A Hidden Markov Model of Horizontal Gene Transfer in Bacteria

A Thesis

Presented in Partial Fulfillment of the Requirements for the
Degree of Master of Science
with a

Major in Probability and Statistics
in CIMAT

by

Rosana Teresa Zenil López

Major Professor: José Miguel Ponciano Castellanos, Ph.D.

November 28, 2008

Acknowledgements

My deep gratitude to Dr. José Miguel Ponciano for his guidance and advice during these 18 months that ended in this thesis.

I would like to thank my committee members Dr. Andrés Christen and Dr. Miguel Nakamura who greatly enriched my knowledge with their exceptional insights into statistics.

I would like to convey thanks to CONACYT for providing the financial means to my studies. My gratitude to CIMAT that provided all the means to study and research. I have to thank also COMECYT for the thesis scholarship that allowed me to finish this work.

To my father Jorge and my mother Ma. del Rosario for their endless love and understanding that allowed me to continue my studies.
To my brother Jorge for being honest and trustful, you have always been my strength.
To my sister Jessica for accepting us as your family.
To my sister Lucía who always has a smile for everyone, you are becoming a wonderful mother.
To my nice Macarena for being so sweet and making me laugh all the time.
To my grandparents Coco and Yeye your love and company have been my inspiration to grow as a better person.
To the López family, thank you for the wonderful times, for all the differences that make you all special and unique.

Contents

List of Figures

ABSTRACT

Horizontal Gene Transfer (HGT) is a general term used in Biology referring to the exchange of genes between two individuals in the same generation. HGT plays an important role in bacterial adaptation and evolution. In harsh environments, the acquisition of genetic material that allows bacterial adaptation occurs through the transmission of circular DNA molecules known as plasmids. In this work, a hierarchical model is proposed to explain and predict the plasmids' population dynamics when HGT is taken into account. This model is an extension of the deterministic and stochastic models presented by Ponciano et al. (2007). The hierarchical model consists of two parts: the first part is a hidden Markov model (HMM) that incorporates the relevant biological processes. This Markov model follows the proportion of plasmid-free cells over many generations. In laboratory experiments called "stability experiments", the growth of the fraction of plasmid-free cells is monitored daily. The data produced by such experiments consists of time series of the observed fraction of plasmid free cells. In order to connect such data sets with the Markov model, the second part of the hierarchical model describes the statistical sampling scheme from the Markov model and here we chose to use a binomial distribution whose success probability changes according to the HMM. We developed intensive computational methods based on Markov Chain Monte Carlo (MCMC) to be able to estimate the model parameters with real, experimental data sets. These parameter estimates were used to estimate the plasmids' persistence times for simulated data sets. The statistical tools developed here were thoroughly tested via simulations. Finally, the methods developed in this work are suitable to estimate persistence times in antibiotic-resistant bacteria populations and they can be easily extended to more biologically realistic scenarios.

1 Introduction

Horizontal Gene Transfer (HGT), the exchange of genes between two individuals in the same generation, plays an important role in bacterial adaptation and evolution. This biological process allows individual bacteria to sample genetic material from a diverse gene pool in a given population. Recently, Gogarten and Towsend (1995) showed that HGT greatly complicates the evolutionary history of many microbial organisms and that their phylogenetic histories usually cannot be elucidated unless HGT is accounted for. In bacteria, HGT has evolved through plasmids, which are extrachromosomal circular DNA molecules capable of autonomous replication. The transmission of plasmids between two individual bacteria occurs through a process known as conjugation. Plasmids can carry genetic material that can be advantageous to their hosts. This is in fact the mechanism through which bacteria acquire genes that code for antibiotics resistance or genes that allow them to cope with harsh environments (Genereux and Bergstrom 1999).

Two important mechanisms govern the dynamics of plasmids in a population: conjugation and segregation. Often, the quick spread of antibiotic resistance in a population occurs because plasmids are quickly spread through conjugation (Genereux and Bergstrom 1999). Segregation occurs if during bacterial fission all of the plasmids present in the mother cell segregate to a single daughter cell. If there are m plasmids in the mother cell for example, the fact that one of the daughter cells may inherit $0, 1, 2, 3, \ldots, m$ plasmids makes of segregation a chance mechanism. This argument is the basis of the stochastic model construction of the plasmid dynamics done by Seneta and Tavaré (1983) using branching processes. In the absence of antibiotic resistance, the advantageous genes carried by the plasmid are no longer necessary and cells that lose the plasmid *via* segregation reproduce faster than the others (Levin and Stewart 1977, 1980, Bergstrom *et al.* 2000, Lili *et al.* 2007, Slater *et al.* 2008). In such cases, the faster spread of the plasmid-free cells has been taken as evidence that hosting a

plasmid imposes a metabolic cost for the bacteria, and ultimately a fitness cost. As a consequence, it is a delicate balance between conjugation, segregation and fitness cost what determines the persistence of plasmids in a natural population (Stewart and Levin 1980, Bergstrom *et al.* 2000). Modeling these key processes when this balance occurs is essential to understand plasmids' persistence (De Gelder *et al.* 2004, Bergstrom *et al.* 2000).

An understanding of the plasmids' dynamics can be critical to design medical treatments. If the plasmids' cost is high enough, in the absence of antibiotics plasmid-free cells will tend to sweep over in a bacterial population. This is the reason why, if a patient is infected with an antibiotic-resistant strain, a medical treatment consists on decreasing or completely discontinuing the use of antibiotics (Genereux and Bergstrom 1999). However the tendency to disappear is not deterministic and seems to fluctuate widely (De Gelder *et al.* 2004, Ponciano *et al.* 2007). There are relevant questions that can only be answered properly through a careful modeling of the plasmids' dynamics. One of them is knowing what are the values of the fitness cost imposed by plasmids that guarantee their disappearance from a population of interest. Also, one may ask what are the consequences of the wide fluctuations reported by De Gelder *et al.* (2007), and what is the expected time to be waited until 99.9% of the bacteria disappear.

To tackle these questions, data is obtained by carrying "segregation experiments". In these experiments, a bacterial population is seeded with 100% plasmid-carrying individuals in an antibiotic-free medium. Then, bacteria are allowed to evolve (duplicate) during 24 hrs and a sample of, say, 50 cells is taken day after day, for many days (*e.g.* 20). By carrying stability experiments with seven strains, De Gelder *et al.* (2004) observed that the estimated fraction of the plasmid-free bacteria increased each day but this increase was "quite slow and erratic" and while

there was an average (of plasmid-free cells) trend upward there was a considerable amount of variability (Joyce *et al.* 2005). Although the amount of variability in the data seem to be higher than what would be expected under pure random sampling from a deterministic trend, most of the plasmids' dynamics models have been built in a deterministic framework using either differential or difference equation systems. Levin and Stewart (1977) modeled the population dynamics of plasmids governed by conjugation and plasmid lost via segregation and found the biological conditions under they persist. Later, in 1979, they considered conditions in which non self-transmissible extrachromosomal elements were maintained. Bergstrom *et al.* (2000) reviewed the plasmids' dynamics literature and proposed that growth and conjugation rates should be functions of the time (measured in generations). These authors also found the mathematical expressions describing precisely the conditions for plasmid persistence expressed as a function of conjugation and segregation rates. Later, De Gelder *et al.* (2004) built a system of difference equations to model the fraction of plasmid-free cells and connected this model to the data via rigorous statistical methods.

The main goal of the present work is to provide practitioners with the tools to answer some of the relevant questions mentioned above by means of a stochastic model. We develop and test the methods to fit and evaluate the proposed model, given data from segregation experiments (De Gelder *et al.* 2004) . As Novozhilov *et al.* (2005) point out, the use of stochastic processes to model HGT is readily justified when it is considered that the biological principles under which evolution and adaptation are understood are essentially stochastic. In 2008, Slater *et al.* wrote that deterministic models "fail to take into account some of the physiological and ecological complexities of plasmids and their hosts in natural environments". If the plasmids' environment is understood as its host, and one considers that the host qualities may widely change from individual bacteria to another, then a model with environmental

uncertainty might be a good choice to incorporate stochasticity in a plasmids' dynamics model (Ponciano *et al.* 2007). Following this argument, the model we propose incorporates environmental stochasticity and extends the hierarchical model proposed by Ponciano *et al.* (2007).

2 Modeling Horizontal Gene Transfer

In this section we present a Markov Model (MM) for HGT in bacteria. This model is an extension of the Hidden Markov Model (HMM) developed by Ponciano *et al.* (2007) in which horizontal plasmid transfer is modelled explicitly. The model's determistic "skeleton" is a difference equation system also described by these authors. In what follows, we first summarize the deterministic model properties and then we characterize the MM in detail.

2.1 Mathematical Model

De Gelder *et al.* (2004) established a system of difference equations to model the growth of the fraction of plasmid free cells in an anti-biotic free medium. They assumed cells can lose plasmids but they cannot acquire it again. The number of plasmid-free cells at time t is denoted as m_t and the number of plasmid-carrying cells is denoted as n_t. In this system, the abundance of plasmid-free cells m_t was given by the sum of the m_{t-1} plasmid-free cells that grew at a rate $2^{1+\sigma}$ and the n_{t-1} plasmid-carrying cells that lost their plasmid at a segregation rate λ and doubled in number from time $t-1$ to time t. The parameter σ was called the selection coefficient, or plasmid cost. If $\sigma > 0$, the selection coefficient represents the advantage in growth speed that plasmid-free cells have over plasmid-carrying cells. On the other hand, the number of plasmid-carrying cells at time t, n_t, was simply given by twice the number of plasmid-carrying at time $t-1$ that did not loose their plasmid. Hence,

$$\begin{aligned} m_t &= 2^{1+\sigma} m_{t-1} + 2\lambda n_{t-1}, \\ n_t &= 2(1-\lambda) n_{t-1}. \end{aligned} \tag{1}$$

Then, plasmid-free cells' proportion x_t was

$$x_t = \frac{m_t}{m_t + n_t}. \tag{2}$$

Using this basic formulation, Ponciano *et al.* (2007) considered a model extension to account for the fact that bacterial cells can acquire plasmid via conjugation.

These authors assumed that the plasmid conjugation rate, denoted by γ, depended linearly not only on the relative abundance of plasmid-carrying and plasmid-free cells, but also on their probability of encounter given by

$$\frac{(1-x_{t-1})}{\theta+(1-x_{t-1})} =: \Phi(x_{t-1}).$$

This modification to the mass-action principle is the well-known Michaelis-Menten formula that has been used before in a similar context (Ponciano *et al.* 2007). Also, in the ecological literature, this functional form has been derived from first principles as a probability of encounter between two individuals of different sex in a population (Dennis, 1989). The parameter θ is interpreted as the fraction of plasmid-carrying cells at which the frequency of conjugations is half its maximum. Hence, the model became

$$\begin{aligned} m_t &= \left(1-\gamma\tfrac{(1-x_{t-1})}{\theta+(1-x_{t-1})}\right)2^{1+\sigma}m_{t-1}+2\lambda n_{t-1}, \\ n_t &= 2(1-\lambda)n_{t-1}+2^{1+\sigma}m_{t-1}\gamma\tfrac{(1-x_{t-1})}{\theta+(1-x_{t-1})}. \end{aligned} \tag{3}$$

From Equation 3, the fraction of plasmid-free cells at time t was found to be

$$x_t = \frac{\left(1-\gamma\frac{(1-x_{t-1})}{\theta+1-x_{t-1}}\right)x_{t-1}2^{1+\sigma}+2\lambda(1-x_{t-1})}{2^{1+\sigma}x_{t-1}+2(1-x_{t-1})}. \tag{4}$$

By carrying a standard stability analysis (Kot 2001) of the difference Equation 4, unique deterministic properties of the model were found. In particular, three equilibrium solutions to equation 4 are possible. The first two were obtained as solutions of the quadratic equation $Ax^2+Bx+C=0$, where $A=2^\sigma-1$, $B=((\theta+1)(2^\sigma-1)+\lambda+2^\sigma\gamma)$ and $C=-\lambda(\theta+1)$. The resulting solutions are $x_1=1$, $x_2=\frac{-B+\sqrt{B^2-4AC}}{2A}$ and $x_3=\frac{-B-\sqrt{B^2-4AC}}{2A}$.

The behavior of the stable solutions and of the growth rate of the fraction of plasmid-free cells depends strongly on the value of σ (see Figure 1). When the parameter σ is too small, plasmid-free cells replicate as fast as plasmid-carrying ones so that the plasmids may never disappear from the population. When this occurs,

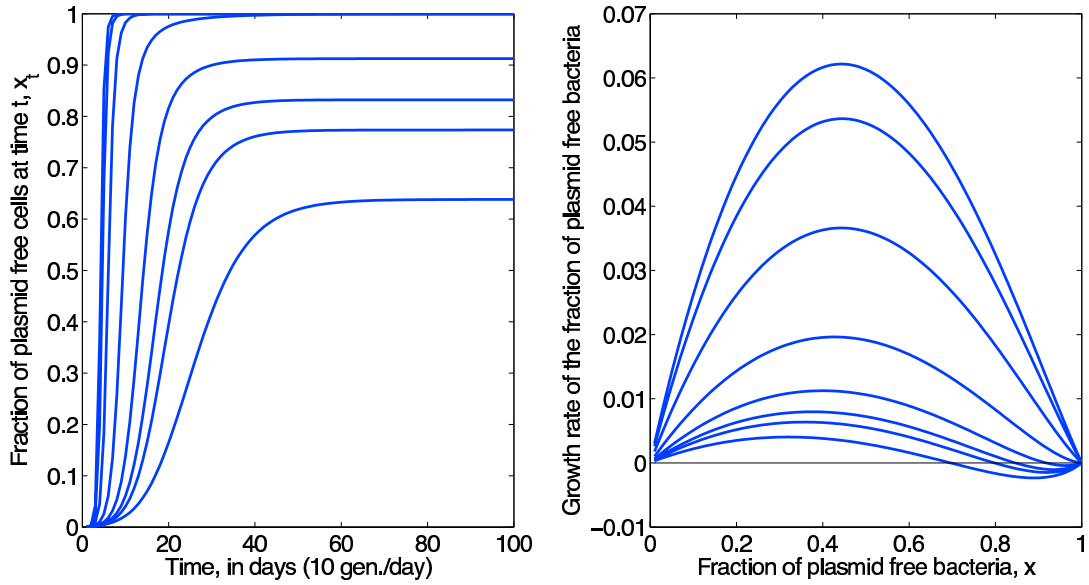

Figure 1: Left panel, behavior of the fraction of plasmid-free cells (x_t) at time t is shown for different values of σ. In the lowest curve $\sigma = 0.055$ and for the uppermost curve $\sigma = 0.41$. Right panel, the growth rate of the fraction of plasmid-free cells ($x_t - x_{t-1}$) is shown. The intersections with the x-axis shows the location of the stable solutions (Taken from Ponciano *et al.* 2007) .

the solution x_2, which is smaller than 1, is stable. The solution $x_1 = 1$ will become stable if and only if

$$\frac{\gamma}{\theta} \leq 1 - \frac{1-\lambda}{2^{\sigma}} \tag{5}$$

holds. Under this last solution all cells lose their plasmid and they become sensitive to antibiotic treatment.

