Linköping University | Department of Computer and Information Science
Master's thesis, 30 ECTS | Statistics and Machine Learning
2021 | LIU-IDA/LITH-EX-A--2021/001--SE

HMM modelling for the spread of the SARS–CoV–2

Martin Benes

Supervisor : Krzysztof Bartoszek
Examiner : Maryna Prus

LINKÖPING UNIVERSITY

Linköpings universitet
SE–581 83 Linköping
+46 13 28 10 00 , www.liu.se

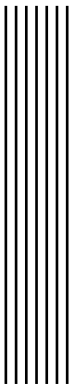
Upphovsrätt

Copyright

Abstract

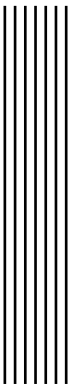
The aim of the project is to develop an HMM for the current spread of the SARS–CoV–2 virus. The HMM could be coupled with a SIR+ based compartmental model for the different types of statistics—confirmed cases, hospitalizations, deaths. The confirmed cases should be treated as a random sample from the whole population of infected and the probability of sampling should try to take into account the different testing strategies.

The aim of the project would be to compare the spread of the virus in different countries (e.g. Czech Republic, Poland, Sweden, Italy, but other depending on the availability of data are possible) through regional (whenever possible) dynamics. For the thesis publicly available COVID–19 connected data will be used.



Acknowledgments

TODO.



Contents

List of Figures

1 Introduction

Motivation

Currently there is an ongoing pandemic of Covid-19, the greatest challenge humans as a species had to face in last decades. Twentieth century introduced epidemiology as a research discipline and enabled spread of infections being viewed from mathematical rather than medical perspective. Neglected field just a few years ago got to the public eye now as the media chase its experts in these days.

Figure 1.1: Press Conference of Federal Chancellery of Austria, March 14, 2020. [1]

Aim

This thesis is presenting a Hidden Markov model of Covid-19 spread and performs a simulation with it to approximately estimate the true situation about infection for regions of several European countries: Czechia, Poland, Sweden and Italy.

Research questions

To construct the model, the characteristics of the Covid-19 disease are investigated from relevant scientific literature and presented in form of probability distributions. Just the model definition already answers question What are the distributions of parameters of Covid-19 - the incubation period, infection fatality ratio, reproduction number and duration of disease?

The Covid-19 statistics come from various sources and methods of measurement. Their correctness can be questioned not only in terms of accuracy, but in some cases even reliability of the source. [2] In other words, to what extent are the collected data used to fit the model reliable?

The simulation with the HMM on different regional data might bring questions such as How much are the reported statistics projected in the results of simulation? Are there visible patterns or similarities between regions? Do introduced restrictions in the region or country correspond with the numbers yielded by the simulation?

The reliability of the simulation results is directly connected with the correspondence of the probabilistic definition of Covid-19 introduced in the thesis with reality. To what extend the results show that the drafted model of the disease is correct?

Delimitations

The infection is modelled on a certain level of reality abstraction, so that many aspects of the infection are simplified. Those not included, but discussed are a single level of infection severeness - no asymptotic cases, a permanent immunity after recovery, no movement of population between regions/countries including incoming infectious population, no vaccination and 100% accuracy of clinical tests.

2 Theory

2.1 Epidemics

Epidemic is an outbreak of disease that freely spreads through population. Various epidemics and pandemics[1] are not only a matter of the modern era, but occur throughout the human history.

Pre-modern epidemics

One of the oldest mentions in literature is influenza epidemic in Persian Babylon in 1103 BC [3], however archaeological discoveries suggest even much older occurrences, such as the one in northeast China from 3000 BC [4].

By far the deadliest was plague pandemic called Black Death from the 14^{th} century with the death toll around 25 mil. people [5], other epidemics caused by *Yersinia pestis* were Justinian's Plague ($540 - 750$ AD) [6], the Second Plague ($14^{\text{th}} - 19^{\text{th}}$ century) and the Third Plague ($1899 - 1940$'s) [7].

Another frequent epidemics were caused by influenza [8], cholera [9], tuberculosis, typhus or smallpox [10], the latter was eradicated in 1980 [11]. Many diseases are endemic such as yellow fever or malaria due to climate-dependent disease vector [12, 13], or Cocoliztli - a group of common diseases that decimated the Aztec population in mid 16^{th} century [14].[2]

Modern epidemics

Regarding the pandemics, 20^{th} and 21^{st} centuries are dominated by the flu - the Spanish flu ($1918 - 1920$), the Russian flu (1977) and the Swine flu (2019) caused by Influenza A/H1N1, the Hong Kong flu ($1968 - 1969$) and the Asian flu ($1957 - 1958$) caused by Influenza A/H2N2 and its descendant A/H3N2 respectively [15]. The Spanish flu by itself directly caused 20 mil. deaths, far more than WWI [16]. There

[1] Epidemic is a general outbreak of disease. Pandemic is an epidemic that affects a significant portion of population of a continent, or worldwide.

[2] Civilizations on American continent developed isolated from the rest of the world for thousands of years and so did their immunity systems, adapted to the pathogens in the environment. Diseases brought by the first European colonizers (called *Cocoliztli*, in Nahuatl/Aztec meaning pest) were something absolutely novel for Americans' immunity systems and Cocoliztli wiped most of the population out.

Since early 1980's and still ongoing there has been a pandemic of a sexually-transmitted virus HIV[3], that causes AIDS[4], a disease that in the last 40 years killed more than 38 mil. people [17].

2.2 SARS-CoV-2

At the end of 2019 an outbreak of novel coronavirus occurred in Wuhan, China, later named *Severe acute respiratory syndrom coronavirus 2* (SARS-CoV-2). The virus shown in the figure 2.1 quickly spreaded around China and abroad and in just a matter of months, most of the world introduced epidemiological restrictions in order to stop the spread.

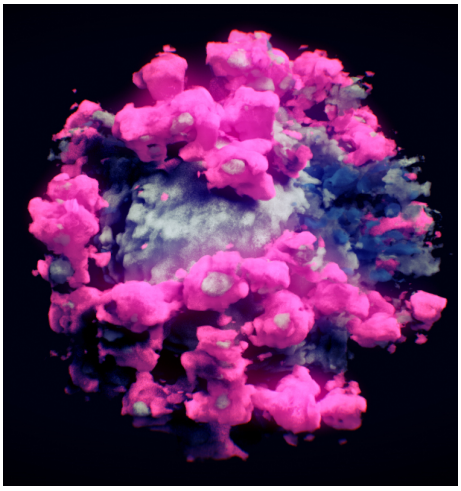


Figure 2.1: SARS-CoV-2 cryo-electron tomography scan [18].

The main symptoms were cough (67.6%), fever (62.2%), shortness of breath (32.4%), fatigue (24.3%), sore throat (21.6%) vomiting or diarrhea (21.6%).

The disease caused by SARS-CoV-2 was called Covid-19. The infected encountered respiratory illness with symptoms such as cough, fever, breathing difficulties, tiredness and loss of taste and smell, however their severity or absence varied significantly between patients. Fatality was significantly different for elderly population over 60 years (< 0.001), as shown in the figure 2.2 [19, 20, 21, 22, 23].

Diagnostics

The method of collection of the data is an important factor for the correct evaluation of the analysis result. There are several broadly used diagnostic tests, lab-based and rapid, used for detecting of presence (past or present) of SARS-CoV-2 virus in the patient's organism. There are two types with regard to what is being detected:

- *diagnostic tests* - virus itself or its parts (spike protein, RNA)
- *antibody tests* - antibodies produced by the host organism as a response to the virus

Diagnostic tests Diagnostic tests detect an active Covid-19 infection. The sample for the test is a nasal or throat swab. After patient recovers, the test yields negative result again.

[3]Human immunodeficiency virus
[4]Acquired immunodeficiency syndrome

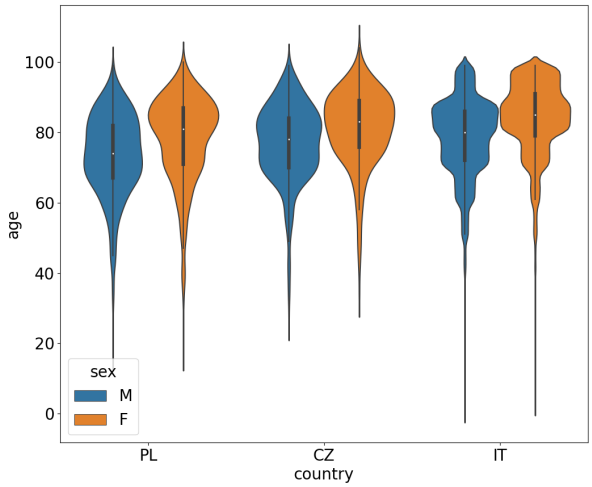


Figure 2.2: Fatality of Covid-19 per age and gender.

