

Analisi spaziale del comportamento delle coorti: mutamenti nell'uso di contraccettivi al primo rapporto in Italia

Riccardo Borgoni, Francesco C. Billari¹

*Max Planck Institute for Demographic Research
Doberaner Strasse 114 D-18057 Rostock, Germania*

Riassunto: In questo lavoro si analizza la dinamica territoriale dell'uso di metodi contraccettivi alla prima esperienza sessuale nel caso delle donne italiane così come emerge dall'*Indagine Nazionale sulla Fecondità* evidenziando in particolare modo l'evoluzione che tale dinamica ha presentato attraverso coorti successive. I dati utilizzati forniscono l'informazione necessaria ad individuare la locazione delle rispondenti. Mediante record linkage ad un dataset geo-referenziato si rappresenta cartograficamente l'evoluzione della propensione all'uso di metodi contraccettivi tramite un sistema informativo geografico. L'approccio modellistico, di tipo bayesiano, si basa su un modello lineare generalizzato misto di tipo logit. La correlazione spaziale è gestita tramite un processo autoregressivo condizionato gaussiano (CAR). L'inferenza è condotta tramite simulazione Monte Carlo di catene markoviane (MCMC). I risultati evidenziano l'esistenza di un marcato trend territoriale che risulta abbastanza stabile nelle varie coorti. A conclusione del lavoro emerge che nell'analisi di indicatori di rischio sociale i metodi messi a disposizione dalla statistica spaziale possono rivelarsi utili e altamente informativi come guida a politiche d'intervento.

Parole chiave: uso di anticoncezionali, modelli lineari generalizzati misti, modelli autoregressivi condizionati (CAR), MCMC, sistemi informativi geografici (GIS).

1. Introduzione

Di recente, nella letteratura soprattutto statunitense, si è sviluppata una forte attenzione ai rischi per la salute legati ad alcuni comportamenti sessuali. Una particolare

¹ Il presente lavoro è stato finanziato dal Max Planck Institute for Demographic Research e nell'ambito del progetto "La ricerca di determinanti del rischio mediante analisi di segmentazione di campioni" cofinanziato dal MIUR. Coordinatore nazionale è Luigi Fabbris. La nota è stata redatta da Riccardo Borgoni per i parr. 3, 4 e 5 e da Francesco C. Billari per i parr. 1, 2 e 6. Gli autori desiderano ringraziare Paola Del Bianco e un anonimo referee. Le opinioni espresse in questo lavoro sono attribuibili agli autori e non riflettono necessariamente quelle dell'Istituto di appartenenza.

enfasi è posta sugli adolescenti, spesso soggetti ad un significativo rischio di contrarre infezioni sessualmente trasmesse, tra cui il virus HIV, nonché alla possibilità di gravidanze indesiderate (Kotchick et al., 2001). Il comportamento sessuale "a rischio" è stato indicato dunque come una delle dimensioni fondamentali per la misurazione dei rischi relativi alla salute dei giovani, anche perché metà dei nuovi casi di infezione HIV nel mondo si manifestano nei giovani di età compresa tra 15 e 24 anni (UNFPA, 2000). Spesso la definizione di comportamento a rischio contiene il riferimento a rapporti sessuali particolarmente precoci (Duberstein, Lindberg et al., 2000), ma l'attenzione è anche rivolta all'utilizzo di metodi contraccettivi in occasione del primo rapporto (Hogan et al., 2000). È stato anche mostrato che l'utilizzo di metodi contraccettivi al primo rapporto ha un'influenza decisiva sulle successive decisioni contraccettive (Ku et al., 1994). L'Italia ad esempio si distingue nel contesto internazionale, sia per una bassa prevalenza dell'uso di metodi contraccettivi al primo rapporto, sia per la permanenza comportamenti contraccettivi "tradizionali" anche in età adulte (Bonarini, 1999; Spinelli et al., 2000). Il caso italiano è caratterizzato, inoltre, da un'età al primo rapporto sessuale decisamente più elevata rispetto agli altri paesi occidentali (Cazzola, 1999). Questo fa sì che l'attenzione sul rischio relativo alla salute riproduttiva dei giovani possa essere concentrata sull'utilizzo di metodi contraccettivi in quell'occasione. Tale indicatore, come indicatore di rischio, può essere insufficiente se preso *sic et simpliciter*. In particolare, la valutazione del rischio è imperfetta quando non sia noto il metodo contraccettivo utilizzato. Il metodo contraccettivo, infatti, può essere non adatto alla prevenzione di malattie sessualmente trasmesse e/o poco sicuro per quanto riguarda la protezione da gravidanze indesiderate. Anche in seguito alle limitazioni dei dati disponibili, in quanto segue ci concentreremo sull'utilizzo di metodi contraccettivi indipendentemente dal tipo di metodo in occasione del primo rapporto sessuale. Tale indicatore fornisce senza dubbio una misura sottostimata per l'esposizione al rischio di gravidanze (potendo il metodo essere di non perfetta efficacia), e decisamente sottostimata per l'esposizione al rischio di malattie sessualmente trasmissibili. D'altra parte, in alcuni casi il primo rapporto può essere stato volontariamente volto alla ricerca di un concepimento, il che rende la sottostima meno grave.

Il contesto sociale in cui si trovano i giovani è fondamentale nello studio delle decisioni contraccettive. L'informazione disponibile nell'ambito della rete sociale di appartenenza, le norme sociali ivi prevalenti, l'informazione fornita dai mezzi di informazione locali, dalle scuole e la disponibilità di contraccettivi sono caratteristiche contestuali di fondamentale importanza (Brewster et al., 1993; Billy et al., 1994; Teitler e Weiss, 2000). L'importanza del territorio di riferimento è dunque primariamente campo di interpretazione dei comportamenti. Ad esempio, soprattutto per le donne, l'età al primo rapporto sessuale varia fortemente sul territorio italiano (Ongaro, 2001; Billari e Borgoni, 2002). Inoltre, le possibili politiche sociali volte a ridurre il rischio hanno un'efficacia maggiore quando pianificate ad livello territoriale partico-

larmente disaggregato, anche se non necessariamente il quartiere: sempre in riferimento al contesto statunitense, Teitler e Weiss (2000) hanno mostrato l'importanza delle istituzioni scolastiche. Mauldon e Luker (1996) hanno messo in evidenza come l'istruzione relativa alla contraccezione aumenta di un terzo gli odds dell'utilizzo di contraccettivi al primo rapporto, ed anche dell'utilizzo di profilattici. Un problema generale per la misura e la valutazione statistica di rischi di salute riproduttiva come quello qui considerato è l'assenza di statistiche ufficiali sul fenomeno. I questi casi, diviene necessario ricorrere a informazioni tratte da indagini campionarie. Se si tiene conto dell'importanza della variazione territoriale del fenomeno, occorre utilizzare apposite tecniche per la sua mappatura, che tengano sotto controllo la variabilità campionaria, possibilmente sfruttando la geo-referenziazione dei dati e l'ipotizzato radicamento territoriale del fenomeno. L'analisi relativa a differenziali geografici nell'uso di metodi contraccettivi al primo rapporto sessuale per le donne italiane è stata condotta sugli stessi dati qui utilizzati in Borgoni e Billari (2002-a, 2002-b). Nei lavori menzionati è stata evidenziata l'esistenza di un gradiente nord-sud nell'uso dei metodi contraccettivi al primo rapporto. Inoltre, mentre l'età al primo rapporto sessuale non sembra essere un predittore statisticamente significativo della propensione all'uso di metodi contraccettivi, è stato mostrato come la coorte dell'intervistata influenzi in modo significativo la probabilità di usare un metodo contraccettivo. Il presente contributo estende i lavori citati in precedenza, approfondendo l'evoluzione temporale (dove il tempo considerato è il tempo delle coorti) della struttura spaziale del fenomeno considerato, nonché la stabilità dei differenziali nel corso del tempo. In particolare, si discute se è ipotizzabile nel caso dell'uso di metodi di contraccezione un processo diffusivo a livello territoriale analogo a quello che sembra essersi verificato per quanto riguarda un'analisi dell'incidenza del primo rapporto sessuale (Billari e Borgoni, 2002). A questo proposito si fornisce una mappa del fenomeno a livello provinciale per tre coorti decennali consecutive.

