ETAS Formulary - Derivations for Gradient-Based Optimization

Christian Grimm
(Christian.Grimm@stat.uni-muenchen.de)

May 16, 2022

Abstract

This document gives a comprehensive summary of formulas and derivations of all partial derivatives needed to implement a gradient-based optimization process for the conventional ETAS model (e.g. Jalilian, 2019) or, in particular, for the ETAS-Anisotropic and ETAS-Incomplete model versions proposed in Grimm et al. (2021) and Grimm et al. (2022), respectively.
Chapter 1 introduces the general notation of ETAS models. Then, Chapter 2 derives the log-likelihood functions for the corresponding model versions. Finally, Chapter 3 and 4 rigorously derive the derivatives of the log-likelihood function by the corresponding model parameters, as needed for the implementation of a gradient-based optimization method.
Note that I use the terms *conventional* and *standard* ETAS model synonymously.

Contents

Chapter 1

Definition and Notation of ETAS Models

In this chapter, two versions of the Epidemic Type Aftershock Sequence (ETAS) model are introduced:

- the *Conventional ETAS* model introduced by Ogata (1988, 1998) and implemented in the R package "ETAS" (Jalilian, 2019) and
- the *ETAS-Incomplete* model as described by (Grimm et al., 2022).

We define both models and derive the corresponding log-likelihood (LL) functions that are needed to optimize the model parameters by maximum likelihood estimation (MLE). Derivatives of the LL-functions, necessary for gradient-based optimization methods, are derived in later chapters.

Notations follow Jalilian (2019), Hainzl (2021) and Grimm et al. (2022), the general derivation of the LL-function for Poisson models is taken from **?**.

1.1 Conventional ETAS Model

In the conventional ETAS approach, the occurrence rate of an earthquake

- with *magnitude* m,
- occurring at *time* t and
- at *location* (x, y)

is modeled by a non-homogeneous Poisson process with intensity function

$$\lambda(t, x, y, m) = f_0(m)\, R_0(t, x, y) \tag{1.1}$$

where

$$f_0(m) = \beta\, e^{-\beta(m-M_c)} \tag{1.2}$$

is the probability density function (pdf) of the frequency-magnitude distribution (FMD) following the Gutenberg-Richter law with parameter β (M_c denotes the pre-defined cut-off magnitude).

The overall event occurrence rate at time t and at location (x, y), denoted by $R_0(t, x, y)$ in Equ. (1.1), is modeled by a superposition of the (time-invariant) *seismic background rate*

$$\mu\, u(x, y)$$

and a sum of the trigger rate contributions $R_0^{trig}(t, x, y, i)$ of all events i that occurred prior to current time t, i.e. with $t_i < t$:

$$R_0(t, x, y) = \mu\, u(x, y) + \sum_{i:t_i<t} R_0^{trig}(t, x, y, i) \tag{1.3}$$

More precisely, the trigger rate contribution of a past event i to the current time t and location (x, y) is modeled by the product of the expected number of offsprings (*aftershock productivity*) of event i,

$$A\, e^{\alpha(m_i-M_c)},$$

and the *temporal* and *spatial trigger functions*

$$g_{c,p}(t - t_i) \qquad \textit{and} \qquad f_{D,\gamma,q}(r_i(x, y), m_i, l_i),$$

modeling the distribution of (relative) occurrence times/ locations of aftershocks triggered by event i. The precise inputs and shapes of the temporal and spatial trigger functions can assume different forms, and are discussed in detail in later chapters. The spatial seismic background rate $u(x, y)$ has no functional form, but rather is a set of values on a grid.
We obtain

$$R_0^{trig}(t, x, y, i) := A\, e^{\alpha(m_i-M_c)}\, g_{c,p}(t - t_i)\, f_{D,\gamma,q}(r_i(x, y), m_i, l_i).$$

to be plugged into Equ. 1.3. Thus, in total, the Standard ETAS model consists of the eight parameters

- β (frequency-magnitude distribution),
- μ (seismic background distribution),
- $A\,,\alpha$ (aftershock productivity),
- $c\,,p$ (temporal trigger function) and
- $D\,,\gamma\,,q$ (spatial trigger function).

1.2 ETAS-Incomplete Model

The *ETAS-Incomplete* model is based on the same "true" frequency-magnitude distribution $f_0(m)$ and spatio-temporal event rate $R_0(t, x, y)$ as introduced for the *Standard ETAS* model case.
However, the main idea of the incompleteness model is that these true (physical) relationships are not accurately identifiable in observed earthquake records due to missing events as a result of the bias of temporal record incompleteness. This incompleteness typically stems from overlapping coda-waves after large main-shock events (so-called *short-term aftershock incompleteness, STAI*), but can also be the result of intense seismic swarm activity.
Fitting the "true" relationships to incomplete data records may therefore lead to significantly biased parameter estimates.

Thus, the approach of the ETAS-Incomplete model is to adapt both the frequency-magnitude distribution f and event rate R to account for record incompleteness and fit these "apparent" relationships to the earthquake catalogs.

Following the *ETASI* model of Hainzl (2021), we define the "apparent" temporal frequency-magnitude distribution

$$\begin{aligned} f(m,t) &= f_0(m)\, N_0(t)\, \frac{e^{-N_0(t)\, e^{-\beta(m-M_c)}}}{1-e^{-N_0(t)}} \\ &= \beta\, e^{-\beta(m-M_c)}\, N_0(t)\, \frac{e^{-N_0(t)\, 10^{-b(m-M_c)}}}{1-e^{-N_0(t)}} \end{aligned}$$

and the "apparent" spatio-temporal event occurrence rate

$$R(t,x,y) = \frac{R_0(t,x,y)}{N_0(t)} \left(1 - e^{-N_0(t)}\right).$$

In both functions, $N_0(t)$ denotes the expected number of events occurring within the small blind time window $[t - T_b, t]$ in the entire target space, i.e.

$$N_0(t) = \int_{t-T_b}^{t} \iint_S R_0(t,x,y)\, dx\, dy\, dt \approx T_b \iint_S R_0(t,x,y)\, dx\, dy.$$

The approximation holds under the assumption that the event rate is approximately constant during the blind time.

In contrast to the Standard ETAS model, the ETAS rate is defined as the product of the apparent FMD and the apparent spatio-temporal event rate, rather than their "true" analogons. By canceling out terms, we obtain the ETAS rate

$$\lambda(t,x,y,m) = f(m,t)\, R(t,x,y) \approx \beta\, e^{-\beta(m-M_c)}\, R_0(t,x,y)\, e^{-N_0(t)\, e^{-\beta(m-M_c)}}$$

1.3 Temporal trigger function

1.3.1 Design and Notation

The *spatial trigger function* $g_{c,p}(t - t_i)$ models the distribution of the (relative) occurrence times $t - t_i$ of direct aftershocks triggered by event i. In both ETAS model versions, we use the *Omori law* to specify the temporal decay of aftershock intensity by defining

$$g_{c,p}(t - t_i) = (t - t_i + c)^{-p}, \qquad t - t_i > 0$$

with parameters c and p.

1.3.2 Temporal restriction T

We may want to restrict temporal triggering to a maximum number of days after the mainshock. In that case, we would set the Omori law equal to 0 for $t - t_i > T$. Howevr, this functionality is currently not implemented.