Currently the most commonly used is a Reverse transcription polymerase chain reaction (RT-PCR). It consists of reverse transcription of viral RNA to DNA and amplification (replication) of the DNA, which happens only with the gene sequence of SARS-CoV-2.[5] Chain reaction activates fluorescent molecules, which indicate the presence of the DNA and hence the virus itself.

The accuracy of the test depends on sampling method and the used kit, average performance over different kits is shown in the table 2.3 [24].

Specificity	**Sensitivity**
98.787	99.545

Figure 2.3: Accuracy of RT-PCR test.

Different method of diagnosing an ongoing Covid-19 infection are antigen tests, that look for viral proteins specific for SARS-CoV-2.

Antibody tests Antibody tests use blood serum and search for antibodies (IgM - early infection, IgG - long term immunity, ...) produced by the immune system as a response to encountered antigens - viral proteins. The testing makes only sense for person that already had gone through the disease, as it measures a developed immunity. It is used for estimation of disease prevalence and infection fatality rate (IFR).

[5]Replication of the DNA is selective thanks to customized primers, marking a start point of polymerase reaction. Polymerase is an enzyme capable of synthesis of DNA that duplicates each of the separated strains of DNA (only if matched by primers) in every reaction cycle. Amount of DNA grows exponentially.

2.3 Epidemiological Modeling

The epidemiology has experienced its first boom several years after so called Spanish flu[6], the first modern pandemic and at the beginning was described by medicine specialists, only later it was understood as an inter-discipline with mathematics and statistics [25].

SIR* model

SIR* models, the most prominent class of epidemiological models, describe disease as a set of states with parameterized transitions between them. Each person in the modelled population has a state, that changes with certain probability according to the chosen model. Simple SIR model has three states and two connection: susceptible **S**, infected **I**, recovered **R**, connected such as S → I (getting sick) and I → R (recovering).

$$\begin{aligned} \frac{dS}{dt} &= -aSI \\ \frac{dI}{dt} &= aSI - bI \\ \frac{dR}{dt} &= bI \end{aligned} \tag{2.1}$$

SIR as a dynamic model can be expressed with differential equations (eq. 2.1). In this expression letter denoting state stand for number of people belonging to it at time t. Each equation describes change in the number of people for the state. Its terms correspond with its connections, e.g. number of people at time t taking connection S → I is aSI.

SIR model describes diseases with lethality 0, no incubation period and permanent immunity. Since neither of this is true for SARS-CoV-2, for modelling we must extend SIR model by additional states: exposed **E** and dead **D**. All the states and connection of SEIDR model are shown in the figure 2.4.

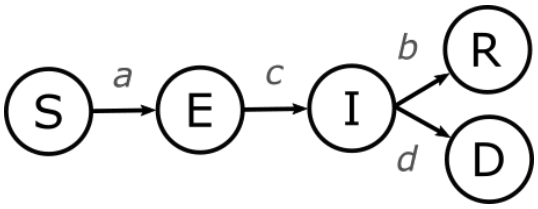


Figure 2.4: SEIDR model with permanent immunity.

As there is no connection R → S, this model assumes permanent immunity as well. The set of differential equations for the SEIDR model is defined by equation 2.2.

The model parameters a, c, b, d are directly connected with the disease characteristics: basic reproduction number R_0, incubation period, symptom duration and infection fatality rate IFR, closely specified in the equation 2.3.

The findings of epidemiology in a form of *theory of happenings* are applicable to wide range of different areas including marketing, malware, culture and others, making it a new and solid area of mathematics [26].

[6]The name *Spanish flu* for the pandemic in late 1910's stems from the fact that media in Spain as one of the few neutral countries informed about the situation and casualties, while countries participating in WWI censored it not to cause a hysteria. Thus, situation in Spain looked much worse than elsewhere.

$$\begin{aligned}
\frac{dS}{dt} &= -aSI \\
\frac{dE}{dt} &= aSI - cE \\
\frac{dI}{dt} &= cE - bI - dD \\
\frac{dR}{dt} &= bI \\
\frac{dD}{dt} &= dD
\end{aligned} \tag{2.2}$$

$$\begin{aligned}
R_0 &= \frac{a}{b+d}S \\
\text{incubation period} &= c^{-1} \\
\text{symptom duration} &= (b+d)^{-1} \\
\text{IFR} &= \frac{d}{b+d}
\end{aligned} \tag{2.3}$$

Bayesian SEIRD* model

Bayesian approach in epidemiological modelling brings uncertainty of parameter values. Instead of single value for the parameter, the model considers parameters a, c, b, d to have a prior distribution representing our best guess supported by prior knowledge. Choice of prior is based on clinical measurements of the infected individuals.

Posterior probability contains both the prior and the information extracted from the data modelled by a likelihood distribution using a Bayes' theorem.

$$\begin{aligned}
\theta &= (a, c, b, d) \\
P(\theta|D) &\propto P(D|\theta)P(\theta)
\end{aligned} \tag{2.4}$$

Connection of parameters a, c, b, d with the disease characteristics as specified in stochastic manner by the equation 2.4 is slightly different for Bayesian approach, the terms on both sides of equation sign are equal by distribution. The alternation is shown in the figure 2.5.

$$\begin{aligned}
R_0 &\overset{d}{=} \frac{a}{b+d}S \\
\text{incubation period} &\overset{d}{=} c^{-1} \\
\text{symptom duration} &\overset{d}{=} (b+d)^{-1} \\
\text{IFR} &\overset{d}{=} \frac{d}{b+d}
\end{aligned} \tag{2.5}$$

2.4 Hidden Markov Models

Hidden Markov model (HMM) is a discrete stochastic model for time-series. It has two main components: latent (unobserved) variables called states and observations, that are to be modelled. Since the model is discrete, both the latent and the observed variables have predefined finite alphabet of symbols (e.g. finite set of numbers) they can contain. HMM with variables with continuous space is called *Kalman filter*.[7]

The parameters of the model are transition and emission probability distributions. One of important assumptions of HMM is stationarity, the distributions are constant in time - what is changing are the states. The figure 2.5 shows the structure, transition probabilities are jointly denoted τ and emission probabilities ε.

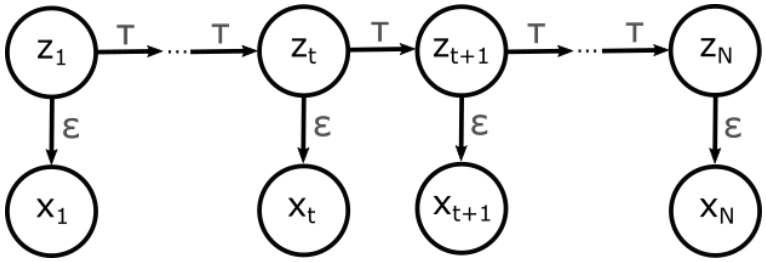


Figure 2.5: Hidden Markov model structure.

Transitional probability

The transitional probabilities are defined between latent variables from time t to time $t+1$ defining the distribution of the transition in each time step. These distributions can be aligned to a *transition matrix*.

Emission probability

The emission probabilities define the observations x_t based on the latent variables z_t. Similarly as the transitional matrix emission distributions can be aligned to *emission matrix*.

Learning of HMM

Analytical approaches of HMM learning are Forward-Backward algorithm and Viterbi. Using transition and emission matrix and the data they compute the probability distributions of the latent variables. Another possible way to learn the parameters of HMM is stochastic using MCMC and numerical optimization connected rather with Bayesian statistics.

Forward-backward The Forward-backward algorithm is used for filtering and smoothing. Filtering uses for estimating of distribution for z_t (state at time t) only observations from the past $x^{0:t}$ and can be used for real time processing. Smoothing estimates distribution of z_t all observations from the $0 : T$, where $t \in 0 : T$. Smoothing cannot be used for real time processing but is more accurate.

Viterbi Neither filtering nor smoothing do not guarrantee that the output will be valid according to the transition and emission matrix, but they simply maximize the total score. The Viterbi algorithm focuses to produce valid output according to the transition and emission

[7] Hidden Markov models and Kalman filters both belong to group of state space models.

matrix, so its output "makes more sense" and it is widely used in some domains, such as natural language processing. For estimation Viterbi uses all the observations from $0 : T$, same as smoothing, but due to the constraints it tends to be less accurate.

Markov Chain Monte Carlo Bayesian approach to HMM uses random simulations from the transition and emission distribution and minimizing of the log likelihood by searching for optimal parameters. There are frameworks that make the method easy to use, such as *Stan* [27].

