Bivariate Count Processes for Earthquake Frequency

Mathieu Boudreault & Arthur Charpentier

Université du Québec à Montréal

charpentier.arthur@uqam.ca

http://freakonometrics.blog.free.fr/

Séminaire GeoTop, January 2012
Figure: Time and distance distribution (to 6,000 km) of large (5<M<7) aftershocks from 205 \textit{M} \geq 7 \text{ mainshocks} \text{ (in sec. and h.)}.
Motivation

“Large earthquakes are known to trigger earthquakes elsewhere. Damaging large aftershocks occur close to the mainshock and microearthquakes are triggered by passing seismic waves at significant distances from the mainshock. It is unclear, however, whether bigger, more damaging earthquakes are routinely triggered at distances far from the mainshock, heightening the global seismic hazard after every large earthquake. Here we assemble a catalogue of all possible earthquakes greater than M5 that might have been triggered by every M7 or larger mainshock during the past 30 years. [...] We observe a significant increase in the rate of seismic activity at distances confined to within two to three rupture lengths of the mainshock. Thus, we conclude that the regional hazard of larger earthquakes is increased after a mainshock, but the global hazard is not.” Parsons & Velasco (2011)

Figure: Number of earthquakes (magnitude exceeding 2.0, per 15 sec.) following a large earthquake (of magnitude 6.5), normalized so that the expected number of earthquakes before and after is 100.
Arthur CHARPENTIER & Mathieu BOUDREULT, Bivariate count processes for earthquake frequency
Number of earthquakes before and after a major one, magnitude of the main event, small events more...
Arthur CHARPENTIER & Mathieu BOUDREULT, Bivariate count processes for earthquake frequency

- Number of earthquakes (magnitude > 2) per 15 sec., average before=100

- Time before and after a major earthquake (magnitude > 6.5) in days

- Red: Same tectonic plate as major one
- Blue: Different tectonic plate as major one
Shapefiles from
http://www.colorado.edu/geography/foote/maps/assign/hotspots/hotspots.html
We look at seismic risks with the eyes of actuaries and statisticians...
Active Volcanoes, Plate Tectonics, and the “Ring of Fire”

Eurasian Plate

North American Plate

Arabian Plate

Indo-Australian Plate

Nazca Plate

Pacific Plate

Cocos Plate

Kermadec-Tonga Trench

Java Trench

East Pacific Rise

Hawaiian “Hot Spot”

American Plate

Antarctic Plate

Agenda

- Motivation (Parsons & Velasco (2011))
- Modeling dynamics
  - AR(1): Gaussian autoregressive processes (as a starting point)
  - VAR(1): multiple AR(1) processes, possible correlated
  - INAR(1): autoregressive processes for counting variates
  - MINAR(1): multiple counting processes
- Application to earthquakes frequency
  - counting earthquakes on tectonic plates
  - causality between different tectonic plates
  - counting earthquakes with different magnitudes
Part 1

Modeling dynamics of counts
(ANSS) [http://www.ncedc.org/cnss/catalog-search.html](http://www.ncedc.org/cnss/catalog-search.html)

Number of earthquakes (Magnitude $\geq 5$) per month, worldwide
(ANSS) http://www.ncedc.org/cnss/catalog-search.html

Number of earthquakes (Magnitude $\geq 5$) per month, in western U.S.
(Gaussian) Auto Regressive processes $AR(1)$

**Definition** A time series $(X_t)_{t\in\mathbb{N}}$ with values in $\mathbb{R}$ is called an $AR(1)$ process if

$$X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t$$

(1)

for all $t$, for real-valued parameters $\phi_0$ and $\phi_1$, and some i.i.d. random variables $\varepsilon_t$ with values in $\mathbb{R}$.

It is common to assume that $\varepsilon_t$ are independent variables, with a Gaussian distribution $\mathcal{N}(0, \sigma^2)$, with density

$$\varphi(\varepsilon) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{\varepsilon^2}{2\sigma^2} \right), \quad \varepsilon \in \mathbb{R}.$$  

Note that we assume also that $\varepsilon_t$ is independent of $X_{t-1}$, i.e. past observations $X_0, X_1, \cdots, X_{t-1}$. Thus, $(\varepsilon_t)_{t\in\mathbb{N}}$ is called the innovation process.
Example: $X_t = \phi_1 X_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim \mathcal{N}(0, 1)$, i.i.d., and $\phi = 0.6$
Example: \( X_t = \phi_1 X_{t-1} + \varepsilon_t \) with \( \varepsilon_t \sim \mathcal{N}(0, 1) \), i.i.d., and \( \phi = 0.6 \)
Example: $X_t = \phi_1 X_{t-1} + \varepsilon_t$: autocorrelation $\rho(h) = \text{corr}(X_t, X_{t-h}) = \phi_h^1$
**Definition** A time series \((X_t)_{t \in \mathbb{N}}\) is said to be (weakly) **stationary** if

