

Einführung in die Mathematische Epidemiologie

Dr. C.J. Luchsinger

3 Zum Aufwärmen

Problem I, Epidemienausbreitung: Feste Population von N Menschen; diskrete Zeit. Zum Zeitpunkt j seien (S_j, I_j, E_j) die Anzahl (Ansteckbare (Susceptibles), Infektiöse, Erholte), wobei damit für jeden Zeitpunkt gelten muss, dass $S_j + I_j + E_j = N$. Die Menschen verhalten sich im Modell unabhängig voneinander. Sei q die Wahrscheinlichkeit, dass ein gegebener Mensch einen Kontakt (genügend eng) mit einem anderen gegebenen Menschen zwischen zwei aufeinanderfolgenden Zeitpunkten hat.

$$S_{j+1} = S_j - (\text{die neu Infizierten}),$$

$P[\text{ein gegebener Mensch aus den Ansteckbaren nicht in diesem Intervall infiziert wird}] = (1 - q)^{I_j}$. Dann gilt

$$S_{j+1} \sim \text{Bin}(S_j, (1 - q)^{I_j}).$$

Diese Verteilung enthält neben q nur (S_j, I_j) als Parameter und ist somit nur von der Gegenwart j abhängig (aber zweidimensional), falls diese bekannt ist. Weiter zurückliegende Werte haben keinen Wert für Vorhersagen. Wir haben hier die sogenannte Markov-Eigenschaft, welche in der Vorlesung AS zentral ist.

Fragestellungen:

- * Wie sollen wir den Genesungsprozess $I \rightarrow E$ modellieren?
- * Gibt es für $j \rightarrow \infty$ noch Ansteckbare?
- * Gibt es nichttriviale Gleichgewichte? Trivial ist zum Beispiel $(N, 0, 0)$.

Problem II, Wachstum einer Population: Wir betrachten das Wachstum einer Population in diskreter Zeit (Zeit z.B. in Generationen gemessen). Z_n sei die Anzahl Individuen zum Zeitpunkt n . Sei X_{nj} die Anzahl Nachkommen von Individuum j zur Zeit n . Jedes Individuum trägt zu jedem Zeitpunkt zum Wachstum der Population bei. Wir nehmen an, dass $X_{nj}, n \geq 0, j \geq 1$ iid (kaum realistisch). Sei z_0 die Grösse der Population zur Zeit 0 (fest). Dann gilt:

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_{nj} + Z_n.$$

Damit hängt Z_{n+1} nur von Z_n , nicht aber von weiter zurückliegenden Zeitpunkten ab (gegeben Z_n) (Markov-Eigenschaft!).

Fragestellungen:

- * Überlappung der Generationen in stetiger Zeit (ist das ein Problem?)
- * Wie schnell wächst Z_n ?
- * Einbau eines parallelen Sterbeprozesses, Immigration?
- * Wann gibt es einen Gleichgewichtszustand?

Problem III, Sich gegenseitig bekämpfende Populationen: Frage an's Publikum: "Infanterie: Kampf bis zur letzten Patrone". Gleiche Qualität und Motivation der Armeen A und B. A hat 10'000 Mann Infanterie, B hat 5'000 Mann Infanterie. Wie wird etwa der Endstand sein?

Problembeschreibung: Gegeben sind zwei Populationen zur Zeit t (stetige Zeit); $x(t) \geq 0$ und $y(t) \geq 0$ ist die Menge oder Anzahl der einzelnen Populationen (z.B. Bakterien oder Infanterie). Diese Populationen bekämpfen sich nun derart, dass folgendes System von deterministischen Differentialgleichungen erfüllt ist (solange $x(t) \geq 0$ und $y(t) \geq 0$):

$$\dot{x} = -\beta y$$

$$\dot{y} = -\alpha x.$$

Offenbar ist α der Parameter, welcher angibt, wie stark die x -Population pro x -Einheit der y -Population zusetzt und β gibt an, wie stark die y -Population pro y -Einheit der x -Population schadet. Dies ist noch ein deterministisches Modell, welches ein stochastisches Analogon hat. Die Markov-Eigenschaft ist offenbar insofern gegeben, als dass nur der jetzige Zeitpunkt mit den jetzigen Zuständen über die Zukunft entscheidet und die Geschichte keine weiteren Hinweise für die Entwicklung in der Zukunft gibt.

