Vorbemerkungen zum folgenden Aufsatz von Kay Rottmann

HEINZ ALTHOFF, BIELEFELD

Zu den Themen der Wiederholung der Stochastik für die Abiturprüfung meines LK 2001 gehörte auch das *Geburtstagsproblem*. In diesem Rahmen bat ich K. R., seinen Mitschülern den Inhalt des Aufsatzes 'Ein Geburtstagsproblem" von G. SCHRAGE (aus Stochastik in der Schule, Heft 2/1992) darzustellen, in dem es um die Wahrscheinlichkeit für das Zusammentreffen dreier Geburtstage bei 15 Personen geht. K. R. überraschte mich dann damit, dass er im zweiten Teil seines Vortrags das Problem auf "mindestens k gemeinsame Geburtstage bei n Personen" erweiterte. Ich ermunterte ihn daraufhin, den folgenden Aufsatz für unsere Zeitschrift auszuarbeiten. Lehrerinnen und Lehrer sollten den Aufsatz vor allem als anregendes Beispiel dafür betrachten, hinreichend begabte Schüler zu motivieren, eine (nicht zu umfangreiche) mathematische Abhandlung anderen verständlich darzustellen, ja sie vielleicht sogar (wie hier geschehen) zu einer eigenen wissenschaftlichen Auseinandersetzung mit einem mathematischen Problem zu "verführen". Kann der Aufsatz von K. R. eventuell eine Kettenreaktion für so etwas auslösen?

Lösung des Geburtstagsproblems für den Fall, dass k von n Personen am gleichen Tag Geburtstag haben

Kay Rottmann, Steinhagen

***Zusammenfassung:** In „Stochastik in der Schule 2/1992" beschreibt Georg Schrage, wie man die Wahrscheinlichkeit dafür berechnet, dass genau drei Personen aus einer Gruppe von n Personen am gleichen Tag Geburtstag feiern. Die dort entwickelte Rekursionsformel soll hier nun so erweitert werden, dass man eine Formel erhält, mit der die Wahrscheinlichkeit berechnet werden kann, dass k Personen am gleichen Tag Geburtstag haben. Des weiteren ist es mit der Formel dann sogar möglich, die Wahrscheinlichkeiten für weitere Konstellationen von Geburtstagen zu berechnen.*

Zunächst sei hier noch einmal die Formel von Georg Schrage hergeleitet. Er benutzt dafür folgenden Ansatz:

W(n, e, z) sei die Wahrscheinlichkeit, dass an z Tagen jeweils genau zwei Personen Geburtstag haben und an e Tagen eine Person; $n = e + 2 \cdot z$ gibt dabei die Gesamtzahl der betrachteten Personen an.

Für *W(n, e, z)* gelten die folgenden Randbedingungen:

(1) $W(1, e, z) = W(1, 1, 0) = 1$

(2) $W(n, n, 0) = W(n-1, n-1, 0) \cdot \dfrac{365-(e-1)}{365}$

(3) $W(n, 0, z) = \dfrac{W(n-1, 1, z-1)}{365}$

Zusätzlich gibt es noch die rekursive Beziehung:

(4) $$W(n, \mathrm{e}, \mathrm{z}) = W(n-1,\ e-1, z) \cdot \frac{366-e-z}{365} + W(n-1, e+1, z-1) \cdot \frac{e+1}{365}$$

Da n in jedem Rekursionsschritt um eins erniedrigt wird, ist das Verfahren zur Berechnung von *W(n, e, z)* endlich, und die Rekursion endet stets bei der Wahrscheinlichkeit *W(1, 1, 0)=1.*

Zur Erläuterung der Rekursion: liegt bei n Personen eine Geburtstagsverteilung vom Typ *(n, e, z)* mit *z>0* und *e>0* vor, so lag bei der Verteilung der ersten *n-1* Geburtstage entweder eine Geburtstagsverteilung vom Typ *(n-1, e-1, z)* vor, dann fällt der n-te Geburtstag auf einen der $365-(e-1)-z = 366-e-z$ noch freien Tage, oder es lag eine Geburtstagsverteilung vom Typ $(n-1, e+1, z-1)$ vor, dann fällt der n-te Geburtstag auf einen der $e+1$ Tage, an denen bereits eine Person Geburtstag hat. Wenn man ausgehend von diesen beiden Möglichkeiten auf die Verteilung *(n, e, z)* kommen möchte, so ergibt sich für die erste Möglichkeit die Wahrscheinlichkeit $W(n-1, e-1, z) \cdot \dfrac{366-e-z}{365}$ und für die zweite die Wahrscheinlichkeit $W(n-1, e+1, z-1) \cdot \dfrac{e+1}{365}$.