- \(\mathbb{E}(X_t)\) is independent of \(t\) \((\mathbb{E}(X_t) = \mu)\)
- \(\text{cov}(X_t, X_{t-h})\) is independent of \(t\) \((\text{cov}(X_t, X_{t-h}) = \gamma(h))\), called **autocovariance** function

**Remark** As a consequence, \(\text{var}(X_t) = \mathbb{E}([X_t - \mathbb{E}(X_t)]^2)\) is independent of \(t\) \((\text{var}(X_t) = \gamma(0))\). Define the **autocorrelation** function \(\rho(\cdot)\) as

\[
\rho(h) := \text{corr}(X_t, X_{t-h}) = \frac{\text{cov}(X_t, X_{t-h})}{\sqrt{\text{var}(X_t)\text{var}(X_{t-h})}} = \frac{\gamma(h)}{\gamma(0)}, \quad \forall h \in \mathbb{N}.
\]

**Proposition** \((X_t)_{t \in \mathbb{N}}\) is a stationary AR(1) time series if and only if \(\phi_1 \in (-1, 1)\).

**Remark** If \(\phi_1 = 1\), \((X_t)_{t \in \mathbb{N}}\) is called a **random walk**.

**Proposition** If \((X_t)_{t \in \mathbb{N}}\) is a stationary AR(1) time series,

\[
\rho(h) = \phi_1^h, \quad \forall h \in \mathbb{N}.
\]
From univariate to multivariate models

**Univariate** gaussian distribution $\mathcal{N}(0, \sigma^2)$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{x^2}{2\sigma^2} \right), \text{ for all } x \in \mathbb{R}$$

**Multivariate** gaussian distribution $\mathcal{N}(0, \Sigma)$

$$\varphi(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d |\det \Sigma|}} \exp \left( -\frac{\mathbf{x}'\Sigma^{-1}\mathbf{x}}{2} \right), \text{ for all } \mathbf{x} \in \mathbb{R}^d.$$  

$\mathbf{X} = \mathbf{A}\mathbf{Z}$ where $\mathbf{A}\mathbf{A}' = \Sigma$ and $\mathbf{Z} \sim \mathcal{N}(0, \mathbb{I})$  

(geometric interpretation)
**Vector (Gaussian) AutoRegressive processes \( VAR(1) \)**

**Definition** A time series \( (X_t = (X_{1,t}, \cdots, X_{d,t}))_{t \in \mathbb{N}} \) with values in \( \mathbb{R}^d \) is called a \( VAR(1) \) process if

\[
\begin{align*}
X_{1,t} &= \phi_{1,1} X_{1,t-1} + \phi_{1,2} X_{2,t-1} + \cdots + \phi_{1,d} X_{d,t-1} + \varepsilon_{1,t} \\
X_{2,t} &= \phi_{2,1} X_{1,t-1} + \phi_{2,2} X_{2,t-1} + \cdots + \phi_{2,d} X_{d,t-1} + \varepsilon_{2,t} \\
&\vdots \\
X_{d,t} &= \phi_{d,1} X_{1,t-1} + \phi_{d,2} X_{2,t-1} + \cdots + \phi_{d,d} X_{d,t-1} + \varepsilon_{d,t}
\end{align*}
\]

or equivalently

\[
\begin{pmatrix}
X_{1,t} \\
X_{2,t} \\
\vdots \\
X_{d,t}
\end{pmatrix} =
\begin{pmatrix}
\phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,d} \\
\phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,d} \\
\vdots & \vdots & \cdots & \vdots \\
\phi_{d,1} & \phi_{d,2} & \cdots & \phi_{d,d}
\end{pmatrix}
\begin{pmatrix}
X_{1,t-1} \\
X_{2,t-1} \\
\vdots \\
X_{d,t-1}
\end{pmatrix} +
\begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t} \\
\vdots \\
\varepsilon_{d,t}
\end{pmatrix}
\]
for all $t$, for some real-valued $d \times d$ matrix $\Phi$, and some i.i.d. random vectors $\epsilon_t$ with values in $\mathbb{R}^d$.

It is common to assume that $\epsilon_t$ are independent variables, with a Gaussian distribution $\mathcal{N}(\mathbf{0}, \Sigma)$, with density

$$\varphi(\epsilon) = \frac{1}{\sqrt{(2\pi)^d | \det \Sigma |}} \exp \left( - \frac{\epsilon' \Sigma^{-1} \epsilon}{2} \right), \quad \forall \epsilon \in \mathbb{R}^d.$$ 

Thus, independent means *time independent*, but can be dependent componentwise.

Note that we assume also that $\epsilon_t$ is independent of $\mathbf{X}_{t-1}$, i.e. past observations $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{t-1}$. Thus, $(\epsilon_t)_{t \in \mathbb{N}}$ is called the innovation process.

