

Epidemiologische Betrachtungen zur Covid-19-Pandemie 2020

8. März 2020

Inhaltsverzeichnis

1 Theorie	1
1.1 Die Reproduktionszahl	1
1.2 Das SIR-Modell	1
1.3 Ermittlung der Modellparameter	2
1.4 Das SEIR-Modell	2
1.5 Ermittlung der Modellparameter	3
2 Daten	4
2.1 Daten zur Fallsterblichkeit	4
2.2 Dynamik der Fallzahlen	4
2.3 Daten zur Basisreproduktionszahl . . .	4
3 Abbildungen	5

Inhaltsangabe. Die Epidemiologischen Grundmodelle SIR und SEIR werden einer Untersuchung unterzogen. Darauf aufbauend werden Beziehungen zwischen den Modellparametern in der Anfangsphase und Methoden zur Ermittlung dieser abgeleitet, das Letztere unter der Prämisse dass hinreichend viele gute Daten zur Verfügung stünden. Eine anschließende Diskussion soll Einblick bringen, wann und ob eine Ermittlung von Parametern ohne fortgeschrittene epidemiologische und statistische Werkzeuge überhaupt möglich ist. Im Weiteren folgt eine Betrachtung von Parametern zu Covid-19 aus aktuellen Arbeiten, auch mit Blick auf Variabilität bezüglich Sommer und Quarantäne. Schließlich soll aufbauend auf all diesem Wissen mit dem SEIR-Modell der Versuch einer Prognose unternommen werden.

1 Theorie

1.1 Die Reproduktionszahl

Die *Basisreproduktionszahl* R_0 ist die Anzahl der Individuen, die ein infiziertes Individuum im Mittel infiziert. Die tatsächliche Anzahl wird allerdings durch die *Nettoreproduktionszahl*

$$R_q = R_0(1 - q). \quad (1.1)$$

beschrieben. Hierbei ist q der immune Anteil der Bevölkerung. Damit die Epidemie zum Stillstand kommt,

muss $R_q < 1$ sein. Aus dem Ansatz $R_q = 1$ ergibt sich

$$q_c := q = 1 - \frac{R_q}{R_0} = 1 - \frac{1}{R_0}. \quad (1.2)$$

Ist dieser Anteil q_c der Bevölkerung immun, bricht die Epidemie also gar nicht erst aus. Man spricht von der *kritischen Immunisierungsschwelle* oder *Schwelle zur Herdenimmunität*.

Für bestimmte Krankheiten lässt sich R_0 abschätzen, wobei dieser Wert allerdings empfindlich vom Verhalten der Bevölkerung und eventueller Quarantäne abhängt.

1.2 Das SIR-Modell

Die Bevölkerung wird aufgeteilt in die Anteile S, I, R . Hierbei sind S die Anfälligen (engl. susceptibles), I die Infizierten (engl. infected) und R die Erholten (engl. recovered). Die Anteile S, I, R sind normiert, so dass $S + I + R = 1$ gilt. Die absolute Zahl ergibt sich durch Multiplikation des jeweiligen Anteils mit der Bevölkerungszahl N .

In einem Zeitschritt von t zu $t + h$ mit $h := \Delta t$ wird es zu einer Zunahme von I kommen. Diese Zunahme ist zunächst davon abhängig, wie viele Infizierte I und Anfällige S es gerade gibt. Sei S fest. Gäbe es zwischen den Anfälligen doppelt so viele Infizierte, würde dies zu der doppelten Zunahme führen. Sei nun umgekehrt I fest. Gäbe es nur halb so viele Anfällige, wäre es nur halb so wahrscheinlich, dass ein Infizierter einen Anfälligen infiziert. Demnach sollte die Zunahme proportional zu I und S sein, also zum Produkt IS . Die Proportionalitätskonstante sei β . Wir machen also den Ansatz

