State Space Models for Regional Epidemiolgical Indicators

Stefan Heyder

October 2023 – Draft v 0.1

Abstract

Short summary of the contents in English...

Zusammenfassung

Kurze Zusammenfassung des Inhaltes in deutscher Sprache...

Publications and Contributions

This thesis consists of mostly unpublished work. During my time as a PhD student I have, however, been fortunate to collarborate with many scientists on problems in mathematical epidemiology with a focus on COVID-19, which resulted in several publications. In this section I want to clarify what my contributions to these publications were and which contributions of the present thesis are new.

[1] J. Bracher et al. "A Pre-Registered Short-Term Forecasting Study of COVID-19 in Germany and Poland during the Second Wave." In: *Nature Communications* 12.1 (1 Aug. 27, 2021), p. 5173. ISSN: 2041-1723. DOI: 10.1038/s41467-021-25207-0. URL: https://www.nature.com/articles/s41467-021-25207-0 (visited on 09/30/2021).

[2] Johannes Bracher et al. "National and Subnational Short-Term Forecasting of COVID-19 in Germany and Poland during Early 2021." In: *Communications Medicine* 2.1 (1 Oct. 31, 2022), pp. 1–17. ISSN: 2730-664X. DOI: 10.1038/s43856-022-00191-8. URL: https://www.nature.com/articles/s43856-022-00191-8 (visited on 11/16/2022).

[3] Jan Pablo Burgard et al. "Regional Estimates of Reproduction Numbers with Application to COVID-19." Aug. 31, 2021. arXiv: 2108.13842 [stat]. URL: http://arxiv.org/abs/2108.13842 (visited on 09/30/2021).

[4] Sara Grundel et al. "How Much Testing and Social Distancing Is Required to Control COVID-19? Some Insight Based on an Age-Differentiated Compartmental Model." In: *SIAM Journal on Control and Optimization* 60.2 (Apr. 2022), S145–S169. ISSN: 0363-0129, 1095-7138. DOI: 10.1137/20M1377783. URL: https://epubs.siam.org/doi/10.1137/20M1377783 (visited on 11/16/2022).

[5] Sara M. Grundel et al. "How to Coordinate Vaccination and Social Distancing to Mitigate SARS-CoV-2 Outbreaks." In: *SIAM Journal on Applied Dynamical Systems* 20.2 (Jan. 1, 2021), pp. 1135–1157. DOI: 10.1137/20M1387687. URL: https://epubs.siam.org/doi/abs/10.1137/20M1387687 (visited on 01/21/2022).

[6] Thomas Hotz et al. "Monitoring the Spread of COVID-19 by Estimating Reproduction Numbers over Time." Apr. 18, 2020. arXiv: 2004.08557 [q-bio, stat]. URL: http://arxiv.org/abs/2004.08557 (visited on 07/20/2020).

[7] K. Sherratt et al. *Predictive Performance of Multi-Model Ensemble Forecasts of COVID-19 across European Nations*. June 16, 2022. DOI: 10.1101/2022.06.16.22276024. URL: http://medrxiv.org/lookup/doi/10.1101/2022.06.16.22276024 (visited on 11/28/2022). preprint.

Du musst bereit sein Dinge zu tun.

— A meme on the internet, 2022.

Acknowledgments

Put your acknowledgments here.

Contents

Chapter 1

Introduction

Chapter 2

Epidemiological considerations

- COVID-19 induced unprecedented interest in epidemiological modelling from all disciplines, but also mathematics
- this chapter highlights challenges that epi modelling brings and what desirable outcomes would be from an applied perspective
- mathematical epidemiology concerns itself with modelling epidemiolocial systems, from small (local outbreaks) to large (epi/pandemics)
- conclusions from analysis only as good as the model and the data are
- dependening on goal and circumstances different methods are applicable
- by its nature, data are observational so causal claims difficult
- in this thesis I focus on models for larger-scale epidemics, techniques would be flexible enough to deal with smaller scale as well, as long as latent states are gaussian

2.1 Objectives of epidemiological modelling

Monitoring

- monitoring is real-time scenario, interested in current developments, i.e. recent past and near future. complicated by potentially slow reporting, data revisions
- informs decision makers on whether measures should be taken
- ForecastHub(s) provide platform that creates ensemble forecast to obtain better predictions [6, 7, 39, 52]

Retrospective Analysis

- evaluation of measures taken, want interpretation as causal as possible
- informs decision makes on which measures were effective and how much
- difficult due to usual reasons: poor data quality, observational data, causual structure difficult, early/late adoption makes timing of measurements difficult
- cite some papers that did this [8, 20, 30]

Scenario Modelling

- concerns itself with modelling the impact that variants, seasonality etc. have in specific scenarios
- find out whether there is already paper of ECDC to cite

2.2 Available data and its quality

- surprising amount of data available, but quality questionable,
- in Germany have data on reported cases and deaths by gender, age group, county, with reporting date of case and for some cases even date of symptom onset
- reporting of cases is regulated by Infektionsschutzgesetz
- parallel dataset for reports of hospitalisations
- have description section from Nowcasting draft here
- descriptive statistics of German COVID-19 data set
- even larger datasets that compile this for europe + EFTA (?) by ECDC or by world (JHU)
- quality of reported case data is potentially too low
  - reporting delays
  - weeakday effects
  - testing regime changing (2G/3G)
  - ...
- data on commuting

2.3 Measures of epidemic spread

This section consists of the ideas published in [23], but has been rewritten to fit better into this thesis.

- not only epidemic spread but also speed of proliferation is of interest, enables forecasts
- measuring speed difficult: data problems ... (look at AK book article)
-

2.3.1 Growth Factor

2.3.2 Reproduction number

2.3.3 Other indicators

2.3.4 Usefulness of indicators

2.4 Dessiderata for epidemiological models

- we want models to be able to include as much data as possible, while still being numerically tractable

Regional dependencies and effects

- German case data are reported on Landkreis level, performing analysis of each individual is not sensible
- inhabitants travel between regions, and measures were taken on on regional level as well
- effects are not really spatial: euclidean distance is not so much of an issue but how closely connected regions are (give some examples)
- also want to account for other regional effects such as different socio-economic settings ...

Temporal correlation

Interpretability

Chapter 3

State space models

State space models are a versatile class of statistical models which allow to model non-stationary time series data and come along with straight-forward interpretation. The main idea of these models is to introduce unobserved **latent states** whose joint distribution is given by a Markov process and model the observed time series conditional on theses states. By exploiting this structure, inference in state space models (SSMs) becomes computationally efficient, i.e. the complexity of algorithms is linear with respect to the number n of time points considered.

An additional advantage, that will become more explicit in Section 3.1, is that SSMs allow to interpret the modeled dynamics of latent states which makes

Definition 3.1 (State Space Model)**.** A **SSM** is a discrete time stochastic process $(X_t, Y_t)_{t=0,\dots,n-1}$ taking values in a measurable space $\mathcal{X} \times \mathcal{Y}$ such that

1. The marginal distribution of the **states** $(X_0, \dots, X_{n-1})$ is a discrete time Markov process, i.e. for $t = 1, \dots, n-1$

$$\mathbf{P}\left(X_t \in B | X_0, \dots, X_{t-1}\right) = \mathbf{P}\left(X_t \in B | X_{t-1}\right) \tag{3.1}$$

   for all measurable $B \subseteq \mathcal{X}$.

2. Conditional on the state X_t and observation Y_{t-1}, Y_t is independent of X_s and Y_{s-1}, $s < t$, i.e.

$$\mathbf{P}\left(Y_t \in B | X_0, \dots, X_t, Y_0, \dots, Y_{t-1}\right) = \mathbf{P}\left(Y_t \in B | X_t, Y_{t-1}\right)$$

   for all measurable $B \subseteq \mathcal{Y}$.

For notational convenience we will write $X_{s:t} = (X_s, \dots, X_t)$ for the vector that contains all states from s to t, dropping the first index if we consider the whole set of observations up to time t, so $X_{:t} = X_{0:t}$. Similarly we set $Y_{s:t} = (Y_s, \dots, Y_t)$ and $Y_{:t} = Y_{0:t}$.

picutre of dependency stru

Remark. Contrary to the standard definition of a SSM, our Definition 3.1 allows Y_t to depend on Y_{t-1}. This is not a limitation of the standard definition: given a SSM of the form in Definition 3.1 we can transform it to the standard form by choosing states $(X_t, Y_t) \in \mathcal{X} \times \mathcal{Y}$ and observations $Y_t \in \mathcal{Y}$ such that the SSM becomes a stochastic process on $(\mathcal{X} \times \mathcal{Y}) \times \mathcal{Y}$.

Additionally, most computations and inferences in this thesis will be conditioned on a single set of observations Y. As such Y may be treated as fixed and Y_t only depends on X_t. The only exception to this is in simulation studies where we sample from the joint distribution of (X, Y).

As the models considered in Chapter 4 will make extensive use of SSMs with this dependency structure we opt to use this non-standard definition here.

In most models we consider in this thesis we use $\mathcal{X} = \mathbf{R}^m$, $\mathcal{Y} = \mathbf{R}^p$ or $\mathcal{Y} = \mathbf{Z}^p$ so that $\mathcal{X}$ is m dimensional and $\mathcal{Y}$ is p dimensional and equip these spaces with the usual Borel σ-Algebras.

Most models that I consider in this thesis will admit densities for the state transitions w.r.t. a common dominating measure $\mu_{\mathcal{X}}$ and similar for the observations w.r.t. a (potentially different) domination measure $\mu_{\mathcal{Y}}$.

check whether there are m that violate this

Notation (Densities, conditional densities). I will use the standard abuse of notation for densities that makes the type of density „obvious“ from the arguments used. This means that $p(x)$ is the density for all states X, $p(x_t|x_{t-1})$ the conditional density of $X_t|X_{t-1}$ and similarly for observations: $p(y|x)$ is the density of all observations Y conditional on all states X.

Note that this notation also implicitly includes the time t and allows for changes in, e.g. , the state transition over time.

When densities stem from a parametric model parametrized by $\theta \in \Theta \subseteq \mathbf{R}^k$ and the dependence of the model on θ is of interest, i.e. because we try to estimate θ, we indicate this by adding a subscript to the densities. If the dependence is not of interest, e.g. because θ is fixed, I will usually omit θ for better readability.

In this notation, the joint density of a parametric SSM factorizes as

$$\begin{aligned} p_\theta(x, y) &= p_\theta(x_0, \dots, x_{n-1}, y_0, \dots, y_{n-1}) \\ &= p_\theta(x_0) \prod_{t=1}^{n-1} p_\theta(x_t|x_{t-1}) \prod_{t=0}^{n-1} p_\theta(y_t|x_t, y_{t-1}), \end{aligned}$$

where $p_\theta(y_0|x_0, y_{-1}) = p_\theta(y_0, x_0)$.

As inferences we make in this thesis depend on the SSM only through the likelihood we identify almost sure versions of (X, Y) with itself, i.e. all equations involving X or Y are understood almost surely.

Given data $(y_t)_{t=0,\dots,n-1}$ that may be modeled with a SSM the practitioner is confronted with several tasks, which provide the structure of this chapter:

1. Choosing a suitable, usually parametric, class of SSMs that include the effects of interest.
2. Fitting such a parametric model to the data at hand by either frequentist or Bayesian techniques.
3. Infer about the latent states X from the observations Y by determining, either analytically or through simulation, the smoothing distribution $X|Y$.

The first step, item 1, requires that the practitioner specifies a joint probability distribution for the states and observations (Section 3.1). Due to the assumed dependency structure this boils down to specifying transition kernels for the states and observations. The setting Definition 3.1 is too abstract to perform inference in so further assumptions on the types of distributions for the latent states and observations are needed. In this chapter we will discuss Gaussian linear state space model (GLSSM) (Section 3.2), where both the posterior distribution and the likelihood are analytically available. For the epidemiological application we have in mind these are however insufficient due to the non-linear behaviour of incidences and the low count per region (Section 2.4). Such observations are better modeled with distributions on the natural numbers, i.e. with a Poisson or negative binomial distribution, leading to the class of logconcave Gaussian state space models (Section 3.3).

Regarding the second step, item 2, a frequentist practitioner will want to perform maximum likelihood inference on θ. While asymptotic confidence intervals for θ can be derived both theoretically and practically [14, Chapter 7], they are, in the context of this thesis, usually of little interest. We choose to view this fitting as an Empirical Bayes procedure and our main practial interest lies in analyzing the posterior distribution $X|Y$.

To obtain the maximum likelihood estimates $\hat{\theta}$ one needs access to the likelihood

$$p(y) = \int_{\mathcal{X}^n} p(x, y)\, \mathrm{d}x, \tag{3.2}$$

which is usually not analytically available. Direct numerical evaluation of Equation (3.2) is hopeless due to the high dimensionality of the state space $\mathcal{X}^n$. Instead we will resort to simulation based inference by importance sampling (see Section 3.4), an alternative would be particle filters [12].

The performance of these simulations depends crucially on constructing distributions that are close to the posterior $p(x|y)$ but are easy to sample from. To this end, we construct suitable Gaussian state space models (Section 3.5) in which sampling from the posterior is analytically possible. This will be a good strategy if the target posterior $p(x|y)$ can be well approximated by a Gaussian distribution — otherwise, we may want to account for multiple modes by considering mixtures of Gaussian state space models or account for heavy tails with t-distributed errors (Section 3.6).

