int p(\theta_i, \psi|y)d\psi = \int p(\theta_i, \psi|y)p(\psi|y)d\psi.$$

Il primo step consiste nello stimare approssimativamente la distribuzione congiunta a posteriori di ψ , pari a

$$\begin{aligned} p(\psi|y) &= \frac{p(\theta, \psi|y)}{p(\theta|\psi, y)} \\ &= \frac{p(y|\theta, y)p(\theta, \psi)}{p(y)} \frac{1}{p(\theta|\psi, y)} \\ &= \frac{p(y|\theta, y)p(\theta|\psi)p(\psi)}{p(y)} \frac{1}{p(\theta|\psi, y)} \\ &\propto \frac{p(y|\theta, \psi)p(\theta|\psi)p(\psi)}{p(\theta|\psi, y)} \\ &\simeq \left[\frac{p(y|\theta, \psi)p(\theta|\psi)p(\psi)}{\tilde{p}(\theta|\psi, y)} \right]_{\theta=\theta^*(\psi)} \\ &= \tilde{p}(\psi|y), \end{aligned} \tag{3.9}$$

dove $\theta^*(\psi)$ è la moda a posteriori di θ dato ψ e $\tilde{p}(\theta|\psi, y)$ è l'approssimazione gaussiana di $p(\theta|\psi, y)$. Il secondo step consiste nello stimare la distribuzione marginale $p(\theta|y)$ e questo può risultare computazionalmente oneroso. Ci sono diversi modi per procedere, ma al fine di ottenere risultati più precisi possibili si opta per il riutilizzo dell'approssimazione di Laplace. Sia $\theta = (\theta_i, \theta_{-i})$ dove θ_i rappresenta l' i -esimo elemento di θ e θ_{-i} il vettore di elementi di θ a meno dell' i -esimo. Allora

$$\begin{aligned}
p(\theta_i|\psi, y) &= \frac{p(\theta|\psi, y)}{p(\theta_{-i}|\theta_i, \psi, y)} \\
&= \frac{p(\theta, \psi|y)}{p(\psi|y)} \frac{1}{p(\theta_{-i}|\theta_i, \psi, y)} \\
&\propto \frac{p(\theta, \psi|y)}{p(\theta_{-i}|\theta_i, \psi, y)} \\
&\simeq \left[\frac{p(\theta, \psi|y)}{\tilde{p}(\theta_{-i}|\theta_i, \psi, y)} \right]_{\theta_{-i}=\theta_{-i}^*(\theta_i, \psi)} = \tilde{p}(\theta_i|\psi, y)
\end{aligned} \tag{3.10}$$

dove $\tilde{p}(\theta_{-i}|\theta_i, \psi, y)$ rappresenta l'approssimazione di Laplace di $p(\theta_{-i}|\theta_i, \psi, y)$, la cui moda è pari a $\theta_{-i}^*(\theta_i, \psi)$. In riferimento alla notazione adottata nella sezione 3.2.2, Il modello LGCP coincide con un Latent Gaussian Model dove si assume distribuzione condizionata Poisson per y_i , e dove $\theta = (\beta_0, u_1, \dots, u_n, v_1, \dots, v_n)^T$ e $\psi = \tau$.

3.4 Applicazione ai passaggi rossoneri

Considerando ancora una volta i passaggi rossoneri in partenza, vengono stimati differenti modelli di Cox.

modello	DIC
Markov + v_{ij}	3251.523
Markov	3252.053
Matern	3283.276
Matern + v_{ij}	3283.381
Besag + v_{ij}	3336.426
Besag	3337.710

Tabella 3.1: DIC in base al modello adottato, stimato attraverso una partizione in celle di 1 m^2

La selezione del miglior modello tra i 6 adottati si basa sull'utilizzo di DIC, e come si evince dalla Tabella 3.1 il modello migliore in termini di DIC si avvale della dipendenza spaziale markoviana. Si noti anche come l'introduzione di v_{ij} per il modello porti ad un esiguo miglioramento. Questo perchè il criterio DIC, analogamente al criterio AIC, tende a sovrastimare il numero di parametri da introdurre all'interno del modello.

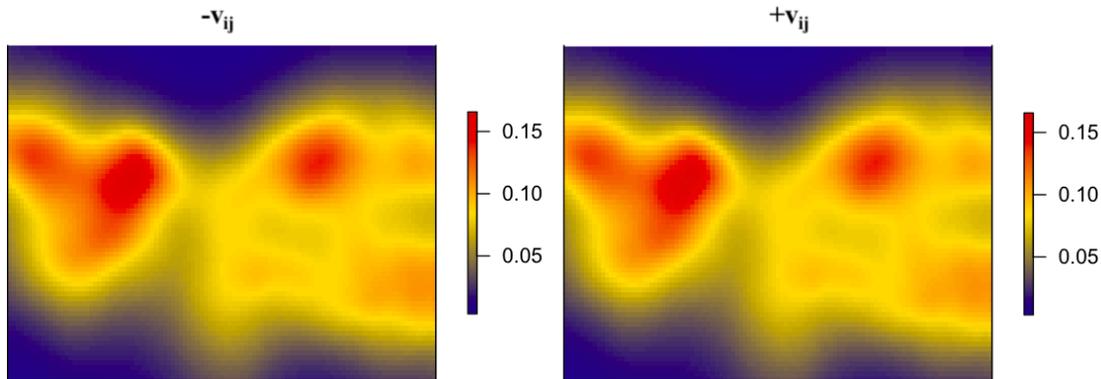


Figura 3.1: Intensità stimata attraverso un modello di Cox, considerando l'intercetta casuale o escludendola dalla modellazione, con dimensione dell'area delle celle pari ad 1 m^2

