Department of Statistics
London School of Economics and Political Science

# Stochastic Models and Methods for the Assessment of Earthquake Risk in Insurance

**A thesis submitted by**

Diego Jiménez-Huerta

**in fulfillment of the requirements for the degree of**

Doctor of Philosophy

London, United Kingdom

May 2009

# Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

# Abstract

The problem of earthquake risk assessment and management in insurance is a challenging one at the interface of geophysics, engineering seismology, stochastics, insurance mathematics and economics. In this work, I propose stochastic models and methods for the assessment of earthquake risk from an insurer's point of view, where the aim is not to address problems in the financial mathematics and economics of risk selection, pricing, portfolio management, and risk transfer strategies such as reinsurance and securitisation, but to enable the latter through the characterisation of the foundation of any risk management consideration in insurance: the distribution of losses over a period of time for a portfolio of risks.

Insurance losses are assumed to be generated by a *loss process* that is in turn governed by an *earthquake process*, a point process marked with the earthquake's hypocentre and magnitude, and a *conditional loss distribution* for an insurance portfolio, governing the loss size given the hypocentre and magnitude of the earthquake, and the physical characteristics of the portfolio as described in the individual policy records.

From the modeling perspective, I examine the (non-trivial) minutiae around the infrastructure underpinning the loss process. A novel model of the earthquake process, a Poisson marked point process with spatial gamma intensity measure on the hypocentral space, and extensions of the Poisson and stress release models through the inclusion of hypocentral location in the mark, are proposed. I discuss the general architectural considerations for constructing the conditional loss distribution, and propose a new model as an alternative to the traditional ground motion attenuation and seismic vulnerability approach in engineering risk assessment. On the actuarial mathematics front, given a fully specified loss process, I address the problem of constructing simulation based and, where possible, analytical approximations to the distribution of portfolio losses over a period of time.

I illustrate the applicability of the stochastic models and methods proposed in this work through the analysis of a residential homeowners property catastrophe portfolio exposed to earthquake risk in California. I construct approximations to the distribution of portfolio losses over a period of time under each of the three models of the earthquake process that I propose, and discuss their relative merits.

# Contents

**Declaration** ii

**Abstract** iii

**Acknowledgments** vi

**1 Introduction** 2
- 1.1 Background . . . 2
- 1.2 The mathematical problem: an insurer's perspective . . . 4
- 1.3 Scope of the thesis . . . 6

**2 The conditional loss distribution** 8
- 2.1 General architectural considerations . . . 8
- 2.2 Construction of a reference hazard measure . . . 10
- 2.3 Modeling of singularities . . . 13
- 2.4 Modeling of the continuous component . . . 15
- 2.5 Derivation of moments . . . 20

**3 Poisson type models of the earthquake process** 22
- 3.1 Introduction . . . 22
- 3.2 The spatial Poisson process . . . 24
- 3.3 The spatial Poisson process with gamma intensity measure . . . 28

**4 A stress release model of the earthquake process** 36
- 4.1 The earthquake process . . . 37
- 4.2 The stress process . . . 37
- 4.3 The earthquake process intensity . . . 38
- 4.4 Stationarity of the stress process . . . 41
- 4.5 Model estimation . . . 44
- 4.6 Simulation, prediction, and diagnostic procedures . . . 48

**5 Earthquake risk assessment in insurance** 51
- 5.1 The aggregate loss distribution . . . 51
- 5.2 Formulæ for selected moments of the aggregate loss under the Poisson and Poisson-gamma earthquake processes . . . 52
- 5.3 Analytical approximations to the aggregate loss distribution . . . 59
- 5.4 Simulation of the loss process . . . 61
- 5.5 Case study . . . 65

**A Notable models in statistical seismology** **86**
A.1 The Hawkes process and the ETAS model . . . . . 86
A.2 Stress release models . . . . . 89

**B The Poisson-gamma and Dirichlet processes** **92**
B.1 The univariate Poisson distribution . . . . . 92
B.2 The multinomial distribution . . . . . 93
B.3 The distribution of multiple independent Poisson variates . . . . 93
B.4 The Beta and Dirichlet distributions . . . . . 95
B.5 The univariate gamma distribution . . . . . 96
B.6 The distribution of multiple independent gamma variates . . . . 97
B.7 Conjugate priors to the Poisson and multinomial distributions . . 100

**C Simulation** **104**
C.1 Simulation of earthquake location in space . . . . . 104
C.2 Simulation of the Dirichlet process . . . . . 105
C.3 Simulation of the stress release model . . . . . 106

**D Sample loss and policy records** **108**

**Bibliography** **114**

# Acknowledgments

I would like to express my deepest gratitude to Prof. Ragnar Norberg, an indefatigable creator and challenger of the questions that underpin science. Without his support, and above all, without his patience, this journey would never have reached a conclusion.

My thanks also go to Dr. Pauline Barrieu, for her support and advice during the latter stages of my doctoral studies, and to my colleagues and friends at Risk Management Solutions, for allowing me to take time off to complete the drafting of this thesis.

This work is dedicated to my parents, my source of intellectual drive, and the giants on whose shoulders I have stood. And to Mariana, my best friend.

We are so constituted that nothing takes us further or leads us higher than the leaps made by our errors. In point of fact we owe the little we have learned to hypotheses that were always hazardous and often absurd, and, as a general rule, less discreet than they are today. They were unwise, perhaps, but they kept alive the ardour for research.

M. Mæterlinck

# Chapter 1

# Introduction

## 1.1 Background

The complex socio-economic development of the world has led to a dramatic increase of losses due to natural and anthropogenic catastrophes: earthquakes, hurricanes, floods, volcanic eruptions, and nuclear accidents to name only a few. It is believed that within the next 50 years, more than a third of the world's population will live in seismically and volcanically active zones. Studies on possible scenarios of earthquakes and losses are a critical issue for decision making in insurance, and are fundamental to the process of designing risk mitigation mechanisms.

Traditional insurance operates on the assumption of independent, frequent, low-consequence risks for which decisions on premiums, estimates of claims, and likelihood of insolvency, can be calculated by using rich historical data. The law of large numbers provides in this case a simple portfolio selection strategy: if the number of independent risks in the portfolio is larger, then, in general, the variance of the average claim is lower and lower premiums can be offered. This increases the demand for insurance, the profits of insurers, the coverage of

potential losses and, ultimately, the stability of the insurance industry.

Traditionally actuaries have had difficulty pricing risks that have low frequency but potential for large severity. Often the prices charged for these risks are determined by underwriting judgment and market forces in the insurance cycle, a practice that may result in inefficiencies potentially leading to market saturation or collapse. A case in point is earthquake risk, the actuarial mathematics of which do not fall within the framework of conventional procedures.

Although large damaging earthquakes occur rarely, they are associated with an extremely high loss potential where all forms of insurance can be affected simultaneously. Earthquakes produce claims that are highly correlated in time and space, depending on the density of buildings and their locations and individual characteristics, and on the mechanics of earthquake occurrence in time and space. The law of large numbers does not operate (in general) and a "more-risk-is-better" strategy could increase the probability of ruin for an insurer unless access to a colossal contingency fund was secured. Then there is the problem of lack of historical data on the occurrence of earthquakes and losses at a particular location. Although rich data may exist on earthquake activity on a regional level, claims data are not so abundant, with many insurance companies lacking the experience of handling claims for a major damaging earthquake. The risk assessment and management problem in the case of earthquake risk is then transformed from a purely statistical one into a challenging one at the interface of geophysics, engineering seismology, stochastics, insurance mathematics and economics, whose important considerations regarding risk selection, premium pricing, portfolio management, and risk transfer strategies such as reinsurance and securitisation, must be well grounded in a solid assessment of earthquake risk.

In this work we propose a unified methodology comprising models and methods for the assessment of earthquake risk from an insurer's point of view. It

is basic to the extent that the aim is not to address the financial mathematics and economics of pricing, portfolio management and risk transfer strategies, but to enable the latter through the characterisation of the foundation of any risk management consideration in insurance: the distribution of losses over a period of time for a portfolio of risks.

## 1.2 The mathematical problem: an insurer's perspective

Assume an insurance company writes earthquake coverage in a given geographic region and has maintained data files for this line of business over a certain period of time. *Policy records* describe the individual objects (buildings) covered under the scheme by ascribing to each object a vector of observed characteristics $c$, including its geographical location, its construction type (wood,unreinforced or reinforced masonry, steel frame, etc.), its occupancy type (residential, commercial, industrial, etc.), and the financial and coverage terms such as estimated value, deductible, coverage limit, and number of reinstatements. *Claims records* specify the policy number which (together with date) identifies uniquely the damaged object and its characteristics at the time of the incident.

Assume there is an earthquake catalog available for a geographic region of interest $\mathbb{L} \subset \mathbb{R}^3$. The catalog describes the generic earthquake by a triple $(T, L, Z)$, where $T \in \mathbb{R}_+$ is its time of occurrence, time being reckoned from the inception of the catalog, $L \in \mathbb{L}$ is its location in space, and $Z \in \mathbb{Z} \subset \mathbb{R}_+$ is its magnitude. We will henceforth refer to $T$ as the *time* and to the pair $(L, Z)$ as the *mark* of the earthquake. We assume $T$, $L$, and $Z$, reside in $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$, $(\mathbb{L}, \mathcal{B}_{\mathbb{L}})$, and $(\mathbb{Z}, \mathcal{B}_{\mathbb{Z}})$, the time, location, and magnitude spaces, respectively, with $\mathcal{B}_{\mathbb{R}_+}$, $\mathcal{B}_{\mathbb{L}}$, and $\mathcal{B}_{\mathbb{Z}}$, the Borel sigma-algebras. The earthquakes $(T_i, L_i, Z_i)$ are assumed to occur at isolated points of time (almost surely) so that they can

be enumerated in chronological order, forming the *earthquake process*, a marked point process $\{(T_i, L_i, Z_i)\}_{i\in\mathbb{N}}$ with associated counting process

$$N(\mathcal{T}, \mathcal{L}, \mathcal{Z}) := \sum_{i=1}^{\infty} 1_{\mathcal{T}\times\mathcal{L}\times\mathcal{Z}}(T_i, L_i, Z_i) = \int_{\mathcal{T}}\int_{\mathcal{L}}\int_{\mathcal{Z}} N(dt, dl, dz),$$

the number of earthquakes occurring in the time interval $\mathcal{T} \in \mathcal{B}_{\mathbb{R}_+}$, where $\mathcal{T} := (r, s]$, $0 \le r < s$, with location in $\mathcal{L} \in \mathcal{B}_{\mathbb{L}}$, and magnitude in $\mathcal{Z} \in \mathcal{B}_{\mathbb{Z}}$.

Let the pair $(Y, c)$ denote the generic insured object with characteristics $c$ and loss $Y$, given the occurrence of an earthquake with location $L$ and magnitude $Z$, and

$$G(y|l, z, c) := \mathbb{P}_c[Y \le y|L = l, Z = z]$$

be its distribution function. Consider the portfolio in force at time $t$, henceforth referred to as the *t-portfolio*, and its $K_t$ constituent risks. The $t$-portfolio loss given $L$ and $Z$ is $R := \sum_{k=1}^{K_t} Y_k$, with distribution

$$\begin{aligned} G_t(r|l, z) &:= \mathbb{P}[R \le r|L = l, Z = z] \\ &= [G(\cdot\,|l, z, c_1) * \cdots * G(\cdot\,|l, z, c_{K_t})]\,(r) \end{aligned}$$

given by the convolution of $G$ over the $t$-portfolio. We shall henceforth refer to $G_t$ as the *conditional loss distribution*.

If the mark $(L, Z)$ for an earthquake occurring at time $T$ is augmented to include $R$, the earthquake process becomes the *loss process*, with associated counting process

$$
\begin{aligned}
N(t, \mathcal{L}, \mathcal{Z}, \mathcal{R}) &= \sum_{i=1}^{\infty} 1_{(0,t]\times\mathcal{L}\times\mathcal{Z}\times\mathcal{R}}(T_i, L_i, Z_i, R_i) \\
&= \int_{(0,t]} \int_{\mathcal{L}} \int_{\mathcal{Z}} \int_{\mathcal{R}} N(ds, dl, dz, dr).
\end{aligned}
$$

Assuming the $t$-portfolio remains fixed during the time interval $(t, u]$, the aggregate loss over $(t, u]$ is given by

$$
X_{t,u} := \int_{(t,u]} \int_{\mathbb{L}} \int_{\mathbb{Z}} \int_{\mathbb{R}_+} rN(d\tau, dl, dz, dr).
$$

Let $\mathcal{F} = \{\mathcal{F}_t\}_{t\geq 0}$ be the filtration to which the loss process is adapted, including information about the marks and times of past events, as well as any information about external variables or processes evolving in time in parallel with the losses. Using the usual suggestive notation, the $\mathcal{F}-$intensity of the loss process is

$$
\Lambda(dt, dl, dz, dr) := \mathbb{E}[N(dt, dl, dz, dr)|\mathcal{F}_{t^-}],
$$

where $\Lambda$ is a measure that admits a representation of the form

$$
\Lambda(dt, dl, dz, dr) = \lambda dt \Delta(dl)\Phi(dz)G_t(dr|l, z),
$$

with $\lambda > 0$, and $\Delta$ and $\Phi$ the location and magnitude distributions, respectively. Thus, specifying a loss process is tantamount to specifying its underlying earthquake process and conditional loss distribution.

## 1.3 Scope of the thesis

In this thesis we examine the (non-trivial) minutiae around the infrastructure underpinning the loss process, and address the problem of constructing approx-

imations to the distribution $P^{t,u}(x) := \mathbb{P}[X_{t,u} \leq x]$.

In Chapter 2, we discuss the general architectural considerations for constructing the conditional loss distribution, and propose a new model as an alternative to the traditional approaches available in the engineering seismology literature, which rely on the availability of ground motion attenuation relations.

In Chapter 3, we propose a novel model of the earthquake process, an application of a spatial Poisson model with gamma intensity measure, along with an extension of the spatial Poisson model that incorporates epicentral location and depth as mark components. We also introduce this feature in a specification of the stress release model of the earthquake process, considered in Chapter 4.

On the actuarial mathematics front, in Chapter 5 we address the problem of constructing simulation based and, where possible, analytical approximations to the loss distribution $P^{t,u}$, and illustrate the applicability of the stochastic models and methods we propose through the analysis of a residential homeowners property catastrophe portfolio exposed to earthquake risk in Southern California. We construct approximations to $P^{t,u}$ under each of the three models of the earthquake process that we propose, and discuss their relative merits.

# Chapter 2

# The conditional loss distribution

## 2.1 General architectural considerations

In this chapter we address the problem of specifying $G_t(r|l,z)$, the distribution of $R$, the loss to the $t$-portfolio given the occurrence of an earthquake with magnitude $Z$ and location $L$. To the extent that $G_t$ is given by the convolution

$$G_t(r|l,z) = \left[G(\,\cdot\,|l,z,c_1) * \cdots * G(\,\cdot\,|l,z,c_{K_t})\right](r),$$

we first focus on the construction of $G(y|l,z,c)$.

**A. The need for a mixed distribution.** Consider a policy/insured object and its record $c$, listing, among other data, its value $v$. Given the occurrence of an earthquake, let $p \geq 0$ denote the probability of no loss occurring. The magnitude of the earthquake, its location relative to the risk, the local site conditions such as soil type, potential for liquefaction and/or landslide, and the

individual characteristics of the object, all stored in $c$, are factors influencing the probability of this event and consequently $p = p(l, z, c)$. This same logic applies to the event of total loss, occurring with probability $q = q(l, z, c) \geq 0$. The magnitude of the earthquake could be so large, and its location relative to the risk in question be so close that, compounded with e.g. soft porous soils prone to liquefaction and a weak construction profile in engineering terms, it would lead to structural damage resulting in a total loss. Finally, between these two ends of the loss spectrum, we have an infinite and uncountable number of possibilities for the size of loss, each occurring with probability zero and governed by a continuous distribution $\overline{G}(y|l, z, c)$, say, with support on $(0, v)$. In mathematical terms, $Y$ has a mixed distribution $G(y|l, z, c)$ with probability density $p$ and $q$, with respect to Dirac measure on 0 and $v$, respectively, and probability density $\overline{g}$ with respect to Lebesgue measure on $(0, v)$, i.e.,

$$\begin{aligned} p(l, z, c) &= G\left(0|l, z, c\right) \\ q(l, z, c) &= G\left(v|l, z, c\right) - G\left(v - |l, z, c\right), \text{ and} \\ \overline{g}\left(y|l, z, c\right) &= \frac{G'\left(y|l, z, c\right)}{1 - p(l, z, c) - q(l, z, c)} \end{aligned}$$

**B. Establishing a reference hazard measure.** The distribution $G$ is a function of earthquake magnitude and location, and of the covariates summarised in $c$. Without reference to a hazard measure that translates the earthquake characteristics and the risk specifics into some form of damage potential, conditioning the size of loss on the occurrence of an earthquake is meaningless.

The traditional approach in engineering seismology is to use the amplitude of ground motion as a reference hazard measure. Given the location and magnitude of an earthquake, the distribution of peak ground motion is estimated through an *attenuation relation* whereby e.g. the peak ground acceleration $A$

at the site is governed by a lognormal distribution $H(a|l,z,c)$ with, say, mean $\mu = \mu(l,z,c)$ and variance $\sigma^2$. Attenuation relations are developed regionally on the basis of ground motion observations for earthquakes with magnitude above a certain threshold. Given a level of peak ground acceleration, a *seismic vulnerability function*, a distribution $G^V(y|a,c)$, say, relating damage to peak ground acceleration is then used to estimate damage at the site. Under this approach, $G$ would take the form

$$G(y|l,z,c) = \int G^V(y|a,c)dH(a|l,z,c). \tag{2.1}$$

For insurance risk assessment purposes, where the aim is not so much to predict ground motion at a particular site as to obtain the probability distribution of insurance losses over relatively long periods, the formulation in (2.1) is not ideal, for a number of reasons. First and foremost, there are the natural robustness issues arising from missing data and measurement and model specification error for each of two different models, one for $G^V$ and one for $H$, with a compounding effect for the uncertainties in each of the latter likely to be reflected in the variance structure of $G$. Then there is the computational burden of calculating (2.1) for each of the risks in the portfolio, a behemoth of a task if more than a few hundred risks are present therein. We propose an alternative to the engineering seismology two-staged approach through the design of a new hazard measure which enables the modeling of $G$ directly, without the need of the 'auxiliary' role played by the attenuation relation. The construction of this new hazard measure is addressed next.

## 2.2 Construction of a reference hazard measure

**A. Energy released by earthquakes and the inverse-square law.** The total energy $E$ released when an earthquake occurs includes energy required to

create new cracks in rock, energy dissipated as heat through friction, and energy $S$ elastically radiated through the earth. Of these, only the latter is susceptible to measurement by ground motion recording instruments. An approximation to $S$ in terms of earthquake magnitude $Z$ is given by the empirical relation (Vere-Jones *et al.* [48])

$$S = 10^{2.4+\frac{3}{4}Z}, \tag{2.2}$$

which is similar to the classical definition of seismic energy and Benioff strain (Kanamori and Anderson [20]), with a changing coefficient considering new seismological data.

A seismic wave loses energy as it propagates through the earth and, as a result, the energy with which an earthquake affects a location is dependent on the distance between the location and the hypocentre. The nature of this relationship is governed by a physical *inverse-square* type law, whereby the amount of energy radiated at a site is inversely proportional to the the square of the distance from the energy source[1].

**B. Amount of energy radiated at the site of a risk.** Taking $S$ as a measure of the energy released upon the occurrence of an earthquake with magnitude $Z$, a measure of the energy radiated at the site of the object with policy record $c$ is given by

$$\epsilon(l, z, c) := \frac{s(z)}{\|l - l^*\|^2},$$

where $s(z) := 10^{2.4+\frac{3}{4}z}$, $\|\cdot\|$ is the Euclidean norm, and the location $l^* \in \mathbb{R}^3$ of the object is stored in $c$. There are two conspicuous features of this formulation.

[1]In more general terms, inverse-square type laws assert that some physical quantity or strength is inversely proportional to the square of the distance from the source of that physical quantity. They appear in e.g. the theories of gravitation, electromagnetism, and acoustics. Newton's law of universal gravitation and Coulomb's law describing the electrostatic force between electric charges are examples of inverse-square laws.

The first has to do with the energy per unit area when $l = l^*$. The energy released by an earthquake radiates from the hypocentre and, at any time, sits on a sphere centred at $l$. When the sphere attains radius $\|l - l^*\|$ and hits the object, the total energy is still $s(z)$ and the area it is uniformly distributed over is proportional to $\|l - l^*\|^2$. Thus, the energy per unit area must, by definition, be infinite at ground 0 at time 0. The second feature is the implicit assumption of isotropy that some would argue is inappropriate. To the extent that the complexities of energy attenuation through an elastic medium are not being accounted for, our formulation is crude. This crudity, however, is offset by the fact that the ultimate goal is not to predict energy or ground motion at a particular site, but to obtain the probability distribution of insurance losses over relatively large periods of time and over a relatively large geographical region. Any sensitivity of the loss distribution to the functional form of $\epsilon(\cdot)$ must then be addressed *vis-à-vis* the corresponding distribution obtained under a formulation of the form (2.1), and the compounding effect of the uncertainties inherent in the attenuation and vulnerability models.

**C. Calculation of distances between earthquakes and risks.** Earthquake catalogs and policy and claims records contain spatial location data in a form akin to spherical coordinates. The latitude and longitude are angles measured in degrees and, in the case of earthquake catalogs, the depth of an earthquake, typically measured in kilometres, can be readily transformed into a spherical radius by subtracting it from 6367km, the radius of the earth. As for the risks in a portfolio, they can be assumed to be 'at sea level' if altitude readings are not available.

Given a typical earthquake catalog entry $(\theta, \phi, h)$, where $\theta$ and $\phi$ are the longitude and latitude in degrees, respectively, and $h$ is the depth in kilometres, the location (in Cartesian coordinates) $l = (l^{(1)}, l^{(2)}, l^{(3)}) \in \mathbb{R}^3$ is obtained from

$(\theta, \phi, h)$ through the relations

$$\begin{aligned} l^{(1)} &= r \sin(\phi^* \frac{\pi}{180}) \cos(\theta \frac{\pi}{180}) \\ l^{(2)} &= r \sin(\phi^* \frac{\pi}{180}) \sin(\theta \frac{\pi}{180}) \\ l^{(3)} &= r \cos(\phi^* \frac{\pi}{180}), \end{aligned}$$

where $r = 6367 - h$ and $\phi^* = (90 - \phi)$. The same logic applies for obtaining $l^* = (l^{*(1)}, l^{*(2)}, l^{*(3)})$, where the longitude and latitude readings are found in the policy record $c$ and the depth $h$ can be assumed to be zero. The squared distance

$$\|l - l^*\|^2 = (l^{(1)} - l^{*(1)})^2 + (l^{(2)} - l^{*(2)})^2 + (l^{(3)} - l^{*(3)})^2$$

is then readily calculated.

## 2.3 Modeling of singularities

Having established $\epsilon = \epsilon(l, z, c)$ as a reference hazard measure, $G$ can be recast as $G(y|\epsilon, c)$, with continuous component $\overline{G}(y|\epsilon, c)$, and probability atoms $p = p(\epsilon, c)$ and $q = q(\epsilon, c)$. In this section we address the estimation of the latter.

**A. The singularity at zero.** First, we define the binary random variable $Y^0 = 1_{\{Y=0\}}$ with probabilities $\mathbb{P}[Y^0 = 1] = p$ and $\mathbb{P}[Y^0 = 0] = 1 - p$. Under this setup,

$$Y^0 \sim \text{Ber}(p)$$

and the problem is to obtain an estimate for $p$ in terms of $\epsilon$ and $c$. A generalised linear model (GLM) of the form

$$\eta(p) = \alpha^0 + \beta^0 \epsilon + \gamma^0 c \tag{2.3}$$

can be defined, where the scalars $\alpha^0$ and $\beta^0$, and the vector $\gamma^0$, are the parameters of the model, with $\eta$ the link function. Several specifications of $\eta$ are possible (e.g. logit, probit, and complementary log-log), with each resulting in a different formulation of (2.3). For estimation and inference considerations we refer the reader to the monograph by McCullagh and Nelder [26], a standard reference in generalised linear modeling.

Parameter estimation is performed on the basis of information available in the claims records $\{\kappa_i\}$. We assume the policy records $\{c_i\}$ contained in the latter include no-claims data from policies in force at the time of an event for which no losses were reported, enabling the construction of the set of observations $\{y_i^0\}$. The observations $\{\epsilon_i = \epsilon(l_i^\diamond, z_i^\diamond, l_i^*)\}$ are also constructed from information available in $\{\kappa_i\}$, with $l_i^*$ the location of the object in claim $i$, and $l_i^\diamond$ and $z_i^\diamond$ the location and magnitude, respectively, of the earthquake generating the claim. The estimates $\hat{\alpha}^0$, $\hat{\beta}^0$ and $\hat{\gamma}^0$, are obtained upon calibration of (2.3), and the estimate of the probability $p$ of a singularity at zero is given by

$$\hat{p} = \eta^{-1}(\hat{\alpha}^0 + \hat{\beta}^0 \epsilon + \hat{\gamma}^0 c).$$

**B. The singularity at the value of the insured object.** The argument for the estimation of $q$ is identical to that of $p$, with one caveat. If $p$ and $q$ were modelled separately and each using the full set of claims records, there would be a possibility of obtaining estimates $\hat{p}$ and $\hat{q}$ such that $\hat{p}+\hat{q} > 1$. To avoid this problem, we define $q^* = \mathbb{P}[Y = v|Y > 0]$ and fit the model to the observations with $Y > 0$. Thus, the random variable $Y^v = 1_{\{Y=v\}} 1_{\{Y>0\}}$, with probabilities $\mathbb{P}[Y^v = 1] = q^*$ and $\mathbb{P}[Y^v = 0] = 1 - q^*$, whereby

$$Y^v \sim \text{Ber}(q^*),$$

is first defined. A model

$$\eta(q^*) = \alpha^v + \beta^v \epsilon + \gamma^v c \tag{2.4}$$

is then specified and calibrated, from where the estimate

$$\hat{q^*} = \eta^{-1}(\hat{\alpha}^v + \hat{\beta}^v \epsilon + \hat{\gamma}^v c)$$

follows. Since $q = \mathbb{P}[Y = v] = \mathbb{P}[Y = v|Y > 0]\mathbb{P}[Y > 0]$, it follows that

$$\hat{q} = \hat{q^*}(1 - \hat{p}).$$

Alternatively, $p$ and $q$ could be modelled together using, for instance, a multinomial logistic model for the 3-category response

$$(I(Y = 0), I(0 < Y < v), I(Y = v)).$$

For estimation and inference considerations we refer to e.g. Agresti [1].

## 2.4 Modeling of the continuous component

In this section we address the modeling of $\overline{G}(y|\epsilon, c)$, the conditional distribution of $Y$ given $0 < Y < v$, which we assume to be specified up to a vector of unknown parameters $\vartheta = \vartheta(\epsilon, c)$. The $j$-th non-central and central ($j \geq 2$) moments of $Y$, given $0 < Y < v$, are

$$m_j^{(0,v)} = m_j^{(0,v)}(\vartheta(\epsilon, c)) := \int_0^v y^j d\overline{G}(y; \vartheta(\epsilon, c)) \tag{2.5}$$

and

$$\overline{m}_j^{(0,v)} = \overline{m}_j^{(0,v)}(\vartheta(\epsilon, c)) := \int_0^v (y - m_1^{(0,v)})^j d\overline{G}(y; \vartheta(\epsilon, c)),$$

respectively. As before, estimation of $\vartheta$ is performed on the basis of $\{\kappa_i\}$. In principle, this exercise is no different to the estimation of $p$ and $q$, the parameters of the distributions of $Y^0$ and $Y^v$, respectively. Whilst the Bernoulli distribution was the obvious choice in the the case of the latter, the specification of $\overline{G}$ is not so apparent. One architectural restriction is the need for $\overline{G}$ to have support on $(0, v)$, but this can be readily overcome by working with an appropriately truncated distribution and so the candidates remain vast.

