SEIR Filter: A Stochastic Model of Epidemics

Martin Šmíd, Luděk Berec, Aleš Antonín Kuběna, Jan Trnka, Vít Tuček, Milan Zajíček

1. Introduction

Whereas fundamental rationale behind mainstream epidemic models, dating back to the works of Ross (X) and Kermack and McKendrick (X) about a century ago, is relatively easy to comprehend, problems arise when we aim to parameterize such models to data for specific epidemics. The source of troubles is here not only absence of clear and consistent protocols for data collection, but also factors that can never be fully intercepted, such as degrees of compliance in many adopted interventions. This is the case also for COVID-19, despite the somewhat paradoxical observation that we now arguably have the best data on any epidemic in the history.

Regarding COVID-19, one of the hottest topics has concerned quantifying effects of various non-pharmaceutical interventions (X, X, X). Related to this are then the problems of how effective testing and tracing need be to compensate for an intervention relaxation (X) and how stringent an intervention should be to compensate for a degree of non-compliance potentially related to it (X). Answering these questions, models correctly handling uncertainty are necessary.

Upon construction of realistic models, various data issues have to be coped with, starting from noise (caused both by the epidemic process itself and the data collection), insignificance (difficulty to distinguish impacts of factors from random fluctuations) or co-linearity (difficulty to distinguish two parameters with sufficient certainty); the most severe difficulty, however, is that relevant data (e.g. numbers of infected) are hidden, are observed only indirectly (through the numbers of positive tests, for instance). Obviously, once any of these phenomena are handled insufficiently, models can provide wrong policy recommendations.

Mathematical statistics has developed tools to handle these issues. However, to our best knowledge, there is no work systematically doing so for compartmental epidemic models. The goal of this paper is to start filling this gap by proposing a general stochastic epidemic modelling framework.

Stochastic epidemic models do obviously exist. Apart from agent-based models (X), they are commonly formulated as Poisson processes (X), branching processes (X) or stochastic differential equations (X). However, many of these studies are theoretical, examining their behavior without providing any link to real data and hence an estimation procedure. On the other hand, many other studies use elaborate filtering methods to estimate parameters of their epidemic

models, but the models are largely deterministic; statistical models here serve just as tool on the way to understand specific epidemics (X, X, X). Here we in a sense bridge these two approaches, precisely formulating and analyzing a general stochastic epidemic model, discussing its highly practical implications, as well as providing an estimation procedure exemplified on the COVID-19 epidemic in the Czech Republic. Due to practical reasons, we develop our model as discrete in both time and state space. Although applicable to a wide class of epidemic models, we call our framework the SEIR Filter, after a model type most appropriate for the current epidemic of COVID-19.

LUDEK: POssible references

[3] diffusion model

[4] markov model + plus estimation procedure together with an estimation procedure is developed.

We show that contrary to many other fitting methods, our framework enables model parameters to be estimated in a straightforward way, as the exact likelihood (or other estimating) function can be derived, thus allowing faster and more reliable parameter estimation. In addition to the tractability of the estimating function, our model allows for closed form formulas for expected future compartment sizes and the reproduction number. In addition, we provide simple criteria for vanishing and explosion of the epidemic, as well as for bounds that limit expected epidemic sizes given stationary imports. As we demonstrated by stylized examples, many applications of the model are possible, including derivation of implicit formulas regarding compensation measures to intervention relaxation or non-compliance, and of optimal control strategies.

As we have premised, we regard discrete-time discrete-space stochastic models as the most practical. One of the models closest to our one is that by [6] (or [7]). When estimating its parameters, the authors use Monte Carlo simulation to evaluate likelihood functions. We argue and demonstrate that the parameters of these models can be estimated more straightforward way, as the likelihood (or other estimating) function can be computed exactly, allowing for quicker and more reliable estimation.

From the point of mathematical statistics, we model the epidemics (possibly together with some ralated variables) by a partially observed inhomogeneous heteroskedastic vector autoregression process. In line with the usual practice [**?**], we assume over-dispersed probability distributions both of the infections and the compartment transitions; thanks to this, we are able to handle realistically not only the actual values of interest, but also the uncertainty associated with them. Moreover, having a standard statistical output of the parameters estimation, we are able to answer questions concerning significance (via P-values), possible co-linearity (via the estimator's correlations) and the hidden compartments values (via the estimate of state-space distribution).

To demonstrate usefulness of our framework, we apply it to the actual COVID-19 pandemic data from the Czech Republic. In particular, we consider four age cohorts, and we create model for incidence, admission to- and release from hospitals and deaths. To estimate the model's parameters, we use several partially overlapping datasets; ability to create statistically correct estimates

based on them is one of the greatest advantages of our model. We demonstrate both in-sample and out-of-sample prediction ability of our model and, as an example of its possible use, we compare three strategies of vaccination: no vaccination, vaccination without preference and the oldes first strategy.

The paper is organized as follows. After a rigorous probabilistic formulation of the model (Section 2), we discuss its basic probabilistic properties (Section 3), its autonomous sub-models and reproduction number (Section 4) and asymptotic properties (Section 5). Next, we introduce an age-cohort version of the model (Section 6) and suggest a way of optimal control of the epidemics (Section 7). Further, we discuss estimation of the model (Section 8). Next, weand demonstrate its usage by its application to the COVID pandemics in Czech Republic 2020 (Section **??**). Finally, we conclude the paper (Section **??**).

2. Model Definition

Assume a population of size $s \in \mathbb{N}$, where s is large. Each individual of the population is either susceptible, or finds himself in one of the compartments $S_1, \dots, S_k$. Let $I_t \in \mathbb{N}_0^k$, $t \in \mathbb{N}_0^+$, be a possibly hidden external inflow (import) of individuals into the compartments and let $Z_t \in \mathbb{R}^p$, $t \in \mathbb{N}_0^+$, be an observed exogenous process.

For any $t \in \mathbb{N}_0$, let $X_t = (X_t^1, \dots, X_t^k) \in \mathbb{N}_0^k, t \in \mathbb{N}_0$, be a possibly hidden stochastic process of the compartment sizes which we define later. Let

$$Y_t \in \mathbb{R}^n, \qquad Y_t = FX_t + \epsilon_t, \qquad t \in \mathbb{N}_0,$$

be a process of observations where F is a deterministic $n \times k$ matrix with rank n, and ϵ_t is a random errors vector.

Denote $(\mathcal{F}_t)_{t\geq 0}$ and $(\mathcal{G}_t)_{t\geq 0}$ the filtrations induced by (X, Y, I, Z), by (Y, Z), respectively – these filtrations may be seen as information flows, $(\mathcal{F}_t)$ representing all the information and $(\mathcal{G}_t)$ the observable one.