3.1 Modelling epidemiological dessiderata with state space models

3.2 Linear Gaussian state space models

- joint model is gaussian
- filtering distribution obtained by Kalman filter
- smoothing distribution obtained by Kalman smoother
- variants: sqrt filter / precision filter
- gaussian likelihood analytically available, MLE can be found by numerical methods (gradient descent or EM, depending on problem)
- computation is efficient: linear in time dimension n
- Y_{t+1} may also depend on Y_t as we will target the conditional distribution anyways

GLSSM are the working horses of most methods used in this thesis because they are analytically tractable and computationally efficient. Indeed for fixed dimension of states m and observations p the runtime of algorithms that we consider in this thesis is $\mathcal{O}(n)$.

Definition 3.2 (GLSSM)**.** A GLSSM is a state space model where states obey the transition equation

$$X_{t+1} = A_t X_t + u_t \varepsilon_{t+1} \qquad t = 0, \dots, n-1, \tag{3.3}$$

and observations obey the observation equation

$$Y_t = B_t X_t + v_t + \eta_t \qquad t = 0, \dots, n. \tag{3.4}$$

Here $A_t \in \mathbf{R}^{m \times m}$ and $B_t \in \mathbf{R}^{p \times m}$ are matrices the specify the systems dynamics. The **innovations** ε_{t+1} and **measurement noise** η_t are independent from one another and from the starting value $X_0 \sim \mathcal{N}(\mathbf{E}X_0, \Sigma_0)$.

Furthermore, $\varepsilon_{t+1} \sim \mathcal{N}(0, \Sigma_t)$ and $\eta_t \sim \mathcal{N}(0, \Omega)$ are centered Gaussian random variables and $u_{t+1}, t = 1, \dots, n-1$, $v_t, t = 0, \dots, n$ are deterministic biases.

The defining feature of a GLSSM is that its joint distribution is Gaussian.

Lemma 3.1. *A state space model can be written as a GLSSM if and only if its joint distribution is Gaussian.*

Proof. □

technicailty: distringuish b a.s. version

As the joint distribution of (X, Y) is Gaussian, so are conditional distributions of states given observations. Two such distributions are of interest: the **filtering distribution** is the conditional distribution of X_t given all observations until time t, that is $Y_{0:t}$. When $t < n$ this is distinct from the **smoothing distribution**, i.e. the distribution of X_t given all observations Y.

Note that the filtering distr does not specify a valid joi tribution for the states, bu smoothing does.

Both distributions may be obtained efficiently using the celebrated Kalman filter and smoother algorithms .

cite correclty

Notice that the Kalman filter calculates the likelihood $p(y)$ while filtering — this is possible because of the dependency structure of the state space model — this makes inference via maximum likelihood possible in GLSSMs.

To ensure numerical stability in these algorithms, the square root filter and smoother [36] may be used, see also [50] for an accessible introduction to it and other variants.

The Kalman smoother computes the marginal distributions $X_t|Y$ for $t = 0, \dots, n-1$ and, owing to the Markov structure of the states, these are enough to specify the joint distribution $X|Y$, allowing to simulate from it.

Algorithm 1: Kalman filter

Input: observations $y = (y_0, \ldots, y_n)$, GLSSM
Output: filtered expectations $\hat{X}_{t|t}$, covariance matrices $\Xi_{t|t}$, likelihood $p(y)$
Initialization;
$\hat{y}_{0|-1} = B_0 \hat{x}_{0|-1} + v_t$;
$\Psi_{0|-1} = B_0 \Sigma_0 B_0^T + \Omega_0$;
Prediction;
Filter;

Algorithm 2: Kalman smoother

Input: observations $y = (y_0, \ldots, y_n)$, GLSSM
Output: filtered expectations $\hat{X}_{t|t}$ and covariance matrices $\Xi_{t|t}$
Initialization;
$\hat{y}_{0|-1} = B_0 \hat{x}_{0|-1} + v_t$;
$\Psi_{0|-1} = B_0 \Sigma_0 B_0^T + \Omega_0$;
Prediction;
Filter;

The modeling capacity of GLSSM is, however, limited: most interesting phenomena follow neither linear dynamics nor are well modeled by a Gaussian distribution. Nevertheless, linearization of non-linear dynamics suggests that GLSSMs may have some use as approximations to these more complicated phenomena, provided they are sufficiently close to Gaussian models, e.g. unimodal and without heavy tails. We start to move away from linear Gaussian models by allowing observations that are non-Gaussian.

3.3 Logconcave Gaussian state space models

- replace gaussian observations with log concave observations
- motivation for logconcave distributions: posterior has unique mode, because up to constants $\log p(x|y) = -\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) + \log p(y|x)$ so $\log p(x|y)$ is concave
- not restricted to same type of distribution per time step (though in ISSSM it will be)
- Laplace approximation sensible for these types of models: single mode
- special case: exponential family distributions

The distribution of observations is never Gaussian - all statisticians may hope for is that the data-generating mechanism is close enough to a Gaussian distribution that inferences made carry over. For epidemiological models, Gaussian distributions are appropriate if incidences are high, e.g. during large outbreaks in a whole country. When case numbers are small, the discrete nature of incidences is better captured by a distribution on $\mathbf{N}_0$, and standard distributions used are the Poisson and negative binomial distributions, see . Both the Poisson and negative binomial belong to the class of exponential family distributions. As such, their densities have a simple structure, allowing only for a linear interaction between the natural parameter and the densities argument. We refer to [9] for a comprehensive treatment of exponential families and use their definitions throughout this chapter.

Algorithm 3: Forwards filter, backwards smoother [21, Proposition 1]

Input: TODO
Output: TODO
Something;

Definition 3.3 (exponential family)**.** Let μ be a σ-finite measure on $\mathbf{R}^p$ and denote by

$$\Theta = \left\{\theta \in \mathbf{R}^p : \int \exp\left(\theta^T x\right)\, \mathrm{d}\mu(y) < \infty\right\}$$

the set of parameters θ such that the moment-generating function of μ is finite. For every $\theta \in \Theta$

$$p_\theta(y) = Z(\theta)^{-1} \exp(\theta^T y)$$

defines a probability density with respect to the measure μ, where

$$Z(\theta) = \int \exp\left(\theta^T x\right)\, \mathrm{d}\mu(y)$$

is the normalizing constant. We call both the densities p_θ and induced probability measures

$$\mathbf{P}_\theta(A) = \int_A p_\theta(y)\, \mathrm{d}\mu(y),$$

for measurable $A \subset \mathbf{R}^p$, a **standard exponential family**.

Conversely, let $\mathbf{P}_\theta, \theta \in \Theta$ be a given parametric family of probability measures on some space $\mathcal{Y}$ that is absolutely continuous with respect to a common dominating measure μ. Suppose there exists a reparametrization $\eta : \Theta \to \mathbf{R}^p$, a statistic $T : \mathcal{Y} \to \mathbf{R}^p$ and functions $Z : \Theta \to \mathbf{R}$, $h : \mathcal{Y} \to \mathbf{R}$ exist, such that

$$p_\theta(y) = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\mu} = Z(\theta)h(y) \exp\left(\eta(\theta)^T T(y)\right),$$

then we call $\mathbf{P}_\theta, \theta \in \Theta$ and $p_\theta, \theta \in \Theta$ a ***p*-dimensional exponential family**. If $\eta(\theta) = \theta$ we call θ the canonical parameter. If $T(y) = y$, we call y the canonical observation. By reparametrization (in θ) and sufficiency (in y) every p-dimensional exponential family can be written as an equivalent standard exponential family, see the elaborations in [9, Chapter 1].

necessary?

Definition 3.4 (curved exponential family)**.**

Exponential families have the attractive property that they are log-concave in their parameters. As such the Fisher-information is always positive semidefinite, which will be crucial in defining surrogate Gaussian models in .

later section

Lemma 3.2 (log-concavity of exponential family distributions)**.** *Let $p_\theta, \theta \in \Theta$ be a natural dimensional exponential family and Θ open in $\mathbf{R}^p$. In this case $\theta \to \log p_\theta(y)$ is concave for every $y \in \mathbf{R}^p$.*

Proof. As $\log p_\theta(y) = -\log Z(\theta) + \theta^T y$ it suffices to show that $\log Z(\theta)$ is convex. However, $\log Z(\theta)$ is the cumulant generating function of the base measure μ which is known to be convex . □

more reasoning? differentia integral, check if dominate gence applies, or look for a

We now generalize definition 3.2 to allow for non-gaussian observations by replacing the observation equation eq. (3.4) by more general exponential families.

Definition 3.5 (Logconcave state space model (LCSSM))**.** A **LCSSM** is a state space model where states obey the transition equation

$$X_{t+1} = A_t X_t + u_t \varepsilon_{t+1}$$

and the conditional distribution of Y_t given X_t comes from an exponential family with respect to a base measure μ_t, i.e.

$$p(y_t|x_t) = h_t(y_t) Z_t(x_t) \exp\left(\eta_t(x_t)^T T_t(y_t)\right)$$

for suitable functions h_t, Z_t, η_t, T_t.

LCSSM only logconcave ob tions would suffice, then E an instance of this

Remark. To simplify notation we will drop in our notation the dependence of h, Z, and T on t and assume that the base measure μ_t is the same for all relevant t.

As in the previous chapter, after having observed Y, one is interested in the conditional distribution of states X, given Y. If the observations are not Gaussian, this is a difficult task as the distribution is not analytically tractable. Instead approximations, e.g. the Laplace approximation (LA) (see **??**, or simulation based inference, e.g. importance sampling or MCMC-methods, are used. Similarly, fitting hyperparameters θ by maximum likelihood inference becomes more difficult as evaluating $\ell(\theta) = p(y) = \int p(x, y)\,\mathrm{d}x$ is not analyically avaiable, thus requiring numerical or simulation methods for evaluation and gradient descent or EM-techniques for optimization (.

really EM?

s to the LA section

In this thesis we will use importance sampling methods, which are the focus of the next section.

Under appropriate assumptions on the observation distributions, the LA is given by a (possibly degenerate) GLSSM [15, 51], though the derivations and computations are more involved in the nonlinear case or if the observation densities are not log-concave.

Theorem 3.1 (LA of LCSSM)**.**

graph to modelling chap-

The Poisson distribution arises from the law of small numbers: if there is a large population where every individual has, independently, a small probability of becoming infected in a small window of time then the total number of infections in that window of time is well approximated by the Poisson distribution. Indeed, the law of small numbers remains valid for small dependencies [4, 45]. However, incidences observed from the SARS-CoV-2 epidemic tend to follow a negative binomial distribution [10].

3.4 Importance Sampling

ro

Importance sampling is a simulation technique that allows to approximate expectations by sampling from a tractable approximation, the proposal, to the measure of interest, the target, and weighting samples according to their importance. As the user has freedom in the choice of approximation (except for some technical conditions), importance sampling also acts as a variance reduction technique with better approximations resulting in smaller Monte-Carlo variance. Thus the role that importance sampling plays is twofold: first it allows to perform Monte-Carlo integration even if sampling from the target is not possible, and second it allows to do so in an efficient way by choosing, to be defined precicely below, the approximation in an optimal way.

Alternative approaches to importance sampling for performing inference in SSMs include Markov chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC). Recall from Chapter that this inference concerns three objectives: maximum likelihood estimation, i.e. evaluation and optimization of the likelihood, the posterior distribution $X_{:n}|Y_{:n}$ and prediction of future states and observations. Let us give a concise comparison of these alternative approaches, weighing their advantages and disadvantages over importance sampling, in particular for the SSMs that this thesis deals with.

k handbook

MCMC is a simulation technique that allows to generate correlated samples from a distribution by constructing a Markov chain that has as its invariant distribution the desired distribution. In the most general method, Metropolis Hastings MCMC, one needs access to the density of the sought distribution up to a constant to simulate a step in the Markov chain. While this method is very general, it fails in high dimensions and a lot of current research in MCMC methods deals with this curse of dimensionality .

ething

ntages MCMC over IS

SMC [12], sometimes called a particle filter, uses sequential importance sampling to provide a particle approximation to the filtering distributions $X_t|Y_{:t}$, essentially decomposing the problem into a n importance sampling steps. To avoid particle collapse, SMC is usually equipped with a resampling step once the effective sample size of the current set of particles drops below a specified level. Once the final filtering distribution $X_n|Y_{:n}$ is approximated, the smoothing distribution may be obtained in several ways ... .

Chopin

Conveniently, SMC allows us to approximate the likelihood $\ell(\theta)$ for a single parameter by a single pass of the particle filter. However, the discrete nature of resampling makes the approximated likelihood non-continuous, complicating maximum likelihood inference. [12, Chapter 14.3] discusses several strategies: the first amounts to importance sampling of the order as discussed in this thesis, where one fixes a reference parameter θ_0 to perform importance sampling with $p_{\theta_0}(x|y)$ against $p_\theta(x|y)$. The second strategy only works in the univariate case and consists of approximating the non-continuous inverse CDFs appearing in the resampling step by continuous ones. Under j stochastic gradient ascent

This chapter proceeds with a general treatment of importance sampling. Subsequently, we will focus our attention on LCSSMs and how we can exploit their structure to perform importance sampling efficiently.