The behavior of stable solutions is shown in Figure 1, where different stable solutions for different values of the parameter σ are plotted. Also in Figure 1, the growth rate of plasmid-free cells' fraction is shown. The different equilibria are the points that intersect the x-axis.

The actual stability experiment carried by De Gelder *et al.* (2004), De Gelder

et al. (2007) and Ponciano *et al.* (2007), showed that the behavior of the fraction of plasmid-free cells might not be deterministic. As shown in Figure 2, the variability of the observations appears to be bigger than would be expected under a simple binomial process, like the one used to find the ML estimates of the deterministic model by De Gelder *et al.* (2004). To model this extra variability we assumed that the growth process of the proportion of plasmid-free cells is a stochastic process. Then, the growth rate of plasmid-free cells becomes a random variable. This model is explained in detail in the next section.

2.2 Stochastic Model

A stochastic model for the growth of the proportion of plasmid-free cells can include two different sources of variability: demographic stochasticity and environmental noise. Demographic stochasticity refers to the variability given by the random loss and acquisitions of plasmids in the population, whereas environmental noise is related to "the effect of external factors on the individuals in the population (Ponciano *et al.* 2005). A State-Space Model (SSM) has been shown to be an adequate tool to model different sources of randomness (Dennis *et al.* 2006). For instance, Clark and Bjørnstad (2004) and Dennis *et al.* (2006) have showed that SMM are suitable to estimate adequately different sources of variability: the process uncertainty consisting of environmental noise and/or demographic stochasticity and observation error given by data. In the experiment mentioned before, a SSM would integrate observational error (related to the fact that we do not know the total number of plasmid-free cells) and process error (related to the uncertainty of the growth of plasmid-free cells).

The model equation in (4) does not consider process error. To introduce process variability, Ponciano *et al.* (2007) proposed that the fitness cost σ might be drawn from a continuous probability distribution denoted in their model as S_t, where $S_t \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \tau^2)$. Here we make the same assumption. According to the distinctions

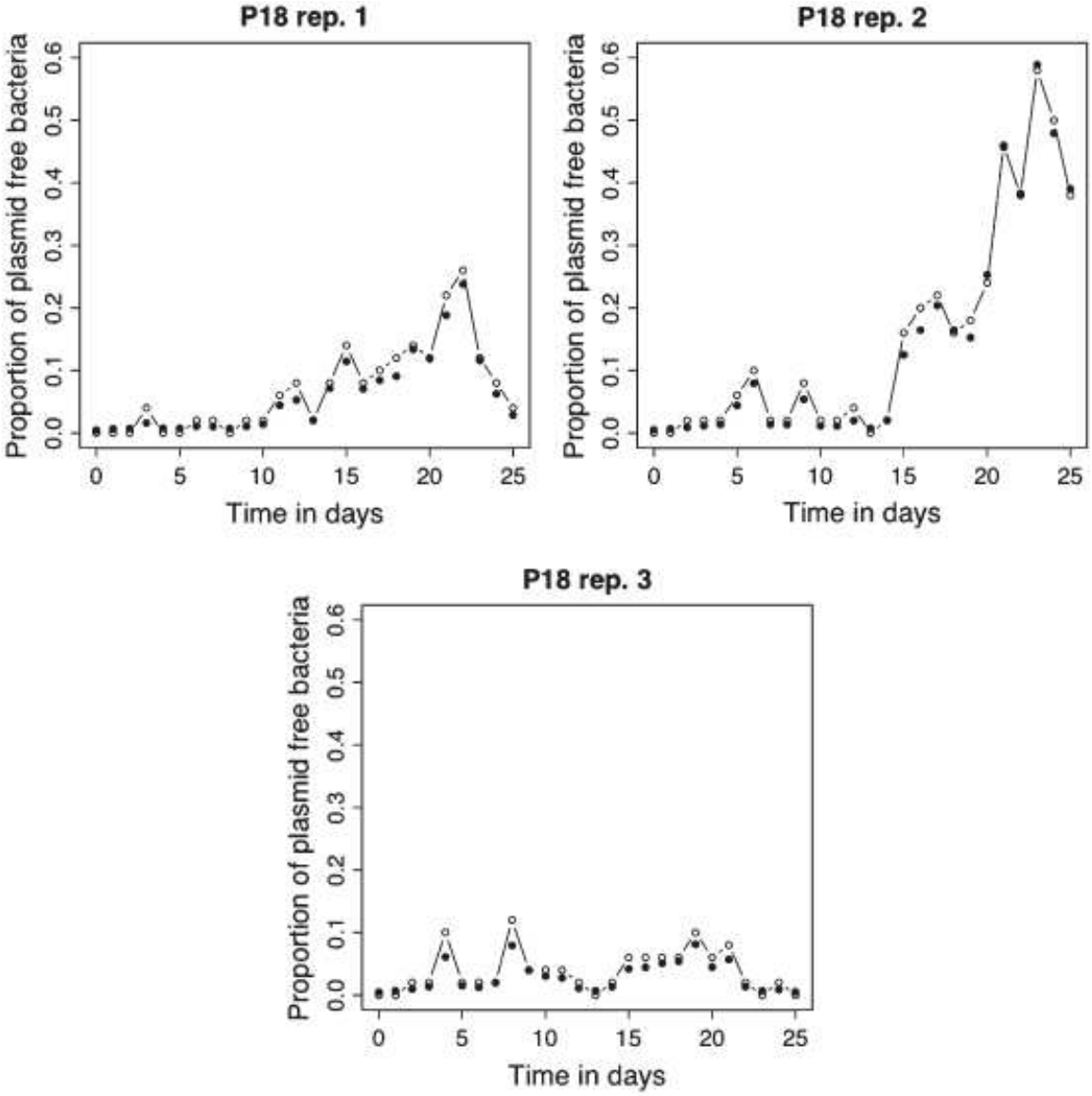

Figure 2: The proportion of plasmid-free cells for bacteria P18, where variability seems bigger than expected. Taken from Ponciano *et al.* (2007).

mentioned above regarding the different types of process variability, including fitness as a random variable amounts to specify a model for environmental noise. Ponciano *et al.* (2007) emphasized that due to theoretical and experimental background it is feasible to consider that "the occurrence of compensatory mutations and/or a variable host-dependent plasmid fitness cost would dramatically alter the plasmid loss dynamics". In other words, there would be periods where loss of plasmids would be severe

and other periods where their loss would not be as severe. Therefore, a stochastic formulation for the proportion of plasmid-free cells seems biologically reasonable.

Lewontin and Cohen (1969) showed that environmental noise can be modeled by assuming that the growth rate is a normal random variable. In stochastic process theory, the introduction of environmental noise can be derived as follows: suppose that population growth is modeled with a density-dependent branching process model, like a standard Galton-Watson process with non-overlaping generations, where the mean of the offspring distribution depends on the current population density. Let Z_t denote the total number of individuals at time t. Then, $Z_t = \sum_{i=1}^{t} N_i$ where $N_i \sim g(m, v)$, $m = E(N_i)$ and $v = V(N_i)$. This model not only assumes that all the individuals are identical with respect to the offspring distribution, but also that from year to year the offspring distribution remains the same. Environmental noise can be introduced in this model by simply letting the mean of this offspring distribution be a random variable that changes from year to year $m = m(t)$ (see Tier and Hanson 1981).

By introducing environmental stochasticity in the deterministic model in (4), the fraction of plasmid-free cells X_t becomes a random variable that is a function of S_t and the realization of X at time $t-1$, X_{t-1}. Therefore, $(X_t, \quad t \geq 0)$ is an unobservable Markov process (Hidden Markov Model).

The relationship between S_t and X_t is

$$X_t = \frac{\left(1 - \gamma \frac{(1-X_{t-1})}{\theta+1-X_{t-1}}\right) X_{t-1} 2^{1+S_t} + 2\lambda(1 - X_{t-1})}{2^{1+S_t} X_{t-1} + 2(1 - X_{t-1})} \tag{6}$$

$$= g(S_t). \tag{7}$$

Here, X_t is a Markov chain, where t represents the number of generations. The transition density function for this Markov chain is found by using the change of

variable theorem:

$$f_{X_t|X_{t-1}}(x_t|x_{t-1}) = \frac{1}{\tau\sqrt{2\pi}}\left|\frac{\partial g^{-1}(x_t)}{\partial x_t}\right| exp\left(-\frac{(g^{-1}(x_t)-\mu)^2}{2\tau^2}\right) I_{(\lambda,\infty)}(x_t) \qquad (8)$$

where

$$\begin{aligned} g^{-1}(x_t) &= \frac{1}{ln(2)} ln\left(\frac{(1-x_{t-1})(x_t-\lambda)(\theta+1-x_{t-1})}{x_{t-1}(\theta+1-x_{t-1})(1-x_t)+x_{t-1}(x_{t-1}-1)\gamma}\right) \\ \frac{\partial g^{-1}(x_t)}{\partial x_t} &= \frac{1}{ln(2)}\left(\frac{1}{x_t-\lambda}+\frac{\theta+1-x_{t-1}}{\theta+1-x_{t-1}-(1-x_{t-1})\gamma}\right). \end{aligned}$$

The function $g(\dot{})$ linking X_t with S_t is invertible and this property allow us to write explicitly the transition probability (8). On the other hand S_t is a normal random variable and this allow us to easily simulate the process X_t as it is shown in Figure 3.

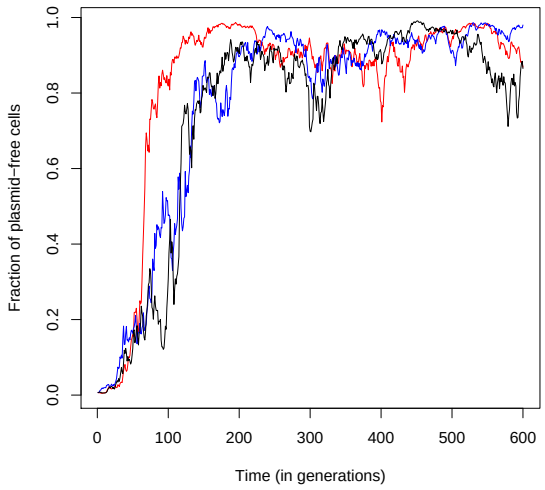

Figure 3: Simulation of three realizations of $(X_t, \quad t \geq 0)$.

The stochastic dynamics model above is characterized by three unique Markov model properties: the transition probability distribution function, its stationary probability distribution function (if it exists) and the plasmids' persistence times.

The transition probability distribution $f_{X_t|X_{t-1}=x_{t-1}}(x_t|x_{t-1})$ looks like a Beta distribution when $X_{t-1} = x_{t-1}$ is set near to 1 and (5) holds. It is possible to fit a

Beta distribution with parameters (α, β) by estimating α and β using the method of moments. To do that, we first approximate the first and second moments of the chain via a second order Taylor series expansion as follows:

$$X_t|X_{t-1} = x_{t-1} \approx g(\mu) + \frac{\partial g(S_t)}{\partial S_t}(S_t - \mu) + \frac{1}{2}\frac{\partial^2 g(S_t)}{\partial S_t^2}(S_t - \mu)^2. \tag{9}$$

Then, using properties of expectation and variance we have:

$$\begin{aligned}
m = E(X_t|X_{t-1} = x_{t-1}) &\approx g(\mu) + \frac{1}{2}\frac{\partial^2 g(S_t)}{\partial S_t^2}E((S_t - \mu)^2) \\
&\approx g(\mu) + \frac{\partial^2 g(S_t)}{\partial S_t^2}\tau^2, \qquad \text{and} \\
v = Var(X_t|X_{t-1} = x_{t-1}) &\approx \left(\frac{\partial g(S_t)}{\partial S_t}\right)^2 Var(S_t - \mu) + \frac{1}{4}\frac{\partial^2 g(S_t)}{\partial S_t^2}Var(S_t - \mu)^2 \\
&\approx \left(\frac{\partial g(S_t)}{\partial S_t}\right)^2 \tau^2 \frac{1}{4}\frac{\partial^2 g(S_t)}{\partial S_t^2}(Var(S_t^2) + 4\mu^2 Var(S_t)) \\
&\approx \left(\frac{\partial g(S_t)}{\partial S_t}\right)^2 \tau^2 \frac{1}{4}\frac{\partial^2 g(S_t)}{\partial S_t^2}(2(\mu^2 + \tau^2)\tau^3 + 4\mu^2\tau^2).
\end{aligned}$$

Finally, the approximated moment parameter estimates are

$$\tilde{\alpha} = \tilde{\beta}\frac{m}{1-m}, \qquad \text{and} \tag{10}$$

$$\tilde{\beta} = m\frac{(1-m)^2}{v + m - 1}. \tag{11}$$

In Figure 4, $x_{t-1} = 0.9$ was fixed and one forward step of the chain was simulated. The histogram represents simulations and the curve is a Beta density with parameters $(\tilde{\alpha}, \tilde{\beta})$.

The stationary law is one of the interesting properties but calculating it for this process is not an easy task. Properties for an ergodic chain (aperiodicity and irreducibility, Karlin 1975) need to be proved in the continuous state space for the fraction of plasmid-free cells. Another approach could be using a diffusion process approximation which is again not an easy job (Meyn and Tweedie 1993). However simulations of the process showed that a stationary law for this Markov chain might

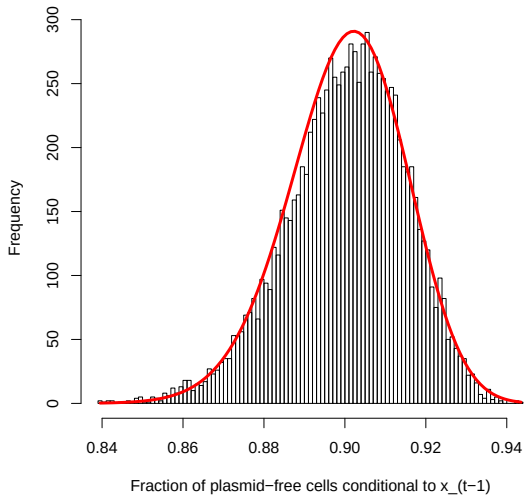

Figure 4: Simulations of X_t conditional to $X_{t-1} = 0.9$. A Beta distribution (solid line) was fitted to the simulated data using the method of moments.

exist under certain conditions. The mode of this stationary law seems to be close to 0.9.

