

Seuchenausbreitungsmodell und Bernoullikette

von Andrew J. Baczkowski – Bearbeitung: K. Krug, Coburg

Zusammenfassung: Die Benutzung von Mikrocomputern befähigt Studenten, Eigenschaften eines einfachen Modells über die Ausbreitung von Epidemien zu untersuchen, das auf der Binomialverteilung beruht.

Das Modell der Bernoulli-Kette ist ein einfaches Modell für die Ausbreitung von Epidemien. Nichtsdestoweniger kann es eine breite Palette von interessanten Ergebnissen liefern, von denen einige dem Verhalten von allgemeineren Epidemiemodellen entsprechen. Es ist von uns erfolgreich als Studienprojekt auf mehreren unterschiedlichen Leistungsniveaus eingesetzt worden. In der Vergangenheit haben andere das Bernoullimodell dazu benutzt, um die Ausbreitung von Krankheiten in kleinen Populationen, z.B. Familien, aufzuzeigen. Die Benutzung eines Mikrocomputers aber macht es möglich, das Modell auf Grund von Simulationen an größeren Populationen zu untersuchen.

Das Modell

Das Buch von Bailey (1964) enthält eine Beschreibung des Begriffes der Bernoulli-Kette. Auch viele deutsche Schulbücher behandeln die Bernoulli-Kette der Länge n als einfache Wiederholung eines Bernoulli-Experimentes (z.B. Stochastik Grundkurs Cornelsen, Stochastik Leistungskurs Ehrenwirth, Wahrscheinlichkeitsrechnung und Statistik Klett). Lassen Sie uns zunächst kurz einen Infizierten als eine Person definieren, die die Fähigkeit zur Übertragung einer Krankheit besitzt und einen Ansteckbaren als eine Person, die in der Lage ist, die Krankheit zu erlangen. Man nehme nun an, daß es zur Zeit t in einer Population $I(t)$ infizierte und $S(t)$ ansteckbare Personen gibt, wobei jede ansteckbare Person mit der Wahrscheinlichkeit q von jeder infizierten Person nicht angesteckt wird.

Bei $I(t)$ Infizierten ist unter der Voraussetzung stochastischer Unabhängigkeit für jede ansteckbare Person die Wahrscheinlichkeit nicht infiziert zu werden $q^{I(t)}$; dagegen ist die Wahrscheinlichkeit für eine Ansteckung $1 - q^{I(t)}$. Angenommen die $S(t)$ ansteckbaren Personen sind der Reihe nach stochastisch unabhängig voneinander, dann ist zur Zeit $t+1$ die Zahl der Infizierten $I(t+1)$, wozu eine Binomialverteilung mit den beiden Parametern

$$N = S(t), \quad p = 1 - q^{I(t)}$$

gehört. Die Zahl der zur Zeit $t+1$ für die Krankheit ansteckbaren Personen wird durch

$$S(t+1) = S(t) - I(t+1)$$

angegeben und die $I(t)$ zur Zeit t infizierten Personen schließen sich der Gruppe der Personen an, die die Krankheit gehabt haben und nicht länger ansteckbar oder infiziert sind. Die Zahl der zu jedem diskreten Zeitpunkt durch die Epidemie infizierten Personen wird auf diese Weise als Kette von Binomialverteilungen dargestellt.

Das Modell setzt voraus, daß q für alle ansteckbaren Personen und zu allen Zeiten t eine Konstante ist. Zwei andere entscheidende Annahmen sind, erstens die Unabhängigkeit der Ansteckung durch die erkrankten Personen, und zweitens die Unabhängigkeit der Reaktion durch jede ansteckbare Person. Man kann mit einem nur begrenzten Wissen in Statistik die Stichhaltigkeit dieser Annahmen an Epidemien im wirklichen Leben studieren.

Auf einer fortgeschritteneren Stufe kann der Verlauf einer Epidemie untersucht werden: Eine Möglichkeit ist die Variation des Parameters q ; hierbei ist eine feste Anfangszahl $S(0)$ an ansteckbaren Personen vorauszusetzen sowie $I(0)=1$ anzunehmen, d. h. zu Beginn von einem Infizierten auszugehen. Eine realistischere Möglichkeit wäre, die Wahrscheinlichkeit q , sich nicht zu infizieren als fest anzunehmen, die Epidemie mit $I(0)=1$ zu starten und eine Reihe von Werten für $S(0)$ zu untersuchen. Akzeptiert man die erste Vorgehensweise, so ergibt sich eine effizientere Nutzung der Rechenzeit, da q nur Werte zwischen Null und Eins annimmt.

