Combining quantitative data with logic-based specifications for parameter inference

Combining quantitative data with logic-based specifications for parameter inference.

Paul Piho, Jane Hillston

Imperial College London, University of Edinburgh

Abstract. Continuous time Markov chains are a common mathematical model for a range of natural and computer systems. An important part of constructing such models is fitting the model parameters based on some observed data or prior domain knowledge. In this paper we consider the problem of fitting model parameters with respect to a mix of quantitative data and qualitative data formulated as temporal logic formulae. Our approach works by defining a set of conditions that capture the dynamics inferred by the quantitative data. This allows for a straightforward way to combine the information from the quantitative and qualitative knowledge into one parameter inference problem via rejection sampling.

1 Introduction

Quantitative models like continuous time Markov chains facilitate the understanding of processes and phenomena from a variety of areas like performance modelling, epidemiology and biology. A large part of constructing these models deals with fitting the model parameters using existing data or prior domain knowledge. In the context of formal methods such parameter inference problems have been considered in [4, 5] where logic specifications are used to specify the constraints on the behaviour of the model. The emphasis is on identifying parameters for which a given logical formula holds or parameters where the probability of satisfying the formula is maximised. These ideas can be used effectively in the area of control where the logic specifications are used to define the requirements for a successful controller [17]. Outside of formal methods, the inference problems are generally based on fitting the model parameters using existing observed data. The observed data can come in the form of quantitative data, like measured time trajectories, or qualitative data [14].

In this work we concentrate on parameter inference for models where the knowledge about the behaviour of the system comes in different forms. Firstly, we have observed time trajectories corresponding to a part of the modelled system's behaviour. Secondly, we have aspects of the system behaviour that are not directly measured but where the prior knowledge about the system's behaviour can be given through a logic-based specification. We combine those two sources of information into a single parameter inference problem. The benefit such integration brings is that lack of time-series data about parts of the system's behaviour can be compensated via a logic-based high level specification.

To that end, we implement a rejection sampling-based algorithm where samples of parameters from a defined prior distribution are accepted as posterior samples if the corresponding model trajectories are close enough to the quantitative data and also satisfy the logic specification. The decision on whether or not model trajectories are close to the available quantitative data is made in two ways. One follows standard approximate Bayesian computation [15] by defining a discrepancy measure between statistics of the quantitative data and simulated model trajectories. Secondly, we propose an approach where the observed data is translated into a logical specification capturing the temporal dynamics of the data. The resulting specification is then combined with the logical specification corresponding to the qualitative data. We then study the resulting posterior distributions empirically by considering the quantities corresponding to expected satisfaction probability of a logical specification over the recovered approximate posterior distributions.

We start by giving an overview of the related work in Section 2 and the required background in Section 3. In Section 4 we introduce the main contribution of the paper — a method for performing inference based on both quantitative and qualitative data in the form of logical specifications. In Section 5 we present computational experiment results for two illustrative examples — a rumour spread and a client server model. Finally, we end in Section 6 with conclusions and further work.

2 Related work

Several recent papers present work on parameter fitting for non-stochastic models of dynamics from qualitative data [12, 14, 13]. All of these consider deterministic ordinary differential equation models. In [12] the qualitative observations are formalised as inequality constraints on the model output. The parameter identification is then treated as a constrained optimisation problem through a heuristically constructed objective function. In [13] this work is extended to give a Bayesian formulation of parameter inference from qualitative data. The authors of [14] take an alternative approach where the best quantitative representation of the qualitative observations in the form of categorical data is found via optimal scaling methods. The found quantitative representation is referred to as surrogate data. This approach is based on an assumption that the qualitative data comes in the form of ordered categorical data, like "low", "high" and "very high", directly corresponding to quantitative trajectories. In the context of formal methods, the authors of [4] propose a method of fitting model parameters based on a logic specification by framing the problem in a Bayesian optimisation framework. However, integration of quantitative and qualitative data remains an open issue.

3 Background

In this section we present an overview of the necessary background topics. Specifically, we introduce the main models of study, called *parametric continuous time Markov chains* and the related statistical model checking and inference problems.

3.1 Parametric continuous time Markov chains

In this paper we consider discrete-state stochastic models, namely continuous time Markov chains. We give the following definition, common in the verification literature [2].

Definition 1. *A continous time Markov chain is a model of the form* $\mathcal{M} = \{S, \mathbb{Q}, A, L\}$ *where*

- $S = \{s_0, s_1, \cdots, s_n\}$ *is a finite set of states.*
- $\mathbb{Q}$ *is the transition rate matrix. In particular,* $\mathbb{Q}$ *is a* $|S| \times |S|$ *matrix such that the off-diagonals are non-negative and the diagonal elements*

$$a_{ii} = -\sum_{i \neq j} a_{ji} \tag{1}$$

- $L : S \to 2^A$ *is a labelling function mapping a state* $s \in S$ *to a subset of atomic propositions* A *that hold for* s*.*

An infinite path of a CTMC $\mathcal{M}$ is a sequence $s_0 t_0 s_1 t_1 \ldots$ where the transition rate $a_{i,i+1} > 0$ and $t_i > 0$ for all $i \geq 0$. The state s_0 denotes the initial state of the CTMC and is assumed to be fixed. The time t_i represents the time CTMC spends in a given state. The times t_i, called holding times, are drawn from the exponential distribution with rate parameter $-a_{ii}$. A trace of $\mathcal{M}$ is the mapping of path $s_0 t_0 s_1 t_1 \ldots$ through the labelling function L.

