



Understanding the numerical issues in and instability of fits based on the Laplace Approximation

Patrick Mogensen¹, Andreas Noack¹

¹PumasAI

Introduction

It is often suggested that Laplace's Approximation and direct handling of censored observations (M3 in the terminology of [1]) is hard to fit. Recently, this was also investigated in [2]. The papers cited in [2] all use simulations to study the issue. However, the issues seen with censored data are not censoring specific. To better understand the issues we look at the underlying expressions in the approximation.

The typical way to see a problematic case with Laplace's Method and censored data is to see decreasing negative loglikelihood values during estimation and a divergent gradient. Ultimately, the gradient elements go to extreme values and optimization halts. A typical error mode is one where just a single subject contributes these large values.

In this poster we at the approximation itself and look for the problematic terms causing failure. Remember, the reason why censored data and "M3" itself is often blamed for the issues rather than Laplace's Approximation is that FOCE cannot be used for this class of models. The problems cannot arise in FOCE based handling of censored data as $-H$ is approximated in a way that is always positive definite.

Results

- Expressions can be derived for various types of models without much effort.
- Terms should sum to a negative number and this depends on sampling times, model parameters, and log prior (of random effects) contributions.
- Only one subject needs to fail this condition for the entire Laplace Approximation to fail.
- Figure 1 shows that in simulations of the proportional error model the issue can easily be provoked under the right conditions (samples at early time points) with error rates nearing 100% (200 replications).
- Figure 2 shows the regions in which a sample at some time point (first axis) and observed value (second axis) leads to negative or positive contributions to $-H$.
- Equations and contour plots explain fraction of failed fits.

Penalized Log-Likelihood

$$l_{p,1} = -\frac{(y_j - m(j))^2}{2\pi f(\eta)^2} \quad l_{p,2} = -\frac{\log(2\pi f(\eta)^2)}{2} \quad l_p(\eta) = \left(\sum_{j=1}^{n_i} l_{p,1} + l_{p,2} \right) - \frac{\eta^2}{2\omega^2}$$

Hessian of penalized Log-Likelihood (Gaussian Error Model)

$$\frac{d^2 l_{p,1}}{d\eta^2} = \frac{f(\eta_i)(f''(\eta_i)R(\eta_i)^2 + f(\eta_i)R(\eta_i)m''(\eta_i) + m'(\eta_i)^2)}{f(\eta_i)^4} - \frac{4f(\eta_i)f'(\eta_i)R(\eta_i)m'(\eta_i) - 3f'(\eta_i)^2 R(\eta_i)^2}{f(\eta_i)^4}$$

$$\frac{d^2 l_{p,2}}{d\eta^2} = \frac{f'(\eta_i)^2 - f(\eta_i)f''}{f(\eta_i)^2} \quad \frac{d^2 \log p(\eta_i)}{d\eta_i^2} = -\frac{1}{\omega^2}$$

Hessian of penalized Log-Likelihood (One Compartment)

$$\frac{d^2 l_{p,1,j}^{add}}{d\eta^2} = \frac{R_j(\eta)m_j''(\eta) - m_j'(\eta)^2}{\sigma^2}$$

$$\frac{d^2 l_{p,2,j}^{add}}{d\eta^2} = 0$$

$$\frac{d^2 l_{p,1,j}^{prop}}{d\eta^2} = -\frac{1}{\sigma^2} \frac{y_j}{m_j(\eta)} k(\eta) t_j \left(\frac{y_j}{m_j(\eta)} (2k(\eta)t_j + 1) - (k(\eta)t_j + 1) \right)$$

$$\frac{d^2 l_{p,2,j}^{prop}}{d\eta^2} = k(\eta) t_j$$

Sign conditions

$$\frac{d^2 l_{p,1,j}^{add}}{d\eta^2} < 0 \Leftrightarrow \frac{y_j}{m_j(\eta)} < \frac{2k(\eta)t_j - 1}{k(\eta)t_j - 1}$$

$$\frac{d^2 (l_{p,1,j}^{prop} + l_{p,2,j}^{prop})}{d\eta^2} < 0 \Leftrightarrow \sigma^2 < \frac{y_j}{m_j(\eta)} \left(\frac{y_j}{m_j(\eta)} (2k(\eta)t_j + 1) - (k(\eta)t_j + 1) \right)$$

Methods

To study the problematic behavior of Laplace without focusing on censored data we simulate data from a **proportional error model** and show the problematic behavior emerges. The model is shown below:

$$m_i(t) = \tilde{D} \times e^{-\frac{CL_i t}{V}}, k(\eta) = \frac{CL_i}{V}$$

Where $\tilde{D} = \frac{D}{V}$, $CL_i = CL \times e^{\eta_i}$, $\eta_i \sim N(0, \omega)$, and $D = 200$, $V = 7$, $CL = 2.9$, $\omega^2 = 0.9$. The error model defines the following sampling distribution

$$y_{ij} \sim N\left(m_i(t_j), \sqrt{(m_i(t_j) * \sigma)^2}\right)$$

The approximate marginal loglikelihood is then defined as

$$\log(p(y_i)) \approx \log(p(y_i|\hat{\eta}_i)) + \log(p(\hat{\eta}_i)) - \frac{1}{2} \log \det -H(\hat{\eta}_i) + \frac{1}{2} \log(2\pi)$$

Where $\hat{\eta}_i$ is the mode of the penalized conditional loglikelihood also known as the empirical bayes estimates. Simulations were done in Pumas v2.6.1 [3,4].

