

UQ – Abgabe1 Gruppe D (Bernhard, Ivanov, Schröppel)

Das SEIR-Modell (susceptible-exposed-infected-removed model) bezeichnet man in der mathematischen Epidemiologie als Ansatz zur Beschreibung der Ausbreitung von ansteckenden Krankheiten. Dabei teilt das Modell die Bevölkerung in verschiedene Gruppen, wobei jeder Mensch nur in einer Gruppierung zu einer bestimmten Zeit sein kann:

Susceptible (S): Anteil der Anfälligen

Exposed(E): Anteil der Exponierten, infiziert aber noch nicht infektiös

Infected(I): Anteil der Infektiösen

Removed(R): Anteil der Erholten

$S + E + I + R = N$ (Bevölkerung/Anzahl an betrachteten Menschen)

Das SEIR-Modell wird durch folgendes System von gewöhnlichen Differentialgleichungen

$$dS/dt = -\beta S(I/N),$$

$$dE/dt = \beta S(I/N) - \alpha E,$$

$$dI/dt = \alpha E - \gamma I,$$

$$dR/dt = \gamma I.$$

mit folgenden Modellparametern:

t: Zeit in Wochen

β : Transmissionrate

γ : Erholungsrate

α : Übergangsrate

I_0 : Anzahl anfänglicher Infizierter

unter folgenden Modelannahmen:

- Es findet eine gute Durchmischung der Bevölkerung statt, so dass jedes mit jedem anderen Individuum in Kontakt kommen kann.
 - Die Werte sind kontinuierlich. Das Modell beschreibt keine diskreten Zahlen von Menschen.
 - Die Krankheit führt entweder zur vollständigen und lebenslangen Immunität oder zum Versterben.
 - Jedes anfällige Individuum ist gleich anfällig für die Krankheit
 - Ein Nachwuchs neuer Anfälliger und natürliches Versterben tritt nicht auf.
- beschrieben.

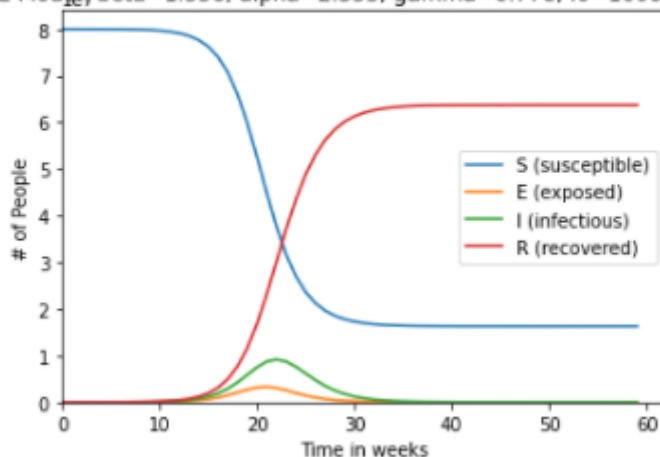
Problem

Das Verhalten des SEIR-Modells hängt vor allem von den Modellparametern ab. Da wir diese Werte jedoch nicht exakt bestimmen können, sind die Ergebnisse/Prognosen des Modells nicht exakt. Das Problem dabei ist, dass wir nicht genau wissen, wie empfindlich das Modell gegenüber Änderungen der Parameter reagiert beziehungsweise bei welchen Parametern das SEIR Modell besonders empfindlich und somit auch fehleranfällig ist. Um eine Aussage darüber zu tätigen wie genau unsere Simulationen sind, müssen wir eine Sensitivitätsanalyse durchführen. Im folgenden soll daher untersucht werden, wie stark das SEIR-Modell mit den durch das RKI bereitgestellten Parameterwerten für die Covid-19 Pandemie auf Änderungen der Parameter reagiert.

Zu Beginn wurden daher zunächst die ODEs welche das Modell beschreiben implementiert. Mithilfe eines Integrators wurden die Ergebnisse des Modells für die gegebenen Parameter Beta, Alpha, Gamma, I₀ für eine Populationsgröße von 80 Millionen für einen Zeitraum von 60 Wochen berechnet. Das Ergebnis kann anschließend geplottet werden um den zeitlichen Verlauf der Anzahl an Menschen innerhalb einzelner Compartments darzustellen.

```
plot_seir(params=(BETA, ALPHA, GAMMA, I_0), n=N, weeks=WEEKS)
```

SEIR ODE Model, beta=1.556, alpha=2.333, gamma=0.778, I₀=1000, N=80000000



Für einen ersten Eindruck über die Bedeutung der einzelnen Parameter, wurden diese nacheinander einzeln abgeändert um anschließend die Ergebnisse (Plots) zu vergleichen.

