Analyse der Coronastatistiken. Teil 2

Hans-Gert Gräbe, Leipzig

Version vom 13. April 2020

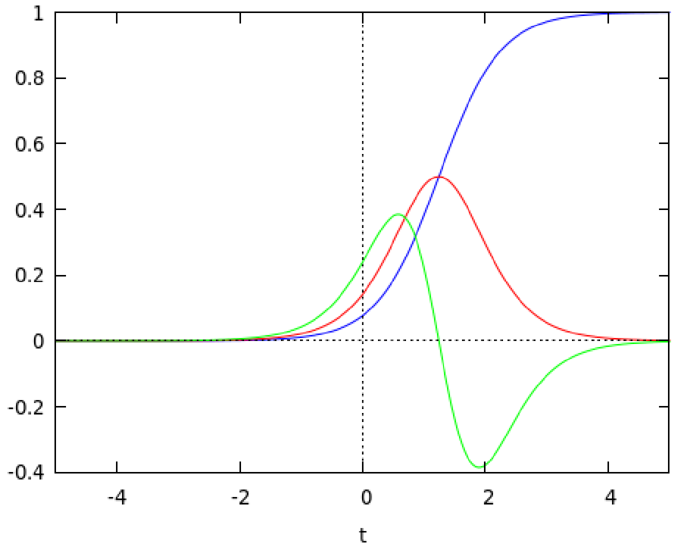
Dieser Text ist eine Fortschreibung des ersten Teils. Die dortigen Beschreibungen der allgemeinen Rahmenbedingungen werden als bekannt vorausgesetzt.

1 Logistische Funktion

Generell ist ein Modell auf der Basis einer *Logistischen Funktion*

$$u(t) = \frac{K}{1 + C \cdot \exp(-rt)} \tag{L.1}$$

die anerkanntere Form der Modellierung der Ausbreitung einer Infektion, siehe dazu den entsprechenden Wikipedia-Eintrag.



Logistische Kurve $u(t) = \frac{1}{1+12\exp(-2t)}$ (blau) sowie deren erste (rot) und zweite Ableitung (grün) Ableitung

K steht dabei für die Sättigungsgrenze $\lim_{t\to\infty} u(t)$ und C ist üblicherweise als $C = \frac{K}{u(0)} - 1$ angeschrieben, was sich unmittelbar aus der Umstellung der Formel für $u(0)$ nach C ergibt. Der Wendepunkt dieser Funktion und damit das Maximum der ersten Ableitung liegt als

Nullstelle der zweiten Ableitung bei $t_0 = \frac{\log(C)}{r}$, was sich unmittelbar mit Maxima berechnen lässt. Dies ist gerade der Median der Kurve, sollte also mit dem oben berechneten m zusammenfallen.

Derartige Funktionen lassen sich aber deutlich schlechter schätzen als Funktionen, die sich wie oben auf einfache Weise auf einen polynomialen Zusammenhang reduzieren lassen, da sie inhärent transzendent sind. Siehe hierzu aber die Arbeit von (Engel 2010) und die Modellierung mit GeoGebra in (Elschenbroich 2020).

In (Engel 2010) wird insbesondere darauf hingewiesen, dass sich mit einer guten Schätzung von K die anderen beiden Parameter mit einem linearen Fitting bestimmen lassen. Wir setzen dazu $C = \exp(c)$, womit sich Formel $(L.1)$ zu

$$\log\left(\frac{K}{u(t)} - 1\right) = c - rt \tag{L.2}$$

umstellen lässt. Der Parameter K ist dabei manuell zu schätzen, so dass die gefittete Kurve möglichst gut auf die Daten passt.