We assume that

$$\mathbb{E}(\epsilon_{t+1}|\mathcal{F}_t) = 0, \qquad \operatorname{var}(\epsilon_{t+1}|\mathcal{F}_t) = \operatorname{diag}(\Gamma_t(X_t, X_t^2)), \qquad t \in \mathbb{N}_0,$$

where Γ_t is a $\mathcal{G}_t$-measurable affine linear function (i.e. $\Gamma_t(x, y) = \gamma_{t,0} + \gamma_{t,1}x + \gamma_{t,2}y$ for some $\gamma_{t,0} \in \mathbb{R}^k$ and $\gamma_{t,1}, \gamma_{t,2} \in \mathbb{R}^{k\times k}$ where all $\gamma_{t,0}, \gamma_{t,1}$ and $\gamma_{t,2}$ are $\mathcal{G}_t$-measurable).

We define X recursively: We let X_0 to be a possibly random vector and, for any $t \in \mathbb{N}$, we put

$$X_{t+1} = I_t + N_{t+1} + M_{1,t+1} + \cdots + M_{k,t+1}.$$

Here, $N_{t+1} \in \mathbb{N}_0^k$ is the inflow of domestically infected individuals such that $N_{t+1}|\mathcal{F}_t \sim \mathrm{CPo}(A_tX_t, L)$ where $A_t = (\alpha_t^{ij})_{1\leq i,j\leq k}$, $A_t \in \mathcal{G}_t$ is a random matrix (the notation $A_t \in \mathcal{G}_t$ means that A_t is a $(\mathcal{G}_t)$-adapted process) and, for any

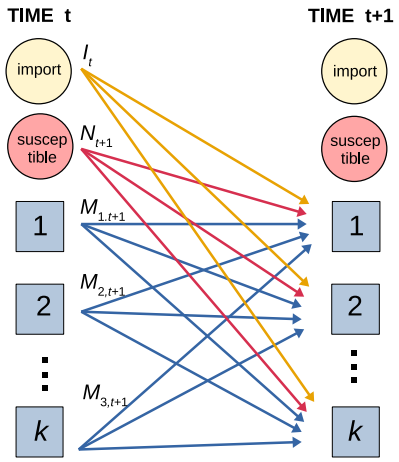
vector x, $\mathrm{CPo}(x, L)$ stands for a vector of independent Compound Poisson variables with the intensities given by x and the embedded distribution L. Observe that, by basic properties of Compound Poisson distribution,

$$\frac{\mathrm{var}(N^i_{t+1}|\mathcal{F}_t)}{\mathbb{E}(N^i_{t+1}|\mathcal{F}_t)} = v \stackrel{\mathrm{def}}{=} \frac{\mathbb{E}L^2}{\mathbb{E}L}$$

(with $\frac{0}{0} \stackrel{\mathrm{def}}{=} \frac{\mathbb{E}L^2}{\mathbb{E}L}$), $t \geq 0, 1 \leq i \leq k$. Further, for any i, $M_{i,t+1} \in \mathbb{N}_0^k$ is the distribution[1] of those, who found themselves in compartment i at t, between the compartments at $t+1$. In particular, $M^j_{i,t+1}$ (the j-th component of $M_{i,t+1}$) is the number of the individuals who transited from the compartment j to the compartment i between t to $t+1$. Further we assume $M_{i,t+1}$, for any i, to follow Dirichlet Multinomial (DM) conditional distribution with parameters $\left(X^i_t, \frac{p^{1,i}_t}{c_i}, \ldots, \frac{p^{k,i}_t}{c_i}\right)$ where $P_t = (p^{ij}_t)_{1\leq i,j\leq k} \in \mathcal{G}_t$ is a "mean" transition matrix and $(c_1, \ldots, c_k)$ are deterministic dispersion parameters.

Remark 1. If we assumed the individuals to change their state according to P_t with the transitions being conditionally independent given $\mathcal{F}_t$, then we would get $M_{i,t+1}|\mathcal{F}_t$ as Multinomial with parameters X^i_t and $P^{(i)}_t$ where, for any matrix Σ, $\Sigma^{(i)}$ is its i-th column. In practice, however, the assumption of the conditional independence, is unrealistic, as the course of infection differs between individuals and (due to mutations) tends to cluster. The standard way of coping with this situation is using DM distribution.

The flow of individuals between states are illustrated by the following chart.



Finally, we assume $N_{t+1}, M_{1,t+1}, \ldots, M_{k,t+1}, \epsilon_{t+1}$ to be mutually conditionally independent given $\mathcal{F}_t$ (which, in words, means that all the dependence

[1] Here, we do not mean a probability distribution, but the numbers of individuals having ended in individual compartments.

between the inflows, the transitions and the observation can be explained by the state of the system at t). Consequently,

$$X_{t+1}|\,\mathcal{F}_t \sim \bigcirc_{1\le i\le k}\mathrm{DM}\left(X_t^i, \frac{1}{c_i}P_t^{(i)}\right)\circ \mathrm{CPo}(A_tX_t, L)\circ\delta(I_t),$$

where $\bigcirc$ and $\circ$ stand for the summation of (mutually) independent random vectors.

Remark 2. The variability of the individual infectiousness may be naturally reflected by the choice of L. To demonstrate it, assume that only a single compartment (labeled I) is infectious, that all the new infections fall to a single compartment (labeled E), and the number of risk contacts of each infectious individual is Poisson. Let $t \ge 0$ and denote $N_{t+1,i}$ the number of the infections, caused by the i-th individual at t.

If the intensity of the contact distribution and the contagion probability were the same for all, equal to c, p, respectively, then it would be $N_{t+1,i} \sim \mathrm{Po}(cq)$, $i \in \mathbb{N}$; consequently, $N_{t+1}^E \sim \mathrm{Po}(\lambda X_t^I)$, $\lambda = cq$, so we may put $\alpha_t^{EA} = \lambda$ and $L = \delta_1$, having $\mathbb{E}(N_{t+1}^E|\mathcal{F}_t) = \mathrm{var}(N_{t+1}^E|\mathcal{F}_t) = \lambda X_t$.

