Gemeinsame Geburtstage

FRIEDRICH BARTH UND RUDOLF HALLER, MÜNCHEN

Zusammenfassung: *In der Schulbuchliteratur wird am Problem des Doppelgeburtstags unter dem Stichwort »Geburtstagsparadoxon« gerne gezeigt, wie unreflektierte Intuition in die Irre führen kann. Die Vaterschaft für dieses Geburtstagsproblem ist unklar und wird gelegentlich RICHARD VON MISES (1939) zugesprochen.*

Im Folgenden wird einerseits an Hand des Doppelgeburtstagsproblems untersucht, wie die Formulierung einer Aufgabe das Ergebnis beeinflussen kann, andererseits wird die Fragestellung des Doppelgeburtstags auf Mehrfachgeburtstage erweitert. Dabei erkennt man, wie eine scheinbar geringfügige Veränderung zu einem erheblichen Mehraufwand bei der Lösung führen kann. Damit soll dem Lehrer ein Hintergrundwissen vermittelt werden, auch um auf nahe liegende Schülerfragen eingehen zu können.

1 Doppelgeburtstage

Am Beispiel der Frage nach Doppelgeburtstagen zeigen wir, wie ähnlich klingende Formulierungen eines Problems zu verschiedenen Lösungen führen können.

Problem 1. Wie wahrscheinlich ist es, dass unter *n* Personen mindestens zwei Personen am gleichen Kalendertag Geburtstag haben?
Problem 2. Wie wahrscheinlich ist es, dass unter *n* Personen genau zwei Personen am gleichen Kalendertag Geburtstag haben, die anderen an paarweise verschiedenen Kalendertagen?
Problem 3. Wie wahrscheinlich ist es, dass unter *n* Personen mindestens eine Person ist, die mit mir am selben Kalendertag Geburtstag hat?
Problem 4. Wie viele Personen muss man im Mittel nach ihrem Geburtstag fragen, bis man einen Doppelgeburtstag erhält?[1]
Problem 5. Wie groß muss eine Gruppe von Personen sein, damit im Mittel mindestens ein Doppelgeburtstag auftritt?

Wie üblich setzen wir voraus, dass die Geburtstage gleichmäßig über 365 Tage verteilt sind, wodurch Schaltjahre ausgeschlossen werden. Auch Mehrlingsgeburten wie Zwillinge, Drillinge usw. werden ausgeschlossen, weil in einer Bevölkerung mit vielen Mehrlingsgeburten die Wahrscheinlichkeiten für Mehrfachgeburtstage anwachsen.

Zu Problem 1. Dieses auch als *Geburtstagsparadoxon* benannte Problem ist das bekannteste Geburtstagsproblem. Die überraschende Antwort auf die Frage besagt, dass schon bei 23 Personen die Wahrscheinlichkeit für einen Doppelgeburtstag 50,7 % ist, d. h., dass es bereits ab 23 Personen günstig ist, darauf zu wetten, dass mindestens zwei von ihnen gemeinsam Geburtstag feiern können. Ab $n = 47$ ist die Wahrscheinlichkeit größer als 95 %, nämlich 95,5 %. Ab dieser Anzahl von Personen ist es also ein signifikantes Ereignis, wenn es keinen Kalendertag gibt, an dem mindestens zwei Personen Geburtstag feiern. Ab $n = 366$ tritt ein Doppelgeburtstag sicher auf. (Barth/Haller 1998, S. 97 f.).

Zu Problem 2. Es gibt 365 Möglichkeiten für den Tag des Doppelgeburtstags. Die zwei Personen lassen sich auf $\binom{n}{2}$ Arten auswählen. Die restlichen $n - 2$ Personen werden einzeln auf die restlichen 364 Tage verteilt. Dafür gibt es $364 \cdot 363 \cdot \ldots \cdot (365 - (n - 2))$ Möglichkeiten. Man hat also insgesamt

$$365 \cdot \binom{n}{2} \cdot 364 \cdot 363 \cdot \ldots \cdot (367 - n) = \binom{n}{2} \frac{365!}{(366 - n)!}$$

günstige Fälle, woraus man für die gesuchte Wahrscheinlichkeit $\binom{n}{2} \frac{365!}{365^n\,(366 - n)!}$ erhält.

Diese Wahrscheinlichkeit wächst echt monoton mit n bis $n = 28$, wo sie den Wert 38,64… % erreicht, und fällt dann wieder echt monoton auf 0 bei $n = 367$. Ab 367 hat sie immer den Wert 0, weil dann mindestens ein Dreifachgeburtstag oder zwei Doppelgeburtstage auftreten.

Es gibt also kein n, bei dem es sich lohnt, auf genau einen Doppelgeburtstag zu wetten.

Zu Problem 3. Wir betrachten das Gegenereignis $\overline{B}$ = „Keine der n Personen hat mit mir am selben Kalendertag Geburtstag“. Da für jede dieser n Personen die restlichen 364 Tage zur Verfügung stehen, erhält man

$$P(B) = 1 - \frac{364^n}{365^n}.$$

$P(B)$ wächst echt monoton mit n von $\frac{1}{365}$ bis 1. Es gibt keine Beschränkung für die Gruppengröße n.

