

Besetzungen und Geburtstage

FRIEDRICH BARTH UND RUDOLF HALLER, MÜNCHEN

Zusammenfassung: Das so genannte Geburtstagsparadoxon gilt als klassisches Beispiel dafür, dass naive Intuition bei Fragen nach Wahrscheinlichkeiten leicht in die Irre führen kann.

Mehrfachgeburtstage sind Sonderfälle der allgemeineren Fragestellung nach der Anzahl mehrfach besetzter Zellen, wenn man eine bestimmte Anzahl von Objekten auf vorgegebene Zellen verteilt. Auch außerhalb der Mathematik spielt diese Frage eine Rolle, u. a. in der Physik. Viele Autoren gehen bei der Lösung von Geburtstagsproblemen im Wesentlichen so vor wie von Mises, ohne jedoch ihn zu erwähnen. Der Aufsatz möchte seine Gedanken wieder in Erinnerung bringen.

Im Folgenden wird eine Antwort von RICHARD VON MISES auf einen Teilaspekt dieses Problems dargestellt, die nur auf elementaren Methoden basiert. Dem Lehrer wird damit die Möglichkeit angeboten, das Geburtstagsproblem in eine allgemeinere Fragestellung einzubetten. Ferner soll zum Nachdenken über die Interpretation von Wahrscheinlichkeit und Erwartungswert angeregt werden. Die im Aufsatz verwendete Kombinatorik ist anspruchsvoll, übersteigt aber sicher nicht die Fähigkeiten eines leistungsfähigen Schülers der Oberstufe.

1 Ein Geburtstagsproblem als Einstieg

Geburtstagsparadoxon heißt vielfach die Tatsache, dass es bereits ab 23 Personen günstig ist, darauf zu wetten, dass mindestens zwei von ihnen gemeinsam Geburtstag feiern können (Barth/Haller 1998, S. 97f.). Der Ursprung dieser Fragestellung liegt nach DONALD ERVIN KNUTH (*1938) im Dunkeln. Ihm zufolge wurde das Problem um 1930 unter Mathematikern informell diskutiert (Knuth 1998, S. 513). Als früheste Publikation hierzu ist uns das Lehrbuch von WILLIAM FELLER (1950, Section 2.3) bekannt. Manche Autoren, so z. B. PERSI DIACONIS und FREDERICK MOSTELLER (1989, Seite 857) sehen den Ursprung dieses Problems in dem Aufsatz *Über Aufteilungs- und Besetzungs-Wahrscheinlichkeiten*, den RICHARD EDLER VON MISES (1883–1953) in Istanbul 1939 veröffentlichte (Mises 1939). Denn dieser Aufsatz beginnt tatsächlich mit einem – allerdings weitergehenden – Geburtstagsproblem:

»In dem mathematischen Bureau einer Versicherungsgesellschaft stellte sich gelegentlich einer Umfrage heraus, dass von den 60 Angestellten drei den gleichen Kalendertag zum Geburtstag hatten. Dies erschien als

ein sehr ungewöhnliches, seltenes Zusammentreffen, und man versuchte auf verschiedene Weise zahlenmäßig die geringe Wahrscheinlichkeit eines solchen Ereignisses zu bestimmen. Es wurden mir Berechnungen vorgelegt, die eine Wahrscheinlichkeit von wenigen Tausendsteln ergaben! Als Grundannahme galt selbstverständlich die, dass für jede der $n = 60$ Personen die gleiche Wahrscheinlichkeit z^{-1} besteht, an einem der $z = 365$ Kalendertage des Jahres geboren zu sein.«¹



1932

Richard Edler von Mises
*19.4.1883 Lemberg
†14.7.1953 Boston

Es handelt sich also um die Frage, wie groß die Wahrscheinlichkeit ist, dass unter n Personen mindestens einmal mindestens drei am selben Kalendertag Geburtstag feiern können. VON MISES geht auf diese spezielle Eingangsfrage nicht näher ein. Stattdessen widmet er sich, wie aus dem Titel seiner Arbeit hervorgeht, anderen mit dieser Fragestellung zusammenhängenden Problemen. Sie ergeben sich, wenn man n unterscheidbare Elemente auf

z Zellen verteilen will. Die von ihm gewonnenen Ergebnisse wendet VON MISES dann auf Fragen bezüglich von Mehrfachgeburtstagen an.

2 Wahrscheinlichkeit einer Besetzung

Seiner Arbeit legt VON MISES das Zufallsexperiment zugrunde, dass n unterscheidbare Elemente auf z nummerierte Zellen per Zufall verteilt werden. Jede der z^n möglichen Verteilungen hat somit gleiche Wahrscheinlichkeit. In § 4 betrachtet er zwei mögliche Fragestellungen, von denen er in seinem Aufsatz aber nur die zweite weiter behandelt.

Erstens:

Gesucht ist die Wahrscheinlichkeit einer Verteilung, bei der per Zufall n unterscheidbare Elemente auf z nummerierte Zellen so verteilt werden, dass genau e_j Elemente in der j -ten Zelle liegen, $1 \leq j \leq z$.

Diese Wahrscheinlichkeit nennt VON MISES *Aufteilungswahrscheinlichkeit*, »deren Untersuchung ein klassisches, für grosse Teile der physikalischen Sta-

tistik grundlegendes Problem bildet^{1,2} heute unter dem Namen *Maxwell-Boltzmann-Statistik* bekannt.

Zweitens:

Gesucht ist die Wahrscheinlichkeit einer Verteilung, bei der per Zufall n unterscheidbare Elemente auf z nummerierte Zellen so verteilt werden, dass genau a_i Zellen mit genau i Elementen besetzt sind, $0 \leq i \leq n$.

Diese Wahrscheinlichkeit nennt VON MISES *Besetzungswahrscheinlichkeit*; sie ist das eigentliche Thema seiner Arbeit. 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.

