The Age of Infection: A Semi-Markov Framework for Developing Mechanistic Models of Malaria Epidemiology

John M. Henry

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:
David L. Smith, Chair
Elizabeth M. Halloran
Robert C. Reiner

Program Authorized to Offer Degree:
Quantitative Ecology and Resource Management

University of Washington

Abstract

The Age of Infection:
A Semi-Markov Framework for Developing
Mechanistic Models of Malaria Epidemiology

John M. Henry

Chair of the Supervisory Committee:
Professor David L. Smith
Health Metrics Sciences

Malaria is an epidemiologically complex disease which poses a significant burden on humanity, contributing an estimated 643,000 deaths in 2019 alone [1]. Infection with one cohort of parasites does not prevent concurrent infection with others [2], making prevalence alone an incomplete measure of infection in a population. Further, immunity is slow to develop in response to exposure and is only partially protective [3]. Therefore, it acts to both suppress disease in individuals and to mask infections from detection and treatment, which allows for longer periods of transmission to mosquitoes. In this way past exposure modifies the bias in the observed prevalence and incidence data collected from each cohort of hosts. Natural variability in the system further obfuscates the connection between data and the process which generates it. Determining the transmission and burden of malaria becomes incredibly difficult without the use of theoretical models to connect the observed data to the latent states and parameters which guide our basic understanding and policy

decisions, such as the distribution of the number of infections in each host and the detectable fraction of infections. Currently existing mechanistic models typically fall into one of two categories: simple and transparent, but deterministic and not descriptive enough to connect to available data [4]; or incredibly detailed individual-based simulation frameworks which are realistic but difficult to implement or calibrate [5,6].

Here, we propose a different approach of intermediate complexity which embraces the probabilistic structure of the system. In chapter 1, we start by demonstrating how the distribution of the multiplicity of infection modeled above is impacted by access to treatment, a major factor for the state of the system. In chapter 2, using data from controlled human malaria infections, we show how infection age is statistically predictive of parasitemia, and therefore detection, fever rates, and transmission rates. These relationships imply that infection age, a relatively simple quantity to model, can help us build predictive models for the latent and highly variable parasite densities. Chapter 3 explores this through simulation, and develops a theoretical framework for tracking the probability densities using simple ODEs. Chapter 4 expands on this, and by treating the outcomes of interest as generalized linear models of our semi-Markov model of infection age, we obtain simple ODE models for these probability density function tracking variables, with closed form equations relating them to observables such as theoretical true prevalence, detection, and fever-prompted treatment (using the results from chapter 1 along the way), all while incorporating immunity as a covariate. Finally in chapter 5 we turn our attention to fitting these models to time series

data of observables, and ultimately develop an algorithm for fitting without any probabilistic simulation of the underlying stochastic process. It is our hope that this series of papers inspires others to build upon and push the new tools developed here, and use them to better improve future efforts toward malaria eradication everywhere.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I of course have to thank my advisor, boss, and mentor Dave Smith. Without his encouragement and support, I would not be where I am today.

I would also like to acknowledge everyone in the QERM program and my coworkers at IHME. Both have exposed me to the ideas of ecological and epidemiological analysis, and forced me to think more seriously with uncertainty in the data we collect.

My parents, sister, brother-in-law, and nephews. They kept me connected to home, even when the pandemic kept us apart.

And Mack, because a dedication isn't enough.

DEDICATION

To my partner Mack, and our cats Oliver and Mabel

INTRODUCTION

Mathematical models of malaria epidemiology have played a pivotal role in our understanding of the disease since our identification of its source. After the discovery that the pathogen was transmitted by mosquitoes, Ronald Ross detailed a mathematical model which described in simple terms the implications for two interacting populations, one host and one vector, written in modern notation as follows:

$$\frac{dx}{dt} = a(1-x)z - bx$$
$$\frac{dz}{dt} = c(1-z)x - dz$$

where a, b, c, d are positive constants and x, z are respectively the proportion of humans and mosquitoes infected with malaria. a and c describe the rate of interaction and conversion from an infected individual of one type to another, while b and d represented either demographic turnover or recovery from infection. As a historical note, Ross denoted the parameter a (the force of infection per infected mosquito to an individual human) as h to represent the "happenings rate" - the average rate at which a transition occurs from uninfected to infected [1, 2].

This theoretical model allowed for further refinement and analysis by George Macdonald, who both described the qualitative behaviors of the dynamics (giving rise to the first notion of the basic reproduction number) and replaced these abstract rates with a set of entomological parameters which could be measured [3]. This allowed for two major advancements: easier parameter estimation for fitting and understanding where our uncertainties remain, and identification of entomological parameters which have the largest impact on transmission.

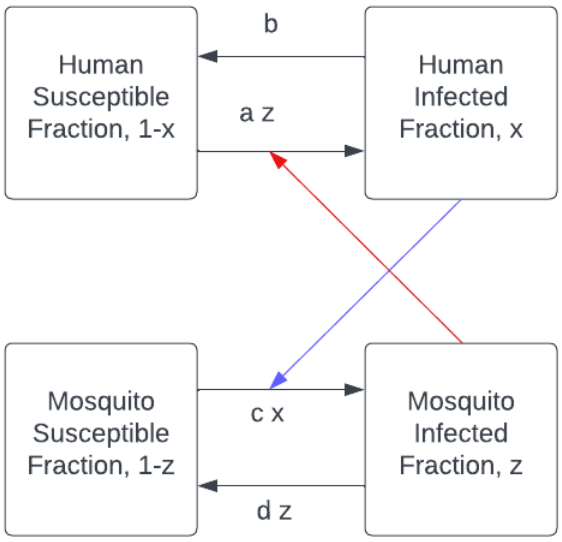


Figure 1: Compartment model diagram for abstract Ross-Macdonald model

Fitting remains a difficult task for ODE models broadly, so finding other ways of measuring the parameters directly allowed for scientists and policymakers to implement the models computationally.

$$\frac{dx}{dt} = mabz(1-x) - rx$$
$$\frac{dz}{dt} = acx(1-z) - gz$$

Further, this new parameterization identified adult mosquito lifespan as being a particularly potent parameter to target. While modifying other parameters tended to achieve a modest reduction in prevalence and the *basic reproduction number* (the expected number of secondary infections resulting from a single infectious individual in a totally susceptible population), the model suggested reducing mosquito lifespan would result in exponential re-

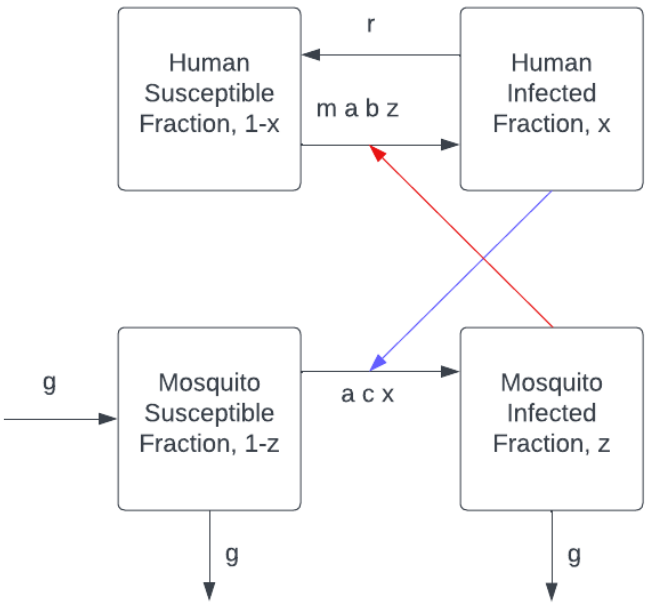


Figure 2: Compartment model diagram for reparameterized Ross-Macdonald model

turns in prevalence reduction, as seen in the new representation for the reproduction number implied by this model:

$$R_0 = \frac{ma^2bc}{gr}e^{-g\nu}$$

where m is the ratio of humans to mosquitoes; a is the human biting rate; b and c are respectively transmission efficiencies from mosquito-to-human and human-to-mosquito; g is the mortality rate of mosquitoes ($1/g$ gives the average lifespan under a model of constant mortality rates); and ν is the extrinsic incubation period, that is the time for a mosquito to become infectious after becoming infected by a human bloodmeal. As reducing the expected lifespan $1/g$, we see an exponential response to the reproduction number.

As Macdonald was publishing his analysis, malaria eradication was being planned around

a program of insecticide spraying, primarily through the use of DDT. At least one previous large-scale intervention, to protect workers from yellow fever and malaria during the construction of the Panama Canal, had resulted in a significant reduction of disease and construction was completed ahead of schedule. The DDT-based strategy was implemented in regions around the world, resulting in the elimination of malaria from some areas and resurgence in others [4, 5]. The exact factors that caused some areas to maintain elimination and others to rebound are a subject of debate, but one major factor includes economic development and the development of infrastructure to reduce standing water for mosquito breeding grounds and openings in housing for mosquitoes to enter. In the discussion around these factors, quantitative reasoning led to reparameterization of the basic reproduction number to measure the effects of interventions and the search for qualitative explanations for why malaria appeared to remain "sticky" in some areas in response to eradication efforts [5].

The phenomenon of backwards bifurcations, described in figure 3, was identified as a possible mechanism driving this behavior. That is, local conditions can allow for both an endemic and disease-free equilibrium to be stable simultaneously under the correct local conditions, and large scale interventions could be all that is needed. The factors that drive these sorts of bifurcations to appear in models are varied, but one such mechanism is a unique feature of malaria parasites with its own complicated mathematical history - superinfection.

Superinfection is the phenomenon in which a single human host is infected with multiple concurrent clonal lineages of parasites, originating from distinct sporozoites from infected mosquitoes. Unlike before, individuals don't just receive a binary value of infected or uninfected, but an integer counting the number of infections they currently are harboring. As detailed in [6], the mathematical model was correctly described by Macdonald but incorrectly translated to equations. After several corrections, a new state space emerged for the process, as detailed in figure 4. Under an equilibrium assumption of infection, the recovery rate was directly impacted by exposure due to an accumulation of independent infections to clear. This results in the approximation used in the Garki project [7]. This model was a major contribution for not only modeling superinfection, but also because it modeled immunity as

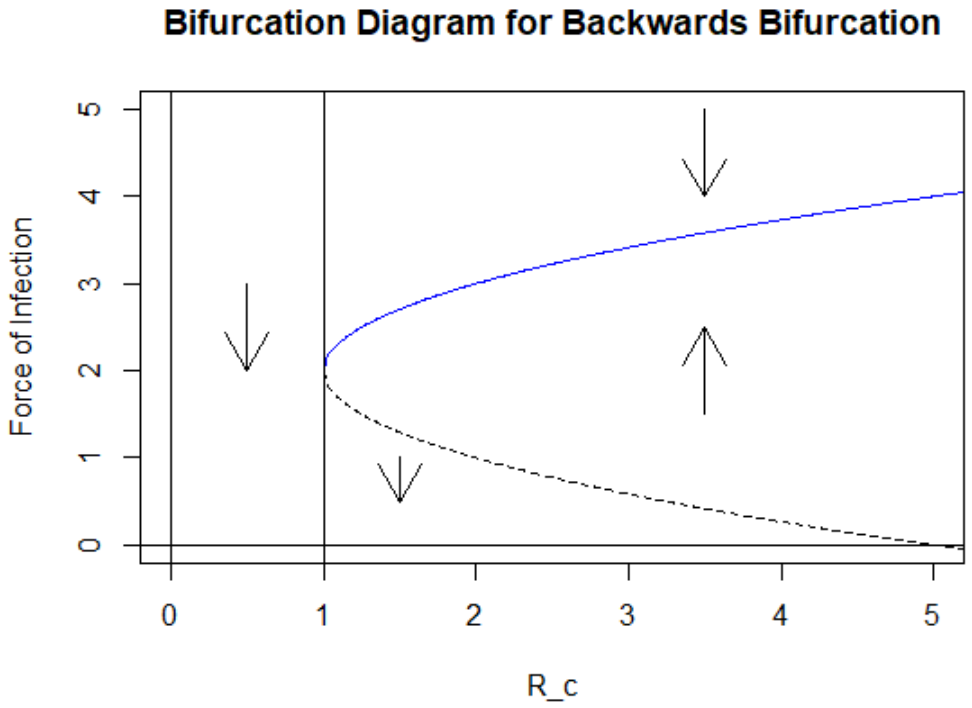


Figure 3: Bifurcation diagram showing qualitative features of model with backwards bifurcation; below some threshold of reproduction number, only a disease-free equilibrium exists. Beyond this point, two stable equilibria exists - one disease-free, and one endemic.

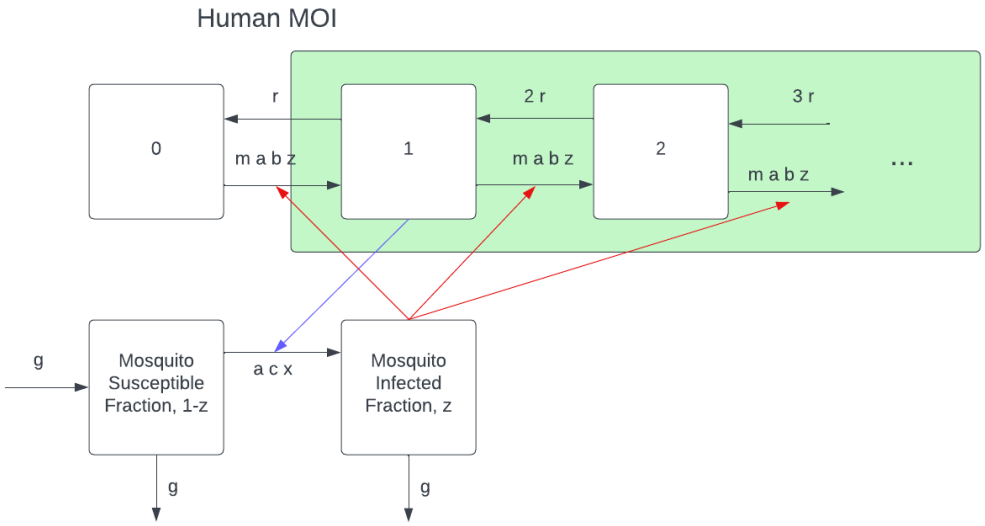


Figure 4: State space of a model with human superinfection and simple mosquito infection. The distribution of the MOI in humans under this model is Poisson.

it develops slowly with exposure and acts to clear infections more quickly.

Later, a version of this superinfection model was simplified under a different equilibrium assumption in [8] - a quasiequilibrium in which the distribution of the multiplicity of infection is Poisson with a dynamically changing intensity parameter M, which follows a relatively simple equation:

$$\frac{dM}{dt} = mabz - rM$$

the prevalence x is then given by the probability of having at least one infection (see the shaded green box in figure 3):

$$x = 1 - e^{-M}$$

This strategy of modeling the stationary distribution and tracking its behavior with a single variable is one we will use often here.

While research continues on understanding how the dynamics of these relatively simple compartment-based models are impacted by complicating realistic factors such as delays caused by heterogeneity by age cohort (resulting in partial differential equations), incubation periods (resulting in delay differential equations), behavior in response to external random environmental fluctuations (resulting in stochastic differential equations), and inhomogeneous mixing (resulting in integro-differential equation) to name a few, much of policy research has moved on to reckon with building entirely stochastic individual based models in order to increase perceived realism and build in probabilistic behavior. These models are a powerful way to incorporate proposed strategies by policymakers in a way which gives output with measures of uncertainty, and are therefore essential tools for modern epidemiology. They are also tools for investigating the impact of proposed mechanisms of human-parasite interactions within the body, and how they change with exposure. However, these models tend to become so complex as to make the parameter space high dimensional and hard to fit to data; the output is stochastic and therefore must be run many times, which can be computationally costly; and the mechanism is opaque, hidden under within black box of how individual probabilistic behavior results in population level metrics.

Here, we propose a different approach embracing the probabilistic structure of the system. In chapter 1, we start by demonstrating how the distribution of the multiplicity of infection modeled above is impacted by access to treatment, a major factor for the state of the system. In chapter 2 we show how infection age is statistically predictive of parasitemia, and therefore detection, fever rates, and transmission rates in data from controlled human infections. These relationships imply that infection age, a relatively simple quantity to model, can help us build predictive models for the latent and highly variable parasite densities. Chapter 3 explores this through simulation, and develops a theoretical framework for tracking the probability densities using simple ODEs. Chapter 4 expands on this, and by treating the outcomes of interest as generalized linear models of our semi-Markov model of infection age, we obtain

simple ODE models for these probability density function tracking variables, with closed form equations relating them to observables such as theoretical true prevalence, detection, and fever-prompted treatment (using the results from chapter 1 along the way), all while incorporating immunity as a covariate. Finally in chapter 5 we turn our attention to fitting these models to time series data of observables, and ultimately develop an algorithm for fitting without any probabilistic simulation of the underlying stochastic process. It is our hope that this series of papers inspires others to build upon and push the new tools developed here, and use them to better improve future efforts toward malaria eradication everywhere.

Chapter 1

A HYBRID MODEL FOR THE EFFECTS OF TREATMENT AND DEMOGRAPHY ON MALARIA SUPERINFECTION

Malaria continues to have a significant impact on the world, and is responsible for over 400,000 deaths in 2015 alone [9]. Over the past century a repertoire of interventions has been developed to prevent its spread and reduce its impact on those afflicted, but the parasite continues to adapt and persist. Because malaria dynamics and control are so complex, mathematical and statistical models play a significant role in the strategic implementation of these existing tools to maximize their impact. One of the historically difficult aspects of malaria and other parasitic diseases to model is the process of superinfection, wherein multiple distinct cohorts of pathogens infect the host simultaneously [10, 8]. The standard SIS system usually used to model similar disease transmission using a constant recovery rate from a single infected class does not capture this type of heterogeneity, and may therefore be a poor approximation to the real system.

There has been a long history of model development to address the complex dynamics of a pathogen capable of superinfection, such as the parasite which causes malaria. Mathematical models were developed with different disease states to track the proportion of individuals with n infections, each of which is still susceptible to further infection [7, 6, 11]. This same problem has been approached in the context of the accumulation and clearance of macroparasites in their hosts [12]. Additionally, the development of strain-dependant immunity, host heterogeneity, drug resistance, and a cap on superinfection have been explored in previous models [13, 14, 15, 16, 17]. However, to the author's understanding there have been no models which investigate the impact of demographic turnover and access to treatment on the pathogen distribution.

Care must be taken in modeling the distribution of the multiplicity of infection (MOI), that is the fraction of the population that has exactly n "broods" or cohorts of parasites. This distribution strongly affects the population mean recovery rate, which can be seen as a weighted sum of the rate at which individuals are recovering from n infections, weighted by the proportion of individuals with n infections. The shape of this distribution can be profoundly altered by factors such as seasonality and case management, which in turn can have significant effects on the mean rate of recovery for the population and the prevalence of the disease. Under George Macdonald's original assumption that infection status is no barrier to reinfection, MOI can be modeled as an $M/M/\infty$ queue [7, 11]. Though Macdonald's written description of the system was correct, he mistakenly presented it mathematically as an M/M/1 queue [6]. This system vastly overestimates an individual's time spent infected, with the average becoming infinite if the force of infection exceeds the rate of recovery. Macdonald's model and its description generated some confusion, substantial discussion, and a series of approximating models [18, 19].

Since the historical debate over defining the problem in the relatively simple case of non-interacting strains was settled, there have been discussions around the impact of the strains interacting indirectly through immune selection [20]. Klein et al developed a model which suggested that it was the initial immune response against the most dominant strain which dictated whether or not a chronic low density infection would take place [21]. This indicates diversity could lead to some infections being significantly shortened due to common antigens with other infections. In addition, the existence of strong cross immunity may also result in the suppression of a similar but more effective type of antigen being produced. However, this type of model would assume knowledge of the number and distribution of the variants and relative cross strain reactivity in the parasite population. Previous work has been done on the interaction of strain specific immunity and superinfection [14]. Therefore to isolate the effects of treatment in particular, we will work under Macdonald's assumption of non-interacting strains and we will leave investigation of these more complex interactions with treatment to be explored in future studies.

For similar reasons, the change in parasitemia within hosts is not explicitly modeled here. This is not to say that it is unimportant for a transmission model; parasite densities are strongly related to severe disease, treatment-seeking behaviors, acquisition of immunity, and transmission rates to mosquitoes. Rather, it is because of the complexities inherent in making a good model with this particular feature. Parasitemia is dependent on the species of parasite, can vary across individuals with age and history of exposure, and it is highly heterogeneous throughout the course of infection. In addition, including immunity in any model which lacks strain diversity could lead to misleading behavior in parasite composition and density because immunity is so intimately tied with strain variability and density [22]. Therefore to reduce data complexity necessary to parameterize the model and increase tractability of results, we don't model the parasitemia across infected individuals in this study.

However, keeping the previous caveats in mind we have two advantages in the present model. First, we are able to relax the strong assumptions of the original Ross-Macdonald model with superinfection of no demographic turnover and no treatment. This is discussed in the subsequent paragraph. Second, and more importantly, we can obtain a model which relates the rate of recovery of a single infection [23], the rate of demographic turnover in a population, and the treatment rate to the force of infection given an estimate of prevalence. The model also provides the entire expected distribution of the complexity of infection in a population. Because the total parasitemia ranges more than 10 orders of magnitude in a human host through the course of an infection, this information cannot currently be obtained through direct epidemiological studies as only the dominant strain will likely be detected [24]. Inference from the distribution itself has implications for understanding and forecasting the burden of the disease, the efficacy of treatment, and the development of immunity in a population.

The condition we relax is a strong steady state condition on the population wherein individuals experience no demographic turnover or external forces that influence their susceptibility or infection status such as treatment. Critically, finite lifespans and the availability

of treatment will affect the MOI distribution. Including their interactions ad hoc in ordinary differential equation (ODE) models may lead to inconsistent transition rates, and even inconsistent equilibria, when compared to the original assumptions of the stochastic model. This is an assumption we wish to relax in this paper to explore the consequences, as many populations living in endemic areas now have some access to antimalarials.

In the present study we consider the system with demography and treatment, with and without transient chemoprotection, and present a new way of modeling the statics and dynamics of the distribution of MOI and its nonlinear effects on the population mean recovery rate and resulting prevalence. In addition, we show that this distribution is stable in the sense that any initial distribution with finite mean converges in distribution to the stationary distribution. We then show the impact of an extended protective period post-treatment, which has the effect of inducing a zero-inflation in the MOI distribution. Using these results and their resulting dynamics, we then extend the Ross-Macdonald model to include this heterogeneity and explore the effects of mortality and treatment on prevalence under these assumptions.

1.1 Model with Demography and Treatment, without Chemoprotection

1.1.1 Model Description and Setup

Consider a Poisson process $X_t(\lambda)$ of new infections in an individual person with intensity λ, the force of infection (FOI). We will assume λ may be time-varying, but it is independent of the current number of infections as per Macdonald's original assumption. In addition, we will have a second Poisson process $Y_t(rX_t)$ of recovery with an intensity rX_t. Again, we make no assumption about r aside from it being independent of X_t or Y_t. Then the M/M/∞ queue can be represented by a third Poisson process, defined to be the difference of the two: $Z_t := X_t - Y_t$. This process counts the current number of infections in an individual, and its

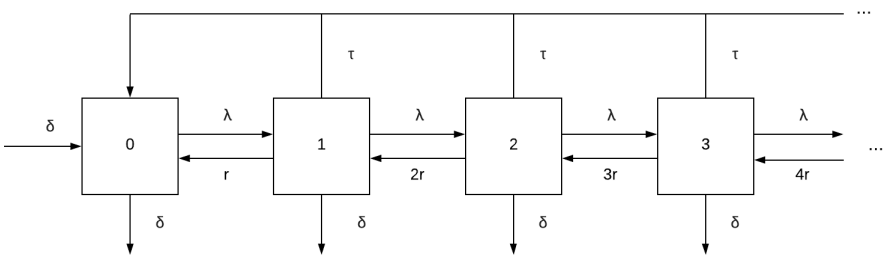


Figure 1.1: State space of human infections without chemoprotection. Numbers indicate the number of infections in an individual and the letters denote transition rates between states.

range is on the non-negative integers. To add demography and treatment, which both act to decrease from the infected classes and increase the susceptible class, we need additional Poisson processes $D_t(\delta)$ and $T_t(\tau)$, where δ and τ are respectively the mean birth/death rate (assumed equal to maintain constant population sizes) and the mean treatment rate. Treatment only acts on those with at least one infection, and because births are assumed to replace deaths, demography also has a net effect of removing infected individuals and introducing susceptible individuals. Therefore for brevity we can replace the two processes with a single Poisson process M_t with rate μ, where $\mu := \delta + \tau$ is the expected rate at which either death or treatment occurs.

The master equations for the system can be built by considering the state space in Figure 1:

$$\frac{dx_0}{dt} = \delta - \delta x_0 - \lambda x_0 + r x_1 + \tau(1 - x_0)$$

$$\frac{dx_n}{dt} = \lambda x_{n-1} + r(n+1)x_{n+1} - (\lambda + rn + \tau + \delta)x_n$$

which we can rewrite using the combined treatment and demographic process with parameter μ:

$$\frac{dx_0}{dt} = -\lambda x_0 + r x_1 + \mu(1 - x_0)$$

$$\frac{dx_n}{dt} = \lambda x_{n-1} + r(n+1)x_{n+1} - (\lambda + rn + \mu)x_n$$

As this is an upper triangular infinite system of linear ODEs, it can't be solved iteratively; instead we will introduce a generating function of the form

$$G(t, s) = \sum_{n=0}^{\infty} x_n s^n$$

which encodes the infinite variables as coefficients in its power series, simplifying further analysis. We can directly substitute:

$$\frac{\partial G}{\partial t} = \frac{dx_0}{dt} + \sum_{n=1}^{\infty} \frac{dx_n}{dt} s^n$$

$$= \mu - (\mu + \lambda)x_0 + r x_1 + \sum_{n=1}^{\infty} [-(\mu + \lambda + rn)x_n + \lambda x_{n-1} + r(n+1)x_{n+1}]s^n$$

This simplifies to the following partial differential equation (PDE) for the probability generating function (PGF) G:

$$\frac{\partial G}{\partial t} - r(1 - s)\frac{\partial G}{\partial s} = \mu - (\mu + \lambda(1 - s))G$$

Because G is a probability generating function, a natural boundary condition is imposed wherein $\lim_{s \to 1^-} G(t, s) = 1$ owing to the fact that the sum of the coefficients of the power series is equal to the sum of all the probabilities.

1.1.2 Statics

Note that if we let $\mu = 0$ then we obtain the equation whose solution is the generating function of a Poisson distribution, aligning with known results [7]. Including demography

or treatment forces the stationary distribution to belong to a different class. To find this class, we will start by solving for the stationary distribution whose generating function can be found by setting $\frac{\partial G}{\partial t} = 0$, leading to the following ODE:

$$\frac{dG}{ds} - \left(\frac{\beta}{1-s} + \alpha\right) G = -\frac{\beta}{1-s}$$

with the condition that $\lim_{s \to 1^-} G = 1$ which is true for any probability generating function.

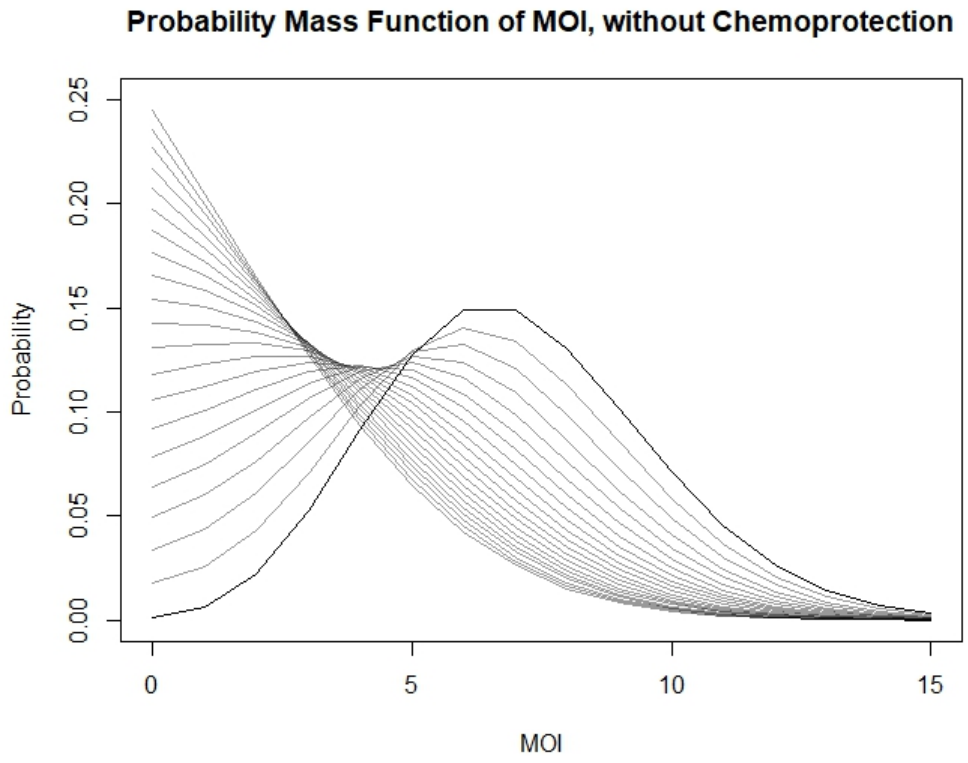


Figure 1.2: Plots of the PMF of the stationary distribution of MOI with varying values of treatment and demography, corresponding to the alternative hyper-Poisson (AHP) distribution. α, which is proportional to the force of infection, is held constant at 7. We vary values of β, defined here to be the ratio of the rate of treatment or death to the rate of recovery. β is varied from 0 to 2 in increments of .1. The bold line represents a Poisson distribution, $\beta = 0$. Increasing β suppresses MOI and increases the probability of having fewer infections, and it increasingly resembles a geometric distribution

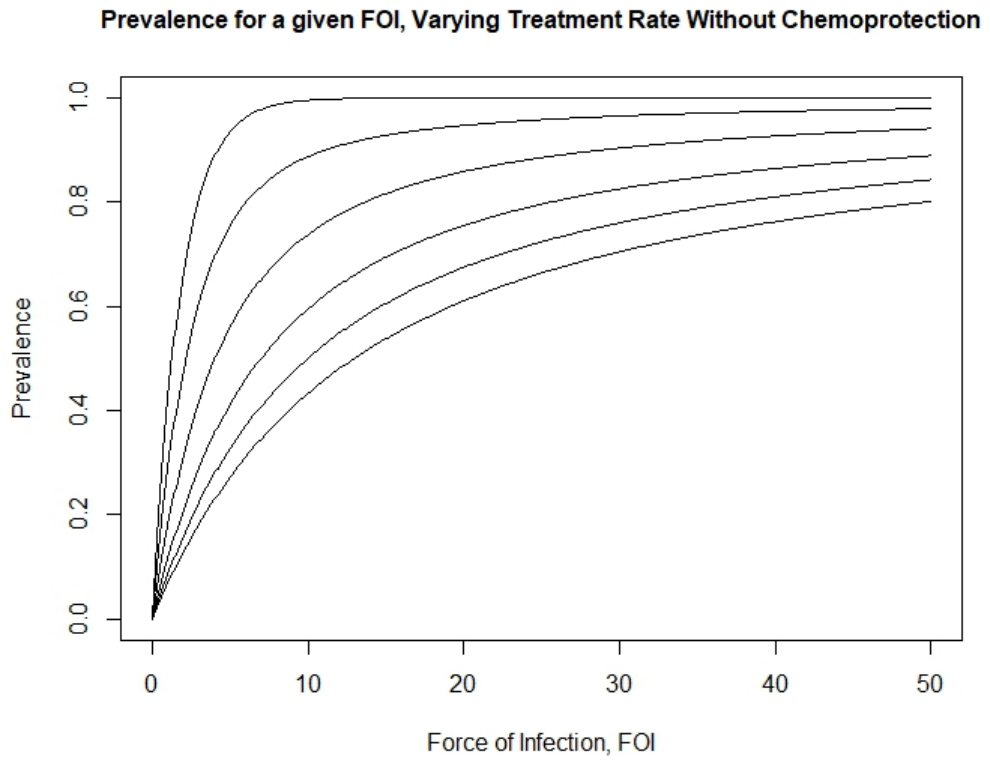


Figure 1.3: Plot of the prevalence as a function of annual FOI, excluding short term protection from treatment. The curves represent the expected prevalence for increasing FOI with set treatment rates. Prevalence monotonically decreases with increasing treatment rate. From top to bottom, the curves represent treatment rates of 0, 1, 3, 6, 9, and 12 per year.

In the static case, the stationary distribution is parameterized by two parameters: $\alpha = \lambda/r$ corresponding to the expected number of infections per recovery, and $\beta = \mu/r$ which is the expected number of demographic events or treatments per recovery. As the parameters are independent of s, this equation can be solved with the application of an integrating factor:

$$G(z) = Ce^{\alpha s}(s-1)^{-\beta} + \beta e^{\alpha(s-1)}(\alpha(s-1))^{-\beta}\gamma(\beta, \alpha(s-1))$$

where $\gamma(a, x)$ is the lower incomplete gamma function. Because any PGF has the property that the left limit as s approaches 1 should equal 1 and the term next to the integration

constant has a singularity at 1, the constant must be zero. Therefore the PGF for the steady state distribution of MOI is given by

$$G(s) = \beta e^{\alpha(s-1)}(\alpha(s-1))^{-\beta}\gamma(\beta, \alpha(s-1))$$

Replacing γ with its power series representation yields

$$G(s) = \Phi(1; \beta+1; \alpha(s-1))$$

where $\Phi(a; b; x)$ is Kummer's confluent hypergeometric function, defined by

$$\Phi(a; b; x) = \sum_{n=0}^{\infty} \frac{a^{(n)}}{b^{(n)}} \frac{x^n}{n!}$$

where $y^{(n)}$ is the rising factorial of y. This is the PGF of the alternative hyper-Poisson distribution, $AHP(\alpha, \beta+1)$ [25], whose probability mass function (PMF) is

$$P(N=n) = \frac{\alpha^n}{(\beta+1)^{(n)}}\Phi(1+n; \beta+1+n; -\alpha)$$

This distribution is a more general form of a Poisson distribution with β acting as a dispersal parameter. As $\beta > 0$ for any model that includes demography or treatment, the distribution is overdispersed compared to a Poisson distribution. A plot of the distribution for varying β is given in Figure 2. Note that despite qualitative similarities, it is distinct from the negative binomial distribution in both form and derivation - and in many cases due to its similarity to the Poisson distribution, as it is in the same family, it is sometimes a more appropriate alternative overdispersed distribution.

Knowing the distribution of MOI is sufficient to get an estimate of the expected prevalence. Prevalence in this model is the complement of the proportion of individuals with zero infections, which is equal to the PGF evaluated at zero. Figure 3 shows plot of the prevalence as a function of the force of infection under varying treatment rates using this conversion. As expected, prevalence increases with FOI and decreases with treatment.

1.1.3 Dynamics

Now that we know the stationary distribution, we can turn our attention to the dynamics of MOI. In order to proceed, note that the equilibrium distribution will always be AHP-distributed with α and β as being (possibly time-varying) parameters for the distribution. As these parameters depend on the three original parameters λ, r, and μ, they satisfy their own ODEs which will be derived here. This invariance of the class of AHP distributions under the dynamics can be seen by noting the set of PGFs corresponding to the class of AHP distributions constitute the only equilibrium solutions to the PDE, and once the system has entered this class it will not leave it without external perturbation.

To begin we go back to our PDE for the PGF and take a partial derivative with respect to s:

$$\frac{\partial^2 G}{\partial t \partial s} + r\frac{\partial G}{\partial s} - r(1-s)\frac{\partial^2 G}{\partial s^2} = -(\mu + \lambda(1-s))\frac{\partial G}{\partial s} + \lambda G$$

Plugging in $s = 1$ gives us the governing equation for the first factorial moment, also equal to the first raw moment:

$$\frac{dm_1}{dt} = \lambda - (\mu + r)m_1$$

we then repeat the previous process with a second partial derivative to find an equation for the second factorial moment:

$$\frac{\partial^3 G}{\partial t \partial s^2} + 2r\frac{\partial^2 G}{\partial s^2} - r(1-s)\frac{\partial^3 G}{\partial s^3} = -(\mu + \lambda(1-s))\frac{\partial^2 G}{\partial s^2} + \lambda\frac{\partial G}{\partial s}$$

from which we can derive the equation for the second raw moment

$$\frac{dm_2}{dt} = \lambda + (2\lambda + r)m_1 - (\mu + 2r)m_2$$

using the relationship between the second raw moment and the second factorial moment,

that is

$$E[X(X-1)] = m_2 - m_1$$

We then can match parameters to moments to determine the relationship between the transient moments and the transient parameter values that codetermine the exact distribution.

For this we note that for a random variable $X \sim AHP(\alpha, \beta+1)$,

$$E[X] = \frac{\alpha}{\beta+1}$$

and the variance is

$$Var(X) = \frac{\alpha}{\beta+1}\left(1 + \frac{\alpha}{\beta+1}\frac{\beta}{\beta+2}\right)$$

Matching moments allows us to get a functional relationship between the dynamics of the moments and the shape of the distribution. Doing so gives us

$$m_1 = \frac{\alpha}{\beta+1}$$

and

$$m_2 - m_1^2 = \frac{\alpha}{\beta+1}\left(1 + \frac{\alpha}{\beta+1}\frac{\beta}{\beta+2}\right)$$

solving the system in terms of α and β gives us

$$\alpha(m_1, m_2) = \frac{m_1(m_2 - m_1)}{2m_1^2 + m_1 - m_2}$$

$$\beta(m_1, m_2) = \frac{2m_2^2 - 2m_1 - 2m_1^2}{2m_1^2 + m_1 - m_2}$$

In particular, if we assume the system reached the limiting PMF at some point in the past without subsequent perturbation, then this combined with the PMF of the AHP distribution gives us a method of modeling the prevalence with a much more manageable system of two

ODEs for the two moments necessary to parameterize the transient distribution:

$$\frac{dm_1}{dt} = \lambda - (\mu + r)m_1$$

$$\frac{dm_2}{dt} = \lambda + (2\lambda + r)m_1 - (\mu + 2r)m_2$$

which we can convert back to prevalence x by noting that prevalence is the complement of the probability of having zero infections:

$$\alpha(m_1(t), m_2(t)) = \frac{m_1(m_2 - m_1)}{2m_1^2 + m_1 - m_2}$$

$$\beta(m_1(t), m_2(t)) = \frac{2m_2^2 - 2m_1 - 2m_1^2}{2m_1^2 + m_1 - m_2}$$

$$x(t) := 1 - x_0(t) = 1 - \Phi(1; \beta(t) + 1; -\alpha(t))$$

Note that we can also find $x_n(t)$, the proportion of individuals with n infections, by considering the PMF of the stationary distribution:

$$x_n(t) = \frac{\alpha(t)^n}{(1 + \beta(t))^{(n)}} \Phi(1 + n, \beta(t) + 1 + n; -\alpha(t))$$

with the same definitions as before.

In addition, this means that if we are given an explicit time-dependent functional form for the FOI λ, we would be able to find the closed form solution for prevalence over time by solving the inhomogeneous system of two linear equations for the moments iteratively through the use of integrating factors. That is, in the language of queueing theory, we could find the transient solution for the PMF of states given the initial distribution is AHP-distributed. This assumption will be relaxed and a more general argument for the behavior of solutions will be made in the next section.

1.2 Model with Demography and Treatment, with Chemoprotection

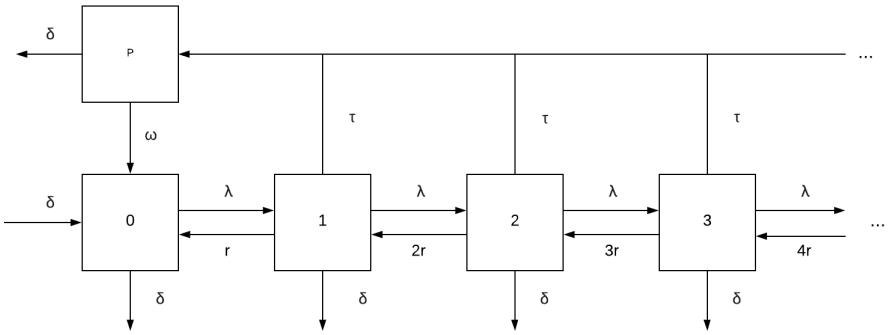


Figure 1.4: State space of human population with chemoprotection. Numbers indicate the number of infections in an individual and the letters denote transition rates between states, and the p compartment denotes the protected class

Now we will consider the same system as before but we further assume that treatment confers some temporary protection from subsequent infection. That is, once someone is treated they enter a protected state as the treatment will linger in the body and protect against new infections for an average of $1/\omega$ units of time before the individual returns to a susceptible state. The state space along with the transition rates are given in figure 4. The distribution of MOI conditioned on not being protected will remain AHP-distributed because adding this compartment does not change the queuing system - it adds a "waiting room" between bouts in the queue. Repeating the analysis done before by introducing a PGF results in another AHP-distribution, but with a zero inflation introduced by the proportion of individuals in the protected class. Furthermore, we can also repeat the stability analysis on the PDE/ODE system and see that it is also stable in the sense that any initial distribution with finite mean will asymptotically approach the zero-inflated AHP distribution. Treatment with waning protection therefore has three major effects on the distribution of MOI: decreasing the expected value, inducing overdispersion relative to a Poisson distribution, and zero-inflation. Each of these effects can individually correspond with an overall

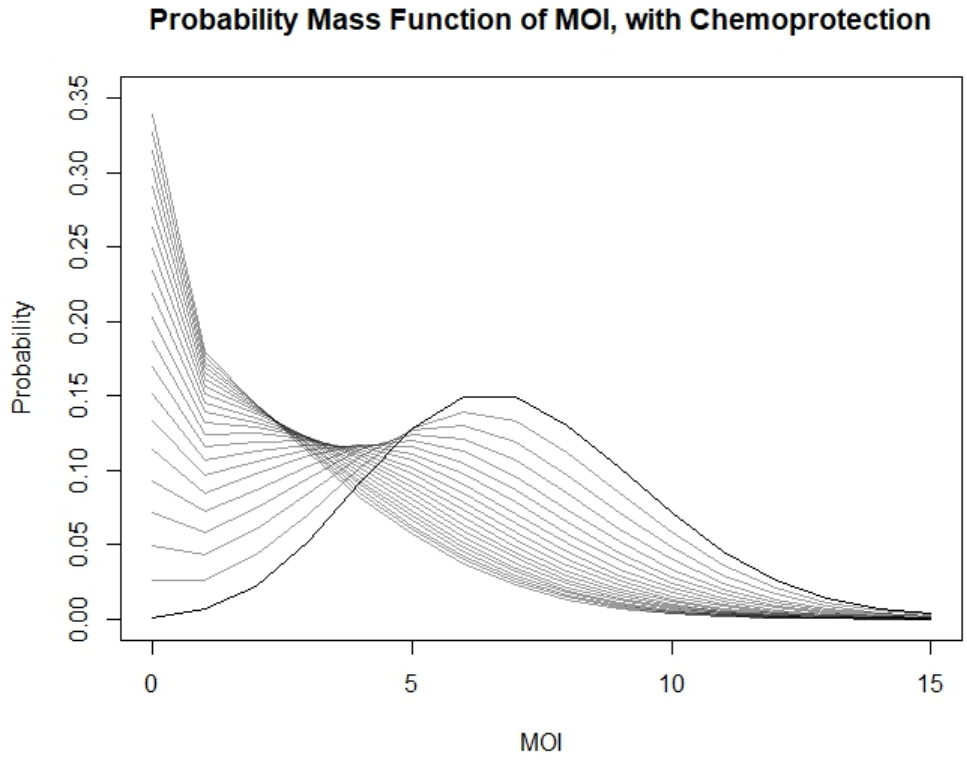


Figure 1.5: Plots of the PMF of the stationary distribution of MOI with varying values of treatment and demography, corresponding to the zero-inflated alternative hyper-Poisson (AHP) distribution. Note the distribution looks the same as in figure 2, but with increased probability mass at zero due to the fraction of the population in a protected state. Mean protection time ($1/\omega$) is set to 30 days.

decrease in prevalence. See figures 5 and 6 for plots of the PMF of MOI and prevalence curves, respectively.

