Master's Thesis in requirement for partial completion

The University of Melbourne

Jacob Cumming

September 2023

Contents

List of Figures

Chapter 1

Epidemiological Modelling

In order to study the behaviour and characteristics of disease spread and eradication, compartmental epidemiological models have been developed. They seek to simplify the dynamics of a disease down to a mathematically representable form. Inference on these models allow for an understanding of how the modelled disease spreads, and allows an assessment of how effective differing disease interventions (such as treatments or vaccinations) may be without the need for large long term trials. Models can also simulate various scenarios such as increases or decreases in viral transmission.

Simple compartmental disease models assume individuals can be only be in one of a finite number of states (which are called compartments). These compartments usually correspond to a state of disease. Some simple common compartments include:

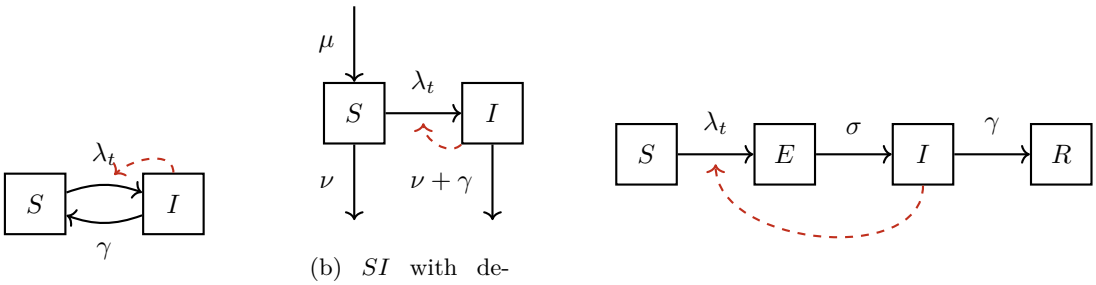
- S - Susceptable: at risk of contracting the disease
- E - Exposed: contracted the disease but not yet transmitting it
- I - Infectious (also called Infected): at risk of transmitting the disease
- R - Recovered: neither at risk of contracting or transmitting the disease.

The number of people in each compartment at time t is a (possibly non-deterministic) function of time t, which we indicate as a subscript t (eg. S_t is the number of susceptibles at time t). Models are routinely described by the compartments they contain. For example, an SIS model, is a model with the susceptable and infectious compartments. Recovering from the infection leaves you susceptible to reinfection (for example most sexually transmitted diseases (Keeling and Rohani 2008, p. 56)), and is graphically depicted in Figure 1.1a.

Furthermore we can also include demography into a compartmental model. Diseases that infect the individual until the time of death such as bovine spongiform encephalopathy (BSE commonly known as mad cow disease) may be modelled using an SI model with demography (birth and death rates), as depicted in Figure 1.1b (Hagenaars, Donnelly, and Ferguson 2006).

Childhood diseases such as varicella (chickenpox) which give lifetime immunity after infection can be modelled using an $SEIR$ model (see Figure 1.1c), particularly when modelling a local outbreak setting (for example Zha et al. 2020 used the $SEIR$ model to model a school outbreak of varicella). Not including demograpy is usually appropriate when disease induced mortality is low.

The number (and names) of compartments can be extended and configured as needed, and compartments could be added for vaccinated individuals, quarantined individuals and so on.



(a) SIS model

(b) SI with demography model

(c) $SEIR$ model

Figure 1.1: Some simple model schematics, with varying numbers of compartments: S (susceptable), E (exposed), I (infectious) and R (recovered). The force of infection λ_t is usually a function of I_t, depicted by the dashed red lines. μ and ν are natural birth and death rates respectively. γ is the rate of progression out of the infectious state. In each of these models the physical interpretation differs slightly. In the SIS and $SEIR$ models, it is the rate at which individuals move from infectious to susceptable again or into lifelong immunity, whereas in the SI with demography model, it can be interpreted as the increase to the rate of death attributable to disease induced mortality. σ is the rate of progression from a state of latent infection to becoming infectious.

By convention N_t (often simply N in models with a closed population) is the total number of individuals in the model, the sum of all compartments.

1.1 Deterministic ODE models

Diseases are often simulated as deterministic ordinary differential equations. For the examples below we assume that the force of infection λ_t is proportional to the number of people in I, such that $\lambda_t := \beta \frac{I_t}{N_t}$. β can be interpreted as the average number of people that an individual interacts with per day in a way such that disease would be spread in that interaction per unit of time t. This is sometimes called the effective contact rate. Therefore since $\frac{I_t}{N_t}$ is the probability that a randomly selected individual is infectious, $\beta \frac{I_t}{N_t}$ can be interpreted as the average number of people that a person interacts with each day who are infectious in a way that they would pass on the infection. In different diseases β varies dramatically, as some diseases need prolonged exposure or sexual contact to transmit, whereas some are very highly transmittable, and so will have very low β. Implicitly there is also a (sometimes poor) assumption of complete uniformly random mixing of people. Note that for this thesis we assume that β is frequency dependent (people interact with the same number of people regardless of population size) as opposed to density dependent (people interact with a number of people proportional to population size, in which case $\lambda := \beta I_t$.).

The SIS model depicted in Figure 1.1a, the ordinary different equations (ODEs) governing the model could be

$$\frac{\mathrm{d}S_t}{\mathrm{d}t} = -\lambda S_t + \gamma I_t = -\beta \frac{I_t}{N} S_t + \gamma I \tag{1.1}$$

$$\frac{\mathrm{d}I_t}{\mathrm{d}t} = \lambda S_t - \gamma I_t = \beta \frac{I_t}{N} S_t - \gamma I. \tag{1.2}$$

With a stated assumption that population size is closed, equation 1.1 fully describes the model.

The system of ODEs that describe the SI with demography model is

$$\frac{\mathrm{d}S_t}{\mathrm{d}t} = \mu N_t - \lambda_t S_t - \nu I_t = \mu N_t - \beta \frac{I_t}{N_t} S_t - \nu I \tag{1.3}$$

$$\frac{\mathrm{d}I_t}{\mathrm{d}t} = \lambda_t S_t - (\gamma + \nu) I_t = \beta \frac{I_t}{N_t} S_t - (\gamma + \nu) I. \tag{1.4}$$

Here is it important to note that N_t is not necessarily constant.

Finally, the system of ODEs that describe the $SEIR$ model is

$$\frac{\mathrm{d}S_t}{\mathrm{d}t} = -\lambda_t S_t - \nu I \qquad = -\beta \frac{I_t}{N} S_t + \gamma I_t \tag{1.5}$$

$$\frac{\mathrm{d}E_t}{\mathrm{d}t} = \lambda_t S_t - \omega E_t \qquad = \beta \frac{I_t}{N} S_t - \omega I_t \tag{1.6}$$

$$\frac{\mathrm{d}I_t}{\mathrm{d}t} = \omega E_t - \gamma I_t \tag{1.7}$$

$$\frac{\mathrm{d}R_t}{\mathrm{d}t} = \gamma I_t \tag{1.8}$$

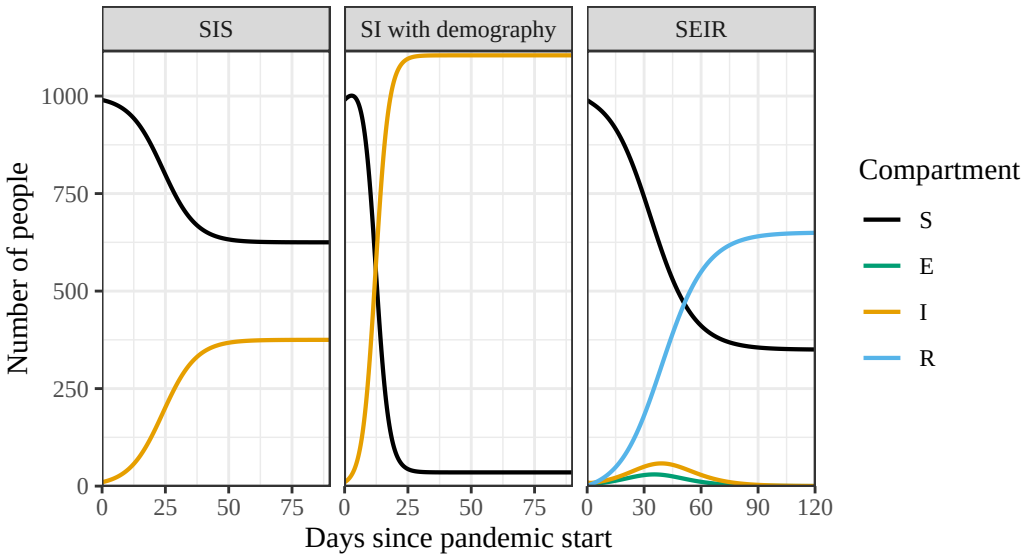


Figure 1.2: Solutions to the ordinary differential equations describing the models depicted in Figure 1.1. The initial infectious population was $I_0 = 10$, with $S_0 = 990$. In the $SEIR$ model, $E_0 = R_0 = 0$. For all models $\beta = 0.4$. For the SIS and SI model with demography $\gamma = 1/4$. For the SI model with demography $\mu = 0.012$, and $\nu = 0.0012$. For the $SEIR$ model, $\gamma = 1/90$, and $\sigma = 1/2$.

After specifying the initial for each compartment the ordinary differential equations have a deterministic output, such as in 1.2.

1.2 Stochastic models

Motivating the form of the stochastic model

To establish a relationship between the deterministic and stochastic disease models, we first need to establish Poisson point processes and their properties.

Definition 1.1 (Poisson Point Process)**.** $\{\mathcal{N}(t)\}_{t\geq 0}$ *is a* (stationary) Poisson point process *with intensity* λ *if*

1. $\mathcal{N}(0) = 0$
2. $\mathcal{N}(t_1) - \mathcal{N}(t_0), \mathcal{N}(t_2) - \mathcal{N}(t_1), \ldots, \mathcal{N}(t_n) - \mathcal{N}(t_{n-1})$ *are independent for* $0 \leq t_0 < t_1 < \cdots < t_{n-1} < t_n$
3. $\mathcal{N}(t_2) - \mathcal{N}(t_1) \sim \text{Pois}(\lambda(t_2 - t_1)), 0 \leq t_1 < t_2$.

Deterministic ODE models are appropriate to study a kind of aggregate disease spread behaviour, and well approximate real world behaviour when the numbers in each compartment are large, however at the start of an epidemic, when the number of infected individuals is small the behaviour of the epidemic may vary significantly. It is possible that if the average number of people that an infectious individual infects near the beginning of the epidemic (formally referred to as R_0) is close to 1, then the disease may die out or become stable. Under the deterministic SIS model described by equations 1.1 and 1.2, consider the model at time t^* the instantaneous rate at which S is decreasing is $\beta \frac{I_{t^*}}{N} S_{t^*}$. In other words, one individual leaves the S compartment every $\beta \frac{I_{t^*}}{N} S_{t^*}$ units of time. We can consider a Poisson point process $\{\mathcal{N}_1(t - t^*)\}_{t\geq t^*}$ with intensity $\beta \frac{I_{t^*}}{N} S_{t^*}$ corresponding to the count of the number of individuals who have left S and entered I t units of time since t^*.

$$\frac{\mathrm{d}\mathbb{E}(\mathcal{N}_1(t^*))}{\mathrm{d}t^*} = \lim_{\delta \to 0} \frac{\mathbb{E}\left(\mathcal{N}(t^* + \delta) - \mathcal{N}(t^*)\right)}{\delta} = \frac{\beta \frac{I_{t^*}}{N} S_{t^*} (t^* + \delta - t^*) \beta \frac{I_{t^*}}{N} S_{t^*}}{\delta} = \beta \frac{I_{t^*}}{N} S_{t^*}.$$

Under the same deterministic formulation of the model, the instantaneous rate into S at t^* is γI_{t^*}. Therefore as above we can construct a Poisson point process $\{\mathcal{N}_2(t - t^*)\}_{t\geq t^*}$ with rate γI_{t^*} describing the number of recoveries from I to S, with $\frac{\mathrm{d}\mathbb{E}(\mathcal{N}_2(t^*))}{\mathrm{d}t^*} = \gamma I_{t^*}$. Combining the two processes, we can see that the rate of change in the average number of people in S is

$$\frac{\mathrm{d}\mathbb{E}(\mathcal{N}_2(t^*) - \mathcal{N}_1(t^*))}{\mathrm{d}t^*} = \frac{\mathrm{d}\mathbb{E}(\mathcal{N}_2(t^*)) - \mathrm{d}\mathbb{E}(\mathcal{N}_1(t^*))}{\mathrm{d}t^*} = -\beta \frac{I_t}{N} S_t + \gamma I = \frac{\mathrm{d}S_t}{\mathrm{d}t}.$$

Therefore we can create a stochastic model where the local average in each compartment matches the behaviour of the ODE model at the same state. We do this first by formulating the model as a random vector $\{\mathbf{C}_t\}_{t\geq 0} = \{C_1(t), C_2(t), \ldots, C_n(t)\}_{t\geq 0}$ where $C_i : \mathbb{R} \to \mathbb{N} \cup \{0\}$, is the number of people in compartment C_i, and for any fixed t, $\{C_1(t), C_2(t), \ldots, C_n(t)\}$ is a random variable describing the state of the model. For example in a model with S and I compartments, $\{\mathbf{C}_t\}_{t\geq 0} := \{S_t, I_t\}_{t\geq 0}$. $\{\mathbf{C}_t\}_{t\geq 0}$ is a continuous time Markov chain (see Definition 3.3) with transition kernel corresponding to the rates of the model. For example, in the SI model with demography in Figure 1.1b the transition rates are:

- $\{s, i\}$ to $\{s + 1, i\}$ has rate $\mu(s + i)$
- $\{s, i\}$ to $\{s - 1, i\}$ has rate νs

- $\{s, i\}$ to $\{s-1, i+1\}$ has rate $\beta \frac{i}{i+s} s$
- $\{s, i\}$ to $\{s, i-1\}$ has rate $(\nu + \gamma) i$.