The state of the CTMC remains constant between jumps. Thus, in addition to the trace we can give the trajectory of a CTMC as a left-continuous function $\omega : \mathbb{R}_{\geq 0} \to S$ that is constant for the duration of the holding time and then jumps to the next state. We can simulate the CTMC via, for example, Gillespie's stochastic simulation algorithm [8]. Finally, we give a definition of the parametric extension of continuous time Markov chains.

Definition 2. *A parametric continuous time Markov chain over parameter vector* θ *is a model of the form* $\mathcal{M}_\theta = \{S, \mathbb{Q}_\theta, A, L\}$ *where* S*,* A *and* L *are defined as before. Additionally,*

- $\theta = (\theta_1, \cdots, \theta_k)$ *is the vector of parameters taking values in some domain* $D \subset \mathbb{R}^k_{\geq 0}$*.*
- $\mathbb{Q}_\theta$ *is the parametric transition rate matrix such that each entry in* $\mathbb{Q}_\theta$ *depends polynomially on* $\theta_1, \cdots, \theta_k$*.*

In particular, $\mathcal{M}_\theta$ *defines a family of continuous time Markov chains with parameters* θ *varying in a domain* D*.*

CTMC-based models are often expressed in a high-level modelling language such as stochastic process algebras [9, 7], generalised stochastic Petri nets [3] or Chemical Reaction Network (CRN) models [1]. For simplicity we consider the latter option defined as follows:

Definition 3. *A chemical reaction network model is defined by a*

- *a vector of population variables* $\mathbf{X} = (X_1, X_2, \cdots, X_n) \in \mathbb{N}^n$ *counting the number of different agents in the system.*
- *a set of reaction rules in the form*

$$r_1 X_1 + \cdots + r_n X_n \xrightarrow{\tau(\mathbf{X},\theta)} s_1 X_1 + \cdots + s_n X_n$$

where r_i *and* s_i *are the counts of agents consumed (respectively produced) in a reaction.* τ *is the rate function depending on the state of the model* $\mathbf{X}$ *and a vector of model parameters* θ*.*

The dynamics of the reaction network are interpreted as a parametric CTMC, or a CTMC if the model parameters θ *are fixed, where each state in the CTMC corresponds to a vector of counts.*

Example 1. In order to illustrate the ideas let us consider the following susceptible-infected-recovered (SIR) model defined as a CRN

$$S + I \xrightarrow{k_I} I + I \qquad\qquad I \xrightarrow{k_R} R$$

where S gives the number of susceptible, I the number of infected and R the number of recovered individuals in the system. The first type of transition corresponds to infected and susceptible individuals interacting, resulting in the number of infected individuals increasing and the number of susceptible decreasing. In particular, the susceptible agent turns into an infected one. The second type of transition corresponds to recovery of an infected individual and results in the number of infected decreasing and the number of recovered increasing. The states of the underlying CTMC keep track of the counts of different individuals in the system. The model parameters are given by the vector (k_I, k_R) resulting in transition rates $k_I \times S \times I$ and $k_R \times I$ for infection spread and recovery respectively.

3.2 Statistical model checking

An important problem in statistical model checking is estimating the probability that a model satisfies a given logical specification [10]. In this paper we assume that the logical properties are defined in the time-bounded fragment of metric interval temporal logic (MiTL) [11]. MiTL is a linear temporal logic for continuous time trajectories with the syntax given by

$$\varphi ::= \mathit{true} \mid \mu \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 U_{[T_1,T_2]} \varphi_2 \tag{2}$$

where $U_{[T_1,T_2]}$ denotes the time-bounded until operator. The atomic propositions are inequalities on population variables of a CRN model. A MiTL formula is interpreted over a real-value function of time $\mathbf{x}(t)$ that corresponds to a trajectory of a CTMC. The time-bounded *until* is then defined as follows: $\mathbf{x}, t \models \varphi_1 U_{[T_1,T_2]} \varphi_2$ if and only if there exists a time t_1 in the interval $[t+T_1, t+T_2]$ such that $\mathbf{x}, t_1 \models \varphi_2$ and for all t_0 in the interval $[t, t_1]$ we have $\mathbf{x}, t_0 \models \varphi_1$. Other temporal modalities can then be defined using the *until* operator. For example, time-bounded *eventually* and *always* are given by $F_{[T_1,T_2]}\varphi \equiv \mathit{true}\ U_{[T_1,T_2]}\varphi$ and $G_{[T_1,T_2]}\varphi \equiv \neg F_{[T_1,T_2]}\neg\varphi$ respectively.

One of the quantities of interest is the satisfaction probability with respect to a MiTL formula. This is defined as follows.