Für $n = 23$ ergibt sich 6,1 %, also ein erheblich kleinerer Wert als oben. Soll es vorteilhaft sein, auf das Eintreten des Ereignisses B zu setzen, dann muss die Gruppe sehr viel größer sein: Aus

$P(B) = 1 - \frac{364^n}{365^n} \geq 50\ \%$ erhält man

$$n \geq \frac{\ln 0{,}5}{\ln \frac{364}{365}} = 252{,}6\ldots, \text{ also } n_{\min} = 253.$$

Zu Problem 4. Sei A_k = „Die k-te befragte Person ist die erste, deren Geburtstag mit einem Geburtstag der $k - 1$ zuvor befragten Personen übereinstimmt“. Zur Berechnung der Wahrscheinlichkeit $P(A_k)$ überlegen wir:

Die Wahrscheinlichkeit, dass der Geburtstag der zweiten Person von dem der ersten befragten Person verschieden ist, ist $\frac{364}{365}$.

Die Wahrscheinlichkeit, dass der Geburtstag der dritten Person von denen der ersten zwei befragten Personen verschieden ist, ist $\frac{363}{365}$.

Für $k \leq 366$ ist die Wahrscheinlichkeit, dass der Geburtstag der $(k - 1)$-ten Person von allen Geburtstagen der vorhergehenden $k - 2$ befragten Personen verschieden ist, gleich $\frac{365 - (k - 2)}{365}$.

Die Wahrscheinlichkeit, dass der Geburtstag der k-ten Person mit dem Geburtstag einer der vorhergehenden $k - 1$ befragten Personen übereinstimmt, ist $\frac{k - 1}{365}$.

Also ist für $k \leq 366$

$$P(A_k) = \frac{364}{365} \cdot \frac{363}{365} \cdot \ldots \frac{365 - (k - 2)}{365} \cdot \frac{k - 1}{365}.$$

Für $k \geq 367$ ist $P(A_k) = 0$, weil ja spätestens bei der 366. Befragung ein Doppelgeburtstag auftreten muss.

Für den Erwartungswert $E(Z)$ der Zufallsgröße Z = „Anzahl der nötigen Befragungen bis zum ersten Doppelgeburtstag“ erhält man demzufolge

$$\sum_{k=2}^{366} k \cdot P(A_k) = \sum_{k=2}^{366} \left(\frac{k(k - 1)}{365} \cdot \prod_{m=3}^{k} \frac{365 - (m - 2)}{365} \right)$$

Der Computer liefert $E(Z) = 24{,}6\ldots$ Man muss also im Mittel etwa 25 Personen befragen, bis man auf einen Doppelgeburtstag stößt.

Zu Problem 5. Richard Edler von Mises (1883 bis 1953) hat 1939 gezeigt, dass eine Gruppe mindestens 29 Personen umfassen muss, damit man im Mittel einen Tag mit genau einem Doppelgeburtstag bei ihr erwarten kann. Zur Begründung verweisen wir auf unseren Aufsatz Barth/Haller 2012.

Die vorstehenden Ergebnisse zeigen (wieder einmal), wie genau man eine Fragestellung formulieren und lesen muss. Geringfügige Unterschiede in der Fragestellung führen zu unterschiedlichen Ergebnissen, wie der Vergleich zeigt. Dies sollte auch zur Vorsicht bei Interpretationen mahnen!

2 Dreifachgeburtstage

Die naheliegende Verallgemeinerung von Problem 1, wie wahrscheinlich es ist, dass unter n Personen mindestens drei Personen am gleichen Kalendertag Geburtstag haben, ist sehr viel schwieriger zu beantworten. RICHARD VON MISES nimmt diese Frage zum Anlass für seine Untersuchungen über Besetzungszahlen (Mises 1939, Barth/Haller 2012). Mit Hilfe dieser Besetzungszahlen lässt sich dieses Problem und auch das k-facher Geburtstage lösen.

Die Besetzungszahl N gibt die Anzahl der Möglichkeiten dafür an, dass n Elemente auf z Zellen so verteilt werden, dass genau a_i Zellen mit genau i Elementen besetzt sind. Man beachte, dass es bei dieser Fragestellung keine Rolle spielt, *welche* der z Zellen mit jeweils genau i Elementen besetzt sind, sondern nur, *wie viele* Zellen mit genau i Elementen besetzt sind. Außerdem spielt die Reihenfolge der Elemente in einer Zelle keine Rolle. Für N gilt:

$$N(n; z \parallel a_0 \mid a_1 \mid \ldots \mid a_n)$$

$$= \frac{n!}{0!^{a_0} \cdot 1!^{a_1} \cdot 2!^{a_2} \cdot 3!^{a_3} \cdot \ldots \cdot n!^{a_n}} \cdot \frac{z!}{a_0! a_1! a_2! a_3! \cdot \ldots \cdot a_n!}$$

$$= n! \cdot z! \cdot \left(\prod_{i=0}^{n} a_i! \cdot i!^{a_i} \right)^{-1}$$

$$\text{mit} \quad \sum_{i=0}^{n} a_i = z \quad \text{und} \quad \sum_{i=0}^{n} a_i \cdot i = n$$

Die Wahrscheinlichkeit einer Besetzung erhält man, wenn man die Besetzungszahl N durch z^n dividiert.

