

Theoretischer Rahmen einer ökonometrischen Analyse der deutschen Corona-Data

von Georg Quaas (vom 4. April 2020, überarbeitet und ergänzt am 3. Mai 2020)

Das klassische SIR-Modell

Auf der Basis von täglichen Daten wird das klassische epidemiologische Modell (im Weiteren nach der englischen Bezeichnung CEM genannt) durch folgende Gleichungen (Hethcote 2000: 604) formuliert:

$$(1) \Delta S = -\beta IS/N$$

$$(2) \Delta I = \beta IS/N - \Delta R$$

$$(3) \Delta R = \gamma I$$

S , I und R sind die Zahlen für die Größe von drei überlappungsfreien Klassen der Bevölkerung, S ist die Zahl der Menschen, die für das Virus empfänglich sind (zu Beginn der Krankheit und ohne Informationen über die Anzahl der immunen Menschen ist sie identisch mit der Anzahl der Bevölkerung), I ist die Zahl der infektiösen Menschen (ihr Anfangswert liegt in der Nähe von null, zum Beispiel = 1) und R ist die Zahl der geheilten Menschen mit Immunität (diese Zahl ist für die ersten 10 Tage der empirischen Untersuchung auf null gesetzt worden).

Die Logik hinter diesem Modell wird deutlich, wenn man die Proportionen der beiden Bevölkerungsgruppen mit und ohne aktive Krankheitserreger in der Gesamtpopulation betrachtet: $i = I/N$ und $s = S/N$. Diese Proportionen können als Wahrscheinlichkeiten interpretiert werden, in der Bevölkerung vom Umfang N zufällig auf einen Infektiösen oder zufällig auf einen für das Virus Empfänglichen zu stoßen. Die Wahrscheinlichkeit, dass diese beiden Personen sich treffen, hat die Wahrscheinlichkeit $i \cdot s$ (Hamer 1906). Das epidemiologische Modell geht davon aus, dass die Häufigkeit von Infektionen mit der Größe der Bevölkerung zunimmt, so dass sich mit Ausnahme von β der mittlere Term der Gleichung (2) ergibt, dem wir mit H abkürzen:

$$(4) \quad i s N = \frac{ISN}{N^2} = \frac{IS}{N} = H$$

Aus praktischer Sicht kann H als die Anzahl der Möglichkeiten interpretiert werden, sich in einer Population mit dem Umfang N anzustecken, es ist sozusagen die abstrakte Risikosituation, der man in einer Seuche ausgesetzt ist. Der Parameter β stellt den tatsächlichen Ansteckungsprozess unter dem Gesichtspunkt dar, wie viele Menschen ein Infektiöser pro Zeiteinheit (hier: pro Tag) im Durchschnitt ansteckt. Das hängt von mehreren Faktoren ab, die im Modell nicht explizit enthalten sind, wie z. B. die Bevölkerungsdichte, die Anzahl der täglichen Interaktionen und das allgemeine Verhalten der Bevölkerung (Hygiene, Händeschütteln usw.). Einige dieser Faktoren können pragmatisch beeinflusst werden, so dass politische Entscheidungsträger Zugang zum laufenden Prozess haben. Der Effekt von gesundheitspolitischen Maßnahmen wird durch β mit einer Zeitverzögerung indiziert. Im Falle des neuartigen Coronavirus gingen die Medien zunächst davon aus, dass jede infizierte Person während ihrer infektiösen Phase drei weitere Menschen infiziert. Das RKI meldet in

seinem „Profil“ des Virus‘ eine Basisreproduktionszahl R_0 zwischen 2 und 2,5. Erste Erfahrungen deuten darauf hin, dass die infektiöse Phase etwa 10 Tage dauert, so dass im Durchschnitt ein anfängliches $\beta = 0.3$ festgelegt werden kann – solange keine genaueren Schätzungen verfügbar sind.

