Semi-parametric estimates of the long-term background trend, periodicity, and clustering effect in crime data

庄 建倉 モデリング研究系 准教授

【Abstract】

Past studies have shown that crime behaviors are clustered. This study proposes a spatiotemporal Hawkes-type point-process model, which includes a background component with daily and weekly periodization and a clustering component that is triggered by previous events, for describing the occurrences of violence or robbery related to crimes in the city of Castellon, Spain, during 2012 and 2013. A nonparametric method, called stochastic reconstruction, is used to estimate each component, including daily and weekly periodicity of background rate, spatial background rate, long-term background trend, and the spatial and temporal response function in the triggering component, of the conditional intensity of the model .

【Data】

This study investigates the data of violence or robbery related to crimes in the city of Castellon, Spain, during 2012 and 2013 (Figure 1).

【Model formulation】

We use a space-time point-process model to describe the data, which is completely specified by a conditional intensity function

$$\lambda(t,x,y)=\mu_0\,\mu_t(t)\,\mu_d(t)\,\mu_w(t)\,\mu_b(x,y)+A\int_{-\infty}^{t}\iint_{R^2} g(t-s)\,f(x-u,y-v)\,N(ds\times du\times dv),$$

where where t (day) and (x,y) (km) denote time and location, respectively, $\mu_t(t)$, $\mu_d(t)$, and $\mu_w(t)$ represent the trend term, the daily periodicity, the weekly periodicity in the temporal components of the background rate, respectively, $\mu_b(x,y)$ represents the spatial homogeneity of the background rate, μ_0 and A are constants, and $g(t-s)f(x-u,y-v)$ represents the intensity of the subprocess triggered by an event previously occurring at location (u,v) and time s. In the above, the average values of μ_t, μ_d, μ_w, and μ_b are all normalized to 1, g and f are normalized as probability densities.

【Estimation method and algorithm: Stochastic reconstruction】

We estimate μ_t, μ_d, μ_w, μ_b, g and f non-parametrically by using the stochastic reconstruction method proposed in [1,2,3,4]. Given a realization of the point process, $\{(t_i,x_i,y_i):\ i=1,2,\cdots,n\}$, These functions in the background component can be reconstructed in the following way:

$$\hat{\mu}_t(t)\propto\sum_i w_i^{(t)}\,I(t_i\in[t-\Delta,t+\Delta]),\qquad w_i^{(t)}=\frac{\mu_t(t_i)\,\mu_b(x_i,y_i)}{\lambda(t_i,x_i,y_i)}$$

$$\hat{\mu}_d(t)\propto\sum_i w_i^{(d)}\,I(t_i-\lfloor t_i\rfloor\in[t-\Delta,t+\Delta]),\qquad w_i^{(d)}=\frac{\mu_d(t_i)\,\mu_b(x_i,y_i)}{\lambda(t_i,x_i,y_i)}$$

$$\hat{\mu}_w(t)\propto\sum_i w_i^{(w)}\,I\left(t_i-14\times\left\lfloor\frac{t_i}{14}\right\rfloor\in[t-\Delta,t+\Delta]\right),\qquad w_i^{(d)}=\frac{\mu_w(t_i)\,\mu_b(x_i,y_i)}{\lambda(t_i,x_i,y_i)}$$

$$\hat{\mu}_b(x,y)\propto\sum_i\varphi_i\,Z_h(x-x_i,y-y_i),\qquad \varphi_i=\frac{\mu_0\,\mu_t(t_i)\mu_d(t_i)\mu_w(t_i)\,\mu_b(x_i,y_i)}{\lambda(t_i,x_i,y_i)}$$

$$\hat{g}(t)\propto\sum_{ij}\rho_{ij}\,I(t_j-t_i\in[t-\Delta,t+\Delta]),\qquad \rho_{ij}=\frac{Ag(t_j-t_i)h(x_j-x_i,y_j-y_i)}{\lambda(t_j,x_j,y_j)},\qquad \text{for } j<i,$$

$$\hat{f}(x,y)\propto\sum_{ij}\rho_{ij}\,I(x_j-x_i\in[x-\Delta_x,x+\Delta_x]),$$

where Z_h is the Gaussian kernel function with bandwidth h. In calculation, we applied kernel smoothing to all the functions expect μ_b, the one that is already smoothed, and took correction of edge effect into account. Once the above functions are estimated, we can update μ and A through maximizing the likelihood function

$$\log L=\sum_{i=1}^{n}\log\lambda(t_i,x_i,y_i)-\int_0^T\iint_S\lambda(s,u,v)\,du\,dv\,ds$$