Zur Beantwortung der 2. Fragestellung bezeichnen wir die Anzahl der Besetzungen, bei denen genau a_i Zellen mit genau i der n Elemente besetzt sind, mit $N(n; z \parallel a_0 \mid a_1 \mid a_2 \mid \dots \mid a_n)$, $0 \leq i \leq n$. Dabei unterliegen die $a_i \in \mathbb{N}_0$ folgenden Einschränkungen.

- Höchstens z Zellen können belegt werden, d. h.,

$$(1) \quad 0 \leq a_i \leq z.$$

- Genau z Zellen stehen zur Verfügung, d. h.,

$$(2) \quad \sum_{i=0}^n a_i = z.$$

- Es werden insgesamt n Elemente verteilt, d. h.,

$$(3) \quad \sum_{i=0}^n a_i \cdot i = n.$$

- Bei n gegebenen Elementen können höchstens $\frac{n}{i}$ Zellen mit i Elementen besetzt werden. Da die $a_i \in \mathbb{N}_0$ sind, muss man ggf. die größte Ganze von $\frac{n}{i}$ nehmen, d. h.,

$$(4) \quad a_i \leq \left\lfloor \frac{n}{i} \right\rfloor \text{ für } 1 \leq i \leq n.$$

Die gesuchte Wahrscheinlichkeit erhält man mittels Division von N durch z^n .

Die Abzählung für N nehmen wir, angelehnt an VON MISES, schrittweise vor.

1. Schritt: Wir besetzen die a_i Zellen mit jeweils i Elementen der Reihe nach. Zunächst bleiben a_0 Zellen leer, dann werden a_1 Zellen mit genau einem Element besetzt, die folgenden a_2 Zellen mit jeweils genau zwei Elementen, usw. bis hin zu den a_k Zellen, die mit jeweils genau k Elementen zu besetzen sind. Damit sind alle n Elemente untergebracht, und man

erhält folgendes spezielle z -Tupel der Besetzungszahlen der Zellen.

$$\begin{array}{ccccccc} & & & & n & & \\ & & & & \overline{\hspace{1.5cm}} & & \\ 00 \dots 0 & 11 \dots 1 & 22 \dots 2 & \dots & kk \dots k & & \\ \underbrace{\hspace{1.5cm}} & \underbrace{\hspace{1.5cm}} & \underbrace{\hspace{1.5cm}} & & \underbrace{\hspace{1.5cm}} & & \\ a_0 & a_1 & a_2 & & a_k & & \\ & & & & z & & \end{array}$$

Wie viele solcher z -Tupel gibt es? Permutiert man die Elemente dieses speziellen z -Tupels, dann erhält man Permutationen mit Wiederholungen. Jede dieser Permutationen beschreibt aber dieselbe Anzahl von Zellen mit $0, 1, \dots, k$ Elementen. Die Anzahl dieser z -Tupel ist $\frac{z!}{a_0! \cdot a_1! \cdot \dots \cdot a_k!}$. Da $a_i! = 0! = 1$ ist, falls $k < n$ ist und i die Ungleichung $k + 1 \leq i \leq n$ erfüllt, kann man $\frac{z!}{a_0! \cdot a_1! \cdot \dots \cdot a_k!}$ auch als $\frac{z!}{a_0! \cdot a_1! \cdot \dots \cdot a_n!}$ schreiben.

2. Schritt: Nachdem festgelegt wurde, wie viele Elemente in den Zellen zu liegen kommen, müssen nun die zugehörigen Elemente aus den n unterscheidbaren Elementen dafür ausgewählt werden. Dazu ordnen wir die n Elemente als n -Tupel an, von denen es $n!$ gibt. Um das spezielle z -Tupel des 1. Schritts zu füllen, gliedern wir jedes dieser n -Tupel nun folgendermaßen. Die ersten a_1 Elemente werden jedes für sich genommen und der Reihe nach auf die Zellen verteilt, die nach den a_0 leeren Zellen kommen; die folgenden $2a_2$ Elemente werden zu a_2 Paaren zusammengefasst und auf die folgenden a_2 Zellen verteilt; die folgenden $3a_3$ Elemente werden zu a_3 Tripeln zusammengefasst und auf die folgenden a_3 Zellen verteilt, usw. Eine Permutation der Elemente, die in der gleichen Zelle liegen, ändert die Besetzung nicht.

Somit gibt es $\frac{n!}{1!^{a_0} \cdot 2!^{a_2} \cdot \dots \cdot k!^{a_k}}$ Möglichkeiten für die Verteilung der n Elemente auf die Zellen des z -Tupels des 1. Schritts. Formal lässt sich dieser Ausdruck auch schreiben als $\frac{n!}{0!^{a_0} \cdot 1!^{a_1} \cdot 2!^{a_2} \cdot \dots \cdot n!^{a_n}}$, da ja einerseits $0!^{a_0} = 1$ ist, und außerdem $i!^i = i!^0 = 1$ ist, falls $k < n$ ist und i die Ungleichung $k + 1 \leq i \leq n$ erfüllt.

Die Anzahl aller Besetzungen, also den Wert für N , erhält man demnach, wenn man die in den beiden Schritten gefundenen Ausdrücke miteinander multipliziert, weil jeder Fall des 1. Schritts mit jedem Fall aus dem 2. Schritt kombiniert werden kann. Dies ergibt die **Besetzungszahl** dafür, dass n Elemente auf z Zellen so verteilt werden, dass genau a_i Zellen mit genau i Elementen besetzt sind:

$$\begin{aligned} N(n; z \parallel a_0 \mid a_1 \mid a_2 \mid \dots \mid a_n) &= \\ &= \frac{n!}{0!^{a_0} \cdot 1!^{a_1} \cdot 2!^{a_2} \cdot 3!^{a_3} \cdot \dots \cdot n!^{a_n}} \cdot \frac{z!}{a_0! \cdot a_1! \cdot a_2! \cdot a_3! \cdot \dots \cdot a_n!} \end{aligned}$$

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

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

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

Beispiel. An kleinen Zahlen wollen wir das obige Abzählverfahren beispielhaft erläutern.