The last property refers to persistence times. Let $T_k^{x_0} = \min\{t > 0 : X_t \geq k | X_0 = x_0\}$ be the first time at which the Markov chain X_t is greater or equal than k given that the chain started at x_0. Then, it is possible to simulate these stoping times using the transition pdf, as it is shown in Figure 5.

2.3 The Hierarchical Model

A hierarchical model $\underline{Y}$ is a statistical model that has random components and fixed parameters. They have been widely used in Ecology and Biology. For example, realistic structures using hierarchical models for capture-recapture problems were worked by George and Robert (1992) and Basu and Ebrahimi (2001). Problems related to fisheries stock assessment were treated by Meyer and Millar (1999). Later, Gelfand *et al.* (2005) tackled a problem of geospatial models of species and habitats

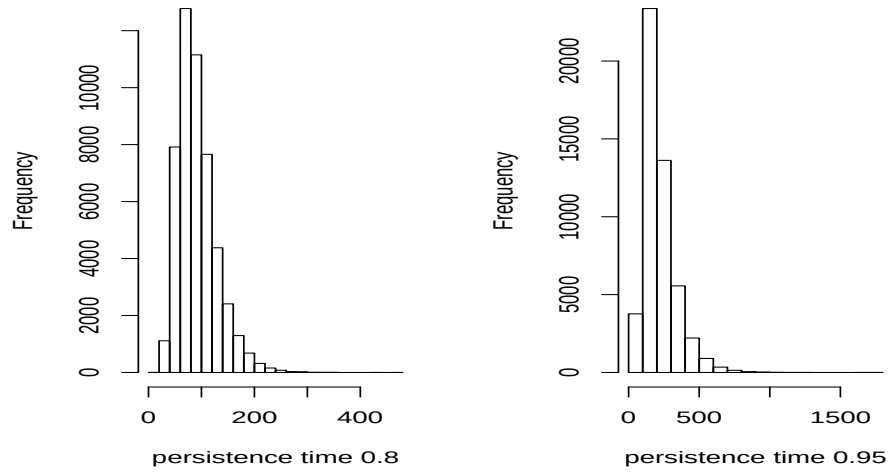

Figure 5: Left panel, the histogram of 50000 simulations of $T_{0.80}^{0.006}$ shows that at most in 300 generations the fraction of plasmid-free cells hits for the first time 0.8. Right panel, the histogram of 50000 simulations of $T_{0.95}^{0.006}$ shows that at most in 1000 generations the fraction of plasmid-free cells hits for the first time the boundary of 0.95.

using a hierarchical model. Clark and Bjørnstand (2004) used this type of models to estimate different sources of variability in population time series (environmental and process variability) and Ponciano *et al.* (2007) for the fraction of plasmid-free cells without conjugation. Finally, Lele *et al.* (2007) review many of the applications of these models in Ecology. Hierarchical models are usually specified using two equations called the observation and the process model equations. These are typically written as

$$\begin{array}{lrcl} \text{observation equation} & \underline{Y} & \sim & g(\underline{y}|\underline{X},\phi), \\ \text{process equation} & \underline{X} & \sim & f(\underline{x}|\theta). \end{array} \tag{12}$$

The observation equation contains the specification of the statistical sampling model by which the observations $\underline{Y}$ are obtained. This model is governed by a set of parameters, say ϕ. The sampling model is used to connect the data with the stochastic process model of interest. The process is usually denoted by $\underline{X}$ and is governed by

the density $g(\underline{x}|\theta)$. The paramers θ usually are the parameters of biological relevance.

The experiment described in De Gelder *et al.* (2004) gives the basic elements to propose a suitable hierarchical model. In three separated colonies, stability experiments were performed three times. Each colony (along with a certain antibiotic) was inoculated in 5ml of LB[1]. After 24hr, these cultures were washed to remove all antibiotics. A fraction of the 4.88 μl were transferred to 5ml of LB and they were grown for 10 generations, this represents time 0 (it is assumed that in 24 hrs, ten generations of bacteria span).

During 24hr they were incubated in a rotary shaker, then 4.88 μl of the cultures were transferred each 24 hr. to fresh 5 ml LB. The fraction of plasmid-free cells was obtained by replica, picking 50 colonies per culture at random and counting how many of them were free of plasmid through comparison of their genomic fingerprints.

Then observed data are three replica of the same process $\underline{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iq})$, $i = 1, 2, 3$ coming from a hidden Markov process of length q: $\underline{X}_i = (X_{1i}, X_{2i}, \ldots, X_{iq})$. The hierarchical model proposed includes observational and process error of the experiment and can be written as follows:

$$\begin{aligned} Y_{i,t} &\sim g(y_{i,t}|50, X_{i,t}), \qquad \textit{and} \\ X_{i,t}|X_{i,t-1} = x_{i,t-1} &\sim f_{X_{i,t}|X_{i,t-1}}(x_{i,t}|x_{i,t-1}), \end{aligned} \tag{13}$$

where $g(y_{it}|50, X_{i,t}) = \binom{50}{y_{i,t}}(X_{i,t})^{y_{i,t}}(1 - X_{i,t})^{50-y_{i,t}}$ is a binomial distribution and $f_{X_{i,t}|X_{i,t-1}}(x_{i,t}|x_{i,t-1})$ is as in Equation (8) for $i = 1, 2, 3$. The model in (13) considers environmental and demographic stochasticity, hoping that inferences based on such model will be adequate and complete. Inferences based on model (13) will be pre-

[1]LB: Lysogeny broth, a medium composed by peptides, vitamins, trace elements and minerals that is used for the growth of bacteria.

sented in the following sections, where point estimates using MCMC methods will be discussed.

3 Statistical Inference

The statistical inference for the fraction of plasmid-free cells will be based on the hierarchical model (13). Obtaining parameter and persistence time estimates is the main goal of the present statistical analysis. Using a posterior distribution (using a Bayesian approach) and a likelihood function the hierarchical model proposed in the last section is discussed.

3.1 Bayesian inference

Bayesian statistical inference is based on the analysis of posterior distributions of a vector of parameters. The posterior distributions are formed by the product of two important functions: The likelihood function and the joint prior distribution for the vector of parameters. In our case, the likelihood function and the posterior distribution for the vector $(\underline{X}_i, \mu, \tau, \lambda, \theta, \gamma)$ are used to estimate the fraction of plasmid-free cells and persistence times. The procedure to achieve this is described as follows:
For the Markov chain, let $(\mu, \tau^2, \lambda, \theta, \gamma)$ be the set of parameters of $\underline{X}_i = (X_{i1}, \ldots, X_{iq})$ that govern the hidden Markov process of length q. Variable X_t is latent (unobserved) since it only can be known through indirect observational schemes (Clark and Bjørnstad 2004). In our case, we have a sample of size 50 only. Let $\underline{y}_i = (y_{i1}, \ldots, y_{iq})$ be the set of observations that come from the i-th replica of the time series experiment which accounts for the error in the process due to the plasmid-free bacteria sampling method. The joint posterior distribution for the parameters and the process is

$$\pi(\mu, \tau^2, \lambda, \theta, \gamma, \underline{X}_1, \underline{X}_2, \underline{X}_3 | \underline{y}_1, \underline{y}_2, \underline{y}_3) \propto \prod_{i=1}^{3} \underbrace{g(\underline{y}_i | \underline{X}_i)}_{\text{Likelihood}} \times \underbrace{\pi(\underline{X}_1, \underline{X}_2, \underline{X}_3, \mu, \tau^2, \lambda, \theta, \gamma)}_{\text{Prior}}, \tag{14}$$

where the joint prior distribution can be written as:

$$\begin{aligned}\pi(\underline{X}_1, \underline{X}_2, \underline{X}_3, \mu, \tau^2, \lambda, \theta, \gamma) &= \pi(\mu)\pi(\tau^2)\pi(\lambda|\underline{X}_1, \underline{X}_2, \underline{X}_3)\pi(\gamma|\underline{X}_1, \underline{X}_2, \underline{X}_3, \theta)\pi(\theta) \\ &\quad \times \pi(x_{01})\pi(x_{02})\pi(x_{03}) \\ &\quad \times \prod_{i=1}^{3}\prod_{t=1}^{q} f_{X_{it}|X_{i,t-1},\mu,\tau^2,\lambda,\theta,\gamma}(x_{it}|x_{i,t-1}, \mu, \tau^2, \lambda, \theta, \gamma).\end{aligned}$$

From Equation (6) the Hidden Markov process has transition distribution defined by (14) if:

$$\begin{aligned} 0 < \lambda < \min_{1\leq t\leq q}\{X_{it}\} = f_1(\underline{X}_i) \\ \text{and if} \qquad\qquad (15) \\ 0 < \gamma < \min_{1\leq t\leq q}\left\{\tfrac{(\theta+1-X_{i,t-1})(1-X_{it})}{1-X_{i,t-1}}\right\} = f_2(\underline{X}_i, \theta) \quad \text{for } i = 1, 2, 3.\end{aligned}$$

Bayesian inference requires elicitation of prior distributions. For hierarchical models, elicitation of priors is not an easy task since priors for all the parameters and the process need to be settled. The hidden Markov model is a process that by definition cannot be observed and therefore, priors for unobservable quantities are difficult to elicit. Although the influence of the prior on the final inference can be strong and lead to different conclusions when working with different prior distributions (Lele and Dennis 2007), if subjective priors are used, methods to evaluate the robustness of the inferences to changes in the prior distribution exist (Ruggieri 2000).

Most of our unobservable variables are quantities between 0 and 1 or they are related through inequalities (15), so the elicitation of prior distribution for our model (14) was done as follows:

1. $\pi(\mu) = \frac{1}{10} I_{(0,10)}(\mu)$,
2. $\pi(\tau^2) = \frac{1}{60} I_{(0,60)}(\tau^2)$,
3. $\pi(\lambda|\underline{X}_1, \underline{X}_2, \underline{X}_3) = \frac{1}{\min_{i=1,2,3}\{f_1(\underline{X}_i)\}} I_{(0,\min_{i=1,2,3}\{f_1(\underline{X}_i)\})}(\lambda)$,
4. $\pi(\theta) = I_{(0,1)}(\theta)$,

5. $\pi(\gamma|\underline{X}_1, \underline{X}_2, \underline{X}_3, \theta) = \frac{1}{\min_{i=1,2,3}\{f_2(\underline{X}_i,\theta)\}} I_{(0,\min_{i=1,2,3}\{f_2(\underline{X}_i,\theta)\})}(\gamma)$,

6. $\pi(x_0) = I_{(0,1)}(x_0)$.

Using the above priors the resulting posterior distribution does not have a closed simple form, therefore computational tools are needed to sample from it. This was achieved using two different MCMC implementations to be explained in Section 4.

3.2 Likelihood Inference

A likelihood function is a function from which inferences about populations are done using random samples. Inferences will be meaningful as long as this population is correctly specified. The likelihood function is a function that is proportional to the probability of observing the sample $\underline{y}$ from a population with parameter θ. Sprott (2000) discussed that "the likelihood function ranks the plausibility of all possible values of θ by how probable they make the observed $\underline{y}$'. The likelihood function also allows comparison of plausibilities of different values of θ for the value of the observed sample $\underline{y}$.

For our model the likelihood function for the hierarchical model (13) is:

$$L(\mu, \tau^2, \lambda, \theta, \gamma; \underline{y}_1, \underline{y}_2, \underline{y}_3) = \int_{[0,1]^q}\int_{[0,1]^q}\int_{[0,1]^q}\prod_{i=1}^{3} g(y_{i,t}|50, X_{i,t}) \times \times f_{X_{i,t}|X_{i,t-1}}(x_{i,t}|x_{i,t-1}) d\underline{X}_1 d\underline{X}_2 d\underline{X}_3. \quad (16)$$

An efficient and numerically accurate maximization of this high-dimensional integral is difficult to achieve. Here, as it will be seen, MCMC computational tools will be used to overcome this problem. Maximum likelihood estimates will be obtained numerically and will be used to estimate the plasmids' persistence time.

4 Computer intensive statistical methods

4.1 ARMS within Gibbs Sampler

Markov Chain Monte Carlo (MCMC) algorithms were first described by Metropolis *et al.* (1953) and later improved by Hastings (1970). It is a general procedure to simulate from any probability distribution which is especially useful when its explicit form is not available (Robert and Casella 2005). The MCMC algorithm assures that the Markov chain's stationary law is the probability distribution of interest (see appendix A.1).

A very well-known example of MCMC is the Gibbs Sampler (Gelfand and Smith 1990, Casella and George 1992). The Gibbs sampler allows to sample from a distribution without having its explicit form. This algorithm is based on the fact that it is sufficient to know the full conditional distribution of each parameter to determine a joint distribution of the parameters. This computational technique uses sampling from each parameter's full conditional distribution at each step and then replacing this new sampled value in another parameter's full conditional distribution. In our model full conditional distributions are easy to write and have an important property: log-concavity.

Adaptive Rejection Sampling (ARS) is an often used methodology to sample from the conditional posterior distributions needed to carry the Gibbs sampler. To estimate the model parameters, Ponciano *et al.* (2007) used an ARS variant, Adaptive Rejection Metropolis Sampling (ARMS) within Gibbs to sample from the full conditionals in (17) that had the log-concavity property. Here, we first attempted to obtain samples from the full posterior (14) by using this method.

Adaptive Rejection Sampling (ARS) was proposed by Gilks and Wild (1992).

It is useful when we want to simulate from distributions whose logarithm is a concave function. This algorithm is based on building a concave hull that covers the log-density($h(\cdot)$). The concave hull is formed by intersections of lines that are tangent to a set of ordinates. This set of ordinates is obtained from a set of abscissae T_k of size k that are selected at the beginning of the algorithm and have the property of being within the support of the target density (Figure 6).

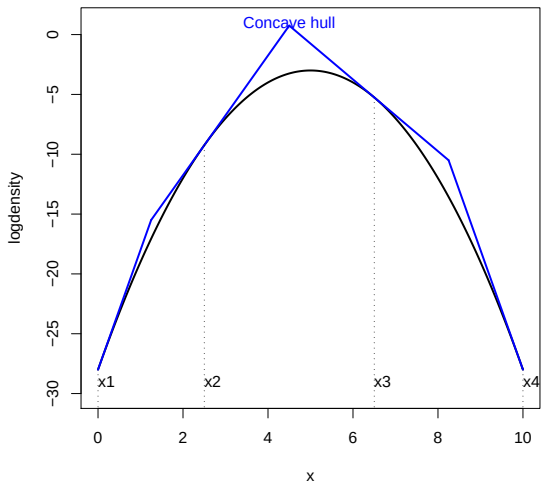

Figure 6: Log-concave density and concave hull. Using $k = 4$ a concave hull can be built. The hull covers completely the log-density.