Suppose we have a function $h : \mathcal{X} \to \mathbf{R}$ whose integral

$$\zeta = \int_{\mathcal{X}} h(x)\,\mathrm{d}x$$

we want to compute. Furthermore, suppose that we can write

$$\int_{\mathcal{X}} h(x)\,\mathrm{d}x = \int_{\mathcal{X}} f(x)\,\mathrm{d}\mathbf{P}(x) = \mathbf{P}(f)$$

for a probability measure $\mathbf{P}$ and function $f : \mathcal{X} \to \mathbf{R}$. Let $\mathbf{G}$ be another measure on $\mathcal{X}$ such that $f\mathbf{P}$ is absolutely continuous with respect to $\mathbf{G}$, $f\mathbf{P} \ll \mathbf{G}$, and let $v = \frac{\mathrm{d}f\mathbf{P}}{\mathrm{d}\mathbf{G}}$ be the corresponding Radon-Nikodym derivative. Then

$$\zeta = \int_{\mathcal{X}} h(x)\,\mathrm{d}x = \int_{\mathcal{X}} f(x)\,\mathrm{d}\mathbf{P}(x) = \int_{\mathcal{X}} v(x)\,\mathrm{d}\mathbf{G}(x)$$

which suggests to estimate ζ by Monte-Carlo integration:

$$\hat{\zeta} = \frac{1}{N}\sum_{i=1}^{N} v(X_i)$$

for $X_i \overset{\text{i.i.id}}{\sim} \mathbf{G}$, $i = 1, \ldots, N$. Here we call $\hat{\zeta}$ the importance sampling estimate of ζ.

A classical result is that the minimum MSE proposal $\mathbf{G}^*$ has a closed form, which can be shown by a simple application of Jensen's inequality.

Proposition 3.1 ([12, Proposition 8.2]). *[minimum MSE proposal] The proposal $\mathbf{G}^*$ that minimizes the MSE of importance sampling is given by*

$$\mathbf{G}^* = \frac{|f|}{\mathbf{P}(|f|)}\mathbf{P}$$

Unfortunately, this optimality result has no practical use, indeed if f is positive we'd have to obtain $\mathbf{P}(f)$, the overall target of our endeavor. Additionally, sampling from $\mathbf{G}^*$ is not guaranteed to be practically feasible.

If one is not interested in a particular h but rather in an approximation of $\mathbf{P}$, and $\mathbf{P}$ is absolutely continuous with respect to $\mathbf{G}$, then one may view

$$\hat{\mathbf{P}}_N = \frac{1}{N}\sum_{i=1}^{N} v(X_i)\delta_{X_i}$$

as a particle approximation of $\mathbf{P}$, in the sense that for sufficiently well behaved test functions f, $\mathbf{P}(f) \approx \hat{\mathbf{P}}_N(f)$. In this setting [1] shows that the random measure $\hat{\mathbf{P}}_N$ converges to $\mathbf{P}$ at rate $\mathcal{O}\left(\frac{1}{N}\right)$ in an appropriate metric. check

To perform importance sampling one must be able to evaluate the weights v. In a Bayesian setting this is usually infeasible: if $\mathbf{P}$ is a posterior distribution then the integration constant of its density is intractable. In this case one can usually evaluate the weights up to a constant, i.e. $w(x) \propto_x \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{G}}(x)$ is available. The missing constant is then $\int w(x)\,\mathrm{d}\mathbf{G}$ which is itself amenable to importance sampling. This leads to the self-normalized importance sampling weights $W_i = \frac{w(X_i)}{\sum_{i=1}^{N} w(X_i)}$ and Monte Carlo estimates $\hat{\zeta} = \sum_{i=1}^{N} W_i f(X_i)$ and particle approximation $\hat{\mathbf{P}}_N = \sum_{i=1}^{N} W_i \delta_{X_i}$.

In both cases one can show that once second moments of w with respect to $\mathbf{G}$ exist the Monte-Carlo estimates are consistent and asymptotically normal at the usual rates, see [12, Chapter 8]. However, the finite sample variance of $\hat{\zeta}$, and thus the practical performance of the procedure, depends on the variance of $w \cdot f$ under $\mathbf{G}$, and thus on the proposal $\mathbf{G}$. [1] show that the expected mean squared TV distance between $\hat{\mathbf{P}}_N$ and $\mathbf{P}$ may be bounded, up to a constant, by the second check if true

moments of w . In addition, they provide bounds that involve the KL-divergence

more extensive discuss

$$\mathcal{D}_{\mathrm{KL}}(\mathbf{P}||\mathbf{G}) = \int \log \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{G}} \,\mathrm{d}\mathbf{P},$$

fostering the intuition that $\mathbf{G}$ should be close to $\mathbf{P}$ for the particle approximation $\hat{\mathbf{P}}_N$ to be close to $\mathbf{P}$.

erenceable defs in this

To judge the convergence of importance sampling several criteria are discussed in the literature. The classic effective sample size (ESS)[31]

$$\mathrm{ESS} = \frac{1}{\sum_{i=1}^{N} W_i^2}$$

arises from the asymptotic efficiency of importance sampling estimates and is easy to interpret, though care has to be taken in particular settings . Assessing convergence through the variance of $\hat{\mathbf{P}}_n$ is, while natural, flawed [11] and should be avoided. As a remedy [11] suggest the heuristic $q_N = \mathbf{E}Q_N$ where

ature

$$Q_N = \max_{1 \le i \le N} W_i.$$

This judges whether importance sampling has collapsed to just a few particles and is amenable to Monte-Carlo integration.

roblems/remedies of high nal importance sampling

In the context of this thesis, importance sampling serves as the main tool to facilitate inference in SSMs, both for maximum likelihood estimation and access to the smoothing distribution. As the dimension of the state space is $n \cdot m^2$, which will usually be large, we have to perform importance sampling efficiently, exploiting the dependence structure.

here to put this

As the likelihood of a general state space model is neither analytically nor numerically tractable one has to resort to Monte-Carlo techniques. Recall that the likelihood is a high-dimensional integral of the form

$$\ell(\theta) = p_\theta(y) = \int p_\theta(y, x) \,\mathrm{d}x = \int p_\theta(y|x) p_\theta(x) \,\mathrm{d}x = \mathbf{E} p_\theta(y|X).$$

By the standard law of large numbers we can approximate $\ell(\theta)$ by

$$\hat{\ell}(\theta) = \frac{1}{N} \sum_{i=1}^{N} p_\theta(y|X^i)$$

for $N \in \mathbf{N}$ samples $X^i \overset{\text{i.i.id}}{\sim} p(x)$. However, the variance of $\hat{\ell}(\theta)$ is likely to be very high if samples X^i are drawn from the prior distribution $p(x)$ as they are not informed by the observations y. As $p_\theta(x|y) \propto p_\theta(x, y)$ a more promising approach would be to use samples $X^i \sim p_\theta(x|y)$, but this distribution is usually not available.

While bayesian computational approaches such as MCMC are able to generate (approximate) samples from this posterior distribution, importance sampling tries to find a distribution close to the target and re-weighs samples to ensure unbiased estimates of $\ell(\theta)$.

- importance sampling as a variance reduction technique
- importance sampling as a technique to make intractable distributions tractable
- importance sampling vs. other methods:
  - vs. ABC
  - vs. MCMC
  - vs. INLA (isn't this MCMC?)
- measuring how good IS performs: ESS and other measures
- related results regarding performance of IS (Chatterje, Agapiou)

3.4.1 Laplace approximation (LA)

Laplace approximation (LA) goes back to Laplace [32] who invented the technique to approximate moments of otherwise intractable distributions. Since [53, 54] rediscovered its use to approximate posterior means and variances, it has been a staple method for statisticians .

cite something

The method is based on a second-order Taylor series expansion of the log target density $\log p(x)$ around its mode $\hat{x}$, i.e. matching mode and curvature. Assuming the density is sufficiently smooth and the mode exists and is unique, we have

$$\log p(x) \approx \log p(\hat{x}) + \underbrace{\nabla_x \log p(\hat{x})}_{=0}(x - \hat{x}) + \frac{1}{2}(x - \hat{x})^T H (x - \hat{x})$$

where H is the Hessian of $\log p$ evaluated at the mode. As $\log p(\hat{x})$ does not depend on x, the right-hand side can be seen (up to additive constants) as the density of a Gaussian distribution with mean $\hat{x}$ and covariance matrix $\Sigma = -H^{-1}$.

If p is log-concave in x, H is guaranteed to be negative semidefinite and the LA yields an actual Gaussian distribution.

what to do if we cannot gu this

The main advantage of the LA is that fast to obtain and, for sufficiently well-behaved distributions on a moderate dimensional space, provides reasonably high ESS. Additionally, the Newton-Raphson iterations to find the mode and Hessian are robust and require no simulation, unlike the other methods discussed further below. For the SSMs we consider in this thesis, the numerical methods can be implemented using the Kalman filter and smoother [15, 51], even in the degenerate case where H is indefinite [27].

incorporate some of the cit from How good is your LA

However, as the LA is a local approximation, it may be an inappropriate description of the global behavior of the target, see example 3.2 for a breakdown of LA. Additionally, even if LA works in principle, its ESS will usually degenerate quickly once the dimension increases whereas the cross-entropy method (CE-method) and efficient importance sampling (EIS) do so at a slower pace.

- approximate at mode, problematic if posterior is not unimodal (but then gaussian approximation probably not worth it)
- can be solved by KF, even if non log-concave ()

think about whether to ge to non LC, problem might sqrt filter does not work

3.4.2 CE-method

To provide a global approximation to the target, the CE-method[46, 47] selects from a family $(\mathbf{G}_\psi)_{\psi\in\Psi}$ of proposals the one that minimizes the Kullback Leibler divergence (KL-divergence) to the target. Thus, the CE-method finds ψ_{CE} which solves the following optimization problem

$$\begin{aligned}\psi_{\text{CE}} &= \text{argmin}_{\psi\in\Psi}\, \mathcal{D}_{\text{KL}}\left(\mathbf{G}^* || \mathbf{G}_\psi\right) \\ &= \text{argmin}_{\psi\in\Psi} \int \log \frac{\mathrm{d}\mathbf{G}^*}{\mathrm{d}\mathbf{G}_\psi}\, \mathrm{d}\mathbf{G}^*\end{aligned}$$

where $\mathbf{G}^*$ is either the optimal proposal from Proposition 3.1 or $\mathbf{P}$. If $\mathbf{G}^*$ and $\mathbf{G}_\psi$ possess densities g^* and g_ψ w.r.t. some common measure μ, the same for all ψ, we may reformulate the optimization problem to maximize the cross-entropy between g^* and g_ψ

$$\begin{aligned}\psi_{\text{CE}} &= \text{argmin}_{\psi\in\Psi} \int g^*(x) \log g^*(x)\, \mathrm{d}\mu(x) - \int g^*(x) \log g_\psi\, \mathrm{d}\mu(x) \\ &= \text{argmax}_{\psi\in\Psi} \int g^*(x) \log g_\psi(x)\, \mathrm{d}\mu(x), \end{aligned} \tag{3.5}$$

as the first integral does not depend on ψ.

As $\mathbf{G}^*$ is usually intractable, so is ψ_{CE}. However, the integral eq. (3.5) is amenable to importance sampling: Given a proposal $\mathbf{G}$, we may estimate it by

$$\hat{\psi}_{\text{CE}} = \text{argmax}_{\psi\in\Psi} \frac{1}{N} \sum_{i=1}^{N} \log g_\psi(X_i) \tag{3.6}$$

where $X_1, \dots, X_N \overset{\text{i.i.id}}{\sim} \mathbf{G}$.

An attractive property of the CE-method is that if $\mathbf{G}_\psi$ form an exponential family with natural parameter $\psi \in \mathbf{R}^p$, the optimal ψ_{CE} only depends on certain moments of $\mathbf{G}^*$. Indeed, for $\log g_\psi(x) = \log h(x) + \log Z(\psi) + \psi^T T(x)$ we have

$$\int g^*(x) \log g_\psi(x) \, \mathrm{d}\mu(x) = \mathbf{E} \log h(X) + \log Z(\psi) + \psi^T \mathbf{E} T(X)$$

for $X \sim \mathbf{G}^*$. As $\log Z(\psi)$ is the cumulant-generating function of $\mathbf{G}_\psi$it is, under appropriate regularity conditions, smooth. Thus the optimal ψ_{CE} solves

$$\frac{\nabla_\psi Z(\psi_{\text{CE}})}{Z(\psi_{\text{CE}})} = -\mathbf{E} T(X).$$

Given $\mathbf{E}T(X)$, this system of equations can, in many cases, be solved analytically or by gradient descent algorithms.

While $\mathbf{E}T(X)$ is usually not available, it is itself amenable to importance sampling. Given a proposal $\mathbf{G}$ we may estimate $\mathbf{E}T(X)$ by $\hat{\mathbf{G}}_N T = \sum_{i=1}^N W^i T(X^i)$ for $X_1, \dots, X_N \overset{\text{i.i.id}}{\sim} \mathbf{G}$ and auto-normalized importance sampling weights W^i and in turn estimate ψ_{CE} by $\hat{\psi}_{\text{CE}}$ solving

$$\frac{\nabla_\psi Z(\hat{\psi}_{\text{CE}})}{Z(\hat{\psi}_{\text{CE}})} = -\frac{\hat{G}_N T}{\hat{G}_N w}.$$

Thus $\hat{\psi}_{\text{CE}}$ is a Z-estimator, and we can analyze its asymptotic behavior using standard results.