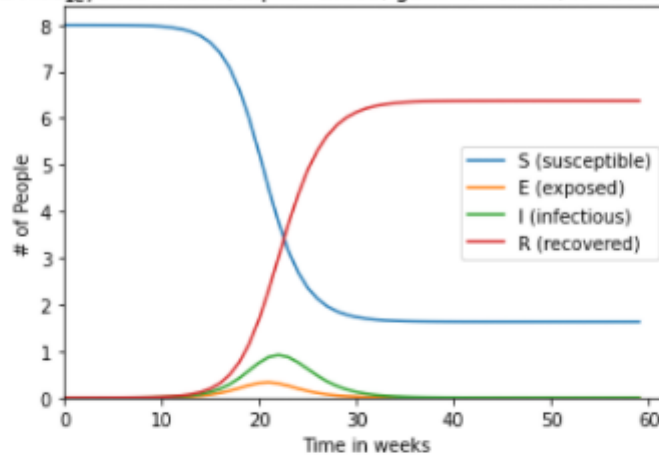
- Bedeutung und Einfluss β (Transmissionrate)
Eine höhere Transmissionsrate führt zu einem früheren und absolut höheren Maximum an Infizierten Menschen. Auch die Zahl der Menschen die sich zum Ende der Zeit innerhalb des Removed Compartments befinden ist höher. Eine Verdopplung des Parameters β hat bereits sehr starke Auswirkungen auf das Ergebnis. Nahezu die gesamte Population ist bereits nach etwa 15 Wochen entweder genesen oder verstorben. Anschließend findet praktisch kein Infektionsgeschehen mehr statt. Bereits eine Erhöhung der Transmissionsrate um 20% zeigt die anfangs beschriebenen Effekte im Plot deutlich sichtbar.

```

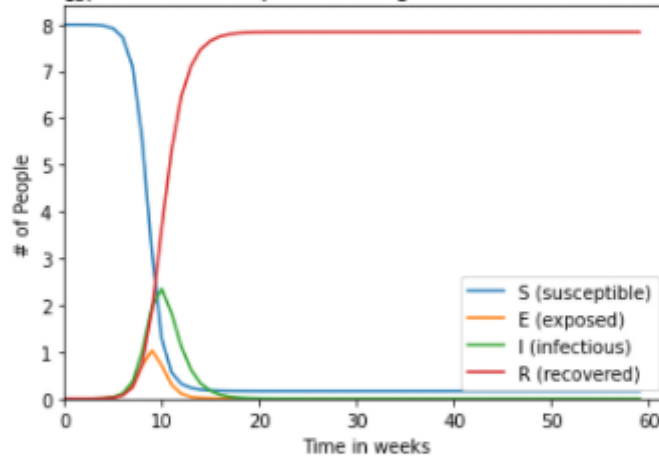
plot_seir(params=(BETA, ALPHA, GAMMA, I_0), n=N, weeks=WEEKS)
plot_seir(params=(2 * BETA, ALPHA, GAMMA, I_0), n=N, weeks=WEEKS)
plot_seir(params=(1.2 * BETA, ALPHA, GAMMA, I_0), n=N, weeks=WEEKS)

```

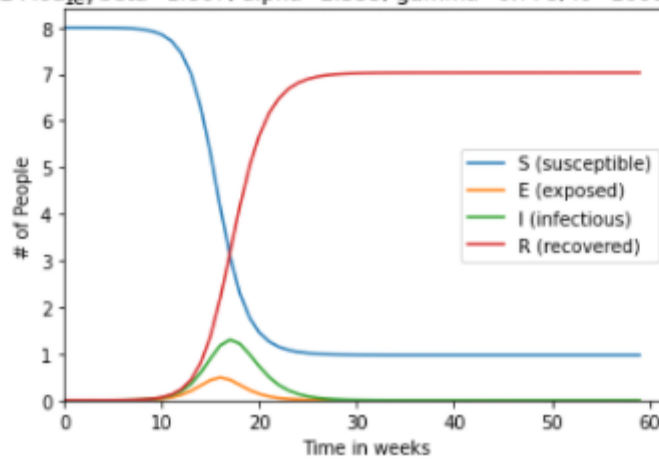
SEIR ODE Model, beta=1.556, alpha=2.333, gamma=0.778, I0=1000, N=80000000



SEIR ODE Model, beta=3.111, alpha=2.333, gamma=0.778, I0=1000, N=80000000



SEIR ODE Model, beta=1.867, alpha=2.333, gamma=0.778, I0=1000, N=80000000

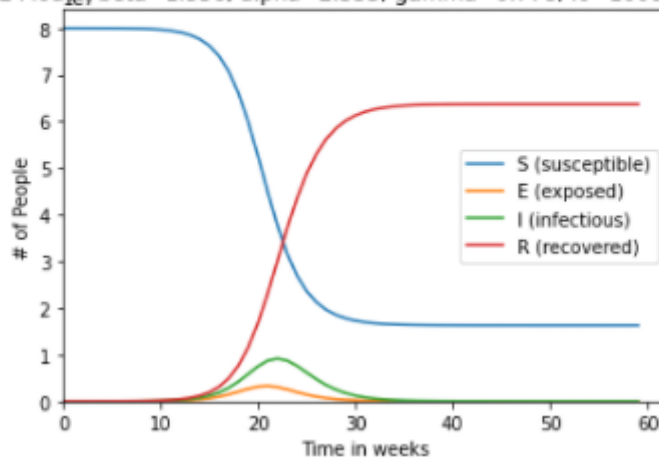


- Bedeutung und Einfluss α (Übergangsrate)

