Angewandte Stochastik

Dr. C.J. Luchsinger

10 Epidemienmodelle - Outbreak

Wir werden in diesem Kapitel auch ein paar philosophische Fragen, Einwände und dazugehörige Gegenargumente zur Modellierung behandeln.

10.1 SIR - determinisch und nichtlinear

Bekanntlich besucht uns im Winter regelmässig Influenza A. Die (potentielle) Epidemie beginnt mit ein paar wenigen Fällen. Dann kann eine regelrechte Explosion der Krankheitsfälle stattfinden. Nach ein paar Wochen hat es dann wieder nur ein paar wenige Fälle; die Epidemie klingt aus. Wir wissen zudem, dass ein paar Menschen gar nicht infiziert werden. Wir sind am Anfang einer solchen Epidemie daran interessiert zu wissen, wie viele Personen infiziert werden, wie viele Menschen maximal gleichzeitig infiziert sein werden und wann die Epidemie etwa vorüber sein wird. MedizinerInnen der Gesundheitsbehörde werden erste Vorhersagen machen, bei welchen Sie sich auf die Erfahrungen von anderen Ländern stützen oder von vergangenen Jahren. Aber solche Vorhersagen können sehr unpräzise sein, wenn man nicht berücksichtigt, was genau vor sich geht. Hier können mathematische Modelle Abhilfe schaffen: mathematische Modelle erlauben die Beschreibung des Verhaltens von Systemen, welche zu kompliziert sind, um sie allein in natürlicher Sprache zu beschreiben. Weiter kann man versuchen, damit das Verhalten in der Zukunft vorher zu sagen. Die Sprache der Mathematik (Differentialgleichungen und Stochastische Prozesse) ist universell in dem Sinne, dass wir damit Prozesse der Physik, Chemie, Biologie, Medizin, Ingenieur-Wissenschaften und Wirtschafts-Wissenschaften beschreiben können. Diese Modelle müssen jedoch zuerst aufgestellt werden:

* nachdem ein Phänomen in *natürlicher Sprache* beschrieben wurde,
* muss die *relevante Information* in die *Sprache der Mathematik* übersetzt werden.

Dies nennt man Modellieren. Sobald das Modell steht, kann der ganze Apparat der (reinen und angewandten) Mathematik eingesetzt werden, um das Modell zu analysieren. Die Folgerungen, welche man dann zieht, müssen danach wieder in die natürliche Sprache retourübersetzt werden. In obigem Beispiel wird also zuerst einE MedizinerIn einem/einer MathematikerIn ganz genau erklären müssen, wie der heutige Wissensstand über die Ausbreitung von Influenza A ist.

Erfahrene BeraterInnen aus allen Gebieten wissen es: Dies ist ein ganz wichtiger Prozess. Warum? MathematikerInnen werden nicht angestellt um zu wissen, sondern um zu denken und hinterfragen (kleine Nebenbemerkung: braucht Nerven und Grösse bei der Fachperson und es muss schon einE guteR MathematikerIn sein). Wenn eine Fachperson einer quantitativ geschulten Person das Problem erklären muss (mit allen möglichen Gegenfragen), wird die Fachperson selber viel dazulernen. Zudem muss man sich bewusst sein, dass das "Wissen" ausserhalb der Mathematik unterschiedlich gesichert ist. Es geistern viele Halbwahrheiten, Vermutungen und richtige Theorien herum. MathematikerInnen, welche die kindliche Frage "Wieso isch das so?" gerne stellen, können hier klärende Fragen stellen.

In unserem Beispiel kann das Problem folgendermassen beschrieben werden: ein ganz kleiner Anteil der Bevölkerung des uns interessierenden Landes ist mit einem Virus infiziert und kann andere Personen infizieren. Die bereits betroffene Bevölkerung hat die Infektion aufgelesen, indem sie in ferne Länder gereist ist oder Kontakt mit Personen aus diesen Ländern gepflegt hat (Touristen, Geschäftsleute, Flüchtlinge). Die Infektion wird eventuell weitergegeben, falls der Kontakt eng genug ist und die potentiell neu infizierte Person empfänglich für die Infektion ist. Danach kann die neu infizierte Person selber weitere Personen infizieren. Nach einer gewissen Zeit sind die Personen wieder gesund; sie können a) keine weiteren Personen mehr infizieren und sind b) selber nicht mehr empfänglich. Dies mag die Beschreibung der MedizinerIn sein. Obige Fakten sind in der Schulmedizin akzeptiert.

