Gerüchte und Vehler

Über das Laster, beim Programmieren (etwa eines Gerüchtemodells) Fehler zu machen, und was man daraus lernen kann

Von Brian Hayes

Der »bekennende Essay« ist in der Mathematik oder anderen Wissenschaften kein verbreitetes Genre. Nur wenige von uns mögen sich zu ihren Fehlern bekennen oder sie ins Rampenlicht rücken. Eine rühmliche Ausnahme bildet Donald E. Knuth von der Universität Stanford. Während seiner zehnjährigen Arbeit am Schriftsatzprogramm Tex hat der heute emeritierte Professor akribisch genau alle seine Fehler aufgezeichnet und die Liste dann mit detaillierten Kommentaren veröffentlicht.

Ich hab Knuths Mut zur öffentlichen Fehlerbeichte bewundert – und möchte seinem Beispiel hier folgen. Dabei hoffte ich, »Don« Knuth sogar auf diesem Gebiet zu übertrumpfen: Ich könne noch mehr Fehler machen als er. Leider fiel ich auch darin hinter ihn zurück. Knuths publizierte Fehlerliste umfasst 900 Einträge, während ich nur eine beschämende Hand voll Fehler vorweisen kann. Er benötigte allerdings zehn Jahre Arbeit, um diese hohe Zahl zu erreichen.

Knuth merkt an, dass ihm die Liste nicht nur bei der Beseitigung von Bugs, also Programmierfehlern geholfen hat, sondern auch dabei, sich »selbst besser kennen zu lernen«. Auch ich glaube, durch Erfahrung und Konfrontation mit meiner eigenen Fehlbarkeit einiges an Selbsterkenntnis hinzugewonnen zu haben. Und es wäre erfreulich, wenn ich davon ausgehen könnte, andere durch meine Ausführungen davor zu bewahren, die gleichen Fehler zu begehen. Tatsächlich jedoch mag ich nicht so recht daran glauben – und womöglich ist es nicht einmal eine gute Idee.

Die Geschichte beginnt mit einigen Anekdoten anlässlich der Recherche zu meinem Artikel über die Lambert-W-Funktion. Ich suchte nach einer Studie mit dem kuriosen Titel »Gerüchte mit allgemeinen Anfangsbedingungen«, verfasst von Selma Belen und C. E. M. Pearce von der University of Adelaide (Australien) und publiziert im Journal »Anziam« (siehe Literaturhinweise). Als ich das Paper endlich aufgetrieben hatte, war mein Artikel bereits im Druck. Dies war sehr schade, denn Belen und Pearce beschreiben eine aufschlussreiche Anwendung der W-Funktion in einem Kontext, den ich sehr interessant fand. Hier der Anfang ihres Artikels:

»Die stochastische Theorie der Gerüchte – mit ihren Gruppen von Nichtwissern, Verbreitern und Unterdrückern – nahm ihren Anfang mit der Studie von Daley und Kendall. Darin wurde auch das interessanteste Ergebnis auf diesem Gebiet präsentiert: dass, wenn es anfangs nur einen Verbreiter gibt, der Anteil der Bevölkerung, die das Gerücht niemals zu Ohren bekommt, sich nahezu sicher dem Wert von 0,203188 der gesamten Bevölkerungsgröße annähert, während diese gegen unendlich geht. Dieses Ergebnis errechnet sich auch nach dem stochastischen Modell von Maki und Thompson; leider wurde der darin auf Grund eines Druckfehlers genannte Wert von 0,238 auch in weiteren darauf basierenden Publikationen zitiert.«

Der Spaß, ein Gerücht zu verbreiten

Ich war fasziniert, dass ein Gerücht, wenn es sich wieder legt, »nahezu sicher« ein Fünftel der Leute niemals erreicht hat. Warum erreicht es nicht irgendwann alle? Und diese Zahl 0,203188, mit ihren hochpräzisen sechs Nachkommastellen, wo kam *die* her?

Ich habe im Artikel genug gelesen, um die Details dieses Modells zu verstehen. Die Grundannahme ist, wie ich herausfand, dass die Verbreitung eines Gerüchts nur dann Spaß macht, wenn man es selbst kennt, die Zuhörer aber nicht. Es verschafft kein so prickelndes Gefühl, abgeschmackte News zu verkaufen. Im Sinne der obigen drei Gruppen verbreiten Leute ein Gerücht so lange, wie sie Nichtwissern begegnen, die es begierig aufnehmen. Danach werden die Verbreiter zu Unterdrückern, die das Gerücht zwar kennen, aber kein Interesse daran haben, es weiterzugeben. Die Modelle von Daley-Kendall und Maki-Thompson vereinfachen diese sozialen Prozesse. Beide gehen von einer sorgfältig »durchmischten« Bevölkerung aus, in der sich Leute stets zufällig und mit gleich bleibender Wahrscheinlichkeit treffen. Eine weitere Vereinfachung besteht darin, dass sich nur jeweils zwei Leute treffen, niemals größere Gruppen. Diese Paar-Treffen unterliegen strikten Regeln:

