

Variational Inference Framework for Compartmental Disease Transmission Models

GitHub: <https://github.com/epidemic-research-team/>

June 28, 2020

1 Introduction

2 Model Goals Overview

The goal of this research is to utilise Variational Inference methods in order to increase the computational efficiency of the [?] paper.

- Approximate the posterior
- Apply VI methods on non-linear ODE system
- Evaluate the results

3 Literature Review

- Hensman, 2016 - Variational Bayesian non-Parametric Inference for Infectious Disease Models
- Blei, 2016 - Variational Inference: A Review for Statisticians

4 Method

4.1 ODE system model

The compartmental transmission model from [?] has a form as shown in Figure 1 [?, p.5].

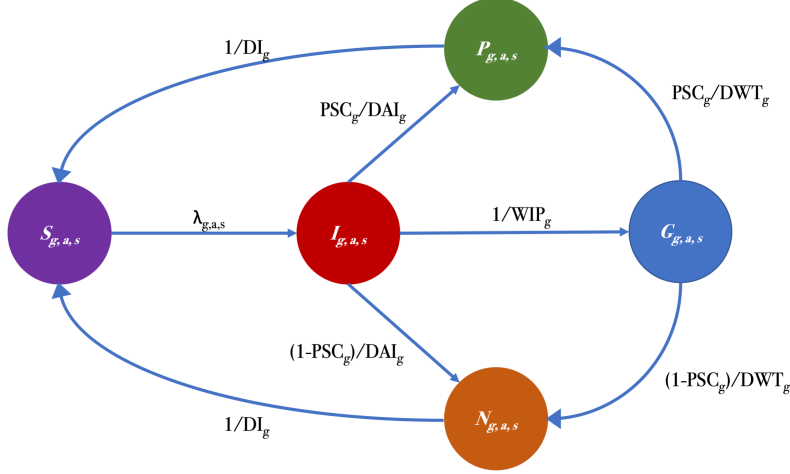


Figure 1: A compartmental Covid-19 transmission model.

We modify the original model by removing the indices g and s that indicate gender and level of social activity and retaining only the age related groups a , 19 age-groups in total. The system of ODEs is formulated as follows

$$\dot{S}_a = -\lambda_a(t)S_a + (P_a + N_a)/DI + \frac{1}{r}S_{a-1} - \frac{1}{r}S_a + \quad (1)$$

$$\frac{1}{R}(S_{19} + I_{19} + G_{19} + P_{19} + N_{19}) \times \\ \delta_1(a)(\pi_1\delta_1(s) + \pi_2\delta_2(s) + \pi_3\delta_3(s) + \pi_4\delta_4(s))$$

$$\dot{I}_a = \lambda_a(t)S_a - (1/WIP + 1/DAI)I_a + \frac{1}{r}I_{a-1} - \frac{1}{r}I_a \quad (2)$$

$$\dot{G}_a = I_a/WIP - G_a/DWT + \frac{1}{r}G_{a-1} - \frac{1}{r}G_a \quad (3)$$

$$\dot{P}_a = PSC(I_a/DAI + G_a/DWT) - P_a/DI + \frac{1}{r}P_{a-1} - \frac{1}{r}P_a \quad (4)$$

$$\dot{N}_a = (1 - PSC)(I_a/DAI + G_a/DWT) - N_a/DI + \frac{1}{r}N_{a-1} - \frac{1}{r}N_a \quad (5)$$