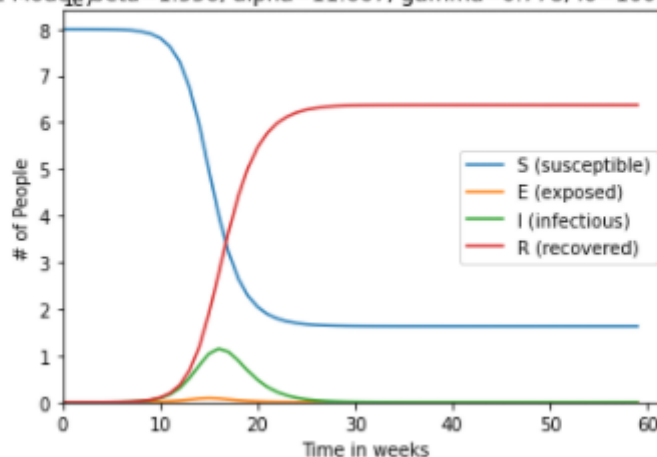
Eine Erhöhung der Übergangsrate (also eine Reduktion des Zeitfensters zwischen Ansteckung und der Fähigkeit andere anzustecken) führt zu ähnlichen Effekten wie eine höhere Transmissionsrate (früherer und höherer Peak an Infizierten). Allerdings bleibt die Anzahl der Menschen die zum Ende der Pandemie im Compartment Removed sind nahezu unverändert. Außerdem fällt sofort auf, dass eine Verhundertfachung der Übergangsrate in etwa den selben Effekt auf das Modell hat, wie die Erhöhung der Transmissionsrate um 20%. Das Modell reagiert also relativ schwach auf Änderungen der Übergangsrate wohingegen Änderungen der Transmissionsrate großen Einfluss haben.

```
plot_seir(params=(BETA, ALPHA, GAMMA, I_0), n=N, weeks=WEEKS)
plot_seir(params=(BETA, 5*ALPHA, GAMMA, I_0), n=N, weeks=WEEKS)
```

SEIR ODE Model: beta=1.556, alpha=2.333, gamma=0.778, I0=1000, N=80000000



SEIR ODE Model: beta=1.556, alpha=11.667, gamma=0.778, I0=1000, N=80000000



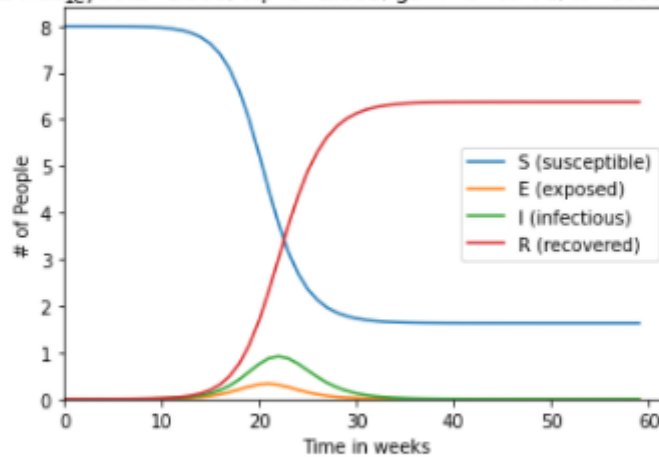
- Bedeutung und Einfluss γ (Erholungsrate)

Wird die Erholungsrate erhöht, führt dies zu gegenteiligen Effekten wie eine Erhöhung der Transmissionsrate. Während sich das Maximum der Menschen die zu einem Zeitpunkt infiziert sind reduziert und nach hinten geschoben wird, sinkt auch die Anzahl der Menschen die gegen Ende der Pandemie zum Removed Compartment gehören. Dies hat zur Folge, dass auch die Anzahl der weiterhin Anfälligen (S-Compartment) relativ hoch bleibt. Auch hier fällt wieder auf, dass ähnlich wie bei der Transmissionsrate bereits eine zwanzigprozentige Erhöhung das Ergebnis des Modells