The ARS algorithm requires the target log-density $h(\cdot)$ to be continuous, concave and differentiable in all its support D. Furthermore, D has to be a connected set. ARS is performed through the following steps:

1. Select an abscissae set in the density's support D: $T_k = (x_1, x_2, \ldots, x_k)$.

2. Calculate the intersections of the tangents in the ordinates $h(x_j)$:

$$z_j = \frac{h(x_{j+1}) - h(x_j) - x_{j+1}h'(x_{j+1}) + x_j h'(x_j)}{h'(x_j) - h'(x_{j+1})}.$$

3. Calculate a concave hull: $x \in [z_{j-1}, z_j]$

$$u_k(x) = h(x_j) + (x - x_j)h'(x_j)$$

and its exponential

$$s_k = \frac{\exp(u_k)(x)}{\int_D \exp(u_k)(x)}.$$

4. Simulate x^* from s_k, which is a mixture of exponential distributions with parameters $u_k(x) = h(x_j) + (x - x_j)h'(x_j)$.

5. Simulate u from a Uniform distribution in $(0, 1)$, accept x^* if

$$u \leq \exp(h(x^*) - u_k(x^*)),$$

else do $T_{k+1} = T_k \cup \{x^*\}$.

Gilks and Wild (1994) noted that using ARS, it is still possible to simulate from quasi log-concave distributions (see Figure 7) if an extra Metropolis-Hastings step is added to the algorithm described above. In the last step, if $u \leq \exp(h(x^*) - u_k(x^*))$ is true, then accept x^* with probability equal to the following Metropolis-Hastings ratio:

$$\rho(x^{(i)}, x^*) = min\left(1, \frac{f(x^*)min(f(x^{(i)}), s_k(x^{(i)}))}{f(x^{(i)})min(f(x^*), s_k(x^*))}\right),$$

where $log(f(\cdot)) = h(\cdot)$. If $u \leq \exp(h(x^*) - u_k(x^*))$ is not true or if x^* is not accepted, then T_{k+1} is set as $T_k \cup \{x^*\}$.

For our model equation (14), the ARMS algorithm was implemented within a Gibbs Sampler where Kernels (full conditionals) of each parameter were (considering 1 replica to simplify notation):

1. For τ^2: $p(\tau^2|\underline{X}, \mu, \lambda, \theta, \gamma) \propto \pi(\tau^2) \prod_{t=1}^{q} f_{X_t|X_{t-1},\mu,\tau^2,\lambda,\theta,\gamma}(x_t|x_{t-1}, \mu, \tau^2, \lambda, \theta, \gamma)$.

2. For λ: $p(\lambda|\underline{X}, \mu, \tau^2, \theta, \gamma) \propto \pi(\lambda|\underline{X}) \prod_{t=1}^{q} f_{X_t|X_{t-1},\mu,\tau^2,\lambda,\theta,\gamma}(x_t|x_{t-1}, \mu, \tau^2, \lambda, \theta, \gamma)$.

3. For γ: $p(\gamma|\underline{X}, \mu, \tau^2, \lambda, \theta) \propto \pi(\gamma|\underline{X}, \theta) \prod_{t=1}^{q} f_{X_t|X_{t-1},\mu,\tau^2,\lambda,\theta,\gamma}(x_t|x_{t-1}, \mu, \tau^2, \lambda, \theta, \gamma)$.

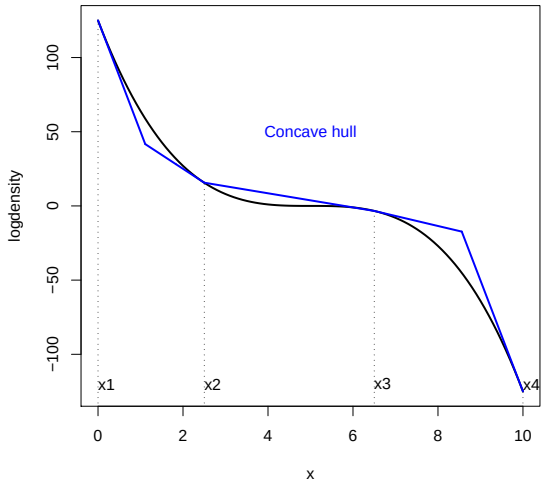

Figure 7: Quasi log-concave density and concave hull.

4. For θ:

$$p(\theta|\underline{X},\mu,\tau^2,\lambda,\gamma) \propto \pi(\theta)\pi(\gamma|\underline{X},\theta)\prod_{t=1}^{q} f_{X_t|X_{t-1},\mu,\tau^2,\lambda,\theta,\gamma}(x_t|x_{t-1},\mu,\tau^2,\lambda,\theta,\gamma). \tag{17}$$

5. For X_0: $p_{X_0|X_{t>0}\underline{Y},\mu,\tau^2,\lambda,\gamma,\theta}(x_0|x_{t>0}\underline{y},\mu,\tau^2,\lambda,\gamma,\theta) \propto$
$\propto \pi(\lambda)\pi(x_0)f_{X_1|X_0,\mu,\tau^2,\lambda,\gamma,\theta}(x_1|x_0,\mu,\tau^2,\lambda,\gamma,\theta)g(y_0|x_0).$

6. For X_t, $0 < t < q$: $p_{X_t|X_{j\neq t}\underline{Y},\mu,\tau^2,\lambda,\gamma,\theta}(x_t|x_{j\neq t}\underline{y},\mu,\tau^2,\lambda,\gamma,\theta) \propto$
$\propto \pi(\lambda|\underline{X})f_{X_t|X_{t-1},\mu,\tau^2,\lambda,\gamma,\theta}(x_t|x_{t-1},\mu,\tau^2,\lambda,\gamma,\theta) \times$
$\times f_{X_{t+1}|X_t,\mu,\tau^2,\lambda,\gamma,\theta}(x_{t+1}|x_t,\mu,\tau^2,\lambda,\gamma,\theta)g(y_t|x_t).$

7. For X_q: $p_{X_q|X_{q-1},\underline{y}\mu,\tau^2,\lambda,\gamma,\theta}(x_q|x_{q-1}\mu,\tau^2,\lambda,\gamma,\theta) \propto$
$\propto \pi(\lambda|\underline{X})f_{X_q|X_{q-1},\mu,\tau^2,\lambda,\gamma,\theta}(x_q|x_{q-1},\mu,\tau^2,\lambda,\gamma,\theta)g(y_q|x_q).$

Before carrying this Gibbs sampler and deciding whether to use ARS or ARMS, the log-concavity of each full conditional was checked. As it is shown in Figure 8, the log-density of parameter μ its log-density is completely concave and the log-density for $x_{1,2}$ is quasi log-concave.

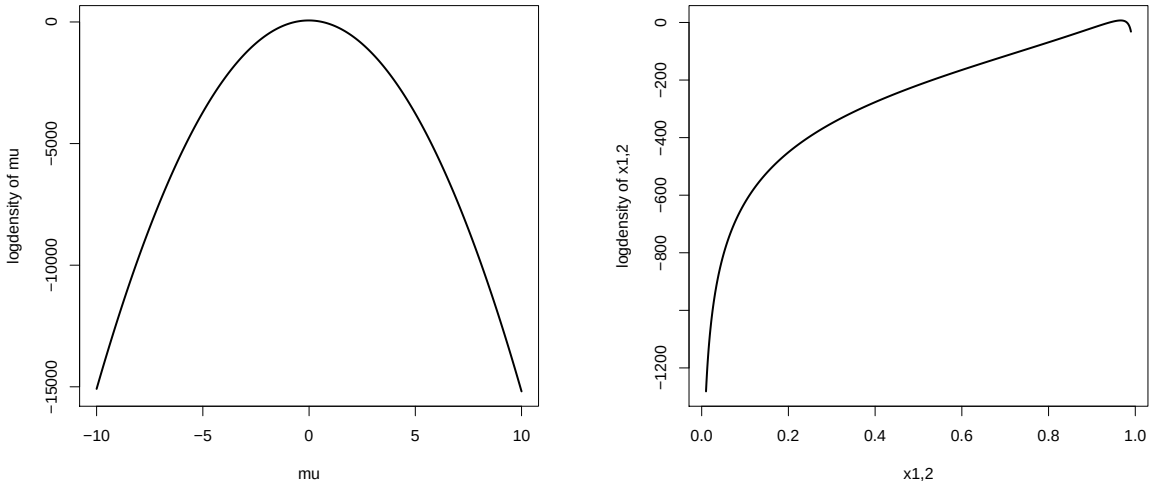

Figure 8: Log-densities for μ (left) and $x_{1,2}$ (right).

To draw B independent samples from the posterior distribution the above Kernels were sampled sequentially at each step. The values $(\underline{x}_i^{(0)}, \mu^{(0)}, \tau^{2(0)}, \lambda^{(0)}, \gamma^{(0)}, \theta^{(0)})$ were set at arbitrary initial values. A sample from the marginal posterior distribution of any parameter is obtained simply by taking the sequence of that parameter from the total sample and high-dimensional integration is not necessary (Meyer and Millar 1999).

4.2 t-walk

The second MCMC method implemented in this work is the t-walk algorithm. The t-walk is a general purpose scale-independent MCMC algorithm that maintains two independent points to sample from a joint posterior distribution. Proposed by Christen and Fox (2007), this algorithm solves the problem of tuning a proposal distribution for a MCMC. It changes the objective distribution $\pi(x)$ for $f(x, x') = \pi(x)\pi(x')$ and is different from running two chains simultaneously. This general MCMC algorithm has the advantage that is not an adaptive algorithm. It uses four different moving

strategies designed to preserve the chain's homogeneity and ergodic properties.

The t-walk's design is based on a general distribution $q\{(y, y')|(x, x')\}$ that proposes the chain moves according to:

$$(y, y') = \begin{cases} (x, h(x, x')) & \text{with probability } 0.5 \\ (h(x, x'), x') & \text{with probability } 0.5 \end{cases},$$

where $h(x, x')$ is a random variable associated with one the four moves described below. In the t-walk algorithm only some of the coordinates of the points (x, x') are moved. If the length of x is n, n independent Bernoulli random variables ϕ_j, $j = 1, \ldots, n$ are simulated. If $\phi_j = 1$ then the x_j coordinate will be updated according to $h(\cdot, \cdot)$.

The first of the four moves is called the walk-move. This move is closely related to a random walk and it allows to sample from distributions that are weakly correlated. This move is defined as:

$$h_w(x, x')_j = x_j + \phi_j(x_j - x'_j)z_j,$$

where z_j are real random variables i.i.d. with density $f_w(\cdot)$.

The second move is the traverse-move and it helps to sample from distributions whose variables are strongly correlated. The traverse-move is defined as:

$$h_t(x, x')_j = \begin{cases} x'_j + \beta(x'_j - x_j) & \phi_j = 1 \\ x_j & \phi_j = 0 \end{cases},$$

where β is a positive random variable with density $f_t(\cdot)$.

The final two are the hop-move and the blow-move. These two moves help to maintain the irreducibility of the chain and to guarantee its convergence. The

hop-move can be defined as follows:

$$h_h(x,x')_j = \begin{cases} (x_j + z_j \frac{\sigma(x,x')}{3}) & \phi_j = 1 \\ x_j & \phi_j = 0 \end{cases},$$

where $\sigma(x,x') = \max_j \phi_j |x_j - x'_j|$ and $z_j \sim N(0,1)$.

Finally, the blow move is:

$$h_b(x,x')_j = \begin{cases} x'_j + \sigma(x,x') z_j & \phi_j = 1 \\ x_j & \phi_j = 0 \end{cases},$$

z_j and $\sigma(x,x')$ are defined as before. It must be noticed that the difference between the hop and the blow moves is that they start from different initial points.

By mixing the set of the four standard Metropolis-Hastings Kernels K_α associated with the four moves h_α (and $\alpha \in \{w,t,h,b\}$) the convergence of t-walk algorithm is assured (Fox and Christen 2007).

The t-walk is not a sequential algorithm. It allows to sample from all the parameters in a posterior distribution at once. This property can be useful when some conditions need to hold simultaneously among the parameters, as it is the case with our model (see eq. 15).

The t-walk algorithm is particularly useful when the target distribution is highly multimodal and other MCMC variant is difficult to calibrate (*i.e.* finding a good proposal distribution). However, it is by definition a Metropolis-Hastings algorithm and as such it inherits one of its difficulties, namely choosing initial points that allow a quick convergence.

4.3 The Prior Feedback algorithm

It is possible to take advantage from MCMC methods to perform Likelihood inference. Robert (1993) proposed a method called Prior Feedback to obtain maximum likelihood (ML) estimates. Prior Feedback is a method that exploits the statistical properties of the problem and finds ML estimates as a limit of bayes estimates. The Prior Feedback algorithm updates the parameters using repeatedly the observed data as a new prior each time. By doing that, the posterior distribution converges to a fixed point, identical to the ML estimate.

Let $f(x|\theta) = c(x)\exp(\theta x - \phi(\theta))$ be a density from a k-dimensional exponential family. Let x be a single observation from $f(x|\theta)$. The conjugate prior $\pi(\theta|x_0, \lambda) \propto \exp(\theta x_0 - \lambda\phi(\theta))$ depends on x_0 and $\lambda \in R^+$ parameters.

The quantity of interest is a continuous function of the parameter $h(\theta)$. A good estimate for this quantity is $E_\pi(h(\theta)|x)$ since it depends on the observed data. The prior expectation of $h(\theta)$ is:

$$E(h(\theta)|x_0, \lambda) = \int h(\theta)\pi(\theta|x_0, \lambda)d\theta.$$

Then the updated posterior expectation will be

$$E_\pi(h(\theta)|x_1, \lambda) = E_\pi(h(\theta)|x_0 + x, \lambda).$$

Continuing in this fashion, the Prior Feedback Estimate (PFE), denoted by $\delta^*_\lambda(x)$, will be associated to a fixed point x^* such that

$$\delta^*_\lambda(x) = E_\pi(h(\theta)|x^*, \lambda) = E_\pi(h(\theta)|x^* + x, \lambda). \tag{18}$$

This estimate can be interpreted as the one whose prior density agrees with the information given by the data x (neutral distribution). Robert proved that $\delta^*_\lambda(x)$ converges to ML estimate of θ when λ goes to infinity (see Appendix A).