Theorem 3.2 (consistency of $\hat{\psi}_{\text{CE}}$)**.**

his with non-natural pa-

y conditions? + consis-

Theorem 3.3 (asymptotic normality of $\hat{\psi}_{\text{CE}}$)**.** *Suppose that $\nabla_\psi \log Z$ is locally Lipschitz around ψ_{CE}, T is square integrable w.r.t. $\mathbf{G}$ and the Fisher information $I(\psi_{CE})$ is positive definite. Then*

$$\sqrt{N}\left(\hat{\psi}_{CE} - \psi_{CE}\right) \to \mathcal{N}(0, V)$$

where

$$V = -\frac{1}{(\mathbf{G}w)^2} I(\psi_{CE})^{-1} Cov_{\mathbf{G}}\left(w(T - \nabla_{\psi_{CE}} \log Z(\psi_{CE}))\right) I(\psi_{CE})^{-1}.$$

Moreover $\mathbf{G}(w(T - \nabla_\psi \log Z(\psi_{CE}))) = 0$, so we may estimate V consistently by

$$\hat{V} = I(\hat{\psi}_{CE})^{-1} \frac{\sum_{i=1}^N w_i^2 \left(T(X^i) - \nabla_\psi \log Z(\hat{\psi}_{CE})\right)\left(T(X^i) - \nabla_\psi \log Z(\hat{\psi}_{CE})\right)^T}{\left(\sum_{i=1}^N w_i\right)^2} I(\hat{\psi}_{CE})^{-1}.$$

Proof. We check that the conditions of the central limit theorem for Z-estimators [56, Theorem 5.21] are fulfilled. Let $\dot{z}(\psi) = \nabla_\psi \log Z(\psi)$ and consider the estimating equations for ψ_{CE}

$$x \mapsto f_\psi(x) = \nabla_\psi \left(w(x) \log g_\psi(x)\right) = w(x)T(x) - w(x)\dot{z}(\psi).$$

As $\|f_{\psi_1}(x) - f_{\psi_2}(x)\| = w(x) \|\dot{z}(\psi_1) - \dot{z}(\psi_2)\|$ for all $\psi_1, \psi_2 \in \Psi$ and $\mathbf{G}w^2(x) < \infty$ the local Lipschitz condition required by [56, Theorem 5.21] is fulfilled.

riangle equation (squared

Furthermore, as wT is square integrable w.r.t. $\mathbf{G}$, we have

$$\mathbf{G}\|f_\psi\|^2 \le \mathbf{G}w^2 \|\dot{z}(\psi)\|^2 + 2\|\dot{z}(\psi)\|\mathbf{G}\|wT\| + \mathbf{G}\|wT\|^2 < \infty.$$

Additionally $\psi \mapsto \mathbf{G}f_\psi = (\mathbf{G}w)\dot{z}(\psi) + \mathbf{G}wT$ is differentiable everywhere, with Jacobian $(\mathbf{G}w)\ddot{z}(\psi)$, where $\ddot{z}(\psi) = \partial_\psi \dot{z}(\psi)$ is the Hessian of the cumulant generating function.

Under the proposed regularity conditions we see that

$$\mathbf{G}f_{\psi_{\text{CE}}} = \mathbf{P}\left(\dot{z}(\psi_{\text{CE}}) + T\right) = \dot{z}(\psi_{\text{CE}}) + \mathbf{P}T = 0,$$

by definition of ψ_{CE}, so $\text{Cov}_{\mathbf{G}}\left(w(T - \nabla_{\psi_{\text{CE}}} \log Z(\psi_{\text{CE}}))\right) \text{Cov}_{\mathbf{G}}(f_{\psi_{\text{CE}}}) = \mathbf{G} f_{\psi_{\text{CE}}} f_{\psi_{\text{CE}}}^T$. By [56, Theorem 5.21] the asymptotic covariance matrix is

$$V(\psi_{\text{CE}}) = \frac{\ddot{z}(\psi_{\text{CE}})^{-1}}{\mathbf{G}w} \left(\mathbf{G} f_{\psi_{\text{CE}}} f_{\psi_{\text{CE}}}^T\right) \frac{\ddot{z}(\psi_{\text{CE}}{}^{-1})}{\mathbf{G}w},$$

which finishes the proof. □

Example 3.1 (univariate Gaussian)**.** ...

The form of the asymptotic covariance matrix is that of the sandwich estimator , corrected for the importance sampling with $\mathbf{G}$. This is not surprising: the CE-method performs maximum likelihood estimation where the data X_i come from the misspecified $\mathbf{P}$. Additionally, we have to correct the variance for performing importance sampling with $\mathbf{G}$, instead of sampling directly from $\mathbf{P}$.

If $\mathbf{G}_\psi$ do not form an exponential family, $\hat{\psi}_{\text{CE}}$ will still be consistent and asymptotically normal, provided the usual regularity conditions for M-estimators apply.

The CE-method is routinely used for estimating failure probabilities for rare events [25] and has been applied to Bayesian inference [16, 17] and optimal control problems [29, 58].

cite Huber

sub or superskript

cite more

3.4.3 Efficient importance sampling

EIS[41] provides an alternative to the CE-method. Instead of minimizing the KL-divergence between the target $\mathbf{G}^*$ and $\mathbf{G}_\psi$, EIS aims at minimizing the variance of the logarithm of importance sampling weights. Thus, EIS finds ψ_{EIS} which solves

$$\begin{aligned}\psi_{EIS} &= \text{argmin}_{\psi\in\Psi} \text{Var}_{\mathbf{P}} (\log w_\psi) \\ &= \text{argmin}_{\psi\in\Psi} \mathbf{P}(\log w_\psi - \mathbf{P} \log w_\psi)^2,\end{aligned}$$

where $\log w_\psi = \log p - \log g_\psi$. As $\mathbf{P} \log w$ is usually intractable as well, we include it in the optimization problem, utilizing the fact that the mean is the minimizer of the squared distance functional. Here the unnormalized weights $w \propto \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{G}}$ may be used, as the unknown integration constant gets absorbed by the unknown mean. In total, EIS solves

$$\psi_{\text{EIS}}, \lambda_{\text{EIS}} = \text{argmin}_{\psi\in\Psi,\lambda\in\mathbf{R}} \mathbf{P} \left(\log p - \log g_\psi - \lambda\right)^2.$$

Similar to the CE-method this problem is analytically intractable, and so we resort to importance sampling with a proposal $\mathbf{G}$, estimating ψ_{EIS} by

$$\hat{\psi}_{\text{EIS}} = \text{argmin}_\psi \sum_{i=1}^N W^i \left(\log p(X^i) - \log g_\psi(X^i) - \lambda\right)^2,$$

where $X^1, \ldots, X^N \overset{\text{i.i.id}}{\sim} \mathbf{G}$. This weighted least squares problem has many desirable properties: estimation is quick and numerically stable.

Under the usual regularity conditions allowing to differentiate under the integral, the estimating equations for ψ_{EIS} read

$$\begin{aligned}-2\mathbf{P}\left(\left(\log p - \log g_\psi - \lambda\right) \nabla_\psi \log g_\psi\right) &= 0 \\ -2\mathbf{P}\left(\log p - \log g_\psi - \lambda\right) &= 0\end{aligned}$$

3.4.4 Comparison

- discuss variances from CLT
- discuss targets
- numerical stability
- time complexity

Example 3.2 (Failure of LA)**.** Consider the Gaussian scale mixture $\mathbf{P} = \frac{1}{2}\left(\mathcal{N}(0,1) + \mathcal{N}(0,\varepsilon^{-2})\right)$ with mode $x^* = 0$. The LA is $\mathbf{G}_{\text{LA}} = \mathcal{N}\left(0, \frac{1}{\varepsilon^2 - \varepsilon + 1}\right)$, whose variance goes to 1 as ε goes to 0, so the LA will miss close to $\frac{1}{2}$ of the total mass. For ε small enough, the variance of the LA will be smaller than $\frac{1}{2\varepsilon^2}$, whence the second moment of the weights is infinite and importance sampling fails.

The CE-method minimizes the KL-divergence between $\mathbf{P}$ and $\mathbf{G}_\psi$, is given by $\mathbf{G}_{\text{CE}} = \mathcal{N}(0,\sigma^2)$, where $\sigma^2 = \frac{1}{2}\left(1+\varepsilon^{-2}\right)$ is the variance of $\mathbf{P}$. As $\sigma^2 > \frac{1}{2}\varepsilon^{-2}$, the weights have finite second moment, and importance sampling is consistent.

f for $\frac{1}{2}$ to appendix

3.5 Gaussian importance sampling for state space models

Most models in this thesis can be viewed as an inverse problem of the form

$$\begin{aligned} \mathbf{R}^{n\cdot m} \ni X &\sim \mathcal{N}(\mu, \Sigma) \\ Y|X \sim Y|BX &\sim p(y|s) \end{aligned}$$

and the state space formulation allows for efficient computation of, e.g., $p(y|x)$. To perform importance sampling for the smoothing distribution $p(x|y)$ we want to have close tractable approximations, that also depend on few parameters, ideally only $\mathcal{O}(n)$ many.

h about laplace approxi-
or the posterior

In total we want to perform importance sampling with proposal distributions $g(x|z)$ given by Gaussian linear models of the form

$$\begin{aligned} X &\sim \mathcal{N}(\mu, \Sigma) \\ Z &= BX + \eta \\ \eta &\sim \mathcal{N}(0, \Omega). \end{aligned}$$

The dependency structure of the state space model implies that Ω should be a blockdiagonal matrix with at most $n \cdot m^2$ many non-zero entries. If, additionally, the observations y_t are conditionally independent given x_t, i.e. if $p(y_t|s_t) = \prod_{i=1}^{p} p(y_t^i|s_t^i)$, then Ω is a diagonal matrix with only $\mathcal{O}(n \cdot m)$ many non-zero entries.

The proposal distribution $g(x|z)$ is then parameterized by the synthetic observations z and the entries of Ω and we denote this set of parameters by $\psi = (z, \Omega)$. The following results on this distribution will be useful when analysing Gaussian importance sampling.

other subsection

The Laplace approximation chooses ψ_{LA} such that the mode of $g(x|z)$ and the curvature at the mode match that of the true posterior, while the CE-method and EIS choose ψ_{CE} and ψ_{EIS} the solutions to associated optimization problems.

This means that we can treat all three methods in the same framework, facilitating comparison between the resulting three importance sampling proposals.

3.5.1 Gaussian smoothing proposals

In this section, we analyze the properties of Gaussian proposals for importance sampling in SSMs that exploit the available Markov property of states. As mentioned in the introduction to this section, these proposals are conditional distributions $\mathbf{P}^{X|Z=z}$ where $\mathbf{R}^m \ni X \sim \mathcal{N}(\mu, \Sigma)$ and $Z = BX + \eta \in \mathbf{R}^p$ where $\eta \sim \mathcal{N}(0, \Omega)$ is independent of X. Standard results from linear regression theory imply that the conditional distribution in question is again a Gaussian distribution, $X|Z = z \sim \mathcal{N}(\bar{\mu}, \bar{\Sigma})$ with mean

$$\begin{aligned} \bar{\mu} &= \mu + \Sigma B^T \left(B\Sigma B^T + \Omega\right)^{-1}(z - B\mu), && (3.7) \\ &= \bar{\Sigma}\left(\Sigma^{-1}\mu + B^T\Omega^{-1}z\right) && (3.8) \end{aligned}$$

and covariance matrix

$$\begin{aligned} \bar{\Sigma} &= \Sigma - \Sigma B^T \left(B\Sigma B^T + \Omega\right)^{-1} B\Sigma && (3.9) \\ &= \left(\Sigma^{-1} + B^T\Omega^{-1}B\right)^{-1}. && (3.10) \end{aligned}$$

Note that Equations (3.7) and (3.9) are more general, requiring only $B\Sigma B + \Omega$ be invertible, while the others require both Σ and Ω to be invertible, see [12, Lemma 7.1] for further discussion.

Proposition 3.2 (Exponential family of smoothing distribution)**.** *Suppose Ω is invertible. In this case the family of conditional distributions $X|Z = z$ parameterized by z and Ω form an exponential family*

$$p(x|z) = h(x) \exp\left(\langle \eta, T(x) \rangle - A(\eta)\right)$$

where the parameters are

$$\eta = (\eta_1, \eta_2) = \left(\bar{\Sigma}^{-1}\bar{\mu}, -\frac{1}{2}\Omega^{-1}\right)$$

and

$$\begin{aligned} h(x) &= \frac{1}{\sqrt{(2\pi)^m \det \Sigma}} \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x\right) \\ A(\eta) &= \frac{1}{2}\left(\log\det\left(I - \Sigma\ \mathrm{diag}\left(2\eta_2\right)\right) + \bar{\mu}^T\bar{\Sigma}^{-1}\bar{\mu}\right) \\ &= \frac{1}{2}\log\det\left(I - \Sigma\ \mathrm{diag}\left(2\eta_2\right)\right) + \frac{1}{2}\eta_1^T\left(\Sigma^{-1} - \mathrm{diag}(2\eta_2)\right)\eta_1 \\ T(x) &= \left(x, xx^T\right). \end{aligned}$$

Note that $\eta_1 = \Sigma^{-1}\mu + B^T\Omega^{-1}z \in \Sigma^{-1}\mu + \operatorname{im} B^T$ making the exponential family curved if $\operatorname{rank} B < m$.

fix Ω and diag ω here

probably cite something ab curved exponential familie Brown1986Fundamentals

3.5.2 Analysis of optimal parameters

Theorem 3.4 (Optimal EIS proposal)**.** *Let $p(x)$ be some density and consider importance sampling by exponential family proposals with densities*

$$q_\psi(x) = h(x) \exp\left(\langle \psi, S(x) \rangle - A(\psi)\right)$$

with natural parameter $\psi \in \mathbf{R}^k$, base measure h, sufficient statistic S and log-partition function A. The parameter $\hat{\psi}$ that minimizes the variance of log importance sampling weights $\log w_\psi(x) = \log p(x) - \log q_\psi(x)$ is given by

$$\begin{aligned} \hat{\psi} &= \operatorname{argmin}_\psi \mathit{Var}\left(\log w_\psi(X)\right) \\ &= \mathit{Cov}(S(X))^{-1}\,\mathit{Cov}\left(S(X), \log\frac{p(X)}{h(X)}\right) \end{aligned}$$

where $X \sim p$.

