Bachelor Thesis Marginal-Sampling

Michael Fedders

January 20, 2021

Table of Contents

Introduction

Model

We assume that we can describe a (biological) process through a function $x(t, \theta)$ with time t and unknown model parameters θ. Through (un)voluntary limitations the measured data is not x but

$$\bar{y} = c + h\big(x(t, \theta)\big) + \varepsilon$$

where

- $\bar{y}$ is the measured data
- c is an offset parameter
- h is the observation function
- ε is a noise - for now $\epsilon \sim \mathcal{N}(0, 1/\lambda)$

Standard approach

The standard approach is to use a data set D to determine the model parameters θ and the offset c and noise parameter λ with Bayes theorem:

$$p(\theta, c, \lambda \mid D) = \frac{p(D \mid \theta, c, \lambda) \cdot p(\theta, c, \lambda)}{p(D)}.$$

We can then use Markov chain Monte Carlo (MCMC) methods to proportionally sample the posterior distribution with the product of likelihood $p(D \mid \theta, c, \lambda)$ and prior $p(\theta, c, \lambda)$. We will call this way **FP-approach** from now on

Hierarchical approach

For Maximum Likelihood methods it was shown that it can be faster to first derive the model parameter θ and then in a second step the noise and transformation (e.g. offset, scaling) parameter.

$\Longrightarrow$ We would like to do the same for the posterior sampling

Hierarchical approach

For this **MP-approach** we have to calculate the marginalised likelihood first:

$$p(D \mid \theta) = \iint p(D \mid \theta, c, \lambda) \cdot p(\theta, c, \lambda) \, \mathrm{d}c \, \mathrm{d}\lambda.$$

Hierarchical approach

For this **MP-approach** we have to calculate the marginalised likelihood first:

$$\begin{aligned} p(D \mid \theta) &= \iint p(D \mid \theta, c, \lambda) \cdot p(\theta, c, \lambda) \, \mathrm{d}c \, \mathrm{d}\lambda \\ &= \iint p(D \mid \theta, c, \lambda) \cdot p(c, \lambda) \, \mathrm{d}c \, \mathrm{d}\lambda \cdot p(\theta) \end{aligned}$$

Hierarchical approach

We can then use that

$$p(\theta \mid D) = \frac{p(D \mid \theta) \cdot p(\theta)}{p(D)}$$

to proportionally sample from the marginalized posterior $p(\theta \mid D)$, again with MCMC methods.

MCMC

maybe an extra frame to explain MCMC and Parallel Tempering

Gaussian noise

Normal-Gamma prior

The normal-gamma prior depends on 4 shape parameters, $\mu, \kappa, \alpha, \beta$ and has the following structure:

$$\begin{aligned} p(c, \lambda) &= f(c, \lambda \mid \mu, \kappa, \alpha, \beta) \\ &= \mathcal{N}(c \mid \mu, 1/(\kappa\lambda)) \cdot \Gamma(\lambda \mid \alpha, \beta). \end{aligned}$$

Likelihood

We recall that the measurements are defined as $\overline{y} = c + h + \epsilon$. We assume independent noise for different points in time $t_1, \dots, t_N,\ N \in \mathbb{N}$. Therefore the likelihood is

$$p(D \mid \theta, c, \lambda) = \left(\frac{\lambda}{2\pi}\right)^{N/2} \cdot \exp\left(-\frac{\lambda}{2}\sum_{k=1}^{N}(\overline{y}_k - (c + h_k))^2\right)$$

Marginal Likelihood

The integral which we have to solve is defined as

$$p(D \mid \theta) = \int_{\mathbb{R}\times\mathbb{R}_+} p(D \mid \theta, c, \lambda) p(c, \lambda) \, d(c, \lambda) \tag{1}$$

Marginal Likelihood

The integral which we have to solve is defined as

$$p(D \mid \theta) = \int_{\mathbb{R}\times\mathbb{R}_+} p(D \mid \theta, c, \lambda) p(c, \lambda)\, d(c, \lambda) \tag{2}$$

$$= \int_{\mathbb{R}_+} \int_{\mathbb{R}} \left(\frac{\lambda}{2\pi}\right)^{N/2} \exp\left(-\frac{\lambda}{2} \sum_{k=1}^{N} (\overline{y}_k - (c + h_k))^2\right) \tag{3}$$

$$\cdot \frac{\beta^\alpha \sqrt{\kappa}}{\Gamma(\alpha)\sqrt{2\pi}} \lambda^{\alpha - 1/2} \exp\left(-\frac{\lambda}{2}\left(\kappa(c-\mu)^2 + 2\beta\right)\right)\, dc\, d\lambda \tag{4}$$

Marginal Likelihood

With an exponential integration formula

$$\int_{\mathbb{R}} \exp(-a \cdot c^2 + b \cdot c - d)\, dc = \sqrt{\frac{\pi}{a}} \cdot \exp\left(\frac{b^2}{4a} - d\right)$$

Marginal Likelihood

With an exponential integration formula

$$\int_{\mathbb{R}} \exp(-a \cdot c^2 + b \cdot c - d)\, dc = \sqrt{\frac{\pi}{a}} \cdot \exp\left(\frac{b^2}{4a} - d\right)$$

the substitution of $\varphi(\lambda) = C \cdot \lambda$ and the definition of the gamma-function we conclude with the following form:

$$\frac{(\beta/C)^\alpha}{\Gamma(\alpha)(2\pi C)^{\frac{N}{2}}} \cdot \sqrt{\frac{\kappa}{N+\kappa}} \cdot \Gamma\left(\frac{N}{2} + \alpha\right).$$

Distribution of c and λ

To sample c and λ in a second step we need to derive their distribution from the integrand of the marginalized likelihood. We have

$$\lambda \propto \text{Gamma}(\alpha' = \alpha + N/2, \beta' = C)$$

and

$$c \propto \mathcal{N}\left(\mu' = \frac{\left(\sum_{k=1}^{N} \overline{y_k} - h_k\right) + \kappa\mu}{N + \kappa}, \hat{\lambda} = \lambda(N + \kappa)\right)$$

Conversion Reaction model

For $k_1, k_2 \in \mathbb{R}$ we consider the following ordinary differential equation (ODE):

$$\frac{dX_1(t)}{dt} = k_2 X_2 - k_1 X_1 \qquad \frac{dX_2(t)}{dt} = k_2 X_1 - k_1 X_2.$$

We want to observe X_2.