**A. General estimation framework.** Once more, we operate in the GLM framework, where the aim is to 'regress' $Y$, given $0 < Y < v$, on $\epsilon$ and $c$ in a way such that

$$\eta(\mu) = \alpha + \beta\epsilon + \gamma c,$$

where $\mu = \mu(\vartheta) = \mathbb{E}[Y|0 < Y < v]$ and, as before, $\alpha$, $\beta$ and $\gamma$ are parameters, with $\eta$ a link function. A standard GLM specification would define the model so as to contain a precision (or dispersion) parameter $\varphi = \varphi(\vartheta)$ allowing a reparameterisation $\overline{G}(y; \mu, \varphi)$ of $\overline{G}$ whereby, given the observations $\{y_i\}$, $\{\epsilon_i\}$, and $\{c_i\}$, the log-likelihood is given by

$$\ln L(\alpha, \beta, \gamma, \varphi) = \sum_i \ln \overline{g}(y_i; \mu_i, \varphi),$$

from where the maximum likelihood estimate (MLE)

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\varphi}) = \underset{(\alpha,\beta,\gamma,\varphi)\in\mathbb{R}^4}{\arg\max} \ln L(\alpha, \beta, \gamma, \varphi)$$

follows, allowing the estimation of $\vartheta$ through $\hat{\vartheta} = \hat{\vartheta}(\hat{\mu} = \hat{\mu}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}), \hat{\varphi})$.

The gamma, Weibull (if the shape parameter is known), log-normal, and Pareto distributions belong to the exponential family of distributions and could be parameterised in the form $\overline{G}(y; \mu, \varphi)$, constituting likely candidates for $\overline{G}$. Another case in point is the beta distribution, which has two further desirable properties: it has a bounded support and a versatile shape accommodating a variety of uncertainties. We present this model next.

**C. The beta model.** Assume $\overline{G}$ is beta distributed, whereby the density of $\overline{G}$ with respect to Lebesgue measure on $(0, v)$ is

$$\overline{g}(y) = \frac{1}{\mathrm{B}(\alpha', \beta')} \left(\frac{y}{v}\right)^{\alpha'-1} \left(\frac{v-y}{v}\right)^{\beta'-1} \left(\frac{1}{v}\right) 1_{(0,v)}(y),$$

where $\alpha' = \alpha'(\epsilon, c) > 0$, $\beta' = \beta'(\epsilon, c) > 0$, and $\mathrm{B}(\cdot, \cdot)$ is the beta function. The first three central moments of $Y$, given $0 < Y < v$, are

$$\begin{aligned} m_1^{(0,v)} &= v \left[\frac{\alpha'}{\alpha' + \beta'}\right], \\ \overline{m}_2^{(0,v)} &= v^2 \left[\frac{\alpha'\beta'}{(\alpha' + \beta')^2(\alpha' + \beta' + 1)}\right], \text{ and} \\ \overline{m}_3^{(0,v)} &= v^3 \left[\frac{2(\beta' - \alpha')\sqrt{1 + \alpha' + \beta'}}{\sqrt{\alpha'\beta'}(2 + \alpha' + \beta')}\right]. \end{aligned} \tag{2.6}$$

Because $v$ varies between different insured objects we model the proportion of loss out of the total value instead. We introduce the change of variable $U := \frac{Y}{v}$ and work with the density

$$\overline{g}(u) = \frac{1}{B(\alpha', \beta')} u^{\alpha'-1}(1-u)^{\beta'-1} 1_{(0,1)}(u), \tag{2.7}$$

a beta distribution on the unit interval indexed by $\alpha'$ and $\beta'$. Ferrari and Cribari-Neto [10] have proposed a regression model where the response is beta

distributed using a parameterisation of (2.7) that is indexed by mean and dispersion parameters. The highlights of this model are presented below.

Let $\mu = \frac{\alpha'}{\alpha'+\beta'}$ and $\varphi = \alpha' + \beta'$. Under this reparameterisation $\alpha' = \mu\varphi$, $\beta' = (1-\mu)\varphi$, and (2.7) can be recast as

$$\overline{g}(u) = \frac{\Gamma(\varphi)}{\Gamma(\mu\varphi)\Gamma((1-\mu)\varphi)} u^{\mu\varphi-1}(1-u)^{(1-\mu)\varphi-1} 1_{(0,1)}(u),$$

where $0 < \mu < 1$, $\varphi > 0$, and $\Gamma(\cdot)$ is the gamma function. The linear model is obtained by assuming

$$\eta(\mu) = \alpha + \beta\epsilon + \gamma c.$$

Note that $\mu$ varies as a function of the explanatory variables, while the dispersion parameter $\varphi = \alpha' + \beta'$ remains constant. The log-likelihood function based on the observations $\{u_i\}$, $\{\epsilon_i\}$, and $\{c_i\}$, with $i \in \{1, \ldots, n_\kappa\}$, is

$$\ln L(\alpha, \beta, \gamma, \varphi) = \sum_{i=1}^{n_\kappa} \ln \overline{g}(u_i; \mu_i, \varphi),$$

where $\mu_i = \eta^{-1}(\alpha + \beta\epsilon_i + \gamma c_i)$. Denoting $u_i^* := \ln[\frac{u_i}{1-u_i}]$ and $\mu_i^* := \psi(\mu_i\varphi) - \psi((1-\mu_i)\varphi)$, with $\psi(\cdot)$ the digamma function, the score function is given by

$$(U_{\alpha,\beta,\gamma}(\alpha, \beta, \gamma, \varphi)^t, U_\varphi(\alpha, \beta, \gamma, \varphi))^t,$$

where

$$U_{\alpha,\beta,\gamma}(\cdot) := \varphi X^t T(u^* - \mu^*),$$

with $X = \{(1, \epsilon_i, c_i)\}$, $T = \text{diag}\{\frac{1}{\frac{d}{d\mu_i}\eta(\mu_i)}\}$, $u^* = \{u_i^*\}$ and $\mu^* = \{\mu_i^*\}$, and

$$U_\varphi(\cdot) := \sum_{i=1}^{n_\kappa} \left[\mu_i(u_i^* - \mu_i^*) + \ln(1-u_i) - \psi((1-\mu_i)\varphi) + \psi(\varphi)\right].$$

The maximum likelihood estimates of $\alpha$, $\beta$, $\gamma$, and $\varphi$, are obtained by finding the root of the equations $U_{\alpha,\beta,\gamma}(\alpha,\beta,\gamma,\varphi)=0$ and $U_{\varphi}(\alpha,\beta,\gamma,\varphi)=0$, respectively, and do not have closed-form. Fisher's information matrix is obtained as follows. Define

$$\begin{aligned} w_i &:= \varphi(\psi'(\mu_i\varphi)+\psi'((1-\mu_i)\varphi))\frac{1}{\eta'(\mu_i)^2}, \\ \varrho_i &:= \varphi(\psi'(\mu_i\varphi)\mu_i-\psi'((1-\mu_i)\varphi)(1-\mu_i)), \text{ and} \\ d_i &:= \psi'(\mu_i\varphi)\mu_i^2+\psi'((1-\mu_i)\varphi)(1-\mu_i)^2-\psi'(\varphi), \end{aligned}$$

where $\psi'(\cdot)$ is the trigamma function. Let $W=\mathrm{diag}\{w_i\}$, $\varrho=\{\varrho_i\}$, and $D=\mathrm{diag}\{d_i\}$. Further, let

$$\begin{aligned} K_{\alpha,\beta,\gamma} &:= \varphi X^tWX, \\ K_{(\alpha,\beta,\gamma),\varphi} &:= K^t_{\varphi,(\alpha,\beta,\gamma)}=X^tT\varrho, \text{ and} \\ K_{\varphi} &:= \mathrm{tr}(D). \end{aligned}$$

Fisher's information matrix is given by

$$\mathcal{I}=\mathcal{I}(\alpha,\beta,\gamma,\varphi)=\begin{pmatrix} K_{\alpha,\beta,\gamma} & K_{(\alpha,\beta,\gamma),\varphi} \\ K_{\varphi,(\alpha,\beta,\gamma)} & K_{\varphi} \end{pmatrix}.$$

Assuming the usual regularity conditions for maximum likelihood estimation are satisfied, the estimator $(\hat\alpha,\hat\beta,\hat\gamma,\hat\varphi)$ is consistent

$$(\hat\alpha,\hat\beta,\hat\gamma,\hat\varphi)\xrightarrow{p}(\alpha,\beta,\gamma,\varphi),$$

and asymptotically normally distributed

$$(\hat\alpha,\hat\beta,\hat\gamma,\hat\varphi)\xrightarrow{d}\mathrm{N}\left((\alpha,\beta,\gamma,\varphi),\mathcal{I}^{-1}\right).$$

The estimates of $\alpha'$ and $\beta'$ are given by

$$\begin{aligned}\hat{\alpha}' &= \eta^{-1}(\hat{\alpha} + \hat{\beta}\epsilon + \hat{\gamma}c)\hat{\varphi} \\ \hat{\beta}' &= \hat{\varphi} - \hat{\alpha}.\end{aligned}$$

## 2.5 Derivation of moments

**A. The moments of $R$.** On the basis of a fully specified model for the conditional loss distribution $G$, the $j$-th non-central and central ($j \geq 2$) moments of $R = \sum_{k=1}^{K_t} Y_k$, the conditional loss for the $t$-portfolio comprising $K_t$ risks, are given by

$$m_j^R = m_j^R(\{\vartheta_k\}) := \int r^j dG_t(r; \{\vartheta_k\}) \tag{2.8}$$

and

$$\overline{m}_j^R = \overline{m}_j^R(\{\vartheta_k\}) := \int (r - m_1^R)^j dG_t(r; \{\vartheta_k\}), \tag{2.9}$$

respectively. We assume $Y_k$, $k = 1, \ldots, K_t$, are independent with mean, variance, and third central moment

$$\begin{aligned}m_{1,k} = m_{1,k}(\vartheta_k) &:= \mathbb{E}[Y_k], \\ \overline{m}_{2,k} = \overline{m}_{2,k}(\vartheta_k) &:= \mathbb{V}[Y_k], \text{ and} \\ \overline{m}_{3,k} = \overline{m}_{3,k}(\vartheta_k) &:= \mathbb{E}[(Y_k - m_{1,k})^3],\end{aligned}$$

from where it follows that $m_1^R = \sum_k m_{1,k}$, $\overline{m}_2^R = \sum_k \overline{m}_{2,k}$, and $\overline{m}_3^R = \sum_k \overline{m}_{3,k}$. Below we derive analytical expressions for $m_{1,k}$, $\overline{m}_{2,k}$, and $\overline{m}_{3,k}$.

**B. Analytical expressions for the mean, variance, and third central moment of $Y_k$.** The $j$-th non-central moment of $Y_k$ is

$$m_{j,k} := (1 - p_k - q_k) \int y^j d\overline{G}(y; \vartheta_k) + q_k (v_k)^j, \tag{2.10}$$

where $m_{j,k} = m_{j,k}(p_k, q_k, \vartheta_k)$. From (2.5), (2.10) can be recast as

$$m_{j,k} = (1 - p_k - q_k) m_{j,k}^{(0,v_k)} + q_k (v_k)^j,$$

with $m_{j,k}^{(0,v_k)} := m_j^{(0,v)}(\vartheta_k)$, from where $m_{1,k}$ follows trivially. To obtain $\overline{m}_{2,k}$ we use the identity $\mathbb{V}[Y_k] = \mathbb{E}[Y_k^2] - \mathbb{E}[Y_k]^2$, from where it follows that

$$\overline{m}_{2,k} = m_{2,k} - m_{1,k}^2$$

and, in a similar way, using the identity $\mathbb{E}[(Y_k - \mathbb{E}[Y_k])^3] = \mathbb{E}[Y_k^3] - 3\mathbb{V}[Y_k]\mathbb{E}[Y_k] - \mathbb{E}[Y_k]^3$,

$$\overline{m}_{3,k} = m_{3,k} - 3\overline{m}_{2,k} m_{1,k} - m_{1,k}^3.$$

Under the beta model presented in Section 2.4, $m_{1,k}^{(0,v_k)}$, $\overline{m}_{2,k}^{(0,v_k)}$, and $\overline{m}_{3,k}^{(0,v_k)}$ are given in (2.6). The first central moment of $Y_k$ is given by

$$m_{1,k} = (1 - p_k - q_k)m_{1,k}^{(0,v_k)} + q_k(v_k).$$

Since $\mathbb{V}[Y_k^{(0,v_k)}] = \mathbb{E}[{Y_k^{(0,v_k)}}^2] - \mathbb{E}[Y_k^{(0,v_k)}]^2$, it follows that

$$m_{2,k}^{(0,v_k)} = \overline{m}_{2,k}^{(0,v_k)} + (m_{1,k}^{(0,v_k)})^2,$$

from where $m_{2,k}$ and $\overline{m}_{2,k}$ follow readily. The argument for $\overline{m}_{3,k}$ is similar. Using the third central moment identity it follows that

$$m_{3,k}^{(0,v_k)} = \overline{m}_{3,k}^{(0,v_k)} + 3\overline{m}_{2,k}^{(0,v_k)} m_{1,k}^{(0,v_k)} + (m_{1,k}^{(0,v_k)})^3,$$

from where $m_{3,k}$ and $\overline{m}_{3,k}$ follow.

# Chapter 3

# Poisson type models of the earthquake process

## 3.1 Introduction

Estimation and forecasting of *geophysical risk* has been the subject of intense scientific effort in both the seismological and statistical communities. Stochastic models with an increasing component of physical reasoning have been slowly gaining acceptance over the past two decades, giving rise to *statistical seismology*, a subject that aims to bridge the gap between physics-based models without statistics, and statistics-based models without physics. For an illustration of the range of issues now coming under the statistical seismology heading see [49].

From the statistical side, a fundamental impulse to statistical seismology has been the development of the theory of stochastic point processes, whose realisations can be represented as a family of Dirac measures in time or space or both. Vere-Jones has long advocated (see e.g., [41] and [42]) the theory of stochastic point processes as the natural framework for the development of geophysical risk models. The point process context is relevant so long as the

origin times can be treated as time instants, with which other variables can be associated. One then has a choice of treating the process as a point process in time and space, or as a marked point process in time, the mark for each event containing information about spatial and other parameters it is desired to include in the study.

Natvig and Tvete [28] have proposed a methodology aiming at predicting earthquake occurrence in time and space. Their statistical methods are in sharp contraposition to the point process framework, the interest being to predict event ocurrence in coarse geographic areas and time windows, as opposed to time instants and points in space.

In this and the next chapter we consider models of the process of earthquake occurrence. We assume that at the base of everything there is a filtered probability space $(\Omega, \mathcal{F} = \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ and operate in the marked point process framework introduced in Section 1.2, where we assume that the earthquake process is intensity driven and that, using the usual suggestive notation,

$$\Lambda(dt, dl, dz) := \mathbb{E}[N(dt, dl, dz)|\mathcal{F}_{t^-}]$$

is its intensity, with $\{\mathcal{F}_t\}$ a filtration to which the process is adapted. The latter includes information about the marks and times of (strictly) past events, and possibly about external variables or processes evolving in time in parallel with the earthquake process.

Vere-Jones [44] suggests that despite the massive literature on earthquake prediction, there are very few models from which intensities or probabilities can be calculated, and which at the same time have been tested on data, enjoy some semblance of credibility and yield a significant sharpening of the risk over the Poisson model. Some of the best-known and more successful attempts are reviewed in Appendix A.

In this chapter we formulate two Poisson type models of the earthquake process. In Section 3.2 we present a model whereby the earthquake process is Poissonian, setting the scene that will later allow us to assess the relative merits of more complex formulations. In this model, the intensity of occurrence of an earthquake is constant through time, with a deterministic time-independent marking mechanism. In Section 3.3 we propose a Bayesian formulation: a Poisson model with gamma intensity measure where both the intensity of occurrence of an earthquake and the marking mechanism are time-independent (*a priori*), albeit stochastic. The purpose of the latter model is to capture earthquake clustering in space, a pervasive feature of earthquake occurrence.

In both Poissonian models, independence plays a central role that leads to elegant analytical expressions for moments and representation results. Interesting extensions appear when the assumption of completely independent marks is removed, and ways in which either the marks can influence the future development of the process, or the current state of the process can influence the distribution of marks, or both, is considered. This is the case in, for instance, the stress release model presented in Chapter 4 or some of the ETAS models reviewed in Appendix A.

## 3.2 The spatial Poisson process

**A. Notation preliminaries.** If $X$ is a random variable, then we say $\mathcal{L}(X)$ denotes the probability law of $X$. The conditional law of $X$, given the random variable $Y$ (or the sigma-algebra it generates) is denoted by $\mathcal{L}(X|Y)$. We may also occasionally write $\mathcal{L}(X|Y = y)$. In general we have $\mathcal{L}(X, Y) = \mathcal{L}(X)\,\mathcal{L}(Y|X)$. In particular, $\mathcal{L}(X, Y) = \mathcal{L}(X)\,\mathcal{L}(Y)$ signifies that $X$ and $Y$ are independent. Some standard probability laws will have given names, e.g. $\mathcal{L}(N) = \mathrm{Po}(N; \lambda)$ means that $N$ has a Poisson distribution with mean $\lambda$ and

$\mathcal{L}(\Lambda) = \mathrm{Ga}(\Lambda; \alpha, \beta)$ means $\Lambda$ has a gamma distribution with shape parameter $\alpha$ and inverse scale parameter $\beta$.

**B. The earthquake process.** Let $\mu$ be Lebesgue measure on $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$, $\Lambda^L$ a positive finite measure on $(\mathbb{L}, \mathcal{B}_{\mathbb{L}})$, and $\Phi$ a probability measure on $(\mathbb{Z}, \mathcal{B}_{\mathbb{Z}})$. Define the product measure $\Lambda$ on $(\mathbb{R}_+ \times \mathbb{L} \times \mathbb{Z}, \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{L}} \otimes \mathcal{B}_{\mathbb{Z}})$ by specifying that, for all $\mathcal{T} \times \mathcal{L} \times \mathcal{Z} \in \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{L}} \otimes \mathcal{B}_{\mathbb{Z}}$,

$$\Lambda(\mathcal{T}, \mathcal{L}, \mathcal{Z}) := (\mu \times \Lambda^L \times \Phi)(\mathcal{T} \times \mathcal{L} \times \mathcal{Z}) = \mu(\mathcal{T})\Lambda^L(\mathcal{L})\Phi(\mathcal{Z}).$$

We assume that, for all $\mathcal{T} \times \mathcal{L} \times \mathcal{Z} \in \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{L}} \otimes \mathcal{B}_{\mathbb{Z}}$,

$$\mathcal{L}\left(N(\mathcal{T}, \mathcal{L}, \mathcal{Z})\right) = \mathrm{Po}\left(N(\mathcal{T}, \mathcal{L}, \mathcal{Z}); \Lambda(\mathcal{T}, \mathcal{L}, \mathcal{Z})\right),$$

and that $N(\mathcal{T}, \mathcal{L}, \mathcal{Z})$ and $N(\mathcal{T}', \mathcal{L}', \mathcal{Z}')$ are independent for any disjoint sets $\mathcal{T} \times \mathcal{L} \times \mathcal{Z}, \mathcal{T}' \times \mathcal{L}' \times \mathcal{Z}' \in \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{L}} \otimes \mathcal{B}_{\mathbb{Z}}$. The convolution property (see Appendix B, (B.5)) of the Poisson distribution ensures that $N$ is a well defined random measure, with the indexed family $\{N(\mathcal{T}, \mathcal{L}, \mathcal{Z}); \mathcal{T} \times \mathcal{L} \times \mathcal{Z} \in \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{L}} \otimes \mathcal{B}_{\mathbb{Z}}\}$ satisfying the Kolmogorov consistency conditions. We write

$$\mathcal{L}\left(N\right) = \mathrm{Po}\left(N; (\mathbb{R}_+ \times \mathbb{L} \times \mathbb{Z}, \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{L}} \otimes \mathcal{B}_{\mathbb{Z}}, \Lambda)\right). \tag{3.1}$$

We further assume that the *ground process* $\{T_i\}_{i \in \mathbb{N}}$ of occurrence times, with associated counting process

$$N(\mathcal{T}) := N(\mathcal{T}, \mathbb{L}, \mathbb{Z}),$$

the number of earthquakes occurred in the time interval $\mathcal{T}$, is independent of the marks process $\{(L_i, Z_i) = (L, Z)(T_i)\}_{i \in \mathbb{N}}$, that the locations $\{L_i\}$ are mutually independent and are governed by a probability measure $\Delta(\mathcal{L}), \forall \mathcal{L} \in$

$\mathcal{B}_{\mathbb{L}}$, and that the magnitudes $\{Z_i\}$ are mutually independent, governed by the probability measure $\Phi$, and independent of the locations $\{L_i\}$. With such a setting, the earthquake process is a generalized stationary compound Poisson process whereby the mean measure $\Lambda$ can be recast as

$$\Lambda(\mathcal{T}, \mathcal{L}, \mathcal{Z}) = \mu(\mathcal{T})\lambda\Delta(\mathcal{L})\Phi(\mathcal{Z}),$$

where $\mu(\mathcal{T})\lambda$, with $\lambda := \Lambda^L(\mathbb{L})$, is the mean measure of the Poisson ground process. To verify this observe that

$$\begin{aligned}
\Lambda(\mathcal{T}, \mathcal{L}, \mathcal{Z}) &= \mu(\mathcal{T})\Lambda^L(\mathcal{L})\Phi(\mathcal{Z}) \\
&= \mu(\mathcal{T})\Lambda^L(\mathbb{L})\frac{\Lambda^L(\mathcal{L})}{\Lambda^L(\mathbb{L})}\Phi(\mathcal{Z}) \\
&= \mu(\mathcal{T})\lambda\Delta(\mathcal{L})\Phi(\mathcal{Z}),
\end{aligned}$$

where $\Delta(\mathcal{L}) := \frac{\Lambda^L(\mathcal{L})}{\Lambda^L(\mathbb{L})}$. Define $N((0,s],\cdot,\cdot) := N(s,\cdot,\cdot)$ and let

$$\{\mathcal{F}_t = \sigma\{N(s,\mathcal{L},\mathcal{Z}); 0 < s \leq t, \mathcal{L} \in \mathcal{B}_{\mathbb{L}}, \mathcal{Z} \in \mathcal{B}_{\mathbb{Z}}\},\}_{t>0} \tag{3.2}$$

be the natural filtration generated by the counting process (trivially, $N(0,\mathcal{L},\mathcal{Z}) = 0$). The intensity of the latter, given $\mathcal{F}_{t^-}$, is

$$\mathbb{E}[N(dt,dl,dz)|\mathcal{F}_{t^-}] = \Lambda(dt,dl,dz) = dt\lambda\Delta(dl)\Phi(dz),$$

from where it follows that the intensity of occurrence of an earthquake is independent of the time and past history of the earthquake process.

**C. Specification of the location and magnitude distributions.** There is a host of conceivable candidates for the location and magnitude distributions $\Delta$ and $\Phi$, respectively. The distribution of earthquake sizes, for instance, has been

the subject of considerable research (see e.g. Utsu [40] for a thorough review). It is known empirically that magnitudes of earthquakes follow roughly a (left truncated) exponential distribution

$$\phi(z) = \gamma \exp\left(-\gamma(z - z_0)\right), \;\; z > z_0, \tag{3.3}$$

with $z_0$ a minimum magnitude threshold[1], known in the seismology literature as the *Gutenberg-Richter law* (Gutenberg [11]). Several alternatives to the latter have been proposed: the Pareto, left truncated gamma, and tapered Pareto models (see, e.g., Kagan and Schoenberg [19]) to name a few.

The distribution of earthquake location in space has been explored by, e.g., Musmeci and Vere-Jones [27] and Rathbun [37], where kernel density estimators have been applied to the modelling of the epicentral location of earthquakes. In the light of the large number of observations usually found in earthquake catalogs, an avenue of action is to avoid a particular choice of parametric model, or of kernel, in favour of a nonparametric one, using the empirical distribution of a sample of $N(\tau)$ observed locations $L_1, \ldots, L_{N(\tau)}$,

$$\Delta_{N(\tau)}(\mathcal{L}) := \frac{N(\tau, \mathcal{L}, \mathbb{Z})}{N(\tau)}, \;\; \forall \mathcal{L} \in \mathcal{B}_{\mathbb{L}},$$

as an estimate of the latter.

**D. Estimation.** The estimation of the earthquake process amounts to obtaining estimates of $\lambda$, $\Delta(\mathcal{L}), \forall \mathcal{L} \in \mathbb{L}$, and of any parameters appearing in the

[1]The minimum magnitude threshold $z_0$ must be carefully selected to ensure there is historical consistency, or *completeness*, in the catalog of events used for model calibration. The completeness of a catalog is determined, in most part, by a seismic network's capability of consistently detecting earthquakes of a certain magnitude over a period of time. For low magnitude earthquakes, for example, it is commonly the case that there is an apparent increase in seismic activity over time which could be entirely due to improvements in a seismic network's station density and capabilities. The quality of the waveforms, phase picks, hypocentral locations and magnitudes also vary considerably, with the most recently recorded events having the best determined parameters. For this reason only the largest shocks (typically $Z \geq 4$ or 5) are considered, the reason being that earthquakes of this size or larger are almost certainly regarded as being consistently detectable. See Musmeci and Vere-Jones [27], and Ogata et.al. [35].

specification of $\Phi$ (e.g. $\gamma$, under the assumption that $\Phi$ has density (with respect to Lebesgue measure on $(\mathbb{Z}, \mathcal{B}_{\mathbb{Z}})$) $\phi(z)$, given by (3.3)). As a result of the independence assumptions, the estimation of each of $\lambda$, $\Delta$, and $\Phi$, can be done separately, using classical methods.

For the estimation of $\lambda$, consider $\mathcal{F}_\tau$ and the occurrence times

$$T_1, T_2, \ldots, T_{N(\tau)} \in (0, \tau\,].$$

The probability that $N(\tau) = n$ and the $n$ events occurring in intervals $[t_i, t_i + dt_i)$, $i = 1, ..., n$, is

$$\mathbb{P}[N(\tau) = n, T_i \in [t_i, t_i + dt_i), i = 1, \ldots, n] = e^{-\lambda\tau} \prod_{i=1}^{n} \lambda dt_i,$$

from where the likelihood function and maximum likelihood estimator of $\lambda$,

$$L\left(\lambda | N(\tau), T_1, T_2, \ldots, T_{N(\tau)}\right) := e^{-\lambda\tau}\lambda^{N(\tau)}$$

and $\hat{\lambda} = \frac{N(\tau)}{\tau}$, respectively, follow readily.

## 3.3 The spatial Poisson process with gamma intensity measure

In this section we propose a new model of the earthquake process, a Poisson marked point process with random intensity measure obtained by assuming that the spatial measure $\Lambda^L$ in the Poisson process introduced in the previous section is stochastic. Under this model both the intensity of occurrence of an earthquake and the marking mechanism are time-independent (*a priori*), albeit stochastic. Through the stochastic spatial intensity the proposed model captures earthquake clustering in space, one of the most pervasive features of

catalog data.

The spatial gamma and Dirichlet measures are introduced in paragraph A, and a representation result for the gamma measure is obtained. Particular care is taken to ensure that the measures are well defined and for this reason the workings on an arbitrary finite partition of the location space are made explicit, before extending the measure(s) to a bona-fide sigma algebra. The model is then formulated in paragraph B, defining the Poisson model intensity measure $\Lambda$ as a random measure underpinned by the gamma measure $\Lambda^L$. The posterior distribution of the gamma measure is defined in paragraph C where, again, care is taken to show the workings of the posterior measure on an arbitrary finite partition of space before extending it. Finally, the posterior distribution of the earthquake process and its intensity is derived in paragraph D.