Infatti, come si evince dalla Figura 3.1 la stima a posteriori dell'intensità risulta pressoché identica, in presenza o in assenza di v_{ij} , e quindi considerarla all'interno del modello non sembra essere sensato, o per lo meno non porta a variazioni rilevanti. Tra i processi LGCP il migliore quindi risulta il modello che presenta come struttura di correlazione spaziale GMRf senza componente v_{ij} . Si noti anche come, a differenza della stima di intensità basata sul Kernel la stima dell'intensità risulta più regolare e omogenea all'interno dell'area di gioco, mentre rispetto all'intensità stimata mediante approcci GLM e GAM la stima bayesiana risulta più clusterizzata.

3.5 Il derby di Milano

3.5.1 Due squadre, quattro processi

Il 15 ottobre 2017 si è giocato il 219° derby di Milano. Nonostante l'elevata differenza del numero di passaggi tra le due squadre (478 dei rossoneri contro i 326 neroazzurri) ed un maggior possesso palla (60% Milan), la partita si è conclusa con il risultato di 3 a 2 in favore della squadra di Spalletti. Le formazioni ufficiali e i cambi effettuati sono riportati in Figura 3.2.

Si consideri 4 eventi di interesse osservati durante i 90 minuti di gioco. Si assume che i 4 corrispondenti processi siano tra loro indipendenti. La corrispondente stima della funzione di intensità è utile a comprendere le tattiche adottate dai due club. Nello specifico P_M indica la posizione di partenza dei passaggi del Milan (analizzata nei capitoli precedenti); P_I indica la posizione di partenza dei passaggi dell'Inter; A_M indica il punto d'arrivo dei passaggi del Milan; A_I indica il punto d'arrivo dei passaggi dell'Inter. Visti i risultati ottenuti per il processo P_M , si



Figura 3.2: Formazioni dei due club, con sostituzioni e voti squadra e singoli giocatori

decide di stimare i rimanenti processi attraverso altri 3 modelli LGCP seguendo la procedura di Illians. In termini di DIC, la struttura di covarianza spaziale selezionata è la dipendenza markoviana, per tutti e tre i processi rimanenti. Il termine v_{ij} non porta a differenze significative in termini di media a posteriori di Λ_{ij} e intermini di DIC. Quindi, per parsimonia, si decide di ometterlo.

Osservando la PCF riportata in Figura 3.3, si ha la conferma dell'inadeguatezza dei processi disomogenei per i processi P_I e A_I . Tuttavia, per i punti di arrivo dei passaggi rossoneri sembra essere presente una debole forma di correlazione solo tra punti a distanza di 1 o 2 metri.

3.5.2 Analisi dei residui in contesto Bayesiano

Grazie alla procedura di stima adottata, è possibile ricavare la distribuzione a posteriori dei residui del modello. Sia $\Lambda_{ij}^* = \Lambda_{ij}|z, \tau, y$ il numero di eventi attesi a posteriori della cella S_{ij} . Allora, il residuo a posteriori è pari a

$$r_{ij} \doteq n(\bar{u} \cap S_{ij}) - \Lambda_{ij}^* = y_{ij} - \Lambda_{ij}^*.$$

È possibile ricavare la distribuzione a posteriori di r_{ij} essendo nota la distribuzione a posteriori di Λ_{ij} . In figura 3.5 vengono riportati i valori attesi dei residui

