Bayesian Analysis Procedure - version 2

Alexandre Y. Péré

September 1, 2023

1 Methods

Notes:
- for the rest of the document, a bold letter indicates a vector.

1.1 Terminology

Let the general definition of a Delay-Differential Equations (DDE) model be:

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = f_i(t, \boldsymbol{X}(t), \boldsymbol{X}(t-\tau)|\boldsymbol{\theta}), \qquad t \in [t_0, t_{max}], i = 1, \dots, I$$

where τ denotes a constant delay, so that the rate of change of state X_i depends on both the present state $\boldsymbol{X}(t)$ and a past state $\boldsymbol{X}(t-\tau)$. The subscript i indexes the different state variables of interest, and $\boldsymbol{\theta}$ is the (unknown) vector of the parameters for the DDE model. We must bear in mind that this parameter vector is different for each treated mouse, as it uniquely characterises its treatement response, and hence we denote with $\boldsymbol{\theta}_j$ the parameter vector that characterises the j-th mouse. The experimentally observed tumour evolution for the j-th mouse is denoted by $\boldsymbol{y}_j$, and each element is the tumour volume observed at a given time.

In the context of the immune response model, there are 5 state variables so $\boldsymbol{X} \in \mathbb{R}^5$ and i ranges from 1 to 5. There are 21 parameters and thus $\boldsymbol{\theta} \in \mathbb{R}^{21}$. The exact size of the observation matrix (where each row is an observation vector $\boldsymbol{y}_j$) depends on the investigated treatment but on average is 10×10.

We consider the problem of parameter estimation in a hierarchical Bayesian model, hence we seek to estimate the probability distribution of the parameter vector $\boldsymbol{\theta}$ for each treated mouse. The use of a hierarchical model enables us the formulate that the parameter vector is sampled from an population-level distribution characterised by the (also unknown) hyperparameters $\boldsymbol{\phi}$. The objective is hence to find the distribution of both $\boldsymbol{\theta}_j \ \forall j$ and $\boldsymbol{\phi}$.

1.2 Defining the prior distributions

In a hierarchical Bayesian model, the posterior distribution is defined as follows [1]:

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{y}) \propto p(\boldsymbol{\phi})p(\boldsymbol{\theta}|\boldsymbol{\phi})p(\boldsymbol{y}|\boldsymbol{\theta})$$

This expression is a product of the hyperprior $p(\boldsymbol{\phi})$, the population distribution $p(\boldsymbol{\theta}|\boldsymbol{\phi})$ and the likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$. In the following part, we explain in more details how these distributions are defined.

The population distribution

The hypothesis is that the individual physiological properties (which are effectively contained in the parameter vector $\boldsymbol{\theta}$) are sampled from a hyperdistribution that is a bimodal Gaussian mixture model $\mathcal{B}$ characterised by $\boldsymbol{\phi}$, defined below. Let Λ be a random variable, and define $f(\lambda)$ as its probability density function. Denote the probability density function of the positive normal distribution (truncated at 0) by $g(\lambda; \mu, \sigma)$. Then:

$$\Lambda \sim \mathcal{B}(\boldsymbol{\phi}) \Leftrightarrow f(\lambda) = w \cdot g(\lambda; \mu_1, \sigma_1) + (1-w) \cdot g(\lambda; \mu_2, \sigma_2)$$

such that the distribution is characterised by 5 hyperparameters contained in the vector $\boldsymbol{\phi}$: two means (μ_1, μ_2) and two standard deviations (σ_1, σ_2), as well as w, the mixture weight. The truncated normal is necessary to ensure that all parameters are positive (since negative ones do not make biological sense). This is a general definition and can be extended to a multi-dimensional random variable $\boldsymbol{\Lambda}$. There would then be 5 hyperparameters for each dimension of $\boldsymbol{\Lambda}$. In the context of the immune response problem, the hypothesis is that the vector $\boldsymbol{\theta} \in \mathbb{R}^{21}$ is sampled from $\mathcal{B}$, so $\boldsymbol{\phi} \in \mathbb{R}^{105}$. The high-dimensionality of this vector will be reduced for practical purpose, see Section 1.4.