3 Data

The model is defined (that means its transition and emission matrices) using results of clinical measurements of patients' disease characters: incubation and symptom period, results of molecular and antibody tests, but also tracing information (for estimation of R_0) and others.

The input data for the model are the statistical information of the Covid-19 infection: counts of positively tested individuals, number of deceased on the disease etc.

3.1 Covid-19 Statistics

The statistics are usually reported by national or regional authorities, responsible for publishing them - government institutions or public agencies - ministries of health, statistical offices or regional hygienic offices.

The data can come in daily or weekly records with country-wise, regional, subregional (district-wise) or municipal administrative unit granularity. In seen data sources the attributes are very similar.

Data attributes

Following are the common attributes in the Covid-19 data [28].

Tests First statistics is number of performed tests. This statistic should ideally contain only number of diagnostic tests of individuals done to confirm their infection and repeated tests confirming recovery or antibody tests should be excluded. However some countries publish overall number of all tests performed regardless of type.

Confirmed The tests can be seen as an sample over population and confirmed cases is the number of positive tests. If the probabilities of infected and healthy getting tested are different, the test sample is biased.

The value of confirmed should ideally resemble the total infected, but since tests do not cover the whole population, it is influenced by number of tests and the sample bias.

Deaths Number of deceased is a count of people whose deaths is a direct consequence of infection Covid-19. This statistic has been a subject matter of the dispute (dying with/for Covid-19) as some countries overestimated it by including deceased whose real reason was not Covid-19, but a previous often critical condition or the infection was clearly not connected with their *causa mortis*, e.g. car accident [29, 30, 31].

Hospitalized Another reported number is number of hospitalized patients positive for Covid-19. Hospitalization can be in several modes coming with raising severeness of the infection, the data often contain additional current number of patients at intensive care unit and with connected ventilator, although often researches uses more detailed information of hospitalization.

Decision of hospitalization is often based on patient's state. Many researches are performed on sample of hospitalized people. Thus if asymptomatic and mild cases are eliminated from the sample, measurements on such sample might be skewed.

Prevalence Prevalence is the percentage of infected in population. The number can be either estimated real-time with molecular tests or backwards with antibody testing. The latter allows more careful sample selection and better results. Antibody testing can be under certain conditions performed even during active infection.

Fatality ratio Fatality ratio is a percentage of how deadly an infection is. It is derivated from prevalence and deaths. Dependent on how the prevalence is estimated we distinguish *case fatality ratio* (CFR, equation 3.1) and *infection fatality ratio* (IFR, equation 3.2).

$$\text{CFR} = \frac{\text{Number of deaths}}{\text{Confirmed by tests}} \tag{3.1}$$

Figure 3.1: CFR equation.

$$\text{IFR} = \frac{\text{Number of deaths}}{\text{Truly infected}} \tag{3.2}$$

Figure 3.2: IFR equation.

Recovered Recovered is number of confirmed patients that did undergo the disease and on the given day received first negative test confirming their recovery.

Data sources

Czechia The official data for Czechia are published by MZ ČR[1]. Most statistics cover the whole epidemic (since March 2020) and as for now (4th February) contains (in *daily* time slots) following data attributes [19]:

- **Country**: RT-qPCR + antigen tests
- **District**: deaths, tests, hospital capacities and stock states
  - **Per age group**: incidence, prevalence, hospitalized, vaccinated
  - **Cases with age and gender**: confirmed, deaths

[1] Ministerstvo Zdravotnictví České Republiky/The Ministry of Health of the Czech Republic

- **Municipality**: confirmed

The fetching of the data are implemented in Python package `covid19czechia` [32]. The usage is shown in the listing 3.1.

```
import covid19czechia as CZ
x = CZ.covid_deaths()
```

Listing 3.1: covid19czechia: usage example

Poland The responsible institution to publish the data for Poland is MZ RP[2]. Until October 10 2020, the regional data of confirmed and deaths were published via Twitter as daily updates. Deaths were reported as cases with gender and age. Via government webpage one could only acquire current counts in regions [33].

At the moment regional data between October 10 2020 to November 23 2020 are not published on either of the official sources mentioned. Since November 23 2020, MZ RP started to publish daily a CSV file on their webpage with regional counts (without gender or age information).

Currently data between January 20 2021 to February 28 2021 (today) are missing [34].

- **Country**: tests, recovered, hospitalized, quarantined
- **Region/municipality**: confirmed, deaths

The package `covid19poland` contains data collected both webscraped from Twitter and fetched from the MZ official webpages [20]. The sample code is in the list 3.2.

```
import covid19poland as PL
x = PL.covid_deaths()
```

Listing 3.2: covid19poland: usage example

Sweden Sweden's official Covid-19 statistics are managed by FOHM[3], which publishes information of current situation on weekly basis in PDF reports. FOHM also provides XLSX with more detailed data such as daily deaths, confirmed, intensive care unit cases and applied vaccines, and weekly confirmed and deaths per municipality [35].

- **Country**: deaths, icu, confirmed
- **Region**: icu (weekly), vaccines, tests - antibody
- **Municipality** (weekly): confirmed, deaths

The package `covid19sweden` contains data collected from the XLSX [36].

```
import covid19sweden as SE
x = SE.deaths()
```

Listing 3.3: covid19sweden: usage example

Covid-19 Data Hub Since the Covid-19 data sources are publishing data in different formats, there are many projects collecting and unifying the data to make the access to them easy; *Covid-19 Data Hub* of Guidotti and Ardia used in this thesis is one of them [37].

[2]Ministerstwo Zdrowia Rzeczypospolitej Polskiej/Ministry of Health of the Republic of Poland
[3]Folkhälsomyndigheten/The Public Health Agency of Sweden

Data transformation

SIR model uses values in $[0;1]$, so confirmed and recovered are normalized by test size `tests[t]` and deaths by population size `POP`.

The confirmed incidence is transformed to $[0;1]$ by normalizing by population size.

For various data sources having different data formats certain data transformations unique per each source has to be applied in order to mae a successful modelling on data with unified and homogenous format.

Administrative division The administrative units for regional data are based on what division is used in the data of Covid-19 statistics, published by each of the countries. All 4 countries as EU members have regions with NUTS[4] codes for statistical purposes, table 3.3 specifies what level do the data have [38].

Country	Division	Notes
Czechia	NUTS-3	
Italy		
Poland	NUTS-2	PL91 and PL92 aggregated into PL9.
Sweden	NUTS-3	

Figure 3.3: Administrative divisions used in data.

Timestep size The minimal time step defined in the statistics is a day, although Sweden publishes some data only weekly. As the epidemic is changing slowly, it might be sufficient to estimate the time dependent parameters with fixed-sized windows. This accelerates the computation, but brings additional issues regarding alignment, because some changes might be on the edge of two windows. If the window is small enough, the problem is negligible.

3.2 Demographical Statistics

The demographical data of mortality and deaths have been acquired from the Eurostat.[5] The violinplots of age distribution of mortality per country and gender are shown in the figure 3.4.

Regarding the shape, Poland and Czechia, two countries with communist regimes before 1989, looks quite similar, mean life expectancy is slightly higher for both men and women. Distribution of male mortality in Poland have heavier tail towards younger ages, both male and female in Poland has a small bubble in mortality in age group $0-4$ years. Sweden and Italy are very similar to each other, but different from Poland and Czechia - both male and female life expectancy is higher (*hypothesis test* **TODO**) and there is a lot of women who live longer than 90 years.

TODO: *Visual analysis of the population plot.*

- Italy has slightly older population than other, lot of very old people.
- Sweden has most equally distributed population over ages.

The population distribution by age is shown in the figure 3.5.

[4]Nomenclature of territorial units for statistics (NUTS) is a European standard encoding of regions.
[5]European Statistical Office (Eurostat) is an EU institution responsible for data managing and publishing.