- Jedes Mal, wenn ein Verbreiter auf einen Nichtwisser trifft, wird der Nichtwisser zum Verbreiter, während der ursprüngliche Verbreiter ein solcher bleibt.
- Trifft ein Verbreiter auf einen Unterdrücker, so wird der Verbreiter selbst zum Unterdrücker.
- Für den Fall, dass sich zwei Verbreiter treffen, unterscheiden sich die Modelle. In der Daley-Kendall-Variante werden beide zu Unterdrückern. Im Maki-Thomson-Modell wird einer der beiden Verbreiter zum Unterdrücker, während der andere Verbreiter bleibt.
- Alle anderen Interaktionen (Nichtwisser-Nichtwisser, Unterdrücker-Unterdrücker) haben keinen wechselseitigen Einfluss.

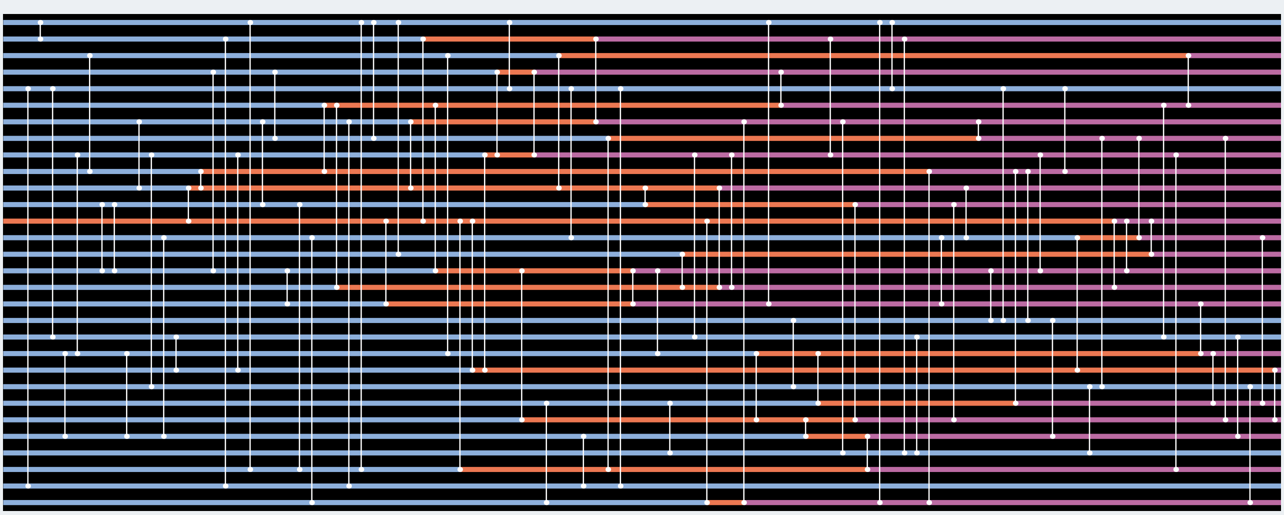
Die Modellregeln legen nahe, warum Gerüchte selbstbegrenzend wirken: Anfangs werden neue Verbreiter aus dem Riesenreservoir an Nichtwissern rekrutiert, wobei sich das Gerücht in einem Teil der Bevölkerung ausbreitet. Nimmt die Zahl der Verbreiter jedoch stark zu, treffen sie zunehmend aufeinander und werden zu Unterdrückern. Da der Übergang vom Nichtwisser zum Verbreiter und vom Verbreiter zum Unterdrücker jeweils unumkehrbar ist, ergibt sich zwangsläufig, dass das Gerücht irgendwann verebbt. Am Ende sind alle Verbreiter zu Unterdrückern geworden. Nicht ganz so klar ist, warum der letzte Verbreiter verschwindet, bevor das Reservoir an Nichtwissern gänzlich erschöpft ist – oder warum der nachhaltig nichtinformierte Anteil 0,203188 der Gesamtbevölkerung umfasst.

Wie eine ansteckende Krankheit

Gerüchtesimulationen sind stark verwandt mit Modellen nichttödlicher ansteckender Krankheiten, bei denen die drei Gruppen aus ansteckbaren, angesteckten und geheilten Fällen bestehen. Doch es gibt einen grundlegenden Unterschied zwischen Gerüchten und Epidemien: Bei den Gerüchtemodellen ist nicht nur die »Krankheit« ansteckend, sondern auch ihre »Heilung«, da sowohl die Verbreitung als auch die Unterdrückung kommunikativ übertragbar ist.