Per agevolare la lettura, riportiamo di seguito la struttura del lavoro. Nel paragrafo 2 introduciamo i dati utilizzati. Nel paragrafo 3 presentiamo una breve rassegna sullo stato dell'arte per quanto riguarda i metodi statistici e gli strumenti informatici recentemente introdotti per l'analisi di dati geo-referenziati. L'approccio metodologico, di impostazione bayesiana, è introdotto nel paragrafo 4. Il paragrafo 5 presenta i risultati dell'analisi condotta, mentre il paragrafo 6 conclude il lavoro con alcune riflessioni.

2. I dati utilizzati

I dati utilizzati per questo lavoro provengono dall'Indagine Nazionale sulla Fecondità (INF-2), un'indagine campionaria contenente domande sulla situazione attuale e sulla biografia dell'intervistato, con questionario etero-somministrato (De Sandre et al.,

1997). L'indagine ha fatto parte di un programma implementato dalla Population Activities Unit (PAU) della Commissione Economica per l'Europa delle Nazioni Unite (UN/ECE) allo scopo di raccogliere informazioni comparabili a livello internazionale.

La strategia di campionamento, curata dall'ISTAT, ed il lavoro sul campo dell'indagine, svolto dalla società di rilevazione Metron, sono descritti in dettaglio da Zannella et al. (1997). Il disegno campionario adottato è a tre stadi (comuni, sezioni elettorali, individui), ed il campione complessivo è composto da 6030 unità, con due campioni indipendenti per donne e uomini. In particolare, sono state intervistate 4824 donne (di cui il 58.1% proviene dalla lista base senza sostituzioni) e 1206 uomini (il 51.7% proveniente dalla lista base), nati fra il 1946 e il 1975. Le interviste, condotte tra il novembre 1995 e il gennaio 1996, contenevano domande riguardanti una serie di eventi bio-demografici, rilevando al contempo le età in cui questi erano stati sperimentati. Le domande utilizzate in questo lavoro sono contenute nella sezione "Regolazione della fecondità". Agli intervistati è stato chiesto "Al fine di evitare quesiti successivi non pertinenti, posso chiederle: ha mai avuto rapporti sessuali completi?" (risposte possibili: sì, no, non risponde). Nel caso di risposta affermativa, veniva chiesto "A che età ha avuto il primo rapporto sessuale completo?" (con possibilità di non risposta) e "In quella occasione è stato fatto qualcosa per evitare un concepimento?" (risposte possibili: sì, no, non sa, non risponde). Purtroppo, non è stato richiesto il tipo di metodo, la cui rilevazione è stata limitata alle ultime quattro settimane prima dell'intervista (Bonarini, 1999). Ai fini dell'analisi qui riportata è stato fatto riferimento al sotto insieme di soggetti che avevano sperimentato il primo rapporto sessuale, la gran parte del campione (5279 individui). Dal data set sono stati inoltre eliminati tutti i soggetti che non hanno risposto sulla sperimentazione del primo rapporto sessuale (95 casi). I dati sono di qualità soddisfacente: Cazzola (1999, pag. 318) analizza le mancate risposte alla domanda sull'età al primo rapporto e conclude che "l'ipotesi di una distorsione nelle risposte come fattore esplicativo di qualche rilievo non trova particolari riscontri".

Essendo interessati in questo lavoro essenzialmente a problematiche spaziali, e dovendo collocare gli intervistati sulla base del loro domicilio nelle età attorno al primo rapporto sessuale, abbiamo utilizzato come riferimento la risposta alla domanda "In quale comune lei ha vissuto per la maggior parte del tempo fino ai suoi 15 anni?", con risposta codificata dall'ISTAT. Si tratta dunque di un comune che non necessariamente è quello di residenza ma che riflette presumibilmente in modo superiore il contesto in cui è stato vissuto il primo rapporto sessuale. Sono stati eliminati i casi di residenti all'estero e quelli di cui non era noto il comune di residenza, e ciò ha ridotto l'insieme dei dati a 5107 osservazioni. I comuni sono poi stati successivamente aggregati a livello provinciale². Dato lo scopo essenzialmente metodologico del

² Rivellini e Zaccarin (1999, p. 653) osservano che il comune, come livello contestuale, ha il vantaggio di essere l'unità naturale dato il disegno di campionamento dell'Inf-2 (anche se nel nostro caso la

presente lavoro, abbiamo deciso di concentrare l'attenzione sulla parte femminile del campione, riducendo l'insieme dei dati a 4038 osservazioni. Infine, volendo tenere conto dell'età in cui il primo rapporto ha avuto luogo, si è deciso di eliminare dall'analisi tutte quelle unità (32) per cui non era disponibile quest'informazione. Il campione finale utilizzato consiste di 4006 osservazioni. La risposta alla domanda sui metodi contraccettivi è stata codificata in modo binario per discriminare coloro che avevano fatto uso di un qualche tipo di contraccezione, da coloro che non ne avevano usato alcuno. Si è deciso di includere nella classe di coloro che non hanno usato metodi anche le persone che hanno risposto "non so" (40 casi).

3. Software e metodi statistici per l'analisi territoriale di fenomeni sociali: una breve rassegna

I fenomeni sociali presentano spesso un'elevata variazione geografica. Questo può essere attribuito al fatto che la locazione in cui gli individui maturano le loro scelte è solitamente una proxy di fattori economici e socio-culturali. Ignorare questo aspetto a favore di analisi in cui solo le caratteristiche individuali siano prese in considerazione può indurre una perdita di informazione relativa al contesto e la possibilità di incorrere in una classica "fallacia ecologica". Prima di procedere con l'analisi effettuata, ci sembra quindi opportuno soffermarsi su una breve descrizione dello "stato dell'arte", per quanto riguarda da un lato le metodologie statistiche idonee all'analisi di dati spaziali, e dall'altro le risorse software disponibili.

L'analisi di dati spaziali ha avuto un grosso impulso nell'ultimo ventennio. Riferimenti bibliografici fondamentali nell'ambito della statistica per processi spaziali sono i testi di Ripley (1981, 1988), Diggle (1983) e Cressie (1993), ma molti rilevanti contributi sono stati apportati nel corso del tempo da autori che hanno lavorato in diversi contesti applicativi come la demografia (Woods e Rees, 1986) o l'epidemiologia (Elliot et al. 1992, 2000, Lawson, 2001).

Ai progressi riscontrati in ambito teorico si è accompagnato in anni recenti uno sviluppo parallelo del software. Molti pacchetti commerciali come Matlab o S-Plus o liberamente accessibili via Internet come R presentano librerie dedicate. Si pensa qui in particolare alle librerie gratuitamente accessibili via Internet *Spatial* (Venables e Ripley, 1999) e *GeoS/GeoR* (Ribeiro e Diggle, 1999) per S-Plus e R o a estensioni commerciali come la libreria *SpatialStat* per S-Plus. Allo stesso tempo hanno fatto la loro

corrispondenza non è perfetta), e che esso costituisce il "livello territoriale più prossimo - almeno rispetto ad altre possibili aggregazioni territoriali - con cui l'individuo entra in contatti nelle decisioni e nelle scelte quotidiane del suo corso di vita". Le autrici utilizzano anche il livello provinciale, facendo riferimento ad altre analisi del comportamento riproduttivo (Egidi e Zaccarin, 1996; Zaccarin, 1996).

comparsa anche pacchetti unicamente dedicati all'analisi spaziale come Variowin (Pannatier, 1996).

L'evoluzione degli strumenti hardware cui abbiamo assistito nell'ultimo decennio ha poi reso possibile la gestione di grandi basi di dati in cui l'informazione territoriale viene registrata assieme a quella relativa all'accadimento degli eventi (si pensi in particolare a survey condotte su specifici argomenti come quella ricordata nella sezione precedente o a dati di censimento o, ancora, a dati provenienti da registri in cui i soggetti sono seguiti nel tempo e nello spazio per gli eventi che riguardano il loro corso di vita). La geo-referenziazione degli eventi anche a livelli di dettaglio elevati rende possibile analisi territoriali particolarmente ricche.