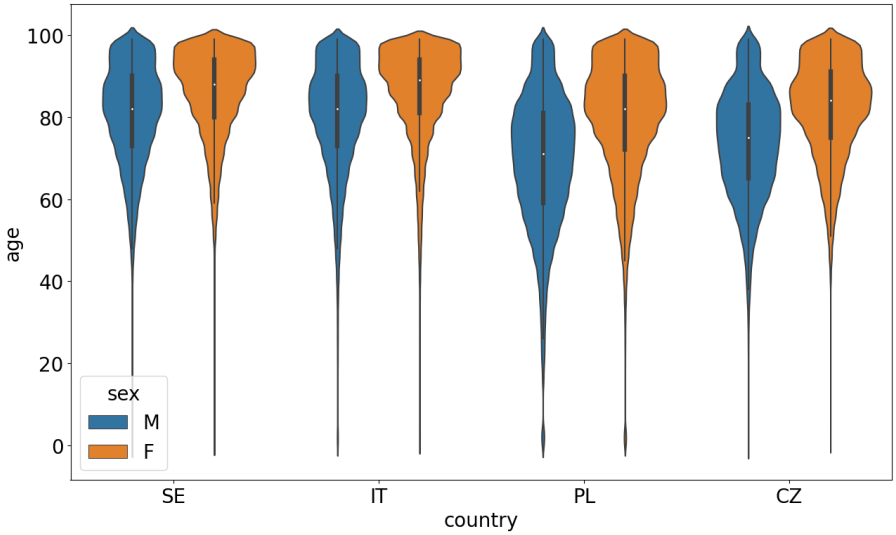


Figure 3.4: Mortality per countries.

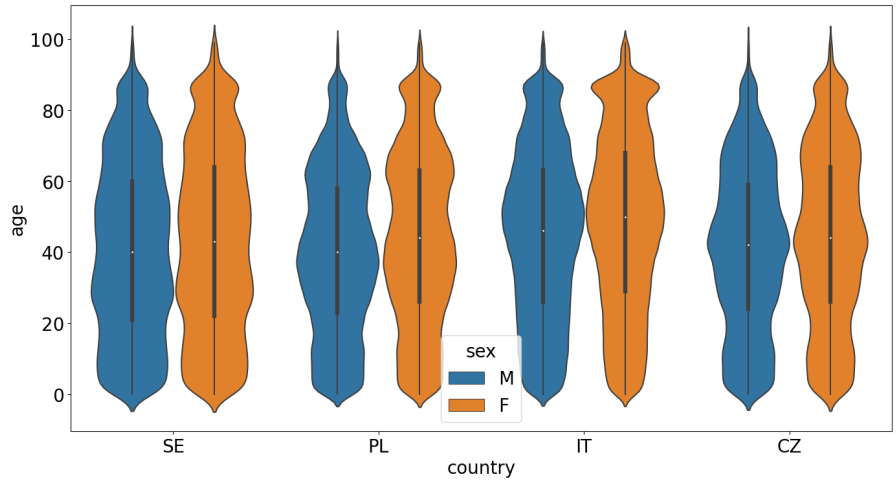


Figure 3.5: Country populations in 2020.

4 Method

4.1 Model

HMM consisting of **transition** and **emission** models. Constructing them requires research of parameters of disease.

SARS-CoV-2 parameters

The characteristics of SARS-CoV-2 infection are needed to be able to model the outbreak. As SEIDR is used, the objectives are distributions of following variables

- Duration of incubation period
- Duration of disease since symptoms
- Reproduction number R_0
- Infection fatality rate (IFR) - investigated in age groups separately

There is several methods to acquire these characteristics

- Clinical measurements = (anonymized) information about hospitalized patients - *incubation period*, *duration of symptoms*
- Antibody tests = presence of antibodies in organism, signs that person had the disease - *prevalence*, *IFR*
- Tracing = reconstruction of the infection transmission graph in the population by detecting contacts of positively tested - *serial interval*, *reproduction number*

Incubation period A research measuring the incubation period length cited even by WHO[1] in precausion recommendation [39] estimates the median incubation to be 5.1 days, although 95% of all cases experiencing $2.2 - 11.5$ days and 50% of all cases experiencing $3.8 - 6.7$ days.

The paper also estimated several parametric distributions to the data, shown in the figure 4.1. The best one fitting to the data selected using lowest MSE of its quantiles to the data quantiles is $\Gamma(5.807, 0.948)$, the results are shown in the table 4.2 [40].

[1] World Health Organization

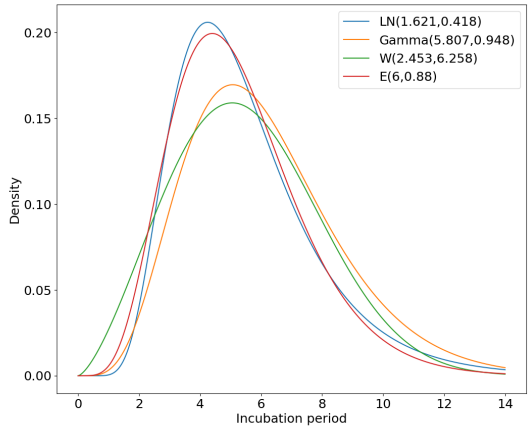


Figure 4.1: Estimated distributions of incubation period length [40].

Distribution	$\mathrm{LN}(1.621, 0.418)$	$\Gamma(5.807, 0.948)$	$\mathrm{W}(2.453, 6.258)$	$\mathrm{E}(6, 0.88)$
MSE	0.438798	0.427651	0.666146	1.022750

Figure 4.2: Incubation period distributions' goodness-of-fit by quantile MSE.

$$P_{X_d}(i) = \int_i^{i+1} f_X(t)\,dt = F_X(i+1) - F_X(i), i = 0, 1, 2, \ldots \tag{4.1}$$

If the distribution is to be modelled in using transition matrix, we need to discretize the distribution to get probability of symptom onset per day since exposure. Using equation 4.1 we get distribution from the figure 4.3.

The domain of the random variable x is limited to $i \in \{0, 1, \ldots, 20\}$, as less than 0.01% of cases had incubation period longer than 20 days. The probabilities can be found in `data/distr/incubation.csv`. Similarly looking distribution was reported by [41] too.

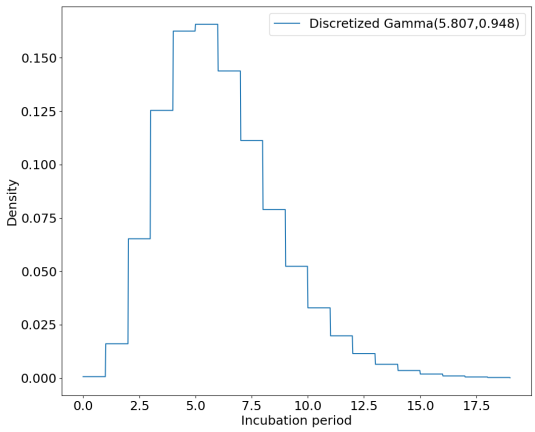


Figure 4.3: Discretized incubation period length distribution.

Disease duration A proper description of disease is far more complicated than just its duration, usually to evaluate disease dynamics, epidemiological research estimates serial interval, attack rate, reproduction number, incubation period and branches the disease based on symptomatic and asymptomatic patients into scenarios.

The model designed for this thesis will simplify the disease dynamic into disease duration, only using two scenarios:

- Symptomatic - infectiousness and symptoms occur at the same time
- Asymptomatic - symptoms do not occur, but infectiousness does. The individuals in this scenario have lower probability to go and get themselves tested for coronavirus.

The scenarios probabilities for a patient depends on age of the patient, symptoms are more likely with older patients and from the literature [42] was created the table 4.4.

Age group	**Asymptomatic**	95% **CI**
Total	0.308	$0.077 - 0.538$
$0 - 15$	0.6	0.4
$16 - 64$	0.45	0.55
$65+$	0.3	0.7

Figure 4.4: Probabilities of scenarios per age groups [42].

Following data shown in the figure 4.5 are 129 hospitalized patients diagnosed with COVID-19. Of those 69% were also at ICU[2] and out of them 91% had to be connected to mechanical ventilation. Immunosuppressed was 23% of the patients. The data contains only hospitalized patients, thus the sample is biased [43].

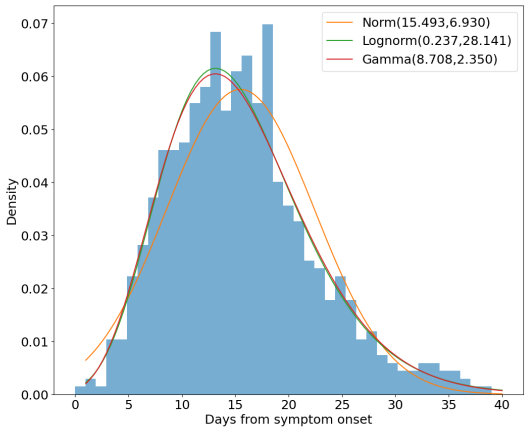


Figure 4.5: Estimated distributions of symptom period length.

The distribution fitting to the data the best seems to be lognormal or gamma. Using AIC from the equation 4.2 that finds the best distribution $\hat{M}$ it is analytically determined, that the best fitting distribution is $\Gamma(4.545, 0.293)$, all results are shown in the table 4.6.