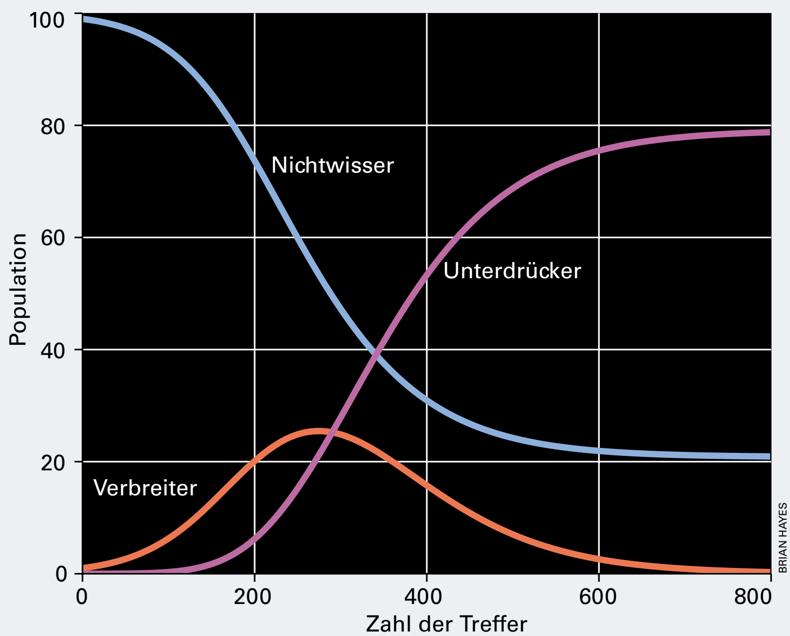
Ich war neugierig, das Gerüchtemodell in Aktion zu sehen, und schrieb daher ein kleines Computerprogramm. Dabei ging ich von einer Bevölkerung von 1000 Individuen aus, die jeweils Nichtwisser, Verbreiter oder Unterdrücker sein konnten. Am Anfang gab es nur einen Verbreiter, alle anderen waren Nichtwis- ▷

Wie entsteht ein Gerücht? Und wie verebbt es wieder?



Entstehen und Verebben eines Gerüchts sind hier vorgeführt. Von links nach rechts läuft die Zeit, jede horizontale Linie entspricht einem Mitglied der Population. Die Farben repräsentieren den Status der Person: Blau steht für *Nichtwisser*, die von dem Gerücht nichts gehört haben; Rot für *Verbreiter* des Gerüchts und Violett für *Unterdrücker*, die das Gerücht zwar kennen, aber kein Interesse daran haben, es weiterzugeben. Jede weiße vertikale Verbindung zeigt ein Treffen zwischen zwei zufällig ausgewählten Personen. Verbreiter geben das Gerücht an Nichtwisser weiter, die dadurch zu Verbreitern werden. Wenn sich zwei Verbreiter treffen, werden beide zu Unterdrückern. Ein Verbreiter, der einen Unterdrücker trifft, wird ebenfalls zum Unterdrücker. Die Verbreitung beginnt mit einem einzigen Verbreiter inmitten von Nichtwissern. Danach werden immer mehr zu Verbreitern, die schließlich alle zu Unterdrückern werden. Sieben der 30 Individuen der Gruppe hören niemals von dem Gerücht. (Alle Grafiken wurden von einem Programm erzeugt, das keinen der hier diskutierten Fehler enthält.)

Dynamik der Gerüchteverbreitung



Die Dynamik der Gerüchteverbreitung wird erkennbar, wenn man Mittelwerte aus 100 000 Programmläufen eines Modells bildet – in diesem Fall mit einer Gruppe von 100 Personen. Die Zahl der Verbreiter erreicht ein Maximum und sinkt später auf null, sodass die Gruppe schließlich nur noch aus Unterdrückern und Nichtwissern besteht. Eine grundlegende Vorhersage dieses Modells ist, dass die am Ende verbleibende Zahl von Nichtwissern etwa 0,203, also 20,3 Prozent der Gesamtgruppe umfasst.

Wiederhole
wähle ein zufälliges X aus der Menge der Verbreiter in der Bevölkerung;
wähle ein zufälliges Y aus der gesamten Bevölkerung;
wenn Y Nichtwisser ist,
dann wandle Y in einen Verbreiter um;
oder falls Y bereits ein Verbreiter ist,
dann wandle sowohl X als auch Y in einen Unterdrücker um;
oder falls Y bereits ein Unterdrücker ist,
dann wandle X in einen Unterdrücker um;
bis es keine Verbreiter mehr gibt.

