Bayesian Parameter Estimation in gPROMS

Lucian Gomoescu , Process Systems Enterprise Ltd and Imperial College London

Summary

The objective of this work is to **apply Bayesian inference on parameters** of a simple viral infection dynamics model, with **non-linear parameter dependence**, in order to determine if the parameter values used in generating the artificial experimental data are included in the bulk of posterior parameter distribution.
A model for viral infection dynamics [1] was developed in gPROMS, then artificial noisy-data points were created from the simulation results obtained with a set of known parameter values and initial conditions.
Parameter posterior distribution was estimated using an ensemble Markov chain Monte Carlo (MCMC) sampling technique [2] and a uniform parameter prior distribution.

Motivation

In the Bayesian approach, the inference problem is formulated using Bayes theorem. On one hand, this presents the following advantages:

* prior knowledge can be incorporated;
* there is no limitation with respect to shape of inferred variable distribution.

... on the other hand, the methods required for the calculation of posterior distribution are very computationally expensive (many likelihood function evaluations).

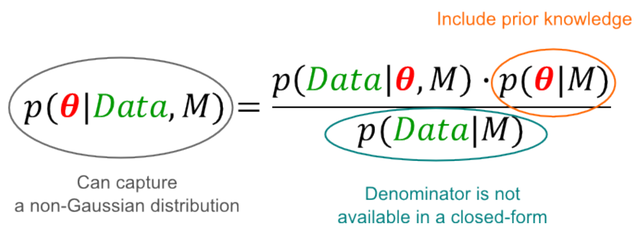

Figure 1: Bayes theorem applied to parameter estimation.

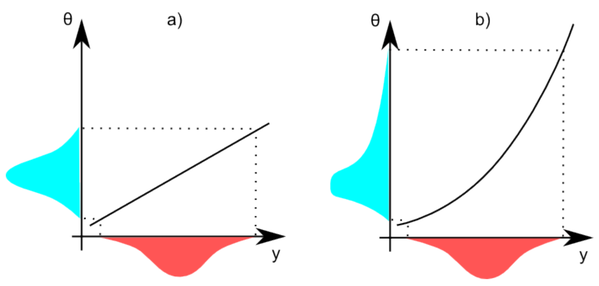

Figure 2: In case of models with non-linear parameter dependence (b), the degeneracy from normality manifests as skewness or/and multi-modality.

Calibration Data

In order to test if the "true" parameter values are included in the bulk of parameter posterior distribution, artificial calibration data was generated through simulation of the model with parameter values from table 1 and addition of Gaussian noise.
The number of uninfected cells, infected cells, antibodies and virus particles were recorded every 6 hours for 30 days.
The errors were applied to measured variables and considered normally distributed around recorded value with a spread given by a standard deviation equal to $0.3 \cdot \tilde{Y}$, where $\tilde{Y}$ is any of the measured variables.

Parameter	Value	Units
a	0.4	day^{-1}
b	0.5	day^{-1}
β	2×10^{-7}	day^{-1}
c	10	day^{-1}
d	0.102	day^{-1}
k	100	day^{-1}
λ	1×10^{5}	day^{-1}
p	5×10^{-6}	–
s	0.3	–
u	5	day^{-1}

Table 1: Parameters used to generate the calibration data. Initial conditions were: $x(0) = 10^6$, $y(0) = 0$, $z(0) = 100$, and $v(0) = 10$.

Acknowledgements

This research received financial support from the European Union's Horizon 2020 research and innovation programme under grant agreement No 675585 (Marie-Curie ITN "SyMBioSys"). Author is a Marie-Curie Early Stage Researcher at Process Systems Enterprise Ltd and a PhD student at Imperial College London, being involved within SyMBioSys project.

References

[1] M. A. Nowak and R. M. May. *Virus dynamics: Mathematical principles of immunology and virology*. Oxford University Press, 1st edition, 2000.

[2] Daniel Foreman-MacKey. emcee: The MCMC Hammer. *Chicago Journals*, 125(925):306–312, March 2013.

[3] John Skilling. Nested sampling for general bayesian computation. *Bayesian Analysis*, 1(4):833–860, 2006.

