
BOLLETTINO UNIONE MATEMATICA ITALIANA

Sezione A – La Matematica nella Società e nella Cultura

LUCA LA ROCCA

**Stima bayesiana non parametrica della funzione
di intensità con applicazione alla valutazione
della pericolosità sismica di alcune zone
sismogenetiche italiane**

*Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. 8-A—La
Matematica nella Società e nella Cultura (2005), n.3-1, p. 561–564.*

Unione Matematica Italiana

[<http://www.bdim.eu/item?id=BUMI_2005_8_8A_3-1_561_0>](http://www.bdim.eu/item?id=BUMI_2005_8_8A_3-1_561_0)

L'utilizzo e la stampa di questo documento digitale è consentito liberamente per motivi di ricerca e studio. Non è consentito l'utilizzo dello stesso per motivi commerciali. Tutte le copie di questo documento devono riportare questo avvertimento.

*Articolo digitalizzato nel quadro del programma
bdim (Biblioteca Digitale Italiana di Matematica)
SIMAI & UMI*

<http://www.bdim.eu/>

Stima bayesiana non parametrica della funzione di intensità con applicazione alla valutazione della pericolosità sismica di alcune zone sismogenetiche italiane

LUCA LA ROCCA

Questa tesi si occupa di inferenza non parametrica sulla funzione di intensità da un punto di vista bayesiano. Tale approccio ha una sua tradizione nell'ambito della cosiddetta *analisi della sopravvivenza*, anche perché permette di considerare osservazioni censurate da destra assieme a osservazioni esatte; inoltre, come la tesi mostra, si tratta di un metodo che consente di dare una valutazione di pericolosità sismica non parametrica e variabile nel tempo in un'assegnata zona sismogenetica. In effetti la valutazione probabilistica della pericolosità sismica è una sfida moderna della sismologia alla quale il paradigma bayesiano fornisce un ambito naturale di definizione e l'Italia in particolare è un paese dove nei secoli si è registrata un'intensa attività sismica, con il vantaggio che una lunga tradizione di studi ha reso disponibile un ricco e affidabile catalogo sismico. Alcuni dati storici sono analizzati in questa tesi alla luce della più recente zonazione sismogenetica.

Il lavoro svolto si fonda sulla proposta di una nuova classe di leggi iniziali non parametriche per la distribuzione incognita di una famiglia di variabili positive, ottenuta costruendo la funzione di intensità come mistura di convoluzione di una densità di probabilità con un processo di Poisson composito. In primo luogo si danno condizioni sufficienti affinché la legge iniziale sia ben definita e la funzione di intensità a priori quasi certamente liscia. Poi si suggerisce una procedura di scelta degli iperparametri, doverosamente invariante per cambiamento di scala, che consente di assegnare una funzione di intensità attesa costante, controllando al contempo la variabilità a priori. Successivamente, introducendo opportune variabili latenti, si propone un'approssimazione di tipo MCMC della legge finale, la si implementa e la si valida su alcuni insiemi di dati notevoli. Infine si mostra come, con minimi cambiamenti, quanto sopra si estenda al caso del modello a rischi proporzionali.

La tesi include una rassegna dei modelli probabilistici comunemente impiegati per la valutazione della pericolosità sismica, così come una rassegna della letteratura che ha ispirato la classe di leggi iniziali proposta. In tale letteratura la funzione di intensità iniziale è modellata con l'ausilio di processi gamma pesati, invece che di processi di Poisson composti, cosicché i metodi di questa tesi si adattano a fornire una soluzione approssimata per le corrispondenti leggi finali, nella misura in cui processi gamma pesati si lasciano approssimare da processi di Poisson composti.

1. – Inferenza sulla funzione di intensità.