▷ ser. Kern des Programms war die folgende, auf dem Daley-Kendall-Modell (bei der aufeinander treffende Verbreiter beide zu Unterdrückern werden) basierende Prozedur (siehe mein kleines Programm links unten).

Wenn es keine Verbreiter mehr gibt, kann sich nichts mehr ändern. An diesem Punkt endet das Programm und gibt Auskunft darüber, wie groß der Anteil der Bevölkerung ist, der nichts von dem Gerücht gehört hat. Dieser Anteil, Theta (griechisch: Θ) genannt, sollte 0,203188 betragen. Doch mein Programm errechnete, im Mittel bei über 1000 Durchläufen, Werte zwischen 0,28 und 0,29 – eine beträchtliche Diskrepanz.

An diesem Punkt möchte ich innehalten, um zu gestehen, dass ich Mist gebaut habe. Bevor Sie weiterlesen, könnten Sie einmal versuchen, die Bugs in meinem links stehenden Programm zu finden – oder sogar selbst ein Programm zu schreiben.

Die Programmierung von Computern lehrt – zumindest nach meiner eigenen Erfahrung – Bescheidenheit. Im Prinzip könnte die von mir beobachtete Diskrepanz auf einen Fehler in den Ergebnissen zurückführbar sein, die im Artikel publiziert sind. Dies war jedoch nicht meine erste Vermutung. Ich überprüfte meinen eigenen Programmcode – in der vollen Überzeugung, einige fahrlässige Fehler zu finden. Das Programm könnte einmal zu viel oder zu wenig die (Wiederholungs-)Schleife durchlaufen haben, dabei versagt haben, eine Variable auf den neuesten Stand zu bringen, oder einen Feldindex falsch berechnet haben. Das Problem, dachte ich, war nicht ein Tipp-, sondern ein Denkfehler.

Ich kannte eine Ungenauigkeit im Programm: Die Individuen X und Y wurden in einer Weise gewählt, dass beide dieselbe Person sein konnten – was das seltsame Szenario ergibt, dass jemand ein Gerücht an sich selbst weitergibt (»Pssst! Habe ich schon davon gehört?«). Als ich diesen Fehler beseitigt hatte, entdeckte ich einen neuen. Eine Variable namens *Verbreiterzähler* wurde bei jedem Durchlaufen der Schleife vergrößert oder verkleinert – je nach Ausgang des Treffens von X und Y. Wenn diese Variable den Wert null erreicht, endet das Programm. Nach jeder Verbreiter-Verbreiter-Interaktion zog ich die Zahl 2 von *Verbreiterzähler* ab – mit möglicherweise fatalen Folgen, falls X und Y identisch sind. Dies war ein schwer wiegender Fehler, der korrigiert werden musste. Doch als dies geschehen war, änderte sich der Wert von Θ immer noch nicht, er blieb bei 0,285.

Mir kam eine andere Idee. Belen und Pearce hatten besonders darauf hingewiesen, dass ihre Resultate nur dann gelten, wenn die Populationsgröße gegen unendlich tendiert. Vielleicht würde meine Abweichung verschwinden, wenn ich eine größere Ausgangspopulation wählte. Ich untersuchte unterschiedliche Größen und kam zu folgenden Ergebnissen (siehe Tabelle rechts oben).

Tippfehler und Denkfehler

Der Trend ist richtig: Es gab eine kleinere Zahl verbleibender Nichtwisser, wenn sich die Population vergrößerte. Doch die Kurve schien bei Werten über 1000 flach auszulaufen, und es schien unwahrscheinlich, dass Θ jemals 0,203 erreichen würde.

Dennoch fand ich es der Mühe wert, noch größere Populationen zu testen. Doch dafür benötigte ich ein schnelleres Programm. Ich schrieb eine neue, einfachere Version, in der ich auf das Feld für die Individuen verzichtete und lediglich die Zahl der Personen in jeder der drei Kategorien erfasste. Mit dieser Strategie konnte ich Populationen bis zu 100 Millionen Individuen testen. Der Wert für Θ indes blieb beharrlich bei 0,285.

Als ich die Verteilung der Θ-Werte bei einzelnen Programmläufen (im Gegensatz zum Mittelwert vieler Läufe) verfolgte, kam mir ein anderer Gedanke. Die meisten Ergebnisse für Θ lagen zwischen 0,25 und 0,35, doch es gab einige Ausreißer, in denen 99 Prozent der Population nie von dem Gerücht hörten. Ich ahnte, was da vor sich ging. Gesetzt den Fall, dass beim ersten Kontakt X das Gerücht an Y weitergibt und beim zweiten Kontakt durch Zufall wieder dasselbe X und Y ausgewählt werden – dann würde das Gerücht bereits hier stoppen, nachdem nur zwei Leute erreicht waren. Könnte es sein, dass bei Ausschluss dieser Ausreißer der Mittelwert auf 0,203 sinkt? Ich versuchte es, doch die Antwort war Nein.