We can interpret each transition as a seperate events, each behaving as independent Poisson point processes until the time of the first transition. Therefore at time t^* we have the Poisson point processes:

- $\{\mathcal{E}_1(t)\}_{t \geq 0}$: the number of births into S after time t^* with intensity μN_{t^*}
- $\{\mathcal{E}_2(t)\}_{t \geq 0}$: the number of deaths in S after time t^* with intensity νS_{t^*}
- $\{\mathcal{E}_3(t)\}_{t \geq 0}$: the number of infections after time t^* with intensity $\beta \frac{I_{t^*}}{N_{t^*}} S_{t^*}$
- $\{\mathcal{E}_4(t)\}_{t \geq 0}$: the number of deaths from I after time t^* with intensity $(\nu + \gamma) I_{t^*}$.

Theorem 1.1 (Sums of Independent Poisson Point Processes)**.** *Given independent Poisson point processes* $\{\mathcal{N}_1(t)\}_{t \geq 0}, \{\mathcal{N}_2(t)\}_{t \geq 0}, \ldots, \{\mathcal{N}_n(t)\}_{t \geq 0}$, *with intensities* $\lambda_1, \lambda_2, \ldots, \lambda_n$,

$$\{\mathcal{N}(t)\}_{t \geq 0} := \{\mathcal{N}_1(t) + \mathcal{N}_2(t) + \cdots + \mathcal{N}_n(t)\}_{t \geq 0}$$

is a Poisson point process with intensity $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.

Proof. We show that $\{\mathcal{N}(t) := \{\mathcal{N}_1(t) + \mathcal{N}_2(t) + \cdots + \mathcal{N}_n(t)\}_{t \geq 0}\}$ meets each component of Definition 1.1.

1. $\mathcal{N}(0) := \mathcal{N}_1(0) + \mathcal{N}_2(0) + \cdots + \mathcal{N}_n(0) = 0$ since $\mathcal{N}_i(0) = 0$ by definition of a Poisson point process.
2. We show that $\mathcal{N}(t_1) - \mathcal{N}(t_0), \mathcal{N}(t_2) - \mathcal{N}(t_1), \ldots, \mathcal{N}(t_n) - \mathcal{N}(t_{n-1})$ with $0 \leq t_0 < t_1 < \cdots < t_n$ are independent.

   $$\mathcal{N}(t_i) - \mathcal{N}(t_{i-1}) = \underbrace{[\mathcal{N}_1(t_i) - \mathcal{N}_1(t_{i-1})]}_{X_{i1}} + \underbrace{[\mathcal{N}_2(t_i) - \mathcal{N}_2(t_{i-1})]}_{X_{i2}} + \cdots + \underbrace{[\mathcal{N}_n(t_i) - \mathcal{N}_n(t_{i-1})]}_{X_{in}}$$

   X_{ik} is independent of $X_{j\ell}$ for $k \neq \ell$ since $\mathcal{N}_k$ and $\mathcal{N}_\ell$ are independent processes. X_{ik} is independent of X_{jk} for $i \neq j$ by the second property of Definition 1.1. Therefore all X_{ik} are independent of $X_{j\ell}$ for $i \neq j$, and all j, k. Hence $\mathcal{N}(t_1) - \mathcal{N}(t_0), \mathcal{N}(t_2) - \mathcal{N}(t_1), \ldots, \mathcal{N}(t_n) - \mathcal{N}(t_{n-1})$ with $0 \leq t_0 < t_1 < \cdots < t_n$ are independent.
3. For fixed $t_1 < t_2$, and $i \in \{1, 2, \ldots, n\}$,

   $$\mathcal{N}_i(t_2) - \mathcal{N}_i(t_1) \sim \text{Pois}((t_2 - t_1)\lambda_i).$$

   Consider the associated moment generating function of $\mathcal{N}_i(t_2) - \mathcal{N}_i(t_1)$,

   $$M_i(z) := \exp(\lambda_i (t_2 - t_1)(\exp(z) - 1)).$$

   Therefore the moment generating function of

   $$\mathcal{N}(t_2) - \mathcal{N}(t_1) = [\mathcal{N}_1(t_2) - \mathcal{N}_1(t_1)] + [\mathcal{N}_2(t_2) - \mathcal{N}_2(t_1)] + \cdots + [\mathcal{N}_n(t_2) - \mathcal{N}_n(t_1)]$$

   is

   $$M(z) := \prod_{i=1}^{n} M_i(z) = \exp[(\lambda_1 (t_2 - t_1) + \lambda_2 (t_2 - t_1) + \cdots + \lambda_n (t_2 - t_1))(\exp(z) - 1)].$$

   Therefore $\mathcal{N}_1(t) + \mathcal{N}_2(t) + \cdots + \mathcal{N}_n(t) \sim \text{Pois}((\lambda_1 + \lambda_2 + \cdots + \lambda_n)t)$ by the uniqueness of the moment generating function.

□

Theorem 1.2 (Time to First Event in Poisson Point Process)**.** *Given a Poisson point process* $\{\mathcal{N}(t)\}_{t\geq 0}$ *with intensity* λ*, let* $\tau = \inf\{t|\mathcal{N}(t_0+t)-\mathcal{N}(t_0)=1, t>0\}$*.* $\tau \sim \text{Exp}(\lambda)$ *for* $t_0 \geq 0$

Proof.

$$\Pr(\tau > x) = \Pr(\mathcal{N}(t_0+x)-\mathcal{N}(t_0)=0) = \frac{(\lambda x)^0 e^{-\lambda x}}{0!} = e^{-\lambda x}$$

□

By Theorem 1.1 and Theorem 1.2,

$$\{\mathcal{E}(t)\}_{t\geq 0} := \{\mathcal{E}_1(t)+\mathcal{E}_2(t)+\mathcal{E}_3(t)+\mathcal{E}_4(t)\}_{t\geq 0}$$

is a Poisson point process with intensity

$$\mu N_{t^*} + \nu S_{t^*} + \beta \frac{I_{t^*}}{N_{t^*}} S_{t^*} + (\nu+\gamma) I_{t^*},$$

and the time to the next event is random variable distributed

$$\text{Exp}(\mu N_{t^*} + \nu S_{t^*} + \beta \frac{I_{t^*}}{N_{t^*}} S_{t^*} + (\nu+\gamma) I_{t^*}).$$

Theorem 1.3 (Probability of ith Poisson Process Generating the Next Event)**.** *Consider independent Poisson point processes*

$$\{\mathcal{N}_1(t)\}_{t\geq 0},\ \{\mathcal{N}_2(t)\}_{t\geq 0},\ \ldots,\ \{\mathcal{N}_n(t)\}_{t\geq 0}$$

having intensities $\lambda_1, \lambda_2, \ldots, \lambda_n$*. For fixed* t_0*, let* $\tau_i := \inf\{t|\mathcal{N}(t_0+t)-\mathcal{N}(t_0)=1\}$*. Then*

$$\Pr(\min_i \tau_i = \tau_j) = \frac{\lambda_j}{\sum_{i=1}^n \lambda_i}.$$

Proof. By Theorem 1.2, $\tau_i \sim \text{Exp}(\lambda_i)$. Therefore

$$\begin{aligned}
\Pr(\min_i \tau_i = \tau_j) &= \int_0^\infty \Pr(\{\tau_i = x\} \cup \bigcup_{j\neq i} \{\tau_j > x\})\,\mathrm{d}x \\
&= \int_0^\infty \Pr(\{\tau_i = x\} \cup \bigcup_{j\neq i} \{\tau_j > x\})\,\mathrm{d}x \\
&= \int_0^\infty \Pr(\tau_i = x) \times \prod_{j\neq i} \Pr(\tau_j > x)\,\mathrm{d}x && \text{(by independence)} \\
&= \int_0^\infty \lambda_i \exp(-\lambda_i x) \times \prod_{j\neq i} \exp(-\lambda_j x)\,\mathrm{d}x \\
&= \lambda_i \int_0^\infty \exp(-(\sum_{i=1}^n \lambda_j)x)\,\mathrm{d}x \\
&= \lambda_i \left[\frac{\exp(-(\sum_{i=1}^n \lambda_j)x)}{\sum_{i=1}^n \lambda_j}\right]_0^\infty \\
&= \frac{\lambda_i}{\sum_{i=1}^n \lambda_j}
\end{aligned}$$

□

1.3 Doob-Gillespie Algorithm

All of this leads naturally to a common method of simulating the stochastic model. The Doob-Gillespie algorithm (often just called the Gillespie algorithm) is an algorithm that simulates a stochastic realisation of a model given a set of starting conditions.

Algorithm 1 The Doob-Gillespie Algorithm

1: Initialise time $t \leftarrow 0$ and initial state of the model $\mathbf{C}(0) := \{C_1(0), C_2(0), \ldots, C_n(0)\}$
2: **while** termination condition not met **do**
3: Calculate intensities λ_i for all possible events $\mathcal{E}_i$
4: Calculate total intensity $\lambda = \sum_i \lambda_i$
5: Generate $\Delta t \sim \text{Exp}(\lambda)$
6: Choose event E_i with probability $\frac{\lambda_i}{\lambda}$
7: Update time $t \leftarrow t + \Delta t$
8: Update state of $\mathbf{C}(t + \delta t) \leftarrow \mathbf{C}(t) +$ change in state due to event $\mathcal{E}_i$
9: **end while**

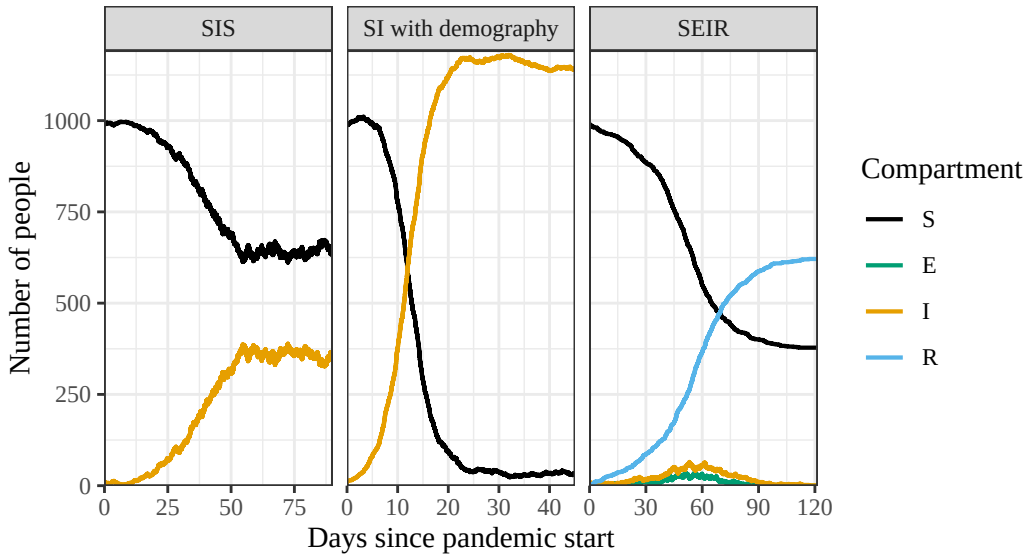


Figure 1.3: Exact stochastic simulations of the 3 different models using Algorithm 1. The parameters used were identical to those in Figure 1.2

1.4 τ-leaping

τ-leaping exploits the local Poisson point process like behaviour of epidemiological models. Consider the SIS model, when $S_t = I_t = 10000$. Events happen at a very high rate, meaning the Δt found in each step of the Doob-Gillespie algorithm will be very small, but the rates also change a negligible amount after each event (compare $\gamma \times 10000$ to $\gamma \times 10001$ or $\gamma \times 9999$). Therefore we can approximate the number of events in a short time period τ as a Poisson point process

with the total intensity $\lambda = \sum_i \lambda_i$ at time t, with the probability of any one event having the same probability as above of $\frac{\lambda_i}{\lambda}$. Therefore we have the following algorithm.

