

12 Robuste Inzidenzzahlen

Mit Mathematik kann man die Inzidenz stabilisieren, allerdings muss man auf Transparenz und Verständlichkeit achten, um eine breite Akzeptanz in der Bevölkerung zu bekommen.

Bei der Steuerung der (nicht-pharmazeutischen) Maßnahmen zur Bekämpfung der COVID-19 Pandemie wird standardmäßig als Orientierung die *7-Tage Inzidenz* der bestätigten Neuinfektionen (Summe der letzten 7 Tage, pro 100.000 Einwohner, auf Grundlage von Meldedaten im Zeitraum ab dem Berichtstag)³

$$\text{7-Tage Inzidenz} = \frac{\text{Anzahl der bestätigten Neuinfektionen der letzten 7 Tage}}{\frac{\text{betrachtete Einwohnerzahl}}{100.000}}$$

verwendet. Es werden also alle gemeldeten Neuinfektionen der jeweils zurückliegenden 7 Tage addiert und die Summe wird durch die betrachtete Einwohnerzahl (z.B. Stadt, Landkreis, ...) geteilt und mit 100.000 multipliziert.

Um mit diesem Inzidenzwert die Pandemie zu steuern, könnte man zum Beispiel es als Ziel setzen, diese Inzidenz jede Woche um 20 % zu senken⁴. Vielmehr wurden allerdings absolute Schwellwerte (z.B. 7-Tage-Inzidenz = 50, 100, usw.) verwendet, um Maßnahmen hinzuzufügen oder zu lockern (siehe z.B. die bundeseinheitliche „Notbremse“-Regelungen⁵).

Während die einen dies als ein geeignetes Frühwarnsystem betrachten, geben andere die mangelnde Robustheit der Inzidenz-Zahlen zu bedenken. So würde bei vermehrter Testung diese Inzidenz ansteigen; allerdings dann in der Regel durch die Nachverfolgung von mehr Fällen auch wieder schneller fallen. Ebenso geht die Inzidenz als Index implizit von einer Gleichverteilung der Infektionen in der getesteten Gruppe aus, was nicht immer der Fall sein muss, da sich vermehrt Leute mit einem Verdacht selber testen. Was kann man tun?

Zunächst kann man die Berechnung der Inzidenz selbst oder vergleichbarer Indizes durch eine *Ausgleichsrechnung* stabilisieren, anstatt einfach nur die Summe der letzten 7 Tage zu nehmen, um die sog. 'Wochen-Saisonalität' (Wochenend-Effekt) auszugleichen. Neben der Stabilisierung ist ein weiterer Vorteil, dass man den Wert für 'heute' erhält und nicht die Summe (also eine Art Mittelwert) der letzten 7 Tage. Zudem könnte man auch den gewählten Indexwert für einige Tage im voraus abschätzen.

³Robert Koch-Institut: COVID-19-Dashboard, <https://corona.rki.de/>

⁴<https://www.wa.de/politik/corona-notbremse-inzidenz-michael-meyer-hermann-markus-lanz-zdf-hamburg-90483279.html>

⁵<https://www.bundesregierung.de/breg-de/suche/bundesweite-notbremse-1888982>

Ausgleichsrechnung

Wir wollen im folgenden das Verfahren der Ausgleichsrechnung an einem konkreten Beispiel durchrechnen. Für das Beispiel wechseln wir von der Inzidenz zur sog. *Verdopplungszeit*, die aufgrund des Trends in den Daten geschätzt werden soll. Ende April 2020 war der Trend in den Daten schon klar abfallend (effektive Reproduktionszahl < 1 seit Ende März 2020) und die Verdopplungszeit wurde deutlich länger verglichen mit den Schätzungen in den ersten Wochen der Pandemie (wo wir im exponentiell-wachsenden Trend waren).

Im Zuge der Corona-Pandemie meldete die Johns-Hopkins-Universität (JHU) im Zeitraum 21.-26. April 2020 für Deutschland folgende Daten:

Datum	18.04	19.04	20.04	21.04	22.04	23.04	24.04
I	143.342	145.184	147.065	148.291	150.648	153.129	154.999

wobei I die Gesamtzahl der bestätigten Infektionen bezeichnet.

