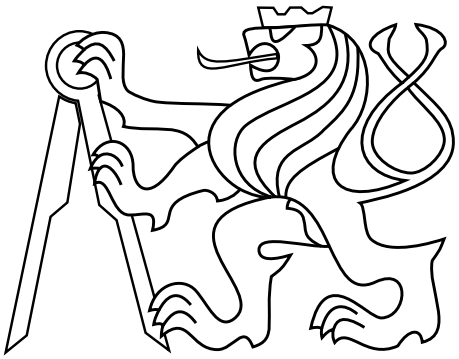
Master Thesis

# Bayesian Parameter Estimation of State-Space Models with Intractable Likelihood

Bc. Tomáš Kala
Supervisor: Ing. Kamil Dedecius, PhD.

April 2019



Department of Computer Science
Faculty of Electrical Engineering
Czech Technical University in Prague

# ZADÁNÍ DIPLOMOVÉ PRÁCE

## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Kala** Jméno: **Tomáš** Osobní číslo: **434690**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra počítačů**

Studijní program: **Otevřená informatika**

Studijní obor: **Bioinformatika**

## II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

**Bayesovské odhadování parametrů stavových modelů při nedostupné věrohodnostní funkci**

Název diplomové práce anglicky:

**Bayesian parameter estimation of state-space models with intractable likelihood**

Pokyny pro vypracování:

Stavové modely představují velmi populární formalismus vhodný pro popis celé řady různých náhodných procesů, od časových řad po aplikace v teorii řízení. Pokud tyto modely neobsahují statické parametry, lze pro jejich odhad použít např. Kalmanův filtr a jeho varianty, dále particle filtraci aj. Pokud ovšem statické parametry obsahují, tyto filtry zpravidla nekonvergují a nezbývá, než přikročit k optimalizačním technikám typu maximalizace věrohodnosti či particle Markov chain Monte Carlo. Další komplikace nastávají, pokud navíc není věrohodnostní funkce pro pozorovanou veličinu dostupná, nebo je nepřesná či příliš komplikovaná. Diplomová práce je specificky zaměřena poslední zmíněnou problematiku.
Specifické pokyny
1. Seznamte se s metodami pro odhadování stavových modelů pomocí kalmanovské filtrace a sekvenční Monte Carlo filtrace. Nastudujte problematiku statických parametrů a jejich odhadu.
2. Proveďte rešerši ohledně využití daných metod v oblasti bioinformatiky, například v modelování buněčných procesů.
3. Seznamte se s metodami ABC - Approximate Bayesian Computation a jejich využití ve filtraci stavových modelů.
4. Navrhněte vhodný způsob odhadování statických parametrů stavových modelů s využitím metod ABC.
5. Experimentálně (na vhodném problému z oblasti molekulární biologie) a případně teoreticky ověřte vlastnosti navržené metody, diskutujte její vlastnosti a navrhněte další možné směry výzkumu.

Seznam doporučené literatury:

1] C. C. Drovandi, A. N. Pettitt, and R. A. McCutchan, “Exact and approximate Bayesian inference for low integer-valued time series models with intractable likelihoods,” Bayesian Anal., vol. 11, no. 2, pp. 325–352, 2016.
[2] S. Martin, A. Jasra, S. S. Singh, N. Whiteley, P. Del Moral, and E. McCoy, “Approximate Bayesian Computation for Smoothing,” Stoch. Anal. Appl., vol. 32, no. 3, pp. 397–420, 2014.
[3] T. B. Schön, A. Svensson, L. Murray, and F. Lindsten, “Probabilistic learning of nonlinear dynamical systems using sequential Monte Carlo,” Mech. Syst. Signal Process., vol. 104, pp. 866–883, May 2018.
[4] C. Andrieu, A. Doucet, and R. Holenstein, “Particle Markov chain Monte Carlo methods,” J. R. Stat. Soc. Ser. B (Statistical Methodol., vol. 72, no. 3, pp. 269–342, Jun. 2010.
[5] K. Dedecius, “Adaptive kernels in approximate filtering of state-space models,” Int. J. Adapt. Control Signal Process., vol. 31, no. 6, pp. 938–952, Jun. 2017.