Algorithm 2 τ-Leaping Algorithm

Initialise time $t \leftarrow 0$ and initial state of the model $\mathbf{C}(0) := \{C_1(0), C_2(0), \ldots, C_n(0)\}$
while termination condition not met **do**
    Calculate intensities λ_i for all possible events $\mathcal{E}_i$
    Calculate total intensity $\lambda = \sum_i \lambda_i$
    Choose a suitable time step τ (this can be deterministic or adaptive)
    Calculate Poisson random variable $X \sim \text{Poisson}(\lambda)$
    for i in 1 to X **do**
        Choose event E_i with probability $\frac{\lambda_i}{\lambda}$
        Update state of $\mathbf{C}(t+\tau) \leftarrow \mathbf{C}(t) +$ change in state due to event $\mathcal{E}_i$
    end for
    Update time $t \leftarrow t + \tau$
end while

Chapter 2

Malaria and Malaria Models

2.1 Malaria

Malaria kills around 600,000 people each year, with over 75% of deaths occuring in children under 5 years old (World Health Organization 2022).

Plasmodium Life Cycle

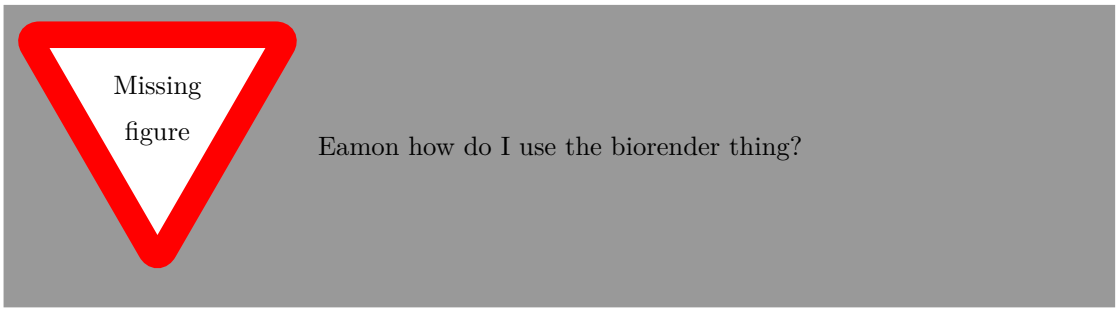


Figure 2.1: The malaria lifecycle.

Malaria is a vector borne disease, needing both human (or other vertebrate) and mosquito hosts to complete it's lifecycle (Figure 2.1). Six species of the unicellular parasite able to infect humans (Milner 2018). Although *Plasmodium falciparum* is responsible for around 90% of total human malaria deaths, outside of Africa *Plasmodium vivax* is the leading cause of malaria infection (Zekar and Sharman 2023; Adams and Mueller 2017). Sporozoites (a stage of the malaria parasite) enter the human blood stream via the skin after the female mosquito has a blood meal. From the blood stream they proceed to enter into the liver. Once a hepatocyte (liver cell) is invaded, the parasite will undergo asexual reproduction into up to 40,000 merozoites per hepatocyte, which are released into the blood stream . These merozoites then bind to, and invade erythrocytes (red blood cells), once again reproducing 16-32 fold in a process called schizogony. At this point, the erythrocyte membrane is ruptured, allowing for *Plasmodium* to invade new erythrocytes. Eventually, the merozoites undergo sexual differentiation, resulting in the sequestration and maturation of male and female gametocytes in the bone marrow, until they are released into the blood stream to be consumed by a mosquito during a blood feed where

it matures into sporozoites ready to reinfect a new vertebrate host when the mosquito next takes a blood feed (Cowman et al. 2016).

Illness, Treatment, and Immunity

The most common symptom of malaria infection in persons without natural or acquired immunity is fever. After treatment, fever will usually subside over a few days. In severe cases, malaria can lead to anemia, cerebral malaria (coma), and respiratory distress (Cowman et al. 2016).

In a population with stable malarial infection, immunity increases with age, with the proportion of severe cases negligible after age 10, and asymptomatic infection being the dominant infection type beyond age 15 (Cowman et al. 2016).

Control and Eradication

Widespread use of DDTs in the mid 20th century led to significant successes in some countries towards the control and eradication of malaria. In the 1980s and 1990s, drug resistant malaria led to a doubling of malaria-attributable death. Currently, the control techniques include insecticide treated bed nets, and a mixture of antimalarial drugs (Cowman et al. 2016).

Plasmodium vivax

Unlike *P. falciparum*, *P. vivax* has hypnozoites, which are a dormant liver stage of the parasite. These can remain dormant for weeks and even months, leading to recurrent infections and illness, possibily until the conditions for transmission are more favourable. In subtropical/temperate areas, the incubation periods can be between 8-12 months, compared to 3-4 weeks in tropical regions. Price et al. 2020. *P. vivax* also has lower levels of the blood stage parasite during infection, which means diagnosis is more difficult, and it has an increased proportion of asymptomatic cases (Adams and Mueller 2017).

It is likely that death and severe disease attributable to *P. vivax* has been traditionally underestimated. In view of recent evidence, the old notion that *P. vivax* is benign has become untenable (Cowman et al. 2016).

Motivating Malaria Models

Levels of asymptomatic cases and latent parasite (in the case of *P. vivax*) are impossible or difficult to experimentally determine without mass testing. By creating a model of the disease, and calibrating the model so that it simulates symptomatic case levels reported by health authorities, it is possible to estimate these previously 'hidden' levels. Furthermore, modelling malaria allows for modelling the effect of public health interventions such as mass treatment or testing, in order to determine an estimate of how effective the intervention may be, before large amounts of money are spent on trials.

2.2 Malaria Models

Ross-Macdonald

Modelling malaria presents an additional challenge, as the disease is transmitted from mosquito to human and human to mosquito, rather than having direct human to human transmission.

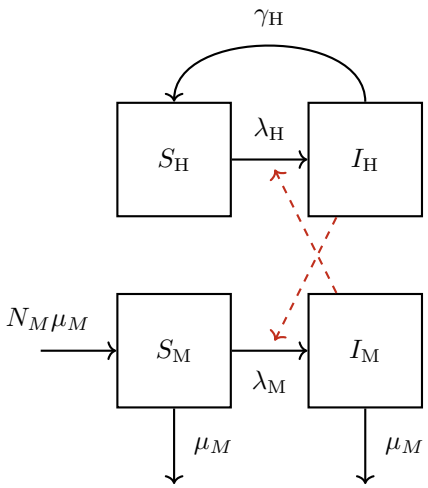


Figure 2.2: A simple Ross-Macdonald malaria model schematic, as described by Aron and May 1982. S_{H} and I_{H} are the number of susceptable and infected humans respectively, and S_{M} and I_{M} are the number of susceptable and infected mosquitos. The rate of human infection (λ_{H}) is dependant on I_{M}, and the rate of human infection (λ_{M}) is dependant on I_{H}.

The most simple Ross-Macdonald model is depicted in figure 2.2. The ODEs for this model are

$$\begin{aligned}
\frac{\mathrm{d}S_{\mathrm{H}}}{\mathrm{d}t} &= \gamma_{\mathrm{H}} I_{\mathrm{H}} - bT_{\mathrm{HM}} I_{\mathrm{M}} \frac{S_{\mathrm{H}}}{N_{\mathrm{H}}} \\
\frac{\mathrm{d}I_{\mathrm{H}}}{\mathrm{d}t} &= bT_{\mathrm{HM}} I_{\mathrm{M}} \frac{S_{\mathrm{H}}}{N_{\mathrm{H}}} - \gamma I_{\mathrm{H}} \\
\frac{\mathrm{d}S_{\mathrm{M}}}{\mathrm{d}t} &= N_{\mathrm{M}}\mu_M + \gamma_{\mathrm{M}} I_{\mathrm{M}} - bT_{\mathrm{MH}} S_{\mathrm{M}} \frac{I_{\mathrm{H}}}{N_{\mathrm{H}}} - S_{\mathrm{M}}\mu_M \\
\frac{\mathrm{d}I_{\mathrm{M}}}{\mathrm{d}t} &= bT_{\mathrm{MH}} S_{\mathrm{M}} \frac{I_{\mathrm{H}}}{N_{\mathrm{H}}} - \gamma_{\mathrm{M}} I_{\mathrm{M}}
\end{aligned}$$

where b is the biting rate per mosquito, and T_{HM} is the probability of tranmission to a human given a bite by an infectious mosquito, with T_{MH} being vice-versa. Note that it is $\frac{I_{\mathrm{H}}}{N_{\mathrm{H}}}$ in the mosquito dynamics. Biologically this is assuming the number of blood meals a mosquito takes per day is invariant to the size of the human population. Mosquitos don't 'recover' from malaria due to their short lifespans, but the births and deaths are mathematically equivalent to assuming that the rate of 'recovery' amongst mosquitos is $\mu_{\mathrm{M}} I_{\mathrm{M}}$ per unit time, with no population dynamics.

A Ross-Macdonald style model simplifies the lifecycle of malaria to the following four steps (Smith et al. 2012):

1. Malaria is transmitted to human (or vertebrate) via a blood feed.
2. Malaria proliferates in the human host until it circulates in the peripheral blood
3. A mosquito then takes a blood feed, ingesting the pathogen
4. Malaria develops within the mosquito host, progressing to its salivary glands, able to infect a human.

Models of *P. Vivax* Malaria

White Model

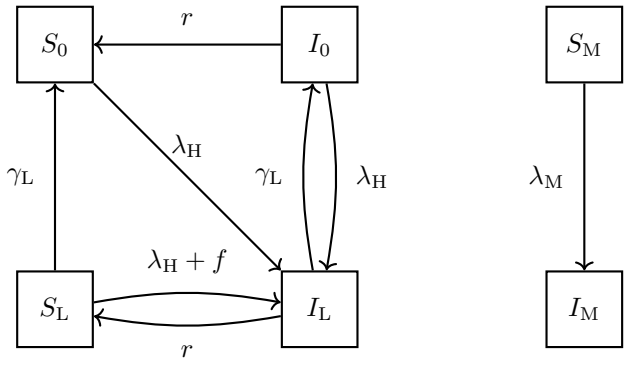


Figure 2.3: Diagram for *P. vivax* model in a tropical setting described by White et al. 2016. S and I are the number of susceptable and infected humans and mosquitos (denoted by subscript M). $\lambda_H = mabI_M$ and $\lambda_M = ac(I_0 + I_L)$

The White model - described in White et al. 2016 (tropical model) and depicted in Figure 2.3 - is characterised by the following ordinary differential equations:

$$
\begin{aligned}
\frac{\mathrm{d}S_0}{\mathrm{d}t} &= -\lambda_H S_0 + rI_0 + \gamma_L S_L \\
\frac{\mathrm{d}I_0}{\mathrm{d}t} &= -\lambda_H I_0 - rI_0 + \gamma_L I_L \\
\frac{\mathrm{d}S_L}{\mathrm{d}t} &= -\lambda_H S_L + rI_L - fS_L - \gamma_L S_L \\
\frac{\mathrm{d}I_L}{\mathrm{d}t} &= \lambda_H(S_0 + I_0 + S_L) - rI_L + fS_L - \gamma_L I_L \\
\frac{\mathrm{d}S_M}{\mathrm{d}t} &= g - \lambda_M(pS_M - (1-p)I_M) - gS_M \\
\frac{\mathrm{d}I_M}{\mathrm{d}t} &= \lambda_M(pS_M - (1-p)I_M) - gI_M. \qquad (I_0 + I_L = \text{total number of bloodstage infections})
\end{aligned}
$$

It modifies the Ross-Macdonald models, to capture the differences in disease progression between *P. vivax* and *P. falciparum*. In particular, the White model includes the dormant liver stage that is unique to *P. vivax*.

The model is comprised of six compartments:

1. S_0 **(Susceptible Individuals - No Latent Hypnozoite Liver Stage Infection)**: People in this compartment have no form of malarial infection. These people are susceptible to new malarial infections, and are infected into compartment I_L (with both blood and liver stage parasites) with rate λ_H.

2. I_L **(Infected Individuals - Both Blood Stage and Latent Hypnozoite Liver Stage Infection)**: Individuals in this compartment have both an active blood-stage infection,

and latent hypnozoite infection in the liver. They can progress to either I_0 through the clearance of liver stage infection with rate γ_L, or to S_L through clearance of blood stage infection with rate r.