Nachdem wir also eine Beschreibung des Ausbreitungsmechanismus' in natürlicher Sprache erhalten haben, wollen wir diesen Mechanismus modellieren (mit Differential-Gleichungen).

Wir machen hierbei folgende Annahmen; dabei übersetzen wir obige Erklärungen sukzessive von der natürlicher Sprache in die Sprache der Mathematik:

Wir führen drei Variablen $S(t), I(t)$ und $R(t)$ ein, wobei diese Variablen die Anteile der Gesamtbevölkerung von Empfänglichen (Susceptibles, $S(t)$), Infektiösen (Infectives, $I(t)$) und Entfernten (Removed, $R(t)$) zur Zeit t angeben. Wir wollen dies präzisieren: Mit "susceptibles" meinen wir Menschen, welche mit dem Virus infiziert werden können, wenn sie Pech haben; mit "infectives" meinen wir kranke Menschen, welche das Virus noch weitergeben können; mit "removed" meinen wir sowohl Personen, welche von Anfang an immun waren, oder isoliert wurden, oder gestorben sind oder nach Krankheit wieder genesen und jetzt erst immun sind (von der mathematischen Modellierung her ist das alles gleich - vom realen Erleben her natürlich nicht!). Damit haben wir ganz sicher $\forall t \geq 0$, dass $S(t) + I(t) + R(t) = 1$ und $0 \leq S(t), I(t), R(t) \leq 1$. Wir führen jetzt zwei Parameter ein: λ und μ; $\lambda, \mu > 0$. Wir geben damit an, mit welcher *Rate* (im stochastischen Analogon wird das die Rate des Poisson-Prozesses sein) die Menschen potentiell infektiöse Kontakte machen (λ) und mit welcher Rate (stochastisch: exponential-verteilte Zeit bis gesund) die Menschen genesen (μ). Eine Infektion kann nur stattfinden, wenn ein Empfänglicher einen genug engen Kontakt mit einem Infektiösen hat. Die Rate, mit der der Anteil der Emfänglichen schrumpft, kann deshalb mit $\lambda S(t)I(t)$ modelliert werden. Des weiteren können nur Infektiöse genesen. Damit steigt der Anteil der Entfernten an der Gesamtbevölkerung mit Rate $\mu I(t)$. Diese Modellierungsannahmen fassen wir in folgendem System von Differential-Gleichungen zusammen:

$$\begin{aligned} \frac{dS(t)}{dt} &= -\lambda S(t)I(t) \\ \frac{dI(t)}{dt} &= \lambda S(t)I(t) - \mu I(t) = I(t)[\lambda S(t) - \mu] \\ \frac{dR(t)}{dt} &= \mu I(t). \end{aligned} \tag{SIR}$$

Die Ableitungen auf der rechten Seite summieren auf 0; dies sollte auch sein, da ja $S(t) + I(t) + R(t) = 1 \ \forall t \geq 0$.

Eine Warnung: Die Gleichungen (SIR) mit *diesen* Raten sind in obigem Text kurz begründet worden. Wir haben auch den Poisson-Prozess und die exponential-verteilte

Genesungszeit erwähnt. Dass das Modell dann zwingend wie in (SIR) angegeben aussehen *muss*, ist nicht richtig. Wohl liegt (SIR) auf der Hand und leuchtet ein. Aber im Gegensatz zu stochastischen Analoga (siehe später), ist es im deterministischen Umfeld nicht derart zwingend, wie die Differential-Gleichungen aussehen müssen. Bei der stochastischen Modellierung ist dies anders: sobald wir dort (präzise) von Poisson-Prozessen und exponential-verteilten Zufallsgrössen sprechen, sind die Modelle festgelegt.