Proof. □

formultae this, consider ex sumptions

Remark (Optimal Gaussian proposal). As the family of Gaussian distributions $\mathcal{N}(\mu, \Sigma)$ form an exponential family with natural parameter $\psi = \left(\Sigma^{-1}\mu, -\frac{1}{2}\Sigma^{-1}\right)$ and sufficient statistic $S(x) = \left(x, xx^T\right)$, Theorem 3.4 implies that the optimal EIS Gaussian proposal involves up to fourth order moments of p.

As a consequence we expect EIS to produce proposals that are more robust to skewness and heavier than Gaussian tails than the Laplace approximation .

which is validated by simul in section ...

3.5.3 Analysis of convergence (?)

Additionally, each iteration of the CE and EIS method may be seen as performing M-estimation and as such the one step estimates ψ_{CE} and ψ_{EIS} are, in the limit as the number of samples M goes to ∞, asymptotically normally distributed.

Analyzing the multi-step behavior of these iterative estimates is more complex, as we want to keep a fixed seed, i.e. common random numbers, to ensure numerical convergence. Thus the distribution of the second iterate conditional on the first iterate depends only the conditional distribution of the common random numbers given the first iterate, which is intractable.

check

n der vaart

Theorem 3.5 (Consistency of importance sampling estimates)**.**

asymptotic covariances

Theorem 3.6 (Asymptotic normality of importance sampling estimates)**.**

ive procedures are M-rs, so a single step is (in of samples $N \to \infty$), me regularity conditions, tically normal, compare tic variances

Proof. □

this in a sensible way, EIS more numerically

3.6 Accouting for multimodality and heavy tails

Performing importance sampling with the Gaussian models discussed so far will work well only if the smoothing distribution $p(x|y)$ is well approximated by a Gaussian distribution. However, a Gaussian distribution is a very specific kind of distribution, in particular, it is an unimodal distribution and has light tails .

correct wording

If the smoothing distribution violates any of these assumptions, importance sampling with the models presented so far is likely to fail, i.e. requiring large sample sizes for both finding the optimal importance sampling parameter $\hat{\psi}$ as well as the final importance sampling evaluation.

There are however techniques to keep most of the computational efficiency discussed in the above sections to address both multimodality as well as heavy tails.

We start with heavier than gaussian tails: the textbook example of a heavy tailed distribution is the multivariate t-distribution with density

....

for degrees of freedom $\nu > 1$, location μ and scale matrix Σ. When $\nu > 2$ then this distribution has mean μ and if $\nu > 3$ it has covariance matrix ? .

The main properties necessary to facilitate Gaussian importance sampling strategies above are that the distribution $p(x|y)$ is analytically tractable and simulation from it is possible. These properties still hold for the multivariate t-distribution and, in fact, for the even larger class of elliptical distributions:

correct book

Theorem 3.7 (Conditional distribution of elliptical distributions)**.**

As one can readily see from Theorem 3.7 the parameters of the smoothing distribution $p(x|y)$ if $p(x, y)$ follows an elliptical distribution is again elliptical and its parameters only depend on quantities that are computed by the Kalman smoother.

some models with heavy

3.7 Maximum likelihood estimation in SSMs

3.8 Simulation studies

Chapter 4

Analysis of selected models

4.1 Spatial reproduction number model

1. essentially the Regional model presented in ECMI

4.2 Regional growth factor model

4.3 Nowcasting hospitalizations

4.3.1 Context

Judging the severity of the COVID-19 epidemic has been an ongoing challenge since its inception. As immunization against COVID-19 rose, strict enforcement of social distancing rules eased and testing regimes became less strict, case incidences became a less reliable and harder to interpret indicator of epidemic severity. Instead more direct indicators of morbidity, such as the number of deaths and ICU admissions and occupancy have come to the fore. But these indicators are late due to the substantial delays between infection and occurence. An alternative indicator that captures the morbidity caused by COVID-19 but is earlier than the others is the number of hospitalisations of positive COVID-19 cases.

While hospitalisations occur earlier, they still come with substantial delay between the infection and subsequent admission to hospital. Additional difficulties arise due to delays in reporting, i.e. the time it takes until the hospital reports the new case to the national health authorities. The problem of accounting for delays in reporting for occurred, but not yet reported events has been termed **nowcasting**, i.e. forecasting of the indicator at time "now". Predicting the number of hospitalisations is thus a mixture of both forecasting — which reported COVID-19 cases will end up in the hospital — and nowcasting — which cases have yet to be reported — and we will use the term nowcasting in this paper to mean this predictive mixture. In this section we focus on the situation in Germany where data on hospitalisations has been available since April 2021 provided by the German federal health care authorithy, the Robert Koch-Institut (RKI), via Github [42]. In these data the number of hospitalisations is linked to the date of reporting of the associated case, so the term of nowcasting is accurate: we are interested in the "true" value of the indicator today, that will only be observed after a long delay. While this association requires a careful interpretation of the indicator (see Section 4.3.4) it was, besides case incidences and ICU occupancy, one of the main official indicators in Germany informing countermeasures in 2021 and so there is merit in nowcasting it.

The extent of delays is visible in Figure 4.1: the reported number of hospitalisations will roughly double over the course of twelve weeks. By the aforementioned reporting scheme of hospitalisations there are two reporting dates for a single hospitalised case: the reporting date of the case, i.e. the date when local health authorities were made aware of the positive test, and the reporting date of the hospitalisation, i.e. when the hospitalisation was reported to the RKI. This induces a double weekday effect in the reporting delays which we make visible in Figure 4.2.

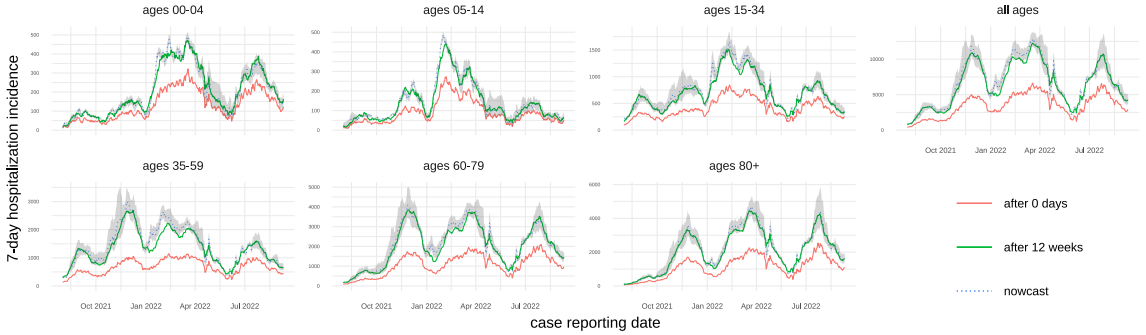

Figure 4.1: Germany's 7-day hospitalisation incidence changes due to various delays such as time to hospitalisation and delays in reporting. This figure shows the extent of these delays: incidences reported at the present date (red lines) severely underestimate the hospitalisation incidence (green solid lines) that is reported after 3 months. Our nowcasting model (blue dotted lines, 95% prediction intervals in shaded gray) deals with this problem by predicting the hospitalisation incidence based on past cases and their delays to hospitalisation.

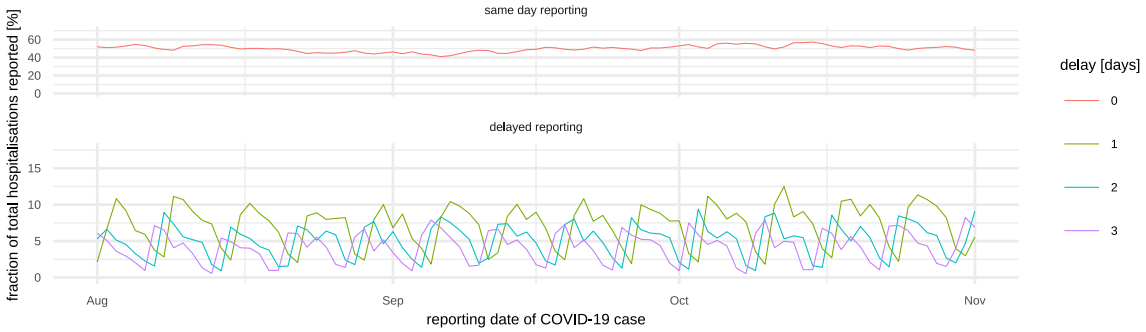

Figure 4.2: The hospitalisation incidence contains a double weekday effect owed to the reporting of both the COVID-19 case and the subsequent hospitalisation. While the weekday effect of the case reporting date is somewhat mitigated by summing over 7 day periods, the weekday effect of reporting date of the hospitalisation is still present in the data. It is most pronounced for hospitalisations that are reported with delays, i.e. where the case reporting date does not match the reporting date of the hospitalisation.

Compared to other approaches in the COVID-19 NowcastHub, that tended to exclusively focus on modelling the delay distribution with parametric and non-parametric models, our model sidesteps this complex delay structure by decomposing delayed hospitalisations into weekly chunks (Figure 4.4) and incorporating case data. As cases and hospitalisations are explicitly linked by the case reporting date we forecast the number of hospitalisations in each chunk based on the current incidences and past fractions of hospitalisations in a comparable weekly chunk. We additionally quantify uncertainty by prediction intervals that are informed by the past performance of our model. This makes our model straightforward to understand, easy to implement and fast to run.

The origin of nowcasting lie in accounting for incurred, but not reported claims in the actuarial sciences [28], delays in reporting for AIDS [33, 57] and other infectious diseases [19]. Popular statistical approaches include methods from survival analysis [33] and generalized linear regression [57]. In the survial analysis setting one commonly models the reverse time discrete hazard parametrically and assumes multinomial sampling of the final number of cases, potentially accounting for overdispersion. This has been studied with frequentist [35] and Bayesian [3, 24] methods. The generalized linear regression approach has origins in the chain ladder model from actuarial sciences [40] and models the observed counts in the reporting triangle by a Poisson or negative binomial distribution. For both approaches, available covariates can be incorporated in a straightforward way. In the setting of real-time nowcasting, it is often beneficial to incorporate epidemic dynamics into the model, this can be achieved by splines [24, 55] or by a latent process of infections [34].

Nowcasting methods have wide application in accouting for reporting delays [35], early outbreak detection [5, 48], and, in the recent COVID-19 epidemic, improving real-time monitoring of epidemic outbreaks [2, 3, 22, 49]. Evaluating a forecasting model in a real-time public health setting is advantageous as it avoid hindsight bias [13], however nowcasting approach may have difficulties with bias and properly calibrated uncertainty if used in a real-time setting. This includes rapidly changing dynamics [22, 55], both of the delay distribution and the underlying epidemic, retrospective changes in data [35] and long delays with few observed cases [37].

To avoid the aforementioned hindsight bias one can make their predictions publicly available in real-time [6, 39]. For the hospitalisations in Germany, Thomas Hotz and I have participated in the German COVID-19 NowcastHub [38] since November 2021 where nowcasts are available in a public Github repository [26] with the "ILM-prop" model. The ideas, especially the model and the "double-weekday effect", discussed this section are based on this model. However, the "ILM-prop" model is based on simple point estimates for the proportion of hospitalisations per reported case, neglecting regularization over time. In this thesis we extend this model to the SSM setting of this thesis and investigate if the increased model complexity results in improved performance.

4.3.2 Data

To predict the number of hospitalisations we consider the reporting process of both reported COVID-19 cases and reported hospitalisations. Recall that the reporting date of a COVID-19 case is shared for both the case and its hospitalisation, i.e. the case and hospitalisation are linked through this date.

As hospitalisations are only available as 7-day rolling sums, we use 7-day rolling sums for daily reported incidences as well. To avoid dealing with the double weekday effect of both reporting date of the case and reporting date of the hospitalisation (see Figure 4.2) we divide the future hospitalisations we wish to predict into chunks of one week, which gets rid of the weekday effect for the hospitalisations. This is depicted in Figure 4.4. Our prediction of each of these weekly chunks then consists of the fraction of hospitalisations of reported cases in the past.

We use publicly available data from the German national health authority (RKI) on daily reported COVID-19 cases [43] and weekly reported hospitalisations [42]. Both datasets are updated on a daily basis.

COVID-19 cases are described by their date of reporting, i.e. the date that the local health authorities were made aware of the case. For a fraction (63 %) of cases the date of symptom onset is also reported. Due to delays in the process from infection to reporting – e.g. the time it takes to get tested, evaluate the test and report the result to local health authorities – the date of reporting is, for most cases, some days after symptom onset (median delay: 3 with interquartile range $[2, 6]$). As the date of symptom onset is not known for a substantial amount of incident cases, and is not reported for hospitalised cases, we focus our analysis on the date of reporting.

Hospitalisations are associated with the *reporting date of the corresponding case* and no information is available on the actual date of hospitalisation. In addition, hospitalisations are only published as weekly sums over the past seven days. This means that the number of hospitalisations reported for today consists of all hospitalisations that correspond to cases that have a *case reporting date* in the past seven days. In particular if the case reporting date of a hospitalised case is today the case will *not* count towards todays hospitalisation count. The reporting date of hospitalisation is not available in the dataset, but can be inferred by comparing datasets from consecutive days.

Daily incident cases and weekly hospitalisations are reported by federal state and age group (00-04, 05-14, 15-34, 35-59, 60-79, 80+). Incident cases are additionally reported by county and sex.

In line with the structure of the data provided by the RKI we let $H^a_{t,d}$ be the number of weekly hospitalisations in age group a with case reporting date $t-1, \dots, t-7$ that are known on day $t+d$, aggregating over all states. Accordingly we define $I^a_{t,d}$ to be the number of weekly incident cases in age group a with reporting date $t-1, \dots, t-7$ that are known on day $t+d$. Finally we reconstruct the reporting triangles for weekly hospitalisations (Figure 4.4) by differencing the $H^a_{t,d}$ for fixed t: $h^a_{t,d} = H^a_{t,d} - H^a_{t,d-1}$, setting $H^a_{t,-1}$ to 0 by convention. We recover the reporting triangle $i^a_{t,d}$ for incident cases in the same manner.