With the same notation as before and letting p represent the proportion of individuals in the protected class, the full model with the protected state can be written as

$$\frac{dm_1}{dt} = \lambda - (\mu + r)m_1$$

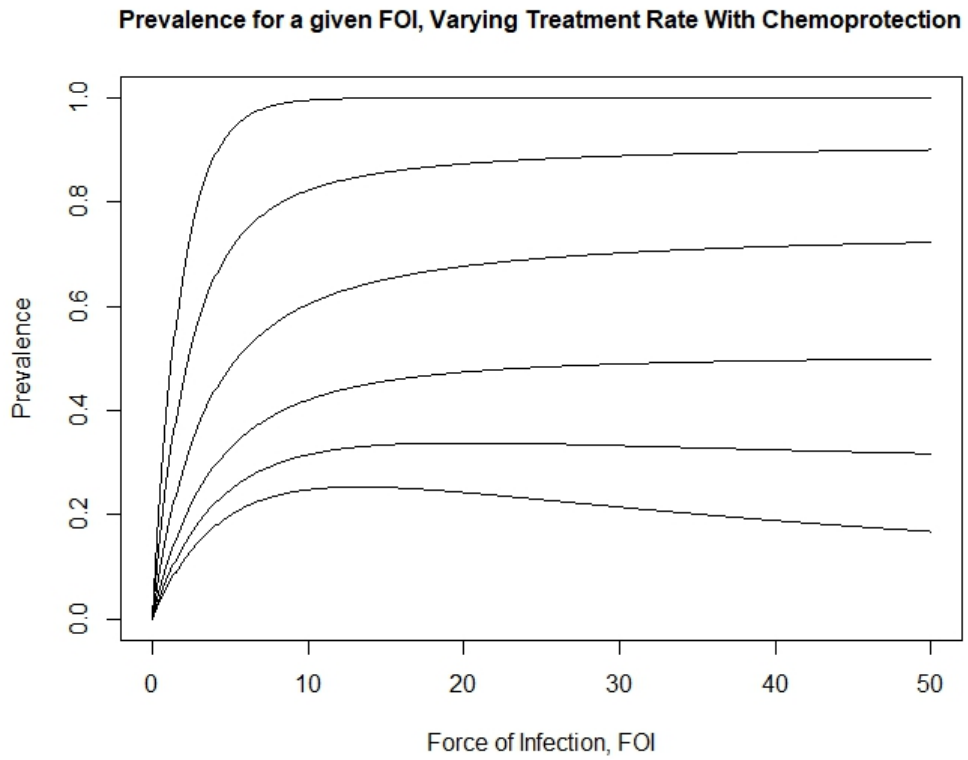


Figure 1.6: Plot of the prevalence as a function of annual FOI, including short term protection from treatment. Average duration of protection is set to 30 days. The curves represent the expected prevalence for increasing FOI with set treatment rates. Prevalence monotonically decreases with increasing treatment, but at high treatment rates prevalence will drop with increasing FOI due to the increase to the protected class. From top to bottom, the curves represent treatment rates of 0, 1, 3, 6, 9, and 12 per year.

$$\frac{dm_2}{dt} = \lambda + (2\lambda + r)m_1 - (\mu + 2r)m_2$$

$$\frac{dp}{dt} = \tau x - \omega p$$

where

$$\alpha = \frac{m_1(m_2 - m_1)}{2m_1^2 + m_1 - m_2}$$

$$\beta = \frac{2m_2^2 - 2m_1 - 2m_1^2}{2m_1^2 + m_1 - m_2}$$

$$x(t) = 1 - [(1-p)\Phi(1; \beta + 1; -\alpha) + p]$$

$$= (1-p)\left[1 - \Phi(1; \beta + 1; -\alpha)\right]$$

Note that x has a nice physical interpretation in this form - it is the probability that an individual has at least 1 infection given that they are not protected, times the probability they are not protected.

1.3 Pseudoequilibrium Model

Now that we have transient behavior established, we can go one step further and include a pseudoequilibrium analysis. That is, rather than having to deal with variables which parameterize the distribution we can assume the distribution approaches equilibrium on timescales that are much smaller than the rate at which the force of infection changes. This was implicitly made in the model developed for the Garki project [7], wherein the rate of recovery was given as the reciprocal of the expected duration of infection. This gives us the correct mean behavior, at the sacrifice of information on higher order moments. This simplified model captures the behavior of the full model assuming the force of infection does not change too rapidly or heterogeneously in the population, and if the system is close to its equilibrium distribution. Because the system is asymptotically stable (as proven in the appendix) this assumption is justified, and the system becomes much simpler to investigate and simulate. This would correspond to an endemic region with little seasonality

in transmission.

If we denote the time spent infected as X, then the expression for the expectation is as follows:

$$E[X] = \frac{e^{\lambda/r} - 1}{\lambda}$$

It's useful to compare this to the case without superinfection. Because each infection has an exponentially-distributed time to clearance, without superinfection the expected time spent infected is $1/r$. In the case with superinfection, we can see first that it is monotonically increasing with respect to the force of infection. In addition, for small FOI the Taylor expansion of the exponential shows that the expected time is approximately $1/r$:

$$E[X] = \frac{\lambda/r + O(\lambda^2)}{\lambda} = 1/r + O(\lambda)$$

Now we can consider the same simplification with treatment and demography. In order to consider this case, we need to know how this expectation changes when introducing exponentially-distributed treatment and demography. Although the distribution of the time to recovery starting at a single infection is not known in closed form, its Laplace transform is - and we can exploit that to determine the impact of these perturbations [26]. Denoting the time until treatment (or death) with T, then the total duration of infection D is $D = \min\{X, T\}$. That is, the infection will end if an individual either recovers naturally or gets treated or dies, whichever comes first. Then we can write down the following:

$$P(D \leq t) = P(X \leq t) + P(T \leq t) - P(X \leq t \ \& \ T \leq t)$$

$$= P(X \leq t) + P(T \leq t) - P(X \leq t)(T \leq t)$$

where this step uses the independence of the natural recovery and treatment/demography processes. As these are exactly the CDFs of the respective variables, we get

$$F_D(t) = F_X(t) + F_T(t) - F_X(t)F_T(t)$$

Because treatment and death are exponentially distributed, the minimum of the two is also exponentially distributed with a rate equal to the sum of their rates. We will again denote this new rate as μ. Knowing this allows us to plug in the CDF of this exponential distribution for F_T:

$$F_D(t) = F_X(t) + 1 - e^{-\mu t} - (1 - e^{-\mu t})F_X(t)$$

$$= 1 - e^{-\mu t} + e^{-\mu t}F_X(t)$$

Using the fact that the integral of the complement of the CDF is equal to the expectation for any random variable, we get

$$E[D] = \int_0^\infty (1 - F_D(t))dt$$

$$= \int_0^\infty e^{-\mu t} - e^{-\mu t}F_X(t)dt$$

$$= 1/\mu - \mathcal{L}F_X(\mu)$$

where here $\mathcal{L}$ denotes the Laplace transform. Finally, using the relationship between the Laplace transform and the integral of a function gives us a formula for the expected duration of infection as a function of the Laplace transform of X:

$$E[D] = \frac{1 - E[e^{-\mu X}]}{\mu}$$

In proposition 4.1 from the paper of Guillemin et al, the Laplace transform of the duration spent infected (there called the "congestion duration") starting with a single infection is given

by

$$E[e^{-\mu X}] = \frac{r}{r+\mu}\frac{\Phi(\mu/r;\mu/r+2;-\lambda/r)}{\Phi(\mu/r;\mu/r+1;-\lambda/r)}$$

which, plugging in, gives us the expected duration of an infection with superinfection and treatment

$$E[D] = \frac{1-\frac{r}{r+\mu}\frac{\Phi(\mu/r;\mu/r+2;-\lambda/r)}{\Phi(\mu/r;\mu/r+1;-\lambda/r)}}{\mu}$$

Although this rate looks complicated it is consistent; the limit as μ goes to zero is exactly the rate found previously and implemented in the Garki model [7].

Before plugging this in to a model of infection and recovery, we must also determine the proportion of people who were treated rather than died or recovered, as they will enter the protected class. This can by found again through a probabilistic argument:

$$P(\text{Treat}) = P(\text{Treat before Die } |T \leq X)P(T \leq X)$$

That is, the probability of being treated to end the infection is the product of the probabilities of ending the treatment through either treating or dying given that an individual won't recover naturally, times the probability that the individual won't recover naturally. The first probability is the probability that one exponential random variable is less than another, and in this case is $\tau/(\tau+\delta) = \tau/\mu$. The second probability is more interesting, since again with superinfection X is not exponentially-distributed. Again we can write this probability out:

$$P(T \leq X) = \int_0^\infty P(T \leq t)P(X = t)dt$$

$$= \int_0^\infty F_T(t)f_X(t)dt$$

$$= \int_0^\infty (1 - e^{-\mu t}) f_X(t) dt$$

$$= 1 - E[e^{-\mu X}]$$

which shows that the Laplace transform of a random variable X can also be interpreted as the probability that a random variable defined over the nonnegative reals is less than or equal to an exponential random variable with rate μ. Note that this quantity as a probability is bounded between 0 and 1, and as μ increases to infinity this probability decreases to 0. Multiplying those probabilities, we see

$$P(\text{Treat}) = \frac{\tau}{\mu}(1 - E[e^{-\mu X}]) = \tau E[D]$$

Therefore, we can complete the simplified model by replacing the simple rate of recovery r from infected to susceptible with the reciprocal of this duration of infection as follows:

$$\frac{dx}{dt} = \lambda(1 - x - p) - \frac{x}{E[D]}$$

$$\frac{dp}{dt} = \frac{P(\text{Treat})}{E[D]} x - \omega p$$

which we found can be written as

$$\frac{dx}{dt} = \lambda(1 - x - p) - \frac{\mu}{1 - \frac{r}{r+\mu} \frac{\Phi(\mu/r; \mu/r+2; -\lambda/r)}{\Phi(\mu/r; \mu/r+1; -\lambda/r)}} x$$

$$\frac{dp}{dt} = \tau x - \omega p$$

This reduces the model down to just the two equations that will adequately describe what proportion of individuals infected in a single variable, x, given that the distribution of MOI which determines the rate of recovery is transiently at (or very close to) its steady state.

1.4 Connecting to Ross-Macdonald Dynamics

Now that we have fully established the system of infections and recoveries on the human side, we can couple the human system with a vector system to observe the transmission dynamics. This is the hybridization step, wherein we reinterpret the probability distribution of MOI as a frequency distribution for a large population. This coupling with simple vector dynamics gives us a Ross-Macdonald style model. Ross-Macdonald models are SIS compartment models with vector-based transmission used to model malaria transmission. There is no single canonical form for a Ross-Macdonald model (see for example the review by Smith et al. for an overview of the history and main developments in Ross-Macdonald models [19]), but as an example we chose the one implemented by Smith and McKenzie because there is only one additional equation for mosquitoes without any delay dynamics, but the extrinsic incubation period is still taken into account [27].

In this model the force of infection is proportional to the number of infectious mosquitoes. That is, $\lambda = \frac{M}{H}abz$ where $\frac{M}{H}$ is the ratio of mosquitoes to humans, a is the rate at which mosquitoes bite humans, b is the proportion of infectious bites which result in human infections, and z is the infectious fraction of mosquitoes. Replacing this expression for the force of infection in the original system for the PGF of the MOI G, we get the following PDE/ODE system:

$$\frac{\partial G}{\partial t} = r(1-s)\frac{\partial G}{\partial s} + \delta + \tau(1-p) - (\delta + \tau + \frac{M}{H}abz(1-s))G$$

$$\frac{dp}{dt} = \tau x - \omega p$$

$$\frac{dz}{dt} = acx(e^{-g\nu} - z) - gz$$

with

$$x(t) = (1-p)(1-G(t,0))$$

where x is the proportion of humans infected, c is the probability an infectious human will infect a susceptible mosquito during a blood meal, g is the death rate of mosquitoes, and ν is the extrinsic incubation period. As done before, we include the following boundary condition for the PGF G: $\lim_{s\to 1^-} G(t,s) = 1$. This system will hold for any biologically reasonable initial distribution of MOI and fraction of infectious mosquitoes.

However if our initial MOI is AHP-distributed, then the dynamics can be entirely captured by a system of four ODEs:

$$\frac{dm_1}{dt} = \frac{M}{H}abz - (\tau + \delta + r)m_1$$

$$\frac{dm_2}{dt} = \frac{M}{H}abz + \left(2\frac{M}{H}abz + r\right)m_1 - (\tau + \delta + 2r)m_2$$

$$\frac{dp}{dt} = \tau x - \omega p$$

$$\frac{dz}{dt} = acx(e^{-g\nu} - z) - gz$$

where

$$x(t) = (1-p)\left[1 - \Phi(1;\beta+1;-\alpha)\right]$$

$$\alpha(t) = \frac{m_1(m_2 - m_1)}{2m_1^2 + m_1 - m_2}$$

$$\beta(t) = \frac{2m_2^2 - 2m_1 - 2m_1^2}{2m_1^2 + m_1 - m_2}$$

c is the probability an infectious human will infect a susceptible mosquito during a blood meal, g is the death rate of mosquitoes, and ν is the extrinsic incubation period. These ODEs are derived in the same manner as before, this time tracking the extra compartment.

This ODE system is a bit different from most implementations of the Ross-Macdonald model, but similar to the one introduced by Nåsell [8]. The first two equations determine the moments of the distribution of MOI while the third determines the degree of zero-inflation. These are converted to current prevalence by taking the complement of the probability an individual has zero infections. This in turn affects the rate of recruitment of mosquitoes to the infectious class in the fourth equation, which is assumed proportional to the force of infection acting on the distribution of susceptible people.

It's important to recognize that this will hold if the initial distribution is in the null space of the linear operator described in the previous section, which corresponds to the stationary AHP distribution. However, because we proved any distribution asymptotically approaches the class of AHP distributions, knowing malaria has been endemic in a region for a long period of time without strong perturbation to the structure of the system is enough to make this a reasonable assumption for modeling local transmission.

Finally, we can also couple the pseudoequilibrium simplification to transmission dynamics in an analogous way to achieve a set of three simpler, albeit still nonlinear, ODEs:

$$\frac{dx}{dt} = \frac{M}{H}abz(1-x-p) - \frac{\mu}{1-E[e^{-\mu X}]}x$$

$$\frac{dp}{dt} = \tau x - \omega p$$

$$\frac{dz}{dt} = acx(e^{-g\nu} - z) - gz$$

where

$$E[e^{-\mu X}] = \frac{r}{r+\mu}\frac{\Phi(\mu/r;\mu/r+2;-\frac{Mabz}{rH})}{\Phi(\mu/r;\mu/r+1;-\frac{Mabz}{rH})}$$

The interpretation of this formulation is much more straightforward than the previous two versions - we no longer need information about the entire distribution. In particular, this form makes it clear that the distribution of MOI only affects the rate of recovery and it is therefore much more directly comparable to other Ross-Macdonald models.

1.5 Discussion

In this paper we introduced a model of malaria superinfection which properly accounts for treatment and demography, two epidemiologically-relevant features of malaria which influence the distribution of MOI and have nonlinear effects on the mean recovery rate. As the proportion of the human population with a single infection represents the only group which can recover naturally at a given time, these perturbations to the recovery rate can propagate through to differentially affect the modeled effect size of drugs on transmission. In particular, this effect will be highly context-dependent - a drug will have a stronger measured effect on prevalence in areas with a high force of infection than predicted in a model without superinfection.

In addition, temporary protection conferred by drugs can have a major impact on the prevalence, as shown in comparing figures 3 and 6. This "waiting room" for the infection queue has the effect of pulling down the equilibrium prevalence, and nonlinearly impacting the influence of FOI on prevalence. This has interesting and somewhat troubling implications for malaria control in areas with high treatment rates - reducing the effective FOI through any intervention appears to have the opposite intended effect on overall prevalence. However, it's important to be cognizant of the fact that this result is under the assumption that treatment clears all strains. This particularly strong assumption may be modified in future studies to investigate the propagation of drug resistance under a model with superinfection. The model also implies an issue with using equilibrium prevalence to estimate the force of infection; namely, that chemoprotection appears to force a non-monotonic (and therefore

non-invertible) relationship between the two, creating ambiguity in translating between them. However, we also see where and when this function curves down in higher treatment areas, and which regions of the parameter space we would likely be safe using this relationship.

Many modern agent based models (ABMs) in use today (for a review of many in use, see [28]) allow for superinfection with much more of the biological complexity intact. The analysis presented here is meant to be complementary to such models. Assuming uniform within host dynamics, the expected behavior of the prevalence under the Ross-Macdonald model should follow the differential equations derived here; therefore, this model can be used to validate the simulation. The impact of superinfection will be better captured by ABMs, but due to their complexity the relative contribution of superinfection and the mechanism by which it impacts prevalence is much more opaque. The analysis here provides insight into the mechanism, a functional form between the different parameters and prevalence, and improved understanding of the dynamic relationship between the states and parameters.

In particular the present model elucidates the subtle interaction between the complexity of infection and the rate of natural recovery. As a consequence of the assuming infections clear independently, the mean population recovery rate is dependent on the proportion of individuals with a single infection - therefore the distribution of MOI strongly influences the recovery rate. Excluding demographics and allowing for variable force of infection (FOI, here denoted λ) allows hybridization of the transmission system given the distribution is Poisson, and it was modeled elegantly by Nåsell with two relatively simple ODEs [8]. However, introducing demography and treatment requires a different approach as they can fundamentally alter the process generating the distribution of MOI.

While this model represents an advance and generalization of past results, it is important to keep in mind the underlying assumptions when interpreting the model predictions in real transmission systems. Because infection events are modeled as a simple Poisson process, an infectious bite can only result in a single infection. This allows the original M/M/∞ model to be "tridiagonal" in the sense that an individual can only increase or decrease the number of infections by one at a time. However in areas of high transmission intensity, a bite from a

superinfected mosquito may establish multiple infections from the same blood feeding event [29]. Parasite diversity can also contribute to transmission; for example, deletion of HRP-2, the target of common rapid diagnostic tests, can decrease the rate of detection and subsequent treatment of infections [30]. There is also evidence to suggest a host who becomes anemic due to periods of high parasitemia may be less prone to infection due to the efflux of iron from the liver [10]. All of these break Macdonald's original assumption of independence of the infections, and generalizations which account for these could have important implications for malaria control and elimination strategies. As a consequence of ignoring parasite densities we have assumed no impact of superinfection on an individual's gametocytemia, and therefore their infectiousness to mosquitoes. The relationship could be important for the feedback between humans and mosquitoes [29]; this can be explored in future studies by considering different functional responses between MOI and net infectiousness to mosquitoes.

Despite these simplifying assumptions, this model serves two main purposes. First, it provides a consistent framework with respect to the underlying assumptions of the stochastic process for investigating the impact of drugs and demographic turnover on the prevalence of any pathogen which can superinfect its host. Second, it provides an analytic demonstration of how the distribution of MOI under the effects of superinfection is inextricably connected to the structure of the rest of the model, as well as some tools for investigating the impact of this kind of heterogeneity using more easily measured parameters. This connection and the dynamics that it generates increases our understanding of the subtle factors involved in systems of pathogens which superinfect their host and has the potential to improve inference on the complexity of infection in a population. However, future work is needed to explore the more complex biological interactions at play.

Chapter 2

INFECTION AGE AS A PREDICTOR OF EPIDEMIOLOGICAL METRICS FOR MALARIA

2.1 Background

Despite great progress in recent decades, malaria from *Plasmodium falciparum* infection continues to claim approximately 435 thousand lives each year [31]. Deaths represent only part of the overall burden, as an estimated 194 million cases occurred in 2017 alone [31] despite the existence of cures and preventative measures. Malaria therefore represents a major source of avertible disease burden, with a major challenge being how to efficiently allocate resources to people in need. Assessments of potential targeted interventions to efficiently reduce prevalence depend on detailed knowledge of the epidemiology in the region of interest gained from local research studies.

Epidemiological surveys of malaria typically report some combination of asexual parasite counts, clinical incidence, prevalence, and rates of fever to compare age-specific patterns of disease and transmission among populations [32, 33]. However these multifaceted data paint a complex picture, with patterns in routine clinical surveillance data that are difficult to interpret due to ambiguity in the causes of observed trends. Therefore it is important to characterize patterns which appear in first infections in order to determine how they may be altered with different levels of past exposure. To this end, this study aims to establish time since infection as a predictor of patent infection probability by microscopy, asexual parasite densities, gametocyte densities, fever probabilities, and transmission efficiency. This relationship can be leveraged in future modelling studies which can track theoretical estimates of infection age given rates of exposure, and incorporate the effects of immunity to see how these baseline patterns in individuals may be altered to lead to the population level patterns

seen in data.

Previous research on the time course of infection has described the highly volatile trajectories of parasitaemia in a single individual over time, whose counts can jump orders of magnitude over the course of a day [34]. Studies focused on infection durations have estimated that the average time to the last observed patent infection ranges from around 100 days to over 1000 days [35, 36, 37], with at least one confirmed infection persisting for over a decade [23]. Patterns in the spikes and troughs of parasite counts have been investigated for evidence of mechanisms driving patterns such as VAR gene switching in parasite densities [37] or blood cell age preferences of the parasites [38]. Insight derived from these studies are valuable, but it is difficult to interpret these mechanisms' role in the efficacy of a particular intervention, or to scale to the impact on country or continent level estimates of burden. Asexual parasitaemia is a standard covariate for estimating the malaria attributable fraction of fever [32] and the heterogeneous relationship between gametocytaemia and transmission efficiency [39, 40] at the population level, but there remain lingering issues of identifiability regarding the impact of immunity.

Due in part to widely varying histories of past exposure among individuals in a population, observational studies of the relationship between quantities such as prevalence, malaria attributable fever, and per capita transmission rate may be misleading if they do not take into account the heterogeneous effects of immunity. Innate immune responses may differ in individuals depending on exposure, age, and other possible underlying conditions, and adaptive immune responses may differ based on an individual's history of past exposure. The effects of immunity limit an infection, reducing parasite population growth and eventually clearing infections. The impact of adaptive immunity on subsequent infections can further be decomposed into five categorical effects: pre-erythrocytic immunity, anti-(asexual) parasitic immunity, parasite tolerance, anti-gametocytic immunity, and transmission blocking immunity. Pre-erythrocytic immunity slows or prevents the establishment of an infection before or in the liver. Anti-parasitic immunity acts to reduce the blood stage parasitaemia, which is correlated with disease. Parasite tolerance modifies the relationship between para-

sitaemia and disease; it is measured as a reduction in the likelihood of fever and other clinical symptoms for a given parasitaemia. Anti-gametocytic immunity reduces gametocyte densities. Gametocytes infect mosquitoes, and their densities are correlated with transmission efficiency. Transmission blocking immunity reduces transmission efficiency for a given gametocytaemia, analogous to parasite tolerance [40]. This is mediated through immune effectors which target gametocyte-specific antigens, compounding the impact of antiparasite immunity which reduces the number of asexual parasites which produce gametocytes. Malaria epidemiology and immunity can therefore be understood as a set of cascading consequences of parasitaemia (Fig. 1). As these different modalities will impact fever, patency, and transmission rates, a direct translation from parasitaemia or prevalence to other epidemiological measures without previous exposure taken into account may be difficult to establish at best and lead to spurious patterns at worst.

Understanding how *P. falciparum* infections develop in the absence of acquired immunity is thus key to understanding and interpreting malaria data in which immunity has modified baseline patterns. Baseline data are difficult to obtain in malaria-endemic areas, but data describing some malaria infections in non-immune individuals is available from the historical data describing carefully monitored, deliberate malaria infections used to induce a fever to treat neurosyphilis, called malariatherapy [34, 41]. The malariatherapy data used in this study was initiated through either intravenous injection with asexual parasites directly or mosquito bite with sporozoites. They have previously been used to study the duration of single infections [42]. This is difficult to estimate from longitudinal studies due to the relatively common occurrence of superinfection, in which individuals are infected with multiple cohorts of parasites simultaneously, and unknown past exposure. Some recent studies have used genetic data to follow individual infections and estimate the multiplicity of infection [24] but such studies can only detect genetically distinct strains sporadically, and infections with the highest densities at the time of measurement mask the presence of lower density infections. Some malariatherapy patients were infected several times, and their records have been used to study the effects of adaptive immunity by comparing the difference between ho-

mologous and heterologous challenge; effects of immunity to homologous challenge appear to be present after one or two infections, but it may take more exposure for strain transcending immunity to occur [22].

In this study, focus was placed on population-averaged patterns in parasite densities over the time course of the infection in individuals with no prior exposure as a reference for uncomplicated malaria with no effects of previously acquired immunity in adults using a sample from the malariatherapy patient data. Statistical relationships were established between average parasitaemia and epidemiological measures such as fever rates and transmission efficiency in relation to recent exposure in immunologically naïve individuals (Fig. 1). Given an infection has not been cleared, it can either be patent or subpatent; given patency, individuals will have some measurable asexual parasitaemia. This parasitaemia is used as a measure of severity of disease, and therefore risk of symptoms such as fever. Asexual parasites also produce gametocytes after some maturation period, previously estimated to be 9-12 days [43, 44], which persist with a short half-life. In turn, gametocytaemia is used as a predictor for transmission efficiency [40]. This framework was used as a lens through which to statistically view the aggregate data, demonstrating that the age of an infection can be used as a potentially powerful surrogate for estimating these hard to measure and dynamic quantities.

2.2 Methods

The data set consists of 316 adult patient records, 258 males and 58 females of unreported ages, between 1941 and 1954 in the American south. The participants were patients with neurosyphilis being treated with malaria parasites, which were intended to induce a fever and effective immune response against spirochetes to mitigate outcomes of neurosyphilis. The infections were initiated through injection with either sporozoites to induce a liver infection or merozoites to directly induce blood stage infection. Results were aggregated across both inoculation methods, as previous results have shown little difference in measured quantities in this study between the two groups or by dose administered beyond the prepatent

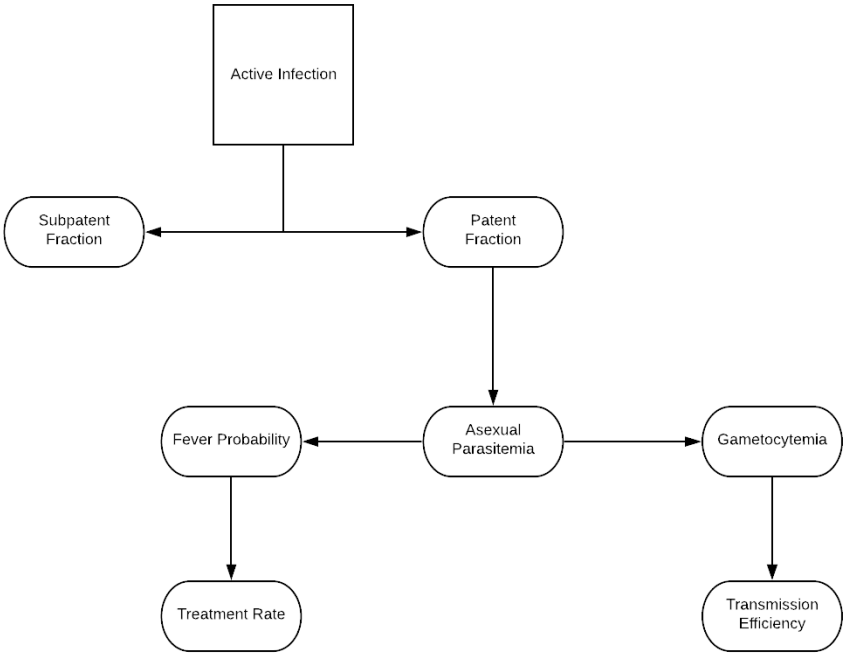


Figure 2.1: Schematic breakdown of infections. Given infections are active, some fraction will be subpatent (ie, undetectable by light microscopy) and the rest will be patent. Given the parasitaemia is patent, we can estimate the parasitaemia which will inform fever probability today and gametocytaemia roughly 9 days in the future. Fever rates presumably are correlated with treatment rates as symptomatic individuals are much more likely to seek treatment, and gametocytaemia is positively correlated with the transmission efficiency per mosquito bite.

period, which was not investigated here [45]. Infection age was counted from the first day of patency of an infection. Little clinical difference in outcomes aside from latency before first measurement was noticed, so both types of exposure were included here. Once patent by microscopy, daily measurements were taken of asexual and sexual stage parasites, and body temperature if there was an apparent fever. If symptoms of malaria became severe or if parasite densities were too high, patients were given treatment inadequate to cure malaria but sufficient to reduce parasitaemia. Treatment was not standardized and included a range of 28 different drug regimens. When there were detectable gametocytes in the blood, mosquito feedings were performed to determine the transmission efficiency from human to mosquito. Once the infection had been subpatent for some time, full treatment was given to clear the parasites entirely. In a subset of patients, secondary infections were initiated through either homologous or heterologous challenge. No deaths were reported.

For the estimate of duration of patent infection, we excluded treated cases (treated n=189). For all other estimation, we included infections that were treated until the day they were first treated, where they were truncated (n=299). On any day that an individual had patent gametocytaemia, a mosquito feeding was performed and after the estimated extrinsic incubation period the mosquitoes were dissected to determine the fraction which became sporozoite positive (n=2029 observations). The subsequent infection challenges that occurred in some patients were excluded, as we were interested in the course of first infections.

The primary features of interest included infection duration, patent fraction over time, asexual parasitaemia over time, gametocytaemia over time, fever risk associated with parasitaemia, and transmission efficiency associated with gametocytaemia. The fits of the relationships between all the observable quantities are summarized in table 1. For patency, asexual parasitaemia, and gametocytaemia, piecewise linear or generalized linear fits were performed and summarized. The distributions of asexual parasitaemia and gametocytaemia were also represented in violin plots aggregated by month to show general trends. Daily means and variances appeared to have a relationship, so a power law was fit. Logistic regressions were performed to translate daily average parasitaemia to fever risk, and smoothed

gametocytaemia to transmission efficiency. The degree of zero inflation in the transmission efficiency and the beta-fitted histograms of transmission efficiency across binned levels of gametocytaemia were also plotted to emphasize the overdispersion of the relationship.

2.3 Results

The average duration of the infection, restricted to the subset of patients who were untreated during the entire infection (n=110), was estimated to be 130 days (Fig. 2b). Here the duration was defined as the age of infection on the last day with patent asexual parasitaemia by microscopy, which was followed by a sequence of parasite negative observations and the cessation of measurement. We compared exponential, gamma, weibull, and lognormal survival curves, with delta AIC (23.3, 0, 1.7, 6.8, respectively) and delta BIC (20.6, 0, 1.7, 6.8, respectively) confirming exponential as a poorest fit and gamma and weibull as being the best candidates. The plotted blue curve is the gamma survival curve, with shape parameter 2.058 and rate parameter .015 per day.

For the analyses that follow all infections were included, in particular observations of all infections until they were ended by treatment. This allowed a reduction of the bias introduced by removing all infections with higher parasitaemia which tended to be treated at higher rates.

Asexual parasite densities followed a three-phase pattern (Fig. 2). For the first six days of a patent infection, parasite densities had increased geometrically. This was followed by a sharp decline between days 6-20. After that, here referred to as the chronic phase, parasite densities and the fraction patent on a given day slows to a shallower nearly linear trend in time. Among patent infections, the log10 transformed daily average declined linearly (Fig. 2a). In the chronic phase, infections sometimes had subpatent periods before spikes in parasitaemia occurred again. The proportion of peristent infections which were patent on a given day declined steadily (Fig. 2c). Older infections tend to spend a signficant fraction of time at submicroscopic densities in the blood with occasional bouts above that theshold of detectability, somewhere around 88 parasites per cmm of blood [46]. The increasing

variance in points around the fitted average was due in part to the sample size in the daily averages decreasing as individuals either receive treatment or recover. Additional variation had occurred to a smaller fraction of those with persisting infection remaining patent, so many of the later plotted points are representative of a small number of individuals with late spikes in parasitaemia.

A strong relationship between asexual parasitaemia and gametocytaemia became apparent. The log10 of the average parasitaemia and gametocytaemia appeared to be shifted and scaled versions of one another (Fig. 3a). Lagged average parasitaemia as a linear predictor of gametocytaemia was therefore explored. The optimal lag was determined to be 9 days, which minimized the standard deviation of the residuals (see the trough of Fig. 3c). The flat nature of the standard deviation as a function of the lag around days 8-12 suggests that other lags may be nearly as good of fits, which is consistent with the estimated maturation period of gametocytes [43]. Gametocyte densities in the chronic phase had also declined linearly (Fig 3b). After accounting for the lag, gametocyte densities were approximately 10-fold lower than parasite densities.

In addition to patterns in average asexual parasitaemia and gametocytaemia, patterns in the distribution across all individuals on a given day were investigated (Fig. 4). Monthly violin plots of the asexual parasitaemia and gametocytaemia after log-transformation appeared to maintain their shape while shifting down as infections aged. A power law relationship between mean and variance of both asexual parasitaemia and gametocytaemia for patent infections was quantified (Fig 4b,d). Power laws are often known to exist in higher density regions [47], although a decrease in variance may occur in measurements near the threshold of detectability by light microscopy as lower measurements are likely to be recorded as subpatent and therefore not included in daily measurements.

Log10 daily measurements of parasitaemia were then fit as a predictor of the proportion of individuals with fever, conditioned on patent parasitaemia. The best fit was a sigmoid function from the GLM with logit link (Fig. 5). There was a different relationship between parasite densities and fever in the first five days of patency compared to the rest of the

infection. Note the five purple points laying above the sigmoid, with the leftmost two in particular being strong outliers. These points represent the first five days of infection, with the days being ordered from left to right. This would imply parasitaemia is a poor predictor in the first few days of infection, and in particular fever may come days before high parasitaemia. Unsurprisingly, fever was also a function of the age of the infection. The fit appears to follow the data very well, though underestimates the first few days of infection as expected.

Finally the relationship between gametocytaemia and transmission efficiency was quantified. Transmission efficiency was measured as the fraction of mosquitoes that developed sporozoites after feeding on individuals with patent gametocytaemia. The relationship between gametocytaemia and transmission efficiency was consistent but noisy. To fit the data, the data generating process was modeled as a mixture process, a zero-inflated beta-binomial distribution. Smoothing individual measurements across log10 gametocytaemia by averaging over measurements with similar gametocytaemia resulted in the blue points in Fig. 6a. A logistic regression was performed on the blue points, with weights proportional to the number of measurements used in the average. Analogous to Fig. 5b, Fig. 6b shows the log linear fit of gametocytaemia filtered through the sigmoid in green.

The beta-binomial interpretation had allowed for quantification of overdispersion, with zero-inflation added due to the large abundance of zeroes. Interestingly, the amount of zero inflation had appeared to decrease with increasing gametocytaemia, as shown in Fig. 6c. Conditioning on a nonzero number of mosquitoes counted, the histograms were plotted in binned gametocytemia with beta distributions fit in Fig. 6d-i. As expected, increasing gametocytaemia had shifted likelihood to higher transmission efficiency but with a large amount of variability. Therefore even highly gametocytemic patients often infected less than expected mosquitoes, or even none at all, during a particular feeding. This could have been partially explained by a result of fairly small samples of mosquitoes feeding successfully per patient per day, though another mechanism cannot be ruled out such as differences in sex ratios of gametocytes or strain-specific differences that are accentuated in transmission more than disease states [48].

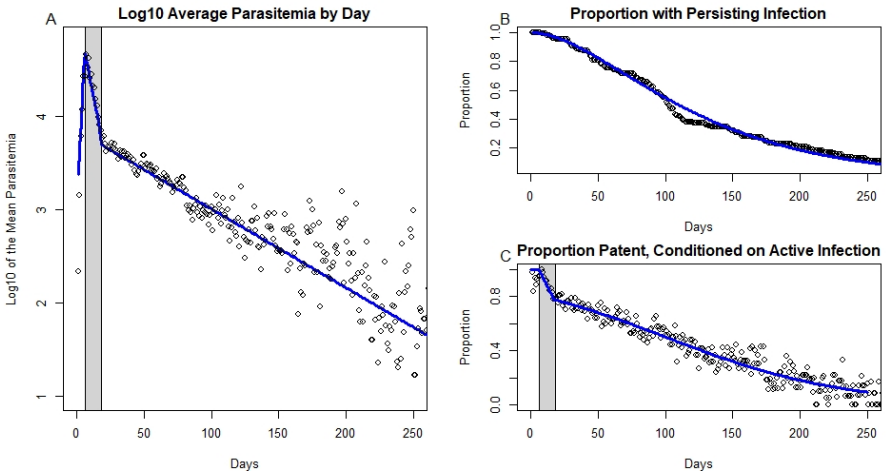


Figure 2.2: Plots of parasitaemia, duration, and patency. Panel A shows a scatter plot of the log10 of the daily average parastemia among patent infections conditioned on continued infection. The blue lines are a three piecewise linear fit in three parts, with a gray shaded region representing the middle of the three. Panel B shows a plot of the empirical survival function of infections that had positive measures of parasitaemia on or after that day of infection. The blue curve represents the best fit gamma survival curve, that is the complement of the corresponding gamma CDF. Finally panel C is a plot of the proportion patent conditioned on continued infection. The gray shaded region is the same highlighted in panel A, showing that during the initial growth phase nearly all infections remain patent; then patency drops about 20 percent over a short time, then it slowly decays to around 10 percent by day 250.

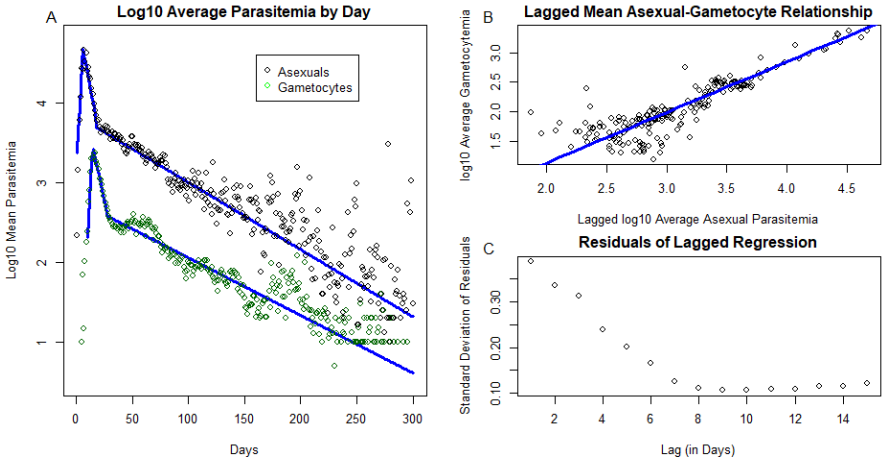


Figure 2.3: Panel A shows two time series of points, the top representing the log10 of the daily mean parasitaemia on a given day across all continued patent infections as in Fig. 2 panel A and the bottom green time series represents the corresponding log10 daily mean gametocytaemia. A fit to predict gametocytaemia from lagged parasitaemia was performed across many lags, and the optimal lag was determined by the minimum of the standard deviation of the lagged residuals shown in panel C. The linear fit at the optimal lag of 9 days in presented in panel B.

2.4 Discussion

It was shown that asexual parasite densities were strongly predicted by the age of infection, the relationship between parasite densities and fever and gametocytaemia and the relationship between gametocytaemia and infectiousness were quantified. In particular, estimation of the expected asexual parasitaemia was determined as a function of infection age, from which every other quantity can be estimated. The diagram in Fig. 1 shows pathways through intermediate quantities to translate from infection age to outcomes of interest. As there is also have a relationship between the mean and variance of asexual parasitaemia and gametocytaemia, fitting a family of distributions allows for estimation of full distributions of fever rates and transmission efficiencies as well. Therefore this represents a potentially powerful framework for future estimation.

This analysis suggests it may be possible to translate knowledge of a history of exposure

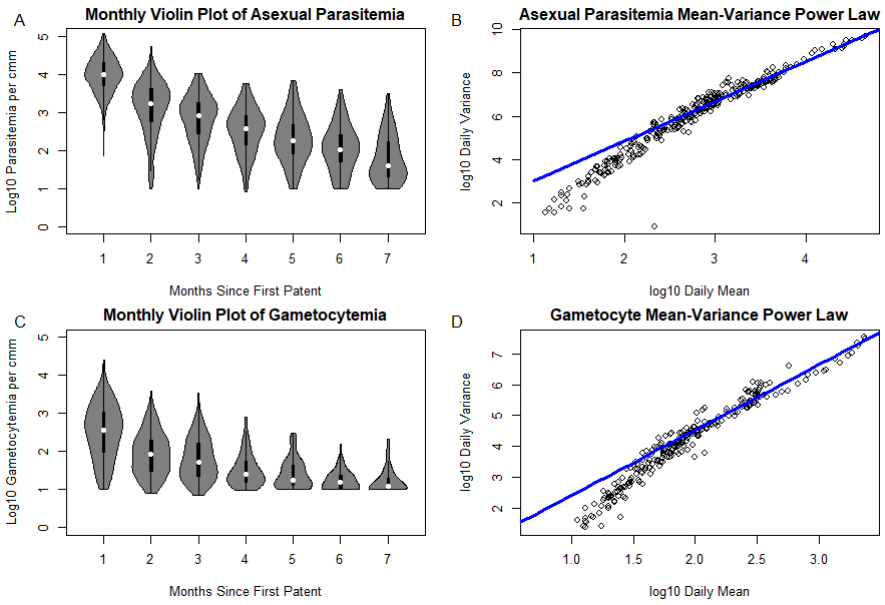


Figure 2.4: Panel A and C represent violin plots of the distributions across individuals of respectively the log10 daily parasitaemia and log10 daily gametocytaemia across all individuals, aggregated across 30 day periods for compactness. Despite being log transformed, the distribution shapes appear to shift down with infection age but maintain the same general shape, suggesting a power-law relationship between the mean and variance of parasitaemia. Power laws are then fit to log10 transformed daily mean and variances in panels B and D, restricted to those above the estimated sensitivity of light microscopy (88 parasites per cmm blood).

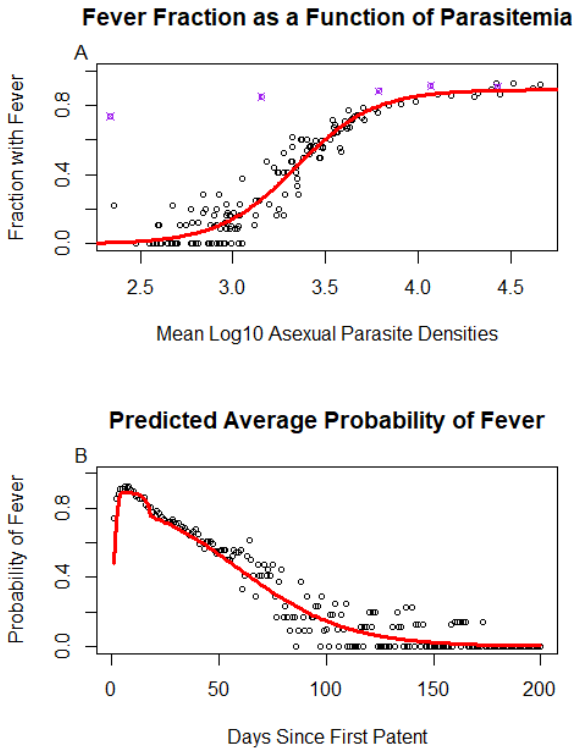


Figure 2.5: Panel A shows a logistic fit of the log10 of daily mean asexual parasitaemia to the fraction of individuals with objective fever. The five points in purple above the logistic curve are the first 5 days of infection, suggesting that many febrile individuals got primary fevers before high density infections. Panel B shows a time series of the daily fraction of individuals with active infections who have fever. The red curve represents the transformed piecewise fit of parasitaemia transformed through the fitted logistic curve.