Now consider a more realistic situation in which the infectiousness randomly varies between individuals. A standard way of modeling this situation is assuming, for each i, that the intensity λ_i of $N_{t+1,i}$ is chosen from Gamma distribution, implying that $N_{t+1,i}$ is negative binomial (see [8]). In particular, once $\lambda_i \sim \Gamma(k,\theta)$, and $N_{t+1,i}|\lambda_i = \mathrm{Po}(\lambda_i)$, we are getting that $N_{t+1,i} \sim \mathrm{NB}(k,p)$, $p = \frac{\theta}{1+\theta}$, with

$$\mathbb{E}N_{t+1,i} = \kappa \stackrel{\mathrm{def}}{=} \theta k, \qquad \mathrm{var}(N_{t+1,i}) = \kappa v, \qquad v = 1 + \frac{\kappa}{k}. \tag{1}$$

As the Negative Binomial distribution can be represented by a Compound Poisson one and as the sum of independent Compound Poisson distributions is Compound Poisson, we have that N_{t+1}^E is Compound Poisson[2] and it follows from (1) that

$$\mathbb{E}(N_{t+1}^E|\mathcal{F}_t) = \kappa X_t, \qquad \mathrm{var}(N_{t+1}^E|\mathcal{F}_t) = \kappa X_t v$$

Remark 3. [2] claim that the total number T of persons infected by a single COVID-infectious individual is $T \sim \mathrm{NB}(K,P)$ where $K \doteq 0.1$ and P is such that $\mathbb{E}T = R_0$ where R_0 is the basic reproduction number. As R_0 of COVID is generally assumed to be around 2.5 and $\mathbb{E}T = K\frac{P}{1-P}$, it follows that $P \doteq \frac{25}{26}$. Consequently, the variance-to-mean ratio is $\tilde{v} \stackrel{\mathrm{def}}{=} \frac{1}{1-P} \doteq 26$. Assuming that the individual is infectious for f days and that his contacts are restricted by a factor β, we may, in light of the Compound Poisson reformulation of T, conclude that the number $N_{t+1,i}$ of daily infected is Compound Poisson with

[2] In particular, $N_{t+1,i} = \mathrm{CPo}(k\ln(1+\theta), \mathrm{Log}(p))$, where Log is the Logarithmic distribution, see [8]). Thus, $N_{t+1}^E \sim \mathrm{CPo}(X_t^I k\ln(1+\theta), \mathrm{Log}(p))$ so we may put $\alpha_t^{EA} = k\ln(1+\theta)$, $L = \mathrm{Log}(p)$..

$\mathbb{E}N = \frac{\beta}{f}\mathbb{E}T, \mathrm{var}(N) = \frac{\beta}{f}\mathrm{var}(T)$, i.e. $N_{t+1,i}$ and consequently N^E_{t+1} has the same variance-to-mean ratio as T, i.e. $v = \tilde{v} \doteq 26$.

Remark 4. For any $x \in \mathbb{N}_0, c \geq 0$ and $p \in [0,1]^k$ such that $\sum_{i=1}^k p^i = 1$, we have

$$\mathbb{E}\left(\mathrm{DM}(x, \frac{p}{c})\right) = px = \mathbb{E}\left(\mathrm{Multinomial}(x,p)\right)$$

and

$$\mathrm{var}\left(\mathrm{DM}(x, \frac{p}{c})\right) = [\mathrm{diag}(p) - pp^T]\frac{x+c}{1+c}x = \mathrm{var}(\mathrm{Multinomial}(x,p))\frac{x+c}{1+c},$$

from which it is clear that $\mathrm{DM}(x, \frac{p}{c}) \to \mathrm{Multinomial}(x,p)$ as $c \to \infty$ and that the deviation of the Multinomial variance matrix grows with decreasing c.

3. Model Properties

By probability calculus, we get that

$$\mathbb{E}(X_{t+1}|\mathcal{F}_t) = \mathbb{E}(T_t X_t + I_t|\mathcal{F}_t) = \mathbb{E}\left(\sum_{i=1}^k M_{t+1,i} + N_{t+1} + I_t \middle| \mathcal{F}_t\right)$$
$$\sum_{i=1}^k P_t^{(i)} X_t^i + (\mathbb{E}L)A_t X_t + I_t = T_t X_t + I_t,$$

where, for any matrix Σ, $\Sigma^{(i)}$ denotes its i-th column, and

$$T_t \stackrel{\mathrm{def}}{=} P_t + B_t, \qquad B_t = (\beta_t^{i,j})_{1 \leq i,j \leq k} \stackrel{\mathrm{def}}{=} (\mathbb{E}L)A_t, \qquad t \geq 0. \tag{2}$$

Consequently, for any $t, s \in \mathbb{N}_0, t > s$,

$$\mathbb{E}(X_t|\mathcal{F}_s) = \mathbb{E}(T_{s,t-1}X_s + \sum_{\theta=s}^{t-1} T_{\theta+1,t-1} I_\theta|\mathcal{F}_s)$$
$$= \mathbb{E}(T_{s,t-1}|\mathcal{F}_s)X_s + \sum_{\theta=s}^{t-1} \mathbb{E}(T_{\theta+1,t-1} I_\theta|\mathcal{F}_s),$$

where, for any matrix process Σ, $\Sigma_{s,t} \stackrel{\mathrm{def}}{=} \Sigma_t \times \cdots \times \Sigma_s$ with $\Sigma_{s,s-1} \stackrel{\mathrm{def}}{=} E$ where E is the identity matrix.

In the special case that

$$B_\tau \equiv B_s, \qquad P_\tau \equiv P_s, \qquad s \leq \tau \leq t, \tag{3}$$

we have

$$\mathbb{E}(X_t|\mathcal{F}_s) = T_s^{t-s} X_s + \sum_{\tau=s}^{t-1} T_s^{t-\tau-1} \mathbb{E}(I_\tau|\mathcal{F}_s),$$

and

$$\mathbb{E}(X_t|\mathcal{G}_s) = T_s^{t-s}\mathbb{E}(X_s|\mathcal{G}_s) + \sum_{\tau=s}^{t-1} T_s^{t-\tau-1}\mathbb{E}(I_\theta|\mathcal{G}_s)$$

If, in addition, $\mathbb{E}(I_\theta|\mathcal{G}_s) \equiv \mu$ for some $\mu \in \mathcal{G}_s$ and $(E - T_s)$ is invertible, the latter formula simplifies to

$$\mathbb{E}(X_t|\mathcal{G}_s) = T_s^{t-s}\mathbb{E}(X_s|\mathcal{G}_s) + (E - T_s)^{-1}(E - T_s^{t-s})\mu.$$

As for variance, we have

$$\begin{aligned}
\operatorname{var}(X_{t+1}|\mathcal{F}_t) &= \sum_{i=1}^{k} \operatorname{var}(M_{t+1,i}|\mathcal{F}_t) + \operatorname{var}(N_{t+1}|\mathcal{F}_t) + \operatorname{var}(I_t|\mathcal{F}_t) \\
&= \sum_{1\le i\le m} \operatorname{var}\left(\mathrm{DM}\left(X_t^i, \frac{1}{c_i}P_t^{(i)}\right)\right) + \operatorname{var}(\mathrm{CPo}(A_tX_t, L)) + 0 \\
&= \sum_{i\le t\le m} \frac{X_t^i + c_i}{1 + c_i} X_t^i[\operatorname{diag}(P_t^{(i)}) - P_t^{(i)}(P_t^{(i)})^T] + \operatorname{diag}(vB_tX) \\
&= \Lambda_t(X_t, X_t^2)
\end{aligned}$$

where

$$\Lambda_t(x, y) \stackrel{\text{def}}{=} \sum_{i=1}^{k} \frac{y^i + x^i c_i}{1 + c_i}[\operatorname{diag}(P_t^{(i)}) - P_t^{(i)}(P_t^{(i)})^T] + \operatorname{diag}(vB_tx)$$