From the Figure 1, there are 5 non-overlapping compartments: S is for

individuals who are at risk of Covid-19 infection, I is for infected individuals, G is for infected who developed serious symptoms and required ICU admission, P is for those who have recovered, and are seropositive and immune, N is for individuals who are recovered, immune but seronegative. Movements between compartments are occurring at per capita rates specified by the following parameters: **PSC** is the probability of becoming seropositive, **WIP** is the Covid-19 incubation period, **DAI** is the duration of asymptomatic (i.e. without Covid-19) infection symptoms, **DWT** is the duration of treatment for Covid-19, and **DI** is the duration of immunity. Subscripts denote stratification of parameters: for example, \dot{I}_a means that in our model this variable is age-dependent. Finally, λ is the **force of infection** and is defined as

$$\lambda_a(t) = \sum_{a'} \left\{ M(t) \frac{I_{a'}(t)}{S_{a'}(t) + I_{a'}(t) + G_{a'}(t) + N_{a'}(t) + P_{a'}(t)} \right\} \quad (6)$$

with M a time-varying social mixing matrix, constructed by combining social information from the BBC Pandemic project¹ and from the Google mobility reports.

4.2 Model Inputs

Let X be a time dependent vector that includes the 5 states of the compartmental model, with $X_a(t) = [S_a(t), I_a(t), G_a(t), P_a(t), N_a(t)]$ and $t \in \{1, 2, \dots, T\}$. Also, assume that the observations of the model for time t are $D_a(t)$ as the number of infected.

4.3 Priors

Next step, we replicate work on a Bayesian modelling framework that will involve the treatment the parameters of the non-linear system of ODE equations describing the epidemic as unknown random variables. In total, Gareth et al. model [?] has 7 parameters with their relevant priors (Table 1)

¹For more information about the data and the construction of the social mixing matrix the reader should refer to the Social Mixing Matrix Literature Review and an Intuitive Approach for the \mathcal{R}_0

Parameter	Interpretation	Prior
Average incubation period	WIP	$\mathcal{Ga}(k_{WIP}, \theta_{WIP})$
Average duration of treatment	DWT	$\mathcal{Ga}(k_{DWT}, \theta_{DWT})$
Average duration of asymptomatic	DAI	$\mathcal{Ga}(k_{DAI}, \theta_{DAI})$
Average duration of immunity	DI	$\mathcal{U}(\alpha_{DI}, \beta_{DI})$
Probability of seroconversion	PSC	$\mathcal{Be}(\alpha_{PSC}, \beta_{PSC})$
Observation error for incidence	σ	$inv\mathcal{Ga}(k_\sigma, \theta_\sigma^{-1})$

Table 1: Non-linear ODE Model Parameter Priors

4.4 Likelihood

The likelihood function of the model has the following form

$$\begin{aligned} \mathcal{L}(params, X_a(1), \dots, X_a(T); D_a(1), \dots, D_a(T)) = \\ \Pi_a \Pi_t p(D_a(t) | X_a(t), params) \\ \Pi_a \Pi_t \mathcal{N}(D_a(t); \mu, \sigma^2) \end{aligned} \quad (7)$$

with $X_a(t)$ is the vector of states ($\mathcal{R}^{5 \times 19 \times T}$), the σ is the inverse Gamma as shown in Table 1

$$\sigma = \frac{\theta_\sigma^{-k_\sigma}}{\Gamma(k_\sigma)} \left(\frac{1}{x}\right)^{k_\sigma+1} \exp\left(-\frac{\theta_\sigma^{-1}}{x}\right) \quad (8)$$

$params$ stands for the parameters in Table 1, and the mean μ is a $\mathcal{R}^{19 \times T}$ matrix in the form

$$\mu_{a,t} = \frac{1}{WIP} \times \frac{I_a(t)}{S_a(t) + I_a(t) + G_a(t) + N_a(t) + P_a(t)} \times 1000 \quad (9)$$