**Definition** A time series $(\mathbf{X}_t)_{t \in \mathbb{N}}$ is said to be (weakly) stationary if

- $\mathbb{E}(\mathbf{X}_t)$ is independent of $t \ (= : \mu)$
- $\text{cov}(\mathbf{X}_t, \mathbf{X}_{t-h})$ is independent of $t \ (= : \gamma(h))$, called autocovariance matrix
Remark As a consequence, \( \text{var}(X_t) = \mathbb{E}((X_t - \mathbb{E}(X_t))'[X_t - \mathbb{E}(X_t)]) \) is independent of \( t (=: \gamma(0)) \). Define finally the autocorrelation matrix,

\[
\rho(h) = \Delta^{-1} \gamma(h) \Delta^{-1}, \quad \text{where} \quad \Delta = \text{diag} \left( \sqrt{\gamma_{i,i}(0)} \right).
\]

Proposition \( (X_t)_{t \in \mathbb{N}} \) is a stationary AR(1) time series if and only if the \( d \) eigenvalues of \( \Phi \) should have a norm lower than 1.

Proposition If \( (X_t)_{t \in \mathbb{N}} \) is a stationary AR(1) time series,

\[
\rho(h) = \Phi^h, \quad h \in \mathbb{N}.
\]
Statistical inference for AR(1) time series

Consider a series of observations $X_1, \cdots , X_n$. The likelihood is the joint distribution of the vectors $\mathbf{X} = (X_1, \cdots , X_n)$, which is not the product of marginal distribution, since consecutive observations are not independent ($\text{cov}(X_t, X_{t-h}) = \phi^h$). Nevertheless

$$L(\phi, \sigma; (X_0, \mathbf{X})) = \prod_{t=1}^{n} \pi_{\phi,\sigma}(X_t|X_{t-1})$$

where $\pi_{\phi,\sigma}(\cdot|X_{t-1})$ is a Gaussian density.

Maximum likelihood estimators are

$$(\hat{\phi}, \hat{\sigma}) \in \text{argmax} \log L(\phi, \sigma; (X_0, \mathbf{X}))$$
**Poisson distribution - and process - for counts**

$N$ as a Poisson distribution is $\mathbb{P}(N = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ where $k \in \mathbb{N}$.

If $N \sim \mathcal{P}(\lambda)$, then $\mathbb{E}(N) = \lambda$.

$(N_t)_{t \geq 0}$ is an homogeneous Poisson process, with parameter $\lambda \in \mathbb{R}^+$ if:
- on time frame $[t, t + h]$, $(N_{t+h} - N_t) \sim \mathcal{P}(\lambda \cdot h)$
- on $[t_1, t_2]$ and $[t_3, t_4]$ counts are independent, if $0 \leq t_1 < t_2 < t_3 < t_4$, $(N_{t_2} - N_{t_1}) \perp \perp (N_{t_4} - N_{t_3})$
Poisson processes and counting models

Earthquake count models are mostly based upon the Poisson process (see Utsu (1969), Gardner & Knopoff (1974), Lomnitz (1974), Kagan & Jackson (1991)), Cox process (self-exciting, cluster or branching processes, stress-release models (see Rathbun (2004) for a review), or Hidden Markov Models (HMM) (see Zucchini & MacDonald (2009) and Orfanogiannaki et al. (2010)).

See also Vere-Jones (2010) for a summary of statistical and stochastic models in seismology. Recently, Shearer & Starkb (2012) and Beroza (2012) rejected homogeneous Poisson model,
**Thinning operator** \( \circ \)

Steutel & van Harn (1979) defined a thinning operator as follows

**Definition** Define operator \( \circ \) as

\[
p \circ N = Y_1 + \cdots + Y_N \quad \text{if } N \neq 0, \text{ and } 0 \text{ otherwise},
\]

where \( N \) is a random variable with values in \( \mathbb{N} \), \( p \in [0, 1] \), and \( Y_1, Y_2, \cdots \) are i.i.d. Bernoulli variables, independent of \( N \), with \( \mathbb{P}(Y_i = 1) = p = 1 - \mathbb{P}(Y_i = 0) \). Thus \( p \circ N \) is a compound sum of i.i.d. Bernoulli variables.

Hence, given \( N \), \( p \circ N \) has a binomial distribution \( \mathcal{B}(N, p) \).

Note that \( p \circ (q \circ N) \overset{L}{=} [pq] \circ N \) for all \( p, q \in [0, 1] \).

