Bachelor Thesis Marginal-Sampling

Michael Fedders

28.05.2020

Table of Contents

Introduction / Motivation

Motivation

In the area of applied mathematics and especially while we are working with model based assumptions we have to deal with uncertainties. Mainly there are two possible points:

- during observation of the model there may be a loss of dimension
- measuring variables can add a noise

Mathmatical model

In this setting we will focus on ODE models. The original model can be described through

$$\frac{dx(t,\theta)}{dt} = f(x(t,\theta),\theta), \quad x(t_0,\theta) = x_0(\theta)$$

with x as our state vector, t as the time and θ representing parameters from our system, which are unknown at the start.

Observation function

We can not analyse the function directly but have to use

$$y(t_k, \theta) = \tilde{h}(h(x(t_k, \theta), \theta))$$

and, assuming additive noise

$$\overline{y}_k = y(t_k, \theta) + \epsilon_k$$

Measure uncertainties

We can group this data observation in three steps.

Measure uncertainties

We can group the data interpretation in three steps.

- observation function $h(x(t_k, \theta))$

Measure uncertainties

We can group the data interpretation in three steps.

- observation function $h(x(t_k, \theta))$
- relative experimental data through $y_k = \tilde{h}(h(t_k, \theta))$

Measure uncertainties

We can group the data interpretation in three steps.

- observation function $h(x(t_k, \theta))$
- relative information $y_k = \tilde{h}(h(t_k, \theta))$

For example:

$$y(t_k, \theta, c) = c + h(x(t_k, \theta), \theta) \qquad y(t_k, \theta, s) = s \cdot h(x(t_k, \theta), \theta)$$

Measure uncertainties

We can group the data interpretation in three steps.

- observation function $h(x(t_k, \theta))$
- relative information $y_k = \tilde{h}(h(t_k, \theta))$

For example:

$$\boxed{y(t_k, \theta, c) = c + h(x(t_k, \theta), \theta)}$$

Measure uncertainties

We can group the data interpretation in three steps.

- observation function $h(x(t_k, \theta))$
- relative experimental data through $y_k = \tilde{h}(h(x(t_k, \theta)))$
- adding the noise $\overline{y}_k = y + \varepsilon_k$

Measure uncertainties

We can group the data interpretation in three steps.

- observation function $h(x(t_k, \theta))$
- relative experimental data through $y = \tilde{h}(h(x(t_k, \theta))$
- adding the noise $\overline{y}_k = y + \varepsilon_k$

For example:

$$\varepsilon_k \sim \mathcal{N}(0, \sigma^2) \qquad \qquad \varepsilon_k \sim \text{Laplace}(0, \sigma)$$

Measure uncertainties

We can group the data interpretation in three steps.

- observation function $h(x(t_k, \theta))$
- relative experimental data through $y = \tilde{h}(h(x(t_k, \theta))$
- adding the noise $\overline{y}_k = y + \varepsilon_k$

For example:

$$\boxed{\epsilon_k \sim \mathcal{N}(0, \sigma^2)}$$

Measure uncertainties

We can group the data interpretation in three steps.

- observation function $h(x(t_k, \theta))$
- relative experimental data through $y = \tilde{h}(h(x(t_k, \theta))$
- adding the noise $\overline{y}_k = y + \varepsilon_k$

In total we get:

$$\boxed{\overline{y}_k = c + (h(x(t_k, \theta)) + \varepsilon_k \text{ with } \varepsilon_k \sim \mathcal{N}(0, \sigma^2)}$$

Bayes Theorem

We now want to use data to estimate our parameters. To do so we will use Bayes Theorem:

Bayes Theorem

We now want to use data to estimate our parameters. To do so we will use Bayes Theorem:

$$p(\theta, c, \sigma^2 \mid D) = \frac{p(D \mid \theta, c, \sigma^2) \cdot p(\theta)\, p(c)\, p(\sigma^2)}{p(D)}$$

Marginalization of the posterior

But from a process standpoint we are only interested in the distribution of θ and therefore would like to only calculate $p(\theta \mid D)$ which has less dimensions to be sampled.

Marginalization of the posterior

To get this result we take the expected value from both sides over the parameters introduced by the noise ε and the relative factor $\tilde{h}$ and receive:

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

with

$$p(D \mid \theta) = \int_{\mathbb{R}_+} \int_{\mathbb{R}} p(D \mid \theta, c, \sigma^2) p(c) p(\sigma^2) \, dc \, d\sigma^2$$

Personal progress

Calculation

By inserting the likelihood function we receive

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}} \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \left(\frac{\overline{y}_k - (c + h_k)}{\sigma} \right)^2 \right) p(c)\, p(\sigma^2)\, dc\, d\sigma^2$$

Choice of priors

One main choice here is, which priors $p(c)$ and $p(\sigma^2)$ we use, as they are defining the distribution of c and σ^2.