**A. The spatial gamma and Dirichlet measures.** Let $\alpha$ be a positive finite measure on $(\mathbb{L}, \mathcal{B}_{\mathbb{L}})$. Define a random measure $\Lambda^L$ on $(\mathbb{L}, \mathcal{B}_{\mathbb{L}})$ by assuming that, for all $\mathcal{L} \in \mathcal{B}_{\mathbb{L}}$,

$$\mathcal{L}(\Lambda^L(\mathcal{L})) = \mathrm{Ga}(\Lambda^L(\mathcal{L}); \alpha(\mathcal{L}), \beta),$$

and that $\Lambda^L(\mathcal{L})$ and $\Lambda^L(\mathcal{L}')$ are independent for any disjoint sets $\mathcal{L}, \mathcal{L}' \in \mathcal{B}_{\mathbb{L}}$. The convolution property of the gamma distribution ensures that $\Lambda^L$ is a well defined random measure, with the indexed family $\{\Lambda^L(\mathcal{L});\, \mathcal{L} \in \mathcal{B}_{\mathbb{L}}\}$ satisfying the Kolmogorov consistency conditions. We write

$$\mathcal{L}\left(\Lambda^L\right) \;=\; \mathrm{Ga}\left(\Lambda^L; (\mathbb{L}, \mathcal{B}_{\mathbb{L}}, \alpha), \beta\right) \;. \tag{3.4}$$

Let $\Delta$ be the random probability measure obtained upon norming $\Lambda^L$ to a

probability:

$$\Delta(\mathcal{L}) := \frac{\Lambda^L(\mathcal{L})}{\Lambda^L(\mathbb{L})}, \ \forall \mathcal{L} \in \mathcal{B}_{\mathbb{L}}. \tag{3.5}$$

(This is well defined: $\Lambda^L(\mathbb{L}) > 0$ almost surely since $\alpha(\mathbb{L}) > 0$.) From Appendix B, Section B.6, we know that, for any measurable partition $\{\mathcal{L}_1, \ldots, \mathcal{L}_k\}$ of $\mathbb{L}$, we have

$$\mathcal{L}(\Delta(\mathcal{L}_1), \ldots, \Delta(\mathcal{L}_k)) = \mathrm{Dir}(\Delta(\mathcal{L}_1), \ldots, \Delta(\mathcal{L}_k); \alpha(\mathcal{L}_1), \ldots, \alpha(\mathcal{L}_k)) .$$

These joint multidimensional distributions determine the random measure $\Delta$ uniquely, and we call this measure the *Dirichlet* measure on $(\mathbb{L}, \mathcal{B}_{\mathbb{L}})$ with parameter $\alpha$, written

$$\mathcal{L}(\Delta) = \mathrm{Dir}(\Delta; (\mathbb{L}, \mathcal{B}_{\mathbb{L}}, \alpha)) . \tag{3.6}$$

This construction gives the following representation result (the infinite dimensional extension of (B.20)):

$$\begin{aligned} \mathcal{L}\left(\Lambda^L\right) &= \mathrm{Ga}\left(\Lambda^L; (\mathbb{L}, \mathcal{B}_{\mathbb{L}}, \alpha), \beta\right) \\ &\Leftrightarrow \\ \mathcal{L}\left(\Lambda^L(\mathbb{L}), \Delta\right) &= \mathrm{Ga}\left(\Lambda^L(\mathbb{L}); \alpha(\mathbb{L}), \beta\right) \mathrm{Dir}(\Delta; (\mathbb{L}, \mathcal{B}_{\mathbb{L}}, \alpha)) . \end{aligned} \tag{3.7}$$

Thus, the gamma measure can be constructed as the product of a gamma variable and a Dirichlet measure.

**B. The earthquake process I.** Let $\mu$ be Lebesgue measure on $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$, $\Lambda^L$ a gamma random measure as in (3.4), and $\Phi$ a probability measure on $(\mathbb{Z}, \mathcal{B}_{\mathbb{Z}})$. Define the random product measure $\Lambda$ on $(\mathbb{R}_+ \times \mathbb{L} \times \mathbb{Z}, \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{L}} \otimes \mathcal{B}_{\mathbb{Z}})$ by assuming that, for all $\mathcal{T} \times \mathcal{L} \times \mathcal{Z} \in \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{L}} \otimes \mathcal{B}_{\mathbb{Z}}$,

$$\Lambda(\mathcal{T}, \mathcal{L}, \mathcal{Z}) := (\mu \times \Lambda^L \times \Phi)(\mathcal{T} \times \mathcal{L} \times \mathcal{Z}) = \mu(\mathcal{T})\Lambda^L(\mathcal{L})\Phi(\mathcal{Z}).$$

We assume that earthquake magnitude is independent of time and of location and that $\Lambda$ drives the earthquake generating mechanism in a way such that, conditional on $\Lambda$, $N$ is a Poisson measure as in (3.1),

$$\mathcal{L}\left(N|\Lambda\right) = \mathrm{Po}\left(N; \left(\mathbb{R}_+ \times \mathbb{L} \times \mathbb{Z}, \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{L}} \otimes \mathcal{B}_{\mathbb{Z}}\right), \Lambda\right).$$

The gamma distribution is the natural conjugate prior to the Poisson distribution and so, for all $\mathcal{T} \times \mathcal{L} \in \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{L}}$,

$$\mathcal{L}\left(\Lambda^L(\mathcal{L})|N\right) = \mathrm{Ga}\left(\Lambda^L(\mathcal{L}); \alpha(\mathcal{L}) + N(\mathcal{T}, \mathcal{L}), \beta + \mu(\mathcal{T}\right),$$

where we define $N(\cdot, \cdot, \mathbb{Z}) := N(\cdot, \cdot)$, which is equivalent to (using (3.7))

$$\begin{aligned}\mathcal{L}(\Lambda^L(\mathbb{L}), (\Delta(\mathcal{L}), \Delta(\mathcal{L}^c)) \,|\, N) \;=\; & \mathrm{Ga}\left(\Lambda^L(\mathbb{L}); \alpha(\mathbb{L}) + N(\mathcal{T}, \mathbb{L}), \beta + \mu(\mathcal{T})\right) \\ & \times \mathrm{Dir}\left(\Delta(\mathcal{L}), \Delta(\mathcal{L}^c); \alpha(\mathcal{L}), \alpha(\mathcal{L}^c)\right).\end{aligned}$$

**C. Posterior distribution of the gamma measure.** The posterior distribution of $\Lambda^L$, given $N$ observed in the time interval $(0, \tau]$, is obtained as follows (trivially, for all $\mathcal{L} \in \mathcal{B}_{\mathbb{L}}$, $N(0, \mathcal{L}) = 0$). Partition $(0, \tau]$ into $(t_{h-1}, t_h]$, $h = 1, \ldots, j$ where $0 =: t_0 < t_1 < \cdots < t_j := \tau$, and denote the length of the $h$-th interval by $\tau_h := t_h - t_{h-1}$, $h = 1, \ldots, j$. Partition $\mathbb{L}$ into $\{\mathcal{L}_1, \ldots, \mathcal{L}_k\}$,

and denote $\alpha_i := \alpha(\mathcal{L}_i)$, $i = 1, \ldots, k$. The joint distribution of

$$\{N_{hi} := N((t_{h-1}, t_h], \mathcal{L}_i, \mathbb{Z});\, h = 1, \ldots j,\, i = 1, \ldots, k\}$$

and

$$\{\Lambda_i := \Lambda^L(\mathcal{L}_i);\, i = 1, \ldots, k\}$$

is given by (the product of)

$$\mathcal{L}(\Lambda_i;\, i = 1, \ldots, k) \;=\; \prod_{i=1}^{k} \mathrm{Ga}(\Lambda_i;\, \alpha_i; \beta)$$

and

$$\mathcal{L}\left(N_{hi};\, h = 1, \ldots j,\, i = 1, \ldots, k \,|\, \Lambda_i;\, i = 1, \ldots, k\right) \;=\; \prod_{i=1}^{k} \prod_{h=1}^{j} \mathrm{Po}\left(N_{hi};\, \tau_h\, \Lambda_i\right) .$$

The conditional distribution of $\{\Lambda_i;\, i = 1, \ldots, k\}$, given $\{N_{hi};\, h = 1, \ldots j,\, i = 1, \ldots, k\}$, is proportional to their joint distribution:

$$\begin{aligned}
&\mathcal{L}\left(\Lambda_i;\, i = 1, \ldots, k \,|\, N_{hi};\, h = 1, \ldots j,\, i = 1, \ldots k\right) \\
&\quad \propto \prod_{i=1}^{k} \left( \mathrm{Ga}(\Lambda_i;\, \alpha_i, \beta) \prod_{h=1}^{j} \mathrm{Po}\left(N_{hi};\, \tau_h\, \Lambda_i\right) \right) \\
&\quad \propto \prod_{i=1}^{k} \left( \Lambda_i^{\alpha_i - 1}\, e^{-\beta\, \Lambda_i}\, \Lambda_i^{\sum_{h=1}^{j} N_{hi}}\, e^{-\sum_{h=1}^{j} \tau_h \Lambda_i} \right) \\
&\quad \propto \prod_{i=1}^{k} \mathrm{Ga}\left( \Lambda_i;\, \alpha_i + \sum_{h=1}^{j} N_{hi},\, \tau + \beta \right) . \qquad (3.8)
\end{aligned}$$

Since this holds for all partitions of $(0, \tau]$ and in view of the convolution property (B.15) of the gamma distribution, (3.8) is also the conditional distribution of

$\{\Lambda_i;\, i = 1, \ldots, k\}$, given $\mathcal{F}_\tau$ as defined in (3.2):

$$\mathcal{L}\left(\Lambda_i;\, i = 1, \ldots, k \,|\, \mathcal{F}_\tau\right) \,=\, \prod_{i=1}^{k} \mathrm{Ga}\left(\Lambda_i;\, \alpha_i + \sum_{h=1}^{j} N_{hi},\, \tau + \beta\right) .$$

By the extension theorem for measures, we obtain

$$\mathcal{L}\left(\Lambda^L \,|\, \mathcal{F}_\tau\right) \,=\, \mathrm{Ga}\left(\Lambda^L;\, (\mathbb{L}, \mathcal{B}_\mathbb{L}, \alpha(\cdot) + N((0,\tau], \cdot)),\, \tau + \beta\right) . \tag{3.9}$$

**D. The earthquake process II.** The model and the results above can be recast by use of (3.7). Assume that

$$\mathcal{L}(\Lambda^L(\mathbb{L}), \Delta) \,=\, \mathrm{Ga}\left(\Lambda^L(\mathbb{L});\, \alpha(\mathbb{L}), \beta\right) \mathrm{Dir}\left(\Delta;\, (\mathbb{L}, \mathcal{B}_\mathbb{L}, \alpha)\right) \tag{3.10}$$

and that

$$\mathcal{L}(N|(\Lambda^L(\mathbb{L}), \Delta)) = \mathrm{Po}\left(N;\, \left(\mathbb{R}_+ \times \mathbb{L} \times \mathbb{Z}, \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_\mathbb{L} \otimes \mathcal{B}_\mathbb{Z}, \mu \times \Lambda^L(\mathbb{L})\Delta \times \Phi\right)\right) \tag{3.11}$$

Then

$$\begin{aligned} \mathcal{L}\left(\Lambda^L(\mathbb{L}), \Delta \,|\, \mathcal{F}_\tau\right) \;&=\; \mathrm{Ga}\left(\Lambda^L(\mathbb{L});\, N((0,\tau]) + \alpha(\mathbb{L}), \tau + \beta\right) \\ &\quad \times \mathrm{Dir}\left(\Delta; (\mathbb{L}, \mathcal{B}_\mathbb{L}, N((0,\tau], \cdot) + \alpha(\cdot))\right) \end{aligned} \tag{3.12}$$

and

$$\mathcal{L}\left(N|\Lambda, \mathcal{F}_\tau\right) = \mathrm{Po}\left(N; \mu \times \Lambda^L(\mathbb{L})\Delta \times \Phi\right),$$

from where it follows that, for all $\mathcal{T} \times \mathcal{L} \times \mathcal{Z} \in \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_\mathbb{L} \otimes \mathcal{B}_\mathbb{Z}$, with $\mathcal{T} := (\tau, s]$,

$$\begin{aligned}
\mathbb{E}\left[N(\mathcal{T},\mathcal{L},\mathcal{Z})|\mathcal{F}_\tau\right] &= \mathbb{E}\left[\mathbb{E}\left[N(\mathcal{T},\mathcal{L},\mathcal{Z})|\Lambda,\mathcal{F}_\tau\right]|\mathcal{F}_\tau\right] \\
&= \mathbb{E}\left[\mu(\mathcal{T})\Lambda^L(\mathbb{L})\Delta(\mathcal{L})\Phi(\mathcal{Z})|\mathcal{F}_\tau\right] \\
&= \mu(\mathcal{T})\mathbb{E}\left[\Lambda^L(\mathbb{L})|\mathcal{F}_\tau\right]\mathbb{E}\left[\Delta(\mathcal{L})|\mathcal{F}_\tau\right]\Phi(\mathcal{Z}) \\
&= \mu(\mathcal{T})\cdot\frac{\alpha(\mathbb{L})+N\left(\tau\right)}{\tau+\beta}\cdot\Delta_\tau(\mathcal{L})\cdot\Phi(\mathcal{Z}),
\end{aligned}$$

where

$$\begin{aligned}
\Delta_\tau(\mathcal{L}) &:= \mathbb{E}\left[\Delta(\mathcal{L})|\mathcal{F}_\tau\right] \\
&= \frac{\alpha(\mathcal{L})+N\left(\tau,\mathcal{L}\right)}{\alpha(\mathbb{L})+N\left(\tau\right)} \\
&= p_\tau^\Delta\Delta_0(\mathcal{L})+(1-p_\tau^\Delta)\Delta_{N(\tau)}(\mathcal{L}),
\end{aligned} \tag{3.13}$$

with $p_\tau^\Delta := \frac{\alpha(\mathbb{L})}{\alpha(\mathbb{L})+N(\tau)}$, $\Delta_0(\mathcal{L}) := \frac{\alpha(\mathcal{L})}{\alpha(\mathbb{L})}$, and $\Delta_{N(\tau)}(\mathcal{L}) = \frac{N(\tau,\mathcal{L})}{N(\tau)}$, is a Bayes estimate of $\Delta(\mathcal{L})$. The expectation (3.13) follows readily upon recognising that

$$\mathcal{L}\left(\Delta(\mathcal{L})|\mathcal{F}_\tau\right) = \mathrm{Be}\left(\alpha(\mathcal{L})+N(\tau,\mathcal{L}),\alpha(\mathcal{L}^c)+N(\tau,\mathcal{L}^c)\right).$$

Note that if $\alpha(\mathbb{L})$ is large, relative to $N(\tau)$, little weight is given to the observations. On the other hand, if $\alpha(\mathbb{L})$ is small, relative to $N(\tau)$, little weight is given to the prior estimate $\Delta_0$. As $\alpha(\mathbb{L})$ tends to zero, the Bayes estimate $\Delta_\tau$ converges to the empirical distribution function $\Delta_{N(\tau)}$.

In particular, we have

$$\mathbb{E}\left[N(d\tau,dl,dz)|\mathcal{F}_{\tau^-}\right] = d\tau\cdot\frac{\alpha(\mathbb{L})+N\left(\tau^-\right)}{\tau+\beta}\cdot\Delta_{\tau^-}(dl)\cdot\Phi(dz).$$

Since the factor $d\tau$ annihilates the left limit, we can use the right-continuous version of the intensity of the process and so

$$\mathbb{E}\left[N(d\tau, dl, dz)|\mathcal{F}_{\tau^-}\right] = d\tau \cdot \frac{\alpha(\mathbb{L}) + N\left(\tau\right)}{\tau + \beta} \cdot \Delta_\tau(dl) \cdot \Phi(dz).$$

# Chapter 4

# A stress release model of the earthquake process

A distinctive feature of the Poisson type models proposed in the preceding chapter is their time-independent nature. In this chapter we will consider a model relating to the physical processes causing earthquakes, capturing the way in which seismic activity, with its periods of quiescence and activation, depends on the dynamical change of the underlying stress-field of a seismic region. The physical essence of this model is Reid's elastic rebound theory of major earthquakes, which postulates that elastic stress in a seismically active region accumulates due to movement of tectonic plates and is released when the stress exceeds the strength of the medium. The stress release model was first proposed by Vere-Jones [42] to represent the deterministic build-up of stress within a region and its stochastic release through earthquakes, further developing Knopoff's [23] stochastic Markov model for the occurrence of main-sequence earthquakes.

The model has been applied to historical earthquake catalogs from China (Zheng and Vere-Jones [51] and Liu et al. [25]), Persia and Japan (Zheng and Vere-Jones [52]), Taiwan (Bebbington and Harte [5]), and Italy (Rotondi and

Varini [38]). In all these instances, the aim has been to model the ground process $\{T_i\}_{i\in\mathbb{N}}$ of occurrence times or the marked point process $\{(T_i, Z_i)\}_{i\in\mathbb{N}}$. In this chapter we include earthquake location as a mark component and consider a stress release model of the marked point process $\{(T_i, Z_i, L_i)\}_{i\in\mathbb{N}}$.

## 4.1 The earthquake process

We operate in the marked point process framework introduced in Section 1.2 where, in the interest of simplicity of notation, we define

$$N(t, \mathcal{Z}) := N(t, \mathcal{Z}, \mathbb{L}), \text{ and } N(t) := N(t, \mathbb{Z}, \mathbb{L}).$$

## 4.2 The stress process

The general assumption is that the probabilities of earthquakes occurring within the seismic region (or on the fault) in question are determined by an unobserved state variable that increases linearly between earthquakes and decreases instantaneously when an event occurs. We assume that, at any time $t > 0$,

$$X_t := X_0 + \rho t - S_t,$$

where $X_0$ is the initial stress level, represents the balance between the accumulated tectonic stress in a region, building up linearly at a fixed rate $\rho > 0$, and the total amount of stress released through earthquakes,

$$S_t := \sum_{i=1}^{\infty} 1_{(0,t]}(T_i) s^*(Z_i) = \int_0^t \int_{z_0}^{\infty} s^*(z) N(ds, dz),$$

where $s^*(z) := 10^{\frac{3}{4}(z-z_0)}$ is (up to a constant of proportionality) the stress released upon the occurrence of an earthquake with magnitude $z$, and $z_0$ is an

appropriate minimum magnitude threshold.

It is implicit in the formulation of the model that earthquakes lower the regional stress and hence reduce the likelihood of immediately subsequent events, limiting its validity to sequences of main events with magnitude above a threshold $z_0$, usually taken to be 5 or 6. This is not considered a major drawback since it is the main events that carry the majority of tectonic information and are of primary practical concern. The build up and release of stress from smaller events follows a much more complex mechanism; they tend to occur in clusters resulting primarily from perturbations of a near critical system where the stress released does not simply dissipate, it moves down the fault and concentrates in sites nearby, typically at the ends of the rupture, increasing the level in the stress field. (See, e.g., Bebbington and Harte [5], and King et.al. [22]).

## 4.3 The earthquake process intensity

The stochastic mechanism governing the earthquake process and the stress process must specify how the former is driven by the latter. Letting

$$\{\mathcal{F}_t = \sigma\{N(s, \mathcal{Z}, \mathcal{L}), X_s; 0 < s \leq t, \mathcal{Z} \in \mathcal{B}_{\mathbb{Z}}, \mathcal{L} \in \mathcal{B}_{\mathbb{L}}\}\}_{t>0}$$

denote the filtration generated by the stress and counting processes, the stress release model assumes that, using the usual suggestive notation,

$$\mathbb{E}\left[N(dt, dz, dl)|\mathcal{F}_{t^-}\right] := \lambda(X_{t^-}, z, l)dt\, dz\, dl,$$

whereby at any time $t$ when the current level of stress is $X_{t^-}$, the intensity of occurrence of an earthquake with mark in $dz \times dl$ is $\lambda(X_{t^-}, z, l)dzdl$, independent of the time and past history, but dependent on the current stress. The intensity of occurrence of an earthquake is $\lambda(X_{t^-}) = \int_{\mathbb{L}} \int_{z_0}^{z(t)} \lambda(X_{t^-}, z, l)dzdl$, where

$$z(t) = z_0 + \frac{4}{3}\log_{10}(X_{t^-}) \tag{4.1}$$

is the upper bound for the magnitude of an earthquake occurring at time $t$, (obviously $\lambda(X_{t^-}, z, l) = 0$ for $z > z(t)$). The intensity of occurrence of an earthquake of magnitude $z$ is

$$\lambda(X_{t^-}, z) = \int_{\mathbb{L}} \lambda(X_{t^-}, z, l)dl,$$

and the intensity of occurrence of an earthquake is

$$\lambda(X_{t^-}) = \int_{z_0}^{z(t)} \lambda(X_{t^-}, z)dz.$$

The ratio

$$f(z, l|X_{t^-}) := \frac{\lambda(X_{t^-}, z, l)}{\lambda(X_{t^-})}$$

is the joint density of the mark $(Z, L)$ of an earthquake occurring at time $t$. We assume the location of an earthquake is independent of the stress level triggering it and, as a result, of its magnitude. In consequence,

$$f(z, l|X_{t^-}) = \phi(z|X_{t^-})\delta(l), \tag{4.2}$$

where $\phi$ is the density of $Z$ and $\delta$ is the density of $L$, and

$$\lambda(X_{t^-}, z, l) = \lambda(X_{t^-})\phi(z|X_{t^-})\delta(l). \tag{4.3}$$

The intensity function (4.3) determines the probability structure of $N$ uniquely (see, e.g., Daley and Vere-Jones [7], Proposition 7.3.IV) and is the key not just to the likelihood analysis, and hence to fitting and testing the model, but also to simulation and prediction.

**A. Specification of the intensity function.** A commonly used parametric model for the intensity $\lambda(\cdot)$ proposed by Zheng and Vere-Jones in [51] and [52] specifies that

$$\lambda(X_{t^-}) = \exp(\mu + \nu X_{t^-}), \tag{4.4}$$

where $\mu$ and $\nu$ are fixed parameters, the latter positive.

A problem arising is that the stress process $X$ is not directly observable. More specifically, the initial stress $X_0$ plays a special role since it is an unknown random variable generated by the stationary distribution $F_X$ of $X$ (see discussion in Section 4.4). If we should take account of this for maximum likelihood estimation purposes, we would be loading an unbearable burden on ourselves. Firstly, the likelihood would be the integral of the conditional likelihood, given $X_0$, with respect to the stationary distribution. Secondly, the stationary distribution does not in general admit an explicit expression and would therefore render the likelihood maximization infeasible. To evade such problems we choose to work in the conditional distribution, given the initial stress value, which amounts to letting $X_0$ be a parameter. A great gain in tractability against a small sacrifice of information. Upon absorbing $\nu X_0$ into $\mu$ and introducing new parameters $a = \mu + \nu X_0$, $b = \nu\rho$, and $c = \frac{1}{\rho}$, we arrive at the intensity function $\lambda(X_{t^-}) = \exp\left(a + b\left[t - cS_{t^-}\right]\right)$. Thus, we will henceforth be working with the parametric intensity function

$$\lambda(t) := \lambda(X_{t^-}) = \exp\left(a + b\left[t - cS_{t^-}\right]\right), \tag{4.5}$$

where $\theta = (a, b, c)$ is the parameter vector.

**B. Specification of the mark distribution.** There is a vast number of candidates for the magnitude and location distributions in (4.2) (we refer to the discussion in Section 3.2 C). Considering the restriction imposed by the upper bound $z(t)$, we suggest a truncated exponential distribution with support $(z_0, z(t)]$,

$$\phi(z|t) := \phi(z|X_{t^-}) = \frac{\gamma \exp\left(-\gamma(z - z_0)\right)}{1 - \exp\left(-\gamma\left(z(t) - z_0\right)\right)}.$$

Note that this is effectively a truncated version of the Gutenberg-Richter law introduced in (3.3).

Our proposed model for the distribution of spatial location is the empirical distribution of a sample of $N(\tau)$ observed locations $L_1, \ldots, L_{N(\tau)}$,

$$\Delta_{N(\tau)}(\mathcal{L}) := \frac{N(\tau, \mathbb{Z}, \mathcal{L})}{N(\tau)}, \ \forall \mathcal{L} \in \mathcal{B}_{\mathbb{L}}.$$

## 4.4 Stationarity of the stress process

**A. Background.** With the specification above, the pair $(X, N)$ is a Markov process, as is $X$ taken alone, since it consists of a non-random drift and stochastic jumps governed by an intensity that depends only on the current state of $X$ itself. The existence of a stationary distribution for the stress process then becomes an issue.

Elastic stress in a seismically active region accumulates due to movement of tectonic plates which are subject to forces working in different directions. Relative to the geological timescale, our short historical time horizon is infinitesimal and so we assume that both the tectonic forces and the conditions of the earth's crust are constant. Thus, one should expect the stress process to possess a stationary distribution. It is in general not a simple matter to determine whether a given process possesses a stationary distribution. For a brief account of the vast

theory devoted to this problem we refer to Davis [8]. The ergodicity and existence of moments of the stress release model have been explored by Vere-Jones ( [47] and [43]), Zheng [50], and Last [24], who have considered conditions for the existence of a stationary distribution and finiteness of its moments. Some results have been obtained by Borovkov and Vere-Jones in [4] for a special stress release model with exponential intensity. We shall not elaborate on this issue, but point out that a stationary distribution does not always exist and that this fact must be taken into account in our specification of the model parameters and in their inferences.

**B. Intuition behind the existence of a stationary regime.** Intuitively, the existence of a stationary regime suggests that if $X_t$ becomes too small, the rate of stress input $\rho$ should exceed the mean rate of stress released due to the occurrence of earthquakes, and similarly that for $X_t$ large the reverse inequality should hold. To verify that this is indeed the case, assume that $X$ possesses a stationary distribution, which we denote by $F_X$. Let $H$ be a continuously differentiable function such that $\mathbb{E}[H(X_t)] = \int_0^\infty H(x)dF_X(x)$ exists and is finite. By the direct backward construction, conditioning on what happens in a small time interval $(0, h)$,

$$
\begin{aligned}
&\mathbb{E}\left[H(X_h)|\, X_0 = x\right] \\
&\quad = (1-\lambda(x)h)H(x+\rho h) + \int_0^{z_x} \lambda(x,z)h\, dz\, H(x - s^*(z)) + o(h) \\
&\quad = (1-\lambda(x)h)(H(x) + H'(x)\rho\, h) + \int_0^{z_x} \lambda(x,z)h\, dz\, H(x - s^*(z)) + o(h) \\
&\quad = H(x) + H'(x)\rho\, h - \int_0^{z_x} \lambda(x,z)[H(x) - H(x - s^*(z))]\, dz\, h + o(h),
\end{aligned}
$$

where $z_x$ denotes the maximum possible magnitude for an earthquake occurred when the current level of stress is $x$. Integrating $\mathbb{E}\left[H(X_h)|\, X_0 = x\right]$ with respect to $F_X$, using

$$\int_0^\infty \mathbb{E}\left[\,H(X_h)|\,X_0 = x\right] dF_X(x) = \mathbb{E}\mathbb{E}\left[\,H(X_h)|\,X_0\right] = \mathbb{E}[H(X_h)],$$

we obtain

$$\begin{aligned}\mathbb{E}[H(X_h)] &= \int_0^\infty H(x)dF_X(x) + \int_0^\infty H'(x)dF_X(x)\,\rho\,h \\ &\quad - \int_0^\infty \int_0^{z_x} \lambda(x,z)[H(x) - H(x - s^*(z))]dz\,dF_X(x)\,h + o(h).\end{aligned}$$

Since $X_h$ is distributed according to $F_X$, the term on the left cancels against the first term on the right, and we arrive at

$$\rho \int_0^\infty H'(x)dF_X(x) = \int_0^\infty \int_0^{z_x} \lambda(x,z)[H(x) - H(x - s^*(z))]\,dz\,dF_X(x).$$

This relationship holds for all $H$ satisfying the stated conditions ($H$ is continuously differentiable and $\mathbb{E}[H(X_t)]$ exists and is finite) and, in principle, this determines $F_X$. It suffices to consider functions of the form $H(x) = e^{-\eta x}$ since they produce the Laplace transform,

$$\hat{F}_X(\eta) = \int_0^\infty e^{-\eta x} dF_X(x),$$

which determines $F_X$ uniquely. We find

$$\rho\eta \int_0^\infty e^{-\eta x} dF_X(x) = \int_0^\infty \int_0^{z_x} \lambda(x,z)\left[e^{\eta s^*(z)} - 1\right] dz\,e^{-\eta x} dF_X(x). \qquad (4.6)$$

One cannot hope to extract from this relationship an explicit expression for $\hat{F}_X$ (or $F_X$) except possibly for very simple specifications of the intensity function $\lambda(x,z)$. Upon differentiating both sides of (4.6) with respect to $\eta$ and setting

$\eta = 0$, we obtain

$$\rho = \int_0^\infty \int_0^{z_x} s^*(z)\lambda(x,z)\, dz\, dF_X(x),$$

from where it follows that, in the stationary state, the build-up of stress per time unit equals the expected stress release per time unit, as expected on intuitive grounds. Observe that, since the stress is not affected by the location of an earthquake in our model, we have omitted the location in the above and worked with $\lambda(x,z)$ as the relevant intensity function.