As before, for modelling using transition matrix we discretize the distribution to get daily probabilities using equation 4.1, the result is shown in the figure 4.3 and in the file `data/symptoms.csv`. Similar to ours are also the results of [44].

[2]Intensive care unit

$$\begin{aligned} \text{best model } \hat{M} &\equiv \underset{m \in M}{\operatorname{argmin}} \text{ AIC} \\ \text{AIC} &= 2k - 2 \ln\Big[\text{likelihood}(\cdot)\Big] \end{aligned} \tag{4.2}$$

Distribution	$\mathcal{N}(15.4942, 6.9272^2)$	$log\mathcal{N}(0.5142, 13.8266^2)$	$Gamma(4.545, \frac{1}{3.409})$
AIC	4635.0654	4670.95	4594.3844

Figure 4.6: Duration of symptoms distributions' goodness-of-fit by AIC.

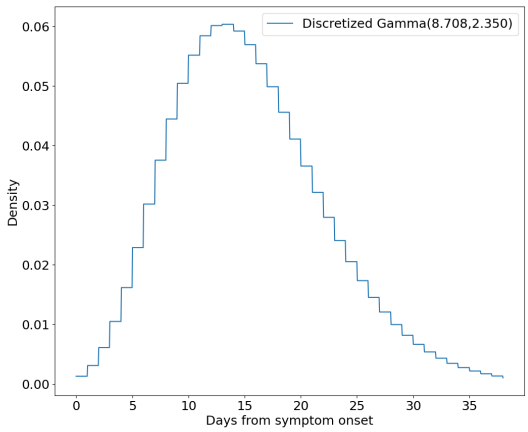


Figure 4.7: Discretized duration of symptoms distribution [43].

Hospitalized cases are either risky patients or patients with severe symptoms of the disease, measuring characteristics only on hospitalized people is biased, as in the case of estimate 4.7) - different publications state duration of symptoms for patients with milder Covid-19 within 10 days [45, 46].

Duration of disease is also dependent on sample type - usually tissue from upper respiratory specimens is used, but it can be measured from various samples: there are studies measuring SARS-CoV-2 presence of Covid-19 positive patients from rectal swabs, where it turns out the viral persistence is longer (than usual nasopharyngeal swab) [47]. In addition, symptoms negatively affecting digestive system has also occurred in some cases [48].

Generation time / serial interval Generation period $w(t)$ is an experimentally measured characteristic of the disease - time between infection of two successive cases. The serial interval on the other hand is the time between symptoms onset of two successive cases [49].

Paper [50] estimates serial interval to be $\Gamma(\alpha, \beta)$ distribution with mean $\mu = 4.55$ and standard deviation $\sigma = 3.3$. When using formula for expected value and variance of gamma distributed random variable, the serial interval has distribution $\Gamma(1.901, 0.41781)$, as shown in the equation 4.3.

The distribution both continuous and per-day discretized is visualized in the figure 4.8.

Reproduction number Basic reproduction number, average count of new infection generated by single infected individual, is a tricky statistic to estimate, as it is computed from

$$\begin{aligned} \text{Serial interval} &\sim \Gamma(\alpha, \beta) \\ E[\text{Serial interval}] &= \frac{\alpha}{\beta} = 4.55 \\ V[\text{Serial interval}] &= \frac{\alpha}{\beta^2} = 3.3^2 = 10.89 \\ \alpha &= 1.901,\ \beta = 0.41781 \end{aligned} \tag{4.3}$$

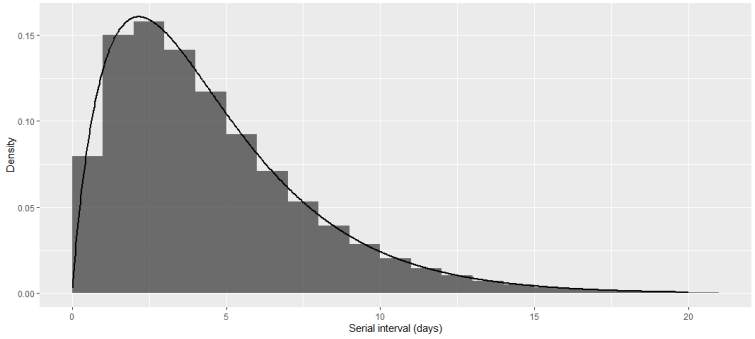


Figure 4.8: Serial interval distribution [50].

incidence, which is also unobserved. There are several methods used in literature that approximates R_0, either time-varying (changing over time) or overall [51].

Their strategy is estimation using clinical measurements of disease characteristics - incubation period, serial interval and infectious period. Most current research papers and WHO estimates overall R_0 of SARS-CoV-2 to be 2 - 4 [52, 53, 50, 54]. However the conditions of the environment (e.g. sufficient precausion of people) change the R_0 significantly.

The method introduced by [55] is minimizing objective function from the equation 4.4. Here p_{ij} denotes probability, that case i was caused by case j, t_i and t_j are time points (days) of infection of i and j respectively.

$$\begin{aligned} p_{ij} &= \frac{w(t_i - t_j)}{\sum_{i \neq k} w(t_i - t_k)} \\ R_j &= \sum_i p_{ij} \end{aligned} \tag{4.4}$$

The above mentioned algorithm is implemented by R package `EpiEstim` and was used to estimate R0 over the incidence confirmed by tests in the data. The result aggregated per months is shown in the figure 4.9.

Infection fatality rate Fatality rates are estimated from prevalence and death counts. While case fatality ratio (CFR) uses molecular test results and thus can be estimated in real time with the disease, infection fatality ratio uses true prevalence, measured with antibody testing, that is in case of Covid-19 mostly performed several weeks after the patient's recovery to be reliable, although it was shown that AgM test can turn positive already during the infection [24, 56]. The values estimated by [57] are shown in the table 4.10. Very similar results are presented by [58] and [59].

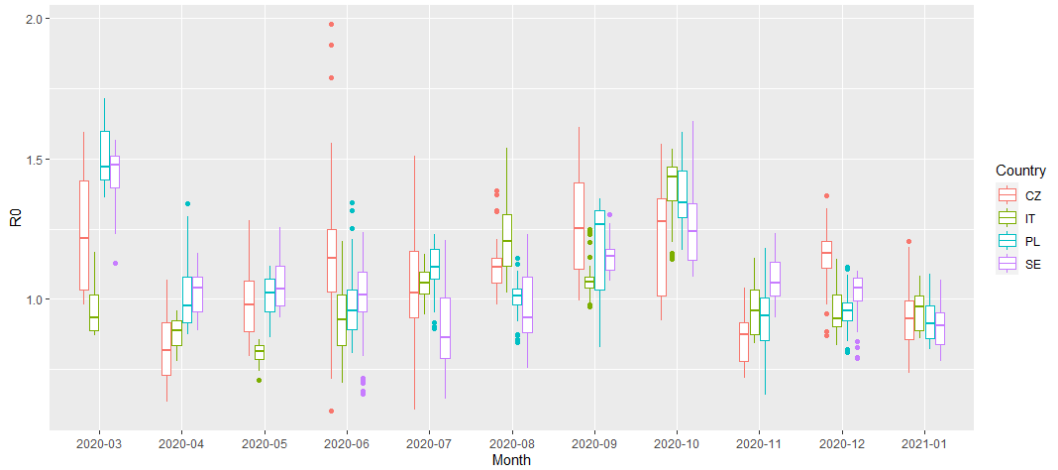


Figure 4.9: $R0$ monthly estimates using confirmed cases.

Age group	IFR estimate	Credible interval
$5-9$	0.0016	$[0; 0.019]$
$10-19$	0.00032	$[0; 0.0033]$
$20-49$	0.0092	$[0.0042; 0.016]$
$50-64$	0.14	$[0.096; 0.19]$
$>= 65$	5.6	$[4.3; 7.4]$
Total	0.64	$[0.38; 0.98]$

Figure 4.10: Infection fatality rate estimates [57].

Transition model

The transition model denotes the transition of latent variables between times d and $d+1$. Model parameters can be either defined as a scalar (meaning single value for given time point) or a vector if the parameter value differ significantly for different groups - either duration of infectiousness for both symptomatic and asymptomatic progress of disease.

The disease progress is projected in the transition model using an SEIRD model. If a Bayesian definition is used, vector parameters, e.g. incubation, infection or immunity periods, are expressed as random variables with appropriate prior distributions. The structure of the model including the transition parameters a, c, b, d is shown in the figure 2.4.

Parameters can be also time-dependent, which takes into account that pandemic characteristics change over time. To lower the computational costs, a time unit for parameter values might differ from the data time unit, for $n = 7$ shown in the figure 4.11.

Figure 4.11: Parameter time slots, $n = 7$.