$$I_{t+h} = I_t + h\beta I_t S_t. \quad (1.3)$$

Man kann auch $h = 1$ setzen. Für die Anfälligen gilt dementsprechend $S_{t+h} = S_t - h\beta I_t S_t$. Nun muss man schließlich noch beachten, dass sich Infizierte mit der Zeit erholen. Die Abnahme der Infizierten ist proportional zur Zahl der Infizierten, unter der Annahme dass die Zahl aus Altinfektionen und Neuinfektionen besteht. Die Proportionalitätskonstante sei γ . Nun ergibt sich insgesamt das Modell

$$S_{t+h} = S_t - h\beta I_t S_t, \quad (1.4)$$

$$I_{t+h} = I_t + h\beta I_t S_t - h\gamma I_t, \quad (1.5)$$

$$R_{t+h} = R_t + h\gamma I_t. \quad (1.6)$$

Dies ist ein diskretes dynamisches System, gegeben durch ein nichtlineares System von Differenzgleichungen.

chungen. Umformung der jeweiligen Gleichung bringt

$$\frac{S_{t+h} - S_t}{h} = -\beta I_t S_t \quad (1.7)$$

$$\frac{I_{t+h} - I_t}{h} = \beta I_t S_t - \gamma I_t, \quad (1.8)$$

$$\frac{R_{t+h} - R_t}{h} = \gamma I_t. \quad (1.9)$$

Für $h \rightarrow 0$ ergibt sich

$$S' = -\beta IS, \quad (1.10)$$

$$I' = \beta IS - \gamma I, \quad (1.11)$$

$$R' = \gamma I. \quad (1.12)$$

Dies ist ein kontinuierliches dynamisches System, gegeben durch ein nichtlineares System von Differentialgleichungen. Aufgrund von $S + I + R = 1$ gilt

$$S' + I' + R' = 0. \quad (1.13)$$

Man bezeichnet β als *Erkrankungsrate* und γ als *Genesungsrate*.

Angenommen es ist $I \approx 0$, aber $I > 0$. Unter welchem Umstand gibt es dann keine Neuerkrankungen? Dazu muss $I' = 0$ sein. Dieser Ansatz führt zu

$$0 = \beta IS - \gamma I \Leftrightarrow 0 = \beta S - \gamma \Leftrightarrow S = \frac{\gamma}{\beta}. \quad (1.14)$$

Aufgrund von $R + S \approx 1$ ist $q = R = 1 - S$. Einsetzen von q in Gleichung (1.1) bringt

$$R_q = R_0(1 - q) = R_0 S = R_0 \frac{\gamma}{\beta}. \quad (1.15)$$

Weil $R_q = 1$ sein muss, ergibt sich die Beziehung

$$R_0 = \frac{\beta}{\gamma}. \quad (1.16)$$

Interessant ist nun das Grenzverhalten

$$R(\infty) := \lim_{t \rightarrow \infty} R(t), \quad (1.17)$$

das eine Aussage macht, wie viele Individuen die Erkrankung bei Abwesenheit von Quarantäne durchmachen werden. Wie sich aus einer numerischen Simulation ergibt, führt ein Startzustand $S \approx 1$ und $I \approx 0$ zu $R(\infty) > q_c$. Bei $R_0 = 3$ ist z. B. $R(\infty) = 94\%$, obwohl $q_c = 66\%$ ist. Siehe Abb. 1.

1.3 Ermittlung der Modellparameter

Für das SIR-Modell sind die Parameter sicherlich bestimmbar, indem die numerische Lösung mit Messdaten verglichen wird. Variation der Parameter β und γ führt zu unterschiedlichen Fehlerquadratsummen, von diesen wird die minimale ermittelt.

Unabhängig vom genannten Ansatz folgen noch analytische Betrachtungen, womit sich die Parameter unter Umständen mit weniger Aufwand ermitteln lassen.