Jméno a pracoviště vedoucí(ho) diplomové práce:

**Ing. Kamil Dedecius, Ph.D., ÚTIA AV ČR**

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **04.02.2019** Termín odevzdání diplomové práce: **24.05.2019**

Platnost zadání diplomové práce: **20.09.2020**

| Ing. Kamil Dedecius, Ph.D.<br>podpis vedoucí(ho) práce | podpis vedoucí(ho) ústavu/katedry | prof. Ing. Pavel Ripka, CSc.<br>podpis děkana(ky) |
|---|---|---|

## III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

. | Datum převzetí zadání

Podpis studenta

**Abstract**

Abstract in English

**Abstrakt**

Abstract in Czech

# Author statement for graduate thesis:

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

Prague, date ..... .....

signature

# Acknowledgements

# Contents

# Chapter 1

# Introduction

Probabilistic modelling arises in a wide variety of situations. Often, the measurements one uses to perform inference have been carried out with an unknown error. Frequently, one also does not have access to a correct model for the particular situation — the true model is either unknown, or such model is even impossible to formulate.

In the former case, one naturally assumes a random error associated with the observations, and attempts to infer something from the data while accounting for this randomness.

In the latter case, one has no choice but to work with a given, although possibly simplified model, purely because of insufficient domain knowledge. Connected with such a model is some degree of uncertainty about its parameters. It is often beneficial to think of these parameters as random variables themselves, in accordance with the Bayesian methodology (Robert, 2007). Such formulation allows to formulate one's prior beliefs about the parameter values, and then updating them upon receiving new observations.

In this thesis, we work with state-space models (SSMs) consisting of a sequence of observed random variables $y_t$ indexed by discrete time $t = 1, \ldots, T$, which have been generated by a latent random process $x_t,\ t = 1, \ldots, T$. The distribution of $x_t$ and $y_t$ is assumed to be parameterized by a static parameter $\theta$. Our goal is to perform posterior inference about this parameter, given a sequence of observations $y_t$. Furthermore, we assume that the likelihood function of the SSM is intractable and must be approximated. This assumption is well-grounded, as the likelihood is only available in severely restricted cases, to be discussed in Chapter 3, together with a formal definition of the SSM.

Our contribution is twofold. First, we show how to apply the Approximate Bayesian Computation (ABC) methodology (Rubin et al., 1984; Pritchard et al., 1999) to obtain an estimate of the likelihood even under a misspecified observation model. Our formulation allows for arbitrary kernel functions with automatically determined widths, unlike the simple accept-reject routine typically discussed in the literature. Second, we apply the resulting model to the genetic auto-regulation process in prokaryotes. Such situation is suitable for a state-space model with a possibly misspecified observation model, as all attempts to model such a complex system are necessarily simplified. To quote the famous statistician George E. P. Box, *"all models are wrong, but some are useful"* (Box, 1979). This statement is particularly true for such situations, and it is our hope that our model is indeed useful.

The rest of the thesis is organized as follows:

In Chapter 2, we discuss the related work. First, we review some classical works on Markov Chain Monte Carlo (MCMC) methods. We also discuss how these can be applied to state-space models with an intractable likelihood function. Next, we overview the literature on ABC methods and how these could be used to obtain a suitable likelihood estimate even when the observation model is incorrect. Finally, we discuss the application of related methods in bioinformatics and molecular biology.

In Chapter 3, we describe the assumed form of a state-space model. We show how one would implement a sampler to approximately infer the static parameters given a sequence of observations. We also show that in this basic form, such sampler is unusable, since it relies on the evaluation of the likelihood function of the observed sequence, which is intractable (up to certain special cases). We then describe how this likelihood can be estimated using the particle filter (Doucet et al., 2001) without affecting the asymptotical properties of the sampler.