Priors for parameters a, b, c, d can be estimated using the clinically measured characteristics of Covid-19. The formulas for parameters come from the SEIRD model.

Incubation period Parameter c_t carries the information about incubation period in a form of a probability of transition $E \rightarrow I$. Incubation period is derived from c_t by the equation 4.5.

$$\text{Incubation period} \overset{d}{=} c_t^{-1} \tag{4.5}$$

Samples of c_t are acquired using a simulation from the incubation period distribution and follow-up transformation defined by the equation 4.5. Then, the distribution fitting to the samples is used as a prior. In case of parameter c representing transition $E \rightarrow I$ it turns out to be distribution from equation 4.6. The samples and the distribution are shown in the figure 4.12.

$$\frac{c - 0.035}{2.545} \sim Beta(3.478, 51.059) \tag{4.6}$$

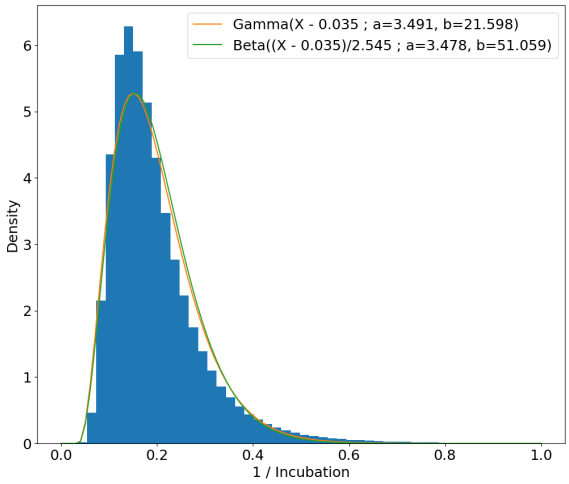


Figure 4.12: Parameter c estimated.

Duration of symptoms Parameter b_t stands for probability of recovering $P(I \rightarrow R)$ and it is connected to duration of symptoms. If deaths are excluded and only surviving infected (so called recovered), we can define the parameter b_t according to the equation 4.7.

$$\text{Symptom duration of recovered} \overset{d}{=} b_t^{-1} \tag{4.7}$$

Infection fatality rate Parameter d_t represents probability $P(I \rightarrow D)$ is directly connected with b_t. In order to relate both d_t and b_t to the symptom duration and each other, the infection fatality rate (IFR), whose distribution has been measured as well, is used. These distributions are related to known disease characteristics as specified in the equation 4.8.

$$\begin{aligned} b_t &\overset{d}{=} (1 - \text{IFR}) \cdot \text{Symptom duration}^{-1} \\ d_t &\overset{d}{=} (\text{IFR}) \cdot \text{Symptom duration}^{-1} \end{aligned} \tag{4.8}$$

Similarly as before, samples of b_t and d_t are produced by simulation from the distributions in the equation 4.8. Then the distribution is fitted to the draws as shown in the figures 4.13 and 4.14, the final distributions are specified by the equation 4.9.

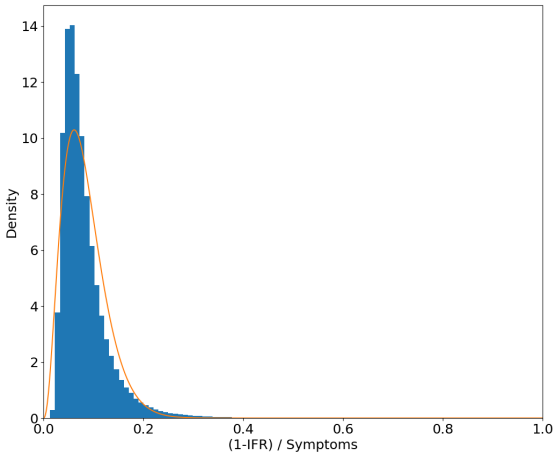


Figure 4.13: Estimate for parameter b.

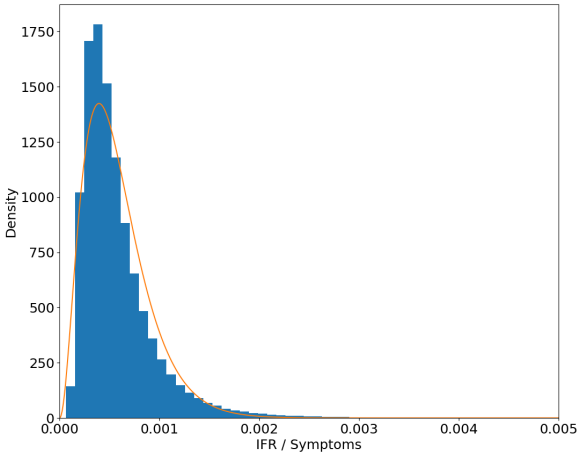


Figure 4.14: Estimate for parameter d.

$$\begin{aligned} b &\sim Beta(3.1514, 37.7134) \\ d &\sim Beta(3.4024, 5438.4883) \end{aligned} \tag{4.9}$$

Reproduction number Parameter a_t represents the infection rate, the probability of transition $S \to E$ and is directly connected to reproduction number as specified by the equation 4.10.

$$\text{Reproduction}_t \overset{d}{=} \frac{a_t}{d_t + b_t} \tag{4.10}$$

Together equations 4.10 and 4.8 imply that the distribution of a_t defined according to equation 4.11.

$$a_t \overset{d}{=} \frac{\text{Reproduction}_t}{\text{Symptom duration}_t} \tag{4.11}$$

The best fitting distribution to draws of parameter a is specified by the equation 4.12 and shown in the figure 4.15.

$$a \sim Weibull(1.836352, 0.365743) \tag{4.12}$$

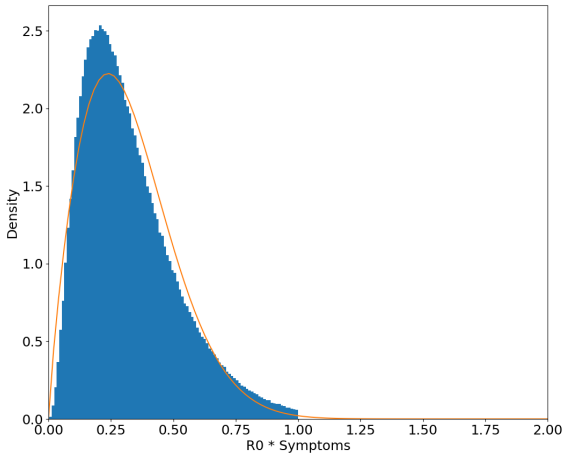


Figure 4.15: Estimate for parameter a.

Emission model

Active infection is measured by tests, certain percentage of whose turns out to be positive. Simplest distribution to model percentage of positively tested individual is Bernoulli, as shown in equation 4.13 with parameter p interpreted as ratio of positive tests.

$$\begin{aligned} \text{Infected} \mid \text{Tested} &\sim Bernoulli(p) \\ P(\text{Infected} = x_i \mid \text{Tested} = p_i) &= p_i^{x_i}(1-p_i)^{1-x_i} \end{aligned} \tag{4.13}$$

The ratio of positive tests over months for each of the countries is shown in the figure 4.16.

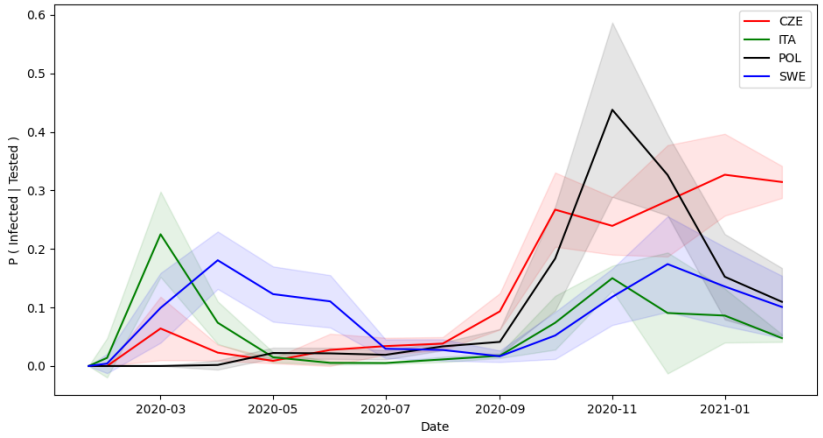


Figure 4.16: Ratio of positive tests over time in Poland, Sweden, Czechia and Italy.