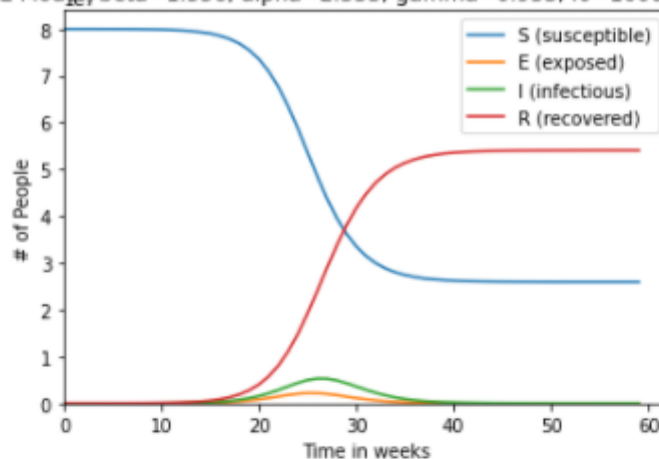
stark verändert. Bei einer Verdopplung des Werts ($\gamma > 1$) kann praktisch kein Krankheitsverlauf mehr erkannt werden.

```
plot_seir(params=(BETA, ALPHA, GAMMA, I_0), n=N, weeks=WEEKS)
plot_seir(params=(BETA, ALPHA, 1.2 * GAMMA, I_0), n=N, weeks=WEEKS)
plot_seir(params=(0.8 * BETA, ALPHA, GAMMA, I_0), n=N, weeks=WEEKS)
plot_seir(params=(BETA, ALPHA, 2 * GAMMA, I_0), n=N, weeks=WEEKS)
```

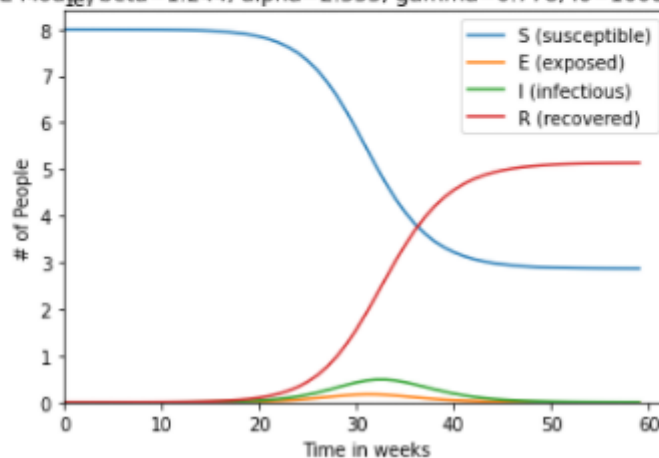
SEIR ODE Model, beta=1.556, alpha=2.333, gamma=0.778, I0=1000, N=80000000



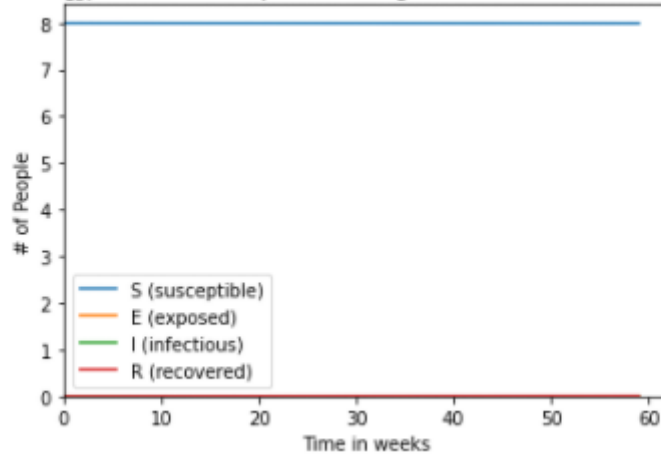
SEIR ODE Model, beta=1.556, alpha=2.333, gamma=0.933, I0=1000, N=80000000



SEIR ODE Model, beta=1.244, alpha=2.333, gamma=0.778, I0=1000, N=80000000



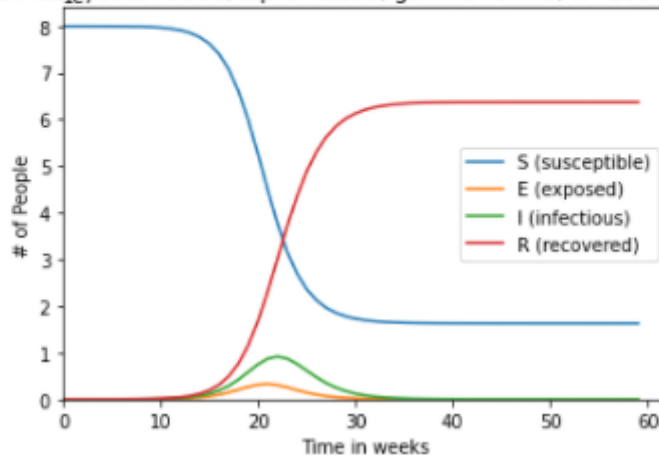
SEIR ODE Model, $\beta=1.556$, $\alpha=2.333$, $\gamma=1.556$, $I_0=1000$, $N=80000000$