Model

Model description:
The uninfected cells within an analysed tissue are replenished from external sources with the constant rate, λ, follow a logistic growth together with infected cells, and die at a rate $d \cdot x$.
Free viruses infect cells at a rate $\beta \cdot x \cdot v$ and decay with rate $u \cdot v$ in absence of any immune system reaction.
Infected cells release new free viruses at a rate $k \cdot y$ and die with a rate $(d + a)y$.
As a response to viral infection, the immune system produces antibodies, which remove free viruses with a rate $p \cdot v \cdot z$ and their number reaches a saturation level. In absence of free virus particles, the antibodies would decay at a rate $b \cdot z$.

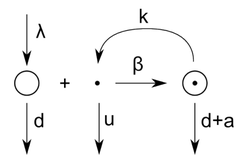

Figure 3: Basic viral infection mechanism [1].

Mathematical model:

$$\dot{x} = \lambda + r \cdot x(1 - \frac{x+y}{K}) - d \cdot x - \beta \cdot x \cdot v \quad (1)$$

$$\dot{y} = r \cdot y(1 - \frac{x+y}{K}) + \beta \cdot x \cdot v - (d+a)y \quad (2)$$

$$\dot{z} = \frac{c \cdot v \cdot z}{1 + s \cdot z} - b \cdot z \quad (3)$$

$$\dot{v} = k \cdot y - u \cdot v - p \cdot v \cdot z \quad (4)$$

Notations:
x, y, z, v - number of uninfected cells, infected cells, antibodies, respectively free viruses within analysed tissue;
a - additional death rate constant of infected cells;
b - antibodies decay rate in absence of stimulation;
β - infection effectiveness;
c - antibodies responsiveness;
d - uninfected cells death rate constant;
k - free viruses generation rate constant;
λ - replenishment rate of uninfected cells from external sources;
K - maximum proliferation capacity;
p - free viruses removal rate constant;
r - proliferation rate constant;
s - antibodies saturation level;
u - virus death rate constant.

Conclusions

- A Bayesian approach to parameter estimation is easily implemented in gPROMS due to already existing capability of likelihood calculation, based on specified experiments and the stochastic model of measurement errors.
- The credibility and the confidence regions may have different size and shape, in particular when the parameter uncertainty is high.
- The methods for Bayesian inference overcome the problem of getting stuck in local minima of the likelihood function.
- MCMC methods are computationally-intensive and, up to date, there is no convergence criterion, in the sense that the entire posterior thoroughly exploration is not guaranteed.
- gPROMS may keep to an acceptable level the wall-clock time required for Bayesian parameter estimation.
- As an observation, the distribution of parameters, or their estimators (in classical approach), tend to a multivariate unimodal Gaussian as more information-rich data is assimilated in the inference process, therefore, in such situations, the two approaches come in agreement. The reason for this agreement is given, on one hand, by the increasing linearisation of model output-parameter dependence around maximum-likelihood point (see fig. 2) and, on the other hand, by the decreasing influence of prior distribution over the posterior distribution.

Future Work

The Monte Carlo methods based on Markov chains construction have difficulties in case of multi-modal posterior distribution because ergodicity is low and the coverage of all modes implies an extremely high number of iterations.
Nested Sampling is another family of sampling techniques that **will be tested**. It was introduced in 2006 by John Skilling [3] and was designed for the estimation of the marginal likelihood; therefore, it has to cover the entire sampling space efficiently.
Besides handling well multi-modality, nested sampling presents also the advantage that its results are the marginal likelihood and posterior distribution, hence the **model selection and parameter estimation can be realized at once**.