Robert also showed that this method can be very useful to obtain the ML estimates when the likelihood function is not from the exponential family. Also, he compared this method with Simulated Annealing (SA) in the sense that the prior distribution plays, in this algorithm, the role of the temperature. Robert (1993) claimed that it is easier for this algorithm to control priors since they have an statistical interpretation, whereas for SA it is always difficult to calibrate temperatures.

Later Robert and Titterington (1998) proposed a prior feedback method coupled with a Gibbs sampler for a Normal and Poisson hierarchical models to obtain the maximum likelihood estimates of the model parameters. The basic idea of this computational method is based on elevating to a high enough power the likelihood function so that the effect of the prior distribution vanishes. In the context of HMM they implemented the Gibbs sampler with an increasing number of replica of the sample of size m, asumming these replica are independent. The theoretical justification is that the prior feedback estimate will converge to a Dirac mass located at the maximum likelihood estimates as m goes to infinity, since the m-th posterior distribution of parameter θ is

$$\pi^{(m)}(\theta) \propto \pi(\theta)L(\theta)^m.$$

In the former method, m can be interpreted as the temperature of a SA algorithm but regularity conditions need to hold. In particular, $L(\theta)$ needs to be bounded, which is not trivial. In fact, to implement this computational method adequately, Robert and Titterington had to introduce a modification of the likelihood function for the Normal and Poisson cases because they were not bounded.

In the following section, a generalization of MCMC methods for obtaining maximum likelihood will be reviewed.

4.4 The Data Cloning algorithm

Lele *et al.* (2007) proposed to take advantage of MCMC methods to avoid high-dimensional integration and maximize likelihood functions to obtain ML estimates for hierarchical models. Their method is called Data Cloning (DC) and as the Prior Feedback algorithm, it is a MCMC method that converges to ML estimates as a variable k goes to infinity.

When the prior distribution is positive over the space parameter the posterior distribution of the hierarchical model (12) is

$$\pi^{(1)}(\theta,\phi|\underline{y}) = \frac{\int\{g(\underline{y}|\underline{X},\phi)f(\underline{X}|\theta)d\underline{X}\}\pi(\theta,\phi)}{h(\underline{y})},$$

where $h(\underline{y}) = \int g(\underline{y}|\underline{X},\phi)f(\underline{X}|\theta)\pi(\theta,\phi)d\underline{X}d\theta d\phi$. Now,if $\pi^{(1)}(\theta,\phi|\underline{y})$ is prior, the updated posterior is

$$\begin{aligned}\pi^{(2)}(\theta|\underline{y}) &= \frac{\int\{g(\underline{y}|\underline{X},\phi)f(\underline{X}|\theta)d\underline{X}\}\pi^{(1)}(\theta,\phi)}{h^2(\underline{y})} \\ &\quad \frac{\left\{\int g(\underline{y}|\underline{X},\phi)f(\underline{X}|\theta)d\underline{X}\right\}^2\pi(\theta,\phi)}{h^2(\underline{y})} \\ &\quad \frac{\left\{L(\theta,\phi;\underline{y})\right\}^2\pi(\theta,\phi)}{h^2(\underline{y})}.\end{aligned}$$

If we continue this procedure k times, k-th posterior distribution is

$$\pi^{(k)} = \frac{\left\{L(\theta,\phi;\underline{y})\right\}^k\pi(\theta,\phi)}{h^k(\underline{y})}.$$

On other hand, if the experiment was to be carried k times and by happenstance the results were always the same, then the likelihood for hierarchical model (12) would be

$$\left\{\int g(\underline{y}|\underline{X}\phi)f(\underline{X}|\theta)d\underline{X}\right\}^k = \left\{L(\theta,\phi;\underline{y})\right\}^k. \tag{19}$$

When $(\hat{\theta}_{ML},\hat{\phi}_{ML})$ are the maximum likelihood estimates for all (θ,ϕ),

$$L(\theta,\phi;\underline{y}) \leq L(\hat{\theta}_{ML},\hat{\phi}_{ML};\underline{y}),$$

and if k goes to infinity:

$$\frac{\pi^{(k)}(\theta,\phi|\underline{y})}{\pi^{(k)}(\hat{\theta},\hat{\phi}|\underline{y})} = \frac{\{L(\theta,\phi;\underline{y})\}^k}{\{L(\hat{\theta},\hat{\phi};\underline{y})\}^k} = \begin{cases} 0 & \text{if} \quad (\theta,\phi) \neq (\hat{\theta}_{ML},\hat{\phi}_{ML}), \\ 1 & \text{if} \quad (\theta,\phi) = (\hat{\theta}_{ML},\hat{\phi}_{ML}). \end{cases}$$

Lele proved that as $k \to \infty$, $\pi^{(k)}(\theta,\phi|\underline{y})$ converges in distribution to a multivariate normal with mean the ML estimates and variance a k-th part of Fisher's information matrix $(MVN([\hat{\theta},\hat{\phi}]', \frac{1}{k}I^{-1}(\hat{\theta},\hat{\phi})))$. This asymptotic result makes this methodology better than SA because it allow us to build confidence intervals easily, which is one the main goals of doing inference, by the method of Wald.

Here, the k-th posterior, using (13)and (16), is

$$\begin{aligned} \pi^{(k)}(\mu,\tau^2,\lambda,\theta,\gamma;\underline{y}_1,\underline{y}_2,\underline{y}_3) \quad \propto \quad & \left\{L(\mu,\tau^2,\lambda,\theta,\gamma;\underline{y}_1,\underline{y}_2,\underline{y}_3)\right\}^k \times \\ & \times\pi(x_0)\pi(\mu)\pi(\tau^2)\pi(\lambda)\pi(\theta)\pi(\gamma|\theta). \end{aligned} \tag{20}$$

5 Results and Discussion

In this thesis we proposed a new stochastic model for horizontal plasmid transfer along with the inference tools necessary to analyze real data from the widespread stability experiments (De Gelder *et al.* 2004, De Gelder *et al.* 2007, De Gelder *et al.* 2008). One of the most important contributions of this work is to present plasmids' persistence time as a random variable and the methodology to estimate its most important properties. Furthermore, the methodology presented here was thoroughly tested via simulations. In what follows we present first the results with the Gibbs sampler. Using the ARMS algorithm we show that for this particular case the use of the Gibbs sampler is in fact inefficient. Next, we present the results obtained from the t-walk algorithm when one and three replicated time series of the growth of the fraction of the plasmid-free cells are available. Then, we evaluate via simulations the quality of the persistence times' estimates. Finally, we illustrate how to carry maximum likelihood estimation via data cloning for our hierarchical model (13).

Simulations from the stochastic process (6) can be used to test thoroughly the Bayesian methodology presented here. If many (*e.g.* 1,000) sets of three replicated time series (as the ones obtained in the stability experiments) are generated according to the model with a set of parameter values, and if each time the modes of the marginal posterior distributions are recorded, then a histogram of these modes can be drawn. If the methodology works properly, the histograms of the *a posteriori* modes are expected to be centered around the true values, for long enough time series (that is, time series containing enough information about how the process behaves). Ponciano *et al.* (2007) carried that simulation exercise using the Gibbs sampler for each simulation from a stochastic model without horizontal transfer. They showed that indeed the histograms of the marginal *a posteriori* modes were centered around the true generating values. However, as we show next, examining the marginal posterior distribution of the parameters for a single, very long simulation, can shed some light

on the qualities of the inference. Also, as we will demonstrate, it is not trivial to pursue such simulation exercise for the Stochastic Horizontal Transfer model. Hence, here we limited ourselves to examine the results with a single simulation.

Three time series of y_{it} with $i = 1, 2, 3; \quad t = 1, 2 \ldots 600$ were simulated. In order to use ARMS within the Gibbs sampler to estimate the parameters for this simulated time series, one needs to sample sequentially the individual components of the posterior distribution (14). A few (two or three) parameters can be jointly sampled in some cases as shown in Robert and Casella (2005). Here, according to the conditions (15) we would need to jointly sample the parameters in the vector $(\underline{X}_i, \lambda, \theta, \gamma)$, which contains most of the parameters of interest. To do that an MCMC is required, and therefore the Gibbs sampler becomes extremely inefficient: in order to sample a single point one needs to achieve stationarity of the MCMC algorithm for this multidimensional chain. Using the full conditionals described in section 3, the elements in $(\underline{X}_i, \lambda, \theta, \gamma)$ were sampled one at a time and not jointly. Independent samples for each of the parameters were obtained using ARMS on these conditionals and letting the Gibbs sampler to run 2×10^6 times. A lag of 20,000 iterations was taken between each sample (Figure 9), until accumulating 100 samples. The maximum *a posteriori* estimates (MAP) for the parameters θ and γ are not close to their respective real values and the posterior distribution is highly bimodal (Figure 9). As it will be seen, the results with the t-walk sampler under the same simulation conditions are different, and the posterior distribution for both θ and γ are this time unimodal, close to the true generating value. This indicates that one-at-a-time sampling from the vector $(\underline{X}_i, \lambda, \theta, \gamma)$ may not be adequate.

To avoid using an MCMC algorithm inside the Gibbs sampler, the t-walk algorithm was used. Using one simulated time series y_t, $t = 1, 2 \ldots 600$, a hundred samples from each of the parameters were obtained taking a lag of size 20,000 (at

stationarity). Because of the lag size (20,000) these samples are approximately independent (Figure 10). The resulting histograms of these samples are multimodal (Figure 10), due to the total multimodality of the posterior distribution. These posterior distributions also illustrate the ability of the t-walk algorithm to sample from all the modes. A hundred *a posteriori* unobservable time series of the hidden Markov process $(X_t, t = 0, 1, \ldots, 600)$ were sampled independently using the t-walk algorithm and all of them were very close to true values as it is shown in Figure 11.

The stability experiments are usually triplicated, and hence, a very plausible data set consists of three independent time series of the growth of the fraction of plasmid free cells. That is, in contrast to what is usual in macro-ecological studies, samples from various realizations of the stochastic population process are obtained. These replicated realizations increase the amount of information available about the properties of the stochastic process under study. As the number of simulated replicated time series is increased, unimodality of the posterior distribution and the likelihood function seems to be achieved. Three replicas $y_{it}, i = 1, 2, 3; t = 1, 2, \ldots 600$ were simulated and the t-walk is run. Histograms of an independent sample of size 100 of each of the marginal posterior distributions were drawn (Figure 12) and they appear unimodal. Due to the increased amount of information, the posterior distributions modes are easier to identify. Note that the mode of the parameter λ is not close to the true value as the other modes seem to be. That means that even 600 iterations of the process do not contain enough information about the parameter λ. It may be possible that it is difficult to observe a single plasmid loss by segregation. In other words, segregation might not appear; its value, 6.851×10^{-5} indicates that we would expect 6 segregations to occur in 100,000 generations (the segregation rate used to simulate data), so that this information might not be observed in 600 generations of the process. However, note that the drawn histograms represent only one possible maximum *a posteriori* estimate for the parameters. To account for the variability of

the maximum *a posteriori* estimate of the segregation parameter λ (and of all of the parameters as well) it would be necessary to carry a parametric bootstrap simulation exercise as the one suggested at the beginning of this discussion and carried by Ponciano *et al.* (2007) for a one parameter simpler model.

The posterior distribution of the vector $(X_t, t = 0, 1, \ldots, 600)$ is well centered around the true unobserved process (Figure 13). Each of the samples of each element of $(X_t, t = 0, 1, \ldots, 600)$ were taken at lag sizes of 12,000 MCMC iterations, as suggested by the Integrated Autocorrelated Time value (Figure 14).

Simulations of size 50,000 of the persistence time $T_{0.8}^{0.006}$ using real parameter values were compared to those using the posterior modes of the histograms in Figure 12. The resulting histograms (Figure 15) show that their distributions are practically the same for persistence times smaller than 150 generations (approximately 15 days). The estimated persistence times are bigger than real ones for generations larger than 150. That can be seen in the probability plot (Figure 15) for $T_{0.8}^{0.006}$ (that is, when we are measuring the time until 80% of the bacteria cells are free from plasmids).

The same procedure was done to estimate $T_{0.006}^{0.95}$. In that case, the estimated and real stopping times distributions were very similar. In that case, the distribution of the stopping time is much better estimated (Figure16). Around $X_t = 0.80$ the variance of the process is much higher than the variance of the process at $X_t = 0.90$. This would explain why, with the same parameter estimates, the quality of the inference on the distribution of the stopping times varies so much between $T_{0.8}^{0.006}$ and $T_{0.95}^{0.006}$. In order to estimate the variability around the estimated distribution of the stopping times the values simultaneous 95 % credible intervals of all the parameters could be used to simulated the lower and upper stopping times' "credibility distributions".

The Bootstrap algorithm to evaluate the properties of the bayesian *a posteriori* modes would work as follows: Let $\theta = (\mu, \tau^2, \lambda, \theta, \gamma)$ be the set of parameters to be estimated and $\hat{\theta}^{(0)}$ the set of MAP estimates generated by the original three simulated time series $y_{it}^{(0)}$. Then, using $\hat{\theta}^{(0)}$ we could simulate B hidden Markov chains $x_{it}^{(1)}, x_{it}^{(2)}, \ldots, x_{it}^{B}$. Using the binomial sampling of (13) with each HMM we then would obtain $3 \times B$ new time series $y_{it}^{(1)}, y_{it}^{(2)}, \ldots, y_{it}^{(B)}$. The $3 \times B$ time series $y_{it}^{(j)}, i = 1, 2, 3; j = 1, 2, \ldots B$, would then be the new input for the B t-walk algorithms that would allow us to obtain B MAP estimates $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots, \hat{\theta}^{(B)}$. These bootstraped MAP estimates could then be drawn in histograms and frequentist properties (such as MSE, coverage and bias) could be obtained for each one of them (Figure 17).

Finally, to carry maximum likelihood inference, the Data Cloning algorithm was tried. To find the ML estimates a big enough value of k is needed so the estimates converge. This value can be found by repeatedly doubling the number of clones until the ML estimates do not change. If the data contains enough information about the parameters (that is, the likelihood profile for the parameter is not flat), then the variance of the k^{th} marginal posterior distribution for each parameter should decrease as k increases. If that does not happen for a given parameters, that means that the likelihood function does not contain enough information for that particular parameter. For our model, we chose $k = 20$. The maximum likelihoods estimates are shown in the following table:

k	μ	τ^2	λ	θ	γ
10	0.07454193	0.08044456	6.381449×10^{-4}	0.1493229	0.01276289
20	0.07521905	0.0803876	6.116462×10^{-4}	0.1493229	0.01275640

When we choose ML estimates with 20 clones, the estimated distribution of persistence times was shifted to the right of the real distribution of persistence times, as shown in Figure 18. The difference in the estimated versus the true distribution

increases in the tails. Also, we found that Markov chains provided by Data Cloning should be extremely large (more than 200,000 steps) in order to estimate adequately variability of the estimates when using 20 clones in this model. We did not pursue an investigation on the qualities of the ML estimates as $k > 20$, due to numerical difficulties (insufficient memory) to estimate Fisher's information matrix in those cases.