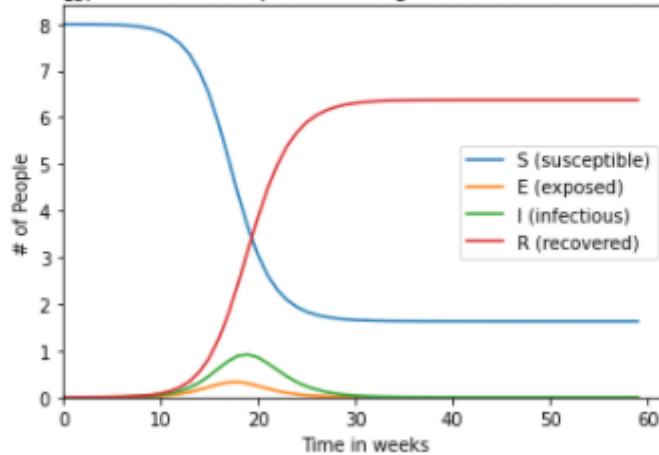
- Bedeutung und Einfluss I_0 (Anzahl anfänglicher Infizierter)
Die Zahl der Anfänglich Infizierten Personen scheint geringen Einfluss auf das Modell zu haben. Selbst eine Verhundertfachung des Werts verändert den Verlauf kaum.

```
plot_seir(params=(BETA, ALPHA, GAMMA, I_0), n=N, weeks=WEEKS)
plot_seir(params=(BETA, ALPHA, GAMMA, 5 * I_0), n=N, weeks=WEEKS)
```

SEIR ODE Model, $\beta=1.556$, $\alpha=2.333$, $\gamma=0.778$, $I_0=1000$, $N=80000000$



SEIR ODE Model, $\beta=1.556$, $\alpha=2.333$, $\gamma=0.778$, $I_0=5000$, $N=80000000$



Insgesamt entscheidend für das Modell scheinen die Transmissionsrate sowie die Erholungsrate zu sein. Während die Transmissionsrate bestimmt wie viele Menschen sich mit der Krankheit anstecken bestimmt die Erholungsrate wie schnell die Menschen wieder genesen und somit keine weiteren Menschen mehr anstecken können. Diese zwei Faktoren scheinen entscheidend für den Pandemieverlauf zu sein.

Da es nicht möglich ist abzuschätzen bei welchem Parameter letztendlich Abweichungen zwischen Annahme und Realität auftreten, wurden im Anschluss Samples für den Parametervektor $x = (\beta, \alpha, \gamma, I_0)^T$ innerhalb des Hyperwürfels

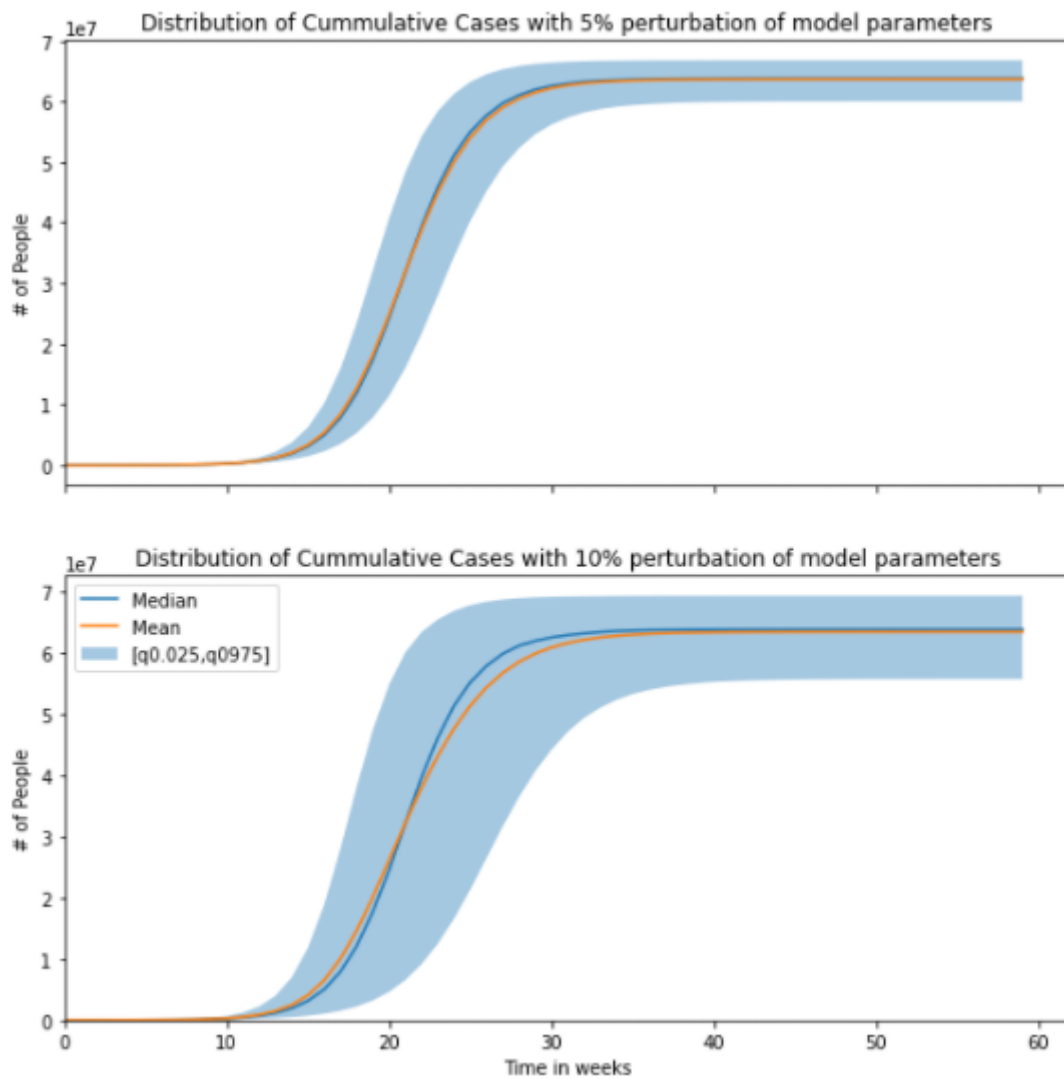
$$H = [1 - p * \beta; 1 + p * \beta] \times [1 - p * \alpha; 1 + p * \alpha] \times [1 - p * \gamma; 1 + p * \gamma] \\ \times [1 - p * I_0; 1 + p * I_0]$$