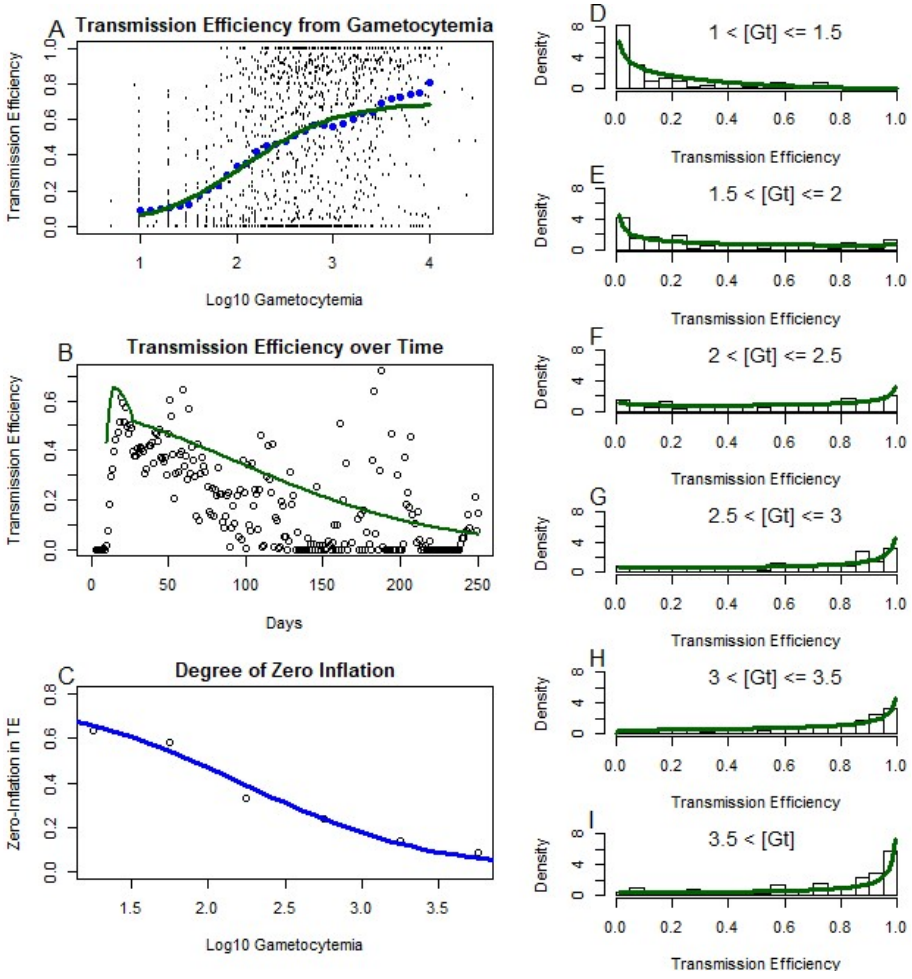


Figure 2.6: The black points in panel A represent the transmission efficiency, measured through the fraction of fed mosquitoes who developed sporozoites, against the log10 gametocytaemia measured on that day for each individual on every day they were gametocyte positive. Treated infections were not included in this analysis. Blue points are a rolling average of transmission efficiency as a function of log10 gametocytaemia, and the green curve is a sigmoid curve fit to these points with weights given by the number of points included in their rolling average. Panel B shows the daily average transmission efficency as a function of infection age, with the green curve representing the measured log10 daily average gametocytaemia composed with the sigmoid fit in panel A, showing good qualitative agreement. As the relationship between gametocytaemia and transmission efficiency is highly heterogeneous and zero-inflated, zero-inflated beta distributions were fit to binned values of gametocytaemia to quantify this. Panel C shows the degree of zero inflation, that is the fraction of mosquito feedings resulting in no infections at all, as a function of log10 gametocytaemia. The blue sigmoid curve was fit to these points, showing the apparent zero-inflation decreases with increasing gametocytaemia. Finally, panels D-I are histograms of transmission efficiency for a given range of gametocytaemia conditioned on nonzero measurements, with beta distributions fit to each. Despite the high degree of heterogeneity, density can be seen to aggregate to the right with increasing gametocytaemia.

to epidemiologically important quantities which are strongly correlated to the age of the infection if acquired immunity can also be estimated. Averaging these conditional rates on the probability an individual has had an infection for some duration across all present infection ages weighted by the fraction of the population who has had an infection for that duration gives expected population-level metrics in a given transmission setting. The basic idea follows the law of the unconscious statistician. If for example one were to estimate some observable X which depends on the concentration of the pathogen p, the following could be computed:

$$E[X_p] = \int_0^\infty P(p = p_0)E[X_{p_0}|p_0]dp_0$$

Examples of such X would include fever rates and transmission efficiency, as each would reasonably increase on average with pathogen density. However, this would require knowledge of the probability that an individual in the population has a pathogen concentration of p. This is rarely known and difficult to measure directly, as surveys often only measure incident cases which tend to have higher parasitaemia, and detection rates themselves depend on pathogen densities. However due to the observation that malaria appears to exhibit strong average parasite density patterns with respect to the age of an infection, it would be possible to parameterize p (and therefore X) through the infection age α. This would give

$$E[X_{p(\alpha)}] = \int_0^\infty P(\alpha = \alpha_0)E[X_{p(\alpha_0)}|\alpha_0]d\alpha_0$$

One advantage to this formulation would be to allow one to estimate the probability of having an infection of age α through the use of standard age of infection dynamics given estimates of exposure [49]. The resulting distribution could give a reasonable estimate on the unknown probability given patterns in rates of exposure, allowing one to compute the desired expectation.

Past work on malaria attributable fever [32], which estimated malaria attributable fraction of total fever based on parasitological survey data of children, is analogous to the

parasitaemia-to-fever risk regression done here. This interpretation would allow for the logistic regression to assign a probability to any child with parasite density measurements and a fever to determine how likely it is that the fever they have is attributable to malaria. The data was restricted to relatively young children, so the effects of adaptive immunity could be largely ignored as they were here. However to apply those results to an entire population, the regression would need to be reworked across measurements of all age groups as their past exposure and developed immunity will modify parasite densities and fever tolerance. Additionally, exposure history may vary dramatically from location to location, so the method would require an enormous amount of data and each location with its unique history of exposure could have a very different estimate from even other locations with the same current day prevalence. The analysis which was presented here shows that if one is instead able to estimate how long ago individuals were most recently infected based on a history of exposure from historical prevalence data, it would be possible to estimate a malaria attributable fraction of fever in the absence of parasite density surveys given a reasonable model of immunity.

Although parasitaemia appears to be a very good predictor of fever after the first week or so of patent parasitaemia, the fever rate was seen to be consistently higher than predicted by parasitaemia in the early days of patency. This would suggest the notion of a difference between primary fever, caused at the beginning of an infection, and a secondary fever, correlated strongly with parasitaemia and occurring later in the infection. This could be a consequence of the inflammatory cascade early in infection which is subsequently tempered by anti-inflammatory responses as the immune response has matured. An uncontrolled early inflammatory response is often found in severe cases of malaria [50, **?**], so mortality in cases may be closely correlated with this early stage of infection.

Analogous to the fever and asexual parasitaemia relationship, the infection reservoir and its impact on estimation of the human-to-mosquito transmission potential in environments with seasonal transmission may be largely impacted by the additional heterogeneity presented here. In addition to the overdispersion shown in the translation from gametocytaemia to

transmission efficiency, the infection-age dependent patterns of infectivity may be leveraged to improve our understanding of which locations may be a "source" or "sink" for malaria transmission, and how a location may switch from one to another based on the history of recent exposure and the age group of individuals in question [51].

The fit for transmission efficiency as a function of gametocyte densities had appeared to be consistently above the data. This was due to a technical difference in the fitting procedure compared to fever. While fever was predicted from daily average parasitaemia, transmission efficiency was predicted from a function of the gametocytaemia measurements and not to the averaged time series data directly. Combined with the fact that most of the gametocytaemia measurements in the time series are from the top half the fitted sigmoid and therefore filtered through a concave function, Jensen's inequality would guarantee that plugging in the mean to the function rather than taking an average of the data filtered through the sigmoid is expected to be an overestimate. However either through computing the conditional expectation mentioned above or simulating draws of gametocytaemia on a given day then converting through the sigmoid function, this problem would be avoided completely.

The power law relationships demonstrated between the mean and the variance, which suggests evidence of the well-known Taylor's Law from ecology [52], could be used for practical computations here. This relationship is often seen between the means and variances of populations across different spatial regions, but would apply here as each human host can be imagined as an independent habitat for the parasite populations. One of the significant advantages of it is it would allow one to use the relatively simple pattern in mean parasitaemia over time to obtain a similar pattern in variance over time, and therefore parametrically could describe a wide class of two-parameter distributions for parasitaemia as a function of infection age through moment matching. This allows for the propagation of uncertainty of estimates through the relationships in a way which would circumvent the issues of Jensen's inequality mentioned above.

It is necessary to highlight several limitations to the extensibility of these observations.

The atypical immunological states of the patients considered here were a clear concern. Naturally questions could be asked about the application of trends found in adults with neurosyphilis to otherwise healthy individuals in endemic settings. Further, all subjects were adults with presumably fully developed immune systems, and therefore their response may differ from children in endemic settings as well. However, none of these individuals have previously had exposure to malaria and therefore represent a sort of baseline for trends in first exposure, even if the exact values of the parameter values are not perfectly representative.

A conscious decision was made here to not limit the analysis to a single strain of *P. falciparum*. Strain-specific differences may play a large role in overall transmission dynamics [37], but often genetic information of strain diversity is limited in a particular setting. Additionally, no simple mapping between strain and pathogenicity or specific parasitaemia profiles is known to exist. Inflammatory signalling would also change in response to recent exposure [50], possibly altering the baseline relationship between fever and infections. Often in endemic settings individuals will have multiple infections simultaneously, so any mapping would also need to account for pairwise interactions or work on an assumption of independence. Handling the possibility of superinfection on this age of infection relationship would warrant further investigation in the future.

The five modalities of immunity as well as any age dependent trends will vary over time and impact all of the statistics presented here. With exposure, parasitaemia and gametocytaemia will decline; separate from that, higher parasitaemias can be tolerated before a fever develops, and transmission given a set gametocytaemia may decline. For these reasons, a static mapping from prevalence to fever rates or transmitting fraction in the absence of information on the history of exposure may be a poor representation of the epidemiological reality. Patterns of exposure (seasonality, source/sink dynamics, human travel patterns, etc) should play a large role in developing a dynamic mapping, which coincides with the understanding that malaria is a very heterogeneous disease by location. Given an understanding of the patterns of exposure, it may be possible to estimate the likely distribution of immune states in the population and take that into account for estimation of quantities of interest.

In light of these statistical relationships, if one knows an infection age distribution one can obtain estimates of fever rates and transmission potential in the absence of immunity. If a direct measure of these quantities is available, this may be able to act as a counterfactual for measuring the impact of immunity; if a model of immunity is included, one can obtain estimates of the quantities. In both cases, an infection age distribution is a crucial piece. Therefore, it places particular emphasis on the importance in determining such infection age distributions. Subsequent work is aimed to provide a model-based approach for constructing reasonable families of distributions of the age of infection given an exposure history.

Difficulty interpreting data arises in part from the extreme range of unknown previous exposure history across locations. Exposure has been measured at levels varying from no bites to more than a thousand bites by infectious mosquitoes, per person, per year; transmission efficiency, detection, and clinical manifestations of malaria depend on previous exposure and acquired immunity; exposure is often seasonal, and highly heterogeneous across individuals; and acquired immunity to malaria develops slowly, varies by exposure, protects poorly, and has poor memory [53, 22]. This prompts many studies to focus on the prevalence and outcomes in children, who can be reasonably assumed to have had little previous exposure [54]. Often patterns in children appear to more closely match trends in exposure intensity, but this leads to an identifiability problem: differences may arise due to the acquisition of immunity directly, or due to differences in the developing immune systems of children. Observed age-dependent patterns are likely a combination of both of these effects. Integrating the statistical patterns demonstrated here with mechanistic models of immunity may help to address this.

2.5 Conclusions

Patterns relating the age of malaria infections to patency, parasitaemia, gametocytaemia, fever rates, and transmission efficiency were quantified and described. These patterns can be leveraged in population-level models of disease transmission to obtain dynamic estimates of each of these quantities. Future investigations can use these models to determine the impact

of public health interventions on fever and transmission.

Table 2.1: Fitted Relationships between Infection Age and Quantities of Interest

Quantity of Interest	**Fitted Equation of Kernel**
Patent Fraction, D(α)	$\begin{cases} 1 & \text{if } \alpha \leq 6 \\ 1.12 - .02\ \alpha & \text{if } 6 < \alpha \leq 18 \\ (1 + e^{-1.52+0.0151\ \alpha})^{-1} & \text{if } \alpha > 18 \end{cases}$
$\log_{10}$ Asexual parasitaemia, P(α)	$\begin{cases} 0, & \text{if } \alpha \leq 0 \\ 3.10 + .278\ \alpha, & \text{if } 0 < \alpha \leq 6 \\ 5.12 - .0743\ \alpha, & \text{if } 6 < \alpha \leq 18 \\ 3.85 - .00843\ \alpha, & \text{if } \alpha > 18 \end{cases}$
Fever Probability, F(P(α))	$\frac{.859\ e^{3.45\ P(\alpha)}}{58200 + e^{3.45\ P(\alpha)}}$
$\log_{10}$ gametocytaemia, G($L_9P(\alpha)$)	$-0.684 + .892\ L_9P(\alpha), \quad L_9P(\alpha) = P(\alpha - 9)$
Transmission Efficiency, c(G(α))	$\frac{.683\ e^{2.14\ G(\alpha)}}{131 + e^{2.14\ G(\alpha)}}$

Patent fraction consisted of two piecewise-linear fits and a generalized linear fit at each of the transitions mentioned in the text. Log parasitaemia likewise is piecewise linear. Fever probability and transmission efficiency are logistic functions of their predictors, which are linear functions of infection age.

Chapter 3

MEMORY AND MALARIA: A PROBABILISTIC APPROACH TO MALARIA INFECTION AND IMMUNITY

3.1 Background

The epidemiology of *Plasmodium falciparum* malaria presents a unique set of challenges due to the complex dynamics of infection and immunity [55]: infections are long and variable, so parasite densities, detection, disease and infectiousness vary with the age of infection (AoI) [56]; acquired and innate immunity modify infection and the risk and severity of disease; but acquired immunity is weak, has a poor memory, and accumulates slowly with age and exposure. Immunity does not prevent reinfection or superinfection, so malaria infections can be complex – the number of distinct parasite lineages is called the multiplitiy of infection (MoI). Further, a person can experience multiple bouts of malaria (*i.e.* disease) over a lifetime. Malaria immunity varies with the age, intensity of exposure, and the development of immunity with cumulative exposure, over a lifetime.

Many complex features of malaria epidemiology are made more coherent when seen as emergent properties of the biological needs of the malaria parasite. Sexual reproduction occurs in the mosquito, and outcrossing occurs only if the multiplicity of infection (MoI) is greater than one. Frequent outcrossing, in turn, supports a genetically diverse parasite population, which is partly responsible for the features of malaria epidemiology. One challenge for malaria epidemiology is how to predict, given an individual's history of exposure and infection, the likely outcomes of the next infection. Framed broadly, outcomes are affected by host infection status (*e.g.* parasite densities), acquired immunity resulting from the history of exposure, other host factors (anemia, genetics, nutrition, *etc.*), parasite factors

(genotype and phenotype), and case management with anti-malarial drugs. The difficulty of understanding and quantifying outcomes is increased by the challenge of simultaneously measuring exposure, infection, disease, and immunity (*i.e.* through serology) in enough settings to characterize the relationships. This has been done only rarely, so the contemporary view of malaria epidemiology has been pieced together from a heterogeneous set of studies conducted over roughly 140 years. A synthesis has been attempted through development of mathematical models of malaria infection and immunity, but most of these studies have focused on a single dimension of the complexity [57]. Individual-based simulation models have managed to replicate the basic patterns in models that simultaneously deal with the complex time course of infections, superinfection, and waxing and waning immunity [28]. While IBMs are a useful computational tool for many purposes, they are often presented as black boxes due to the complexity of their design and the probabilistic nature of their output. Here, we present a simpler and more tractable probabilistic approach to modeling exposure, infection, immunity, parasite densities, detection, fever, and infectiousness in a cohort of humans as it ages.

The first challenge for modeling malaria is the long, variable time course of infection. After the infective bite and a short period in the liver, parasites begin a cycle of asexual replication in red blood cells, all while their numbers are increasing geometrically. Immunity to malaria contributes to inhibit parasite replication, after which the infection begins a variable chronic phase and the eventual end of the infection. Untreated malaria infections can, nevertheless, persist for a few weeks to several years. The time course of malaria infections is highly variable over the time course of an infection, and the features of those infections vary over a lifetime. After parasite densities peak, parasite densities and the probability of detecting parasites tends to decline, so accurate measurement of the distribution of the duration of infections has been difficult. Parasite densities in blood fluctuate over several orders of magnitude, and only a fraction of blood is examined. The sensitivity of parasite counts or other estimates of parasite density and malaria prevalence varies across diagnostic methods. Available data thus measure the duration of patency, but evidence for very long

carriage times comes from infections that have recrudesced after persisting for decades. In the simple infections observed in malaria therapy patients *i.e.* MoI = 1), most aspects of infections were well described statistically by the AoI, the time elapsed since the infective bite for a parasite cohort: the proportion still active, the proportion still detectable, the mean and variance of logged parasite densities, the risk of fever and disease, and infectiousness (see chapter 2). In the following, we thus develop models around the AoI.

A second challenge is superinfection. Acquired immunity to malaria does not provide complete protection to infection after subsequent exposure, so multiple, distinct parasite lineages accumulate within a single host, a phenomenon called super-infection. It is also possible for a single infective bite to transmit multiple distinct parasite clones. A statistic to measure superinfection is the MoI. Understanding parasite density distributions in complex infections (*i.e.*, those with MoI>1) presents a daunting challenge, in part, because of the detection problem. With highly variable infections, high-density infections on sampling can mask low density, and the complexity of infections may be partially revealed by sampling the same person repeatedly over several days.

A third challenge is immunity. The effects of acquired malaria immunity accumulate with exposure over a lifetime. In malaria-endemic areas, even the first infection in life can be modified by immunity, as protection conferred by maternal antibodies at birth wanes with age. Malaria outcomes are modified by immunity in several ways and at multiple distinct temporal scales, depending partly on the patterns of exposure [55]. The most basic manifestation of immunity, which we have already discussed, is that parasite densities and the duration of infections are limited. The features of each successive infection are modified by immunity, which reflects the whole history of exposure. The effects of immunity include reductions in parasite densities, reductions in the risk and severity of disease, and reductions in transmission [54, 58]. These effects manifest as cohorts age, so some of these effects could be confused with other effects of age: for example, exposure rates and blood volume increase with body size. When exposure is highly variable, either because of sporadic outbreaks or periods of intermittent control, acquired protective immune responses can wane. After a

long period of exposure, the effects of immunity are apparently dulled, possibly because of exposure to parasites that are genetically different from those previously seen.

Fourth, the patterns of exposure vary enormously over space and time because of mosquito ecology, malaria control, and a large set of endogenous and exogenous factors driving mosquito populations and malaria transmission. Mosquito population ecology is mainly responsible for geographical variability in the overall patterns of exposure as well as differences in the seasonality and inter-annual variability. The resulting differences in exposure affect development of acquired immunity in response to exposure as well as the degree of protection passed on to infants by their mothers.

Given all these factors, the challenge for malaria is to translate exposure, infection, and immunity into relevant and observable outcomes, including models for parasite detection and parasite counts, models for infectiousness, fever and disease, and the consequences of curing infections after treatment with anti-malarial drugs. Parasite densities and detection, fever and disease, and infectiousness all depend on host age and the history of exposure. Malaria is understood mainly through observational epidemiology using a diverse set of metrics: the entomological inoculation rate, the force of infection (FoI), parasite counts, malaria prevalence (by various metrics), and serology are intermediate variables to study subjective and objective fever, severe disease, anemia, and infectiousness.

Because of the complexity, summarized above, a quantitative synthesis of malaria epidemiology must be developed around some sort of mathematical model. A practical challenge, from the mathematical perspective, is how to represent the multi-dimensional state space for infection and immunity. The memory problems has led to a proliferation of models that address various aspects of malaria, including simple models exploring various aspects of transmission [57], and comprehensive and synthetic studies using individual-based simulation models of malaria infections [59, 60, 37]. While these models have the advantage of being flexible and extensible, they require intensive computation and the models are difficult to calibrate, to understand, and to evaluate critically. What is needed is a parsimonious synthesis.

$$\zeta_0 \underset{r}{\overset{h}{\rightleftarrows}} \zeta_1 \underset{2r}{\overset{h}{\rightleftarrows}} \zeta_2 \underset{3r}{\overset{h}{\rightleftarrows}} \zeta_3 \underset{4r}{\overset{h}{\rightleftarrows}} \cdots$$

$$\begin{aligned} d\zeta_0/dt &= -h\zeta_0 + r\zeta_1 \\ d\zeta_i/dt &= -(h+r)\zeta_i + h\zeta_{i-1} + r(i+1)\zeta_{i+1} \end{aligned}$$

Figure 3.1: A diagram of the queuing model $M/M/\infty$ (top), and the master equations (bottom), in which ζ_i denotes the proportion of the population with MoI $= i$. MoI increments by one if an infection occurs at rate h and decrements by one if an infection clears at rate ri.

Here, we present a new probabilistic modeling framework for malaria that uses the AoI as a unifying concept. The approach links exposure to the distributions of the MoI, the AoI, and to variables describing cumulative exposure that are used as a basis for modeling the effects of immunity. We translate the epidemiological state space into a model for parasite densities and detection in complex infections, to gametocyte densities and infectiousness, to fever and disease, modification by immunity, and treatment and chemoprotection. We have fit one particular version of this model to the malaria therapy data, but the approach can be extended to any model of parasites that is complex enough to describe parasite densities over the time-course of an infection. The probabilistic approach that tracks the infection status of human birth cohorts as they age. The probabilistic approach provides a highly rigorous set of formulas and computational algorithms to serve as an *in silico* laboratory. Using this appraoch, we show that malaria epidemiology, including parasite densities, can be understood in terms of the age of the youngest infection (AoY), and using the AoY, we present a synthesis of observational malaria epidemiology.

3.2 A Probabilistic Cohort Model

Here we model malaria epidemiology in a host cohort as it ages by extending analysis of a queuing model for super-infection called $M/M/\infty$ (Fig 3.1), which has played an important role in malaria epidemiology [7, 6, 61]. A core assumption of the model is that infections clear independently, and in the spirit of that model, we also assume that parasite densities

over the time courses of two or more parasite infections are also independent. To model parasite densities, have derived formulas describing the distribution of the AoI, and we use the co-distribution of the AoI and the MoI as a basis for understanding malaria parasite densities and outcomes in observational studies.

In the following, we let a denote the age of a host birth cohort as it ages over time. Similarly, let α denote the AoI: the age of a parasite cohort, or the time since the infective bite that initiated an infection. The parasite cohort includes all the progeny of the parasites in a single infective mosquito bite.

We derive formulas describing the full distribution of the MoI, the AoI, and true prevalence contingent upon exposure, which we model using the FoI. We extend the model by including a variable to track cumulative exposure, and we present a family of models for parasite densities as a function of the AoI and cumulative exposure. From these we derive formulas that describe the full distribution of parasite densities. Finally, we filter these through a model of detection to model parasite density distributions, including the estimated prevalence, estimated MoI, and the mean $\log_{10}$ parasite counts, and other outcomes (see Fig 3.2).

3.2.1 Exposure and Infection

Differences in the rate of exposure by age and season are modelled as a time-dependent FoI, defined as the time-dependent average hazard rate for new infections in the population. To make the analysis as general and applicable as possible, we have adopted a trace-driven simulation approach: the FoI was passed to algorithms as a function $h(t)$ (Supplement 1). In doing so, we are explicitly reducing the FoI to a model input, though in many studies with mechanistic models, the FoI is a term of interest, computed as the result of a dynamical process reflecting parasite transmission from humans to mosquitoes, and blood feeding by mosquitoes. Such trace-driven simulations are one way to decouple different parts of a simulation when one aspect of a simulation can be replaced by an external input (trace) from other parts of the simulation on which it depends [62]. By using trace-driven simulations,

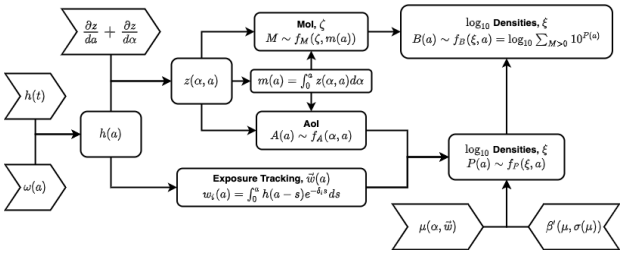


Figure 3.2: An overview of the probabilistic cohort model. We pass the daily FoI in the population, $h(t)$, and an age weight, $\omega(a)$, and we model exposure as an age-specific, daily FoI for a particular cohort as it ages, denoted $h(a)$. With a model for parasite clearance, we compute density of all parasite cohorts of age α, distributed among a cohort of hosts as it ages, $z(\alpha, a)$, and from this, we compute random variables describing the distribution of the MoI (M with mean $m(a)$) and the distribution of the AoI ($A \sim f_A(\alpha, a)$). Using variables that track cumulative and recent exposure in the cohort, $\vec{w}$, we formulate a random variable describing ξ, mean $\log_{10}$ parasite densities, as a function of the AoI and exposure, $\mu(\alpha, \vec{w})$. We model the variance in ξ as a function of the mean, $\sigma(\mu)$ and we modify a $\beta-$distribution, denoted β', to give properly bounded values for ξ. We can thus develop a random variable describing $\log_{10}$ parasite densities among the cohorts, P, and for parasite densities in complex infections, B.

these models can be used to address questions about malaria under almost any pattern of exposure without the need to specify exactly how that pattern of exposure might have arisen.

The Force of Infection We developed a simple but flexible set of functions to facilitate construction of "realistic" functions to model exposure (Fig 3.3, Supplement 1). We let $\bar{h}$ be a parameter that sets the overall level of exposure. We assume exposure has a seasonal pattern, $S(t)$, and a temporal trend, $T(t)$: $h(t) = \bar{h} \cdot S(t) \cdot T(t)$.

Motivated by the evidence for age-dependent exposure, the rate of exposure in the cohort is assumed to change with age. We use the age-weighting function, $\omega(a)$, consistent with the available evidence. Each cohort is indexed by its birth day, τ. The FoI at age a for the cohort born on day τ, which is at time $t = a + \tau$, is:

$$h_\tau(a) = \bar{h} \cdot \omega(a) \cdot S(a + \tau) \cdot T(a + \tau). \tag{3.1}$$

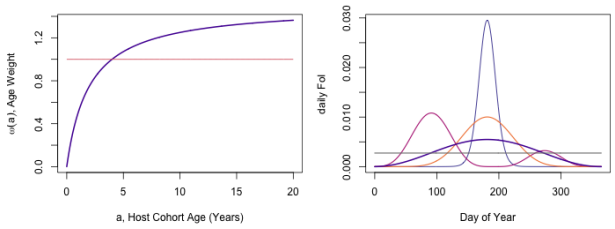


Figure 3.3: FoI trace functions developed to simulate exposure combine four elements: 1) a parameter that sets the overall intensity, $\bar{h}$; 2) functions that modify exposure by age, $\omega(a)$; 3) functions that describe the seasonal pattern, $S(t)$; and 4) and a long-term trend, $T(t)$. The left panel illustrates representative functions for age-modification, and the right panel shows representative seasonal patterns.

We can thus compare exposure patterns for cohorts of different ages, of the same age at different times. Supplement 1 includes a highly configurable and extensible library of FoI functions.

Parasite Cohort Dynamics Let $z(\alpha, a)$ denote the density of parasite cohorts of age α in a host cohort of age a. We assume infections clear at the constant rate r. Since infections in $M/M/\infty$ are independent, we can track the dynamics for the AoI of all parasite cohorts with the equation,

$$\frac{\partial z}{\partial a} + \frac{\partial z}{\partial \alpha} = -rz, \tag{3.2}$$

with the boundary condition $z_\tau(a, 0) = h_\tau(a)$. We note that the age of the host birth cohort sets an upper limit for the parasite cohort, so $0 < \alpha < a$. The solution to Eq. 3.2, which describes density of infections of age α in a host cohort of age a, is given by the formula:

$$z_\tau(\alpha, a) = h_\tau\left(a - \alpha\right) e^{-r\alpha}. \tag{3.3}$$

Under the assumptions of $M/M/\infty$, all the parasite in the infection cohorts are distributed independently among hosts.

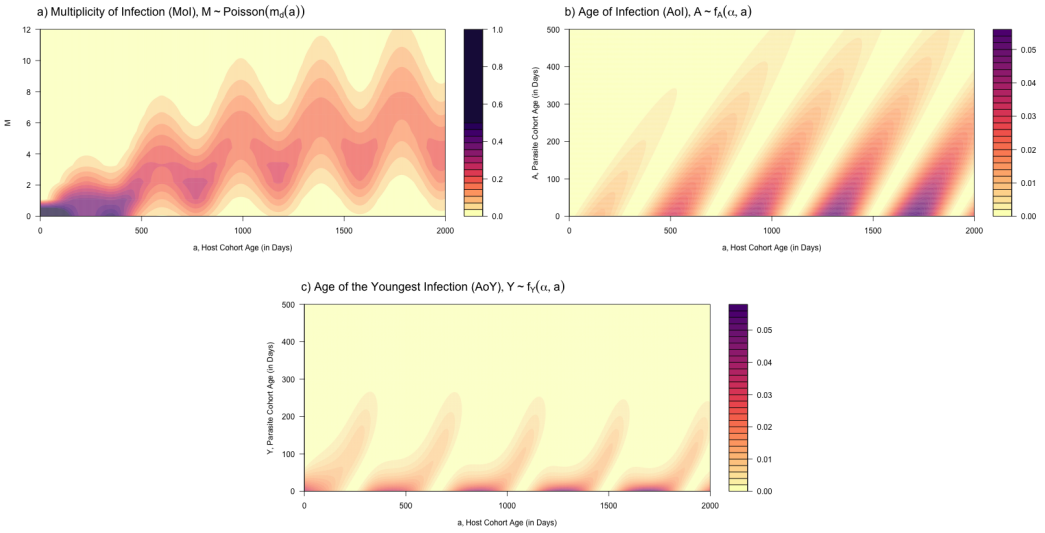


Figure 3.4: The distribution of the a) multiplicity of infection (MoI), b) the age of infection (AoI), and c) age of the youngest infection (AoY) in a cohort as it ages from birth up through 5 years old. The FoI was sinusoidal, with a population mean aFoI of 5 infections, per year.

Age and Multiplicity of Infection Importantly, the mean MoI in the cohort at age a is the integral of this density function:

$$m_\tau(a) = \int_0^a z_\tau(\alpha, a) d\alpha \tag{3.4}$$

Elsewhere, it has been shown that the distribution of the MoI converges to a Poisson [8]. Since the MoI is zero at birth (a degenerate Poisson distribution), the distribution of the MoI remains Poisson as the cohort ages (in this model), and mean MoI, $m_\tau(a)$ completely describes the distribution of MoI:

$$M_\tau(a) \sim f_M\left(\zeta; m_\tau\left(a\right)\right) = \text{dpois}(\zeta, m_\tau(a)) \tag{3.5}$$

Since f_M is Poisson, true prevalence (p) is the proportion with MoI greater than 0, which can be computed directly from $M_\tau(a)$:

$$p_\tau(a) = 1 - e^{-m_\tau(a)} \tag{3.6}$$

While $z_\tau(\alpha, a)$ describes the density of infections in the population, a random variable describing the AoI, A, within hosts in a particular cohort is:

$$A_\tau(a) \sim f_A(\alpha; a, \tau) = \frac{z_\tau(\alpha, a)}{m_\tau(a)} \tag{3.7}$$

From a trace function, we can directly compute and visualize the full probability distribution function for the AoI and the MoI in a cohort as it ages (Fig 3.4). Supplement 1 presents the derivations for these formulas and functions to compute all these quantities for a cohort of age a born on day τ directly from $h(t)$.

An important feature of this model is that the distribution of the AoI is affected by the seasonal pattern and age, but not by its overall intensity. In this system, patterns tend to be driven by the youngest infections, which has a different pattern (Fig 3.4b,c).

3.2.2 Exposure Tracking Variables

While the variables A and M describe the infection state space, a new set of variables is needed to model the history of exposure and its effects on most aspects of malaria.

We let w_i be a variable that tracks exposure, and we let δ_i be parameter describing the rate of decay of memory, δ_i:

$$w_i(a; \tau) = \int_0^a h_\tau(a - s) e^{-\delta_i s} ds \tag{3.8}$$

where in generial, $w_i(0, \tau) = 0$. For w_0, we set $\delta = 0$, and we get a variable that describes cumulative exposure. For $\delta_i > 0$, the variables track recent exposure over time scales characterized by $1/\delta_i$.

At the other extreme, we may wish to model maternal protection. To do so, we introduce a variable describing the waning effects after birth:

$$w_m(a) = w_m(0)e^{-\delta_m a} \tag{3.9}$$

We note that the initial level of maternal antibodies could vary with time, depending on the mother's history of exposure.

In the section below, we show how this set of variables $\vec{w}$ can be used to modify various aspects of the relationship between the AoI parasite densities. Collectively, the decay rates determine the time scales of the effects of recent exposure on parasite densities, and we can model anti-parasitic immunity as a function of one or more of these variables, generically denoted $\vec{w}_\tau(a)$.

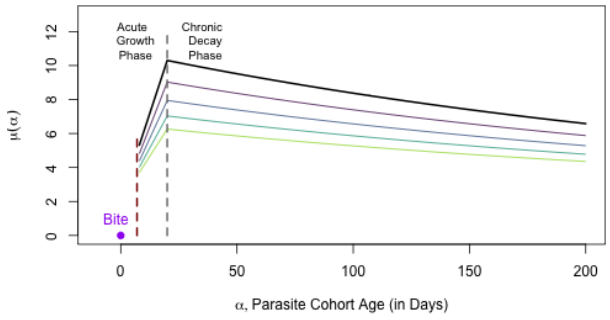


Figure 3.5: Expected parasite densities

3.2.3 Parasite Densities

Parasite densities in an infection are often measured clinically and reported as parasite counts, and either these counts or the prevalence of infection is used as a covariate in most observational studies. Early studies of malaria parasite counts observed enormous variability in a single host over time as well as among hosts [63, 64]. Evidence describing parasite densities improved in quality, granularity, and quantity during malaria therapy, when pa-

tients with neuro-syphilis were deliberately infected with malaria and carefully monitored [41]. More recently, a vast amount of detail about the early stages of infection has been learned through controlled human malaria infections (CHMI), where patients are deliberately exposed, often as part of clinical trials for developmental phase drugs or vaccines [65]. Here, we present a model for parasite densities and counts using the concept of the AoI, and we fit it to the malaria therapy data using maximum likelihood estimation.

To model parasite densities (denoted $\mu(\alpha, \vec{w})$), we proceed in several steps: first, we formulate a model for mean $\log_{10}$ parasite densities in a parasite cohort as a function of its AoI and cumulative exposure; second, we develop a family of probability distributions describing the number of infected red blood cells (iRBCs) in an infection as a function of the mean; third, we combine these two models, and we develop a random variable describing parasite density distributions in a parasite cohort of age α in a host cohort of age a born on day τ (denoted $P_\tau(a)$); and fourth, we develop a new random variable that sums parasite densities over the MoI to describe total iRBCs in an infection, $B_\tau(a)$.

Mean $\log_{10}$ Parasite Densities In malaria, across settings, the prevalence of infection with *P. falciparum* varies with respect to age: it rises to a plateau sometime before age five where it remains until about age ten before it starts to decline into adulthood, when it is approximately a third of the plateau [66]. Declining parasite densities in older adolescents and adults, and the resulting reduction in detection and prevalence, are called anti-parasite immunity. To model anti-parasitic immunity we developed a class of variables that tracks cumulative exposure and the history of recent exposure, and we use it to modify the relationship between the AoI and parasite densities. Following a previous analysis (see Chapter 2), we assume that the densities of each parasite cohort are a function of its age, α, and the history of exposure ($\vec{w}$). Generically, we let $\mu(\alpha, \vec{w}, \ldots)$ denote the average $\log_{10}$ parasite densities in a simple infection as a function of the AoI and one or more variables describing cumulative exposure, recent exposure, or maternal protection.

We developed and analyzed one particular model for parasite densities that matches

the features of malaria documented across multiple previous studies (Fig 3.6). The basic features of the function are: after an infectious bite, parasites spend at least 6 days in the liver [67]; parasites enter the blood on or after day 7 of the infection and the asexual blood-stage parasites begin a period of geometric growth; parasite populations become patent (*i.e.* detectable) around day 14; the geometric growth phase of an infection ends around day $D = 20$ days when parasite densities peak, and the infection begins a chronic phase during which parasite densities fluctuate overall but tend to decline as the infection ages. Infections can persist for several months, or even years; the duration of infections is highly variable. The model is described by the equation:

$$\mu(\alpha, w) = \begin{cases} \text{NA} & \text{if } 0 \leq \alpha < 7 \\ l + F_w(\alpha, \vec{w})(b - l)\frac{\alpha}{D} & \text{if } 7 \leq \alpha \leq D \\ l + F_w(\alpha, \vec{w})(b - l)e^{-s_\alpha(\alpha - D)} & \text{if } \alpha \geq D \end{cases} \tag{3.10}$$

A key feature of the model is thus that infections have two distinct phases: an acute growth phase when parasite densities are increasing geometrically (for $\alpha < D$); and a chronic decline phase (Fig. 3.6a; the specific functional form is found in Supplement 2).

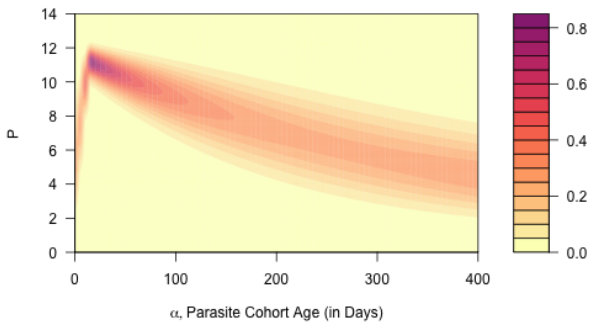


Figure 3.6: The surface plots the model using the MLE best-fits to the malaria-therapy data, with the detection process described in Eq. 3.16 (Methods).

The functional form describing the effects of immunity is flexible, so it is possible to test ideas about the effects and effect sizes of exposure on parasite densities. The software to compute and apply functions of the form $\mu(\alpha, \vec{w}, \ldots)$ was designed to be modular, so the framework would easily accommodate alternative functional forms (Supplement 2).

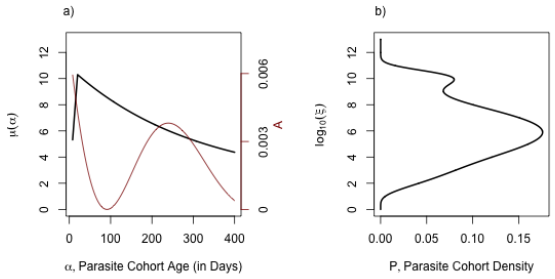


Figure 3.7: a) the expected value for parasite densities, $\mu(\alpha)$ (black, Eq. 3.10) and the distribution of AoI, A (red, Eq. 3.7) for one particular cohort at 5 years old. b) The distribution of parasite densities, P integrates parasite density distributions (Fig 3.6), with expected value $\mu(\alpha)$, over the distribution of the AoI. Ordinarily, we plot parasite density distributions with the axes switched.

Probability Density Functions In the malaria therapy data, the variance in parasite counts was a power-function of the mean, suggesting that the variance in logged parasite would be a simple linear function of the logged mean. We thus developed a realistic probability distribution function that explicitly modeled the variance as a function of the mean.

Total parasite densities must be bounded between 1 and the number of iRBCs, which is a function of host age. To model $\log_{10}$ iRBCs for a parasite cohort of age α, denoted ξ, we modified a *beta* distribution to: 1) return a value between 0 and the $\log_{10}$ maximum number of parasites, denoted $\hat{b}(a)$; 2) to accept the mean $\log_{10}$ densities; and 3) to compute the variance as a constrained function of the mean. The function that computes the variance was constrained by the shape of the β distribution (Supplement 2). A random variable describing $\log_{10}$ parasite densities in the parasite infection cohorts indexed by the AoI is

denoted P:

$$P_\tau(a) \sim f_P(\xi; a, \tau) = \beta'(\xi; \mu(A_\tau(a), \vec{w}_\tau(a)), \hat{b}(a)) \tag{3.11}$$

We note that the framework could easily accommodate other continuous probability distribution function families with finite support.

Complex Infections To model parasite densities in complex infections, some assumption about interactions among parasites within a host is required. Do they compete for resources or immune-free space, interfere directly, facilitate one another or fluctuate independently? While there are many compelling candidate models that could be developed and tested, we have analyzed parasite densities under the simple and parsimonious assumption, consistent with the other assumptions of $M/M/\infty$, that parasite densities fluctuate independently. Computing total densities for potentially complex infections thus entails adding up the parasite densities of all parasite cohorts, the number of which is given by the distribution of the MoI.

Let B denote a random variable describing $\log_{10}$ parasite densities in a host cohort of age a:

$$B_\tau(a) \sim f_B(\xi; a, \tau) = \log_{10} \sum_{M>0} 10^{P_\tau(a)} \tag{3.12}$$

This equation can be understood through the pseudo-code to draw random values of the parasite densities from individual hosts in a cohort of age a born on day τ (Algorithm 1).

We have also developed functions to compute the PDFs and CDFs for parasite density distributions in complex infections, as a convolution over a distribution of MoI (Fig 3.8c, Fig 3.9, Supplement 2). We developed functions to draw random variates, and we verified the convolution-based algorithms that computed the PDFs and CDFs for parasite density distributions in complex infections with fixed MoI and for B for any host cohort at any age, from the FoI (Fig 3.9, Supplement 2).

Since B is $\log_{10}$ transformed value of the sum values of 10^P, it follows that B is close to

Algorithm 1 Parasite densities for complex infections

$\zeta_i \sim f_M(\zeta; m_\tau(a))$ ▷ MoI
if $\zeta_i > 0$ **then**
    $\vec{w} = \vec{w}_\tau(a)$ ▷ Immunity
    for j in 1 to ζ_i **do**
        $\alpha_j \sim f_A(\alpha; a, \tau)$ ▷ AoI
        $\xi_j \sim f_P(\xi; \mu(\alpha_j, \vec{w}), a, \tau)$ ▷ Parasite density
    end for
    $\xi_i = \log_{10} \sum_j 10^{\xi_j}$ ▷ Total parasite density
else
    $\xi_i = 0$
end if

the maximum value of P, and we can bound the sum (Fig 3.9b):

$$\max(\{P\})\, B \log_{10} M + \max(\{P\}) \tag{3.13}$$

It follows that the sums of parasite densities are closer to extreme value statistics. Below, we show how B can be understood in terms of the age of the youngest infection, not the AoI.

A potential problem with this model formulation is that while the parasite densities for individual clones are bounded well below $10^{\hat{b}(a)}$, total parasite densities summed over multiple clones need not be. To deal with extreme chance events in the upper tail, the mean of the distribution of parasite densities was bounded well below the total number of iRBCs, such that computed parasite densities exceeding 1% of $10^{\hat{b}(a)}$ are extremely rare. If parasite densities in a patient were close to the absolute maximum, catastrophic health risks would be quite likely. Hyperparasitemia, a form of severe malaria defined by having more that 5% infected RBCs in immunologically-naïve individuals, is associated with most acute clinical symptoms.