Si suppone di avere osservato i tempi di sopravvivenza $t_1, \dots, t_n$ assieme alle corrispondenti indicatrici di censura $o_1, \dots, o_n$: se $o_i = 1$ l'evento $\{T_i = t_i\}$ è stato osservato e t_i è *esatto*, mentre se $o_i = 0$ si è osservato l'evento $\{T_i > t_i\}$ e t_i è *censurato da destra*. Le variabili aleatorie positive $T_1, \dots, T_n$ costituiscono il modello statistico dell'esperimento e si assume, come d'abitudine, che esse siano indipendenti e identicamente distribuite secondo la densità di probabilità $\rho(t) \exp\{-\int_0^t \rho(s)ds\}$, $t \geq 0$, dove ρ è la funzione di intensità incognita. Se il meccanismo di censura è *non informativo*, ovvero indipendente dal processo di osservazione, la verosimiglianza del modello è data dall'espressione

$$(1) \quad L(\rho) = \prod_{i=1}^n \rho(t_i)^{o_i} \exp\left\{-\int_0^{t_i} \rho(s)ds\right\},$$

dove, affinché ρ sia una valida funzione di intensità, occorre e basta che si abbia

$$(2) \quad \rho \geq 0, \quad \exists t > 0 : \int_0^t \rho(s)ds < \infty, \quad \int_0^\infty \rho(s)ds = \infty.$$

Da un punto di vista bayesiano si deve prima costruire un processo stocastico ρ che soddisfi le condizioni (2) e poi calcolarne la distribuzione condizionata ai dati; i primi tentativi in tal senso risalgono ai lavori pionieristici di Dykstra e Laud [1] e Padgett e Wei [4].

2. – Valutazione della pericolosità sismica.

L'interesse per la funzione di intensità risiede in generale nella sua interpretazione come densità condizionale istantanea della probabilità di accadimento, in quanto

$$(3) \quad \rho(s) = \lim_{h \downarrow 0} \frac{1}{h} \mathcal{P}(s \leq T_i \leq s+h \mid T_i \geq s), \quad s \geq 0,$$

dove $\mathcal{P}$ è condizionata a ρ stessa, avendo assunto un punto di vista bayesiano. Se poi in particolare $T_1, \dots, T_n$ sono gli intertempi di una sequenza di eventi sismici registrata in una certa zona sismogenetica, l'ultimo dei quali chiaramente censurato da destra, allora la funzione di intensità attesa a posteriori

$$(4) \quad \hat{\rho}(s) = E[\rho(s) \mid T_1 = t_1, \dots, T_{n-1} = t_{n-1}, T_n > t_n], \quad s \geq 0,$$

consente di calcolare come $\hat{\rho}(t_n)$ il cosiddetto *rischio geofisico*, ovvero il numero atteso condizionale istantaneo di eventi per unità di tempo, vale a dire l'intensità stocastica del processo di conteggio degli eventi al tempo di osservazione.

Si rinvia senz'altro al lavoro di rassegna di Vere-Jones [5] per approfondimenti sulla nozione di rischio geofisico, sulla sua valutazione probabilistica e più in generale sull'uso dei processi di conteggio come modelli per sequenze di eventi sismici. Nella tesi si sono analizzati dati reali gentilmente concessi in anteprima dal Gruppo di lavoro per la redazione della mappa di pericolosità sismica prevista dall'Ordinanza PCM 3274 del 20 marzo 2003 (<http://zonesismiche.mi.ingv.it/>).

3. – Una nuova classe di leggi iniziali.

Si propone di costruire la funzione di intensità incognita come

$$(5) \quad \rho(t) = \xi_0 k_0(t) + \sum_{j=1}^{\infty} \xi_j k(t - \sigma_j), \quad t \geq 0,$$

dove $\xi_0, \xi_1, \xi_2, \dots$ sono variabili indipendenti positive, con $\xi_1, \xi_2, \dots$ identicamente distribuite, $\sigma_0 = 0$ e $\sigma_j = \tau_1 + \dots + \tau_j$, $j \geq 1$, con $\tau_1, \tau_2, \dots$ indipendenti da $\xi_0, \xi_1, \xi_2, \dots$ e identicamente distribuite secondo la legge esponenziale di media q^{-1} , mentre k_0 è una funzione reale positiva definita sulla semiretta positiva e integrabile in un intorno dell'origine e infine k è una densità di probabilità sulla retta.

La formula (5) può riscriversi come $\rho(t) = \xi_0 k_0(t) + \int_{(0, \infty)} k(t-s) \mu(ds)$, $t \geq 0$, dove μ è il processo di Poisson composito con tempi di salto σ_j e ampiezze di salto ξ_j ; si vede allora come la costruzione proposta riprenda le misture introdotte da Lo e Weng [3] e sviluppate da James [2] con μ processo *gamma pesato*.