Figure 1: Simulation Exercise

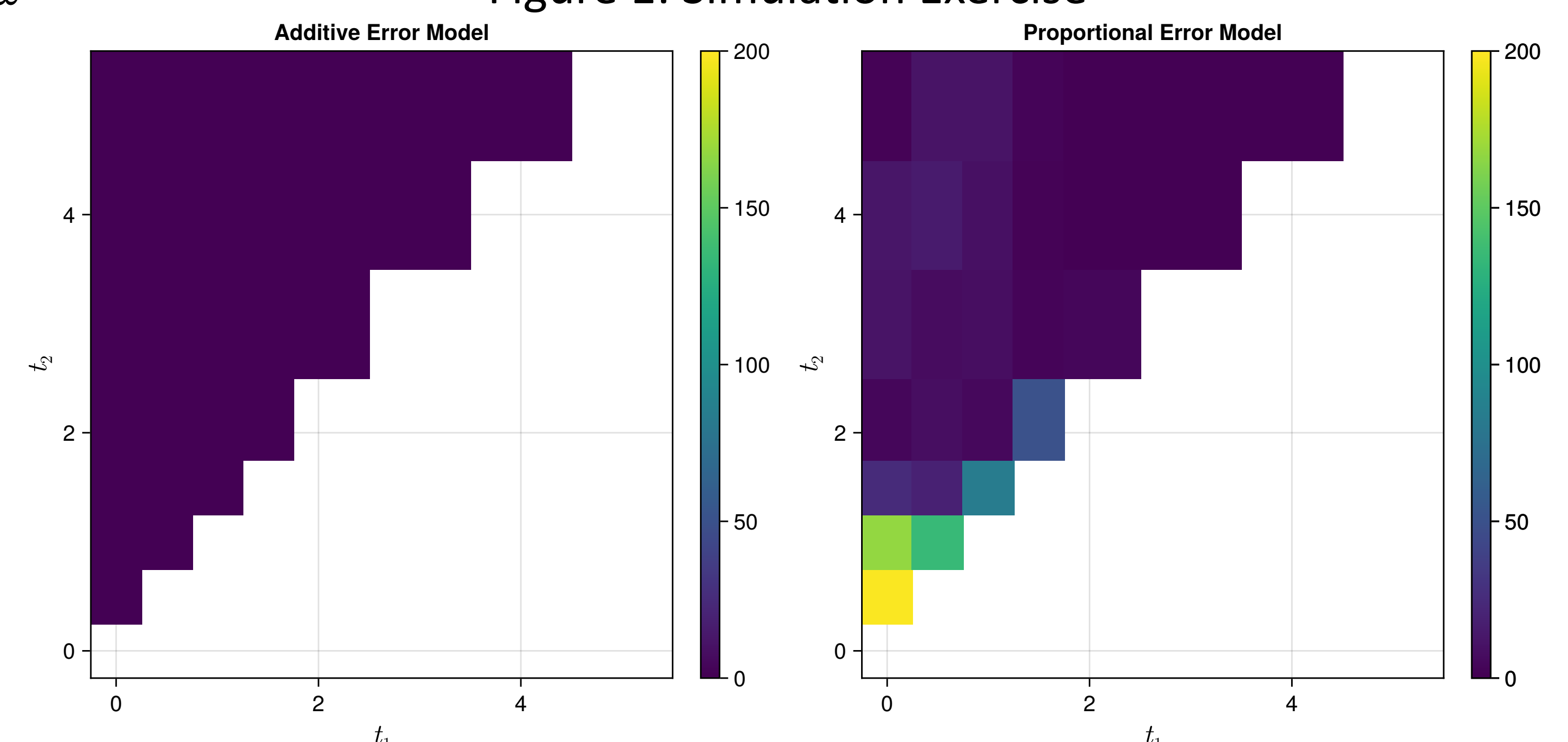
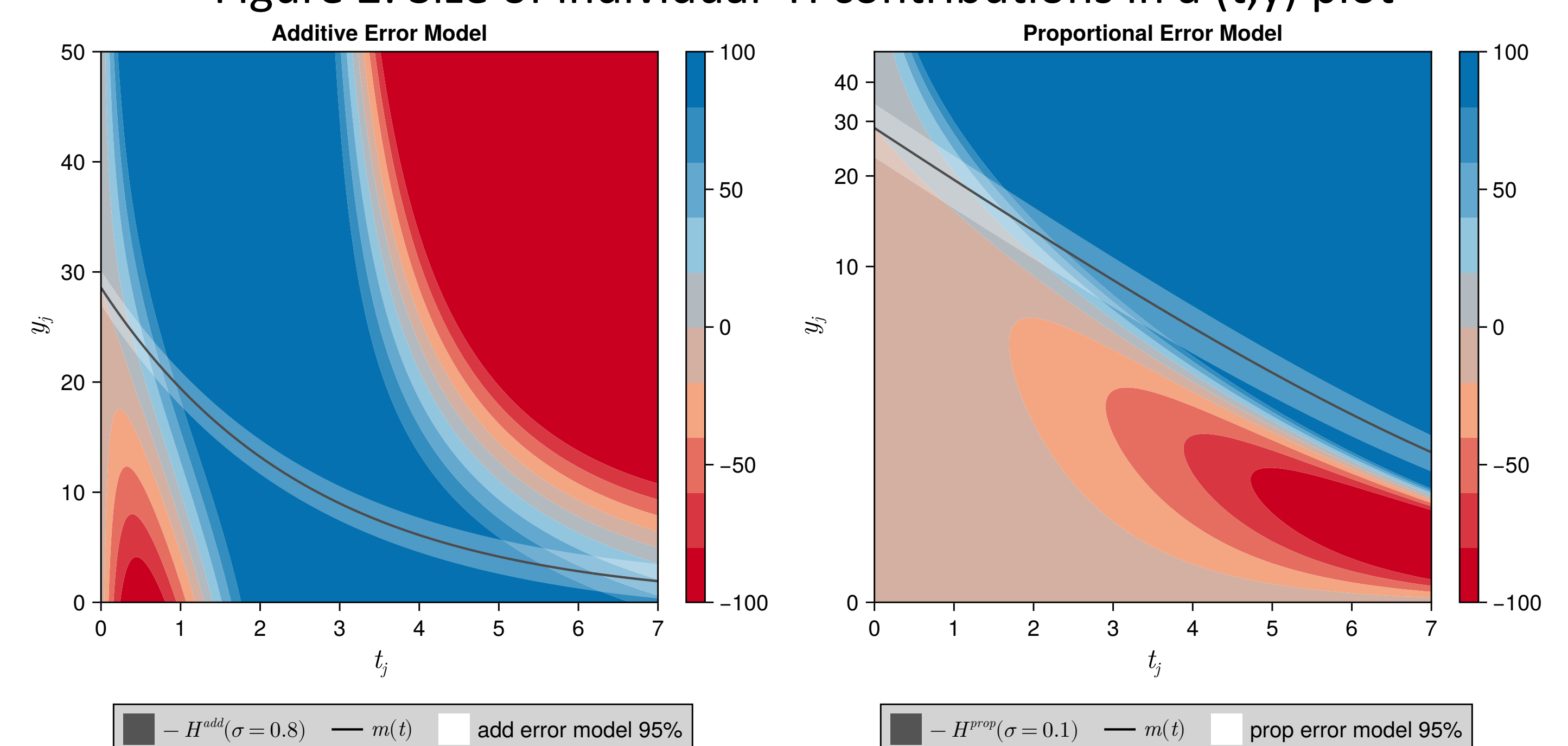