Am Anfang steigt die Zahl der Infektionen exponentiell. Die Kurve $I(t)$ ist in dieser Phase also beschrieben durch $I = I_0 e^{\lambda t}$, wobei man λ nach Linearisierung aus einer Regression ermitteln kann. Es gilt $I' = \lambda I$. Für die Dgl. (1.11) ergibt sich die Faktorisierung

$$I' = \beta IS - \gamma I = (\beta S - \gamma)I. \quad (1.18)$$

Am Anfang der Phase ist in guter Näherung $S \approx 1$, vor allem bei großer Bevölkerungszahl. Demnach gilt

$$\lambda = \beta - \gamma. \quad (1.19)$$

Setzt man $\gamma = \beta/R_0$ ein, gelangt man zu

$$\lambda = \beta \left(1 - \frac{1}{R_0}\right) = \beta q_c. \quad (1.20)$$

Nun ist es allerdings schwierig, Daten für I zu bekommen. Was die WHO täglich herausgibt, ist die Anzahl der bestätigten Fälle, das entspricht $C := 1 - S = R + I$. Die Zahl der WHO ist nicht die Zahl der tatsächlichen Fälle, sollte aber zu ihr proportional sein. Der Proportionalitätsfaktor ergibt sich aus der ermittelten Fallsterblichkeit.

Trotzdem können wir an λ gelangen. Integration von Dgl. (1.12) bringt nämlich

$$R = \int_0^t \gamma I dt = \gamma I_0 \int_0^t e^{\lambda t} dt \quad (1.21)$$

$$= \frac{\gamma}{\lambda} I_0 e^{\lambda t} = \frac{\gamma}{\lambda} I. \quad (1.22)$$

Daher ist

$$C = R + I = \frac{\gamma}{\lambda} I + I = \left(\frac{\gamma}{\lambda} + 1\right) I_0 e^{\lambda t}. \quad (1.23)$$

Eine Regression der Daten zu C ergibt folglich die gleiche Vermehrungskonstante λ .

1.4 Das SEIR-Modell

Das SEIR-Modell stellt eine Verfeinerung des SIR-Modells dar, indem berücksichtigt wird, dass sich ein Individuum in der Phase E befinden kann, in welcher es

zwar schon infiziert ist, jedoch noch nicht infektiös. Die Gleichungen sind

$$S' = -\beta IS, \quad (1.24)$$

$$E' = \beta IS - aE, \quad (1.25)$$

$$I' = aE - \gamma I, \quad (1.26)$$

$$R' = \gamma I. \quad (1.27)$$

Entsprechend gilt wieder

$$S + E + I + R = 1, \quad (1.28)$$

$$S' + E' + I' + R' = 0. \quad (1.29)$$

Der Kehrwert $1/a$ ist die mittlere *Latenzzeit*, also die mittlere Zeit, welche ein Individuum in der Phase E verbringt. Die mittlere Latenzzeit kann je nach Krankheit kleiner oder größer als die Inkubationszeit sein.

Gleichung (1.16) bleibt auch im SEIR-Modell gültig, denn der Ansatz $I' = 0$ und $E' = 0$ führt wieder zu Gleichung (1.14).

1.5 Ermittlung der Modellparameter

Schon bekannte Abschätzungen für Parameter kann man auch der Literatur entnehmen, das betrifft z.B. R_0 und a . Verbleibende Parameter versucht man bezüglich einem Vergleich der numerischen Lösung mit den Messdaten zu gewinnen. Dies geschieht wie üblich mit der Methode der kleinsten Quadrate. Zur Suche des Minimums kann man einfach brachial absteigendes Hill-Climbing ansetzen. Mit wachsender Anzahl freier Parameter könnte das Problem allerdings konkav werden (d. h. weitere ungewünschte lokale Minima enthalten), ich habe mich damit noch nicht näher beschäftigt.

Man würde auch gerne in Erfahrung bringen, ob man für vorhandene Messdaten überhaupt in der Lage ist, eine Vorhersage zu treffen, bzw. welche Genauigkeit dabei erreicht wird. Auch ohne große Kenntnisse in Statistik gibt es dafür eine Methodik. Nämlich lässt sich das Modell auch zur *Erzeugung* von Messdaten benutzen, wobei man entsprechend etwas statistisches Rauschen einfügt, wie es auch bei den echten Messdaten vorhanden ist. Für die erzeugten Messdaten sind nun die ursprünglichen wahren Parameter bekannt. Demnach können die ermittelten mit den wahren Parametern verglichen werden.

Es folgen wieder analytische Betrachtungen, mit welchen sich Parameter unter Umständen mit weniger Rechenaufwand ermitteln bzw. abschätzen lassen.