Computer-Algorithmus

Als Folge von Binomialverteilungen ist das Modell relativ leicht zu programmieren. Um binomiale Zufallsgrößen zu erzeugen, wird vom Bernoulli-Modell Gebrauch gemacht. Bilde

$$I(t+1) = X_1 + X_2 + \dots + X_n,$$

dabei sind die X_i Indikatoren, die den Wert 1 mit der Wahrscheinlichkeit p und 0 mit der Wahrscheinlichkeit $1-p$ annehmen und $n = S(t)$. Wenn U eine Zufallszahl zwischen 0 und 1 ist, dann setze $X_1=1$, falls U kleiner als p ist, sonst setze $X_1=0$. Dies wiederholt man für X_2, X_3 usw. Ein Basic-Algorithmus, der Werte $INEW = I(t+1)$ und $SNEW = S(t+1)$ aus den gegebenen Werten $IOLD = I(t)$ und $SOLD = S(t)$ erzeugt, ist :

```

10  REM IOLD UND SOLD ENTHALTEN DIE WERTE VON I(t) UND S(t)
20  INEW = 0
30  P = 1 - Q^IOLD
40  FOR K = 1 TO SOLD
50  U = RND(1)
60  IF U < P THEN INEW = INEW + 1
70  NEXT K
80  SNEW = SOLD - INEW
90  REM INEW UND SNEW ENTHALTEN DIE WERTE VON I(t+1) und S(t+1)

```

Der obige Algorithmus kann dann schleifenweise ausgeführt werden, so daß eine Folge von Werten $I(1), S(1); I(2), S(2);$ usw. erhalten werden kann. Man muß achtgeben, wenn bei einem Durchlauf $S(t) = 0$ oder $I(t) = 0$ ist, um ein Abstürzen des Algorithmus zu verhindern.

Gesamtzahl der infizierten Personen

Die Gesamtzahl, der während einer Epidemie infizierten Personen ist durch $T = S(0) - S(\infty)$ gegeben, wobei $S(\infty)$ die Zahl der übriggebliebenen nichtinfizierten Personen ist, sobald einmal $I(t)$ den Wert Null annimmt. Für jede beliebige Simulation der Epidemie kann der Wert T bestimmt werden. Wiederholt man dies oft, so kann ein Diagramm der Verteilung von T für verschiedene q gezeichnet werden. Tabelle 1 beschreibt das Diagramm, das man im Falle $I(0) = 1$ und $S(0) = 100$ erhielt.

Tabelle 1: Diagrammverläufe von T für verschiedene q Werte

$q < 0,900$	einen Höchstwert bei 100
$0,900 < q < 0,950$	Verteilung in J-Form, mit häufigstem Wert
$0,950 < q < 0,985$	bimodale Form, mit den häufigsten Werten 0 und nahe 100
$0,988 < q$	Verteilung in J-Form, mit häufigstem Wert 0

Abbildung 1 und 2 zeigen mit den Anfangswerten $I(0)$ und $S(0)$ für eine Reihe von q Werten den Mittelwert und die Standardabweichung für die Gesamtzahl aller infizierten Personen bei 5000 Simulationen.