1.3.3 Temporal integral $\int_{T_1}^{T_2} g_{c,p}(t - t_i)\,dt$

It holds

$$\begin{aligned}
\int_{T_1}^{T_2} g_{c,p}(t - t_i)\,dt &= \int_{T_1}^{T_2} (t - t_i + c)^{-p}\,dt \\
&= \frac{1}{1-p}\left[(t - t_i + c)^{1-p}\right]_{T_1}^{T_2} \\
&= \frac{1}{1-p}\left((T_2 - t_i + c)^{1-p} - ((T_1 - t_i)_{\geq 0} + c)^{1-p}\right).
\end{aligned}$$

1.4 Spatial trigger function (Spatial Kernel)

1.4.1 Design and Notation

The *spatial kernel* $f_{D,\gamma,q}(r_i, m_i, l_i)$ models the distribution of the 2-dimensional locations (x,y) at which direct aftershocks of event i occur. We introduce here a novel design of the spatial kernel, that works in both the Standard and ETAS-Incomplete model and that may generally assume two shapes,

$$f_{D,\gamma,q}(r_i(x,y), m_i, l_i) := \begin{cases} \frac{q-1}{D\,\exp(\gamma(m_i - M_c))}\left(1 + \frac{\pi\, r_i(x,y)^2}{D\,\exp(\gamma(m_i - M_c))}\right)^{-q} & (isotropic) \\ \frac{q-1}{D\,\exp(\gamma(m_i - M_c))}\left(1 + \frac{2\, l_i\, r_i(x,y) + \pi\, r_i(x,y)^2}{D\,\exp(\gamma(m_i - M_c))}\right)^{-q} & (anisotropic). \end{cases} \tag{1.4}$$

Herein, the inputs of f are defined as follows:

- $r_i(x, y) := dist(x, y, i)$ is the point-to-point distance between a potential aftershock location (x, y) and the coordinates (x_i, y_i) of the triggering event i (isotropic case) or the shortest point-to-line distance of (x, y) to the estimated rupture segment of triggering event i (anisotropic case),
- m_i is the magnitude of the triggering event i and
- l_i is the rupture length of the triggering event i (only needed in the anisotropic case).

The kernel is constrained by the parameters D and γ that control the magnitude-dependent width of the kernel, and parameter q that describes the exponential decay of the function. The kernel is constructed in a way that they serve as a pdf, with

$$\int_0^\infty f(r, m_i, l_i)\, dr = \int_0^\infty \int_0^\infty f_{D,\gamma,q}(r_i(x, y), m_i, l_i)\, dx\, dy = 1.$$

Note that the isotropic kernel is a special case of the anisotropic one, if rupture length $l_i = 0$, i.e. if no extension of the rupture along a line segment is assumed. Denoting

$$E_\gamma(m_i) := \exp(\gamma(m_i - M_c)),$$

we can therefore work with the more general anisotropic function

$$f(r_i(x, y), m_i, l_i) = \frac{q-1}{D\, E_\gamma(m_i)} \left(1 + \frac{2\, l_i\, r_i(x, y)\; +\; \pi\, r_i(x, y)^2}{D\, E_\gamma(m_i)}\right)^{-q}.$$

1.4.2 Spatial restriction R

Following Grimm et al (2021), it may be useful to restrict the spatial kernel to a maximum distance $r_i(x, y) \leq R$. In order to retain a pdf that integrates to 1 over the restricted space, $f(r_i(x, y), m_i, l_i)$ needs to be normalized by the integral till distance R,

$$1 - \left(1 + \frac{2\, l_i\, R\; +\; \pi\, R^2}{D\, \exp(\gamma(m_i - M_c))}\right)^{1-q},$$

and we obtain the final kernel

$$f_{D,\gamma,q}(r_i(x, y), m_i, l_i) = \begin{cases} \frac{q-1}{D\, E_\gamma(m_i)}\; \frac{f^{inn}_{D,\gamma}(r_i(x,y), m_i, l_i)^{-q}}{1 - f^{inn}_{D,\gamma}(R, m_i, l_i)^{1-q}} & (r \leq R) \\ 0 & (r > R) \end{cases}$$

with the *inner function*

$$f^{inn}_{D,\gamma}(r_i(x, y), m_i, l_i) = 1 + \frac{2\, l_i\, r_i(x, y)\; +\; \pi\, r_i(x, y)^2}{D\, E_\gamma(m_i)}.$$

1.4.3 Spatial integral $\iint_{S_i(\tilde{r})} f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy$

Both ETAS model versions require the computation of the integral of the "true" event rate $R_0(t,x,y)$ over the target space region S,

$$\iint_S R_0(t,x,y)\, dx\, dy = \mu \iint_S u(x,y)\, dx\, dy + \sum_{i:t_i<t} \iint_S R_0^{trig}(t,x,y,i)\, dx\, dy$$

$$= \mu \iint_S u(x,y)\, dx\, dy + \sum_{i:t_i<t} A\, e^{\alpha(m_i - M_c)}\, g_{c,p}(t-t_i) \iint_S f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy$$

The integral of the spatial background, $\iint_S u(x,y)\, dx\, dy$, is always computed numerically, since $u(x,y)$ has no functional expression.
The integral of the spatial kernel,

$$\iint_S f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy,$$

can be computed analytically for specific spatial areas $S_i(\tilde{r})$ covering all points $(x,y) \in \mathbb{R}^2$ up to a fixed distance $\tilde{r} \geq 0$ to the point coordinates (x_i, y_i) (isotropic case) or the rupture segment (anisotropic case) of a triggering event i:

$$S_i(\tilde{r}) := \{(x,y) \in \mathbb{R}^2 \,|\, r_i(x,y) \leq \tilde{r}\}.$$

Note that this area describes a circle in the case of an isotropic spatial kernel, and a box parallel to the rupture segment, closed by semi-circles on both sides, in the case of the anisotropic spatial kernel.

Due to identical values of $f_{D,\gamma,q}(r_i(x,y), m_i, l_i)$ along the (isotropic or anisotropic) contour lines with constant distance $r_i(x,y)$, one can convert the two-dimensional integral into a one-dimensional one integrating over the distance:

$$\iint_{S_i(\tilde{r})} f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy = \int_0^{\tilde{r}} (2\,\pi\, r + 2\, l_i)\, f_{D,\gamma,q}(r, m_i, l_i)\, dr$$

where the factor $2\,\pi\, r + 2\, l_i$ is the length of the isotropic ($l_i = 0$) or anisotropic ($l_i > 0$; two sides parallel to rupture segment + two semi-circular closings) contour lines for a given distance r to the trigger point source (isotropic case) or triggering rupture segment (anisotropic case).

Using the substitution rule for integrals with

$$u := u(r) := f_{inn}(r, m_i, l_i) = 1 + \frac{2\, l_i\, r \,+\, \pi\, r^2}{D\, E_\gamma(m_i)}$$

$$\frac{d}{d\,r} u(r) = \frac{2\,\pi\, r + 2\, l_i}{D\, E_\gamma(m_i)}$$

in step (*) we obtain for the spatial integral (up to a distance smaller or equal to the spatial restriction, i.e. $\tilde{r} \leq R$)

$$\begin{aligned}
&\int_0^{\tilde{r}} (2\,\pi\,r + 2\,l_i)\, f_{D,\gamma,q}(r, m_i, l_i)\, dr \\
&\quad = \int_0^{\tilde{r}} \frac{(q-1)\,(2\,\pi\,r + 2\,l_i)}{D\,E_\gamma(m_i)} \, \frac{f_{inn}(r_i, m_i, l_i)^{-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}}\, dr \\
&\quad \overset{(*)}{=} \int_0^{\tilde{r}} \frac{(q-1)\,(2\,\pi\,r + 2\,l_i)}{D\,E_\gamma(m_i)} \, \frac{u^{-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}} \, \frac{du}{(2\,\pi\,r + 2\,l_i)\,/(D\,E_\gamma(m_i))} \\
&\quad = \int_{u(0)}^{u(\tilde{r})} (q-1)\, \frac{u^{-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}}\, du \\
&\quad = \left[\frac{q-1}{1-q} \, \frac{u^{1-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}} \right]_{u(0)}^{u(\tilde{r})} \\
&\quad = \left[-\, \frac{f_{inn}(r, m_i, l_i)^{1-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}} \right]_0^{\tilde{r}} \\
&\quad = -\, \frac{f_{inn}(\tilde{r}, m_i, l_i)^{1-q} - 1}{1 - f_{inn}(R, m_i, l_i)^{1-q}} \\
&\quad = \frac{1 - f_{inn}(\tilde{r}, m_i, l_i)^{1-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}}.
\end{aligned}$$

If we integrate up to the spatial restriction distance R, i.e. $\tilde{r} = 1$, follows

$$\iint_{S_i(R)} f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy = \frac{1 - f_{inn}(R, m_i, l_i)^{1-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}} = 1.$$

confirming that f is a pdf.