Ich war aufgeschmissen. Beim Debugging erreichte ich den Punkt, an dem man beginnt, die Funktionsfähigkeit seines Zufallszahlengenerators anzuzweifeln – oder sogar des Compilers. Zufällig hatte auch Knuth bei seiner Arbeit an Tex einige Compilerbugs gefunden. Doch diese Spur führte ins Nichts.

Alles war meine Schuld

Mangels besserer Ideen wandte ich mich dem anderen Modell der Gerüchte-Verbreitung, dem von Maki-Thompson, zu. Wie oben beschrieben, unterscheidet es sich von dem Daley-Kendall-Modell darin, dass bei einer Begegnung zwischen zwei Verbreitern lediglich einer von ihnen zum Unterdrücker wird. Die Änderung meines Programms erfolgte im Handumdrehen. Als ich es laufen ließ, lieferte es als Ergebnis Θ = 0,203. Nun war ich nicht nur aufgeschmissen, sondern auch noch ziemlich durcheinander.

Mein Bekenntnis geriet zunehmend zum Charaktertest. Es gab einen Moment – er kam in der Dunkelheit der Nacht –, in dem ich zu glauben wagte, dass vielleicht *ich* am Ende richtig lag und der Rest der Welt eine Schraube locker hatte. Ich nahm mir noch einmal den ersten Absatz der Studie von Belen und Pearce vor. Ich erkannte, dass die Sache stimmig war, und konnte die Zahlen mit meinen in Einklang bringen – freilich allein durch die Annahme, dass die australischen Autoren alles auf den Kopf gestellt haben: Die Zahl 0,203, die sie als Ergebnis für das Daley-Kendall-Modell nannten, gehört in Wahrheit zu dem von Maki-Thompson. Und die Zahl 0,238, die sie als Schreibfehler bezeichneten, war genau das: ein Zahlendreher von 0,283 – was nahe genug an 0,285 lag, dem Θ-Wert, den ich ermittelt hatte für das Daley-Kendall-Modell …

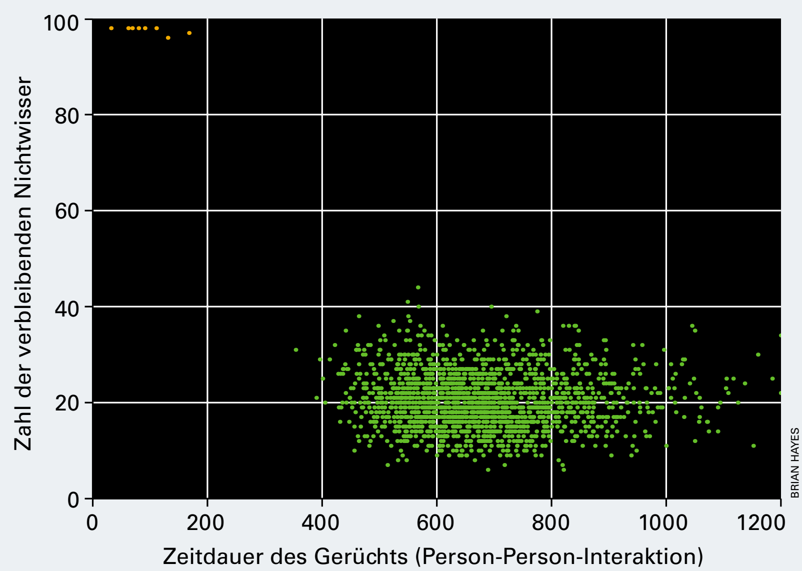
Am Morgen hatte sich der Anfall gelegt, doch es blieb ausweglos. Ich wusste, dass ich die Probleme vermutlich lösen könnte, wenn ich in die Bücherei ginge und mir die von Belen und Pearce zitierten Quellen ansähe. Doch das erschien mir zu unsportlich. Ich hätte versuchen können, die Korrektheit des einen oder anderen Ergebnisses zu beweisen. Doch wenn ich mir schon selbst nicht zutraute, eine korrektes Programm zu schreiben, wie konnte man mir dann vertrauen, dass ich einen korrekten Beweis vorlege?