Acht Elemente sollen auf sechs Zellen so verteilt werden, dass eine Zelle einfach, zwei Zellen doppelt und eine Zelle dreifach besetzt sind; zwei Zellen bleiben also leer, d. h., $a_0 = 2, a_1 = 1, a_2 = 2, a_3 = 1, a_4 = \dots = a_8 = 0$.

1. Schritt: Die Besetzung der sechs Zellen in aufsteigender Reihenfolge erzeugt das Sextupel

$$\begin{array}{cccccc} & n=8 & & & & \\ \underline{00} & \underline{1} & \underline{22} & \underline{3} & & \\ \underline{2} & \underline{1} & \underline{2} & \underline{1} & & \\ \hline & z=6 & & & & \end{array}$$

Insgesamt gibt es $\frac{6!}{2! \cdot 1! \cdot 2! \cdot 1!} = 180$ solcher Sextupel.

Eines davon ist z. B. das Sextupel **2 0 1 0 3 2**.

2. Schritt: Ein mögliches Oktupel der acht Elemente ist $a b c d e f g h$. Jedes dieser 8! Oktupel wird folgendermaßen eingeteilt: Das erste Element steht allein ($a_1 = 1$), dann folgen zwei Paare ($a_2 = 2$) und schließlich ein Tripel ($a_3 = 1$), z. B. $a (b c)(d e)(f g h)$. Permutationen innerhalb der Zellen liefern dieselbe Besetzung; so beschreibt z. B. $a (c b)(e d)(h g f)$ dieselbe Besetzung wie $a (b c)(d e)(f g h)$. Daher gibt es nur $\frac{8!}{1! \cdot 2! \cdot 2! \cdot 3!} = 1680$ Möglichkeiten für die Verteilung der acht Elemente auf die Zellen des Sextupels des 1. Schritts.

Somit ergeben sich für die gestellte Aufgabe insgesamt $180 \cdot 1680 = 302400$ mögliche Besetzungen. Diese Zahl ergibt sich auch als $N(8; 6 \parallel 2 \mid 1 \mid 2 \mid 1)$

$$= \frac{8!}{0!^2 \cdot 1!^1 \cdot 2!^2 \cdot 3!^1 \cdot 4!^0 \cdot 5!^0 \cdot 6!^0 \cdot 7!^0 \cdot 8!^0} \cdot \frac{6!}{2! \cdot 1! \cdot 2! \cdot 1! \cdot 0! \cdot 0! \cdot 0! \cdot 0!}$$

Wir haben oben für die Besetzungszahl N eine explizite Formel hergeleitet. KAY ROTTMANN (*1981) gibt zur Berechnung von N ein Rekursionsverfahren an, bemerkt aber dazu: »Es ist zwar relativ leicht, ein Computerprogramm [...] zu schreiben, allerdings stößt ein solches schnell an [...]« die Leistungsgrenzen eines Computers. Wie aufwändig dieses Rekursionsverfahren ist, sieht man sogar schon an dem Beispiel mit kleinen Zahlen, das ROTTMANN selbst bringt (Rottmann 2002). Ein ähnliches Vorgehen

schildert GERD RIEHL (2006), bemerkt aber auch, dass für große n ein Computer schnell überfordert ist. Als Ausweg schlägt er ein Verfahren vor, das mit we-niger Aufwand Näherungswerte liefert.

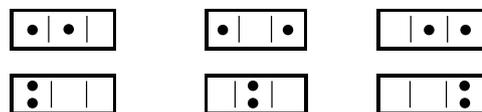
3 Wie realisiert man eine zufällige gleichwahrscheinliche Besetzung?

VON MISES gibt keine Realisierung für das in 2 angeführte Zufallsexperiment an. Man könnte folgendermaßen vorgehen:

In einer Urne liegen z Kugeln, die mit den Zahlen von 1 bis z beschriftet sind. Ferner gibt es z von 1 bis z nummerierte Zellen. Außerdem stehen n unterscheidbare Objekte zur Verfügung. Zieht man beim k -ten Zug die Kugel mit der Nummer i , dann legt man das k -te Objekt in die Zelle mit der Nummer i . Insgesamt wird n -mal mit Zurücklegen gezogen. Auf diese Weise entsteht eine Zufallsbesetzung der Zellen, bei der jede Besetzung die gleiche Wahrscheinlichkeit hat.

Vorsicht! Weil es nur auf die Anzahl der Objekte in einer Zelle ankommt, könnte man meinen, dass die Unterscheidung der Objekte im obigen Zufallsexperiment nicht nötig ist. Dass dem nicht so ist, zeigt das folgende Gegenbeispiel.

Es sei $z = 3$ und $n = 2$. Unterscheidet man die Objekte nicht, dann gibt es die folgenden sechs Möglichkeiten, die drei Zellen zu besetzen:



Dabei ist allerdings jede der Besetzungen aus der ersten Zeile doppelt so wahrscheinlich wie eine aus der zweiten Zeile, weil die ersteren auf zwei Arten zustande kommen können, die letzteren aber nur auf eine Art. So entsteht die erste Besetzung der ersten Zeile durch die Züge 1–2 und 2–1, wohingegen die erste der zweiten Zeile nur durch den Zug 1–1 entstehen kann. Somit ist bei diesem Vorgehen die Gleichwahrscheinlichkeit der Besetzungen nicht gegeben.

4 Wie wahrscheinlich sind gleiche Ziffern in einer Zufallszahl?

Als Anwendung der komplizierten Formel für die Besetzungszahl N betrachten wir n -stellige Zufallszahlen³, wobei uns nur interessiert, wie oft jede der zehn Ziffern auftritt.

Durch n -maliges Ziehen mit Zurücklegen aus einer Urne mit $z = 10$ von 0 bis 9 nummerierten Kugeln erzeugt man eine n -stellige Zufallszahl.

1. *Frage:* Mit welcher Wahrscheinlichkeit hat eine n -stellige Zufallszahl mindestens zwei gleiche Ziffern?