È argomento degli ultimi anni l'introduzione di pacchetti statistici dedicati alla rappresentazione e analisi di dati geo-referenziati denominati Sistemi Informativi Geografici (GIS, *Geographical Information System*). Un'estesa presentazione delle potenzialità dei GIS per l'analisi quantitativa è fornita da Fotheringham et al. (2000) e una rassegna delle possibili prospettive future dell'uso dei GIS nell'analisi dei comportamenti umani è data ad esempio da Kwan (2000). In realtà, se in fase di presentazione delle informazioni e dei risultati della modellazione i GIS risultano essere particolarmente efficaci consentendo mappe anche altamente dettagliate, in termini di analisi statistica gli strumenti disponibili all'interno di questi pacchetti sono poco evoluti e spesso sostanzialmente limitati alla possibilità di calcolare statistiche di tipo descrittivo. Questo richiede allo statistico che si occupa di analizzare le dinamiche territoriali di fenomeni sociali due passi di analisi. Una prima fase in cui i dati sono analizzati tramite pacchetti statistici usuali e una seconda fase in cui i risultati sono importati nel GIS per la loro rappresentazione. Questa seconda fase, allorché le analisi siano fatte a livello individuale può richiedere opportuni *linkage* fra data set differenti. Nel corso del tempo si è assistito quindi al tentativo di integrare le potenzialità dei GIS con quelle di metodi statistici avanzati per le analisi spaziali. Questo processo ha attraversato differenti fasi (Anselin, 2000). Inizialmente si è assistito al tentativo di estendere alcune potenzialità dei GIS ai pacchetti statistici e viceversa. Successivamente, per corredare il GIS di funzioni statistiche più avanzate, si è ricorsi alla scrittura di macro e funzioni nei linguaggi di programmazione presenti nei sistemi informativi geografici più diffusi. Oggi sono finalmente disponibili interfacce fra alcuni pacchetti statistici e alcuni sistemi informativi geografici tra i più diffusi, ci si riferisce qui in particolare alle interfacce tra S-Plus e ArcView GIS o Grassland (Bao et al., 2000) o R e Grass.

In realtà molti metodi statistici per l'analisi di dati spaziali si sono diffusi solo recentemente anche per effetto della loro natura computazionale che ha richiesto, preliminarmente, lo sviluppo di strumenti hardware adeguati. Tali metodi non sono ancora implementati in pacchetti statistici standard. Ciò rende inevitabile un'analisi che si articoli nei due passi sopra ricordati. Questo approccio è stato seguito nelle analisi

riportate nel paragrafo 5 di questo lavoro dove, risultati preliminarmente ottenuti usando un pacchetto specifico per l'analisi bayesiana di dati spaziali, sono stati importati in un GIS per la loro rappresentazione. In particolare in questo lavoro si è fatto ricorso ad ArcView/GIS (ESRI, 1996-a). In realtà, occorre dire, che il ricorso al GIS può rivelarsi utile non solo in fase di rappresentazione dei risultati, ma anche per la preparazione dei dati e l'elaborazione di informazioni rilevanti da passare poi al software statistico. Nella nostra analisi, ad esempio, il software utilizzato, *BayesX* (Lang e Brezger, 2000), richiedeva un file con le coordinate dei confini di ogni provincia per costruire la matrice delle adiacenze necessarie per la specificazione del processo autoregressivo condizionale (si veda il paragrafo successivo). Ovviamente non è un'operazione semplice codificare a mano questi confini nel caso di oltre 100 province e questa operazione diviene sicuramente impossibile nel momento in cui l'analisi si sposti ad un livello di dettaglio maggiore (si pensi ad un'analisi condotta a livello comunale). Quest'informazione può essere invece ottenuta dal GIS. Nel nostro caso, uno script in Avenue, il macrolinguaggio di ArcView/GIS (ESRI, 1996-b), è stato utilizzato per ricavare il file richiesto.

4. I metodi utilizzati

In questa sezione, dopo una breve rassegna delle motivazioni soggiacenti all'uso di metodi computazionali intensivi in ambito bayesiano (paragrafo 4.1), sono descritte le tecniche statistiche utilizzate nel lavoro (i cui risultati sono analizzati in dettaglio nella sezione 5). In particolare nel paragrafo 4.2 è presentato il modello statistico, e nel paragrafo 4.3 gli algoritmi utilizzati nelle procedure inferenziali.

4.1 Inferenza bayesiana e stima Monte Carlo Markov Chain

L'approccio seguito nel presente articolo è di tipo bayesiano. Come è noto, la caratteristica fondamentale dell'inferenza bayesiana è quella di combinare l'informazione a priori di cui il ricercatore dispone con l'evidenza contenuta nei dati. La prima è formalizzata tramite un'opportuna distribuzione (a priori), mentre la seconda è tipicamente fornita dalla funzione di verosimiglianza. La combinazione delle due funzioni tramite il teorema di Bayes fornisce una nuova legge, detta distribuzione a posteriori o finale, che può essere vista come l'aggiornamento delle informazioni a priori in base ai dati.

Supponendo che l'obiettivo sia fare inferenza rispetto ad un parametro $\theta \in \Theta$, dove Θ rappresenta lo spazio parametrico eventualmente multidimensionale, indicata con $\pi(\theta)$ la distribuzione a priori, con x l'informazione campionaria, e con $f(x|\theta)$ la distribuzione campionaria, la distribuzione a posteriori è data dall'espressione

$\pi(\theta|x) = f(x|\theta)\pi(\theta)/m(x)$ dove $m(x) = \int_{\Theta} f(x|\theta)d\Pi(\theta)$ è denominata distribuzione pre-

dittiva e $\Pi(\theta)$ è la funzione di ripartizione della distribuzione a priori. La distribuzione a posteriori o alcune sue caratteristiche costituiscono lo strumento principale dell'inferenza bayesiana. Ad esempio misure di sintesi come la media, la mediana o la moda forniscono stime puntuali dei parametri di interesse, mentre intervalli di probabilità basati su tale distribuzione forniscono stime intervallari.

Il ricorso alla distribuzione a posteriori determina due ordini di problemi. Il primo è un problema di robustezza dell'inferenza condotta. Si può infatti pensare che il risultato inferenziale dipenda strettamente dalla prior utilizzata. Il ricorso ad a priori piatte, cioè con la massa di probabilità estremamente dispersa sul supporto è stato indicato come un metodo che generalmente fornisce una buona robustezza dei risultati ottenuti (Berger, 1985). Il secondo problema riguarda aspetti più propriamente computazionali. I momenti della legge a posteriori possono non essere immediati da calcolare, spesso gli integrali coinvolti risultano essere altamente multidimensionali e perfino la legge a posteriori, se si escludono pochi casi essenzialmente riconducibili all'analisi delle famiglie coniugate, può non essere espressa in forma chiusa. L'integrazione numerica delle funzioni di interesse è uno dei metodi cui spesso si ricorre per risolvere problemi di questo tipo, ma anche questa via diviene progressivamente sempre più difficilmente percorribile al crescere della dimensione dello spazio parametrico coinvolto.

Un'altra possibilità per il calcolo della distribuzione a posteriori e delle sue caratteristiche è il ricorso all'integrazione stocastica. In particolare nel presente lavoro si è fatto riferimento a metodi Monte Carlo Markov Chain (Gilks, Richardson e Spiegelhalter, 1996), MCMC d'ora in poi. L'idea di base di questi metodi è quella di approssimare la distribuzione a posteriori dei parametri per mezzo di un campione simulato ed approssimare le caratteristiche di tale distribuzione (gli stimatori dei parametri di interesse), tramite le corrispondenti statistiche campionarie. Differentemente dai metodi di integrazione Monte Carlo basati su simulazioni di campioni indipendenti e identicamente distribuite, le generazioni casuali ottenute via MCMC, si basano sul campionamento una catena di Markov che ammette la legge a posteriori come sua distribuzione limite. Le osservazioni risultano fra loro dipendenti. In via di principio, come hanno osservato molti autori (si veda per esempio Gelman, 1996), non ci sono sostanziali problemi nel considerare nell'inferenza osservazioni correlate. Questo può però indurre dei *mixing* rallentati della catena, e quindi richiedere la simulazione di catene molto lunghe. In questo caso la gestione dei risultati può diventare onerosa in termini computazionali. Una soluzione proposta è quella di salvare e utilizzare per l'inferenza solo un'osservazione ogni k simulate (*thinning*). Scegliere il passo k in termini dell'autocorrelazione presente nella catena consente di ridurre al contempo il grado di correlazione presente nel campione effettivamente uti-

lizzato per l'inferenza. Inoltre, potendo la catena trovarsi lontana dalla sua legge limite durante le prime iterazioni, è prassi comune eliminare dai calcoli per determinare le caratteristiche della legge a posteriori le prime M osservazioni generate (tempo di *riscaldamento* o *burn-in*). Ovviamente se la catena generata è particolarmente lunga l'effetto di queste prime osservazioni tenderà ad essere trascurabile.