Auch im SEIR-Modell stimmen die Vermehrungskonstanten von I und C für den Ansatz $I = I_0 e^{\lambda t}$ überein, denn man bekommt

$$\begin{aligned} C &= 1 - S = R + E + I = \frac{\gamma}{\lambda} I + \frac{\lambda I + \gamma I}{a} + I \\ &= \left(\frac{\gamma}{\lambda} + \frac{\lambda + \gamma}{a} + 1 \right) I = \frac{C(0)}{I(0)} I. \end{aligned}$$

Zur Berechnung von β benötigen wir eine entsprechende Beziehung wie (1.19), diese gestaltet sich nun allerdings etwas komplizierter. Die Auffindung dieser Beziehung gestaltet sich auch ein klein wenig schwieriger. Die Gleichungen müssen hierzu so zurechtgeformt werden, dass I wieder herausgekürzt werden kann. Eine solche Form gewinnt man unvermittelt aus der letzten Gleichung für C . Ableiten dieser bringt nämlich

$$C' = -S' = \frac{C(0)}{I(0)} I' = \frac{C(0)}{I(0)} \lambda I. \quad (1.30)$$

In Verbindung mit Dgl. (1.24) gewinnt man

$$\beta IS = \frac{C(0)}{I(0)} \lambda I, \quad (1.31)$$

und somit

$$\beta S = \frac{C(0)}{I(0)} \lambda = \left(\frac{\gamma}{\lambda} + \frac{\lambda + \gamma}{a} + 1 \right) \lambda, \quad (1.32)$$

bzw.

$$\beta S = \left(\frac{\lambda}{a} + 1 \right) (\lambda + \gamma). \quad (1.33)$$

Wegen $S \approx 1$ in sehr guter Näherung setzt man wieder $S = 1$ ein, außerdem $\gamma = \beta/R_0$. Umformung nach β führt schließlich zur gewünschten Gleichung

$$\beta \left(\frac{a}{\lambda + a} - \frac{1}{R_0} \right) = \lambda. \quad (1.34)$$

Diese Beziehung lässt sich auch wieder bezüglich der kritischen Immunisierungsschwelle ausdrücken:

$$\beta \left(q_c - \frac{\lambda}{\lambda + a} \right) = \lambda. \quad (1.35)$$

Als Nebenresultat können wir das Verhältnis von Infektiösen zu aufgetretenen Fällen abgreifen:

$$\frac{I}{C} = \frac{\lambda}{\beta} = \left(\frac{a}{\lambda + a} - \frac{1}{R_0} \right). \quad (1.36)$$

Mit Dgl. (1.26) bekommt man auch noch

$$\frac{I}{E} = \frac{a}{\lambda + \gamma} = \frac{a}{\lambda + \beta/R_0}. \quad (1.37)$$

2 Daten

2.1 Daten zur Fallsterblichkeit

Tabelle 1: Bestätigte Covid-19-Fälle
nach Daten der WHO.
Kumuliert, am 2. März 2020.

Land	Fälle	Tote	Verhältnis
VR China	80134	2914	3.6 %
Iran	978	54	5.5 %
Italien	1689	35	2.1 %
Japan	960	12	1.2 %
Südkorea	4212	22	0.52 %
Rest	975	6	0.61 %
Welt	88948	3043	3.4 %
bereinigt*	6147	31	0.50 %

*Welt abzüglich VR China, Iran und Italien.

2.2 Dynamik der Fallzahlen

Deutschland

Tabelle 2: Fälle in Deutschland
nach Daten des RKI.

Tag	Datum	Fälle	neue
0	24. Feb.	16	0
1	25. Feb.	18	2
2	26. Feb.	21	3
3	27. Feb.	26	5
4	28. Feb.	53	27
5	29. Feb.	66	13
6	01. Mär.	117	51
7	02. Mär.	150	33
8	03. Mär.	188	38
9	04. Mär.	240	52
10	05. Mär.	349	109
11	06. Mär.	534	185
12	07. Mär.	684	150
13	08. Mär.	847	163
14	09. Mär.	1112	265

Regression für 24. Feb. bis 06. Mär.:

$$C = C_0 e^{\lambda t}, \quad C_0 = 12.83, \quad \lambda = 0.3353$$

Regression für 24. Feb. bis 09. Mär.:

$$C_0 = 13.38, \quad \lambda = 0.3251$$

2.3 Daten zur Basisreproduktionszahl

Tabelle 3: Ermittelte Werte für R_0 mit
Intervall der Unsicherheit.