erstellt. Es wurden dabei jeweils Werte für die einzelnen Parameter gewählt die bis zu p Prozent vom ursprünglich angenommenen Wert abweichen. Für die Werte $p=5\%$ und $p=10\%$ wurden jeweils 1000 Samples erstellt. Die einzelnen Werte liegen dabei gleichverteilt innerhalb der durch den Hyperwürfel definierten Intervallen.

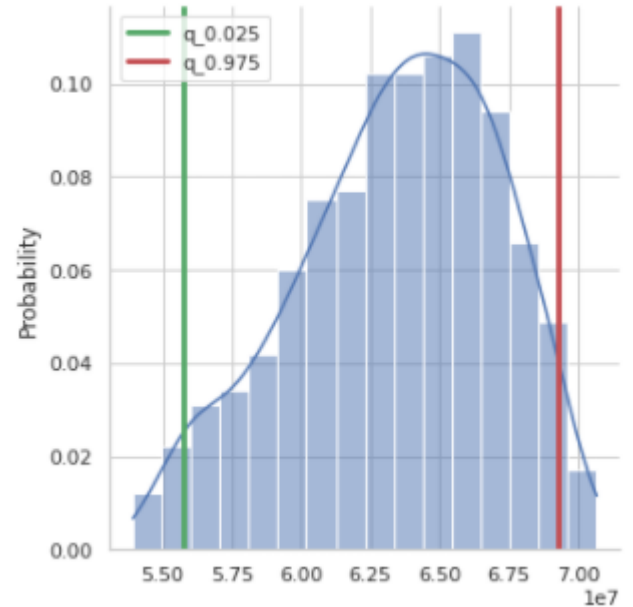
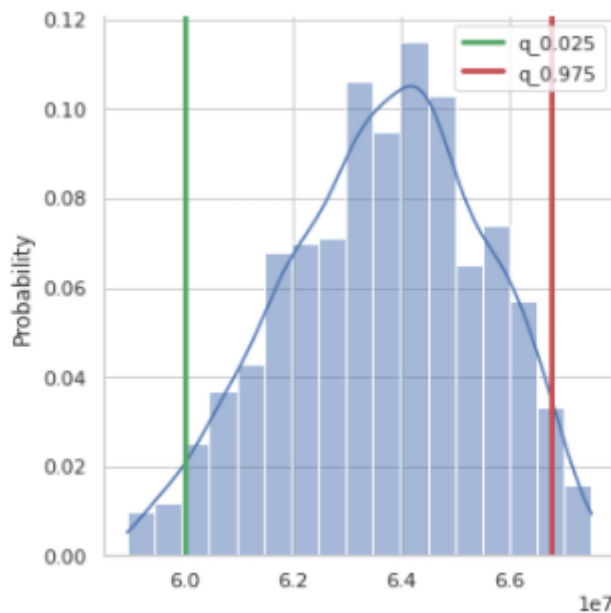
Zu Beginn der nachfolgenden Untersuchungen wurde das SEIR Modell zum SEIRC Modell ergänzt. Der Wert C beschreibt dabei die Anzahl der Kumulativen Fälle d.h. die Anzahl an Personen die aktuell Infiziert sind oder zu einem früheren Zeitpunkt infiziert waren und bereits dem Removed Compartment angehören. Die Anzahl der Personen die pro Zeiteinheit zu den Kumulativen Fällen addiert werden entspricht der Anzahl an Personen die sich zu diesem Zeitpunkt infizieren und in das Infected Compartment rutschen.