Fragestellungen:

* Wer gewinnt bei gegebenen Anfangswerten; z.B. $x(0) = 5, \alpha = 1, y(0) = 3, \beta = 2$?

* Welche Kombination von $x(0), \alpha, y(0), \beta$ entscheidet über Sieg und Niederlage?

* Wieviel hat die Siegerpopulation am Schluss noch übrig?

Schlussfolgerungen aus diesem Beispiel:

1. Bei kleinen Populationen wird der Zufall wichtiger werden und ein stochastisches Modell ist dann adäquater.

2. Die Antwort auf die zweite Frage hat für die Militärwissenschaften grosse Konsequenzen, welche bereits Clausewitz (1780-1831, "Vom Krieg") bekannt waren: Konzentration der Kräfte - klotzen nicht kleckern (Guderian) - Guerillataktik. Der Name dieses Prinzips ist das sogenannte N^2 -Gesetz von Lanchester (1868-1946). Mehr dazu in Kapitel 18 in der Vorlesung AS.

4 Exponentielles und geometrisches Wachstum, logarithmische Skala

auch im negativen...

5 Bernoulli- und Poisson-Prozesse

Dieses Kapitel ist nicht nur für das Verständnis der nachfolgenden Kapitel zentrale Voraussetzung. Die Bernoulli-Prozesse werden selber in der Extremwert-Theorie (Versicherungsmathematik, Risk-Management) eingesetzt; der Poisson-Prozess ist als Modell für den radioaktiven Zerfall und wie wir in 5.4 sehen werden in zahlreichen Anwendungsgebieten (Versicherungsmathematik, Operations Research, Volkswirtschaftslehre, Epidemiologie und weiteren) wichtiges Arbeitswerkzeug. Beeindruckend werden vor allem die analogen Resultate in diskreter Zeit (Bernoulli-Prozesse) und stetiger Zeit (Poisson-Prozesse) sein.

5.1 Bernoulli-Prozesse [diskrete Zeit - zum Beispiel von Tag zu Tag]

Von Vlsg WTS repetieren wir aus "Ausgewählte Verteilungen":

Bernoulli Be(p)

X kann 2 Werte annehmen: 0 und 1. $P[X = 1] = p$ (Erfolg) und $P[X = 0] = 1 - p$ (Misserfolg). $E[X] = p$ und $V[X] = p(1 - p)$.

$$P[X = x] = p^x(1 - p)^{1-x}$$

Binomial Bin(n,p)

Seien X_i , $1 \leq i \leq n$, n iid Be(p)-Zufallsgrößen. Sei $Y := \sum_{i=1}^n X_i$. Dann hat Y per Definitionem die Binomialverteilung mit Parametern n und p ; Bin(n,p). $E[Y] = np$ und $V[Y] = np(1 - p)$.

$$P[Y = y] = \binom{n}{y} p^y (1 - p)^{n-y}$$

Geometrisch Ge(p)

Seien wieder X_i , $1 \leq i \leq n$, n iid Be(p)-Zufallsgrößen. Wir interessieren uns für den Zeitpunkt des ersten Erfolgs (zum Beispiel erste/nächste Infektion): $Z := \min\{i | X_i = 1\}$.