Dann gab es noch die experimentelle Methode: Ich hätte tausend Freiwillige suchen, sie sorgfältig mit den Daley-Kendall-Regeln vertraut machen und inmitten von ihnen ein Gerücht ausstreuen können. Am Ende jedoch setzte ich auf eine weitere Computersimulation. Ich entschied mich, ein Programm zu schreiben, das ein reales Experiment so ähnlich wie möglich nachahmt und das alle Vorkommnisse im zu Grunde liegenden Modell ohne Vereinfachungen und Optimierungen reproduziert. Im Kopf hatte ich dabei eine Gruppe von Leuten, die wie Moleküle eines Gases umherirren, zufällig aufeinander stoßen und bei diesen Zufallskollisionen das Gerücht weitergeben. Ein solches System wollte ich simulieren.

Mein erstes Programm, das explizit jedes Mitglied der Gruppe berücksichtigte, kam dem Ziel schon recht nahe. Ich musste jedoch eine Veränderung zur Erhöhung der Recheneffizienz vornehmen: Da Interaktionen, in denen keiner der beiden Partner ein Ver- ▷

Bevölkerung	Θ
10	0,354
100	0,296
1000	0,286
10000	0,285
100000	0,285

Wie lange lebt ein Gerücht?



Die Lebensdauer eines Gerüchts und sein Reichweiten-Erfolg innerhalb der Gruppe sind nicht stark miteinander korreliert. Jeder Punkt entspricht einem einzelnen Lauf des Gerüchteprogramms. Die Punkte zeigen, wie lange das Gerücht lebt und wie viele Personen am Ende Nichtwisser bleiben. Der Anteil an Nichtwissern ist weit gehend unabhängig davon, ob das Gerücht nach 500 Programmläufen oder erst nach über 1000 abklingt. In der oberen linken Ecke finden sich einige Ausnahmen (orange gekennzeichnet), in denen Gerüchte bereits am Anfang verebben und nur sehr wenige Leute erreichen.

Begegnungen in der Gerüchteküche

	Nw	V	U
Nw	Nw ← Nw V ← V U ← U	Nw ← Nw – 1 V ← V + 1 U ← U	Nw ← Nw V ← V U ← U
V	Nw ← Nw – 1 V ← V + 1 U ← U	Nw ← Nw V ← V – 2 U ← U + 2	Nw ← Nw V ← V – 1 U ← U + 1
U	Nw ← Nw V ← V U ← U	Nw ← Nw V ← V – 1 U ← U + 1	Nw ← Nw V ← V U ← U

BRIAN HAYES

Alle im Gerüchtemodell möglichen Begegnungen sind in dieser Matrix zusammengefasst. Man kann damit auch die Wahrscheinlichkeit der Treffen bestimmen. Wenn Länge und Breite der Reihen und Spalten proportional zur Zahl der Mitglieder in den jeweiligen Untergruppen gewählt werden (Nichtwisser (Nw), Verbreiter (V) und Unterdrücker (U)), dann entspricht die zufällige Wahl von zwei Individuen der Wahl eines zufälligen Punkts im Diagramm. Je nachdem, in welchem der neun Kästen der gewählte Punkt liegt, entscheidet sich, was als Nächstes geschieht. Die drei Zuordnungsbedingungen innerhalb jeder Box (der Pfeil steht für »wird«) bewirken Veränderungen in der Größe der Untergruppen. Zwei der neun Kästen repräsentieren Nichtwisser-Verbreiter-Interaktionen; zwei weitere stehen für Verbreiter-Unterdrücker-Interaktionen; für Verbreiter-Verbreiter-Begegnungen gibt es nur einen einzigen Kasten.

▷ breiter war, die Gerüchteverbreitung nicht beeinflussen, schien es mir verschwenderisch, sie mitberechnen zu lassen. Ich vermied diese Verschwendung, indem ich immer einen Verbreiter als ersten Partner einer Begegnung wählte. Dies erschien mir ein völlig harmloser Eingriff zur Effizienzsteigerung zu sein.

Doch dann nahm ich den Eingiff wieder zurück. In einer dritten Programmvariante wählte ich zufällig zwei Individuen aus der gesamten Gruppe aus, prüfte, ob es sich nicht um ein und dieselbe Person handelte, und ließ sie dann gemäß den Daley-Kendall-Regeln interagieren. Es schien, so dachte ich, ein sinnloses Unterfangen zu sein, und dabei würde sicherlich genau das Gleiche herauskommen wie bei den Läufen der anderen Programme, nur nach mehr Rechenzeit. Zu meinem Erstaunen errechnete das neue Programmen einen Θ-Wert von 0,203.Wenn Sie bereits ahnen, an welchem Punkt mein Verstand auf Abwege geriet, möchte ich Ihnen gratulieren.

Meine eigene Erleuchtung setzte ein, als ich mir alle neun Begegnungsmöglichkeiten zwischen Nichtwissern, Verbreitern und Unterdrückern klar machte.