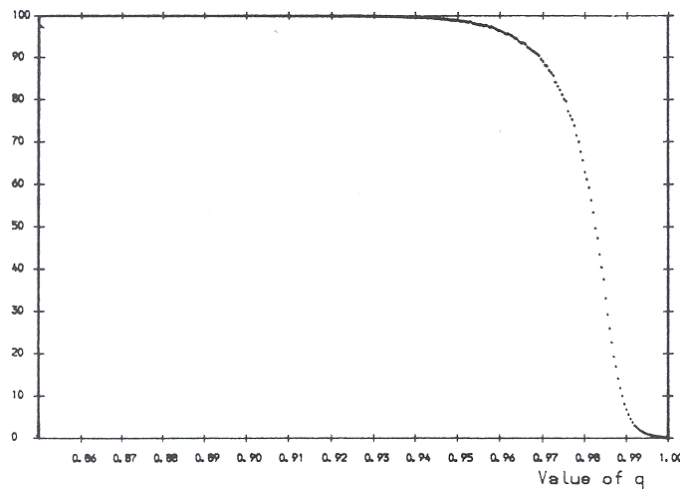


Abb. 1: Mittelwert der infizierten Personen

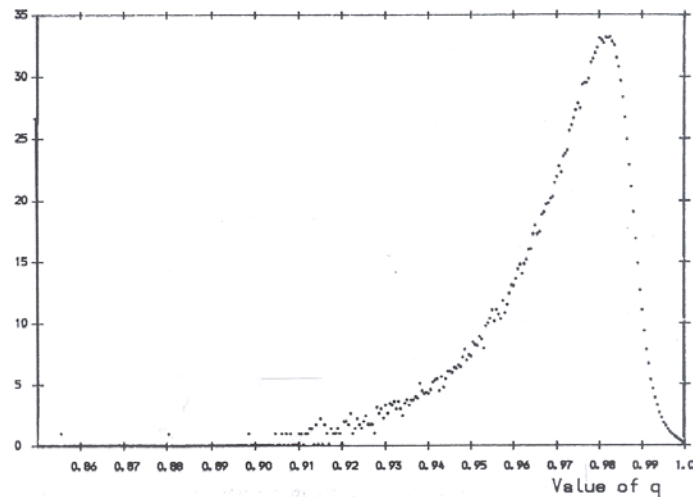


Abb. 2: Standardabweichung der Zahl der infizierten Personen

Für q kleiner als 0,85 ist der Mittelwert gleich 100 mit einem Standardfehler nahe Null. Wenn q von 0,93 bis 1 wächst, nimmt der Mittelwert auf Null ab, mit einer Standardabweichung, die bei 0,983 maximal ist. Nur ein kleiner Bereich von q Werten ergibt für die Gesamtzahl der infizierten Personen einen von 0 oder 100 abweichenden Wert. Variiert man q für die gegebenen Werte von $I(0)$ und $S(0)$ über diesen vergleichsweise kleinen Bereich von Werten, so liefert dies erheblich unterschiedliche Epidemieverläufe.

Dauer der Epidemie

Die Dauer der Epidemie kann definiert werden als die Zeit, bis $S(t)$ oder $I(t)$ gleich Null wird. Durch wiederholte Simulationen einer Epidemie kann ein Diagramm der Dauer erstellt und für verschiedene Werte von q näher untersucht werden. Wir richten wieder die Aufmerksamkeit auf den Fall $I(0)=1$ und $S(0)=100$.

Für Werte von q nahe bei Null ist die Epidemiedauer gleich Eins, wobei alle Personen infiziert werden. Wächst q geringfügig, so treten als Epidemiezeiträume Eins und Zwei auf. Für q zwischen 0,03 und 0,7 entspricht die Dauer immer Zwei. Wächst q weiter, so kommen die Zeiträume Zwei und Drei vor; die Dauer verschiebt sich überwiegend nach Drei für q zwischen 0,8 und 0,9. Danach beschreibt Tabelle 2 die Gestalt des festgestellten Diagramms für die Epidemiedauer.

Tabelle 2 : Diagrammformen für die Epidemiedauer mit $S(0)=100$

$q=0,9$	Dauer ist gleich 3 oder 4
$q=0,915$	Zeiträume 3 und 4 ungefähr gleich wahrscheinlich
$q=0,93$	Zeiträume sind 3,4 und 5 mit häufigstem Wert 4
$q=0,94$	J-förmige Verteilung mit häufigstem Wert 4 und Zeiträumen 4, 5, 6
$q=0,95$	häufigster Wert verschiebt sich nach 5
$q=0,96$	häufigster Wert 6, einige wenige Zeiträume 1 tauchen auf
$q=0,97$	häufigster Wert 8, typische maximale Dauer 12
$q=0,98$	Verteilung mit zwei Häufungswerten, davon einer bei 1
$q=0,985$	wie bei 0,98
$q=0,99$	J-förmige Verteilung mit häufigstem Wert 1, maximale Dauer 15
$q=0,995$	wie vorher, aber die maximale Dauer liegt bei 5
$q=0,999$	wie vorher, aber die maximale Dauer ist gleich 2