Further

\[
\mathbb{E}(p \circ N) = p\mathbb{E}(N) \quad \text{and} \quad \text{var}(p \circ N) = p^2\text{var}(N) + p(1 - p)\mathbb{E}(N).
\]
(Poisson) **Integer AutoRegressive processes** \(\text{INAR}(1)\)

Based on that thinning operator, **Al-Osh & Alzaid (1987)** and **McKenzie (1985)** defined the integer autoregressive process of order 1:

**Definition** A time series \((X_t)_{t \in \mathbb{N}}\) with values in \(\mathbb{R}\) is called an **INAR(1)** process if

\[
X_t = p \circ X_{t-1} + \varepsilon_t,
\]

where \((\varepsilon_t)\) is a sequence of i.i.d. integer valued random variables, i.e.

\[
X_t = \sum_{i=1}^{X_{t-1}} Y_i + \varepsilon_t, \text{ where } Y_i's \text{ are } i.i.d. \mathcal{B}(p).
\]

Such a process can be related to Galton-Watson processes with immigration, or physical branching model.
Let \( X_{t+1} \) be defined as:

\[
X_{t+1} = \sum_{i=1}^{X_t} Y_i + \varepsilon_{t+1},
\]

where \( Y_i \)'s are i.i.d. \( B(p) \).
Proposition $\mathbb{E}(X_t) = \frac{\mathbb{E}(\varepsilon_t)}{1 - p}$, $\text{var}(X_t) = \gamma(0) = \frac{p\mathbb{E}(\varepsilon_t) + \text{var}(\varepsilon_t)}{1 - p^2}$ and 

$$\gamma(h) = \text{cov}(X_t, X_{t-h}) = p^h.$$ 

It is common to assume that $\varepsilon_t$ are independent variables, with a Poisson distribution $\mathcal{P}(\lambda)$, with probability function

$$\mathbb{P}(\varepsilon_t = k) = e^{-\lambda} \frac{\lambda^k}{k!}, k \in \mathbb{N}.$$ 

Proposition If $(\varepsilon_t)$ are Poisson random variables, then $(N_t)$ will also be a sequence of Poisson random variables.

Note that we assume also that $\varepsilon_t$ is independent of $X_{t-1}$, i.e. past observations $X_0, X_1, \cdots, X_{t-1}$. Thus, $(\varepsilon_t)_{t \in \mathbb{N}}$ is called the innovation process.

Proposition $(X_t)_{t \in \mathbb{N}}$ is a stationary INAR(1) time series if and only if $p \in [0, 1)$.

Proposition If $(X_t)_{t \in \mathbb{N}}$ is a stationary INAR(1) time series, $(X_t)_{t \in \mathbb{N}}$ is an homogeneous Markov chain.
\[ \pi(x_t, x_{t-1}) = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}) = \sum_{k=0}^{x_t} \mathbb{P} \left( \sum_{i=1}^{x_t-1} Y_i = x_t - k \right) \cdot \mathbb{P}(\varepsilon = k). \]
Inference of Integer AutoRegressive processes \textit{INAR}(1)

Consider a Poisson INAR(1) process, then the likelihood is

$$L(p, \lambda; X_0, \mathbf{X}) = \left[ \prod_{t=1}^{n} f_t(X_t) \right] \cdot \frac{\lambda^{X_0}}{(1 - p)^{X_0} X_0!} \exp \left( -\frac{\lambda}{1 - p} \right)$$

where

$$f_t(y) = \exp(-\lambda) \sum_{i=0}^{\min\{X_t, X_{t-1}\}} \frac{\lambda^{y-i}}{(y-i)!} \left( \frac{Y_{t-1}}{i} \right)^{p^i(1 - p)^{Y_{t-1} - y}}, \text{ for } t = 1, \cdots, n.$$ 

Maximum likelihood estimators are

$$(\hat{p}, \hat{\lambda}) \in \arg\max \log L(p, \lambda; (X_0, \mathbf{X}))$$
Multivariate Integer Autoregressive processes \( MINAR(1) \)

Let \( N_t := (N_{1,t}, \ldots, N_{d,t}) \), denote a multivariate vector of counts.

**Definition** Let \( P := [p_{i,j}] \) be a \( d \times d \) matrix with entries in \([0, 1]\). If \( N = (N_1, \ldots, N_d) \) is a random vector with values in \( \mathbb{N}^d \), then \( P \circ N \) is a \( d \)-dimensional random vector, with \( i \)-th component

\[
[P \circ N]_i = \sum_{j=1}^{d} p_{i,j} \circ N_j,
\]

for all \( i = 1, \ldots, d \), where all counting variates \( Y \) in \( p_{i,j} \circ N_j \)'s are assumed to be independent.

Note that \( P \circ (Q \circ N) \overset{\mathcal{L}}{=} [PQ] \circ N \).