4.5 Posterior

The posterior of a Bayesian model has the form

$$p(z|X, y) \propto p(y|X, \zeta)p(z) \quad (10)$$

by replacing the likelihood with Equation 7 and the priors (z) with Table 1, we solve for

$$p(z|X, y) \propto \Pi_a \Pi_t \mathcal{N}(D_a(t)|\mu, \sigma^2) \times \Pi_{i=prior} z_i \quad (11)$$

more analytically

$$\begin{aligned} p(\zeta|X, D) &\propto \\ \Pi_a \Pi_t \mathcal{N}(D_a(t)|\mu, \sigma^2) &\times \Pi_{i=prior} z_i = \\ \Pi_a \Pi_t \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(D_{a,t} - \mu_{a,t})^2}{\sigma^2}\right) \times \right. \\ &\left(\frac{1}{WIP} \times \frac{I_a(t)}{S_a(t) + I_a(t) + G_a(t) + N_a(t) + P_a(t)} \times 1000 \right) \times \\ &\frac{\theta_\sigma^{-k_\sigma}}{\Gamma(k_\sigma)} \left(\frac{1}{x}\right)^{k_\sigma+1} \exp\left(-\frac{\theta_\sigma^{-1}}{x}\right) \times \\ &\frac{1}{\Gamma(k_{WIP})\theta^{k_{WIP}}} x^{k_{WIP}-1} \exp^{-\frac{x}{\theta_{WIP}}} \times \\ &\frac{1}{\Gamma(k_{DWT})\theta^{k_{DWT}}} x^{k_{DWT}-1} \exp^{-\frac{x}{\theta_{DWT}}} \times \\ &\frac{1}{\Gamma(k_{DAI})\theta^{k_{DAI}}} x^{k_{DAI}-1} \exp^{-\frac{x}{\theta_{DAI}}} \times \\ &\frac{1}{\beta_{DI} - \alpha_{DI}} \times \\ &\frac{x^{\alpha_{PSC}-1} (1-x)^{\beta_{PSC}-1}}{B(\alpha_{PSC}, \beta_{PSC})} \end{aligned} \quad (12)$$

apart from the complexity of the Equation 12, the most difficult part is to find the value that normalises the posterior or thhe marginal distribution of $p(D)$

$$\begin{aligned}
p(D) = & \quad (13) \\
& \int_{\mu} \int_{\sigma} \int_{WIP} \int_{DWT} \int_{DAI} \int_{DI} \int_{PSC} \Pi_a \Pi_t \mathcal{N}(D_a(t) | \mu, \sigma^2) \times \Pi_{i=priors} z_i = \\
& \Pi_a \Pi_t \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(D_{a,t} - \mu_{a,t})^2}{\sigma^2}\right) \times \right. \\
& \left. \left(\frac{1}{WIP} \times \frac{I_a(t)}{S_a(t) + I_a(t) + G_a(t) + N_a(t) + P_a(t)} \times 1000 \right) \right] \times \\
& \frac{\theta_{\sigma}^{-k_{\sigma}}}{\Gamma(k_{\sigma})} \left(\frac{1}{x}\right)^{k_{\sigma}+1} \exp\left(-\frac{\theta_{\sigma}^{-1}}{x}\right) \times \\
& \frac{1}{\Gamma(k_{WIP})\theta^{k_{WIP}}} x^{k_{WIP}-1} \exp^{-\frac{x}{\theta_{WIP}}} \times \\
& \frac{1}{\Gamma(k_{DWT})\theta^{k_{DWT}}} x^{k_{DWT}-1} \exp^{-\frac{x}{\theta_{DWT}}} \times \\
& \frac{1}{\Gamma(k_{DAI})\theta^{k_{DAI}}} x^{k_{DAI}-1} \exp^{-\frac{x}{\theta_{DAI}}} \times \\
& \frac{1}{\beta_{DI} - \alpha_{DI}} \times \\
& \frac{x^{\alpha_{PSC}-1} (1-x)^{\beta_{PSC}-1}}{B(\alpha_{PSC}, \beta_{PSC})} d\mu d\sigma dWIP dDWT dDAI dDI dPSC
\end{aligned}$$

4.6 Variational Inference

To solve the inference problem and compute the density over parameters for the epidemiological model we start with the Bayesian formula as in Equation 10. The complexity of the marginal distribution (Equation 13) means that we have two ways to compute the results, use MCMC sampling methods (Gibb's, Hamiltonian) or use approximate methods