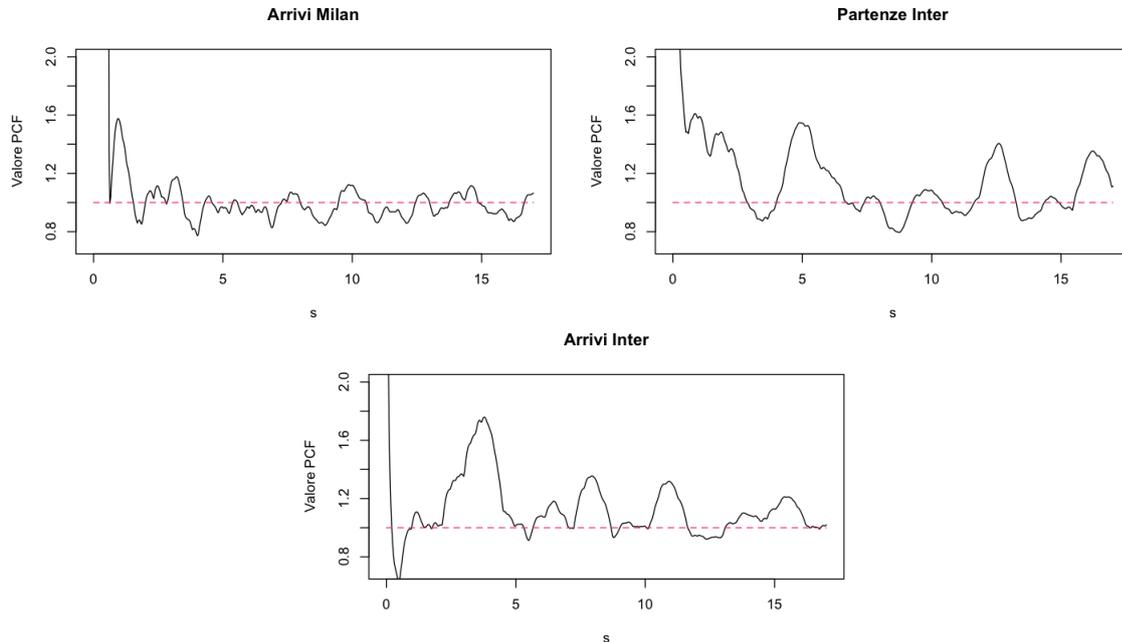


Figura 3.3: Funzione di intensità a posteriori per i 4 processi

al quadrato, al fine di rendere la rappresentazione grafica più chiara. Sapendo che il numero di eventi osservati in celle di 1 m^2 varia tra 0 e 7, si noti come i residui siano compresi nell'intervallo $[0, 20]$. Si osservano migliori performance per i modelli associati all'Inter, in quanto i residui variano in $[0, 3.5]$ e in quanto i valori più elevati sono disposti casualmente all'interno della regione per i processi P_I e A_I .

Si dubita della bontà di adattamento dei modelli LGCP associati a P_M e A_M . Se si confrontano i residui di P_M con quelli precedentemente mostrati con i modelli GLM e GAM, non vi sono differenze significative. Tuttavia, è noto che i modelli LGCP tengono conto della dipendenza tra i punti e, in aggiunta, valori elevati dei residui si osservano nelle zone del campo coincidenti alle zone di intensità a posteriori elevata. Quindi, tra i modelli stimati si considera per A_M e P_M come miglior modello il processo logaritmico di Cox.

3.5.3 Tattiche individuate

In Figura 3.4 vengono riportate le stime a posteriori del valore atteso di Λ_{ij} di tutti e 4 i processi.