4.2 Un modello logit bayesiano a tre stadi

Di seguito si utilizza un modello bayesiano a tre stadi di gerarchia. Nel primo stadio del modello, si suppone che la variabile di risposta, Y , sia una variabile indicatrice per l'evento analizzato che assume valore 1 nel caso che un metodo contraccettivo sia stato utilizzato dall'intervistata durante il suo primo rapporto sessuale e 0 altrimenti. Si assume inoltre che Y sia distribuita secondo una distribuzione di Bernoulli di media p e che tale parametro dipenda da un predittore lineare tramite una funzione logit.

Come è stato ampiamente discusso da molta letteratura soprattutto di tipo epidemiologico nell'ambito di analisi sul raggruppamento delle patologie (*disease clustering* si veda ad esempio Elliot e Wakefield 2001) le unità statistiche tendono spesso a raggrupparsi territorialmente, ma anche temporalmente, in riferimento all'evento considerato. Sono state date varie definizioni di *clustering*. Alcune di queste, si richiamano direttamente ad aspetti meramente statistici, come una concentrazione di frequenza difficilmente attribuibile all'effetto del caso (Knox, 1989) o legata ad elevate superfici di rischio in aree o periodi limitati (Wakefield e altri 2000). Knox (1989) definisce in modo più qualitativo e generale un *cluster*. Un *cluster* è quindi caratterizzato dal raggruppamento delle occorrenze di un determinato fenomeno dovuto a meccanismi sociali o biologici o al fatto che esso è legato in modo latente ad altri fattori o circostanze. Tale definizione può evidentemente travalicare l'ambito epidemiologico del *disease clustering* per estendersi a qualunque fenomeno sociale. Campionare gli individui secondo caratteristiche territoriali, come spesso avviene nelle survey, può indurre quindi correlazione fra le osservazioni o un'eterogeneità che può non essere adeguatamente spiegata dalle esplicative considerate.

Per tenere conto di questa fonte di variabilità uno o più effetti casuali possono essere introdotti nel modello. In particolare, nel nostro caso, si può pensare che i log odds varino sul territorio senza una struttura spaziale sottostante, ma solamente per effetti locali. Questa assunzione porta in modo naturale a specificare una componente di varianza all'interno del modello. In realtà, l'assunzione di indipendenza ed uguale distribuzione, per altro presente nella letteratura, non sembra essere molto coerente con un clustering di tipo spaziale in fenomeni relativi alla sfera sociale. In effetti spesso tale clustering appare attribuibile a fattori culturali ed economici che ragionevolmente si possono supporre essere più simili per aree vicine che per aree fra loro distanti. Questo richiede, in altri termini, che gli effetti fra le locazioni siano in qual-

che modo correlati. Nel secondo stadio del modello, quindi, due disturbi casuali sono inseriti nel predittore lineare. Il primo costituito da una variabile casuale, U , che sia assume indipendente e identicamente distribuita per tutte le locazioni e il secondo, S , di tipo strutturato rappresentato cioè da un processo spaziale correlato. Formalmente il modello viene specificato come:

$$\text{logit}(p_{gi}) = X_{gi}'\beta + U_g + S_g \quad g = 1, \dots, G, \quad (1)$$

dove G rappresenta il numero di locazioni considerate (nel nostro caso il numero delle province italiane pari a 103) e il deponente i fa riferimento alla generica i -esima unità campionaria. Per la componente non strutturata si suppone che il vettore $U' = \{U_1, \dots, U_G\}'$ abbia una distribuzione a priori normale di media 0 e matrice di varianza e covarianza $\sigma^2 \mathbf{I}$ (con \mathbf{I} matrice identità e $\sigma^2 > 0$ ignota). Per la componente strutturata $S' = \{S_1, \dots, S_G\}'$, invece, si assume che il modello a priori sia dato da un campo gaussiano markoviano o autoregressione condizionata gaussiana (Besag e alt. 1991). In tal caso indicando con S_{-g} l'insieme degli effetti escluso quello sulla g -esima locazione si assume che $S_g | S_{-g} \sim N(\sum_r w_{gr} S_r, \tau_g^2)$ dove w_{gr} e τ_g^2 sono definiti

in termini della matrice di precisione di \mathbf{S} . Più specificamente è stata utilizzata una forma limite denominata *regressione gaussiana intrinseca* (Besag e Kooperberg, 1995) in cui gli unici vicini rilevanti per la determinazione della legge condizionata sono le locazioni adiacenti. Il sistema dei pesi è tale per cui un peso unitario è attribuito ad ogni vicino e uno nullo a tutte le altre locazioni. In tal caso, indicando con ∂g l'insieme dei vicini della provincia g e con n_g la sua cardinalità, si ha $E(S_g | \partial g) = \sum_{r \in \partial g} S_r / n_g$ e $\text{Var}(S_g | \partial g) = \tau^2 / n_g$, espressioni che rendono palese la struttura markoviana del processo.

Infine, nel terzo stadio del modello, vengono definite le distribuzioni a priori degli effetti fissi³ e degli iperparametri (*hyperprior*). In particolare si è ipotizzato sia per σ^2 che τ^2 una distribuzione gamma inversa altamente dispersa, ma propria (parametro di scala $b=1$ e di forma $a=0.005$). Per gli effetti fissi l'usuale assunzione di una apriori piatta è stata ipotizzata anche nel nostro caso.

4.3 Caratteristiche dell'inferenza MCMC

Come ricordato in precedenza lo strumento fondamentale dell'inferenza bayesiana è costituito dalla distribuzione a posteriori che lega insieme le assunzioni fatte sul fenomeno considerato (la distribuzioni a priori) e l'evidenza dei dati (verosimiglianza).

³ Il termine "fissi" va qui interpretato nel senso bayesiano. In tale contesto, diversamente dalla terminologia usata nell'ambito dei modelli misti in un contesto non bayesiano, tutti i parametri sono delle variabili casuali. Quello che contraddistingue gli effetti fissi dai casuali è sostanzialmente la presenza di un grado di gerarchia superiore (Gelman e alt. 1995).

Per modelli complessi tale distribuzione è spesso analiticamente o perfino numericamente intrattabile. I metodi di simulazione Monte Carlo si basano sul principio di campionare la distribuzione a posteriori, utilizzando poi i campioni simulati per approssimare la legge a posteriori e alcune sue caratteristiche (momenti) rilevanti per l'inferenza. In particolare i metodi MCMC si basano sulla simulazione delle osservazioni da una catena di Markov le cui probabilità di transizione convergono alla distribuzione a posteriori. Il campione ottenuto quindi, dopo l'eliminazione di un numero iniziale di osservazioni che si possono pensare necessaria alla catena per raggiungere la convergenza, è utilizzato per l'inferenza. Per il modello considerato si è inoltre ipotizzato che:

- date le esplicative e l'insieme dei parametri, le osservazioni sono indipendenti
- le distribuzioni a priori per gli effetti fissi, casuali e le hyperprior sono mutualmente indipendenti.

In tal caso, indicando con K il numero di esplicative nel modello, con $p(x|y)$ la legge condizionata di volta in volta considerata e con $L_{gi}(y_{gi})$ il contributo alla verosimiglianza dell' i -esima unità nella g -esima provincia, per $i=1, \dots, n_g$ e $g=1, \dots, G$, la distribuzione a posteriori si fattorizza come:

$$p(s, u, \tau^2, \sigma^2, \beta | y_1, \dots, y_n) \propto \prod_{g=1}^G \prod_{i=1}^{n_g} L(y_{gi} | s, u, \beta) \prod_{g=1}^G p(u_g | \sigma^2) p(\sigma^2) \prod_{g=1}^G p(s_g | s_{-g}, \tau^2) p(\tau^2) \prod_{j=1}^K p(\beta_j).$$

L'inferenza MCMC si basa quindi sulla simulazione di campioni da ciascuna distribuzione condizionata (*full conditional distribution*).