## 4.5 Model estimation

**A. Maximum likelihood estimation.** Assume $(a,b,c,\gamma) \in \mathbb{R}^4$ is our parameter vector. Taking our stand at time $\tau$ (the present time), the likelihood of the observations $(T_i, Z_i)$, $i = 1, \ldots, N_{(\tau)}$, is (see Proposition 7.3.III in [7])

$$L(a,b,c,\gamma) = \prod_{i=1}^{N(\tau)} \lambda_{a,b,c}(T_i)\, \phi_{c,\gamma}(Z_i|T_i) \exp\left(-\int_0^\tau \lambda_{a,b,c}(t)dt\right), \qquad (4.7)$$

where the subscripts in $\lambda_{a,b,c}(\cdot)$ and $\phi_{c,\gamma}(\cdot)$[1] mean that functions $\lambda$ and $\phi$ are specified up to a vector of parameters $(a,b,c)$ and $(c,\gamma)$, respectively. To the extent that $X_0$ is not directly estimable (it is confounded with $a$), the dependence of $\phi_{c,\gamma}(z|t)$ on the latter (through $(z(t))$ renders the maximisation of the likelihood impossible. To overcome this obstacle, consider the log likelihood

$$\ln L(a,b,c,\gamma) = \sum_{i=1}^{N(\tau)} \ln\left(\lambda_{a,b,c}(T_i)\, \phi_{c,\gamma}(Z_i|T_i)\right) - \int_0^\tau \lambda_{a,b,c}(t)dt, \qquad (4.8)$$

[1]Note that the magnitude density $\phi$ depends on $c = \frac{1}{\rho}$ through $z(t)$, where $z(t) = z_0 + \frac{4}{3}\log_{10}(X_0 + \rho t - S_{t-})$.

and the likelihoods

$$\ln L_1(a,b,c) := \sum_{i=1}^{N(\tau)} \ln \lambda_{a,b,c}(T_i) - \int_0^\tau \lambda_{a,b,c}(t)dt, \tag{4.9}$$

and

$$\ln L_2(c,\gamma) := \sum_{i=1}^{N(\tau)} \ln \phi_{c,\gamma}(Z_i|T_i), \tag{4.10}$$

where $\ln L = \ln L_1 + \ln L_2$. Letting $\gamma$ be a nuisance parameter, and treating the marks as a set of given values, about whose structure or distribution we have no information, $\ln L_1$ can be maximised to obtain the MLE of $(a,b,c)$, given by

$$(\hat{a},\hat{b},\hat{c}) = \underset{(a,b,c)\in\mathbb{R}^3}{\arg\max} \ln L_1(a,b,c).$$

Assuming the usual regularity conditions are satisfied, the MLE is consistent

$$(\hat{a},\hat{b},\hat{c}) \xrightarrow{p} (a,b,c), \tag{4.11}$$

and asymptotically normally distributed

$$(\hat{a},\hat{b},\hat{c}) \xrightarrow{d} \mathrm{N}\left((a,b,c), \mathcal{I}^{-1}(a,b,c)\right), \tag{4.12}$$

where $\mathcal{I}(a,b,c)$ is Fisher's information matrix.

To obtain an estimate of $\gamma$ we proceed as follows. From (4.1) we know that, for all $i = 1,\ldots,N(\tau)$, $z(T_i) \geq Z_i$. This constraint is tantamount to requiring that, for all $i$, the stress released by an earthquake of magnitude $Z_i$ occurring at time $T_i$, be less than or equal to the amount of stress available at time $T_i$, this is, for all $i$, $s(Z_i) \leq X_0 + \rho T_i - S_{T_i^-}$, from where it follows that

$$X_0 \geq s(Z_i) - \rho T_i + S_{T_i^-}, \ \forall i, \tag{4.13}$$

and a lower bound for $X_0$ is given by

$$X_0^* := \max_{i=1,\ldots,N(\tau)} \left\{0; s(Z_i) - \rho^* T_i + S_{T_i^-}\right\},$$

where $\rho^* = \frac{1}{\tilde{c}}$. Letting $X_{T_i}^* := X_0^* + \rho^* T_i - S_{T_i^-}$, for all $i$, the upper limit $z(T_i)$ of the support of $\phi(Z_i|T_i)$ can be calculated using (4.1) and so an estimator of $\gamma$ follows upon maximisation of the likelihood function

$$\ln L^*(\gamma) := \sum_{i=1}^{N(\tau)} \ln \phi_\gamma \left(Z_i|T_i\right),$$

with

$$\hat{\gamma} = \arg\max_{\gamma \in \mathbb{R}} \ln L^*(\gamma).$$

**B. Bayesian approach to estimation.** The maximum likelikhood approach described above relies on letting $\gamma$ be a nuisance parameter, treating the marks as a set of given values, and estimating $X_0^*$ to enable the maximum likelihood estimation of $\gamma$. A Bayesian approach to estimation is an alternative with the advantage that these restrictions are not required. Consider the original specification

$$\lambda(t) = \exp(\mu + \nu[X_0 + \rho t - S_{t^-}])$$

and the model parameters $(\mu, \nu, \rho, \gamma, X_0)$. The joint posterior density of the latter is given by

$$\pi(\xi, \rho, X_0|\{T_i, Z_i\}) = \frac{1}{K_\omega} \prod_{i=1}^{N(\tau)} \lambda\left(T_i\right) \phi\left(Z_i|T_i\right) \exp\left(-\int_0^\tau \lambda(t)dt\right) \pi(\xi)\pi(\rho, X_0), \tag{4.14}$$

where $\xi = (\mu, \nu, \gamma)$, $\pi(\xi)$ and $\pi(\rho, X_0)$ are prior densities, and $K_\omega$ is the normalizing constant. The parameters $\mu$, $\nu$, and $\gamma$, are assumed independent and so

$$\pi(\xi) = \pi(\mu)\pi(\nu)\pi(\gamma).$$

On the contrary, $\rho$ and $X_0$ are related through the restriction (4.13). Define

$$X_\rho := \max_{i=1,\ldots,N(\tau)} \left\{0; s(Z_i) - \rho T_i + S_{T_i-}\right\}$$

and let $X_U$ be a suitable upper bound for the stress level (under the assumption of a stationary regime, the stress will not become too large before the occurrence of an event and so one could take, e.g., $X_U = s(Z^*)$, the stress level corresponding to a suitably large magnitude $Z^*$ for the seismic region in question). The support for the density $\pi(X_0|\rho)$ is then $(X_\rho, X_U)$, and the prior of $(\rho, X_0)$ is given by

$$\pi(\rho, X_0) = \frac{1}{K_\rho}\pi(X_0)1_{(X_\rho, X_U)}\pi(\rho),$$

where $K_\rho = \int_{X_\rho}^{X_U} \pi(X_0)dX_0$.

Conditional on $X_0$, maximum likelihood estimates obtained on the basis of (4.7) can be used to obtain the variability ranges of $\mu$, $\nu$, $\rho$, and $\gamma$, by varying $X_0$ along $(0, X_U)$ (which is equivalent to assuming that $\pi(X_0)$ is uniform on $(0, X_U)$), thus enabling the specification of $\pi(\mu)$, $\pi(\nu)$, $\pi(\rho)$, and $\pi(\gamma)$, where the support of the latter three should be positive. To the extent that the variability ranges depend on the data through maximum likelikhood estimates, however,

this procedure is not genuinely Bayesian.

Since the normalizing constant $K_\omega$ in (4.14) cannot be obtained easily, Markov Chain Monte Carlo methods can be used to generate a sample from the posterior distribution, upon which parameter inferences can be based. A Bayes estimate of $(\mu, \nu, \rho, \gamma, X_0)$, for instance, is given by (assuming quadratic loss) the mean of the posterior distribution. We shall not elaborate further on this aspect here or in the case study presented in Section 5.5, but refer to Rotondi and Varini [38] who have followed this approach, and to Holden, Natvig *et al.* [14], for further considerations in Bayesian modelling in the statistical seismology context.

## 4.6 Simulation, prediction, and diagnostic procedures

**A. Simulation.** The specification of the stress release model in terms of an intensity function has the major advantage that the latter can be used as the basis of simulation procedures, which are a key component in evaluating the numerical characteristics of the model and in the important task (in our case) of model-based prediction of insurance losses.

The intuition behind simulation of space-time processes is as follows. The first step is to simulate a constant rate Poisson process over a space by time rectangle. The second step, referred to as *thinning*, involves working through the points of the latter, one by one, and making a comparison between a uniform random variable $U$ on $(0, M)$, where $M$ is some upper bound for the intensity, and the intensity $\lambda(s)$. The point is selected if $U \leq \frac{\lambda(s)}{M}$, otherwise rejected. The main limitation of this method is the requirement that the intensity have an upper bound, which is often not the case for extensions to more complex history-dependent intensities. In such a case, if the process is to be simulated

over a finite time interval $(0, T)$, the bound of the intensity should hold not only over $(0, T)$, but also over all histories of the process up to time $T$. To meet this difficulty, Ogata [29] suggested a sequential variant of the process outlined above, requiring only a local boundedness condition on the intensity. For a survey of the principal approaches to point process simulation and of the theoretical principles on which these are based, we refer to Daley and Vere-Jones [7], from where we extract a variant of Ogata's thinning approach that can be used for simulation of the stress release model we propose. This algorithm is presented in Appendix C.

**B. Prediction.** Suppose we wish to predict some functional $G$ of the process, starting from time $t$ and assuming the model is specified through its intensity. Then the thinning method can be used, with the origin shifted to the current time $t$, to simulate the behaviour of the process in $[t, t+u)$ for any $u > 0$. From the simulation, a value can be obtained for the quantity $G$ of interest. Repeating the simulation, with the same initial history, the empirical distribution of $G$ can then be constructed.

**C. Diagnostic procedures.** Apart from likelihood ratio and associated AIC procedures, which can be applied to point process models much as to other stochastic process models, testing and diagnostic techniques for point process models have been slow to emerge. One of the first effective techniques was developed by Ogata [30] for temporal point processes, and depends on a famous theorem first clearly stated and proved by Papangelou [36]: under the random time change $t \mapsto \Lambda(t)$, where $\Lambda(t) := \int_0^t \lambda(s)ds$ is the compensator of $N$, the transformed process

$$\tilde{N}(t) := N\left(\Lambda^{-1}(t)\right)$$

is a Poisson process with unit rate. Ogata suggests to plot the cumulative frequency of points in transformed time and examine for deviations.

Residuals can be constructed from the fact that, when the model is correctly specified, the *innovation* or *error process*

$$I(t) := N(t) - \Lambda(t)$$

is a martingale, with $\mathbb{E}[I(t)|\mathcal{F}_t] = 0$ (see, e.g., Karr [21]). The adequacy of the fitted model can then be assessed by inspecting whether the residuals

$$R(t) := N(t) - \int_0^t \hat{\lambda}(s)ds,$$

where $\hat{\lambda}$ represents the fitted intensity, are effectively close to zero. Various plots and transformations of $R(t)$ are useful diagnostics for a fitted point process model. For further detail we refer to Baddeley, *et al.* [2] and Schoenberg [39].

# Chapter 5

# Earthquake risk assessment in insurance

In this chapter we address the problem of earthquake risk assessment from the point of view of an insurer. The aim is not to address questions in the financial mathematics and economics of risk selection, pricing, portfolio management, and risk transfer strategies such as reinsurance and securitisation, but to enable the latter through the characterisation of the foundation of any risk management consideration in insurance: the distribution of losses over a period of time for a portfolio of risks. On the actuarial mathematics front, we address the problem of constructing simulation based and, where possible, analytical approximations to the latter.

## 5.1 The aggregate loss distribution

Consider the $t$-portfolio and the aggregate loss in the time interval $(t, u]$

$$X_{t,u} = \int_{(t,u]} \int_{\mathbb{L}} \int_{\mathbb{Z}} \int_{\mathbb{R}_+} rN(d\tau, dl, dz, dr) \tag{5.1}$$

introduced in Section 1.2. The distribution $P^{t,u}(x) = \mathbb{P}[X_{t,u} \leq x]$ does not admit a closed form representation, and so any insight into its probability structure goes hand-in-hand with simulation in all but the special cases when the earthquake process is Poissonian. In the latter case, analytical expressions for the first three central moments of $X_{t,u}$ can be obtained, which allow us to formulate closed-form approximations of $P^{t,u}$. We comment on this approach next, and defer the discussion on simulation to Section 5.4.

## 5.2 Formulæ for selected moments of the aggregate loss under the Poisson and Poisson-gamma earthquake processes

Below we derive analytical expressions for the mean, variance, and third central moment, given $\mathcal{F}_t$, of $X_{t,u}$ under the assumption that the earthquake process is Poisson or Poisson-gamma. We assume that the $t-$portfolio remains fixed throughout $(t, u]$.

**A. Moments under the Poisson model.** Consider the Poisson model defined in Section 3.2 and the moments of $R = \sum_{k=1}^{K_t} Y_k$ defined in Section 2.5. We make the dependence of $m_j^R$ and $\overline{m}_j^R$ on $(l, z)$ explicit and define

$$m_j^R(l) := \int_{z_0}^{\infty} \int_{\mathbb{R}_+} r^j dG_t(r|l, z)\Phi(dz), \quad j = 1, 2, 3. \tag{5.2}$$

The mean, variance, and third central moment measures of the Poisson random measure $N(dt, dl, dz, dr)$, are all equal to the $\mathcal{F}_t-$intensity

$$dt\lambda\, \Delta(dl)\, \Phi(dz)\, dG_t(r|l, z).$$

The first moment of $X_{t,u}$ is

$$\begin{aligned} M_1^{t,u} &:= \mathbb{E}\left[X_{t,u}\middle|\mathcal{F}_t\right] \\ &= \int_{(t,u]}\int_{\mathbb{L}}\int_{z_0}^{\infty}\int_{\mathbb{R}_+} r\mathbb{E}\left[N(d\tau,dl,dz,dr)|\mathcal{F}_t\right] \\ &= \int_{(t,u]}\int_{\mathbb{L}}\int_{z_0}^{\infty}\int_{\mathbb{R}_+} r dG_t(r|l,z)\Phi(dz)\lambda\Delta(dl)\,d\tau \\ &= \lambda(u-t)\int_{\mathbb{L}} m_1^R(l)\Delta(dl). \end{aligned}$$

The variance of $X_{t,u}$ is

$$\begin{aligned} \overline{M}_2^{t,u} &:= \mathbb{V}\left[X_{t,u}\middle|\mathcal{F}_t\right] \\ &= \int_{(t,u]}\int_{\mathbb{L}}\int_{z_0}^{\infty}\int_{\mathbb{R}_+} r^2\mathbb{V}\left[N(d\tau,dl,dz,dr)|\mathcal{F}_t\right] \\ &= \int_{(t,u]}\int_{\mathbb{L}}\int_{z_0}^{\infty}\int_{\mathbb{R}_+} r^2 dG_t(r|l,z)\Phi(dz)\lambda\Delta(dl)\,d\tau \\ &= \lambda(u-t)\int_{\mathbb{L}} m_2^R(l)\Delta(dl). \end{aligned}$$

Finally, letting $\mu^3$ denote the third central moment operator, the third central moment of $X_{t,u}$ is

$$\begin{aligned} \overline{M}_3^{t,u} &:= \mu^3\left[X_{t,u}|\mathcal{F}_t\right] \\ &= \int_{(t,u]}\int_{\mathbb{L}}\int_{z_0}^{\infty}\int_{\mathbb{R}_+} r^3\mu^3\left[N(d\tau,dl,dz,dr)|\mathcal{F}_t\right] \\ &= \int_{(t,u]}\int_{\mathbb{L}}\int_{z_0}^{\infty}\int_{\mathbb{R}_+} r^3 dG_t(r|l,z)\Phi(dz)\lambda\Delta(dl)\,d\tau \\ &= \lambda(u-t)\int_{\mathbb{L}} m_3^R(l)\Delta(dl). \end{aligned}$$

**B. Moments under the Poisson-gamma model.** Consider the Poisson-gamma model defined in Section 3.3. Conditional on $\Lambda$, the random measure $N(dt, dl, dz, dr)$ is Poisson, with mean, variance, and third central moment, given by the (random) intensity measure

$$\Lambda(dt, dl, dz, dr) = dt\Lambda^L(dl)\,\Phi(dz)\,dG_t(r|l, z).$$

*A. The conditional moments.* Consider the first three moments of $X_{t,u}$, conditional on the gamma measure $\Lambda^L$. The first conditional moment is

$$\begin{aligned} M_1(\Lambda^L) &:= \mathbb{E}\left[X_{t,u}|\Lambda^L\right] \\ &= \int_{(t,u]}\int_{\mathbb{L}}\int_{z_0}^{\infty}\int_{\mathbb{R}_+} r\mathbb{E}\left[N(d\tau, dl, dz, dr)|\Lambda^L\right] \\ &= \int_{(t,u]}\int_{\mathbb{L}}\int_{z_0}^{\infty}\int_{\mathbb{R}_+} rdG_t(r|l, z)\Phi(dz)\Lambda^L(dl)d\tau \\ &= (u-t)\int_{\mathbb{L}} m_1^R(l)\Lambda^L(dl). \end{aligned}$$

The second conditional moment is

$$\begin{aligned} M_2(\Lambda^L) &:= \mathbb{E}\left[X_{t,u}^2|\Lambda^L\right] \\ &= \mathbb{V}\left[X_{t,u}|\Lambda^L\right] + M_1^2(\Lambda^L) \\ &= \int_{(t,u]}\int_{\mathbb{L}}\int_{\mathbb{Z}}\int \mathbb{V}\left[rN(d\tau, dl, dz, dr)|\Lambda^L\right] + M_1^2(\Lambda^L) \\ &= (u-t)\int_{\mathbb{L}}\int_{z_0}^{\infty}\int_{\mathbb{R}_+} r^2 dG_t(r|l, z)\Phi(dz)\Lambda^L(dl) + M_1^2(\Lambda^L) \\ &= (u-t)\int_{\mathbb{L}} m_2^R(l)\Lambda^L(dl) + (u-t)^2\left(\int_{\mathbb{L}} m_1^R(l)\Lambda^L(dl)\right)^2. \end{aligned}$$

Finally, the third conditional moment is

$$
\begin{aligned}
M_3(\Lambda^L) &:= \mathbb{E}\left[X_{t,u}^3|\Lambda^L\right] \\
&= \mu^3\left[X_{t,u}|\Lambda^L\right] + 3\mathbb{V}\left[X_{t,u}|\Lambda^L\right] M_1(\Lambda^L) + M_1^3(\Lambda^L), \qquad (5.3)
\end{aligned}
$$

where the third central moment of $X_{t,u}$, given $\Lambda^L$, is

$$
\begin{aligned}
\mu^3\left[X_{t,u}|\Lambda^L\right] &= \int_{(t,u]}\int_{\mathbb{L}}\int_{z_0}^{\infty}\int_{\mathbb{R}_+} r^3\mu^3[N(d\tau,dl,dz,dr)|\Lambda^L] \\
&= \int_{(t,u]}\int_{\mathbb{L}}\int_{z_0}^{\infty}\int_{\mathbb{R}_+} r^3 dG_t(r|l,z)\Phi(dz)\Lambda^L(dl)d\tau \\
&= (u-t)\int_{\mathbb{L}} m_3^R(l)\Lambda^L(dl),
\end{aligned}
$$

allowing (5.3) to be recast as

$$
\begin{aligned}
M_3(\Lambda^L) &= (u-t)\int_{\mathbb{L}} m_3^R(l)\Lambda^L(dl) \\
&\quad + 3(u-t)^2\int_{\mathbb{L}} m_2^R(l)\Lambda^L(dl)\int_{\mathbb{L}} m_1^R(l)\Lambda^L(dl) \\
&\quad + (u-t)^3\left(\int_{\mathbb{L}} m_1^R(l)\Lambda^L(dl)\right)^3.
\end{aligned}
$$

*B. The unconditional moments.* The goal is now to obtain analytical expressions for the moments of $X_{t,u}$, given $\mathcal{F}_t$. The first moment is

$$
\begin{aligned}
M_1^{t,u} &:= \mathbb{E}\left[X_{t,u}|\mathcal{F}_t\right] \\
&= \mathbb{E}[\mathbb{E}\left[X_{t,u}|\Lambda^L]|\mathcal{F}_t\right] \\
&= \mathbb{E}[M_1(\Lambda^L)|\mathcal{F}_t] \\
&= (u-t)\int_{\mathbb{L}} m_1^R(l)\mathbb{E}[\Lambda^L(dl)|\mathcal{F}_t] \\
&= (u-t)\int_{\mathbb{L}} m_1^R(l)\frac{\alpha_t(dl)}{\beta_t},
\end{aligned}
$$

where $\alpha_t(dl) := \alpha(dl) + N(t, dl, (z_0, \infty))$ and $\beta_t := \beta + N(t)$ are the shape and inverse scale parameters of the posterior gamma measure $\Lambda^L$.

The second moment is

$$
\begin{aligned}
M_2^{t,u} &:= \mathbb{E}[X_{t,u}^2|\mathcal{F}_t] \\
&= \mathbb{E}[\mathbb{E}[X_{t,u}^2|\Lambda^L]|\mathcal{F}_t] \\
&= \mathbb{E}[M_2(\Lambda^L)|\mathcal{F}_t] \\
&= (u-t)\int_{\mathbb{L}} m_2^R(l)\mathbb{E}[\Lambda^L(dl)|\mathcal{F}_t] \\
&\quad + (u-t)^2\mathbb{E}\left[\left(\int_{\mathbb{L}} m_1^R(l)\Lambda^L(dl)\right)^2\middle|\mathcal{F}_t\right]. \qquad (5.4)
\end{aligned}
$$

To obtain a workable explicit expression for the expectation in (5.4), note that

$$
\begin{aligned}
\mathbb{E}\left[\left(\int_{\mathbb{L}} m_1^R(l)\Lambda^L(dl)\right)^2\middle|\mathcal{F}_t\right] &= \mathbb{V}\left[\int_{\mathbb{L}} m_1^R(l)\Lambda^L(dl)\middle|\mathcal{F}_t\right] \\
&\quad + \left(\mathbb{E}\left[\int_{\mathbb{L}} m_1^R(l)\Lambda^L(dl)\middle|\mathcal{F}_t\right]\right)^2 \\
&= \int_{\mathbb{L}} m_1^{R^2}(l)\frac{\alpha_t(dl)}{\beta_t^2} + \left(\int_{\mathbb{L}} m_1^R(l)\frac{\alpha_t(dl)}{\beta_t}\right)^2,
\end{aligned}
$$

result which follows from the fact that the variance of the posterior gamma measure $\Lambda^L$ is $\frac{\alpha_t(\cdot)}{\beta_t^2}$, allowing us to recast the second moment as

$$\begin{aligned} M_2^{t,u} &= (u-t)\int_{\mathbb{L}} m_2^R(l)\frac{\alpha_t(dl)}{\beta_t} \\ &\quad +(u-t)^2\left[\int_{\mathbb{L}} {m_1^R}^2(l)\frac{\alpha_t(dl)}{\beta_t^2} + \left(\int_{\mathbb{L}} m_1^R(l)\frac{\alpha_t(dl)}{\beta_t}\right)^2\right]. \end{aligned}$$

The third moment is

$$\begin{aligned} M_3^{t,u} &:= \mathbb{E}\left[X_{t,u}^3|\mathcal{F}_t\right] \\ &= \mathbb{E}[\mathbb{E}\left[X_{t,u}^3|\Lambda^L\right]|\mathcal{F}_t] \\ &= \mathbb{E}\left[M_3(\Lambda^L)|\mathcal{F}_t\right]. \end{aligned}$$

This requires the evaluation of the expectations

$$\mathbb{E}\left[\int_{\mathbb{L}} m_2^R(l)\Lambda^L(dl)\int_{\mathbb{L}} m_1^R(l)\Lambda^L(dl)\middle|\mathcal{F}_t\right] \tag{5.5}$$

and

$$\mathbb{E}\left[\left(\int_{\mathbb{L}} m_1^R(l)\Lambda^L(dl)\right)^3\middle|\mathcal{F}_t\right]. \tag{5.6}$$

Let $m_i = \int_{\mathbb{L}} m_i^R(l)\Lambda^L(dl)$, $i = 1, 2, 3$. Then (5.5) can be recast as

$$\begin{aligned}\mathbb{E}\left[m_2\, m_1|\mathcal{F}_t\right] &= \int_l\int_{l'} m_2^R(l)m_1^R(l')\mathbb{E}[\Lambda^L(dl)\Lambda^L(dl')|\mathcal{F}_t] \\ &= \int_l\int_{l'} m_2^R(l)m_1^R(l')\mathbb{E}[\Lambda^L(dl)|\mathcal{F}_t]\mathbb{E}[\Lambda^L(dl')|\mathcal{F}_t] \qquad (5.7)\\ &\quad + \int_l m_2^R(l)m_1^R(l)\mathbb{V}[\Lambda^L(dl)|\mathcal{F}_t], \qquad (5.8)\end{aligned}$$

a consequence of the fact that

$$\mathbb{E}[\Lambda^L(dl)\Lambda^L(dl')|\mathcal{F}_t] = \mathbb{E}[\Lambda^L(dl)|\mathcal{F}_t]\mathbb{E}[\Lambda^L(dl')|\mathcal{F}_t] + \mathbb{V}[\Lambda^L(dl)|\mathcal{F}_t]\delta_{\{l,l'\}},$$

where $\delta$ is Kronecker's delta. If $dl \cap dl' = \emptyset$, $\Lambda^L(dl)$ and $\Lambda^L(dl')$ are independent, and the double integral in (5.7) can be recast as

$$\int_{\mathbb{L}} m_2^R(l)\frac{\alpha_t(dl)}{\beta_t}\int_{\mathbb{L}} m_1^R(l)\frac{\alpha_t(dl)}{\beta_t}.$$

The variance of the posterior gamma measure $\Lambda^L$ is $\frac{\alpha_t(\cdot)}{\beta_t^2}$, and so the integral in (5.8) is

$$\int_{\mathbb{L}} m_2^R(l)m_1^R(l)\frac{\alpha_t(dl)}{\beta_t^2}.$$

The third central moment of the posterior gamma measure $\Lambda^L$ is $\frac{2\alpha_t(\cdot)}{\beta_t^3}$, in consequence, recalling that $\mu^3$ denotes the third central moment operator,

$$\begin{aligned}\mu^3\left[m_1\middle|\mathcal{F}_t\right] &= \int_{\mathbb{L}} {m_1^R}^3(l)\mu^3\left[\Lambda^L(dl)\middle|\mathcal{F}_t\right] \\ &= \int_{\mathbb{L}} {m_1^R}^3(l)\frac{2\alpha_t(dl)}{\beta_t^3}.\end{aligned}$$

Finally, the identity

$$\mathbb{E}[m_1^3|\mathcal{F}_t] \quad = \quad \mu^3[m_1|\mathcal{F}_t] + 3\mathbb{V}[m_1|\mathcal{F}_t]\mathbb{E}[m_1|\mathcal{F}_t] + \mathbb{E}[m_1|\mathcal{F}_t]^3$$

allows us to evaluate (5.6) and to recast the expression for $M_3^{t,u}$ as

$$\begin{aligned} M_3^{t,u} \quad = \quad & (u-t)\mathbb{E}\left[m_3\middle|\mathcal{F}_t\right] + 3(u-t)^2\mathbb{E}\left[m_2\middle|\mathcal{F}_t\right]\mathbb{E}\left[m_1\middle|\mathcal{F}_t\right] \\ & + 3(u-t)^2\int_{\mathbb{L}} m_2^R(l)m_1^R(l)\frac{\alpha_t(dl)}{\beta_t^2} + (u-t)^3\mathbb{E}[m_1^3|\mathcal{F}_t]. \quad (5.9) \end{aligned}$$

The variance and third central moment of $X_{t,u}$, given $\mathcal{F}_t$, follow readily from the non-central moments. The former is

$$\overline{M}_2^{t,u} = M_2^{t,u} - {M_1^{t,u}}^2,$$

and the latter is

$$\overline{M}_3^{t,u} = M_3^{t,u} - 3\overline{M}_2^{t,u}M_1^{t,u} - {M_1^{t,u}}^3.$$

## 5.3 Analytical approximations to the aggregate loss distribution

There exist various approximations of the aggregate loss distribution in classical risk theory (see e.g. Daykin *et al.* [9]). The most well-known ones are the normal approximation and its refinements (the normal power (NP) approximation and the Edgeworth expansions), the gamma approximation and the Esscher approximation. Chaubey *et al.* [6] have introduced the inverse Gaussian (IG) and the gamma-IG approximation. The underlying principle is to approximate

the aggregate loss distribution by a function that uses the mean, variance, coefficient of skewness, and for some approximations (Edgeworth and gamma-IG) the coefficient of kurtosis of the aggregate loss.