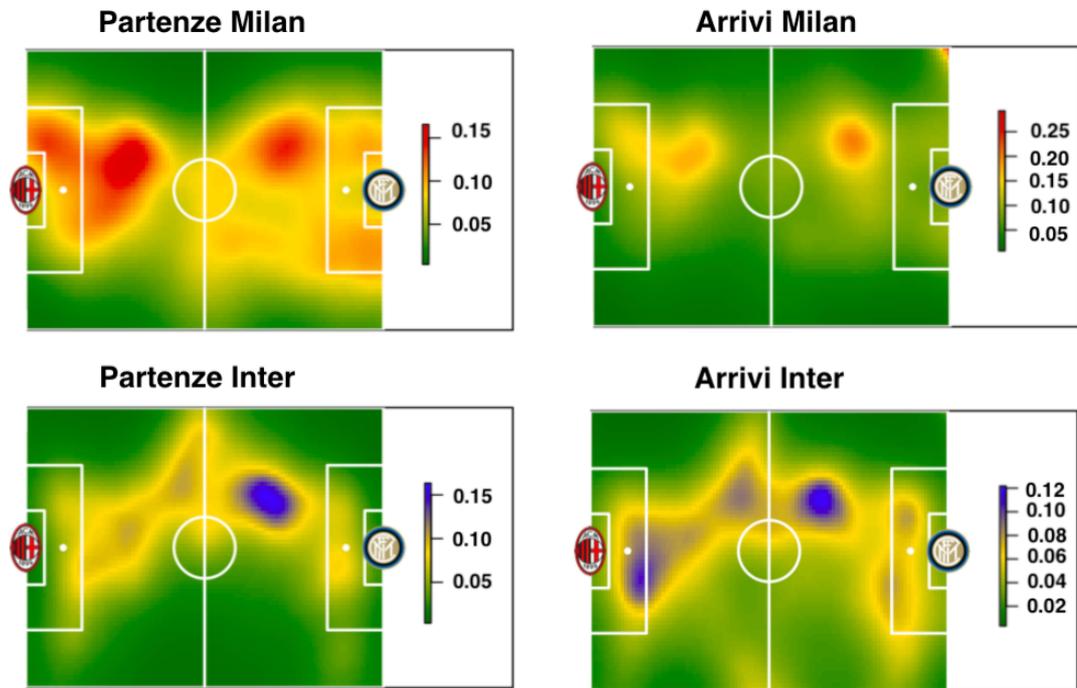


Figura 3.4: Funzione di intensità a posteriori per i 4 processi

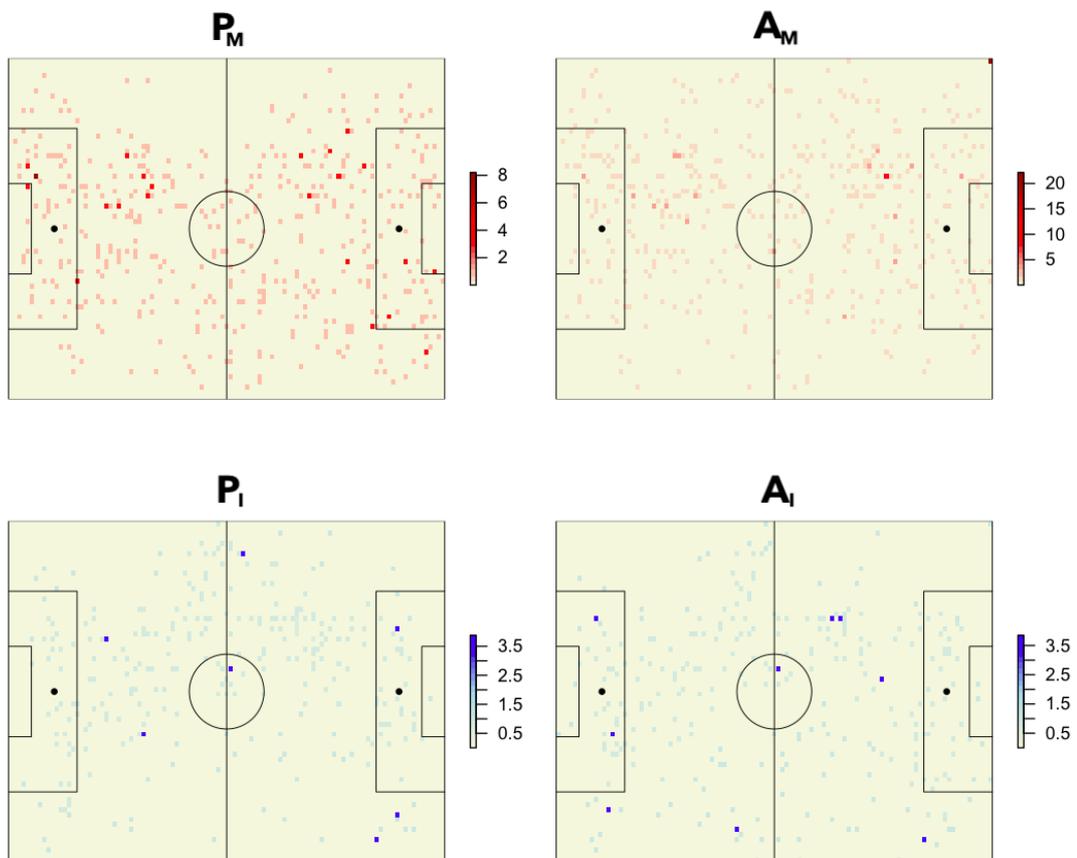


Figura 3.5: Rappresentazione della media a posteriori dei residui dei 4 modelli bayesiani

Il grafico relativo ai passaggi in partenza del Milan presenta una zona più marcata nella trequarti campo difensiva, più precisamente sul centro-sinistra. Da ciò si evince che, in questa partita, il Milan ha improntato una tattica di gioco volta a effettuare passaggi in spazi brevi e nelle zone centrali del campo, affidandosi ai giocatori con la maggior propensione ad impostare il gioco, quali Bonucci (difensore centrale) e Biglia (regista di centrocampo). Gli arrivi dei passaggi dei rossoneri ricalcano lo schema delle partenze, con un punto di spicco nella trequarti campo offensiva, intorno al limite dell'area avversaria. Questo suggerisce che i giocatori del Milan hanno cercato i propri attaccanti o centrocampisti offensivi poco fuori dall'area, mostrando una certa difficoltà ad entrare nel cuore della difesa dell'Inter. L'abilità del sopracitato Bonucci nel lancio lungo può essere evidenziata da entrambi i grafici: infatti il difensore ed ex-capitano del Milan ha tentato più volte il lancio in avanti, in cerca delle punte, specialmente André Silva.

La *heatmap* che rappresenta i passaggi in partenza dei giocatori dell'Inter evidenzia un pattern di gioco maggiormente focalizzato allo sviluppo dell'azione sulla fascia destra. Come già visto per il Milan, la zona più "calda" è la trequarti campo difensiva, con maggior propensione ad un'impostazione sul centro-destra: si parte da Skriniar (difensore centrale destro) o da Vecino (centrocampista centrale destro), per poi "allargare" il gioco sul terzino destro D'Ambrosio o sull'esterno alto Candreva, con l'obiettivo di effettuare un cross in area di rigore a beneficio di Icardi (punta centrale) o Perisic (esterno alto di sinistra). Il grafico degli arrivi supporta questa interpretazione, poiché evidenzia una zona all'interno dell'area del Milan verso la quale il pallone è arrivato spesso a destinazione. Questo sfrutta le caratteristiche di Icardi, attaccante molto forte in area di rigore e pronto a finalizzare l'azione con tiri verso la porta avversaria. Confrontando le due tattiche di gioco, si notano delle sostanziali differenze: l'Inter ha effettuato meno passaggi ma ha trovato più spesso gli attaccanti con manovre brevi, volte a finalizzare l'azione in maniera più efficace; il Milan ha improntato il gioco attraverso un susseguirsi di passaggi corti tra compagni di squadra, come dimostrato dal numero di passaggi dei rossoneri (68% in più rispetto ai "cugini" nerazzurri) e dalla funzione di intensità distribuita nella fascia centrale del campo, pur non trovando spesso la via dell'area di rigore, come precedentemente riportato.

Capitolo 4

Limiti e possibili miglioramenti

In questo ultimo capitolo vengono descritti limiti affrontati durante l'analisi, con un focus in particolare sui punti coincidenti. Vengono proposti possibili miglioramenti, in particolare citando una componente rilevante in ambito calcistico: il tempo.

4.1 Punti coincidenti

Sia $\bar{u} = \{u_1, \dots, u_n\}$ la realizzazione del processo di punto di interesse. Allora $u_i \neq u_j, \forall i, j = 1, \dots, n$. Ad ogni posizione u è associata l'osservazione di uno e un solo evento. Tuttavia nella pratica si osservano punti coincidenti, a causa dalla limitata precisione con cui i dati vengono raccolti. L'imposizione di questa assunzione è legato alla rappresentazione "pixel" della regione R di interesse.