Gli algoritmi usati per le simulazioni MCMC sono descritti in dettaglio in Fahrmeir e Lang (2001). Qui ne riportiamo solo una breve descrizione utile anche per fornire un'idea dello stato dell'arte in tale ambito. Per l'aggiornamento dei valori dei parametri fissi del modello e dell'intercetta casuale non strutturata è stato usato un algoritmo proposto da Gamerman (1997). L'idea fondamentale è quella di inserire ad ogni passo della procedura dei minimi quadrati iterati ponderati (IWLS), usualmente utilizzata per ottenere le stime di massima verosimiglianza nell'ambito dei modelli lineari generalizzati (McCullagh e Nelder, 1989), una generazione via algoritmo Metropolis-Hestings. Per l'aggiornamento della componente strutturata (CAR) l'algoritmo utilizzato è stato originariamente proposto da Knorr-Held (1999) nell'ambito dei modelli lineari generalizzati dinamici. L'algoritmo mira a migliorare il mixing e la convergenza delle catene simulate aggiornando, ad ogni iterata, un intero blocco di valori. Questo tipo di algoritmi traggono notevoli vantaggi computazionali dalla struttura sparsa della matrice di correlazione, struttura che riflette le relazioni di vicinato fra le locazioni. Per ulteriori sviluppi e dettagli su metodi computazionali altamente efficienti per la generazione di campi gaussiani markoviani si veda Rue (2001). Infine le full conditional per i parametri di varianza degli effetti spaziali sono

ancora delle gamma inverse e quindi l'aggiornamento dei valori può essere fatto via Gibbs Sampler (Fahrmeir e Lang, 2001).

5. L'applicazione: aspetti metodologici e risultati

L'analisi relativa a differenziali geografici nell'uso di metodi contraccettivi al primo rapporto sessuale per le donne italiane è stata condotta, sugli stessi dati qui utilizzati, in Borgoni e Billari (2002-a). In quel lavoro è stato evidenziato come, controllando per l'età al primo rapporto sessuale (e per la coorte di appartenenza dell'intervistata), esista un gradiente (negativo) nord-sud nell'uso dei metodi contraccettivi. In altre parole, si ha un rischio del non uso di metodi marcatamente crescente dal nord al sud della penisola. Questo evidenzia un'importante eterogeneità a livello spaziale fra le diverse province italiane, che riflette, tra l'altro, quanto rilevato sull'età al primo rapporto sempre su dati INF-2 (Ongaro, 2001, Billari e Borgoni, 2002). L'età al primo rapporto sessuale risulta invece non avere un effetto significativo. Non ci dilungheremo ulteriormente su quest'analisi in questa sede.

Analoghi risultati relativamente all'esistenza di un'eterogeneità non osservata altamente significativa a livello territoriale e di un effetto dell'età al primo rapporto praticamente inesistente sulla predisposizione alla contraccezione sono stati evidenziati in Borgoni e Billari (2002-b) dove, unitamente ad un'analisi sugli effetti di processi di selezione nel campione, sono state analizzate, tramite un modello ad equazioni simultanee, alcune determinanti dell'uso di contraccettivi al debutto sessuale. In Borgoni e Billari (2002-a) viene mostrato come la coorte di appartenenza dell'intervistata influenza in modo significativo la probabilità di usare un metodo contraccettivo. In particolare, l'uso di metodi contraccettivi al primo rapporto aumenta drasticamente per le coorti più giovani. In questo lavoro, ci si è quindi posti l'obiettivo di approfondire l'evoluzione temporale della struttura spaziale del fenomeno considerato e la stabilità dei differenziali nel corso del tempo. In particolare ci si è proposti di investigare se è ipotizzabile nel caso dell'uso di metodi di contraccezione alla prima esperienza sessuale un processo diffusivo a livello territoriale analogo a quello che sembra essersi verificato per quanto riguarda un'analisi dell'incidenza del primo rapporto sessuale. (Billari e Borgoni, 2002).

Abbiamo dunque ripartito il data set in tre sotto insiemi relativi alla coorti delle nate negli 1946-55, 1956-65 e 1966-75. Il modello (1) è stato applicato separatamente per ciascuna coorte. In particolare, per ogni coorte è stato stimato un modello che includeva la variabile età unitamente ai due effetti casuali spaziali strutturato e non strutturato. I modelli sono stati stimati via MCMC. Un'analisi preliminare è stata effettuata per valutare la correlazione tra i campioni generati per ogni modello relativo a ciascuna coorte.

Tabella 5.1: Stima degli effetti fissi per ogni modello relativo alle tre coorti considerate.

Variabile	Media	Dev. Stand.	1° decile	mediana	9° decile		Variazione odd (%)
<i>Coorte 1946-1955</i>							
Intercetta	-0.666	0.157	-0.868	-0.665	-0.469		
Età 18-20	0.041	0.180	-0.191	0.041	0.272		4.2%
Età 21-24	-0.167	0.192	-0.416	-0.168	0.081		-15.4%
Età oltre 25	0.091	0.229	-0.200	0.088	0.384		9.5%
<i>Coorte 1956-1965</i>							
Intercetta	0.036	0.113	-0.106	0.036	0.177		
Età 18-20	-0.250	0.136	-0.425	-0.251	-0.073	-	-22.1%
Età 21-24	-0.412	0.173	-0.634	-0.412	-0.193	-	-33.8%
Età oltre 25	-0.250	0.217	-0.531	-0.245	0.024		-22.2%
<i>Coorte 1966-1975</i>							
Intercetta	0.534	0.124	0.372	0.535	0.689		
Età 18-20	0.431	0.154	0.233	0.433	0.630	+	53.9%
Età 21-24	0.292	0.193	0.042	0.293	0.541	+	33.8%
Età oltre 25	-0.099	0.358	-0.557	-0.086	0.349		-9.4%

Per le coorti nate nei decenni 1946-55 e 1956-65, le catene presentavano una convergenza più lenta e nel primo caso una più alta correlazione delle osservazioni. Per i modelli relativi a queste due coorti sono state quindi implementate due catene di 160000 iterazioni con un tempo di riscaldamento pari a 10000 campionamenti. Per la coorte più vecchia inoltre si è scelto un passo di thinning di una osservazione ogni 50 anziché di una ogni 30 come nel caso della coorte 1956-65. Per la generazione più giovane invece l'analisi preliminare ha evidenziato una correlazione minore ed è quindi stato implementato un esperimento MCMC con 95000 ricampionamenti, burn-in di 5000 iterazioni e passo di thinning pari a 1 osservazione ogni 30. I risultati sono riportati in tabella 5.1. È interessante notare come, mentre per la coorte più anziana nessuna classe della variabile età risulta significativamente diversa da quella di riferimento, la situazione cambi passando alle coorti successive. In particolare nella coorte 1956-1965 si nota un effetto negativo delle classi di età successive alla classe di riferimento (14-17 anni), mentre nella generazione più recente l'effetto è di segno opposto. Si potrebbe arguire che quest'evidenza sia attribuibile anche ad un mutamento nelle determinanti della pratica contraccettiva e ad un'affermazione della tendenza all'uso di pratiche contraccettive a fini preventivi nei confronti di malattie infettive come l'HIV manifestatesi in modo preoccupante negli ultimi quindici - venti anni, i primi casi registrati in Italia risalgono, infatti, al 1982 (Galli e Rezza, 1998). Conclusioni di questo tipo per altro necessiterebbero di un'analisi più approfondita anche in termini dei metodi contraccettivi utilizzati, informazione non disponibile

nell'indagine considerata. Venendo ad aspetti più propriamente spaziali, l'analisi effettuata è sintetizzata in figura 5.1. In essa sono riportate, nella prima colonna, le mappe relative alla stima (media a posteriori) degli effetti spaziali strutturati sul log odd, mappe (a), (c) e (e), e nella seconda colonna, mappe (b), (d), (f), le tematizzazioni delle loro significatività per ognuna delle tre coorti considerate. Ciascuna riga della tabella di grafici è relativa ad una coorte. Nella prima riga, mappe (a) e (b), sono riportate le tematizzazioni relative alla coorte 1946-55, nella seconda riga, mappe (c) e (d) sono riportate le tematizzazioni relative alla coorte 1956-65 e, infine nella terza riga, mappe (e) e (f), quelle relative alla coorte 1966-75. Le tematizzazioni sono state ottenute utilizzando gli stessi valori soglia per definire i colori utilizzati nella mappa in figura 1-a riportata in Borgoni e Billari (2002-a). La mappa di ogni coorte risulta quindi direttamente comparabile con la mappa relativa all'andamento dell'intera popolazione.