Valgono i due risultati seguenti, in virtù dei quali ρ è ben definita e infinitamente liscia, se tale è k_0 e per esempio le ξ_j seguono una distribuzione gamma mentre k è una densità normale con media nulla.

TEOREMA 1. – Se $E[\xi_0] < \infty$ e $E[\xi_1] < \infty$, le traiettorie del processo stocastico ρ definito dall'equazione (5) sono quasi certamente valide funzioni di intensità.

TEOREMA 2. – Se k_0 e k sono r volte differenziabili con continuità sui loro rispettivi domini di definizione e per ogni $i \in \{0, \dots, r\}$ la derivata i -esima $k^{(i)}$ della densità k è integrabile sulla retta e tale che $|k^{(i)}(x)| \downarrow 0$, per $x \rightarrow -\infty$, allora le traiettorie del processo stocastico ρ definito dall'equazione (5) sono quasi certamente r volte differenziabili con continuità sulla semiretta positiva.

È poi possibile calcolare la funzione di intensità attesa a priori

$$(6) \quad E[\rho(t)] = E[\xi_0]k_0(t) + qE[\xi_1]K(t), \quad t \geq 0,$$

dove $K(t) = \int_{-\infty}^t k(s)ds$, e porla identicamente pari a un valore assegnato $r_0 > 0$, prendendo $E[\xi_0] = E[\xi_1] = r_0 q^{-1}$ e $k_0 = q(1 - K)$. Di qui, con ulteriori calcoli che

consentono di controllare la variabilità a priori, si sviluppa una procedura di scelta degli iperparametri che doverosamente risulta essere invariante per cambiamento di scala.

4. – Calcolo approssimato della legge finale.

Una volta costruito il processo stocastico ρ la sua distribuzione condizionata ai dati può calcolarsi per mezzo della formula di Bayes. In particolare, se la funzione di intensità attesa a posteriori è l'oggetto di interesse, questa si scrive come

$$(7) \quad E[\rho(s) \mid (t_i, o_i)_{i=1}^n] = \frac{E[\rho(s)L(\rho)]}{E[L(\rho)]}, \quad s \geq 0.$$

Ai fini pratici, almeno nel caso in questione, è necessario ricorrere a una qualche approssimazione numerica della formula (7). Tuttavia, in virtù della natura discreta del processo di Poisson composito, ovvero dell'espressione (5), l'implementazione di un'approssimazione di tipo MCMC è quasi immediata. In particolare, con l'ausilio del linguaggio-ambiente R per il calcolo statistico (<http://www.r-project.org/>), si è implementato nella tesi un algoritmo *à la Gibbs* che, oltre tutto, si adatta agevolmente al caso del modello a rischi proporzionali.

BIBLIOGRAFIA

- [1] DYKSTRA R.L. e LAUD P., *A Bayesian nonparametric approach to reliability*, The Annals of Statistics, **9** (1981), 356-367.
- [2] JAMES L.F., *Bayesian calculus for gamma processes with applications to semiparametric intensity models*, Sankhyā, Series A, **65** (2003), 196-223.
- [3] LO A.Y. e WENG C.-S., *On a class of Bayesian nonparametric estimates: II. Hazard rate estimates*, Annals of the Institute of Statistical Mathematics, **41** (1989), 227-245.
- [4] PADGETT W.J. e WEI L.J., *A Bayesian nonparametric estimator of survival probability assuming increasing failure rate*, Communications in Statistics, Part A—Theory and Methods, **10** (1981), 49-63.
- [5] VERE-JONES D., *Forecasting earthquakes and earthquake risk*, International Journal of Forecasting, **11** (1995), 503-538.

Dipartimento di Scienze Sociali, Cognitive e Quantitative,
Università degli Studi di Modena e Reggio Emilia
e-mail: larocca.luca@unimore.it

Dottorato in Statistica Matematica (sede amministrativa:
Università degli Studi di Pavia) - Ciclo XV

Direttore di ricerca: dott.ssa Renata Rotondi, Consiglio Nazionale delle Ricerche,
Istituto di Matematica Applicata e Tecnologie Informatiche, Sezione di Milano