Definition 4. *Let $\mathcal{M}_\theta$ be pCTMC over parameter vector $\theta \in D \subset \mathbb{R}^k_{>0}$ and φ be a temporal logic formula. The satisfaction probability associated with φ is the probability*

$$f_\varphi(\theta) = P(\varphi = \mathit{true} \mid \mathcal{M}_\theta). \tag{3}$$

Statistical model checking provides a way to estimate the defined satisfaction probability and relies on analysing a set of simulated trajectories of the model $\mathcal{M}_\theta$ for different parameter values θ. Supposing $\mathcal{D}_\theta$ is a set of N trajectories for θ we can give a Monte Carlo estimate of the satisfaction probability

$$f_\varphi(\theta) \approx \bar{f}_\varphi(\theta) = \frac{1}{|N|} \sum_{t \in \mathcal{D}_\theta} \mathbb{1}[t \models \varphi]. \tag{4}$$

In particular, we take a mean over the trajectories in $\mathcal{D}_\theta$ with respect to the indicator function that returns 1 if the trajectory satisfies the temporal logic formula φ and 0 otherwise. Note that this corresponds to the maximum likelihood estimate of the parameter of a Bernoulli distribution. Thus the variance of the estimate is given by

$$Var(\bar{f}_\varphi(\theta)) = \frac{\bar{f}_\varphi(\theta)(1 - \bar{f}_\varphi(\theta))}{N}. \tag{5}$$

The size of the simulated data set $\mathcal{D}_\theta$ can then be chosen such that the variance of the estimate gives acceptable confidence for the intended application. An alternative to statistical methods for estimating the defined satisfaction probability are numerical approximation methods [6] which suffer greatly from state space explosion.

3.3 Approximate parameter inference

Parameter inference for stochastic systems in general, and continuous time Markov chains in particular, is a difficult problem. Supposing we have a pCTMC model $\mathcal{M}_\theta$ and a set of observed data $\mathcal{D}$. The aim of parameter inference is to describe the distribution over parameters θ such that the pCTMC $\mathcal{M}_\theta$ offers a good

model for the data. In the Bayesian setting we are looking to find the posterior distribution over model parameters θ

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

given the observed data $\mathcal{D}$. The posterior is proportional to the product of the likelihood $p(\mathcal{D}|\theta)$ and the prior $p(\theta)$. In the context of parameter inference of stochastic systems like CTMCs the likelihood term is usually computationally intractable. Because of this, the methods that rely on computing the likelihood are infeasible. Thus, likelihood-free methods such as approximate Bayesian computation [15] and synthetic likelihood methods [16] are commonly used to identify model parameter values such that the data simulated by the stochastic model resembles the observed data.

In this section we concentrate on approximate Bayesian computation (ABC). The idea is to simulate the model for different parameters and compare the outcomes with the observed data. Let us consider the observed data $\mathcal{D}$ and the data $\mathcal{D}_\theta$ simulated from a model $\mathcal{M}_\theta$ with a parameter value θ. In this paper we think of both $\mathcal{D}$ and $\mathcal{D}_\theta$ as sets of time-trajectories. The first step in such methods is to reduce the observed data set $\mathcal{D}$ to appropriate summary statistics. Let $\mathcal{D} = \{z_1, \cdots, z_n\}$ denote the observed data set and let the function $S_n : \mathbb{R}^n \to \mathbb{R}^d$ represent the chosen vector of summary statistics. The likelihood $\mathcal{L}(\theta)$ is approximated by

$$L(\theta) = p(S_n(\mathcal{D})|\theta)$$

which is the likelihood of data $\mathcal{D}$ under the summary statistic S_n. This likelihood however is also not known and in practice $L(\theta)$ is approximated further.

The method we consider is based on the summary statistics $S_n(\mathcal{D}_\theta)$ computed from the simulated data $\mathcal{D}_\theta$. In particular, consider the standard ABC rejection sampler. This method relies on defining a discrepancy measure

$$\Delta_\theta = \|S_n(\mathcal{D}) - S_n(\mathcal{D}_\theta)\|$$

for some norm $\|\cdot\|$. ABC methods proceed to draw samples θ from the prior distribution and simulate the corresponding summary statistics $S_n(\mathcal{D}_\theta)$. The parameters θ sampled from the prior $p(\theta)$ are retained as samples from the posterior $p(\theta|\mathcal{D})$ if the corresponding summary statistics $S_n(\mathcal{D})$ and $S_n(\mathcal{D}_\theta)$ are within a chosen distance ϵ of each other. In other words, θ is retained as a posterior sample if the acceptance criterion $\Delta_\theta < \epsilon$ is satisfied.

The notable problem with this approach is the sensitivity of the results to the choice of the discrepancy measure Δ_θ and the threshold. In this paper we work with the discrepancy measure defined in terms of the Euclidean distance between the summary statistics. The choice of the threshold value is generally less obvious and several values need to be tried to find a performing one. Setting the value too high will result in a rejection sampler that discriminates poorly between different parameter values of the model. Setting the value too low results in computation time blowing up as too many samples are rejected.

Another common class of likelihood-free inference methods is synthetic likelihood methods where the intractable likelihood $p(S_n(\mathcal{D}|\theta)$ is approximated by

a normal likelihood. The likelihood is then used in a Markov chain Monte Carlo scheme like the Metropolis-Hastings algorithm. Such methods are out of the scope of this paper and are left for future work.