Conversion Reaction model

For both approaches we sampled 50 independent runs with 10.000 steps each with the Adaptive Metropolis sampler from pyPESTO.

Laplacian noise

Laplacian likelihood

Now we make the assumption that

$$\epsilon_k \sim \mathit{Laplace}(0, \sigma),\ \ \sigma \in (0, \infty)$$

i.e. it has a Laplace distribution. The new likelihood has the following form:

$$\begin{aligned} p(D \mid \theta, c, \sigma) &= \prod_{k=1}^{N} \text{Laplace}\left(\overline{y}_k \mid c + h_k, \sigma\right) \\ &= \prod_{k=1}^{N} \frac{1}{2\sigma} \cdot \exp\left\{-\frac{|\overline{y}_k - c - h_k|}{\sigma}\right\} \end{aligned}$$

Marginalisation Integral

The integral we receive is

$$\iint p(D \mid \theta, c, \sigma) p(c) p(\sigma) \, \mathrm{d}c \, \mathrm{d}\sigma \tag{5}$$

$$= \int_0^\infty \int_{-\infty}^\infty \prod_{k=1}^N \frac{1}{2\sigma} \cdot \exp\left\{ -\frac{|c - (\overline{y}_k - h_k)|}{\sigma} \right\} p(c) p(\sigma) \, \mathrm{d}c \, \mathrm{d}\sigma \tag{6}$$

Calculation

For calculation-reasons we renumber $\overline{y}_k$ and h_k so that $y_k - h_k$ are ordered from smallest to biggest, i.e. $\overline{y}_1 - h_1$ is the smallest number, $\overline{y}_N - h_N$ the biggest. Then we choose $b_0 = -\infty, b_i = \overline{y} - h_i (i = 1, \ldots, N), b_{N+1} = \infty$. Now we can split up the integral in the following parts:

$$\int_0^\infty \sum_{i=0}^{N} \int_{b_i}^{b_{i+1}} \frac{1}{2\sigma} \exp\left\{ -\frac{\sum_{k=1}^{N} |c - (\overline{y}_k - h_k)|}{\sigma} \right\} p(c) p(\sigma) \, \mathrm{d}c \, \mathrm{d}\sigma \tag{7}$$

Calculation

To remove the absolute value, we introduce the index $R_{k,i}$ which is defined like this:

$$r_{k,i} = \begin{cases} 1 & \text{if } k \leq i \\ -1 & \text{else} \end{cases}$$

Calculation

$$\int_0^\infty \frac{1}{2\sigma} \sum_{i=0}^{N} p(\sigma) \int_{b_i}^{b_{i+1}} \exp \underbrace{\left\{ -\frac{\sum_{k=1}^{N} r_{k,i}(c - (\overline{y}_k - h_k))}{\sigma} \right\}}_{=(*)} \mathrm{d}c \, \mathrm{d}\sigma \tag{8}$$

$$\text{with } (*) = \frac{-c(i - (N - i)) + \sum_{k=1}^{i} \overline{y}_k - h_k - \sum_{i+1}^{N} \overline{y}_k - h_K}{\sigma} \tag{9}$$

$$= \int_0^\infty \frac{1}{2\sigma} p(\sigma) \sum_{i=0}^{N} \exp \left\{ \frac{\sum_{k=1}^{i} \overline{y}_k - h_k - \sum_{i+1}^{N} \overline{y}_k - h_K}{\sigma} \right\} \tag{10}$$

$$\cdot \int_{b_i}^{b_{i+1}} e^{-c(2i-N)} \, \mathrm{d}c \, \mathrm{d}\sigma \tag{11}$$

$$\int_0^\infty \frac{1}{2\sigma} \sum_{i=1}^{N-1} \exp\left\{ \frac{\sum_{k=1}^{i} \overline{y}_k - h_k - \sum_{i+1}^{N} \overline{Y}_k - h_k}{\sigma} \right\} \tag{12}$$

$$\cdot \frac{1}{N-2i} \cdot \left(e^{-b_{i+1}(2i-N)} - e^{-b_i(2i-N)} \right) \mathrm{d}\sigma \tag{13}$$

$$+ \int_0^\infty \frac{1}{2\sigma} \exp\left\{ \frac{-\sum_{k=1}^{N} \overline{y}_k - h_k}{\sigma} \right\} \underbrace{\int_{-\infty}^{b_1} e^{Nc}\, \mathrm{d}c}_{\frac{1}{N} e^{Nb_1}} \mathrm{d}\sigma \tag{14}$$

$$+ \int_0^\infty \frac{1}{2\sigma} \exp\left\{ \frac{\sum_{k=1}^{N} \overline{y}_k - h_k}{\sigma} \right\} \underbrace{\int_{b_N}^{\infty} e^{-Nc}\, \mathrm{d}c}_{\frac{1}{N} e^{-Nb_N}} \mathrm{d}\sigma \tag{15}$$