Semplificando la notazione, sia $\{X_n\}_{n \in \mathbb{N}}$ una successione di variabili aleatorie t.c. $X_n \sim Bin(n, p_n)$, e t.c.

$$\lim_{n \rightarrow +\infty} np_n = \lambda \in (0, +\infty).$$

Allora

$$\lim_{n \rightarrow \infty} \binom{n}{x} p_n^x (1 - p_n)^{n-x} = \frac{e^{-\lambda} \lambda^x}{x}. \quad (4.1)$$

Di conseguenza

$$X_n \xrightarrow{D} Y \sim Poisson(\lambda).$$

La teoria dei processi di punto si appoggia su questo assunto [18]. Se esiste una partizione infinitesimale di R t.c. all'interno di ogni cella du vi sia uno e un solo

punto con probabilità $p(du)$ allora

$$n(\bar{U} \cap B) \sim \text{Poisson} \left(\int_B \lambda(u) du \right)$$

$\forall B \subset R$, con R regione di interesse. Nel caso in cui $\exists u_i = u_j, i \neq j$, allora il limite (4.1) non convergerebbe alla distribuzione Poisson, e di conseguenza

$$n(\bar{U} \cap B) \not\sim \text{Poisson} \left(\int_B \lambda(u) du \right).$$

Per ovviare al problema, esistono 3 soluzioni differenti:

- eliminare tutti i punti coincidenti a meno di uno rappresentante, perdendo così informazione;
- aggiungere "rumore" alle coordinate spaziali coincidenti, prestando attenzione che non si ottengano nuovi punti che cadono al di fuori di R ;
- considerare (ove possibile) una terza dimensione, che sia spaziale o temporale, controllando che non vi siano punti coincidenti.

Durante l'analisi spaziale associata a questo elaborato, è stato riscontrato il problema dei punti duplicati. Questo avviene perchè la precisione di misurazione delle coordinate arriva all'ordine dei centimetri. La soluzione adottata è stata quella di aggiungere una sporcatura che fosse di molto minore dell'unità di misura dell'intensità $\lambda(u)$. Come detto in precedenza, questa unità di misura è pari a $(\text{m}^2)^{-2} = \text{m}^{-4}$. Di conseguenza si decide di aggiungere rumore alle coordinate spaziali mediante la generazione casuale di valori la cui distribuzione è $Unif(-10^{-10}, +10^{-10})$.

4.2 Processi di punto multitypo

Un'altra assunzione imposta durante l'analisi del seguente elaborato è l'indipendenza dei 4 processi associati ai passaggi osservati durante il derby. Similmente ai modelli pooling, no pooling e pooling parziale, i processi LGCP stimati sono modelli no pooling, in grado di descrivere solo la dinamica associata ai processi marginali P_M, P_I, A_M e A_I rispettivamente.

Per tener conto del legame tra punti di diversa natura (detti **multitypo**), è possibile sfruttare modelli statistici concettualmente simili ai modelli pooling parziale.

Sia $m = 1, \dots, M$ indicatore dell' m -esimo tipo di punto considerato in analisi. Sia $Y = (U^{(1)}, \dots, U^{(M)})$ processo di punto congiunto, con $U^{(m)}$ processo marginale associato all' m -esimo tipo di punto. Allora il processo di punto multitypo omogeneo coincide con il processo omogeneo marginale per ogni tipo di punto, dove

$$E[n(U^{(M)} \cap B)] = \lambda_m |B|$$

e dove l'intensità di Y è $\lambda = \sum_{m=1}^M \lambda_m$.

Per quanto riguarda il modello Poisson disomogeneo si ottiene una stima dell'intensità pari a

$$\lambda(u, m) = \exp[b(u, m) + \theta^T Z(u, m)],$$

dove l'intensità è in funzione dello spazio e del tipo di punto, come accade per la baseline $b(u, m)$ e per le covariate $Z(u, m)$. Ponendo $b(u, m) = \alpha(m)$, si impone un'intercetta diversa al variare del tipo di punto considerato. La stima dell'intensità del processo, senza considerare la tipologia di punto, è pari a $\lambda(u) = \sum_{m=1}^M \lambda(u, m)$.

Questo tipo di modellistica è utile nel caso in cui l'interesse primario sia la stima del rischio che si verifichi nello spazio uno specifico punto di tipo m . Infatti, sia $r(m|u)$ il rischio di osservare l'evento di tipo m in posizione u , allora si dimostra che

$$r(m|u) = \frac{p(m|u)}{p(m_0|u)}$$

dove $p(m|u) = \lambda(u, m)/\lambda(u)$ è la probabilità di osservare il punto di tipo m nella posizione u e m_0 rappresenta una classificazione del punto di riferimento.

Il processo di Poisson disomogeneo multitypo non è stato adottato in contesto calcistico in quanto, come dimostrato in precedenza, è fondamentale tenere in considerazione nella fase di modellazione la dipendenza spaziale tra i punti, e in quanto non si è interessati a $r(m|u)$.