Abbildung 3 gibt die Mittelwerte der Epidemiedauer für q -Werte zwischen 0 und 1 an. Die Abbildung zeigt, daß der Mittelwert im großen Bereich der q -Werte von 0,05 bis 0,7 gleich Zwei ist. Wächst dann q , so nimmt auch der Mittelwert zu. Bei $q=0,985$ ist die durchschnittliche Epidemiedauer ca. 9,5 bei einer Standardabweichung von etwa 6. Für größere q fällt der Mittelwert auf Eins zurück. Eine Epidemiedauer von Eins entspricht dem Fall, daß entweder alle Personen sofort infiziert werden oder niemand wird infiziert, und die Ausbreitung der Krankheit hört sofort auf.

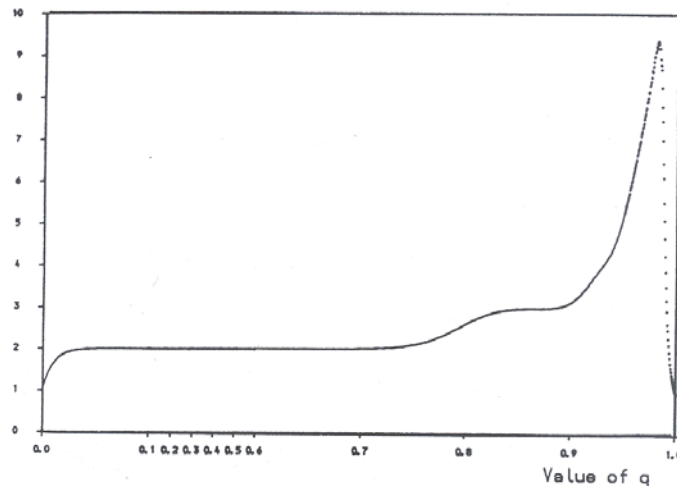


Abb.3: Graph des Mittelwertes der Epidemiedauer

Von Interesse ist die Frage, ob die Epidemie beendet ist, weil $S(t) = 0$ ist und es keine weiteren ansteckbaren Personen gibt, oder weil $I(t) = 0$ ist und es keine weiteren infizierten Personen zur Weitergabe der Krankheit gibt. Das ist ein Arbeitsgebiet, das sich für weitere Untersuchungen durch Studenten eignet.

Deterministisches Modell

Ein deterministisches Analogon zum Bernoulli-Modell entsteht beim Blick auf den Mittelwert $E(I(t+1))$ von $I(t+1)$, der für gegebenes $I(t)$ und $S(t)$ durch

$$E(I(t+1)) = S(t) \times (1 - q^{I(t)})$$

gegeben ist. Wir nehmen an, daß $I(t+1)$ gleich diesem Mittelwert ist, so daß

$$I(1) = S(0) \times (1 - q^{I(0)}), \quad I(2) = S(1) \times (1 - q^{I(1)}) \text{ usw. .}$$

Für gegebenes $S(0)$ und $I(0)=1$ kann dieser Fall nun für eine Reihe von q Werten untersucht werden. Ein Rechner kann dazu benutzt werden, um das typische Verhalten des Modells zu zeigen.

Tabelle 3 gibt Werte von $I(t)$ für verschiedene q Werte, beginnend mit $S(0)=100$ und $I(0)=1$, wieder. Für kleine q Werte hört das Verfahren schnell auf. Für dazwischenliegende Werte von q , die aber kleiner als 0,99 sind, klettert die Zahl der infizierten Personen $I(t)$ auf einen Höchstwert, bevor sie auf Null fällt. Für q Werte größer als 0,99 ist das Ergebnis eine sich lang hinziehende Epidemie mit $I(t)$ kleiner als Eins für alle Zeiten t .