R_0	Intervall	Quelle
2.6	1.5 bis 3.5	Ferguson für Wuhan [2]
2.2	1.4 bis 3.9	Qun Li für Wuhan [3]

3 Abbildungen

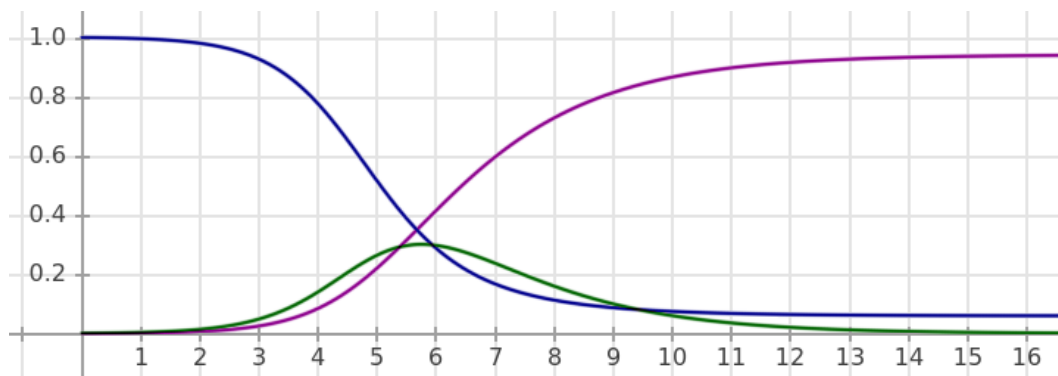


Abbildung 1: Simulation zum SIR-Modell. S in blau, I in grün, R in magenta. (→[Link](#))

Anfangswerte: $S := 0.999, I := 0.001, R := 0$.

Parameter: $\beta := 2$ und $\gamma := \beta/R_0$ mit $R_0 := 3$.

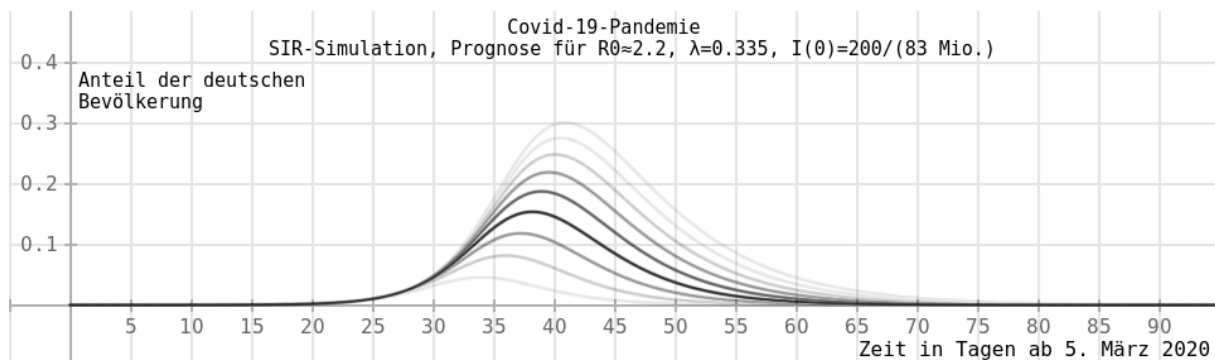


Abbildung 2: Naive Prognose für den Anteil der infektiösen Bevölkerung in Deutschland.

Szenario, falls die Vermehrungskonstante λ so hoch bleiben sollte wie bisher.

Kurven für R_0 von 1.4 bis 3.0, die schwarze Kurve für $R_0 = 2.2$.