Bemerkung: Liegen höchstens k Elemente in einer Zelle, dann ist $a_{k+1} = \ldots = a_n = 0$. Damit kann man $N(n; z \parallel a_0 \mid a_1 \mid a_2 \mid \ldots \mid a_k \mid 0 \mid \ldots \mid 0)$ verkürzt als $N(n; z \parallel a_0 \mid a_1 \mid a_2 \mid \ldots \mid a_k)$ schreiben. Entsprechend kann man in der Formel dann Faktoren weglassen, die den Wert 1 haben, wie $a_i! \cdot i!^{a_i}$ für $k + 1 \le i \le n$.

Löst man sich von der Geburtstagseinkleidung, dann handelt es sich um das folgende Problem:

> Gesucht ist die Wahrscheinlichkeit des Ereignisses A, dass n Elemente per Zufall auf z Zellen so verteilt werden, dass in mindestens einer Zelle mindestens drei Elemente zu liegen kommen.

Wir betrachten das Gegenereignis $\overline{A}$ = „In jeder Zelle liegen höchstens zwei Elemente". Das ist nur möglich für $n \le 2z$.

Zur Abzählung der für $\overline{A}$ günstigen Fälle gehen wir schrittweise vor und zerlegen $\overline{A}$ in disjunkte Ereignisse, deren Besetzungszahlen wir mit der von MISES'schen Formel ermitteln. In einfachen Fällen lassen sie sich aber auch ohne diese Formel bestimmen. Das Vorgehen hierzu zeigen wir weiter unten an einem konkreten Beispiel mit den kleinen Werten $n = 5$ und $z = 7$. Das dortige Vorgehen lässt sich zwar ohne weiteres verallgemeinern, was jedoch zu großem Aufwand führt. Die Besetzungszahlen sind aber geradezu ein Beispiel für das Vorgehen der Mathematik: Man entwickelt eine Formel oder einen Kalkül, um so zu vermeiden, in jedem Einzelfall erneut Überlegungen anstellen zu müssen.

1. Schritt: Genau $n = a_1$ Zellen sind einfach belegt und keine Zelle mehrfach.

$$N(n; z \parallel z - n \mid n) = \frac{z!}{(z-n)!} \quad \text{für} \quad n \le z\,.$$

2. Schritt: Genau a_2 Zellen sind doppelt belegt, die anderen mit keinem bzw. genau einem Element.

$$N(n; z \parallel z - (n - a_2) \mid n - 2a_2 \mid a_2) =$$

$$\frac{n! \cdot z!}{2!^{a_2} \cdot (z - (n - a_2))! \cdot (n - 2a_2)! \cdot a_2!} \quad \text{für} \quad 1 \le a_2 \le \left[\frac{n}{2}\right].$$

Für $n \le 2z$ erhält man damit für $P(A)$ den Ausdruck

$$1 - \frac{1}{z^n} \sum_{a_2=0}^{\left[\frac{n}{2}\right]} N(n; z \parallel z - (n - a_2) \mid n - 2a_2 \mid a_2)$$

Für $n > 2z$ ist $P(A) = 1$, weil dann in mindestens einer Zelle mindestens drei Elemente liegen müssen.

Die Berechnung von $P(A)$ ist sehr aufwändig, wie auch eine Aufgabe zeigt, die GEORG SCHRAGE 1990 in seinem Aufsatz *Ein Geburtstagsproblem* behandelt: An einem Institut der Universität Dortmund haben drei der fünfzehn Institutsmitglieder am 28. Juli Geburtstag. Die Wahrscheinlichkeit, dass unter 15 Personen mindestens einmal mindestens drei einen gemeinsamen Geburtstag feiern können, errechnet sich gemäß der obigen Formel wie folgt.[2]

$$1 - \frac{1}{365^{15}} \sum_{a_2=0}^{7} N(15; 365 \parallel 350 + a_2 \mid 15 - 2a_2 \mid a_2) =$$

$$\begin{aligned} 1 - \frac{1}{365^{15}} (&N(15; 365 \parallel 350 \mid 15 \mid 0) + \\ &N(15; 365 \parallel 351 \mid 13 \mid 1) + \\ &N(15; 365 \parallel 352 \mid 11 \mid 2) + \\ &N(15; 365 \parallel 353 \mid 9 \mid 3) + \\ &N(15; 365 \parallel 354 \mid 7 \mid 4) + \\ &N(15; 365 \parallel 355 \mid 5 \mid 5) + \\ &N(15; 365 \parallel 356 \mid 3 \mid 6) + \\ &N(15; 365 \parallel 357 \mid 1 \mid 7)) \end{aligned}$$