Bemerkung: In Modell (SIR) ist implizit die Annahme eingebaut, dass die Menschen unendlich schnell sich wieder *mischen*. Sobald ein kleiner Anteil von S zu I gewechselt hat, kommt er voll in der ersten Übergangsrate $\lambda S(t)I(t)$ zum Einsatz!

Nachdem wir das Modell aufgestellt haben, wollen wir es analysieren. Wir setzen voraus, dass $I(t) > 0$ für alle $t \geq 0$. Von der zweiten Gleichung sehen wir, dass der Anteil der Infektiösen zunimmt solange

$$\lambda S(t) - \mu > 0;$$

andernfalls nimmt dieser Anteil ab. Falls nun $S(t)$ fast 1 ist (z.B. am Anfang einer Epidemie) dann sind $I(t)$ und $R(t)$ fast 0, also gibt es kaum Personen, welche die Infektion einschleppen und kaum jemand ist von Anfang an immun. Dann ist es sogar so, dass der Anteil der Infektiösen zunimmt, falls

$$\lambda - \mu > 0$$

oder $\lambda/\mu > 1$ (das Gegenteil geschieht falls $\lambda/\mu < 1$; was wenn $\lambda/\mu = 1$?). Offenbar ist λ/μ eine kritische Grösse, um zu Beurteilen, wie sich diese Epidemie entwickeln wird. Wir nennen sie (nicht ganz zufällig) R_0; $R_0 = \lambda/\mu$.

Definition 10.1 [Basic Reproduction Ratio (R_0), allgemein] R_0 ist allgemein in einem Modell die durchschnittliche Anzahl Personen, welche eine infizierte Person selber infiziert, solange sie infektiös ist, unter der Annahme, dass jeder Kontakt mit einer empfänglichen Person ist (optimales Umfeld; hier $S(t) \doteq 1$).

Frage: Weshalb gilt (vom stochastischen her argumentierend) in (SIR)

$$R_0 = \frac{\lambda}{\mu} \; ?$$

Eine Frage der Mediziner haben wir hiermit bereits beantwortet: Falls $R_0 < 1$ wird es keinen "Outbreak" geben. Ein paar Menschen werden allenfalls noch infiziert werden, dann ist die "Epidemie" vorbei. Falls $R_0 > 1$, so gibt es ein starkes Ansteigen der Anzahl Infizierten. Der "Peak" wird erreicht sein, sobald $\lambda S(t) - \mu = 0$, warum? Die einzigen stationären Lösungen sind $(S, I, R) = (c, 0, 1 - c)$ mit $c \in (0, 1)$. Die Entwicklung von (SIR) kann in einem Computer gut simuliert werden. Wir haben mit "Mathematica" zwei Simulationen mit unterschiedlichen Startwerten $(S(0), I(0), R(0))$ durchgeführt. Man findet Sie in den Diagrammen online: Diagramm 10.2 und 10.3.

Diagramm 10.2 zeigt eine typische Situation, in der $S(0) = 0.9, I(0) = 0.01$ und $R(0) = 0.09$. Die Parameter haben Werte $\lambda = 5, \mu = 3$. Wir beobachten einen vollen Ausbruch und weniger als 40% ($S(\infty)$) der Empfänglichen werden *nie* Influenza A bekommen (in unserem Modell). Diagramm 10.3 beschreibt die Situation, wo rechtzeitig ein Impfprogramm lanciert wurde. Mit gleichem Anteil Infizierten am Anfang wie bei Diagramm 10.2 haben wir $S(0) = 0.66, I(0) = 0.01$ und $R(0) = 0.33$, mit λ und μ unverändert. Also ist $1/3$ der Bevölkerung geimpft worden (oder war schon immun/isoliert). Es gibt immer noch einen (kleinen) Ausbruch. Aber wenn man die beiden Diagramme vergleicht, sieht man, dass das Impfprogramm viel bewirkt hat.

Wie viele Menschen wird man *sinnvollerweise (!)* impfen, damit ein Ausbruch nicht stattfindet (Annahme: $R(0) = 0$ ohne Impfprogramm)? Kleiner Tip: das ist *nicht* 100%! Warum? Worauf ist beim Impfprogramm zu achten?