3. I_0 **(Infected Individuals - Blood-Stage Infection Only)**: Those in this compartment have a blood-stage infection with no latent hypnozoite infection in the liver. They are be reinfected into I_L with rate λ_H, relapse with rate f. Blood-stage infection is cleared (moving into compartment S_0) with rate r.

4. S_L **(Susceptable Individuals - Blood-Stage Infection Only)**: Those in this compartment have latent hypnozoite infection in the liver without blood-stage infection. They get novel infection through a mosquito bite into I_L with rate λ_H, or hypnozoite activation with rate f. This means that those in S_L move to compartment I_L with total rate $\lambda_H + f$. Alternatively the hypnozoites are cleared from the liver (moving to compartment S_0) with rate γ_L.

5. S_M **(Susceptable Mosquitoes)**: Suspectable mosquitoes become infectious at rate $\lambda_M p$. They die at rate $g + \lambda_M(1-p)$. Since there is a constant mosquito population assumption, mosquitoes are born into this state at rate $g + \lambda_M$.

6. I_M **(Infectious Mosquitoes)**: Infectious mosquitos die at rate $g + \lambda_M(1-p)$.

$\lambda_H := mabI_M$ where m is the number of mosquitos per human (held constant since there is no birth or death in the human dynamics), a is the mosquito biting rate, and b is the probability that a human bitten by an infectious mosquito develops an infection.

$\lambda_M := ac(I_0 + I_L)$ where a is defined above, and c is the probability that a mosquito bite on an infectious mosquito causes the mosquito to become infectious. g can be interpreted as the natural birth/death rate for mosquitos. p is then the proportion of mosquitos that survive long enough after the initial infection that the parasite matures enough in the mosquito before becoming infectious to new susceptible humans. Under the assumption that time until parasite transmissability after infection in a mosquito is a constant n days, and that mosquitoes naturally die at rate g, $p = e^{-gn}$. To see this let $V \sim \text{Exp}(g)$, represent the lifespan of the mosquito. $\Pr(V > n) = 1 - F_V(n) = 1 - (1 - e^{-gn}) = e^{-gn}$.

$\lambda_M(1-p)$ can be interpreted as an additional rate of death, where of the mosquitos that would develop malaria after a bite, a proportion $1-p$ die instantly. This applies to both the susceptible and infectious mosquitoes. Presumably this approximates a model where mosquitoes are moved to an 'exposed' compartment for n time, after initial infection, however no justification is given in (White et al. 2016) for this additional parameter n. A more straightforward SI model could be constructed that absorbs c and n into the single parameter c^*, such that it becomes the proportion of mosquito bites on blood stage infectious humans that result in mosquito infection where the mosquito does not die before becoming infectious. With steady mosquito population, the mosquito dynamics would now be characterised by

$$\frac{\mathrm{d}I_M}{\mathrm{d}t} = \lambda_M^* S_M - gI_M \quad \text{where } \lambda_M^* := ac^*(I_0 + I_L).$$

By modelling both liver and bloodstage infection, blood stage infections from relapses can be captured in the dynamics, meaning it is possible to analyse case number data that may be confounded by relapses as well as novel infections.

This model does not account for continual depletion of liver stage parasites which would vary the rate of relapse over time (through clearance or relapse). It also does not directly model any interventions or case importations. The lack of population dynamics means the model may only

be useful on a small time scale. Finally, it doesn't account for any importation of disease from an outside area, so if $S_0 = 1$, *P. vivax* is presumed permanently eradicated.

Champagne Model

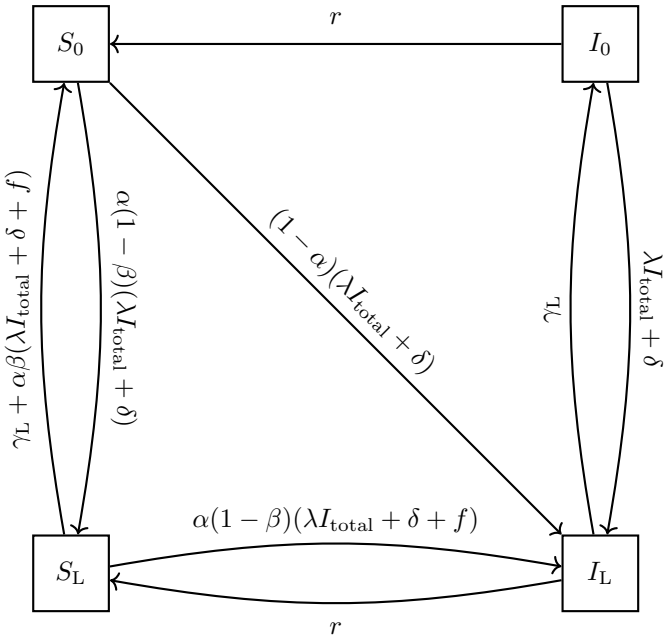


Figure 2.4: Diagram for *P. vivax* model described by Champagne et al. 2022. $I_{\mathrm{total}} = I_0 + I_{\mathrm{L}}$. Since the mosquito dynamics have been removed, λ now not has no dependencies on the number of infectious mosquitos.

The Champagne model - described in (Champagne et al. 2022) and diagrammatically depicted in figure 2.4 - both simplifies and extends the White model. The model assumes human to human transmission, removing mosquito dynamics, and extends it by adding in a rate of imported cases and treatment of malarial infection. It is characterised by the system of ordinary differential equations

$$\begin{aligned}
\frac{\mathrm{d}I_{\mathrm{L}}}{\mathrm{d}t} =& (1-\alpha)(\lambda I_{\mathrm{total}} + \delta)(S_0 + S_{\mathrm{L}}) + (\lambda I_{\mathrm{total}} + \delta)I_0 + (1-\alpha)fS_{\mathrm{L}} - \gamma_{\mathrm{L}} I_{\mathrm{L}} - rI_{\mathrm{L}} \\
\frac{\mathrm{d}I_0}{\mathrm{d}t} =& -(\lambda I_{\mathrm{total}} + \delta)I_0 + \gamma_{\mathrm{L}} I_{\mathrm{L}} - rI_0 \\
\frac{\mathrm{d}S_{\mathrm{L}}}{\mathrm{d}t} =& -(1-\alpha(1-\beta))(\lambda I_{\mathrm{total}} + \delta + f)S_{\mathrm{L}} + \alpha(1-\beta)(\lambda I_{\mathrm{total}} + \delta)S_0 - \gamma_{\mathrm{L}} S_{\mathrm{L}} + rI_{\mathrm{L}} \\
\frac{\mathrm{d}S_0}{\mathrm{d}t} =& -(1-\alpha\beta)(\lambda I_{\mathrm{total}} + \delta)S_0 + (\lambda I_{\mathrm{total}} + \delta)\alpha\beta S_{\mathrm{L}} + \alpha\beta f S_{\mathrm{L}} + \gamma_{\mathrm{L}} S_{\mathrm{L}} + rI_0
\end{aligned}$$

where $I_{\mathrm{total}} := I_0 + I_{\mathrm{L}}$.

The compartments have the same interpretation as in the White model, however the rates between compartments are significantly modified.

The new parameters are λ : the rate of infection, δ : importation rate, α : proportion of those infected but cleared of blood stage infections (through treatment), and β : a further proportion are also cleared of liver stage parasites, given that they were also cleared of blood stage infection (radical cure). In other words, the proportion of infected individuals $\alpha\beta$ are completely cured from liver and blood stage parasites. The model assumes treatment clears infection instantaneously. Individuals in S_{L} who relapse or get a new infection are assumed to be cured with the same proportions as new infections from S_0, but individuals in I_0 who are superinfected are assumed not to seek treatment.

In contrast to the White model, the Champagne model allows analysis of potential treatment interventions, or how much of an impact limiting the importation rate might have on case numbers (through border control/testing). Although the lack of mosquito dynamics simplifies the model and it's running, it is unrealistic. The model still has some of the same problems as the White model, such as not incorporating hypnozoite depletion rates and a lack of population dynamics, meaning all analytic results are done assuming the system is at equilibrium.

Chapter 3

Bayesian Parameter Estimation

TODO:

1. make all single digit numbers words

In classical statistics, $\mathbf{\Theta}$ is fixed, and the data Y is generated from a distribution depending on $\mathbf{\Theta}$. Frequentist estimators such as the maximum likelihood estimator, or least squares estimator as previously seen are point estimates. In contrast, inference under a Bayesian framework assumes that $\mathbf{\Theta}$ is also a random variable. This means although inference may be done on the same model, the Bayesian statistician will more often than not end up with a probability distribution describing $\mathbf{\Theta}$ given Y.

3.1 Monte Carlo Integration

Let us assume we have a random variable $X \sim p$, and we want to integrate $I = \int_{\mathcal{X}} f(x)p(x)\,\mathrm{d}x$ with relation to some known density p and function f. We can numerically approximate I by taking n i.i.d. samples $\{x_1, x_2, \ldots, x_n\}$ from the distribution specified by p, and approximating I by $I_n := \frac{1}{N}\sum_{i=1}^{n} f(x_i)$. By the strong law of large numbers, under specific conditions (for example $\sigma_f := \mathrm{Var}(f(X)) < \infty$ is sufficient) we have that $I_n \to I$ almost surely as $n \to \infty$. Furthermore by the central limit theorem we know that $\sqrt{n}(I_n - I) \sim N(0, \sigma_f)$ as $n \to \infty$. Intuitively, Monte Carlo integration works efficiently because the support set we numerically integrate over are likely to be points of higher probability density that integrating numerically over a set of uniformly distributed points.

3.2 Accept-Reject sampling methods

Accept-Reject methodology fundamentally relies on the theorem that $X \sim f$ is equivalent to simulating $(X, U) \sim \mathrm{Unif}\{(x, u)|0 < u < f(x)\}$. (Theorem 2.15 pg 47 Casella Roberts). In the case where a direct sampling method doesn't exist, we can use this theorem to our advantage. For example, if $a \leq X \leq b$ almost surely, then we can sample the pairs $X^* \sim \mathrm{Unif}(a, b)$, and then $U^* \sim \mathrm{Unif}(0, M)$, where $M \geq \sup_{a<x<b} f(x)$. To then get out samples to follow the distribution of (X, U) we then reject all samples (x_i, u_i) where $u_i > f(x_i)$. The distribution of X can then be described as follows

$$\begin{aligned}\Pr(X \le x) &= \Pr(X^* \le x | U^* \le f(X^*)) \\ &= \frac{\Pr(X^* \le x, U^* \le f(X^*))}{\Pr(U^* \le f(X^*))} \\ &= \frac{\int_a^x f(y)\,\mathrm{d}y/((b-a)*M)}{1/((b-a)*M)} \\ &= \int_a^x f(y)\,\mathrm{d}y \\ &= F(x)\end{aligned}$$

For example in finding samples from the distribution below, we generate 100 random pairs (X^*, U^*)

There is no need for the distribution to be uniform over the rectangle, and to decrease the number of rejections we can generate points uniformly under a function $h(x)$ that more closely resembles our target distribution as long as $f(x) \le h(x)$ for all x. If $h(x) = Mg(x)$ where $g(x)$ is a probability distribution function with known method of sampling, we can use this to generate via $Y \sim g$ and $U|Y \sim \text{Unif}(0, Mg(x))$ then we reject all pairs where $U|Y > f(Y)$.