4 Inference

In this section we discuss the main contribution of this paper — a method for performing inference with quantitative and qualitative information about the system's behaviour. The Bayesian inference problem we are looking to solve is constructing the posterior for model parameters θ under the evidence from observed data $\mathcal{D}$ and logical specification φ. From the Bayes' theorem we know that the posterior of interest is proportional to likelihood multiplied by the prior:

$$p(\theta|\mathcal{D},\varphi) \propto p(\mathcal{D},\varphi|\theta)p(\theta)$$

The likelihood $p(\mathcal{D},\varphi|\theta)$ in the case of continuous time Markov chain models is generally intractable. Further, the likelihood $p(\mathcal{D},\varphi|\theta)$ which considers both quantitative and qualitative observations is more complex than considering the quantitative and qualitative observations independent from each other. In particular, it is clear that conditional on the parameter θ the likelihood of $\mathcal{D}$ and φ are not independent. From the law of conditional probability we know that either

$$p(\mathcal{D},\varphi|\theta) = p(\mathcal{D}|\theta,\varphi)p(\varphi|\theta)$$

or

$$p(\mathcal{D},\varphi|\theta) = p(\varphi|\theta,\mathcal{D})p(\mathcal{D}|\theta)$$

It is not clear how to treat the conditional likelihoods $p(\mathcal{D}|\theta,\varphi)$ and $p(\varphi|\theta,\mathcal{D})$ corresponding to the likelihoods of observed data given parameter θ and specification φ holding and the likelihood of φ holding given observed data $\mathcal{D}$. Neither the ABC methods or synthetic likelihood methods mentioned in Section 3.3 can be applied directly to the problem at hand. In the case of ABC the challenge is defining a set of summary statistics from the available (quantitative and logical) data and a corresponding discrepancy measure. In the case of synthetic likelihood methods we also have a problem with defining a set of summary statistics that are approximately Gaussian.

In order to overcome these challenges we propose the following approach. The quantitative data in the form of observed trajectories is translated into a logical specification $\varphi_{\mathcal{D}}$. The joint likelihood $p(\varphi_{\mathcal{D}},\varphi|\theta)$ can then be estimated via statistical model checking methods. The main difficulty is in constructing the specification $\varphi_{\mathcal{D}}$ from the qualitative data $\mathcal{D}$ such that the posterior $p(\varphi_{\mathcal{D}},\varphi|\theta)$ approximates well the posterior $p(\mathcal{D},\varphi|\theta)$. Note that this problem of choosing a suitable way to transform the quantitative observations into logical specification is in spirit similar to choosing a set of summary statistics and a discrepancy measure in ABC. Once we have an appropriately constructed logical specification we can use rejection sampling to sample from the approximate posterior $p(\theta|\varphi,\varphi_D)$.

4.1 Quantitative data

The most trivial method for transforming the observed trajectories to a logical specification is to define a discrepancy measure $\Delta_\theta = \|S_n(\mathcal{D}) - S_n(\mathcal{D}_\theta)\|$ and a threshold ϵ as done for ABC. We then consider a property that is satisfied when the discrepancy is less than ϵ. Suppose then that φ denotes the logical specification corresponding to qualitative information. When implementing the rejection sampling we only keep the samples θ for which both the discrepancy measure is below a threshold and the logical specification for qualitative information is satisfied.

Alternatively we can convert the quantitative data into the same logic, MiTL, as the qualitative specification. Suppose $\mathcal{D} = \{z_1, \cdots, z_n\}$ are time trajectories observed at a finite number of points. Each trajectory z_i consists of a set

$$\{(y_0^i, t_0), (y_1^i, t_1), \ldots (y_m^i, t_m)\}$$

where y_j^i is a measured vector of data points at time t_j. As with the approximate Bayesian computation we consider the summary statistics $S_n(\mathcal{D}) = \{s_1, \cdots, s_m\}$ at times $t_0, \cdots, t_m$. Recall that $\omega(t)$ is the function that maps a CTMC to its summary statistic at time t. We say that the trajectory is close to a point (s_j, t_j) if for a small time ϵ_t and a time interval $[t_j - \epsilon_t, t_j + \epsilon_t]$ there exists a time $t \in [t_j - \epsilon_t, t_j + \epsilon_t]$ such that $d(\omega(t), s_j) < \delta$ where d is a distance measure defined between the state space of the CTMC model and summary statistic of the observed trajectories. For chosen tolerances ϵ_t and δ this can be encoded as a MiTL specification in the following way

$$F_{[t_j - \epsilon_t, t_j + \epsilon_t]}\{d(\omega(t), s_j) < \delta\}. \tag{6}$$

That is, eventually in the given time interval the discrepancy between the trajectory and the summary statistic from observations is less than δ. A useful special case is when $\epsilon_t = 0.0$ which gives

$$F_{[t_j, t_j]}\{d(\omega(t), s_j) < \delta\}. \tag{7}$$

meaning at time t_j the inequality $d(\omega(t), s_j) < \delta$ holds. In the rest of the paper we are going to work with this special case.

Let us denote this formula by φ_i. The full specification corresponding to the observed time trajectories is then given by the conjunction $\varphi_D = \bigwedge_{i \in \{1, \cdots, m\}} \varphi_i$ corresponding to the model trajectories close to the observed quantitative data. This can be taken together with the logical specification φ for the qualitative data. In particular, the assumption is that $p(\mathcal{D}, \varphi|\theta)$ is approximately proportional to $p(\varphi_D, \varphi|\theta)$.

Similarly to ABC we can then implement rejection sampling from the posterior

$$p(\theta|\varphi_D, \varphi) \propto p(\varphi_D, \varphi|\theta)p(\theta) \tag{8}$$

by randomly sampling from the prior $p(\theta)$ and accepting those samples for which the model $\mathcal{M}_\theta$ satisfies the constructed logical specification. Note that the rejection sampling will only accept samples if there is part of the parameter space

covered by the prior where both the satisfaction probability to both φ_D and φ are non-zero.