$$= 0{,}0033\,.$$

Wesentlich aufwändiger wird die Rechnung, wenn man die Frage lösen will, die RICHARD VON MISES in seinem Artikel aufwirft. Dort handelt es sich um die Frage, wie wahrscheinlich ein Dreifachgeburtstag bei 60 Personen ist. BERNHARD BERCHTOLD hat unter www.mathematik.ch ein Programm zur Lösung von Dreiergeburtstagsproblemen veröffentlicht. Mit $z = 365$ und $n = 60$ liefert dieses Programm $P(A) = 0{,}20723 \approx 20{,}7\,\%$, also keineswegs die „Wahrscheinlichkeit von wenigen Tausendsteln“, wie von Mises in seinem Artikel berichtete (Barth/Haller 2012). Bei $n = 88$ Personen ist die Wahrscheinlichkeit zum ersten Mal größer als 50 %, nämlich 51,1 %. Ab dieser Anzahl von Personen lohnt es sich, darauf zu wetten, dass mindestens drei Personen am gleichen Tag Geburtstag feiern. Ab $n = 145$ ist die Wahrscheinlichkeit sogar größer als 95 %, nämlich 95,2 %. Ab dieser Anzahl von Personen ist es also ein signifikantes Ereignis, wenn es keinen Kalendertag gibt, an dem mindestens drei Personen Geburtstag feiern.

VON MISES stellt fest, dass man im Mittel bei einer Gesamtheit von 103 Personen erwarten kann, dass einmal drei Personen den gleichen Geburtstag haben.

Wie das Beispiel von SCHRAGE zeigt, ist die Berechnung der Geburtstagskoinzidenzen sehr aufwändig. Bei kleinen Werten von z und n ist eine direkte Berechnung durch Bestimmung aller Fälle möglich, sodass man ohne die komplizierten Besetzungszahlen auskommt. Dies zeigen wir für $z = 7$ und $n = 5$, indem wir alle $7^5 = 16807$ möglichen Besetzungen explizit darstellen. Die Zahlen 7 und 5 passen übrigens zur Frage, wie viele von fünf Personen denselben Wochentag als Tag ihrer Geburt haben.

- Alle fünf Elemente belegen verschiedene Zellen:

$$\left.\begin{array}{l}1111100\\ \cdots\\ 0011111\end{array}\right\}\ 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 = 2520 = N(5;\, 7 \,\|\, 2 \,|\, 5 \,|\, 0 \,|\, 0 \,|\, 0 \,|\, 0)$$

  Für das erste Element gibt es 7 Zellen, für das zweite 6, …, für das fünfte (7 – 4) = 3 Zellen.

- Genau eine Zelle ist doppelt belegt:

$$\left.\begin{array}{l}2111000\\ \cdots\\ 0001112\end{array}\right\}\ \binom{5}{2} \cdot 7 \cdot 6 \cdot 5 \cdot 4 = 8400 = N(5;\, 7 \,\|\, 3 \,|\, 3 \,|\, 1 \,|\, 0 \,|\, 0 \,|\, 0)$$

  Man wählt die beiden Elemente aus den fünf Elementen aus, die eine Zelle doppelt belegen sollen, und verteilt dann die vier „neuen“ Elemente *Paar*, drei *Einer* auf die sieben Zellen; das geht auf $7 \cdot 6 \cdot 5 \cdot 4$ Arten.

- Genau zwei Zellen sind doppelt belegt:

$$\left.\begin{array}{l}2210000\\ \cdots\\ 0000122\end{array}\right\}\ \binom{5}{2} \cdot \binom{3}{2} \cdot \frac{1}{2!} \cdot 7 \cdot 6 \cdot 5 = 3150$$

$$= N(5;\, 7 \,\|\, 4 \,|\, 1 \,|\, 2 \,|\, 0 \,|\, 0 \,|\, 0)$$

  Man wählt zuerst das erste Paar, dann aus den restlichen drei Elementen das zweite Paar. Dabei kommt jede Aufteilung doppelt vor, also ist durch 2! zu teilen. Die drei „neuen“ Elemente *Paar*, *Paar*, *Einer* sind dann auf die sieben Zellen zu verteilen.

- Genau eine Zelle ist dreifach, keine Zelle doppelt belegt:

$$\left.\begin{array}{l}3110000\\ \cdots\\ 0000113\end{array}\right\}\ \binom{5}{3} \cdot 7 \cdot 6 \cdot 5 = 2100 = N(5;\, 7 \,\|\, 4 \,|\, 2 \,|\, 0 \,|\, 1 \,|\, 0 \,|\, 0)$$

  Man wählt zuerst das Tripel aus und verteilt dann die drei „neuen“ Elemente *Tripel*, *Einer*, *Einer* auf die sieben Zellen.