Z ist eine Zufallsgrösse auf den natürlichen Zahlen ohne die Null. Z hat per Definitionem eine geometrische Verteilung $\text{Ge}(p)$ mit Parameter p . Es gilt: $E[Z] = 1/p$ und $V[Z] = (1-p)/p^2$.

$$P[Z = z] = p(1-p)^{z-1}$$

*

* *

Seien jetzt B_1, B_2, \dots iid $\text{Be}(p)$ -Zufallsgrössen mit:

$$P[B_1 = 1] = p = 1 - P[B_1 = 0] \in (0, 1).$$

Als Motivation stelle man sich einen infektiösen Menschen vor, der diversen Menschen begegnet. Die Begegnung kann eng sein (Infektion) oder lose (keine Infektion). Die Zeit wird gemessen in Anzahl Kontakten der Person. Mit $N_n := \sum_{i=1}^n B_i$ (**dem Bernoulli-Prozess**) können wir dann modellieren, wieviele Personen neu von diesem einen Menschen infiziert wurden.

Wir wollen die Folge der Zufallsgrössen N_n untersuchen. In Satz 5.1 sind einige (zum Teil *sehr* einfache) Eigenschaften aufgelistet. Die Beweise sind (wo überhaupt notwendig) als freiwillige Übung empfohlen.

Satz 5.1 [Eigenschaften des Bernoulli-Prozesses N_n] Für den Bernoulli-Prozess N_n gilt (wir vereinbaren noch $N_0 := 0$):

a) [**Binomialverteilte Zuwächse**] mit natürlichen Zahlen $r; j > k > 0$:

$$P[N_j - N_k = r] = P[\text{Bin}(j - k, p) = r];$$

b) [**unabhängige Zuwächse**] mit natürlichen Zahlen $j_2 > k_2 \geq j_1 > k_1$:

$$N_{j_1} - N_{k_1}$$

und

$$N_{j_2} - N_{k_2}$$

sind unabhängig;

c) [**lineare erwartete Zuwächse**] mit natürlichen Zahlen $j > k$:

$$E[N_j - N_k] = p(j - k);$$

d) [**uniforme Verteilung der Ereignisse**] die Verteilung der "Punkte" (Indizes j , wo $B_j = 1$), gegeben $N_n = m$ ist uniform; genauer: Sei $N_n = m$ gegeben. Die verschiedenen, zufälligen Teil-Mengen $A(\omega)$ von Kardinalität m von $\{1, 2, \dots, n\}$ wo

$$j \in A(\omega) \Leftrightarrow B_j(\omega) = 1,$$

sind (aus Symmetriegründen) alle gleichwahrscheinlich. Die Wahrscheinlichkeit ist

$$\frac{1}{\binom{n}{m}}.$$

e) [**Markov-Eigenschaft**] Die Folge der B_i ist eine Markov-Kette auf $\{0, 1\}$ mit Übergangsmatrix

$$\begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix},$$

die Folge der N_n ist eine vergängliche Markov-Kette auf $\{0, 1, 2, 3, \dots\}$ mit $p_{ii} = 1 - p$ und $p_{i,i+1} = p$ für $i \geq 0$.

f) [**Zeit bis erstes Ereignis**] Die Verteilung des Indexes des ersten Ereignisses ist geometrisch auf $j \in \{1, 2, \dots\}$: Wir definieren mit $T_1 := \min\{j | B_j = 1\}$ die Zeit (**T**ime) des ersten Ereignisses. Dann gilt:

$$P[T_1 = j] = p(1 - p)^{j-1};$$

von der Repetition her haben wir $E[T_1] = 1/p$.

g) [**Zeit für nachfolgende Ereignisse**] Sei $j \geq 2$. Wir definieren mit $T_j := \min\{k | B_k = 1, k > T_{j-1}\}$ die Zeit des j -ten Ereignisses. Dann gilt (definiere für diese kompakte Aussage $T_0 := 0$):

$$T_j - T_{j-1} \sim Ge(p)$$

für $j \geq 1$. Die Folge der $(T_j - T_{j-1})_{j \geq 1}$ ist eine iid $Ge(p)$ -Folge!

h) [Gedächtnislosigkeit der geometrischen Verteilung Sei G eine geometrisch verteilte Zufallsgrösse. Dann gilt für natürliche Zahlen $n > m > 0$:

$$P[G > n | G > m] = P[G > n - m].$$

Wir werden in 5.2 sehr schöne, analoge Resultate in stetiger Zeit herleiten. Zuerst wollen wir ein Paradoxon besprechen, welches in 5.2 noch eindrucksvoller sein wird.