We show the empirical surivival function of hospitalisations for a fixed date in Figure 4.3. We observe that delays have long tails, with most cases reported after 12 weeks (84 days), except for

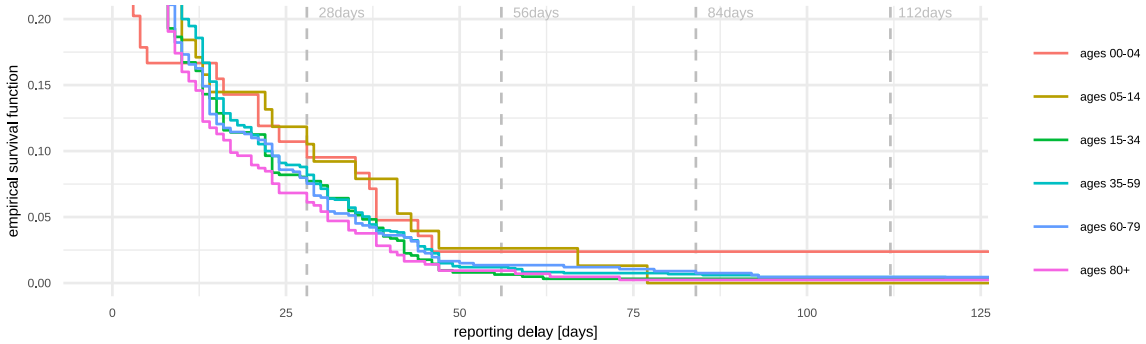

Figure 4.3: Survival function of reporting delays of weekly hospitalisations $H_{t,d}$ with case reporting date 01 September 2021. The delay distribution has long tails with a non-neglibgible fraction of observed delays longer than eight weeks, in some age groups even twelve weeks.

the youngest age group. After such a long delay between infection and hospitalisation we deem it unlikely that hospitalisation is due to COVID and disregard all longer delays accordingly. Given such long delays, it does not suffice to nowcast only todays hospitalisations, but also for dates in the past to monitor hospitalisation, i.e. observe current trends; we thus nowcast for all delays $d = 0, \dots, 28$.

4.3.3 Model

More formally, denote by $h_{t,d}$ the number hospitalisations with reporting date t that are known d days later. Unfortunately we only observe

$$H_{t,d} = \sum_{s=t-6}^{t} h_{s,d+(t-s)},$$

i.e. a weekly sum of reported hospitalisations. On day T our goal is to predict $H_{t,D}$ for large delays D and days $t \leq T$, of course it suffices to predict $H_{t,D} - H_{t,T-t}$ and add the known $H_{t,T-t}$ to this prediction. We rewrite this into weekly telescoping sum

$$H_{t,D} - H_{t,d} = (H_{t,d+7} - H_{t,d}) + (H_{t,d+14} - H_{t,d+7}) + \cdots + (H_{t,D} - H_{t,d+7K}),$$

where $K = \lfloor (D-d)/7 \rfloor$, reducing the task at hand to predict hospitalisations in the k-th week ahead, $H_{t,d+7k} - H_{t,d+7\cdot(k-1)}$, $k = 1, \dots, K$. To leverage known reported incidences, rewrite this as

$$\underbrace{\frac{H_{t,d+7k} - H_{t,d+7\cdot(k-1)}}{I_{t,d}}}_{=:p_{t,d,k}} I_{t,d}$$

where $I_{t,d}$ is the 7-day case incidence with reporting date t known at time $t+d$, i.e. the incidenct case analouge of $H_{t,d}$.

Assuming that the proportions $p_{t,d,k}$ change slowly over time t we estimate them by

$$\widehat{p_{t,d,k}} = \frac{H_{t-7k,d+7k} - H_{t-7k,d+7\cdot(k-1)}}{I_{t-7k,d}} = p_{t-7k,d,k} \tag{4.1}$$

and finally predict

$$\widehat{H_{t,D}} = H_{t,d} + I_{t,d}\,(\widehat{p_{t,d,1}} + \cdots + \widehat{p_{t,d,K}})\,. \tag{4.2}$$

As hospitalisation is affected by age, we perform this procedure for all available age groups separately and finally aggregate over all age groups to obtain a nowcast for all age groups combined.

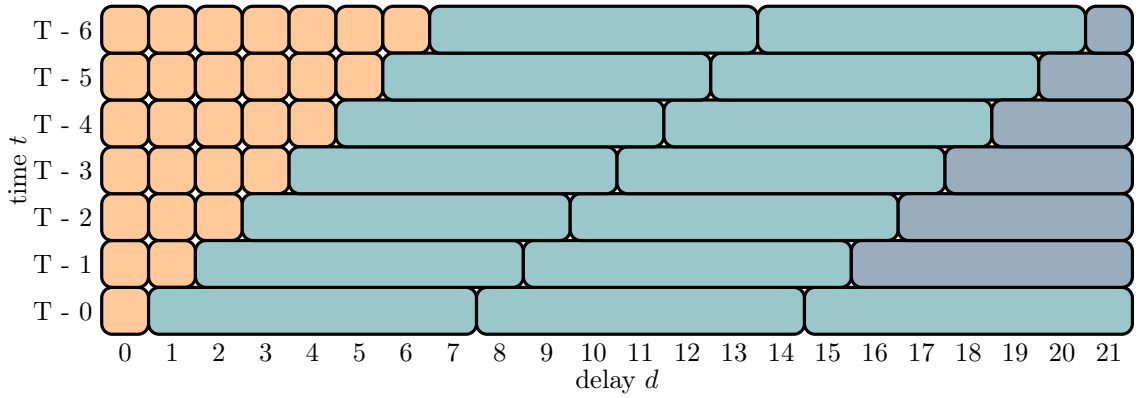

Figure 4.4: Decomposition of the daily reported hospitalisation incidences into the known incidences, i.e. the **reporting triangle**, and the future weekly increments. The last increment might not be a weekly one, but we expect few cases to occur for such long delays.

This describes our point nowcast for 7-day hospitalisations. To obtain uncertainty intervals we fit a normal (age groups 00-04 and 05-14) or lognormal (all other age groups) distribution to the past performance of our model. We chose these distributions based on explorative analysis and believe that these should be seen as heuristics rather than as a matter of fact, which is in line with the philosophy of our model to be as simple as possible.

Denote by $\hat{H}_{t,D,s}$ the nowcast made for date t on date $s \geq t$. Starting with date $t + D$ the definite $H_{t,D}$ is known and we can estimate the absolute prediction error $\varepsilon_{t,s} = H_{t,D} - \hat{H}_{t,D,s}$ and the relative prediction error $\eta_{t,s} = \log(H_{t,D} - H_{t,s-t}) - \log\left(\hat{H}_{t,D,s} - H_{t,s-t}\right)$. For the nowcast for date t made on date s we estimate the standard deviation $\hat{\sigma}$ of $\varepsilon_{t-D-i,s-D-i}$ or $\eta_{t-D-i,s-D-i}$ (age groups 00-04, 05-14 and others respectively), $i = 0, \ldots, 27$ by its empirical counterpart. The estimated predictive distribution which informs our prediction intervals is then $\mathcal{N}(\hat{H}_{t,D,s}, \sigma^2)$ (age groups 00-04 and 05-14) or $\mathcal{LN}\left(\log\left(\hat{H}_{t,D,s} - H_{t,s-t}\right), \sigma^2\right) + H_{t,s-t}$ (all other agr groups).

4.3.4 Discussion

Before evaluating the predictive performance of our model we investigate how the fraction of hospitalisations after one up to four weeks changes over time across different age groups. Figure 4.5 shows that these fractions are changing slowely over time, especially in the older age groups. Due to smaller numbers of infections and hospitalisations reported in the younger age groups these fractions vary more strongly, occassionally dropping to 0. Across all age groups we observe a steady decline from October 2021 to December 2021 with a steeper drop in fraction of hospitalisations starting with January 2022. The former period corresponds to a time of mandatory testing at the workplace which may improve ascertainment of asymptomatic and less severe cases. The latter effect is most recognizable in the 35-59 age group and coincides with the time that the Omicron variant became dominant in Germany [44]. Additionally there is no visually discernible weekday effect present in Figure 4.5.

In Figure 4.1 we depict the nowcasts produced from our model including 95% prediction intervals, whose lengths are based on the past performance of our model. Except for the period from January to April 2022, the model produces reasonable nowcasts with prediction intervals that have sensible widths. In the aforementioned period the nowcasts overpredict the final hospitalisations, except for the oldest age group, and, after a transitionary period, have larger uncertainty.

To investigate the quality of point predictions we display the time-evolution of absolute (AEP) and relative errors of predictions (REP, $\log_{10}$-scale) across all age groups in Figure 4.6. From this figure one can infer that the point nowcasts produced by our model tend to slightly overpredict the final number of hospitalisations. Indeed, the interquartile range of REPs for all age groups and dates combined spans $[-1.56, 8.33]$, demonstrating the same tendency. The highest REPs occured in October / November 2021 and January/February 2022; the first corresponding to introduction

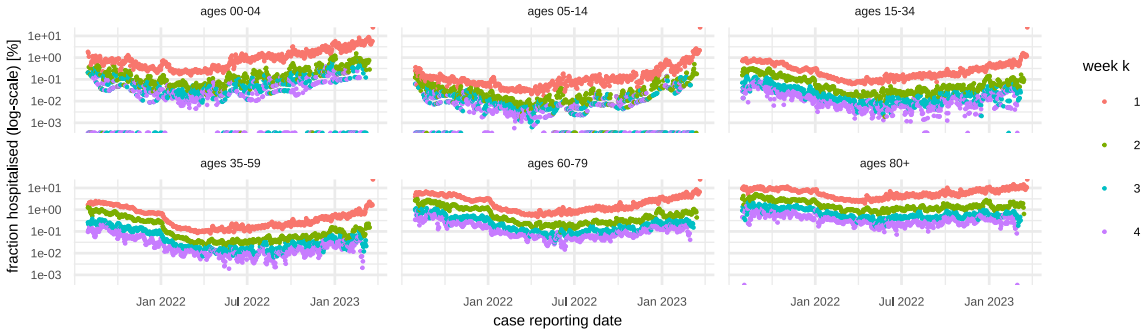

Figure 4.5: We show the fractions of hospitalisations in the k-th week after case reporting date t of initially reported cases in different age groups, i.e. $p_{t,0,k} = \left(H_{t,7k} - H_{t,7\cdot(k-1)}\right)/I_{t,0}$. Note the log-scale of the y-axis. During periods of low incidence, e.g. July – September, we find large fluctations, but no discernable weekly pattern. With rising case numbers the fractions stabilise and decrease in most age groups. This might be due to changes in testing regime detecting less severe cases. As changes occur on slow time scales, estimating these fractions by Eq. (4.1) is a promising approach.

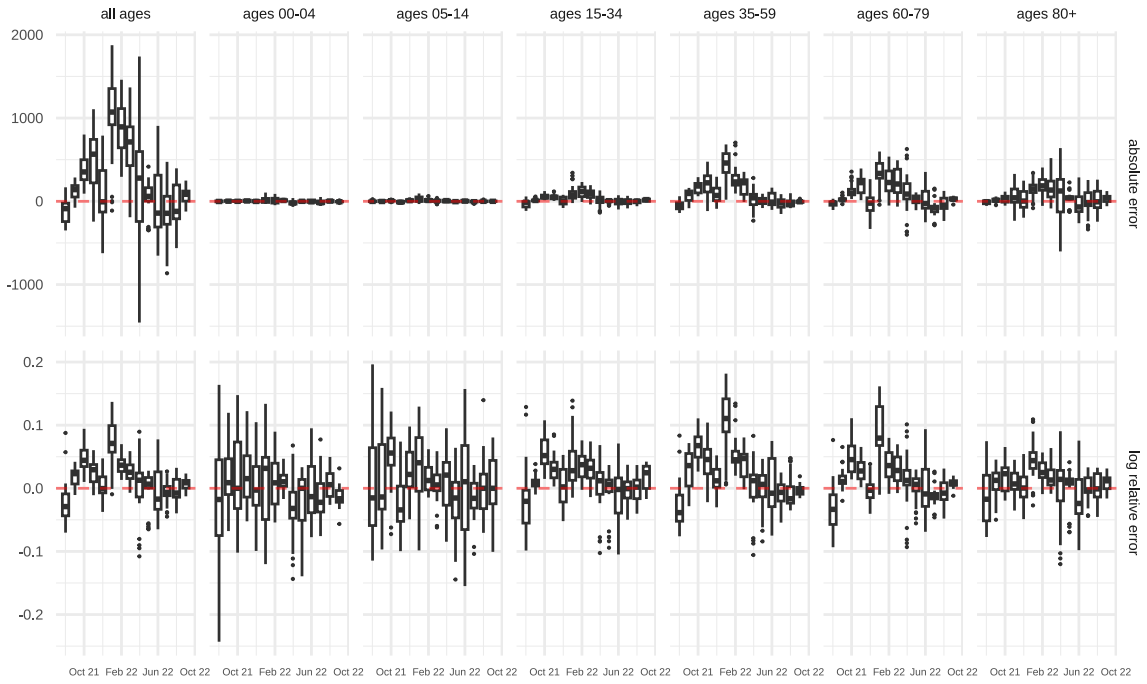

Figure 4.6: We show relative and absolute errors of prediction of our model for same day nowcasts by month of forecast and selected age groups. Relative errors are displayed on the log10 scale, i.e. as $\log_{10}(\text{predicted}) - \log_{10}(\text{actual})$. Up to December 2021 the model performs well, especially in the older age groups where most hospitalisations occur. The sharp increase in cases in January 2022 coupled with a lower probability of hospitalisation, most likely due to the appearance of the Omicron variant in Germany, lead to overpredictions across all age groups.