(note that Λ_t is linear in x, y). Consequently,

$$\mathbb{E}\left[\begin{array}{c} X_{t+1} \\ Y_{t+1} \end{array}\middle|\, \mathcal{F}_t\right] = \begin{bmatrix} E \\ F \end{bmatrix}(T_tX_t + I_t), \tag{4}$$

$$\operatorname{var}\left(\begin{array}{c} X_{t+1} \\ Y_{t+1} \end{array}\middle|\, \mathcal{F}_t\right) = \begin{bmatrix} E \\ F \end{bmatrix} \Lambda_t(X_t, X_t^2) \begin{bmatrix} E \\ F \end{bmatrix}^T + \operatorname{diag}\begin{pmatrix} 0_k \\ \Gamma_t(X_t, X_t^2) \end{pmatrix}, \qquad t \ge 0.$$

4. Sub-epidemics and Reproduction Number

We say that the subset of compartments $D = \{s_1, \dots, s_m\}$ is *subepidemic* if, for any t and any $i \in D$ and $j \notin D$, $\beta_t^{ij} \equiv \beta^{ji} \equiv 0$, and $p_t^{ij} \equiv 0$. In words this means that, for any $i \in D$, the i-th compartment does not increase through direct infection, the infection does not depend on the compartment, and it is impossible to get to the state i once being outside D.

Let, after a possible re-ordering, $m \in \mathbb{N}$ be such that $\{1, \dots, m\}$ is subepidemic (such m always exists because it can be always put to k). For any vector $x \in \mathbb{R}^k$, denote $\overline{x}$ its restriction to $(1, \dots, m)$ and, for any matrix $A \in \mathbb{R}^{k\times k}$, denote $\overline{A}$ its restriction to $(1, \dots, m) \times (1, \dots, m)$.

Observe that $\overline{X}$ follows a slightly modified version of our model, namely

$$\overline{X}_{t+1}\big|\,\mathcal{F}_t \sim \bigcirc_{1\le i\le m}\mathrm{DM}^-(\overline{X}^i_t, \frac{1}{c}\overline{P}^{(i)}_t, c)\circ \mathrm{CPo}(\overline{B}_t\overline{X}_t, L)\circ\delta(\overline{I}_t).$$

where, for any $x\in\mathbb{N}_0, \alpha\in\mathbb{R}^m_+$ and $c>0$, $\mathrm{DM}^-(x,\alpha,c)$ is the marginal distribution of the first m components of $\mathrm{DM}\left(x, \left[c-\sum_{i=1}^m \alpha^i\right]\right)$. (by the aggregation property of DM).

For any t, we define the reproduction number r_t (of a subepidemic $\{1,\dots,m\}$) as

$$r_t \stackrel{\mathrm{def}}{=} \sum_{\tau=t}^{\infty}\mathbf{1}^T\mathbb{E}(B_\tau\overline{P}_{t,\tau-1}|\mathcal{F}_{t-1})\pi_t, \qquad \pi_t = \mathbb{E}\left\{\nu\left(\overline{N}_t+\overline{I}_{t-1}\right)\big|\,\mathcal{F}_{t-1}\right\}.$$

where ν is unit normalization of a vector. Observe that r_t complies with the usual definition of reproduction number as it equals to the conditional expectation (w.r.t. $\mathcal{F}_{t-1}$) of the infections caused by an individual having arrived at t. To see it, note that π_t is the conditional distribution of the state in which a randomly chosen newcomer (the one brought by the import or by the infection) finds himself at t, and observe that, for each newcomer at t, the expected number of those infected by him at $t+1$ is given by the sum of the components of $\overline{B}_t\pi_t$, the expected number infected at $t+1$ is given by the sum of components of $\overline{B}_{t+1}\overline{P}_t\pi_t$ etc.

If $\mathcal{F}_t\neq\mathcal{G}_t$ (i.e. X is not fully observed), then the reproduction number has to be estimated, most naturally by its conditional expectation with respect to the known information:

$$\tilde{r}_t \stackrel{\mathrm{def}}{=} \mathbb{E}(r_t|\mathcal{G}_t) = \sum_{\tau=t}^{\infty}\mathbf{1}^T\mathbb{E}(\overline{B}_\tau\overline{P}_{t,\tau-1}\pi_t|\mathcal{G}_{t-1}).$$

In the special case of $\overline{B}_\tau\equiv\overline{B}_{t-1}, \overline{P}_\tau\equiv\overline{P}_{t-1}$, $\tau\ge t$, with $\rho(\overline{P}_{t-1})<1$ where ρ is the spectral radius, the formula simplifies to

$$\tilde{r}_t = \mathbf{1}^T\overline{B}_{t-1}\left(\sum_{i=0}^{\infty}\overline{P}^i_{t-1}\right)\mathbb{E}(\pi_t|\mathcal{G}_{t-1}) = \mathbf{1}^T\overline{B}_{t-1}(E-\overline{P}_{t-1})^{-1}\mathbb{E}(\pi_t|\mathcal{G}_{t-1}).$$

Note that, once $\overline{N}$ and/or $\overline{I}$ are possibly not observed, there could be difficulties computing $\mathbb{E}(\pi_t|\mathcal{G}_{t-1})$ – yet the estimate $\mathbb{E}(\pi_t|\mathcal{G}_{t-1})\doteq\nu(\overline{B}_{t-1}\mathbb{E}(\overline{X}_{t-1}|\mathcal{G}_{t-1})+\mathbb{E}(\overline{I}_{t-1}|\mathcal{G}_{t-1}))$ seems a straightforward choice, it is generally not unbiased due to the normalization. This problem, however, vanishes if the imports and new infections all fall into a single state (typically called exposed and labeled E), in which case $\pi_t\equiv(1,0,\dots,0)^T$.

5. Asymptotic Behavior

Keep assuming that $\{1,\dots,m\}$ is subepidemic. The next Proposition states conditions for vanishing, explosion and "stationary" behavior of the subepidemic.