Das Inspektionsparadoxon, 1. Teil (diskrete Zeit)

Stellen Sie sich vor, Sie wollen in einer Krankenstation eine Inspektion machen, um herauszufinden, wieviele Menschen eine bestimmte Krankheit haben. Konkret wollen Sie schauen, ob wie sonst üblich (\mathcal{H}_0) jede hundertste Person die Krankheit hat oder ob der Anteil hier kleiner beziehungsweise grösser ist. Sie wollen nicht die 1000 Patienten durchkontrollieren, sondern gehen mal vorerst zufällig an eine Stelle (zwischen 500 und 501), gehen von dort nach links bis zum ersten Patienten mit der Krankheit, und dann von der gleichen Ausgangsstelle nach rechts bis zur ersten Krankheit. Vom (krankheitsfreien) Zwischenintervall, das Sie erhalten, können Sie hieraus in der Tat auf die Proportion der Kranken erste Rückschlüsse ziehen. Wie geht das? Vorsicht: es heisst Inspektions**paradoxon**.

5.2 Poisson-Prozesse [stetige Zeit]

Von Vlsg WTS repetieren wir aus "Ausgewählte Verteilungen":

Poisson $Po(\lambda)$

Eine Zufallsgrösse V ist poissonisch, wenn Sie Werte auf den natürlichen Zahlen inklusive 0 annimmt und zwar mit folgenden Wahrscheinlichkeiten:

$$P[V = v] = e^{-\lambda} \frac{\lambda^v}{v!}.$$

Es gilt: $E[V] = V[V] = \lambda$.

(Negativ-) Exponential $Exp(\lambda)$

Eine Zufallsgrösse X mit Dichtefunktion

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0,$$

heisst exponentialverteilt mit Parameter λ ; $Exp(\lambda)$. Die gleiche Wahl des Parameters wie bei der Poissonverteilung ist *nicht* zufällig! Es gilt $E[X] = 1/\lambda$ und $V[X] = 1/\lambda^2$ (vgl. mit der $Ge(p)$ -Verteilung!).

*

* *

Man kann wegen Satz 5.1 g) einen Bernoulli-Prozess konstruieren, indem man "einfach eine Folge von iid $Ge(p)$ -Zufallsgrössen nimmt und aneinanderhängt!" Wir machen mal in stetiger Zeit das selbe einfach mit iid exponentialverteilten Zufallsgrössen - mit Erfolg:

Definition 5.2 [Poisson-Prozess] Sei $(T_i)_{i \geq 1}$ eine Folge von iid exponentialverteilten Zufallsgrössen mit Parameter $\lambda > 0$. Wir definieren $\tau_n := \sum_{i \geq 1}^n T_i$ die Folge der Zeitpunkte von Ereignissen. Wir nennen

$$X_t := \sum_{n \geq 1} I[\tau_n \leq t]$$

einen Poisson-Prozess mit Parameter (oder Intensität, Rate) λ .

Man vergleiche jeweils die Resultate in Satz 5.3 mit denen aus Satz 5.1:

Satz 5.3 [Eigenschaften des Poisson-Prozesses X_t] *Der in Definition 5.2 konstruierte Prozess X_t hat folgende Eigenschaften:*

a) [**Poisson-verteilte Zuwächse**] *mit natürlicher Zahl r und reellen Zahlen $t > s \geq 0$:*

$$P[X_t - X_s = r] = P[Po(\lambda(t - s)) = r];$$

Warnung: $\lambda(t - s)$, also Rate mal Zeitdauer und nicht nur λ !