4.2 Expected satisfaction probability

The assumption in the previous section that $p(\mathcal{D}, \varphi|\theta)$ is approximately proportional to $p(\varphi_D, \varphi|\theta)$ is difficult to verify in practice. Moreover, as discussed we do not have access to the true posterior $p(\theta|\mathcal{D}, \varphi)$. Thus, in order to study the resulting posteriors we compare the expected satisfaction probabilities calculated over the approximate posterior distributions $p(\theta|\varphi_D)$, $p(\theta|\varphi)$ and $p(\theta|\varphi_D, \varphi)$

Suppose φ is a logical property and $p(\theta|\mathcal{D})$ is a posterior distribution. We can compute the expected satisfaction probabilities over the distribution as

$$\int_\theta f_\varphi(\theta) p(\theta|\mathcal{D}) d\theta. \tag{9}$$

For example, with respect to the posterior arising from the logical specification for quantitative data we have

$$\mathbb{E}_{\varphi_D}\left[f_\varphi(\theta)\right] = \int_\theta f_\varphi(\theta) p(\theta|\varphi_D) d\theta. \tag{10}$$

Clearly, analytical solution of this will be infeasible as neither the satisfaction probability function $f_\varphi(\theta)$ nor the posterior density $p(\theta|\varphi_D)$ have analytical forms. However the expectation can be estimated from the posterior samples $\mathcal{S}$ by giving the following Monte Carlo estimate

$$\mathbb{E}_{\varphi_D}\left[f_\varphi(\theta)\right] \approx \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} f_\varphi(s). \tag{11}$$

The satisfaction probability $f_\varphi(s)$ requires another Monte Carlo estimate as described in Section 3.2. Thus, for example, we can study the change in expected satisfaction probability of properties φ and φ_D when going from posterior $p(\theta|\varphi)$ to $p(\theta|\varphi, \varphi_D)$. This serves as a proxy for the trade-off being made when attempting to fit a posterior distribution with respect to both quantitative and qualitative information.

5 Results

In this section we present parameter inference results for two examples. Namely, a rumour spread model [4] and a client-server model, both defined as CRNs. In both cases the quantitative data considered are samples of trajectories at a given set of times. Both approaches of describing the time trajectory as a logical property presented in the previous section require us to specify a threshold parameter and are analogous to each other in the context of rejection sampling.

5.1 Rumour spread model

To illustrate the proposed method we are going to first consider a simple rumour spread model. This model is a variant of the commonly studied susceptible-infected-recovered model from epidemiology. The agents in the model can be in one of three states — ignorant I, spreading S or repressing R. As a chemical reaction network the model is given as follows.

$$S + I \xrightarrow{k_I} S + S$$
$$S + S \xrightarrow{k_R} S + R$$
$$R + S \xrightarrow{k_R} R + R$$

In particular, the rumour spreads from spreading individuals to ignorant. Only novel rumours are deemed worthy of sharing and thus if two spreaders meet, one of them will stop spreading the rumour. Finally, a repressing individual will quash the rumour if they meet a spreading individual, and turn them into a repressor. We assume two parameters k_I and k_R governing the rates at which ignorants become spreaders and spreading individuals are converted to repressing ones. The initial state of the model is given by 10 ignorants and 5 spreaders.

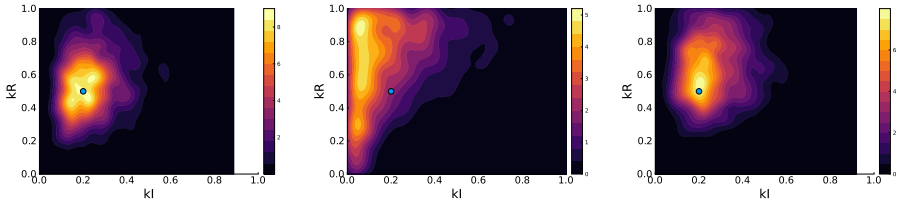
The experimental set up is as follows. For the quantitative data we assume that only one of the species in the chemical reaction network is directly observed. In particular, suppose only the number of repressors is quantitatively measured. The quantitative observations are taken to be the mean values of 1000 trajectories simulated from the model with fixed parameters at 6 equally spaced time-points in the time-interval $[0.0, 5.0]$. We repeat the experiment for 3 different parameter combinations from the half-open intervals $k_I \in (0.0, 1.0]$ and $k_R \in (0.0, 1.0]$. Namely, we consider $\{k_I = 0.2, k_R = 0.5\}$, $\{k_I = 0.4, k_R = 0.8\}$ and $\{k_I = 0.1, k_R = 0.1\}$. We then construct the discrepancy measure and MiTL-based logic specifications for the quantitative data as described in Section 4.