- Eine Zelle ist dreifach, eine Zelle doppelt belegt:

$$\left.\begin{array}{l}3200000\\ \cdots\\ 0000023\end{array}\right\}\ \binom{5}{3} \cdot \binom{2}{2} \cdot 7 \cdot 6 = 420 = N(5;\, 7 \,\|\, 5 \,|\, 0 \,|\, 1 \,|\, 1 \,|\, 0 \,|\, 0)$$

  Man wählt das Tripel aus; das Paar bleibt übrig. Die beiden „neuen“ Elemente *Tripel*, *Paar* sind auf die sieben Zellen zu verteilen.

- Eine Zelle ist vierfach belegt:

$$\left.\begin{array}{l}4100000\\ \cdots\\ 0000014\end{array}\right\}\ \binom{5}{4} \cdot 7 \cdot 6 = 210 = N(5;\, 7 \,\|\, 5 \,|\, 1 \,|\, 0 \,|\, 0 \,|\, 1 \,|\, 0)$$

  Man wählt das Quadrupel aus; der Einer bleibt übrig. Die beiden „neuen“ Elemente *Quadrupel*, *Einer* sind auf die sieben Zellen zu verteilen.

- Eine Zelle ist fünffach belegt:

$$\left.\begin{array}{l}5000000\\ \cdots\\ 0000005\end{array}\right\}\ 7 = N(5;\, 7 \,\|\, 6 \,|\, 0 \,|\, 0 \,|\, 0 \,|\, 0 \,|\, 1)$$

  Man wählt für das Quintupel eine der sieben Zellen aus.

Daraus ermittelt man z. B.

P(„Mindestens eine Zelle ist mindestens dreifach belegt“) = P(„An mindestens einem Wochentag sind mindestens drei der fünf Personen geboren“)

$$= \frac{1}{16807}(2100 + 420 + 210 + 7) = \frac{2737}{16807} = 16{,}3\,\%$$

Über das Gegenereignis ergibt sich für diese Wahrscheinlichkeit der Wert

$$1 - \frac{1}{16807}(2520 + 8400 + 3150) = 83{,}7\,\%,$$

was dem oben hergeleiteten Ausdruck

$$1-\frac{1}{7^5}\sum_{a_2=0}^{\left[\frac{5}{2}\right]} N(5;7\,\|\,7-(5-a_2)\,|\,5-2a_2\,|\,a_2)=$$

$$1-\frac{1}{7^5}\,(N(5;7\,\|\,2\,|\,5\,|\,0)+N(5;7\,\|\,3\,|\,3\,|\,1)+$$

$$+N(5;7\,\|\,4\,|\,1\,|\,2\,|\,0))$$

entspricht.

3 Mehrfachgeburtstage

Ermutigt durch den Erfolg beim Dreifachgeburtstag gehen wir Problem 1 der k-fach-Geburtstage für $k \in \mathbb{N}$ an. Wie in **2** betrachten wir zunächst ein Besetzungsproblem:

Gesucht ist die Wahrscheinlichkeit des Ereignisses A_k, dass n Elemente auf z Zellen so verteilt werden, dass in mindestens einer Zelle mindestens k Elemente zu liegen kommen.

Wir betrachten das Gegenereignis $\overline{A}_k$ = „Keine der z Zellen ist mit mehr als $k-1$ Elementen besetzt." Für $\overline{A}_k$ gilt

$$P(\overline{A}_k)=\frac{1}{z^n}\sum_{\forall a_i} N(n;z\,\|\,a_0\,|\,a_1\,|\,\ldots\,|\,a_{k-1}\,|\,0\,|\,\ldots\,|\,0)\,.$$

Dabei ist über alle a_i zu summieren, die Lösungen des Diophantischen Gleichungssystems

$$\text{(I)}\ \sum_{i=0}^{k-1} a_i = z\,,\quad \text{(II)}\ \sum_{i=0}^{k-1} a_i\cdot i = n$$

sind. Von diesen Lösungen sind all diejenigen auszuschließen, die die folgenden Bedingungen *nicht* erfüllen:

(1) $0 \le a_0 \le z-1$, da mindestens eine Zelle Elemente enthalten muss.

(2) $0 \le a_{k-1} \le \min\left\{z,\left[\frac{n}{k-1}\right]\right\}$, da es höchstens $\left[\frac{n}{k-1}\right]$ Zellen mit jeweils $k-1$ Elementen geben kann.

$$\text{(3)}\ 0 \le a_i \le \min\left\{z,\ \frac{n-\sum\limits_{j=1,\,j\ne i}^{k-i} a_j\cdot j}{i}\right\}\ \text{für } 1\le i\le k-2.$$

Die Anzahl a_i der Zellen mit genau i Elementen erhält man, wenn man die in allen anderen Zellen liegenden $\sum\limits_{j=1,\,j\ne i}^{k-1} a_j\cdot j$ Elemente von n wegnimmt und den Rest durch i teilt. Da es sich um eine tatsächliche Besetzung handelt, muss diese Division aufgehen.