Für $n > 10$ ist die gesuchte Wahrscheinlichkeit 1. Es sei also im Folgenden $n \leq 10$. Zur Beantwortung der Frage betrachtet man das Gegenereignis, dass alle Ziffern verschieden sind. Die Anzahl der hierfür günstigen Fälle ist

$$N(n; 10 \parallel 10 - n \mid n) = \frac{n! \cdot 10!}{(10 - n)! \cdot n!} = \frac{10!}{(10 - n)!}$$

In diesem einfachen Fall kann man das Ergebnis auch durch direktes Abzählen ohne Verwendung der Formel für die Besetzungszahlen erhalten: An der ersten Stelle können 10 Ziffern stehen, an der zweiten nur noch 9, ..., an der n -ten Stelle nur noch $10 - (n - 1)$. Das ergibt

$$10 \cdot 9 \cdot \dots \cdot (10 - (n - 1)) = \frac{10!}{(10 - n)!} \text{ } n\text{-stellige verschieden-ziffrige Zufallszahlen.}$$

Die gesuchte Wahrscheinlichkeit hat also den Wert $1 - \frac{10!}{10^n(10 - n)!} = 1 - \left(1 - \frac{1}{10}\right)\left(1 - \frac{2}{10}\right)\dots\left(1 - \frac{n-1}{10}\right)$. Für eine 7-stellige Zufallszahl ergibt sich damit die Wahrscheinlichkeit

$$1 - \left(1 - \frac{1}{10}\right)\left(1 - \frac{2}{10}\right)\left(1 - \frac{3}{10}\right)\left(1 - \frac{4}{10}\right)\left(1 - \frac{5}{10}\right)\left(1 - \frac{6}{10}\right) = 1 - 0,939520 = 94,0 \%$$

Bei einer 7-stelligen Telefonnummer kann man also ziemlich sicher sein, dass mindestens eine Ziffer mindestens zweimal auftritt, wenngleich die Telefonnummern keine Zufallszahlen im strengen Sinne sind.

2. *Frage:* Mit welcher Wahrscheinlichkeit enthält eine 7-stellige Zufallszahl

- genau eine Ziffer dreifach und keine weitere mehrfach,
- genau zwei Ziffern doppelt und keine weitere mehrfach,
- genau eine Ziffer dreifach und genau eine Ziffer doppelt und keine weitere mehrfach?

$$a) \frac{1}{10^7} N(7; 10 \parallel 5 \mid 4 \mid 0 \mid 1) = \frac{7!}{0!^5 1!^4 2!^0 3!^1} \cdot \frac{10!}{5! \cdot 4! \cdot 0! \cdot 1!} = \frac{1.058.400}{10^7} = 10,6 \%$$

$$b) \frac{1}{10^7} N(7; 10 \parallel 5 \mid 3 \mid 2 \mid 0) = \frac{7!}{0!^5 1!^3 2!^2 3!^0} \cdot \frac{10!}{5! \cdot 3! \cdot 2! \cdot 0!} = \frac{3.175.200}{10^7} = 31,8 \%$$

$$c) \frac{1}{10^7} N(7; 10 \parallel 6 \mid 2 \mid 1 \mid 1) = \frac{7!}{0!^6 1!^2 2!^1 3!^1} \cdot \frac{10!}{6! \cdot 2! \cdot 1! \cdot 1!} = \frac{1.058.400}{10^7} = 10,6 \%$$

Erstaunlicherweise sind die Ergebnisse bei a und c gleich. Der Wert von b ist exakt dreimal so groß wie der von a.

Anhand von Zufallszahlentabellen kann man diesen Sachverhalt leicht überprüfen.

5 Wie viele Zellen werden im Mittel mindestens dreifach besetzt?

Die in der Überschrift gestellte Frage lässt sich beantworten, wenn man den Erwartungswert der Anzahl genau s -fach besetzter Zellen kennt. VON MISES berechnet ihn in § 1 seiner Arbeit.⁴

Auf z nummerierte Zellen sollen n unterscheidbare Elemente per Zufall verteilt werden. X_s sei die Zufallsgröße »Anzahl der genau s -fach besetzten Zellen«, $0 \leq s \leq n$. Gesucht ist die mittlere Anzahl genau s -fach besetzter Zellen, also der Erwartungswert $E(X_s)$.

Ehe wir uns der VON MISES'schen Lösung zuwenden, betrachten wir zum besseren Verständnis ein einfaches

Beispiel. Für $z = 3$ und $n = 5$ gibt es $z^n = 3^5 = 243$ Besetzungen. Diese zerfallen in fünf Gruppen:

- Alle fünf Elemente werden in eine Zelle gelegt;

Typ: `abcde` | | | | |

Anzahl der Besetzungen: Da man die Zelle auf $\binom{3}{1}$ Arten auswählen kann, gibt es 3 Besetzungen.

Das ist die Besetzungszahl $N(5; 3 \parallel 0 \mid 0 \mid 0 \mid 0 \mid 1)$.

- Eine Zelle wird vierfach, eine andere einfach besetzt;

Typ: `abcd` | `e` | | | | |

Anzahl der Besetzungen: Man wählt zunächst die Zelle, die vierfach besetzt wird; das geht auf $\binom{3}{1}$ Arten. Dann wählt man die Zelle, die einfach besetzt wird; das geht auf $\binom{2}{1}$ Arten. Nun bestimmt man die vier Elemente für die erste Zelle, das geht auf $\binom{5}{4}$ Arten. Für das Einzelement gibt es noch $\binom{1}{1}$ Möglichkeiten. Also besteht Typ 2 aus

$\binom{3}{1} \cdot \binom{2}{1} \cdot \binom{5}{4} \cdot \binom{1}{1} = 30$ Besetzungen. Das ist die Besetzungszahl $N(5; 3 \parallel 1 \mid 1 \mid 0 \mid 0 \mid 1)$.

- Eine Zelle wird dreifach, eine andere doppelt besetzt;

Typ: `abc` | `de` | | | | |

Anzahl der Besetzungen: *Mutatis mutandis* erhält man analog zu 2: Typ 3 besteht aus $\binom{3}{1} \cdot \binom{2}{1} \cdot \binom{5}{3} \cdot \binom{2}{2} = 60 = N(5; 3 \parallel 1 \mid 0 \mid 1 \mid 1)$ Besetzungen.