For the qualitative information let us suppose the following condition:

$$G_{[0,1.0]}\{I > S\}. \tag{12}$$

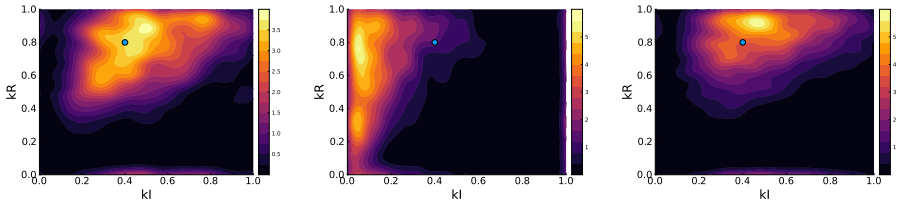
That is, in the time-interval $[0.0, 1.0]$ the number of ignorants is greater than the number of spreaders. Clearly, this indirectly constrains the rate at which the rumour spreads. Note that we have now said something about all three populations in the model.

The parameter for the MiTL formula construction and rejection criterion for the approximate Bayesian inference construction is chosen such that the rejection rates are similar between the two conditions. In particular, we fix the acceptance criterion parameter for the ABC as $\epsilon = 1.0$. With that in mind, choosing the parameters for the MiTL formula construction to be $\delta = \sqrt{\frac{1}{6}}$ and $\epsilon_t = 0.0$ ensures that a parameter accepted based on the MiTL formula is also accepted by the Euclidean metric-based condition. In this case this selection of parameters also happens to give a similar acceptance rate between the two methods.



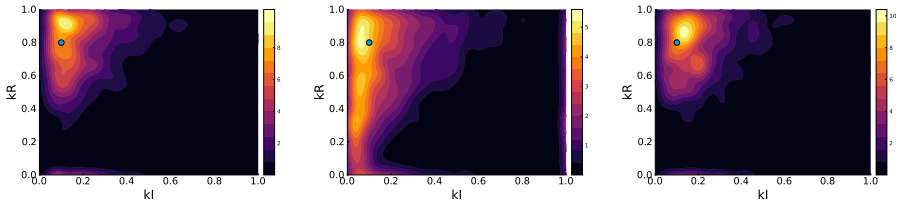
(a) Quantitative. (b) Qualitative. (c) Quantitative and qualitative.

Fig. 1: Plot of the kernel density estimator based on accepted parameters for the rumour spread model. The quantitative data was sampled from the model with parameters $k_I = 0.2$ and $k_R = 0.5$. (a) shows the approximate posterior based on quantitative information only. (b) shows the posterior based on qualitative information only. (c) shows the posterior for the combined information as found by method in Section 4.



(a) Quantitative. (b) Qualitative. (c) Quantitative and qualitative.

Fig. 2: Same as Figure 1 but with the quantitative data sampled from the model with parameters $k_I = 0.4$ and $k_R = 0.8$.



(a) Quantitative. (b) Qualitative. (c) Quantitative and qualitative.

Fig. 3: Same as Figure 1 but with the quantitative data sampled from the model with parameters $k_I = 0.1$ and $k_R = 0.8$.

Figures 1, 2 and 3 show the distributions resulting from the rejection sampling based on both quantitative and qualitative information as well as each separately. For each posterior we gathered 1000 samples. As discussed in Section 4.2 we assess the method by comparing the recovered posterior distribution when only quantitative, only qualitative and both sets of information are used.

Table 1: Expected satisfaction probabilities under different combinations of logical specifications.

Parameters	Formula	Expect. sat. w.r.t $p(\theta\|\varphi_D)$	$p(\theta\|\varphi)$	$p(\theta\|\varphi, \varphi_D)$
$(0.2, 0.5)$				
MiTL	φ	0.37	0.63	0.39
	φ_D	0.021	0.009	0.017
Discrepancy measure	φ	0.39	0.63	0.41
	φ_D	0.033	0.015	0.028
$(0.4, 0.8)$				
MiTL	φ	0.20	0.64	0.21
	φ_D	0.11	0.048	0.12
Discrepancy measure	φ	0.21	0.64	0.23
	φ_D	0.12	0.056	0.13
$(0.1, 0.8)$				
MiTL	φ	0.62	0.63	0.61
	φ_D	0.11	0.077	0.11
Discrepancy measure	φ	0.64	0.63	0.64
	φ_D	0.13	0.095	0.13

In particular, Table 1 shows the trade-offs that are made to make the posterior agree with both types of data as closely as possible. The trade-off is smaller in the cases where the two posteriors $p(\theta|\varphi)$ and $p(\theta|\varphi_D)$ have the most overlap. Here, this would be the parameter combination $k_I = 0.1$ and $k_R = 0.8$. It can also be seen that there is very little difference between the two methods for taking into account the quantitative trajectories. However, the interpretation or construction of the MiTL method is simpler. In our case the parameter was chosen to give similar results with the discrepancy measure-based condition. However, the MiTL formula construction parameter can easily be interpreted as the measurement noise or uncertainty.

5.2 Client server model

As the second example we consider the client-server model implemented as the following chemical reaction network

$$\begin{aligned} \mathit{Client} + \mathit{Server} &\xrightarrow{k_{req}} \mathit{ClientThink} + \mathit{Server} \\ \mathit{ClientThink} &\xrightarrow{k_{thk}} \mathit{Client} \\ \mathit{Server} &\xrightarrow{k_{brk}} \mathit{ServerBroken} \\ \mathit{ServerBroken} &\xrightarrow{k_{fix}} \mathit{Server} \end{aligned}$$

The model consists of two types of agents, clients and servers. The clients request data from the server, receive the requested data and perform some independent

action with the data. The servers in addition to serving the clients are susceptible to failure. The initial state of the model is given by 20 clients and 3 servers.