The prior is modelled with Beta distribution as shown in the equation 4.14.

$$\text{Tested} \sim Beta(\alpha, \beta),\ \alpha, \beta > 0 \tag{4.14}$$

The prior is constructed based on ratio of performed tests in population $\tau = \frac{\#\text{Tests}}{\#\text{Population}}$ and

Posterior On given day in given administrative unit with population N there are N' infected people. A test is performed on T-sized sample $\{x_1, ..., x_T\}$ ($x_i \in \{0, 1\}$), T' infected individuals are confirmed. Ratio of positive tests from data $\frac{T'}{T}$ is denoted $\bar{x}$. It is possible to find known distribution of posterior $Infected|Tested$ since Beta is a conjugate prior for Bernoulli model, as proven by the equation 4.15.

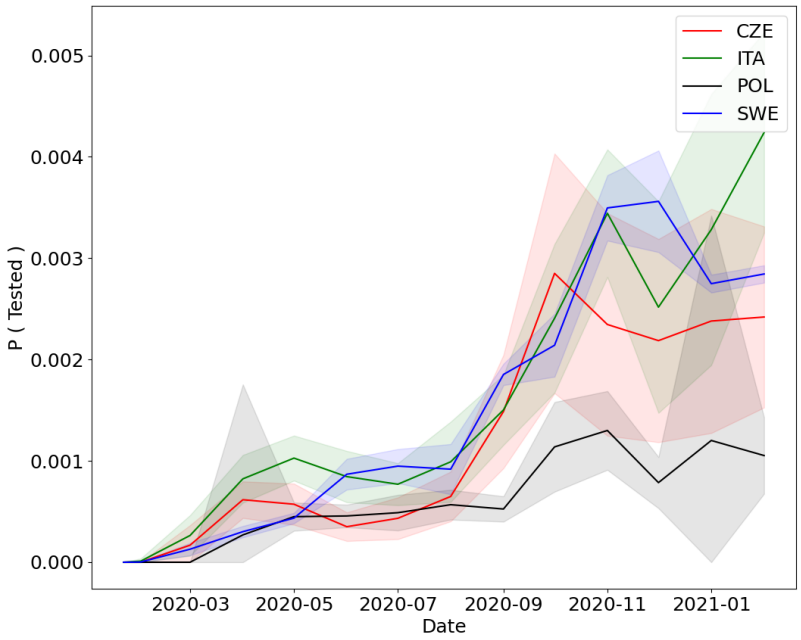


Figure 4.17: Ratio of tests over population in Poland, Sweden, Czechia and Italy.

$$
\begin{aligned}
&P(Tested = p|Infected = \vec{x}) \overset{\perp}{\propto} P(Tested)\prod_{i=1}^{N} P(Infected = x_i|Tested) = \\
&= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} p^{\alpha-1}(1-p)^{\beta-1} p^{\sum_{i=1}^{N} x_i}(1-p)^{N-\sum_{i=1}^{N} x_i} \propto \\
&\propto p^{\alpha-1}(1-p)^{\beta-1} p^{N\bar{x}}(1-p)^{N(1-\bar{x})} = \\
&= p^{(\alpha+N\bar{x})-1}(1-p)^{(\beta+N-N\bar{x})-1} \propto Beta(\alpha' = \alpha + N\bar{x}, \beta' = \beta + N - N\bar{x})
\end{aligned}
\tag{4.15}
$$

5 Results

6 Discussion

6.1 Results

6.2 Method

Estimation of reproduction number

Ratio of positive tests in test sample as already said does not reflect the true prevalence of the infection of the population, as it depends on the number of performed tests and testing strategy. To evaluate the reliability of the R0 estimated, we might use the same method to estimate R0 using daily performed tests treated as incidence input. A simple correlation of the output with the R0 estimate gives a good picture about dependence of the estimate on the number of tests.

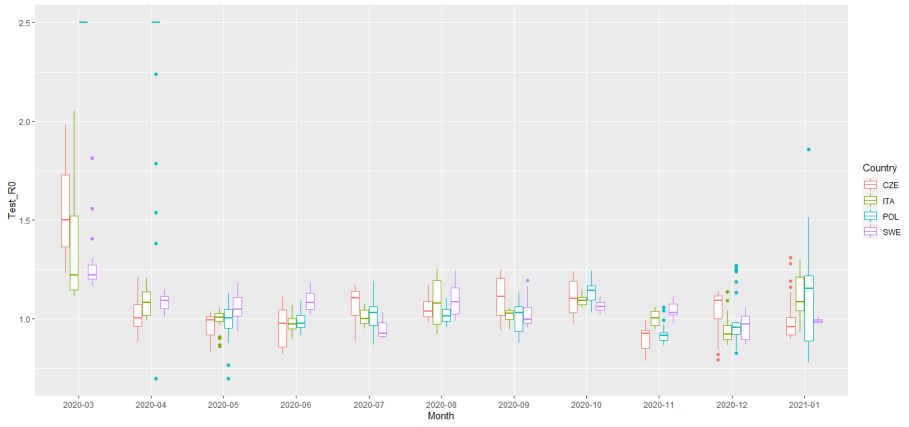
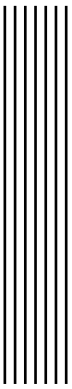


Figure 6.1: R0 estimated on tests (treated as incidence).

6.3 The work in a wider context

7 Conclusion



Bibliography

[1] Bundesministerium für Finanzen. *Pressekonferenz am 14.03.2020*. Mar. 2020. URL: https://www.flickr.com/photos/159530260@N03/49657522158/.

[2] Radek Kosarzycki. *Teraz już nikt nieprawidłości nie znajdzie. Rząd zmienił sposób raportowania zakażeń koronawirusem*. Nov. 2020. URL: https://spidersweb.pl/2020/11/mz-zmiana-raportowania-nowych-zakazen.html.

[3] A. Mouritz. *"The Flu" A brief history of influenza in U.S. America, Europe, Hawaii*. 1921.

[4] Owen Jarus. *Gruesome Find: 100 Bodies Stuffed into Ancient House*. July 2015. URL: https://www.livescience.com/51662-100-bodies-found-prehistoric-house.html.

[5] Jenny Howard. *Plague was one of history's deadliest diseases—then we found a cure*. July 2020. URL: https://www.nationalgeographic.com/science/article/the-plague.

[6] John Horgar. *Justinian's Plague (541-542 CE)*. Dec. 2014. URL: https://www.ancient.eu/article/782/justinians-plague-541-542-ce/.

[7] B. Bramanti and K. et.al. Dean. *The Third Plague Pandemic in Europe*. Apr. 2019. DOI: 10.1098/rspb.2018.2429.

[8] C. Potter. "A history of influenza". In: *Journal of Applied Microbiology* 91.4 (2001), pp. 572–579. DOI: 10.1046/j.1365-2672.2001.01492.x.

[9] Paul A. Blake. *Historical Perspectives on Pandemic Cholera*. Mar. 1994. DOI: 10.1128/9781555818364.ch18.

[10] Didier Raoult, Theodore Woodward, and J.Stephen Dumler. "The history of epidemic typhus". In: *Infectious Disease Clinics of North America* 18.1 (2004). Historical Aspects of Infectious Diseases, Part I, pp. 127–140. ISSN: 0891-5520. DOI: 10.1016/S0891-5520(03)00093-X.

[11] J. et.al. Breman. *The Confirmation and Maintance of Smallpox Eradication*. 1980.

[12] A. Barrett and S. Higgs. "Yellow Fever: A Disease that Has Yet to be Conquered". In: 52 (Jan. 2007), pp. 209–229. DOI: 10.1146/annurev.ento.52.110405.091454.

[13] Gelband H. Panosian C. *Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance*. 2004. DOI: 10.17226/11017.

[14] R. Acuna-Soto and D. et.al. Stahle. "Megadrought and Megadeath in 16th Century Mexico". In: (2004). DOI: 10.3201/eid0804.010175.

[15] Edwin D. Kilbourne. "Influenza Pandemics of the 20th Century". In: 12 (Jan. 2006), pp. 9–14. DOI: 10.3201/eid1201.051254.

[16] K. Kuszewski and L. Brydak. "The epidemiology and history of influenza". In: *Biomedicine Pharmacotherapy* 54.4 (2000), pp. 188–195. ISSN: 0753-3322. DOI: 10.1016/S0753-3322(00)89025-3.

[17] World Health Organization. "HIV/AIDS". In: (Nov. 2020). URL: https://www.who.int/news-room/fact-sheets/detail/hiv-aids.

[18] Nanographics GmbH. *High resolution renderings of SARS-CoV-2 Cryo-ET*. 2021. URL: https://nanographics.at.

[19] Ministry of Health of Czechia. *Onemocnění aktuálně*. 2021. URL: https://onemocneni-aktualne.mzcr.cz/covid-19.