b) [**unabhängige Zuwächse**] *mit reellen Zahlen $j_2 > k_2 \geq j_1 > k_1$:*

$$X_{j_1} - X_{k_1}$$

und

$$X_{j_2} - X_{k_2}$$

sind unabhängig;

c) [**lineare erwartete Zuwächse**] *mit reellen Zahlen $t > s$:*

$$E[X_t - X_s] = \lambda(t - s);$$

d) [**uniforme Verteilung der Ereignisse**] *Gegeben $X_t = r$, so haben die Ereignisse $(\tau_1, \tau_2, \dots, \tau_r)$ die gleiche Verteilung wie die geordnete Stichprobe vom Umfang r aus der $U[0, t]$ -Verteilung.*

e) [**Gedächtnislosigkeit der exponentiellen Verteilung**] *Sei T eine exponentiell verteilte Zufallsgrösse. Dann gilt für reelle Zahlen $t > s > 0$:*

$$P[T > t | T > s] = P[T > t - s].$$

Mithilfe von Satz 5.3 a) und d) kann man auch "rückwärtsgerichtet" einen Poissonprozess konstruieren:

Beweis: in der Vorlesung Angewandte Stochastik.

Das folgende Lemma beinhaltet zwei wichtige Eigenschaften von exponentialverteilten Zufallsgrößen. Sie sind wichtige Grundlage für Simulationen.

Lemma 5.4 *Seien $(X_1, X_2, X_3, \dots, X_n)$ n unabhängige Zufallsgrößen, wo X_i exponentialverteilt mit Parameter μ_i ist. Wir definieren $Y := \min(X_1, X_2, X_3, \dots, X_n)$, $\mu := \sum_{i=1}^n \mu_i$. Dann gelten*

- a) $\mathbb{P}[Y \leq t] = 1 - e^{-\mu t}$,
- b) $\mathbb{P}[X_i = Y] = \mu_i / \mu$.

Bemerkung zu Lemma 5.4 a) sagt aus, dass Y wieder exponentialverteilt ist mit Parameter μ ; in b) wird gesagt, dass die Wahrscheinlichkeit, dass die i -te Zufallsgröße die kleinste ist, dem Anteil des i -ten Parameters an der Gesamtrate μ entspricht.

Beweis von Lemma 5.4: Wir beweisen dieses Lemma vorerst für den Fall $n = 2$.

a) Wegen der Unabhängigkeit können wir folgern:

$$\begin{aligned} \mathbb{P}[Y \leq t] &= 1 - \mathbb{P}[\{X_1 > t\} \cap \{X_2 > t\}] = 1 - \mathbb{P}[X_1 > t] \mathbb{P}[X_2 > t] \\ &= 1 - e^{-(\mu_1 + \mu_2)t}. \end{aligned}$$

b)

$$\begin{aligned} \mathbb{P}[X_1 < X_2] &= \int_0^\infty \mathbb{P}[X_1 < X_2 | X_2 = x] e^{-\mu_2 x} \mu_2 dx \\ &= \int_0^\infty (1 - e^{-\mu_1 x}) e^{-\mu_2 x} \mu_2 dx \\ &= \int_0^\infty e^{-\mu_2 x} \mu_2 dx - \int_0^\infty e^{-(\mu_1 + \mu_2)x} \mu_2 dx \\ &= 1 - \frac{\mu_2}{\mu_1 + \mu_2} = \frac{\mu_1}{\mu_1 + \mu_2}. \end{aligned}$$

Wie verallgemeinert man jetzt dieses Resultat auf den Fall $n > 2$?