3.3 Parasite Counts and Detection

3.3.1 Sampling Blood

To model some important features of malaria epidemiology, including parasite transmission and parasite-based diagnostics, the variable depends on the densities of parasites in a small fraction of total blood, either taken up by a mosquito in the blood meal, or taken up in a needle and examined by light microscopy or PCR. Parasites are thus often not detected, even when they are present. To understand the relationship between true and observed parasite densities, we have developed a model for parasite counts and detection based on the total fraction of blood examined, q. For light microscopy, where approximately 1 μl, of blood is drawn and a fraction of it examined. If we let $s \approx 10^6$ denote the number of RBCs, then the fraction of RBCs examined is approximately:

$$q(a, s) \approx 10^{s-\hat{b}(a)}. \tag{3.14}$$

A critical question is the shape of the distribution of counts conditioned on parasite densities. If parasite densities were sampled several times and counted, would the counts data follow a Poisson distribution, negative binomial distribution, or something else? In the sections that follow, we assume the distribution is negative binomial. (We use the `mu` and `size` parameterization.) The $\log_{10}$ density of parasites in a sample, $\hat{\xi}$, is thus a function of $\log_{10}$ parasite densities in blood, ξ:

$$\text{nb}(\hat{\xi}; \text{mu} = q(a, s) \cdot 10^{\xi}, \text{size} = s). \tag{3.15}$$

We can use different size parameters for different sampling processes.

Sampling Parasites

To relate the latent variables describing parasite infections to data, we model detection by light microscopy. For light microscopy, sources of sampling error include variation in parasite densities in the sampled blood and during counting, which can be exaggerated by differences among light microscopists. Similar concerns apply to PCR. There are good reasons to believe parasite densities in a blood sample are not Poisson. In comparing parasite counts made by different light microscopists for the same slide, spatial heterogeneity of parasites on the blood slide suggest a negative binomial model for counts is best. In the computational framework, the model for counts and detection is modular, and we have formulated models for the negative binomial and Poisson distributions with and without zero inflation (Supplement 2).

The following narrative is for the negative binomial model for counts and detection, which we treat separately. As q is the fraction of blood volume examined (Eq. 3.14), the expected number of parasites counted is $q \cdot 10^B$, but since only positive counts are tabulated, the distributions are modified slightly. We model parasite counts, C as a negative binomial distribution (here, we use the parameterization with `mu` and `size` in R):

$$C \sim f_C(\hat{\xi}; \xi) = \text{nb}(\hat{\xi}; \text{mu} = q \cdot 10^{\xi}, \text{size} = s_B). \tag{3.16}$$

Using the model above, we fit the model for parasite density distributions and detection to the malaria-therapy data using maximum likelihood estimation, including the parameter s_B (Supplement 3). Sampling error for this function is enormous (Fig 3.10).

In applying this model for counts to complex infections, we are left with a choice: should we model detection of individual parasite cohorts, or aggregate parasite densities in an infection? We show the analysis for a model that treats individual parasite cohorts separately, consistent with our other assumptions. We note that the observed prevalence of infection, $\hat{p}_\tau(a) = \Pr(C > 0)$, the probability of observing at least one parasite (conditioned on being infected), will be different based on the sampling model as well as the sensitivity of the method (*e.g.*, by light microscopy or PCR, rapid diagnostic tests do not count parasites).

The probability of counting at least one parasite, denoted D, conditioned on true prevalence, which we write as $D|p$ (or equivalently, $D|p = 1$):

$$D|p \sim 1 - \prod_{M>0} C(0, q10^P, s_B), \tag{3.17}$$

so the observed prevalence is

$$\hat{p} \sim p_\tau(a)D. \tag{3.18}$$

The first part of Eq. 3.18 is the probability a host is infected, and the second part is the probability parasites are detected, given infection. We note that D and $\hat{p}$ are boolean variables.

Like observed prevalence, the observed MoI, $\hat{M}$ is conditioned on being infected and detected:

$$\hat{M}|\hat{p} \sim \sum_{M>0} 1 - C(0, q10^P, s_B). \tag{3.19}$$

We note that $\hat{M} \geq 1$, since we have conditioned on detection.

Total counts are an estimate of total parasite densities. Since zero counts are excluded, we get a distribution of counts, conditioned on detection:

$$\hat{B}|\hat{p} \sim \log_{10} \sum_{M>0} C(\hat{\xi}, q10^P, s_B) \tag{3.20}$$

In Supplement 2, we compare different approaches to modeling detection, using modular computation for parasite counts, estimated MoI, and the estimated prevalence, given various models of detection.

3.4 Discussion

We developed a probabilistic framework to show how the patterns of exposure and the history of exposure are related to parasite densities, fever, transmission, treatment, and the standard metrics used. The complexity of malaria infection and immunity is at least partly

due to three main factors: a long and variable infection that changes with the AoI; the lack of sterile protection leading to an increase in the MoI, and immunity that waxes and wanes over various time scales. The complexity is amplified further by the outcomes that must be considered: detection, fever and disease, and infectiousness, all of which are direct consequences of exposure or parasite densities. Here, we have shown that the full epidemiological state space for infection can be represented concisely in terms of the distribution of the MoI and the AoI. We have proposed a flexible, semi-mechanistic framework for modeling the immune state space immunity and the effects of exposure over a range of time scales. We have shown how the distribution of parasite densities, detection, and most relevant outcomes can be modeled using simple functional forms. Notably, most of the features of this system can be understood and approximated in terms of the mean AoI and the mean AoY. The fully defined model has a large number of parameters, but the parametric complexity of the model is partly defined by the large number of outcomes. To model each outcome, we must define the quantitiy in terms of a functional form and at least one parameter.

To describe these phenomena exactly, we have we developed several linked sub-models, each one of which is parametrically simple. The dynamics of infection under the model $M/M/\infty$ have one fixed parameter (r) and an context-dependent input parameter describing the FoI in a population as it ages. The most parametrically rich is the model of parasite densities which was fit to malaria-therapy data with nine parameters: five parameters describing the expected value (two parameters are bounds, one describes time in the liver, one describes parasite cohort age at peak densities, and one describes the decline in parasite densities with the parasite cohort's age; three parameters describe the variance; two parameters describing detection, of which one is context dependent (*i.e.* fraction of blood examined). Anti-parasite immunity has one new variable and two new parameters. Clinical immunity entails two variables and four parameters: maternal protection takes two parameters; cumulative incidence requires two; and clinical protection one more. Fever prevalence associated with secondary fever required three parameters. Gametocyte densities require three parameters. The probability a mosquito would become infected requires two more. Treatment and

cure is modeled with two new parameters.

In presenting this model, we emphasis that the claims are not specific inferential claims but general ones. The general claim is that models of this sort naturally replicate the sorts of quantitative patterns that have been observed in observational malaria epidemiological studies. The conclusions are not about any specific mechanism, but that this is a useful way of modeling malaria epidemiology. The family of distributions is naturally close to malaria, and we have not attempted to show how immunity is driving the pattern over the time-course of an infection. We have made some minor modifications they would be suitable to make comparisons about the and to weight specific inferential claims through model building and model comparison. To put it another way, the comparisons we are making demonstrate that the models are in the right neighborhood. Since the models are a simple extension of available evidence, it is not surprising that they should come close.

The exact mathematical approach we have adopted here serves to provide at least one tractable analysis, to justify the overall approach. While this exact framework – working from exact solutions to queuing models to probability distributions – is difficult to extend analytically, we are thus developing in parallel a computational framework for robust computation where we can examine specific inferential claims.

We can, nevertheless, make some conclusions about malaria epidemiology. The first is that the non-linear relationship between age and prevalence relationships in malaria is likely partly explained by an interaction between the steady decline in parasite densities and the non-linear process of detection. This natural inflection point provides at least one reasonable explanation for the shape of the curves.

The second is that the medical approach to understanding malaria epidemiology, which relies heavily on parasite densities, is likely to overlook the importance of malaria incidence and infections in the growth phase. There is a narrow window of time, probably less a week, in which the growth phase of a parasite is likely to cause disease, but when parasite densities are comparatively low. Arguably, severe disease is often associated with these growth-phase infections. To distinguish these two paths, we have named fever associated with this acute

growth-phase "primary fever." There is also a heavy load of fever associated with the chronic phase, which we have called "secondary fever."

3.5 Conclusion

We have introduced a new modeling framework which bridges the gap between stochastic individual-level behavior and population-level observable quantities. This is a novel simulation method for infectious disease and sidesteps a lot of the complexity and uncertainty inherent in the mechanisms that control parasitemia, while still accounting for immunity, treatment, and superinfection. However as with all simulation models, this complexity and stochasticity can hinder our ability to isolate mechanism almost as much as the system it is meant to model. future directions of this work include working through expected behavior of the model analytically for ease of qualitative interpretation of mechanism; exploring the effects of disease-prompted treatment on age-prevalence curves; and comparing the evolutionary dynamics of the parasite with strain-specific immunity.

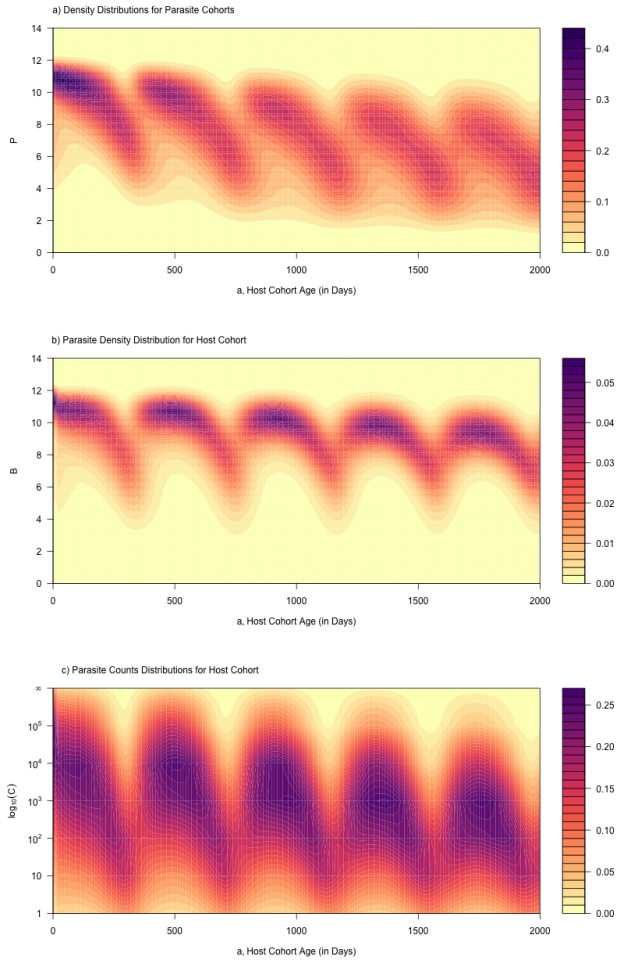


Figure 3.8: Parasite densities and counts for the same cohort illustrated in Fig. 3.4. a) The distribution of $\log_{10}$ parasite densities across all cohorts of parasites, $\log_{10} P$; b) The distribution of $\log_{10}$ parasite densities in the host cohort as it ages, $\log_{10} B$; and c) The distribution of $\log_{10}$ parasite counts, conditioned on counting at least one parasite, $\log_{10} C$. Compare the patterns in panels a) and b) here to the patterns in panels b) and c) in Fig 3.4. Mean $\log_{10}$ parasite densities in complex infections can be understood in terms of the AoY. Parasite densities are declining slightly because of anti-parasitic immunity.

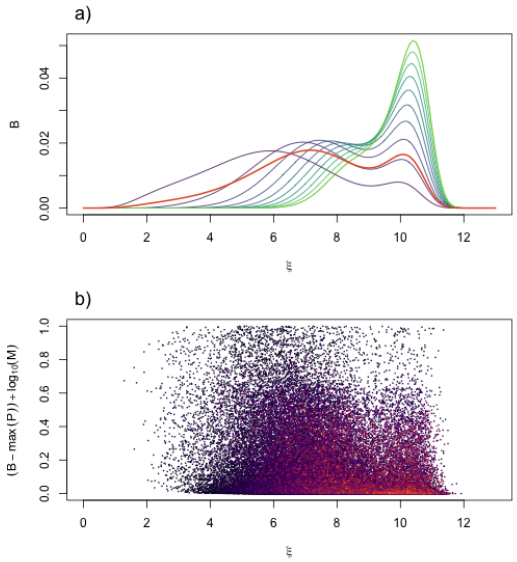


Figure 3.9: a) Parasite density distribution, $B_\tau(a)$, for one particular host cohort (orange); the distribution of parasite densities for that same host cohort, stratified by the MoI from 1 up to ten (purple to green). Importantly, the sum of parasite densities from two or more infections, a convolution, is close to the maximum, so parasite densities in complex infections are something like an extreme value statistic. b) An illustration of Eq. 3.13, showing the distribution of $(B - \max(\{P\})) \,/\, \log_{10}(M)$ is bounded between 0 and 1 for 10^5 random infections.

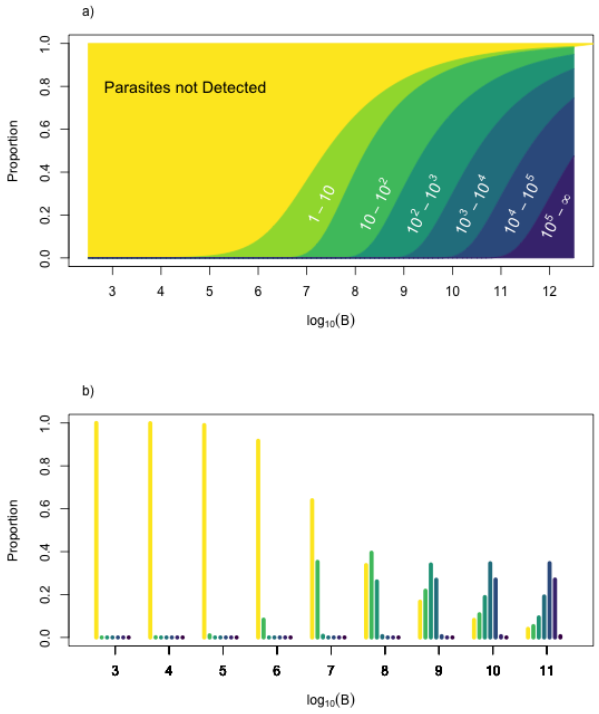


Figure 3.10: Sampling error for the detection process. a) For values of $\log_{10}$ parasite densities (x-axis), vertical slices show the proportion of parasite counts in a range defined by powers of 10 (y-axis). To make this plot, we assume the proportion of blood examined is $q = 10^{-6}$, and the size parameter was fitted to the malaria-therapy data. b) For each integer value of B, we show a histogram of a vertical slice of the figure above it. Each one is a histogram, binned by powers of 10 (colors and categories match the top figure) for $\log_{10}(B)$ for integer values from 3 to 11.

Chapter 4

GENERALIZED LINEAR MODELS OF A SEMI-MARKOV PROCESS: A NEW FRAMEWORK FOR MULTISCALE MODELING FOR MALARIA EPIDEMIOLOGY

4.1 Background

Heterogeneity at different scales is difficult to reconcile in current mathematical models of malaria epidemiology. Typically, scales are either modeled distinctly (as in population transmission models and within-host models) or all together in a complex individual based model. Individual based models play a central role in current quantitative research and policy [28], however they suffer from their complexity which makes it difficult to isolate mechanism responsible for observed patterns or fit the large number of parameters to data. Further, mechanisms at one scale affect mechanisms at other scales, so models focusing on just a single scale are of limited utility when analyzing complex policy questions.

For example, treatment-seeking behaviors play a huge role in public health. Individual infections vary in intensity over time within hosts and across populations, and such heterogeneity leads to potentially large implications for transmission at a population level. For example parasite densities, a strongly predictive quantity of detection and disease in an individual, can span orders of magnitude over the course of several days [35, 56]. Just knowing the average parasitemia in a human host stratum is not enough to make inferences for the overall population, as the average of a function of parasitemia is not the same as a function of the average parasitemia. This means that the entire distribution of parasitemia is helpful for understanding transmission, not just the expected value. Therefore to model the distribution of parasitemia over time in an individual, a stochastic process can be defined.

Here we extend the probabilistic framework introduced in chapter 3, and more rigorously

consider the dynamics without resorting to simulation. For ease of analysis we will assume independence of the infections clearing rates conditional on a host's immune state, allowing us to easily convert from statistical properties of individual infections to observable quantities at a population level such as the likelihood a person chosen at random will test positive by light microscopy or be febrile due to at least one infection independently causing a fever. Unfortunately due to the complex nature of malaria infections, stochastic processes which model the parasitemia directly are typically complex and require some assumption about the immune response's mechanism of control of parasite densities, which is itself a field of current research [53, 22]. Therefore here we model a statistically observed pattern (outlined in Chapter 2) of the distribution of parasitemia as predicted by the time since infection. This in turn allows us to use heterogeneous individual information into a model of what we could expect to see at a population level, using only relatively simple ODEs.

4.2 Methods

We will first recall the model introduced in chapter 3. The unnormalized infection age density of an infection chosen at random is given by

$$\begin{aligned} \frac{\partial z}{\partial a} + \frac{\partial z}{\partial \alpha} &= -rz \\ z(a,0) &= h(a) \end{aligned}$$

where $z(a,\alpha)$ is the density of *individual infections* of age a and infection age α, making this an age-cohort model.

The solution is given by

$$z(a,\alpha) = h(a-\alpha)e^{-r\alpha}$$

As a host cannot have an infection older than they are, the domain is $0 \leq \alpha \leq a$. To normalize this density we can integrate across the possible infection ages:

$$N(a) = \int_0^a z(a, \alpha) d\alpha$$

Differentiating the expression, we find that the normalizing constant is exactly equal to the expected number of infections given a host is of age a, modeled through the mean of the M/M/∞ queueing process of superinfection m [8]:

$$\begin{aligned} \frac{dm}{da} &= h(a) - rm \\ m(0) &= 0 \end{aligned}$$

The PDF of the infection age density is then

$$\begin{aligned} f_A(a, \alpha) &= \frac{z}{m} \\ &= \frac{h(a-\alpha)}{m} e^{-r\alpha}, \\ \frac{dm}{da} &= h(a) - rm(a) \end{aligned}$$

which allows us to model the entire distribution of parasitemia using a single ODE for m and a relatively simple equation to transform it into a density.

Now we will build on this basic model structure to produce models of the quantities we care about. In order to do this we will build a statistical model as if A_a, the age of an infection chosen at random from a population, is a random variable we have observed. That is, we consider detection and fever as response variables in a generalized linear model where parasite densities are the predictor:

$$g(\hat{\mu}_t) = c_0 + c_1 P_t$$

where g is a link function for the GLM. Typical link functions include the identity function, the natural log, or the logit. Here we are primarily interested in logit link functions, as fever and detection are logistic functions of parasitemia.

Note here P_t, the distribution of log parasitemia in an active infection chosen at random, is a stochastic process, meaning we are defining our expected value in the GLM as a function of this stochastic process:

$$\hat{\mu}_t = g^{-1}(c_0 + c_1 P_t)$$

Further, our log parasitemia is modeled as a linear function of time since infected:

$$P_t = c_2 - c_3 A_t$$

where A_t is the age of infection in an individual. We can therefore define the following estimator:

$$\begin{aligned}
\hat{\mu}_t &= g^{-1}(c_0 + c_1(c_2 - c_3 A_t)) \\
&= g^{-1}(c_0 + c_1 c_2 - c_1 c_3 A_t) \\
&= g^{-1}(\kappa - C A_t)
\end{aligned}$$

where $\kappa := c_0 + c_1 c_2$ and $C := c_1 c_3$ are respectively the effective intercept and slope of the linear models. Conditioning on A_t allows us to produce a more easily modeled pro-

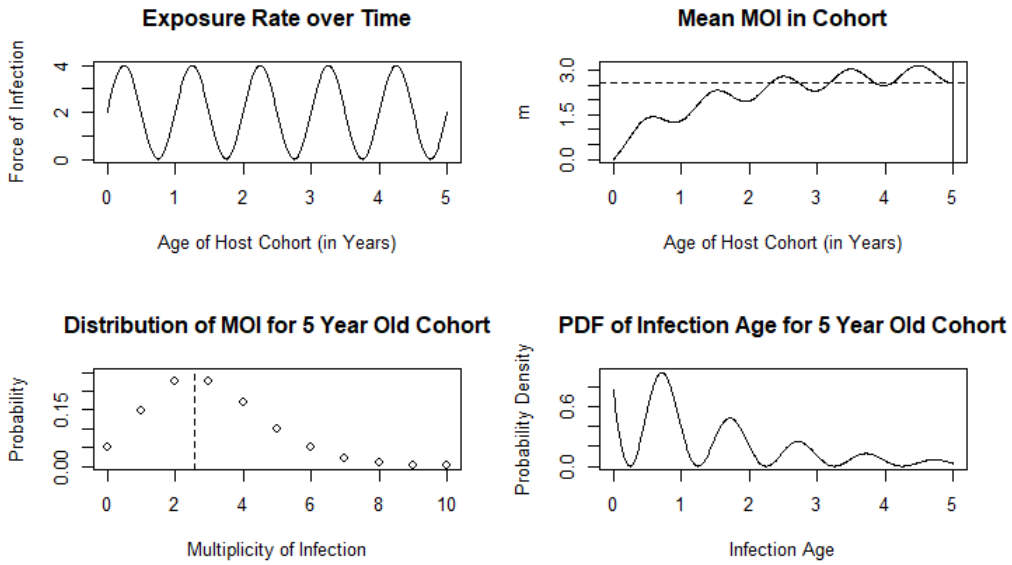


Figure 4.1: Top left: Example of seasonal exposure rate for a cohort of children from age 0 to 5 years, born at the beginning of a wet season. Top right: Expected number of infections in an individual chosen at random from a cohort of children from age 0 to 5 years. Bottom left: The distribution of the multiplicity of infection in the 5 year old cohort; vertical line represents the expected value, as also marked by the horizontal dotted line in previous plot. Bottom right: The probability density function of the infection age of an *infection* chosen at random (noting that individual hosts may have multiple independent infections concurrently).

cess of infection occurence and clearance, and thus we can define our response variable as a homeomorphism of the semi-Markov aging process. In particular this transformation is trivially "lumpable" per the definition in [68], and by the same paper's results this transformed semi-Markov process is again a semi-Markov process.

4.3 Results

4.3.1 No Treatment

Recall the PDE from Chapter 3 which gave the unnormalized density of infection age:

$$\begin{aligned}\frac{\partial z}{\partial a} + \frac{\partial z}{\partial \alpha} &= -rz \\ z(a,0) &= h(a) \\ z(0,\alpha) &= \delta_0(\alpha)\end{aligned}$$

The solution is given by

$$z(a,\alpha) = h(a-\alpha)e^{-r\alpha}$$

To normalize the density, we note that for each fixed age a the total density is given by

$$\begin{aligned}m &= \int_0^a z(a,\alpha)d\alpha \\ &= \int_0^a h(a-\alpha)e^{-r\alpha}d\alpha\end{aligned}$$

As suggested by the notation, this m is also the expected MOI in the cohort, that is the solution to the following ODE:

$$\begin{aligned}\frac{dm}{da} &= h(a) - rm \\ m(0) &= 0\end{aligned}$$

Therefore the normalized density can be computed and tracked by just this single ODE:

$$\begin{aligned} y &:= \frac{z}{m} \\ &= \frac{h(a-\alpha)}{m} e^{-r\alpha} \\ \frac{dm}{da} &= h(a) - rm \\ m(0) &= 0 \end{aligned}$$

This is summarized in figure 4.1. An example signal representing a seasonal exposure signal $h(a)$ is plotted in the top right, the solution to our mean MOI equation in terms of this signal is given in the top right, the full distribution for that given mean is in the bottom left, and finally combining those as in the equations above we obtain the PDF of infection age of an infection chosen at random.

Although interesting in itself, the above infection age density is primarily of interest due to its statistical relationships to log10 parasitemia, detection, and fever. This is demonstrated empirically in chapter 2, and summarized in figure 4.2. As both fever and detection follow the same statistical pattern, detection only is described in figure 4.2. The predictor we have is the infection age density; we can convert this to an estimate for log10 parasitemia per cmm of blood as shown in the top right; and detection is a sigmoidal function of parasitemia, as shown in the bottom left. Applying these two functions and applying a quasiequilibrium assumption to the detection PDF, we obtain the predicted density of detection probability in the bottom right. This section describes this procedure without treatment or effects of immunity.

As each of the transformations in figure 4.2 are smooth and invertible, we can obtain an expression for our outcome of interest in the bottom right as X, representing either deteciton or fever rates per infection. Applying the two transformations in terms of our infection age, we get

$$
\begin{aligned}
E[P|A] &= c_0 - c_1 A \\
E[X|P] &= \sigma(c_2 + c_3 P) \\
\hat{X} &:= E[X|A] \\
&= E[X|E[P|A]] \\
&= \sigma(c_2 + c_3(c_0 - c_1 A)) \\
&= \sigma(k - cA)
\end{aligned}
$$

where $k := c_2 + c_0 c_3$, $c := c_3 c_1$, and

$$\sigma(x) = \frac{e^x}{1 + e^x}$$

is a standard sigmoid function. Therefore, the outcome we are interested in is a sigmoidal function of a linear function of infection age.

Now to understand the dynamics of the transformed semi-Markov process of infection age, we will rewrite our dynamics in terms of the moment generating function (MGF). Recall that the MGF of a random variable X is defined to be

$$G_X(s) = E[e^{sX}]$$

and can be thought of as a projection of the PDF (if it exists) onto an exponential basis. Analogously, the PDE for the characteristic function is a projection of the PDF onto a Fourier basis; the latter will always exist for a random variable while the former may not. We choose here for notational convenience to work with an exponential basis, though it should be noted that all of the analysis could be repeated with a Wick rotation in which all s in our PDE for the MGF are replaced with is, where i is the imaginary unit.

The PDE for the moment generating function of the estimator is given by

$$\begin{aligned}\frac{\partial G_{\hat{X}}}{\partial a} &= \frac{h(a)}{m}\left[e^{-s\sigma(k)} - G_{\hat{X}}\right] - cs\frac{\partial G_{\hat{X}}}{\partial s} + cs\frac{\partial^2 G_{\hat{X}}}{\partial s^2} \\ G_{\hat{X}}(a,0) &= 1 \\ \lim_{s\to\infty} G_{\hat{X}}(a,s) &= 0\end{aligned}$$

(see Appendix 4). Note the diffusion term appears here as a result of the fact that the sigmoid function $\sigma(x)$ satisfies

$$\frac{d\sigma}{dx} = \sigma(x)(1-\sigma(x))$$

and the each s-derivative of the moment generating function brings down a copy of $\sigma(\tilde{k}+\tilde{c}A_a)$ from the exponent. Interestingly this implies that in our moment generating function space, our dynamics appear to act as a diffusion in our exponential basis and thus methods from linear advection-diffusion theory can be applied here. Further details applying these methods are given in the appendix, but here we will just present results and then apply the result to develop an ODE model for the mean.

The stationary CDF is given by

$$F_{\hat{X}}(x) = e^{-kr/C}x^{\frac{r}{C}}(1-x)^{-\frac{r}{C}}$$

and its PDF is

$$\begin{aligned}f_{\hat{X}}(x) &= \frac{r}{C}e^{-\kappa r/C}x^{\frac{r}{C}-1}(1-x)^{-\frac{r}{C}-1} \\ &\propto x^{\frac{r}{C}-1}(1-x)^{-\frac{r}{C}-1}\end{aligned}$$

This family of distributions is parameterized by a single parameter, r/C. With a quasiequi-

librium assumption, the mean of the estimator can be show to follow

$$\frac{dm_1}{da} = \frac{h}{m}\sigma(k) - \left[\frac{h}{m} + C(1-\sigma(k))\right] m_1$$

(see Appendix 4) and thus the entire distribution can be tracked by this single linear ODE.

Remember that this is the per-infection expected probability of detection or fever (depending on the model used). Notice that under the dynamics of m_1, any initial condition inside the unit interval will remain there indefinitely as the rate of recovery will be strictly larger than the rate of infection. The equilibrium of this equation is

$$x^* = \frac{\frac{h}{m}\sigma(k)}{\frac{h}{m} + C(1-\sigma(k))}$$

which rises with exposure to a maximum value of $\sigma(k)$, the value of $\hat{X}$ when infection age is 0. Further, C dictates the sensitivity of x to changes in infection age by determining the saturation rate.

Further, equilibrium implies that m is also at equilibrium and equal to h/r. This gives

$$x^* = \frac{r\sigma(k)}{r + C(1-\sigma(k))}$$

This shows that the higher the rate of recovery is, the higher the per-infection rate of disease and detection. This makes sense; lower recovery rate will lead to higher prevalence, and in particular a higher prevalence of old infections. These infections are less likely to cause disease or be detected. Conversely, increasing the rate of recovery means fewer infections and younger infections, which are typically easier to detect and more severe.

Further, we have a full distribution around this expectation. This allows us to make unbiased estimates of transformations of our random variables. For example, to get the total detection rate we have the following relationship:

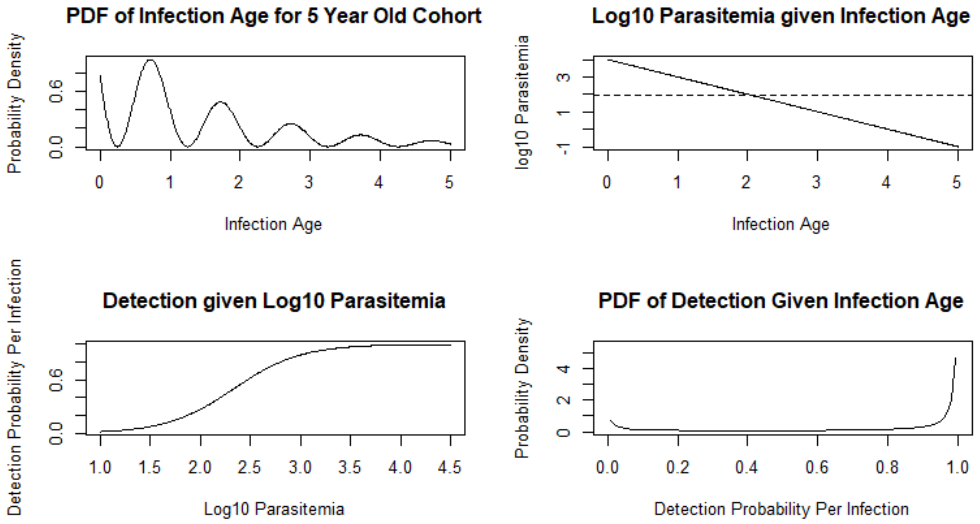


Figure 4.2: Top right: Probability density function of infection age for an infection chosen at random from the cohort of 5 year old hosts. Top right: The loglinear relationship between parasitemia and infection age; Bottom left: The logistic relationship between log10 parasitemia and probability of detecting an infection; Bottom right: Stationary probability density function of the distribution of detection rates. Note this is a transformation of the distribution on the top left through the smooth and invertible relationships described by the previous two plots.

$$PR_{tot} = 1 - e^{-mD}$$

where D is our estimator of detection given infection age.

Plugging in, we get the following system of equations for the dynamics, and a set of functions which translate the state variables into observables:

$$\begin{aligned}
\frac{dm}{da} &= h - rm \\
\frac{dD}{da} &= \frac{h}{m}\sigma(k_D) - \left[\frac{h}{m} + c_D(1 - \sigma(k_D))\right] D \\
\frac{dF}{da} &= \frac{h}{m}\sigma(k_F) - \left[\frac{h}{m} + c_F(1 - \sigma(k_F))\right] F \\
PR_{true} &= 1 - e^{-m} \\
PR_{tot} &= 1 - e^{-mD} \\
F_{tot} &= 1 - e^{-mDF}
\end{aligned}$$

which, to our knowledge, is the first simple ODE model which incorporates information at an individual-level to affect observable population-level metrics.

4.3.2 Immunity as a Covariate

So far, we have assumed that all infections follow the same infection age to parasitemia relationship. However this is clearly not the case, as immunity acts to suppress parasitemia as cumulative exposure is gained. Assuming antiparasite immunity wanes at some average rate ω and is stimulated up to some saturating rate by exposure h,

$$\frac{dI}{da} = \frac{Ah}{1+Bh} - \omega I$$

Treating this as a deterministic covariate to our parasitemia such that

$$log(E[P|A, I]) \sim c_0 - c_1 A - c_2 I$$

We can again follow the same steps and show that the equations are modified by the covariate as follows:

$$\begin{aligned}
\frac{dm}{da} &= h - rm \\
\frac{dD}{da} &= \frac{h}{m}\sigma(k_D - \kappa_D I) - \left[\frac{h}{m} + c_D(1 - \sigma(k_D - \kappa_D I))\right] D \\
\frac{dF}{da} &= \frac{h}{m}\sigma(k_F - \kappa_F I) - \left[\frac{h}{m} + c_F(1 - \sigma(k_F - \kappa_F I))\right] F \\
\frac{dI}{da} &= \frac{Ah}{1+Bh} - \omega I \\
PR_{true} &= 1 - e^{-m} \\
PR_{tot} &= 1 - e^{-mD} \\
F_{tot} &= 1 - e^{-mDF}
\end{aligned}$$

Importantly, we assumed here that immunity was gained changed the statistical properties of an infection without changing the susceptibility to, or the duration of, an infection. However subpatency will occur sooner and disease is less likely given higher levels of immunity, giving the age-prevalence curve implied by the model the characteristic bend downward with age.

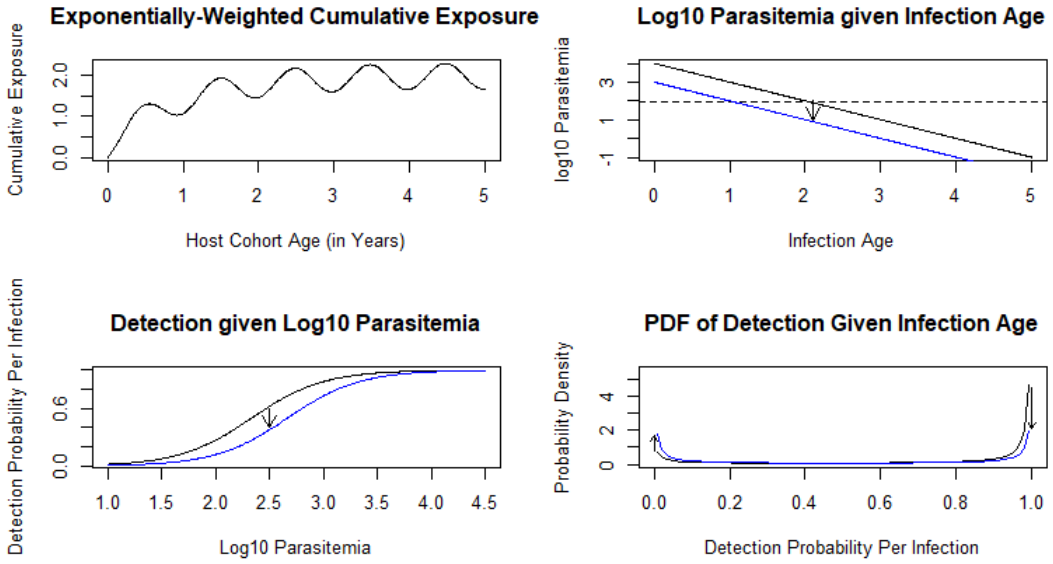


Figure 4.3: Top right: Exponentially-weighted cumulative exposure in a human host cohort over 5 years. This is used as a very simple model of immunity, assuming it follows a first order autoregressive process. Top right: The effect of anti-parasite immunity. Here it is assumed to reduce parasitemia by a set amount on average, indicated by the shifted blue line. Bottom left: Detection given the reduced parasitemia demonstrated in the top right. Bottom right: Finally propagated to the density of detection rates, we see that more immunity results in a larger likelihood of low detection probabilities.

4.3.3 Fever-Prompted Treatment

One major reason to model disease is to use disease rates as a surrogate for treatment-seeking. Access to treatment can have massive effects on the prevalence of malaria, but the decision to seek care is a complex and human process dependent on many factors. By modeling fever, we can provide some estimate of a treatment rate responsive to changes in exposure and immunity.

$$\begin{aligned}
\log(E[\tau|A,I]) &= \tau_0 + c_F \text{logit}(E[F|A]) \\
&= \tau_0 + c_F(k - c_{1,F}A - c_I I) \\
\log(E[\tau|A]) &= \tilde{\tau}_0 - \tilde{c}_A A - \tilde{c}_I I \\
E[\tau|A,I] &= e^{\tilde{\tau}_0 - \tilde{c}_A A - \tilde{c}_I I}
\end{aligned}$$

Therefore, we see that treatment follows an exponential function of infection age and immunity, both depressing the treatment rate as they increase.

The exponential transformation is a simpler transformation than the sigmoidal transformation, and this case is also covered in the appendix (see subsection "Estimators Conditioned on Infection Age"). As before, we can derive a PDE for the MGF of the estimator:

$$\frac{\partial G_{\hat{X}}}{\partial a} + \frac{\tilde{c}_A}{m} s \frac{\partial G_{\hat{X}}}{\partial s} = \frac{h}{m}\left[e^{se^{\tau_0 - \tilde{c}_I I}} - G_{\hat{X}}\right]$$

Taking a derivative with respect to s and setting $s = 0$ for the MGF, we get an equation for the first moment:

$$\frac{\partial^2 G_{\hat{X}}}{\partial a \partial s} + \frac{\tilde{c}_A}{m}\frac{\partial G_{\hat{X}}}{\partial s} + \frac{\tilde{c}_A}{m} s \frac{\partial^2 G_{\hat{X}}}{\partial s^2} = \frac{h}{\tilde{c}_A}[e^{\tau_0 - \tilde{c}_I I} e^{s e^{\tau_0 - \tilde{c}_I I}} - \frac{\partial G_{\hat{X}}}{\partial s}]$$

$$\frac{dm_1}{da} + \frac{\tilde{c}_A}{m} m_1 = \frac{h}{\tilde{c}_A}[e^{\tau_0 - \tilde{c}_I I} - m_1]$$

$$\frac{dm_1}{da} = \frac{h}{\tilde{c}_A} e^{\tau_0 - \tilde{c}_I I} - \left[\frac{\tilde{c}_A}{m} + \frac{h}{\tilde{c}_A}\right] m_1$$

where m_1 is the first moment (that is, the expectation) of our variable $\hat{X}$. Noting that this is to model our expected per-infection treatment rate, we can set that to τ and see that

$$\frac{d\tau}{da} = \frac{h}{\tilde{c}_A} e^{\tau_0 - \tilde{c}_I I} - \left[\frac{\tilde{c}_A}{m} + \frac{h}{\tilde{c}_A}\right] \tau$$

and therefore we have another first order autoregressive model for the treatment rate.

Treatment will be assumed to be statistically uniform across the population conditional on their immune status, but independent of MOI directly. This allows us to apply the framework introduced in chapter 1, showing that the MOI distribution is shifted to follow an Alternative Hyperpoisson Distribution [25, 56]. Therefore, the probabilistic model has the following model structure:

$$
\begin{aligned}
\frac{dm_1}{da} &= h - (\tau + \delta + r)m_1 \\
\frac{dm_2}{da} &= h + (2h + r)\, m_1 - (\tau + \delta + 2r)m_2 \\
\frac{dp}{da} &= \tau x - \omega p \\
\frac{dD}{da} &= \frac{h}{m}\sigma(k_D - \kappa_D I) - \left[\frac{h}{m} + c_D(1 - \sigma(k_D - \kappa_D I))\right] D \\
\frac{dF}{da} &= \frac{h}{m}\sigma(k_F - \kappa_F I) - \left[\frac{h}{m} + c_F(1 - \sigma(c_F - \kappa_F I))\right] F \\
\frac{dI}{da} &= \frac{Ah}{1 + Bh} - \omega I \\
\frac{d\tau}{da} &= \frac{h}{\tilde{c}_A} e^{\tau_0 - \tilde{c}_I I} - \left[\frac{\tilde{c}_A}{m} + \frac{h}{\tilde{c}_A}\right] \tau \\
\alpha &= \frac{m_1(m_2 - m_1)}{2m_1^2 + m_1 - m_2} \\
\beta &= \frac{2m_2^2 - 2m_1 - 2m_1^2}{2m_1^2 + m_1 - m_2} \\
PR_{true} &= (1 - p)[1 - \Phi(1; \beta + 1; -\alpha)] \\
PR_{tot} &= (1 - p)[1 - \Phi(1; \beta + 1; -\alpha D)] \\
F_{tot} &= (1 - p)[1 - \Phi(1; \beta + 1; -\alpha DF)]
\end{aligned}
$$

where $\Phi(a; b; z)$ is Kummer's Confluent Hypergeometric function and p is the proportion chemoprotected for an average of $1/\omega$ units of time.

These equations are significantly different than those derived through standard compartment models. Each of these variables represents a probabilistic process, sometimes requiring two moments (as in the first two moments of the distribution of MOI, m_1, m_2) to specify and otherwise requiring just one (as in the proportion detectable or febrile, D, F). The prevalence here is not a state variable, but rather computable from the state variables. This view allows us to represent the system in an expanded state space, separating the dynamics of the system from the observables we use to measure them.

4.3.4 Transmission

Finally, we turn to the last piece of information described in chapter 2 - gametocytes and transmission. Gametocytes are the sexual stage of the parasite during the erythrocytic stage of *P. falciparum*, and their densities are directly relevant for predicting the distribution of transmission efficiency to mosquitoes. The transmission efficiency here is interpreted as the probability that an uninfected mosquito bites an infected human host and becomes infected.

From chapter 2, we noted that log10 gametocyte density given the history of parasitemia was

$$E[G_a|P_s, s \leq a] = \begin{cases} -\infty & \text{if } a - \delta < 0 \\ g_0 + cL_\delta P_a & \text{if } a - \delta \geq 0 \end{cases}$$

that is, we assume that there are no gametocytes before some lag of length δ (so the log density is defined to be $-\infty$), and after that delay gametocytes are equal proportional to some linear function of *lagged* parasitemia. Notationally, $L_\delta f(x) := f(x - \delta)$ represents the lag operator with lag equal to δ. From this, we obtain an expression for our transmission efficiency given infection age:

$$E[TE_a|A_s, s \leq a] = \begin{cases} 0 \text{ if } a - \delta < 0 \\ \sigma(E[G_a|P_s, s \leq a]) \quad \text{if } a - \delta \geq 0 \end{cases}$$

$$= \begin{cases} 0 \text{ if } a - \delta < 0 \\ \sigma(E[G_a|E[P_s|A_s], s \leq a]) \quad \text{if } a - \delta \geq 0 \end{cases}$$

$$= \begin{cases} 0 \text{ if } a - \delta < 0 \\ \sigma(g_0 + cL_\delta E[P_a|A_a]) \quad \text{if } a - \delta \geq 0 \end{cases}$$

$$= \begin{cases} 0 \text{ if } a - \delta < 0 \\ \sigma(\tilde{g}_0 + \tilde{c}L_\delta A_a) \quad \text{if } a - \delta \geq 0 \end{cases}$$

Therefore, all of the math above holds just as it did for fever, though this time our transmission efficiency depends on a scaled and lagged version of our infection age:

$$\frac{dTE}{da} = \frac{h_\delta}{m_\delta}\sigma(k_{TE} - \kappa_{TE}I_\delta) - \left[\frac{h_\delta}{m_\delta} + c_{TE}(1 - \sigma(k_{TE} - \kappa_{TE})I_\delta)\right]TE$$

We can use this to determine a distribution for the transmission efficiency, and compare this qualitative description to the statistical patterns observed in chapter 2, as illustrated by figure 4.4.

