Nochmals: Geburtstage und Maschinenausfälle *)

von D. WILKIE

übersetzt von Karl Röttel

BISSELL [1] bringt eine ziemlich detaillierte Abhandlung des Koinzidenzproblems am Beispiel des wiederholten Ausfallens in einer Gruppe von $n = 100$ Maschinen. Dies ist eines von vielen ähnlichen Problemen, deren wohl bekanntestes das Geburtstagsproblem ist. Er wählt dieses Problem aus, um das überraschende Ergebnis von 50 % Wahrscheinlichkeit dafür zu nennen, daß wenigstens 2 Personen einer Gruppe von 23 denselben Geburtstag besitzen.

Es ist lehrreich, dieses Resultat in einer Klasse zu veranschaulichen, indem die Schüler gebeten werden, ihren Geburtstag zu nennen und – wenn es zu wenig Personen in der Klasse sind – den der Eltern oder anderer Verwandter. Tatsächlich kann auf diese Weise die gesamte Wahrscheinlichkeitsfunktion experimentell bestimmt und mit den berechneten Werten verglichen werden, die man aus der Formel (vgl. Anmerkung 1)

$$P = 1 - \frac{n!}{(n-r)!\,n^r} \qquad (1)$$

erhält.

In (1) bedeuten

P: Wahrscheinlichkeit für zwei oder mehr Koinzidenzen,

n: Zahl der Objekte in der Gruppe (hier 365 Tage),

r: Anzahl der Versuche (hier die Zahl der Personen, deren Geburtstage verglichen werden).

Viele Taschenrechner sind der Gleichung (1) für hohe Werte von n (z.B. $n > 68$) nicht gewachsen. Ein Ausweg besteht darin, veröffentlichte Tabellen zu verwenden (beispielsweise reicht der Literaturhinweis [2] bis zu 200! und lg(1000!)). Ein anderer Ausweg ist

*) Originalartikel in "Teaching Statistics" (1981) Heft 1, Band 3: Birthdays and Breakdowns Revisited.

es, die Stirlingsche Formel für große n einzusetzen, die in der Form

$$n! = \sqrt{2\pi n}\cdot n^n \cdot e^{-n} \qquad (2)$$

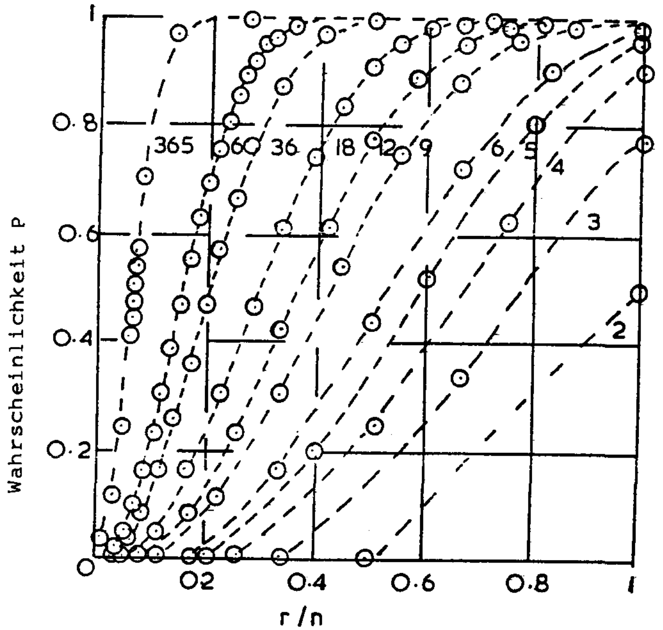
geschrieben werden kann. Setzt man Gleichung (2) in (1) ein, erhält man

$$P = 1 - \left(\frac{n}{n-r}\right)^{n-r+\frac{1}{2}} \cdot e^{-r}. \qquad (3)$$

Außerdem kann das Koinzidenzproblem in verschiedenen Glücksspielen, im Ausfall von Teilen in Maschinenanlagen und im Zusammentreffen von Heißstellen in Brennelementen der Kernreaktoren vorkommen.

Aus diesem Grunde ist es nützlich, P für eine große Zahl von Werten für n auszurechnen. Dies wurde durchgeführt für n vom niedrigsten Wert 2 (für $n = 1$ kann es keine Koinzidenzen geben) bis 365 und für r/n im Bereich von 0 bis 1 (denn bei $\frac{r}{n} > 1$ wird $P = 1$).