□

Lemma 5.5 [Superposition und Thinning] a) $X_1(t), X_2(t), \dots, X_n(t)$ seien n unabhängige Poisson-Prozesse mit Raten $\lambda_1, \lambda_2, \dots, \lambda_n$. Dann ist

$$\sum_{i=1}^n X_i(t)$$

ein Poisson-Prozess mit Rate $\sum_{i=1}^n \lambda_i$.

b) Sei $p \in (0, 1)$ (eine "Eliminationswahrscheinlichkeit"). Jedes Ereignis (jeder Punkt) des Poissonprozesses $X_1(t)$ wird unabhängig vom Prozess $X_1(t)$ und von den anderen Zeitpunkten mit Wahrscheinlichkeit p eliminiert (nicht berücksichtigt) und mit Wahrscheinlichkeit $1 - p$ gelassen. Der so konstruierte Prozess \hat{X}_1 ist ein Poissonprozess der Rate $(1 - p)\lambda_1$.

Beweis von Lemma 5.5: a) folgt aus Lemma 5.4 a) (siehe Konstruktion/Definition des Poisson-Prozesses; warum?) und b) ist als freiwillige Übung gedacht.

□

Das Inspektionsparadoxon, 2. Teil (stetige Zeit)

Wenn wir den Poissonprozess vor der Zeit 0 fortsetzen, so erhalten wir ein noch eindrücklicheres Paradoxon als in diskreter Zeit. Dies deshalb, weil die Wahrscheinlichkeit, dass es gerade am Inspektionsort einen Punkt hat, bei stetigen Wahrscheinlichkeitsdichten gleich 0 ist. Also ist hier *bei einer Inspektion* die Länge des erwarteten Intervalles ohne Zwischenpunkt ("nach links, nach rechts") gleich

$$\frac{2}{\lambda};$$

dies ist zu vergleichen mit der erwarteten Zwischenzeit zwischen 2 Ereignissen (nur nach Vorne schauen!) der Länge

$$\frac{1}{\lambda}.$$

5.3 Diskrete und stetige Analoga

Wir haben gesagt, dass die geometrische Verteilung das diskrete Analogon der Exponential-Verteilung ist. Die graphische Darstellung der Wahrscheinlichkeitsfunktion bzw. der Dichte gibt einen ersten Hinweis darauf:

Des weiteren erhalten wir bei beiden Zufallsgrößen einen vergleichbaren Ausdruck für

$$P[X > t] :$$

Die Gedächtnislosigkeit beider Verteilungen sind ein weiterer Hinweis. Wir wollen jetzt, wo wir diese Verteilungen besser kennen, das zwingende noch genauer anschauen:

5.4 Modellierung in stetiger Zeit bei abzählbaren Zustandsräumen

Wir werden jetzt allgemein erklären, wie man die Resultate aus 5.2 zur Modellierung und Simulation verwenden kann. In Kapitel 6 folgt danach die Anwendung davon. Mehr über Simulationen auf Rechnern findet man in Kapitel 8 von "Introduction to Stochastic Calculus Applied to Finance" von D. Lamberton und B. Lapeyre (siehe dazu www.math-jobs.com/lit.html – – – > Options & Futures).

Das allgemeine Vorgehen:

- 1. Wann findet wieder ein Ereignis statt? (Lemma 5.5 a) bzw. Lemma 5.4 a) mit aktueller Gesamtrate)**
- 2. Was ist bei Punkt 1. passiert? (Lemma 5.4 b) mit Anteilen)**
- 3. Neu einstellen, neue Gesamtrate goto 1.**

Wenn man nur die Zeitpunkte des Ereignisses anschaut, so haben wir eigentlich einen normalen Poisson-Prozess bis zum allerersten Ereignis. Dann ändert sich die Rate. Dann haben wir wieder einen Poisson-Prozess (mit neuer Rate) bis zum nächsten Ereignis und so weiter und so fort.

Wenn in der Literatur von einem stochastischen Modell und von einer Rate gesprochen wird, dann ist damit (zumindest bei stetiger Zeit und abzählbaren Zustandsräumen) diese obige Modellierung gemeint. Man kann schon andere Übergangsraten (anders als Exp) einbauen - z.B. mit Time Lag bis zum nächsten Ereignis - aber dann muss man dies explizit angeben. Das gleiche gilt auch für allfällige Abhängigkeiten. Das obige ist also der Default wenn nichts anderes geschrieben steht.