Please note:
For the rest of this manuscript, we will assume that all spatial integrals can be solved analytically over the specific region $S_i(\tilde{r})$ and will use the simplified notation S for any spatial constraints.

For more general spatial windows S, the area can be partitioned into many narrow pieces between any event and the target window boundary. The integral over each of these segments can be well approximated by the accordingly weighted analytical integral as derived above. The according methodology is called "radial partitioning" and is described in more detail by Jalilian (2019).

Chapter 2

Derivation of Log-Likelihood Functions

2.1 Conventional ETAS Model

The expected number of events occurring in the entire space-time-magnitude target window $[T_1, T_2] \times S \times [M_c, \infty)$ is

$$\begin{aligned}
&\int_{M_c}^{\infty} \int_{T_1}^{T_2} \iint_S \lambda(t, x, y, m)\, dx\, dy\, dt\, dm \\
&= \int_{M_c}^{\infty} f_0(m) dm \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt \qquad (2.1) \\
&= \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt
\end{aligned}$$

since $\int_{M_c}^{\infty} f_0(m) dm = 1$ holds by construction of the PDF.

Thus, the probability of observing $N = n$ events in the remaining space-time target window, assuming a Poisson distribution with expected number

$$\mathbb{E}(N) = \lambda_{[T_1,T_2]\times S} := \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt$$

is

$$P(N = n) = e^{-\lambda_{[T_1,T_2]\times S}} \frac{\left(\lambda_{[T_1,T_2]\times S}\right)^n}{n!}.$$

Furthermore the probability density of a specific event i with (t_i, x_i, y_i, m_i) is

$$d_i = \frac{\lambda(t_i, x_i, y_i, m_i)}{\lambda_{[T_1,T_2]\times S}}.$$

Due to independence of observations, and $n!$ possible timely reorders of an observed event sample $\tilde{E} = \{(t_i, x_i, y_i, m_i)|i = 1, ..., n\}$, the likelihood of getting the very observed sample $\tilde{E}$ is

$$\begin{aligned}
L(\theta) &= n!\ P(n)\ \prod_{i=1}^{n}\ d_i \\
&= n!\ e^{-\lambda_{[T_1,T_2]\times S}}\ \frac{\left(\lambda_{[T_1,T_2]\times S}\right)^n}{n!}\ \prod_{i=1}^{n} \frac{\lambda(t_i, x_i, y_i, m_i)}{\lambda_{[T_1,T_2]\times S}} \\
&= e^{-\lambda_{[T_1,T_2]\times S}}\ \prod_{i=1}^{n} \lambda(t_i, x_i, y_i, m_i) \\
&= e^{-\int_{T_1}^{T_2} \iint_S R_0(t,x,y)\, dx\, dy\, dt}\ \prod_{i=1}^{n} f_0(m_i)\, R_0(t_i, x_i, y_i).
\end{aligned}$$

with $\theta = \{\beta, \mu, A, \alpha, c, p, D, \gamma, q\}$.

Given an earthquake catalog with N recorded events in the target time-space window, the log-likelihood function evaluates to

$$\begin{aligned}
LL(\theta) = ln\,(L(\theta)) &= \sum_{j=1}^{N} ln\,(f_0(m_i)\, R_0(t_i, x_i, y_i)) - \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt \\
&= LL_0(\beta) + LL_1(\theta) - LL_2(\theta)
\end{aligned} \tag{2.2}$$

with

$$\begin{aligned}
LL_0(\beta) &= \sum_{j=1}^{N} ln\,(f_0(m_j)) \\
LL_1(\theta) &= \sum_{j=1}^{N} ln\,(R_0(t_j, x_j, y_j)) \\
&= \sum_{j=1}^{N} ln \left(\mu\, u(x_j, y_j) + \sum_{i:t_i<t_j} A\, e^{\alpha(m_i - M_c)}\ g_{c,p}(t_j - t_i)\, f_{D,\gamma,q}(r_i(x_j, y_j), m_i, l_i) \right) \\
LL_2(\theta) &= \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt \\
&= \mu\,(T_2 - T_1) \iint_S u(x, y)\, dx\, dy \\
&+ \sum_{i:t_i<t} A\, e^{\alpha(m_i - M_c)} \left(\frac{(T_2 - t_i + c)^{1-p} - ((T_1 - t_i)_{\geq 0} + c)^{1-p}}{1-p} \right) \left(\frac{1 - f_{inn}(\tilde{r}, m_i, l_i)^{1-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}} \right)
\end{aligned} \tag{2.3}$$

Note that, while the outer sum in LL_0 and LL_1 only includes target events (i.e. events $j = 1, ..., N$ with $(t_j, x_j, y_j, m_j) \in [t_0, t_1] \times S \times [m_0, \infty))$, the sum within $R_0(t,x,y)$ may include trigger contributions from complementary events outside of the space-time-magnitude target window.

The optimal set of model parameters θ is found by maximizing $LL(\theta)$. For the use of gradient-based solvers, we need to know the derivatives of $LL(\theta)$ by the respective parameters. Derivations are shown in later chapters.

2.2 ETAS-Incomplete Model

In the ETAS-Incomplete model, it is not as easily seen that we can waive the magnitude part in the integral of the time-space-magnitude intensity function $\lambda(t,x,y,m)$ over the respective target ranges $[T_1,T_2]\times S\times[M_c,\infty)$ such as

$$\int_{M_c}^{\infty}\int_{T_1}^{T_2}\iint_S \lambda(t,x,y,m)\,dx\,dy\,dt\,dm = \int_{T_1}^{T_2}\iint_S R(t,x,y)\,dx\,dy\,dt.$$

Indeed, it holds

$$\begin{aligned}
&\int_{M_c}^{\infty}\int_{T_1}^{T_2}\iint_S \lambda(t,x,y,m)\,dx\,dy\,dt\,dm \\
&\quad= \int_{M_c}^{\infty}\int_{T_1}^{T_2}\iint_S \beta\,e^{-\beta(m-M_c)}\,R_0(t,x,y)\,e^{-N_0(t)\,e^{-\beta(m-M_c)}}\,dx\,dy\,dt\,dm \\
&\quad= \int_{T_1}^{T_2}\left(\int_{M_c}^{\infty}\beta\,e^{-\beta(m-M_c)}\,e^{-N_0(t)\,e^{-\beta(m-M_c)}}\,dm\right)\left(\iint_S R_0(t,x,y)\,dx\,dy\right)dt
\end{aligned}$$

with

$$\int_{M_c}^{\infty}\beta\,e^{-\beta(m-M_c)}\,e^{-N_0(t)\,e^{-\beta(m-M_c)}}\,dm = \left[\frac{e^{-N_0(t)\,e^{-\beta(m-M_c)}}}{N_0(t)}\right]_{M_c}^{\infty} = \frac{1-e^{-N_0(t)}}{N_0(t)}$$

and consequently, using the aproximation

$$N_0(t) = \int_{t-T_b}^{t}\iint_S R_0(\tilde{t},x,y)\,dx\,dy\,d\tilde{t} \approx T_b\iint_S R_0(t,x,y)dxdy$$

the integral evaluates to

$$\begin{aligned}
\int_{M_c}^{\infty}\int_{T_1}^{T_2}\iint_S \lambda(t,x,y,m)\,dx\,dy\,dt\,dm &= \int_{T_1}^{T_2}\iint_S R(t,x,y)\,dx\,dy\,dt \\
&= \int_{T_1}^{T_2}\frac{1-e^{-N_0(t)}}{N_0(t)}\iint_S R_0(t,x,y)\,dx\,dy\,dt \\
&= \int_{T_1}^{T_2}\frac{1-e^{-T_b\iint_S R_0(t,x,y)\,dx\,dy}}{T_b\iint_S R_0(t,x,y)\,dx\,dy}\iint_S R_0(t,x,y)\,dx\,dy\,dt \\
&= \int_{T_1}^{T_2}\frac{1-e^{-T_b\iint_S R_0(t,x,y)\,dx\,dy}}{T_b}dt
\end{aligned}$$

