

Analyse der Coronastatistiken. Teil 2

Hans-Gert Gräbe, Leipzig

Version vom 04.06.2020

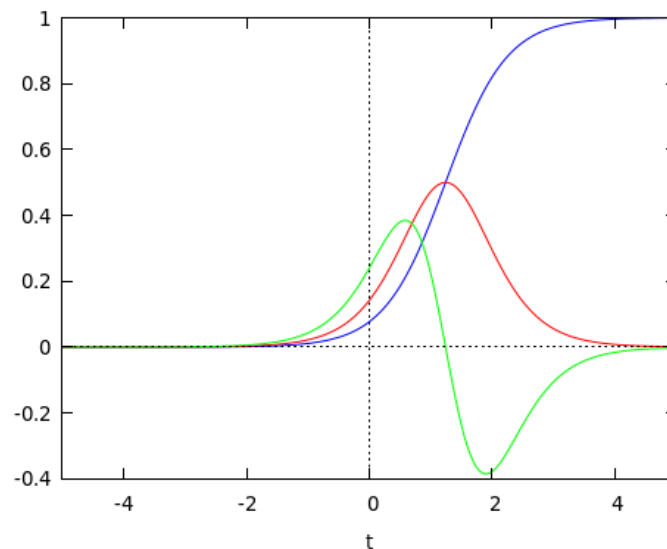
Dieser Text ist eine Fortschreibung des ersten Teils. Die dortigen Beschreibungen der allgemeinen Rahmenbedingungen werden als bekannt vorausgesetzt.

1 Logistische Funktion

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

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

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 (grün) Ableitung

K steht dabei für die Sättigungsgrenze $\lim_{t \rightarrow \infty} u(t)$ 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 mit $u(t_0) = \frac{1}{2} K$.

Derartige Funktionen lassen sich deutlich schlechter schätzen als Glockenkurven, die im Teil 1 dieser Reihe betrachtet wurden und sich durch Logarithmieren auf einfache Weise auf einen polynomialen Zusammenhang reduzieren lassen. 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 dann doch mit einem linearen Fitting bestimmen lassen. Wir transformieren dazu (L.1) in die Formel

$$l(t) = \frac{K}{1 + \exp(-r(t - m))}, \quad (\text{L.2})$$

indem $C = \exp(rm)$ ersetzt wird. Weiter logarithmieren wir (L.2) und erhalten als neuen Schätzzusammenhang

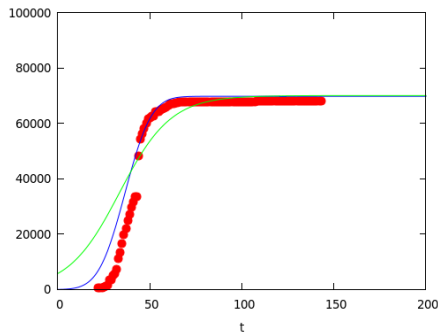
$$\log \left(\frac{K}{l(t)} - 1 \right) = -r(t - m). \quad (\text{L.3})$$

Der Parameter K ist dabei als Sättigungsgrenze vorab manuell zu schätzen, so dass die gefittete Kurve möglichst gut auf die Daten passt. Wir vergleichen unten diese Ergebnisse mit den Schätzungen auf Basis der Fehlerfunktion aus Teil 1.

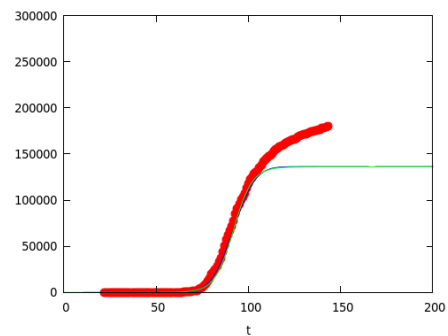
Im Skript ist dazu eine Funktion `lFit(G,K0)` implementiert, der eine Liste G von Datenpunkten und der Schätzer für K übergeben werden, wobei in G zusätzlich vorab alle Datenpunkte mit $y_t \leq 10$ ausgefiltert sind. Weiter wird eine Funktion `selectData(G,von,bis)` definiert, mit der aus einer Datenreihe entsprechenden Daten mit $von < t < bis$ selektiert werden können, um die Modellierungssituation vom 10.04.2020 = Tag 101 nachzustellen. In einer zweiten Rechnung verwenden wir diese Funktion, um zu sehen, welche Datenbereiche einer Kurve für eine Schätzung von $l(t)$ besonders gut geeignet sind.