4.4 Discussion

The probabilistic models derived here represent a departure from compartment models in a significant way. We were able to write equations which describe the fraction of a population with detectable and febrile infections, and just as importantly the proportion undetectable or asymptomatic, as well as a dynamic treatment rate and immunity which modifies all of these relationships. Further, these equations represent the expected behavior of the more complex

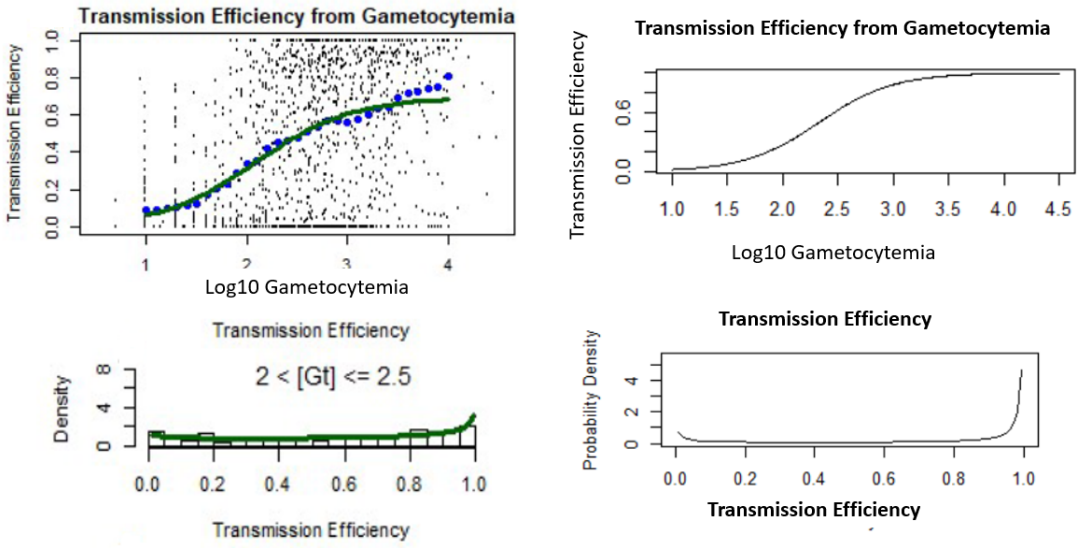


Figure 4.4: Empirical relationship between log10 gametocytemia in malaria therapy patients and transmission efficiency (top left), with a histogram of transmisison efficiency and fitted beta distribution in green (bototm left). The panels on the right mirror this behavior, but in the GLM framework of infection age presented here. Qualitatively, there is good agreement between the distribution of transmission efficiency.

probabilistic model introduced in chapter 3. By mapping the dynamics of the simulation framework onto ODEs of the mean behavior of the system, fitting and exploring mechanism analytically become possible.

The hybrid mechanistic and statistical nature of this framework allows one to consider other covariates on parasitemia, for example coinfection with other pathogens or individual health states such as nutrition and hemaglobinopathies. The covariates can be static or dynamic, allowing for stratification into different categories for a static covariate (eg presence/absence of sickle cell trait) or a dynamic covariate (eg coinfection with HIV). Drawing from the strength of the mechanistic side, we can further consider other hypothetical mechanisms and compare the statistical evidence for them through the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) implied by the resulting likelihoods.

However, the complex structure requires one to be careful about modifications. For example, including treatment in a system will alter the distribution of the multiplicity of infection (see chapter 1) and affect much of the downstream analysis, as infections are no longer clearing independently. If one infection causes symptoms, then it is more likely that treatment will be sought and all active infections could be cleared. Therefore it requires careful consideration of the underlying distributions before changes are made in the ODEs derived here.

We believe this modeling framework will allow us to think outside the compartment, and consider statistical patterns' implications on the known or assumed mechanisms generating the data. Further, the formulation within a state space paradigm allows us to separate the observation process from the true state. This follows the work of [69, 70, 71] in ecology. While the simulation-based approach to fitting does allow for arbitrary complexity, the simplicity and amenability of the presented method to be combined with other deterministic or stochastic models for other components of the system (e.g. mosquito ecology and human movement) allows for a modular approach to modeling a system.

It is our hope that others will build from this example to examine other predictors to decrease the variability inherent in infection age, include predictors with different dynamic

ranges, and also include other response variables of interest such as hemoglobin concentration or exposure to different phenotypes of parasite. This method may also be used in tandem with simulation models to predict interesting average behaviors, and to explore which mechanisms cross-scale heterogeneity influences the most.

4.5 Conclusions

Here we simultaneously modeled the statistics at an individual-level through a simple stochastic process and the expected behavior at the population level. This provides a new possible toolbox for multi-scale modeling of infectious disease.

Chapter 5

A SIMULATION-FREE MAXIMUM LIKELIHOOD METHOD FOR PARTIALLY OBSERVED MARKOV MODELS IN EPIDEMIOLOGY

5.1 Background

Recent developments in the statistical methodology for epidemiology have provided an exciting new computational toolbox for practitioners to apply to better understand the spread of disease. In particular, partially observed Markov processes (POMP), also known as state space models, have proven to be a powerful framework for fitting prevalence data to model parameters of interest such as the force of infection and the rate of recovery [69, 70, 71]. POMP treats measurement error and process error separately, allowing for more transparent error propagation in estimation. Therefore to achieve the maximum likelihood estimates of the parameters, one needs to be able to either evaluate or simulate from each of the likelihood functions. POMP is very flexible but most implementations have a bottleneck in efficiency due to their need to rely on Monte Carlo simulations to estimate the likelihood function.

The purpose of this paper is to describe an approximate method of modelling and fitting of epidemiological processes to time series data in the spirit of POMP in a way which is analytically tractable to circumvent the need for Monte Carlo simulations. This represents special cases of models common to epidemiology, so gains in efficiency come at the cost of the generality of POMP methods to nonlinear models.

The philosophy behind this method is to consider the state space for individuals, use this to construct the state space for a population, then describe the functional which maps the state of the population to a particular measure of the system (e.g., the proportion of the population with an infection). Here we will consider several simple examples in epidemiology

under the assumption that the data collected is prevalence over time. This implies that individuals are in one of two possible states: susceptible (state '0') or infected (state '1'). This fully represents the state space for an individual in terms of the data as measured. For a population consisting of N individuals, we have a series of N binary random variables which may switch their state at any point in time according to a Poisson process. Finally, the functional on the population state space is the proportion infected; that is,

$$X_t := \frac{1}{N} \sum_{i=1}^{N} 1_{X_{t,i}=1}$$

where $1_{X_{t,i}=1}$ is the indicator variable which is 1 iff $X_{t,i} = 1$. This takes a possible state of the system given by the collection of N binary random variables $\{X_{t,i}\}_{i=1}^{N}$ and maps them onto a number between 0 and 1 in increments of $\frac{1}{N}$. This defines the realized prevalence in a population. Therefore the models under consideration will describe the rules for how individuals move from one state to another.

In compartment models, some of these states split the uninfected and infected states into multiple compartments. A simple example is the SIR model, where 'R' is a separate uninfected category. A more complex version is the M/M/∞ queueing model of superinfection, which split the infected category into a countably infinite set of states which describe the number of infections in an individual. In each of these cases, we can define $X_{t,i}$ as an indicator function that an individual chosen at random in these models is in the infected class

We will consider here the linear case of estimation, with some consideration given to approximations of nonlinear systems. There are two major reasons to restrict ourselves to linear methods here. First, many nonlinear systems with stable equilibria can be well approximated by their linearized counterparts and so the simplification is well justified. The second reason is more interesting based on the structure of the stochastic models; nonlinear models distort the path of each trajectory in a non-uniform manner, and so the distribution

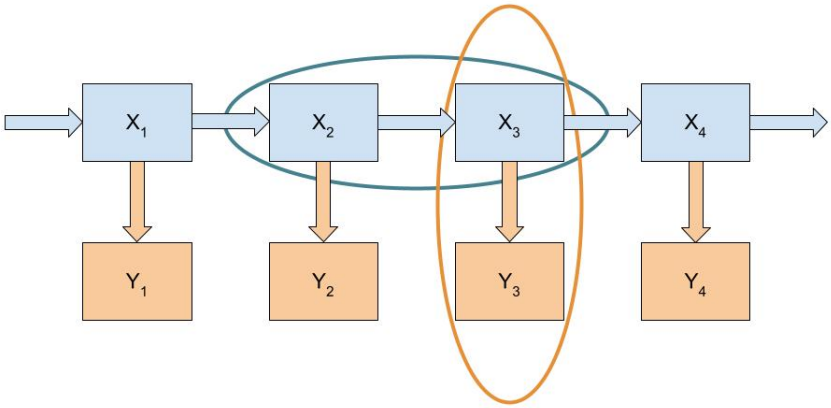


Figure 5.1: Schematic of the computation of the dependence between observations. Blue squares represent latent states and orange squares represent the observation of the state. Observations are assumed to only have dependence through the underlying mechanism. The blue oval represents a latent state transition, and the orange oval represents an observation.

of all trajectories becomes much harder to ascertain and is typically not a named distribution. The linear dynamics discussed here result in beta and gamma distributed noise about some mean value, with correlation structure determined by the assumed mechanism.

5.2 Methods

Following the POMP paradigm, we assume two levels to our model. First, a mechanistic model which transitions latent states from one time to the next (as shown in the blue boxes and arrows in figure 5.1); and second, there is an observation process (orange boxes) which allows us to connect observational data with the state parameters. The likelihood can then be broken down into simple components parts, applying a Markov assumption from the mechanism to decompose the likelihood into the initial state of the system, state transi-

tion probabilities conditioned on the previous assumed state, and observation probabilities conditioned on the current observation. Concretely,

$$P(Y_1, ..., Y_n, X_1, ..., X_n|\Theta) = P(X_0|\Theta) \prod_i P(Y_i|X_i)P(X_i|X_{i-1})$$

Typically, the transition density $P(X_i|X_{i-1})$ is approximated through simulation through a method such as particle filtering [cite] which can be computationally costly. However in the simple cases presented here, it can be modeled exactly. Further, if we know the transition density, we can marginalize out the state variables to obtain the joint likelihood of just the observations:

$$P(Y_1, ..., Y_n|\Theta) = P(X_0) \prod_i \int_\Omega P(Y_i|X_i = x) \left[\int_\Omega P(X_i = x|X_{i-1} = y, \Theta)dy \right] dx$$

Therefore the mathematical goal here is to obtain both an expression for the transition density $P(X_i = x|X_{i-1} = y)$ and the posterior predictive distribution for the observations given the state $P(Y_i|X_i = x)$.

SIS Dynamics and the Beta-Binomial Model

5.3 Results

5.3.1 SIS Dyanamics and the Beta-Binomial Model

Model Definition

$$dX_t = [\beta(1 - X_t) + rX_t]dt + 2\sqrt{\frac{\beta r X_t(1 - X_t)}{\beta(1 - X_t) + rX_t}}dW_t$$

Focusing on the noise term, we get

$$2\sqrt{\frac{\beta r X_t(1-X_t)}{\beta(1-X_t)+rX_t}}dW_t = 2\sqrt{\frac{\beta r X_t(1-X_t)}{\beta+(r-\beta)X_t}}dW_t$$
$$= 2\sqrt{\frac{rX_t(1-X_t)}{1+\left(\frac{r}{\beta}-1\right)X_t}}dW_t$$

Note that X_t will likely be near its equilibrium, so if we replace the prevalence term with its equilibrium value, we get

$$2\sqrt{\frac{rX_t(1-X_t)}{1+\left(\frac{r}{\beta}-1\right)X_t}} \approx 2\sqrt{\frac{rX_t(1-X_t)}{1+\frac{r-\beta}{\beta}\frac{\beta}{r+\beta}}}$$
$$\approx 2\sqrt{r}\sqrt{X_t(1-X_t)\left(1-\frac{r-\beta}{r+\beta}\right)}$$
$$= 2\sqrt{\frac{2r\beta}{r+\beta}}\sqrt{X_t(1-X_t)}$$

On the other hand, if either parameter is much larger than the other, then X_t will typically be close to 0 or 1 and so the noise term will be close to zero. Therefore the above approximation is likely to be good in general. Therefore we can get an approximate SDE of the form

$$dX_t = (h+r)\left[\frac{h}{h+r} - X_t\right]dt + 2\sqrt{\frac{2rh}{r+h}}\sqrt{X_t(1-X_t)}dW_t$$

Note that using the traditional method, the SDE formed would be

$$dX_t = (h+r)\left[\frac{h}{h+r} - X_t\right]dt + \sqrt{h(1-X_t)+rX_t}dW_t$$

which has the same drift but a different scaling on the noise. When $\beta \approx r$, they get noise that scales with $\sqrt{\beta}$. In the model presented here, we get $2\sqrt{\frac{2r\beta}{r+\beta}}\sqrt{X_t(1-X_t)}$, and when the two parameters are approximately equal then the equilbrium is around 1/2. Replacing $X_t = 1/2$, we get the noise scales with $\sqrt{r} \approx \sqrt{\beta}$ by assumption, so the noise agrees. Therefore while they will likely agree to a good approximation for many parameter values, the new model has behavior which avoids this pathological behavior.

Further, while the old formulation does not have a simple closed form solution, the approximation of the new model is known to have a closed form solution with stationary beta distributions. This implies that random testing of a group of individuals, seen as a Binomial random variable for the number infected when conditioned on the prevalence, will instead be modeled as a BetaBinomial random variable when taking into account the uncertainty associated with prevalence. The beta stationary distribution can also be used to construct an explicit likelihood function for the parameters.

Equilibrium Distribution

At equilibrium, our estimate of the state of the system is given by the stationary solution of the Kolmogorov Forward Equation implied by the SDE (see Appendix 5). The equilibrium distribution is given by

$$\text{Beta}\left(\frac{N}{4(1-x^*)}, \frac{N}{4x^*}\right)$$

where x^* is our prevalence at equilibrium. Note from the ODE for x that $x^* = \frac{h}{h+r}$. We can further see that the population size determines how tightly the solutions are distributed around the expected value; larger population sizes imply a tighter distribution around the mean.

Statistical Algorithm

Using a quasiequilibrium assumption, we are able to reduce the number of degrees of freedom in our model to just two by assuming X_t is *transiently* beta distributed. The two degrees of freedom are determined by the first two transient moments, which follow

$$\begin{pmatrix} \frac{dE[X_t]}{dt} \\ \frac{dE[X_t^2]}{dt} \end{pmatrix} = \begin{pmatrix} h \\ 0 \end{pmatrix} + \begin{pmatrix} -(h+r) & 0 \\ 2\left(h + \frac{4rh}{N(r+h)}\right) & -2\left(h + r + \frac{4r\beta}{N(r+h)}\right) \end{pmatrix} \begin{pmatrix} E[X_t] \\ E[X_t^2] \end{pmatrix}$$

These are linear ODEs which can be solved to give

$$\begin{aligned} E[X_{t_{i+1}}] &= E[X_{t_i}]\rho_i + \mu_i(1-\rho_i) \\ E[X_{t_{i+1}}^2] &= \frac{\mu+\omega}{1/2+\omega}(E[X_{t_i}]-\mu_i)\rho_i(1-\rho_i^{1+2\omega_i}) + \mu_i\frac{\mu_i+\omega_i}{1+\omega_i}(1-\rho_i^{2(1+\omega_i)}) + E[X_{t_i}^2]\rho_i^{2(1+\omega_i)} \\ \mu_i &= \frac{h(t_i)}{h(t_i)+r} \\ \rho_i &= e^{-(h(t_i)+r)(t_{i+1}-t_i)} \\ \omega_i &= \frac{4\mu_{t_i}(1-\mu_{t_i})}{N} \end{aligned}$$

Physically, μ is the instantanous equilibrium value for the system, ρ gives the average rate at which the system is approaching the equilibrium deterministically, and ω gives a measure of the degree of uncertainty that pushes a trajectory off its expected path. Note that as population sizes become large, this term becomes small and the equation acts nearly deterministically.

We can further apply the method of moments to transform the difference equations into one which takes initial beta parameters (α_0, β_0) and quickly computes all future beta parameters in terms of the parameters of the mechanistic model:

$$\alpha_i = \left(\frac{E[X_{t_i}](1 - E[X_{t_i}])}{V(X_{t_i})} - 1 \right) E[X_{t_i}]$$
$$\beta_i = \left(\frac{E[X_{t_i}](1 - E[X_{t_i}])}{V(X_{t_i})} - 1 \right) (1 - E[X_{t_i}])$$

where

$$V(X_{t_i}) = E[X_{t_i}^2] - E[X_{t_i}]^2$$

This difference equation allows us to assume parameters for the initial beta distributed density (α_0, β_0), then immediately compute the subsequent distributions parameterized by (α_i, β_i). Therefore we have the approximate transition densities defined through this linear difference equation. We can now integrate this beta distribution of the state out to obtain posterior predictive distributions for the observations, which conditional on the state follow a binomial distribution:

$$\begin{aligned} Y_{t_i} | X_{t_i} &\sim \text{Binom}(n, X_{t_i}) \\ X_{t_i} &\sim \text{Beta}(\alpha_{t_i}, \beta_{t+i}) \\ \implies Y_{t_i} &\sim \text{BetaBinom}(n, \alpha_{t_i}, \beta_{t_i}) \end{aligned}$$

That is, our posterior predictive distributions are beta-binomially distributed with the full distribution across time being fully determined by the model parameters Θ and the initial state α_0, β_0, which will be adjoined to Θ as additional parameters to be fit.

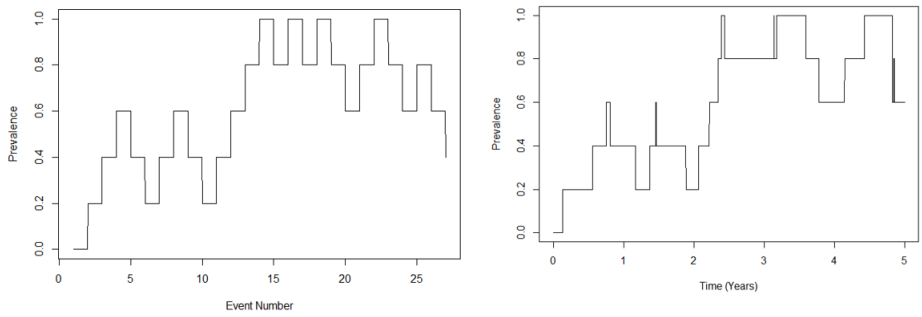


Figure 5.2: Example of random time change of trajectory. A trajectory of the embedded Markov chain (left) and the random time changed version (right).

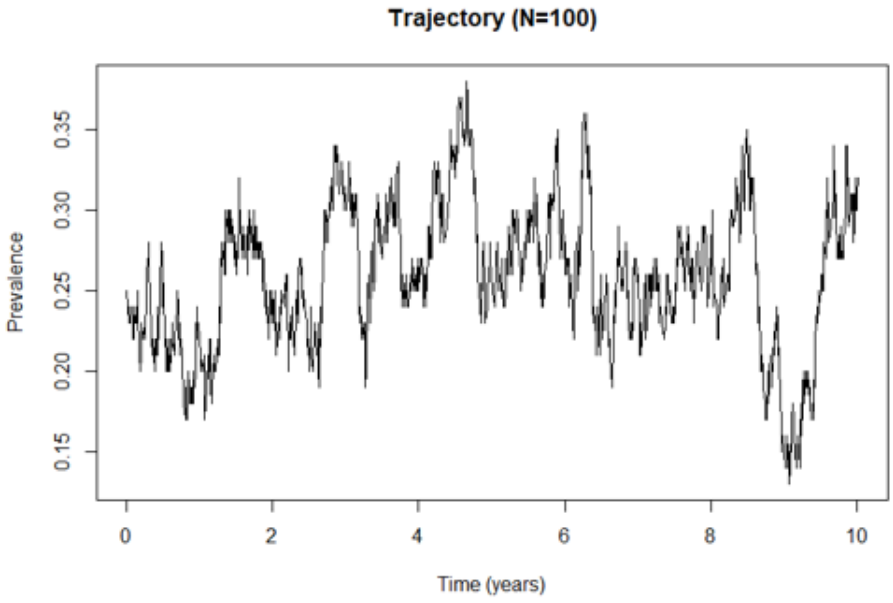


Figure 5.3: Example trajectory of a single population trajectory of prevalence in a population of 100 for 10 years under the SIS model.

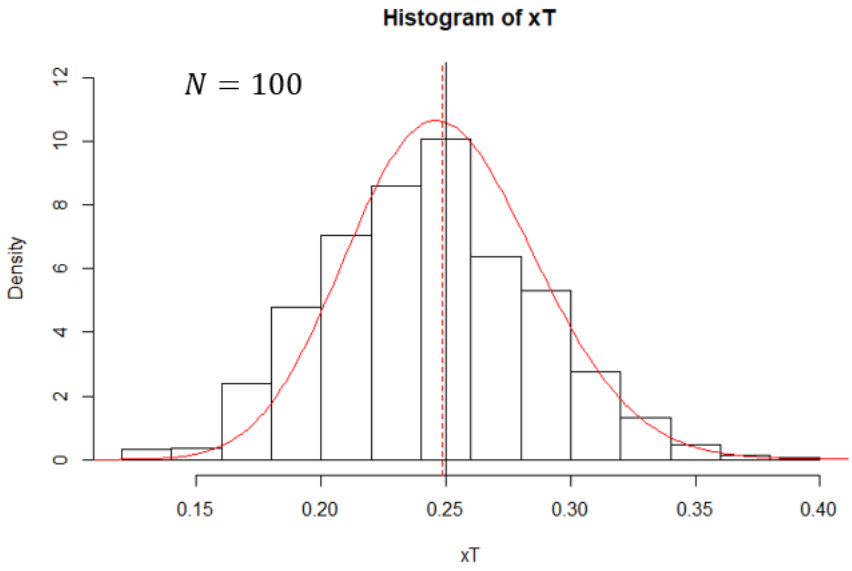


Figure 5.4: Histogram of many population trajectories of prevalence in a population of 100 at the end of 10 years. We can see the beta distribution computed in the above text fits very closely with the simulated output.

5.3.2 Superinfection Dynamics and the Negative-Binomial Model

Model Definition

Now we can consider the more complex situation in which infections accumulate and clear independently, corresponding to the classic M/M/∞ queueing model. The value followed here is not just prevalence, but the distribution of the number of infections in individuals in a population given infections initiate and are cleared independently of one another. An individual therefore is not bounded to one of two states, but a countable state space. If the number of infections in individual i at time t is denoted $M_{i,t}$, then the average number of infections per person in the population is

$$\bar{M}_t := \frac{1}{N} \sum_{i=1}^{N} X_{i,t}$$

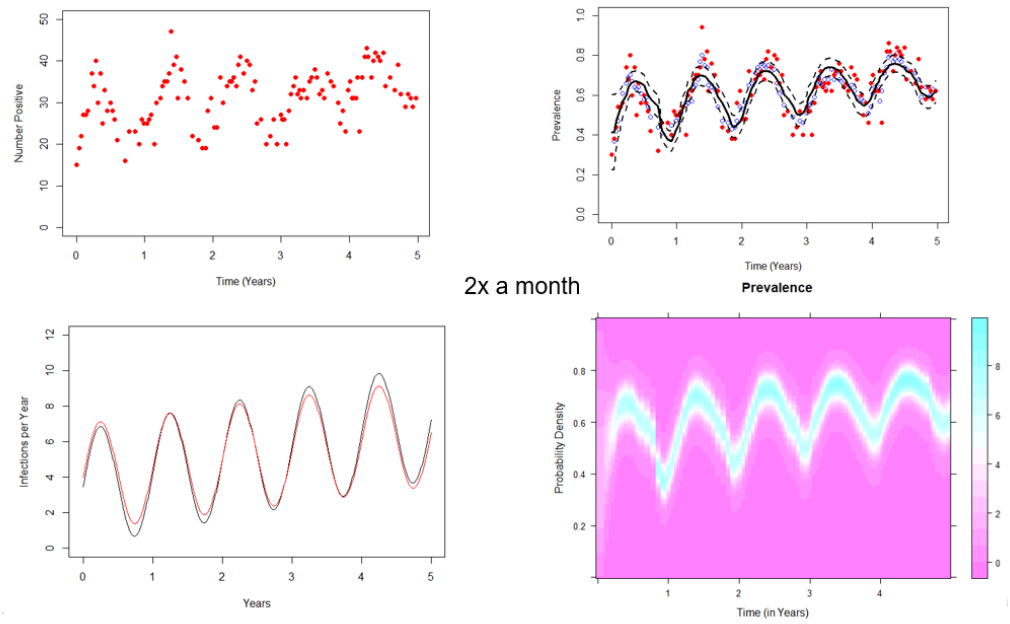


Figure 5.5: Example of algorithm fitting to sinusoidal exposure rate with a linear trend. Top left shows the simulated data, sampled twice a month with measurement uncertainty; top right shows the mean predicted fit to prevalence (solid black) with the latent true values plotted in blue, dotted line shows $\pm$ 1 standard deviation; bottom left shows the true (black) and estimated (red) force of infection; and the bottom right shows a heat plot of the posterior estimate of prevalence values.

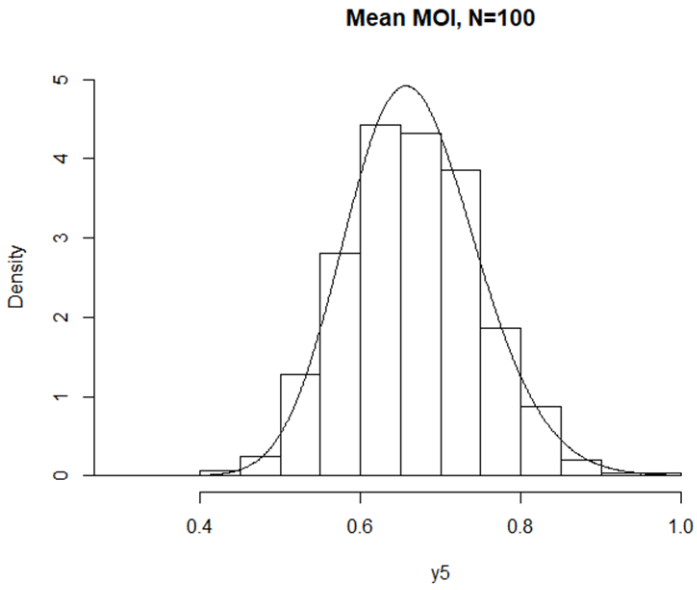


Figure 5.6: The stationary distribution of the superinfection SDE; histogram from simulation, curve predicted from Kolmogorov forward equation.

As before, we will want to pass to the embedded discrete time Markov chain in order to apply the same method as before. In a population of size N, the time before the next infection initiates is exponentially distributed with rate

$$\begin{aligned} E[\Delta t_I | \bar{M}_t] &\sim \Delta t_i \\ &\sim Exp(hN) \end{aligned}$$

as they are all independent and all individuals are susceptible to infection in this model. The time before the next recovery given people have a Poisson number of infections with average value $\bar{M}_t$, the time until the next recovery for a person chosen at random with $M_{t,i}$ infections is

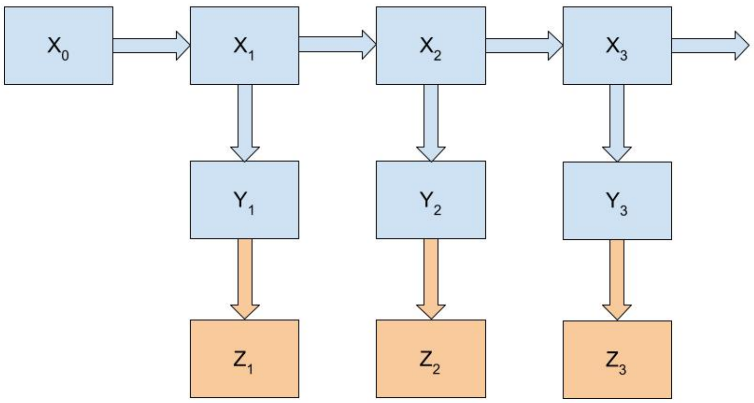


Figure 5.7: The expanded state space schematic for the model with superinfection. X_i represents the state of the mean MOI, Y_i is the posterior predictive of the MOI, and Z_i is the observed prevalence.

$$E[\Delta t_R | M_{t,i}] \sim Exp(rM_{t,i})$$

For a population of size N, the next time to recovery is

$$\begin{aligned}
E[\Delta t_R | M_{t,i}, i = 1, ..., N] &\sim \min_{i=1,...,N} Exp(rM_{t,i}) \\
&\sim \min_{i=1,...,N} Exp(r \sum_{i=1}^{N} M_{t,i}) \\
&\sim Exp(r \sum_{i=1}^{N} M_{t,i}) \\
&\sim Exp(rN \frac{1}{N} \sum_{i=1}^{N} M_{t,i}) \\
&\sim Exp(rN\bar{M}_t) \\
&= E[\Delta t_R | \bar{M}_t]
\end{aligned}$$

The next time to either a recovery or infection is given by

$$\begin{aligned}
E[\Delta t_i | \bar{M}_t] &= \min \left\{ E[\Delta t_I | \bar{M}_t], E[\Delta t_R | \bar{M}_t] \right\} \\
&\sim Exp(N(h + r\bar{M}_t))
\end{aligned}$$

Next, we can look at the state transitions. Given an event occurred, the probability that the event was an infection is

$$P\left(\bar{M}_{i+1} = \bar{M}_i + \frac{1}{N}\middle|\bar{M}_i\right) = P(E[\Delta t_I|\bar{M}_i] < E[\Delta t_R|\bar{M}_i])$$
$$= \frac{h}{h + r\bar{M}_i}$$

and similarly,

$$P\left(\bar{M}_{i+1} = \bar{M}_i - \frac{1}{N}\middle|\bar{M}_i\right) = \frac{r\bar{M}_i}{h + r\bar{M}_i}$$

as these are the only two transitions that can occur consecutively. Therefore the expected change is

$$\begin{aligned} E[\Delta\bar{M}_i|\bar{M}_i] &:= E[\bar{M}_{i+1} - \bar{M}_i] \\ &= \left(\frac{1}{N}\right) P\left(\bar{M}_{i+1} = \bar{M}_i + \frac{1}{N}\middle|\bar{M}_i\right) - \frac{1}{N} P\left(\bar{M}_{i+1} = \bar{M}_i - \frac{1}{N}\middle|\bar{M}_i\right) \\ &= \frac{h - r\bar{M}_i}{N[h + r\bar{M}_i]} \end{aligned}$$

Next, we can determine the variance in this change. As in the simple infection case, each change has the same maginitude of $\frac{1}{N}$ and so

$$E[(\Delta\bar{M}_i)^2|\bar{M}_i] = \frac{1}{N^2}$$

and therefore the conditional variance is

$$\begin{aligned} Var(\Delta\bar{M}_i|\bar{M}_i) &= E[(\Delta\bar{M}_i)^2|\bar{M}_i] - E[\Delta\bar{M}_i|\bar{M}_i]^2 \\ &= \frac{1}{N^2} - \left(\frac{h - r\bar{M}_i}{N[h + r\bar{M}_i]}\right)^2 \\ &= \frac{1}{N^2}\left(\frac{[h + r\bar{M}_i]^2 - (h - r\bar{M}_i)^2}{[h + r\bar{M}_i]^2}\right) \\ &= \frac{4hr\bar{M}_i}{N^2[h + r\bar{M}_i]^2} \end{aligned}$$

As before, we will define our normalized variable

$$\begin{aligned} \bar{Z}_i &= \frac{\Delta\bar{M}_i - E[\Delta\bar{M}_i|\bar{M}_i]}{\sqrt{Var(\Delta\bar{M}_i|\bar{M}_i)}} \\ &= \frac{\Delta\bar{M}_i - \frac{h - r\bar{M}_i}{N[h + r\bar{M}_i]}}{\sqrt{\frac{4hr\bar{M}_i}{N^2[h + r\bar{M}_i]^2}}} \end{aligned}$$

Setting our timescale to be $\Delta t_i = \frac{1}{N[h + rM_i]}$, we can simplify the expression to be

$$\bar{Z}_i = \frac{\Delta\bar{M}_i - [h - r\bar{M}]\Delta t}{2\sqrt{hr\bar{M}_i}\Delta t}$$

$$dm_t = r[\frac{h(t)}{r} - Z_t]dt + \sqrt{\frac{2r}{N}}\sqrt{m_t}dW_t$$

Equilibrium Distribution

The equilibrium distribution can again be found by considering the Kolmogorov forward equation and solving for the stationary solution (see Appendix 5):

Statistical Algorithm

Defining $\mu = \frac{h}{r}$, $\rho = e^{-rt}$, we have

$$E[\bar{M}_{t+s}] = E[\bar{M}_s]\rho + \mu(1-\rho)$$
$$E[\bar{M}^2_{t+s}] = \left(2\mu + \frac{1}{N}\right)\rho(1-\rho)E[\bar{M}_s] + \mu\left(\mu + \frac{1}{N}\right)(1-\rho)^2 + \rho^2 E[\bar{M}^2_s]$$

The MOI will follow a Poisson distribution with mean $\bar{M}_t$. $\bar{M}_t$ follows an SDE, whose stationary distribution is Gamma. Given this Gamma distribution with shape α and rate β, the resulting MOI distribution is NB(α, $\frac{1}{1+\beta}$). Above we have the first and second moments. Applying the MoM relations,

$$\alpha = \frac{E[\bar{M}_t]^2}{Var(\bar{M}_t)}$$
$$\beta = \frac{E[\bar{M}_t]}{Var(\bar{M}_t)}$$

so we obtain a posterior NB distribution:

$$NB\left(\frac{E[\bar{M}_t]^2}{Var(\bar{M}_t)}, \frac{Var(\bar{M}_t)}{Var(\bar{M}_t) + E[\bar{M}_t]}\right)$$

The probability of having at least one infection for each individual is then

$$
\begin{aligned}
P(M_t \geq 1) &= 1 - P(M_t = 0) \\
&= 1 - \left(1 - \frac{Var(\bar{M}_t)}{Var(\bar{M}_t) + E[\bar{M}_t]}\right)^{\frac{E[\bar{M}_t]^2}{Var(\bar{M}_t)}} \\
&= 1 - \left(\frac{E[\bar{M}_t]}{Var(\bar{M}_t) + E[\bar{M}_t]}\right)^{\frac{E[\bar{M}_t]^2}{Var(\bar{M}_t)}}
\end{aligned}
$$

Our measurement equation is again a binomial distribution with the probability of detection being equal to the prevalence, that is

$$
Y_t \sim \text{Binom}\left(N_{\text{sample}},\ 1 - \left(\frac{E[\bar{M}_t]}{Var(\bar{M}_t) + E[\bar{M}_t]}\right)^{\frac{E[\bar{M}_t]^2}{Var(\bar{M}_t)}}\right)
$$

5.3.3 SIR Dynamics and a Multivariate Gamma Distribution

Model Definition

$$
\begin{pmatrix} dS_\tau \\ dI_\tau \end{pmatrix} = \begin{pmatrix} -R_0 S_\tau I_\tau \\ R_0 S_\tau I_\tau - I_\tau \end{pmatrix} d\tau + \sqrt{\frac{R_0 S_\tau I_\tau}{5N(R_0 S_\tau + 1)}} \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} dW_{1,\tau} \\ dW_{2,\tau} \end{pmatrix}
$$

We can compare this to the classic stochastic derivation from the Poisson process interpretation:

$$
\begin{pmatrix} dS_\tau \\ dI_\tau \end{pmatrix} = \begin{pmatrix} -R_0 S_\tau I_\tau \\ R_0 S_\tau I_\tau - I_\tau \end{pmatrix} d\tau + \begin{pmatrix} -\sqrt{\beta SI} & 0 \\ \sqrt{\beta SI} & -\sqrt{\gamma I} \end{pmatrix} \begin{pmatrix} dW_{1,\tau} \\ dW_{2,\tau} \end{pmatrix}
$$

Interestingly, this formulation may make approximate analysis of outcomes of interest such as the distribution of the asymptotic proportion infected over the course of the epidemic.

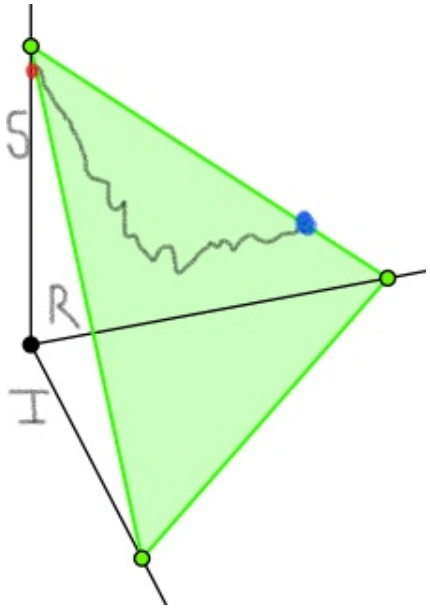


Figure 5.8: A single trajectory of the SDE for prevalence in an SIR model. Each trajectory requires $S + I + R = 1$ and all are nonnegative, so the state space is given by the 2-simplex plotted in green.

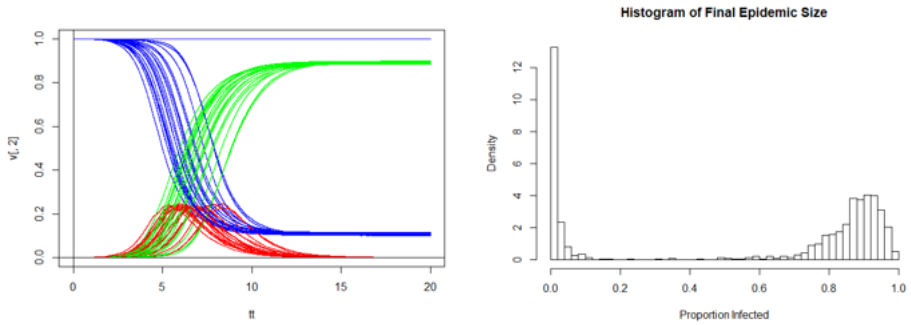


Figure 5.9: Simulation of many trajectories of the SIR SDE model (Left) and a histogram of the total density of individuals infected at the end of the epidemic (Right)

Figure 5.8 shows an example of a trajectory on the simplex. Any point on the interior of the simplex will move toward the I axis before turning and landing on the $S - R$ axis as the disease dies out. This acts as an absorbing boundary for the SDE, with the marginal density on the boundary asymptotically approaching the total epidemic size density of interest.

5.3.4 The Shape of Noise

Here we have shown that the structure of the noise term in the SDE for the functionals of the system, such as the proportion infected, is as intimately tied to the assumed mechanism of the underlying stochastic process as the drift term. We have assumed here that all of the stochasticity is internal, due to randomness inherent in the infection and recovery process. Other studies involving SDEs often focus on the effect of external environmental noise in a system, for example small perturbations due to weather or human movement and demography, which result in an additive noise term which is either unscaled with the state or approximately scaled according to the underlying Poisson transitions.

However, adding unscaled noise to a model can result in nonsensical answers, such as negative prevalence by adding white noise to a standard SIS model. Therefore it is reasonable to believe the scaling derived in the methods here is in some sense a natural choice; the noise term resulted in a beta density over the unit interval for the SIS model, and a gamma density over the real numbers for the $M/M/\infty$ model. As environmental noise is assumed to not approach zero as the population size gets large, we can suggest a model of the form

$$dp_t = [h(1 - p_t) - rp_t]dt + [\sigma_{\text{env}} + \mathcal{O}(\frac{1}{N})\sqrt{p_t(1 - p_t)}dW_t$$

where σ_{env} is a residual environmental noise not reduced with sample size. Therefore this method suggests the mechanism will suggest the shape the noise takes as it propagates through a model.

5.4 Discussion

Just like plucking a guitar string at different parts along its length does not result in the same tone due to the inherent structure of the oscillator, external noise propagating through a mechanistic model does not necessarily result in noise of the same magnitude everywhere. Importantly, uncertainty is quelled in the SIS model when populations are close to the boundary because in the underlying stochastic process of individuals switching between infected and uninfected states, the degrees of freedom in this model are maximized when $x = 1/2$ and minimized when $x \in \{0, 1\}$. This suggests that 1. the state space itself has a sort of geometry that must be taken into account before considering how noise is added to the model; and 2. 'fitting an ODE to data' is an ill-posed problem in this setting, as many different stochastic models with distinct uncertainty profiles may result in the same mean behavior described by that ODE.

Here we have developed SDE models for key descriptive functionals such as prevalence directly from relatively simple stochastic processes of infection. This is distinct from previous methods which tend to consider external noise as adding a white noise process without considering the structure of the state space. Here the uncertainty is scaled appropriately to prevent unrealistic behavior such as negative prevalence. Further, quasi-stationary assumptions made here simplify the transition equations significantly, allowing for direct solution and the development of the algorithms presented here.

There is also the distinct advantage in this method to apply different sampling strategies in case people aren't chosen randomly for testing, which is typical for data streams such as clinical incidence. This will be a biased sample, but may still be able to be reconciled with prevalence survey data by considering the joint likelihood of both observations with different observation functions. This is a potentially major advancement to combine different kinds of observations under a coherent framework. Additionally, sample sizes and sampling frequencies don't need to be uniform even within the same data stream. This allows further flexibility with the structure of studies.

5.5 Conclusions

Here we have introduced a simulation-free method for estimating the force of infection and rate of recovery under different mechanistic assumptions. The work presented here represents a push toward a consideration of the interplay of mechanism and uncertainty inherent in measuring, modeling, and predicting epidemiological systems.

BIBLIOGRAPHY

[1] Ross R. The prevention of malaria. London John Murray; 1911.

[2] Ross R. Some quantitative studies in epidemiology. Nature. 1911;87:466–467.

[3] Macdonald G. The analysis of the sporozoite rate. Trop Dis Bull. 1952;.

[4] Cohen JM, Smith DL, Cotter C, et al. Malaria resurgence: a systematic review and assessment of its causes. Malaria Journal. 2012;.

[5] Smith DL, Cohen JM, Chiyaka C, et al. A sticky situation: the unexpected stability of malaria elimination. Philos Trans R Soc Lond B Biol Sci. 2013;doi:10.1098/rstb.2012.0145.

[6] Fine PEM. A Problem in Formulating a Problem. Tropical Diseases Bulletin. 1975;72:475–488.

[7] Dietz K, Molineaux L, Thomas A. A malaria model tested in the African savannah. Bulletin of the World Health Organization. 1974;50:347–357.

[8] Nasell I. Hybrid models of tropical infections. 1985; p. 206.

[9] Mortality G, of Death Collaborators C. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. 2016;388:1459–1544. doi:10.1016/S0140-6736(16)31012-1.

[10] Portugal S, Drakesmith H, Mota M. Superinfection in malaria: Plasmodium shows its iron will. EMBO Reports. 2011;12. doi:10.1038/embor.2011.213.

[11] Macdonald G. The analysis of infection rates in diseases in which superinfection occurs. Tropical Diseases Bulletin. 1950;47:907–915.

[12] Rosà R, Pugliese A. Aggregation, stability, and oscillations in different models for host-macroparasite interactions. Theoretical Population Biology. 2002;61:319–334. doi:10.1006/TPBI.2002.1575.

[13] Picard P. A semi-hybrid model for malaria with limited superinfection. Advances in Applied Probability. 1998;30:1027–1057. doi:10.1239/aap/1035228206.

[14] Milligan PJM, Downham DY. Models of superinfection and acquired immunity to multiple parasite strains. Journal of Applied Probability. 1996;33:915–932. doi:10.2307/3214973.

[15] Nowak MA, May RM. Superinfection and the evolution of parasite virulence. Proceedings of the Royal Society of London Series B: Biological Sciences. 1994;255:81–89. doi:10.1098/rspb.1994.0012.

[16] Weiss GH, Aron JL. Note on the formulation of a stochastic model of superinfection. Mathematical Biosciences. 1983;67:213–223. doi:10.1016/0025-5564(83)90101-3.

[17] Klein EY, Smith DL, Laxminarayan R, Levin S. Superinfection and the evolution of resistance to antimalarial drugs. Proceedings Biological sciences. 2012;279:3834–42. doi:10.1098/rspb.2012.1064.

[18] Nedelman J. Inoculation and recovery rates In the malaria model of Dietz, Molineaux, and Thomas. Mathematical Biosciences. 1984; p. 209–233.

[19] Smith DL, Battle KE, Hay SI, Barker CM, Scott TW, McKenzie FE. Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. PLoS Pathogens. 2012;8:e1002588. doi:10.1371/journal.ppat.1002588.

[20] Recker M, Nee S, Bull PC, Kinyanjui S, Marsh K, Newbold C, et al. Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. Nature. 2004;429:555–558. doi:10.1038/nature02486.