Diese Simulationen geben uns also eine Vorstellung, wie solche Epidemien ablaufen. Überraschend ist m.E., dass das Maximum (über die Zeit) an Infizierten gar nicht so hoch ist (das liegt am relativ grossen μ).

Dies war nur ein einführendes Beispiel - wir wollen nicht allzu lange dabei verweilen. Trotzdem sei noch angefügt, wie man in der Praxis fortfahren würde. Man will überprüfen, ob dieses Modell nahe genug an der Realiät ist. Was "Nahe sein" konkret heisst, ist von der Fragestellung abhängig! Auf jeden Fall wird man dazu Daten brauchen.

* Wir müssen wissen, wieviele Menschen am Anfang bereits immun sind ($R(0)$). Dies ist aber anders von Fall zu Fall, je nach dem welches Virus untersucht wird. Weiter ist dies abhängig von der genetischen und immunologischen Situation der uns interessierenden Bevölkerung.

* Die Genesungsrate (μ) ist ebenfalls abhängig von der genetischen und immunologischen Situation der uns interessierenden Bevölkerung und der Behandlung, welche kranken Leuten zu Teil wird.

* Die Kontaktrate (λ) hängt davon ab, wie einfach die Infektion von Person zu Person übertragen werden kann und wie mobil die Menschen sind. Erinnern Sie sich an's Mittelalter: Es dauerte Jahre, bis eine Epidemie von einem Ende in Europa zum anderen kam.

Man kann versuchen, solche Grössen zu schätzen oder von früheren Epidemien oder vom Ausland her Parameter übernehmen. Falls erwartet werden kann, dass die vorhergesagte Entwicklung der Epidemie und der tatsächliche Verlauf einigermassen übereinstimmen, können Gegenmassnahmen mit Model (SIR) diskutiert werden: μ kann kaum beeinflusst werden, ausser dass man infizierten Personen Quarantäne verordnen kann. Man kann wie bereits besprochen ein Impfprogramm *vor* Ausbruch der Epidemie starten. Damit steigern wir $R(0)$ und senken $S(0)$. Da die Entwicklung von $I(t)$ von $\lambda S(t) - \mu$ abhängt, ist eine Milderung zu erwarten. Individuell können Personen möglichen Infektionen vorbeugen, indem Sie eher zu Hause bleiben, öffentliche Verkehrsmittel meiden und diverse

andere Massnahmen ergreifen. Dies senkt λ. Falls die vorhergesagte Entwicklung und die tatsächliche Entwicklung vollkommen auseinanderklaffen, müssen die Annahmen überprüft werden. Aber mathematische Modelle sollten nicht zu kompliziert gemacht werden, weil sonst die Analyse zu schwierig wird (denken Sie an PDE's; jedeR StudentIn des ersten Semesters kann problemlos partielle DGL aufstellen, welche die besten MathematikerInnen kaum lösen können!). Neben der Schwierigkeit der Analyse darf weiter nicht unterschätzt werden, dass man die Resultate auch interpretieren können muss. Bei (SIR) war dies sehr einfach (und sehr schön, oder?). Und die Moral von dä Gschicht:

MAKE MATHEMATICAL MODELS AS SIMPLE AS POSSIBLE

(BUT *NEVER* SIMPLER)!

EinE MedizinerIn könnte fragen, wie gross die Wahrscheinlichkeit ist, dass eine Epidemie ausbricht, falls *eine einzige Person* das Virus in die Population einführt. In der realen Welt ist es ja sehr gut möglich, dass das Environment zwar sehr gut ist (oben: R_0 sehr gross), aber per Zufall trifft die infizierte Person einfach gerade niemanden, bevor sie wieder genesen ist (vgl. Lemma 8.4 b)). Solche Fragen können nicht mit Modell (SIR) beantwortet werden. Hierfür brauchen wir einen vollständig neuen Ansatz:

10.2 Ein stochastisches Analogon zu SIR für die Initialphase

Zu vergleichen mit dem Geburts- und Todesprozess (pure Birth and Death-Process) in 8.4 - es ist das selbe.