The model parameters are the various rates with which the transitions occur. There are four rate parameter k_{req}, k_{thk}, k_{brk} and k_{fix}. We are going to assume that two of them, namely k_{thk}, k_{brk} are fixed at 0.1 and 0.2 respectively. In the following we use the methods previously described to fit the remaining two parameters.

Let us assume that the data available for parametrising the model are the time-series observations of the available servers. Furthermore, let us assume the following logical formula

$$G_{[0,10.0]}\{Client > ClientThink\}. \tag{13}$$

This means, in the time interval $[0, 10.0]$, we always have more clients ready to request data from the server than thinking. In this case again we have quantitative and qualitative information about different aspects of the system dynamics.

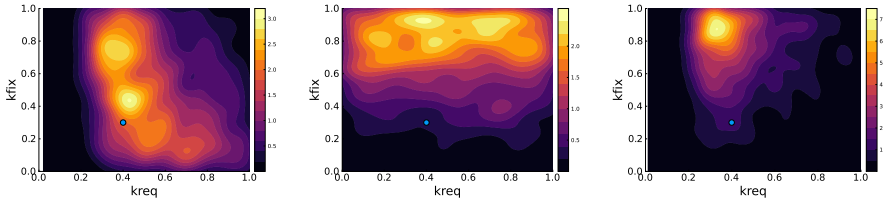
The threshold for the discrepancy measure and MiTL formula construction parameter were again chosen by hand: they were fixed as $\epsilon = 3.0$ and $\{\delta = 2.05, \epsilon_t = 0.0\}$ respectively to give similar acceptance rates. The experiments were repeated for three parametrisations of the pCTMC model. Namely, $\{k_{req} = 0.4, k_{fix} = 0.3\}$, $\{k_{req} = 0.2, k_{fix} = 0.6\}$ and $\{k_{req} = 0.8, k_{fix} = 0.8\}$.

Figures 4, 5 and 6 show the distributions resulting from the rejection sampling based on both quantitative and qualitative information as well as each separately. In this case the effect of taking the quantitative and qualitative information together has a stronger effect on the resulting posteriors than in the previous case-study. As discussed in Section 4.2 we assess the method by comparing the recovered posterior distribution when only quantitative, only qualitative and both sets of information are used. Similarly to the previous examples, Table 2 gives an overview of the trade-off in satisfaction relative to the quantitative and qualitative data when attempting to satisfy both.

6 Conclusions

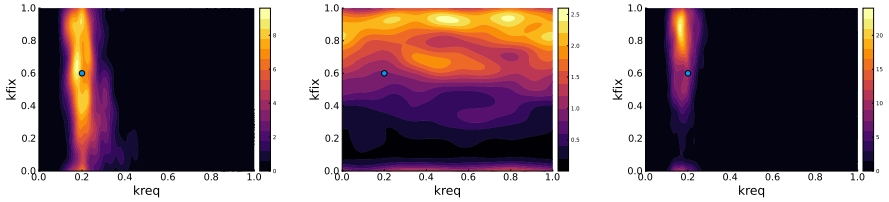
In this paper we presented a method for performing inference based on a combination of quantitative and qualitative data. Our approach was to construct a rejection sampling criterion such that the accepted samples correspond to trajectories that are close to the quantitative data and also satisfy the logical property. To achieve that, we provided a way of converting the quantitative data into a set of logical formulae. The benefit of this logic-based construction over standard approximate Bayesian rejection sampling acceptance criteria given in terms of discrepancy measures is that this construction can easily be interpreted in terms of measurement noise or uncertainty.

We studied the proposed method for two examples — a rumour spread model, which is based on the standard susceptible-infected-recovered model, and a simple client-server model. In both case we supplemented the quantitative data which gave information about one aspect of the model with quantitative data



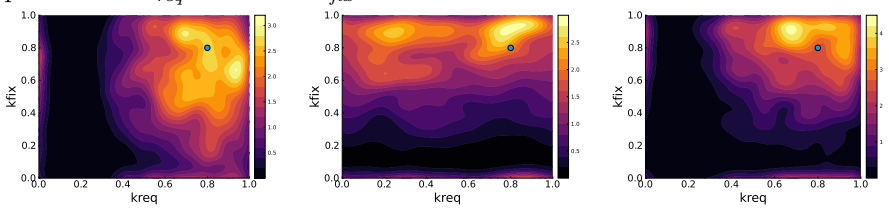
(a) Quantitative. (b) Qualitative. (c) Quantitative and qualitative.

Fig. 4: Plot of the kernel density estimator based on accepted parameters for the client server model. The quantitative data was sampled from the model with parameters $k_{req} = 0.4$ and $k_{fix} = 0.3$. (a) shows the approximate posterior based on quantitative information only. (b) shows the posterior based on qualitative information only. (c) shows the posterior for the combined information as found by method in Section 4.



(a) Quantitative. (b) Qualitative. (c) Quantitative and qualitative.

Fig. 5: Same as Figure 4 with the quantitative data sampled from the model with parameters $k_{req} = 0.2$ and $k_{fix} = 0.6$.



(a) Quantitative. (b) Qualitative. (c) Quantitative and qualitative.

Fig. 6: Same as Figure 4 with the quantitative data sampled from the model with parameters $k_{req} = 0.8$ and $k_{fix} = 0.8$.

given as a logic specification. We saw that the logic-based specification successfully supplements the quantitative information. It allows us to refine the posterior compared to the rejection sampling ABC with only quantitative data.