Proposition 5. *(i) If $\overline{T}_t \leq S$ component-wise, where S is deterministic with $\sigma \overset{def}{=} \rho(S) < 1$, and if $\mathbb{E}\overline{I}_t = o(t^{-\alpha})$ for some $\alpha > 0$, then $\overline{X}_t \to 0$ almost sure. Here, ρ denotes the spectral radius of a matrix.*
(ii) If $\overline{T}_t \geq R$ where R is deterministic irreducible with $\varrho \overset{def}{=} \rho(R) > 1$ and either $\mathbb{E}\overline{X}_0 \neq 0$ or $\mathbb{E}\overline{I}_\tau \neq 0$ for some τ, then $\|\mathbb{E}\overline{X}_t\| \to \infty$.
(iii) If $\mathbb{E}\overline{I}_t \equiv \mu$ for some μ and $R \leq \overline{T}_t \leq S$ such that $\sigma \overset{def}{=} \rho(S) < 1$, then

$$\liminf_t \mathbb{E}\overline{I}_t \geq (E-R)^{-1}\mu, \qquad \limsup_t \mathbb{E}\overline{I}_t \leq (E-S)^{-1}\mu$$

Proof. (i) We have

$$\mathbb{E}\overline{X}_t = \mathbb{E}(\mathbb{E}(\overline{X}_t|\mathcal{G}_0)) = \mathbb{E}(\overline{T}_{0,\tau-1}X_0 + \sum_{\theta=0}^{t-1}\overline{T}_{\theta+1,t-1}\mathbb{E}(\overline{I}_\theta|\mathcal{G}_0))$$

$$\leq \mathbb{E}(S^tX_0 + \sum_{\theta=0}^{t-1}S^{t-\theta-1}\mathbb{E}(\overline{I}_\theta|\mathcal{G}_0)) \leq a_t + b_t, \qquad a_t = S^t\mathbb{E}\overline{X}_0, \qquad b_t = \sum_{\theta=0}^{t-1}S^{t-\theta-1}\mathbb{E}\overline{I}_\theta$$

Thanks to the sub-unit spectral radius of S, we have $a_t \to 0$. Further, by the non-negativity of H and the properties of convergence, there exists $c \in \mathbb{R}^m_+$ such that $\mathbb{E}\overline{I}_t \leq c(t+1)^{-1}$. Thus, for any ς fulfilling $\sigma < \varsigma < 1$, we get, after re-indexing the sum,

$$b_t = \sum_{\tau=0}^{t-1}S^\tau\mathbb{E}\overline{I}_{t-\tau-1} \leq \sum_{\tau=0}^{t-1}S^\tau c\frac{1}{(t-\tau)^\alpha} = \underbrace{\frac{1}{t^\alpha}}_{\to 0} \times \underbrace{\sum_{\tau=0}^{t-1}(\varsigma^{-1}S)^\tau c}_{\to(E-\varsigma^{-1}S)^{-1}c} \underbrace{\left(\frac{\varsigma^{\tau/\alpha}}{t-\tau}\right)^\alpha}_{\leq d} \to 0;$$

the second convergence holding because $\rho(\varsigma^{-1}S) = \frac{\sigma}{\varsigma_t} < 1$, the upper bound d existing as $f(\tau) \overset{\text{def}}{=} \frac{t\varsigma^{\tau/\alpha}}{t-\tau}$ increases in $\tau = t-1$ and its derivative has only a single root, so we have $f(\tau) \leq \max(f(0), f(t-1)) = \max(1, \varsigma^{\frac{t-1}{\alpha}}t) \leq d \overset{\text{def}}{=} \max(1, \frac{1}{e\varsigma^{1/\alpha}|\alpha^{-1}\ln\varsigma|})$ on $[0, t-1]$. Finally, thanks to the non-negativity of $\overline{X}$, convergence of $\mathbb{E}\overline{X}_t$ suffices for a.s. convergence of $\overline{X}_t$.

(ii) Let $\mathbb{E}\overline{X}_0 \neq 0$ and $\varrho > 1$. As R is irreducible non-negative ϱ is its eigenvalue and the corresponding eigenvector x is positive by the Perron-Frobenius Theorem. Further, by the irreducibility of T, there exists n such that $y \overset{\text{def}}{=} R^n\mathbb{E}\overline{X}_0 > 0$ component-wise, so there exist $e > 0$ such that $y \geq ex$. Thus

$$\mathbb{E}\overline{X}_t \geq R^t\mathbb{E}\overline{X}_0 \geq R^{t-n}y \geq eR^{t-n}x$$

norm of which converges to infinity. The proof for $\mathbb{E}\overline{I}_\tau \neq 0$ is analogous.

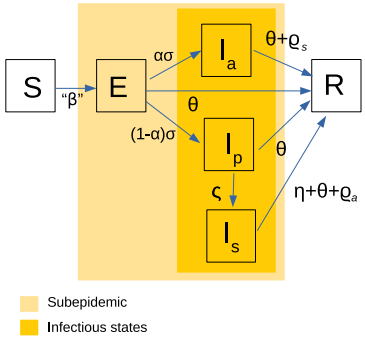
(iii)

$$\mathbb{E}\overline{X}_t = \sum_{\tau=0}^{t-1}\overline{T}_{\tau,t-2}\mu + \overline{T}_{0,t-1}\mathbb{E}\overline{X}_0 \leq \sum_{\tau=0}^{t-1}S^\tau\mu + S^{t-1}\mathbb{E}\overline{X}_0 \to (\sum_{\tau=0}^{\infty}S^\tau)\mu = (E-S)^{-1}\mu$$

and similarly for R.

□

Example 6. Say there are five states E – exposed, I_a – infectious asymptomatic, who will never show symptoms, I_p – infectious pre-symptomatic, who will later show symptoms, I_s – infectious symptomatic, and R – removed, which includes the recovered, the dead, and the infectious isolated. We index the states by e, a, p, s, r. For simplicity we assume $v = 1$ which means that the new infections are Poisson rather than Compound Poisson. All the infectious states are equally infectious, i.e. $\beta_t^{ex} = \beta_t$, $x \in \{a, p, s\}$, where β is a $\mathcal{G}_t$-adapted process. The probability that the exposed transits to $\{a, p\}$ is σ, the probability of completely asymptomatic course is α, the probability of transition from I_p to I_s is ς. Further, the probability of ending I_a or I_s, by natural causes (recovery, end of infectiousness, death in case of I_s) is ϱ_a, ϱ_s, respectively. Finally, the probability that a symptomatic individual isolates himself is η and the probability that the individual finding herself in state i is isolated is θ_t^i for some $\mathcal{G}_t$-adapted process θ^i. The situation is illustrated on the following Figure:



If we neglect (small) joint probabilities of natural exits from the infectious states and the isolations, we get

$$P_t = \begin{bmatrix} 1-\sigma-\theta_t^e & 0 & 0 & 0 & 0 \\ \alpha\sigma & 1-\varrho_a-\theta_t^a & 0 & 0 & 0 \\ (1-\alpha)\sigma & 0 & 1-\varsigma-\theta_t^p & 0 & 0 \\ 0 & 0 & \varsigma & 1-\varrho_s-\eta-\theta_t^s & 0 \\ \theta_t^e & \theta_t^a+\varrho_a & \theta_t^p & \theta_t^s+\eta+\varrho_s & 1 \end{bmatrix}, \qquad B_t = \begin{bmatrix} 0 & \beta_t & \beta_t & \beta_t & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Clearly, we can put $m = 4$ (the first four states form a sub-epidemic), getting

$$\overline{T}_t = Q + \beta_t C - \mathrm{diag}(\theta_t), \qquad Q = \begin{bmatrix} 1-\sigma & 0 & 0 & 0 \\ \alpha\sigma & 1-\varrho_a & 0 & 0 \\ (1-\alpha)\sigma & 0 & 1-\varsigma & 0 \\ 0 & 0 & \varsigma & 1-\varrho_s-\eta \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}..$$

Further we assume $\theta_t \overset{\text{def}}{=} \theta_t^e = \theta_t^a = \theta_t^p = \theta_t^s$, which yields, by the well known rule, $\rho(\overline{T}_t) = \rho(Q + \beta_t C) - \theta_t$. We consider two ways of decreasing the spectral

radius: decreasing the infection rate β_t (typically by wide counter-epidemic measures) and increasing the isolation rate θ_t (e.g. by strengthening the testing and tracing capacity).