The plasmids' loss dynamics and persistence times were effectively estimated using the MAP and ML estimates for the hierarchical model (13). Given that including process noise in the model that considers HGT (6) is a more realistic model, we think that the methodology presented here is suitable for modeling the plasmids' dynamics for bacterial strains where conjugation is thought to be important factor. Conjugation might play an important role in spatially complex bacterial communities, like biofilms (Fox *et al.* 2008). There, this gene exchange process allows the generation and maintenance of diversity as well as multiple resistance to harsh environments. The model we propose here could then be extended to a spatially explicit, multivariate model to explain the spatial plasmid dynamics. This is an area of intense research in biology, yet stochastic and statistical models for such systems are generally lacking (Lili *et al.* 2007). We believe that the model presented in this work provides more flexibility and enough realism to answer the relevant questions about plasmids' population dynamics.

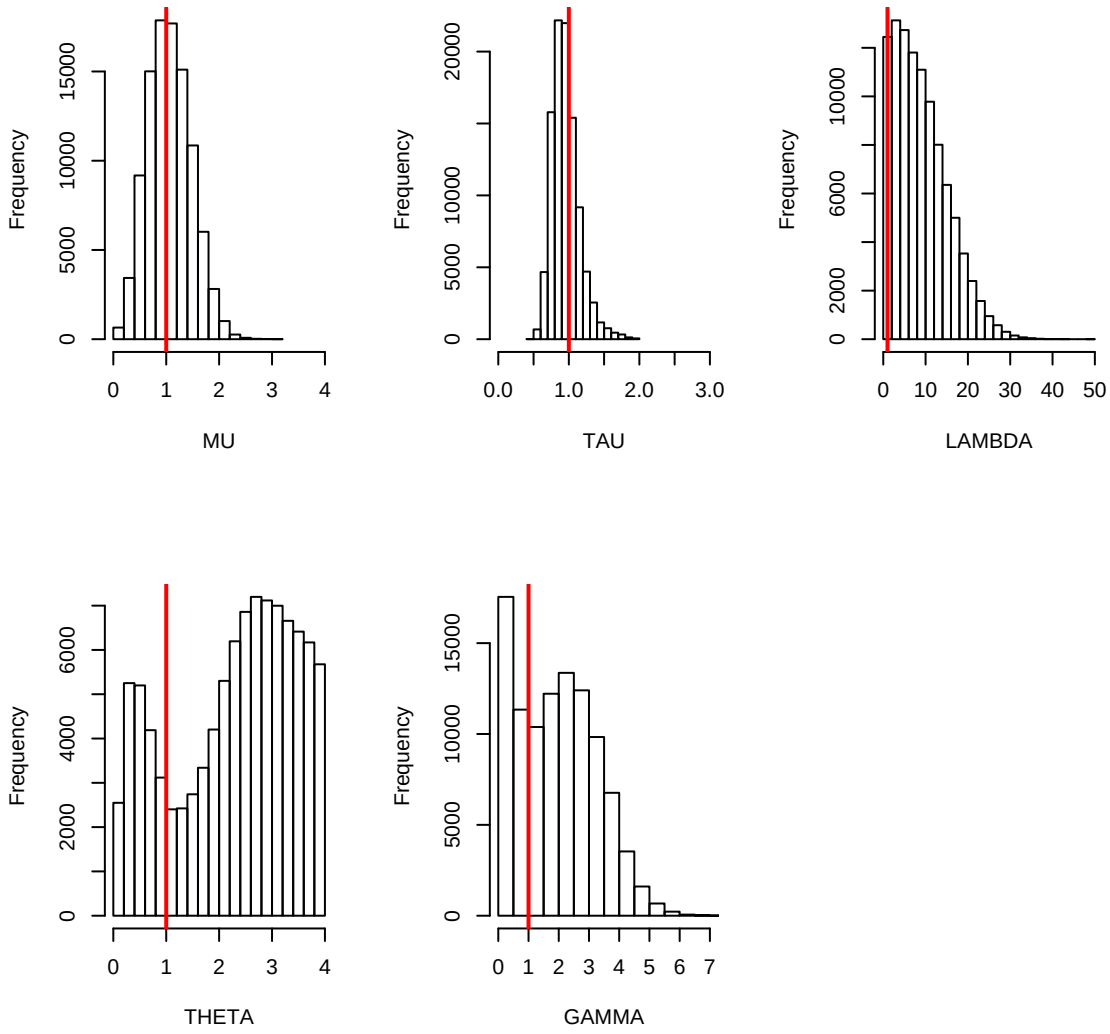

Figure 9: 100 independent samples of the marginal posterior distributions for each one of the parameters $(\mu, \tau^2, \lambda, \theta, \gamma)$, using ARMS within Gibbs. The true values are represented in the vertical line.

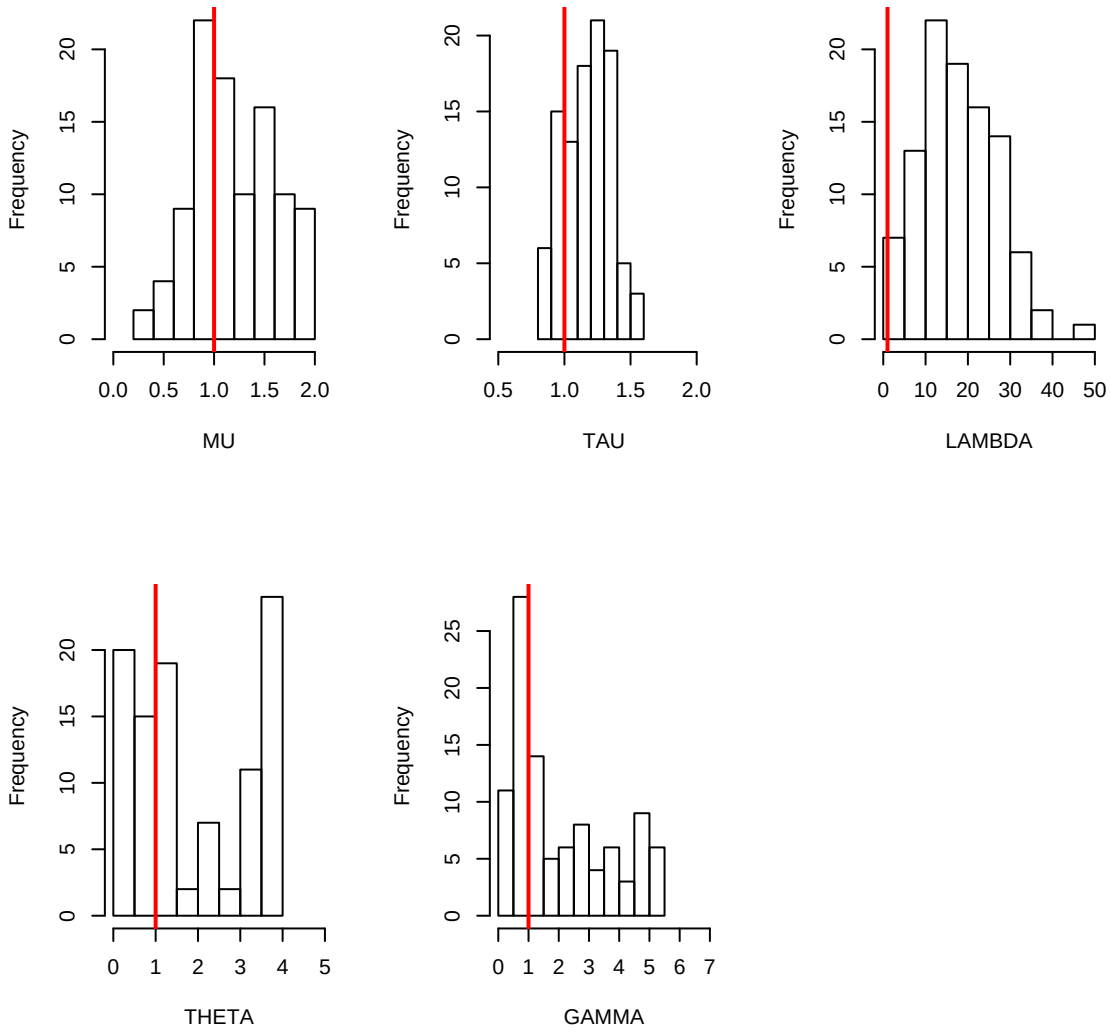

Figure 10: 100 independent samples of the marginal posterior distributions for each one of the parameters $(\mu, \tau^2, \lambda, \theta, \gamma)$, using the t-walk with one time series. The true values are represented with the vertical line.

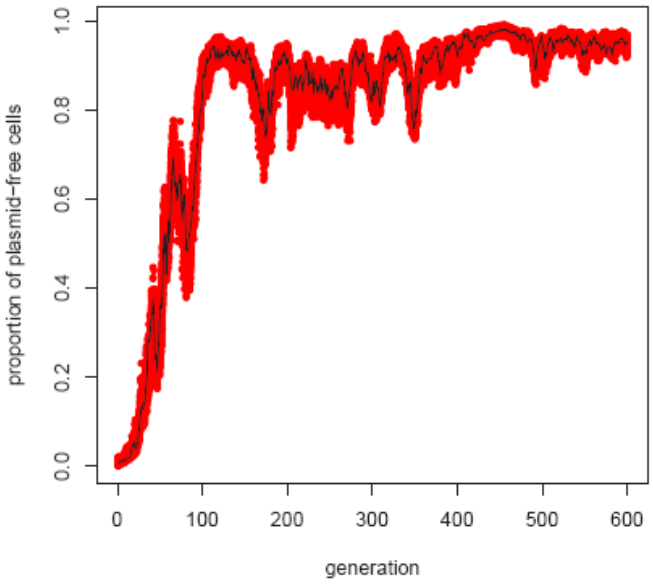

Figure 11: 100 independent samples for the hidden Markov process X_t (grey points) using the t-walk with one time series. The true value of the Markov chain is shown in black.

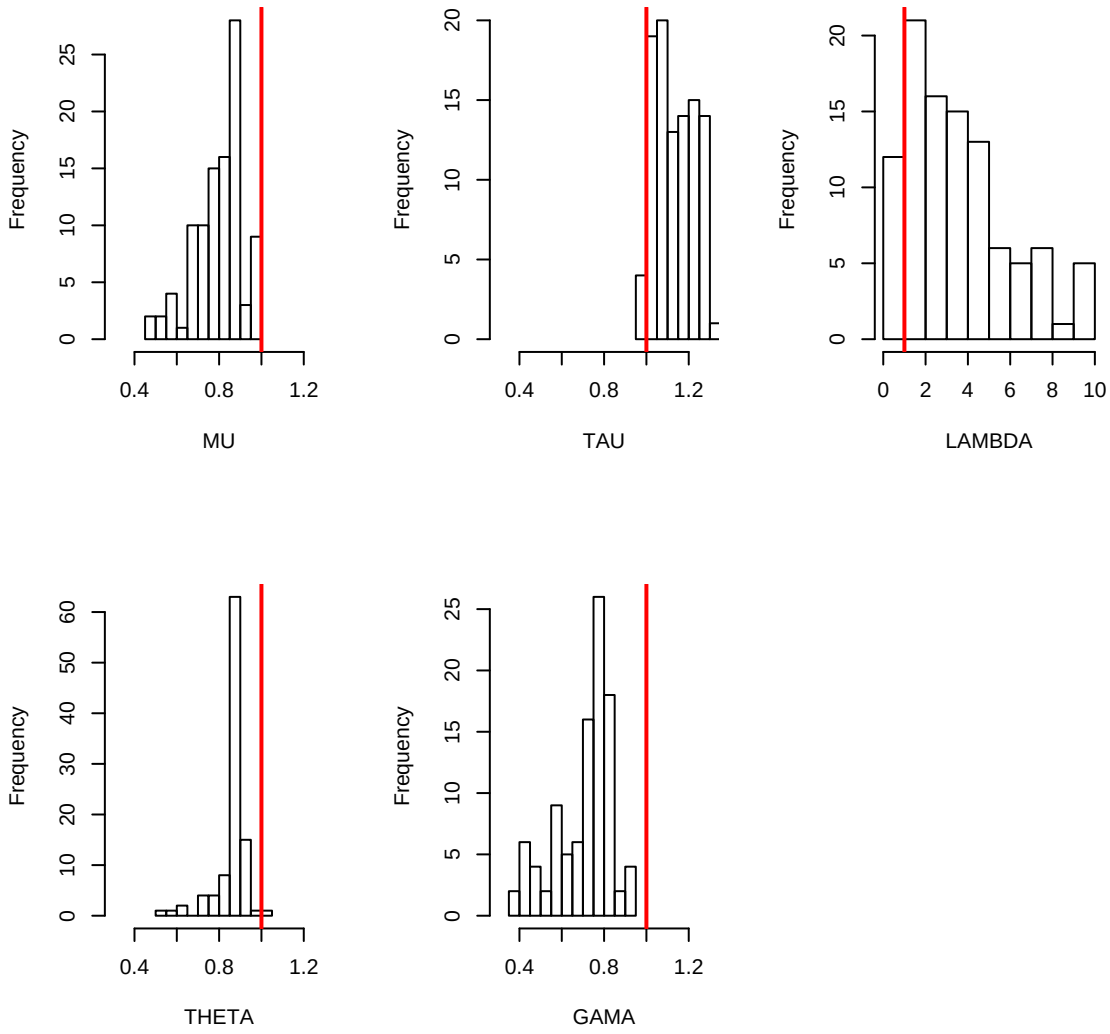

Figure 12: 100 independent samples of the marginal posterior distributions for each one of the parameters $(\mu, \tau^2, \lambda, \theta, \gamma)$, using the t-walk with three time series. The true values are represented with the vertical line.