Die Ergebnisse sind in Figur 1 dargestellt, wobei aus Gründen der Übersichtlichkeit nicht alle Punkte für höhere n-Werte aufgenommen wurden. Aus eben diesem Grund sind auch die Punkte durch Kurven verbunden, obwohl nur ganze Zahlenwerte für r möglich sind.



Figur 1. Koinzidenzwahrscheinlichkeiten bei r Versuchen in einer Gruppe von n Elementen.

Der Fall $n = 2$ findet Anwendung beim Werfen einer Münze. Die zwei Elemente der Gruppe sind "Zahl" und "Wappen". Ein Wurf liefert $P = 0$, da keine Koinzidenzen möglich sind. Drei Würfe liefern $P = 1$, da man wenigstens 2mal Zahl oder 2mal Wappen erhält (dieser Fall, für den $r/n = 1,5$ ist, ist im Graphen nicht enthalten). Bei 2 Würfen kann eines von 4 Ereignissen eintreten, nämlich 2mal Zahl, 2mal Wappen, Zahl und Wappen oder Wappen und Zahl, woraus hergeleitet werden kann, daß $P = 1/2$ für ein Doppel (2 Z oder 2 W) ist.
Der Lehrer kann Anwendungen für weitere Werte von n betrachten. So ist $n = 6$ beim Würfel verwendbar. Das zweimalige Werfen eines Würfels oder das gleichzeitige Werfen eines Paares von Würfeln (d.h. $r = 2$ und $r/n = 1/3$) hat die Wahrscheinlichkeit 0,166... bzw. 1/6 für das Auftreten irgendeines Doppels. Deshalb ist die Wahrscheinlichkeit für ein bestimmtes Doppel (etwa den Sechserpasch) ein Sechstel der genannten Wahrscheinlichkeit, also 1/36, ein Ergebnis, das leicht auf andere Weise gefunden werden kann. Für andere Werte von r bis hinauf zu 6, etwa $r = 4$, gibt die Figur 1 die Wahrscheinlichkeiten an, zwei- oder mehrmals irgendeine Augenzahl im Rahmen der r Würfe zu erhalten, wobei also die Wiederholungen nicht unmittelbar aufeinander folgen müssen. Diese Ergebnisse können experimentell überprüft werden.
Der Fall $n = 20$ trifft für einen ungeschickten Pfeilwerfer zu, vorausgesetzt, wir lassen "bulls", Doppelzählungen, Dreifachzählungen und Fehlwürfe unberücksichtigt (vgl. Anmerkung 2). Deutliche Abweichung von der erwarteten Verteilung ist ein Beweis für Geschicklichkeit und kann experimentell geprüft werden, indem die Zahl der Würfe von 2 bis 20 variiert wird. Der Lehrer könnte die gleichen Überlegungen auf das "zufällige" Zuspätkommen von Schülern anwenden lassen!
Angesichts des überraschenden Ergebnisses des Geburtstagsproblems lohnt es sich, das Verhalten der r/n-Werte, die jeweils 50% Wahrscheinlichkeit ergeben, für alle n-Werte zu untersuchen. Diese Werte können aus einer vergrößerten Abwandlung der Figur 1 abgelesen werden. Die Ergebnisse sind in Figur 2 dargestellt in einer Form, daß ein näherungsweise linearer Zusammenhang entsteht, der aber auch eine theoretische Grundlage hat.

Der Graph zeigt, daß für $n = 9$ durchschnittlich ungefähr 4 Versuche erforderlich sind, bis ein Zusammentreffen eintritt, nur ungefähr 5 für $n = 18$ und 7 für $n = 36$. Der Wert für r/n fällt rapide, wenn n zunimmt, und ist nur 0,06 bei $n = 365$. Eine brauchbare Approximation für Figur 2 ist $r/\sqrt{n} = 1,25$.
Obwohl es möglich ist, r/n aus Figur 1 für jede vorgelegte Wahrscheinlichkeit zu entnehmen, ist eine rechnerische Lösung nützlich. Wenn wir in Gleichung (3)