Wir führen zunächst die Schätzung aus Teil 1 für die Zahl $p(t)$ der positiv getesteten Personen für die Tage 30...100 erneut aus (Werte m_0 , s_0 und $K = \sqrt{\pi} \exp(c) s$) und verwenden diesen Schätzer für die Berechnungen einer Parameterschätzung für $l(t)$. Die Schätzungen liegen nahe beieinander, aber deutlich unter der weiteren Entwicklung der Daten, wie auch schon aus der Schätzung für die Hubei-Provinz im Teil 1 zu erwarten war. Siehe dazu die letzte Spalte sowie die beiden Diagramme.

Land	m_0	s_0	K	m	r	$p(143)$
Deutschland	91.83	13.40	136 280	91.91	0.196	179 710
Italien	86.68	15.38	156 683	85.90	0.196	228 658
Österreich	87.17	10.62	12 860	126.30	0.091	16 436
Spanien	90.57	12.12	171 096	110.14	0.158	234 824
China (Hubei)	36.52	15.95	69 772	33.58	0.007	68 135



Szenario für China (Hubei)

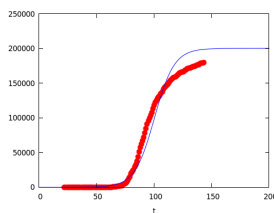


Szenario für Deutschland

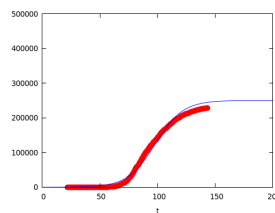
Neben dem schlechten Schätzer für K sehen wir an den Daten von China mit deutlich früherem Maximum der Epidemie, dass auch die Wahl des Schätzintervalls großen Einfluss auf die Güte der Schätzung hat. Beste Ergebnisse werden erzielt, wenn Daten aus dem „linearen Teil“ rund um $t = m$ verwendet werden.

Dies ist in der folgenden Übersicht für die logistische Funktion ausgeführt. Als neuer Schätzer für K wird ein aus $p(143)$ geschätzter Wert verwendet, als Schätzintervall etwa der Bereich $m - 25 < t < m + 25$. In der folgenden Tabelle sind Schätzungen aus entsprechenden Zeitreihen der positiv Getesteten für verschiedene Länder gegenübergestellt, die durch entsprechende Parameteradjustierungen auf Daten bis zum Tag 143 (22. Mai 2020) gewonnen wurden:

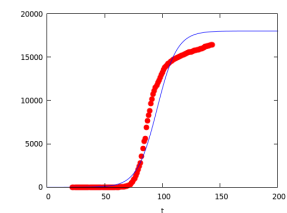
Land	K	von	bis	r	m
Deutschland	200 000	70	120	0.116	100.34
Italien	250 000	70	120	0.056	103.15
Österreich	18 000	70	120	0.116	95.30
Spanien	270 000	70	120	0.118	100.81
China (Hubei)	70 000	22	62	0.213	42.51
Schweden	50 000	100	150	0.048	128.96



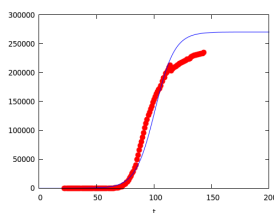
Deutschland



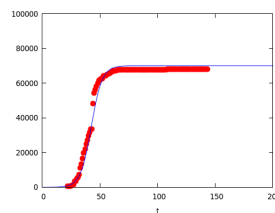
Italien



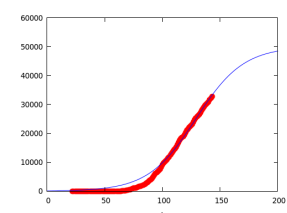
Österreich



Spanien



China (Hubei)



Schweden

In den Diagrammen ist die (etwas zeitversetzte) Abflachung der Kurven nach dem Lock-down in verschiedenen europäischen Ländern deutlich zu sehen, dessen Wirkung in das hier verwendete zeithomogene Modell natürlich nicht eingehen kann. Dies bleibt gesondert zu untersuchen.

2 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.