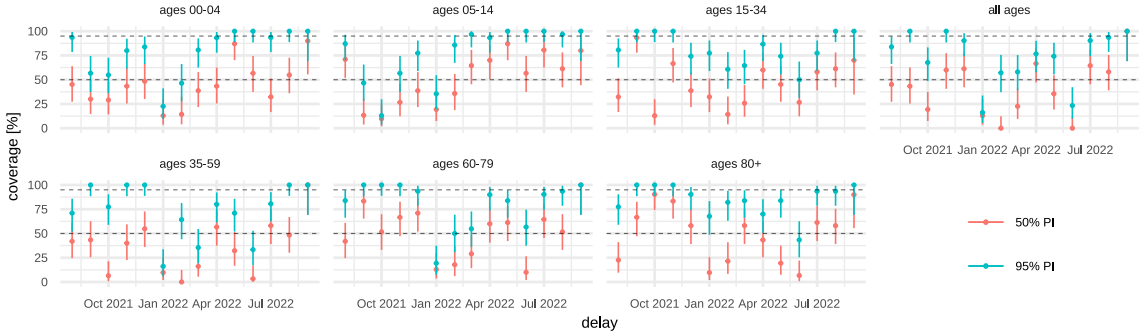

Figure 4.7: Empirical coverage of 50% and 95% prediction intervals (PI) based on same-day nowcasts for dates 2021-08-01 to 2022-09-10 (406 dates) for which the true amount of hospitalisations after 12 weeks is known as of the writing of this paper. We also display pointwise 95% binomial confidence intervals for the coverage. Given the difficulties of real-time forecasting [13] we deem the coverages good, except for the transitionary period in the end of 2021 where changing testing schemes and the change from Delta to Omicron cause our model to be overly confident. Coverage is generally better in the older age groups.

of mandatory testing at the workplace and the second to the arrival of the Omicron variant in Germany. In both circumstances the number of cases rose while hospitalisations did not increase proportionally, a similar effect to the one observed in Figure 4.5.

We further quantified uncertainties in estimaton by uncertainty intervals based on an assumption of a (log)-normal distribution for the errors with standard deviation based on past performance of our model. In Figure 4.7 we show the coverages of the 50% and 95% prediction interval across all age groups and delays for the whole time period of our study. For most age groups the 50% prediction interval has close to nominal coverage, while the 95% intervals have less than nominal coverage.

As our goal is to capture all of the uncertainty in this prediction, we chose to assume a sensible distribution for the prediction, a normal distribution for the two young age groups and a log-normal distribution for all other age groups. This has the advantage of producing more honest, wider, prediction intervals than those based on parametric distributions. The estimated standard deviation will also account for periods of low coverage, such as January 2022, albeit only after the maximum delay of 12 weeks.

We base our choice of 12 weeks of delay on the empirical survival function displayed in Figure 4.3. One could, however, argue for shorter maximum delay such as 6 weeks because time from reported infection to hospitalisation is much shorter, on the order of ≈ 10 days [18], so hospitalisations after this (shorter) period are unlikely to be due to the acute infection with SARS-CoV-2. This would have two main benefits: The model would adapt faster to changing circumstances and the indicator nowcasted describes the severity of the epidemic more appropriately.

The main advantage of our model over established nowcasting approches is its simplicity, making it easy to understand, straightforward to implement and, once the reporting triangles for incidences and hospitalisation are created, fast to run; taking only $< XX$ minutes on a standard notebook(**TODO: check!**).

The problem of nowcasting hospitalisations is different from previously studied nowcasting settings in several ways. At the time of nowcast a large fraction of hospitalisations are not only unobserved, but are yet to occur - in this sense the nowcast is more accurately termed a forecast. As the date of hospitalisation is not known, the hospitalisations are associated by the date of reporting of the COVID-19 case, creating the double-weekday effect displayed in Figure 4.2. While daily updated data on hospitalisations are available, these consist only of moving weekly aggregates, consecutive observations are strongly auto-correlated.

We sidestep all of these issues by splitting the hospitalisations to nowcast into weekly chunks, incorporating leading indicators of hospitalisation – the weekly reported case incidences – and modelling the number of hospitalisations to come in each chunk by binomial thinning of incidences. Let us stress that this approach is only possible in the special situation where case and hospitalisation

are explicitly linked, however we believe that incorporating leading indicators into nowcasting models is a promising approach.

An additional advantage of our model is that the hospitalisation probabilities can further be analysed, e.g. by investigating association between the publicly available vaccination rates and the probability of hosptialisation and delay to hospitalisation. Sudden changes in these fractions, as observed in Figure 4.5, can also hint towards worse model performance, especially if this change can be attributed to changing probability of hospitalisation due to new variants or changing testing regime.

Real time forecasting of epidemiological indicators is a difficult task [13], in particular quantifying uncertainty [6]. To test our model under real-time circumstances we submitted daily nowcasts to the German COVID-19 NowcastHub [38] since Novemer 2021. In the nowcasting context, [33] goes to great lengths to account for overdispersion due to changes in delay distribution, introducing gamma and Dirichlet priors and explicitly modelling trends. Such an approach would also be feasible for our approach, e.g. model incidences by an appropriate Poisson or negative binomial distribution and, conditional on incidences, model hospitalisations by a binomial distribution. As this increases the complexity of our model and relies on the assumed distributions being sensible we opted for another approach.

Regarding the indicator we stress that its value on a given date does not represent the current occupancy of hospitals in Germany with COVID-19 patients but is rather an approximation to the morbidity caused by COVID-19 on that date. The reason for this discrepancy is that hospitalisations are attributed to the reporting date of the associated case, not that of hospitalisation. While the reporting date of the hospitalisation can be recovered from the publicly available data, the date of hospitalisation cannot. Additionally, no information on the duration of stay is available, making it impossible to create an indicator for the occupancy of hospitals based solely on data provided by the RKI.

Implicit in all of these approaches is an assumption of "stationarity", i.e. that future reported hospitalisations will behave as they did in the past. Thus, all of these approaches might still be insufficient if circumstances change drastically, for example introduction of new testing schemes (school, 3G at workplace), changes in the delay distribution due to new variants, or hospitals close to capacity taking longer to process cases.

In summary, because models usually only capture a small part of the highly dynamic data-generating process, we believe that uncertainty in such circumstances should not come from unrealistic parametric assumptions but rather be based on past model performance. Given the discussed difficulties and the changing epidemiological dynamics in the period studied, the observed errors of prediction (Figure 4.6) and coverages of prediction intervals (Figure 4.7) are satisfying.

In this paper we provide a straight-forward model for nowcasting hospitalisations associated with COVID-19 in Germany. By leveraging known incident cases, we can estimate fractions of hospitalisations in weekly chunks which in turn avoids a complicated model of the two weekday effects present in the data. As the circumstances of the epidemic are changing constantly, e.g. vaccination coverage, testing regimes and emerging variants, we based uncertainty not on parametric assumptions but on the past performance of our model, assuming a (log)normal predictive distribution. We contributed nowcasts based on this model since November 2021 to the German COVID-19 NowcastHub [38], a collaborative platform collecting and aggregating such nowcasts from multiple research groups. The performance of the nowcasts in this Hub and presented in this paper (Figure 4.6 and Figure 4.7) are, regarding the simplicity of the model and the highly dynamic situation, quite satisfying.

There are multiple extensions to our model worth investigating. Firstly hospitalisations are also available at the federal state level so nowcasting on a spatial scale is naturally of interest due to hetereogeneity in immunisation status and testing regimes across states. However, splitting hospitalisations into six age groups and 16 federal states will result in small numbers with larger variability which in turn increases variabliity in estimates $p_{t,d,k}$ and thus predictions which may require some regularisation thus increasing the models complexity. Secondly, in a similar vein, modeling the temporal evolution of hospitalisation probabilities by smooth functions, e.g. splines [49, 55], may help in early detection of changing circumstances and thus lead to better forecasts. Thirdly our uncertainty intervals account for the variance of past performance, but Figure 4.6 suggests that there is substantial bias in periods of changing circumstances which could be incorporated into our model in a straightforward way. Finally to predict the future course of the epidemic forecasting

hospitalisations for dates t that lie in the future is of interest, which could be accomplished if one has a model that produces forecasts for incidences for all age groups.

Chapter 5

Discussion

Appendix A

Implementation in Python

Appendix B

Proofs (?)

Bibliography

[1] S. Agapiou et al. *Importance Sampling: Intrinsic Dimension and Computational Cost*. Jan. 14, 2017. DOI: 10.48550/arXiv.1511.06196. arXiv: 1511.06196 [stat]. URL: http://arxiv.org/abs/1511.06196 (visited on 04/03/2023). preprint.

[2] Andrei R. Akhmetzhanov. "Estimation of Delay-Adjusted All-Cause Excess Mortality in the USA: March-December 2020." In: *Epidemiology and Infection* (2021). DOI: 10.1017/s0950268821001527. pmid: 34210370.

[3] Matthias An Der Heiden and Osamah Hamouda. "Schätzung Der Aktuellen Entwicklung Der SARS-CoV-2- Epidemie in Deutschland – Nowcasting." In: *Epidemiologisches Bulletin* (Apr. 22, 2020). In collab. with Robert Koch-Institut and Robert Koch-Institut. DOI: 10.25646/6692.4. URL: https://edoc.rki.de/handle/176904/6650.4 (visited on 09/01/2020).

[4] Richard Arratia, Larry Goldstein, and Louis Gordon. "Poisson Approximation and the Chen-Stein Method." In: *Statistical Science* 5.4 (1990), pp. 403–424. ISSN: 0883-4237. JSTOR: 2245366. URL: https://www.jstor.org/stable/2245366 (visited on 01/11/2024).

[5] Leonardo Soares Bastos et al. "A Modelling Approach for Correcting Reporting Delays in Disease Surveillance Data." In: *Statistics in Medicine* (2019). DOI: 10.1002/sim.8303.

[6] J. Bracher et al. "A Pre-Registered Short-Term Forecasting Study of COVID-19 in Germany and Poland during the Second Wave." In: *Nature Communications* 12.1 (1 Aug. 27, 2021), p. 5173. ISSN: 2041-1723. DOI: 10.1038/s41467-021-25207-0. URL: https://www.nature.com/articles/s41467-021-25207-0 (visited on 09/30/2021).

[7] Johannes Bracher et al. "National and Subnational Short-Term Forecasting of COVID-19 in Germany and Poland during Early 2021." In: *Communications Medicine* 2.1 (1 Oct. 31, 2022), pp. 1–17. ISSN: 2730-664X. DOI: 10.1038/s43856-022-00191-8. URL: https://www.nature.com/articles/s43856-022-00191-8 (visited on 11/16/2022).

[8] Jan M. Brauner et al. "Inferring the Effectiveness of Government Interventions against COVID-19." In: *Science* 371.6531 (Feb. 19, 2021), eabd9338. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abd9338. URL: https://www.science.org/doi/10.1126/science.abd9338 (visited on 07/06/2023).

[9] Lawrence D. Brown. *Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory*. Lecture Notes-Monograph Series v. 9. Hayward, Calif: Institute of Mathematical Statistics, 1986. 283 pp. ISBN: 978-0-940600-10-2.

[10] Stephen Chan et al. "Count Regression Models for COVID-19." In: *Physica A: Statistical Mechanics and its Applications* 563 (Feb. 1, 2021), p. 125460. ISSN: 0378-4371. DOI: 10.1016/j.physa.2020.125460. URL: https://www.sciencedirect.com/science/article/pii/S0378437120307743 (visited on 01/09/2024).

[11] Sourav Chatterjee and Persi Diaconis. "The Sample Size Required in Importance Sampling." In: *The Annals of Applied Probability* 28.2 (Apr. 1, 2018). ISSN: 1050-5164. DOI: 10.1214/17-AAP1326. URL: https://projecteuclid.org/journals/annals-of-applied-probability/volume-28/issue-2/The-sample-size-required-in-importance-sampling/10.1214/17-AAP1326.full (visited on 05/26/2023).

[12] Nicolas Chopin and Omiros Papaspiliopoulos. *An Introduction to Sequential Monte Carlo*. Springer Series in Statistics. Cham, Switzerland: Springer, 2020. 378 pp. ISBN: 978-3-030-47844-5.

[13] Angel N. Desai et al. "Real-Time Epidemic Forecasting: Challenges and Opportunities." In: *Health Security* 17.4 (Aug. 2019), pp. 268–275. ISSN: 2326-5094. DOI: 10.1089/hs.2019.0022. URL: https://www.liebertpub.com/doi/abs/10.1089/hs.2019.0022 (visited on 10/19/2022).

[14] J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. 2nd ed. Oxford Statistical Science Series 38. Oxford: Oxford University Press, 2012. 346 pp. ISBN: 978-0-19-964117-8.

[15] James Durbin and Siem Jan Koopman. "Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models." In: *Biometrika* 84.3 (Sept. 1, 1997), pp. 669–684. ISSN: 0006-3444. DOI: 10.1093/biomet/84.3.669. URL: https://doi.org/10.1093/biomet/84.3.669 (visited on 02/06/2023).