Bemerkung: Bedingung (3) ist in der Praxis unbrauchbar, weil man ja schon alle a_j mit Ausnahme des einzuschränkenden a_i kennen müsste. Dies kann man vermeiden, indem man (3) zu (3*) abschwächt:

$$\text{(3*)}\ 0 \le a_i \le \min\left\{z,\ \left[\frac{n-\sum\limits_{j=i+1}^{k-1} a_j\cdot j}{i}\right]\right\}$$

für $1\le i\le k-2$.

Man betrachtet dazu nur die Zellen, die mehr als i Elemente enthalten, d. h., man muss nur die a_j kennen, für die $j > i$ ist. Da die Division nicht aufgehen muss, muss man die größte Ganze aus dem Quotienten nehmen. Diese ist dann nur eine obere Schranke für a_i.

Eine weitere Abschwächung von (3*) führt auf eine Bedingung, die man mit (2) verbinden kann zu

$$\text{(3**)}\ 0 \le a_i \le \min\left\{z,\left[\frac{n}{i}\right]\right\}\ \text{ für } 1\le i\le k-1$$

Dabei braucht man über die a_j mit $j \ne i$ keinerlei Information. Dafür muss man aber all die a_i, die keine Lösung des Diophantischen Gleichungssystems sind, durch mühsames Probieren ausschließen.

Die Lösung des Diophantischen Gleichungssystems wird mit zunehmendem k immer aufwändiger. Wir zeigen das Vorgehen für $n = 5$, $z = 3$ und $k = 4$.

(I) $a_0 + a_1 + a_2 + a_3 = 3$

(II) $a_1 + 2a_2 + 3a_3 = 5$

(I) $a_0 = 3 - a_1 - a_2 - a_3$

(II) $a_1 + 2a_2 + 3a_3 = 5$

Zunächst suchen wir alle Tripel (a_1, a_2, a_3), die die Gleichung II erfüllen. Systematisches Probieren lässt sie finden. Wegen Bedingung (3**) ist

$$a_3 \le \min\left\{3,\left[\frac{5}{3}\right]\right\}=1\,,$$

also kann a_3 nur 0 oder 1 sein. Ebenso ergibt sich, dass a_2 nur 0, 1 oder 2 sein kann und a_1 nur 0, 1, 2 oder 3.

Die Bedingung (3*) fordert für $1 \le i \le 2$

$$0 \le a_i \le \min\left\{3,\ \left[\frac{n-\sum\limits_{j=i+1}^{3} a_j\cdot j}{i}\right]\right\},$$ woraus man erhält:

$$0 \le a_2 \le \min\left\{3, \left[\frac{5-3a_3}{2}\right]\right\} = \begin{cases} 2 \text{ für } a_3 = 0 \\ 1 \text{ für } a_3 = 1 \end{cases} \quad \text{und}$$

$$0 \le a_1 \le \min\left\{3, \left[\frac{5-3a_3-2a_2}{1}\right]\right\}$$

$$= \begin{cases} 1 \text{ für } a_3 = 0 \wedge a_2 = 2 \\ 3 \text{ für } a_3 = 0 \wedge a_2 = 1 \\ 3 \text{ für } a_3 = 0 \wedge a_2 = 0 \\ 0 \text{ für } a_3 = 1 \wedge a_2 = 1 \\ 2 \text{ für } a_3 = 1 \wedge a_2 = 0 \end{cases}$$

Damit ergeben sich fünf mögliche Lösungstripel und aus (I) das zugehörige a_0:

$(0, 1, 1), a_0 = 1$; $(1, 2, 0), a_0 = 0$;

$(2, 0, 1), a_0 = 0$; $(3, 0, 0), a_0 = 0$;

$(3, 1, 0), a_0 = -1$.

Da $a_0 = -1$ nicht möglich ist, verbleiben vier Tripel, die aber Gleichung II erfüllen müssen. Bei (3, 0, 0) trifft dies nicht zu, sodass das Diophantische Gleichungssystem die folgenden drei Lösungsquadrupel besitzt:

(0, 1, 2, 0), (0, 2, 0, 1) und (1, 0, 1, 1).

Damit erhält man schließlich

$$P(\bar{A}_4) = \frac{1}{3^5}\,(N(5; 3 \,\|\, 0 \mid 1 \mid 2 \mid 0)$$

$$+ N(5; 3 \,\|\, 0 \mid 2 \mid 0 \mid 1) + N(5; 3 \,\|\, 1 \mid 0 \mid 1 \mid 1))$$

$$= \frac{1}{243}\left(\frac{5! \cdot 3!}{2!^2 \cdot 2!} + \frac{5! \cdot 3!}{3! \cdot 2!} + \frac{5! \cdot 3!}{2! \cdot 3!}\right)$$

$$= \frac{1}{243}\,(90 + 60 + 60) = \frac{70}{81}$$

und damit $P(A_4) = \frac{11}{81} = 13{,}6\ \%$.