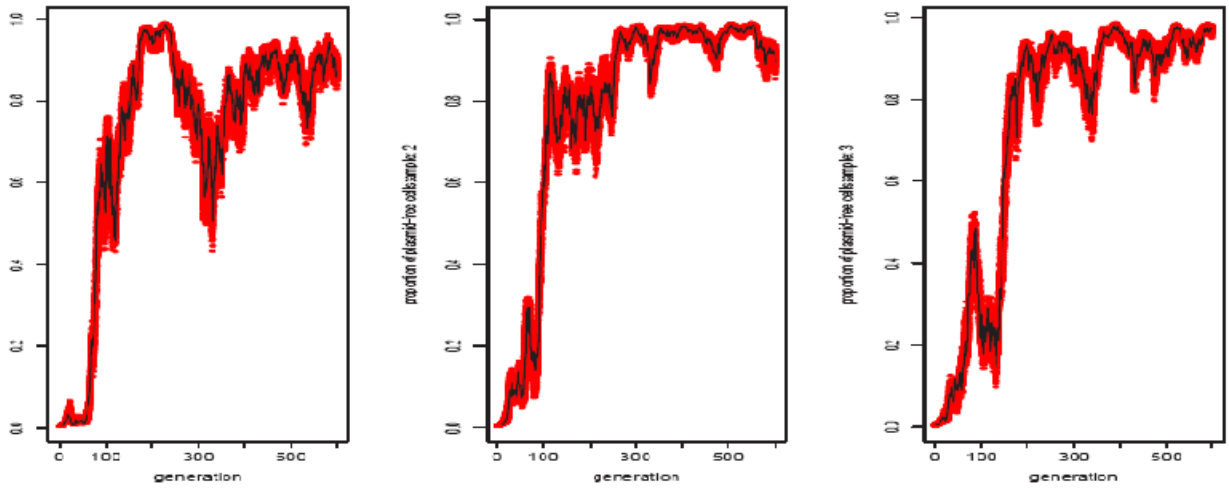

Figure 13: 100 independent samples for the three hidden Markov processes x_{it} (grey points) using the t-walk with three time series. The true value of the three Markov chains are shown in black.

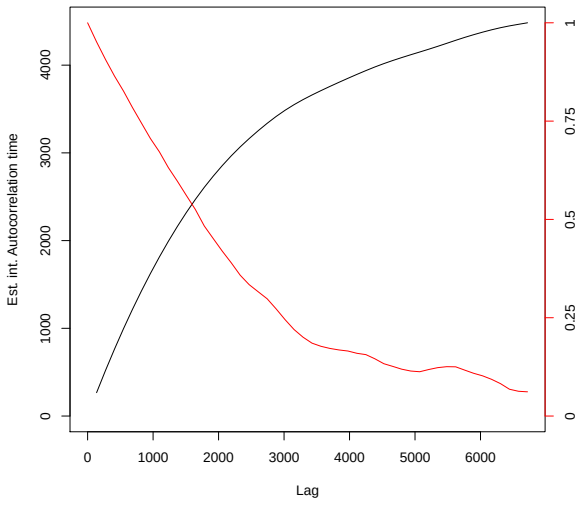

Figure 14: Estimation of the Integrated Autocorrelated Time of the t-walk MCMC samples of the posterior distribution (14).

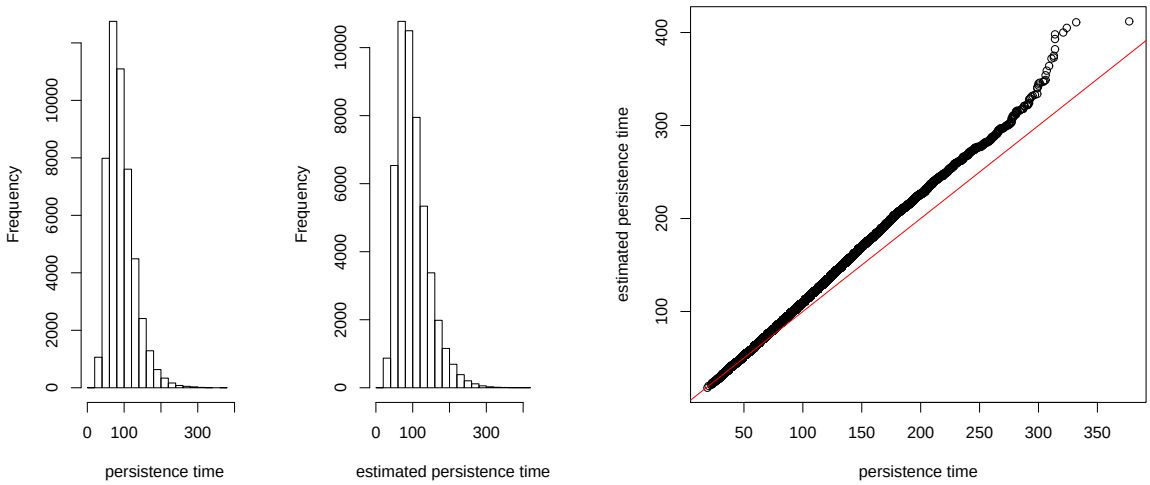

Figure 15: The histograms of 50,000 persistence times ($T^{0.8}$) simulated with real parameter values ($\mu = 0.11, \tau^2 = 0.0625, \lambda = 6.851044\times^{-5}, \theta = 0.25, \gamma = 0.02443239;$) and then with the estimated parameter values($\mu = 0.093431, \tau^2 = 0.06678281, \lambda = 1.266791 \times 10^{-4}, \theta = 0.2203061, \gamma = 0.01855009$). On the right, the probability plot, of the simulated persistence time distribution is shown.

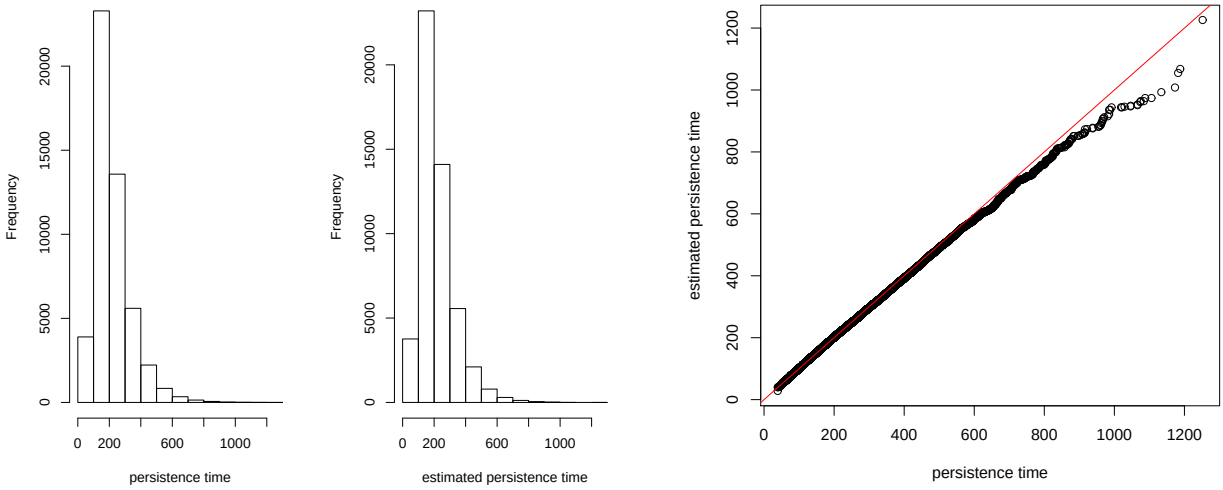

Figure 16: The histograms of 50,000 persistence times ($T^{0.95}$) simulated with real parameter values ($\mu = 0.11, \tau^2 = 0.0625, \lambda = 6.851044\times^{-5}, \theta = 0.25, \gamma = 0.02443239;$) and then with the estimated parameter values($\mu = 0.093431, \tau^2 = 0.06678281, \lambda = 1.266791 \times 10^{-4}, \theta = 0.2203061, \gamma = 0.01855009$). On the right, the probability plot of the simulated persistence time distribution is shown.

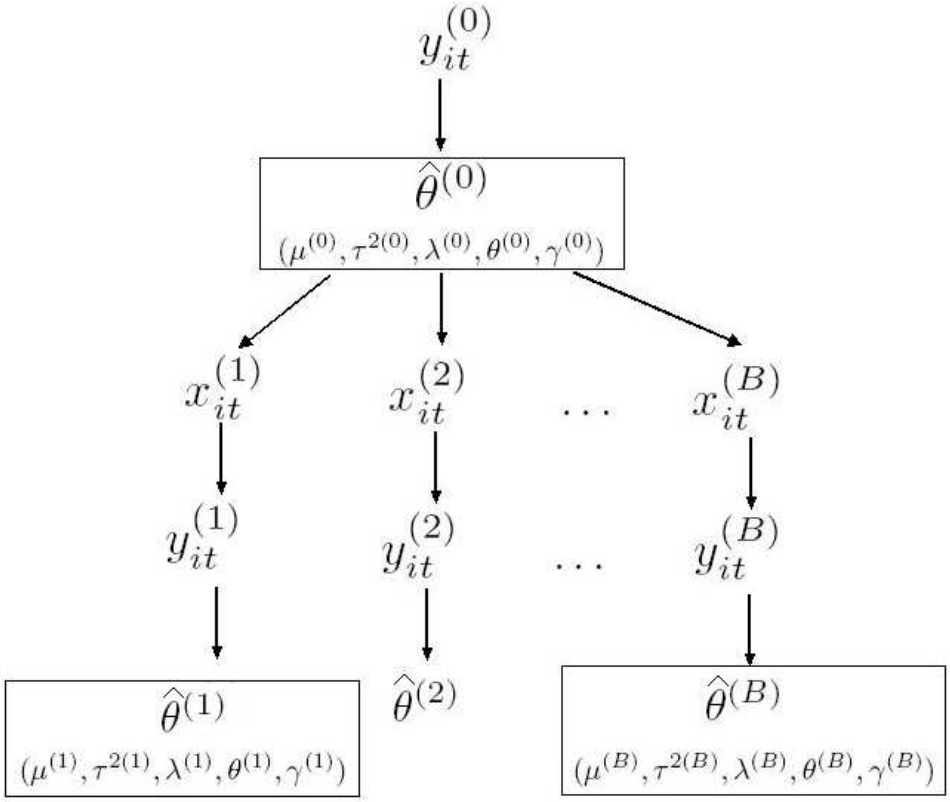

Figure 17: The bootstrap methodology for model (13). It is possible to give bootstraped interval estimates for all the parameters using this algorithm.

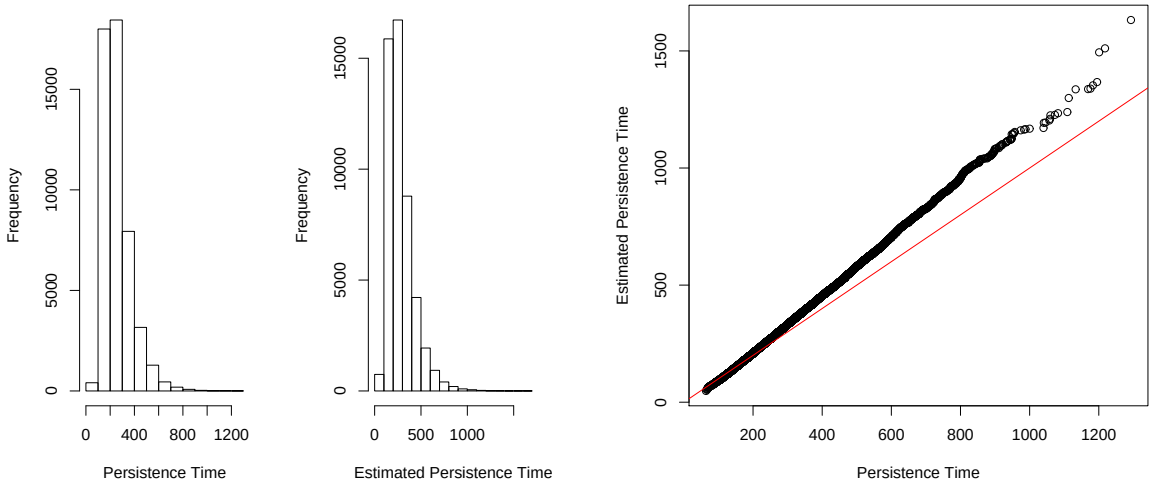

Figure 18: The histograms of 50,000 persistence times ($T^{0.95}$) simulated with real parameter values ($\mu = 0.11, \tau^2 = 0.0625, \lambda = 6.851044\times^{-5}, \theta = 0.25, \gamma = 0.02443239;$) and then with the estimated parameter values($\mu = 0.07521905, \tau^2 = 0.0803876, \lambda = 6.116462 \times 10^{-4}, \theta = 0.1493229, \gamma = 0.01275640$) with $k = 20$ clones. On the right, the probability plot of the simulated persistence time distribution is shown.

6 Conclusions

The stochastic dynamics model (6) effectively considered the three relevant process that determine plasmids' persistence: Segregation, conjugation and fitness cost. Furthermore, in the present work, the stochastic process efficiently modeled the daily changes in plasmid cost by proposing the fitness cost as a random variable. The growth rate of the fraction of plasmid-free cells became a random variable as it was expressed as a function of the fitness cost. This stochastic population dynamics model also allowed us to effectively model the plasmids' persistence time via simulations. This model included environmental stochasticity, a property that reflects how external factors influence the behavior of cells in plasmid loss dynamics.

Intensive computational methods were very useful to analyze simulated data. The first method used was ARMS within Gibbs sampler which failed to sample correctly from the posterior distribution of the vector of parameters since it samples parameters sequentially. The posterior distribution of the vector of parameters needed to be sampled at once because conditions (15) need to be stated *a priori*. The second method used was the t-walk considering one simulated time series from (12). Multimodality of the posterior distribution was suspected, so extensive validation of the parameter estimates needed to be done. Then, three simulated time series were obtained and the t-walk was run. This last methodology yielded a posterior distribution with a unique mode (Figure 12). The Maximum a Posteriori parameter (MAP) estimates were used to compare the estimated vs. the true distribution of plasmids' persistence times. Simulated persistence times using MAPs from the t-walk with three time series were very close to the simulated persistence times using the real parameter values. The approximation is particularly good for true persistence times of 95%. This result was expected since the simulated time series from which the MAPs were calculated were very long, and therefore they contained enough information on the parameters of interest. Nonetheless, this simulation exercise is useful because it

indirectly shows that the MCMC does converge.

Further work needs to be done to assess the statistical properties of the MAPs obtained. In particular, a parametric Bootstrap methodology could be used to assess the statistical properties of the Bayesian MAP estimates under different scenarios and parameterizations. This was not done here because very intensive computational resources are needed (for instance, the burn-in time for each t-walk in the Bootstrap might be extremely large). For hierarchical models like the one presented in this thesis, an exhaustive exploration of the effects of different "uninformative" *a priori* distributions also needs to be considered. Also, given that we are dealing with properties of a process that by definition is hidden, it seems difficult to elicit truly subjective prior distributions for the parameters of interest. Thus, the elicitation of adequate prior distributions and the evaluation of the robustness of the methods presented here (Ruggieri 2008) are topics of further research.