An iterative algorithm is designed to estimate both the functions and the relaxation parameters, μ and A, simultaneously.

【Results】

The results show that the background rate of the occurrence process of violence or robbery related to crimes in the city of Castellon, Spain, during 2012 and 2013, includes clear daily and weekly periodicity (Figures 2 and 3). The reconstructed spatial and temporal response functions in the clustering component imply that, once a crime occurs, it likely trigger another crime within the coming 3 days and within 100 meters in distance. The parameter estimates are $\mu=0.771$ and $A=0.029$.

【References】

Zhuang J., Ogata Y. and Vere-Jones D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97: 369-380.

Zhuang J., Ogata Y. and Vere-Jones D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research*, 109, 05301.

Zhuang, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *Journal of the Royal Statistical Society: Series B*, 68(4):635--653.

Mohler, G. O. , M. B. Short, P. J. Brantingham, F. P. Schoenberg & G. E. Tita (2011) Self-exciting point process modeling of crime, *Journal of the American Statistical Association*, 106:493, 100-108,

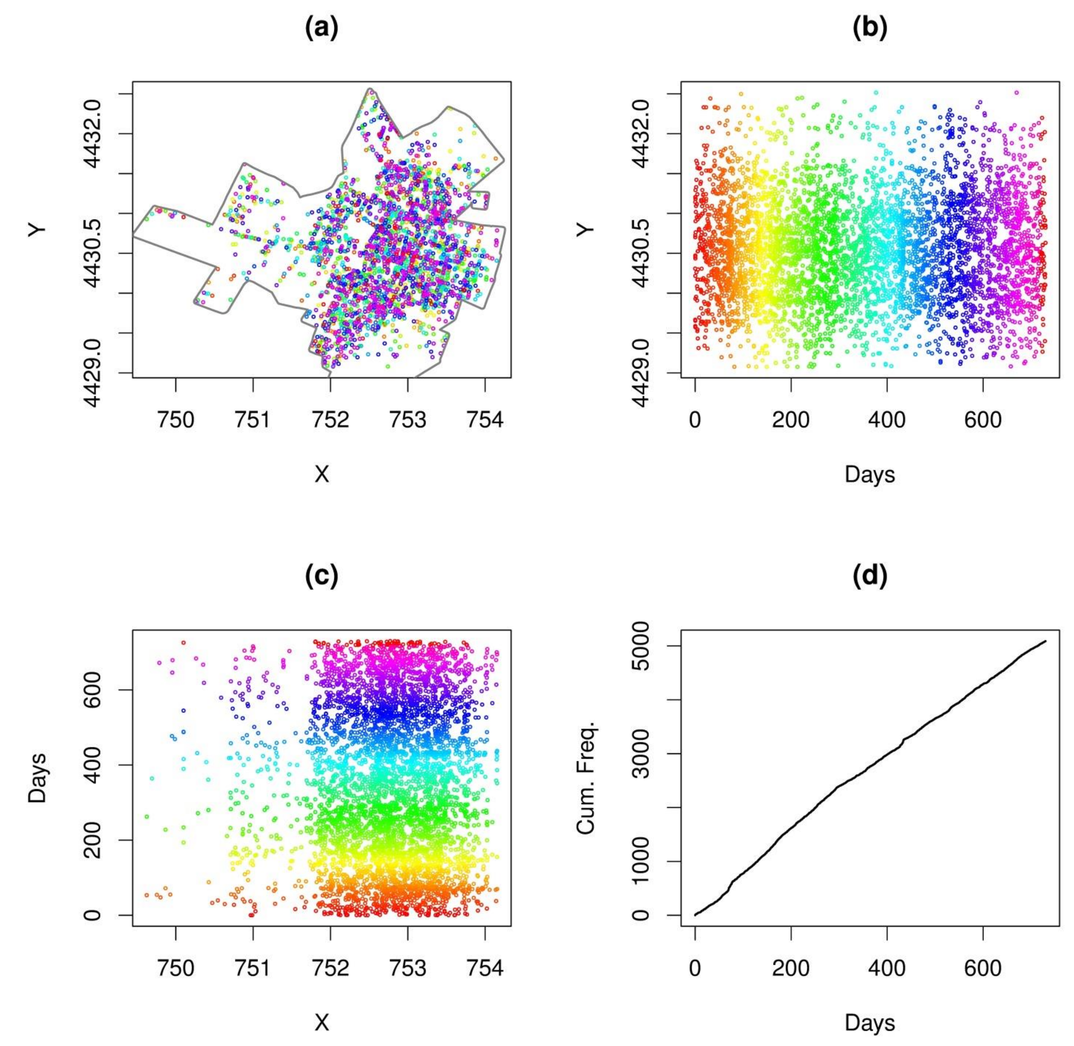

Figure 1. Basic information of the crime data: (a) Spatial locations, (b) y-t space-time plot, (c) t-x space-time plot, and (d) cumulative numbers verse occurrence times.

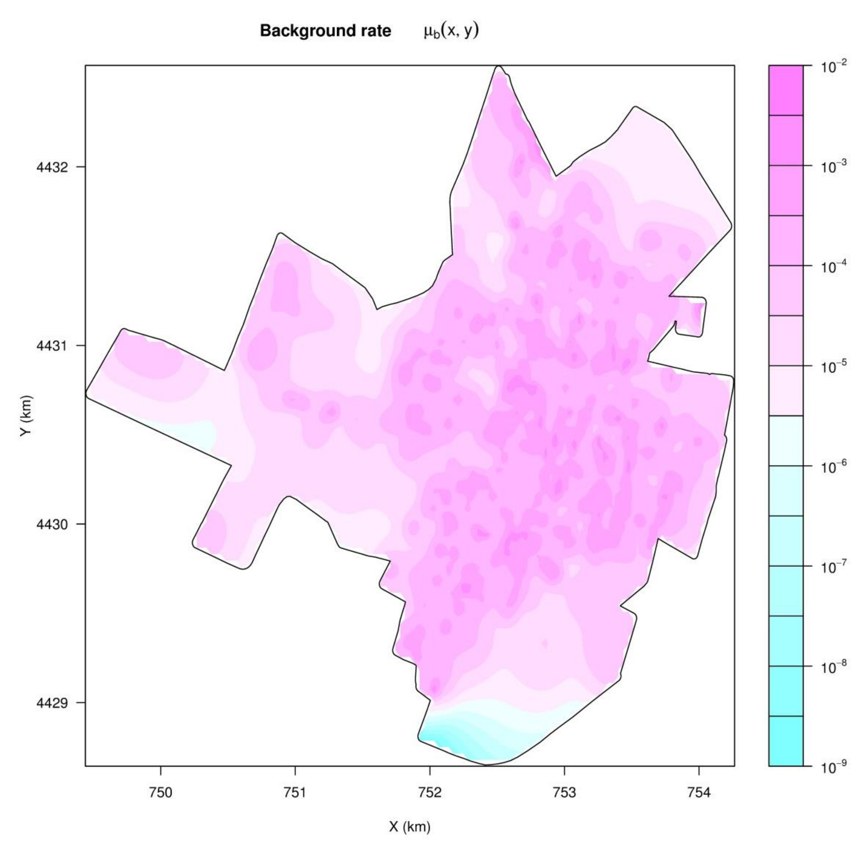

Figure 2. Estimated background rate (unit: events/day/km^2).