Anwendung auf das Geburtstagsproblem

Fragt man nach der Wahrscheinlichkeit, dass unter n Personen mindestens k Personen gemeinsam Geburtstag feiern können, so muss man $z = 365$ setzen. Der Rechenaufwand wird gewaltig! Auch der in **2** eingeschlagene Weg, das Gegenereignis in disjunkte Teilereignisse zu zerlegen, wird immer schwieriger, da man zu leicht den Überblick beim Zerlegen verliert.

Mit Hilfe eines Computerprogramms kann man errechnen, dass für $n = 187$ die Wahrscheinlichkeit, dass an einem Kalendertag mindestens vier Personen Geburtstag feiern, zum ersten Mal größer als 50 % ist, nämlich 50,3 %.

Bruce Levin veröffentlichte für $k \le 15$ eine Tabelle der Werte von n, bei der die Wahrscheinlichkeit für mindestens einen k-fachen Geburtstag den Wert von 50 % überschreitet. Wir bringen sie hier in verkürzter Form.

k	n	$P(A_k)$	k	n	$P(A_k)$
2	22	47,6	9	984	49,86
	23	50,7		985	50,09
3	87	49,9	10	1180	49,90
	88	51,1		1181	50,09
4	186	49,6	11	1384	49,90
	187	50,3		1385	50,10
5	312	49,6	12	1595	49,94
	313	50,1		1596	50,11
6	459	49,9	13	1812	49,94
	460	50,3		1813	50,11
7	622	49,97	14	2034	49,88
	623	50,29		2035	50,03
8	797	49,76	15	2262	49,98
	798	50,03		2263	50,12

Tab.1: Die 50 %-Schwellen für mindestens einen k-fachen Geburtstag. Die Werte in der Spalte $P(A_k)$ sind Prozentangaben.

4 Zum Problem der Gleichverteilung

An zwei einfachen Beispielen legen wir dar, wie sich die vereinfachende Modellannahme der Gleichverteilung der Geburtstage übers Jahr auf die Ergebnisse auswirken kann.

1. Beispiel: Mit welcher Wahrscheinlichkeit fallen die Geburtstage von zwölf verschiedenen Personen in die zwölf Monate?

Lösung:

1) Statt der Gleichverteilung der Geburtstage über das Jahr nehmen wir zunächst an, dass die Wahrscheinlichkeit, in einem bestimmten Monat geboren zu werden, für alle Monate gleich groß ist. Die Wahrscheinlichkeit, dass die i-te Person ($1 \le i \le 12$) im i-ten Monat geboren ist, ist somit für jedes i gleich $\frac{1}{12}$. Dann ist die Wahrscheinlichkeit, dass alle zwölf Personen in verschiedenen Monaten geboren sind, $\frac{1}{12^{12}}$. Da man die Personen aber noch permutieren kann, ergibt sich schließlich als Antwort auf die gestellte Frage der Wert

$$\frac{12!}{12^{12}} = \frac{1.925}{35.831.808} = 5{,}372 \cdot 10^{-5}.$$

2) Berücksichtigen wir hingegen die Anzahl t_i der Tage eines Monats unter Vernachlässigung von Schaltjahren und nehmen wir außerdem an, dass die Geburtstage gleichmäßig über die Kalendertage verteilt sind, dann ist die Wahrscheinlichkeit, dass die i-te Person ($1 \le i \le 12$) im i-ten Monat geboren ist, gleich $\frac{t_i}{365}$. Die Wahrscheinlichkeit, dass alle zwölf Personen in verschiedenen Monaten geboren sind, ist wegen der noch möglichen Permutation der Personen dann

$$\frac{12!}{365^{12}} \prod_{i=1}^{12} t_i = \frac{12!}{365^{12}} \cdot 28 \cdot 31^7 \cdot 30^4 = 5{,}346 \cdot 10^{-5}.$$

3) Will man die Wahrscheinlichkeitsrechnung auf die Wirklichkeit anwenden und kennt man die Wahrscheinlichkeit eines Ereignisses nicht, so ist es üblich, die relative Häufigkeit als Schätzwert für die unbekannte Wahrscheinlichkeit zu verwenden. In diesem Sinne legen wir eine empirische Verteilung von 11.088.533 Geburtstagen gemäß Tabelle 2 auf die zwölf Monate zu Grunde und bezeichnen die relative Häufigkeit gleich mit p_i.[3]

Sei also p_i die Wahrscheinlichkeit, im Monat i geboren zu sein, dann ergibt sich für die Wahrscheinlichkeit, dass alle zwölf Personen in verschiedenen Monaten geboren sind, der Wert

$$p = 12! \cdot p_1 \cdot p_2 \cdot \ldots \cdot p_{12} = 5{,}30253 \cdot 10^{-5}.$$