Next, we give an overview of some approximations of the distribution $P^{t,u}$ of $X_{t,u}$ that can be constructed on the basis of the mean, variance, and third central moment of $X_{t,u}$. Throughout, we work with the coefficient of skewness defined by

$$\gamma^{t,u} := \frac{\overline{M}_3^{t,u}}{\overline{M}_2^{t,u\,\frac{3}{2}}}.$$

**A. The NP approximation.** The NP approximation extends the normal approximation with a correction for the positive skewness of the aggregate loss. Under the NP approximation,

$$P^{t,u}(x) \approx \mathbf{\Phi}\left(\sqrt{\frac{9}{\gamma^{t,u\,2}} + \frac{6\dot{x}}{\gamma^{t,u}} + 1} - \frac{3}{\gamma^{t,u}}\right), \tag{5.10}$$

where $\dot{x} := \frac{x - M_1^{t,u}}{\overline{M}_2^{t,u\,\frac{1}{2}}}$, and $\mathbf{\Phi}$ is the standard normal distribution function.

**B. The gamma approximation.** Following a translation of $x_0$ units from the origin, three moments can be fitted. Under the gamma approximation,

$$P^{t,u}(x) \approx \int_{x_0}^{x} \frac{\beta^\alpha}{\Gamma(\alpha)}(s - x_0)^{\alpha-1} e^{-\beta(s-x_0)}\, ds, \tag{5.11}$$

where $\Gamma(\cdot)$ is the gamma function, $\alpha = \left(\frac{2}{\gamma^{t,u}}\right)^2$, $\beta = \frac{2}{\gamma^{t,u}\overline{M}_2^{t,u\,\frac{1}{2}}}$, and $x_0 = M_1^{t,u} - \frac{2\overline{M}_2^{t,u\,\frac{1}{2}}}{\gamma^{t,u}}$.

**C. The IG approximation.** As in the case of the gamma approximation, three moments can be fitted following a translation of $x_0$ units from the origin. Under the IG approximation,

$$P^{t,u}(x) \approx \int_{x_0}^{x} \frac{\alpha}{\sqrt{2\pi\beta(s-x_0)^3}} \exp\left(-\frac{(\alpha-\beta(s-x_0))^2}{2\beta(s-x_0)}\right) ds, \tag{5.12}$$

where $\alpha = \left(\frac{3}{\gamma^{t,u}}\right)^2$, $\beta = \frac{3}{\gamma^{t,u}\overline{M}_2^{t,u\,\frac{1}{2}}}$, and $x_0 = M_1^{t,u} - \frac{3\overline{M}_2^{t,u\,\frac{1}{2}}}{\gamma^{t,u}}$.

## 5.4 Simulation of the loss process

The loss process, which can be defined in a number of ways and can be seen as a functional of a finite segment of the future of the earthquake process, rarely falls into any general category for which analytic expressions are available. Since, on the other hand, simulation of the earthquake process is relatively straightforward once its intensity function is specified, prediction of the loss process goes hand-in-hand with simulation.

Suppose that our aim is to predict a particular quantity $X_{t,u}$ that can be represented as a functional of the future of the earthquake process over $(t,u]$. So far we have considered $X_{t,u}$ to be the aggregate loss for the $t-$portfolio over $(t,u]$, defined as in (5.1) or by the equivalent representation

$$X_{t,u} := \sum_{i=1}^{N((t,u])} \sum_{k=1}^{K_t} Y_{k,i}. \tag{5.13}$$

This need not be the case, however, and various considerations in portfolio management in insurance will call for alternative specifications. The maximum $t-$portfolio loss occurring from a single event over the period $(t,u]$

$$X_{t,u} = \max_{i\in\{1,\ldots,N((t,u])\}} \sum_{k=1}^{K_t} Y_{k,i}, \tag{5.14}$$

for instance, might be of interest, as might be the $n$-th and subsequent event losses over $(t,u]$

$$X_{t,u} = \sum_{i=n}^{N((t,u])} \sum_{k=1}^{K_t} Y_{k,i}, \tag{5.15}$$

where $X_{t,u} := 0$ if $N((t,u]) < n$.

The aim is to estimate the distribution of $X_{t,u}$. An outline of prediction procedures under each of the proposed earthquake processes follows.

**A. Simulation under the Poisson model.** We simulate earthquake activity and resulting losses over the period $(t,u]$, $N_{\text{sim}}$ times, under the assumption that the earthquake process is Poisson, as defined in Section 3.2. For each $q \in 1, ..., N_{\text{sim}}$, we have a simulated loss $x_q$, calculated on the basis of (5.13). For alternative specifications of the loss, step 9 below must be modified accordingly.

1. Set $q = 1$.

2. Simulate $t_1, t_2, ..., t_i$ according to a Poisson process with rate $\hat{\lambda}$. If there are no points in the simulated process and $q \leq N_{\text{sim}}$, set $x_q = 0$, $q = q+1$, and repeat this step. If $q > N_{\text{sim}}$, go to step 12. Otherwise go to step 3.

3. Simulate the locations $l_1, l_2, ..., l_i$ as a set of i.i.d random variables from the empirical distribution $\hat{\Delta}_{N(t)}$. Simulation from the empirical distribution of spatial locations is considered in Appendix C, Section C.1.

4. Simulate the magnitudes $z_1, z_2, ..., z_i$ as a set of i.i.d random variables following the Gutenberg-Richter law (see (3.3)).

5. Set $j = 1$.

6. Set $k = 1$.

7. Simulate $y_k$ as a random variable from the distribution $G(y|\epsilon_k, c_k)$. Set $k = k + 1$.

8. If $k \leq K_t$, return to step 7.

9. Set $r_j = \sum_k y_k$. Set $j = j + 1$.

10. If $j \leq i$, return to step 6.

11. Set $x_q = \sum_j r_j$ and $q = q + 1$. If $q \leq N_{\text{sim}}$, return to step 2.

12. The output consists of the empirical distribution of $\{x_1, \ldots, x_{N_{\text{sim}}}\}$ and key characteristics of the latter, such as its mean, standard deviation, and selected quantiles.

**B. Simulation under the Poisson-gamma model.** We simulate earthquake activity and resulting losses over the period $(t, u]$, $N_{\text{sim}}$ times, under the assumption that the earthquake process is Poisson-gamma, as defined in Section 3.3. For each $q \in 1, ..., N_{\text{sim}}$, we have a simulated loss $x_q$, calculated on the basis of (5.13). For alternative specifications of the loss, step 10 below must be modified accordingly.

1. Set $q = 1$.

2. Simulate $\lambda$ as a random variable from the distribution

$$\text{Ga}\left(\alpha(\mathbb{L}) + N(t), \beta + t\right).$$

   This is done by setting $\lambda = \sum_p \lambda_p$, where $\{\lambda_p\}$ is the output of the first step of the algorithm described in Appendix C, Section C.2.

3. Simulate $t_1, t_2, ..., t_i$ according to a Poisson process with rate $\lambda$. If there are no points in the simulated process and $q \leq N_{\text{sim}}$, set $x_q = 0$, $q = q+1$, and return to step 2. If $q > N_{\text{sim}}$, go to step 13. Otherwise go to step 4.

4. Simulate the locations $l_1, l_2, ..., l_i$ as a set of i.i.d random variables from the Bayes estimate of the Dirichlet process defined in (3.13). Simulation

from the latter is considered in Appendix C, Section C.2.

5. Simulate the magnitudes $z_1, z_2, ..., z_i$ as a set of i.i.d random variables following the Gutenberg-Richter law (see (3.3)).

6. Set $j = 1$.

7. Set $k = 1$.

8. Simulate $y_k$ as a variate from the distribution $G(y|\epsilon_k, c_k)$. Set $k = k + 1$.

9. If $k \leq K_t$, return to step 8.

10. Set $r_j = \sum_k y_k$. Set $j = j + 1$.

11. If $j \leq i$, return to step 7.

12. Set $x_q = \sum_j r_j$ and $q = q + 1$. If $q \leq N_{\text{sim}}$, return to step 3.

13. The output consists of the empirical distribution of $\{x_1, \ldots, x_{N_{\text{sim}}}\}$ and key characteristics of the latter, such as its mean, standard deviation, and selected quantiles.

**C. Simulation under the stress release model.** We simulate earthquake activity and resulting losses over the period $(t, u]$, $N_{\text{sim}}$ times, under the stress release model of the earthquake process introduced in Chapter 4. For each $q \in 1, ..., N_{\text{sim}}$, we have a simulated loss $x_q$, calculated on the basis of (5.13). For alternative specifications of the loss, step 7 below must be modified accordingly.

1. Set $q = 1$.

2. Simulate a realization of the earthquake process over $(t, u]$ following the thinning algorithm considered in Appendix C, Section C.3. If there are no points in the simulated process and $q \leq N_{\text{sim}}$, set $x_q = 0$, $q = q + 1$, and return to step 2. If $q > N_{\text{sim}}$, go to step 10. Otherwise go to step 3.

3. Set $j = 1$.

4. Set $k = 1$.

5. Simulate $y_k$ as a variate from the distribution $G(y|\epsilon(l_j, z_j, c_k), c_k)$. Set $k = k + 1$.

6. If $k \leq K_t$, return to step 5.

7. Set $r_j = \sum_k y_k$. Set $j = j + 1$.

8. If $j \leq i$, return to step 4.

9. Set $x_q = \sum_j r_j$ and $q = q + 1$. If $q \leq N_{\text{sim}}$, return to step 2.

10. The output consists of the empirical distribution of $\{x_1, \ldots, x_{N_{\text{sim}}}\}$ and key characteristics of the latter, such as its mean, standard deviation, and selected quantiles.

## 5.5 Case study

In this section we illustrate the applicability of the proposed earthquake risk assessment methodology through a case study involving the analysis of a subset of an insurer's residential homeowners book of business. Company XYZ is a large personal lines insurer member of the California Earthquake Authority (CEA), operating in a number of major lines of insurance, including auto, property, life, and commercial. The business under analysis in this example is the residential homeowners property catastrophe business exposed to earthquake risk in Southern California. Although a typical earthquake policy offered by XYZ insures for loss against structural damage, damage to contents, and loss of use, this example will consider losses arising from structural damage only.

**A. The loss records.** Table D.1 in Appendix D shows a sample of XYZ's California residential homeowners property portfolio in force at the time of the 1994 Northridge earthquake, along with the recorded losses resulting from the latter. Each of the loss records provided by XYZ includes a location identifier, the address of the structure in question, which we do not disclose, except for the city, its value, and the loss, if any. No information on the policies' financial structures (such as deductibles and limits) was provided and so we assume that the value entry refers to the property's value, not the sum insured, and that the loss entry refers to structural loss and not the actual claim that has been lodged against XYZ. The geocoding details for each structure have been retrieved on the basis of its address and comprise the structure's location in terms of latitude and longitude, the local soil type, and geological conditions such as the local soil's proneness to liquefaction and landslide. Figures D.1, D.2, and D.3, in Appendix D, display maps for California showing these characteristics.

There are a total of 211 records in the Northridge earthquake reference loss experience provided by XYZ, out of which 109 report no loss. The remaining 102 policies covered structures for which a loss occurred. Figure 5.1 shows the geographical distribution of the structures, along with their loss ratio, defined as the ratio of loss to value. It is not surprising that the largest concentration of high loss ratios correspond to structures in or around the Los Angeles area, with close proximity to the location of the Northridge earthquake epicentre. The geographic distribution of the available structures is suitably disperse, thus providing observations of $\epsilon(l, z, c)$ over a wide range of energy values. Furthermore, in the model considered, with conditional independence between risks and no time dependence on conditional individual loss distributions, the conditional distribution can be estimated from losses generated by the same earthquake or different earthquakes.

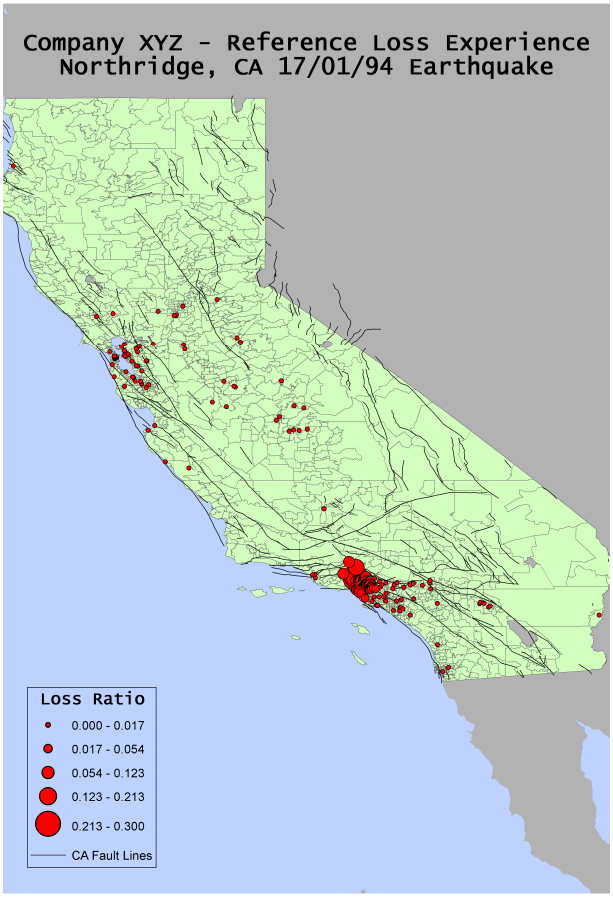


Figure 5.1: *Northridge earthquake reference loss experience for Company XYZ.*

**B. Modeling of the conditional loss distribution.** The conditional loss distribution $G(y|\epsilon, c)$ was introduced in Chapter 2. We first focus on the modeling of the singularity of $G$ at 0, and then proceed to model its continuous component. Observe that the largest loss ratio recorded was 0.3001, and so there is no data to support the modeling of a singularity at the value of the building. This, however, might not be the case if the policies' financial structures were to be modelled.

Consider the random variable

$$Y^0 \sim \text{Ber}(p)$$

| Conditional Loss Distribution<br>Logistic regression model of the singularity at zero | | | | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Std. Error* | *z-stat* | *p-value* |
| $\alpha^0$ | $-3.33$ | 0.48 | $-6.95$ | $< 0.001$ |
| $\beta^0$ | $8.12 \times 10^{-5}$ | $1.22 \times 10^{-5}$ | 6.64 | $< 0.001$ |
| Null deviance ($D_0$) | 292.28 on 210 degrees of freedom | | | |
| Residual deviance ($D_1$) | 75.47 on 209 degrees of freedom | | | |
| $\mathbb{P}[D_0 - D_1 > 216.80] < 0.001$ | | | | |
| AIC | 79.47 | | | |

Table 5.1: *Estimation results for the conditional loss distribution. Logistic regression model of the singularity at zero.*

and the GLM

$$\eta(p) = \alpha^0 + \beta^0 \epsilon + \gamma^0 c,$$

defined in Section 2.3, paragraph A. In this example, $c = (ST, LIQ, LL)$, where $ST$ denotes soil type, $LIQ$ denotes the soil's proneness to liquefaction, and $LL$ denotes the soil's proneness to landslide[1]. Various GLM configurations, comprising different covariate arrangements and link functions were investigated and the best, in terms of parameter significance, likelihood ratio test, and AIC, was found to be

$$\eta(p) = \alpha^0 + \beta^0(\epsilon * ST * LIQ), \tag{5.16}$$

where * denotes the product operator and $\eta$ the logit link. The model estimation results, obtained using the `glm()` function in `R`, are summarised in Table 5.1.

To model the continuous component $\overline{G}(y|\epsilon, c)$ of $G$, we considered the beta

[1] $ST$, $LIQ$, and $LL$ are all dimensionless quantitities. $ST$ ranges from 1 - 4, with 1 representing rock and 4 representing soft soil; $LIQ$ ranges from 1-5, with 1 representing very low and 5 representing very high proneness to liquefaction; and $LL$ ranges from 1-5, with 1 representing very low and 5 representing very high proneness to landslide. For further detail please refer to figures D.1, D.2, and D.3, in Appendix D, where maps for California showing these characteristics are displayed.

| Conditional Loss Distribution<br>Beta regression model of continuous component | | | | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Std. Error* | *z-stat* | *p-value* |
| $\alpha$ | $-2.5698$ | 0.1726 | $-14.89$ | $< 0.001$ |
| $\beta$ | $3.7645 \times 10^{-6}$ | $4.048 \times 10^{-7}$ | 9.30 | $< 0.001$ |
| $\gamma$ | $-0.1497$ | $3.883 \times 10^{-2}$ | $-3.86$ | $< 0.001$ |
| $\varphi$ | 35.4585 | 5.1233 | | |
| Null deviance ($D_0$) | 210.85 on 101 degrees of freedom | | | |
| Residual deviance ($D_1$) | 67.79 on 99 degrees of freedom | | | |
| $\mathbb{P}[D_0 - D_1 > 143.05] < 0.001$ | | | | |
| AIC | -409.73 | | | |
| Pseudo-$R^2$ | 0.49 | | | |

Table 5.2: *Estimation results for the conditional loss distribution. Beta regression model of the continuous component.*

regression model introduced in Section 2.4, paragraph C, whereby

$$Y^{(0,v)} \sim \text{Be}\left(\alpha'(\epsilon, c), \beta'(\epsilon, c)\right),$$

and

$$\eta(\mu) = \alpha + \beta\epsilon + \gamma c,$$

where $\mu = \mathbb{E}[Y^{(0,v)}]$. As before, we assume $c = (ST, LIQ, LL)$. The best model configuration, in terms of parameter significance, likelihood ratio test, AIC, and pseudo-$R^2$ (the sample correlation coefficient between $\eta(\hat{\mu}_i)$ and $\eta(u_i)$, with $\hat{\mu}_i$ and $u_i$ the fitted and observed loss ratios, respectively), was found to be

$$\eta(\mu) = \alpha + \beta(\epsilon * ST * LIQ) + \gamma(ST * LIQ), \tag{5.17}$$

with $\eta$ the logit link. The model estimation results, obtained using function `betareg()` in package `betareg` in `R`, are summarised in Table 5.2.

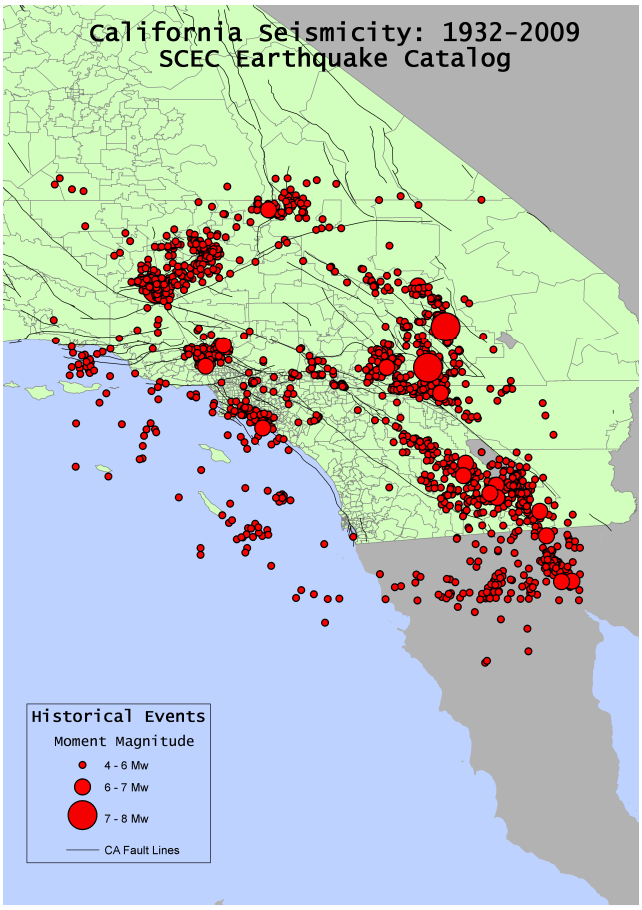


Figure 5.2: *Southern California seismicity from 1932 - 2009. Events with magnitude greater than 4.*

Note that models (5.16) and (5.17) are non-hierarchical in the sense that they include interactions without the correponding main effects and lower-order interactions.

**C. Modeling of the earthquake process.** For the purpose of this illustration we have used the earthquake catalog compiled by the Southern California Earthquake Center (SCEC), comprising earthquakes occurred over the time span 1 January 1932 - 22 February 2009[2], limited to the region $30 - 36^\circ$ N,

[2]Generally speaking, historical consistency in event detection and measurement requires that only a limited amount of years' worth of data be used. For a brief summary of the various 'eras' in the SCEC catalog please see `http://www.data.scec.org/catalog_search/date_mag_loc.php`. Musmeci and Vere-

$115 - 120°$ W, and with magnitude greater than 4. There are a total of 1740 such events, shown in Figure 5.2.

The estimated annual rate rate of occurrence is

$$\hat{\lambda} = \frac{1740}{77.1452} = 22.55,$$

and maximum likelihood estimation of $\gamma$ in the exponential density

$$\phi(z) = \gamma \exp\left(-\gamma(z-4)\right), \quad z > 4$$

results in the estimate $\hat{\gamma} = 2.30$ (with standard error = 0.055). The boundaries of the earthquake location space

$$\mathbb{L} = [-2641, -2256] \times [-4897, 4442] \times [3313, 3743]$$

were obtained upon conversion of the spherical coordinates in the earthquake catalog into Cartesian coordinates (see Section 2.2, paragraph C.). The space $\mathbb{L}$ was divided into $602,602$ grid cubes, each corresponding to a volume of 125 $\text{km}^3$. Simulation of the empirical distribution of earthquake location (in the case of the Poisson and stress release models) and of the Dirichlet process (in the case of the Poisson-gamma model) was performed on the basis of this grid, as was the computation of the analytical expressions of the moments of $X_{t,u}$ formulated in Section 5.2.

In the case of the Poisson-gamma model, our baseline results are under the assumption of a total prior mass $\alpha(\mathbb{L}) = 0.5$ placed uniformly across the grid cubes, with the inverse-scale parameter of the gamma measure $\beta$ set at 0.02217. The uniformity assumption corresponds to a non-informative Dirichlet prior

Jones [27] point out that in such circumstances modelling is clearly more of an art than a science. However, the historical catalogues provide a unique, and in most cases the only, record of regional earthquake activity over periods of time of the order of hundreds of years.

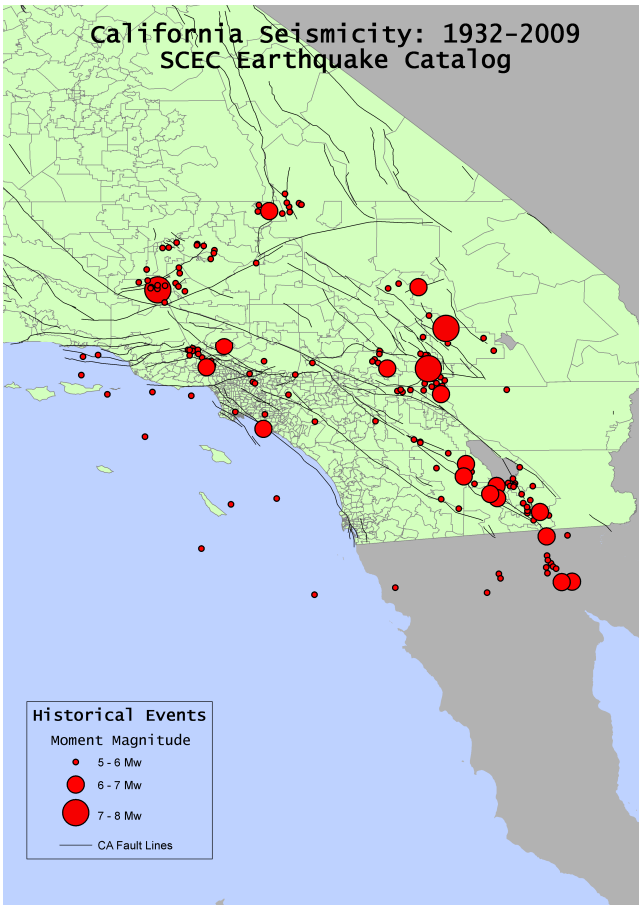


Figure 5.3: *Southern California seismicity from 1932 - 2009. Events with magnitude greater than 5.*

specification, while the prior $\alpha$ mass is arbitrarily set to 0.5 (which, relative to the 1740 observed events, is negligible) to give virtually all of the weight to the empirical distribution function in the Bayes estimate of the Dirichlet process (see 3.13).

For the implementation of the stress release model we set $z_0 = 5$. There were a total of 171 events with magnitude 5 or greater occurring in the time span and region under consideration, shown in Figure 5.3. The intensity function

$$\lambda(t) = \exp\left(a + b\left[t - cS_{t^-}\right]\right),$$

| Stress Release Model | | |
|---|---|---|
| | Intensity function | |
| *Parameter* | *Estimate* | *Std. Error* |
| $a$ | -4.9975 | $07.5 \times 10^{-2}$ |
| $b$ | $1.7911 \times 10^{-5}$ | $3.6046 \times 10^{-10}$ |
| $c$ | 54.18 | 0.2944 |
| | Magnitude distribution | |
| *Parameter* | *Estimate* | *Std. Error* |
| $\gamma$ | 2.0574 | 0.1721 |
| Lower bound for $X_0$ ($X_0^*$) | 195.5118 | |

Table 5.3: *Maximum likelihood estimation results for the intensity function and magnitude distribution of the stress release model of earthquake occurrence in Southern California.*

and the magnitude distribution

$$\phi(z|t) = \frac{\gamma \exp\left(-\gamma(z-5)\right)}{1 - \exp\left(-\gamma\left(z(t)-5\right)\right)}, \;\; 5 < z \leq z(t),$$

were estimated following the maximum likelihood procedure set forth in Section 4.5, paragraph A. The parameter estimates and standard errors are shown in Table 5.3, and the fitted intensity function is shown in Figure 5.4.