So in the end we do

1. Generate $X \sim g$
2. Accept if $U \le f(X)/Mg(X)$, otherwise repeat 1.

If f is computationally expensive, but bounded below by g_l, we can

1. Generate $X \sim g$
2. Accept if $U \le g_l(X)/Mg(X)$
3. Accept if $U \le f(X)/Mg(X)$, otherwise repeat 1.

3.3 Essentials for MCMC

Definition 3.1 (Random Walk)**.** Random walk *(MCSM 6.1) is a sequence* $\{X_n\}$ *such that*

$$X_{n+1} = X_n + \varepsilon_n,$$

and ε_n *is independent from* $X_1, \ldots, X_n$*.*

Definition 3.2 (Transition Kernel)**.** *A* transition kernel *is a function* K *defined on* $\mathcal{X} \times \mathcal{B}(\mathcal{X})$ *such that*

1. $\forall x \in \mathcal{X},\ K(x, \cdot)$ *is a probability measure;*
2. $\forall A \in \mathcal{B}(\mathcal{X}),\ K(\cdot, A)$ *is measureable.*

(MCSM 6.2)

Definition 3.3 (Markov Chain)**.** *Given a transition kernel* K*, a sequence of* $X_0, X_1, \ldots, X_n, \ldots$ *of random variables is a* Markov chain *denoted by* (X_n) *if for any* t*,*

$$P(X_{k+1} \in A | X_0, X_1, \ldots, X_k) = P(X_{k+1} \in A | X_0, \ldots, X_k)$$

(MCMS 6.4)

Definition 3.4 (φ-irreducible)**.** *Given a measure φ, the Markov chain (X_n) with transition kernel K is* φ-irreducible *if, for every $B \in \mathcal{B}(\mathcal{X})$ with $\varphi(A) > 0$, there exists n s.t. $K^n(x, A) > 0$ for all $x \in \mathcal{X}$. (MCSM 6.13)*

Definition 3.5 (Atom)**.** *The Markov chain (X_n) has an* atom *$\alpha \in \mathcal{B}(\mathcal{X})$ if there exists an associated nonzero measure ν such that*

$$K(x, A) = \nu(A)$$

(MCSM 6.18)

Definition 3.6 (Harris Recurrent)**.** *A set A is* Harris recurrent *if $P_x(\eta_A = \infty) = 1$ for all $x \in A$. The chain (X_n) is Harris recurrent if there exists a measure ψ such that (X_n) is ψ-irreducible and for every set A with $\psi(A) > 0$, A is Harris recurrent. (MCSM 6.32)*

Definition 3.7 (Invariant (Measure))**.** *A σ-finite measure π is* invariant *for the transition kernel $K(\cdot, \cdot)$ if*

$$\pi(B) = \int_{\mathcal{X}} K(x, B)\pi(\mathrm{d}x), \forall B \in \mathcal{B}(\mathcal{X})$$

(MCSM 6.35) and π is a stationary distribution *when π is a probability measure. If such a π exists for a ψ-reducible chain, the chain is* positive*.*

Definition 3.8 (Ergodic)**.** *For a positive Harris recurrent chain (X_n), with invariant distribution π, an atom α is* ergodic *if*

$$\lim_{n\to\infty} |K^n(\alpha, \alpha) - \pi(\alpha)| = 0$$

(MCSM 6.47)

3.4 Motivating the MH/Gibbs Algorithm

use some of the MC theory to arrive at the acceptance probability in MH

3.5 Metropolis Hastings

Definition 3.9 (Markov Chain Monte Carlo Method)**.** *A* Markov chain Monte Carlo method *for simulating a distribution f is any method producing an ergodic Markov chain whose stationary distribution is f. (MCSM 7.1)*

Given a proposal distribution q, and a function f proportional to the posterior distribution, the Metropolis Hastings algorithm is as follows:

1. `Initialise` X_0
2. `Generate` $Y \sim q(y|x)$
3. `Calculate` $\alpha = min\left(\frac{f(y)q(x|y)}{f(x)q(y|x)}, 1\right)$ 4. $X_{t+1} = \begin{cases} Y, & \texttt{with probability } \alpha \\ X_t, & \texttt{else} \end{cases}$

If q is symmetric (that is $q(y|x) = q(x|y)$ for all x, y), then α simplifies down to $\frac{f(y)}{f(x)}$, and the algorithm is simply called the Metropolis algorithm.

As an example we can consider a series of coin tosses from a weighted coin. Let p be the probability of a head. We assume that $p \sim \text{Unif}(0, 1)$. After 10 tosses six heads could be observed. Let $H \sim \text{Binom}(10, p)$ be the number of heads that are tossed.

$$\begin{aligned}\Pr(p \in \mathrm{d}p | H = 6) &\propto \Pr(H = 6 | p \in \mathrm{d}p) \Pr(p \in \mathrm{d}p) \\ &= \binom{10}{6} p^H (1-p)^{10-H} \\ &\propto p^6 (1-p)^4\end{aligned}$$

Consider two different (symmetric) proposal distributions $q(p_i|p_{i-1})$:

- $p_i^* \sim N(p_{i-1}, 1/12)$
- $p_i^* \sim \mathrm{Unif}(p_{i-1} - 1/2, p_{i-1} + 1/2)$.

Both are symmetric, and so

$$\begin{aligned}\alpha &= \min\left(\frac{f(p_i)q(p_{i-1}|p_i)}{f(p_{i-1})q(p_i|p_{i-1})}, 1\right) \\ &= \min\left(\frac{{p_i}^6(1-p_i)^4}{p_{i-1}^6(1-p_{i-1})^4}\mathbb{I}(p_i > 0), 1\right) \\ &= \min\left(\left(\frac{p_i}{p_{i-1}}\right)^6\left(\frac{1-p_i}{1-p_{i-1}}\right)^4\mathbb{I}(p_i > 0), 1\right)\end{aligned}$$

The effect of the choice of proposal distribution is shown in figure 3.1.

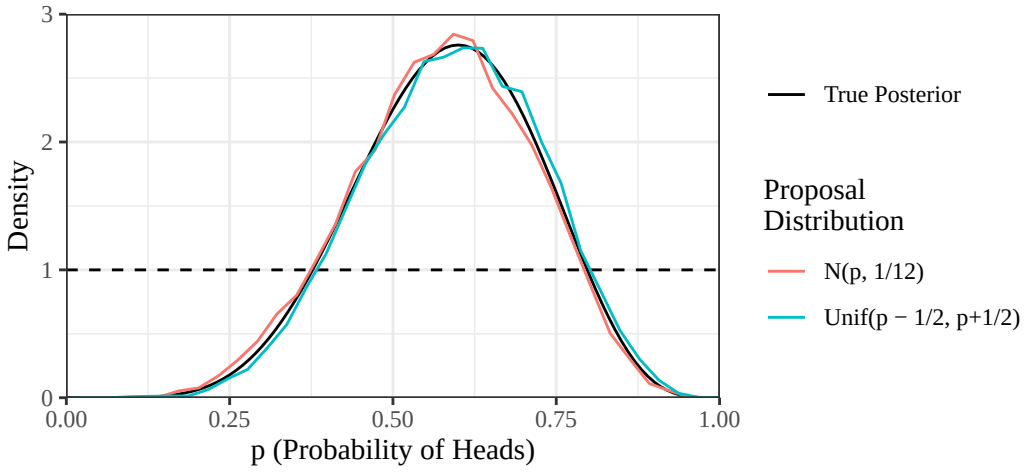


Figure 3.1: 30,000 samples from the posterior distribution of p using the Metropolis Hastings algorithm. It was assumed that $p \sim \mathrm{U}(0,1)$ and $H \sim \mathrm{Binom}(10, p)$, given $H = 6$. A uniform and normal proposal distributions were compared.

For another example we can consider a simple SIS epidemiological model, with

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \gamma I - \beta SI \quad \frac{\mathrm{d}I}{\mathrm{d}t} = \beta SI - \gamma I,$$

and a population of 25,000. Previous pandemics can be used to fix $\beta = 0.5$. At day 0, it is assumed that there is 1 infected person. Given 1184 new cases on day 30, and $\gamma \sim \mathrm{Gamma}(2, 6)$ we can use the Metropolis Hastings algorithm to sample from the posterior distribution of γ. This will depend on the choice of likelihood.Expand this explanation See figure 3.2

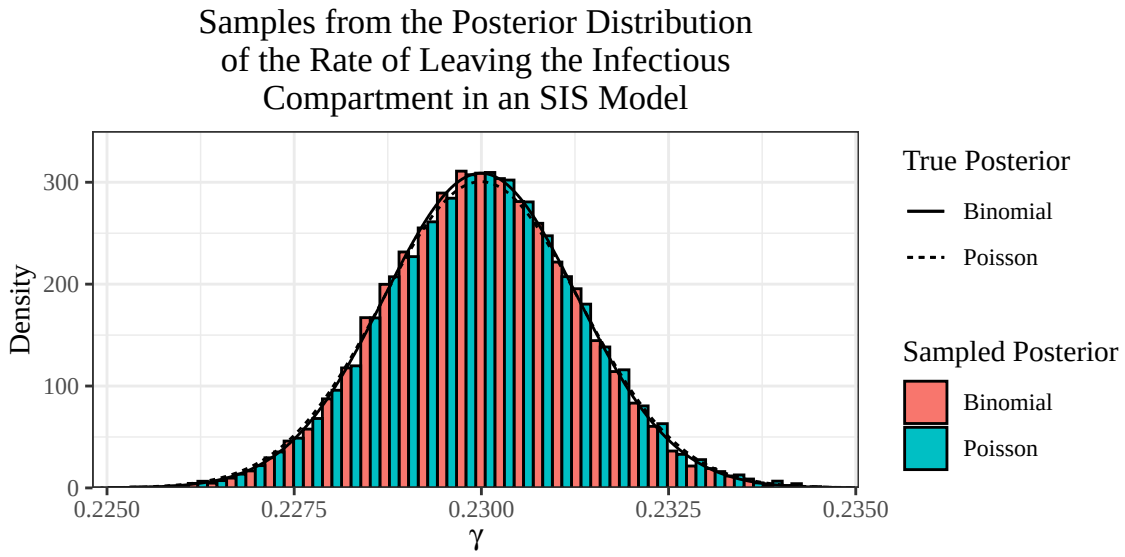


Figure 3.2: Using a basic SIS model, using incident changing likelihood function.

3.6 Gibb's Sampling

For a multivariate case ($\boldsymbol{\Theta} = \{\theta_1, \theta_2, \ldots, \theta_n\}$), it can be analytically or computationally intractable to calculate the posterior (or even something proportional to the posterior) distribution. In this case it may be possible to calculate $\Pr(\theta_i \in \mathrm{d}\theta | \theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \theta_{i+2}, \ldots, \theta_n, \mathrm{x})$ where x is the data we are fitting the parameters to. Prove Gibbs produces sample from the posterior distribution if this is in my final paper

In this case we can use Gibbs sampling to sample from our posterior distribution. The basic algorithm is as follows:

`Initialise for` $\boldsymbol{\Theta} = \{\theta_1, \theta_2, \ldots, \theta_n\}$
1. `For` $i \in \{1, 2, \ldots, n\}$ `sample` $\theta_i^* \sim \theta_i | \theta_1^*, \theta_2^*, \ldots, \theta_{i-1}^*, \theta_{i+1}, \ldots, \theta_n, \mathbf{x}$
2. `Let` $\boldsymbol{\Theta}^* = \{\theta_1^*, \theta_2^*, \ldots, \theta_n^*\}$
3. `Append` $\boldsymbol{\Theta}^*$ `to the chain of parameters, and let` $\boldsymbol{\Theta} := \boldsymbol{\Theta}^*$

3.7 Diagnostics for Metropolis Hastings

include things like autoregression, thinning, burn-in etc.

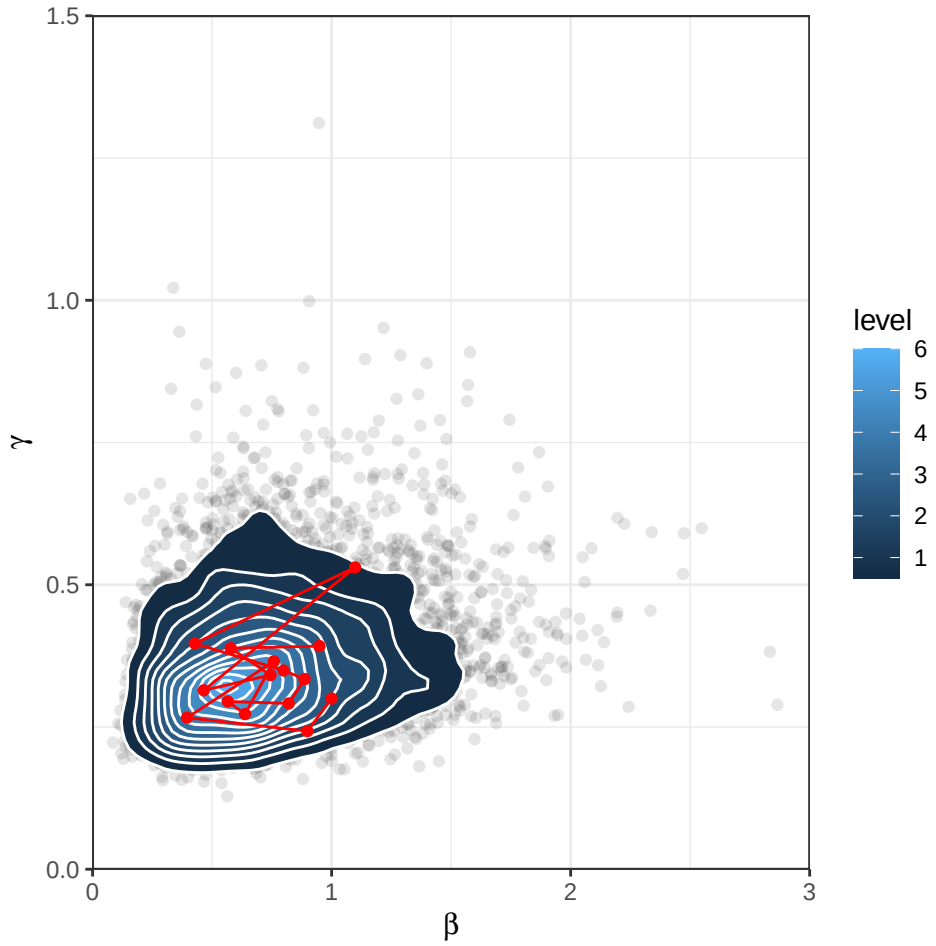


Figure 3.3: Using Gibbs on beta and gamma given an 'R_0' observation

Chapter 4

Approximate Bayesian Computing

TODO:

1. define p-norm

The previous chapter assumes that the likelihood function is known. In reality as models become complicated, the likelihood function often has no closed form or becomes computationally intractable to derive. This is the motivation behind likelihood free techniques.