Further work can be done to investigate the effects of attributing weight to different sources of information. For example, more reliable pieces can be given

Table 2: Expected satisfaction probabilities under different combinations of logical specifications.

Parameters	Formula	Expect. sat. w.r.t		
		$p(\theta\|\varphi_D)$	$p(\theta\|\varphi)$	$p(\theta\|\varphi,\varphi_D)$
$(0.4, 0.3)$				
MiTL	φ	0.07	0.12	0.12
	φ_D	0.18	0.11	0.19
Discrepancy measure	φ	0.06	0.12	0.12
	φ_D	0.17	0.10	0.20
$(0.2, 0.6)$				
MiTL	φ	0.08	0.12	0.12
	φ_D	0.08	0.015	0.010
Discrepancy measure	φ	0.08	0.12	0.12
	φ_D	0.08	0.014	0.09
$(0.8, 0.8)$				
MiTL	φ	0.09	0.12	0.12
	φ_D	0.63	0.42	0.66
Discrepancy measure	φ	0.09	0.12	0.12
	φ_D	0.67	0.44	0.7

more weight during the parameter fitting process. This can improve accuracy of the model and allow information with different certainties to be used more effectively. Even in the case of information defined in a logic specification we may want to assign an uncertainty to the specification. In addition we are looking to provide more rigorous justification for the method used in this paper. Finally, further work can be done to improve the accuracy and robustness of the results.

References

1. Anderson, D. F., and Kurtz, T. G. *Continuous Time Markov Chain Models for Chemical Reaction Networks*. Springer New York, New York, NY, 2011, pp. 3–42.
2. Aziz, A., Sanwal, K., Singhal, V., and Brayton, R. K. Verifying continuous time markov chains. In *Computer Aided Verification, 1996, Proceedings* (1996), R. Alur and T. A. Henzinger, Eds., vol. 1102 of *Lecture Notes in Computer Science*, Springer, pp. 269–276.
3. Balbo, G. Introduction to generalized stochastic petri nets. In *Formal Methods for Performance Evaluation, 7th International School on Formal Methods for the Design of Computer, Communication, and Software Systems, SFM 2007, Bertinoro, Italy, May 28-June 2, 2007, Advanced Lectures* (2007), M. Bernardo and J. Hillston, Eds., vol. 4486 of *Lecture Notes in Computer Science*, Springer, pp. 83–131.
4. Bortolussi, L., and Sanguinetti, G. Learning and designing stochastic processes from logical constraints. *Log. Methods Comput. Sci. 11*, 2 (2015).
5. Ceska, M., Dannenberg, F., Paoletti, N., Kwiatkowska, M., and Brim, L. Precise parameter synthesis for stochastic biochemical systems. *Acta Informatica 54*, 6 (2017), 589–623.

6. Chen, T., Diciolla, M., Kwiatkowska, M. Z., and Mereacre, A. Time-bounded verification of ctmcs against real-time specifications. In *Formal Modeling and Analysis of Timed Systems - 9th International Conference, FORMATS 2011, Aalborg, Denmark, September 21-23, 2011. Proceedings* (2011), U. Fahrenberg and S. Tripakis, Eds., vol. 6919 of *Lecture Notes in Computer Science*, Springer, pp. 26–42.
7. Georgoulas, A., Hillston, J., and Sanguinetti, G. Proppa: Probabilistic programming for stochastic dynamical systems. *ACM Trans. Model. Comput. Simul. 28*, 1 (2018), 3:1–3:23.
8. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry 81*, 25 (1977), 2340–2361.
9. Hillston, J. *A compositional approach to performance modelling.* PhD thesis, University of Edinburgh, UK, 1994.
10. Legay, A., Lukina, A., Traonouez, L., Yang, J., Smolka, S. A., and Grosu, R. Statistical model checking. In *Computing and Software Science - State of the Art and Perspectives*, B. Steffen and G. J. Woeginger, Eds., vol. 10000 of *Lecture Notes in Computer Science*. Springer, 2019, pp. 478–504.
11. Maler, O., and Nickovic, D. Monitoring temporal properties of continuous signals. In *FORMATS/FTRTFT 2004, Grenoble, France, September 22-24, Proceedings* (2004), vol. 3253 of *Lecture Notes in Computer Science*, Springer, pp. 152–166.
12. Mitra, E. D., Dias, R., Posner, R. G., and Hlavacek, W. S. Using both qualitative and quantitative data in parameter identification for systems biology models. *Nature Communications 9*, 1 (Sept. 2018).
13. Mitra, E. D., and Hlavacek, W. S. Bayesian inference using qualitative observations of underlying continuous variables. *Bioinformatics 36*, 10 (02 2020), 3177–3184.
14. Schmiester, L., Weindl, D., and Hasenauer, J. Statistical inference of mechanistic models from qualitative data using an efficient optimal scaling approach. *bioRxiv* (2019).
15. Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface 6*, 31 (2009), 187–202.
16. Wood, S. N. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature 466*, 7310 (2010), 1102–1104.
17. Xu, Z., Wu, B., and Topcu, U. Control strategies for covid-19 epidemic with vaccination, shield immunity and quarantine: A metric temporal logic approach. *PLOS ONE 16*, 3 (03 2021), 1–20.