The goodness of fit of the model was assessed through residual analysis. The random time change theorem introduced in Section 4.6, paragraph C, suggests that, if the compensator used for the random time change

$$t_i \mapsto \Lambda(t_i) = \int_0^{t_i} \lambda(t)dt$$

is that of the true model, then the transformed process will be unit-rate Poisson, whereas if the wrong compensator is used, the transformed process will show some systematic departure from the unit-rate Poisson process. Figure 5.5 shows a plot of the cumulative number of events versus the transformed times

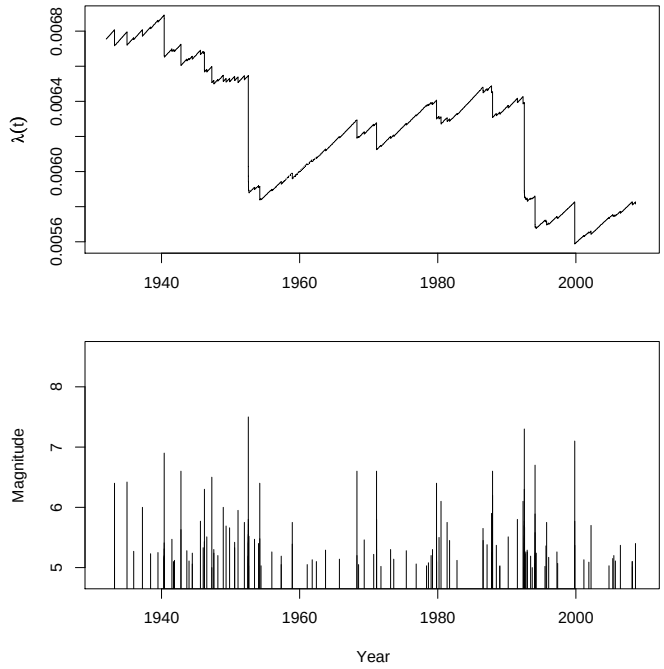


Figure 5.4: *Fitted intensity for the stress release model and historical seismicity in Southern California (1932 - 2009).*

$$t_i \mapsto \hat{\Lambda}(t_i) = \int_0^{t_i} \hat{\lambda}(t)dt, \;\; i = 1, \ldots, 171,$$

where no systematic departure from the unit-rate Poisson assumption is evident, suggesting that the intensity function $\lambda(t)$ has been specified correctly. This is further supported by the structure of the residuals

$$R(t_i) := N(t_i) - \int_0^{t_i} \hat{\lambda}(t)dt, \quad i = 1, \cdots, 171,$$

shown in Figure 5.6.

**D. Estimation of the aggregate loss distribution.** The reference exposure provided by XYZ is a subset of its Southern California residential lines portfolio in force on 1 March 2009, an excerpt of which is shown in Table D.2in Appendix D. It comprises policies covering 83 distinct locations with a com-

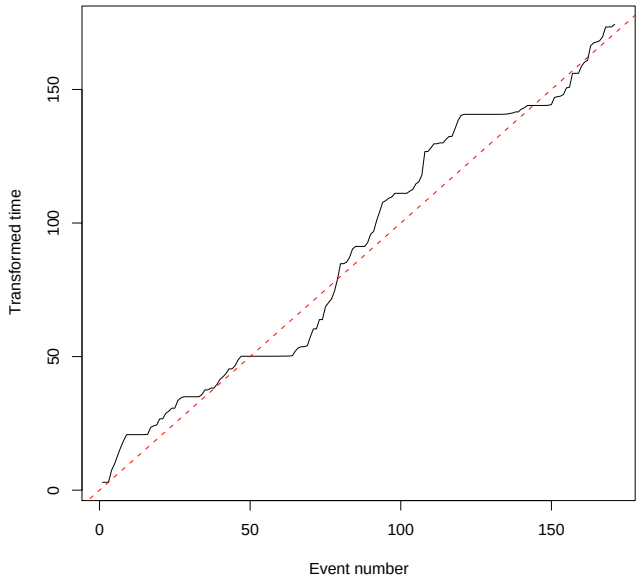


Figure 5.5: *Residual analysis of the stress release model. Cumulative number of events versus the transformed times* $t_i \mapsto \hat{\Lambda}(t_i)$.

bined value of 615, 704, 000 USD. We assumed this to be the $t$-portfolio, where we set $t$ equal to 00:00 hours on 1 March 2009, and $K_t = 83$.

We set $u$ equal to 23:59 on 28 February 2010 and the distribution of the aggregate loss

$$X_{t,u} = \sum_{i=1}^{N((t,u])} \sum_{k=1}^{K_t} Y_{k,i} = \int_{(t,u]} \int_{\mathbb{L}} \int_{z_0}^{\infty} \int_{\mathbb{R}_+} r N(d\tau, dl, dz, dr)$$

was estimated on the basis of simulation of 10,000 $(t, u]$ periods of earthquake activity in Southern California, in accordance with the simulation algorithms for the Poisson, Poisson-gamma, and stress release models, introduced in Section 5.4. We also calculated the mean, variance, and third central moment of $X_{t,u}$ under the Poisson and Poisson-gamma models, in accordance with the formulæ derived in Section 5.2. Our findings are summarised in Table 5.4, where we show selected quantiles and moments of the distribution $P^{t,u}$ of $X_{t,u}$. The *exceedance*

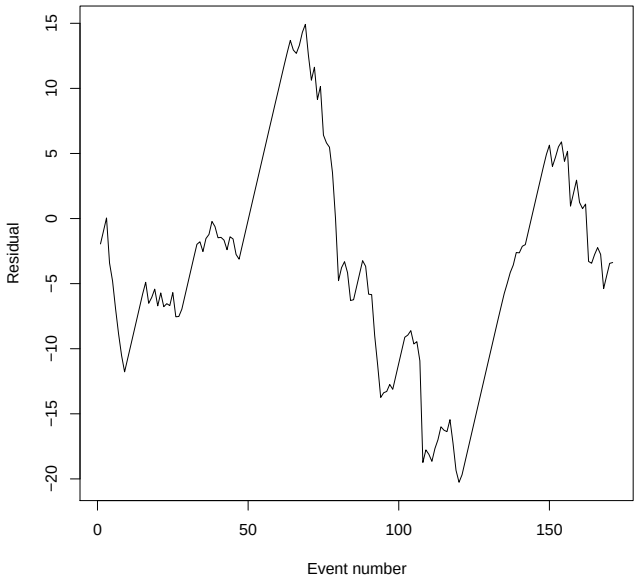


Figure 5.6: *Residual analysis of the stress release model. Events versus residuals.*

*probability* refers to $1 - P^{t,u}(x)$ and so, for example, under the Poisson model,

$$1 - P^{t,u}(46.69) = 0.005.$$

The tail of the simulated distribution for each of the different models is shown in Figure 5.7, and the various analytical approximations to the latter are shown in figures 5.12, 5.13, and 5.14 (for the stress release model, the approximations are calculated on the basis of simulated moments).

**E. Differences in modeled results.** The differences in the estimated losses under the Poisson, Poisson-gamma, and stress release models, are seemingly moderate. We shall hypothesize as to the reasons behind the following conspicuous features of the modeled results (we refer to Table 5.4):

1. Estimated coefficient of skewness of $X_{t,u}$ under the Poisson-gamma model, which is the largest, followed by its counterparts under the Poisson and

| Southern California Residential Portfolio<br>Distribution of the aggregate loss<br>(01 March 2009 - 28 February 2010) | | | |
|---|---|---|---|
| | Stress release | Poisson | Poisson-gamma |
| *Exceedance probability* | *Loss (USD million)* | | |
| 0.05 | 8.38 | 7.77 | 7.53 |
| 0.02 | 19.43 | 17.12 | 14.70 |
| 0.01 | 35.16 | 32.58 | 23.78 |
| 0.005 | 57.39 | 46.69 | 48.97 |
| 0.004 | 66.27 | 53.14 | 53.59 |
| 0.002 | 79.82 | 73.51 | 79.12 |
| 0.001 | 96.20 | 86.51 | 95.00 |
| 0.0005 | 113.02 | 117.43 | 132.46 |
| Probability of no loss | 0.69 | 0.45 | 0.47 |
| Mean (USD million) | 1.87 | 1.85† | 1.76† |
| Std. dev. (USD million) | 7.57 | 6.90† | 7.05† |
| Coefficient of skewness | 8.84 | 10.20† | 11.85† |

† On the basis of analytical evaluation of moments

Table 5.4: *Distribution of the aggregate loss over the period 01 March 2009 - 28 February 2010. Except where indicated otherwise, all figures are estimates on the basis of 10,000 simulated* $(t, u]$ *periods.*

stress release models, in that order.

2. Larger estimated mean and standard deviation under the stress release model. More generally, evidence of first order stochastic dominance of the aggregate loss distribution under the stress release model (over the counterparts under the Poisson and Poisson-gamma models) up to the c. 100 million USD level, after which the stochastic dominance is reverted.

The Poisson and Poisson-gamma models share the same magnitude distribution, and so the probability of observing very large aggregate losses under the Poisson-gamma model is likely to be driven by a combination of the stochastic frequency and Dirichlet-driven earthquake location components of the model. To the extent that the prior mass $\alpha(\mathbb{L}) = 0.5$ is negligible (relative to $N(t) = 1740$), the Bayes estimate of the Dirichlet process is almost entirely

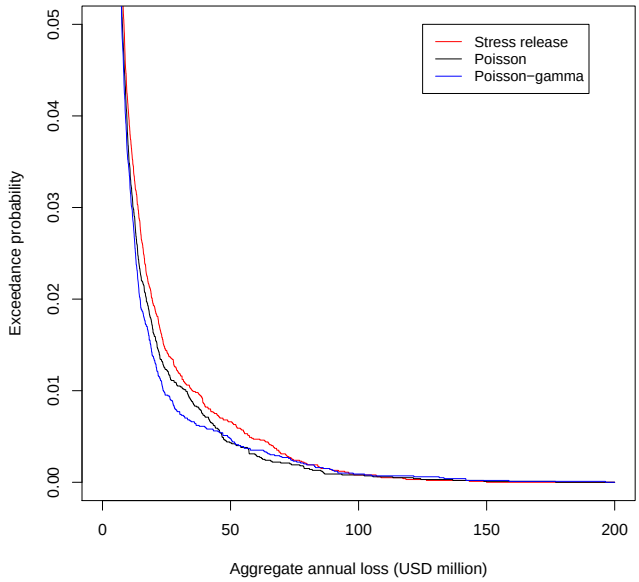


Figure 5.7: *Exceedance probability* $(1 - P^{t,u}(x))$ *curves estimated on the basis of 10,000 simulated* $(t, u]$ *periods for the Poisson, Poisson-gamma, and stress release models of the earthquake process.*

determined by the empirical distribution of locations, and so we think it is reasonable to hypothesize that the differences are mainly due to the increased variability introduced by the presence of a stochastic intensity measure. In turn, the differences are seemingly mild, which would suggest that there is no strong evidence of spatial clustering (feature which would be uncovered by the gamma intensity measure). However, this is likely to be influenced by the relatively concentrated $t$-portfolio (see Figure 5.8) and so this should not be regarded as conclusive evidence.

We tested the sensitivity of the estimated aggregate loss distribution to changes in the specification of $\alpha(\mathbb{L})$. In the first sensitivity analysis we placed a prior mass $\alpha(\mathbb{L}) = \frac{1}{2}N(t)$ distributed uniformly across the $602,602$ grid cubes. The impact of the prior distribution, which has a weight of $\frac{1}{3}$ on the Bayes estimate, is evident in the estimated losses shown in Figure 5.9: prior mass is being

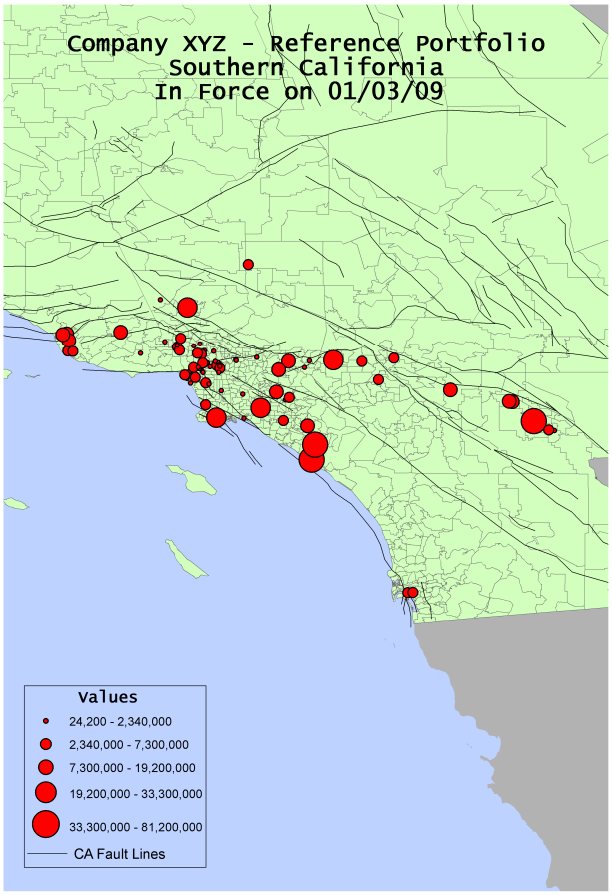


Figure 5.8: *Southern California reference exposure for Company XYZ. Portfolio in force on 01 March 2009.*

wasted in spatial regions where there is no empirical evidence of earthquake activity, and one could argue that knowledge of the location of earthquake faults should be used to inform the specification of the prior $\alpha$ measure; to the extent that this knowledge is not being incorporated in the specification of the prior, the latter is misspecified. In our second sensitivity analysis, we constructed a prior $\alpha$ measure as follows. We assumed the observed faults to dip vertically 13 km (twice the average earthquake depth, assuming fault rupture occurs half-way through its vertical length) under the surface of the earth, and placed a prior mass $\alpha(\mathbb{L}) = N(t)$ uniformly across the grid cells intersecting the fault planes.

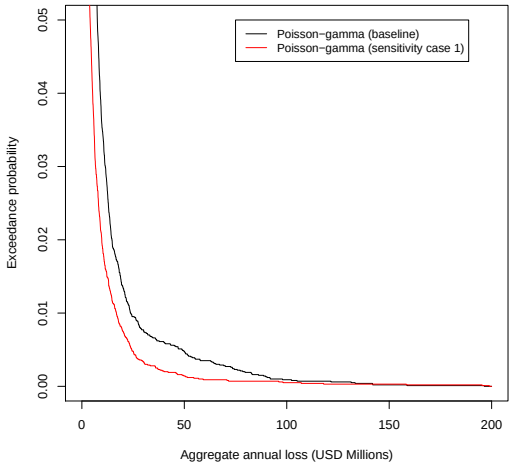


Figure 5.9: *Analysis of sensitivity of the estimated aggregate loss distribution to the specification of the prior $\alpha$ measure under the Poisson-gamma model. Case when, a priori, $\alpha(\mathbb{L}) = \frac{1}{2}N(t)$ is uniformly distributed.*

Having $\alpha(\mathbb{L}) = N(t)$ results in the Bayes estimate of the Dirichlet process being a mixture of the prior and empirical distributions of location, in equal parts. The resulting estimated aggregate loss distribution under this assumption is shown in Figure 5.10, along with the baseline estimate. To the extent that the prior $\alpha$ measure was specified on the basis of observed fault locations, there is a closer agreement between both estimates: the prior $\alpha$ measure (and resulting normed probability measure $\frac{\alpha(\cdot)}{\alpha(\mathbb{L})}$) is already largely defined by empirical evidence.

We now turn our attention to the second conspicuous feature referred to above. The stress release and Poisson models share exactly the same distribution of spatial location, and so, most likely, the differences in the loss estimates are due to the assumptions on the frequency and magnitude components of the models. The estimated mean annual number of occurrences under the stress release model (on the basis of 10,000 simulated $(t, u]$ periods) is 2.48, with a 95% C.I. given by $(2.45, 2.51)$, and so there is strong statistical evidence to

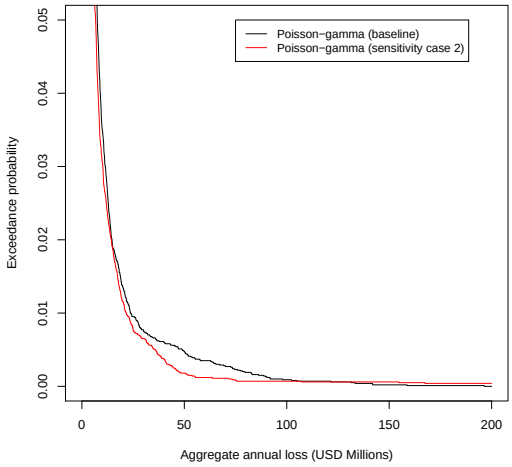


Figure 5.10: *Analysis of sensitivity of the estimated aggregate loss distribution to the specification of the prior $\alpha$ measure under the Poisson-gamma model. Case when, a priori, $\alpha(\mathbb{L}) = N(t)$ is uniformly distributed on fault planes.*

suggest an increased frequency of event occurrence over the Poisson or Poisson-gamma models (the average annual rate of events with magnitude 5 or above is $\frac{171}{77.14} = 2.22$). This in turn helps to explain the larger estimated mean, standard deviation and, in general, larger quantiles of loss up to the c.100 million USD mark: relative to the Poisson or Poisson-gamma models, there is an increased frequency of 'low-magnitude' events which prevent the underlying stress level becoming too large and thus potentially triggering a very large magnitude event that would result in a very large portfolio loss. Under the Poisson and Poisson-gamma models, on the other hand, there are no restrictions on the size (magnitude) of events, as the latter is independent of the current time or past history of the process. To get an insight into the sensitivity of this condition to the specification of the lower bound $X_0^*$, we simulated event occurrence under the stress release model, and the resulting aggregate losses, under three different assumptions for the value of the initial level of stress. We considered $X_0 = 500, 750$, and 1000, corresponding to levels of stress required to make

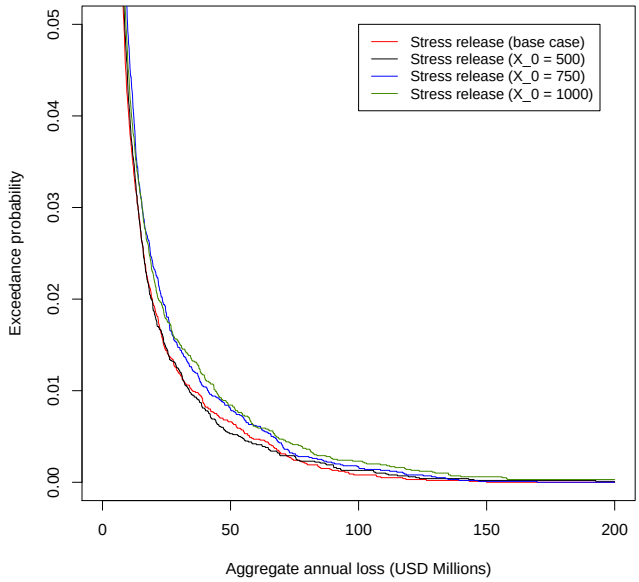


Figure 5.11: *Sensitivity scenarios under the stress release model for the distribution of the aggregate loss over the period 01 March 2009 - 28 February 2010.*

the occurrence of events of magnitude 8.6, 8.83, and 9, respectively, possible. The results are shown in Figure 5.11 and Table 5.5. Under the most extreme scenario, the 'restriction' on very large magnitude event occurrence disappears as a result of the larger value of $X_0$.

The baseline estimated losses would suggest that, for underwriting purposes (e.g. premium rating and management of accumulations), the stress release model produces more conservative loss estimates, while for reinsurance management purposes, the Poisson and Poisson-gamma models would be more conservative. This would only be valid for the one year period $(t, u]$. Should the time span considered be larger than a year (as is typically the case for capital management considerations such as securitisation of earthquake risk), the distribution of losses predicted by the stress release model may be completely different. Furthermore, this would in general have to be considered in relation

| Southern California Residential Portfolio<br>Distribution of the aggregate loss<br>(01 March 2009 - 28 February 2010) | | | | |
|---|---|---|---|---|
| | | SRM scenario | | |
| | Baseline | $X_0 = 500$ | $X_0 = 750$ | $X_0 = 1000$ |
| *Exceedance probability* | | *Loss (USD million)* | | |
| 0.05 | 8.38 | 8.67 | 9.79 | 9.15 |
| 0.02 | 19.43 | 19.10 | 23.37 | 21.84 |
| 0.01 | 35.16 | 33.90 | 41.20 | 44.07 |
| 0.005 | 57.39 | 53.62 | 66.31 | 67.92 |
| 0.004 | 66.27 | 62.34 | 70.31 | 77.16 |
| 0.002 | 79.82 | 87.21 | 91.84 | 106.03 |
| 0.001 | 96.20 | 108.67 | 117.56 | 132.54 |
| 0.0005 | 113.02 | 123.07 | 132.12 | 156.89 |

Table 5.5: *Sensitivity scenarios under the stress release model for the distribution of the aggregate loss over the period 01 March 2009 - 28 February 2010. All figures are estimates on the basis of 10,000 simulated* $(t, u]$ *periods.*

to the policies' financial structures in place, such as deductibles and limits, as the latter may have a mitigating effect (to the insurer) on large losses.

**F. On the accuracy of the analytical approximations of the aggregate loss distribution.** The NP approximation is known (see e.g. Daykin *et al.* [9]) to be a fairly accurate approximation if $0 \leq \gamma^{t,u} \leq 1$, with accuracy decreasing as $\gamma^{t,u}$ increases. It is thus not surprising to see that the NP approximation of the empirical distribution of simulated losses is very poor in all three cases, with the approximation on the basis of the simulated mean, variance, and third conditional moment of $X_{t,u}$ under the stress release model being the better of the three owing to the low value (relative to the counterparts under the Poisson and Poisson-gamma models) of the coefficient of skewness (see Figure 5.14). Under the Poisson earthquake process, both the gamma and IG approximations are surprisingly accurate (see Figure 5.12), and this is still the case under the Poisson-gamma model (see Figure 5.13), if to a lesser degree of accuracy owing to the larger coefficient of skewness of $X_{t,u}$ relative to that under the Poisson

model. The accuracy of the gamma and IG approximations is further confirmed under the stress release earthquake process (see Figure 5.14). In the latter case, however, the estimates of the moments on which the approximations are based are simulation based, and hence the potential gains in computational expense that would make the use of approximations attractive in the first place are not as apparent. One could argue, though, that the computational expense of simulation to obtain a required amount of precision for loss estimates corresponding to very low exceedance probabilities (i.e. far away in the tail of the distribution) could be reduced.

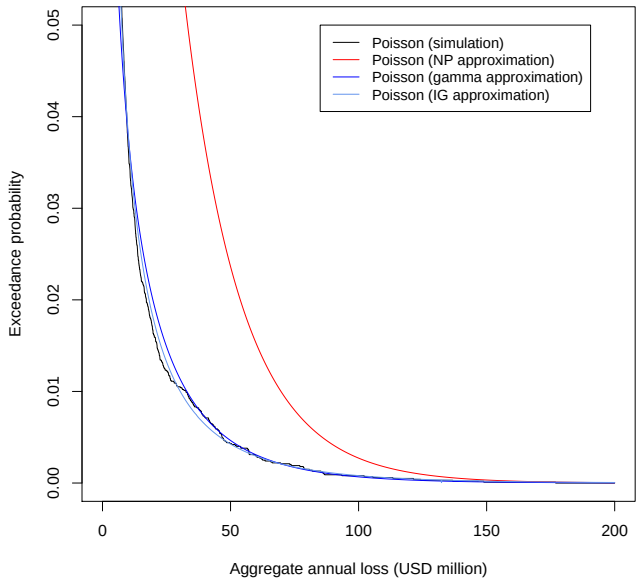


Figure 5.12: *Analytical approximations of* $P^{t,u}$ *under the Poisson model of earthquake occurrence, on the basis of* $M_1^{t,u}$, $\overline{M}_2^{t,u}$, *and* $\overline{M}_3^{t,u}$.

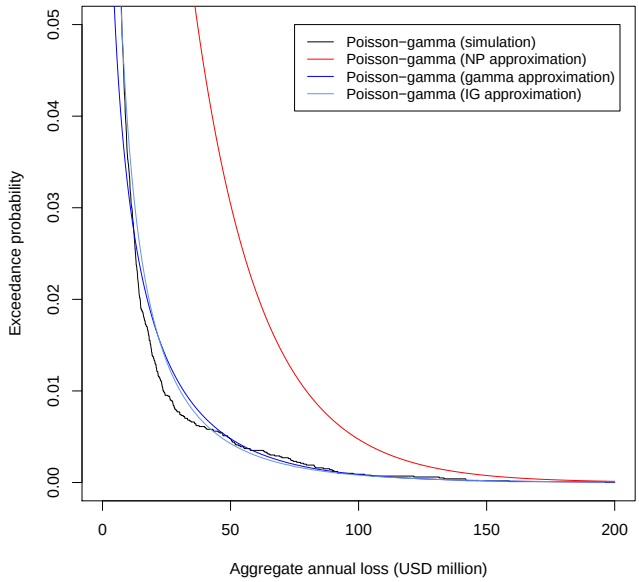


Figure 5.13: *Analytical approximations of $P^{t,u}$ under the Poisson-gamma model of earthquake occurrence, on the basis of $M_1^{t,u}$, $\overline{M}_2^{t,u}$, and $\overline{M}_3^{t,u}$.*

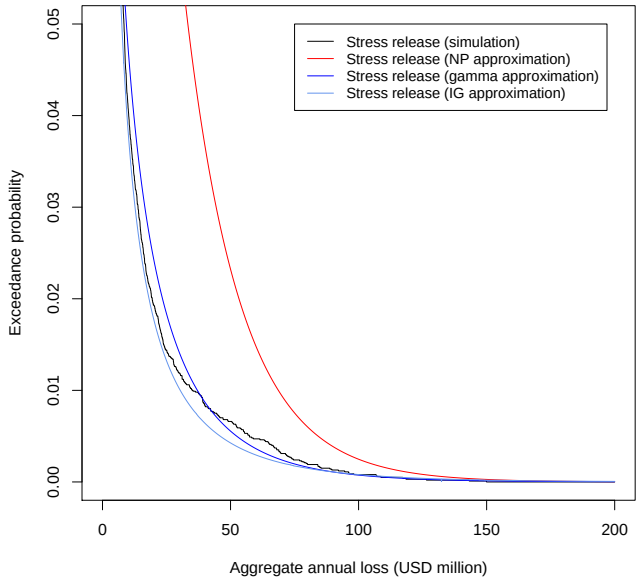


Figure 5.14: *Analytical approximations of $P^{t,u}$ under the stress release model of earthquake occurrence, on the basis of simulated moments.*

# Appendix A

# Notable models in statistical seismology

## A.1 The Hawkes process and the ETAS model

In the early 1970's Hawkes [13] introduced a family of what he referred to as *self exciting* models, which became pioneering examples of the conditional intensity methodology and models of general utility for the description of earthquake catalogs. The conditional intensity takes the form

$$\lambda(t) = \nu + \int_0^t g(t-s)N(ds), \tag{A.1}$$

where $\nu > 0$ is a background immmigration term and $g(u) > 0$ represents the contribution to the conditional intensity after a lag of length $u$, with $\int_0^\infty g(u)du < 1$. This construction captures earthquake clustering, one of the most pervasive features of catalog data: any event can be thought of as the *parent* of a family of later events, referred to as its *offspring*, which ultimately die out but are replenished by the immigration component $\nu$. In practice, $g(u)$ is usually given

a simple parametric form[1]. Ogata [30] used the Pareto-type specification

$$g(t-s) = \frac{K_0}{(t-s+c)^p} e^{\alpha(z-z_0)}, \tag{A.2}$$

by analogy with Omori's law[2]. In this model, known as the Epidemic Type Aftershock-Sequences (ETAS) model, the conditional intensity function takes the form

$$\lambda(t) = \nu + \int_0^t \int_{z_0}^{\infty} \frac{K_0}{(t-s+c)^p} e^{\alpha(z-z_0)} N(ds, dz), \tag{A.3}$$

where $Z$ represents magnitude, $z_0$ is a minimum magnitude threshold and $(\nu, K_0, c, \alpha, p)$ is a parameter vector representing certain characteristics of seismic activity in the region under study; $\alpha$ and $p$, for instance, characterize the temporal pattern of seismicity. The value $p$ indicates the decay rate of aftershocks, and the $\alpha$ value measures an efficiency of magnitude of an earthquake in generating its offspring, or aftershocks.