Further, \( \mathbb{E}(P \circ N) = P \mathbb{E}(N) \), and

\[
\mathbb{E}((P \circ N)(P \circ N)') = P \mathbb{E}(NN')P' + \Delta,
\]

with \( \Delta := \text{diag}(V \mathbb{E}(N)) \) where \( V \) is the \( d \times d \) matrix with entries \( p_{i,j}(1 - p_{i,j}) \).
Definition A time series \((X_t)\) with values in \(\mathbb{N}^d\) is called a \(d\)-variate MINAR(1) process if

\[
X_t = P \circ X_{t-1} + \varepsilon_t
\]  

for all \(t\), for some \(d \times d\) matrix \(P\) with entries in \([0, 1]\), and some i.i.d. random vectors \(\varepsilon_t\) with values in \(\mathbb{N}^d\).

\((X_t)\) is a Markov chain with states in \(\mathbb{N}^d\) with transition probabilities

\[
\pi(x_t, x_{t-1}) = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1})
\]

satisfying

\[
\pi(x_t, x_{t-1}) = \sum_{k=0}^{x_t} \mathbb{P}(P \circ x_{t-1} = x_t - k) \cdot \mathbb{P}(\varepsilon = k).
\]
Parameter inference for MINAR(1)

**Proposition** Let \((X_t)\) be a \(d\)-variate MINAR(1) process satisfying stationary conditions, as well as technical assumptions (called C1-C6 in FRANKE & SUBBA RAO (1993)), then the conditional maximum likelihood estimate \(\hat{\theta}\) of \(\theta = (P, \Lambda)\) is asymptotically normal,

\[
\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma^{-1}(\theta)), \text{ as } n \to \infty.
\]

Further,

\[
2[\log \mathcal{L}(\overline{N}, \hat{\theta}|N_0) - \log \mathcal{L}(\overline{N}, \theta|N_0)] \xrightarrow{d} \chi^2(d^2 + \text{dim}(\lambda)), \text{ as } n \to \infty.
\]
Granger causality with \textit{BINAR}(1)

\[
\begin{pmatrix}
X_{1,t} \\
X_{2,t}
\end{pmatrix}
= \begin{pmatrix}
p_{1,1} & p_{1,2} \\
p_{2,1} & p_{2,2}
\end{pmatrix}
\circ
\begin{pmatrix}
X_{1,t-1} \\
X_{2,t-1}
\end{pmatrix}
+ \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix}, \text{ with } \text{var } \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix} = \begin{pmatrix}
\lambda_1 & \phi \\
\phi & \lambda_2
\end{pmatrix}
\]
Granger causality with $BINA R(1)$

1. $(X_1)$ and $(X_2)$ are instantaneously related if $\epsilon$ is a noncorrelated noise,

$$
\begin{pmatrix}
X_{1,t} \\
X_{2,t}
\end{pmatrix}
= 
\begin{pmatrix}
p_{1,1} & p_{1,2} \\
p_{2,1} & p_{2,2}
\end{pmatrix}
\circ
\begin{pmatrix}
X_{1,t-1} \\
X_{2,t-1}
\end{pmatrix}
+ 
\begin{pmatrix}
\epsilon_{1,t} \\
\epsilon_{2,t}
\end{pmatrix}, \text{ with var } 
\begin{pmatrix}
\epsilon_{1,t} \\
\epsilon_{2,t}
\end{pmatrix} = 
\begin{pmatrix}
\lambda_1 & * \\
* & \lambda_2
\end{pmatrix}
$$
Granger causality with $BINAR(1)$

2. $(X_1)$ and $(X_2)$ are independent, $(X_1) \perp (X_2)$ if $P$ is diagonal, i.e. $p_{1,2} = p_{2,1} = 0$, and $\varepsilon_1$ and $\varepsilon_2$ are independent,

\[
\begin{pmatrix}
X_{1,t} \\
X_{2,t}
\end{pmatrix}_t =
\begin{pmatrix}
p_{1,1} & 0 \\
0 & p_{2,2}
\end{pmatrix}
\begin{pmatrix}
X_{1,t-1} \\
X_{2,t-1}
\end{pmatrix}_{t-1} +
\begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix}_t,
\text{ with var } \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix} =
\begin{pmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{pmatrix}
\]
Granger causality with $Binar(1)$

3. $(N_1)$ causes $(N_2)$ but $(N_2)$ does not cause $(X_1)$, $(X_1) \rightarrow (X_2)$, if $P$ is a lower triangle matrix, i.e. $p_{2,1} \neq 0$ while $p_{1,2} = 0$,

$$
\begin{pmatrix}
X_{1,t} \\
X_{2,t}
\end{pmatrix}
= 
\begin{pmatrix}
p_{1,1} & 0 \\
p_2 & p_{2,2}
\end{pmatrix}
\circ
\begin{pmatrix}
X_{1,t-1} \\
X_{2,t-1}
\end{pmatrix}
+ 
\begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix},
$$