Per tener conto della dipendenza sia tra punti dello stesso tipo che tra punti di diversa natura è opportuno utilizzare il modello di Cox generalizzato. La dinamica aleatoria che guida il processo logaritmico e gaussiano di Cox è realizzazione di un processo stocastico M -variato, indicata con $(G_1(u), \dots, G_M(u))^T$, con $G_m(u) = \log(\Lambda_m(u))$, dove $\Lambda_m(u)$ rappresenta l'intensità stocastica associata all' m -esimo processo. La rappresentazione dei passaggi si formalizza utilizzando un unico modello in cui le m componenti presentano matrice di varianza e covarianza che tenga opportunamente in conto di questo aspetto, oppure in due modelli (uno per squadra) opportunamente specificati, composte da passaggi in partenza (P)

e in arrivo (A). Nel secondo caso, si assume che i punti associati alla squadra rossonera siano indipendenti da quelli neroazzurri. Allora risulterebbe opportuno considerare un processo stocastico bivariato della forma

$$\Lambda_M(u) = [\exp(G_{P_M}(u)), \exp(G_{A_M}(u))]^T = [\Lambda_{P_M}(u), \Lambda_{A_M}(u)]^T,$$

$$\Lambda_I(u) = [\exp(G_{P_I}(u)), \exp(G_{A_I}(u))]^T = [\Lambda_{P_I}(u), \Lambda_{A_I}(u)]^T,$$

indicando così la dipendenza solo tra partenze e arrivi di una stessa squadra. La modellazione (anche di un semplice processo bivariato) può risultare estremamente complicata, in quanto richiede la definizione di nuove strutture di correlazione non descritte, più complesse e indicizzate da un numero maggiore di parametri. Il consiglio è quello di definire processi marginali per ogni tipo di punto studiato, e di considerare un processo m -variato ove ce ne fosse necessità, ovvero nel caso in cui la bontà di adattamento dei modelli marginali risulti non soddisfacente. Il modello pooling raramente viene utilizzato nell'analisi dei processi di punto. Non considerando in nessun modo la diversa classificazione dei punti osservati, la stima dell'intensità porterebbe a conclusioni totalmente fuorvianti e non informative per descrivere il fenomeno di interesse.

4.3 Piastrellatura e SPDE

La discretizzazione del processo LGCP si basa sulla suddivisione del campo in n celle a forma rettangolare e di uguale area. Tuttavia, nulla vieta l'utilizzo di altre forme geometriche e di altre procedure di ripartizioni di R . In generale si parla di piastrellatura, intesa come la procedura di separazione della regione R in sotto-regioni disgiunte, chiamate piastrelle. Tra le più famose vi sono la piastrellatura di Dirichlet e la triangolarizzazione di Delaunay che, a differenza della piastrellatura rettangolare, definiscono parzioni di R in funzione di \bar{u} . Un esempio applicato a P_M è riportato in Figura 4.1.

L'uso di piastrelature più complesse permette l'utilizzo di procedure più sofisticate e precise per la stima del modello LGCP. La procedura di Illians rappresenta un'approssimazione del processo a spazio continuo, partizionando il campo in diverse celle disgiunte, ed ottenendo così un processo a spazio discreto.

Esiste una seconda alternativa, che si serve della triangolarizzazione di Delanay

e della *stochastic partial differential equation* (SPDE) [12,13]. Sia

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau\xi(u)) = W(u)$$

dove Δ rappresenta l'operatore di Laplace, α è detto parametro di liscio, $\kappa > 0$ parametro di scala, τ parametro di controllo della varianza e $W(u)$ processo spaziale white noise gaussiano. La soluzione esatta e stazionaria di SPDE è pari al Gaussian Field $\xi(u)$ con matrice di varianza e covarianza di Matern. Essendo finito il numero di eventi osservati nello spazio, allora $\xi(u)$ è definita attraverso una rappresentazione in G basi associata ad una opportuna triangolarizzazione di R , dove

$$\xi(u) = \sum_{g=1}^G \psi_g(u) \tilde{\xi}_g$$

e dove $\tilde{\xi} = (\tilde{\xi}_1, \dots, \tilde{\xi}_G)^T$ è vettore di pesi con distribuzione gaussiana multivariata a media nulla. Si è deciso di non utilizzare una procedura di stima del processo

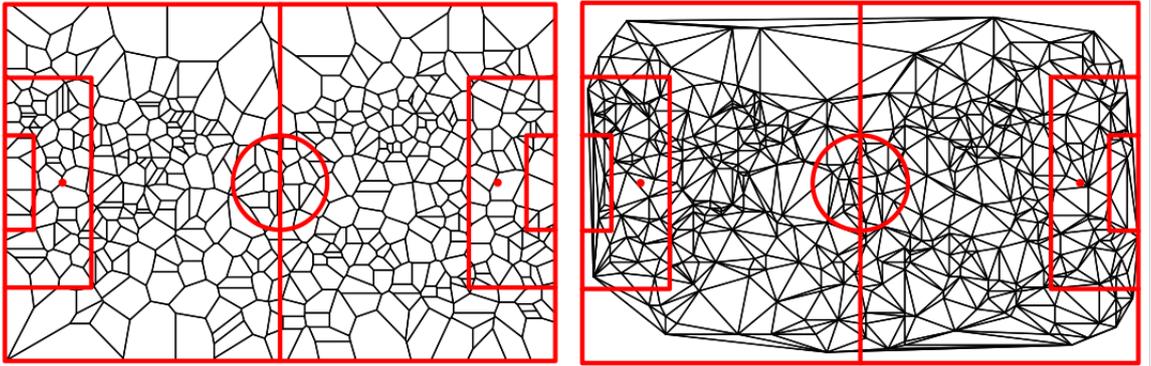


Figura 4.1: Applicazione della piastrellatura di Dirichlet (a sinistra) e triangolarizzazione di Delaunay (a destra), applicato alle partenze rossonere.