The ETAS model has been extended to include spatial location and magnitude as mark components. For a Borel rectangle $\mathcal{K} = \mathcal{L} \times \mathcal{Z}$ of spatial locations and magnitudes, the time-space-magnitude intensity takes the form

$$\lambda(t, l, z) = \phi(z) \left[ \nu(l) + \int_0^t \int_{\mathcal{L}} \int_{\mathcal{Z}} g(t-s, l-l', z) N(ds, dl', dz) \right], \tag{A.4}$$

where $L \in \mathcal{L} \subseteq \mathbb{R}^2$ represents the epicentre, $\nu(l)$ is a spatially distributed immigration term and $g(u, l, z)$ is now a space-time kernel. Musmeci and Vere-

[1] Vere-Jones and Ozaki [46], and Ogata et.al. [34], considered specifying $g(u)$ as a finite sum of Laguerre polynomials but these choices yielded poor-fitting models.

[2] The frequency of aftershocks is well represented by the modified Omori formula

$$n(t) = K(t+c)^{-p},$$

where $t$ is the lapse time from the occurrence of the mainshock and $K$ depends on the magnitude of the mainshock and the lower bound of the magnitude of aftershocks counted, while $p$ is known to be independent of these. See e.g. Ogata [32].

Jones [27] suggested a Gaussian diffusion-type kernel

$$g(u,l,z) = Ae^{\alpha z}e^{-\beta u}\frac{1}{2\pi\sigma_{l^{(1)}}\sigma_{l^{(2)}}u}\exp\left\{-\frac{1}{2u}\left(\frac{l^{(1)2}}{\sigma^2_{l^{(1)}}}+\frac{l^{(2)2}}{\sigma^2_{l^{(2)}}}\right)\right\}, \qquad (A.5)$$

where $l = (l^{(1)}, l^{(2)})$ represents an earthquake's epicentre, $A$ is an overall constant, $e^{\alpha z}$ describes the dependence of the risk on the magnitude of the exciting event, $e^{-\beta u}$ is an exponential damping factor (energy absorbption) and the diffusion constants $\sigma_{l^{(1)}}$,$\sigma_{l^{(2)}}$ control the rates of difussion of risk along the $l^{(1)}$ and $l^{(2)}$ directions, respectively. Rathbun [37] applied the standard bivariate Gaussian kernel to California earthquakes with $Z \geq 5$ between 1932 and 1992.

Kagan [17] suggested other parametric forms based on investigations of the second-order moment features in time and space. Kagan and Jackson [18] took the space-time-magnitude ETAS model one step further by incorporating not only the location but also the orientation of the fault movement. Ogata's most recent studies (see e.g. [33] and [35] ) use a version with spatially dependent-dependent parameters, coupled with a penalized-likelihood approach, to give optimal spatial smoothing.

Vere-Jones [44] points out that the disadvantage of these models, from the earthquake prediction point of view, is that the models have little predictive power. To the extent that $g(\cdot)$ is typically a decreasing function, the highest risk is immediately following a past event. As the local activity increases in a region, so does the risk of further events. This is not to be taken as a criticism of the self-exciting models, it is either a fact of life or a reason to look outside the catalog. Ogata ( [31], [32]) further points out that as the number of data increases or as threshold magnitudes decrease, it becomes difficult for a single ETAS model to represent the seismicity throughout the considered region or volume. This is mainly owing to the fact that significantly different seismicity patterns often take

place, even in neighbouring regions. Conseguently, the ETAS model frequently fits poorly to data with a large number of small events of microearthquakes. It is also possible that the real seismicity for a region may include forms of quiescence and activation not reproduced by the ETAS models, which do not fully capture the way in which seismic activity depends on the dynamical change of the underlying stress-field of an area.

## A.2 Stress release models

The first stress release model was proposed by Vere-Jones [42] to represent the deterministic build-up of stress within a region and its stochastic release through earthquakes. A version of this model is introduced in Chapter 4, to which we refer for background. The general assumption is that the probabilities of events occurring within the region are determined by an unobserved state variable. In the versions of the model that have been developed so far it is assumed that this state variable can be represented by a scalar quantity $X_t$ which increases linearly between events and decreases instantaneously when events occur. Whilst it is not necessary that this quantity be interpreted literally as stress, this is the general character. At any time $t > 0$ and for an initial stress level $X_0$,

$$X_t = X_0 + \rho t - S_t \tag{A.6}$$

represents the balance between the accumulated tectonic stress in a region, building up linearly at a fixed rate $\rho$, and the total amount of stress released through earthquakes,

$$S_t = \int_0^t \int_{z_0}^{\infty} s^*(z) N(ds, dz). \tag{A.7}$$

Under the stress release model, at any time $t$ when the current level of stress is

$X_{t^-}$, the intensity of occurrence of an earthquake with magnitude $z$ is

$$\lambda(t,z) = \Psi(X_{t^-})\phi(z|X_{t^-}), \tag{A.8}$$

where $\Psi(x)$ is a nondecreasing function. Letting $\Psi(x) = K$, for a constant $K$, would result in a Poisson model. Assuming a singularity

$$\Psi(x) = \begin{cases} 0 & x \leq x_c \\ \infty & x \geq x_c, \end{cases} \tag{A.9}$$

where $x_c$ is the fixed crustal strength of the region, results in a *time-predictable* model (see e.g. Vere-Jones [44] and references therein for further detail).

The time-predictable model is a version of the *characteristic earthquake* model which assumes that every major fault or fault-segment is characterised by earthquakes of a fixed size and frequency. The ideal characteristic earthquake sequence would have identical magnitudes, identical fault mechanisms and identical time intervals between successive events. The variability introduced through uncertainties of measurement and the physical process itself has led to the sequence of inter-event times being modelled either as a renewal process, or as a modified renewal process (the time-predictable model) in which the time to the next event is taken to be proportional to the size (in terms of observed stress drop) of the preceding event, the justification being that the time interval represents the time taken to build up the stress along the fault segment to the critical value needed to rupture that segment again.

The stress release model has been used to identify statistically distinct regions, in the sense that each is best fitted by different stress release models. In [51] Zheng and Vere-Jones divided North China into four seismic belts and noted evidence of clustering which relates to some form of action at a distance, i.e. stress transfer and interaction. This motivated a modification of the stress

release model by Liu, Vere-Jones et.al. [25] and Bebbington and Harte [5].

In the linked stress release model a finite number of disjoint spatial regions follow a stress release model, with the additional possibility of positive or negative stress transfers between regions at the time of occurrence of an earthquake. The stress in each region $i$ is represented by the process

$$X_t^{(i)} = X_0^{(i)} + \rho_i t - S_t^{(i)}, \tag{A.10}$$

where

$$S_i(t) = \sum_j \int_0^t \int_{z_0}^{\infty} \theta_{ij} s^*(z) N(ds, dz). \tag{A.11}$$

The integrand in (A.11) corresponds to stress drops resulting from earthquakes originating in region $j$, weighted by the stress transfer coefficient $\theta_{ij}$ measuring the fixed proportion of stress drop, initiated in region $j$, transferred to region $i$. The $\theta_{ij}$ form a matrix of stress transfer coefficients with $\theta_{ii} = 1$ and with off-diagonal elements either positive or negative, reflecting damping or excitation, respectively. The conditional intensity in each region is of the form (A.8), where each region has the exponential risk function $\Psi(x_i) = \exp(\mu_i, +\nu_i x_i)$, with differing parameters indicating that the strength and the tectonic loading rate can differ in each seismic region.

# Appendix B

# The Poisson-gamma and Dirichlet processes

In this appendix a number of representation results for a finite collection of independent Poisson variates and for the Dirichlet distribution are derived. Further, conjugate priors to the Poisson and multinomial distributions are derived. Much of the material presented here is standard. For further detail please refer to e.g. the treatises by Johnson, Kotz and Balakrishnan, [15] and [16].

## B.1 The univariate Poisson distribution

The random variable $N$ is said to be *Poisson distributed* with frequency parameter $\tau\,\lambda > 0$, written

$$\mathcal{L}(N) \;=\; \mathrm{Po}(N;\,\tau\lambda),$$

if its density with respect to the counting measure on the integers is

$$\mathbb{P}[N = n] = \frac{(\tau\,\lambda)^n}{n!}\,e^{-\tau\,\lambda}, \tag{B.1}$$

$n = 0, 1, \ldots$ The parameter $\tau > 0$, which might appear redundant, represents operational time and $\lambda$ represents the frequency of events per time unit.

The Laplace transform of $N$ is

$$L_N(s) = \sum_{n=1}^{\infty} e^{-sn} \frac{(\tau\,\lambda)^n}{n!}\, e^{-\tau\,\lambda} = e^{-\tau\,\lambda\,(1-e^{-s})}. \tag{B.2}$$

## B.2 The multinomial distribution

A random vector $(N_1, \ldots, N_k)$ is said to be *multinomially distributed*, written

$$\mathcal{L}(N_1, \ldots, N_k) \;=\; \mathrm{M}(N_1, \ldots, N_k;\, n,\, \delta_1, \ldots, \delta_k),$$

if its density with respect to $k$-dimensional counting measure on the integers is

$$\mathbb{P}[N_i = n_i\,,\; i = 1, \ldots, k] \;\;=\;\; \frac{n!}{n_1! \cdots n_k!} \prod_{i=1}^{k} \delta_i^{n_i}, \tag{B.3}$$

$n_i \in \{0, \ldots, n\}$, $i = 1, \ldots, k$, $\sum_{i=1}^{k} n_i = n$, where the parameters are as follows: $n$ is integer, $\delta_i > 0$, $i = 1, \ldots, k$, and $\sum_{i=1}^{k} \delta_i = 1$.

## B.3 The distribution of multiple independent Poisson variates

Let $(N_1, \ldots, N_k)$ have independent Poisson entries,

$$\mathcal{L}(N_1, \ldots, N_k) \;=\; \prod_{i=1}^{k} \mathrm{Po}(N_i;\, \tau\,\lambda_i).$$

Its joint distribution is

$$\mathbb{P}\left[N_i = n_i,\, i = 1, \ldots, k\right] \;=\; \prod_{i=1}^{k} \frac{(\tau\,\lambda_i)^{n_i}}{n_i!}\, e^{-\tau\,\lambda_i}. \tag{B.4}$$

Introduce

$$N \;:=\; \sum_{i=1}^{k} N_i\,, \quad \lambda \;:=\; \sum_{i=1}^{k} \lambda_i.$$

By the independence assumption and (B.2),

$$L_N(s) \;=\; \prod_{i=1}^{k} e^{-\tau\,\lambda_i\left(1-e^{-s}\right)} = e^{-\tau\,\lambda\left(1-e^{-s}\right)},$$

hence, by (B.2),

$$\mathcal{L}(N) \;=\; \mathrm{Po}\left(N;\, \tau\,\lambda\right). \tag{B.5}$$

It follows that the Poisson distribution is infinitely divisible.

Set $n \;=\; \sum_{i=1}^{k} n_i$, and rewrite (B.4) as

$$\mathbb{P}\left[N = n,\, N_i = n_i,\, i = 1, \ldots, k\right] \;=\; \frac{(\tau\,\lambda)^n}{n!}\, e^{-\tau\,\lambda}\; \frac{n!}{n_1! \cdots n_k!} \prod_{i=1}^{k} \left(\frac{\lambda_i}{\lambda}\right)^{n_i}.$$

Comparing this with (B.1) and (B.3), we rediscover (B.5) and can moreover state the following representation result (where $\delta_i := \lambda_i/\lambda,\; i = 1, \ldots, k$):

$$\begin{aligned} \mathcal{L}(N_1, \ldots, N_k) \;&=\; \prod_{i=1}^{k} \mathrm{Po}(N_i;\, \tau\,\lambda_i) \\ &\Leftrightarrow \\ \mathcal{L}(N, (N_1, \ldots, N_k)) \;&=\; \mathrm{Po}(N;\, \tau\,\lambda)\, \mathrm{M}(N_1, \ldots, N_k;\, N, \delta_1, \ldots, \delta_k), \end{aligned} \tag{B.6}$$

Thus, a finite collection of independent Poisson variables can be obtained in two

steps: first generate their sum from a Poisson distribution and then generate the individual Poisson variables from a conditional multinomial distribution.

## B.4 The Beta and Dirichlet distributions

A random variable $\Delta$ is said to be *beta distributed* with shape parameters $\alpha > 0$ and $\beta > 0$, written

$$\mathcal{L}(\Delta) \,=\, \mathrm{Be}(\Delta;\, \alpha, \beta),$$

if its density with respect to Lebesgue measure is

$$\frac{1}{\mathrm{B}(\alpha,\beta)}\, \delta^{\alpha-1}\, (1-\delta)^{\beta-1}\, 1_{(0,1)}(\delta), \tag{B.7}$$

where

$$\mathrm{B}(\alpha,\beta) \,:=\, \int_0^1 \delta^{\alpha-1}\, (1-\delta)^{\beta-1}\, d\delta \,=\, \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

is the so-called *beta function*. The moments are easily obtained, e.g.

$$\mathbb{E}\,[\Delta] \,=\, \frac{\alpha}{\alpha+\beta}\,, \qquad \mathbb{V}\mathrm{ar}\,[\Delta] \,=\, \frac{\alpha\,\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}\,.$$

A multivariate version of the beta distribution is defined as follows. The random vector $(\Delta_1, \ldots, \Delta_k)$ is said to be *Dirichlet distributed* with parameter $(\alpha_1, \ldots, \alpha_k)$, written

$$\mathcal{L}(\Delta_1, \ldots, \Delta_k) \,=\, \mathrm{Dir}(\Delta_1, \ldots, \Delta_k;\, \alpha_1, \ldots, \alpha_k),$$

if the joint density of $(\Delta_1, \ldots, \Delta_{k-1})$ with respect to $(k-1)$-dimensional Lebesgue

measure is

$$\frac{1}{\mathrm{B}(\alpha_1, \ldots, \alpha_k)} \prod_{1=1}^{k} \delta_i^{\alpha_i - 1}, \tag{B.8}$$

$\delta_i > 0$, $i = 1, \ldots, k-1$, $\sum_{i=1}^{k-1} \delta_i < 1$, where $\delta_k := 1 - \sum_{i=1}^{k-1} \delta_i$. Here

$$\mathrm{B}(\alpha_1, \ldots, \alpha_k) \;=\; \frac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k} \alpha_i)}$$

is the multidimensional beta function. The redundant variable $\Delta_k := 1 - \sum_{i=1}^{k-1} \Delta_i$ is introduced for the sake of symmetry. It will be demonstrated below that (B.8) is a well defined density. Moreover, it will be demonstrated that, for any partition $\{K_1, \ldots, K_\ell\}$ of $\{1, \ldots, k\}$, the $\Delta_{K_j} := \sum_{i \in K_j} \Delta_i$, $j = 1, \ldots, \ell$, are distributed as

$$\mathcal{L}(\Delta_{K_1}, \ldots, \Delta_{K_\ell}) \;=\; \mathrm{Dir}\left(\Delta_{K_1}, \ldots, \Delta_{K_\ell}; \sum_{i \in K_1} \alpha_i, \ldots, \sum_{i \in K_\ell} \alpha_i\right). \tag{B.9}$$

## B.5 The univariate gamma distribution

A random variable $\Lambda$ is said to be *gamma distributed* with shape parameter $\alpha > 0$ and inverse scale parameter $\beta > 0$, written

$$\mathcal{L}(\Lambda) \;=\; \mathrm{Ga}(\Lambda;\, \alpha, \beta),$$

if its density with respect to Lebesgue measure is

$$\frac{\beta^\alpha}{\Gamma(\alpha)}\, \lambda^{\alpha - 1}\, e^{-\beta\lambda}\, d\lambda\, 1_{(0,\infty)}(\lambda)\,. \tag{B.10}$$

Plainly,

$$\mathbb{E}[\Lambda^r\, e^{-s\,\Lambda}] \;=\; \frac{\beta^\alpha}{\Gamma(\alpha)}\,\frac{\Gamma(\alpha+r)}{(\beta+s)^{\alpha+r}}\,, \qquad r>-\alpha,\; s>-\beta. \tag{B.11}$$

In particular, the Laplace transform of $\Lambda$ is

$$L_\Lambda(s) \;=\; \mathbb{E}[e^{-s\,\Lambda}] \;=\; (1+s/\beta)^{-\alpha},\;\; s>-\beta\,, \tag{B.12}$$

and the first two moments are

$$\mathbb{E}\,[\Lambda] \;=\; \frac{\alpha}{\beta},$$

$$\mathbb{V}\mathrm{ar}\,[\Lambda] \;=\; \frac{\alpha}{\beta^2}.$$

## B.6 The distribution of multiple independent gamma variates

Let the entries of $(\Lambda_1,\ldots,\Lambda_k)$ be independent gamma variates with common scale parameter,

$$\mathcal{L}(\Lambda_1,\ldots,\Lambda_k) \;=\; \prod_{i=1}^{m} \mathrm{Ga}(\Lambda_i;\,\alpha_i,\beta)\,.$$

By (B.10), their joint density is

$$\frac{\beta^{\sum_{i=1}^k \alpha_i}}{\prod_{i=1}^k \Gamma(\alpha_i)}\,\left(\prod_{i=1}^k \lambda_i^{\alpha_i-1}\right)\,e^{-\beta\sum_{i=1}^k \lambda_i}, \tag{B.13}$$

$\lambda_i>0$, $i=1,\ldots,k$. From the independence assumption and (B.12) it follows that the sum,

$$\Lambda \;:=\; \sum_{i=1}^k \Lambda_i,$$

has Laplace transform

$$L_\Lambda(s) \;=\; \prod_{i=1}^{k} L_{\Lambda_i}(s) \;=\; (1+s/\beta)^{-\sum_{i=1}^{k}\alpha_i}\,, \quad s > -\beta\,.$$

Thus, setting

$$\alpha \;:=\; \sum_{i=1}^{k} \alpha_i, \tag{B.14}$$

$$\mathcal{L}(\Lambda) \;=\; \mathrm{Ga}(\Lambda;\, \alpha, \beta). \tag{B.15}$$

It follows that the gamma distribution is infinitely divisible.

Consider the transformation of $(\lambda_1, \ldots, \lambda_k)'$ to

$$\begin{pmatrix} \delta_1 \\ \vdots \\ \delta_{k-1} \\ \lambda \end{pmatrix} = \begin{pmatrix} \lambda_1 / \sum_{i=1}^{k} \lambda_i \\ \vdots \\ \lambda_{k-1} / \sum_{i=1}^{k} \lambda_i \\ \sum_{i=1}^{k} \lambda_i \end{pmatrix}.$$

The inverse transform is

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_{k-1} \\ \lambda_k \end{pmatrix} = \begin{pmatrix} \delta_1\, \lambda \\ \vdots \\ \delta_{k-1}\, \lambda \\ \lambda \left(1 - \sum_{i=1}^{k-1} \delta_i\right) \end{pmatrix}. \tag{B.16}$$

The Jacobi matrix of the inverse transform is

$$J = \begin{pmatrix} \lambda & 0 & 0 & \cdots & 0 & \delta_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & \delta_{k-1} \\ -\lambda & -\lambda & -\lambda & -\lambda & -\lambda & 1 - \sum_{i=1}^{k-1} \delta_i \end{pmatrix}.$$

The determinant of this matrix is (add the first $(k-1)$ rows to the $k$-th row)

$$\det(J) = \det \begin{pmatrix} \lambda & 0 & 0 & \cdots & 0 & \delta_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & \delta_{k-1} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} = \lambda^{k-1}. \tag{B.17}$$

The joint density of

$$\begin{pmatrix} \Delta_1 \\ \vdots \\ \Delta_{k-1} \\ \Lambda \end{pmatrix} = \begin{pmatrix} \Lambda_1 / \sum_{i=1}^{k} \Lambda_i \\ \vdots \\ \Lambda_{k-1} / \sum_{i=1}^{k} \Lambda_i \\ \sum_{i=1}^{k} \Lambda_i \end{pmatrix} \tag{B.18}$$

is obtained upon substituting (B.16) into (B.13) and multiplying by the (absolute value of the) determinant (B.17):

$$\begin{aligned} &\frac{\beta^{\sum_{i=1}^{k} \alpha_i}}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \left( \prod_{i=1}^{k-1} (\delta_i \, \lambda)^{\alpha_i - 1} \right) \left( \lambda \left( 1 - \sum_{i=1}^{k-1} \delta_i \right) \right)^{\alpha_k - 1} e^{-\beta\lambda} \, \lambda^{k-1} \\ &\quad = \frac{\beta^{\sum_{i=1}^{k} \alpha_i}}{\Gamma(\sum_{i=1}^{k} \alpha_i)} \lambda^{\sum_{i=1}^{k} \alpha_i - 1} e^{-\beta\lambda} \times \\ &\qquad \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \left( \prod_{i=1}^{k-1} \delta_i^{\alpha_i - 1} \right) \left( 1 - \sum_{i=1}^{k-1} \delta_i \right)^{\alpha_k - 1}, \end{aligned}$$

$\delta_i > 0$, $i = 1, \ldots, k-1$, $\sum_{i=1}^{k-1} \delta_i < 1$, $\lambda > 0$. Inspection of this expression and

(B.8) and (B.10) gives the following representation result for

$$(\Delta_1, \ldots, \Delta_k) = \frac{1}{\Lambda}(\Lambda_1, \ldots, \Lambda_k)\,, \qquad \Lambda \;=\; \sum_{i=1}^{k} \Lambda_i\,: \tag{B.19}$$

$$\begin{aligned} \mathcal{L}(\Lambda_1, \ldots, \Lambda_k) \quad &= \quad \prod_{i=1}^{k} \mathrm{Ga}(\Lambda_i;\, \alpha_i, \beta) \\ &\Leftrightarrow \\ \mathcal{L}(\Lambda, (\Delta_1, \ldots, \Delta_k)) \quad &= \quad \mathrm{Ga}(\Lambda;\, \alpha, \beta)\, \mathrm{Dir}((\Delta_1, \ldots, \Delta_k); \alpha_1, \ldots, \alpha_k). \end{aligned} \tag{B.20}$$

Thus, Dirichlet variables are obtained by dividing independent gamma variables with common scale parameter by their sum (the scale parameter cancels in these ratios, of course). And a finite collection of independent gamma variables can be obtained as the product of a gamma variable and an independent Dirichlet variable with shape parameters satisfying (B.14).

The previously announced result (B.9) is a simple consequence of (B.19) and (B.20).

## B.7 Conjugate priors to the Poisson and multinomial distributions

Suppose

$$\mathcal{L}(\Delta_1, \ldots, \Delta_k) \;=\; \mathrm{Dir}\,(\Delta_1, \ldots, \Delta_k;\, \alpha_1, \ldots, \alpha_k)\;,$$

and

$$\mathcal{L}\,(N_1, \ldots, N_k \,|\, \Delta_1, \ldots, \Delta_k) \;=\; \mathrm{M}(N_1, \ldots, N_k;\, n, \Delta_1, \ldots, \Delta_k)\,.$$

Recalling (B.3) and (B.8), we find

$$\begin{aligned}
&\mathbb{P}[\Delta_i \in d\delta_i,\, i = 1, \ldots, k-1 \,|\, N_i = n_i,\, i = 1, \ldots, k] \\
&\quad \propto \mathbb{P}[N_i = n_i,\, i = 1, \ldots, k \,|\, \Delta_i = \delta_i,\, i = 1, \ldots, k-1] \,\times \\
&\qquad \mathbb{P}[\Delta_i \in d\delta_i,\, i = 1, \ldots, k-1] \\
&\quad \propto \left( \prod_{i=1}^{k} \delta_i^{n_i + \alpha_i - 1} \right) d\delta_1 \cdots d\delta_{k-1} \,,
\end{aligned}$$

and conclude that

$$\mathcal{L}\,(\Delta_1, \ldots, \Delta_k \,|\, N_1, \ldots, N_k) \;=\; \mathrm{Dir}(\Delta_1, \ldots, \Delta_k;\, N_1 + \alpha_1, \ldots, N_k + \alpha_k). \tag{B.21}$$

Thus, the Dirichlet distribution is the natural conjugate prior to the multinomial distribution.

Suppose

$$\mathcal{L}(\Lambda) \;=\; \mathrm{Ga}(\Lambda;\, \alpha, \beta)\,, \tag{B.22}$$

and

$$\mathcal{L}\,(N \,|\, \Lambda) \;=\; \mathrm{Po}(N;\, \tau\,\Lambda)\,. \tag{B.23}$$

From (B.1) and (B.10) we gather

$$\mathbb{P}[\Lambda \in d\lambda \,|\, N = n] \;\propto\; \mathbb{P}[N = n \,|\, \Lambda = \lambda]\,\mathbb{P}[\Lambda \in d\lambda] \;\propto\; \lambda^{n+\alpha-1}\, e^{-(\tau+\beta)\,\lambda} d\lambda,$$

which means

$$\mathcal{L}\,(\Lambda \,|\, N) \;=\; \mathrm{Ga}(N + \alpha, \tau + \beta)\,. \tag{B.24}$$

Thus, the gamma distribution is the natural conjugate prior to the Poisson distribution.

(A side remark: From (B.11) we obtain

$$\mathbb{P}[N = n] \;=\; \mathbb{E}\left[\frac{(\tau\,\Lambda)^n}{n!}\,e^{-\tau\,\Lambda}\right] \;=\; \binom{n+\alpha-1}{n}\left(\frac{\beta}{\tau+\beta}\right)^{\alpha}\left(\frac{\tau}{\tau+\beta}\right)^{n}, \tag{B.25}$$

$n = 0, 1, ...$ This means that the marginal distribution of $N_i$ is

$$\mathcal{L}(N) \;=\; \mathrm{NB}\left(\alpha, \frac{\beta}{\tau+\beta}\right),$$

the negative binomial distribution.)

Now consider the multivariate doubly stochastic Poisson-gamma model,

$$\mathcal{L}\,(\Lambda_1, \ldots, \Lambda_k) \;=\; \prod_{i=1}^{k} \mathrm{Ga}(\Lambda_i;\, \alpha_i,\, \beta)\,, \tag{B.26}$$

$$\mathcal{L}\,(N_1, \ldots, N_k \,|\, \Lambda_1, \ldots, \Lambda_k) \;=\; \prod_{i=1}^{k} \mathrm{Po}(N_i;\, \tau\,\Lambda_i)\,. \tag{B.27}$$

Due to independence, (B.24) can be applied component-wise to give the posterior distribution

$$\mathcal{L}\,(\Lambda_1, \ldots, \Lambda_k \,|\, N_1, \ldots, N_k) \;=\; \prod_{i=1}^{k} \mathrm{Ga}(\Lambda_i;\, N_i + \alpha_i,\, \tau + \beta). \tag{B.28}$$

Using (B.6) and (B.20), the model can equivalently be cast as

$$\mathcal{L}\,(\Lambda, (\Delta_1, \ldots, \Delta_k)) \;=\; \mathrm{Ga}(\Lambda;\, \alpha, \beta)\ \mathrm{Dir}(\Delta_1, \ldots, \Delta_k; \alpha_1, \ldots, \alpha_k)\,, \tag{B.29}$$

$$\begin{aligned}
&\mathcal{L}\,(N,\,(N_1,\ldots,N_k)\,|\,\Lambda,\,(\Delta_1,\ldots,\Delta_k))\\
&\quad=\,\mathrm{Po}(N;\,\tau\,\Lambda)\,\mathrm{M}\,(N_1,\ldots,N_k;\,N,\Delta_1,\ldots,\Delta_k)\;. \qquad \text{(B.30)}
\end{aligned}$$

By (B.21) and (B.24), the posterior distribution is

$$\begin{aligned}
&\mathcal{L}(\Lambda,(\Delta_1,\ldots,\Delta_k)\,|N,(N_1,\ldots,N_k)\,)\\
&=\mathcal{L}(\Lambda\,|N)\;\mathcal{L}\,(\Delta_1,\ldots,\Delta_k\,|\,(N_1,\ldots,N_k))\\
&=\mathrm{Ga}(\Lambda;\,N+\alpha,\tau+\beta)\mathrm{Dir}(\Delta_1,\ldots,\Delta_k;N_1+\alpha_1,\ldots,N_k+\alpha_k). \qquad \text{(B.31)}
\end{aligned}$$

# Appendix C

# Simulation

## C.1 Simulation of earthquake location in space

Consider earthquake location $L = \left(L^{(1)}, L^{(2)}, L^{(3)}\right) \in \mathbb{L}$ and assume

$$\mathbb{L} := \left[l_{\mathrm{l}}^{(1)}, l_{\mathrm{u}}^{(1)}\right] \times \left[l_{\mathrm{l}}^{(2)}, l_{\mathrm{u}}^{(2)}\right] \times \left[l_{\mathrm{l}}^{(3)}, l_{\mathrm{u}}^{(3)}\right] \subset \mathbb{R}^3,$$

where $[l_{\mathrm{l}}^{(j)}, l_{\mathrm{u}}^{(j)}]$ denotes the support of $L^{(j)}$, $j = 1, 2, 3$, is the seismic region under study. The interest is to simulate observations of $L$ on the basis of the empirical distribution $\Delta_{N(\tau)}$ of the latter, or, equivalently, on the basis of the density estimate (given $l_1, \cdots, l_{N(\tau)}$)

$$\hat{\delta}_{N(\tau)}(l) = \frac{1}{N(\tau)} \sum_{i=1}^{N(\tau)} \left[ \prod_{j=1}^{3} 1_{[-h_j, h_j]} \left( l^{(j)} - \tilde{l}_i^{(j)} \right) \right], \tag{C.1}$$

where $h_j$ is the bandwidth in the $l^{(j)}$ direction, and $\tilde{l}_i^{(j)}$ is the centre of the interval in which $l_i^{(j)}$ falls.