Once there is a "target" spectral radius ρ_0, all the combinations of β and θ yielding $\rho(\overline{T}_t) = \rho_0$ fulfill $\rho_0 + \theta - \rho(Q + \beta_t C) = 0$,which gives a "marginal rate of substitiution" $\theta(\beta)' = -\frac{\partial}{\partial\beta}\rho(Q + \beta_t C)$ of the infectiousness by the isolation, i.e. how much we have to increase the isolation speed when we release the restrictions.

Example 7. Assume there is a fraction ν of the population is non-compliant, which means that, once a restriction on social contacts is imposed, they apply it only partially. Assume that, without restrictions, the population is mixed which means that each individual, compliant or not, has, up to a constant, $(1-\nu)$ contacts with the compliant individuals and ν contacts with the non-compliant ones. Once there is a measure imposed under which the compliant individuals restrict their opportunities to contacts by ϕ, the non-compliant ones do so only to $f(\phi) > \phi$. As a result, the compliant ones will have, up to a constant, $\phi^2(1-\nu)$ contacts with the compliant ones, $\phi f(\phi)\nu$ contacts with the non-compliant ones, while the non-compliant will have $\phi f(\phi)(1-\nu)$ and $f(\phi)^2\nu$ contacts with the compliant, non-compliant, respectively.

Assuming a simple epidemic model with compartments I_c - infected compliant, I_n - infected non-compliant, and R - removed, with the course of infection being the same for both the compartments such that $\beta_t^{1i} = \beta c_i$, $i \in \{1, 2\}$, where β is a constant and c_i is the number of contacts of the i-th sub-population, this gives

$$P_t = \begin{bmatrix} 1-\varrho & 0 & 0 \\ 0 & 1-\varrho & 0 \\ \varrho & \varrho & 1 \end{bmatrix}, \qquad B_t = \begin{bmatrix} \beta\phi^2(1-\nu) & \beta\phi f(\phi)\nu & 0 \\ \beta\phi f(\phi)(1-\nu) & \beta f(\phi)^2\nu & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where ϱ is a removal rate (perhaps consisting of an artificial and a natural part). This gives

$$\overline{T}_t = \beta C + (1-\varrho)E, \qquad C = \begin{bmatrix} \phi^2(1-\nu) & \phi f(\phi)\nu \\ \phi f(\phi)(1-\nu) & f(\phi)^2\nu \end{bmatrix}$$

with

$$\varrho(\overline{T}_t) = \beta\rho(C) + (1-\varrho).$$

As the characteristic polynomial of C is

$$\lambda^2 - \lambda g, \qquad g = g(\phi, \nu) = \phi^2(1-\nu) + f(\phi)^2\nu$$

we clearly have $\rho(C) = g$.

Now say that our goal is to decrease $\rho(\overline{T}_t)$ to a predetermined value r by finding appropriate $\phi = \phi(\nu)$. In order to do so, we have to solve

$$\beta g(\phi(\nu), \nu) + (1-\varrho) = r.$$

Clearly, $\phi(0) = \phi_0 \stackrel{\text{def}}{=} \sqrt{\frac{r-1+\varrho}{\beta}}$. For $\nu > 0$ we get, by the Implicit function theorem,

$$\frac{\partial}{\partial \nu}\phi = \frac{f(\phi(\nu))^2 - \phi(\nu)^2}{2\phi(\nu)(1-\nu) + 2f(\phi(\nu))f'(\phi(\nu))\nu}.$$

Note that the derivative depends neither on r nor on ϱ. Thus we can easily compute how the non-compliance influences strictness of the necessary restrictions. For instance, by the first-order Taylor expansion at $\nu = 0$, we get

$$\phi(\nu) \doteq \phi_0 + \nu\frac{f(\phi_0)^2 - \phi_0^2}{2\phi_0} = \phi_0\left(1 - \frac{\nu}{2}\right) + \nu\frac{f(\phi_0)^2}{2\phi_0}$$

roughly holding for ν close to zero.

6. Cohort Model

In the present Section, we assume the population to be split into r (age) cohorts of sizes $s_1, \ldots, s_r$, $s_1 + \cdots + s_r = s$. The members of each cohort may be either susceptible or belong to one of κ analogous compartments. Naturally assuming that individuals do not migrate between cohorts, we get the overall transition matrix as

$$P_t = \begin{bmatrix} P_t^1 & 0 & \cdots & 0 \\ 0 & P_t^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & P_t^r \end{bmatrix}$$

where P_t^i are $\kappa \times \kappa$ cohort transition matrices, $1 \leq i \leq \kappa$, $t \geq 0$. Note that once there are dispersion parameters $c_1^i, \ldots, c_\kappa^i$ associated with each matrix P_t^i (meaning that, once the j-th compartment of the i-th cohort is of size x, the transfers from the cohort to the cohort's compartments follow $\mathrm{DM}(x, \frac{(P^i)^{(j)}}{c_i^j})$), the dispersion parameters of the overall model are $(c_1^1, \ldots, c_\kappa^1, c_1^2, \ldots c_\kappa^2, c_1^3, \ldots\ldots, c_\kappa^r)$.

We assume that the contagions can happen across cohorts. Namely, the probability of contagion, i.e. the transfer of a susceptible individual to the i-th compartment of the p-th cohort, upon a risk contact with a member of the j-th compartment of the q-th cohort does not depend on p or q, being equal to ϖ_t^{ij}. Further we assume that, on average, a member of the p-th cohort has ν^{pq} risk contacts with the q-th cohort, assuming that the number of contacts with the infectious compartments (those with non-zero $\varpi_t^{i\bullet}$) is equal. Under these assumption, the probability a transfer of a susceptible individual from cohort p into the i-th compartment is roughly $e_t^{pi} \stackrel{\text{def}}{=} b\sum_{q=1}^r \sum_{j=1}^k \nu^{pq}\varpi_t^{ij}\frac{X_{tq}^i}{s_q}$, where X_{tq}^i is the size of the i-th compartment of cohort q and b is a constant. Consequently, the total number the infections in the i-th compartment of the p-th cohort will be $s_p e_t^{pi}$, which gives

$$B_t = Q \otimes C_t$$

where

$$Q = \begin{bmatrix} \nu^{11} & \nu^{12}\frac{s_1}{s_2} & \cdots & \nu^{1r}\frac{s_1}{s_r} \\ \nu^{21}\frac{s_2}{s_1} & \nu^{22} & \cdots & \nu^{2r}\frac{s_2}{s_r} \\ \vdots & \vdots & \ddots & \vdots \\ \nu^{r1}\frac{s_r}{s_1} & \nu^{r2}\frac{s_r}{s_2} & \cdots & \nu^{rr} \end{bmatrix}, \qquad C_t = b \begin{bmatrix} \varpi_t^{11} & \varpi_t^{12} & \cdots & \varpi_t^{1k} \\ \varpi_t^{21} & \varpi_t^{22} & \cdots & \varpi_t^{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \varpi_t^{k1} & \varpi_t^{k2} & \cdots & \varpi_t^{kk} \end{bmatrix}.$$