Wir werden jetzt also fortfahren, indem wir ein Modell entwickeln, welches den Zufall erlaubt: ein sogenanntes stochastisches Modell. Der Nachteil dieses Ansatzes ist, dass wir nicht mehr wie im deterministischen Modell (SIR) in Diagrammen 10.2 und 10.3 die Entwicklung in der Zeit derart kompakt präsentieren können.

Sei $i(t)$ die *Anzahl* der Infizierten zur Zeit t. Die Anzahl der Empfänglichen (Susceptibles) sei so gross, dass wir sie gleich ∞ setzen. Wir entwickeln dieses Modell *ausschliesslich*, um zu klären, ob die Epidemie sich am Anfang entwickelt oder sofort ausstirbt. Fragen der Art: "Was wenn 50 % der Bevölkerung infiziert sind?" sind in diesem Modell nicht sinnvoll - die Anzahl Susceptibles ist ja unendlich gross und kommt als solches gar nicht vor. Wir setzen voraus, dass die Individuen sich unendlich schnell wieder vermischen, wenn eine Infektion stattgefunden hat. Damit ist es unmöglich, dass zwei Personen aufeinandertreffen, welche beide schon infiziert sind. Alle infizierten Individuen haben infektiöse Kontakte entsprechend einem Poisson-Prozess der Rate λ und zwar unabhängig voneinander. Falls wir einbauen wollen, dass ein Anteil $c, c \in (0,1)$, von Anfang an immun ist, können wir wegen Lemma 8.5 b) die Rate des Poisson-Prozesses einfach auf $\lambda(1-c)$ verkleinern. Falls eine Person infiziert ist, bleibt sie infektiös für eine Zeitspanne, welche exponentialverteilt ist mit Parameter μ. Wir haben also folgende sogenannte *Übergangsraten* in diesem Modell:

$$\begin{aligned} i(t) &\to i(t)+1 \text{ mit Rate } \lambda i(t) \\ i(t) &\to i(t)-1 \text{ mit Rate } \mu i(t). \end{aligned} \tag{10.1}$$

Zur Zeit t finden Ereignisse statt mit einer totalen Rate von

$$i(t)(\lambda+\mu)$$

(Lemma 8.4 a) bzw. Lemma 8.5 a); es ist dies die Rate, bis das *nächste* Ereignis stattfindet!). Die Wahrscheinlichkeit, dass solch ein Ereignis eine neue Infektion ist, ist

$$\frac{\lambda}{\lambda+\mu}$$

(Lemma 8.4 b)); die Wahrscheinlichkeit, dass es eine Genesung ist, ist

$$\frac{\mu}{\lambda + \mu}.$$

Intuitiv wird man wegen obiger beider Raten annehmen, dass hier

$$R_0 = \frac{\lambda}{\mu}$$

(vgl. Definition 10.1).

$$R_0 \lessgtr 1$$

trennt in der Tat Entwicklung der Epidemie vom Aussterben der Epidemie. Wieder ist λ/μ die durchschnittliche Anzahl Menschen, welche ein Infizierter infiziert, bis er/sie selber wieder gesund ist.

Preisfrage: Was passiert, wenn $R_0 = 1$?

Bezüglich der Bedeutung von R_0 gibt es jedoch einen fundamentalen Unterschied zwischen deterministischen (SIR) und stochastischen (10.1) Modellen; welchen?

Wenn $R_0 > 1$ ist, gibt es eine positive Wahrscheinlichkeit, dass die Epidemie sich ausbreitet. Doch wie gross ist diese Wahrscheinlichkeit? Wir definieren dazu

$$q := P[\lim_{t\to\infty} i(t) = 0 | i(0) = 1] \leq 1.$$

Wegen der Unabhängigkeit der Kontakte der Infizierten gilt dann

$$P[\lim_{t\to\infty} i(t) = 0 | i(0) = n] = q^n.$$

Damit können wir aber q berechnen. Es gilt nämlich wegen FTW:

$$q = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} q^2.$$

Die beiden Lösungen sind $q = 1$ und $q = \mu/\lambda$. Wie in Kapitel 5 (Treffwahrscheinlichkeiten, Satz 5.3) ist die minimale Lösung die richtige Lösung (diese Aussage ohne Beweis). Man sieht sofort, dass q kleiner wird, wenn R_0 grösser wird. Damit haben wir also die Frage der MedizinerInnen beantwortet.