Im Skript ist das Ganze in einer Funktion `lFit(G,K0)` implementiert, der eine Liste G zu übergeben ist, in der vorab alle Datenpunkte mit $y_t \leq 10$ ausgefiltert werden. Es ergeben sich folgende Schätzungen für die Zahl der (in der Statistik erfassten) infizierten Personen (Stand 10.04.2020):

Land	K	c	r	c/r
Deutschland	125000	18.058	0.198	91.11
Italien	145000	17.445	0.207	84.47
Österreich	13000	23.053	0.268	85.90
Spanien	160000	23.815	0269	88.42
China (Hebei)	70000	4.664	0.103	45.12

2 Die „Verdoppelungsdebatte“

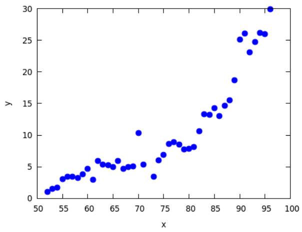
Anfang April 2020 kommt eine Diskussion hoch, dass man die rigiden Beschränkungen erst aufheben könne, wenn „die Verdopplungszeit der Infektionen größer als 14 Tage“ sei.

So meldet zum Beispiel der Deutschlandfunk am 04.04.2020[1]

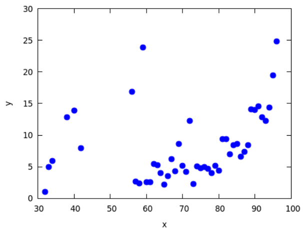
> Die Verdopplungszeit der Ausbreitung von Coronavirus-Infektionen in Deutschland hat sich in den vergangenen Tagen verlangsamt.
>
> Für ganz Deutschland liegt sie nun bei 11,2 Tagen. Die Lage in den Bundesländern ist unterschiedlich. In den großen Flächenländern liegt die Verdopplungszeit in Nordrhein-Westfalen bei 13,1 Tagen, in Baden-Württemberg bei 12,5 Tagen und in Bayern bei 9,7 Tagen. In Berlin sind es inzwischen 12,8 Tage, in Hamburg 12,4. Im Saarland hingegen liegt die Verdoppelungszeit bei 5,5 Tagen, in Sachsen bei 11,0 Tagen.

[1] `https://www.deutschlandfunk.de/covid19-verdopplungszeit-der-coronavirus-infektionen-in.1939.de.html?drn:news_id=1117169`

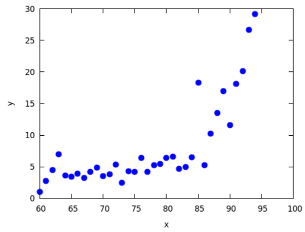
Auch wenn dies nicht immer deutlich wird, bezieht sich die Verdopplungszeit v_t auf die kumulierten Daten und steigt deshalb bereits durch die schiere Masse der Infizierten. Ist $y(t) = mt + n$ ein linearer Zusammenhang, so ergibt sich $v_t = \frac{y(t)}{m}$. Für einen annähernd linearen Zusammenhang kann man also $v_t = \frac{y(t)}{y'(t)}$ als Schätzung nehmen. Die Zahl lässt sich auch aus unseren Daten leicht berechnen: Ist y_t die kumulierte Zahl der Infizierten am Tag t und d_t die Zahl der Neuinfektionen, so ist nach $v_t = \frac{y_t}{d_t}$ Tagen eine Verdopplung der Infizierten erreicht, die Zuwachsrate d_t über diesen Zeitraum als konstant vorausgesetzt. Beide Datenreihen (Stand 10.04.2020) hatten wir schon oben extrahiert, so dass wir eine einfache Funktion `doublePlot(Land)` schreiben können, um die folgenden Plots zu erzeugen:



Italien



Deutschland



Österreich

3 Literatur

- Hans-Jürgen Elschenbroich. Corona: Mathematik & Modellbildung. `https://www.geogebra.org/m/cfammtpe`. 2020.
- Joachim Engel. Parameterschätzen in logistischen Wachstumsmodellen. Stochastik in der Schule 30 (2010) 1, S. 13–18.