4.1 Exact Likelihood Free Bayesian Computation

Let $\boldsymbol{\theta} = (\theta_i)_{i\in\{1,\dots,n\}}$ be a set of parameters with prior distribution π. Given a (random) function f : and $(y_j)_{j\in\{1,\dots,m\}}$ be a set of realised data, generated using $(\theta_i)_{i\in\{1,\dots,n\}}$. The most basic Bayesian computation algorithm is such:
`1. Generate` $(\hat{\theta}_i)_{i\in\{1,\dots,n\}} \sim \pi$
`2. Using` $(\theta_i)_{i\in\{1,\dots,n\}}$`, regenerate the data` $(\hat{x_j})_{j\in\{1,\dots,m\}}$ `using the parameter set` $(\hat{\theta}_i)_{i\in\{1,\dots,n\}}$
`3. If` $(\hat{x_j})_{j\in\{1,\dots,m\}} = (x_j)_{j\in\{1,\dots,m\}}$`, accept` $(\hat{\theta}_i)_{i\in\{1,\dots,n\}}$`, else regenerate` $(\hat{\theta}_i)_{i\in\{1,\dots,n\}}$ `and go to step 2`

For non-discrete cases

Generate parameter(s) from your prior.

Generate the data.

If it matches (or is close enough), add that parameter to your set of accepted parameters. This set will approximate the posterior distribution.

Drawbacks of this approach. Very difficult to estimate multiple parameters to the one outcome parameters. Or if you have a single parameter, generating multiple observations, unlikely to match all of them. Same with multiple parameters to multiple outputs

4.2 Summary Statistics

Distribution of summary statistic

Should be relatively normally distributed

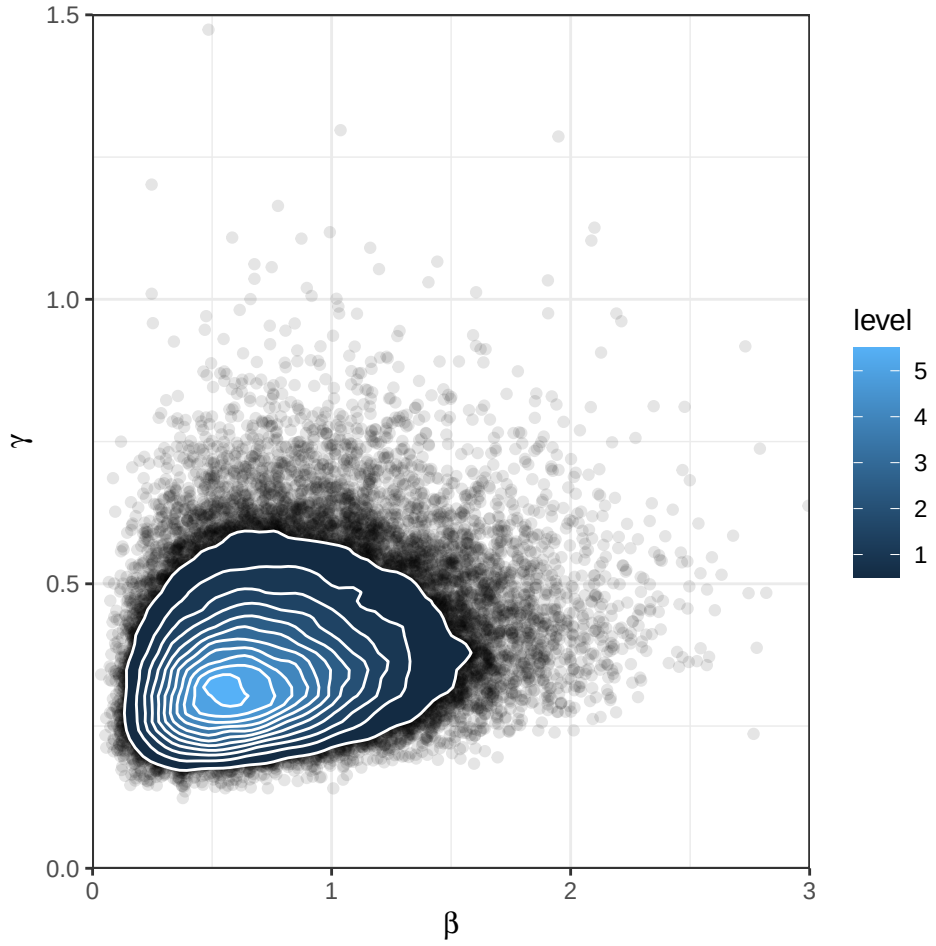


Figure 4.1: Using a basic SIS model as in previous figure using ABC to generate from posterior beta, gamma

4.3 Discrepency

Distribution of discrepency

|Normal - normal| or something so should be normalish?

4.4 Estimating the likelihood from the summary statistic

See handbook ABC 10 or so

Chapter 5

Gaussian Processes

- Prove all pos-sem is cov
- expand on intro to motive the section
- prove exponential quad kern is pos semi-def

5.1 Motivation and Preliminaries

If we knew the distribution of the discrepency function described in the previous chapter across the parameter space, then in order to do approximate Bayesian computation, we would not need to actually run the model in order to obtain a 'sample.' Instead we could draw from the distribution of the discrepency function at fixed $\boldsymbol{\theta}$. distribution itself. Further we could create an approximate likelihood $\mathcal{L}_{\text{ABC}}(\boldsymbol{\theta})$ and use this directly in sampling from the posterior.

This motivates methods of function approximation. To approximate our discrepency, we consider the class of plausible functions, and assign some probability to each of them. Each model simulation gives some more information that we can use to update the probability of each function being the 'true' function. We will consider the classes of functions generated by Gaussian processes.

Definition 5.1 (Kernel Function (symmetric, stationary)). *Any function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a* kernel function. *If $k(x, x') = k(x', x)$, the kernel is called* symmetric.

I think I define kernels above?

Definition 5.2 (Gaussian Process). *A collection of random variables $\{f(x)\}_{x \in \mathcal{X}}$ (where x may be a vector) is a* Gaussian process *if any finite subset of the collection of random variables is multivariate normal distributed. That is, there is a function $m : \mathcal{X} \to \mathbb{R}$ and symmetric kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that for all finite sets $\mathbf{x} := \{x_1, x_2, \ldots, x_n\} \subset \mathcal{J}$, with $f(\mathbf{x}) := [f(x_1), f(x_2), \ldots, f(x_n)]^T$*

$$f(\mathbf{x}) \sim \text{MVN}\left(\begin{bmatrix} m(x_1) \\ m(x_2) \\ \vdots \\ m(x_n) \end{bmatrix}, \mathbf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \ldots & k(x_1, x_n) \\ k(x_2, x_1) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ k(x_n, x_1) & \cdots & \cdots & k(x_n, x_n) \end{bmatrix} \right).$$

Definition 5.3 (Mean and Covariance Function)**.** *We define the* mean function *and* covariance kernels *as*

$$m(x_i) := \mathbb{E}\left[f(x_i)\right]$$

and

$$k(x_i, x_{i'}) := \operatorname{cov}\left(f(x_i), f(x_{i'})\right).$$

Some common examples of Gaussian processes include

- Brownian motion on $\mathbb{R}$:

$$m \equiv 0,$$

and

$$k(s,t) = \min(s,t)$$

- Ornstein Uhlenbeck process with parameters θ and σ:

$$m \equiv 0,$$

and

$$k(s,t) = \frac{\sigma^2}{2\theta}\left(e^{-\theta|t-s|} - e^{-\theta(t+s)}\right)$$

Definition 5.4 (Brownian Motion)**.** $B(t) : \mathbb{R}^+ \to \mathbb{R}$ *is a* Brownian motion *on* $\mathbb{R}$ *if*

1. $B(0) = 0$ *almost surely*
2. $B(t_0), B(t_1) - B(t_0), \ldots, B(t_n - t_{n-1})$ *are independent for all* $t_0 < t_1 < t_2 < \cdots < t_n$
3. $B(t+s) - B(t) \sim N(0, s)$ *for* $s, t \geq 0$
4. $B(t)$ *is continuous almost surely for* $t > 0$.

We can generate novel Gaussian processes with a custom m, and covariance kernel k. This thesis only considers symmetric k that are also stationary - that is $k(x, x')$ can be expressed as a function of $x - x'$.

Definition 5.5 (Positive Semi-Definite Matrix)**.** *An* $n \times n$ *matrix* $\mathbf{A}$ *is* positive semi-definite *if* $\mathbf{v}^T\mathbf{A}\mathbf{v} \geq 0$ *for all* $\mathbf{v} \in \mathbb{R}^n$.

Theorem 5.1 (Sufficient Condition for Positive Semi-Definite)**.** *A symmetric matrix* $\mathbf{A}$ *is positive semi-definite, if (and only if) it's eigenvalues are non-negative.*

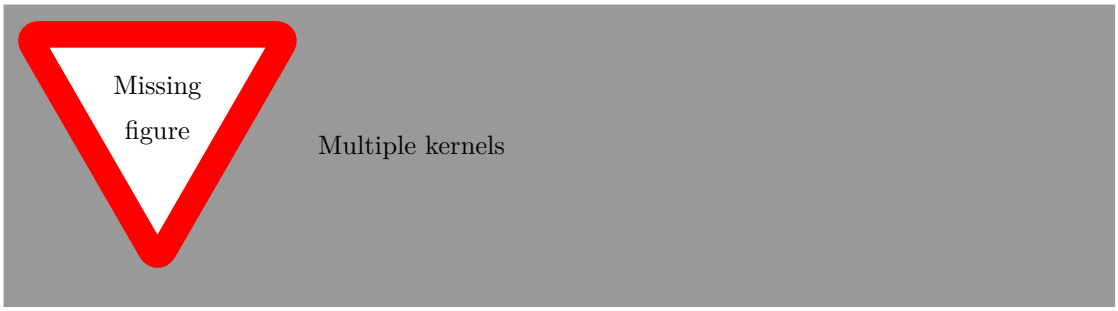
Proof. □

Definition 5.6 (Positive Semi-Definite Kernel)**.** *A kernel* $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *is* positive semi-definite *if the matrix*

$$\mathbf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \ldots & k(x_1, x_n) \\ k(x_2, x_1) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ k(x_n, x_1) & \cdots & \cdots & k(x_n, x_n) \end{bmatrix}$$

is positive semi-definite for any collection of $x_i \in \mathcal{X}$

Theorem 5.2. *All symmetric positive semi-definite matrices are covariance matrices for some set of random variables*



Proof. □

Various classes of covariance kernel exist. The choice of covariance kernel determines how we expect the function to behave, including how 'smooth' we expect the discrepency function to be. We formalise this by the notion of mean square continuity/differentiability.