Choice of priors

First choice: "flat" priors, aka $p(c) = p(\sigma^2) = 1$. The problem with this choice is, that

$$\int_{\mathbb{R}} 1\, dx = \infty \neq 1$$

so we are not using a probability distribution here.

Choice of priors

First choice: "flat" priors, aka $p(c) = p(\sigma^2) = 1$. The problem with this choice is, that

$$\int_{\mathbb{R}} 1 \, dx = \infty \neq 1$$

so we are not using a probability distribution here.

But: We can still argue, that this choice is preserving enough information to make it a valid first choice.

Calculation

If we now insert $p(c) = p(\sigma^2) = 1$ in

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}} \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{\overline{y}_k - (c + h_k)}{\sigma}\right)^2\right) p(c)\, p(\sigma^2)\, dc\, d\sigma^2$$

we can substitute twice and get

$$\frac{\left(\sqrt{N}\right)^{N-4}}{2\left(\sqrt{\pi}\right)^{N-1}} \cdot \left(N \cdot \sum_{k=1}^{N} (h_k - \overline{y_k})^2 - \left(\sum_{k=1}^{N} \overline{y_k} - h_k\right)^2\right)^{-(N-3)/2} \cdot \Gamma\left(\frac{N-3}{2}\right)$$

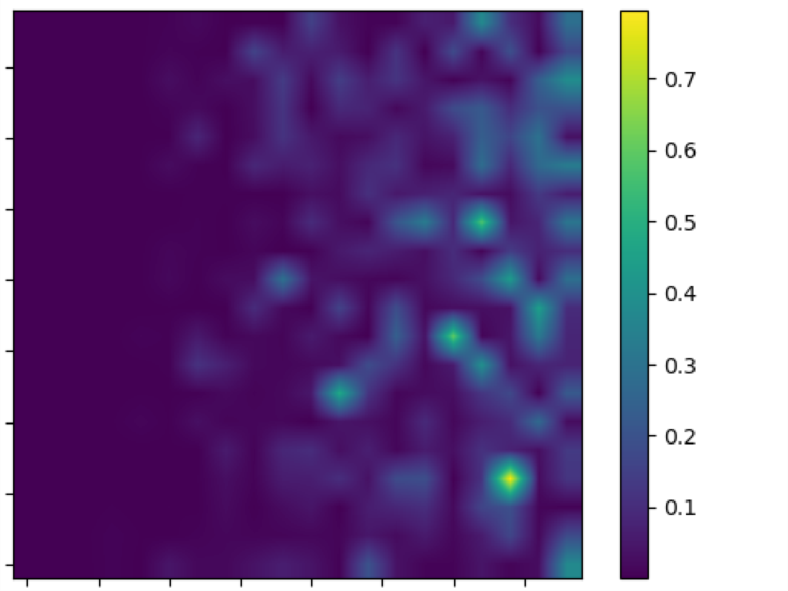
Numerical testing

First test with basic sample regression model:

- $y(t, \theta, c) = c + t$ and
- $t_k = k$

In total we get $\overline{y}_k \sim \mathcal{N}(c + t_k, \sigma^2)$.

Plot



Next steps

Current work

- The next steps for the flat prior choice will be to implement test examples of ODEs to further test the analytical derivation.
- I also started the calculation of the integral in the case of a gaussian prior $c \sim \mathcal{N}(\mu, \hat{\sigma}^2)$.

Gaussian prior

By defining three constants

$$K_1 = \left(\sum_{k=1}^{N} (\overline{y_k} - h_k)\right)^2,\ K_2 = \sum_{k=1}^{N} (\overline{y_k} - h_k)^2,\ K_3 = \sum_{k=1}^{N} (\overline{y_k} - h_k)$$

we can rewrite the integral in the following way:

Gaussian prior

$$\int_0^\infty \frac{1}{(2\pi\sigma^2)^{N/2}} \sqrt{\frac{\sigma^2}{N\hat{\sigma}^2 + \sigma^2}} \cdot p(\sigma^2)$$

$$\cdot \exp\left(\frac{(\mu\hat{\sigma}^2 K_3 - \mu^2 N\hat{\sigma}^2/2 - \hat{\sigma}^2 K_2) \cdot \sigma^2 + (\hat{\sigma}^2)^2 K_1/2 - N(\hat{\sigma}^2)^2 K_2}{\sigma^2 \cdot (N\hat{\sigma}^2 + \sigma^2)} \right) d\sigma^2$$

Thank you for your attention!