First, we need to specify a family of variational distributions \mathcal{Q} that the model would select from in order to approximate the posterior distribution over the parameters. In order to find the best candidate $q(z) \in \mathcal{Q}$, with z as the latent parameters we need to solve the following optimization, using the KL divergence, the distance between the selected approximate distribution $q(z)$ and the posterior $p(z|X, y)$

$$q^*(z) = \arg_{q(z) \in \mathcal{Q}} \min KL(q(z) || p(z|X, y)) \quad (14)$$

where $p(z|X, y)$ is the Equation 12. The complexity of the variational family of distributions \mathcal{Q} affects the results and the computational needs of the model. We break down the KL divergence into components

$$\begin{aligned} KL(q(z) || p(z|y, X)) &= \int q(z) \log \left(\frac{q(z)}{p(z|y, X)} \right) dz \\ &= \int q(z) \log(q(z)) dz - \int q(z) \log(p(z|y, X)) dz \end{aligned} \quad (15)$$

but again, we experience the same complex problem as we have to compute the posterior $\log(p(z|y, X))$. For simplicity we set $KL(q(z) || p(z|y, X))$ as KL_{post} . The next step is to expand the conditional $\log(p(z|y, X))$ according to the Equation 10

$$KL_{post} = \int q(z) \log(q(z)) dz - \int q(z) \log \left(\frac{p(y|z, X)p(z)}{p(y|X)} \right) dz \quad (16)$$

which becomes

$$\begin{aligned} KL_{post} &= \\ &= \int q(z) \log(q(z)) dz - \int q(z) \log(p(y|z, X)) dz \\ &\quad - \int q(z) \log(p(z)) dz + \log(p(y|X)) \end{aligned} \quad (17)$$

the RHS of the equation is a summary of expectations with regards to the approximate distribution $q(z)$; we have the expectation of the log-marginal distribution $q(z)$, the log-likelihood of the model $p(y|z, X)$ or the knowledge we get from our current data, the expectation of the log-transformed prior beliefs about the parameters $p(z)$ and finally the expectation of the log-marginal likelihood $p(y|X)$. In order to solve this system we need to rearrange the terms of the formula,

$$\begin{aligned}
KL_{post} - \log(p(y|X)) &= \\
&= \int q(z) \log(q(z)) dz - \int q(z) \log(p(y|z, X)) dz - \int q(z) \log(p(z)) dz \\
&= \int q(z) \log\left(\frac{q(z)}{p(z)}\right) dz - \int q(z) \log(p(y|z, X)) dz \\
&= KL(q(z)||p(z)) - \mathbb{E}_q(\log(p(y|z, X)))
\end{aligned} \tag{18}$$

the RHS of the equation shows the entropy $KL(q(z)||p(z))$ and the average energy $\mathbb{E}_q(\log(p(y|z, X)))$. By reverting the signs we get

$$\log(p(y|X)) - KL_{post} = \mathbb{E}_q(\log(p(y|z, X))) - KL(q(z)||p(z)) \tag{19}$$

we set the ELBO (taking into account Jensen's inequality), which is our objective function that we need to optimize

$$ELBO(q) = \mathbb{E}_q(\log(p(y|z, X))) - KL(q(z)||p(z)) \tag{20}$$

therefore, we replace Equation 14 with the new objective function

$$q^*(z) = \arg_{q(z) \in \mathcal{Q}} \max ELBO(q) \tag{21}$$

instead of minimizing the posterior KL we instead maximize the ELBO and the final formula becomes

$$\log(p(y)) - KL(q(z)||p(z|y, X)) \geq ELBO(q) \tag{22}$$

In our model, this means that the marginal (Equation 13) of the data y minus the KL difference between a preferred variational distribution of the parameters $q(z)$ and the posterior Equation 12 equals the lower bound. By increasing the ELBO we minimise the KL of the parameters.