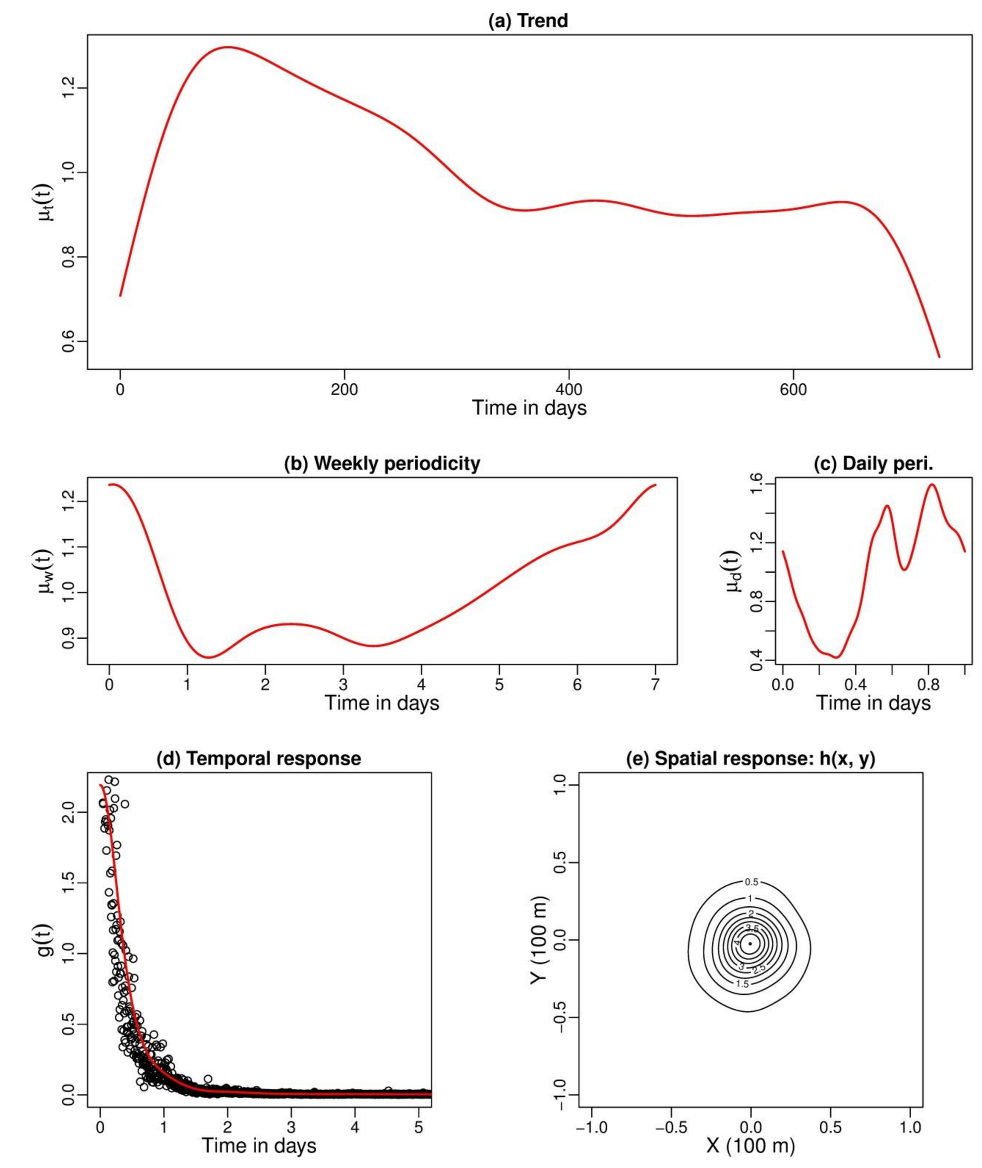

Figure 3. Output results: (a) trend function, (b) weekly periodicity, (c) daily periodicity, (d) temporal response function, and (e) spatial response function.

【Acknowledgment】

This poster is based on joint research work with Prof. Jorge Matteau.