Chapter 4 provides a description of the ABC method, and also how it can be applied to estimate the likelihood even under a misspecified model. We discuss the pros and cons of such approach as well as potential issues and how to address them.

Chapter 5 provides numerical studies, where we apply the model developed in Chapter 4 to several examples and compare it with the model utilizing the particle filter. This chapter also includes the prokaryotic auto-regulation study discussed earlier.

Finally, Chapter 6 concludes the thesis and discusses some possible directions to be investigated in the future.

# Chapter 2

# Related work

In this chapter, we provide a survey of literature relevant to our task. Addressed will be works on the use of Markov Chain Monte Carlo methods for approximate inference, works on approximating the likelihood of state-space models by the particle filter, and on Approximate Bayesian Computation methods. We also provide a section describing the use of the considered models in bioinformatics, focusing on molecular biology and genetics.

## 2.1 Markov Chain Monte Carlo methods

Markov Chain Monte Carlo (MCMC) can be summarized as algorithms designed to simulate random samples from a distribution of interest, which itself is too complicated to sample directly. Assuming the probability density function of this distribution can be evaluated (at least to a multiplicative constant), MCMC methods work by designing a Markov chain whose stationary distribution is the target one.

An attractive property is that the transition distribution of such chain need not resemble the target distribution even closely, and that the problem is relatively unaffected by the dimensionality of the distribution. The downside is a difficulty to determine convergence — for how long should a chain be ran in order to approximately reach this stationary distribution. In addition, one typically requires independent samples from the target distribution, which, however, the Markov chain samples are *not*. Typically, one needs to "thin" the Markov chain samples by keeping every $n$th one to ensure their approximate independence.

Perhaps the best known MCMC algorithm is the Metropolis algorithm (Metropolis et al., 1953), later improved by Hastings (1970). Random samples are iteratively generated from the Markov chain transition distribution, called the proposal distribution in this context. Each such sample is then compared with the previous one, and accepted with a certain probability which ensures that the stationary distribution is indeed the target. The go-to reference for Monte Carlo methods is Robert and Casella (2005). A particularly appealing treatment of MCMC methods with applications towards physics and machine learning can be found in MacKay (2002).

## 2.2 Parameter inference in state-space models

Assuming that the state-space model (SSM) takes the form informally stated in Chapter 1 and more formally given in Chapter 3, if all the parameters of interest are changing in time, that is, the inference is about $x_t$ given $y_t$, one arrives at the task of filtering.

If the transition distribution from state $x_t$ to state $x_{t+1}$ is linear in the states and corrupted by uncorrelated additive noise with mean 0, this task can be solved exactly by the Kalman filter (Kalman, 1960). The resulting filter is then optimal with respect to the mean squared error. An especially nice overview of the Kalman filter connecting it with other linear statistical models is Roweis and Ghahramani (1999).

Once the state transition becomes non-linear, one can use various generalizations of the Kalman filter, such as the extended Kalman filter, which locally linearizes the transition distribution, or the unscented Kalman filter (Julier and Uhlmann, 1997).

In recent years, though, the particle filter (Doucet et al., 2001) has become the most popular alternative due to its particularly simple implementation, appealing asymptotic properties and the fact that it allows for the transition model to be arbitrarily non-linear. The algorithm uses a relatively small number of random samples to approximate the distribution of $x_t$ given $y_1, \ldots, y_t$ at any given time $t$. Since the particle filter is used later in Chapter 3, we postpone a more detailed description there.

On the other hand, if some of the unknown parameters are static, the task becomes more complex. Simply applying MCMC algorithms or other approximations is not possible, as the likelihood function, which is a part of the Metropolis-Hastings algorithm, cannot be evaluated. The paper Andrieu et al. (2010) introduced the idea of using the particle filter to obtain an estimate of the likelihood, which has been shown in Del Moral (2004) to preserve the stationary distribution of the underlying Markov chain. The resulting algorithm is called *Marginal Metropolis-Hastings*. A more recent overview can be found in the tutorial by Schön et al. (2017).