Als einfache Approximation in der Anfangszeit einer Pandemie (oder dem Beginn einer Infektionswelle) kann ein exponentielles Wachstum

$$I(t) = I(t_0) e^{b(t-t_0)}, \quad t > t_0$$

mit $b \in \mathbb{R}$ angenommen werden. Nun soll die *Verdopplungszeit* T_2 am 24. April 2020 in diesem Modell als Maß für die Inzidenz vorhergesagt werden. Es gilt

$$\frac{I(T_2)}{I(t_0)} = 2 \quad \text{d.h.} \quad e^{b(T_2-t_0)} = 2.$$

Somit erhalten wir die Verdopplungszeit $T_2 = t_0 + (\ln 2)/b$.

Wir betrachten nun die logarithmierten Werte der Infektionszahlen

$$L(t) := \ln I(t) = \ln I_0 + b(t - t_0)$$

und berechnen durch (lineare) Ausgleichsrechnung auf einer sinnvollen Zeitskala Polynome, z.B. $p_1(t) = a_0 + a_1(t - t_0)$. Wir bestimmen b und damit T_2 am 24. April durch Ableiten der Approximationen für $L(t)$ und vergleichen das Ergebnis mit dem der JHU (53.0 Tage)! Nummeriert man die Tage aufsteigend durch $t_0 = 0, \dots, t_6 = 6$, erhalten wir für p_1 das Problem $\|A_1 x_1 - v\|_2 \rightarrow \min$, wobei

$$A_1 = \begin{pmatrix} 1 & \dots & 1 \\ 0 & \dots & 6 \end{pmatrix}^\top, \quad x_1 = (a_0, a_1)^\top, \quad v = (\ln I(t_0), \dots, \ln I(t_6))^\top$$

Ferner ist

$$A_1^\top A_1 = \begin{pmatrix} 7 & 21 \\ 21 & 91 \end{pmatrix}, \quad A_1^\top v \approx (83.3772, 250.4968)$$

und wir erhalten als Lösung der *Normalgleichungen* $x_1 \approx (11.8719, 0.0130)^\top$. Ableiten liefert $b \approx p_1'(t_6) = a_1 \approx 0.0130$, womit $T_2 = (\ln 2)/b \approx 53.1438$, also der Wert der JHU.

Verwendet man dagegen ein quadratisches Polynom $p_2(t) = \hat{a}_0 + \hat{a}_1(t - t_0) + \hat{a}_2(t - t_0)^2$, erhält man nach Ableiten $b \approx p_2'(t_6) \approx 2\hat{a}_2(t_6 - t_0) + \hat{a}_1 = 12\hat{a}_2 + \hat{a}_1 \approx 0.0146$, womit $T_2 = (\ln 2)/b \approx 47.48$. Diese (bessere?) Vorhersage weicht stärker von der der JHU ab. Somit wird die JHU vermutlich auch einen linearen Fit mit dem Polynom p_1 verwendet haben.

Robuste Indizes mit Verzögerung

Man ist bestrebt, den Inzidenzwert oder einen entsprechenden Index so zu modifizieren, dass er für den Verlauf der Pandemie und damit für die zu ergreifenden Maßnahmen aussagekräftiger ist. Ein klassisches COVID-19 Modell wurde bereits in Kapitel 10 vorgestellt. Nun gibt es zwar robuste, verlässliche Werte, wie etwa

- Einweisungen in ein Krankenhaus (sog. Hospitalisierung)
- Anzahl COVID-19 Patienten auf Intensivstationen (intensivpflichtige Patienten)
- Anzahl verfügbarer Intensivbetten (engl. intensive care units, ICUs), inklusive der zugehörigen Fachpflegekräfte
- Anzahl der (an/mit COVID-19) verstorbenen Personen

Allerdings haben diese Werte eine große Verzögerung (z.B. 14 Tage) zu den möglichen Maßnahmen (in der Mathematik: 'Steuerungsvariablen') und sind daher als alleiniger Index nur bedingt geeignet.

Eine Lösung wäre ein Index als (gewichtetes) Mittel der verschiedenen Einzelindizes, vergleichbar mit einem Portfolio von Aktien. Auch auf dem Aktienmarkt möchte man durch diese Streuung das Risiko falsch zu liegen reduzieren. Ein Nachteil dabei ist, dass dieses Vorgehen (mit den sozio-ökonomischen Konsequenzen) für die Bevölkerung weniger transparent ist und die genaue Wahl des Portfolios zu einem gewissen Grad willkürlich ist.

Eine weitere Möglichkeit, die Inzidenz stabiler gegenüber der Anzahl der Testungen zu machen, wäre die *Dunkelziffer der Neuinfektionen* abzuschätzen und zu berücksichtigen. Dazu erweitert man das SIR-Modell aus Kapitel 10 um weitere Kompartimente zur Berücksichtigung von robusten Daten (s.o.) und baut einen Dunkelzifferfaktor ω ein. Nun fittet man das erweiterte Modell an den Daten (Hospitalisierungsraten, Sterbezahlen) und findet auf diesem Weg den Faktor ω der am besten zu den Daten passt und verwendet ihn als Schätzung für die Dunkelziffer⁶.