- 4) Eine Zelle wird dreifach, die beiden anderen einfach besetzt;

Typ: $\boxed{abc | d | e}$.

Anzahl der Besetzungen: Man wählt zunächst die Zelle, die dreifach besetzt wird; das geht auf $\binom{3}{1}$ Arten. Dann bestimmt man die drei Elemente für diese Zelle, das geht auf $\binom{5}{3}$ Arten. Die beiden übrigen Elemente kann man auf $2!$ Arten permutieren. Also besteht Typ 4 aus

$$\binom{3}{1} \cdot \binom{5}{3} \cdot 2! = 60 = N(5; 3 \parallel 0 | 2 | 0 | 1) \text{ Besetzungen.}$$

- 5) Zwei Zellen werden doppelt, eine einfach besetzt;

Typ: $\boxed{ab | cd | e}$.

Anzahl der Besetzungen: Man wählt zunächst die Zelle, die einfach besetzt wird; das geht auf $\binom{3}{1}$ Arten. Dann bestimmt man das Element für diese Zelle, das geht auf $\binom{5}{1}$ Arten. Dann wählt man zwei Elemente für die eine doppelt besetzte Zelle aus, das geht auf $\binom{4}{2}$ Arten. Also besteht Typ 4 aus $\binom{3}{1} \cdot \binom{5}{1} \cdot \binom{4}{2} = 90 = N(5; 3 \parallel 0 | 1 | 2)$ Besetzungen.

Es gibt also sechs Zufallsgrößen X_s , nämlich X_0, X_1, X_2, X_3, X_4 und X_5 . Ihre Wahrscheinlichkeitsverteilungen $P(X_s = x)$ erhält man – ggf. durch Addition – aus den obigen fünf Typen.

| x | 0 | 1 | 2 | 3 |
|------------------------|-----|-----|----|---|
| $3^5 \cdot P(X_0 = x)$ | 150 | 90 | 3 | 0 |
| $3^5 \cdot P(X_1 = x)$ | 63 | 120 | 60 | 0 |
| $3^5 \cdot P(X_2 = x)$ | 93 | 60 | 90 | 0 |
| $3^5 \cdot P(X_3 = x)$ | 123 | 120 | 0 | 0 |
| $3^5 \cdot P(X_4 = x)$ | 213 | 30 | 0 | 0 |
| $3^5 \cdot P(X_5 = x)$ | 240 | 3 | 0 | 0 |

Damit lassen sich die Erwartungswerte $E(X_s)$ berechnen.

$$E(X_0) = \frac{1}{3^5} (0 \cdot 150 + 1 \cdot 90 + 2 \cdot 3 + 3 \cdot 0) = \frac{96}{243} = \frac{32}{81} = 0,40.$$

Analog errechnet man

$$E(X_1) = \frac{240}{243} = \frac{80}{81} = 0,99, \quad E(X_2) = \frac{240}{243} = \frac{80}{81} = 0,99,$$

$$E(X_3) = \frac{120}{243} = \frac{40}{81} = 0,49, \quad E(X_4) = \frac{30}{243} = \frac{10}{81} = 0,12$$

und

$$E(X_5) = \frac{3}{243} = \frac{1}{81} = 0,01.$$

Nun wenden wir uns der VON MISES'schen Lösung der Fragestellung zu.

Sei $P(X_s = x)$ die Wahrscheinlichkeit, dass genau x Zellen genau s -fach besetzt sind, dann gilt definitionsgemäß $E(X_s) = \sum x P(X_s = x)$, wobei über alle Werte der Zufallsgröße X_s summiert wird.

VON MISES hat bei seinen Betrachtungen stillschweigend angenommen, dass $z \geq n$ ist, was jedoch nicht sein muss, wie unser obiges kleines Beispiel zeigt. Tatsächlich hängt die Wertemenge W_{X_s} von X_s von n, z und s ab, wobei grundsätzlich gilt: X_s kann nur Werte annehmen, die nicht größer als $\lfloor \frac{n}{s} \rfloor$ sind. Sind nämlich genau x Zellen s -fach besetzt, dann sind $x \cdot s$ Elemente verbraucht. Weil es aber nur n Elemente gibt, muss $x \cdot s \leq n$ oder $x \leq \lfloor \frac{n}{s} \rfloor$ gelten. Da schließlich x nicht-negativ ganzzahlig ist, erhält man $x \leq \lfloor \frac{n}{s} \rfloor$.

Im Einzelnen ergibt sich die folgende Fallunterscheidung:

I. $s = 0$.

- $z \geq n$: $W_{X_0} = \{z - n, \dots, z - 1\}$.

Begründung: Es können höchstens $z - 1$ Zellen leer sein, weil ja mindestens *eine* Zelle besetzt sein muss. Andererseits müssen mindestens $z - n$ Zellen leer bleiben, wenn man n Zellen mit jeweils genau einem Element besetzt.

- $z < n$: $W_{X_0} = \{0, \dots, z - 1\}$.

Begründung: Für den Wert $z - 1$ gilt das oben Gesagte. Es kann jetzt aber auch der Fall eintreten, dass *keine* Zelle leer bleibt, wenn man $z - 1$ Zellen jeweils einfach und die verbleibende Zelle mit $n - (z - 1)$ Elementen besetzt.

II. $1 \leq s \leq n$.

- $z \geq n$: $W_{X_s} = \{0, 1, \dots, \lfloor \frac{n}{s} \rfloor\}$.

Begründung: Für $\lfloor \frac{n}{s} \rfloor$ siehe oben. Es stehen auch genügend Zellen zur Verfügung, weil

$\lfloor \frac{n}{s} \rfloor \leq \frac{n}{s} \leq n \leq z$. Es kann aber auch sein, dass *keine* Zelle s -fach besetzt ist.

- $z < n$.

Hier müssen drei Fälle unterschieden werden.

- 1) $n = z \cdot s$: $W_{X_s} = \{0, 1, \dots, z\}$.

Begründung: Alle z Zellen können s -fach besetzt werden. Es kann aber auch sein, dass *keine* Zelle s -fach besetzt ist.