To this end, we construct a grid to be laid over $\mathbb{L}$ as follows. For $j = 1, 2, 3$,

the grid points in the $l^{(j)}$ direction are

$$l_{i+1}^{(j)} = l_1^{(j)} + i \cdot \Delta l^{(j)}, \;\; i = 0, \ldots, N_j,$$

the grid size in the $l^{(j)}$ direction is given by

$$\Delta l^{(j)} := \frac{\left(l_{\mathrm{u}}^{(j)} - l_1^{(j)}\right)}{N_j},$$

and the grid midpoints in the $l^{(j)}$ direction are

$$\tilde{l}_i^{(j)} = l_1^{(j)} + \frac{i}{2} \cdot \Delta l^{(j)}, \;\; i = 1, \ldots, N_j - 1.$$

Define the set of grid hypercube midpoints

$$\mathbb{S} := \left\{ \left( \tilde{l}_i^{(1)}, \tilde{l}_{j'}^{(2)}, \tilde{l}_k^{(3)} \right) \right\},$$

for all $i = 1, ..., N_1 - 1$, $j' = 1, ..., N_2 - 1$, and $k = 1, ..., N_3 - 1$, and let $\tilde{l}_p$ denote its $p$-th element. Probability weighted sampling from $\mathbb{S}$, on the basis of $\left(\hat{\delta}_{N(\tau)}(\tilde{l}_1), \ldots, \hat{\delta}_{N(\tau)}(\tilde{l}_{N_{\mathrm{grid}}})\right)$, can then be used to generate observations of $L$.

## C.2 Simulation of the Dirichlet process

Consider the partition $\{\mathcal{L}_1, \ldots, \mathcal{L}_{N_{\mathrm{grid}}}\}$ of $\mathbb{L}$, where $\mathcal{L}_p$, is the $p-$th constituent hypercube of the grid defined in C.1, with midpoint given by the $p-$th element of $\mathbb{S}$. Given $\mathcal{F}_t$, an observation of the Dirchlet process can be simulated as follows:

1. Simulate $\{\lambda_p\}$, $p = 1, ..., N_{\mathrm{grid}}$, as a set of independent

$$\lambda_p \sim \mathrm{Ga}\left(\alpha(\mathcal{L}_p) + N(t), \beta + t\right)$$

random variables.

2. Set $\delta_p = \frac{\lambda_p}{\sum_p \lambda_p}$.

3. The output is the vector $(\delta_1, \ldots, \delta_{N_{\text{grid}}})$.

An observation of $L$ can then be simulated using the approach in C.1, using probability weighted sampling from $\mathbb{S}$ on the basis of $(\delta_1, \ldots, \delta_{N_{\text{grid}}})$.

## C.3 Simulation of the stress release model

We assume that the process, specified through its fitted intensity $\hat{\lambda}(s)$, is to be simulated over a finite time interval $(t, u]$, given $\mathcal{F}_t$. For $s > t$, define the *list history*

$$H_s := \{\mathcal{F}_t, t_1, \ldots, t_{N(s)}\},$$

where $N(s)$ is the number of points $t_i$ satisfying $t < t_i < s$. For every $s$ in $(t, u]$, we assume there are given two quantities, a local bound $M(s|H_s)$ for the intensity over a time interval of length $L(s|H_s)$. We take

$$L(s|H_s) = \frac{2}{\hat{\lambda}(s|H_s)}$$

and

$$M(s|H_s) = \hat{\lambda}(s + L(s|H_s)|H_s),$$

where the reasoning behind the latter is that, with high probability, the next event would occur within twice the mean interval length at the start of the interval, and because of the increasing nature of the intensity function, a simple bound is its value at the end of the interval. The algorithm then proceeds as follows.

1. Set $s = t+$, $i = 0$.

2. If $s \geq u$ go to step 8. Otherwise, calculate $L(s|H_s)$ and $M(s|H_s)$.

3. Generate $T$ as an $\text{Exp}\,(M(s|H_s))$ random variable and $U$ as a $\text{Unif}(0,1)$ random variable.

4. If $T > L(s|H_s)$, set $s = s + L(s|H_s)$ and return to step 2. Otherwise go to step 5. (This step is to verify whether the simulated time $T$ corresponds to the ocurrence time of a candidate earthquake. The event that $T > L(s|H_s)$ means no earthquake occurred —according to the model— because otherwise the interocurrence time would be less than or equal to $L(s|H_s)$.)

5. If $T \leq L(s|H_s)$ and $\frac{\hat{\lambda}(s+T|H_s)}{M(s|H_s)} > U$, replace $s$ by $s+T$ and return to step 2. Otherwise go to step 6. (This step is essentially a thinning step to verify whether the candidate earthquake occurring at time $T$ is a 'qualifying' event. If the thinning criteria $\frac{\hat{\lambda}(s+T|H_s)}{M(s|H_s)} > U$ is met, then $T$ doesn't correpond to the occurrence time of an earthquake.)

6. Set $i = i+1$ and $t_i = s+T$. Set $s = t_i$ and generate $z_i$ from the distribution $\phi(z|t_i)$ and $l_i$ from the empirical distribution $\hat{\Delta}_n$.

7. Update the list-history $H$ to $H \cup (t_i, z_i, l_i)$, and return to step 2.

8. The output is the list $\{i; (t_1, z_1, l_1), \ldots, (t_i, z_i, l_i)\}$.

# Appendix D

# Sample loss and policy records

California Residential Lines Portfolio in Force on 17 January 1994
Northridge Earthquake Loss Experience

| | | Claims Records | | | Geocoding | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *LocationID* | *City* | *Value (USD)* | *Loss (USD)* | *Loss Ratio* | *Latitude* | *Longitude* | *Landslide* | *Liquefaction* | *Soil Type* |
| 11937 | REDONDO BEACH | 3,983,900 | 101,462 | 0.0255 | 33.89 | -118.38 | 1.00 | 1.00 | 3.00 |
| 11933 | LOS ANGELES | 5,132,600 | 209,074 | 0.0407 | 34.04 | -118.20 | 1.01 | 1.19 | 2.55 |
| 12075 | PASADENA | 712,400 | 4,265 | 0.0060 | 34.16 | -118.14 | 1.00 | 2.00 | 2.50 |
| 12082 | MONTEREY PARK | 7,148,200 | 83,463 | 0.0117 | 34.06 | -118.13 | 1.05 | 1.93 | 2.49 |
| 11953 | SAN GABRIEL | 1,715,800 | 17,926 | 0.0104 | 34.10 | -118.11 | 1.00 | 2.00 | 2.61 |
| 12109 | COMPTON | 2,795,700 | 34,556 | 0.0124 | 33.90 | -118.23 | 1.00 | 2.00 | 3.00 |
| 12103 | CARSON | 7,065,300 | 44,010 | 0.0062 | 33.83 | -118.27 | 1.00 | 3.04 | 2.91 |
| 12181 | PARAMOUNT | 17,686,800 | 210,849 | 0.0119 | 33.90 | -118.15 | 1.00 | 2.09 | 3.00 |
| 11954 | ARCADIA | 2,067,200 | 23,685 | 0.0115 | 34.13 | -118.04 | 1.00 | 2.00 | 3.00 |
| 12145 | NORWALK | 5,094,400 | 51,433 | 0.0101 | 33.91 | -118.08 | 1.00 | 3.03 | 2.98 |
| 12199 | WHITTIER | 390,200 | 3,701 | 0.0095 | 33.97 | -118.04 | 1.00 | 1.00 | 3.00 |
| 12045 | LONG BEACH | 2,621,600 | 0 | 0.0000 | 33.78 | -118.19 | 1.00 | 1.00 | 2.50 |
| 12242 | LONG BEACH | 13,996,800 | 0 | 0.0000 | 33.77 | -118.18 | 1.00 | 1.05 | 2.51 |
| 11970 | LONG BEACH | 499,600 | 0 | 0.0000 | 33.77 | -118.18 | 1.00 | 1.05 | 2.51 |
| 12160 | LA PALMA | 19,597,900 | 185,187 | 0.0094 | 33.84 | -118.04 | 1.00 | 4.00 | 3.00 |
| 12187 | WEST COVINA | 13,879,200 | 56,588 | 0.0041 | 34.04 | -117.93 | 1.07 | 1.18 | 2.67 |
| 12179 | COVINA | 13,672,900 | 9,291 | 0.0007 | 34.10 | -117.91 | 1.00 | 2.00 | 3.00 |
| 12141 | OXNARD | 9,601,800 | 102,347 | 0.0107 | 34.22 | -119.19 | 2.00 | 3.39 | 3.00 |
| 12121 | VENTURA | 31,160,000 | 259,386 | 0.0083 | 34.27 | -119.19 | 1.17 | 2.00 | 3.00 |
| 12055 | GLENDORA | 4,048,700 | 0 | 0.0000 | 34.14 | -117.86 | 1.00 | 2.00 | 3.00 |
| 12147 | VENTURA | 24,302,300 | 124,775 | 0.0051 | 34.26 | -119.23 | 1.17 | 2.00 | 3.00 |

Table D.1: *Company XYZ Northridge earthquake reference loss experience. Sample records.*

| Southern California Residential Lines Portfolio in Force on 01 March 2009 | | | | | | | |
|---|---|---|---|---|---|---|---|
| Policy Records | | | Geocoding | | | | |
| *LocationID* | *City* | *Value (USD)* | *Latitude* | *Longitude* | *Landslide* | *Liquefaction* | *Soil Type* |
| 12214 | BEVERLY HILLS | 550,200 | 34.06 | -118.39 | 1.00 | 1.58 | 2.76 |
| 11951 | VAN NUYS | 1,500,800 | 34.19 | -118.44 | 1.00 | 2.00 | 3.00 |
| 12119 | LAGUNA NIGUEL | 65,562,500 | 33.51 | -117.73 | 1.63 | 1.56 | 2.33 |
| 11929 | LOS ANGELES | 5,120,900 | 34.08 | -118.30 | 1.00 | 1.06 | 2.48 |
| 12076 | SANTA ANA | 3,278,500 | 33.75 | -117.90 | 1.00 | 3.99 | 3.00 |
| 12231 | LOS ANGELES | 547,400 | 34.02 | -118.44 | 1.00 | 2.81 | 3.00 |
| 12271 | LOS ANGELES | 755,000 | 34.07 | -118.34 | 1.00 | 1.00 | 2.50 |
| 12127 | SANTA MONICA | 724,100 | 34.03 | -118.48 | 1.00 | 1.68 | 2.77 |
| 12130 | PORT HUENEME | 2,778,000 | 34.16 | -119.19 | 1.00 | 3.97 | 3.11 |
| 11987 | RESEDA | 3,096,900 | 34.19 | -118.54 | 1.00 | 2.05 | 3.00 |
| 12069 | LOS ANGELES | 6,957,500 | 34.00 | -118.43 | 1.00 | 3.00 | 3.00 |
| 11971 | LOS ANGELES | 1,590,000 | 34.10 | -118.31 | 1.02 | 2.86 | 2.82 |
| 12240 | MARINA DEL REY | 566,900 | 33.97 | -118.46 | 1.00 | 4.96 | 3.00 |
| 12090 | LOS ANGELES | 1,217,700 | 33.92 | -118.27 | 1.00 | 1.00 | 2.50 |
| 12197 | POMONA | 1,242,200 | 34.06 | -117.78 | 1.34 | 2.03 | 3.01 |
| 12123 | FULLERTON | 580,800 | 33.87 | -117.89 | 1.00 | 2.01 | 3.00 |
| 12157 | CANYON COUNTRY | 33,299,200 | 34.43 | -118.48 | 3.95 | 3.29 | 2.74 |
| 12168 | VENTURA | 11,484,600 | 34.27 | -119.20 | 1.17 | 2.00 | 3.00 |
| 11997 | LOS ANGELES | 1,585,200 | 34.11 | -118.19 | 1.73 | 1.52 | 2.54 |
| 12089 | LOS ANGELES | 3,483,200 | 34.06 | -118.28 | 1.00 | 1.49 | 2.60 |
| 12136 | MOORPARK | 14,045,400 | 34.28 | -118.87 | 1.75 | 2.54 | 2.66 |

Table D.2: *Company XYZ reference portfolio in force on 01 March 2009. Sample records.*

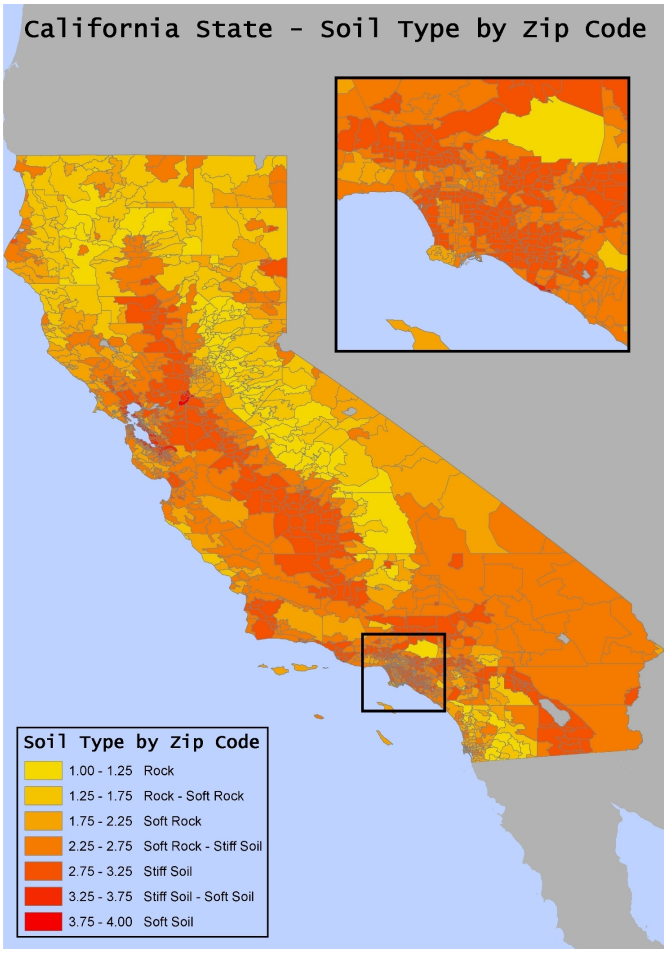


Figure D.1: *California soil type by ZIP code.*

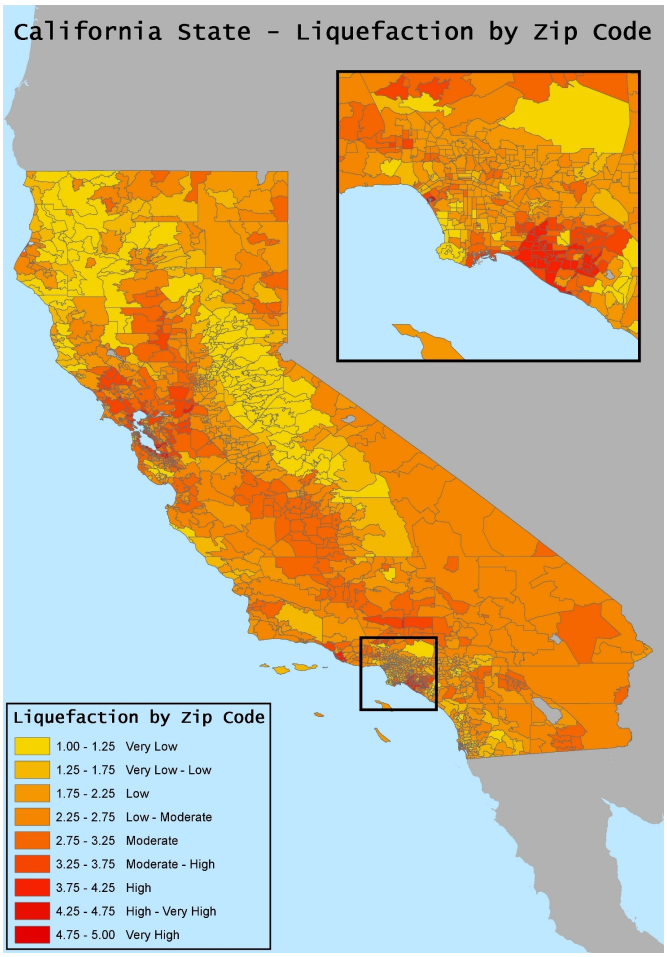


Figure D.2: *California proneness to liquefaction by ZIP code.*

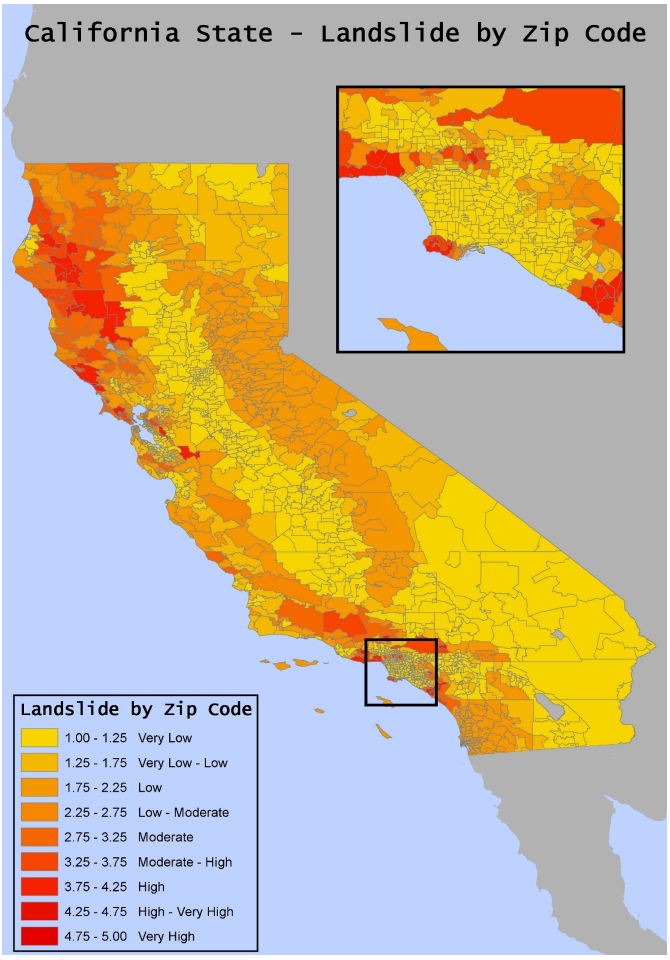


Figure D.3: *California proneness to landslide by ZIP code.*

# Bibliography


[1] Agresti,A., *Categorial Data Analysis, 2nd. Edition*, Wiley, 2002.

[2] Baddeley, A., Turner, R., Moller, J. and Hazelton, M., Residual analysis for spatial point processes (with discussion), *Journal of the Royal Statistical Society*, B,67, Part 5, 617-666, 2005.

[3] Betrò, B. and Ladelli, L., Point process analysis for italian seismic activity, *Applied Stochastic Models and Data Analysis*, 12, 75-105, 1996.

[4] Borovkov, K. and Vere-Jones, D., Explicit formulae for stationary distributions of stress release processes, *J. Appl. Prob.*, 37, 315-321, 2000.

[5] Bebbington, M. and Harte, D., On the statistics of the linked stress release model, *J. Appl. Probab.*, 38A, 176-187, 2001.

[6] Chaubey, Y.P., Garrido, J. and Trudeau, S., On the computation of aggregate claims distributions: some new approximations., *Insur.: Math. Econ.*, 23(3), 215-230, 1998.

[7] Daley, D.J. and Vere-Jones, D., *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*, Springer-Verlag, 2003.

[8] Davis, M.H.A., *Markov Models and Optimization*, Chapman & Hall, 1993.

[9] Daykin, C.D., Pentikäinen, T. and Pesonen, M., *Practical Risk Theory for Actuaries*, Chapman & Hall, 1996.

[10] Ferrari, S.L.P. and Cribari-Neto, F., Beta regression for modelling rates and proportions, *Journal of Applied Statistics*, 31, 799-815, 2004.

[11] Gutenberg, B. and Richter, C.F., Frequency of earthquakes in California, *Bull. Seismol. Soc. Amer.*, 34, 185-188, 1944.

[12] Harding, E.J. and Kendall, D.G. (Eds.), *Stochastic Geometry*, Wiley, 1974.

[13] Hawkes, A.G., Spectra of some self-exciting and mutually exciting point processes, *Biometrika*, 58, 83-90, 1971.

[14] Holden, L., Natvig, B., Sannan, S. and Bungum, H., Modeling spatial and temporal dependencies between earthquakes, Statistical Research Report, Department of Mathematics, University of Oslo, 2000.

[15] Johnson, N.L., Kotz, S. and Balakrishnan, N., *Discrete Multivariate Distributions*, Wiley, 1997.

[16] Johnson, N.L., Kotz, S. and Balakrishnan, N., *Continuous Multivariate Distributions, Volume 1, Models and Applications, 2nd. Edition*, Wiley, 2000.

[17] Kagan, Y.Y., Likelihood analysis of earthquake catalogs, *J. Geophys. Res.*, 106, 135-148, 1991.

[18] Kagan, Y.Y. and Jackson, D.D., New seismic gap hypothesis: Five years after, *J. Geophys. Res.*, 100, 3943-3959, 1995.

[19] Kagan, Y.Y. and Schoenberg, F., Estimation of the upper cutoff parameter for the tapered Pareto distribution, *J. Appl. Probab.*, 38A, 158-175, 2001.

[20] Kanamori, H. and Anderson, D.L., Theoretical basis of some empirical relations in seismology, *Bull. Seismol. Soc. Amer.*, 65, 1073-1095, 1975.

[21] Karr, A.F., *Point processes and their Statistical Inference*, Dekker, 1986.

[22] King, G.C.P., Stein, R.S. and Lin, J., Static stress changes and the triggering of earthquakes, *Bull. Seismol. Soc. Amer.*, 84, 935-953, 1994.

[23] Knopoff, L., A stochastic model for the occurrence of main sequence events, *Rev. Geophys. and Space Phys.*, 9, 175-188, 1971.

[24] Last, G., Ergodicity properties of stress release, repairable system and workload models, *Adv. Appl. Prob.*, 36, 471-498, 2004

[25] Liu, J., Vere-Jones, D., et.al., Coupled stress release model for time-dependent seismicity, *Pure appl. geophys.*, 155, 649-667, 1999.

[26] McCullagh, P. and Nelder, J.A., *Generalized Linear Models*, 2nd. ed., Chapman & Hall/CRC, 1999.

[27] Musmeci, F. and Vere-Jones, D., A space-time clustering model for historical earthquakes, *Ann. Inst. Statist. Math.*, 44, 1-11, 1992.

[28] Natvig, B. and Tvete, I.F., Bayesian hierarchical space-time modeling of earthquake data, *Methodol. Comput. Appl. Probab*, 9, 89-114, 2007.

[29] Ogata, Y., On Lewis' simulation method for point processes, *IEEE Trans. Inf. Theory*, IT-27, 23-31, 1981.

[30] Ogata, Y., Statistical models for earthquake occurrences and residual analysis for point processes, *J. Amer. Statist. Assoc.*, 83, 9-27, 1988.

[31] Ogata, Y., Space-time point-process models for earthquake occurrences, *Ann. Inst. Statist. Math.*, 50, 379-402, 1998.

[32] Ogata, Y., Seismicity analysis through point-process modeling: A review, *Pure appl. geophys.*, 155, 471-501, 1999.

[33] Ogata, Y., Space-time model for regional seismicity and detection of crustal stress changes, *J. Geophys. Res.*, 109, ,2004.

[34] Ogata, Y., Akaike, H. and Katsura, K., The application of linear intensity models to the investigation of causal relations between a point process and another stochastic process, *Ann. Inst. Statist. Math.*, 34, 373-387, 1982.

[35] Ogata, Y., Katsura, K. and Kanemura, M., Modelling heterogenous space-time occurrences of earthquakes and its residual analysis, *Appl. Stat.*, 52, 499-509, 2003.

[36] Papangelou, F., On the Palm probabilities of processes of points and processes of lines. In Harding and Kendall [12], 114 -147.

[37] Rathbun, S.L., Modeling marked spatio-temporal point patterns, *Bulletin of the International Statistical Institute*, 55, 379-396, 1993.

[38] Rotondi, R. and Varini, E., Bayesian analysis of a marked point process: Application in seismic hazard assessment, *Statistical Methods & Applications*, 12, 79-92, 2003.

[39] Schoenberg, F.P., Multidimensional residual analysis of point process models for earthquake occurrence, *J. Amer. Statist. Assoc.*, 98, 789-795, 2003.

[40] Utsu, T., Representation and analysis of the earthquake size distribution: a historical review and some new approaches, *Pure Appl. Geophys.*, 155, 509-535, 1999.

[41] Vere-Jones, D., Stochastic models for earthquake occurrence (with discussion), *Journal of the Royal Statistical Society*, B, 32, 1-62, 1970.

[42] Vere-Jones, D., Earthquake prediction: A statistician's view, *Journal of the Physics of the Earth*, 26, 129-146, 1978.

[43] Vere-Jones, D., On the variance properties of stress-release models, *Austral. J. Statist.*, 30A, 123-135, 1988.

[44] Vere-Jones, D., Forecasting earthquakes and earthquake risk, *International Journal of Forecasting*, 11, 503-538, 1995.

[45] Vere-Jones, D., Some models and procedures for space-time point processes, *Environ. Ecol. Stat.*, DOI 10.1007/s10651-007-0086-0, 2008.

[46] Vere-Jones, D. and Ozaki, T., Some examples of statistical estimation applied to earthquake data, I. Cyclic Poisson and self-exciting models, *Ann. Inst. Statist. Math.*, 34, 189-207, 1982.

[47] Vere-Jones, D. and Ogata, Y., On the moments of a self-correcting process, *J. Appl. Prob.*, 21, 335-342, 1984.

[48] Vere-Jones, D., Robinson, R., and Yang, W.Z., Remarks on the accelerated moment release model; problems of model formulation and estimation, *Geophys. J. Int.*, 144, 517-531, 2001.

[49] Vere-Jones, D., Ben-Zion, Y. and Ziga, R., Statistical seismology, *Pure Appl. geophys.*, 162, 1023-1026, 2005.

[50] Zheng, X., Ergodic theorems for stress release processes, *Stochastic Processes and their Applications*, 37, 239-258, 1991.

[51] Zheng, X. and Vere-Jones, D., Applications of stress-release models to earthquakes from North China, *Pure Appl. Geophys.*, 135, 559-576, 1991.

[52] Zheng, X. and Vere-Jones, D., Further applications of the stochastic stress release model to historical earthquake data, *Tectonophysics*, 229, 101-121, 1994.