LGCP mediante SPDE, visti gli ottimi risultati ottenuti tramite la procedura di Illians.

4.4 Spazio e Tempo

Considerare la componente temporale nel contesto calcistico può fornire informazioni più precise sulle tattiche di gioco, che possono cambiare durante il corso della partita. Assumendo che il processo sia stocastico nello spazio e nel tempo si parla di **processi di punto spazio-temporali**. A differenza della notazione

adottata precedentemente, sia

$$u = (s, t), \text{ con } s = (x, y)$$

dove $s \in S \subset \mathbb{R}^2$, con S regione di interesse e $t \in T \subset \mathbb{R}^+$ con S e T insieme limitati. Mentre s rappresenta le coordinate spaziali, t rappresenta il tempo in cui avviene l'evento. Le osservazioni $u_i = (s_i, t_i)$ rappresentano il punto osservato nella s_i -esima posizione spaziale e nel t_i -esimo tempo. Come in precedenza, $\bar{u} = \{u_1, \dots, u_n\}$ rappresenta il campione osservato e il dominio delle osservazioni è pari a $S \times T$ insieme finito.

La quantità $\lambda(s, t)$ è in funzione delle due componenti di interesse, portando a concludere che

$$n(\bar{U} \cap B) \sim \text{Poisson} \left(\int_{S^*} \int_{T^*} \lambda(s, t) ds dt \right)$$

dove

$$E[n(\bar{U} \cap B)] = \int_{S^*} \int_{T^*} \lambda(s, t) ds dt$$

con $B = S^* \times T^* \subset S \times T$.

E' possibile definire la pair correlation function

$$\rho(u_i, u_j) = \frac{\lambda_2(u_i, u_j)}{\lambda(u_i)\lambda(u_j)}$$

dove la funzione $\lambda_2(\cdot, \cdot)$ rappresenta l'intensità prodotto del processo e

$$\gamma(u_i, u_j) = \lambda_2(u_i, u_j) - \lambda(u_i)\lambda(u_j)$$

rappresenta la funzione di covarianza.

Adottare una modellazione spazio-temporale si dimostra sensato solo nel caso in cui il processo stesso non risulti stazionario in una o entrambe le componenti. Nello specifico, si parla di processo stazionario e isotropico:

- **nello spazio**, se $\lambda(s, t) = \lambda(t)$, e se $\lambda_2((s_i, t), (s_j, t)) = \lambda_2(\|s_i - s_j\|, t)$;
- **nel tempo** se $\lambda(s, t) = \lambda(s)$, e se $\lambda_2((s, t_i), (s, t_j)) = \lambda_2(s, |t_i - t_j|)$;
- **nello spazio e nel tempo** se $\lambda(s, t) = \lambda$, e se $\lambda_2((s_i, t_i), (s_j, t_j)) = \lambda_2(\|s_i - s_j\|, |t_i - t_j|)$.

Modellare congiuntamente la componente stocastica associata allo spazio e quella associata al tempo risulta estremamente complesso, per questo si impone un'assunzione semplificatoria, detta separabilità del primo e secondo ordine. Si parla di **separabilità del primo ordine** se

$$\lambda(s, t) = m(s)\mu(t)$$

per ogni $(s, t) \in S \times T$ e di **separabilità del secondo ordine** se

$$\gamma(u, v) = \gamma_S(h)\gamma_T(k)$$

dove $\gamma_S(u)$ rappresenta la covarianza spaziale isotropica ($h = \|s_i - s_j\|$) e $\gamma_T(u)$ rappresenta la covarianza temporale isotropica ($k = |t_i - t_j|$).

Conclusione

L'obiettivo del seguente elaborato è stato l'utilizzo dei processi di punto spaziale per l'analisi degli eventi osservati nel derby di Milano (15/10/2017) durante i 90 minuti di gioco. L'intensità dei passaggi è stata di primario interesse in quanto dalla corrispondente stima di intensità è stato possibile definire le tattiche di gioco utilizzate dalle due squadre. Nulla vieta l'analisi di altri tipi di evento, quali per esempio i duelli vinti o persi, al fine di identificare le zone con lacune difensive per ogni club.

In prima battuta è stato dimostrato che la disposizione degli eventi all'interno della regione non fosse casuale, per questo sono stati introdotti i processi Poisson Omogenei.

Successivamente, attraverso i processi Poisson Disomogenei, sono state definite possibili forme funzionali associate all'intensità. I parametri che governano la dinamica del processo sono stati stimati via massima verosimiglianza pesata. L'inadeguatezza dei modelli, dettata da una debole forma funzionale imposta all'intensità, è stata mostrata attraverso il test χ^2 e l'analisi dei residui. In aggiunta, i processi Poisson Disomogenei non sono in grado di cogliere la dipendenza tra punti. Dopo aver definito il concetto di correlazione tra punti, e dopo aver dimostrato la presenza di questa nei dati oggetto di studio, è stato presentato il processo log-Gaussiano di Cox, in grado di cogliere questa forma di attrazione e repulsione imponendo che la dinamica latente del processo sia Gaussian Random Field. La sua matrice di varianza e covarianza descrive la dipendenza spaziale tra le coppie di punti.