Daraus erhält man die in (4) stehende rekursive Beziehung.

Nach der Festlegung, dass *W(n, e, z)*=0 für z<0 oder e<0 gelten soll, können die Bedingungen zusammengefasst werden zu:

(5) $W(n, e, z) =$

$$\begin{Bmatrix} 0, \text{falls } z < 0 \text{ oder } e < 0 \\ 1, \text{falls } n = 1 \text{ und } e = 1 \\ \dfrac{W(n-1, e-1, z)(366-z-e)}{365} \\ \dfrac{+W(n-1, e+1, z-1)(e+1)}{365} \end{Bmatrix}$$

Für die Wahrscheinlichkeit von E_3 - mindestens drei Personen haben am gleichen Tag Geburtstag - gilt unter Zuhilfenahme der Formel (5):

$$P(E_3) = 1 - \sum_{i=0}^{\left[\frac{n}{2}\right]} W(n, n-2i, i).$$

Nun soll gezeigt werden, wie die Formel so erweitert werden kann, dass sie auf jedes beliebige Geburtstagsproblem anwendbar ist.

$W(n, a_1, a_2, \ldots, a_k)$ sei die Wahrscheinlichkeit dafür, dass von n Personen an a_1 Tagen genau eine Geburtstag hat, an a_2 Tagen genau zwei Personen usw.

Auch hier bleibt weiterhin der Zusammenhang:

$$n=\sum_{i=1}^{k} i\cdot a_i .$$

Für $W(n, a_1, a_2, \ldots, a_k)$ gelten nun folgende Randbedingungen:

(6) $W(1,1,0,\ldots,0)=1$

(7) $W(n, a_1,\ldots,a_i,\ldots,a_k)=0 \quad \text{für } a_i<0$

Zusätzlich gibt es wieder folgende rekursive Beziehung:

(8) $W(n, a_1,\ldots,a_k)=$

$$W(n-1,a_1-1,\ldots,a_k)\cdot\frac{365-(a_1-1+\sum_{i=2}^{k}a_i)}{365}$$
$$+\sum_{i=2}^{k}W(n-1,a_1,\ldots,a_i-1,\ldots,a_k)\cdot\frac{a_{i-1}+1}{365}$$
$$\text{mit } n=\sum_{i=1}^{k} i\cdot a_i$$

Erläuterung der Rekursion: Wieder wird die Wahrscheinlichkeit einer Situation als Resultat der vorangegangenen Situationen betrachtet. Wenn also die Wahrscheinlichkeit einer Geburtstagsverteilung vom Typ $(n, a_1,\ldots,a_k)$ mit $a_1\geq 0$ und $a_2\geq 0\ldots$ und $a_k\geq 0$ ermittelt werden soll, bedient man sich der Tatsache, dass diese Geburtstagsverteilung aus einem der folgenden Typen hervorging:

(A) die n-te Person hat an einem Tag Geburtstag, an dem noch kein anderer Geburtstag hat. Also lag zuvor eine Geburtstagsverteilung vom Typ $(n-1, a_1-1, a_2,\ldots,a_k)$ vor und von dieser ausgehend gibt es dann $365-(a_1-1+\sum_{i=2}^{k}a_i)$ Möglichkeiten, um zu der Endsituation zu kommen.

(B) die n-te Person hat an einem Tag Geburtstag, an dem mindestens eine andere Person Geburtstag hat. Also lag zuvor eine Geburtstagsverteilung vom Typ $(n-1, a_1,\ldots,a_{i-1}+1, a_i-1,\ldots,a_k)$ vor und von dieser gibt es dann $a_{i-1}+1$ Möglichkeiten, um zu der Endsituation zu kommen.

Nach Division der Anzahl „günstiger" Tage (um zu der Endsituation zu kommen) durch die Anzahl aller möglichen Tage (365) liegen die Wahrscheinlichkeiten vor, um von einer Geburtstagsverteilung für n-1 Personen zur Geburtstagsverteilung für n Personen zu kommen.
Wie oben gilt wieder, dass durch das Vermindern von n in jedem Rekursionsschritt das Verfahren endlich ist und damit immer eine Lösung liefert.