- 2) $n < z \cdot s$: $W_{X_s} = \{0, 1, \dots, \lfloor \frac{n}{s} \rfloor\}$.

Begründung: Für $\lfloor \frac{n}{s} \rfloor$ siehe oben. Es stehen auch genügend Zellen zur Verfügung, weil

$\lfloor \frac{n}{s} \rfloor < \lfloor \frac{z \cdot s}{s} \rfloor = z$. Es kann aber auch sein, dass *keine* Zelle s -fach besetzt ist.

- 3) $n > z \cdot s$: $W_{X_s} = \{0, \dots, z - 1\}$.

Begründung: Es können nicht alle z Zellen s -fach besetzt sein, weil $z \cdot s < n$ ist, und somit Elemente

übrig bleiben. Andererseits können $z - 1$ Zellen s -fach besetzt werden, wenn man die restlichen $n - z \cdot s$ Elemente in die verbleibende Zelle legt. Es kann aber auch sein, dass *keine* Zelle s -fach besetzt ist.

Nun können wir uns der Berechnung von $E(X_s) = \sum xP(X_s = x)$ zuwenden, wobei wir wieder zu unterscheiden haben.

- $z \geq n$: Für diesen von VON MISES betrachteten Fall gilt:

Ist $s = 0$, dann ist

$$E(X_0) = \sum_{x=z-n}^{z-1} xP(X_0 = x) = \sum_{x=1}^{z-1} xP(X_0 = x),$$

weil $P(X_0 = x) = 0$ ist für alle $x < z - n$, da nicht weniger als $z - n$ Zellen leer sein können.

Ist $s > 0$, dann ist

$$E(X_s) = \sum_{x=0}^{\lfloor \frac{n}{s} \rfloor} xP(X_s = x) = \sum_{x=1}^{\lfloor \frac{n}{s} \rfloor} xP(X_s = x).$$

- $z < n$: Für diesen von VON MISES nicht betrachteten Fall gilt:

Ist $s = 0$, dann ist

$$E(X_0) = \sum_{x=0}^{z-1} xP(X_0 = x) = \sum_{x=1}^{z-1} xP(X_0 = x).$$

Ist $s > 0$, dann gilt:

Für $n = z \cdot s$ ist

$$E(X_s) = \sum_{x=0}^z xP(X_0 = x) = \sum_{x=1}^{\lfloor \frac{n}{s} \rfloor} xP(X_s = x).$$

Für $n < z \cdot s$ ist

$$E(X_s) = \sum_{x=0}^{\lfloor \frac{n}{s} \rfloor} xP(X_s = x) = \sum_{x=1}^{\lfloor \frac{n}{s} \rfloor} xP(X_s = x).$$

Für $n > z \cdot s$ ist

$$E(X_s) = \sum_{x=0}^{z-1} xP(X_0 = x) = \sum_{x=1}^{\lfloor \frac{n}{s} \rfloor} xP(X_s = x),$$

weil $P(X_s = x) = 0$ für alle $x > \lfloor \frac{n}{s} \rfloor$, da X_s ja nur Werte bis $\lfloor \frac{n}{s} \rfloor$ annehmen kann.

In allen Fällen ergeben sich erfreulicherweise dieselben Summen, wie sie VON MISES für seinen Fall errechnet hat.

Zur Berechnung dieser Summen schlägt VON MISES den folgenden Weg ein.

Sei p_i die Wahrscheinlichkeit, dass die i -te Zelle genau s -fach besetzt ist, die Besetzung der anderen Zellen hingegen beliebig, also gleich oder verschiedenen von s . Man kann auf $\binom{n}{s}$ Arten die s Elemente auswählen, die die i -te Zelle besetzen. Die restlichen

$n - s$ Elemente kann man auf $(z - 1)^{n-s}$ Arten auf die übrigen $z - 1$ Zellen verteilen. Weil es insgesamt z^n Besetzungen gibt, erhält man

$$p_i = \frac{1}{z^n} \binom{n}{s} (z - 1)^{n-s} = \frac{1}{z^s} \binom{n}{s} \left(1 - \frac{1}{z}\right)^{n-s}.$$

Man beachte: p_i ist unabhängig von i , also sind alle p_i gleich, z. B. gleich p_1 .

Zur **Illustration** berechnen wir p_3 für $z = 5$, $n = 3$ und $s = 2$. Die Besetzungen sind also vom Typ

| | | | | |
|--|--|---|--|--|
| | | 2 | | |
|--|--|---|--|--|

. Wir zählen sie ab: Man wählt zwei Elemente aus den drei Elementen für die Doppelbesetzung aus, das geht auf $\binom{3}{2}$ Arten. Für das verbleibende Element gibt es 4^1 Möglichkeiten. Also ist $p_3 = \frac{1}{125} \binom{3}{2} \cdot 4^1$. Das ist genau der VON MISES'sche Ausdruck für p_3 , nämlich $\frac{1}{5^3} \binom{3}{2} \cdot (5 - 1)^{3-2}$.

Nun berechnet VON MISES die Summe der Wahrscheinlichkeiten derjenigen Besetzungen, bei denen *mindestens* eine Zelle genau s -fach besetzt ist, also $p_1 + p_2 + \dots + p_z = zp_1$.

In der Summe $p_1 + p_2 + \dots + p_z$ wird eine Besetzung, bei der genau t Zellen genau s -fach besetzt sind, genau t -mal gezählt. Also gilt

$$\begin{aligned} p_1 + p_2 + \dots + p_z &= 1 \cdot P(X_s = 1) + 2 \cdot P(X_s = 2) + \dots + z \cdot P(X_s = z) \\ &= \sum_{x=1}^z xP(X_s = x) = E(X_s). \end{aligned}$$

Weil $P(X_s = x) = 0$ ist für alle $x > \lfloor \frac{n}{s} \rfloor$, kann man die letzte Summe auch als $\sum_{x=1}^{\lfloor \frac{n}{s} \rfloor} xP(X_s = x)$ schreiben.