⁶A. Kergaßner, C. Burkhardt, D. Lippold, et al., *Memory-based meso-scale modeling of COVID-19*, Comput. Mech. 66 (2020), 1069–1079. <https://doi.org/10.1007/s00466-020-01883-5>.

Der EPG-Index

Eine weitere Alternative ist der *EPG-Index* (engl. Effective Potential Growth), den wir hier kurz aus einem Artikel zusammenfassen⁷. Wir beginnen dazu mit dem *Infizierten-Verstorbenen-Anteil* (engl. infection fatality ratio, IFR), der den Anteil der Todesfälle unter allen Infizierten angibt:

$$\text{IFR} = \frac{\text{Anzahl der an COVID-19 Verstorbenen in einem Zeitraum}}{\text{Gesamtzahl der COVID-19 Infektionen im gleichen Zeitraum}}$$

Der IFR schätzt also das Sterberisiko bei allen Infizierten (diagnostiziert, asymptomatisch, nicht getestet) ab.

Zunächst nehmen wir einen Schätzwert $\text{IFR} = 1\%$ (in Europa 0.3–3%) und unter Berücksichtigung der gemeldeten (relativ sicheren) Zahlen der Todesfälle schätzen wir den Zeitraum (*mittlere*) *Zeit bis zum Tod* TtD (engl. time to death) auf $\text{TtD}=18$ Tage (Menschen sterben zwischen 15 und 22 Tagen nach dem Auftreten der ersten Symptome)⁸. Die geschätzte Anzahl der infizierten Personen mit COVID-19 E_t zum Zeitpunkt t , lautet:

$$E_t = \frac{d_{t+\text{TtD}}}{\text{IFR}},$$

wobei $d_{t+\text{TtD}}$ die Anzahl der gemeldeten Todesfälle zum Zeitpunkt $t + \text{TtD}$ ist. Damit können wir die Anzahl der Fälle vor $\text{TtD}=18$ Tagen schätzen und den Wert mit der Anzahl der vor 18 Tagen entdeckten Fälle vergleichen, wodurch wir einen diagnostischen Prozentsatz erhalten.

Allerdings ist dies eine unrealistische untere Grenze, da niemand am ersten Tag der Symptome einen PCR-Test durchführt oder zum Arzt geht. Ebenso können externe Faktoren, wie die Verfügbarkeit von PCR-Tests, Sättigung des Gesundheitssystems, usw. die Diagnose verzögern. Diese *Verzögerung bis zur Diagnose* (DD, engl. delay to detection, 4-14 Tage) entspricht der Zeit zwischen dem Auftreten der ersten Symptome und der Meldung durch das Gesundheitssystem. Diese Zeit ist stark länderspezifisch.

Die Autoren Català, Pino, Marchena, et al. schlagen folgende allgemeine Prozedur für die Berechnung des Prozentsatzes der Diagnose (%D) vor:

1. Mithilfe der obigen Formel kann mit TtD und IFR die Anzahl der Fälle geschätzt werden (alternativ mittels eines gefitteten Modells, siehe voriger Abschnitt)
2. Berechnung der zeitlichen Korrelation (Zusammenhang) zwischen gemeldeten Fällen C_t und gemeldeten Todesfällen d_t . Das Maximum der Korrelation bestimmt die Zeit zwischen (Bestätigungs-)Diagnose und Tod (engl. Diagnosis to Death,

⁷Martí Català, David Pino, Miquel Marchena, et al., *Robust estimation of diagnostic rate and real incidence of COVID-19 for European policymakers*, PLOS, January 7, 2021 <https://doi.org/10.1371/journal.pone.0243701>

⁸F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, et al., *Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study*, Lancet 395 (2020), 1054-1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)

DtD, 4-14 Tage). Die Verzögerung in der Diagnose DD ergibt sich als $DD = TtD - DtD$.