with

$$
\text{var}
\begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix}
= 
\begin{pmatrix}
\lambda_1 & \varphi \\
\varphi & \lambda_2
\end{pmatrix}
$$
4. \((N_2)\) causes \((N_1)\) but \((N_{1,t})\) does not cause \((N_2)\), \((N_1) \leftarrow (N_{2,t})\), if \(P\) is a upper triangle matrix, i.e. \(p_{1,2} \neq 0\) while \(p_{2,1} = 0\),

\[
\begin{pmatrix}
X_{1,t} \\
X_{2,t}
\end{pmatrix}_{X_t} =
\begin{pmatrix}
p_{1,1} & * \\
0 & p_{2,2}
\end{pmatrix}_{P}
\begin{pmatrix}
X_{1,t-1} \\
X_{2,t-1}
\end{pmatrix}_{X_{t-1}} +
\begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix}_{\varepsilon_t}, \text{ with var } \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix} = \begin{pmatrix}
\lambda_1 & \varphi \\
\varphi & \lambda_2
\end{pmatrix}
\]
Granger causality with $BINA R(1)$

5. $(N_1)$ causes $(N_2)$ and conversely, i.e. a feedback effect $(N_1)\leftrightarrow(N_2)$, if $P$ is a full matrix, i.e. $p_{1,2}, p_{2,1} \neq 0$

\[
\begin{pmatrix}
X_{1,t} \\
X_{2,t}
\end{pmatrix}
= \begin{pmatrix}
p_{1,1} & * \\
* & p_{2,2}
\end{pmatrix}
\circ
\begin{pmatrix}
X_{1,t-1} \\
X_{2,t-1}
\end{pmatrix}
+ \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix}, \text{ with } \text{var} \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix} = \begin{pmatrix}
\lambda_1 & \varphi \\
\varphi & \lambda_2
\end{pmatrix}
\]
Bivariate Poisson $BINAR(1)$

A classical distribution for $\varepsilon_t$ is the bivariate Poisson distribution, with one common shock, i.e.

$$\begin{cases} 
    \varepsilon_{1,t} = M_{1,t} + M_{0,t} \\
    \varepsilon_{2,t} = M_{2,t} + M_{0,t}
\end{cases}$$

where $M_{1,t}$, $M_{2,t}$ and $M_{0,t}$ are independent Poisson variates, with parameters $\lambda_1 - \varphi$, $\lambda_2 - \varphi$ and $\varphi$, respectively. In that case, $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t})$ has joint probability function

$$e^{-[\lambda_1 + \lambda_2 - \varphi]} \frac{(\lambda_1 - \varphi)^{k_1} (\lambda_2 - \varphi)^{k_2}}{k_1! k_2!} \min\{k_1, k_2\} \sum_{i=0} \binom{k_1}{i} \binom{k_2}{i} i! \left( \frac{\varphi}{[\lambda_1 - \varphi][\lambda_2 - \varphi]} \right)$$

with $\lambda_1, \lambda_2 > 0$, $\varphi \in [0, \min\{\lambda_1, \lambda_2\}]$.

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$
Bivariate Poisson $BINARY(1)$ and Granger causality

For instantaneous causality, we test

$$H_0 : \varphi = 0 \text{ against } H_1 : \varphi \neq 0$$

**Proposition** Let $\hat{\lambda}$ denote the conditional maximum likelihood estimate of $\lambda = (\lambda_1, \lambda_2, \varphi)$ in the non-constrained MINAR(1) model, and $\lambda \perp$ denote the conditional maximum likelihood estimate of $\lambda \perp = (\lambda_1, \lambda_2, 0)$ in the constrained model (when innovation has independent margins), then under suitable conditions,

$$2[\log \mathcal{L}(N, \hat{\lambda} | N_0) - \log \mathcal{L}(N, \hat{\lambda} \perp | N_0)] \xrightarrow{\mathcal{L}} \chi^2(1), \text{ as } n \to \infty, \text{ under } H_0.$$
**Bivariate Poisson BINAR(1) and Granger causality**

For lagged causality, we test

$$H_0 : P \in \mathcal{P} \text{ against } H_1 : P \notin \mathcal{P},$$

where $\mathcal{P}$ is a set of constrained shaped matrix, e.g. $\mathcal{P}$ is the set of $d \times d$ diagonal matrices for lagged independence, or a set of block triangular matrices for lagged causality.