Wir üben das jetzt an ein paar Aufgaben:

Reiner Geburtsprozess (Yule-Prozess)

1. Am Anfang haben wir 1 Parasit [$X_0 = 1$]
2. Alle Parasiten erzeugen iid Nachkommen mit Rate $\lambda > 0$
3. Geben Sie Übergangsraten und Simulationsstrategie an.
4. Geben Sie auch das deterministische Analogon (DGL) an.

Geburtsprozess mit Immigration

1. Am Anfang haben wir 1 Parasit [$X_0 = 1$]
2. Alle Parasiten erzeugen iid Nachkommen mit Rate $\lambda > 0$
3. Dazu gibt es von aussen eine Immigration von Parasiten mit konstanter Rate $\eta > 0$
4. Geben Sie Übergangsraten und Simulationsstrategie an.
5. Geben Sie auch das deterministische Analogon (DGL) an.

Geburts- und Todesprozess

1. Am Anfang haben wir 1 Parasit [$X_0 = 1$]
2. Alle Parasiten erzeugen iid Nachkommen mit Rate λ
3. Parasiten sterben iid mit Rate $\mu > 0$ - der Sterbeprozess ist II vom Geburtsprozess
4. Geben Sie Übergangsraten und Simulationsstrategie an.
5. Geben Sie auch das deterministische Analogon (DGL) an.

”Kleine Nebenbemerkung”: Man beachte, dass es wie immer sehr einfach ist, solche Systeme zu *simulieren* - die theoretische Untersuchung kann beliebig schwierig werden... .

6 Epidemienmodelle - Outbreak

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

6.1 SIR - deterministisch 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 re-tourübersetzt 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 Empfänglichen 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 stochasti-

schen 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 6.1 [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 6.2 und 6.3.

Diagramm 6.2 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 6.3 beschreibt die Situation, wo rechtzeitig ein Impfprogramm lanciert wurde. Mit gleichem Anteil Infizierten am Anfang wie bei Diagramm 6.2 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 Realität 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 5.4 b)). Solche Fragen können nicht mit Modell (SIR) beantwortet werden. Hierfür brauchen wir einen vollständig neuen Ansatz:

6.2 Ein stochastisches Analogon zu SIR für die Initialphase

Zu vergleichen mit dem Geburts- und Todesprozess (pure Birth and Death-Process) in 5.4 - 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 6.2 und 6.3 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 5.5 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:

$$\begin{aligned}i(t) &\rightarrow i(t) + 1 \text{ mit Rate } \lambda i(t) \\i(t) &\rightarrow i(t) - 1 \text{ mit Rate } \mu i(t).\end{aligned}\tag{6.1}$$

Zur Zeit t finden Ereignisse statt mit einer totalen Rate von

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

(Lemma 5.4 a) bzw. Lemma 5.5 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 5.4 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 6.1).

$$R_0 \stackrel{\leq}{\geq} 1$$

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 (6.1) 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

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

Wegen der Unabhängigkeit der Kontakte der Infizierten gilt dann

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

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 von AS (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.

6.3 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 (6.1) *direkt* zusammengehören. Indirekt hängen sie zusammen: Modell (6.1) ist das (lineare) stochastische Modell für die *Anfangsphase* von (SIR). Das deterministische Analogon im Sinne von **a)** von (6.1) ist die DGL

$$\frac{dI}{dt} = \lambda I - \mu I. \quad (6.2)$$

Die Lösung hiervon ist $I(0)e^{(\lambda-\mu)t}$. Weil für $i(t)$ von (6.1) gilt $\mathbb{E}[i(t)] = i(0)e^{(\lambda-\mu)t}$ (ohne Beweis), sind Modell (6.1) und (6.2) analog im Sinne von **a)**.

Wie wird man Modell (6.1) und (6.2) 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!