Tabelle 3 : I(t)-Werte beim deterministischen Modell

t	q = 0,50	q = 0,90	q = 0,95	q = 0,99
0	1	1	1	1
1	50	10	5	1
2	50	58,6	21,5	1
3	0	31,3	40,1	1
4		0,1	22,4	0,9
5		0	1,3	0,9
6			0	0,9
7				0,8

Offensichtlich gibt es eine Art Schwelle oder einen kritischen Wert für q , bei dessen Unterschreiten eine Epidemie vorkommen kann und über dessen Wert sie nicht auftritt. Es kann gezeigt werden, daß keine Zunahme an infizierten Personen vorkommen kann, falls $q^{I(t)}$ kleiner als $1 - I(t)/S(t)$ ist. Setzt man $I(0) = 1$ und $S(0) = 100$ ergibt sich ein kritischer Wert von $q = 0,99$. Eine Zunahme der $I(t)$ bevor sie auf Null zurückfallen, liefert eine Definition dafür, was mit einer Epidemie gemeint ist.

Die darauf gründenden Epidemiemodelle, s. z. B. Nöbauer/Timischl oder Frauenthal, versuchen die Ausbreitung von Infektionskrankheiten in einer Population mit Hilfe mathematischer Modelle zu beschreiben.

Eine Frage, die Studenten hier betrachten können, lautet, ob im deterministischen Modell für $I(t)$ nicht-ganzzahlige Werte erlaubt sein sollten oder nicht.

Diskussion

Das Modell der Bernoulli-Kette hat Statistikstudenten ein herausforderndes und anregendes Arbeitsfeld verschafft. Als Studiengebiet findet es Verwendung wegen seines sehr unterschiedlichen Verhaltens bei Variation des Parameters q . Benützt man einen Mikrocomputer, dann können die Studenten ein Modell einer realen Situation anfertigen und Simulationstechniken in die Praxis umsetzen. Dies kann zu einem Studium anderer Epidemiemodelle führen.

Das oben dokumentierte Epidemiemodell wurde zusammen mit einer Liste möglicher Themen vorgestellt, die die Studenten im Detail untersuchen konnten, so z.B. die Epidemiedauer, die Gesamtzahl der infizierten Personen und das deterministische Modell. Viele Studenten fanden für sich selbst andere interes-

sante Themen, die sie näher untersuchen wollten.

Ein weiteres interessantes Thema wäre die Diskussion von Epidemiekurven, in denen der Anstieg bei den Durchschnittszahlen der infizierten Personen gegen die Zeit aufgetragen wird. Da es sich hier um ein diskretes Zeitmodell handelt, könnte man eher die Differenz der Mittelwerte für aufeinanderfolgende Zeitwerte betrachten, als zu versuchen die Steigung bzw. den Anstieg über Differentiation zu erhalten.

Ein anderer Vorschlag besteht darin, die Endzahl $S(\infty)$ der nichtinfiziert gebliebenen Personen zu betrachten; diese steht für festes $S(0)$ in eindeutiger Beziehung zur Gesamtzahl T der infizierten Personen. Ebenfalls von Interesse wäre eine Untersuchung der Endwerte $I(t)$ und $S(t)$ in Abhängigkeit von q , um zu sehen, welches Verhaltensmuster besteht.

Eine Untersuchung, wie die (anfänglichen) Annahmen das Modell beeinflussen, wäre von Wert. Ein mit der Zeit sich änderndes q , würde dem sich ändernden Verhalten der Menschen entsprechen, das darauf einwirkt, wie die Krankheit weitergegeben wird. Führt man eine Latenzperiode ein, in der ein Infizierter die Infektion nicht übertragen kann, so würde dies Spielraum für weitere Studien verschaffen.

Ein spezielles Modell von Infektionskrankheiten wird bei König behandelt. Im Abschnitt Epidemien wird der Verlauf der Immunschwächekrankheit AIDS diskutiert.

Literatur

- Bailey, N.T.J. (1964) : *The Elements of Stochastic Processes with applications to the Natural Sciences*. New York; Wiley
- Frauenthal, J. C. (1980): *Mathematical Modelling in Epidemiology*. Berlin, Heidelberg, New York: Springer
- König, G. (1991): AIDS und Mathematikunterricht. *ZDM* 23 (6), 207-220
- Nöbauer, W.; Timischl, W. (1979): *Mathematische Modelle in der Biologie, Braunschweig*. Wiesbaden: Vieweg