3. Auswertung des prozentualen Anteils der Diagnose %D basierend auf geschätzten (1.) und gemeldeten (2.) Fällen.

Sobald die Diagnoserate %D bekannt ist, ist es einfach, eine reale Inzidenz zu ermitteln, die nicht mehr von (länderspezifischen) Zeitverzögerungen bis zur Diagnose abhängt. Das Niveau der Diagnose und der realen Inzidenz ist nützlich für politische Entscheidungsträger. Da die Diagnoserate %D stark von dem länderspezifischen Wert für DD abhängt, definieren wir auch die Delay-to-Detection-Diagnoserate (DD-DR) als die Diagnoserate %D, die unter Verwendung einer unterschiedlichen Zeitverzögerung zwischen dem Auftreten von Symptomen und Erkennbarkeit für jedes Land berechnet wird.

Mit Hilfe der DD-DR kann man berechnen, wie viele unentdeckte Personen in den letzten 14 Tagen zur Zahl der Infizierten $I_{t,14}$ hinzugekommen sind. Somit kann man die 14-Tage Befallsrate besser abschätzen. Die Befallsrate A (engl. attack rate) ist der Anteil einer Population, die in einem gewissen Zeitraum an der Infektion erkrankt. Die ECDC (Europäisches Zentrum für die Prävention und die Kontrolle von Krankheiten) benutzt in diesem Zusammenhang den Zeitraum der letzten 14 Tage⁹. Das wollen wir auch so tun. Die 14-Tage-Befallsrate zum Zeitpunkt t ist definiert als

$$A_{t,14} = \frac{\text{exponierte Empfängliche einer Gruppe, die in den letzten 14 Tagen erkrankten}}{\text{exponierte Empfängliche in dieser Gruppe insgesamt, in den letzten 14 Tagen}}$$

Anstatt der (*effektiven*) *Reproduktionszahl* R_t , die man von erweiterten SIR-Modellen abschätzt, verwenden wir als empirische Alternative ρ_t , die Anzahl der heute entdeckten neuen Fälle N_t geteilt durch die Anzahl der neuen Fälle, die vor fünf Tagen entdeckt wurden: N_t/N_{t-5} . Die starken Schwankungen dieser Größen erfordern jedoch, dass wir gemittelte Werte verwenden:

$$\rho_t = \sum_{i=-n_d}^{n_d} \frac{N_{t+i}}{N_{t+\tau+i}}$$

Konkret verwenden wir wegen der ausgeprägten Wochen-Saisonalität Werte über sieben Tage ($n_d = 3$) und die Verzögerung $\tau = 5$ Tage, welches ungefähr die Zeit ist, in der infizierte Personen Symptome entwickeln, falls sie welche entwickeln, d.h. wir nehmen

$$\rho_{t,7} = \frac{N_{t-3} + N_{t-2} + N_{t-1} + N_t + N_{t+1} + N_{t+2} + N_{t+3}}{N_{t-8} + N_{t-7} + N_{t-6} + N_{t-5} + N_{t-4} + N_{t-3} + N_{t-2}}$$

⁹European Centre for Disease Prevention and Control, *Rapid Risk Assessment: Coronavirus disease 2019 (COVID-19) in the EU/EEA and the UK – tenth update*. 11th June, 2020. verfügbar: <https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-coronavirus-disease-2019-covid-19-pandemic-tenth-update>.

Diese Rate ist kleiner als 1, wenn die Zahl der neuen Fälle abnimmt und größer als 1, wenn die Anzahl der Fälle zunimmt. Durch Mittelungsprozesse (oder besser Ausgleichsrechnung, s.o.) kann dieser Wert $\rho_{t,7}$ weiter stabilisiert werden und wird mit $\bar{\rho}_{t,7}$ bezeichnet.

Die Autoren Català, Pino, Marchena, et al. schlagen den folgenden täglichen EPG-Index vor:

$$\text{EPG}_t = \bar{\rho}_{t,7} A_{t,14}.$$

Der EPG-Index ist einfach die Multiplikation der (gemittelten) Wachstumsrate der Krankheit $\bar{\rho}_{t,7}$ mit der Schätzung der Befallsrate $A_{t,14}$, beide ausgewertet zum richtigen Zeitpunkt t in der jüngsten Vergangenheit.

Der ungünstigste Fall ist einer, bei dem sowohl $A_{t,14}$, als auch $\bar{\rho}_{t,7}$ groß sind. Das bedeutet, dass man eine große Population mit der Krankheit und viel Ausbreitung früher. Die beste Situation ist ein niedriger Wert der Geschwindigkeit und eine geringe Anzahl von aktiven Fällen. Eine große Anzahl von $A_{t,14}$, mit niedrigem $\bar{\rho}_{t,7}$ oder ein großes $\bar{\rho}_{t,7}$ mit niedrigem $A_{t,14}$ sind potenziell gefährliche Situationen.