Damit hat VON MISES für $s = 0, 1, \dots, n$

$$E(X_s) = zp_1 = z \cdot \frac{1}{z^s} \binom{n}{s} \left(1 - \frac{1}{z}\right)^{n-s} \text{ oder}$$

$$E(X_s) = \frac{1}{z^{s-1}} \binom{n}{s} \left(1 - \frac{1}{z}\right)^{n-s}.$$

Die VON MISES'sche Formel liefert erfreulicherweise genau die Werte, die wir in unserem obigen einfachen Beispiel explizit berechnet haben. Es gilt nämlich

$$E(X_0) = \frac{1}{3^{-1}} \binom{5}{9} \left(1 - \frac{1}{3}\right)^5 = 3 \cdot \left(\frac{2}{3}\right)^5 = \frac{32}{81}.$$

Analog zeigt man für $s = 1, \dots, 5$ die Übereinstimmung der beiden Berechnungen von $E(X_s)$.

Ergänzend zu RICHARD VON MISES zeigen wir:

$$\sum_{s=0}^n E(X_s) = z$$

Das bedeutet: Die Summe der mittleren Anzahlen der genau s -fach besetzten Zellen ergibt die Gesamtzahl

der Zellen. Das ist nicht erstaunlich, weil bei dieser Summe alle Besetzungsmöglichkeiten von 0 bis n berücksichtigt werden. Also muss diese Summe alle z Zellen erfassen.

$$\begin{aligned} \text{Beweis: } \sum_{s=0}^n E(X_s) &= \sum_{s=0}^n z \cdot \frac{1}{z^s} \binom{n}{s} \left(1 - \frac{1}{z}\right)^{n-s} \\ &= z \cdot \left[\binom{n}{0} \left(\frac{1}{z}\right)^0 \left(1 - \frac{1}{z}\right)^{n-0} + \binom{n}{1} \left(\frac{1}{z}\right)^1 \left(1 - \frac{1}{z}\right)^{n-1} + \dots \right. \\ &\quad \left. \dots + \binom{n}{n} \left(\frac{1}{z}\right)^n \left(1 - \frac{1}{z}\right)^{n-n} \right] \\ &= z \cdot \left[\left(\frac{1}{z} + \left(1 - \frac{1}{z}\right)\right)^n \right] = z \cdot 1^n = z. \end{aligned}$$

In unserem einfachen Beispiel erhält man

$$\sum_{s=0}^5 E(X_s) = \frac{32}{81} + \frac{80}{81} + \frac{80}{81} + \frac{40}{81} + \frac{10}{81} + \frac{1}{81} = \frac{243}{81} = 3.$$

6 Anwendung auf Geburtstagsprobleme

Bei den üblichen Behandlungen des Geburtstagsproblems betrachtet man eine Gruppe von n Personen und berechnet die *Wahrscheinlichkeit* dafür, dass in *dieser* Gruppe ein s -facher Geburtstag auftritt. Interessant ist dann, ab welchem Wert von n diese Wahrscheinlichkeit 50 % überschreitet, weil man dann mit Erfolgsaussicht auf dieses Ereignis wetten kann. VON MISES hingegen betrachtet das Geburtstagsproblem von einer anderen Seite. Er geht von *vielen* Gruppen von n Personen aus und bestimmt den Erwartungswert der Zufallsgröße der Anzahl genau s -fach besetzter Tage. Jetzt ist es interessant zu wissen, *wie viele* Gruppen von n Personen man betrachten muss, damit dieser *Erwartungswert* mindestens 1 ist. In diesem Sinne wendet VON MISES die gefundene Formel für $E(X_s)$ auf das Problem der Istanbul Versicherungsgesellschaft an, dass von den 60 Angestellten drei den gleichen Kalendertag zum Geburtstag haben (siehe 1); denn, so schreibt er in der Einleitung zu seinem Aufsatz:

»Der Erwartungswert der Zufallsgröße X_3 (und nicht irgend eine Wahrscheinlichkeit) ist mit dem Beobachtungsergebnis: 1 dreifach besetzter Platz zu vergleichen.«⁵

Auf die in 1 erwähnte Wahrscheinlichkeit von »wenigen Tausendstel« geht er nicht mehr ein. Wir werden in einem Folgeartikel unter dem Titel *Gemeinsame Geburtstage* dazu Stellung nehmen.

Mit $z = 365$ und $n = 60$ erhält VON MISES

$$E(X_3) = \frac{1}{365^2} \binom{60}{3} \left(1 - \frac{1}{365}\right)^{57} = 0,22,$$

woraus er schließt, dass bei vier bis fünf Gruppen von je 60 Personen im Mittel in einer dieser Gruppen doch ein dreifach besetzter Tag zu erwarten ist⁶, weil die Summe der entsprechenden Erwartungswerte dann den Wert 1 übersteigt.

Ausdrücklich bemerkt VON MISES noch:

»Auf die Veränderlichkeit der Geburtdichte mit der Jahreszeit ist dabei nicht Rücksicht genommen.«

D. h.: Die Verteilung der Geburtstage auf die 365 Tage eines Jahres wird als gleichmäßig angenommen. Damit werden Schaltjahre nicht berücksichtigt. Auch Mehrlingsgeburten wie Zwillinge, Drillinge usw. werden ausgeschlossen, weil in einer Bevölkerung mit vielen Mehrlingsgeburten die Wahrscheinlichkeiten für Mehrfachgeburtstage anwachsen.

Wir berechnen zusätzlich

$$E(X_0) = \frac{1}{365^{-1}} \binom{60}{0} \left(1 - \frac{1}{365}\right)^{60-0} = 309,60$$

$$E(X_1) = \frac{1}{365^0} \binom{60}{1} \left(1 - \frac{1}{365}\right)^{59} = 51,03$$

$$E(X_2) = \frac{1}{365} \binom{60}{2} \left(1 - \frac{1}{365}\right)^{58} = 4,14$$

$$E(X_4) = \frac{1}{365^3} \binom{60}{4} \left(1 - \frac{1}{365}\right)^{56} = 0,0086$$

und schließen daraus: Es ist im Mittel bei 60 Personen zu erwarten, dass an etwa 310 Kalendertagen keine Person Geburtstag hat, an ca. 51 Tagen genau eine Person und an ca. vier Tagen genau zwei Personen; mehr als zwei Geburtstage am gleichen Tag sind im Mittel nicht zu erwarten, weil die zugehörigen Erwartungswerte kleiner als 0,5 sind.