Analogously to the Standard ETAS model case, we therefore obtain the likelihood function

$$L(\theta) = e^{-\int_{M_c}^{\infty} \int_{T_1}^{T_2} \iint_S \lambda(t,x,y,m)\, dx\, dy\, dt\, dm} \prod_{i=1}^{n} \lambda(t_i, x_i, y_i, m_i)$$

and the log-likelihood function

$$LL(\theta) = LL_1(\theta) - LL_2(\theta) \tag{2.4}$$

with

$$\begin{aligned}
LL_1(\theta) &= \sum_{j=1}^{N} ln\left(f(m_j, t_j)\, R(t_j, x_j, y_j)\right) \\
&= \sum_{j=1}^{N} ln\left(\beta\, e^{-\beta(m_j - M_c)}\, R_0(t_j, x_j, y_j)\, e^{-N_0(t_j)\, e^{-\beta(m_j - M_c)}}\right) \\
&= N\, ln(\beta) + \sum_{j=1}^{N} \left(ln(R_0(t_j, x_j, y_j)) - \beta(m_j - M_c) - N_0(t_j)\, e^{-\beta(m_j - M_c)}\right) \\
LL_2(\theta) &= \int_{T_1}^{T_2} \frac{1 - e^{-T_b \iint_S R_0(t,x,y)\, dx\, dy}}{T_b} dt \\
&= \frac{T_2 - T_1}{T_b} - \frac{1}{T_b} \int_{T_1}^{T_2} e^{-T_b \iint_S R_0(t,x,y)\, dx\, dy} dt
\end{aligned} \tag{2.5}$$

and

$$\begin{aligned}
\iint_S R_0(t, x, y)\, dx\, dy = \mu \iint_S u(x, y)\, dx\, dy \\
+ \sum_{i: t_i < t} A\, e^{\alpha(m_i - M_c)}\, g_{c,p}(t - t_i) \iint_S f_{D,\gamma,q}(r_i(x, y), m_i, l_i)\, dx\, dy
\end{aligned}$$

Note that the parameter vector $\theta = \{\beta, T_b, \mu, A, \alpha, c, p, D, \gamma, q\}$ comprises the additional parameter T_b. The sum in LL_1 is again computed over all target events (i.e. events j with $(t_j, x_j, y_j, m_j) \in [t_0, t_1] \times S \times [m_0, \infty)$), whereas the integral in LL_2 may include trigger contributions from complementary events outside of the space-time-magnitude target window.

The optimal set of model parameters θ is found by maximizing $LL(\theta)$. Derivatives of $LL(\theta)$ by the respective parameters, as needed for the use of gradient-based solvers, are derived in later chapters.

Chapter 3

Derivatives of LL-Functions (Conventional ETAS)

According to equations Equ. 2.2, 2.3, the LL-function of the *Standard-ETAS* model is decomposed into the three summands

$$LL(\theta) = LL_0(\beta) + LL_1(\theta) - LL_2(\theta)$$

where

$$LL_0(\beta) = \sum_{j=1}^{N} ln\left(f_0(m_j)\right)$$

only depends on the Gutenberg-Richter parameter β and

$$LL_1(\theta) = \sum_{j=1}^{N} ln\left(R_0(t_j, x_j, y_j)\right)$$

$$LL_2(\theta) = \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt$$

depend on the remaining eight parameters $\theta = \{\mu, A, \alpha, c, p, D, \gamma, q\}$.

3.1 Analytical Estimator for β

The Gutenberg-Richter parameter β appears only separated from the other parameters in $LL_0(\beta)$ and can therefore be optimized analytically by maximizing

$$LL_0(\beta) = \sum_{j=1}^{N} ln\left(f_0(m_j)\right) = N\, ln(\beta) - \beta \sum_{i=1}^{N} (m_i - M_c).$$

Solving the derivative

$$\frac{d}{d\beta} LL_0(\beta) = \frac{N}{\beta} - \sum_{i=1}^{N} (m_i - M_c) = 0$$

leads to the estimator

$$\beta = \frac{N}{\sum_{i=1}^{N}(m_i - M_c)}.$$

3.2 Derivatives of $LL_1(\theta)$

Via chain rule, any derivative of $LL_1(\theta)$ has the form

$$\frac{d\,LL_1(\theta)}{d\,\theta_k} = \sum_{j=1}^{N} \frac{\frac{d}{d\,\theta_k}\,R_0(t_j,x_j,y_j)}{R_0(t_j,x_j,y_j)}$$

with $\theta_k \in \theta \quad (k = 1, ..., 8)$ and

$$\begin{aligned}\frac{d}{d\,\theta_k}\,R_0(t_j,x_j,y_j) &= \frac{d}{d\,\theta_k}\,\mu\,u(x_j,y_j) + \sum_{i:t_i<t_j} \frac{d}{d\,\theta_k}\,R_0^{trig}(t_j,x_j,y_j,i)\\ R_0^{trig}(t_j,x_j,y_j,i) &= A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t_j-t_i)\,f_{D,\gamma,q}(r_i(x_j,y_j),m_i,l_i)\end{aligned}$$

3.2.1 By μ

It holds

$$\frac{d}{d\,\mu}\,R_0(t_j,x_j,y_j) = u(x_j,y_j).$$

3.2.2 By A

It holds

$$\begin{aligned}\frac{d}{d\,A}\,R_0(t_j,x_j,y_j) &= \sum_{i:t_i<t_j} \frac{d}{d\,A}\,R_0^{trig}(t_j,x_j,y_j,i)\\ &= \sum_{i:t_i<t_j} \frac{R_0^{trig}(t_j,x_j,y_j,i)}{A}\end{aligned}$$

3.2.3 By α

It holds

$$\begin{aligned}\frac{d}{d\,\alpha}\,R_0(t_j,x_j,y_j) &= \sum_{i:t_i<t_j} \frac{d}{d\,\alpha}\,R_0^{trig}(t_j,x_j,y_j,i)\\ &= \sum_{i:t_i<t_j} (m_i-M_c)\,R_0^{trig}(t_j,x_j,y_j,i)\end{aligned}$$

3.2.4 By c

It holds

$$\begin{aligned}\frac{d}{d\,c}\,g(t_j-t_i) &= \frac{d}{d\,c}\,(c+t_j-t_i)^{-p}\\ &= (-p)\,(c+t_j-t_i)^{-p-1}\\ &= \frac{-p}{c+t_j-t_i}\,g(t_j-t_i)\end{aligned}$$

and therefore

$$\begin{aligned}\frac{d}{d\,c}\,R_0(t_j,x_j,y_j) &= \sum_{i:t_i<t_j}\frac{d}{d\,c}\,R_0^{trig}(t_j,x_j,y_j,i)\\ &= \sum_{i:t_i<t_j}\frac{-p}{c+t_j-t_i}\,R_0^{trig}(t_j,x_j,y_j,i)\end{aligned}$$

3.2.5 By p

It holds

$$\begin{aligned}\frac{d}{d\,p}\,g(t_j-t_i) &= \frac{d}{d\,p}\,(c+t_j-t_i)^{-p}\\ &= \frac{d}{d\,p}\,e^{ln\left((c+t_j-t_i)^{-p}\right)}\\ &= \frac{d}{d\,p}\,e^{(-p)\,ln(c+t_j-t_i)}\\ &= -ln(c+t_j-t_i)\,g(t_j-t_i)\end{aligned}$$

and therefore

$$\begin{aligned}\frac{d}{d\,p}\,R_0(t_j,x_j,y_j) &= \sum_{i:t_i<t_j}\frac{d}{d\,p}\,R_0^{trig}(t_j,x_j,y_j,i)\\ &= \sum_{i:t_i<t_j}-ln(c+t_j-t_i)\,R_0^{trig}(t_j,x_j,y_j,i)\end{aligned}$$