**Proposition** Let $\hat{P}$ denote the conditional maximum likelihood estimate of $P$ in the non-constrained MINAR(1) model, and $\hat{P}^c$ denote the conditional maximum likelihood estimate of $P$ in the constrained model, then under suitable conditions,

$$2 \left[ \log \mathcal{L}(\mathbf{N}, \hat{P}|\mathbf{N}_0) - \log \mathcal{L}(\mathbf{N}, \hat{P}^c|\mathbf{N}_0) \right] \overset{L}{\to} \chi^2(d^2 - \text{dim}(\mathcal{P})), \text{ as } n \to \infty, \text{ under } H_0.$$

**Example** Testing $(N_{1,t}) \rightarrow (N_{2,t})$ is testing whether $p_{1,2} = 0$, or not.
Proposition Consider a MINAR(1) process with representation
\[ X_t = P \circ X_{t-1} + \epsilon_t, \]
where \((\epsilon_t)\) is the innovation process, with \( \lambda := \mathbb{E}(\epsilon_t) \) and \( \Lambda := \text{var}(\epsilon_t) \). Let \( \mu := \mathbb{E}(X_t) \) and \( \gamma(h) := \text{cov}(X_t, X_{t-h}) \). Then \( \mu = [I - P]^{-1} \lambda \) and for all \( h \in \mathbb{Z} \), \( \gamma(h) = P^h \gamma(0) \) with \( \gamma(0) \) solution of
\[ \gamma(0) = P \gamma(0) P' + (\Delta + \Lambda). \]
Part 2

Application to earthquakes
Multivariate models?

Shapefiles from
http://www.colorado.edu/geography/foote/maps/assign/hotspots/hotspots.html
The dataset, and stationarity issues

We work with 16 (17) tectonic plates,
- Japan is at the limit of 4 tectonic plates (Pacific, Okhotsk, Philippine and Amur),
- California is at the limit of the Pacific, North American and Juan de Fuca plates.

Data were extracted from the Advanced National Seismic System database (ANSS) http://www.ncedc.org/cnss/catalog-search.html
- 1965-2011 for magnitude $M > 5$ earthquakes (70,000 events);
- 1992-2011 for $M > 6$ earthquakes (3,000 events);
- To count the number of earthquakes, used time ranges of 3, 6, 12, 24, 36 and 48 hours;
- Approximately 8,500 to 135,000 periods of observation;
Multivariate models: comparing dynamics

\[
\begin{pmatrix}
X_{1,t} \\
X_{2,t}
\end{pmatrix} = \begin{pmatrix}
p_{1,1} & p_{1,2} \\
p_{2,1} & p_{2,2}
\end{pmatrix} \circ \begin{pmatrix}
X_{1,t-1} \\
X_{2,t-1}
\end{pmatrix} + \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix}
\]
with var \[
\begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix} = \begin{pmatrix}
\lambda_1 & \varphi \\
\varphi & \lambda_2
\end{pmatrix}
\]

Complete model, with full dependence
Multivariate models: comparing dynamics

\[
\begin{pmatrix}
X_{1,t} \\
X_{2,t}
\end{pmatrix} = \begin{pmatrix}
p_{1,1} & 0 \\
0 & p_{2,2}
\end{pmatrix} \circ \begin{pmatrix}
X_{1,t-1} \\
X_{2,t-1}
\end{pmatrix} + \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix}
\] with \[\text{var} \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix} = \begin{pmatrix}
\lambda_1 & \varphi \\
\varphi & \lambda_2
\end{pmatrix}\]

Partial model, with diagonal thinning matrix, no-crossed lag correlation
Multivariate models: comparing dynamics

\[
\begin{pmatrix}
  X_{1,t} \\
  X_{2,t}
\end{pmatrix}
= \begin{pmatrix}
p_{1,1} & 0 \\
0 & p_{2,2}
\end{pmatrix}
\odot \begin{pmatrix}
  X_{1,t-1} \\
  X_{2,t-1}
\end{pmatrix}
+ \begin{pmatrix}
  \varepsilon_{1,t} \\
  \varepsilon_{2,t}
\end{pmatrix}
\quad \text{with var} \begin{pmatrix}
  \varepsilon_{1,t} \\
  \varepsilon_{2,t}
\end{pmatrix}
= \begin{pmatrix}
  \lambda_1 & 0 \\
0 & \lambda_2
\end{pmatrix}
\]

Two independent INAR processes
Multivariate models: comparing dynamics

\[
\begin{pmatrix}
X_{1,t} \\
X_{2,t}
\end{pmatrix} = \begin{pmatrix}
p_{1,1} & 0 \\
0 & p_{2,2}
\end{pmatrix} \circ \begin{pmatrix}
X_{1,t-1} \\
X_{2,t-1}
\end{pmatrix} + \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix}
\text{ with var } \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix} = \begin{pmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{pmatrix}
\]

Two independent INAR processes
Multivariate models: comparing dynamics

\[
\begin{pmatrix}
X_{1,t} \\
X_{2,t}
\end{pmatrix} = \begin{pmatrix}
0 & 0 \\
0 & 0
\end{pmatrix} \circ \begin{pmatrix}
X_{1,t-1} \\
X_{2,t-1}
\end{pmatrix} + \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix}
\]

with \( \text{var} \begin{pmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{pmatrix} = \begin{pmatrix}
\lambda_1 & \varphi \\
\varphi & \lambda_2
\end{pmatrix} \)