It may be convenient to re-parametrize the model, either by putting $b = 1$ and multiplying C (in which case, however, its components cease to be probabilities) and/or by normalizing Q for better comparison with non-cohort models.

7. Optimal Control of the Epidemic

Assume the setting of Example 6 and assume that X_0 is known. Our aim is to minimize the size of the epidemic at time t given that we are ready to pay a given price c_0. We assume that, to achieve infection rate β, a cost $\gamma(\beta)$ has to be paid where γ is a strictly decreasing convex positive function defined on $(0, \beta_0]$ with $\gamma(\beta_0) = 0$ and $\gamma(0-) = \infty$. Further, to achieve the isolation rate θ_t^i in the i-th compartment, the price $\delta(\theta_t^i, X_t^i) \stackrel{\text{def}}{=} d\theta_t^i X_t^i$ has to be paid where d is a constant. This reflects the real-life situation in which the cost of global restrictions does not depend on the infection size while the cost of isolation does, for instance through the number of call-center workers involved in tracing.

Our problem is to find

$$V_0(X_0, c_0) \stackrel{\text{def}}{=} \inf_{\sum_{\tau=0}^{t-1}[\gamma(\beta_\tau)+\delta(\theta_t^k, X_t^k)+\cdots+\delta(\theta_t^k, X_t^k)] \le c_0, \beta, 0 \le \theta \le q, \theta_\tau \in \mathcal{F}_\tau, \beta_\tau \in \mathcal{F}_\tau, 1 \le \tau < t} \mathbb{E}(\mathbf{1}' X_t)$$

where q is the diagonal of Q. The problem may be rewritten by means of Bellman equations

$$V_\tau(x, c) = \inf_{\gamma(\beta)+d\theta^k x^k+\cdots+d\theta^k x^k+y \le c, \beta \ge 0, \theta \ge 0, y \ge 0} \mathbb{E}(V_{\tau+1}(J, y)), \qquad 0 \le \tau < t,$$

$$V_t(x, c) = \mathbf{1}' x,$$

$$J \sim \mathcal{L}(x, \beta, \theta) \stackrel{\text{def}}{=} \bigcirc_{1 \le i \le m} \text{Multinomial}^-(x^i, Q^{(i)} - \Delta_i \theta^i) \circ \text{Po}(\beta C x)$$

where Δ_i is the vector with the unit component on the i-th place and zeros otherwise.

Though the final problem is convex –

$$V_{t-1}(x, c) = \inf_{\gamma(\beta)+d\theta^1 x^1+\cdots+d\theta^m x^m+y \le c, 0 \le \theta \le q, \beta, y \ge 0} \mathbf{1}'(Qx - \text{diag}(\theta)x + \Delta_e \beta \sum_{i \ne e} x^i)$$

– its objective function is not jointly convex in (u, v, x, c), so the convexity of V_{t-1} is not guaranteed. Moreover, as neither the optimal solution nor the expectation of the objective functions of all but the last problem are analytically

tractable, it is necessary to resort to approximations. To this end, we can use the sample mean approximation

$$V_\tau(x,c) \doteq \tilde{V}_\tau(x,c)$$

where $\tilde{V}_t = V_t$ and, recursively,

$$V_\tau(x,c) \doteq \tilde{V}_\tau(x,c) \stackrel{\text{def}}{=} \inf_{\gamma(\beta)+d\theta^k x^k+\cdots+d\theta^k x^k+y\le c,\beta,\theta,y\ge 0} \frac{1}{r}\sum_{i=1}^{r} \tilde{V}_{\tau+1}(J_i,y)$$

where $J_1,\ldots,J_r$ is an i.i.d. sample from $\mathcal{L}(x,\beta,\theta)$.

8. Estimation

For any stochastic process A and integers $s \ge t$, denote $\hat{A}_{s|t} = \mathbb{E}(A_s|\mathcal{G}_t)$. Let $s > t$. When $T_\tau \in \mathcal{G}_t, t < \tau \le s-1$ (which is trivially true if $s = t+1$), we get that

$$\begin{aligned}\begin{bmatrix}\hat{X}_{s|t}\\ \hat{Y}_{s|t}\end{bmatrix} &= \mathbb{E}\left(\mathbb{E}\left(\begin{bmatrix}X_s\\ Y_s\end{bmatrix}\middle|\mathcal{F}_{s-1}\right)\middle|\mathcal{G}_t\right) = \begin{bmatrix}E\\ F\end{bmatrix}\left(T_{s-1}\hat{X}_{s-1|t}+\hat{I}_{s-1|t}\right)\\ &= \begin{bmatrix}E\\ F\end{bmatrix}\left(T_{t,s-1}X_t+\sum_{\theta=t}^{s-1}T_{\theta+1,s-1}\hat{I}_{\theta|t}\right),\end{aligned}$$

$$\begin{aligned}W_{s|t} \stackrel{\text{def}}{=}\ & \operatorname{var}(X_s|\mathcal{G}_t) = \operatorname{var}(\mathbb{E}(X_s|\mathcal{F}_{s-1})|\mathcal{G}_t) + \mathbb{E}(\operatorname{var}(X_s|\mathcal{F}_{s-1})|\mathcal{G}_t)\\ &= \operatorname{var}(T_{s-1}X_{s-1}+I_{s-1}|\mathcal{G}_t) + \mathbb{E}\left(\Lambda_{s-1}(X_{s-1},X_{s-1}^2)\middle|\mathcal{G}_t\right)\\ &= T_{s-1}W_{s-1|t}T_{s-1}^T + 2T_{s-1}\operatorname{cov}(X_{s-1},I_{s-1}|\mathcal{G}_t) + \operatorname{var}(I_{s-1}|\mathcal{G}_t)\\ &\quad + \Lambda_{s-1}(\hat{X}_{s-1|t},\operatorname{diag}(W_{s-1|t})+\hat{X}_{s-1|t}^2)\end{aligned}$$