$$dC/dt = \alpha E.$$

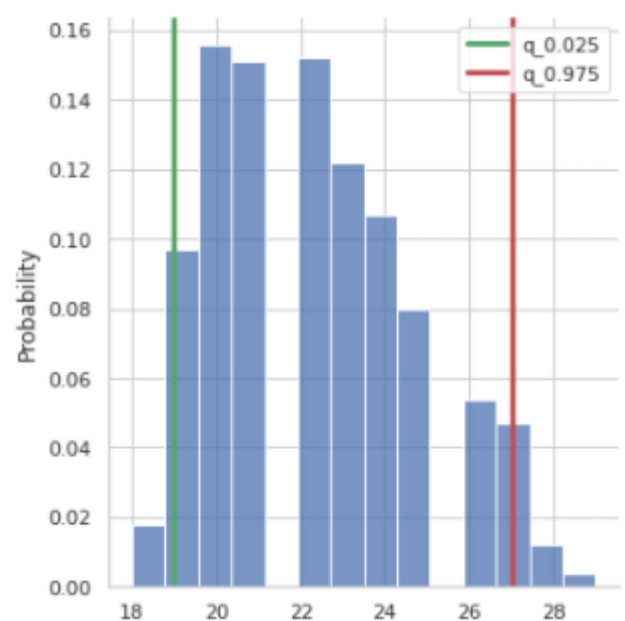
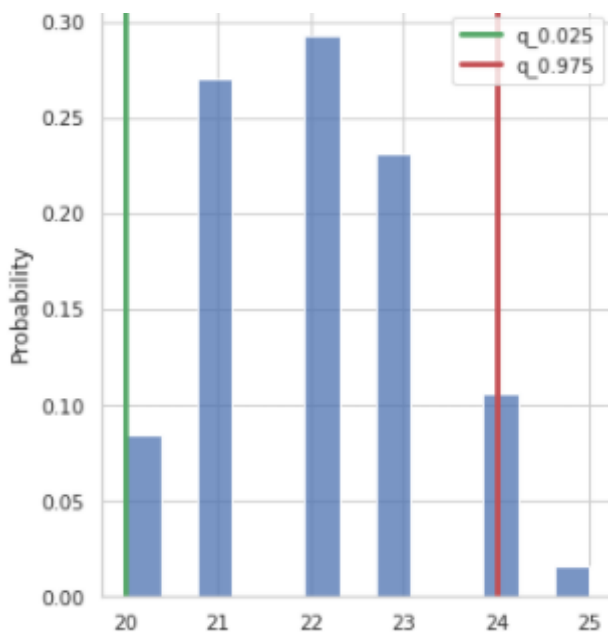
Für jeden der gesampelten Parametervektoren, wird nun das SEIRC-Modell berechnet und das Ergebnis gespeichert. Anschließend wurden für die jeweils 1000 Samples jeweils die Werte der Kumulativen Fälle zu jedem Zeitpunkt betrachtet. Für jede Woche wurden dann die statistischen Maßzahlen Mittelwert, Median, 2,5% Quantil sowie das 97,5% Quantil berechnet und der Zeitliche Verlauf in einem Plot dargestellt.



Im nächsten Schritt wurde die Verteilung der Anzahl kumulativer Fälle zum Ende des betrachteten Zeitraums untersucht. Dafür wurden aus den jeweils 1000 Samples der finale (letzte) Wert der für die Anzahl kumulativer Fälle berechnet wurde, ausgewählt und die empirische Verteilung sowie das 2,5%quantil und das 97,5% Quantil dargestellt.



Abschließend wurde untersucht, wie sich die Wochen in denen die Anzahl der infizierten maximal ist, verteilen. Dafür wurde für jedes der berechneten Ergebnisse diese Woche bestimmt und anschließend ebenfalls die empirische Verteilung sowie das 2,5% Quantil und das 97,5% Quantil geplottet.



Umgesetzt wurde diese formale Lösungsskizze mit der Programmiersprache Python.

Zum Integrieren der Differentialgleichungen des SEIR bzw. SEIRC Modells wurde die Funktion `solve_ivp()` aus der Bibliothek Scipy verwendet. Mit dieser Funktion ist es möglich gewöhnliche Differentialgleichungen mit gegebenen Startwerten (initial values) zu lösen. Es wurde dazu eine Hilfsfunktion `seir_ode(t,y)`, in der die Differentialgleichungen aufgestellt wurden, definiert. Diese Hilfsfunktion kann der Funktion `solve_ivp()` als Argument

übergeben werden um diese Differentialgleichungen zu lösen. Durch die Globale Variable `weeks` wird festgelegt für wie viele Wochen Lösungen für das Modell gefunden werden sollen. Es wird für jede Woche eine Lösung bestimmt.