3.2.6 Notations for Spatial Kernel

For better notation in long formula derivations, we decompose the spatial kernel

$$f_{D,\gamma,q}(r,m,l) = \frac{q-1}{D\,E_\gamma(m_i)}\,\frac{f_{D,\gamma}^{inn}(r,m,l)^{-q}}{1-f_{D,\gamma}^{inn}(R,m,l)^{1-q}}$$

into the numerator term

$$t_{numer} := \frac{q-1}{D\,E_\gamma(m)}\,f_{D,\gamma}^{inn}(r,m,l)^{-q}$$

and the denominator term

$$t_{denom} := 1-f_{D,\gamma}^{inn}(R,m,l)^{1-q},$$

i.e.

$$f_{D,\gamma,q}(r,m,l) = \frac{t_{numer}}{t_{denom}}.$$

The derivative of $f_{D,\gamma,q}(r,m,l)$ by any of the three spatial parameters D,γ,q is computed via the quotient rule

$$(f_{D,\gamma,q}(r,m,l))' = \frac{t'_{numer}\,t_{denom}-t_{numer}\,t'_{denom}}{t_{denom}^2}.$$

Also, in any spatial derivative we assume a distance smaller or equal to the spatial extent, i.e. $r_i(x,y) > R$ (otherwise the spatial kernel is 0, as its derivatives).

3.2.7 By D

Having the inner derivative

$$\frac{d}{dD} f_{D,\gamma}^{inn}(r,m,l) = -\frac{1}{D}\frac{\pi r^2}{D E_\gamma(m)} = -\frac{1}{D}\ \left(f_{D,\gamma}^{inn}(r,m,l) - 1\right)$$

we obtain by the use of product and chain rule in step (*)

$$\begin{aligned}
\frac{d}{dD} t_{numer} &= \frac{q-1}{E_\gamma(m)}\ \ \frac{d}{dD}\left(\frac{1}{D}\ f_{D,\gamma}^{inn}\,(r,m)^{-q}\right) \\
&\overset{(*)}{=} \frac{q-1}{E_\gamma(m)}\Bigg[-\frac{1}{D^2}\ f_{D,\gamma}^{inn}\,(r,m)^{-q} + \ \ldots \\
&\quad \frac{1}{D}\ (-q)\ f_{D,\gamma}^{inn}\,(r,m)^{-q-1}\left(-\frac{1}{D}\right)\ \left(f_{D,\gamma}^{inn}\,(r,m) - 1\right)\Bigg] \\
&= \left[\frac{q-1}{D E_\gamma(m)}\ f_{D,\gamma}^{inn}\,(r,m)^{-q}\right]\frac{1}{D}\left(-1 + q\ f_{D,\gamma}^{inn}\,(r,m)^{-1}\ \ \left(f_{D,\gamma}^{inn}\,(r,m) - 1\right)\right) \\
&= \left(\frac{q-1}{D\, E_\gamma(m_i)}\ f_{D,\gamma}^{inn}(r_i,m_i,l_i)^{-q}\right)\ \frac{1}{D}\ \ \left(q\ \ \left(1 - f_{D,\gamma}^{inn}\,(r,m)^{-1}\right) - 1\right)
\end{aligned}$$

and by the use of chain rule for the normalization term

$$\begin{aligned}
\frac{d}{dD} t_{denom} &= -(1-q)\ f_{D,\gamma}^{inn}\,(R,m)^{-q}\left(-\frac{1}{D}\right)\ \ \left(f_{D,\gamma}^{inn}\,(R,m) - 1\right) \\
&= \frac{1-q}{D}\ f_{D,\gamma}^{inn}\,(R,m)^{-q}\left(f_{D,\gamma}^{inn}\,(R,m) - 1\right).
\end{aligned}$$

By quotient rule, we can now conclude

$$\begin{aligned}
&\frac{d}{dD} f_{D,\gamma,q}(x,y,i) \\
&= \Bigg[t_{numer}\ \frac{1}{D}\ \ \left(q\ \ \left(1 - f_{D,\gamma}^{inn}\,(r,m)^{-1}\right) - 1\right)\ t_{denom}\ \cdots \\
&\qquad - t_{numer}\ \frac{1-q}{D}\ f_{D,\gamma}^{inn}\,(R,m)^{-q}\left(f_{D,\gamma}^{inn}\,(R,m) - 1\right)\Bigg]\ \ /\ t_{denom}^2 \\
&= \frac{t_{numer}}{t_{denom}}\ \frac{1}{D}\ \left[\left(q\ \ \left(1 - f_{D,\gamma}^{inn}\,(r,m)^{-1}\right) - 1\right) - \frac{(1-q)\ \ \left(f_{D,\gamma}^{inn}\,(R,m) - 1\right)}{t_{denom}\ f_{D,\gamma}^{inn}\,(R,m)^{q}}\right] \\
&= f_{D,\gamma,q}(x,y,i)\ \frac{1}{D}\ \left[\left(q\ \ \left(1 - f_{D,\gamma}^{inn}\,(r,m)^{-1}\right) - 1\right) - \frac{(1-q)\ \ \left(f_{D,\gamma}^{inn}\,(R,m) - 1\right)}{t_{denom}\ f_{D,\gamma}^{inn}\,(R,m)^{q}}\right]
\end{aligned}$$

Therefore, it holds

$$\frac{d}{d\,D}\,R_0(t_j,x_j,y_j) = \sum_{i:t_i<t_j}\frac{d}{d\,D}\,R_0^{trig}(t_j,x_j,y_j,i) = \sum_{i:t_i<t_j}\frac{1}{D}\,h_D\,R_0^{trig}(t_j,x_j,y_j,i)$$

with

$$h_D := q\ \left(1 - f^{inn}_{D,\gamma}\,(r,m)^{-1}\right) - 1 - \frac{(1-q)\ \left(f^{inn}_{D,\gamma}\,(R,m) - 1\right)}{t_{denom}\ f^{inn}_{D,\gamma}\,(R,m)^q}.$$

3.2.8 By γ

Having the inner derivative

$$\frac{d}{d\gamma} f^{inn}_{D,\gamma}\,(r,m) = (-(m-m_0))\ \frac{\pi r^2}{D E_\gamma(m)} = (-(m-m_0))\ \left(f^{inn}_{D,\gamma}\,(r,m) - 1\right) \tag{3.1}$$

we obtain by the use of product and chain rule in step (*)

$$\begin{aligned}
&\frac{d}{d\gamma} t_{numer} = \frac{q-1}{D}\ \ \frac{d}{d\gamma}\left(\frac{1}{E_\gamma(m)}\ f^{inn}_{D,\gamma}\,(r,m)^{-q}\right)\\
&\overset{(*)}{=} \frac{q-1}{D}\left[\frac{-(m-m_0)}{E_\gamma(m)}\ f^{inn}_{D,\gamma}\,(r,m)^{-q}\ \ldots\right.\\
&\qquad \left. + \frac{1}{E_\gamma(m)}\ (-q)\ f^{inn}_{D,\gamma}\,(r,m)^{-q-1}\,(-(m-m_0))\ \left(f^{inn}_{D,\gamma}\,(r,m) - 1\right)\right]\\
&= \left[\frac{q-1}{D E_\gamma(m)}\ f^{inn}_{D,\gamma}\,(r,m)^{-q}\right](m-m_0)\left(-1 + q\ f^{inn}_{D,\gamma}\,(r,m)^{-1}\ \ \left(f^{inn}_{D,\gamma}\,(r,m) - 1\right)\right)\\
&= t_{numer}\ (m-m_0)\ \left(q\ \left(1 - f^{inn}_{D,\gamma}\,(r,m)^{-1}\right) - 1\right) \qquad (3.2)
\end{aligned}$$

and by the use of chain rule

$$\begin{aligned}
\frac{d}{d\gamma} t_{denom} &\overset{(*)}{=} -(1-q)\ f^{inn}_{D,\gamma}\,(R,m)^{-q}\,(-(m-m_0))\ \left(f^{inn}_{D,\gamma}\,(r,m) - 1\right)\\
&= (1-q)\ (m-m_0)\ f^{inn}_{D,\gamma}\,(R,m)^{-q}\left(f^{inn}_{D,\gamma}\,(R,m) - 1\right). \qquad (3.3)
\end{aligned}$$