[21] Klein EY, Graham AL, Llinás M, Levin S. Cross-reactive immune responses as primary drivers of malaria chronicity. Infection and immunity. 2014;82:140–51. doi:10.1128/IAI.00958-13.

[22] Doolan DL, Dobaño C, Baird JK. Acquired immunity to malaria. Clinical microbiology reviews. 2009;22:13–36, Table of Contents. doi:10.1128/CMR.00025-08.

[23] Ashley EA, White NJ. The duration of Plasmodium falciparum infections. Malaria Journal. 2014;13:500. doi:10.1186/1475-2875-13-500.

[24] Zhong D, Koepfli C, Cui L, Yan G. Molecular approaches to determine the multiplicity of Plasmodium infections. Malaria Journal. 2018;17:172. doi:10.1186/s12936-018-2322-5.

[25] Kumar CS, Nair BU. An alternative hyper-Poisson distribution. Statistica. 2012;72:357–369. doi:10.6092/issn.1973-2201/3652.

[26] Guillemin F, Simonian A. Transient Characteristics of an M / M / Infinity system Author (s): Fabrice Guillemin and Alain Simonian Source : Advances in Applied Probability , Vol . 27 , No . 3 (Sep ., 1995), pp . 862-888. 1995;27:862–888.

[27] Smith DL, McKenzie FE. Statics and dynamics of malaria infection in Anopheles mosquitoes. Malaria Journal. 2004;3:13. doi:10.1186/1475-2875-3-13.

[28] Smith NR, Trauer JM, Gambhir M, Richards JS, Maude RJ, Keith JM, et al. Agent-based models of malaria transmission: a systematic review. Malaria Journal. 2018;17:299. doi:10.1186/s12936-018-2442-y.

[29] Burkot TR, Graves PM, Cattan JA, Wirtz RA, Gibson FD. The efficiency of sporozoite transmission in the human malarias, Plasmodium falciparum and P. vivax. Bulletin of the World Health Organization. 1987;65:375–80.

[30] Koita OA, Doumbo OK, Ouattara A, Tall LK, Konaré A, Diakité M, et al. False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. Am J Trop Med Hyg. 2012;86:194–198. doi:10.4269/ajtmh.2012.10-0665.

[31] Weiss DJ, Lucas TCD, Nguyen M, Nandi AK, Bisanzio D, Battle KE, et al. Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study. The Lancet. 2019;394:322–331. doi:10.1016/S0140-6736(19)31097-9.

[32] Smith T, Schellenberg JA, Hayes R. Attributable fraction estimates and case definitions for malaria in endemic areas. Statistics in Medicine. 1994;13:2345–2358. doi:10.1002/sim.4780132206.

[33] Dalrymple U, Cameron E, Bhatt S, Weiss DJ, Gupta S, Gething PW. Quantifying the contribution of Plasmodium falciparum malaria to febrile illness amongst African children. eLife. 2017;6. doi:10.7554/elife.29198.

[34] James SP, Nicol WD, Shute PG. A study of induced malignant tertian malaria. Proceedings of the Royal Society of Medicine. 1932; p. 755–759.

[35] Eyles DE, Young MD. The duration of untreated or inadequately treated Plasmodium falciparum infections in the human host. Journal of the National Malaria Society. 1951;.

[36] Sama W, Killeen G, Smith T. Estimating the duration of *Plasmodium falciparum* infection from trials of indoor residual spraying; 2004.

[37] Eckhoff P. P. falciparum infection durations and infectiousness Are shaped by antigenic variation and innate and adaptive host immunity in a mathematical model. PLoS ONE. 2012;7:e44950. doi:10.1371/journal.pone.0044950.

[38] McQueen PG, McKenzie FE. Age-structured red blood cell susceptibility and the dynamics of malaria infections. Proceedings of the National Academy of Sciences of the United States of America. 2004;101:9161–9166. doi:10.1073/pnas.0308256101.

[39] Churcher TS, Trape JF, Cohuet A. Human-to-mosquito transmission efficiency increases as malaria is controlled. Nature Communications. 2015;6. doi:10.1038/ncomms7054.

[40] Ouédraogo AL, Gonçalves BP, Gnémé A, Wenger EA, Guelbeogo MW, Ouédraogo A, et al. Dynamics of the human infectious reservoir for malaria determined by mosquito feeding assays and ultrasensitive malaria diagnosis in Burkina Faso. Journal of Infectious Diseases. 2016;213:90–99. doi:10.1093/infdis/jiv370.

[41] Collins WE, Jeffery GM. A retrospective examination of mosquito infection on humans infected with *Plasmodium falciparum*. The American Journal of Tropical Medicine and Hygiene. 2003;68:366–371. doi:10.4269/AJTMH.2003.68.366.

[42] Bretscher MT, Maire N, Chitnis N, Felger I, Owusu-Agyei S, Smith T. The distribution of Plasmodium falciparum infection durations. Epidemics. 2011;3:109–118. doi:10.1016/j.epidem.2011.03.002.

[43] Hawking F, Wilson ME, Gammage K. Evidence for cyclic development and short-lived maturity in the gametocytes of *Plasmodium falciparum*. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1971;65:549–559. doi:10.1016/0035-9203(71)90036-8.

[44] Churcher TS, Sinden RE, Edwards NJ, Poulton ID, Rampling TW, Brock PM, et al. Probability of transmission of malaria from mosquito to human is regulated by mosquito parasite density in naïve and vaccinated hosts. PLoS Pathogens. 2017;13. doi:10.1371/journal.ppat.1006108.

[45] Collins WE, Jeffery GM, Roberts JM. A retrospective examination of the effect of fever and microgametocyte count on mosquito infection on humans infected with *Plasmodium vivax*. American Journal of Tropical Medicine and Hygiene. 2004;70:638–641. doi:10.4269/ajtmh.2004.70.638.

[46] Joanny F, Löhr SJ, Engleitner T, Lell B, Mordmüller B. Limit of blank and limit of detection of Plasmodium falciparum thick blood smear microscopy in a routine setting in Central Africa. Malaria Journal. 2014;13. doi:10.1186/1475-2875-13-234.

[47] Anderson RM, Gordon DM, Crawley MJ, Hassell MP. Variability in the abundance of animal and plant species. Nature. 1982;296:245–248. doi:10.1038/296245a0.

[48] Tadesse FG, Meerstein-Kessel L, Gonçalves BP, Drakeley C, Ranford-Cartwright L, Bousema T. Gametocyte sex ratio: the key to understanding *Plasmodium falciparum* transmission?; 2019.

[49] Yang CK, Brauer F. Calculation of R0 for age-of-infection models. Mathematical Biosciences and Engineering. 2003;5:79–88. doi:10.1063/1.1478825.

[50] Portugal S, Moebius J, Skinner J, Doumbo S, Doumtabe D, Kone Y, et al. Exposure-dependent control of malaria-induced inflammation in children. PLoS Pathogens. 2014;10. doi:10.1371/journal.ppat.1004079.

[51] Ruktanonchai NW, Smith DL, Leenheer PD. Parasite sources and sinks in a patched Ross–Macdonald malaria model with human and mosquito movement: implications for control. Mathematical Biosciences. 2016;279:90–101. doi:10.1016/j.mbs.2016.06.012.

[52] Taylor LR. Aggregation, variance and the mean. Nature. 1961;189:732–735. doi:10.1038/189732a0.

[53] Crompton PD, Moebius J, Portugal S, Waisberg M, Hart G, Garver LS, et al. Malaria immunity in man and mosquito: insights into unsolved mysteries of a deadly infectious disease. Annual Review of Immunology. 2014;32:157–187. doi:10.1146/annurev-immunol-032713-120220.

[54] Rodriguez-Barraquer I, Arinaitwe E, Jagannathan P, Boyle MJ, Tappero J, Muhindo M, et al. Quantifying heterogeneous malaria exposure and clinical protection in a cohort of Ugandan children. The Journal of infectious diseases. 2016;214:1072–1080. doi:10.1093/INFDIS/JIW301.

[55] Molineaux L. The epidemiology of human malaria as an explanation of its distribution including some implications for its control. Wensdorfer W, McGregor I, editors. Churchill Livingstone; 1988.

[56] Henry JM. A hybrid model for the effects of treatment and demography on malaria superinfection. Journal of Theoretical Biology. 2020;491. doi:10.1016/j.jtbi.2020.110194.

[57] Reiner RC, Perkins TA, Barker CM, Niu T, Chaves LF, Ellis AM, et al. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970-2010. Journal of the Royal Society, Interface. 2013;10. doi:10.1098/RSIF.2012.0921.

[58] Rodriguez-Barraquer I, Arinaitwe E, Jagannathan P, Kamya MR, Rosenthal PJ, Rek J, et al. Quantification of anti-parasite and anti-disease immunity to malaria as a function of age and exposure. eLife. 2018;7. doi:10.7554/ELIFE.35832.

[59] Smith T, Killeen GF, Maire N, Ross A, Molineaux L, Tediosi F, et al. Mathematical modeling of the impact of malaria vaccines on the clinical epidemiology and natural history of *Plasmodium falciparum* malaria: overview. The American journal of tropical medicine and hygiene. 2006;75:1–10. doi:10.4269/AJTMH.2006.75.2$_SUPPL$.0750001.

[60] Griffin JT, Hollingsworth TD, Okell LC, Churcher TS, White M, Hinsley W, et al. Reducing Plasmodium falciparum malaria transmission in Africa: a model-based evaluation of intervention strategies. PLoS medicine. 2010;7. doi:10.1371/JOURNAL.PMED.1000324.

[61] Bailey NTJ. The biomathematics of malaria. 1982; p. 210.

[62] Leemis LM, Park SK. Discrete-event simulation : a first course. 2006; p. 604.

[63] Ross BMR, Thomson D. Some enumerative studies on malarial fever. Proceedings of the Royal Society of London Series B, Containing Papers of a Biological Character. 1910;83:159–173. doi:10.1098/RSPB.1910.0075.

[64] Arzola C, Río JD, Earle M. Observations on the Course of Naturally Acquired Malaria in Puerto Rico. 1939;.

[65] Yap XZ, McCall MBB, Sauerwein RW. Fast and fierce versus slow and smooth: Heterogeneity in immune responses to *Plasmodium* in the controlled human malaria infection model. Immunological Reviews. 2020;293:253–269. doi:10.1111/IMR.12811.

[66] Smith DL, Guerra CA, Snow RW, Hay SI. Standardizing estimates of the *Plasmodium falciparum* parasite rate. Malaria Journal. 2007;6:1–10. doi:10.1186/1475-2875-6-131/FIGURES/6.

[67] Fairley BNH. Sidelights on malaria in man obtained by subinoculation experiments. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1947;40:621–676. doi:10.1016/0035-9203(47)90025-4.

[68] Serfozo RF. Functions of Semi-Markov Processes. SIAM J Appl Math. 1971;.

[69] Bjørnstad ON, Grenfell BT. Noisy clockwork: Time series analysis of population fluctuations in animals. Science. 2001;doi:293:638–643.

[70] King ELIMP A A, Bouma MJ. Inapparent infections and cholera dynamics. Nature. 2008;doi:454:877–880.

[71] A A King FMPR M Domenech de Celles. Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to ebola. Proceedings of the Royal Society of London. 2015;doi:282:20150347.

[72] Robert F, Ntoumi F, Angel G, Candito D, Rogier C, Fandeur T, et al. Extensive genetic diversity of Plasmodium falciparum isolates collected from patients with severe malaria in Dakar, Senegal. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1996;90:704–711. doi:10.1016/S0035-9203(96)90446-0.

[73] Martcheva M, Thieme HR. Progression age enhanced backward bifurcation in an epidemic model with super-infection. Journal of Mathematical Biology. 2003;46:385–424. doi:10.1007/s00285-002-0181-7.

[74] Aron JL, May RM. The population dynamics of malaria. Springer US; 1982.

[75] Ouwens IMD, Lens CE, Fiolet ATL, Ott A, Koehler PJ, Kager PA, et al. Malaria fever therapy for general paresis of the insane: a historical cohort study. European neurology. 2017;78:56–62. doi:10.1159/000477900.

[76] Sama W, Dietz K, Smith T. Distribution of survival times of deliberate *Plasmodium falciparum* infections in tertiary syphilis patients. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2006;100:811–816. doi:10.1016/j.trstmh.2005.11.001.

[77] Afrane YA, Zhou G, Githeko AK, Yan G. Clinical malaria case definition and malaria attributable fraction in the highlands of western Kenya. Malaria Journal. 2014;13:405. doi:10.1186/1475-2875-13-405.

[78] Lee K, Small DS. Estimating the malaria attributable fever fraction accounting for parasites being killed by fever and measurement error. Journal of the American Statistical Association. 2019;114:79–92. doi:10.1080/01621459.2018.1469989.

[79] Guerra CA, Kang SY, Citron DT, Hergott DEB, Perry M, Smith J, et al. Human mobility patterns and malaria importation on Bioko Island. Nature Communications. 2019;10:1–10. doi:10.1038/s41467-019-10339-1.

[80] Buchwald AG, Sorkin JD, Sixpence A, Chimenya M, Damson M, Wilson ML, et al. Association between age and *Plasmodium falciparum* infection dynamics. vol. 188. Oxford University Press; 2019. p. 169–176.

[81] Lamikanra AA, Brown D, Potocnik A, Casals-Pascual C, Langhorne J, Roberts DJ. Malarial anemia: of mice and men. Blood. 2007;110:18–28. doi:10.1182/blood-2006-09-018069.

[82] Kim A, Nemeth E. New insights into iron regulation and erythropoiesis; 2015.

[83] White NJ. Anaemia and malaria 11 Medical and Health Sciences 1108 Medical Microbiology 11 Medical and Health Sciences 1103 Clinical Sciences; 2018.

[84] Wu SL, C HMS, Henry JM, Citron DT, Zhang Q, Compton K, et al.. Vector bionomics and vectorial capacity as emergent properties of mosquito behaviors and ecology; 2019.

[85] Conlan AJK, Rohani P, Lloyd AL, Keeling M, Grenfell BT. Resolving the impact of waiting time distributions on the persistence of measles. Journal of the Royal Society Interface. 2010;7:623–640. doi:10.1098/rsif.2009.0284.

[86] Krylova O, Earn DJD. Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. Journal of the Royal Society Interface. 2013;10. doi:10.1098/rsif.2013.0098.

[87] Feng Z, Gurski KF, Prosper O, Teboh-Ewungkem MI, Grogan M. A mosquito-borne disease model with non-exponentially distributed infection and treatment stages. Journal of Dynamics and Differential Equations. 2020; p. 1–31. doi:10.1007/s10884-020-09863-2.

[88] Helwich M. Durational effects and non-smooth semi-Markov models in life insurance. 2008;.

[89] Shi C, Jiang Y, Zhou T. Queuing models of gene expression: analytical distributions and beyond. Biophysical Journal. 2020;119:1606–1616. doi:10.1016/j.bpj.2020.09.001.

[90] Thieme HR, Castillo-Chavez C. On the role of variable infectivity in the dynamics of the human immunodeficiency virus epidemic. Lecture Notes in Biomathematics. 1989;.

[91] Anderson RM. The epidemiology of HIV infection: variable incubation plus infectious periods and heterogeneity in sexual activity. Journal of the Royal Statistical Society Series A (Statistics in Society. 1988;151:66–98.

[92] Feng Z, Xu D, Zhao H. Epidemiological models with non-exponentially distributed disease stages and applications to disease control. Bulletin of Mathematical Biology. 2007;69:1511–1536. doi:10.1007/s11538-006-9174-9.

[93] Korobeinikov A. Lyapunov functions and global properties for SEIR and SEIS epidemic models. Mathematical Medicine and Biology. 2004;21:75–83. doi:10.1093/imammb/21.2.75.

[94] Lloyd AL. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics; 2001.

[95] Nakata Y, Röst G. Global stability of an SIS epidemic model with a finite infectious period. Differential and Integral Equations. 2018;31:161–172.

[96] Thieme HR, Castillo-Chavez C. How may infection-age-dependent infectivity affect the dynamics of HIV/AIDS?; 1993.

[97] Chowell G, Hyman JM. Mathematical and statistical modeling for emerging and re-emerging infectious diseases. Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases. 2016; p. 1–356. doi:10.1007/978-3-319-40413-4.

[98] Dang YX, Qiu ZP, Li XZ, Martcheva M. Global dynamics of a vector-host epidemic model with age of infection. Mathematical Biosciences and Engineering. 2017;14:1159–1186. doi:10.3934/mbe.2017060.

[99] Smith T, Schellenberg JA, Hayes R. Attributable fraction estimates and case definitions for malaria in endemic. Statistics in Medicine. 1994;13:2345–2358. doi:10.1002/sim.4780132206.

[100] Dalrymple U, Cameron E, Bhatt S, Weiss DJ, Gupta S, Gething PW. Quantifying the contribution of Plasmodium falciparum malaria to febrile illness amongst African children. eLife. 2017;6. doi:10.7554/elife.29198.

[101] Smith T, Vounatsou P. Estimation of infection and recovery rates for highly polymorphic parasites when detectability is imperfect, using hidden Markov models. Satistics in Medicine Statist Med. 2003;22:1709–1724. doi:10.1002/sim.1274.

[102] Cao P, Collins KA, Zaloumis S, Wattanakul T, Tarning J, Simpson JA, et al.. Modelling the population dynamics of Plasmodium falciparum gametocytes in humans during malaria infection; 2019.

[103] Brauer F. Age of infection epidemic models. Springer International Publishing; 2016.

[104] Camponovo F, Lee TE, Russell JR, Burgert L, Gerardin J, Penny MA. Mechanistic within-host models of the asexual *Plasmodium falciparum* infection: a review and analytical assessment. 2021;doi:10.1186/s12936-021-03813-z.

[105] Amrhein L, Harsha K, Fuchs C. A mechanistic model for the negative binomial distribution of single-cell mRNA counts. bioRxiv. 2019; p. 657619. doi:10.1101/657619.

[106] Bershteyn A, Gerardin J, Bridenbecker D, Lorton CW, Bloedow J, Baker RS, et al. Implementation and applications of EMOD, an individual-based multi-disease modeling platform. Pathogens and Disease. 2018;76. doi:10.1093/FEMSPD/FTY059.

[107] Rodriguez-Barraquer I, Arinaitwe E, Jagannathan P, Kamya MR, Rosenthal PJ, Rek J, et al. Quantification of anti-parasite and anti-disease immunity to malaria as a function of age and exposure. eLife. 2018;7. doi:10.7554/ELIFE.35832.

[108] Ukawuba I, Shaman J. Inference and dynamic simulation of malaria using a simple climate-driven entomological model of malaria transmission. PLoS Computational Biology. 2022;18. doi:10.1371/journal.pcbi.1010161.

Appendix A

CHAPTER 1 APPENDIX

Once we know the stationary distribution, we can check its stability in the set of distributions. That is, if we don't start off initially with an AHP distribution or it is perturbed through a non-uniformly acting intervention such as a targeted treatment or vaccine, will the distribution of MOI return to being AHP-distributed? If so, the dynamics above which assume AHP-distributed MOI will be good long-term approximations for any initial distribution. To investigate this, consider the PDE for the PGF:

$$\frac{\partial G}{\partial t} = r(1-s)\frac{\partial G}{\partial s} + \mu - (\mu + \lambda(1-s))G$$

As this is a linear equation, we see we can arrange it to be in the following form:

$$\frac{\partial G}{\partial t} = \mathcal{L}G + \mu$$

where

$$\mathcal{L} := r(1-s)\frac{\partial}{\partial s} - (\mu + \lambda(1-s))$$

We will restrict the domain for $\mathcal{L}$ to the following:

$$\mathcal{D} = \{u \in C^1[0,1], \lim_{s \to 1^-} u = 0\}$$

It's clear that this forms a complete subspace of $C^1[0,1]$ under the L^2 norm.

Now we can define

$$h(t,s) = G(t,s) - G^*(t,s)$$

where G is the generating function for the transient distribution at time t and G^* is the generating function for the stationary distribution found before. In particular this means $G^* \in \{G : \frac{\partial G}{\partial t} = 0\}$. As the left hand limit of each of the functions on the right at 1 is 1 because they are PGFs, the left hand limit of h must be 0. Also, as long as the mean of G is finite, h will be once continuously differentiable. Therefore, $h \in \mathcal{D}$.

Now that we know h is in our domain, we can consider the action of $\frac{\partial}{\partial t}$:

$$\frac{\partial h}{\partial t} = \frac{\partial G}{\partial t} - \frac{\partial G^*}{\partial t}$$

$$= \frac{\partial G}{\partial t}$$

$$= \mathcal{L}G + \mu$$

$$= \mathcal{L}(h + G^*) + \mu$$

$$= \mathcal{L}h + \mathcal{L}G^* + \mu$$

$$= \mathcal{L}h$$

where the last equality holds because G^* being stationary implies that $\mathcal{L}G^* = -\mu$. Now we can investigate the spectrum of $\mathcal{L}|_{\mathcal{D}}$, that is the operator restricted to the subspace that h is in. Setting up the eigenvalue problem, for any eigenfunction $u \in \mathcal{D}$ we get

$$\mathcal{L}u = \sigma u$$

$$r(1-s)\frac{\partial u}{\partial s} - (\mu + \lambda(1-s))u = \sigma u$$

$$\frac{\partial u}{\partial s} = \left(\frac{\mu+\sigma}{r(1-s)} + \frac{\lambda}{r}\right) u$$

The parameters in the equation above for the stationary distribution are constant with respect to s. The solution to this equation is therefore

$$u = e^{\frac{\lambda}{r}s}(1-s)^{-\frac{\mu+\sigma}{r}}$$

where the constant of integration is left out as it is irrelevant here to the span of the eigenfunctions.

The condition that the left hand limit at 1 always equals zero implies that the left hand limit at 1 of the time derivative is also zero. Because the action of the time derivative is equivalent to the action of the linear operator $\mathcal{L}$ in the domain to which u belongs, this implies that the function can't be singular at 1, and in particular

$$\sigma \leq -\mu$$

Therefore, the spectrum of $\mathcal{L}$ is $(-\infty, -\mu]$. This combined with the expression for the time derivative of h yields

$$\frac{\partial h}{\partial t} = \mathcal{L}h$$

$$\frac{\partial h}{\partial t} \leq -\mu h$$

so

$$h \leq h_0 e^{-\mu t}$$

Plugging in the definition of h, we get

$$G - G^* \leq (G_0 - G_0^*)e^{-\mu t}$$

In addition, because PGFs in the unit interval are bounded by 1, h_0 is bounded in absolute value by 1. Using this, we see

$$\lim_{t\to\infty} ||G - G^*||_\infty \leq \lim_{t\to\infty} e^{-\mu t} = 0$$

This implies uniform convergence of any initial PGF to that of an AHP distribution on the unit interval. Therefore any initial distribution with finite mean converges in distribution to an AHP distribution, so it is globally asymptotically stable in this domain as long as the parameters are non-negative and asymptotically constant with $r > 0$ for all time. For this system μ here acts as a sort of bifurcation parameter, where $\mu = 0$ is the boundary between stability and instability. It is still stable at this boundary because any function which is not in the null space of the linear operator will have a strictly negative time derivative, which implies the asymptotic stability of the stationary distribution holds for the standard M/M/∞ queue with its stationary Poisson distribution as well.

Appendix B

CHAPTER 2 APPENDIX

All code used during this study are at the following location:

https://github.com/jmh227/Malaria-Therapy-Data-Analysis The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Appendix C

CHAPTER 3 APPENDIX

Supplemental material and code can be found at:
https://github.com/dd-harp/RAMP-Model-Library/tree/main/Memory

Appendix D

CHAPTER 4 APPENDIX

D.0.1 Analysis of PDE and Normalization

$$
\begin{aligned}
\frac{\partial y}{\partial a} + \frac{\partial y}{\partial \alpha} &= -r(\alpha)y(a,\alpha) \\
&= -\frac{f(\alpha)}{1-F(\alpha)}y(a,\alpha) \\
y(a,0) &= h(a,\alpha) \\
y(a,\alpha) &= h(a-\alpha)(1-F(\alpha)) \\
N(a) &= \int_0^a h(a-\alpha)(1-F(\alpha))d\alpha \\
\frac{dN}{da} &= h(0) + \int_0^a \frac{d}{da}h(a-\alpha)(1-F(\alpha))d\alpha \\
u &:= a-\alpha \\
&= h(0) + \int_0^a \frac{d}{du}h(u)(1-F(a-u))du \\
&= h(0) + h(a) - h(0) - \int_0^a h(u)\frac{d}{du}(1-F(a-u))du \\
&= h(a) - \int_0^a h(u)f(a-u)du \\
&= h(a) - \int_0^a h(a-\alpha)f(\alpha)d\alpha
\end{aligned}
$$

For an exponential distribution, there is a nice relation in that the survival function is a constant multiple of the pdf. That is, for $E \sim \text{Exp}(r)$,

$$
\begin{aligned}
f_E(\alpha) &= re^{-r\alpha} \\
S_E(\alpha) &= 1 - F_E(\alpha) \\
&= e^{-r\alpha}
\end{aligned}
$$

and therefore

$$
\begin{aligned}
N(a) &= \int_0^a h(a-\alpha)e^{-r\alpha}d\alpha \\
\frac{dN}{da} &= h(a) - r\int_0^a h(a-\alpha)e^{-r\alpha}d\alpha \\
&= h(a) - rN
\end{aligned}
$$

In general, this normalizing constant is in fact the expected number of infections in an $M/G/\infty$ queue, which is Poisson-distributed regardless of the time spent infected [physics of the $M/M/\infty$ queue].

D.0.2 Estimators Conditioned on Infection Age

Exponential

This is a relatively simple GLM with a log link:

$$\log(\hat{X}) = k + cY$$

That is, our estimator is of the form

$$\begin{aligned}\hat{X} &:= E[X|Y] \\ &= e^{k+cY}\end{aligned}$$

This is important in estimating quantities such as parasite densities, which decline exponentially with infection age. Or, as stated in chapter 2, log parasitemia declines linearly with infection age.

To understand the dynamics of the induced process of our estimator, we can consider the moment generating function of the process.[1] The moment generating function for a random variable X, when it exists, is defined to be

$$G_X(s) := \int_\Omega e^{sx} f_X(x) dx$$

where Ω is the support of the random variable X. This is clearly defined for negative s (which is equivalent to a Laplace transform of our variable X when the domain $\Omega = \mathbb{R}^+$).

For a stochastic process, we can define the moment generating function as being a function of both time and the new variable s. We can apply age derivatives to both sides and apply the information we have about our infection age density $y(a, \alpha)$:

[1]This is essentially changing the basis to that of exponential functions. This is useful as they are the eigenfunctions of differential operators. Equivalently, we may consider a full Fourier basis which unlike the exponential basis is guaranteed to exist for all random variables and the subsequent analysis will remain the same. For simplicity of notation, we stick with a real-valued basis which exists for the simple dynamics we are considering.

$$
\begin{aligned}
G_{\hat{X}}(a,s) &= E[e^{s\hat{X}}] \\
&= E[e^{se^{k+cY}}] \\
&= \int_0^a e^{se^{k+c\alpha}} y(a,\alpha)d\alpha \\
\frac{\partial}{\partial a}[G_{\hat{X}}m(a)] &= \frac{\partial}{\partial a}\int_0^a e^{se^{k+c\alpha}} z(a,\alpha)d\alpha \\
&= e^{se^{k+ca}} y + \int_0^a e^{se^{k+c\alpha}} \frac{\partial z}{\partial a}(a,\alpha)d\alpha \\
&= \int_0^a e^{se^{k+c\alpha}} \frac{\partial z}{\partial a}(a,\alpha)d\alpha \\
&= \int_0^a e^{se^{k+c\alpha}} \left[-\frac{\partial z}{\partial \alpha} - rz(a,\alpha)\right] d\alpha \\
&= -\int_0^a e^{se^{k+c\alpha}} \frac{\partial z}{\partial \alpha} d\alpha - r\int_0^a e^{se^{k+c\alpha}} z(a,\alpha)d\alpha \\
&= -e^{se^{k+c\alpha}} z(a,\alpha)\Big|_{\alpha=0}^{a} + cs\int_0^a e^{se^{k+c\alpha}+k+c\alpha} z(a,\alpha)d\alpha - rG_{\hat{X}} \\
&= e^{se^k} h(a) + cs\frac{\partial}{\partial s}\int_0^a e^{se^{k+c\alpha}} f_A(a,\alpha)d\alpha - rmG_{\hat{X}} \\
&= cs\frac{\partial G_{\hat{X}}}{\partial s} + e^{se^k} h(a) - rmG_{\hat{X}} \\
\frac{\partial G_{\hat{X}}}{\partial a} m(a) + G_{\hat{X}}[h - rm] &= cs\frac{\partial G_{\hat{X}}}{\partial s} + e^{se^k} h(a) - rmG_{\hat{X}} \\
\frac{\partial G_{\hat{X}}}{\partial a} - \frac{cs}{m}\frac{\partial G_{\hat{X}}}{\partial s} &= \frac{h}{m}\left[e^{se^k} - G_{\hat{X}}\right]
\end{aligned}
$$

This results in a relatively simple inhomogeneous advection equation, which can be solved using a standard method of characteristics argument. Note that this is not well-posed for large s, but for negative s (i.e. the Laplace basis) we can obtain a well-posed equation:

$$
\frac{\partial G_{\hat{X}}}{\partial a} - cs\frac{\partial G_{\hat{X}}}{\partial s} = e^{-se^k} h(a) - rG_{\hat{X}}
$$

with boundary conditions

$$G_{\hat{X}}(0,t) = 1$$
$$\lim_{s\to\infty} G_{\hat{X}}(s,t) = 0$$

which should hold for the Laplace transform of any random variable.

Interestingly, we can see some model behavior dependent on the parameters. For example, the advection direction and speed is determined by the sign of the coefficient of the linear predictor c; the intercept k determines the sensitivity of the model to exposure h; and the rate of recovery r defines a timescale of memory of the exposure signal h.

Sigmoid

$$\begin{aligned}\hat{X} &:= E[X|Y] \\ &\sim \frac{e^{cY+k}}{1+e^{cY+k}} \\ &= \sigma(cY+k)\end{aligned}$$

This is the case for both detection and fever. Using fever as the example, we see that fever depends on log10 parasitemia in a sigmoidal way:

$$F = \sigma(cB+k)$$

or, equivalently, the log odds of fever depend log-linearly on log10 parasitemia:

$$\mathcal{O}_F = cB + k$$

and log10 parasitemia is approximately linearly dependent on infection age:

$$\begin{aligned}\hat{B} &:= E[B|A] \\ &= mA + b\end{aligned}$$

Therefore we can get an estimator for the expected value of the odds of fever:

$$\begin{aligned}\hat{\mathcal{O}}_F &:= E[\mathcal{O}_F|A] \\ &= E[cB + k|A] \\ &= cE[B|A] + k \\ &= c(mA + b) + k \\ &= c_1 A + c_2\end{aligned}$$

Now we can use our estimator of the log odds of fever to get an estimator of the probability of fever:

$$\begin{aligned}\hat{P}_F &:= \sigma(\hat{\mathcal{O}}_F) \\ &= \sigma(c_1 A + c_2)\end{aligned}$$

Interestingly, this is a sort of similarity transformation of the form

$$\sigma \circ E[\cdot\,|A] \circ \sigma^{-1}$$

where in order to get an estimation of the probability an individual is febrile we first map

the quantity into log-odds space, then condition on information from the infection age, then map back into probability space.

$$
\begin{aligned}
G_{\hat{X}}(s,a) &= E[e^{s\hat{X}}] \\
&= \int_0^a e^{s\sigma(c\alpha+k)} f(a,\alpha)d\alpha \\
&= \frac{1}{m}\int_0^a e^{s\sigma(c\alpha+k)} h(a-\alpha)(1-F(\alpha)) \\
mG_{\hat{X}}(s,a) &= \int_0^a e^{s\sigma(c\alpha+k)} h(a-\alpha)(1-F(\alpha))d\alpha \\
\frac{dmG_{\hat{X}}(s,a)}{da} &= \frac{d}{da}\int_0^a e^{s\sigma(c\alpha+k)} h(a-\alpha)(1-F(\alpha))d\alpha \\
&= e^{s\sigma(ca+k)}h(0)(1-F(a)) + \int_0^a e^{s\sigma(c\alpha+k)}(1-F(\alpha))\frac{d}{da}h(a-\alpha)d\alpha \\
&= e^{s\sigma(ca+k)}h(0)(1-F(a)) + \int_0^a e^{s\sigma(c(a-u)+k)}(1-F(a-u))\frac{d}{du}h(u)du \\
&= e^{s\sigma(ca+k)}h(0)(1-F(a)) + h(a)e^{s\sigma(k)} - e^{s\sigma(ca+k)}h(0)(1-F(a)) \\
&- \int_0^a h(u)\frac{d}{du}[e^{s\sigma(c(a-u)+k)}(1-F(a-u))]du \\
&= h(a)e^{s\sigma(k)} - \int_0^a h(u)\frac{d}{du}[e^{s\sigma(c(a-u)+k)}(1-F(a-u))]du \\
&= h(a)e^{s\sigma(k)} - s\int_0^a h(u)\frac{d}{du}[\sigma(c(a-u)+k)]e^{s\sigma(c(a-u)+k)}(1-F(a-u))du \\
&- \int_0^a h(u)e^{s\sigma(c(a-u)+k)}\frac{d}{du}(1-F(a-u))du \\
&= h(a)e^{s\sigma(ca+k)} \\
&+ s\int_0^a h(u)c\sigma(c(a-u)+k)(1-\sigma(c(a-u)+k))e^{s\sigma(c(a-u)+k)}(1-F(a-u))du \\
&- \int_0^a h(u)f(a-u)e^{s\sigma(c(a-u)+k)}du \\
&= h(a)e^{s\sigma(k)} + cs\int_0^a h(a-\alpha)\sigma(c\alpha+k)e^{-s\sigma(c\alpha+k)}(1-F(\alpha))d\alpha \\
&- cs\int_0^a h(a-\alpha)\sigma^2(c\alpha+k)e^{-s\sigma(c\alpha+k)}(1-F(\alpha))d\alpha - \int_0^a h(a-\alpha)f(\alpha)e^{s\sigma(c\alpha+k))}d\alpha
\end{aligned}
$$

There is some simplification that can be made. First, it's easy to see that

$$\begin{aligned}\frac{\partial^n G}{\partial s^n} &= \frac{\partial^n}{\partial s^n}\frac{1}{m}\int_0^a e^{s\sigma(c\alpha+k)}h(a-\alpha)(1-F(\alpha))d\alpha \\ &= \frac{1}{m}\int_0^a \sigma^n(ca+k)e^{s\sigma(ca+k)}h(a-\alpha)(1-F(\alpha))d\alpha\end{aligned}$$

Therefore,

$$\begin{aligned}cs\int_0^a h(a-\alpha)\sigma(c\alpha+k)e^{s\sigma(c\alpha+k)}(1-F(\alpha))d\alpha &= csm\frac{\partial G}{\partial s} \\ cs\int_0^a h(a-\alpha)\sigma^2(c\alpha+k)e^{s\sigma(c\alpha+k)}(1-F(\alpha))d\alpha &= csm\frac{\partial^2 G}{\partial s^2}\end{aligned}$$

Plugging in, we get

$$\frac{dmG_{\hat{X}}}{da} = h(a)e^{s\sigma(k)} + csm\frac{\partial G_{\hat{X}}}{\partial s} - csm\frac{\partial^2 G_{\hat{X}}}{\partial s^2} - \int_0^a h(a-\alpha)f(\alpha)e^{s\sigma(c\alpha+k)}d\alpha$$

$$\begin{aligned}\frac{dmG_{\hat{X}}}{da} &= \frac{dm}{da}G_{\hat{X}} + m\frac{\partial G_{\hat{X}}}{da} \\ &= [h(a) - \int_0^a h(a-\alpha)f(\alpha)d\alpha]G_{\hat{X}} + m\frac{\partial G_{\hat{X}}}{\partial a}\end{aligned}$$

$$\begin{aligned}m\frac{\partial G_{\hat{X}}}{\partial a} &= h(a)e^{s\sigma(k)} - \int_0^a h(a-\alpha)f(\alpha)e^{s\sigma(c\alpha+k)}d\alpha \\ &\quad - \left[h(a) - \int_0^a h(a-\alpha)f(\alpha)d\alpha\right]G_{\hat{X}} + csm\frac{\partial G_{\hat{X}}}{\partial s} - csm\frac{\partial^2 G_{\hat{X}}}{\partial s^2}\end{aligned}$$

In the special case in which Y is exponentially distributed, we have two nice simplifications

due to the fact that $f(\alpha) = re^{-r\alpha} = r(1 - F(\alpha))$:

$$\frac{1}{m}\int_0^a h(a-\alpha)f(\alpha)e^{s\sigma(c\alpha+k)}d\alpha = rG_{\hat{X}}$$
$$\frac{1}{m}\int_0^a h(a-\alpha)f(\alpha)d\alpha = r$$

Plugging in, we see

$$\frac{\partial G_{\hat{X}}}{\partial a} = \frac{h(a)}{m}\left[e^{s\sigma(k)} - G_{\hat{X}}\right] + cs\frac{\partial G_{\hat{X}}}{\partial s} - cs\frac{\partial^2 G_{\hat{X}}}{\partial s^2}$$

First we would like to investigate the equilibrium distribution. Setting the age derivative to zero and noting that at equilibrium $\frac{h}{m} = r$, we get

$$0 = r\left[e^{s\sigma(k)} - G_{\hat{X}}\right] + cs\frac{\partial G_{\hat{X}}}{\partial s} - cs\frac{\partial^2 G_{\hat{X}}}{\partial s^2}$$

First we will make one more substitution, $u := G_{\hat{X}} - 1$. Plugging in,

$$\begin{aligned} 0 &= r\left[e^{s\sigma(k)} - (u+1)\right] + cs\frac{\partial u}{\partial s} - cs\frac{\partial^2 u}{\partial s^2} \\ &= r\left[e^{s\sigma(k)} - 1 - u\right] + cs\frac{\partial u}{\partial s} - cs\frac{\partial^2 u}{\partial s^2} \\ u(0) &= 0 \end{aligned}$$

where the boundary condition follows from knowing that every MGF evaluated at 0 is equal to 1.

First we will consider the homogeneous equation:

$$0 = -ru_0 + cs\frac{\partial u_0}{\partial s} - cs\frac{\partial^2 u_0}{\partial s^2}$$
$$u_0(0) = 0$$

which is a special case of Kummer's differential equation, whose solutions are given by

$$u_0(s) = As_1F_1\left(1 - \frac{r}{c}; 2; s\right)$$

up to an arbitrary constant A, where ${}_1F_1$ is Kummer's confluent hyperogeometric function.

Now we would like to extend this solution to the general equation with the inhomogeneity:

$$-r(e^{s\sigma(k)} - 1) = -ru + cs\frac{\partial u}{\partial s} - cs\frac{\partial^2 u}{\partial s^2}$$

interestingly, we see that the inhomogeneity is in a similar form to the homogeneous solution, namely

$$\begin{aligned} h(s) &:= r(e^{s\sigma(k)} - 1) \\ &= rs_1F_1(1; 2; s\sigma(k)) \end{aligned}$$

Rescaling our independent variable, $z := \sqrt{\sigma(k)}s$, we get

$$-r\frac{z}{\sqrt{\sigma(k)}}\,{}_1F_1(1; 2; \sqrt{\sigma(k)}z) = -ru + cz\frac{\partial u}{\partial z} - \frac{cz}{\sqrt{\sigma(k)}}\frac{\partial^2 u}{\partial z^2}$$

Considering the right hand side as a linear operator

$$\mathcal{L} := -r + cz\frac{\partial}{\partial z} - \frac{cz}{\sqrt{\sigma(k)}}\frac{\partial^2}{\partial z^2}$$

with domain containing the space of twice differentiable functions with boundary condition $u(0) = 0$, we can set up the eigenvalue problem:

$$\begin{aligned}\lambda u &= \mathcal{L}u \\ &= -ru + cz\frac{\partial u}{\partial z} - \frac{cz}{\sqrt{\sigma(k)}}\frac{\partial^2 u}{\partial z^2} \\ 0 &= -(r+\lambda)u + cz\frac{\partial u}{\partial z} - \frac{cz}{\sqrt{\sigma(k)}}\frac{\partial^2 u}{\partial z^2}\end{aligned}$$

We can solve this in the domain (subject to the boundary condition) to obtain the set of eigenvectors:

$$u_\lambda(s) = z_1 F_1(1 - \frac{r+\lambda}{c}; 2; \sqrt{\sigma(k)}z)$$

up to an arbitrary constant multiple. The inhomogeneity is in fact in the span of the eigenvector with eigenvalue $-r$. This allows us to reframe our original problem:

$$-\frac{r}{\sqrt{\sigma(k)}}u_{-r} = \mathcal{L}u$$

that is, we are looking for the function in the domain of our operator whose image is a constant multiple of one of our eigenfunctions. The function u which satisfies that is

$$u(z) = c_1 u_{-r}(z) + c_2 u_0(z/\sqrt{\sigma(k)})$$

that is, a linear combination of the eigenfunction u_{-r} and an element from the null space. Plugging in to verify, we get

$$\begin{aligned}\mathcal{L}u &= \mathcal{L}[c_1 u_{-r} + c_2 u_0] \\ &= c_1 \mathcal{L} u_{-r} + c_2 \mathcal{L} u_0 \\ &= -r c_1 u_{-r}\end{aligned}$$

and the value of c_1 is $\frac{1}{\sqrt{\sigma(k)}}$ to match the inhomogeneity. Therefore the equilibrium to our PDE is given by

$$u(z) = \frac{1}{\sqrt{\sigma(k)}} z_1 F_1(1; 2; \sqrt{\sigma(k)}z) + A z_1 F_1(1 - \frac{r}{c}; 2; z/\sqrt{\sigma(k)})$$

for some constant A. Returning to our original variables, we have

$$\begin{aligned}G^*_{\hat{X}}(s) &= u(\sqrt{\sigma(k)}s) + 1 \\ &= s_1 F_1(1; 2; \sigma(k)s) + A s_1 F_1\left(1 - \frac{r}{c}; 2; s\right) + 1\end{aligned}$$

Subtracting 1 and dividing both sides by s, we get

$$\frac{G_{\hat{X}}(s) - 1}{s} = {}_1F_1(1; 2; \sigma(k)s) + A_1 F_1\left(1 - \frac{r}{c}; 2; s\right)$$

We can also replace the first confluent hypergeometric function by its equivalent simpler form:

$$\frac{G_{\hat{X}}(s)-1}{s} = \frac{e^{\sigma(k)s}-1}{\sigma(k)s} + A_1 F_1\left(1-\frac{r}{c};2;s\right)$$

We can expand each of these terms into their integral representations. Starting with the left-hand side, we have

$$\begin{aligned}\frac{G_{\hat{X}(s)}-1}{s} &= \frac{1}{s}\int_{\sigma(k)}^{1} e^{sx} f_{\hat{X}}(x)dx - \frac{1}{s} \\ &= \int_{\sigma(k)}^{1} \frac{e^{sx}-1}{s} f_{\hat{X}}(x)dx\end{aligned}$$

where we used the fact that $\int_{\sigma(k)}^{1} f_{\hat{X}}(x)dx = 1$. Further,

$$\frac{G_{\hat{X}}(s)-1}{s} = \int_{\sigma(k)}^{1}\left[\int_{0}^{x} e^{su}du\right] f_{\hat{X}}(x)dx$$

Integrating by parts, we have

$$\begin{aligned}
\frac{G_{\hat{X}}(s)-1}{s} &= \int_0^x e^{su} du F_{\hat{X}}(x)\Big|_{\sigma(k)}^1 - \int_{\sigma(k)}^1 e^{su}\left[\int_{\sigma(k)}^x f_{\hat{X}}(u)du\right]dx \\
&= \frac{e^s-1}{s} - \int_{\sigma(k)}^1 e^{sx} F_{\hat{X}}(x)dx \\
&= \int_0^1 e^{sx}dx - \int_{\sigma(k)}^1 e^{sx} F_{\hat{X}}(x)dx \\
&= \int_0^{\sigma(k)} e^{su}du + \int_{\sigma(k)}^1 e^{sx}\left[1-F_{\hat{X}}(x)\right]dx \\
&= \frac{e^{\sigma(k)s}-1}{s} + \int_{\sigma(k)}^1 e^{sx}\left[1-F_{\hat{X}}(x)\right]dx
\end{aligned}$$

Meanwhile on the right hand side, the second term has an integral representation if and only if $\frac{r}{c} \leq 1$:

$$A_1F_1\left(1-\frac{r}{c};2;s\right) = \frac{A}{\beta(1+\frac{r}{c},1-\frac{r}{c})}\int_0^1 e^{us}u^{-\frac{r}{c}}(1-u)^{\frac{r}{c}}du$$

where $\beta(x,y)$ is the beta function.