(we have used linearity of Λ_{s-1}, thanks to which $\mathbb{E}(\Lambda_{s-1}(\hat{X}_{s-1|t},X_{s-1}^2)|\mathcal{G}_t) = \Lambda_{s-1}(\mathbb{E}(X_{s-1}|\mathcal{G}_t),\mathbb{E}(X_{s-1}^2|\mathcal{G}_t))$, and the well known formula $\operatorname{var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$). Consequently,

$$\begin{aligned}V_{s|t} \stackrel{\text{def}}{=}\ & \operatorname{var}\left(\begin{matrix}X_s\\ Y_s\end{matrix}\middle|\mathcal{G}_t\right) = \operatorname{var}\left(\begin{matrix}X_s\\ FX_s+\epsilon_s\end{matrix}\middle|\mathcal{G}_t\right)\\ &= \begin{bmatrix}E\\ F\end{bmatrix}W_{s|t}\begin{bmatrix}E\\ F\end{bmatrix}^T + \operatorname{diag}\begin{pmatrix}0_k\\ \Gamma_{s-1}(\hat{X}_{s-1|t},\operatorname{diag}(W_{s-1|t})+\hat{X}_{s-1|t}^2)\end{pmatrix}\end{aligned}$$

Unfortunately, due to the non-Gaussianity, we have analytical formulas for none of $X_{t|t}$ $I_{t|t}$ and $\mathrm{W_{t|t}}$, so we can formulate neither the likelihood function nor a least square estimate. Two, from the computational point of view equivalent, ways to cope with this are using estimates of the conditional expectation and

variance, or normally approximating the residuals. We go the latter way: in the present Section, we assume that $\left[\begin{smallmatrix} X_{t+1} \\ Y_{t+1} \end{smallmatrix}\right] \Big| \mathcal{F}_t$ is normal with mean given by (4) and

$$\mathrm{var}\left(\begin{matrix} X_{t+1} \\ Y_{t+1} \end{matrix} \Big| \mathcal{F}_t \right) = \begin{bmatrix} E \\ F \end{bmatrix} \Lambda_t(X_t \vee 0, X_t^2) \begin{bmatrix} E \\ F \end{bmatrix}^T + \mathrm{diag} \begin{pmatrix} 0_\mathrm{k} \\ \Gamma_\mathrm{t}(\mathrm{X_t} \vee 0, \mathrm{X_t^2}) \end{pmatrix}$$

Moreover, we assume that $I_t \in \mathcal{G}_t, t \geq 0$ (i.e. the import is observable). Given these assumptions, we have, by the well known formula (see e.g. [1]),

$$\hat{X}_{t|t} \;=\; I_{t-1} \;+\; \hat{X}_{t|t-1} \;+\; K_t \left(Y_t - \hat{Y}_{t|t-1} \right), \qquad K_t \;\stackrel{\mathrm{def}}{=}\; \tilde{V}^{XY}_{t|t-1} (\tilde{V}^{YY}_{t|-1})^{-1}$$

$$\mathrm{var}(X_t|\mathcal{G}_t) \qquad = \qquad \tilde{V}^{XX}_{t|t-1} \qquad - \qquad \tilde{V}^{XY}_{t|t-1} (\tilde{V}^{YY}_{t|t-1})^{-1} \tilde{V}^{YX}_{t|t-1}$$

where $\tilde{V}_{s|t} \stackrel{\mathrm{def}}{=} \mathrm{var}(\left[\begin{smallmatrix} X_s \\ Y_s \end{smallmatrix}\right] |\mathcal{G}_t)$ given the normal approximation. Note that K_t may be seen as a conditional version of the Kalman gain matrix.

If $\mathbb{P}[\hat{X}_t < 0]$ is negligible (which is typically true when modeling large epidemics), then we can neglect truncation in the formula for the variance approximate $\tilde{V}_{t|s} \doteq V_{t|s}$,which further gives

$$K_t \doteq W_{t|t-1} F^T D_t^{-1}, \qquad D_t = F W_{t|t-1} F^T + \mathrm{diag}(\Gamma_t \hat{X}_{t|t-1})$$

$$W_{t|t} \doteq W_{t|t-1} - W_{t|t-1} F^T D_t^{-1} F W_{t|t-1}$$

Assume that $F = F(\Theta_0)$,$P_t = P_t(\Theta_0)$, $B_t = B_t(\Theta_0)$, $\Gamma_t = \Gamma_t(\Theta_0)$ and $I_t = I_t(\Theta_0)$, and $c = c(\Theta_0)$, where $\Theta_0 \in \mathbb{R}^r$ is an unknown parameter. For its estimation, it is possible to use either nonlinear least squares, i.e.

$$\hat{\Theta} = \arg\min \sum_t (Y_t - \hat{Y}_{t|t}(\Theta))^T U_t (Y_t - \hat{Y}_{t|t}(\Theta))$$

where $U_t \in \mathcal{G}_{t-1}$ is a suitable weighting matrix, or

$$\tilde{\Theta} = \arg\min \sum_t \varphi(Y_t - \hat{Y}_{t|t}(\Theta), D_t(\Theta)), \qquad \varphi(x, v) = -\frac{k \ln 2\pi + \ln \det(v) + x^T v^{-1} x}{2}$$

Both these estimators are consistent and asymptotically normal under some conditions, see [5], [**?**], respectively. Verifying these conditions for our model is, however, beyond the scope of this introductory study and remains as topic of a future research.

It should be noted that our proof of Proposition 5 is not valid for the approximate model, as $\overline{X}$ is not necessarily positive given the Normal approximation.

9. Application to The COVID Pandemics in Czech Republic

TBD

References

[1] Morris L. Eaton. *Multivariate Statistics - a Vector Space Approach*, volume 53. Institute of Mathematical Statistics, Ohio, USA, 2007.

[2] Akira Endo et al. Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china. *Wellcome Open Research*, 5, 2020.

[3] Rahman Farnoosh and Mahmood Parsamanesh. Stochastic differential equation systems for an sis epidemic model with vaccination and immigration. *Communications in Statistics-Theory and Methods*, 46(17):8723–8736, 2017.

[4] Gavin J Gibson and Eric Renshaw. Estimating parameters in stochastic compartmental models using markov chain methods. *Mathematical Medicine and Biology: A Journal of the IMA*, 15(1):19–40, 1998.

[5] C. Jacob. Conditional least squares estimation in nonstationary nonlinear stochastic regression models. *The Annals of Statistics*, 38(1):566–597, 2010.

[6] Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo, Fiona Sun, Mark Jit, James D Munday, et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The lancet infectious diseases*, 2020.

[7] Kiesha Prem, Yang Liu, Timothy W Russell, Adam J Kucharski, Rosalind M Eggo, Nicholas Davies, Stefan Flasche, Samuel Clifford, Carl AB Pearson, James D Munday, et al. The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study. *The Lancet Public Health*, 2020.

[8] Mingyuan Zhou and Lawrence Carin. Negative binomial process count and mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):307–320, 2013.