Comparing (3.2) and (**??**) with (**??**) and (**??**), we see that

$$\begin{aligned}
\frac{d}{d\gamma} t_{numer} &= D\ (m-m_0)\ \frac{d}{dD} t_{numer},\\
\frac{d}{d\gamma} t_{denom} &= D\ (m-m_0)\ \frac{d}{dD} t_{denom}.
\end{aligned}$$

Therefore it follows that

$$\frac{d}{d\gamma} f_{D,\gamma,q}(x,y,i) = D\ (m-m_0)\ \frac{d}{dD} f_{D,\gamma,q}(x,y,i)$$

and

$$\frac{d}{d\,\gamma}\, R_0(t_j,x_j,y_j) = \sum_{i:t_i<t_j} \frac{d}{d\,D}\, R_0^{trig}(t_j,x_j,y_j,i) = \sum_{i:t_i<t_j} (m_i - m_0)\, h_D\, R_0^{trig}(t_j,x_j,y_j,i)$$

with h_D as defined in the derivation of the derivative by D.

3.2.9 By q

We obtain by the use of product rule in step (*)

$$\begin{aligned}
&\frac{d}{dq} t_{numer} = \frac{1}{DE_\gamma(m)} \quad \frac{d}{dq}\left((q-1)\; f_{D,\gamma}^{inn}\,(r,m)^{-q}\right) \\
&\overset{(*)}{=} \frac{1}{DE_\gamma(m)} \left[f_{D,\gamma}^{inn}\,(r,m)^{-q} + (q-1)\;\left(-ln\left(f_{D,\gamma}^{inn}\,(r,m)\right)\right)\; f_{D,\gamma}^{inn}\,(r,m)^{-q}\right] \\
&= \left[\frac{q-1}{DE_\gamma(m)}\; f_{D,\gamma}^{inn}\,(r,m)^{-q}\right]\left(\frac{1}{q-1} - ln\left(f_{D,\gamma}^{inn}\,(r,m)\right)\right) \\
&= t_{numer}\;\left(\frac{1}{q-1} - ln\left(f_{D,\gamma}^{inn}\,(r,m)\right)\right)
\end{aligned}$$

and by the use of chain rule

$$\begin{aligned}
\frac{d}{dq} t_{denom} &= -\left(-ln\left(f_{D,\gamma}^{inn}\,(R,m)\right)\right)\; f_{D,\gamma}^{inn}\,(R,m)^{1-q} \\
&= ln\left(f_{D,\gamma}^{inn}\,(R,m)\right)\; f_{D,\gamma}^{inn}\,(R,m)^{1-q}
\end{aligned}$$

By quotient rule, we can now conclude

$$\begin{aligned}
&\frac{d}{dq} f_{D,\gamma,q}(x,y,i) = \frac{d}{dq}\frac{t_{numer}}{t_{denom}} \\
&\quad = \Bigg[t_{numer}\;\left(\frac{1}{q-1} - ln\left(f_{D,\gamma}^{inn}\,(r,m)\right)\right)\; t_{denom}\;\ldots \\
&\quad - t_{numer}\; ln\left(f_{D,\gamma}^{inn}\,(R,m)\right)\; f_{D,\gamma}^{inn}\,(R,m)^{1-q}\Bigg]\;\; / \; t_{denom}^2 \\
&\quad = f_{D,\gamma,q}(x,y,i)\;\left[\frac{1}{q-1} - ln\left(f_{D,\gamma}^{inn}\,(r,m)\right) - \frac{ln\left(f_{D,\gamma}^{inn}\,(R,m)\right)}{t_{denom}}\; f_{D,\gamma}^{inn}\,(R,m)^{1-q}\right]
\end{aligned}$$

Therefore, it holds

$$\frac{d}{d\,q}\,R_0(t_j,x_j,y_j) = \sum_{i:t_i<t_j}\frac{d}{d\,q}\,R_0^{trig}(t_j,x_j,y_j,i) = \sum_{i:t_i<t_j} h_q\,R_0^{trig}(t_j,x_j,y_j,i)$$

with

$$h_q := \frac{1}{q-1} - ln\left(f_{D,\gamma}^{inn}\,(r,m)\right) - \frac{ln\left(f_{D,\gamma}^{inn}\,(R,m)\right)}{t_{denom}}\; f_{D,\gamma}^{inn}\,(R,m)^{1-q}\,.$$

3.3 Derivatives of $LL_2(\theta)$

The second summand of the LL-function is

$$\begin{aligned} LL_2(\theta) &= \int_{T_1}^{T_2} \iint_S R_0(t,x,y)\,dx\,dy\,dt \\ &= (T_2 - T_1)\,\mu \iint_S u(x,y)\,dx\,dy + \sum_{i=1}^{N} A\,e^{\alpha(m_i - M_c)}\,G_{c,p}(T_1,T_2,i)\,F_{D,\gamma,q}(S,i) \end{aligned}$$

with

$$\begin{aligned} G_{c,p}(T_1,T_2,i) &:= \int_{T_1}^{T_2} g(t-t_i)\,dt = \frac{1}{1-p}\left((T_2 - t_i + c)^{1-p} - ((T_1 - t_i)_{\geq 0} + c)^{1-p}\right), \\ F_{D,\gamma,q}(S,i) &:= \iint_S f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\,dx\,dy = \frac{1 - \left(1 + \frac{2\,l_i\,\tilde{r} + \pi\,\tilde{r}^2}{D\,E_\gamma(m_i)}\right)^{1-q}}{1 - \left(1 + \frac{2\,l_i\,R + \pi\,R^2}{D\,E_\gamma(m_i)}\right)^{1-q}} \end{aligned}$$

We obtain the following derivatives.

3.3.1 By μ

It holds

$$\frac{d}{d\,\mu} LL_2(\theta) = (T_2 - T_1) \iint_S u(x,y)\,dx\,dy.$$

3.3.2 By A

It holds

$$\frac{d}{d\,A} LL_2(\theta) = \sum_{i=1}^{N} e^{\alpha(m_i - M_c)}\,G_{c,p}(T_1,T_2,i)\,F_{D,\gamma,q}(S,i)$$

3.3.3 By α

It holds

$$\frac{d}{d\,A} LL_2(\theta) = \sum_{i=1}^{N} A\,(m_i - M_c)\,e^{\alpha(m_i - M_c)}\,G_{c,p}(T_1,T_2,i)\,F_{D,\gamma,q}(S,i)$$

3.3.4 By c

It holds

$$\frac{d}{d\,c} LL_2(\theta) = \sum_{i=1}^{N} A\,e^{\alpha(m_i - M_c)} \left(\frac{d}{d\,c} G_{c,p}(T_1,T_2,i)\right) F_{D,\gamma,q}(S,i)$$

with

$$\frac{d}{d\,c}G_{c,p}(T_1,T_2,i) = \frac{d}{d\,c}\,\frac{1}{1-p}\,\left((T_2 - t_i + c)^{1-p} - ((T_1 - t_i)_{\geq 0} + c)^{1-p}\right)$$
$$= (T_2 - t_i + c)^{-p} - ((T_1 - t_i)_{\geq 0} + c)^{-p}.$$

3.3.5 By p

It holds

$$\frac{d}{d\,p}\,x^{1-p} = \frac{d}{d\,p}\,e^{(1-p)\,ln(x)} = -ln(x)\,x^{1-p},$$

and via quotient rule

$$\frac{d}{d\,p}\,\frac{x^{1-p}}{1-p} = \frac{-ln(x)\,x^{1-p}\,(1-p) - x^{1-p}\,(-1)}{(1-p)^2}$$
$$= x^{1-p}\,\frac{1 - ln(x)\,(1-p)}{(1-p)^2}.$$