Für das Sampling der Parametervektoren aus dem Hyperwürfel wurden zunächst für jeden einzelnen Parameter 1000 Samples generiert die innerhalb der p-Prozentigen Abweichung um den Standardwert liegen. Die gesampelten Werte sollen gleichverteilt in diesem Intervall liegen. Um dies zu erreichen wurden die Samples mithilfe der `Numpy.random.default_rng` Methode `uniform()` erstellt. Die somit erstellten 1000 Samples für jeden Parameter wurden nun zu 1000 Parametervektoren kombiniert.

Auch für die Berechnung der statistischen Maßzahlen wurde die Bibliothek `Numpy` mit den Funktionen `mean()` und `quantile()` verwendet. Der Median als 50% Quantil wurde dabei ebenfalls über die Funktion `quantile()` berechnet.

Für die grafische Darstellung der Pandemieverläufe über die Zeit wurde die Bibliothek `Matplotlib` verwendet. Zur Darstellung der empirischen Verteilung der finalen Anzahl kumulativer Fälle sowie der Woche in der die Anzahl infizierter Personen maximal ist, wurde die Bibliothek `Seaborn` eingesetzt.

Interpretation/Ergebnisse:

Durch Ausprobieren wie das Modell auf Veränderungen einzelner Parameter reagiert, konnte schnell festgestellt werden, dass vor allem die Transmissionsrate `Beta` sowie die Erholungsrate `Gamma` großen Einfluss auf das Modell haben. So hat eine zwanzigprozentige Erhöhung der Transmissionsrate `Beta` einen sehr ähnlichen Effekt auf das Modell wie eine vierhundertprozentige Erhöhung der Parameter `Alpha` oder `I0`. Damit die Vorhersagen des Modells eine hohe Aussagekraft besitzen, ist es daher essentiell die Parameter `Beta` sowie `Gamma` möglichst exakt abzuschätzen.

Wie in der Verteilung der kumulativen Fälle über die Zeit zu erkennen ist, ist das Modell in der Phase, in der die meisten Infektionen auftreten sehr ungenau bzw. es reagiert sehr sensibel auf Änderungen der Parameter. Bei einer Änderung der Parameter um jeweils bis zu 10% liegt die Zahl der kumulativen Fälle bei Woche 20 mit 95%iger Wahrscheinlichkeit zwischen 5 Millionen und 60 Millionen. (Nur ungefähre Werte aus der Grafik abgelesen nicht berechnet). In der Zeit zwischen Woche 15 und 30 ist eine verlässliche Aussage somit kaum mehr möglich. Hier könnte man beispielsweise kleinere Quantils Bänder bestimmen. Damit wäre eine präzisere Vorhersage möglich welche jedoch zu Lasten der Wahrscheinlichkeit geht, dass Sie auch tatsächlich eintritt. Eine weitere Beobachtung bei Betrachtung der Abb. XXX ist die relativ gute Übereinstimmung von Median und Mittelwert. Dies bedeutet, dass die Anzahl der kumulativen Fälle nicht schief verteilt ist (kein Skew). Es ist daher in etwa gleich Wahrscheinlich, dass der tatsächliche Wert über beziehungsweise unterhalb des Durchschnittswerts liegt. In den Wochen 20-30 liegt bei einer Abweichung der Parameter um bis zu 10% liegt der Median leicht über dem Durchschnittswert. Es ist daher wahrscheinlicher, dass der tatsächliche Wert über dem Durchschnittswert liegt.

Relativ gut vorherzusagen ist dagegen die Anzahl der kumulativen Fälle zum Ende der Pandemie. Diese schwankt selbst bei einer Abweichung um bis zu 10% in den Parametern mit 95%iger Wahrscheinlichkeit nur zwischen 55,7mio und 69,3 Mio. Menschen. Bezogen auf die Gesamtpopulation von 80 Mio. ist diese Schwankung relativ gering.

Die Woche in der die Zahl infizierter Personen maximal ist, lässt sich relativ gut eingrenzen. Auch wenn bei einer Abweichung der Parameter um bis zu 10% die Infektionen zwischen Woche 18 und 28 maximal sind, lässt sich aus der Verteilung ableiten, dass diese Woche am wahrscheinlichsten innerhalb des 6. Monats (Woche 20-24) liegt. Variieren die Parameter nur um bis zu 5% kann sogar mit mehr als 95% Sicherheit davon ausgegangen werden, dass dieses Maximum an infizierten zwischen Woche 20 und 24 liegt.