The hyperprior distribution

As we do not have have much information about the distribution of the hyperparameters, we assign a non-informative hyperprior to reflect our ignorance. However, it must first be noted that some hyperparameters have some constraints (the standard deviations have to be positive, and the weight factor must be between 0 and 1), hence we assign a different prior distribution for each type of hyperparameter. We use the following notation: ϕ_μ is a hyperparameter representing a mean, ϕ_σ represents a standard deviation and ϕ_w represents a weight factor.

$$\begin{aligned}\phi_\mu &\sim \mathcal{N}(\mu, 1)\\ \phi_\sigma &\sim \mathcal{N}^+(0, 1)\\ \phi_w &\sim \mathrm{Dir}(1)\end{aligned}$$

In the above definition, $\mathcal{N}(\cdot,\cdot)^+$ is the positive normal distribution (truncated at 0) and $\mathrm{Dir}(\cdot)$ is the 1D Dirichlet distribution.

The likelihood function

The only quantity collected during the experiments is the total tumour volume, defined as the sum of the living tumour volume and the dead tumour volume. In the DDE, this correspond to the sum of the state variables X_4 and X_5. Hence, the likelihood function $\mathcal{L}(\boldsymbol{\theta})$ only considers the solutions for X_4 and X_5, and more precisely their sum that we denote by Y in the following definition.
As we are not familiar with the experimental technics used to probe the tumour volume, we assume that the observed data is normally distributed around the true volume, with standard deviation σ_{err} that is common to all experiments (this assumption might change in the future, once we have more information on the experimental methods). The likelihood can then be defined as follows [2], [3] (for a given tumour evolution $\boldsymbol{y}_j$):

$$\mathcal{L}(\boldsymbol{\theta}_j) = \prod_{t=t_0}^{t_{max}} \frac{1}{\sigma_{err}} \exp\left(-\frac{(y_j(t) - Y_j(t|\boldsymbol{\theta_j}))^2}{2\sigma_{err}^2}\right)$$

1.3 Calculation of the joint, conditional and marginal posterior distributions

First we calculate the joint posterior distribution conditional on each time series individually, ie:

$$p(\theta_j, \phi|y_j) = p(\phi)\mathcal{B}(\theta_j|\phi)\mathcal{L}(\theta_j)$$

We can then calculate each individual marginal probability density (ie. for each vector θ_j) as:

$$p(\theta_j|y_j) = \int p(\theta_j, \phi|y_j)\,\mathrm{d}\phi$$

To find the marginal posterior distribution for the hyperparameter vector $\boldsymbol{\phi}$, we can marginalise out θ by integrating over θ the full joint posterior distribution (ie. accross all time series):

$$p(\phi|y) = \int \prod_{j=1}^{J} p(\theta_j, \phi|y_j)\,\mathrm{d}\theta$$

1.4 Reduction of the computational burden

As the above procedure involves too many parameters (from the values mentioned along the docuement, there are approximately 10 times series of 10 data point each, along with 21 parameters for each times series and 105 hyperparameters), we will have to reduce the computational burden using several methods.

Assumptions

From Christian's sensitivity analysis, the system is mostly sensitive to only three parameters: k_6, d_1 and s_2. Hence, we can set these three parameters to be free parameters, and can fix the others to the average values estimated by GA. This would drastically reduce the dimensionality of the problem.

Analytical Tricks

Under work...

High-performance computational methods

As the estimation of the posterior distribution cannot be evaluated analytically, we will use a Monte Carlo method to sample from the posterior distribution. However, it is likely to be bimodal, a situation that usually traps the traditional Markov chain Monte Carlo (MCMC) method. We are hence considering a modified version of the MCMC developped by Liu et al [2], called Stochastic Approximation Monte Carlo (SAMC), that was specifically designed to be efficient with multi-modal distributions.

References

[1] Andrew Gelman et al. *Bayesian Data Analysis*. 2021.

[2] Baisen Liu, Liangliang Wang, and Jiguo Cao. In: *Monte Carlo Methods and Applications* 24.2 (2018), pp. 117–127. DOI: doi:10.1515/mcma-2018-0010. URL: https://doi.org/10.1515/mcma-2018-0010.

[3] Valderrama-Bahamóndez, Gloria I., and Holger Fröhlich. "MCMC Techniques for Parameter Estimation of ODE Based Models in Systems Biology". In: *Frontiers in Applied Mathematics and Statistics* 5 (2019). ISSN: 2297-4687. DOI: 10.3389/fams.2019.00055. URL: https://www.frontiersin.org/articles/10.3389/fams.2019.00055.