The Data Cloning methodology to find the ML estimates of a hierarchical model is a viable alternative to the current and popular Bayesian methodology for non-linear non-gaussian state-space models when researchers are not willing to use subjective prior distributions and when it is difficult to use "non-informative" priors. Here, Data Cloning was used to obtain ML estimates for the model (6). We observed that a good choice might be using 20 as the number of clones. The resulting ML estimates from this methodology were used to simulate persistence times and it was observed that they performed satisfactorily. To obtain interval estimates more computational power is needed in order to estimate Fisher's information matrix and build Wald's intervals. Another approach to build interval estimates could be using the methodology proposed by Ponciano *et al.* (2008) where a fast and efficient algorithm to obtain the relative profile likelihoods for the parameters of interest is proposed based on Data Cloning. As shown by these authors, such profile likelihood intervals

are more reliable than the asymptotic Wald intervals obtained via Data Cloning, especially for short time series data sets. Again, computational calculations to fully explore the Data Cloning methodology could be intensive since the convergence time of data cloning might be extremely large.

Finally, the hierarchical model (12) not only is a flexible model that can be used to analyze the stability experiments described by De Gelder *et al.* (2004) but also can give the experimentalist a strong basis to understand plasmid growth dynamics. In fact, with this hierarchical model, we show that it is possible to understand simultaneously the sampling process of plasmid-free bacteria and the relevant processes that allows plasmids to spread in the population. Because including process noise in the model that considers HGT (6) is a more realistic model, we think that the methodology presented here is suitable for modeling the plasmids' dynamics for bacterial strains where conjugation is thought to be an important factor. Conjugation might play an important role in spatially complex bacterial communities, like biofilms (Fox *et al.* 2008). There, this gene exchange process allows the generation and maintenance of diversity as well as multiple resistance to harsh environments. The model we proposed here could then be extended to a spatially explicit, multivariate model to explain the spatial plasmid dynamics. This is an area of intensive research in biology, yet stochastic and statistical models for such systems are generally lacking (Lili *et al.* 2007). We believe that the model presented in this work provides more flexibility and enough realism to answer the relevant questions about plasmids' population dynamics.

References

[1] *Basu S., Ebrahimi N.* 2001. Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence. Biometrika **88**: 269-279.

[2] *Bergstrom C.T., Lipsicht M., Levin B.R.* 2000. Natural selection, infectious transfer and existence conditions for bacterial plasmids. Genetics **155**: 1505-1519.

[3] *Casella G., George E.I.* 1992. Explaining the Gibbs sampler. The American Statistician **46 No. 3**: 167-174.

[4] *Christen J.A., Fox C.* 2007. A general purpose scale-independent MCMC algorithm. Comunicación Técnica **I-07-16**. CIMAT.

[5] *Clark J.S., Bjørnstad O.N.* 2004. Population time series: Process variability, observation errors, missing values, lags and hidden states. Journal on Ecology **85**: 3140-3150.

[6] *De Gelder L., Ponciano J.M., Abdo Z., Joyce P. Forney L.J.* 2004. Combining mathematical models and statistical methods to understand and predict the dynamics of antibiotic sensitive mutants in a population of resistant bacteria during experimental evolution. Genetics **168**: 1131-1144.

[7] *De Gelder L., Vandecasteele F., Brown C., Forney L.J., Top E.M.*2005. Plasmid donor affects host range of the promiscuous IncP-1β plasmid pB10 in a sewage sludge micorbial community. Applied Environmental Microbiology **71**: 5309-5317.

[8] *De Gelder L., Ponciano J.M., Joyce P., Top E.M.* 2007. Stability of the broad host range Inc-P-1beta plasmid pB10: no guarantees for a long term association. Microbiology **153**: 452-463.

[9] *De Gelder L., Williams J.J., Ponciano J.M., Sota M., Top E.M.* 2008. Adaptive plasmid evolution results in host range expansion of a broad-host-range plasmid. Genetics **178**: 2179-2190.

[10] *Dennis B., Ponciano J.M., Lele S.R., Taper M.L., Staples D.F.* 2006. Estimating density dependence, process noise and observation error. Ecological Monographs **76**: 323-341.

[11] *Fox R., Zhong X., Krone S.M., Top E.M.* 2008. Spatial structure and nutrients promote invasion of IncP-1 plasmids in bacterial populations. ISME J **2**: 1024-1039.

[12] *Gelfand A.E., Smith A.F.M.* 1990. Sampling-based approaches to calculating marginal densities. Journal of American Statistical Association **85-410**: 398-409.

[13] *Gelfand A.E, Schmidt A.M., Wu S., Silander J.A., Latimer A., Rebelo A.G.* 2005. Explaining species diversity through species level hierarchical modeling. Journal of Applied Statistics **54**: 1-20.

[14] *Genereux D.P., Bergstrom C.T.* 2005. Evolution in action: Understanding antibiotic resistance. Unpublished notes.

[15] *George E.I., Robert C.P.* 1992. Capture-recapture estimation via Gibbs sampling. Biometrika **79**: 677-683.

[16] *Gilks W.R., Wild P.* 1992. Adaptive rejection sampling for Gibbs sampling. Journal on Applied Statistics **41 No. 2**: 337-348.

[17] *Gilks W.R., Best N.G., Tan K.K.C* 1995. Adaptive Rejection Metropolis Sampling within Gibbs Sampling. Journal of Applied Statistics **44 No. 4**: 455-472.

[18] *Gogarten J.P., Townsend J.P.* 2005. Horizontal gene transfer, genome innovation and evolution. Nature Reviews, Microbiology **3**: 679-687.

[19] *Hanson F.B., Tier C.* 1981. A asymptotic solution of the first passage problem for singular diffusion in population biology. Journal on Applied Mathematics **40**: 113-132.

[20] *Hastings W.K.* 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**: 97-109.

[21] *Joyce P., Abdo Z., Ponciano J.M., De Gelder L., Forney L.J., Top E.M.* 2005. Modeling the impact of periodic bottlenecks, unidirectional mutation, and observational error in experimental evolution. Journal on Mathematical Biology **50**: 645-662.

[22] *Karlin S.* 1975. A first course in stochastic processes. San Diego: Academic Press.

[23] *Kot M.* 2001. Elements of Mathematical Ecology. Cambridge.

[24] *Lele S.R., Dennis B., Lutscher F.* 2007. Data cloning: Easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecology Letters **10**: 551-563.

[25] *Lele S.R, Dennis B.* 2008. Bayesian methods for hierarchical models: Are ecologists making a faustian bargain? Ecology In Press.

[26] *Levin B.R., Stewart, F.M.* 1980. The population biology of bacterial plasmids: A priori conditions for the existence of mobilizable nonconjugative factors. Genetics **94**: 425-443.

[27] *Lewontin R.C., Cohen D.* 1969. On population growth in a randomly varying environment. Proceedings of the National Academy of Sciences of the United States of America **62 No. 4**: 1056-1060.

[28] *Lili L.N., Britton N.F., Feil E.J.* 2007. The persistence of parasitic plasmids. Genetics **177**: 399-405.

[29] *Meyer R., Millar R.B.* 1999. Bayesian stock assessment using a state-space implementation of the delay difference model. Canadian Journal on Fish. Aquat. Science **56**: 37-52.

[30] *Meyn S.T., Tweedie R.L.* 1993. Markov chains and stochastic stability. Springer-Verlag.

[31] *Novozhilov A.S., Karev G.P., Koonin E.V.* 2005. Mathematical modeling of evolution of horizontally transferred genes. Molecular Biology Evolution **22**: 1721-1732.

[32] *Ponciano J.M., De Gelder L., Top E.M., Joyce P.* 2007. The population biology of bacterial plasmids: A hidden Markov model approach. Genetics **176**: 957-968.

[33] *Ponciano J.M., Taper M.L., Dennis B., Lele S.R.* 2008. Hierarchical models in ecology: Confidence intervals and hypothesis testing using data cloning. Ecology In Press.

[34] *Ripley B.D., Venables W.N.*1994. Modern applied statistics with S-PLUS. Springer Verlag.

[35] *Robert C.P.* 1993. Prior Feedback: Bayesian tools for maximum likekihood estimation. Journal on Computational Statistics **8**: 279-294.

[36] *Robert C.P., Casella G.* 2005. Monte Carlo statistical methods, 2nd edn. Springer, New York.

[37] *Robert C.P., Titterington D.M.* 1998. Reparameterization strategies for hidden Markov models and Bayesian approaches to maximum likelihood estimation. Statistics and computing **8**: 145-158.

[38] *Ruggeri F., Rios Insua D.* 2000. Robust bayesian analysis. Lecture Notes in Statistics, Springer Verlag.

[39] *Seneta E., Tavaré S.* 1983. Some stochastic models for plasmid copy number. Theoretical Population Biology **23**: 241-256.

[40] *Slater F.R., Bailey M.J., Tett A.J., Turner S.L.* 2008. Progress towards understanding the fate of plasmids in bacterial communities. FEMS Microbiology and Ecology: 1-11.

[41] *Sprott D.A.* 2000. Statistical inference in science. Springer Series in Statistics, Springer-Verlag.

[42] *Stewart F.M., Levin B.R.* 1977. The population biology of bacterial plasmids: A priori conditions for the existence of conjugationally transmitted factors. Genetics **87**: 209-228.

A Convergence of Estimates

A.1 Convergence of MCMC algorithm

The final goal of a MCMC algorithm is to obtain an independent sample from a distribution $\pi(x)$. To run a MCMC algorithm using a Metropolis-Hastings ratio, it is necessary to have a proposal distribution $q(x, y)$ that indicates the probability of moving from the value x to the value y and that this proposal distribution is easy to sample from. While the proposal distribution is being used in the algorithm a Markov chain is being created and this created Markov chain has as stationary law $\pi(x)$. The algorithm performs the following steps:

1. Initialize with a value $X = x_0$ at time $t = 0$.

   For $t = 1, 2, \ldots, B$ repeat the following steps:

2. Sample x_t from $q(x_{t-1}, x_t)$.

3. Accept x_t with probability

$$a(x_{t-1}, x_t) = \min\left(1, \frac{\pi(x_t)q(x_t, x_{t-1})}{\pi(x_{t-1})q(x_{t-1}, x_t)}\right).$$

4. Else, $x_t = x_{t-1}$ and go to step 2.

The transition matrix of the Markov chain is defined by this algorithm as follows:

$$P(x_t|x_{t-1}) = q(x_{t-1}, x_t)a(x_{t-1}, x_t) = m(x_{t-1}, x_t).$$

If $a(x_{t-1}, x_t) < 1$ then, we have $a(x_{t-1}, x_t) = \left(\frac{\pi(x_t)q(x_t, x_{t-1})}{\pi(x_{t-1})q(x_{t-1}, x_t)}\right)$ and $a(x_t, x_t) = 1$. Even more:

$$\begin{aligned}
\pi(x_{t-1})m(x_{t-1}, x_t) &= \pi(x_{t-1})q(x_{t-1}, x_t)a(x_{t-1}, x_t), \\
&= \pi(x_{t-1})q(x_{t-1}, x_t)\left(\frac{\pi(x_t)q(x_t, x_{t-1})}{\pi(x_{t-1})q(x_{t-1}, x_t)}\right), \\
&= \pi(x_t)q(x_t, x_{t-1}), \\
&= \pi(x_t)q(x_t, x_{t-1})a(x_t, x_{t-1}), \\
&= \pi(x_t)m(x_t, x_{t-1}),
\end{aligned}$$

so that $\pi(x)$ is the stationary law of $m(x_{t-1}, x_t)$.

A.2 Convergence of Prior Feedback estimate

Let x be a single observation from $f(x|\theta)$, where $f(x|\theta)$ is a density of a k-th dimensional exponential family. The exponential density is $f(x|\theta) = c(x)\exp(\theta x - \phi(\theta))$. The conjugate prior is $\pi(\theta|x_0, \lambda) \propto \exp(\theta x_0 - \lambda\phi(\theta))$ and it depends on x_0 and $\lambda \in R^+$ parameters.

The maximum likelihood estimate is $\hat{\theta}$ is x obtained from the following equation

$$\epsilon(\hat{\theta}) = \frac{\partial\phi(\theta)}{\partial\theta}\Big|_{\theta=\hat{\theta}} = x.$$

Even more, the mode for the prior distribution is $\tilde{\theta}$ and it can be obtained from

$$\epsilon(\tilde{\theta}) = \frac{\partial\phi(\theta)}{\partial\theta}\Big|_{\theta=\tilde{\theta}} = \frac{x_0}{\lambda}.$$

Then, if we reparameterize using $\epsilon(\theta) = \frac{\partial\phi(\theta)}{\partial\theta}$ we obtain the mean parameterization. Hence, we can transform the *a priori* density as

$$\pi(\theta|x_0, \lambda) \sim \pi^*(\epsilon|x_0, \lambda).$$

The reparameterized density $\pi^*(\epsilon|x_0,\lambda)$ has a mode $\frac{x_0}{\lambda}$ that is equal to its mean. Then

$$\lim_{\lambda\to\infty}\frac{x_0}{\lambda}=\begin{cases}0 & \text{if } x_0 \text{ is fixed,}\\ a^* & \text{if } x_0 \text{ is a function of } \lambda.\end{cases}$$

The prior feedback estimate of $h(\theta)$, when we have the fixed point x^*, will converge to $\lim_{\lambda\to\infty} h\left(\epsilon^{-1}(\frac{x^*}{\lambda})\right)$. But by definition of prior feedback estimate we have that $\delta^*_\lambda(x)=E_\pi(h(\theta)|x^*,\lambda)=E_\pi(h(\theta)|x+x^*,\lambda+1)$ and this implies that for a λ large enough

$$h\left(\epsilon^{-1}\left(\frac{x^*}{\lambda}\right)\right)\approx h\left(\epsilon^{-1}\left(\frac{x^*+x}{\lambda+1}\right)\right).$$

By continuity of $h(\cdot)$ and $\epsilon(\cdot)$, and also by the property of reparameterization it follows that

$$\left(\frac{x^*}{\lambda}\right)\approx\left(\frac{x^*+x}{\lambda+1}\right),$$

so $h(\epsilon^{-1}\left(\frac{x^*}{\lambda}\right))$ converges to $h(\epsilon^{-1}(x))=h(\hat{\theta})$ the maximum likelihood estimate.