10.3 Die Suche nach dem heiligen Gral - sorry, nach *dem richtigen* Modell

Der deterministische Ansatz mit DGL und der stochastische Ansatz mit Markov-Ketten bzw. -Prozessen sind die gängigsten Methoden, mit denen man Phänomene der realen Welt modelliert. Um das Phänomen umfassend zu modellieren, kann man auch mehrere Modelle gleichzeitig untersuchen. Im Normalfall existiert zu jedem stochastischen Modell ein "entsprechendes" deterministisches Modell et vice versa. Man kann versuchen, die beiden Ansätze zu "verlinken": das deterministische Modell kann z.B.

a) der Erwartungswert des entsprechenden stochastischen Prozesses sein, oder/und

b) der Limes im Sinne eines Gesetzes der grossen Zahlen sein.

Es ist *nicht* so, dass (SIR) und Modell (10.1) *direkt* zusammengehören. Indirekt hängen sie zusammen: Modell (10.1) ist das (lineare) stochastische Modell für die *Anfangsphase* von (SIR). Das deterministische Analogon im Sinne von **a)** von (10.1) ist die DGL

$$\frac{dI}{dt} = \lambda I - \mu I. \tag{10.2}$$

Die Lösung hiervon ist $I(0)e^{(\lambda-\mu)t}$. Weil für $i(t)$ von (10.1) gilt $\mathbb{E}[i(t)] = i(0)e^{(\lambda-\mu)t}$ (ohne Beweis), sind Modell (10.1) und (10.2) analog im Sinne von **a)**.

Wie wird man Modell (10.1) und (10.2) im Sinne von **b)** *sinnvoll* verknüpfen? Tip: Skalierung!

Wir haben generell 4 Möglichkeiten, Modelle aufzustellen: linear und nichtlinear, stochastisch und deterministisch. Wann wird man welches wählen?

* stochastisch, linear: Initialphase, solange Möglichkeit des zufälligen Aussterbens real
* stochastisch, nichtlinear: Kleine Gesamtanzahl und Sättigungstendenzen relevant
* deterministisch, linear: Initialphase, sobald zufälliges Aussterben keine Gefahr und Sättigungstendenzen noch nicht relevant
* deterministisch, nichtlinear: sobald zufälliges Aussterben keine Gefahr und Sättigungstendenzen relevant

Übung: bei den bisherigen Modellen waren:

* stochastisch linear:
* stochastisch nichtlinear:
* deterministisch linear:
* deterministisch nichtlinear:

Wie sind die Übergangsraten des stochastischen nichtlinearen Modells von SIR?

Abschliessende Bemerkung über R_0: In allen vier Modellen ist R_0 gleich gewesen und relativ einfach. Dies muss nicht so sein. Im Gegenteil: in der Dissertation und Papers des Autors sind Modelle untersucht, bei denen (im gleichen Modell!) je nach Ausgangslage drei verschiedene "R_0's" relevant sind, um Ausbruch oder Eindämmung der Epidemie zu separieren. Es sind dies, ohne Erklärung zu den einzelnen Parametern (mehr dazu via www.luchsinger-mathematics.ch/cvluchs.html bei Publications (5 & 6)):

$$R_0' := \frac{\lambda\theta}{\mu+\kappa},$$

$$R_1' := \frac{\lambda e \log\theta}{\mu\theta^{\frac{\kappa}{\mu}}}$$

und

$$R_2' := \frac{\lambda}{\kappa}.$$

R_2' entspricht in der Tat "unserem R_0 von oben". R_0' konnte von uns noch interpretiert werden; hingegen fanden wir für R_1' keine brauchbare Interpretation!