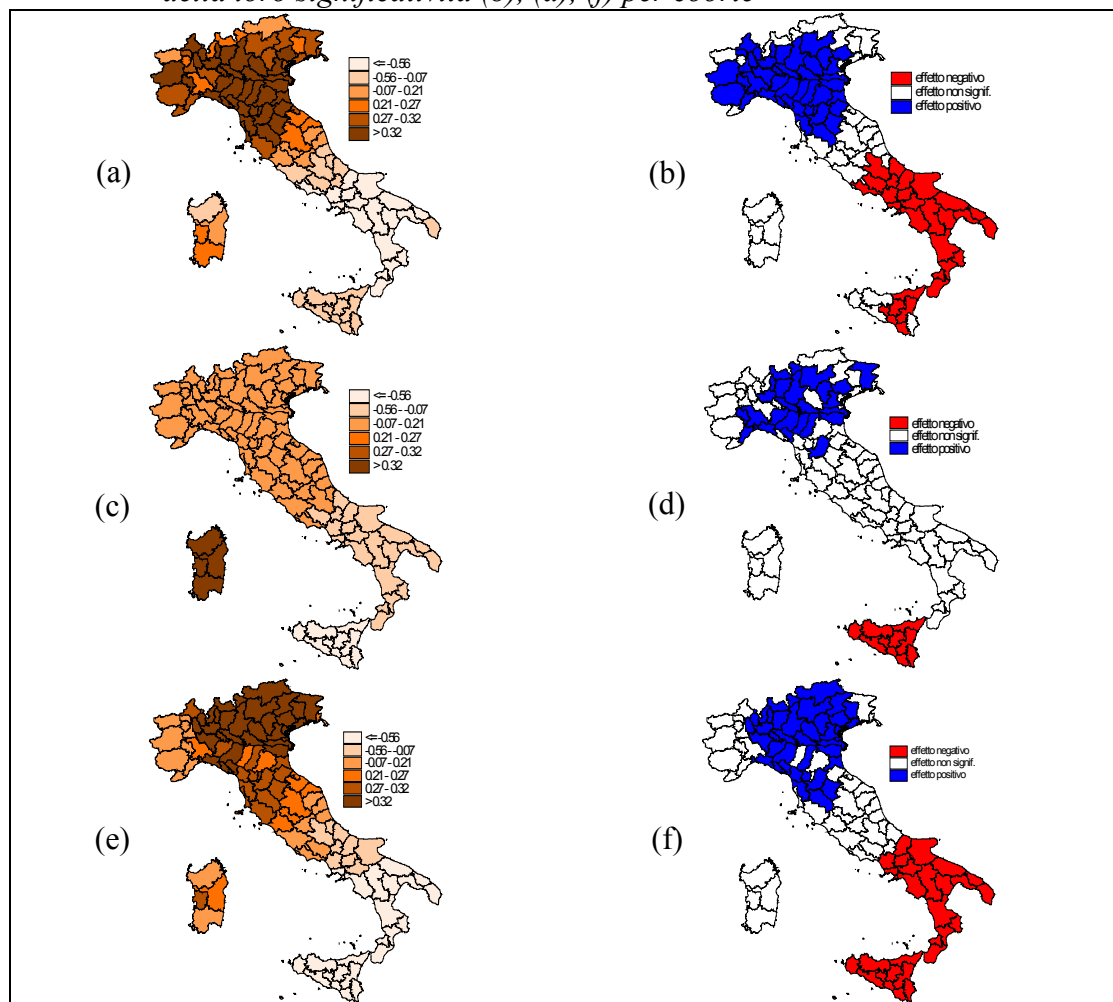
Abbiamo dunque ripartito il data set in tre sotto insiemi relativi alla coorti delle nate negli 1946-55, 1956-65 e 1966-75. Il modello (1) è stato applicato separatamente per ciascuna coorte. In particolare, per ogni coorte è stato stimato un modello che includeva la variabile età unitamente ai due effetti casuali spaziali strutturato e non strutturato. I modelli sono stati stimati via MCMC. Un'analisi preliminare è stata effettuata per valutare la correlazione tra i campioni generati per ogni modello relativo a ciascuna coorte. Per le coorti nate nei decenni 1946-55 e 1956-65, le catene presentavano una convergenza più lenta e nel primo caso una più alta correlazione delle osservazioni. Per i modelli relativi a queste due coorti sono state quindi implementate due catene di 160000 iterazioni con un tempo di riscaldamento pari a 10000 campionamenti. Per la coorte più vecchia inoltre si è scelto un passo di thinning di una osservazione ogni 50 anziché di una ogni 30 come nel caso della coorte 1956-65. Per la generazione più giovane invece l'analisi preliminare ha evidenziato una correlazione minore ed è quindi stato implementato un esperimento MCMC con 95000 ricampionamenti, burn-in di 5000 iterazioni e passo di thinning pari a 1 osservazione ogni 30.

I risultati sono riportati in tabella 5.1. È interessante notare come, mentre per la coorte più anziana nessuna classe della variabile età risulta significativamente diversa da quella di riferimento, la situazione cambi passando alle coorti successive. In particolare nella coorte 1956-1965 si nota un effetto negativo delle classi di età successive alla classe di riferimento (14-17 anni), mentre nella generazione più recente l'effetto è di segno opposto. Si potrebbe arguire che quest'evidenza sia attribuibile anche ad un mutamento nelle determinanti della pratica contraccettiva e ad un'affermazione della tendenza all'uso di pratiche contraccettive a fini preventivi nei confronti di malattie infettive come l'HIV manifestatesi in modo preoccupante negli ultimi quindici - venti anni, i primi casi registrati in Italia risalgono, infatti, al 1982 (Galli e Rezza, 1998). Conclusioni di questo tipo per altro necessiterebbero di un'a-

nalisi più approfondita anche in termini dei metodi contraccettivi utilizzati, informazione non disponibile nell'indagine considerata.

Venendo ad aspetti più propriamente spaziali, l'analisi effettuata è sintetizzata in figura 5.1. In essa sono riportate, nella prima colonna, le mappe relative alla stima (media a posteriori) degli effetti spaziali strutturati sul log odd, mappe (a), (c) e (e), e nella seconda colonna, mappe (b), (d), (f), le tematizzazioni delle loro significatività per ognuna delle tre coorti considerate. Ciascuna riga della tabella di grafici è relativa ad una coorte. Nella prima riga, mappe (a) e (b), sono riportate le tematizzazioni relative alla coorte 1946-55, nella seconda riga, mappe (c) e (d) sono riportate le tematizzazioni relative alla coorte 1956-65 e, infine nella terza riga, mappe (e) e

Figura 5.1: Mappe degli effetti spaziali strutturati (media a posteriori) (a), (c), (e) e della loro significatività (b), (d), (f) per coorte



(f), quelle relative alla coorte 1966-75. Le tematizzazioni sono state ottenute utilizzando gli stessi valori soglia per definire i colori utilizzati nella mappa in figura 1-a riportata in Borgoni e Billari (2002-a). La mappa di ogni coorte risulta quindi direttamente comparabile con la mappa relativa all'andamento dell'intera popolazione.

Risulta immediato osservare come per la coorte centrale (mappe (c) e (d)) gli effetti spaziali siano molto meno evidenti rispetto alle altre due coorti e l'Italia sia caratterizzata da una maggiore omogeneità. Considerando, ad esempio, l'intervallo di probabilità al 90% dato dal quinto e 95-esimo percentile della distribuzione territoriale degli effetti per provincia si ottiene che esso è pari a (-0.93, 0.47) e (-0.96, 0.47) per le coorti 1966-75 e 1946-55 rispettivamente, ma molto più stretto, pari a (-0.73, 0.22), per la coorte centrale. Questo andamento è evidenziato anche dalle stime del parametro di varianza del modello che risulta pari a 0.19, 0.06 e 0.20 rispettivamente per le coorti 1946-55, 1956-65 e 1966-75. In realtà, considerare solo la variabilità strutturata ignorando la componente spazialmente non strutturata può essere talvolta fuorviante o comunque non fornire un'informazione completa. Solamente dal confronto delle due risulta infatti possibile capire quale sia la componente di fondo del fenomeno considerato e quali le aree, se presenti, che si discostano dall'eventuale struttura latente. Per quanto riguarda la componente non strutturata essa presenta una stima a posteriori della varianza pari a 0.007, 0.163 e 0.017 rispettivamente per le tre coorti evidenziando una maggiore eterogeneità spaziale nella coorte centrale; per altro nessuno degli effetti spaziali non strutturati risulta statisticamente significativo in termini dell'intervallo a posteriori a livello di credibilità dell'80%. Anche in un'analisi per coorte quindi la componente strutturata sembra essere quella rilevante negli andamenti territoriali. La stabilità dei differenziali di coorte sembra essere contraria all'idea di una diffusione territoriale dell'utilizzo di contraccettivi al primo rapporto, quindi l'effetto diffusivo notato in altri lavori relativamente al primo rapporto sessuale non sembra essere presente per quello che riguarda le pratiche contraccettive.