[20] Martin Beneš. *Covid19Poland: Web Scraper of COVID-19 data for Poland*. 2021. URL: https://pypi.org/project/covid19poland.

[21] Instituto Nazionale di Statistica. *Impact of Covid-19 Epidemic on Mortality: Causes of Death in Covid-19 Laboratory Confirmed Cases*. July 2020. URL: https://dc-covid.site.ined.fr/en/data/italy/.

[22] European Centre for Disease Prevention and Control. *Clinical characteristics of COVID-19*. Aug. 2020. URL: https://www.ecdc.europa.eu/en/covid-19/latest-evidence/clinical.

[23] M Cascella, M Rajnik, and et. al. Cuomo A. *Features, Evaluation, and Treatment of Coronavirus*. 2020. URL: https://www.ncbi.nlm.nih.gov/books/NBK554776/.

[24] Robert Kubina and Arkadiusz Dziedzic. "Molecular and Serological Tests for COVID-19. A Comparative Review of SARS-CoV-2 Coronavirus Laboratory and Point-of-Care Diagnostics". In: *Diagnostics* 10.6 (2020), p. 434. DOI: 10.3390/diagnostics10060434.

[25] Yutaka Okabe and Akira Shudo. "A Mathematical Model of Epidemics—A Tutorial for Students". In: *Mathematics* 8.7 (2020). ISSN: 2227-7390. DOI: 10.3390/math8071174. URL: https://www.mdpi.com/2227-7390/8/7/1174.

[26] Adam Kucharski. *The Rules of Contagion*. 2020, p. 352. ISBN: 978-1-78816-019-3.

[27] Stan Development Team. *Stan*. Feb. 2021. URL: https://mc-stan.org/.

[28] Bernd Kamps and Christian Hoffmann. *COVID Reference*. Steinhäuser Verlag, Jan. 2021. ISBN: 978-3-942687-53-9. URL: https://amedeo.com/CovidReference06.pdf.

[29] AFP. *UK 'overestimates' coronavirus death toll: Study*. July 2020. URL: https://health.economictimes.indiatimes.com/news/diagnostics/uk-overestimates-coronavirus-death-toll-study/77034966.

[30] Nina Schwalbe. *We could be vastly overestimating the death rate for COVID-19. Here's why*. Apr. 2020. URL: https://www.weforum.org/agenda/2020/04/we-could-be-vastly-overestimating-the-death-rate-for-covid-19-heres-why/.

[31] Patrick Laurie Davis. "Excess deaths, baselines, Z-scores, P-scores and peaks". In: (Oct. 2020). URL: https://arxiv.org/abs/2010.10320.

[32] Martin Beneš. *Covid19Czechia: Web Scraper of COVID-19 data for Czechia*. 2021. URL: https://pypi.org/project/covid19czechia.

[33] Ministry of Health of the Republic of Poland. *Oficjalny profil Ministerstwa Zdrowia*. 2020. URL: https://twitter.com/MZ_GOV_PL.

[34] Ministry of Health of the Republic of Poland. *Koronawirus: informacje i zalecenia*. 2021. URL: https://www.gov.pl/web/koronawirus/pliki-archiwalne-powiaty.

[35] The Public Health Agency of Sweden. *Statistics and analyses*. 2021. URL: https://www.folkhalsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/statistik-och-analyser/.

[36] Martin Beneš. *Covid19Sweden: Web Scraper of COVID-19 data for Sweden*. 2021. URL: https://pypi.org/project/covid19sweden.

[37] Emanuele Guidotti and David Ardia. "COVID-19 Data Hub". In: *Journal of Open Source Software* 5.51 (2020), p. 2376. DOI: 10.21105/joss.02376.

[38] Eurostat. *NUTS Maps*. URL: https://ec.europa.eu/eurostat/web/nuts/nuts-maps.

[39] World Health Organization. *Transmission of SARS-CoV-2: implications for infection prevention precautions*. July 2020. URL: https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions.

[40] Stephen A Lauer, Kyra H Grantz, and Qifang et.al. Bi. "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application". In: *Annals of Internal Medicine* 172.9 (2020), pp. 577–582. DOI: 10.7326/M20-0504.

[41] Xuan Jiang, Simon Rayner, and Min-Hua Luo. "Does SARS-CoV-2 has a longer incubation period than SARS and MERS?" In: *Journal of Medical Virology* 92.5 (2020), pp. 476–478. DOI: 10.1002/jmv.25708.

[42] Hiroshi Nishiura and Tetsuro et.al. Kobayashi. "Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19)". In: *International Journal of Infectious Diseases* 94 (May 2020), pp. 154–155. DOI: 10.1016/j.ijid.2020.03.020.

[43] Jeroen J. A. Van Kampen, David A. M. C. Van De Vijver, and Pieter L. A. et al. Fraaij. "Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19)". In: *Nature Communications* 12.1 (Jan. 2021). DOI: 10.1038/s41467-020-20568-4.

[44] Qingxian Cai, Deliang Huang, and Pengcheng et.al. Ou. "COVID-19 in a designated infectious diseases hospital outside Hubei Province, China". In: *Allergy* 75.7 (2020), pp. 1742–1752. DOI: 10.1111/all.14309.

[45] Roman Wölfel and Victor M. et.al. Corman. "Virological assessment of hospitalized patients with COVID-2019". In: *Nature* 581 (May 2020), pp. 465–469. DOI: 10.1038/s41586-020-2196-x.

[46] Centers for Disease Control and Prevention. "Duration of Isolation and Precautions for Adults with COVID-19". In: (Oct. 2020). URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html.

[47] Yi Xu, Bing Zhu, and Huiying et.al. Liang. "Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding". In: *Nature* 26 (2020), pp. 502–505. DOI: 10.1038/s41591-020-0817-4.

[48] Chaoqun Han, Caihan Duan, and Shengyan et.al. Zhang. "Digestive Symptoms in COVID-19 Patients With Mild Disease Severity: Clinical Presentation, Stool Viral RNA Testing, and Outcomes". In: *The American journal of gastroenterology* 115 (2020), pp. 916–923. DOI: 10.14309/ajg.0000000000000664.

[49] Sonja Lehtinen, Peter Ashcroft, and Sebastian Bonhoeffer. "On the relationship between serial interval, infectiousness profile and generation time". In: (Jan. 2021). ISSN: 1742-5662. DOI: 10.1098/rsif.2020.0756.

[50] Mohammad Aghaali, Goodarz Kolifarhood, and Roya et.al. Nikbakht. *Estimation of the serial interval and basic reproduction number of COVID-19 in Qom, Iran, and three other countries: A data-driven analysis in the early phase of the outbreak*. June 2020. DOI: 10.1111/tbed.13656.

[51] Chong You, Yuhao Deng, Wenjie Hu, Jiarui Sun, Qiushi Lin, Feng Zhou, Cheng Heng Pang, Yuan Zhang, Zhengchao Chen, and Xiao-Hua Zhou. "Estimation of the time-varying reproduction number of COVID-19 outbreak in China". In: *International Journal of Hygiene and Environmental Health* 228 (2020), p. 113555. ISSN: 1438-4639. DOI: 10.1016/j.ijheh.2020.113555.

[52] World Health Organization. *Coronavirus disease 2019 (COVID-19): Situation Report – 46*. Mar. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf.

[53] Kevin Linka, Mathias Peirlinck, and Ellen Kuhl. *The reproduction number of COVID-19 and its correlation with public health interventions*. 2020. DOI: 10.1007/s00466-020-01880-8.

[54] F Najafi, N. Izadi, and S.-S. et.al Hashemi-Nazari. *Serial interval and time-varying reproduction number estimation for COVID-19 in western Iran*. June 2020. DOI: 10.1016/j.nmni.2020.100715.

[55] Jacco Wallinga and Peter Teunis. "Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures". In: *American Journal of Epidemiology* 160.6 (Sept. 2004), pp. 509–516. ISSN: 0002-9262. DOI: 10.1093/aje/kwh255.

[56] World Health Organization. *Estimating mortality from COVID-19*. Aug. 2020. URL: https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19.

[57] Javier Perez-Saez, Stephen A Lauer, and Laurent et.al. Kaiser. "Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland". In: (July 2020). DOI: 10.1016/S1473-3099(20)30584-3.

[58] John P A Ioannidis. "Infection fatality rate of COVID-19 inferred from seroprevalence data". In: *Bulletin of the World Health Organization* 99.1 (Oct. 2020), pp. 19–33. DOI: 10.2471/blt.20.265892.

[59] Gideon Meyerowitz-Katz. "A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates". In: *Intenational Journal of Infectious Diseases* 101 (Dec. 2020), pp. 138–148. DOI: 10.1016/j.ijid.2020.09.1464.