Also gilt:

(9) $W(n, a_1,\ldots,a_i,\ldots,a_k)=$

$$\left\{\begin{array}{c} 0,\text{falls } a_i<0 \text{ für } 1\leq i\leq k \\ 1,\text{falls } n=1, a_1=1, a_i=0 \text{ für } 2\leq i\leq k \\ \frac{1}{365}\cdot \\ \left(\begin{array}{l} W(n-1,a_1-1,a_2,\ldots,a_k)\cdot \\ (365-(a_1-1+\sum_{i=2}^{k}a_i))+ \\ \sum_{i=2}^{k} \begin{array}{l} W(n-1,a_1,\ldots,a_{i-1}+1,a_i-1,\ldots,a_k)\cdot \\ (a_{i-1}+1) \end{array} \end{array}\right) \\ ,\text{sonst} \end{array}\right\}$$

Mit der Formel (9) existiert nun ein Mittel, das einem erlaubt, die Wahrscheinlichkeit für allgemeine Konstellationen von Geburtstagen zu berechnen. So kann durch Festlegen der Anzahl Tage, an denen eine, zwei, ..., k Personen Geburtstag haben, nach Einsetzen in die Formel, natürlich auch hier wieder unter der Bedingung: $n=\sum_{i=1}^{k} i\cdot a_i$, die Wahrscheinlichkeit für solche Konstellationen berechnet werden. Um jedoch die Wahrscheinlichkeit für das Ereignis E_{k+1}, mindestens $k+1$ Personen haben am selben Tag Geburtstag, zu berechnen, geht man zweckmäßig wieder über das Gegenereignis: es gibt keinen Tag, an dem k+1 oder mehr Personen gemeinsam Geburtstag feiern.
Werden nun die Wahrscheinlichkeiten all dieser Konstellationen addiert, dann hat man die Wahrscheinlichkeit für das Gegenereignis:

$$(10)\ P(E_{k+1})=1-\sum_{\mathbb{N}\ni a_i:\sum_{i=1}^{k} i\cdot a_i=n} W(n,a_1,a_2,\ldots,a_k)$$

mit:

(6) $W(1,1,0,\ldots,0)=1$

(7) $W(n, a_1,\ldots,a_i,\ldots,a_k)=0 \quad \text{für } a_i<0$.

Als schwierig erweist sich die Umsetzung der Ergebnisse in die Praxis. Es ist zwar relativ leicht, ein Computerprogramm für das Problem zu schreiben, allerdings stößt ein solches schnell an seine Leistungsgrenzen. Die Rekursionen haben zwar höchstens die Tiefe n, jedoch stellt nicht die „Tiefe" der Rekursionen das Problem dar, sondern die Anzahl der zu berechnenden Wahrscheinlichkeiten, die durch einen Rekursionsschritt hinzukommen. Diese Anzahl ist gleich der Summe der Tage, an denen mindestens eine Person Geburtstag hat, also die Anzahl a_i, für die gilt: $a_i > 0$. Dieser Zusammenhang führt auf ein exponentielles Wachstum der zu berechnenden Wahrscheinlichkeiten und macht damit eine Berechnung leicht zu einer Aufgabe die selbst mit den schnellsten Computern nicht effektiv lösbar ist.

Im Folgenden soll nun noch ein Beispiel die Anwendung der oben beschriebenen Formel zeigen:

Gesucht ist die Wahrscheinlichkeit, dass aus einer Gruppe von 5 Menschen genau 3 am gleichen Tag Geburtstag haben und die verbleibenden 2 an unterschiedlichen Tagen.

Es ist also die folgende Wahrscheinlichkeit gesucht: $W(5;2;0;1)$

$$= \frac{1}{365}\left(W(4;1;0;1)\cdot 363 + W(4;2;1;0)\right)$$

$$= \left(\frac{1}{365}\right)^2 \begin{pmatrix} 363\cdot 364\cdot W(3;0;0;1) \\ +2\cdot 363\cdot W(3;1;1;0) \\ +3W(3;3;0;0) \end{pmatrix}$$

$$= \left(\frac{1}{365}\right)^3 \begin{pmatrix} 363\cdot 364\cdot W(2;0;1;0) \\ +2\cdot 363\cdot \begin{pmatrix} 364\cdot W(2;0;1;0) \\ +W(2;2;0;0)\cdot 2 \end{pmatrix} \\ +2\cdot 3\cdot W(2;2;0;0) \end{pmatrix}$$

$$= \left(\frac{1}{365}\right)^4 (7\cdot 364\cdot 363 + 6\cdot 364)$$

$$\approx 5.2\cdot 10^{-5}$$

Da dieses Verfahren bereits bei leicht größeren Zahlen zu großem Rechenaufwand führt, dürfte eine Simulation in der Praxis deutlich effektiver sein.

KAY ROTTMANN
Queller Str. 15
33803 Steinhagen
kayro@gmx.de