A technical detail of AftFore

Takahiro Omi

A method for forecasting the number of earthquakes with $M > M_t$ in the testing period $[S, T]$ based on the data of earthquakes $\boldsymbol{D} = \{t_i, M_i\}_{i=1}^{N}$ in the testing period $[0, T]$ is described below. Note that we make use of all the observed data including earthquakes below the completeness magnitude.

1. Model description

A joint rate intensity rate of aftershocks at time t after the main shock with magnitude M is modelled by the Omori-Utsu and Gutenberg-Richter laws, given as

$$\lambda(t, M|K, p, c, \beta) = \frac{K}{(t+c)^p} \beta e^{-\beta(M-M_0)}, \tag{1}$$

where K, p, c, and β are parameters and M_0 represents the main shock magnitude. We also consider the detection rate of aftershocks that depends on time and magnitude to consider missing of early aftershocks, given as

$$\Phi(M|\mu(t), \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{M} dx \exp\left[-\frac{\left(x-\mu(t)\right)^2}{2\sigma^2}\right], \tag{2}$$

where $\mu(t)$ is a time-varying parameter that represents the magnitude with 50% detection rate and σ is a parameter representing the magnitude range of partially detected events. To make the following estimation plausible, we decompose the time-varying parameter $\mu(t)$ to the time-varying part $\mu_0(t)$ and the constant term μ_1, $\mu(t) = \mu_0(t) + \mu_1$, and fix the $\mu_0(t)$ to the one estimated by the Bayesian smoothing method proposed in our previous studies (Omi *et al.,* 2013: also see Appendix). In this way, the time-varying parameter $\mu(t)$ is now reduced to a single parameter μ_1. Finally our model is characterized by a parameter set $\theta = \{K, p, c, \beta, \sigma, \mu_1\}$

2. Bayesian Estimation

We here estimate the parameter set θ given the observed aftershock data $\boldsymbol{D}$. In the context of Bayesian statistics, the plausibility of the parameter values given the data is quantified by the posterior probability distribution given by Bayes' theorem as

$$posterior(\theta|\boldsymbol{D}) \propto L(\theta|\boldsymbol{D})prior(\theta), \tag{3}$$

where $L(\theta|\boldsymbol{D})$ and $prior(\theta)$ are the likelihood function and prior probability distribution respectively. If we assume that the observed earthquakes follow the inhomogeneous Poisson process with the intensity rate $\nu_\theta(t, M) = \lambda(t, M|K, p, c, \beta)\Phi(M|\mu(t), \sigma)$, the log-likelihood function can be obtained as

$$\ln L(\theta|\boldsymbol{D}) = \sum_{0<t_i<T} \nu_\theta(t_i, M_i) - \int_{-\infty}^{\infty} dM \int_0^T dt \nu_\theta(t, M). \tag{4}$$

We use independent priors for the p, c, β, and σ parameters, $prior(\theta) = prior(p) \cdot prior(c) \cdot prior(\beta) \cdot prior(\sigma)$. Here the respective prior is given by $N(1.05, 0.13^2)$, $LN(-4.02, 1.42^2)$, $N(1.96, 0.34^2)$, and $LN(-1.61, 1.0^2)$, where N denotes the normal distribution and LN denotes the log-normal distribution based on *Omi et al.*, (2016).

To appropriately account for the estimation uncertainty, we combine the forecasts from many probable parameter sets (Bayesian forecasting). For this purpose, we sample many parameter sets $\{\theta_i\}_{i=1}^m$ from the posterior probability distribution with the Markov chain Monte Carlo method. For our method, we use 1000 parameter sets.

3. Bayesian Forecasting

Given a parameter set θ, the predictive distribution $\mathrm{P}(n|\theta, M_t)$ of the number n of earthquakes with with $M > M_t$ in the testing period $[S, T]$ is the Poisson distribution with mean given by

$$\bar{n} = \int_{M_t}^{\infty} dM \int_0^T dt\, \lambda(t, M|K, p, c, \beta). \tag{5}$$

For the Bayesian forecasting, the predictive distribution $\mathrm{P}(n|\{\theta_i\}_{i=1}^m, M_t)$ is given by

$$\mathrm{P}(n|\{\theta_i\}_{i=1}^m, M_t) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{P}(n|\theta_i, M_t). \tag{6}$$

Appendix . Bayesian smoothing method for the time-varying detection rate

A time-varying detection rate is estimated based on the Bayesian smoothing method. We first discretize the time-varying parameter $\mu(t)$ as $\mu(t) = \mu_i \; (t_{i-1} < t \le t_i)$, where

t_i is the occurrence time of i-th aftershock and we set $t_0 = 0$. Thus the time-varying parameter $\mu(t)$ is now represented by a N-dimensional vector $\boldsymbol{\mu} = \{\mu_i\}_{i=1}^{N}$, where N is the number of observed aftershocks in the learning period.