Es gibt eine vierte Variable, die in dem oben formulierten klassischen Modell keine explizite Rolle spielt, und das ist die Anzahl der Todesfälle D . Es wird oft angenommen (z.B. an der Heiden & Buchholz 2020: 1), dass die Zahl der Todesfälle in der Zahl der Genesenen enthalten ist. Das klingt ein wenig zynisch – aber aus statistischer Sicht ist ein anderer Aspekt wichtig: Wenn die Zahl der Todesfälle empirisch verfügbar ist, sollte das Modell damit ergänzt werden. Sowohl die Zahl der Todesfälle als auch die Zahl der Genesenen verringern die Zahl der infektiösen Menschen; daher kann Gleichung (2) präziser formuliert werden durch:

$$(2a) \Delta I = \beta IS/N - \Delta R - \Delta D$$

Die Anzahl der Genesenen wird durch Gleichung (3) bestimmt, wenn sie nicht bereits vom Datenproduzenten geschätzt worden ist. Für die Zahl der neu gemeldeten Todesfälle gilt die Annahme:

$$(4) \Delta D = \delta I.$$

Der Parameter δ misst die aktuelle Anzahl der neu gemeldeten Todesfälle basierend auf der Anzahl der Infizierten zu einem früheren Zeitpunkt.

Es gibt eine Ähnlichkeit der Gleichung (2a) mit der entsprechenden Gleichung des klassischen *endemischen* Modells mit vitaler Dynamik, die Geburten und Todesfälle berücksichtigt, aber keine krankheitsbedingten Todesfälle (Hethcote 2000: 607). Nach Gleichung (1) wird die Klasse der empfänglichen Personen um die Anzahl der Neuinfektionen verringert, zu denen die (später eintretenden) Todesfälle gehören. Daher darf man diese Gleichung nicht mit Hilfe von (4) ändern. Aber die Zahl der Todesfälle verringert kurzfristig die Zahl der Bevölkerung. Daher ist es im Falle eines plötzlichen und heftigen Seuchenereignisses mit einer erwartungsgemäß hohen Anzahl von Opfern sinnvoll, die Bevölkerungszahl als variabel zu betrachten, indem man die Zahl der Todesfälle berücksichtigt und die Bevölkerungszahl wie alle anderen Variablen als zeitabhängig behandelt wird:

$$(5) S(t) + I(t) + R(t) = N - D(t) = N'(t)$$

Dies ist ein übrigens ein weiterer Unterschied zum klassischen endemischen Modell.

Bei der Anpassung des Modells an die vom Robert-Koch-Institut zur Verfügung gestellten Daten ist es zwingend erforderlich, einzelnen Variablen Zeitverzögerungen – den Merkmalen der zugrunde liegenden Krankheit entsprechend – hinzuzufügen. Zum Beispiel bezieht sich die Zahl der Opfer auf einen Zeitpunkt, der vielleicht 2, 3 oder sogar 4 Wochen zurückliegt. Eine Infektion endet auf jeden Fall entweder mit einer Genesung oder mit dem Tod. Wenn man davon ausgehen könnte, dass sich diese beiden abschließenden Ereignisse auf denselben Zeitpunkt bezögen, würde aus logischen Gründen Folgendes gelten:

$$\delta = (1 - \gamma)$$

Da jedoch die Rekonvaleszenz häufiger auftritt und auch früher als der Tod eines Infizierten registriert wird, haben die Parameter δ und γ unterschiedliche Zeitverzögerungen und sind somit unabhängig voneinander zu bestimmen.

Die Bestimmung der Zeit-Verzögerungen ist vor allem ein empirisches Problem der Virologen. Im Fall von COVID-19 sind nicht alle Merkmale bekannt, mit denen das Modell „gefüttert“ werden muss. Für die unbekannt Parameter kann man in diesem Fall nur plausible Annahmen treffen. Besonders problematisch ist die scheinbar schwer zu bestimmende Infektiositätsdauer. Außerdem sei darauf hingewiesen, dass die vom RKI berichteten Infizierten (besser: die Zahl der von den Gesundheitsbehörden gemeldeten Personen) nicht der Zahl der infektiösen Personen $I(t)$ entspricht, die im klassischen Modell eine Rolle spielen.