The right hand side is then equal to

$$\frac{e^{\sigma(k)s}-1}{s} + \frac{A}{\beta(1+\frac{r}{c},1-\frac{r}{c})}\int_0^1 e^{us}u^{-\frac{r}{c}}(1-u)^{\frac{r}{c}}du$$

We can subtract the exponential terms from each side so that they cancel, and we obtain

$$\int_{\sigma(k)}^1 e^{sx}[1-F_{\hat{X}}(x)]dx = \frac{A}{\beta(1+\frac{r}{c},1-\frac{r}{c})}\int_0^1 e^{us}u^{-\frac{r}{c}}(1-u)^{\frac{r}{c}}du$$

This implies that the two integrands agree on the overlap of their domains, that is

$$1 - F_{\hat{X}}(x) = Ax^{-\frac{r}{c}}(1-x)^{\frac{r}{c}}, \quad \sigma(k) \leq x \leq 1$$

for some constant A. In fact, we can find this constant by noting the survival function on the left must be equal to 1 when $x = \sigma(k)$, the left endpoint of the support:

$$\begin{aligned}
1 &= 1 - F_{\hat{X}}(\sigma(k)) \\
&= A\left(\sigma(k)\right)^{-\frac{r}{c}}\left(1-\sigma(k)\right)^{\frac{r}{c}} \\
&= A\left(\frac{1-\sigma(k)}{\sigma(k)}\right) \\
&= A\left(\frac{1-\frac{e^k}{1+e^k}}{\frac{e^k}{1+e^k}}\right)^{\frac{r}{c}} \\
&= A\left(\frac{1}{e^k}\right)^{\frac{r}{c}} \\
&= Ae^{-kr/c}
\end{aligned}$$

Therefore $A = e^{kr/c}$ and the CDF of the stationary distribution is

$$F_{\hat{X}}(x) = 1 - e^{kr/c}x^{-\frac{r}{c}}(1-x)^{\frac{r}{c}}, \quad \sigma(k) \leq x \leq 1$$

and the PDF is

$$f_{\hat{X}}(x) = \frac{r}{c}e^{kr/c}x^{-\frac{r}{c}-1}(1-x)^{\frac{r}{c}-1}$$

Similarly, we can determine for $c < 0$ the CDF is

$$F_{\hat{X}}(x) = e^{-kr/c} x^{\frac{r}{c}} (1-x)^{-\frac{r}{c}}$$

and its PDF is

$$f_{\hat{X}}(x) = \frac{r}{c} e^{-kr/c} x^{\frac{r}{c}-1} (1-x)^{-\frac{r}{c}-1}$$

More generally, we can define two parameters A and B and parameterize the distribution such that

$$f(x) = |A| B^A x^{-A-1} (1-x)^{A-1}$$

where $A \neq 0$ and $B > 0$, and the support is $(\frac{B}{1+B}, 1)$ if $A > 0$ and $(0, \frac{B}{1+B})$ if $A < 0$.

The expectation of this distribution for $A > 0$ is

$$\begin{aligned}
E[\hat{X}] &= AB^A \int_{\sigma(k)}^{1} x^{-A}(1-x)^{A-1} dx \\
u &= 1 - x \\
du &= -x \\
E[\hat{X}] &= AB^A \int_{0}^{1-\sigma(k)} u^{A-1}(1-u)^{-A} dx \\
&= AB^A \beta_{1-\sigma(k)}(A, 1-A)
\end{aligned}$$

where $\beta_x(a, b)$ is the incomplete beta function.

The second moment can be found analogously:

$$\begin{aligned} E[\hat{X}^2] &= AB^A \int_{\sigma(k)}^{1} x^{-A+1}(1-x)^{A-1}dx \\ &= AB^A \int_{0}^{1-\sigma(k)} u^{A-1}(1-u)^{1-A}du \\ &= AB^A \beta_{1-\sigma(k)}(A, 2-A) \end{aligned}$$

so the variance is

$$\begin{aligned} Var(\hat{X}) &= E[\hat{X}^2] - E[\hat{X}]^2 \\ &= AB^A \left[\beta_{1-\sigma(k)}(A, 2-A) - AB^A \beta_{1-\sigma(k)}(A, 1-A)^2\right] \end{aligned}$$

and the third moment is

$$\begin{aligned} E[\hat{X}^3] &= AB^A \int_{\sigma(k)}^{1} x^{-A+2}(1-x)^{A-1}dx \\ &= AB^A \int_{0}^{1-\sigma(k)} u^{A-1}(1-u)^{2-A}dx \\ &= AB^A \beta_{1-\sigma(k)}(A, 3-A) \end{aligned}$$

There is a nice recurrence property of incomplete beta functions:

$$\beta_x(a, b+1) = \frac{b}{a+b}\beta_x(a, b) + \frac{x^a(1-x)^b}{a+b}$$

This will allow us to rewrite the second and third moments:

$$
\begin{aligned}
E[\hat{X}^2] &= AB^A \left[\frac{1-A}{A+1-A} \beta_{1-\sigma(k)}(A, 1-A) + \frac{(1-\sigma(k))^A \sigma(k)^{1-A}}{A+1-A} \right] \\
&= (1-A)E[\hat{X}] + AB^A (1-\sigma(k))^A \sigma(k)^{1-A} \\
E[\hat{X}^3] &= AB^A \left[\frac{2-A}{A+2-A} \beta_{1-\sigma(k)}(A, 2-A) + \frac{(1-\sigma(k))^A \sigma(k)^{2-A}}{A+2-A} \right] \\
&= \frac{2-A}{2} E[\hat{X}^2] + AB^A \frac{(1-\sigma(k))^A \sigma(k)^{2-A}}{2}
\end{aligned}
$$

this implies we can multiply $E[\hat{X}^2]$ by $\frac{\sigma(k)}{2}$ to get

$$
\frac{\sigma(k)}{2} E[\hat{X}^2] = \frac{\sigma(k)(1-A)}{2} E[\hat{X}] + \frac{AB^A (1-\sigma(k))^A \sigma(k)^{2-A}}{2}
$$

subtracting this from $E[\hat{X}^3]$, we get

$$
\begin{aligned}
E[\hat{X}^3] - \frac{\sigma(k)}{2} E[\hat{X}^2] &= \frac{2-A}{2} E[\hat{X}^2] + AB \frac{(1-\sigma(k))^A \sigma(k)^{2-A}}{2} \\
&\quad - \frac{\sigma(k)(1-A)}{2} E[\hat{X}] - \frac{AB(1-\sigma(k))^A \sigma(k)^{2-A}}{2} \\
&= \frac{2-A}{2} E[\hat{X}^2] - \frac{\sigma(k)(1-A)}{2} E[\hat{X}]
\end{aligned}
$$

solving for $E[\hat{X}^3]$, we see that we have the following relationship between the first three moments:

$$
E[\hat{X}^3] = \frac{2+\sigma(k)-A}{2} E[\hat{X}^2] - \frac{\sigma(k)(1-A)}{2} E[\hat{X}]
$$

This is the important piece to the moment closure problem for our model, in which $A = \frac{r}{c}$. Going back to our PDE for the MGF,

$$\frac{\partial G_{\hat{X}}}{\partial a} = \frac{h}{m}\left[e^{s\sigma(k)} - G_{\hat{X}}\right] + cs\frac{\partial G_{\hat{X}}}{\partial s} - cs\frac{\partial^2 G_{\hat{X}}}{\partial s^2}$$

Taking an s-derivative, we get

$$\begin{aligned}\frac{\partial^2 G_{\hat{X}}}{\partial a \partial s} = \frac{h}{m}\left[\sigma(k)e^{s\sigma(k)} - \frac{\partial G_{\hat{X}}}{\partial s}\right] + c\left[\frac{\partial G_{\hat{X}}}{\partial s} - \frac{\partial^2 G_{\hat{X}}}{\partial s}\right] \\ + cs\left[\frac{\partial^2 G_{\hat{X}}}{\partial s^2} - \frac{\partial^3 G_{\hat{X}}}{\partial s^3}\right]\end{aligned}$$

Plugging in $s = 0$, we see

$$\frac{dm_1}{da} = \frac{h}{m}\left[\sigma(k) - m_1\right] + c\left(m_1 - m_2\right)$$

However, we are left with an unknown second moment. Continuing to a third derivative, we have

$$\begin{aligned}\frac{\partial^3 G_{\hat{X}}}{\partial a \partial s^2} = \frac{h}{m}\left[\sigma(k)^2 e^{s\sigma(k)} - \frac{\partial^2 G_{\hat{X}}}{\partial s^2}\right] + c\left[\frac{\partial^2 G_{\hat{X}}}{\partial s^2} - \frac{\partial^3 G_{\hat{X}}}{\partial s^3}\right] \\ + c\left[\frac{\partial^2 G_{\hat{X}}}{\partial s^2} - \frac{\partial^3 G_{\hat{X}}}{\partial s^3}\right] + cs\left[\frac{\partial^3 G_{\hat{X}}}{\partial s^3} - \frac{\partial^4 G_{\hat{X}}}{\partial s^4}\right]\end{aligned}$$

Again plugging in $s = 0$, we get

$$\begin{aligned}\frac{dm_2}{da} &= \frac{h}{m}\left[\sigma(k)^2 - m_2\right] + c\left(m_2 - m_3\right) + c\left(m_2 - m_3\right) \\ &= \frac{h}{m}[\sigma(k)^2 - m_2] + 2c(m_2 - m_3)\end{aligned}$$

Plugging in the expression found above for m_3, we get

$$\begin{aligned} m_3 &= \frac{2+\sigma(k)-A}{2}m_2 - \frac{\sigma(k)(1-A)}{2}m_1 \\ 2c(m_2-m_3) &= -c(\sigma(k)-A)m_2 + c\sigma(k)(1-A)m_1 \\ \frac{dm_2}{da} &= \frac{h}{m}\sigma(k)^2 - \frac{h}{m}m_2 - c(\sigma(k)-A)m_2 + c\sigma(k)(1-A)m_1 \end{aligned}$$

We know that $A = \frac{h}{cm}$, so plugging in:

$$\begin{aligned} \frac{dm_2}{da} &= \frac{h}{m}\sigma(k)^2 - \frac{h}{m}m_2 - c\sigma(k)m_2 + \frac{h}{m}m_2 + c\sigma(k)(1-\frac{h}{cm})m_1 \\ &= \frac{h}{m}\sigma(k)^2 + \left(c - \frac{h}{m}\right)\sigma(k)m_1 - c\sigma(k)m_2 \\ &= \frac{h}{m}\sigma(k)(\sigma(k)-m_1) + c\sigma(k)(m_1-m_2) \end{aligned}$$

combining with the first moment equation, we have

$$\begin{aligned} \frac{dm_1}{da} &= \frac{h}{m}[\sigma(k)-m_1] + c(m_1-m_2) \\ \frac{dm_2}{da} &= \frac{h}{m}\sigma(k)(\sigma(k)-m_1) + c\sigma(k)(m_1-m_2) \\ &= \sigma(k)\frac{dm_1}{da} \end{aligned}$$

this implies that $m_2 = \sigma(k)m_1 + K$ for some K independent of age.

This is in a sense a linearization of the relationship between the first two moments. If we consider this linearization around one of the points on the boundary of the support, then $K = 0$. This implies

$$\begin{aligned}\frac{dm_1}{da} &= \frac{h}{m}[\sigma(k) - m_1] + c(m_1 - \sigma(k)m_1 - K) \\ &= \frac{h}{m}[\sigma(k) - m_1] + c(1 - \sigma(k))m_1 - Kc \\ &= \frac{h}{m}\sigma(k) - Kc - \left[\frac{h}{m} - c(1 - \sigma(k))\right] m_1\end{aligned}$$

Note that this equation only makes sense when the bracketed term is positive. This is guaranteed anytime

$$\frac{h}{m(1 - \sigma(k))} > c$$

Further if $c < 0$, this is guaranteed. This is the case for detection and probability of fever, and therefore replacing c with $-c$ we get

$$\frac{dm_1}{da} = \frac{h}{m}\sigma(k) + Kc - \left[\frac{h}{m} + c(1 - \sigma(k))\right] m_1$$

which is globally asymptotically stable to its unique equilibrium

$$m_1 = \frac{\frac{h}{m}\sigma(k) + Kc}{\frac{h}{m} + c(1 - \sigma(k))}$$

Note that if $c < 0$, we have the distribution is bounded between 0 and $\sigma(k)$. In order to make that agree, we would have to have here that $K = 0$. This results in the following equation:

$$\frac{dm_1}{da} = \frac{h}{m}\sigma(k) - \left[\frac{h}{m} + c(1 - \sigma(k))\right] m_1$$

with equilibrium

$$\frac{\frac{h}{m}\sigma(k)}{\frac{h}{m} + c(1-\sigma(k))}$$

Note that as $\sigma(k) \to 1$, we have the equilibrium is always equal to 1. This is due to the fact that the pdf reaches a sort of bifurcation, where with negative c and $\sigma(k) \to 1$ leads to a divergent expression for the density. This also makes sense when we think about what would be necessary for this to happen. This would suppose that $\hat{X} = \lim_{k\to\infty} \sigma(k - cA)$, which means any finite change in A will not effect $\hat{X}$ and it will be identically equal to 1.

Appendix E

CHAPTER 5 APPENDIX

E.1 Simple Infection

E.1.1 SDE Derivation

Suppose a population consists of N individuals, each of which are switching between two states labeled 0 and 1. The time spent in 0 is exponentially distributed with rate h, and the time spent in 1 is exponentially distributed with rate r. Then we can describe the proportion in state 1 at time t as

$$\bar{X}_t := \frac{1}{N}\sum_{i=1}^{N} X_{i,t}$$

That is, we add up all of the individuals at time t in state 1 $X_{i,t}$ and divide by the population size. This is the 'true prevalence' of those in state 1.

As $n \to \infty$, the true prevalence approaches the theoretical deterministic model of the following form:

$$\frac{dx}{dt} = h(1-x) - rx$$

as can be shown from the Kolmogorov forward equations for the process. However, this model does not capture inherent stochasticity in finite population sizes and therefore is hard to connect to realistic data collected from a population, even if it is very large. Typically a likelihood function is defined as the squared difference from x and collected data at discrete

time points to fit point estimates of the parameters h and r. However this method ignores the effect of process uncertainty which naturally provide a correlation structure between consecutive measurements and therefore a likelihood function commensurate with randomness in the data generating process.

Therefore let us go back to the stochastic process $\bar{X}_t$. First, we can consider the embedded Markov chain by conditioning on an individual changing. As each individual acts independently and has exponential waiting times to transitioning, we can find the conditional distribution of the time intervals given the current state as follows:

$$\begin{aligned} E[\Delta t_i|\bar{X}_{t_{i-1}}] &:= E[t_i - t_{i-1}|\bar{X}_{t_{i-1}}] \\ &\sim Exp(N[h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}]) \end{aligned}$$

as the next event is determined by the minimum time before one of $(N - \sum_{i=1}^{N} X_{i,t})$ individuals in state 0 potentially transitions to state 1 or one of $\sum_{i=1}^{N} X_{i,t}$ individuals in state 1 potentially transitions to state 0, each of which happens at an exponential rate. Factoring out an N from each gives the rate as written.

Given one of these transitions has occurred, we can consider the expected change in the prevalence:

$$E[\bar{X}_{t_i} - \bar{X}_{t_{i-1}}|\bar{X}_{t_{i-1}}] = \frac{1}{N}p_{01} - \frac{1}{N}p_{10}$$

where p_{01} is the probability of transitioning from state 0 to state 1 given a transition occurred, and p_{10} is the probability of transitioning from state 1 to state 0. The change in prevalence is $\frac{1}{N}$, as a single individual changed their status.

As the probabilities are the probability that the minimum time for some individual to transition from one state is less than the minimum time to transition from the other state,

we can determine them to be

$$\begin{aligned} p_{01} &= \frac{h(1-\bar{X}_{t_{i-1}})}{h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}} \\ p_{10} &= 1 - p_{01} \\ &= \frac{r\bar{X}_{t_{i-1}}}{h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}} \end{aligned}$$

Combining, we get

$$E[\bar{X}_{t_i} - \bar{X}_{t_{i-1}} | \bar{X}_{t_{i-1}}] = \frac{h(1-\bar{X}_{t_{i-1}}) - r\bar{X}_{t_{i-1}}}{N[h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}]}$$

Further, we can compute the variance of these transitions in a similar way.

$$\begin{aligned} Var(\bar{X}_{t_i} - \bar{X}_{t_{i-1}} | \bar{X}_{t_{i-1}}) &= E[((\bar{X}_{t_i} - \bar{X}_{t_{i-1}}) - E[\bar{X}_{t_i} - \bar{X}_{t_{i-1}} | \bar{X}_{t_{i-1}}])^2 | \bar{X}_{t_{i-1}}] \\ &= E[(\bar{X}_{t_i} - \bar{X}_{t_{i-1}})^2] - E[\bar{X}_{t_i} - \bar{X}_{t_{i-1}} | \bar{X}_{t_{i-1}}]^2 \end{aligned}$$

The jump size squared will always be equal to $\frac{1}{N^2}$ whether it increases or decreases, and plugging in our previous expectation we get

$$\begin{aligned} Var(\bar{X}_{t_i} - \bar{X}_{t_{i-1}} | \bar{X}_{t_{i-1}}) &= \frac{1}{N^2} - \left(\frac{h(1-\bar{X}_{t_{i-1}}) - r\bar{X}_{t_{i-1}}}{N[h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}]} \right)^2 \\ &= \frac{4rh\bar{X}_{t_{i-1}}(1-\bar{X}_{t_{i-1}})}{N^2[h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}]^2} \end{aligned}$$

Now we can define a centered random process Z_i as follows:

$$\begin{aligned} E[Z_i|X_{t_{i-1}}] &:= \bar{X}_{t_i} - E[\bar{X}_{t_i}|\bar{X}_{t_{i-1}}] \\ &= \bar{X}_{t_i} - \bar{X}_{t_{i-1}} - E\left[\bar{X}_{t_i} - \bar{X}_{t_{i-1}}|\bar{X}_{t_{i-1}}\right] \\ &= \bar{X}_{t_i} - \bar{X}_{t_{i-1}} - \frac{h(1-\bar{X}_{t_{i-1}}) - r\bar{X}_{t_{i-1}}}{N[h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}]} \end{aligned}$$

Now clearly Z_i forms a martingale sequence with respect to $\bar{X}_{i-1}$ with conditional mean zero and conditional variance equal to the expression found above. Scaling by the standard deviation to obtain a conditional variance of 1,

$$\hat{Z}_i = \frac{\bar{X}_{t_i} - \bar{X}_{t_{i-1}} - \frac{h(1-\bar{X}_{t_{i-1}}) - r\bar{X}_{t_{i-1}}}{N[h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}]}}{2\frac{\sqrt{hr\hat{X}_{t_{i-1}}(1-\bar{X}_{t_{i-1}})}}{N[h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}]}}$$

Now we would like to move our discrete time process back into one in continuous time. multiplying both sides by $\sqrt{\Delta t_i}$,

$$\hat{Z}_i\sqrt{\Delta t_i} = \frac{\bar{X}_{t_i} - \bar{X}_{t_{i-1}} - [h(1-\bar{X}_{t_{i-1}}) - r\bar{X}_{t_{i-1}}]\Delta t_i}{2\sqrt{hr\bar{X}_{t_{i-1}}(1-\bar{X}_{t_{i-1}})\Delta t_i}}$$

Focusing on the lefthand side,

$$\begin{aligned} \hat{Z}_i\sqrt{\Delta t_i} &= \frac{1}{\sqrt{h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}}}\frac{1}{\sqrt{N}}\hat{Z}_i \\ &\approx \frac{1}{\sqrt{h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}}}\mathcal{N}(0,1) \end{aligned}$$

Multiply the first part of the denominator back to the right hand side. Note that for large N,

the scaled sum of the $\hat{Z}_i$ is approximately a standard normal random variable. Multiplying by the square root of the time variable and summing both sides, we have

$$\sum_{i=1}^{N} \mathcal{N}(0, \Delta t_i) \sim \sum_{i=1}^{N} \frac{1}{2} \sqrt{\frac{h(1-\bar{X}_{t_{i-1}}) + r\bar{X}_{t_{i-1}}}{hr\bar{X}_{t_{i-1}}(1-\bar{X}_{t_{i-1}})}} \left[\bar{X}_{t_i} - \bar{X}_{t_{i-1}} - [h(1-\bar{X}_{t_{i-1}}) - r\bar{X}_{t_{i-1}}]\Delta t_i\right]$$

In the limit, we have

$$W_t = \int_0^t \frac{1}{2} \sqrt{\frac{h(1-\bar{X}_s) + r\bar{X}_t}{hr\bar{X}_s(1-\bar{X}_s)}} \left[d\bar{X}_s - [h(1-\bar{X}_s) - r\bar{X}_s]ds\right]$$

Taking derivatives on each side and rearranging, we get an SDE of the form

$$dX_t = [h(1-X_t) + rX_t]dt + 2\sqrt{\frac{hrX_t(1-X_t)}{h(1-X_t) + rX_t}}dW_t$$

Focusing on the noise term, we get

$$\begin{aligned} 2\sqrt{\frac{hrX_t(1-X_t)}{h(1-X_t) + rX_t}}dW_t &= 2\sqrt{\frac{hrX_t(1-X_t)}{h + (r-h)X_t}}dW_t \\ &= 2\sqrt{\frac{rX_t(1-X_t)}{1 + \left(\frac{r}{h} - 1\right)X_t}}dW_t \end{aligned}$$

Note that X_t will likely be near its equilibrium, so if we replace the prevalence term with its equilibrium value, we get

$$
\begin{aligned}
2\sqrt{\frac{rX_t(1-X_t)}{1+\left(\frac{r}{h}-1\right)X_t}} &\approx 2\sqrt{\frac{rX_t(1-X_t)}{1+\frac{r-h}{h}\frac{h}{r+h}}} \\
&\approx 2\sqrt{r}\sqrt{X_t(1-X_t)\left(1-\frac{r-h}{r+h}\right)} \\
&= 2\sqrt{\frac{2rh}{r+h}}\sqrt{X_t(1-X_t)}
\end{aligned}
$$

On the other hand, if either parameter is much larger than the other, then X_t will typically be close to 0 or 1 and so the noise term will be close to zero. Therefore the above approximation is likely to be good in general. Therefore we can get an approximate SDE of the form

$$
\begin{aligned}
dX_t &= [h(1-X_t)+rX_t]dt + 2\sqrt{\frac{2rh}{r+h}}\sqrt{X_t(1-X_t)}dW_t \\
&= (h+r)\left[\frac{h}{h+r}-X_t\right]dt + 2\sqrt{\frac{2rh}{r+h}}\sqrt{X_t(1-X_t)}dW_t
\end{aligned}
$$

Note that using the traditional method, the SDE formed would be

$$
dX_t = (h+r)\left[\frac{h}{h+r}-X_t\right]dt + \sqrt{h(1-X_t)+rX_t}dW_t
$$

which has the same drift but a different scaling on the noise. When $h \approx r$, they get noise that scales with $\sqrt{h}$. In the model presented here, we get $2\sqrt{\frac{2rh}{r+h}}\sqrt{X_t(1-X_t)}$, and when the two parameters are approximately equal then the equilbrium is around $1/2$. Replacing $X_t = 1/2$, we get the noise scales with $\sqrt{r} \approx \sqrt{h}$ by assumption, so the noise agrees. Therefore while they will likely agree to a good approximation for many parameter values, the new model has behavior which avoids this pathological behavior.

Further, while the old formulation does not have a simple closed form solution, the

approximation of the new model is known to have a closed form solution with stationary beta distributions. This implies that random testing of a group of individuals, seen as a Binomial random variable for the number infected when conditioned on the prevalence, will instead be modeled as a BetaBinomial random variable when taking into account the uncertainty associated with prevalence. The beta stationary distribution can also be used to construct an explicit likelihood function for the parameters.

$$dX_t = (h+r)\left[\frac{h}{h+r} - X_t\right]dt + \sqrt{h(1-X_t) + rX_t}dW_t \tag{E.1}$$

$$dX_t = (h+r)\left[\frac{h}{h+r} - X_t\right]dt + 2\sqrt{\frac{hrX_t(1-X_t)}{N[h(1-X_t)+rX_t]}}dW_t \tag{E.2}$$

$$\approx (h+r)\left[\frac{h}{h+r} - X_t\right]dt + 2\sqrt{\frac{2rh}{N(r+h)}}\sqrt{X_t(1-X_t)}dW_t \tag{E.3}$$

Note that if the disease is directly transmitted, the same argument will hold with $h = h_0 X_t$. This implies the SDE for the mean is of the form

$$\begin{aligned} dX_t &= [h_0 X_t(1-X_t) - rX_t]dt + \sqrt{\frac{4h_0 r X_t^2(1-X_t)}{h_0 X_t(1-X_t) + rX_t}}dW_t \\ &= h_0 X_t\left[\left(1 - \frac{r}{h_0}\right) - X_t\right]dt + 2\sqrt{\frac{h_0 r X_t(1-X_t)}{h_0(1-X_t)+r}}dW_t \end{aligned}$$

meaning that the approximate scaling of the noise is not changed significantly, but the mean is now a nonlinear function of the state of the system. It's clear that $X_t = 0$ is an absorbing state, as both the drift and the noise are equal to zero. The bias in the diffusion occurs when

$$\begin{aligned} X_t &= 1 - \frac{r}{h_0} \\ &= 1 - \frac{1}{R_0} \end{aligned}$$

where R_0 is the basic reproduction number from the deterministic theory. Note this is achievable if and only if $R_0 > 1$, otherwise there will be an attracting 'equilibrium' across the absorbing boundary. For any equilibrium contained in $[0, 1)$ (that is, with a finite rate of exposure), the disease will eventually die out as there is a unique absorbing point at zero prevalence.

E.1.2 Stationary Distribution

$$dp_t = [h(1 - p_t) - rp_t]dt + \sigma\sqrt{p_t(1 - p_t)}dW_t$$

We would like to transform to a new variable so that the variance does not depend on our current state. As Ito's formula states

$$\begin{aligned} dp_t &= \mu_t + \sigma_t dW_t \\ df(p_t) &= \left(\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial p} + \frac{\sigma^2}{2} f''(p_t)\right) dt + \sigma_t \frac{\partial f}{\partial x} dW_t \end{aligned}$$

for the associated Fokker-Planck equation to be a diffusion equation with constant diffusion rate, we must have

$$\frac{\partial f}{\partial x} = \frac{1}{\sqrt{x(1-x)}}$$

and therefore

$$f(x) = -2 \arcsin \sqrt{1-x}$$

equivalently,

$$f(x) = 2 \arcsin \sqrt{x}$$

Computing the second derivative, we get

$$\frac{\partial^2 f}{\partial x^2} = \frac{x - \frac{1}{2}}{[x(1-x)]^{3/2}}$$

Defining $Z_t := -2 \arcsin \sqrt{1-p_t}$, we get

$$\begin{aligned} dZ_t &= \left(\frac{h(1-p_t) - rp_t}{\sqrt{p_t(1-p_t)}} + \frac{\sigma^2 p_t(1-p_t)}{2} \frac{p_t - \frac{1}{2}}{[p_t(1-p_t)]^{3/2}} \right) dt + \sigma \sqrt{p_t(1-p_t)} \frac{1}{\sqrt{p_t(1-p_t)}} dW_t \\ &= \left(\frac{h(1-p_t) - rp_t + \frac{\sigma^2}{2}(p_t - \frac{1}{2})}{\sqrt{p_t(1-p_t)}} \right) dt + \sigma dW_t \end{aligned}$$

Note

$$\begin{aligned} f^{-1}(z) &= 1 - \sin^2 \left(-\frac{z}{2} \right) \\ &= \cos^2 \left(-\frac{z}{2} \right) \\ &= \cos^2 \left(\frac{z}{2} \right) \end{aligned}$$

Therefore

$$\begin{aligned} dZ_t &= \left(\frac{h\sin^2(-Z_t/2) - r\cos^2(-Z_t/2) + \frac{\sigma^2}{2}[\cos^2(-Z_t/2) - \frac{1}{2}]}{\cos(-Z_t/2)\sin(-Z_t/2)} \right) dt + \sigma dW_t \\ &= \left(h\tan(-Z_t/2) - r\cot(-Z_t/2) + \frac{\sigma^2}{4}\frac{\cos(-Z_t)}{\cos(-Z_t/2)\sin(-Z_t/2)} \right) dt + \sigma dW_t \\ &= \left(-h\tan(Z_t/2) + r\cot(Z_t/2) + \frac{\sigma^2}{4}\left(\tan(Z_t/2) - \cot(Z_t/2)\right) \right) dt + \sigma dW_t \\ &= \left((\frac{\sigma^2}{4} - h)\tan(Z_t/2) + (r - \frac{\sigma^2}{4})\cot(Z_t/2) \right) dt + \sigma dW_t \end{aligned}$$

The associated Fokker-Planck equation is given by

$$\frac{\partial y}{\partial t} = -\frac{\partial}{\partial x}\left[\left((\frac{\sigma^2}{4} - h)\tan(x/2) + (r - \frac{\sigma^2}{4})\cot(x/2) \right) y \right] + \frac{\sigma^2}{2}\frac{\partial^2 y}{\partial x^2}$$

The equilibrium distribution is then given by

$$\begin{aligned} \frac{\partial y}{\partial x} &= -\frac{2}{\sigma^2}\left[(h - \frac{\sigma^2}{4})\tan(x/2) - (r - \frac{\sigma^2}{4})\cot(x/2) \right] y + C \\ &= \left[(\frac{1}{2} - \frac{2h}{\sigma^2})\tan(x/2) - (\frac{1}{2} - \frac{2r}{\sigma^2})\cot(x/2) \right] y + C \end{aligned}$$

If $h = r$, then we would expect at $\pi/2$ the derivative should be zero; that is, all deterministic forces go to zero at the halfway point when the force "forward" (h) is equal to the force "backward" (r) probability density will flatten to reflect that directional indifference. This will only occur if $C = 0$. An interesting consequence of this argument is also the shape of the beta distribution will reflect whether noise or deterministic forces are dominant; if the deterministic forces dominate, then this middle point will be a mode of the distribution and

represent a maximum. However if noise dominates both forces, then the distribution will be u-shaped.

If we define $a = \frac{1}{2} - \frac{2h}{\sigma^2}$ and $b = \frac{1}{2} - \frac{2r}{\sigma^2}$, the solution is

$$\begin{aligned}
y(x) &= C_1 \exp\left(\int_0^x a\tan(z/2) - b\cot(z/2)dz\right) \\
&= C_1 \exp\left(-2\left[(a+b)\log(\cos(x/2)) + b\log(\tan(x/2))\right]\right) \\
&= C_1 \exp\left(-2(a+b)\log(\cos(x/2))\right)\exp\left(-2b\log(\tan(x/2))\right) \\
&= C_1 \cos(x/2)^{-2(a+b)}\tan(x/2)^{-2b} \\
&= C_1 \cos(x/2)^{-2a}\sin(x/2)^{-2b} \\
&= C_1\left(\frac{(1+\cos(x)}{2}\right)^{-a}\left(\frac{1-\cos(x)}{2}\right)^{-b} \\
&= C_2(1+\cos(x))^{-a}(1-\cos(x))^{-b}
\end{aligned}$$

with domain from $(-\pi, 0)$.

Therefore to find the distribution of PfPR, we must transform the variable such that $PfPR = \cos^2(-Z/2)$. We know that for a monotonic transformation X=f(Z), the density function of X with respect to the density of Z is

$$g_X(x) = g_Z(f^{-1}(x))\left|\frac{df^{-1}}{dx}\right|$$

We know

$$\frac{df^{-1}}{dx} = \frac{1}{\sqrt{x(1-x)}}$$

$$\cos(2\arcsin(\sqrt{1-x})) = 2x - 1$$

Plugging in, this gives

$$\begin{aligned}
g_X(x) &= C_1(1+\cos(2\arcsin(\sqrt{1-x})))^{-a}(1-\cos(\arcsin(\sqrt{1-x})))^{-b}\frac{1}{\sqrt{x(1-x)}} \\
&= C_1(1+(2x-1))^{-a}(1-(2x-1))^{-b}x^{-1/2}(1-x)^{-1/2} \\
&= C_1(2x)^{-a}(2-2x)^{-b}x^{-1/2}(1-x)^{-1/2} \\
&= C_2x^{-a}(1-x)^{-b}x^{-1/2}(1-x^{-1/2}) \\
&= C_2x^{-a-1/2}(1-x)^{-b-1/2} \\
&= C_2x^{(-a+1/2)-1}(1-x)^{(-b+1/2)-1}
\end{aligned}$$

As the density must integrate to 1 on its domain from $(0,1)$, we know the expected prevalence will be beta distributed with parameters

$$\begin{aligned}
\alpha &= -a + 1/2 \\
&= -\left(\frac{1}{2} - \frac{2h}{\sigma^2}\right) + \frac{1}{2} \\
&= \frac{2h}{\sigma^2} \\
\beta &= -b + 1/2 \\
&= -\left(\frac{1}{2} - \frac{2r}{\sigma^2}\right) + \frac{1}{2} \\
&= \frac{2r}{\sigma^2}
\end{aligned}$$

Therefore,

$$
\begin{aligned}
PfPR &\sim \text{Beta}\left(\frac{2h}{\sigma^2}, \frac{2r}{\sigma^2}\right) \\
&\sim \text{Beta}\left(\frac{2h}{\left(2\sqrt{\frac{2rh}{N(r+h)}}\right)^2}, \frac{2r}{\left(2\sqrt{\frac{2rh}{N(r+h)}}\right)^2}\right) \\
&\sim \text{Beta}\left(\frac{N}{4}\left(1+\frac{h}{r}\right), \frac{N}{4}\left(1+\frac{r}{h}\right)\right) \\
&\sim \text{Beta}\left(\frac{N}{4(1-x)}, \frac{N}{4x}\right)
\end{aligned}
$$

Therefore, in a quasi-stationary setting we have a one-parameter model of a non-Gaussian distribution:

$$
\begin{aligned}
\frac{dx}{dt} &= h(1-x) - rx \\
Y &\sim \text{Beta}\left(\frac{N}{4(1-x)}, \frac{N}{4x}\right)
\end{aligned}
$$

where N is the number of individuals being followed and measured periodically. This distribution measures the entirety of the process uncertainty.

$$
\begin{aligned}
E[Y] &= \frac{\frac{N}{4(1-x)}}{\frac{N}{4(1-x)} + \frac{N}{4x}} \\
&= \frac{x}{x + (1-x)} \\
&= x \\
Var(Y) &= \frac{\frac{N^2}{16x(1-x)}}{\left(\frac{N}{4(1-x)} + \frac{N}{4x}\right)^2 \left(\frac{N}{4(1-x)} + \frac{N}{4x} + 1\right)} \\
&= \frac{\frac{1}{x(1-x)}}{\left(\frac{1}{1-x} + \frac{1}{x}\right)^2 \left(\frac{N}{4(1-x)} + \frac{N}{4x} + 1\right)} \\
&= \frac{x(1-x)}{(x + (1-x))^2 \left(\frac{N}{4(1-x)} + \frac{N}{4x} + 1\right)} \\
&= \frac{x(1-x)}{\left(\frac{N}{4(1-x)} + \frac{N}{4x} + 1\right)} \\
&= \frac{x^2(1-x)^2}{\frac{Nx}{4} + \frac{N(1-x)}{4} + x(1-x)} \\
&= \frac{4x^2(1-x)^2}{N + 4x(1-x)}
\end{aligned}
$$

For moderately large N, the variance is approximately

$$Var(Y) \approx \frac{4x^2(1-x)^2}{N}$$

E.1.3 Transient Moments

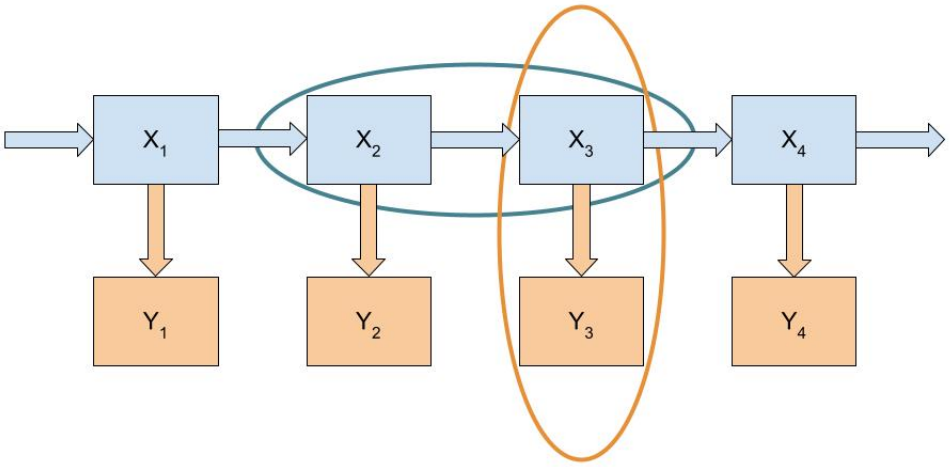
$$
\begin{aligned}
dE[X_t] &= E[dX_t] \\
\frac{dE[X_t]}{dt} &= (h+r)\left(\frac{h}{h+r} - E[X_t]\right) \\
dE[X_t^2] &= E\left[dX_t^2\right] \\
&= E\left[2X_t dX_t + \frac{1}{2}2(dX_t)^2\right] \\
&= E\left[2X_t(h+r)\left(\frac{h}{h+r} - X_t\right)dt + 2X_t\sqrt{\frac{2rh}{N(r+h)}}\sqrt{X_t(1-X_t)}dW_t + 8\frac{rh}{N(r+h)}X_t(1-X_t)dt\right] \\
\frac{dE[X_t^2]}{dt} &= 2hE[X_t] - 2(h+r)E[X_t^2] + 8\frac{rh}{N(r+h)}\left(E[X_t] - E[X_t^2]\right) \\
&= 2\left(h + \frac{4rh}{N(r+h)}\right)E[X_t] - 2\left(h + r + 4\frac{rh}{N(r+h)}\right)E[X_t^2]
\end{aligned}
$$

In matrix notation, we get

$$
\begin{pmatrix} \frac{dE[X_t]}{dt} \\ \frac{dE[X_t^2]}{dt} \end{pmatrix} = \begin{pmatrix} h \\ 0 \end{pmatrix} + \begin{pmatrix} -(h+r) & 0 \\ 2\left(h + \frac{4rh}{N(r+h)}\right) & -2\left(h + r + \frac{4rh}{N(r+h)}\right) \end{pmatrix} \begin{pmatrix} E[X_t] \\ E[X_t^2] \end{pmatrix}
$$

with 'initial' conditions determined by the previous estimated parameter values of the beta initial distribution.

E.1.4 Statistical Methods



E.1.5 MLE with measurement error

Consider the above process p_t. Suppose we have N discretely sampled measurements of the process, $Y_1, Y_2, ..., Y_N$, such that each of N individuals are given a test and the result is a Bernoulli trial; that is, for individual i at time t_k, $E[Y_{i,t_k}|p_{t_k}] \sim Bern(1, X_{t_k})$ and $Y_{t_k} := \sum_{i=1}^{N} Y_{i,t_k} \sim \text{Binom}(N, X_{t_k})$. The correlation between consecutive estimates of prevalence is determined by the process uncertainty.

The joint distribution of the data is

$$P(Y_{t_i}, i \leq N | X_{t_i}, i \leq N, \Theta) = \prod_{i=1}^{N} \int_{x_i} P(Y_{t_i}|X_{t_i}) P(X_{t_i} = x_i|\Theta) dx_i$$

as each measurement is independent conditioned on the current state. This corresponds with the orange oval in the figure.

Note that the ith probability can be written as

$$\int_{x_i} P(Y_{t_i} = y_i | X_{t_i} = x_i) P(X_{t_i} = x_i | \Theta) dx_i$$
$$= \int_{x_i} P(Y_{t_i} = y_i | X_i = x_i) \int_{x_{i-1}} P(X_i = x_i | X_{i-1} = x_{i-1}, \Theta) P(X_{i-1} = x_{i-1}) dx_{i-1} dx_i$$

The one step transition conditioned on the previous value is determined by our SDE. In particular, based on the beta approximation of the distribution of $X_{t_i}|X_{t_{i-1}}$, we only need to know the mean and variance of $X_{t_{i-1}}$ to fully specify the transition density. This inner integral corresponds with the cyan oval in the diagram, propagating our process uncertainty.

$$\int_{x_{n-1}} P(X_n = x_n | X_{n-1} = x_{n-1}, \Theta) P(X_{n-1} = x_{n-1}) dx_{n-1} \approx \frac{1}{B(\alpha_n, h_n)} x_n^{\alpha_n - 1} (1 - x_n)^{\beta_n}$$
$$\alpha_n = f(\alpha_{n-1}, \beta_{n-1}, \Theta)$$
$$\beta_n = g(\alpha_{n-1}, \beta_{n-1}, \Theta)$$

with an initial condition being given by the unknown state at X_1, again assumed to be beta distributed with α_1, β_1 being parameters to be estimated. As each $X_i|X_{i-1}$ is beta distributed and each $Y_i|X_i$ is binomially distributed, this implies

$$Y_i \sim BB(N_i, \alpha_i, \beta_i)$$

Given an $\alpha_1, \beta_1, \Theta$, we can determine the probability of observing each Y_i:

$$
\begin{aligned}
L(\Theta|Y_i, i \leq N) &= \prod_{i=1}^{N} \binom{N_i}{Y_i} \frac{B(Y_i + \alpha_i, N_i - Y_i + \beta_i)}{B(\alpha_i, \beta_i)} \\
&\propto \prod_{i=1}^{N} \frac{B(Y_i + \alpha_i, N_i - Y_i + \beta_i)}{B(\alpha_i, \beta_i)} \\
LL(\Theta|Y_i, i \leq N) &= \sum_{i=1}^{N} \log(B(Y_i + \alpha_i, N_i - Y_i + \beta_i)) - \sum_{i=1}^{N} \log(B(\alpha_i, \beta_i)) \\
\alpha_i &= f(\alpha_{i-1}, \beta_{i-1}, r, h_{i-1}) \\
\beta_i &= g(\alpha_{i-1}, \beta_{i-1}, r, h_{i-1})
\end{aligned}
$$

We can in fact simplify this further through a profile likelihood method for each h_i, optimizing it in terms of $\alpha_{i-1}, \beta_{i-1}, h_{i-1}, r$. Then through this whole procedure, we will be left with just four parameters in this model to optimize: $\alpha_1, \beta_1, h_1, r$. This stays constant regardless of how many data points are included, and a four dimensional optimization problem becomes manageable through standard optimization methods like Broyden's good method.

E.1.6 Conceptual idea:

Individuals are all independent and following a 'binary switch' model. That is, if an individual at time t is uninfected ($X_{i,t} = 0$), then they will wait an exponentially distributed amount of time until they transition with rate β. Similarly, if they are infected ($X_{i,t} = 1$), then they will wait an exponentially distributed amount of time until they transition with rate r.

A population is made up of N of these individuals, $\{X_{i,t}\}_{i=1}^{N}$. Therefore the state of the system is determined jointly by the state of each individual, and is therefore high dimensional.

The data we have consists of the number of tests which came back positive out of some total number of n tests at a given point in time (Y_t). If we assume each test is independent and people are chosen at random, we get $Y_t \sim Binom(n, p_t)$. Therefore to determine the

likelihood of the data we observed, we need to know p_t.