Results

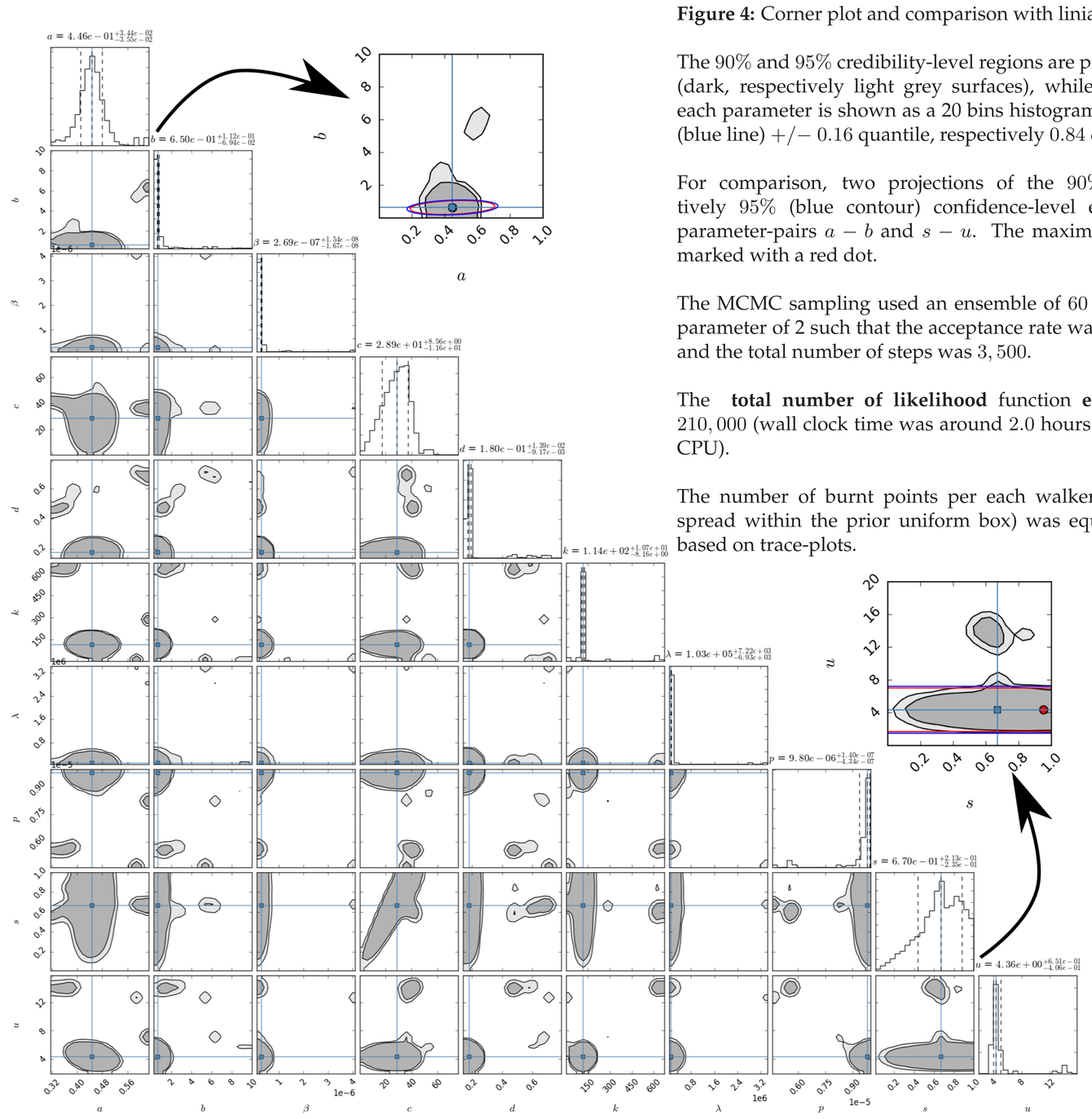

Figure 4: Corner plot and comparison with liniarization around MLE.

The 90% and 95% credibility-level regions are presented as 2D-projections (dark, respectively light grey surfaces), while marginal distribution of each parameter is shown as a 20 bins histogram together with its median (blue line) +/− 0.16 quantile, respectively 0.84 quantile (dashed lines).

For comparison, two projections of the 90% (red contour), respectively 95% (blue contour) confidence-level ellipsoids are shown for parameter-pairs $a - b$ and $s - u$. The maximum likelihood estimate is marked with a red dot.

The MCMC sampling used an ensemble of 60 walkers with a stretching parameter of 2 such that the acceptance rate was on average equal to 0.31 and the total number of steps was $3,500$.

The **total number of likelihood** function **evaluations was equal to** $210,000$ (wall clock time was around 2.0 hours on one core of a i7-5600U CPU).

The number of burnt points per each walker (starting positions were spread within the prior uniform box) was equal to $1,000$ and decided based on trace-plots.

Contact Information

Web http://www.h2020symbiosys.eu/
Email l.gomoescu@psenterprise.com