6. Conclusioni e possibili sviluppi

In questo lavoro abbiamo mostrato come sia possibile ottenere, utilizzando metodi statistici di tipo spaziale, delle mappe di un indicatore di rischio sociale a partire da informazioni tratte da un'indagine campionaria, la quale non sia necessariamente costruita per assicurare la rappresentatività al livello delle unità territoriali analizzate. L'accento è stato posto, più che sull'analisi delle determinanti della pratica contraccettiva alla prima esperienza sessuale, sulle soluzioni metodologiche che consentano, a partire da dati di survey individuali, di evidenziare aree di disagio. L'approccio bayesiano ha permesso di adottare in modo relativamente agevole un modello comples-

so, adatto però al problema affrontato. Il ricorso ad un sistema di distribuzioni a priori gerarchiche ha permesso inoltre di irrobustire l'analisi fornendo un metodo consono all'analisi di campioni caratterizzati da elevata sparsità delle osservazioni.

Nel caso specifico della nostra applicazione, è inevitabile concludere facendo riferimento ai problemi connessi al dato utilizzato. In particolare, l'assenza di riferimenti nell'intervista al metodo contraccettivo utilizzato limita la potenzialità dell'informazione. In futuro, dunque, sarebbe necessario raccogliere tali informazioni, mantenendone la geo-referenziazione. La geo-referenziazione, nel caso di informazioni relative ad eventi del corso di vita, dovrebbe inoltre essere idealmente variabile, nel senso che dovrebbe essere possibile riferire un evento al contesto geografico di appartenenza dell'individuo al momento della sua sperimentazione.

I nostri risultati mostrano come in particolare nel Mezzogiorno le donne siano maggiormente a rischio rispetto alle loro coetanee del centro-nord al momento del loro primo rapporto sessuale in tutte le coorti. Possiamo ipotizzare che minori informazioni, e maggiori pressioni verso il non utilizzo da parte del partner, siano alla base di tali differenziali territoriali. Appare comunque evidente il bisogno di politiche volte all'educazione a metodi contraccettivi, come rilevato dal programma di azione della Conferenza Internazionale su Popolazione e Sviluppo del Cairo e dalle linee guida del Fondo per la Popolazione delle Nazioni Unite (UNFPA, 2000).

Questo andamento territoriale è abbastanza indipendente dalla coorte di appartenenza sebbene una maggiore omogeneità sussista per la coorte delle nate tra il 1956 e il 1965 e quindi l'effetto diffusivo notato altrove per il primo rapporto sessuale non sembra essere presente per quello che riguarda le pratiche contraccettive. Si potrebbe casomai osservare che la maggiore omogeneità osservata per la coorte centrale possa essere attribuita ad un clima di maggior liberismo nei comportamenti sessuali a seguito di variazioni comportamentali o legislative (si pensa qui in particolare a movimenti come quelli del '68 o alla legislazione in tema di aborto) che hanno investito in modo uniforme tutto il paese. Una conferma a questo sembrerebbe venire dall'effetto dell'età sulla probabilità d'uso di un metodo di contraccezione. L'effetto negativo osservato nella coorte centrale sul rischio di non uso potrebbe in effetti essere dovuto a tale atteggiamento maggiormente liberistico nei confronti della sfera della sessualità. Al contrario, la maggiore propensione all'uso per la coorte più giovane può essere collegata alla presa di coscienza della necessità di una sessualità sicura dovuta al diffondersi di malattie infettive sessualmente trasmissibili, presa di coscienza tanto più marcata quanto più matura l'età in cui l'esperienza sessuale viene vissuta per la prima volta. Evidentemente tali argomentazioni andrebbero supportate con un'analisi più approfondita delle determinanti del fenomeno, della diffusione territoriale delle strutture e dei servizi che si riflettono in una diversa accessibilità alla contraccezione, delle tipologie di contraccezione utilizzate, analisi che vanno oltre gli scopi di questo lavoro e che sono rimandate a successivi approfondimenti.

Un ultimo commento può essere fatto sull'importanza dello sviluppo e dell'applicazione dei metodi della statistica spaziale alle analisi statistiche di fenomeni sociali. L'area in cui gli individui vivono e maturano le loro scelte è fondamentale per determinare il tipo di scelta. Un'importante opportunità per questa analisi è data dallo sviluppo della statistica spaziale, i cui metodi si sono sviluppati enormemente negli ultimi venti anni grazie anche allo sviluppo di strumenti informatici sufficientemente potenti da rendere implementabili nella pratica corrente le soluzioni computazionalmente complesse richieste da tale modellistica. Allo stesso tempo, la disponibilità di strumenti software come i GIS ha reso l'analisi di dati spaziali semplice ad un livello esplorativo, e d'immediata lettura in fase di presentazione dei risultati di modelli multivariati, consentendo in particolare l'elaborazione di mappe altamente dettagliate che possono aiutare il *policy maker* nell'individuazione di aree problematiche. Poiché questi strumenti consentono l'individuazione immediata di caratteristiche locali o tendenze territoriali, il loro uso in connessione con metodi di statistica spaziale avanzati potrà rivestire un ruolo centrale, in futuro, nella descrizione ed interpretazione degli andamenti territoriali dei fenomeni sociali.

Riferimenti bibliografici

- ANSELIN, L. (2000) Computing environments for spatial data, *Journal of Geographical Systems*, **2**: 201-220.
- BAO, S., ANSELIN, L., MARTIN, D., STRALBERG, D. (2000) Seamless integration of spatial statistics and GIS: The S-PLUS for ArcView and the S+ Grassland links, *Journal of Geographical Systems*, **2**, 287-306.
- BERGER, J. O. (1985) *Statistical Decision Theory and Bayesian Analysis*, Springer, New York.
- BESAG, J., KOOPEMBERG, C. (1995) On conditional and intrinsic autoregression, *Biometrika*, **82**, 733-746.
- BESAG, J., YORK, J., MOLLIE A. (1991) Bayesian image restoration, with two applications in spatial statistics, *Ann. Inst. Statist. Math.* **43**, 1-59.
- BILLARI, F.C., BORGONI, R. (2002) Spatial profiles in the analysis of event histories: An application to first sexual intercourse in Italy, in corso di stampa *International Journal of Population Geography*.
- BILLY, J., BREWSTER, K., GRADY, W. (1994) Contextual effects on the sexual behaviour of adolescent women, *Journal of Marriage and the Family*, **56**, 387-404.
- BONARINI, F. (1999) L'uso della contraccezione in Italia: dalla retrospezione del 1979 a quella del 1995-96. In DE SANDRE, P., PINNELLI, A., SANTINI, A. (a

- cura di), *Nuzialità e fecondità in trasformazione: percorsi e fattori del cambiamento*, il Mulino, Bologna, 395-411.
- BORGONI, R., BILLARI, F.C. (2002-a) Bayesian spatial analysis of demographic survey data. An application to contraceptive use at first sexual intercourse. Sottoposto per la pubblicazione, Max Planck Institute for Demographic Research, Rostock.
- BORGONI, R., BILLARI, F.C. (2002-b) A multilevel sample selection probit model with an application to contraceptive use, in corso di stampa negli *Atti della XLI Conferenza della Società Italiana di Statistica*, Milano, giugno 2002.
- BREWSTER, K.L., BILLY, J.O., GRADY, W.R. (1993) Social context and adolescent behaviour: The impact of community on the transition to sexual activity, *Social Forces*, **71**: 713-740.
- BROOKS, S. P., GELMAN, A. (1998) General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, **7**, 434-455.
- CAZZOLA, A. (1999) L'ingresso nella sessualità adulta. In De Sandre, P., Pinnelli, A., Santini, A. (a cura di), *Nuzialità e fecondità in trasformazione: percorsi e fattori del cambiamento*, il Mulino, Bologna, 311-326.
- CRESSIE, N.A. (1993) *Statistics for Spatial Data*, Wiley, New York.
- DE SANDRE, P., ONGARO, F., RETTAROLI, R., SALVINI, S. (1997) *Matrimonio e figli: tra rinvio e rinuncia*. il Mulino, Bologna.
- DIGGLE, P.J. (1983) *Statistical Analysis for Spatial Point Patterns*, Academic Press, London
- DUBERSTEIN LINDBERG, L., BOGGESS, S., PORTER, L., WILLIAMS S. (2000) *Teen risk-taking: A statistical portrait*, Urban Institute, Washington, DC.
- EGIDI, V., ZACCARIN, S. (1996) Women's condition and reproductive behaviour: a multilevel approach to analysis of Italian data. In EAPS-IUSSP, *Evolution or Revolution in European Population. European Population Conference*, Franco Angeli, Milano, 49-64.
- ELLIOTT, P., CUZICK, J., ENGLISH, D., STERN, R. (1992) *Geographical and Environmental Epidemiology: methods for small-areas studies*, Medical Publications, Oxford.
- ELLIOTT, P., WAKEFIELD, J. C., BEST, N.G., BRINGS, D. J. (2000) *Spatial Epidemiology: methods and applications* Oxford University Press, Oxford.
- ELLIOT, P., WAKEFIELD, J. (2001) Disease clusters: should they be investigated, and, if so, when and how, *Journal of the Royal Statistical Society, Series A*, **164**, 3-12.
- ESRI (1996-a) *Using ArcView GIS*, Environmental Systems Research Institute, Redlands, CA