[16] Max Ehre et al. "Certified Dimension Reduction for Bayesian Updating with the Cross-Entropy Method." In: *SIAM/ASA Journal on Uncertainty Quantification* 11.1 (Mar. 31, 2023), pp. 358–388. DOI: 10.1137/22M1484031. URL: https://epubs.siam.org/doi/10.1137/22M1484031 (visited on 07/24/2023).

[17] Michael Engel et al. "Bayesian Updating and Marginal Likelihood Estimation by Cross Entropy Based Importance Sampling." In: *Journal of Computational Physics* 473 (Jan. 15, 2023), p. 111746. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2022.111746. URL: https://www.sciencedirect.com/science/article/pii/S0021999122008099 (visited on 07/27/2023).

[18] Christel Faes et al. "Time between Symptom Onset, Hospitalisation and Recovery or Death: Statistical Analysis of Belgian COVID-19 Patients." In: *International Journal of Environmental Research and Public Health* 17.20 (20 Jan. 2020), p. 7560. ISSN: 1660-4601. DOI: 10.3390/ijerph17207560. URL: https://www.mdpi.com/1660-4601/17/20/7560 (visited on 11/14/2022).

[19] C. P. Farrington et al. "A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease." In: *Journal of The Royal Statistical Society Series A-statistics in Society* (1996). DOI: 10.2307/2983331.

[20] Seth Flaxman et al. "Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe." In: *Nature* 584.7820 (Aug. 2020), pp. 257–261. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2405-7. pmid: 32512579. URL: https://www.nature.com/articles/s41586-020-2405-7 (visited on 08/28/2020).

[21] Sylvia Frühwirth-Schnatter. "Data Augmentation and Dynamic Linear Models." In: *Journal of Time Series Analysis* 15.2 (1994), pp. 183–202. ISSN: 1467-9892. DOI: 10.1111/j.1467-9892.1994.tb00184.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.1994.tb00184.x (visited on 06/08/2022).

[22] Felix Günther et al. "Nowcasting the COVID-19 Pandemic in Bavaria." In: *Biometrical Journal* 63.3 (2021), pp. 490–502. ISSN: 1521-4036. DOI: 10.1002/bimj.202000112. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.202000112 (visited on 11/15/2021).

[23] Stefan Heyder and Thomas Hotz. "Measures of COVID-19 Spread." In: *Covid-19 pandisziplinär und international: Gesundheitswissenschaftliche, gesellschaftspolitische und philosophische Hintergründe*. Ed. by Alexander Kraemer and Michael Medzech. Medizin, Kultur, Gesellschaft. Wiesbaden: Springer Fachmedien, 2023, pp. 51–66. ISBN: 978-3-658-40525-0. DOI: 10.1007/978-3-658-40525-0_3. URL: https://doi.org/10.1007/978-3-658-40525-0_3 (visited on 10/21/2023).

[24] Michael Höhle and Matthias An Der Heiden. "Bayesian Nowcasting during the STEC O104:H4 Outbreak in Germany, 2011." In: *Biometrics* 70.4 (2014), pp. 993–1002. ISSN: 1541-0420. DOI: 10.1111/biom.12194. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12194 (visited on 04/15/2020).

[25] Tito Homem-de-Mello. "A Study on the Cross-Entropy Method for Rare-Event Probability Estimation." In: *INFORMS Journal on Computing* (July 20, 2007). DOI: 10.1287/ijoc.1060.0176. URL: https://pubsonline.informs.org/doi/abs/10.1287/ijoc.1060.0176 (visited on 07/07/2023).

[26] *Hospitalization Nowcast Hub*. KITmetricslab, Oct. 31, 2022. URL: https://github.com/KITmetricslab/hospitalization-nowcast-hub (visited on 11/09/2022).

[27] Borus Jungbacker and Siem Jan Koopman. "Monte Carlo Estimation for Nonlinear Non-Gaussian State Space Models." In: *Biometrika* 94.4 (Dec. 1, 2007), pp. 827–839. ISSN: 0006-3444. DOI: 10.1093/biomet/asm074. URL: https://doi.org/10.1093/biomet/asm074 (visited on 03/04/2023).

[28] Kenneth S. Kaminsky. "Prediction of IBNR Claim Counts by Modelling the Distribution of Report Lags." In: *Insurance Mathematics & Economics* 6.2 (Apr. 1, 1987), pp. 151–159. DOI: 10.1016/0167-6687(87)90024-2.

[29] H. J. Kappen and H. C. Ruiz. "Adaptive Importance Sampling for Control and Inference." In: *Journal of Statistical Physics* 162.5 (Mar. 1, 2016), pp. 1244–1266. ISSN: 1572-9613. DOI: 10.1007/s10955-016-1446-7. URL: https://doi.org/10.1007/s10955-016-1446-7 (visited on 07/07/2023).

[30] Yeganeh Khazaei et al. "Using a Bayesian Hierarchical Approach to Study the Association between Non-Pharmaceutical Interventions and the Spread of Covid-19 in Germany." In: *Scientific Reports* 13.1 (Nov. 2, 2023), p. 18900. ISSN: 2045-2322. DOI: 10.1038/s41598-023-45950-2. URL: https://www.nature.com/articles/s41598-023-45950-2 (visited on 11/10/2023).

[31] Augustine Kong, Jun S. Liu, and Wing Hung Wong. "Sequential Imputations and Bayesian Missing Data Problems." In: *Journal of the American Statistical Association* 89.425 (Mar. 1994), pp. 278–288. ISSN: 0162-1459. DOI: 10.1080/01621459.1994.10476469. URL: https://www.tandfonline.com/doi/citedby/10.1080/01621459.1994.10476469 (visited on 03/26/2024).

[32] Pierre Simon Laplace. "Memoir on the Probability of the Causes of Events." In: *Statistical Science* 1.3 (Aug. 1986), pp. 364–378. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/ss/1177013621. URL: https://projecteuclid.org/journals/statistical-science/volume-1/issue-3/Memoir-on-the-Probability-of-the-Causes-of-Events/10.1214/ss/1177013621.full (visited on 03/27/2024).

[33] J. F. Lawless. "Adjustments for Reporting Delays and the Prediction of Occurred but Not Reported Events." In: *Canadian Journal of Statistics* 22.1 (1994), pp. 15–31. ISSN: 1708-945X. DOI: 10.2307/3315826.n1. URL: https://onlinelibrary.wiley.com/doi/abs/10.2307/3315826.n1 (visited on 05/13/2020).

[34] Sarah F. McGough et al. "Nowcasting by Bayesian Smoothing: A Flexible, Generalizable Model for Real-Time Epidemic Tracking." In: *PLOS Computational Biology* (2020). DOI: 10.1371/journal.pcbi.1007735. pmid: 32251464.

[35] Douglas N. Midthune et al. "Modeling Reporting Delays and Reporting Corrections in Cancer Registry Data." In: *Journal of the American Statistical Association* 100.469 (Mar. 1, 2005), pp. 61–70. ISSN: 0162-1459. DOI: 10.1198/016214504000001899. URL: https://doi.org/10.1198/016214504000001899 (visited on 07/21/2020).

[36] M. Morf and T. Kailath. "Square-Root Algorithms for Least-Squares Estimation." In: *IEEE Transactions on Automatic Control* 20.4 (Aug. 1975), pp. 487–497. ISSN: 1558-2523. DOI: 10.1109/TAC.1975.1100994.

[37] Angela Noufaily et al. "Modelling Reporting Delays for Outbreak Detection in Infectious Disease Data." In: *Journal of The Royal Statistical Society Series A-statistics in Society* (2015). DOI: 10.1111/rssa.12055.

[38] *Nowcasts Der COVID-19 Hospitalisierungsinzidenz*. 2022. URL: https://covid19nowcasthub.de/ (visited on 11/09/2022).

[39] Evan L. Ray et al. "Ensemble Forecasts of Coronavirus Disease 2019 (COVID-19) in the U.S." In: *medRxiv* (Aug. 22, 2020), p. 2020.08.19.20177493. DOI: 10.1101/2020.08.19.20177493. URL: https://www.medrxiv.org/content/10.1101/2020.08.19.20177493v1 (visited on 09/02/2020).

[40] Arthur E. Renshaw and Richard J. Verrall. "A Stochastic Model Underlying the Chain-Ladder Technique." In: *British Actuarial Journal* 4.4 (1998), pp. 903–923. DOI: 10.1017/S1357321700000222.

[41] Jean-Francois Richard and Wei Zhang. "Efficient High-Dimensional Importance Sampling." In: *Journal of Econometrics* 141.2 (Dec. 1, 2007), pp. 1385–1411. ISSN: 0304-4076. DOI: 10.1016/j.jeconom.2007.02.007. URL: https://www.sciencedirect.com/science/article/pii/S0304407607000486 (visited on 11/23/2022).

[42] Robert Koch-Institut. *COVID-19-Hospitalisierungen in Deutschland*. Version 2021-10-01. Zenodo, Oct. 1, 2021. DOI: 10.5281/ZENODO.5519056. URL: https://zenodo.org/record/5519056 (visited on 10/01/2021).

[43] Robert Koch-Institut. *SARS-CoV-2 Infektionen in Deutschland*. Version 2022-02-07. Zenodo, Feb. 7, 2022. DOI: 10.5281/ZENODO.4681153. URL: https://zenodo.org/record/4681153 (visited on 02/08/2022).

[44] Robert Koch-Institut. *Wöchentlicher Lagebericht Des RKI Zur Coronavirus-Krankheit-2019 (COVID-19)*. Jan. 2022.

[45] Nathan Ross. "Fundamentals of Stein's Method." In: *Probability Surveys* 8 (none Jan. 1, 2011). ISSN: 1549-5787. DOI: 10.1214/11-PS182. URL: https://projecteuclid.org/journals/probability-surveys/volume-8/issue-none/Fundamentals-of-Steins-method/10.1214/11-PS182.full (visited on 01/11/2024).

[46] Reuven Rubinstein. "The Cross-Entropy Method for Combinatorial and Continuous Optimization." In: *Methodology And Computing In Applied Probability* 1.2 (Sept. 1, 1999), pp. 127–190. ISSN: 1573-7713. DOI: 10.1023/A:1010091220143. URL: https://doi.org/10.1023/A:1010091220143 (visited on 08/05/2023).

[47] Reuven Y. Rubinstein and Dirk P. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. New York, NY: Springer New York, 2004. ISBN: 978-1-4757-4321-0.

[48] Maëlle Salmon et al. "Bayesian Outbreak Detection in the Presence of Reporting Delays." In: *Biometrical Journal* (2015). DOI: 10.1002/bimj.201400159. pmid: 26250543.

[49] Marc Schneble et al. "Nowcasting Fatal COVID-19 Infections on a Regional Level in Germany." In: *Biometrical Journal* 63.3 (Mar. 2021), pp. 471–489. ISSN: 0323-3847, 1521-4036. DOI: 10.1002/bimj.202000143. URL: https://onlinelibrary.wiley.com/doi/10.1002/bimj.202000143 (visited on 09/16/2021).

[50] Wolfgang Schneider. *Der Kalmanfilter Als Instrument Zur Diagnose Und Schätzung Variabler Parameter in Ökonometrischen Modellen*. Arbeiten Zur Angewandten Statistik Bd. 27. Heidelberg: Physica-Verlag, 1986. 490 pp. ISBN: 978-3-7908-0359-4.

[51] Neil Shephard and Michael K. Pitt. "Likelihood Analysis of Non-Gaussian Measurement Time Series." In: *Biometrika* 84.3 (Sept. 1, 1997), pp. 653–667. ISSN: 0006-3444. DOI: 10.1093/biomet/84.3.653. URL: https://doi.org/10.1093/biomet/84.3.653 (visited on 08/10/2023).

[52] K. Sherratt et al. *Predictive Performance of Multi-Model Ensemble Forecasts of COVID-19 across European Nations*. June 16, 2022. DOI: 10.1101/2022.06.16.22276024. URL: http://medrxiv.org/lookup/doi/10.1101/2022.06.16.22276024 (visited on 11/28/2022). preprint.

[53] Luke Tierney and Joseph B. Kadane. "Accurate Approximations for Posterior Moments and Marginal Densities." In: *Journal of the American Statistical Association* 81.393 (Mar. 1986), pp. 82–86. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1986.10478240. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478240 (visited on 03/27/2024).

[54] Luke Tierney, Robert E. Kass, and Joseph B. Kadane. "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions." In: *Journal of the American Statistical Association* 84.407 (Sept. 1989), pp. 710–716. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1989.10478824. URL: https://www.tandfonline.com/doi/full/10.1080/01621459.1989.10478824 (visited on 03/27/2024).

[55] Jan van de Kassteele et al. "Nowcasting the Number of New Symptomatic Cases during Infectious Disease Outbreaks Using Constrained P-Spline Smoothing." In: *Epidemiology (Cambridge, Mass.)* (2019). DOI: 10.1097/ede.0000000000001050. pmid: 31205290.

[56] Aad W. Van der Vaart. *Asymptotic Statistics*. Cambridge: Cambridge University Press, 2000.

[57] Scott L. Zeger, Lai-Chu See, and Peter J. Diggle. "Statistical Methods for Monitoring the AIDS Epidemic." In: *Statistics in Medicine* 8.1 (1989), pp. 3–21. DOI: 10.1002/sim.4780080104.

[58] Wei Zhang et al. "Applications of the Cross-Entropy Method to Importance Sampling and Optimal Control of Diffusions." In: *SIAM Journal on Scientific Computing* 36.6 (Jan. 2014), A2654–A2672. ISSN: 1064-8275. DOI: 10.1137/14096493X. URL: https://epubs.siam.org/doi/abs/10.1137/14096493X (visited on 07/31/2023).

Declaration

Put your declaration here.

Ilmenau, October 2023

Stefan Heyder