Definition 5.7 (Mean Square Continuous)**.** *A function $f : \mathbb{R}^d \to \mathbb{R}$ is* mean square continuous *at $\mathbf{x}$ in the ith direction at if $\mathbb{E}(|f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x})|^2) \to 0$ as $|h| \to 0$, where $\mathbf{e}_i$ is the unit vector with a 1 in the ith coordinate.*

Definition 5.8 (Mean Square Differentiable)**.** *A function $f : \mathbb{R}^d \to \mathbb{R}$ is* mean square differentiable *at $\mathbf{x}$ in the ith direction with derivative $\frac{\partial f(\mathbf{x})}{\partial x_i}$ if $\mathbb{E}\left[\left|\frac{f(\mathbf{x}+h\mathbf{e}_i)-f(\mathbf{x})}{h} - \frac{\partial f(\mathbf{x})}{\partial x_i}\right|^2\right] \to 0$ as $|h| \to 0$, where $\mathbf{e}_i$ is the unit vector with a 1 in the ith coordinate.*

Theorem 5.3. *Brownian motion is mean square continuous, but not mean square differentiable.*

Proof. $(B_{t+h} - B_t)^2 \sim (\sqrt{|h|}Z)^2$ where $Z \sim N(0, 1)$. Therefore $(B_{t+h} - B_t)^2 \sim |h|\chi_1^2 \to 0$ almost surely as $|h| \to 0$, and hence $\mathbb{E}[(B_{t+h} - B_t)^2 = 0]$. Since $\frac{B_{t+h}-B_t}{h} \sim N(0, 1/|h|)$, $\frac{B_{t+h}-B_t}{h}$ does not converge to any valid probability distribution as $|h| \to 0$, as the variance approaches $+\infty$. □

Some common

Theorem 5.4 (Positive Semi-Definiteness of RBF)**.** *The radial basis function $\frac{1}{\sigma^2}\exp(-\frac{(x-x')^2}{2\gamma^2})$ is a positive semi-definite kernel.*

Theorem 5.5 (Bochner's Theorem)**.** *Let k be a stationary kernel function such that $k(x, x') = f(d)$. A function $k : \mathbb{R}^d \to \mathbb{C}$ is the covariance function of a weakly stationary mean square continuous complex-valued random process of $\mathbb{R}^d$ if and only if it can be represented as*

$$k(\tau) = \int_{\mathbb{R}^d} \exp(2\pi i \mathbf{s} \cdot \tau)$$

(Rasmussen and Williams 2008, p. 82)

5.2 Kernel Families

Matérn Class

The Matérn class of kernel gives flexibility over how many times mean square differentiable the realised function is.

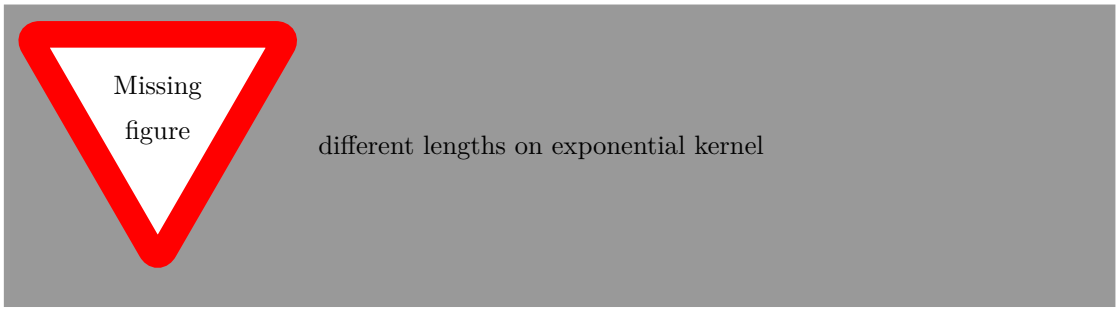


Figure 5.1: a figure about kernel lengths

The Matérn exponential kernel is of the form

$$k_\nu(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}||x - x'||}{\ell} \right)^\nu K_\nu \left(-\frac{\sqrt{2\nu}||x - x'||}{\ell} \right)$$

where K_ν is a modified Bessel function (see Abramowitz and Stegun 2013, p. 374). The general form is not very insightful, but for the most common values of $\nu = 1/2, 3/2$ and $5/2$, we have the forms:

$$k_{1/2}(x, x') = \sigma^2 \exp\left(-\frac{||x - x'||}{\ell} \right)$$

$$k_{3/2}(x, x') = \sigma^2 \left(1 + \frac{\sqrt{3}||x - x'||}{\ell} \right) \exp\left(-\frac{\sqrt{3}||x - x'||}{\ell} \right)$$

$$k_{5/2}(x, x') = \sigma^2 \left(1 + \frac{\sqrt{5}||x - x'||}{\ell} + \frac{5||x - x'||}{3\ell^2} \right) \exp\left(-\frac{||x - x'||^2}{2 * \ell^2} \right)$$

Zero mean processes with a Matérn kernel are n times mean square differentiable, for all $n < \nu$. As $\nu \to \infty$, we get something infinitely differentiable, and specifically we get the squared exponential kernel.

Squared Exponential

The squared exponential kernel is of the form

$$k(x, x') = \sigma^2 \exp\left(-\frac{||x - x'||^2}{2\ell^2} \right)$$

Length parameter

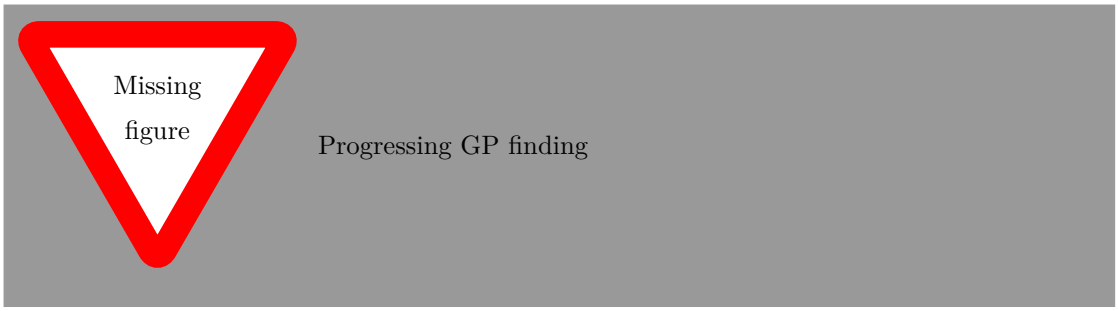
In both of the above kernels, there is a hyperparameter ℓ, which we call the length parameter. This value determines how close two points need to be to be highly correlated. A small value of ℓ leads to a more wiggly function as seen in 5.1.

5.3 Gaussian Process Regression

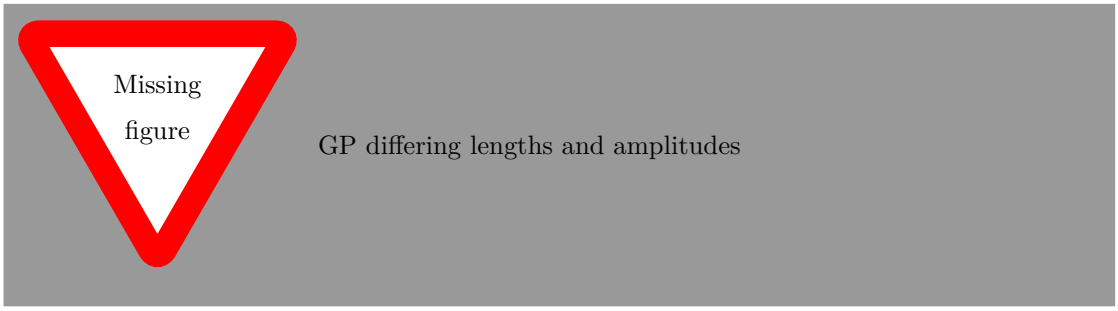
Given a choice of

It asymptotically approaches the squared exponential class.

LOOK AT CHAPTER 4 SKOROKHOD STOCHASTIC I



Progressing GP finding



GP differing lengths and amplitudes

Adding in observation variance

New covariance function

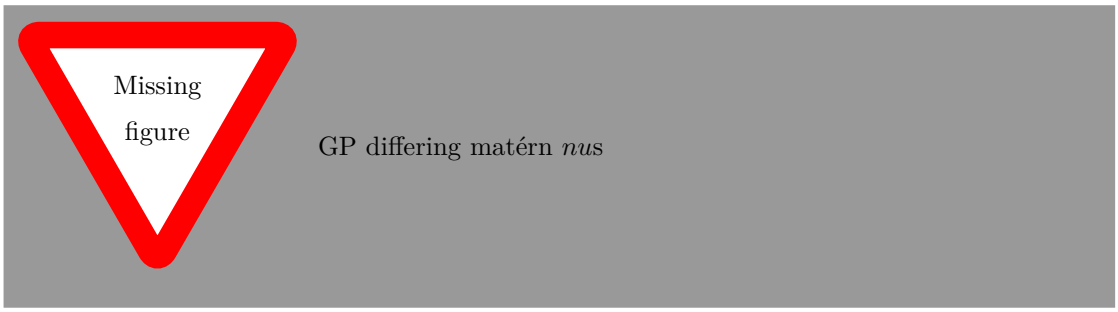
5.4 Differing mean functions

5.5 Gaussian Process Regression

Posterior mean
Posterior variance

5.6 Bayesian Acquisition Functions

- Choosing the next point to sample
- $\arg\min_{\boldsymbol{\theta}} A(\boldsymbol{\theta})$ where A is an acquisition function.



GP differing matérn *nu*s

GP mean functions

- BOLFI paper uses
$$\mu(\boldsymbol{\theta}) - \eta_t \sqrt{\mathrm{v}(\boldsymbol{\theta})}$$
  - $\eta_t := \sqrt{c + 2\ln(t^{d/2+2})}$, and c can be chosen
  - $\mu(\boldsymbol{\theta})$ and $\mathrm{v}(\boldsymbol{\theta})$ are the posterior mean and variance
- Could use expected information
$$(\mu_{\min} - \mu(\boldsymbol{\theta}))\Phi\left(\frac{\mu_{\min} - \mu(\boldsymbol{\theta})}{\sqrt{\mathrm{v}(\boldsymbol{\theta})}}\right) + \sqrt{\mathrm{v}(\boldsymbol{\theta})}\phi\left(\frac{\mu_{\min} - \mu(\boldsymbol{\theta})}{\sqrt{\mathrm{v}(\boldsymbol{\theta})}}\right)$$
  - $\mu_{\min} := \min_{\boldsymbol{\theta}} \mu(\boldsymbol{\theta})$
  - Φ, ϕ CDF and PDF of standard normal

exporation parameter proven by Srinivas et al. 2010

Chapter 6

Methods and Results

TODO:

1. Check for consistent notation

6.1 Description

Model and Synthetic Data Simulations

In order to trial this method we used the model described in Champagne et al. 2022, and graphically depicted in Figure 2.4. Exact model simulations such as in Figure 6.1 were computed using the Doob-Gillespie algorithm (Algorithm 1).

Synthetic data was generated from a single run of the model, using a population size of 1,000, an initial infected population of 10 (with both liver and blood stage infection). The parameters used closely followed those reported in Champagne et al. 2022:

- effective blood stage treatment proportion $\alpha = 0.4$,
- effective liver stage treatment proportion $\beta = 0.4$,
- rate of liver stage disease clearance $\gamma_L = 1/223$ days^{-1},
- importation rate $\delta = 0$ (assumed to be known),
- rate of infection $\lambda = 0.04$ days^{-1},
- rate of relapse $f = 1/72$ days^{-1},
- rate of blood stage disease clearance $r = 1/60$ days^{-1}.

To generate the synthetic data, the model was run for 30 days and at least 15,000 events (infections, recoveries etc.), after which time, it was assumed to have reached equilibrium.

Summary Statistics and Discrepancy Function

The synthetic data (as summary statistics) we measured were:

1. prevalence at the end of the model run - p_{obs},

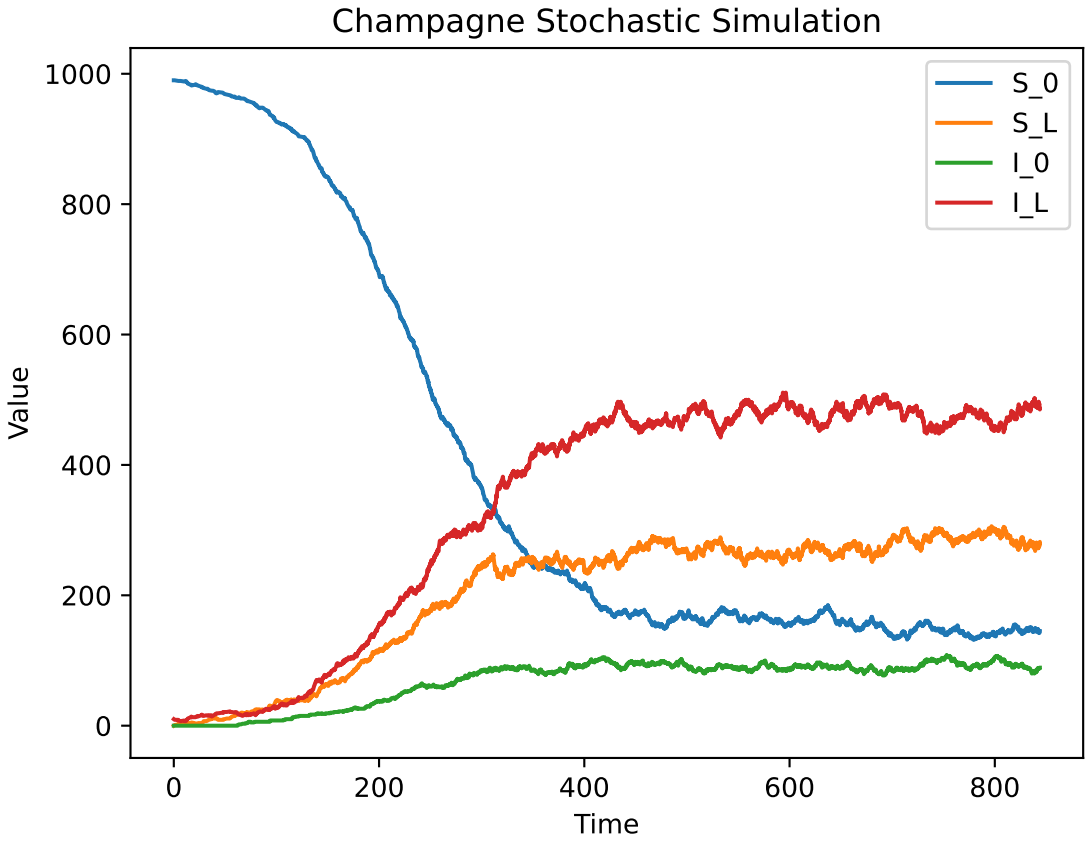


Figure 6.1: A Doob-Gillespie Simulation of the model described by Champagne et al. 2022.