- ESRI (1996-b) *Using Avenue*, Environmental Systems Research Institute, Redlands, CA.
- FAHRMEIR, L., LANG, S. (2001) Bayesian Inference for generalised additive mixed models based on Markov random field priors, *Applied. Statistics*, **50**, 201-220.
- FOTHERINGHAM, A.F., BRUSDON, C., CHARLTON, M. (2000) *Quantitative Geography*, Sage Publications, London.
- GALLI, M., REZZA, G. (1998) L'Aids in Italia, *Le Scienze*, 231.
- GAMERMAN, D. (1997) Sampling from the posterior distribution in generalised linear mixed model, *Statistics and Computing*, **7**, 57-68.
- GELMAN, A., (1996) Inference and monitoring convergence. In: GILKS, W. R., RICHARDSON, S. E SPIEGELHALTER, D. J. (eds), *Monte Carlo Markov Chain in Practice*, Chapman and Hall, London, 131-143.
- GELMAN, A., RUBIN, D. (1992) Inference for iterative simulation using multiple sequences, *Statistical Sciences*, **7**, 457-511.
- GELMAN, A., CARLIN, J. B., STERN, H. S., RUBIN, D. B. (1995) *Bayesian Data Analysis*, Chapman and Hall, London.
- GILKS W.R., RICHARDSON S., SPIEGELHALTER D.J. (eds) (1996) *Monte Carlo Markov Chain in practice*, Chapman and Hall, London.
- HOGAN, D.P., SUN, R., CORNWELL, G.T. (2000) Sexual and Fertility Behaviours of American Females Aged 15-19 Years: 1985, 1990 and 1995, *American Journal of Public Health*, **90**, 1421-1425.
- KNORR-HELD, L. (1999) Conditional prior proposals in dynamic models, *Scandinavian Journal of Statistics*, **26**, 129-144.
- KNOX, G. (1989) Detection of clusters. In ELLIOT, P. (eds) *Methodology of Enquiries into Disease Clustering*, Small Area Health Statistic Unit, London, 17-22.
- KOTCHIK, B.A., SHAFFER A., FOREHAND, R., MILLER, K.S. (2001) Adolescent Sexual Risk Behaviour: A Multi-system Perspective, *Clinical Psychology Review*, **21**, 493-519.
- KU, L., SONENSTEIN, F., PLECK, J. (1994) The dynamics of young men's condom use during and across relationships, *Family Planning Perspectives*, **26**, 246-251.
- KWAN, M. (2000) Analysis of human behaviour in a GIS environment: recent developments and future prospects, *Journal Geographical Systems*, **2**, 85-90.
- LANG, S., BREZGER, A. (2000-b) *BayesX*, University of Munich, Munich.
- LAWSON, A.B. (2001) *Statistical Methods in Spatial Epidemiology*, Wiley New York.
- MCCULLAGH, P., NELDER, J.A. (1989) *Generalised Linear Models*, (2nd ed.) Chapman and Hall, London.
- MAULDON, J., LUKER, C. (1996) The Effects of Contraceptive Education On Method Use at First Intercourse, *Family Planning Perspectives*, **28**, 19-24.

- ONGARO, F. (2001) First sexual intercourse in Italy: a shift towards and ever more personal experience. Articolo presentato alla XXIV IUSSP General Conference, Salvador.
- PANNATIER, Y. (1996) *Variowin: Software for Spatial Data Analysis in 2D*, Springer Verlag, New York.
- RIBEIRO, J.R., DIGGLE, P.J. (1999) GeoS: a statistical library for S-Plus, *Technical Report ST-99-09, Dipartimento di Matematica e Statistica*, Università di Lancaster.
- RIPLEY, B.D. (1981) *Spatial Statistics* Wiley and Sons, New York.
- RIPLEY (1988) *Statistical Inference for Spatial Processes*, University Press, Cambridge.
- RIVELLINI, G., ZACCARIN S. (1999) Comportamenti riproduttivi: biografie individuali e contesto in un'ottica *multilevel*. In DE SANDRE, P., PINNELLI, A., SANTINI, A. (a cura di), *Nuzialità e fecondità in trasformazione: percorsi e fattori del cambiamento*, il Mulino, Bologna, 651-665.
- RUE, H. (2001) Fast sampling of Gaussian Markov random fields, *Journal of the Royal Statistical Society, Series B*, **63**, 325-38.
- SPINELLI, A., FIGÀ TALAMANCA, I., LAURIA, L., and the European Study Group on Infertility and Subfecundity (2000) Patterns of Contraceptive Use in 5 European Countries, *American Journal of Public Health*, **90**, 1403-1408.
- TEITLER, J.O., WEISS, C.C. (2000) Effects of neighbourhood and school environments on transitions to first sexual intercourse, *Sociology of Education*, **73**, 112-32.
- UNFPA (United Nations Population Fund) (2000) *Preventing Infection-Promoting Reproductive Health. UNFPA's response to HIV/AIDS*, United Nations, New York.
- VENABLES, W.N., RIPLEY B.D. (1999) *Modern Applied Statistics with S-Plus*, (terza edizione) Springer, New York.
- WAKEFIELD, J. C., KELSALL, J. E., MORRIS, S. E. (2000): Clustering, cluster detection and spatial variation in risk. In ELLIOTT, P., WAKEFIELD, J. C., BEST, N.G., BRINGS D. J. (eds), *Spatial Epidemiology: methods and applications*, Oxford University Press, Oxford, 128-152.
- WOODS, R., REES, P. (1986) Spatial Demography: themes, issues and progress. In WOODS, R., REES, P. (eds.) *Population Structures and Models: Developments in Spatial Demography*, Allen & Unwin, Sydney.
- ZACCARIN, S. (1996) Un modello multilevel per l'analisi del comportamento riproduttivo. In SIS, *Atti della XXXVIII Riunione Scientifica*, Maggioli Editore, Rimini, 421-428.
- ZANNELLA, F., RINALDELLI, C., DE MARCHIS, A. (1997) Strategia di campionamento e lavoro sul campo. In DE SANDRE P., ONGARO F., RETTAROLI R., SALVINI S. (a cura di), *Matrimonio e figli: tra rinvio e rinuncia*, il Mulino, Bologna, 171-187.

Spatial statistical analysis of cohort behaviour: Changes in contraceptive use at first intercourse in Italy

Summary. *In this paper, we analyse the cohort dynamics in the spatial patterns of the risk of having unprotected first intercourse for Italian women. We rely on georeferenced survey data from the Italian Fertility and Family Survey and we use a bayesian approach and GIS record linkage to depict the spatial pattern of the propensity to use any contraceptive method at first intercourse. We use a logistic regression model with a multilevel structure (in the form of generalised linear mixed model). The spatial pattern is dealt with using an intrinsic Gaussian conditional autoregressive (CAR) error component. Statistical inference is carried out using Markov Chain Monte Carlo simulation. We show that the spatial pattern, with substantial heterogeneity, is stable across birth cohorts, although the effect of age changes in different cohorts. We conclude by arguing that in future research on indicators of social risk, spatial statistics may become particularly important.*

Keywords. *Contraceptive use, generalised linear mixed model, conditional autoregressive model (CAR), MCMC, geographical information system (GIS).*