E' stato possibile stimare il processo LGCP attraverso la discretizzazione del processo e INLA, un algoritmo deterministico per effettuare inferenza Bayesiana che si basa sull'utilizzo della approssimazione di Laplace per il calcolo degli integrali di interesse. Essendo possibile assumere diverse forme di dipendenza tra i punti (markoviana, Besag e Matern sono le tre considerate in questo elaborato) è stato introdotto il criterio di selezione DIC. Si è dimostrato quindi che, stimando 4

processi differenti per punti di partenza e arrivo dei passaggi per ogni squadra, il miglior modello (in termini di DIC) è il modello LGCP con matrice di varianza e covarianza definita sotto ipotesi di Gaussian Markov Random Field. L'analisi dei residui di questi modelli avvalorano l'ipotesi di buona bontà di adattamento. In ultima battuta sono state descritte le tattiche di gioco del Milan e dell'Inter. Estensioni di questo lavoro comprendono l'utilizzo di partizioni del campo più complessa della semplice partizione rettangolare e in funzione dei punti osservati. Questa procedura di suddivisione del campo è detta piastrellatura. In aggiunta, si potrebbe considerare la dipendenza tra i 4 punti del processo, definendo così processi detti multitypo. In ultima battuta, è possibile considerare la componente temporale associata ai punti osservati, al fine di ottenere una rappresentazione sia spaziale che temporale dell'intensità dei passaggi.

Bibliografia

- [1] BADDELEY, A., RUBAK, E., AND TURNER, R. Spatial Point Patterns: Methodology and Applications with R. *CRC Press* (2016).
- [2] BADDELEY, A., TURNER, R., MØLLER, J., AND HAZELTON, M. Residual analysis for spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2005), p. 617–665.
- [3] BADDELEY, A. J., AND MØLLER, J. & WAAGEPETERSEN, R. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 54, 3 (2000), p. 329–350.
- [4] BADDELEY, ADRIAN & TURNER, R. spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software* (2005).
- [5] BERMAN, M., AND TURNER, T. R. Approximating Point Process Likelihoods with GLIM. *Applied Statistics* 41 (1992).
- [6] BESAG, J., AND DIGGLE, P. J. Simple Monte Carlo Tests for Spatial Pattern. *Applied Statistics* (1977), 327.
- [7] BESAG, J., YORK, J., AND MOLLI, A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* (1991).
- [8] BLANGIARDO, M., AND CAMELETTI, M. *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons, (2015).
- [9] DIGGLE, P. A Kernel Method for Smoothing Point Process Data. *Applied Statistics* 34, 2 (1985).
- [10] GUDMUNDSSON, J., AND HORTON, M. Spatio-Temporal Analysis of Team Sports. *ACM Computing Surveys* (2018).

- [11] ILLIAN, J. B., SØRBYE, S. H., RUE, H., AND HENDRICHSEN, D. K. Using INLA To Fit A Complex Point Process Model With Temporally Varying Effects – A Case Study. *Journal of Environmental Statistics* (2012).
- [12] KRAINSKI, E., GÓMEZ-RUBIO, V., BAKKA, H., LENZI, A., CASTRO-CAMILO, D., SIMPSON, D., LINDGREN, F., AND RUE, H. *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. CRC Press, (2018).
- [13] LINDGREN, F., RUE, H., AND LINDSTRÖM, J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach: Link between Gaussian Fields and Gaussian Markov Random Fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 4 (2011), p. 423–498.
- [14] MCFADDEN, J. A. The Mixed Poisson Process. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* (2022).
- [15] MORAGA, P. Species Distribution Modeling using Spatial Point Processes: a Case Study of Sloth Occurrence in Costa Rica. *The R Journal* 12 (2020), p. 293.
- [16] MØLLER, J., SYVERSVEEN, A. R., AND WAAGEPETERSEN, R. P. Log Gaussian Cox processes. *Scandinavian Journal of Statistics* (1998), p. 451–482.
- [17] NEATH, A. A., AND CAVANAUGH, J. E. The Bayesian information criterion: background, derivation, and applications. *WIREs Computational Statistics* 4, 2 (2012), p. 199–203.
- [18] PAPOULIS, A., AND PILLAI, S. U. *Probability, random variables, and stochastic processes*. McGraw-Hill, (2002).
- [19] PAPPALARDO, L., CINTIA, P., ROSSI, A., MASSUCCO, E., FERRAGINA, P., PEDRESCHI, D., AND GIANNOTTI, F. A public data set of spatio-temporal match events in soccer competitions. *Scientific Data* (2019).
- [20] RIPLEY, B. P. The Second-Order Analysis of Stationary Point Processes. *Journal of Applied Probability* (1976).

-
- [21] RUE, H., MARTINO, S., AND CHOPIN, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2009), p. 319–392.
- [22] SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P., AND VAN DER LINDE, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2002), p. 583–639.

Ringraziamenti

Ci tengo a ringraziare il professor Bernardi Mauro, del dipartimento di Scienze Statistiche (Università degli studi di Padova), per avermi fornito i dati.

Ringrazio anche il relatore Aliverti Emanuele, per essere stato una guida importante durante la stesura della tesi.

Ringrazio la mia famiglia, per il supporto fornitomi in tutti questi anni della mia vita.

Ringrazio anche la mia seconda famiglia, i miei coinquilini presenti e passati. Mi hanno sempre spronato ad essere la versione migliore di me stesso.

Ringrazio gli amici incontrati a Padova, compagni di viaggio di questi 5 meravigliosi anni, nella gioia e nel dolore.

Infine ringrazio il direttivo Questi Siamo, i miei fratelli partecipi dei miei ricordi più preziosi.