Beim Geburtstagsproblem fragt man aber nicht nach *genau* dreifach besetzten Zellen, sondern nach *mindestens* dreifach besetzten Zellen. Man muss also alle Erwartungswerte ab $s = 3$ addieren. Mit der

obigen Erkenntnis, dass $\sum_{s=0}^n E(X_s) = z$, erhalten wir

$$\sum_{s=3}^{60} E(X_s) = 365 - E(X_0) - E(X_1) - E(X_2) = 0,23,$$

einen Wert, der nur geringfügig größer als 0,22 ist, sodass die VON MISES'sche Folgerung auch für die Fragestellung nach mindestens einem Dreifachgeburtstag zutrifft.

Weiterhin stellt VON MISES fest, dass der Erwartungswert $E(X_3)$ für $n = 103$ schon annähernd gleich 1 ist,

»d. h., es ist im Durchschnitt anzunehmen, dass in einer Gesamtheit von 103 Personen einmal drei das gleiche Geburtsdatum haben.«

Die Rechnung bestätigt seine Zahl:

$$E(X_3) = \frac{1}{365^2} \binom{103}{3} \left(1 - \frac{1}{365}\right)^{100} = 1,009.$$

Abschließend behauptet VON MISES, dass $E(X_2)$ ungefähr gleich 1 wird für 29 Personen, d. h.,

»wer in seinem Bekanntenkreis rund 29 Geburtstage kennt, wird im Durchschnitt *einen* doppelt besetzten Tag finden.«

Auch hier bestätigt die Rechnung für $n = 28$ und $n = 29$ seine Behauptung:

$$E(X_2) = \frac{1}{365^1} \binom{28}{2} \left(1 - \frac{1}{365}\right)^{26} = 0,96 \text{ und}$$

$$E(X_2) = \frac{1}{365^1} \binom{29}{2} \left(1 - \frac{1}{365}\right)^{27} = 1,03.$$

Das *Geburstagsparadoxon* besagt jedoch, wie in 1 mitgeteilt, dass es in einer Gruppe von 23 Personen günstig ist, darauf zu wetten, dass an mindestens einem Kalendertag mindestens zwei der 23 Personen einen gemeinsamen Geburtstag feiern können. Wir fragen: Wie viele solche Doppelgeburtstage wird es im Mittel bei Gruppen von 23 Personen geben?

$$\text{Aus } E(X_s) = \frac{1}{z^{s-1}} \binom{n}{s} \left(1 - \frac{1}{z}\right)^{n-s}$$

erhalten wir für $n = 23$, $z = 365$ und $s = 2$ den Wert

$$E(X_2) = \frac{1}{365^{2-1}} \binom{23}{2} \left(1 - \frac{1}{365}\right)^{23-2} = 0,65.$$

Berechnen wir analog zu oben

$$\begin{aligned} \sum_{s=2}^{23} E(X_s) &= 365 - E(X_0) - E(X_1) = \\ &= 365 - 342,68 - 21,65 = 0,67, \end{aligned}$$

so ergibt sich auch nur ein geringfügig größerer Wert als 0,65. Wir stellen damit fest:

Ab einer Gruppengröße von $n = 23$ ist es zwar günstig, auf mindestens zwei Personen mit gemeinsamen Geburtstag zu *setzen*. Für $n = 23$ ist aber der Erwartungswert 0,65 für die Anzahl der Doppelgeburtstage aber immer noch kleiner als 1. Erst bei einer Betrachtung von zwei Gruppen mit jeweils 23 Personen ist die Summe der Erwartungswerte größer als 1, d. h., man kann erwarten, dass man mindestens eines dieser Gruppenpaare mit einem Doppelgeburtstag in einer davon findet, wenn man sehr oft zwei Gruppen mit je 23 Personen untersucht. Das ist im Einklang zur Wahrscheinlichkeit $\frac{1}{2}$ dafür, dass das Ereignis »Doppelgeburtstag« bei 23 Personen eintreten wird.

Wie VON MISES gezeigt hat, muss eine einzelne Gruppe mindestens 29 Personen umfassen, damit man im Mittel mindestens einen Tag mit genau einem Doppelgeburtstag bei ihr erwarten kann.

Anmerkungen

- 1 VON MISES verwendet anstelle von n den Buchstaben k und anstelle von z den Buchstaben n .
- 2 VON MISES verweist auf seine *Vorlesungen aus dem Gebiet der angewandten Mathematik*, Bd. 1 Wahrscheinlichkeitsrechnung, S. 409–452.
- 3 »Zufallszahlen« treten im täglichen Leben als PINs, TANs, bei Nummernschlössern, in Passwörtern usw. auf.
- 4 VON MISES hat eine andere Sprechweise in der Wahrscheinlichkeitstheorie als heute üblich. Wir formulieren seine Gedanken sinngemäß unter Verwendung des Begriffs der Zufallsgröße.
- 5 Bei VON MISES heißt es »Besetzungszahl« statt »Zufallsgröße«.
- 6 Diese und die folgenden zitierten Feststellungen finden sich auch in der Einleitung seines Aufsatzes.

Literatur

- Barth F./Haller R. (1998): Stochastik – Leistungskurs. München: Oldenbourg Schulbuchverlag.
- Diaconis, P./Mosteller, F. (1989): Methods für Studying Coincidences. In: *Journal of the American Statistical Association*, 84, 4, S. 853–861.
- Feller, W. (1950): An Introduction to Probability Theory and its Applications. Vol. 1. New York: John Wiley.
- Knuth, D. E. (1998): The Art of Computer Programming. Vol. 3 Sorting and Searching. 2nd edition. Upper Saddle River NJ u. a.: Addison-Wesley.
- 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.

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