2. number of cases in the first week of the epidemic (first month incidence) - m_{obs},
3. a single observation of the number of weekly cases at equilibrium (weekly incidence) - w_{obs}.

Similarly we extracted p, m, and w from each new model run.

Our discrepency function we used was the L_2 norm of the relative differences

$$\mathcal{D}(\alpha, \beta, \gamma_L, \lambda, f, r) = \sqrt{\left(\frac{p - p_{\text{obs}}}{p_{\text{obs}}}\right)^2 + \left(\frac{m - m_{\text{obs}}}{m_{\text{obs}}}\right)^2 + \left(\frac{w - w_{\text{obs}}}{w_{\text{obs}}}\right)^2}.$$

Relative differences were chosen to limit the impact between the scale differences of the summary statistics. No systematic comparison of norms was done.

Parameter	Bound	Unit
Proportion of treatment clearing blood stage disease α	1	
Proportion of treatment clearing liver stage disease β	1	
Rate of liver stage disease clearance γ_L	1/30	1/days
Rate of infection λ	1/10	1/days
Rate of relapse f	1/14	1/days
Rate of blood stage disease clearance r	1/14	1/days

Table 6.1: Target parameters to be estimated and their upper bounds, derived from Champagne et al. 2022; White et al. 2016.

All parameters were given conservative upper bounds after considering values reported in the literature.

Gaussian Process Setup

To approximate the true discrepancy function, a Gaussian process with constant mean $\mu_{\mathcal{GP}}$, and 5/2 Matérn kernel was used with automatic relevance determination - i.e. each parameter $\theta \in \boldsymbol{\theta}$ was scaled by ℓ_θ, such that the kernel is

$$k(\boldsymbol{\theta}_i, \boldsymbol{\theta}'_i) = \sigma_a^2(1 + z_i + \frac{z_i^2}{3}) \exp(-z_i)$$

where

$$z_i = \sqrt{5 \sum_{\theta \in \boldsymbol{\theta}} \left(\frac{\theta_i - \theta'_i}{\ell_\theta}\right)^2}.$$

and σ_O^2 is the observation variance.

Initialisation

50 Latin hypercube samples that were scaled to the upper bounds described in Table 6.1 were taken, and for each set of parameters, two model run discrepencies were taken. The hyperparameters $\sigma_O^2, \sigma_a^2, \ell_\alpha, \ell_\beta, \ell_{\gamma_L}, \ell_\lambda, \ell_f$, and ℓ_r were chosen by maximising the leave one out predictive log-probability as described in Rasmussen and Williams 2008, p. 116.

The acquisition function used was

$$\mu(\boldsymbol{\theta}) - \eta_t \sqrt{\mathrm{v}(\boldsymbol{\theta})}$$

with $\eta_t := \sqrt{2\ln(\frac{t^{2d+2}\pi^2}{3\varepsilon})}$, and $\varepsilon \in (0,1)$ that can be chosen (with a lower epsilon resulting in more exploration), $\mu(\boldsymbol{\theta})$ and $\mathrm{v}(\boldsymbol{\theta})$ are the posterior mean and variance. $\varepsilon = 0.1$ was used.

$$(\mu_{\min} - \mu(\boldsymbol{\theta}))\Phi\left(\frac{\mu_{\min} - \mu(\boldsymbol{\theta})}{\sqrt{\mathrm{v}(\boldsymbol{\theta})}}\right) + \sqrt{\mathrm{v}(\boldsymbol{\theta})}\phi\left(\frac{\mu_{\min} - \mu(\boldsymbol{\theta})}{\sqrt{\mathrm{v}(\boldsymbol{\theta})}}\right)$$

$\mu_{\min} := \min_{\boldsymbol{\theta}} \mu(\boldsymbol{\theta})$ Φ, ϕ

Justification

Using relative differences limits the impact between the scale differences of the summary statistics.

6.2 Running the Model and Our Choices

Although most of the procedure we used closely follows the paper by Gutmann and Cor 2016, there are a few key ways in which we modified the manuscript's method. In particular, we felt the choice of squared exponential kernel was not a good assumption, since it implicitly assumes a high degree of smoothness in the target discrepency function. This is unlikely to be met if the model has any bifurcation points. To compensate for any non-smoothness in the function, the length scale is forced to be set very small. Therefore, although the squared exponential is the most commonly chosen kernel, we used a Matérn kernel with $\nu = 5/2$. This is not as constrained as a squared exponential kernel, but realisations are twice mean square differentiable.

A constant mean was used, rather than the quadratic mean, because the parameters are well bounded, and we don't need to worry about what happens when they go to infinity. Post-hoc mean fitting could be done on the discrepencies sampled space in order to extend the likelihood into regions that haven't been tested without reverting to the mean function.

There seems to be a mistake in Gutmann and Cor 2016, where the exploration parameters is written as $\eta_t := \sqrt{2\ln(\frac{t^{2d+2}\pi^2}{3\varepsilon})}$. This seems to be inherited from Brochu, Cora, and Freitas 2010. We use Brochu's original citation Srinivas et al. 2010 to revise this to $\eta_t := \sqrt{2\ln(\frac{t^{2d+2}\pi^2}{3\varepsilon})}$. [1]

The Gaussian process and Gaussian process regression was implemented using TensorFlow in Python Martín Abadi et al. 2015.

6.3 Results

[1] One Python package that implements BOLFI notes this error, see: `https://github.com/elfi-dev/elfi/blob/dev/elfi/methods/bo/acquisition.py`

Chapter 7

Stuff that could fix stuff

- Gaussian process has gaussian process for variance, run any set of parameters multiple times, have very large length scale, maybe a prior on it
- Add in an artificial variance for the values where the disease dies off
- t-process to allow for sudden drop off at threshold/ bifurcations
- Maybe use matern to allow for sudden jump but I think that allows for very small sudden jumps
- Used a relative distance squared
- Using log-discrepency doesn't allow you to take a limiting argument, and might depend a lot of the size of the error you allow. Should do some calculations to work that out, because the important thing is that the differences have constant ratio
- The use of LC? acquisition function does not encourage a broad searching of the parameter space.
- Effectively just moment matching to either normal or log-normal. Is that a fair assumption?

Bibliography

Abramowitz, Milton and Irene A. Stegun, eds. (2013). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. 9. Dover print.; [Nachdr. der Ausg. von 1972]. Dover books on mathematics. New York, NY: Dover Publ. 1046 pp. ISBN: 978-0-486-61272-0.

Adams, John H. and Ivo Mueller (Sept. 2017). "The Biology of Plasmodium vivax". In: *Cold Spring Harbor Perspectives in Medicine* 7.9, a025585. ISSN: 2157-1422. DOI: 10.1101/cshperspect.a025585. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5580510/ (visited on 03/24/2023).

Aron, Joan L. and Robert M. May (1982). "The population dynamics of malaria". In: *The Population Dynamics of Infectious Diseases: Theory and Applications*. Ed. by Roy M. Anderson. Boston, MA: Springer US, pp. 139–179. ISBN: 978-1-4899-2901-3. DOI: 10.1007/978-1-4899-2901-3_5. URL: https://doi.org/10.1007/978-1-4899-2901-3_5.

Brochu, Eric, Vlad M. Cora, and Nando de Freitas (2010). *A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning*. arXiv: 1012.2599 [cs.LG].

Champagne, Clara et al. (Jan. 2022). "Using observed incidence to calibrate the transmission level of a mathematical model for Plasmodium vivax dynamics including case management and importation". In: *Mathematical Biosciences* 343, p. 108750. ISSN: 00255564. DOI: 10.1016/j.mbs.2021.108750. URL: https://linkinghub.elsevier.com/retrieve/pii/S0025556421001541 (visited on 08/22/2023).

Cowman, Alan F. et al. (2016). "Malaria: Biology and Disease". In: *Cell* 167.3. Type: Review, pp. 610–624. DOI: 10.1016/j.cell.2016.07.055. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84994000411&doi=10.1016%2fj.cell.2016.07.055&partnerID=40&md5=81d9b4c51fe738ac66e0c8561b12c5bf.

Gutmann, Michael U. and Jukka Cor (2016). "Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models". In: *Journal of Machine Learning Research* 17.125, pp. 1–47. ISSN: 1533-7928. URL: http://jmlr.org/papers/v17/15-017.html (visited on 04/28/2024).

Hagenaars, T. J., C. A. Donnelly, and N. M. Ferguson (Apr. 2006). "Epidemiological analysis of data for scrapie in Great Britain". en. In: *Epidemiology and Infection* 134.2, pp. 359–367. ISSN: 0950-2688, 1469-4409. DOI: 10.1017/S0950268805004966. URL: https://www.cambridge.org/core/product/identifier/S0950268805004966/type/journal_article (visited on 03/26/2024).

Keeling, Matthew James and Pejman Rohani (2008). *Modeling infectious diseases in humans and animals*. OCLC: ocn163616681. Princeton: Princeton University Press. 366 pp. ISBN: 978-0-691-11617-4.

Martín Abadi et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: https://www.tensorflow.org/.

Milner, Danny A. (Jan. 2018). "Malaria Pathogenesis". en. In: *Cold Spring Harbor Perspectives in Medicine* 8.1, a025569. ISSN: 2157-1422. DOI: 10.1101/cshperspect.a025569. URL: http://perspectivesinmedicine.cshlp.org/lookup/doi/10.1101/cshperspect.a025569 (visited on 03/24/2023).

Price, R.N. et al. (2020). "Plasmodium vivax in the Era of the Shrinking P. falciparum Map". English. In: *Trends in Parasitology* 36.6, pp. 560–570. ISSN: 1471-4922. DOI: 10.1016/j.pt.2020.03.009.

Rasmussen, Carl Edward and Christopher K. I. Williams (2008). *Gaussian processes for machine learning*. 3. print. Adaptive computation and machine learning. Cambridge, Mass.: MIT Press. 248 pp. ISBN: 978-0-262-18253-9.

Smith, David L. et al. (Apr. 2012). "Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens". en. In: *PLOS Pathogens* 8.4. Publisher: Public Library of Science, e1002588. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1002588. URL: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1002588 (visited on 03/28/2023).

Srinivas, Niranjan et al. (2010). "Gaussian process optimization in the bandit setting: no regret and experimental design". In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML'10. Haifa, Israel: Omnipress, pp. 1015–1022. ISBN: 9781605589077.

White, Michael T. et al. (Mar. 30, 2016). "Variation in relapse frequency and the transmission potential of *Plasmodium vivax* malaria". In: *Proceedings of the Royal Society B: Biological Sciences* 283.1827, p. 20160048. ISSN: 0962-8452, 1471-2954. DOI: 10.1098/rspb.2016.0048. URL: https://royalsocietypublishing.org/doi/10.1098/rspb.2016.0048 (visited on 08/22/2023).

World Health Organization (Dec. 2022). *World malaria report 2022*. en. Tech. rep. Geneva: World Health Organization.

Zekar, Lara and Tariq Sharman (2023). "Plasmodium Falciparum Malaria". eng. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing. URL: http://www.ncbi.nlm.nih.gov/books/NBK555962/ (visited on 03/24/2023).

Zha, Wen-ting et al. (2020). "Research about the optimal strategies for prevention and control of varicella outbreak in a school in a central city of China: based on an SEIR dynamic model". en. In: *Epidemiology and Infection* 148, e56. ISSN: 0950-2688, 1469-4409. DOI: 10.1017/S0950268819002188. URL: https://www.cambridge.org/core/product/identifier/S0950268819002188/type/journal_article (visited on 03/26/2024).