The likelihood function of $\boldsymbol{\mu}$ given the observed magnitude sequence $\boldsymbol{M} = \{M_i\}_{i=1}^{N}$ is given by

$$P_{\beta,\sigma}(\boldsymbol{M}|\boldsymbol{\mu}) = \prod_{i=1}^{N} \beta e^{-\beta(M_i-\mu_i)-\frac{(\beta\sigma)^2}{2}} \Phi(M_i|\mu_i,\sigma), \tag{7}$$

(see *Omi et al.,* 2013). To estimate $\boldsymbol{\mu}$**,** which has the same length as the data, we introduce smoothness constraint that penalizes the time-variation of $\boldsymbol{\mu}$, given as

$$\mathrm{P_V}(\boldsymbol{\mu}) = \prod_{i=3}^{N} \frac{1}{\sqrt{2\pi V}} e^{-\frac{(\mu_i-2\mu_{i-1}+\mu_{i-2})^2}{2V}}, \tag{8}$$

where V is a hyper-parameter that controls the smoothness of $\boldsymbol{\mu}$. From the Bayes' theorem, the posterior probability distribution of $\boldsymbol{\mu}$ given the data $\boldsymbol{M}$ under the hyper parameters $\{\beta,\sigma,V\}$ is given by

$$\mathrm{P_{\beta,\sigma,V}}(\boldsymbol{\mu}|\boldsymbol{M}) \propto \mathrm{P_{\beta,\sigma}}(\boldsymbol{M}|\boldsymbol{\mu})\mathrm{P_V}(\boldsymbol{\mu}). \tag{3}$$

The MAP estimate $\boldsymbol{\mu}^*$ given the hyper-parameters $\{\beta,\sigma,V\}$,
$\boldsymbol{\mu}^* = \arg\max_{\boldsymbol{\mu}} P_{\beta,\sigma,\mathrm{V}}(\boldsymbol{\mu}|\boldsymbol{M})$, can be readily found by using the Newton method.

The Bayesian smoothing method aims to find the MAP estimate $\boldsymbol{\mu}^*$ under the optimal estimates of the hyper-parameters $\{\beta,\sigma,V\}$. The hyper-parameters are optimized by maximizing the posterior probability distribution of the hyper-parameters given as

$$P(\beta,\sigma,V|\boldsymbol{M}) \propto P(\boldsymbol{M}|\beta,\sigma,V)P(\beta,\sigma,V). \tag{3}$$

Here $\mathrm{P}(\boldsymbol{M}|\beta,\sigma,\mathrm{V})$ is the marginal likelihood function,

$$P(\boldsymbol{M}|\beta,\sigma,V) = \int \boldsymbol{d\mu} P_{\beta,\sigma}(\boldsymbol{M}|\boldsymbol{\mu})P_V(\boldsymbol{\mu}), \tag{3}$$

and we approximate it using the Laplace approximation as

$$P(\boldsymbol{M}|\beta,\sigma,V) \approx (2\pi)^{\frac{N}{2}}|-H|^{-\frac{1}{2}} P_{\beta,\sigma}(\boldsymbol{M}|\boldsymbol{\mu}^*)P_V(\boldsymbol{\mu}^*), \tag{3}$$

where $\boldsymbol{\mu}^*$ is the MAP estimate, and H is the Hessian of $\ln P_{\beta,\sigma,\mathrm{V}}(\boldsymbol{\mu}|\boldsymbol{M})$ at $\boldsymbol{\mu} = \boldsymbol{\mu}^*$. $\mathrm{P}(\boldsymbol{M}|\beta,\sigma,\mathrm{V})$ is the prior probability distribution of the hyper-parameters. We employ the priors for the β and σ, and set them to the same one as are employed in Section 2. The hyper-parameters are optimized using the Quasi Newton method, where the gradient is numerically obtained.

References:

T. Omi, Y. Ogata, Y. Hirata, and K. Aihara, “Forecasting large aftershocks within one day after the main shock”, Scientific Reports 3, 2218 (2013).

T. Omi, Y. Ogata, Y. Hirata, and K. Aihara, “Estimating the ETAS model from an early aftershock sequence”, Geophysical Research Letters 41, 850 (2014).

T. Omi, Y. Ogata, Y. Hirata, and K. Aihara, “Intermediate-term forecasting of aftershocks from an early aftershock sequence: Bayesian and ensemble forecasting approaches”, Journal of Geophysical Research: Solid Earth 120, 2561 (2015).

T. Omi, Y. Ogata, K. Shiomi, K. Sawazaki, and K. Aihara, “Automatic aftershock forecasting -A test using real-time seismicity data in Japan”, Accepted in BSSA (2016).