Therefore we obtain

$$\frac{d}{d\,p}LL_2(\theta) = \sum_{i=1}^{N} A\,e^{\alpha(m_i - M_c)}\,\left(\frac{d}{d\,p}G_{c,p}(T_1,T_2,i)\right)\,F_{D,\gamma,q}(S,i)$$

with

$$\frac{d}{d\,p}G_{c,p}(T_1,T_2,i) = \frac{d}{d\,p}\,\frac{1}{1-p}\,\left((T_2 - t_i + c)^{1-p} - ((T_1 - t_i)_{\geq 0} + c)^{1-p}\right)$$
$$= \frac{d}{d\,p}\,\frac{(T_2 - t_i + c)^{1-p}}{1-p} - \frac{((T_1 - t_i)_{\geq 0} + c)^{1-p}}{1-p}$$
$$= (T_2 - t_i + c)^{1-p}\,\frac{1 - ln(T_2 - t_i + c)\,(1-p)}{(1-p)^2} \quad ...$$
$$- ((T_1 - t_i)_{\geq 0} + c)^{1-p}\,\frac{1 - ln((T_1 - t_i)_{\geq 0} + c)\,(1-p)}{(1-p)^2}.$$

3.3.6 Notations for Spatial Integral

Again, for better notation in long formula derivations, we decompose the spatial integral

$$F_{D,\gamma,q}(S,i) = \iint_S f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\,dx\,dy = \frac{1 - \left(1 + \frac{2\,l_i\,\tilde{r} + \pi\,\tilde{r}^2}{D\,E_\gamma(m_i)}\right)^{1-q}}{1 - \left(1 + \frac{2\,l_i\,R + \pi\,R^2}{D\,E_\gamma(m_i)}\right)^{1-q}}$$

into the numerator term

$$t_{numInteg} := 1 - \left(f^{inn}_{D,\gamma}(r,m,l)\right)^{1-q}$$

and the denominator term

$$t_{denom} := 1 - \left(f^{inn}_{D,\gamma}(R,m,l)\right)^{1-q},$$

i.e.

$$F_{D,\gamma,q}(S,i) = \frac{t_{numInteg}}{t_{denom}}.$$

The derivative of $f_{D,\gamma,q}(r,m,l)$ by any of the three spatial parameters D,γ,q is computed via the quotient rule

$$(F_{D,\gamma,q}(S,i))' = \frac{t'_{numInteg}\, t_{denom} - t_{numInteg}\, t'_{denom}}{t^2_{denom}}.$$

Also, in any spatial derivative we assume a distance smaller or equal to the spatial extent, i.e. $r_i(x,y) > R$ (otherwise the spatial integral has reached 1 and the derivative is 0).

3.3.7 By D

From subsection 3.2.7 we obtain

$$\begin{aligned}
\frac{d}{dD} t_{numInteg} &= \frac{1-q}{D}\; f^{inn}_{D,\gamma}(r,m)^{-q} \left(f^{inn}_{D,\gamma}(r,m) - 1\right) \\
\frac{d}{dD} t_{scale} &= \frac{1-q}{D}\; f^{inn}_{D,\gamma}(R,m)^{-q} \left(f^{inn}_{D,\gamma}(R,m) - 1\right)
\end{aligned}$$

and consequently, via quotient rule,

$$\begin{aligned}
&\frac{d}{dD} F_{D,\gamma,q}(S,i) \\
&\quad = \Bigg[\frac{1-q}{D}\; f^{inn}_{D,\gamma}(r,m)^{-q} \left(f^{inn}_{D,\gamma}(r,m) - 1\right)\; t_{scale}\; \ldots \\
&\qquad - t_{numInteg}\; \frac{1-q}{D}\; f^{inn}_{D,\gamma}(R,m)^{-q} \left(f^{inn}_{D,\gamma}(R,m) - 1\right)\Bigg] / t^2_{scale}.
\end{aligned}$$

3.3.8 By γ

From subsection 3.2.8 we obtain

$$\begin{aligned}
\frac{d}{d\gamma} t_{numInteg} &= D\ (m - m_0)\ \frac{d}{dD} t_{numInteg}, \\
\frac{d}{d\gamma} t_{scale} &= D\ (m - m_0)\ \frac{d}{dD} t_{scale}.
\end{aligned}$$

Therefore it follows that

$$\frac{d}{d\gamma} F_{D,\gamma,q}(S,i) = D\ (m - m_0)\ \frac{d}{dD} F_{D,\gamma,q}(S,i).$$

3.3.9 By q

From subsection 3.2.9 we obtain

$$\frac{d}{dq}t_{integ} = ln\left(f_{D,\gamma}^{inn}\left(r,m\right)\right)\ f_{D,\gamma}^{inn}\left(r,m\right)^{1-q} \tag{3.4}$$

$$\frac{d}{dq}t_{scale} = ln\left(f_{D,\gamma}^{inn}\left(R,m\right)\right)\ f_{D,\gamma}^{inn}\left(R,m\right)^{1-q} \tag{3.5}$$

and consequently, via quotient rule,

$$\frac{d}{dq}F_{D,\gamma,q}(r,x,y) = \Big[ln\left(f_{D,\gamma}^{inn}\left(r,m\right)\right)\ f_{D,\gamma}^{inn}\left(r,m\right)^{1-q}\ t_{scale}\ ...$$
$$-\,t_{integ}\ ln\left(f_{D,\gamma}^{inn}\left(R,m\right)\right)\ f_{D,\gamma}^{inn}\left(R,m\right)^{1-q}\Big]/t_{scale}^2. \tag{3.6}$$

Chapter 4

Derivatives of LL-Functions (ETAS-Incomplete)

4.1 Derivatives of $N(t)$

In both LL-summands $LL_1(\theta)$ and $LL_2(\theta)$ occurs the term $N(t)$ representing the overall event rate at time t in the entire target space window,

$$\begin{aligned} N(t) &\approx T_b \iint_S R_0(t,x,y)\,dx\,dy \\ &= T_b \Big(\mu \iint_S u(x,y)\,dx\,dy + \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i)\Big). \end{aligned}$$

with

$$F_{D,\gamma,q}(S,i) = \iint_S f_{D,\gamma,q}(r_i(x,y),m_i,l_i)\,dx\,dy = \frac{1-\left(1+\frac{2\,l_i\,\tilde{r}+\pi\,\tilde{r}^2}{D\,E_\gamma(m_i)}\right)^{1-q}}{1-\left(1+\frac{2\,l_i\,R+\pi\,R^2}{D\,E_\gamma(m_i)}\right)^{1-q}}.$$

4.1.1 By μ

It holds

$$\frac{d}{d\,\mu}\,N(t) = T_b \iint_S u(x,y)\,dx\,dy.$$

4.1.2 By A

It holds

$$\frac{d}{d\,A}\,N(t) = T_b \sum_{i:t_i<t} e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i).$$

4.1.3 By α

It holds

$$\frac{d}{d\,\alpha}\,N(t) = T_b \sum_{i:t_i<t} A\,(m_i - M_c)\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i).$$

4.1.4 By c

By section 3.2, it holds

$$\begin{aligned}\frac{d}{d\,c}\,N(t) &= T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,\left(\frac{d}{d\,c}\,g_{c,p}(t-t_i)\right)\,F_{D,\gamma,q}(S,i)\\ &= T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,\frac{-p}{c+t-t_i}\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i)\end{aligned}$$

4.1.5 By p

By section 3.2, it holds

$$\begin{aligned}\frac{d}{d\,p}\,N(t) &= T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,\left(\frac{d}{d\,p}\,g_{c,p}(t-t_i)\right)\,F_{D,\gamma,q}(S,i)\\ &= T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,\left(-ln(c+t-t_i)\right)\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i)\end{aligned}$$

4.1.6 By D

It holds

$$\frac{d}{d\,D}\,N(t) = T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,\left(\frac{d}{d\,D}\,F_{D,\gamma,q}(S,i)\right)$$

with $\frac{d}{d\,D}\,F_{D,\gamma,q}(S,i)$ as derived in section 3.3.