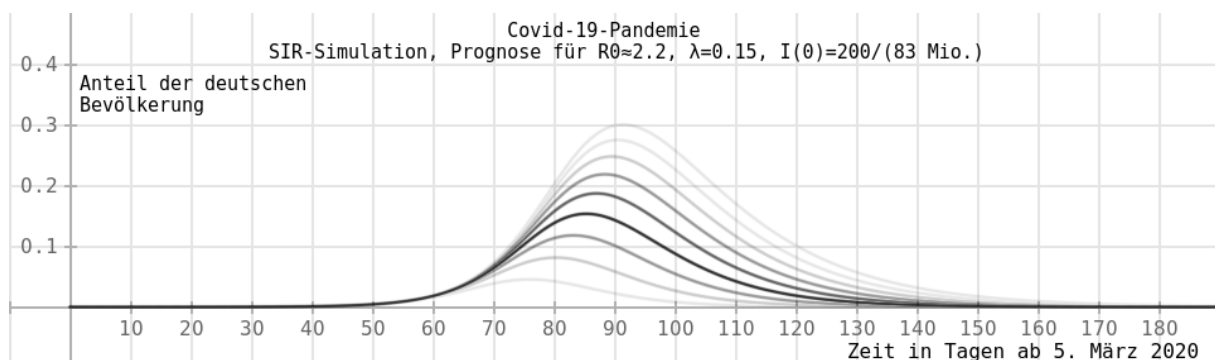


Abbildung 3: Naive Prognose für den Anteil der infektiösen Bevölkerung in Deutschland.

Szenario, falls λ aufgrund von Sommer und Quarantäne absinken wird.

Kurven für R_0 von 1.4 bis 3.0, die schwarze Kurve für $R_0 = 2.2$.

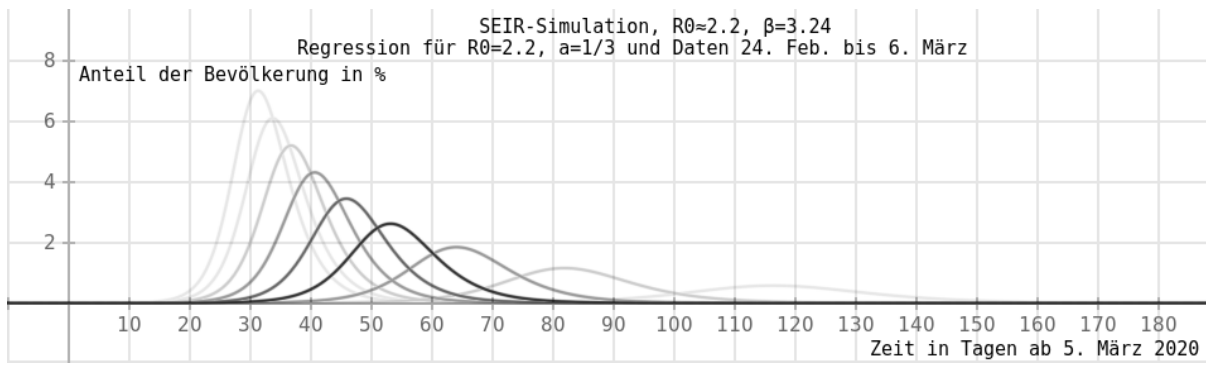


Abbildung 4: Prognose vom 6. März für den Anteil der infektiösen Bevölkerung in Deutschland.

Kurven für R_0 von 1.4 bis 3.0, die schwarze Kurve für $R_0 = 2.2$.

Regression für eine Latenzzeit von drei Tagen.

Literatur

- [1] Matt J. Keeling, Pejman Rohani: »*Modeling Infectious Diseases in Humans and Animals*«. Princeton University Press, 2008.
- [2] Neil M. Ferguson, Natsuko Imai, et al: »*Report 3: Transmissibility of 2019-nCoV*«. WHO Collaborating Centre for Infectious Disease Modelling, MRC Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London (22. Jan. 2020).
- [3] Qun Li, et al: »*Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia*«. In: The New England Journal of Medicine (29. Jan. 2020). DOI: 10.1056/NEJMoa2001316.