## 2.3 Approximate Bayesian Computation

In its original formulation, the method of Approximate Bayesian Computation (ABC) provides a way to approximate the posterior distribution $p(\theta \mid y) \propto f(y \mid \theta)p(\theta)$, assuming that the prior $p(\cdot)$ is fully known, and that the likelihood $f(\cdot \mid \theta)$ can be sampled from, but not evaluated (Rubin et al., 1984; Pritchard et al., 1999). A more recent overview of ABC methods can be found in Marin et al. (2012).

Briefly, the ABC method works by simulating a sample $\hat{\theta}$ from the prior, substituting it to the likelihood, and generating pseudo-observations $\hat{y}$. These are then compared to the real observations $y$, and if they are "similar enough", the sample $\hat{\theta}$ is accepted. Otherwise, it is rejected. The posterior distribution of $\theta$ is then given in terms of these random samples $\hat{\theta}$. This variant is referred to as the accept-reject ABC, for obvious reasons.

In this thesis, we apply the ABC method in place of the particle filter to allow for inference about the static parameter $\theta$ when the likelihood is not available. In addition, the use of ABC allows for a possibly misspecified observation model of the SSM, which is often the case, as one may not possess the necessary domain knowledge or computational power needed for the real model. Such a situation has been considered in Jasra (2015), although only through the use of the accept-reject variant given above.

Since accepting a sufficient number of samples may take a long time, an idea is to measure the distance between the true and pseudo-observations through a kernel function. This formulation would not reject any samples — instead, they would get assigned a lower weight. This is considered for instance in Dedecius (2017), along with a proposed way to automatically tune the kernel width. How to exactly apply the ABC method to our problem will be addressed in Chapter 4 in detail.

## 2.4 Applications to molecular biology

Finally, we review works describing how the framework of SSMs and the parameter inference in those can be applied in the context of bioinformatics, focusing on problems of molecular biology and genetics.

The go-to reference for stochastic modelling in biology is Wilkinson (2011). It contains a broad overview of applications of various probabilistic models to examples from molecular biology and chemistry. Included is a description of the Gillespie algorithm Gillespie (1976, 1977) used to simulate chemical reactions, which we will use in Chapter 5.

A recent application of SSMs to molecular biology can be found in Golightly and J Wilkinson (2011), where the authors use the particle filter to approximate the unknown likelihoods of various biological models. We will implement these examples in Chapter 5 and compare them with the ABC approximation.

The paper d'Alché Buc et al. (2007) views biological networks such as gene regulatory networks or signalling pathways as SSMs, and estimates their parameters. The static parameters of the model are viewed as dynamic states which, however, do not change in time. The unscented Kalman filter is then applied to estimate these "dynamic" parameters. Such approach is simple, as it does not require the use of MCMC algorithms, but comes without the appealing asymptotical properties of MCMC inference.

Wang et al. (2009), Sun et al. (2008) and Zeng et al. (2011) proceed in a similar fashion when estimating the parameters of various biochemical networks. The used models are only mildly non-linear, and so the extended Kalman filter is sufficient, again without any asymptotical guarantees of identifying the true parameters, however.

An interesting approach to learning the structure of a gene regulatory network from a gene expression time series can be found in Noor et al. (2012). First, the particle filter is applied to learn the hidden states of the network. Once these hidden states are known, the LASSO regression is applied to learn a sparse representation of the regulatory network, since each gene is assumed to interact only with a small number of other genes.

# Chapter 3

# Learning the parameters of a state-space model

# Chapter 4

# Approximate Bayesian Computation

# Chapter 5

# Applications

## 5.1 Preliminary: the Gillespie algorithm

## 5.2 Lotka-Volterra model

## 5.3 Prokaryotic auto-regulation model

# Chapter 6

# Conclusion and future work

# Bibliography


C. Andrieu, A. Doucet, and R. Holenstein. Particle markov chain monte carlo methods (with discussion). *Journal of the Royal Statistical Society, Series B*, 72:1–33, 01 2010.

G. E. Box. Robustness in the strategy of scientific model building. In *Robustness in statistics*, pages 201–236. Elsevier, 1979.