4.1.7 By γ

It holds

$$\frac{d}{d\,\gamma}\,N(t) = T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,\left(\frac{d}{d\,\gamma}\,F_{D,\gamma,q}(S,i)\right)$$

with $\frac{d}{d\,\gamma}\,F_{D,\gamma,q}(S,i)$ as derived in section 3.3.

4.1.8 By q

It holds

$$\frac{d}{d\,q}\,N(t) = T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,\left(\frac{d}{d\,q}\,F_{D,\gamma,q}(S,i)\right)$$

with $\frac{d}{d\,q}\,F_{D,\gamma,q}(S,i)$ as derived in section 3.3.

4.1.9 By T_b

It holds

$$\frac{d}{d\,T_b}\,N(t) = \frac{N(t)}{T_b} = \mu\,\iint_S u(x,y)\,dx\,dy + \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i).$$

4.1.10 By β

The parameter β does not occur in $N(t)$, i.e. $\frac{d}{d\,\beta}\,N(t) = 0$.

4.2 Derivatives of $LL_1(\theta)$

For the *ETAS-Incomplete* model, derivatives of the first LL-summand $LL_1(\theta)$ are most easily computed starting from (see Equ. 2.5)

$$LL_1(\theta) = N\,ln(\beta) + \sum_{j=1}^{N} ln(R_0(t_j, x_j, y_j)) - \beta(m_j - M_c) - N(t_j)\, e^{-\beta(m_j - M_c)}$$

In this case, the Gutenberg-Richter parameter β is not isolated from the other nine parameters and therefore the entire parameter set $\theta = \{\beta, T_b, \mu, A, \alpha, c, p, D, \gamma, q\}$ needs to be optimized numerically. The following gradients serve for gradient-based optimization methods.

4.2.1 By $\mu, A, \alpha, c, p, D, \gamma, q$

For any parameter other than β and T_b, we obtain the derivative by chain rule as

$$\frac{d}{d\,\theta_k} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,\theta_k} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,\theta_k} N(t_j)\right)\, e^{-\beta(m_j - M_c)} \tag{4.1}$$

with the inner derivatives $\frac{d}{d\,\theta_k} R_0(t_j, x_j, y_j)$ and $\frac{d}{d\,\theta_k} N(t_j)$ from section section 3.2 and 4.1, respectively, and therefore:

$$\frac{d}{d\,\mu} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,\mu} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,\mu} N(t_j)\right)\, e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,A} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,A} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,A} N(t_j)\right)\, e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,\alpha} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,\alpha} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,\alpha} N(t_j)\right)\, e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,c} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,c} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,c} N(t_j)\right)\, e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,p} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,p} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,p} N(t_j)\right)\, e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,D} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,D} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,D} N(t_j)\right)\, e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,\gamma} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,\gamma} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,\gamma} N(t_j)\right)\, e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,q} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,q} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,q} N(t_j)\right)\, e^{-\beta(m_j - M_c)}$$

4.2.2 By T_b (Blind Time)

Having

$$\frac{d}{d\,T_b}\,N(t) = \frac{N(t)}{T_b}$$

from section 4.1 it holds

$$\frac{d}{dT_b}\,LL_1(\theta) = \sum_{j=1}^{N} -\left(\frac{d}{d\,T_b}N(t_j)\right)\,e^{-\beta(m_j - M_c)}$$

4.2.3 By β (Gutenberg-Richter)

It holds

$$\frac{d}{d\beta}\,LL_1(\theta) = \frac{N}{\beta} + \sum_{j=1}^{N}(m_i - M_c)\,\left(N(t_j)\,e^{-\beta(m_i - M_c)} - 1\right).$$

4.3 Derivatives of $LL_2(\theta)$

The second summand of the LL-function is

$$LL_2(\theta) = \frac{T_2 - T_1}{T_b} - \frac{1}{T_b} \int_{T_1}^{T_2} e^{-N_0(t)} \, dt$$

4.3.1 By $\mu, A, \alpha, c, p, D, \gamma, q$

For any parameter other than T_b, we obtain the derivative by chain rule as

$$\frac{d}{d\,\theta_k} LL_2(\theta) = \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,\theta_k} N_0(t) \right) e^{-N_0(t)} \, dt. \tag{4.2}$$

with the inner derivatives $\frac{d}{d\,\theta_k} N_0(t)$ given in section 4.1:

$$\frac{d}{d\,\mu} LL_2(\theta) = \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,\mu} N_0(t) \right) e^{-N_0(t)} \, dt$$
$$\frac{d}{d\,A} LL_2(\theta) = \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,A} N_0(t) \right) e^{-N_0(t)} \, dt$$
$$\frac{d}{d\,\alpha} LL_2(\theta) = \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,\alpha} N_0(t) \right) e^{-N_0(t)} \, dt$$
$$\frac{d}{d\,c} LL_2(\theta) = \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,c} N_0(t) \right) e^{-N_0(t)} \, dt$$
$$\frac{d}{d\,p} LL_2(\theta) = \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,p} N_0(t) \right) e^{-N_0(t)} \, dt$$
$$\frac{d}{d\,D} LL_2(\theta) = \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,D} N_0(t) \right) e^{-N_0(t)} \, dt$$
$$\frac{d}{d\,\gamma} LL_2(\theta) = \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,\gamma} N_0(t) \right) e^{-N_0(t)} \, dt$$
$$\frac{d}{d\,q} LL_2(\theta) = \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,q} N_0(t) \right) e^{-N_0(t)} \, dt$$

4.3.2 By T_b (Blind Time)

It holds

$$\frac{d}{d\,T_b} \frac{T_2 - T_1}{T_b} = -\frac{T_2 - T_1}{T_b^2}$$

and, via quotient rule,

$$\frac{d}{d\,T_b} \frac{e^{-N_0(t)}}{T_b} = \frac{\left(-\frac{d}{d\,T_b} N_0(t) \right) e^{-N_0(t)} \, T_b - e^{-N_0(t)}}{T_b^2} = -\frac{e^{-N_0(t)}}{T_b^2} \left(T_b \left(\frac{d}{d\,T_b} N_0(t) \right) + 1 \right).$$

We obtain

$$\frac{d}{d\,T_b} LL_2(\theta) = -\frac{T_2 - T_1}{T_b^2} + \frac{1}{T_b^2} \int_{T_1}^{T_2} \left(T_b \left(\frac{d}{d\,T_b}\, N_0(t) \right) + 1 \right) e^{-N_0(t)}\, dt.$$

4.3.3 By β (Gutenberg-Richter)

The parameter β does not occur in $LL_2(\theta)$, therefore

$$\frac{d}{d\,\beta} LL_2(\theta) = 0.$$

Bibliography

C. Grimm, S. Hainzl, M. Käser, and H. Küchenhoff. Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence. *Stochastic Environmental Research and Risk Assessment*, 2, 2022. doi: 10.1007/s00477-022-02221-2.

Christian Grimm, Martin Käser, Sebastian Hainzl, Marco Pagani, and Helmut Küchenhoff. Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model. *Bulletin of the Seismological Society of America*, (Xx), 2021. ISSN 0037-1106. doi: 10.1785/0120210097.

Sebastian Hainzl. ETAS-Approach Accounting for Short-Term Incompleteness of Earthquake Catalogs. *Bulletin of the Seismological Society of America*, 2021. ISSN 0037-1106. doi: 10.1785/0120210146.

Abdollah Jalilian. ETAS: An R package for fitting the space-time ETAS model to earthquake data. *Journal of Statistical Software*, 88(1), 2019. ISSN 15487660. doi: 10.18637/jss.v088.c01.

Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, mar 1988. ISSN 0162-1459.

Yosihiko Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998. ISSN 00203157.