Two (possibly dependent) Poisson processes
Multivariate models: tectonic plates interactions

- For all pairs of tectonic plates, at all frequencies, autoregression in time is important (very high statistical significance);
- Long sequence of zeros, then mainshocks and aftershocks;
- Rate of aftershocks decreases exponentially over time (Omori’s law);
- For 7-13% of pairs of tectonic plates, diagonal BINAR has significant better fit than independent INARs;
- Contribution of dependence in noise;
- Spatial contagion of order 0 (within $h$ hours);
- Contiguous tectonic plates;
- For 7-9% of pairs of tectonic plates, proposed BINAR has significant better fit than diagonal BINAR;
- Contribution of spatial contagion of order 1 (in time interval $[h, 2h]$);
- Contiguous tectonic plates;
- for approximately 90%, there is no significant spatial contagion for $M > 5$ earthquakes
Granger causality $N_1 \rightarrow N_2$ or $N_1 \leftarrow N_2$


Granger Causality test, 3 hours

Granger Causality test, 6 hours
Granger causality $N_1 \rightarrow N_2$ or $N_1 \leftarrow N_2$

Granger causality $N_1 \rightarrow N_2$ or $N_1 \leftarrow N_2$

Multivariate models: frequency versus magnitude

\[ X_{1,t} = \sum_{i=1}^{\infty} 1(T_i \in [t, t+1), M_i \leq s) \text{ and } X_{2,t} = \sum_{i=1}^{\infty} 1(T_i \in [t, t+1), M_i > s) \]

Here we work on two sets of data: medium-size earthquakes \((M \in (5, 6))\) and large-size earthquakes \((M > 6)\).

- Investigate direction of relationship (which one causes the other, or both);
- Pairs of tectonic plates:
  - Uni-directional causality: most common for contiguous plates (North American causes West Pacific, Okhotsk causes Amur);
  - Bi-directional causality: Okhotsk and West Pacific, South American and Nasca for example;
- Foreshocks and aftershocks:
  - Aftershocks much more significant than foreshocks (as expected);
  - Foreshocks announce arrival of larger-size earthquakes;
  - Foreshocks significant for Okhotsk, West Pacific, Indo-Australian, Indo-Chinese, Philippine, South American;
Risk management issues

- Interested in computing $\mathbb{P}\left(\sum_{t=1}^{T} (N_{1,t} + N_{2,t}) \geq n \mid \mathcal{F}_0\right)$ for various values of $T$ (time horizons) and $n$ (tail risk measure);
- Total number of earthquakes on a set of two tectonic plates;
- 100,000 simulated paths of diagonal and proposed BINAR models;
- Use estimated parameters of both models;
- Pair: Okhotsk and West Pacific;
- Scenario: on a 12-hour period, 23 earthquakes on Okhotsk and 46 earthquakes on West Pacific (second half of March 10th, 2011);
### Diagonal model

<table>
<thead>
<tr>
<th>n / days</th>
<th>1 day</th>
<th>3 days</th>
<th>7 days</th>
<th>14 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.9680</td>
<td>0.9869</td>
<td>0.9978</td>
<td>0.9999</td>
</tr>
<tr>
<td>10</td>
<td>0.5650</td>
<td>0.7207</td>
<td>0.8972</td>
<td>0.9884</td>
</tr>
<tr>
<td>15</td>
<td>0.1027</td>
<td>0.2270</td>
<td>0.4978</td>
<td>0.8548</td>
</tr>
<tr>
<td>20</td>
<td>0.0067</td>
<td>0.0277</td>
<td>0.1308</td>
<td>0.4997</td>
</tr>
</tbody>
</table>

### Proposed model

<table>
<thead>
<tr>
<th>n / days</th>
<th>1 day</th>
<th>3 days</th>
<th>7 days</th>
<th>14 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.9946</td>
<td>0.9977</td>
<td>0.9977</td>
<td>1.0000</td>
</tr>
<tr>
<td>10</td>
<td>0.8344</td>
<td>0.9064</td>
<td>0.9712</td>
<td>0.9970</td>
</tr>
<tr>
<td>15</td>
<td>0.3638</td>
<td>0.5288</td>
<td>0.7548</td>
<td>0.9479</td>
</tr>
<tr>
<td>20</td>
<td>0.0671</td>
<td>0.1573</td>
<td>0.3616</td>
<td>0.7256</td>
</tr>
</tbody>
</table>
Some references


Utsu, T. (1969) "Aftershocks and earthquake statistics (I) - Some parameters which characterize an aftershock sequence and their interrelations", Journal of the Faculty of Science of Hokkaido University, 121-195.


The article can be downloaded from [http://arxiv.org/abs/1112.0929](http://arxiv.org/abs/1112.0929)