Geburtsmonat i	Anzahl	p_i	p_i in Prozent
Januar	953.891	0,0860249	8,6
Februar	905.613	0,0816711	8,2
März	1.002.073	0,0903702	9,0
April	949.413	0,0856211	8,6
Mai	977.686	0,0881709	8,8
Juni	934.572	0,0842831	8,4
Juli	934.531	0,0842790	8,4
August	908.572	0,0819379	8,2
September	923.252	0,0832618	8,3
Oktober	879.099	0,0792800	7,9
November	841.728	0,0759097	7,6
Dezember	878.103	0,0791901	7,9
	11.088.533		

Tab. 2: Empirische Verteilung von 11.088.533 Geburtstagen

Fazit: Erwartungsgemäß ist die Wahrscheinlichkeit umso kleiner, je mehr die Verteilung von der Gleichverteilung abweicht. Allerdings sind die Unterschiede minimal.

2. Beispiel: Mit welcher Wahrscheinlichkeit haben zwei Personen im gleichen Monat Geburtstag?

Lösung:

1) Die Geburtstage seien gleichmäßig auf die Monate verteilt. Dann ist die Wahrscheinlichkeit, dass die zweite Person im selben Monat wie die erste geboren ist,

$$p = \frac{1}{12} = 0{,}08333 = 8{,}33\,\%.$$

2) Die Geburtstage seien gleichmäßig auf die Tage verteilt. Dann ist die Wahrscheinlichkeit, dass beide Personen in einem Monat mit a_i Tagen geboren sind, gleich $\left(\frac{a_i}{365}\right)^2$. Für die gesuchte Wahrscheinlichkeit ergibt sich damit

$$p = \left(\frac{28}{365}\right)^2 + 4 \cdot \left(\frac{30}{365}\right)^2 + 7 \cdot \left(\frac{31}{365}\right)^2$$
$$= 0{,}0834002 = 8{,}34\,\%.$$

3) Man legt die empirische Verteilung der 11.088.533 Geburtstage auf die zwölf Monate von oben zu Grunde. Dann erhält man mit den dortigen Bezeichnungen

$$p = p_1^2 + p_2^2 + \ldots + p_{12}^2 = 0{,}0835139 = 8{,}35\,\%.$$

Fazit: Erwartungsgemäß ist die Wahrscheinlichkeit umso kleiner, je stärker die Verteilung einer gleichmäßigen Verteilung ähnelt. Allerdings sind auch hier die Unterschiede minimal.

Abschließend können wir feststellen: Die Annahme der Gleichverteilung verfälscht in den betrachteten Beispielen das Ergebnis nur unwesentlich.

Anmerkungen

1 Klamkin/Newman (1967)

2 SCHRAGE selbst rechnet statt mit den Besetzungszahlen gleich mit Wahrscheinlichkeiten, die man aus den ersteren ja durch Division mit 365^{15} erhält. Ähnlich gehen ROTTMANN (2002) und RIEHL (2006) vor, die k-fache Geburtstage betrachten.

3 Die Daten wurden uns freundlicherweise vom Aktuariat Leben/Kranken der Allianz Deutschland AG zur Verfügung gestellt.

Literatur

Barth F./Haller R. (1998): Stochastik – Leistungskurs. München: Oldenbourg Schulbuchverlag.

Barth F./Haller R. (2012): Besetzungen und Geburtstage. In: *Stochastik in der Schule* 32(3), S. 20–27.

Klamkin, M. S./Newman, D. J. (1967): Extensions of the Birthday Surprise. In: *Journal of Combinatorial Theory,* 3, S. 279–282.

Levin, B.: Exact solutions of the Generalized Birthday Problem. http://oeis.org/A014088/a014088.txt. (Zugriff: 27.7.2011)

Mises, R. von (1939): Dağıtma ve işgal ihtimalleri hakkında – Über Aufteilungs- und Besetzungs-Wahrscheinlichkeiten. In: *İstanbul Üniversitesi Fen Fakültesi Mecmuası – Revue de la Faculté des Sciences de l'Université d'Istanbul,* 4, S. 145–163. [Fehlerhafter] Nachdruck in: *Selected Papers of Richard von Mises.* Vol. 2. S. 313–334. Providence/Rhode Island: American Mathematical Society. 1964.

Riehl, G. (2006): Neues zum Geburtstagsproblem. In: *MNU* 59 (7), S. 439–444.

Rottmann, K. (2002): Lösung des Geburtstagsproblems für den Fall, dass *k* von *n* Personen am gleichen Tag Geburtstag haben. In: *Stochastik in der Schule* 22(2), S. 29–32.

Schrage, G. (1990): Ein Geburtstagsproblem. In: *Mathematische Semesterberichte,* 1990, 37, S. 251–257. Nachdruck in: *Stochastik in der Schule* 12(2), S. 30–36.

Anschriften der Verfasser

Friedrich Barth
Abbachstr. 23
80992 München
e.f.barth@t-online.de

Rudolf Haller
Nederlinger Straße 32a
80638 München
rudolf.haller@arcor.de