F. d'Alché Buc, M. Quach, and N. Brunel. Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference. *Bioinformatics*, 23 (23):3209–3216, 12 2007. ISSN 1367-4803. doi: 10.1093/bioinformatics/btm510. URL `https://doi.org/10.1093/bioinformatics/btm510`.

K. Dedecius. Adaptive kernels in approximate filtering of state-space models. *International Journal of Adaptive Control and Signal Processing*, 31(6):938–952, 2017. doi: 10.1002/acs.2739. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/acs.2739`.

P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems With Applications*, volume 100. 05 2004. ISBN 0387202684. doi: 10.1007/978-1-4684-9393-1.

A. Doucet, A. Smith, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Information Science and Statistics. Springer New York, 2001. ISBN 9780387951461. URL `https://books.google.cz/books?id=uxX-koqKtMMC`.

D. T. Gillespie. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *Journal of Computational Physics*, 22:403–434, Dec. 1976. doi: 10.1016/0021-9991(76)90041-3.

D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977. doi: 10.1021/j100540a008. URL `https://doi.org/10.1021/j100540a008`.

A. Golightly and D. J Wilkinson. Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo. *Interface focus*, 1:807–20, 12 2011. doi: 10.1098/rsfs.2011.0047.

W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi: 10.1093/biomet/57.1.97. URL `http://biomet.oxfordjournals.org/cgi/content/abstract/57/1/97`.

A. Jasra. Approximate bayesian computation for a class of time series models. *International Statistical Review*, 83(3):405–435, 2015. doi: 10.1111/insr.12089. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12089`.

S. J. Julier and J. K. Uhlmann. A new extension of the kalman filter to nonlinear systems. pages 182–193, 1997.

R. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering (ASME)*, 82D:35–45, 01 1960. doi: 10.1115/1.3662552.

D. J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, New York, NY, USA, 2002. ISBN 0521642981.

J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder. Approximate bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180, 2012.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

A. Noor, E. Serpedin, M. N. Nounou, and H. N. Nounou. Inferring gene regulatory networks via nonlinear state-space models and exploiting sparsity. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9:1203–1211, 2012.

J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. Population growth of human y chromosomes: a study of y chromosome microsatellites. *Molecular biology and evolution*, 16(12):1791–1798, 1999.

C. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2005. ISBN 0387212396.

S. Roweis and Z. Ghahramani. A unifying review of linear gaussian models. *Neural Comput.*, 11(2):305–345, Feb. 1999. ISSN 0899-7667. doi: 10.1162/089976699300016674. URL `http://dx.doi.org/10.1162/089976699300016674`.

D. B. Rubin et al. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, 12(4):1151–1172, 1984.

T. Schön, A. Lindholm, L. Murray, and F. Lindsten. Probabilistic learning of nonlinear dynamical systems using sequential monte carlo. *Mechanical Systems and Signal Processing*, 03 2017. doi: 10.1016/j.ymssp.2017.10.033.

X. Sun, L. X. Jin, and M. Xiong. Extended kalman filter for estimation of parameters in nonlinear state-space models of biochemical networks. *PLoS ONE*, 3:1220 – 4, 2008.

Z. Wang, X. Liu, Y. Liu, J. Liang, and V. Vinciotti. An extended kalman filtering approach to modeling nonlinear dynamic gene regulatory networks via short gene expression time series. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*, 6:410–9, 07 2009. doi: 10.1109/TCBB.2009.5.

D. Wilkinson. *Stochastic Modelling for Systems Biology, Second Edition*. Chapman & Hall/CRC Mathematical and Computational Biology. Taylor & Francis, 2011. ISBN 9781439837726. URL `https://books.google.cz/books?id=G3BaHtBrW68C`.

N. Zeng, Z. Wang, Y. Li, M. Du, and X. Liu. Inference of nonlinear state-space models for sandwich-type lateral flow immunoassay using extended kalman filtering. *IEEE Transactions on Biomedical Engineering*, 58:1959–1966, 2011.