Aus der Matrix (links) lässt sich fast alles erkennen. Sie offenbart die gesamte Struktur und Funktion des Modells. Wenn man die Breiten und Längen der Spalten und Reihen so wählt, dass sie proportional zur Größe der drei Untergruppen sind, gibt die Fläche jeder der neun Kästen die Wahrscheinlichkeit der jeweiligen Paar-Begegnung wieder. Wählt man zwei Personen zufällig, so ist dies äquivalent dazu, einen beliebigen Punkt in dem Diagramm zu wählen. Das Ergebnis der Interaktion ist dann davon abhängig, in welchem der neun Kästen der gewählte Punkt liegt. (Ich übergehe wieder das – eine kleine Korrektur erfordernde – Problem, das entsteht, wenn eine Person ein Gerücht sich selbst erzählt.)

Zwischen Verbreitern und Nichtwissern

Um zu verstehen, wo mein Fehler lag, reicht es, den einfachsten Fall zu betrachten. Darin sind alle drei Untergruppen gleich groß und alle neun Begegnungsvarianten haben dieselbe Wahrscheinlichkeit – nämlich 1/9. Die Kästen an den vier Ecken des Diagramms entsprechen Situationen, in denen keiner der Beteiligten ein Verbreiter ist und die den Gerüchtestatus somit nicht beeinflussen. Zwei weitere Kästen entsprechen Nichtwisser-Verbreiter-Treffen, die daher eine Wahrscheinlichkeit von 2/9 aufweisen. Zwei weitere Kästen stehen für Verbreiter-Unterdrücker-Begegnungen, die ebenfalls mit einer Wahrscheinlichkeit von 2/9 eintreten. Es gibt jedoch nur eine einzige Box, in der Verbreiter auf Verbreiter stoßen, was daher mit einer Wahrscheinlichkeit von 1/9 geschieht. Das Entscheidende ist, dass Nichtwisser-Verbreiter-Begegnungen und Verbreiter-Unterdrücker-Begegnungen jeweils doppelt so wahrscheinlich sind wie Verbreiter-Verbreiter-Treffen.

Betrachten wir nun, was in meinem ersten Programm für das Daley-Kendall-Modell geschah. Dadurch, dass ich immer einen Verbreiter als Ausgangspunkt wählte, beschränkte ich mich auf Begegnungen in der mittleren Reihe des Diagramms, und diese drei Kästen wurden mit jeweils gleicher Wahrscheinlichkeit gewählt. Dadurch ergaben sich Verbreiter-Verbreiter-Treffen doppelt so häufig, wie es ihnen zustand, und die Gerüchteküche hörte vorzeitig auf zu brodeln.

Aus der Sicht der Wahrscheinlichkeitstheorie handelt es sich um einen grundlegenden Irrtum; die Fälle wurden nicht richtig gezählt. Ein professioneller Programmierer würde einen anderen Fehlergrund nennen. Ich hatte

gegen das altbewährte Motto verstoßen: »Beginne nicht mit der Optimierung eines Programms, ehe du es zu Ende geschrieben hast.«

Doch auch nachdem ich diesen Fehler erkannt hatte, blieb ich weiterhin verwirrt. Ausgerechnet der Algorithmus, von dem ich jetzt wusste, dass er für das Daley-Kendall-Modell nicht richtig war, lieferte das korrekte Ergebnis für das Maki-Thompson-Modell. Um das Rätsel zu lösen, machte ich mich nun schließlich auf den Weg in die Bibliothek. Ich wollte herauszufinden, was die Originalautoren tatsächlich geschrieben hatten.

Daley und Kendall sind Daryl J. Daley, derzeit an der Australian National University, und David G. Kendall, ein renommierter Statistiker und Wahrscheinlichkeitstheoretiker an der Cambridge University in England. Ihr 1965 publizierter Artikel ist ein Meisterstück klarer Darstellung, der mir all mein Stochern im Dunkeln erspart hätte – und daher bin ich froh, ihn nicht früher gefunden zu haben. Die Wahrscheinlichkeitsberechnungen erfolgten sehr sorgfältig (es findet sich ein Faktor 1/2 bei der Verbreiter-Verbreiter-Interaktion). Weiterhin gibt es eine Herleitung der mysteriös präzisen Zahl 0,203188.