Given the state of the system, we see that

$$\begin{aligned} p_t|\{X_{i,t}\}_{i=1}^N &= \sum_{i=1}^N P(X_{i,t}=1)P(\text{Test } X_{i,t}) \\ &= \frac{1}{N}\sum_{i=1}^N X_{i,t} \end{aligned}$$

assuming each individual is equally likely to be chosen. We know the rules for how each individual changes, which allows us to construct a model for the population. Further, we see that the parameter we need, p_t, is indifferent to the identity of each individual. Therefore we can condition instead on the sufficient statistic:

$$p_t \Big| \sum_{i=1}^N X_{i,t} = \frac{1}{N}\sum_{i=1}^N X_{i,t}$$

which contains significantly less information than the state of the full system.

The dynamics of p_t can be determined on a finite state space over the nonnegative integers from 0 to N scaled by N. For large N, the dynamics are given approximately by the following SDE:

$$dp_t = (\beta + r)\left(\frac{\beta}{\beta + r} - p_t\right)dt + 2\sqrt{\frac{2r\beta}{N(\beta + r)}}\sqrt{p_t(1-p_t)}dW_t$$

No matter what initial distribution we start with, this SDE tends to a beta distribution. Therefore we can approximate the probability of p_t with a beta distribution, meaning we only need to determine the behavior of the parameters (a_t, b_t).

From the SDE, we can derive the behavior of the first two moments:

$$\begin{pmatrix} \frac{dE[X_t]}{dt} \\ \frac{dE[X_t^2]}{dt} \end{pmatrix} = \begin{pmatrix} \beta \\ 0 \end{pmatrix} + \begin{pmatrix} -(\beta + r) & 0 \\ 2\left(\beta + \frac{4r\beta}{N(r+\beta)}\right) & -2\left(\beta + r + \frac{4r\beta}{N(r+\beta)}\right) \end{pmatrix} \begin{pmatrix} E[X_t] \\ E[X_t^2] \end{pmatrix}$$

from which we can moment match to obtain (a_t, b_t). Further these are linear equations which can be solved exactly, meaning that if we know the mean and variance at any point we can project p_t as far into the future as we'd like. Therefore given initial a_0, b_0 for a beta distribution, the solutions to the above equations and the relationship with the parameters give us functions f and g such that

$$a_{t_{i+1}} = f(a_{t_i}, b_{t_i}, \Theta)$$
$$b_{t_{i+1}} = g(a_{t_i}, b_{t_i}, \Theta)$$

with a_0, b_0 being contained in Θ.

We can solve these sequentially. The first can be solved with an integrating factor. Using the initial condition $E[X_{t+s}|X_s] = X_s$, we have

$$E[X_{t+s}|X_s] = X_s e^{-(\beta+r)s} + \frac{\beta}{\beta + r}(1 - e^{-(\beta+r)s})$$
$$E[X_{t+s}] = E[X_s]e^{-(\beta+r)s} + \frac{\beta}{\beta + r}(1 - e^{-(\beta+r)s})$$

by the law of total expectation.

The second moment equation is given by

$$\begin{aligned}\frac{dE[X_t^2]}{dt} &= 2\left(\beta + \frac{4r\beta}{N(r+\beta)}\right)E[X_t] - 2\left(\beta + r + \frac{4r\beta}{N(r+\beta)}\right)E[X_t]^2 \\ &= 2\left(\beta + \frac{4r\beta}{N(r+\beta)}\right)\left[X_t e^{-(\beta+r)t} + \frac{\beta}{\beta+r}(1-e^{-(\beta+r)t})\right] - 2\left(\beta + r + \frac{4r\beta}{N(r+\beta)}\right)E[X_t^2] \\ &= f(t) - 2\left(\beta + r + \frac{4r\beta}{N(r+\beta)}\right)E[X_t^2]\end{aligned}$$

where $f(t)$ represents the entire first term. This equation can again be solved with an integrating factor:

$$E[X_{t+s}^2|X_s] = e^{-2(\beta+r+\frac{4r\beta}{N(r+\beta)}))t}\left[\int f(s)e^{2(\beta+r+\frac{4r\beta}{N(r+\beta)}))s}ds + X_s^2\right]$$

focusing on the integral,

$$\begin{aligned}\int f(s)e^{2(\beta+r+\frac{4r\beta}{N(r+\beta)}))s}ds &= 2\left(\beta + \frac{4r\beta}{N(r+\beta)}\right)\int\left[X_s e^{-(\beta+r)t} + \frac{\beta}{\beta+r}(1-e^{-(\beta+r)t})\right]e^{2(\beta+r+\frac{4r\beta}{N(r+\beta)}))s}ds \\ &= 2\left(\beta + \frac{4r\beta}{N(r+\beta)}\right)X_s\int e^{(\beta+r+\frac{8r\beta}{N(r+\beta)})s}ds \\ &+ 2\left(\beta + \frac{4r\beta}{N(r+\beta)}\right)\frac{\beta}{\beta+r}\left[\int e^{2(\beta+r+\frac{4r\beta}{N(r+\beta)})s}ds - e^{(\beta+r+\frac{8r\beta}{N(r+\beta)})s}ds\right] \\ &= 2\left(\beta + \frac{4r\beta}{N(r+\beta)}\right)\frac{X_s}{\beta+r+\frac{8r\beta}{N(r+\beta)}}\left[e^{(\beta+r+\frac{8r\beta}{N(r+\beta)})s} - 1\right] \\ &+ 2\left(\beta + \frac{4r\beta}{N(r+\beta)}\right)\frac{\beta}{\beta+r}\left[\frac{e^{2(\beta+r+\frac{4r\beta}{N(r+\beta)})s}-1}{2(\beta+r+\frac{4r\beta}{N(r+\beta)})} - \frac{e^{(\beta+r+\frac{8r\beta}{N(r+\beta)})s}-1}{\beta+r+\frac{8r\beta}{N(r+\beta)}}\right]\end{aligned}$$

Plugging in and distributing the exponential on the outside, we have

$$
\begin{aligned}
E[X_{t+s}^2|X_s] = & 2\frac{\beta + \frac{4r\beta}{N(r+\beta)}}{\beta + r + \frac{8r\beta}{N(r+\beta)}} E[X_t] e^{-(\beta+r)t} \left[1 - e^{-(\beta+r+\frac{8r\beta}{N(r+\beta)})t}\right] \\
& + \frac{\beta + \frac{4r\beta}{N(r+\beta)}}{r + \beta + \frac{4r\beta}{N(r+\beta)}} \frac{\beta}{\beta + r} \left(1 - e^{-2(\beta+r+\frac{4r\beta}{N(\beta+r)})t}\right) \\
& - 2\frac{\beta + \frac{4r\beta}{N(r+\beta)}}{r + \beta + \frac{8r\beta}{N(r+\beta)}} \frac{\beta}{\beta + r} e^{-(\beta+r)t} \left(1 - e^{-(\beta+r+\frac{8r\beta}{N(\beta+r)})t}\right) + E[X_t^2] e^{-2(\beta+r+\frac{4r\beta}{N(r+\beta)})t}
\end{aligned}
$$

$\mu = \frac{\beta}{\beta+r}$ is the instantaneous equilbrium point, $\rho_i = e^{-(\beta+r)\Delta t_i}$ is the one step autocorrelation, and $\omega = \frac{4r\beta}{N(\beta+r)^2} = \frac{4\mu(1-\mu)}{N}$ is our variance scaling, simplifying our notation considerably:

$$
\begin{aligned}
E[X_{t+s}|X_s] &= \frac{\mu + w}{\frac{1}{2} + w} E[X_t] \rho (1 - \rho^{1+2w}) + \frac{\mu + w}{1 + w} \mu (1 - \rho^{2(1+w)}) - \frac{\mu + w}{\frac{1}{2} + w} \mu \rho (1 - \rho^{1+2w}) + E[X_t^2] \rho^{2(1+w)} \\
&= \rho (1 - \rho^{1+2w}) \frac{\mu + w}{\frac{1}{2} + w} (E[X_t] - \mu) + \mu \frac{\mu + w}{1 + w} (1 - \rho^{2(1+w)}) + E[X_t^2] \rho^{2(1+w)}
\end{aligned}
$$

Combining with the previous result, we have an affine linear transformation between subsequent values of the first two raw moments:

$$
\begin{aligned}
E[X_{t_{i+1}}] &= \rho E[X_{t_i}] + (1 - \rho)\mu \\
E[X_{t_{i+1}}^2] &= \rho (1 - \rho^{1+2w}) \frac{\mu + w}{\frac{1}{2} + w} (E[X_{t_i}] - \mu) + \mu \frac{\mu + w}{1 + w} (1 - \rho^{2(1+w)}) + \rho^{2(1+w)} E[X_{t+1}^2]
\end{aligned}
$$

From this it's clear the long term values are respectively

$$
\begin{aligned}
E[X_\infty] &= \frac{\beta}{\beta + r} \\
E[X_\infty^2] &= \frac{\beta}{\beta + r} \frac{N\beta(\beta + r) + 4r\beta}{N(\beta + r)^2 + 4r\beta} \\
Var(X_\infty) &= E[X_\infty^2] - E[X_\infty]^2 \\
&= \mu \frac{\mu + w}{1 + w} - \mu^2 \\
&= \frac{\mu^2 + w\mu - \mu^2(1 + w)}{1 + w} \\
&= \frac{w\mu(1 - \mu)}{1 + w} \\
&= \frac{4x^2(1 - x)^2}{N + 4x(1 - x)}
\end{aligned}
$$

where x represents the equilibrium prevalence. This is consistent with our equilibrium calculation of the variance from the beta distribution.

Assuming the initial distribution before our first measurement p_0 is beta distributed as well, we can set a_0, b_0 to be additional parameters to be estimated. Therefore given the parameters, we can project p_0 through p_{t_n}. This means p_{t_i} consists of a correlated sequence of beta random variables parameterized by β_{t_i}, r, a_0, b_0.

Because we have each p_{t_i}, we can condition each Y_{t_i} on its corresponding state and obtain the probability for that observation:

$$
P(Y_{t_i}|\Theta) = \int P(Y_{t_i}|p_{t_i}, \Theta) P(p_{t_i}, \Theta) dp_{t_i}
$$

As p_{t_i} is beta distributed with parameters a_{t_i} and b_{t_i}, and $Y_{t_i}|p_{t_i}$ is binomially distributed with parameters n and p_{t_i}, this implies we have

$$P(Y_{t_i}|\Theta) \sim BetaBinom(n, a_{t_i}, b_{t_i})$$

Each observation, when conditioned on the state of the system, is independent, so the joint probability is the product of the individual probabilities (the overall correlation structure comes from the process itself and the dependence of each a_i, b_i on the previous parameters). This provides our likelihood for the parameters:

$$\begin{aligned}\mathcal{L}(\Theta|\{Y_{t_i}\}_{i=1}^n) &= \prod_{i=1}^n \binom{n}{Y_{t_i}} \frac{B(Y_{t_i}+a_{t_i}, N-Y_{t_i}+b_{t_i})}{B(a_{t_i}, b_{t_i})} \\ &\propto \prod_{i=1}^n \frac{B(Y_{t_i}+a_{t_i}, N-Y_{t_i}+b_{t_i})}{B(a_{t_i}, b_{t_i})}\end{aligned}$$

and our log likelihood is

$$\log\left(\mathcal{L}(\Theta|\{Y_{t_i}\}_{i=1}^n)\right) = \sum_{i=1}^n \log(B(Y_{t_i}+a_{t_i}, N-Y_{t_i}+b_{t_i})) - \log(B(a_{t_i}, b_{t_i}))$$

Part of the reason this approximation works so well is the linearity of the dynamics and the resulting stability of the stationary distribution. However, this method can be extended to other systems which exhibit asymptotic stability. For example, if we were to fit the above model and find that our estimates of β_{t_i} were positively correlated with p_{t_i} as would be expected in a directly transmitted disease, then we will have different dynamics with a different stationary distribution.

This method is meant to approximate the results of methods such as those based on particle filters, but without any stochastic simulation needed to approximate the likelihood. This converts the problem of fitting a stochastic process to data into a deterministic optimization problem.

E.2 Superinfection

Now we can consider the more complex situation in which infections accumulate and clear independently, corresponding to the classic M/M/∞ queueing model. The value followed here is not just prevalence, but the distribution of the number of infections in individuals in a population given infections initiate and are cleared independently of one another. An individual therefore is not bounded to one of two states, but a countable state space. If the number of infections in individual i at time t is denoted $M_{i,t}$, then the average number of infections per person in the population is

$$\bar{M_t} := \frac{1}{N}\sum_{i=1}^{N} X_{i,t}$$

As before, we will want to pass to the embedded discrete time Markov chain in order to apply the same method as before. In a population of size N, the time before the next infection initiates is exponentially distributed with rate

$$\begin{aligned} E[\Delta t_I|\bar{M_t}] &\sim \Delta t_i \\ &\sim Exp(hN) \end{aligned}$$

as they are all independent and all individuals are susceptible to infection in this model. The time before the next recovery given people have a Poisson number of infections with average value $\bar{M_t}$, the time until the next recovery for a person chosen at random with $M_{t,i}$ infections is

$$E[\Delta t_R|M_{t,i}] \sim Exp(rM_{t,i})$$

For a population of size N, the next time to recovery is

$$\begin{aligned} E[\Delta t_R | M_{t,i}, i = 1, ..., N] &\sim \min_{i=1,...,N} Exp(rM_{t,i}) \\ &\sim \min_{i=1,...,N} Exp(r \sum_{i=1}^{N} M_{t,i}) \\ &\sim Exp(r \sum_{i=1}^{N} M_{t,i}) \\ &\sim Exp(rN \frac{1}{N} \sum_{i=1}^{N} M_{t,i}) \\ &\sim Exp(rN\bar{M}_t) \\ &= E[\Delta t_R | \bar{M}_t] \end{aligned}$$

The next time to either a recovery or infection is given by

$$\begin{aligned} E[\Delta t_i | \bar{M}_t] &= \min \left\{ E[\Delta t_I | \bar{M}_t], E[\Delta t_R | \bar{M}_t] \right\} \\ &\sim Exp(N(h + r\bar{M}_t)) \end{aligned}$$

Next, we can look at the state transitions. Given an event occurred, the probability that the event was an infection is

$$\begin{aligned} P\left(\bar{M}_{i+1} = \bar{M}_i + \frac{1}{N} \middle| \bar{M}_i \right) &= P(E[\Delta t_I | \bar{M}_i] < E[\Delta t_R | \bar{M}_i]) \\ &= \frac{h}{h + r\bar{M}_i} \end{aligned}$$

and similarly,

$$P\left(\bar{M}_{i+1} = \bar{M}_i - \frac{1}{N}\middle|\bar{M}_i\right) = \frac{r\bar{M}_i}{h + r\bar{M}_i}$$

as these are the only two transitions that can occur consecutively. Therefore the expected change is

$$\begin{aligned} E[\Delta\bar{M}_i|\bar{M}_i] &:= E[\bar{M}_{i+1} - \bar{M}_i] \\ &= \left(\frac{1}{N}\right) P\left(\bar{M}_{i+1} = \bar{M}_i + \frac{1}{N}\middle|\bar{M}_i\right) - \frac{1}{N} P\left(\bar{M}_{i+1} = \bar{M}_i - \frac{1}{N}\middle|\bar{M}_i\right) \\ &= \frac{h - r\bar{M}_i}{N[h + r\bar{M}_i]} \end{aligned}$$

Next, we can determine the variance in this change. As in the simple infection case, each change has the same maginitude of $\frac{1}{N}$ and so

$$E[(\Delta\bar{M}_i)^2|\bar{M}_i] = \frac{1}{N^2}$$

and therefore the conditional variance is

$$\begin{aligned} Var(\Delta\bar{M}_i|\bar{M}_i) &= E[(\Delta\bar{M}_i)^2|\bar{M}_i] - E[\Delta\bar{M}_i|\bar{M}_i]^2 \\ &= \frac{1}{N^2} - \left(\frac{h - r\bar{M}_i}{N[h + r\bar{M}_i]}\right)^2 \\ &= \frac{1}{N^2}\left(\frac{[h + r\bar{M}_i]^2 - (h - r\bar{M}_i)^2}{[h + r\bar{M}_i]^2}\right) \\ &= \frac{4hr\bar{M}_i}{N^2[h + r\bar{M}_i]^2} \end{aligned}$$

As before, we will define our normalized variable

$$\begin{aligned}\bar{Z}_i &= \frac{\Delta\bar{M}_i - E[\Delta\bar{M}_i|\bar{M}_i]}{\sqrt{Var(\Delta\bar{M}_i|\bar{M}_i)}} \\ &= \frac{\Delta\bar{M}_i - \frac{h-r\bar{M}_i}{N[h+r\bar{M}_i]}}{\sqrt{\frac{4hr\bar{M}_i}{N^2[h+r\bar{M}_i]^2}}}\end{aligned}$$

Setting our timescale to be $\Delta t_i = \frac{1}{N[h+r\bar{M}_i]}$, we can simplify the expression to be

$$\bar{Z}_i = \frac{\Delta\bar{M}_i - [h - r\bar{M}]\Delta t}{2\sqrt{hr\bar{M}_i}\Delta t}$$

we will multiply both sides by the square root of N and sum up to N, starting with the left:

$$\begin{aligned}\sqrt{N}\sum_{i=1}^{N}\bar{Z}_i &= \sqrt{N}\sum_{i=1}^{N}\bar{Z}_i \\ &\sim \mathcal{N}(0, N)\end{aligned}$$

and the right hand side is

$$\sum_{i=1}^{N}\frac{\Delta\bar{M}_i - [h - r\bar{M}]\Delta t}{2\sqrt{\frac{hr\bar{M}_i}{h+r\bar{M}_i}\Delta t}}$$

Multiplying both sides by the square root of Δt_i, we get

$$\sqrt{h + r\bar{M}_i}\mathcal{N}(0, \Delta t_i N) = \sum_{i=1}^{N}\frac{\Delta\bar{M}_t - [h - r\bar{M}]\Delta t}{2\sqrt{\frac{hr\bar{M}_i}{h+r\bar{M}_i}}}$$

In the limit, we have for each time t

$$W_t = \int_0^t \frac{d\bar{M}_t - [h - r\bar{M}_t]dt}{2\sqrt{\frac{hr\bar{M}_t}{h+r\bar{M}_t}}}$$

Differentiating and rearranging, we have

$$d\bar{M}_t = [h - r\bar{M}_t]dt + 2\sqrt{\frac{hr\bar{M}_t}{h + r\bar{M}_t}}dW_t$$

Note that the noise scales with the mean, and so a mean of 0 results in the only possible transition for nonzero parameter values is to increase. This naturally creates the necessary zero-flux boundary that we need.

Further, we know that $\bar{M}_t$ is of order $\frac{h}{r}$ due to the drift, and so:

$$\begin{aligned}\frac{hr\bar{M}_t}{h + r\bar{M}_t} &= \frac{r\bar{M}_t}{1 + \frac{r}{h}\bar{M}_t} \\ &\approx \frac{r\bar{M}}{2}\end{aligned}$$

Plugging in this approximation, we get

$$d\bar{M}_t \approx r\left[\frac{h}{r} - \bar{M}_t\right]dt + \sqrt{\frac{2r}{N}}\sqrt{\bar{M}_t}dW_t$$

The advantage to this approximation is that this approximate SDE has a known Gamma stationary distribution. Therefore as in the simple infection model, we have a simplified superinfection model with a known stationary distribution. Further, the observation process for counting MOI in an individual is Poisson. Therefore the observed number of infections per

individual would be approximately a Poisson-Gamma mixture, that is a Negative Binomial random variable.

This SDE is in fact known in financial mathematics as the Cox-Ingersoll-Ross (CIR) model, where it is used to model the evolution of interest rates.

We can determine the equations for the first two moments as follows:

$$\begin{aligned}
E[d\bar{M}_t] &= E\left[r\left[\frac{h}{r} - \bar{M}_t\right]dt + \sqrt{\frac{2r}{N}}\sqrt{\bar{M}_t}dW_t\right] \\
&= r\left[\frac{h}{r} - E[\bar{M}_t]\right]dt \\
d\bar{M}_t^2 &= \left[r\left(\frac{h}{r} - \bar{M}_t\right)\left(2\bar{M}_t\right) + \frac{1}{2}\left(\frac{2r}{N}\bar{M}_t\right)\right]dt + (2\bar{M}_t)\sqrt{\frac{2r}{N}}\sqrt{\bar{M}_t}dW_t \\
E[d\bar{M}_t^2] &= \left[2hE[\bar{M}_t] - 2rE[\bar{M}_t^2] + \frac{r}{N}E[\bar{M}_t]\right]dt \\
&= 2r\left[\left(\frac{h}{r} + \frac{1}{2N}\right)E[\bar{M}_t] - E[\bar{M}_t^2]\right]dt
\end{aligned}$$

At equilibrium, we have

$$\begin{aligned}
E[\bar{M}]^* &= \frac{h}{r} \\
E[\bar{M}^2]^* &= \left(\frac{h}{r} + \frac{1}{2N}\right)\frac{h}{r} \\
Var(\bar{M})^* &= E[\bar{M}^2]^* - \left(E[\bar{M}]^*\right)^2 \\
&= \left(\frac{h}{r}\right)^2 + \frac{h}{2rN} - \left(\frac{h}{r}\right)^2 \\
&= \frac{h}{2rN}
\end{aligned}$$

which is typically quite small. For example, in a population where people are infected once a year and recover from one infection every 200 days, $\frac{h}{r} \approx 1.3$. The variance is then this quantity divided by twice the population size. For a population of 100 people, the variance

would then be .0065, so the standard deviation would be .08. However if we have just 10 people, the variance is significantly higher at .065 and standard deviation .25. This implies even small sample sizes can give significant information. We can also get estimates for prevalence by noting the relationship between MOI and prevalence is $x = 1 - e^{-m}$, so the expected prevalence is

$$\begin{aligned} E[x] &= 1 - E[e^{-m}] \\ &= 1 - \int_0^\infty e^{-y} P(m = y) dy \\ &= 1 - G_m(-1) \\ &= 1 - \left(1 + \frac{1}{\beta}\right)^{-\alpha} \end{aligned}$$

that is, the moment generating function for the MOI evaluated at -1 given the MOI is gamma distributed. In fact, moment matching yields

$$\begin{aligned} \alpha &= \frac{\left(E[\bar{M}]^*\right)^2}{Var(\bar{M}^*)} \\ &= \frac{\left(\frac{h}{r} + \frac{1}{2N}\right)^2 \left(\frac{h}{r}\right)^2}{\frac{h}{r}\frac{1}{2N}} \\ &= \frac{2h}{r} N \left(\frac{h}{r} + \frac{1}{2N}\right)^2 \\ &\approx 2 \left(\frac{h}{r}\right)^3 N \end{aligned}$$

for large N, and

$$\begin{aligned}\beta &= \frac{E[\bar{M}]^*}{Var(\bar{M}^*)} \\ &\approx 2\left(\frac{h}{r}\right)^2 N\end{aligned}$$

and so we have

$$\begin{aligned}E[x] &= 1 - \left(1 + \frac{Var(\bar{M})}{E[\bar{M}]}\right)^{\frac{E[\bar{M}]^2}{Var(\bar{M})}} \\ &\approx 1 - \left(1 + \frac{1}{2\left(\frac{h}{r}\right)^2 N}\right)^{-2\left(\frac{h}{r}\right)^3 N}\end{aligned}$$

as $N \to \infty$, we'll have

$$E[x] = 1 - e^{-\frac{h}{r}}$$

so we recover the expected value which makes sense as the variance of m goes to zero while its mean remains $\frac{h}{r}$.

These two equations form a 2 by 2 linear system of equations which can be solved to uniquely define the mean and variance of the distribution:

$$\begin{aligned}\frac{d}{dt}\begin{pmatrix} E[\bar{M}_t] \\ E[\bar{M}_t^2] \end{pmatrix} &= \begin{pmatrix} h \\ 0 \end{pmatrix} - \begin{pmatrix} r & 0 \\ -2h - \frac{r}{N} & 2r \end{pmatrix} \begin{pmatrix} E[\bar{M}_t] \\ E[\bar{M}_t^2] \end{pmatrix} \\ \frac{d}{dt}\vec{M}_t &= h\vec{e_1} - A\vec{M}_t\end{aligned}$$

Solving the first equation, we have

$$E[\bar{M}_{t+s}|\bar{M}_s] = \bar{M}_s e^{-rt} + \frac{h}{r}(1 - e^{-rt})$$
$$E[\bar{M}_{t+s}] = E[\bar{M}_s]e^{-rt} + \frac{h}{r}(1 - e^{-rt})$$

We can plug this into the second equation:

$$\begin{aligned}
\frac{dE[\bar{M}_t^2]}{dt} &= (2h + \frac{r}{N})E[\bar{M}_t] - 2rE[\bar{M}_t^2] \\
&= (2h + \frac{r}{N})\left(E[\bar{M}_s]e^{-rt} + \frac{h}{r}(1 - e^{-rt})\right) - 2rE[\bar{M}_t^2] \\
&= f(t) - 2rE[\bar{M}_t^2] \\
E[\bar{M}_t^2] &= e^{-2rt}\left[\int_0^t f(s)e^{2rs}ds + E[\bar{M}_s^2]\right] \\
\int_0^t f(s)e^{2rs}ds &= (2h + \frac{r}{N})E[\bar{M}_s]\int_0^t e^{rs}ds + \left(2\frac{h}{r} + \frac{1}{N}\right)h\int_0^t (1 - e^{-rs})e^{2rs}ds \\
&= \left(2\frac{h}{r} + \frac{1}{N}\right)E[\bar{M}_s][e^{rt} - 1] + \left(2\frac{h}{r} + \frac{1}{N}\right)h\left[\frac{1}{2r}\left(e^{2rt} - 1\right) - \frac{1}{r}\left(e^{rt} - 1\right)\right] \\
&= \left(2\frac{h}{r} + \frac{1}{N}\right)E[\bar{M}_s]\left[e^{rt} - 1\right] + \left(\frac{h}{r} + \frac{1}{2N}\right)\frac{h}{r}\left(e^{2rt} - 1 - 2e^{rt} + 2\right) \\
&= \left(2\frac{h}{r} + \frac{1}{N}\right)\left(e^{rt} - 1\right)E[\bar{M}_s] + \left(\frac{h}{r} + \frac{1}{2N}\right)\frac{h}{r}(e^{2rt} - 2e^{rt} + 1) \\
&= \left(2\frac{h}{r} + \frac{1}{N}\right)\left(e^{rt} - 1\right)E[\bar{M}_s] + \left(\frac{h}{r} + \frac{1}{2N}\right)\frac{h}{r}(e^{rt} - 1)^2
\end{aligned}$$

Multiplying by e^{-2rt}, we get

$$e^{-2rt}\int_s^{t+s} f(s)e^{2rs}ds = \left(2\frac{h}{r} + \frac{1}{N}\right)e^{-rt}(1 - e^{-rt})E[\bar{M}_s] + \left(\frac{h}{r} + \frac{1}{2N}\right)\frac{h}{r}(1 - e^{-rt})^2$$

Combining with the final term, we have

$$E[\bar{M}_t^2] = \left(2\frac{h}{r} + \frac{1}{N}\right) e^{-rt}(1 - e^{-rt})E[\bar{M}_s] + \left(\frac{h}{r} + \frac{1}{2N}\right)\frac{h}{r}(1 - e^{-rt})^2 + e^{-2rt}E[\bar{M}_s^2]$$

Defining $\mu = \frac{h}{r}$, $\rho = e^{-rt}$, we have

$$\begin{aligned} E[\bar{M}_{t+s}] &= E[\bar{M}_s]\rho + \mu(1 - \rho) \\ E[\bar{M}_{t+s}^2] &= \left(2\mu + \frac{1}{N}\right)\rho(1 - \rho)E[\bar{M}_s] + \mu\left(\mu + \frac{1}{N}\right)(1 - \rho)^2 + \rho^2 E[\bar{M}_s^2] \end{aligned}$$

The MOI will follow a Poisson distribution with mean $\bar{M}_t$. $\bar{M}_t$ follows an SDE, whose stationary distribution is Gamma. Given this Gamma distribution with shape α and rate β, the resulting MOI distribution is NB(α, $\frac{1}{1+\beta}$). Above we have the first and second moments. Applying the MoM relations,

$$\begin{aligned} \alpha &= \frac{E[\bar{M}_t]^2}{Var(\bar{M}_t)} \\ \beta &= \frac{E[\bar{M}_t]}{Var(\bar{M}_t)} \end{aligned}$$

so we obtain a posterior NB distribution:

$$NB\left(\frac{E[\bar{M}_t]^2}{Var(\bar{M}_t)}, \frac{Var(\bar{M}_t)}{Var(\bar{M}_t) + E[\bar{M}_t]}\right)$$

The probability of having at least one infection for each individual is then

$$
\begin{aligned}
P(M_t \geq 1) &= 1 - P(M_t = 0) \\
&= 1 - \left(1 - \frac{Var(\bar{M}_t)}{Var(\bar{M}_t) + E[\bar{M}_t]}\right)^{\frac{E[\bar{M}_t]^2}{Var(\bar{M}_t)}} \\
&= 1 - \left(\frac{E[\bar{M}_t]}{Var(\bar{M}_t) + E[\bar{M}_t]}\right)^{\frac{E[\bar{M}_t]^2}{Var(\bar{M}_t)}}
\end{aligned}
$$

Our measurement equation is again a binomial distribution with the probability of detection being equal to the prevalence, that is

$$
Y_t \sim \text{Binom}\left(N_{\text{sample}},\ 1 - \left(\frac{E[\bar{M}_t]}{Var(\bar{M}_t) + E[\bar{M}_t]}\right)^{\frac{E[\bar{M}_t]^2}{Var(\bar{M}_t)}}\right)
$$

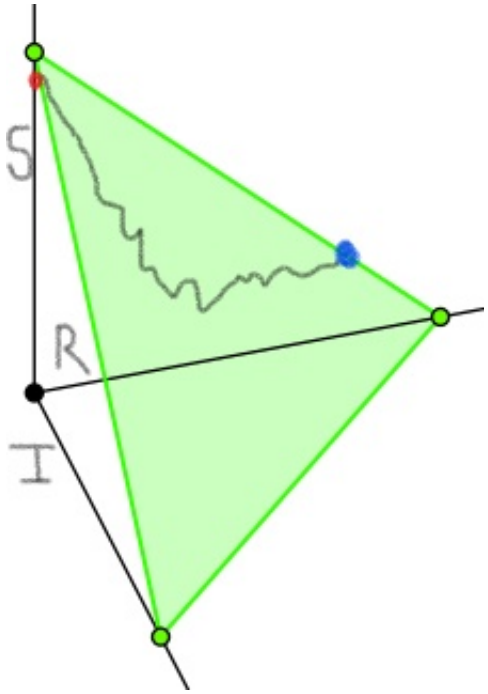


Figure E.1: A single trajectory of the SDE for prevalence in an SIR model. Each trajectory requires $S + I + R = 1$ and all are nonnegative, so the state space is given by the 2-simplex plotted in green.

E.3 SIR Model

Here we will consider this procedure for the classic SIR model for a directly transmitted disease in a closed population of constant size. Unlike the SIS model, this will result in a bivariate distribution on a 2-simplex as knowing the proportions in any pair of states will uniquely determine the third. Additionally, there is a unique absorbing state which the system will tend toward and so no nontrivial stationary distribution can occur without waning immunity or population turnover. Here we will consider S and I, the proportion in the susceptible and infected compartments.

We can also consider the embedded Markov chain by conditioning on an event occurring. The state changes possible are infection going from S to I at a rate βI and recovery going from I to the removed state at a rate γ. This implies that the expected time step between

events is

$$\begin{aligned} E[\Delta t_{i,S}|S_{t_{i-1}}, I_{t_{i-1}}] &= E[t_{i,S} - t_{i-1,S}|S_{t_{i-1}}, I_{t_{i-1}}] \\ &\sim Exp(N[\beta I_{t_{i-1}} S_{t_{i-1}} + \gamma I_{t_{i-1}}]) \end{aligned}$$

The embedded Markov chain then has the following mean and covariance structure:

$$\begin{aligned} E[\Delta I_{t_i}|I_{t_{i-1}}, S_{t_{i-1}}] &= \frac{1}{N}\frac{\beta I_{t_{i-1}} S_{t_{i-1}} - \gamma I_{t_{i-1}}}{\beta I_{t_{i-1}} S_{t_{i-1}} + \gamma I_{t_{i-1}}} \\ &= \frac{1}{N}\frac{\beta S_{t_{i-1}} - \gamma}{\beta S_{t_{i-1}} + \gamma} \\ E[\Delta S_{t_i}|I_{t_{i-1}}, S_{t_{i-1}}] &= -\frac{1}{N}\frac{\beta I_{t_{i-1}} S_{t_{i-1}}}{\beta I_{t_{i-1}} S_{t_{i-1}} + \gamma I_{t_{i-1}}} \\ &= -\frac{1}{N}\frac{\beta S_{t_{i-1}}}{\beta S_{t_{i-1}} + \gamma} \\ E[\Delta I_{t_i}^2|I_{t_{i-1}}, S_{t_{i-1}}] &= \frac{1}{N^2} \\ E[\Delta S_{t_i}^2|I_{t_{i-1}}, S_{t_{i-1}}] &= \frac{1}{N^2}\frac{\beta S_{t_{i-1}}}{\beta S_{t_{i-1}} + \gamma} \\ E[\Delta I_{t_i}\Delta S_{t_i}|I_{t_{i-1}}, S_{t_{i-1}}] &= -\frac{1}{N^2}\frac{\beta S_{t_{i-1}}}{\beta S_{t_{i-1}} + \gamma} \end{aligned}$$

with cancellation of the $I_{t_{i-1}}$ allowed if and only if the epidemic is active (ie, $I_{t_{i-1}} > 0$), otherwise the process will be stopped. We can now use these moments to compute the variances and covariances:

$$
\begin{aligned}
Cov(\Delta I_{t_i}, \Delta I_{t_i} | I_{t_{i-1}}, S_{t_{i-1}}) &= Var(\Delta I_{t_i} | I_{t_{i-1}}, S_{t_{i-1}}) \\
&= E[\Delta I_{t_i}^2 | I_{t_{i-1}}, S_{t_{i-1}}] - E[\Delta I_{t_i} | I_{t_{i-1}}, S_{t_{i-1}}]^2 \\
&= \frac{1}{N^2} - \frac{1}{N^2}\left(\frac{\beta S_{t_{i-1}} - \gamma}{\beta S_{t_{i-1}} + \gamma}\right)^2 \\
&= \frac{1}{N^2}\frac{4\beta\gamma S_{t_{i-1}}}{(\beta S_{t_{i-1}} + \gamma)^2} \\
Cov(\Delta S_{t_i}, \Delta S_{t_i} | I_{t_{i-1}}, S_{t_{i-1}}) &= Var(\Delta S_{t_i} | I_{t_{i-1}}, S_{t_{i-1}}) \\
&= E[\Delta S_{t_i}^2 | I_{t_{i-1}}, S_{t_{i-1}}] - E[\Delta S_{t_i} | I_{t_{i-1}}, S_{t_{i-1}}]^2 \\
&= \frac{1}{N^2}\frac{\beta S_{t_{i-1}}}{\beta S_{t_{i-1}} + \gamma} - \frac{1}{N^2}\left(-\frac{\beta S_{t_{i-1}}}{\beta S_{t_{i-1}} + \gamma}\right)^2 \\
&= \frac{1}{N^2}\frac{\beta^2 S_{t_{i-1}}^2 + \beta\gamma S_{t_{i-1}}}{(\beta S_{t_{i-1}} + \gamma)^2} - \frac{1}{N^2}\frac{\beta^2 S_{t_{i-1}}^2}{(\beta S_{t_{i-1}} + \gamma)^2} \\
&= \frac{1}{N^2}\frac{\beta\gamma S_{t_{i-1}}}{(\beta S_{t_{i-1}} + \gamma)^2} \\
Cov(\Delta S_{t_i}, \Delta I_{t_i} | I_{t_{i-1}}, S_{t_{i-1}}) &= E[\Delta S_{t_i}\Delta I_{t_i} | I_{t_{i-1}}, S_{t_{i-1}}] - E[\Delta I_{t_i} | I_{t_{i-1}}, S_{t_{i-1}}]E[\Delta S_{t_i} | I_{t_{i-1}}, S_{t_{i-1}}] \\
&= -\frac{1}{N^2}\frac{\beta S_{t_{i-1}}}{\beta S_{t_{i-1}} + \gamma} + \frac{1}{N^2}\frac{\beta S_{t_{i-1}} - \gamma}{\beta S_{t_{i-1}} + \gamma}\frac{\beta S_{t_{i-1}}}{\beta S_{t_{i-1}} + \gamma} \\
&= \frac{-\beta^2 S_{t_{i-1}}^2 - \beta\gamma S_{t_{i-1}} + \beta^2 S_{t_{i-1}}^2 - \beta\gamma S_{t_{i-1}}}{N^2(\beta S_{t_{i-1}} + \gamma)^2} \\
&= \frac{-2\beta\gamma S_{t_{i-1}}}{N^2(\beta S_{t_{i-1}} + \gamma)^2}
\end{aligned}
$$

Redefining the common quantity

$$\sigma^2 := \frac{\beta\gamma S_{t_{i-1}}}{N^2(\beta S_{t_{i-1}} + \gamma)^2}$$

we can write the covariance matrix as

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$$

Note that the determinant of this covariance matrix is zero. This makes sense, as if we consider the $(\Delta S, \Delta I)$ plane, the only possible outcomes are $(0, -1)$ for a recovery and $(-1, 1)$ for an infection. This means that knowing the value of either ΔX or ΔY will uniquely determine the value of the other and therefore the value of the random vector. Therefore the rank of the covariance matrix must be 1, and this is a one dimensional noise process.

We can further obtain information about the noise process from the covariance matrix. Diagonalizing, we get

$$\begin{aligned} \Sigma &= SDS^{-1} \\ &= \sigma^2 \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} \frac{2}{5} & \frac{1}{5} \\ -\frac{1}{5} & \frac{2}{5} \end{pmatrix} \\ &= 5\sigma^2 \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{2}{5} & \frac{1}{5} \\ -\frac{1}{5} & \frac{2}{5} \end{pmatrix} \end{aligned}$$

From this form we see we have a similarity transformation between our covariance matrix and a scaled projection onto the y-axis. Therefore we will have a one dimensional projection of a two dimensional Brownian motion as the noise process. Further, we can use this to compute the square root of our covariance matrix. There is a unique positive square root, given by

$$\Sigma^{1/2} = \sqrt{5}\sigma \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{2}{5} & \frac{1}{5} \\ -\frac{1}{5} & \frac{2}{5} \end{pmatrix}$$
$$= \frac{\sqrt{5}}{5}\sigma \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$$

Therefore the process will consist of the following:

$$E\left[\begin{pmatrix} \Delta S_{t_i} \\ \Delta I_{t_i} \end{pmatrix} \Big| \begin{pmatrix} S_{t_{i-1}} \\ I_{t_{i-1}} \end{pmatrix} \right] = \begin{pmatrix} \frac{\beta S_{t_{i-1}}}{\beta S_{t_{i-1}} + \gamma} \\ \frac{\beta S_{t_{i-1}} - \gamma}{\beta S_{t_{i-1}} + \gamma} \end{pmatrix} + \sqrt{\frac{\beta\gamma S_{t_{i-1}}}{5}} \frac{1}{N(\beta S_{t_{i-1}} + \gamma)} \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} \Delta W_{1,t_i} \\ \Delta W_{2,t_i} \end{pmatrix}$$

where the two noise processes $W_{1,t}, W_{2,t}$ are independent.

Now we will incorporate our timescale back into the process as before. The expected value of our subordinator (time process) was

$$E[\Delta t_i | S_{t_{i-1}}, I_{t_{i-1}}] = \frac{1}{N(\beta I_{t_{i-1}} S_{t_{i-1}} + \gamma I_{t_{i-1}})}$$

Setting this as our timescale and taking limits as before, we get

$$\begin{pmatrix} dS_t \\ dI_t \end{pmatrix} = \begin{pmatrix} -\beta S_t I_t \\ \beta S_t I_t - \gamma I_t \end{pmatrix} dt + \sqrt{\frac{\beta\gamma S_t I_t}{5N(\beta S_t + \gamma)}} \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} dW_{1,t} \\ dW_{2,t} \end{pmatrix}$$

Note again that our fluctuations are scaled by a quantity which is defined if and only if $I_t > 0$, and is set to 0 if $I_t = 0$ as the system is at equilibrium. We can rearrange slightly by dividing both the numerator and denominator of the noise scaling factor by γ to get

$$\begin{pmatrix} dS_t \\ dI_t \end{pmatrix} = \begin{pmatrix} -\beta S_t I_t \\ \beta S_t I_t - \gamma I_t \end{pmatrix} dt + \sqrt{\frac{\beta S_t I_t}{5N(R_0 S_t + 1)}} \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} dW_{1,t} \\ dW_{2,t} \end{pmatrix}$$

where $R_0 = \frac{\beta}{\gamma}$ is the classic basic reproduction number.

We can define a dimensionless time so that $\tau = \gamma t$, which sets our timescale to be relative to the average duration of an infection:

$$\begin{pmatrix} dS_\tau \\ dI_\tau \end{pmatrix} = \begin{pmatrix} -R_0 S_\tau I_\tau \\ R_0 S_\tau I_\tau - I_\tau \end{pmatrix} d\tau + \sqrt{\frac{R_0 S_\tau I_\tau}{5N(R_0 S_\tau + 1)}} \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} dW_{1,\tau} \\ dW_{2,\tau} \end{pmatrix}$$

We can compare this to the classic stochastic derivation from the Poisson process interpretation:

$$\begin{pmatrix} dS_\tau \\ dI_\tau \end{pmatrix} = \begin{pmatrix} -R_0 S_\tau I_\tau \\ R_0 S_\tau I_\tau - I_\tau \end{pmatrix} d\tau + \begin{pmatrix} -\sqrt{\beta SI} & 0 \\ \sqrt{\beta SI} & -\sqrt{\gamma I} \end{pmatrix} \begin{pmatrix} dW_{1,\tau} \\ dW_{2,\tau} \end{pmatrix}$$

Although the drift terms are the same, the noise is clearly scaled differently. One oddity of the new method is all stochasticity disappears if the susceptible population is depleted entirely, whereas the traditional approach allows for noise in a recovery-only process to continue. Another difference is the scaling with population size; in the Poisson process interpretation, the noise scales linearly with the mean and is invariant to the population size and so the infinite population limit will still result in the same scaled noise rather than converging to the deterministic limit. Finally the covariance matrix of the classic method is full rank, implying a two dimensional correlated diffusion instead of a reduced one-dimensional diffusion as we have in our current model.

A further reduction of the new model can be done by defining the new diffusion $Z_\tau :=$

$W_{1,\tau} - 2W_{2,\tau}$ which is a diffusion following the SDE

$$dZ_\tau = \sqrt{5} dW_t$$

Then this allows us to write

$$\begin{pmatrix} dS_\tau \\ dI_\tau \end{pmatrix} = \begin{pmatrix} -R_0 S_\tau I_\tau \\ R_0 S_\tau I_\tau - I_\tau \end{pmatrix} d\tau + \sqrt{\frac{R_0 S_\tau I_\tau}{N(R_0 S_\tau + 1)}} \begin{pmatrix} 1 \\ -2 \end{pmatrix} dW_\tau$$

where the noise is captured by a single diffusion.

If we focus just on the I_τ equation for fixed S_τ, we see that it is of the form

$$dI_\tau = -\mu I_\tau d\tau + \sigma \sqrt{I_\tau} dW_\tau$$

where $\mu(S_\tau) = R_0 S_\tau - 1$ and $\sigma(S_\tau) = -2\sqrt{\frac{R_0 S_\tau}{N(R_0 S_\tau + 1)}} dW_\tau$. Initially $\mu > 0$ if $R_0 > \frac{1}{S_0}$, but then it is guaranteed to drop below zero and stay there indefinitely. Note that there is an analogous equation for S_τ given I_τ. Therefore the conditional processes are mean-reverting to the origin with noise proportional to their current level.