Figure 2: Size of individual $-H$ contributions in a (t,y) plot



Conclusions and Discussions

- A negative definite Hessian of the conditional loglikelihood function is not guaranteed
- Interplay between error model, sampling times, and model assumptions affects the risk of failure
- Curvature varies across sampling time and is thus affected by study design including sparseness of data
- Diverging gradients of the marginal loglikelihood occurs generally; not just for the censored data models
- Future work will further qualify the exact conditions for a broader class of models and sources of the failure including the potential to design the study appropriately in terms of sampling times according to expected ranges of parameters. It may also be possible to regularize or otherwise avoid total failure due to just one subject with sparse data.

References

- Stuart L Beal. "Ways to fit a PK model with some data below the quantification limit". In: Journal of pharmacokinetics and pharmacodynamics 28 (2001), pp. 481–504.
- Marie Wijk et al. "A Pragmatic Approach to Handling Censored Data Below the Lower Limit of Quantification in Pharmacokinetic Modeling". In: CPT: Pharmacometrics & Systems Pharmacology (2025).
- Chris Rackauckas et al. "Accelerated predictive healthcare analytics with pumas, a high performance pharmaceutical modeling and simulation platform". In: BioRxiv (2020), pp. 2020–11. <https://tutorials.pumas.ai/>
- <https://docs.pumas.ai/v2.6.1/>