Fehler führen zur Selbsterkenntnis

Die sechsstellige Genauigkeit stammt nicht aus der Simulation tatsächlicher Fälle – wie jene, die ich selbst mit einer anderen, auf kontinuierlichen Differenzialgleichungen basierenden Version des Modells entwickelt hatte. Die Zahl Θ ist vielmehr eine Lösung der Gleichung

$$\Theta\, e^{2(1-\Theta)} = 1.$$

(Diese Gleichung erinnert an die Lambert-W-Funktion, $W e^{w}$.) Maki und Thompson sind Daniel P. Maki und Maynard Thompson von der Indiana University. Gerüchteverbreitung haben sie 1973 in ihrem Lehrbuch »Mathematical Models and Applications« abgehandelt. Sie beschreiben den Gerüchte-Transferprozess als Telefongespräch, wobei sie sich auf Anrufe von Verbreitern beschränken. Wegen dieser Asymmetrie fließt nur die mittlere Reihe des Diagramms links oben in ihre Berechnungen ein – und mein erstes Programm war in der Tat eine korrekte Implementierung dieses Modells. (So hab ich wenigstens *etwas* richtig hinbekommen.) Es ist mehr oder minder Zufall, dass Maki und Thompson den gleichen Wert für Θ ermittelten wie Daley und Kendall: Ihre Verbreiter-Verbreiter-Interaktionen sind doppelt so wahrscheinlich, wirken sich aber nur halb so stark aus.

Der Artikel von Belen und Pearce, der mich zu diesen Ausführungen veranlasst hat, verdient ebenfalls weitere Kommentare: Der Ausdruck »generelle Anfangsbedingungen« im Titel bezieht sich auf Gerüchte, die nicht von einem einzigen, sondern von einer Vielzahl von Verbreitern gestreut werden. Man ist versucht zu glauben, dass das Gerücht bei einer hinreichenden Zahl von Verbreitern die gesamte Gesellschaft erreichen muss. Belen und Pearce zeigen jedoch, dass dies nicht so ist. Sie ermittelten den Anteil der ursprünglichen Nichtwisser, die auch nach Verbreitung des Gerüchts Nichtwisser geblieben sind, und stellten fest, dass dieser Anteil mit der Anzahl der anfänglichen Gerüchteverbreiter ansteigt. Das Maximum beträgt 1/e, also rund 0,368. Mit anderen Worten: Je mehr Leute die Nachricht verbreiten, desto mehr Leute gibt es, die sie nicht mitbekommen. Der Grund ist einfach: Die hohe Zahl von Verbreitern führt dazu, dass diese sich schnell gegenseitig unterdrücken.

Auf dem Gebiet der mathematischen Modellierung von Gerüchten sind inzwischen zahlreiche Studien erschienen – mit unterschiedlichen Ansätzen. Ich hatte bislang noch keine Gelegenheit, in diesen Modellen etwaigen Fehlern nachzuspüren.

Fehler führen zur Selbsterkenntnis – was nicht immer willkommen ist. Doch für die meisten von uns besteht die einzige Möglichkeit, uns nie zu irren, darin, nie irgendetwas zu versuchen. Henry Petroski hat sehr eloquent über die notwendige Rolle von Fehlern und vom Scheitern geschrieben. In seinen Worten ist das Fallen ein Teil des Wachsens. Wenn wir Fehler machen, scheint es zweckdienlich, sie an die Öffentlichkeit zu bringen und ihre Ursachen zu diskutieren. Sie wirken dann weniger demütigend.

Allerdings nur ein wenig. Für solche Bekenntnisse gibt es keinen Sündenerlass. Statt »Gehe und irre dich nicht mehr« zitiert Knuth den Ratschlag des dänischen Aphorismendichters Piet Hein: »Irre dich immer wieder, aber jedes Mal ein bisschen weniger.« Ich ziehe ein anderes Motto vor. Es stammt von Samuel Beckett: »Scheitere wieder. Scheitere besser.« ◁

AUTOR UND LITERATURHINWEISE

Brian Hayes ist Mathematiker und Senior Writer des American Scientist (bhayes@amsci.org).

Rumours with general initial conditions. Von Selma Belen und C. E. M. Pearce in: Anziam (The Australia and New Zealand Industrial and Applied Mathematics Journal), Bd. 45, S. 393, 2004

The errors of Tex. Software-practice & experience. Von Donald E. Knuth. Bd. 19, S. 607, 1989. Mit Ergänzungen nachgedruckt in: Literate Programming, Stanford, Kalifornien: Center for the Study of Language and Information, S. 243, 1992

Mathematical models and applications, with emphasis on the social, life, and management sciences. Von Daniel P. Maki und Maynard Thompson. Englewood Cliffs, New Jersey, Prentice-Hall, 1973

Stochastic rumours. Von D. J. Daley und D. G. Kendall in: Journal of the Institute of Mathematics and Its Applications, Bd. 1, S. 42, 1965

Weblinks zum Thema finden Sie bei www.spektrum.de unter »Inhaltsverzeichnis«.