Die Kontaktrate

Der Kernmechanismus in diesem Modell ist die Übertragung des Virus von einer Person auf eine andere. Sie wird durch die Kontaktrate erfasst:

“If β is the average number of adequate contacts (i.e., contacts sufficient for transmission) of a person per unit time, then $\beta I/N = \beta i$ is the average number of contacts with infectives per unit time of one susceptible, and $(\beta I/N)S = \beta Nis$ is the number of new cases per unit time due to the $S = Ns$ susceptibles.” (Hethcote 2000: 602)

Multiplizieren wir die Kontaktrate β mit der Expositionszeit $T (=1/\gamma)$, das ist nach Hethcote (605) die durchschnittliche Dauer der infektiösen Phase eines Infizierten), erhalten wir die Zahl der Personen R , die im Durchschnitt von einem infektiösen Infizierten angesteckt werden (die replacement number - Zahl der Ersatzmänner, -frauen und -kinder, die oft auch „Reproduktionszahl“ genannt wird). Bedauerlicherweise werden bei Hethcote beide Zahlen mit R abgekürzt, die Zahl der Genesenen und die Reproduktionszahl, so dass der Leser gehalten ist, den Kontext zu beachten.

“The replacement number R is the actual number of secondary cases from a typical infective, so that after the infection has invaded a population and everyone is no longer susceptible, R is always less than the basic reproduction number R_0 .” (A.a.O.: 604)

Nach einer Weile ist ein Infizierter (i) genesen oder (ii) tot. In beiden Fällen ist er nicht mehr ansteckend. Es wird im Fall von COVID-19 angenommen, dass genesene Menschen immun sind, zumindest für eine Weile.

Die Sterblichkeitsrate

Nach Angaben des chinesischen Emergency Response Epidemiology Team (2020) beträgt die Fallsterblichkeitsrate 2,3 Prozent der bestätigten Fälle und die Dauer der Welle (vom Beginn der massiven Ausbreitung bis zum vermeintlichen Ende) wird im Durchschnitt auf 15 Tage geschätzt (a.a.O.: 115).

Das Response Team betrachtet die erste Welle der Epidemie als beendet. Mit dieser Annahme können die Todesrate und die Dauer der Krankheit wie folgt definiert werden:

“Case fatality rates were calculated as the total number of deaths (numerator) divided by the total number of cases (denominator), expressed as a percent. Observed time was summarized using person-days (PD) and mortality was calculated as the number of deaths (numerator) divided by the total observed time (denominator), expressed per 10 PD.”
(a.a.O.: 114-115)

In einer dynamischen Situation ist die Beurteilung der Todesrate komplizierter, da das tödliche Ereignis X Tage nach der Diagnose der Krankheit eintritt. Division der aktuellen Zahl der Todesfälle durch die aktuelle Zahl der Infizierten führt zu einer irreführenden Unterschätzung der Sterberate am Beginn einer Epidemie. Daraus folgt, dass der Nenner die Anzahl der infizierten Personen sein sollte, die X Tage zuvor beobachtet wurden.

Die Reproduktionszahl

Siehe Studie 1 auf dieser Website: www.forschungsseminar.de

Literatur

an der Heiden, M; Buchholz U (2020): Modellierung von Beispielszenarien der SARS-CoV-2-Epidemie 2020 in Deutschland. | DOI 10.25646/6571.2

Hamer, W. H. (1906): Epidemic Disease in England, Lancet, 1, pp. 733-739.

Hethcote, Herbert W. (2000): The Mathematics of Infectious Diseases, in: SIAM Review, Vol. (Dez., 2000), S. 599-653.

[Täglicher] Lagebericht des RKI

The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team (2020): The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China, 2020, China CDC Weekly, Vol. 2, No. 8, S. 113-122.

Daten

Robert Koch-Institut: [Tägliche] Fallzahlen in Deutschland... [vom 4.3.2020 bis zum aktuellen Rand]