$$1 - P = \left(1 - \frac{r}{n}\right)^{-n+r-\frac{1}{2}} \cdot e^{-r}$$

die natürlichen Logarithmen beider Seiten bilden und die Entwicklung

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} \cdots$$

verwenden, erhalten wir mit einigen Umformungen

$$\begin{aligned} \ln(1-P) &= \left(-n+r-\tfrac{1}{2}\right)\left(-\frac{r}{n} - \frac{r^2}{2n^2} - \frac{r^3}{3n^3} - \cdots\right) - r \\ &= -\frac{r^2}{2n}\left[\frac{r-1}{r} + \frac{2r-3}{6r}\left(\frac{r}{n}\right) + \frac{r-2}{6r}\left(\frac{r}{n}\right)^2 + \frac{2r-5}{20r}\left(\frac{r}{n}\right)^3 + \cdots\right] \\ &= -\frac{r^2}{2n}, \text{ wenn } n \gg r \gg 1. \end{aligned}$$

Deshalb ist

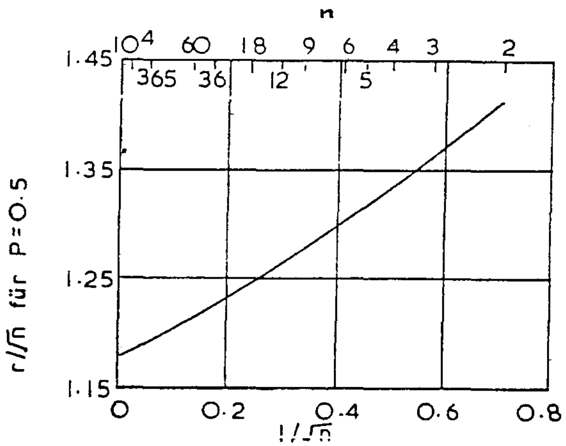
$$\frac{r}{\sqrt{n}} = \sqrt{\ln\left(\frac{1}{1-P}\right)^2} \quad \text{für } n \gg r \gg 1. \qquad (4)$$

Gleichung (4) ist verwendet worden, um $r/\sqrt{n}$ für einzelne Werte von P zu berechnen (Tabelle 1). Man sieht, daß $r/\sqrt{n} = 1,1774$ für $P = 0,5$ ist - in Übereinstimmung mit den Daten von Figur 2.
Obwohl sie für große n hergeleitet wurde, liefert Gleichung (4) für einen weiten Bereich von n-Werten befriedigende Ergebnisse, die sich höchstens um 1 vom genauen Wert unterscheiden, der aus Gleichung (1) errechnet wurde. Dabei ergeben sich für kleines P zu niedrige, für großes P zu hohe Werte.
So bildet Gleichung (4) eine schnelle und genaue Methode, die Wahrscheinlichkeiten von Koinzidenzen zu berechnen.

Tabelle 1.

Werte von $r/\sqrt{n}$ für große n zu vorgegebenen Wahrscheinlichkeiten

P	$r/\sqrt{n}$ für $n \gg r \gg 1$
0	0
0.0001	0.01414
0.001	0.0447
0.01	0.1418
0.05	0.320
0.1	0.459
0.2	0.668
0.3	0.8576
0.4	1.0108
0.5	1.1774
0.6	1.3537
0.7	1.5517
0.8	1.794
0.9	2.146
0.95	2.448
0.99	3.035
0.995	3.255
0.998	3.526
0.999	3.717
0.9995	3.899
0.9999	4.292



Figur 2. Anzahl der Versuche als Funktion der Größe der Gruppe für P = 0,5.

Tabelle 2.

Werte von r für verschiedene n und P

n	Nächst höhere ganze Zahl von r für P =					
	0.1	0.3	0.5	0.7	0.9	0.95
6	2	3	4(3)	4	5(6)	6
9	2	3	4	5	7	7(8)
12	3(2)	4(3)	5	6	8	8(9)
18	3(2)	4	6(5)	7	9(10)	10(11)
36	4(3)	6	8	10	13	15
60	5(4)	7	10	13	17	19
365	10(9)	17	23	30	41	47(48)

Die Werte, die Gleichung (4) liefert, sind in Klammern angegeben, wenn sie von denen aus Gleichung (1) abweichen.

Literaturhinweise:

[1] BISSELL, A.F.: Breakdowns and Birthdays. Teaching Statistics 2.1 (1980), p. 15. Übersetzung in 'Stochastik in der Schule' Band 1 / Nummer 3, S. 15 - 18.

[2] PEARSON, E.S. and HARTLEY, H.O.: Biometrika Tables for Statisticians, Vol. 1. Cambridge University Press.

<u>Anmerkung 1:</u>

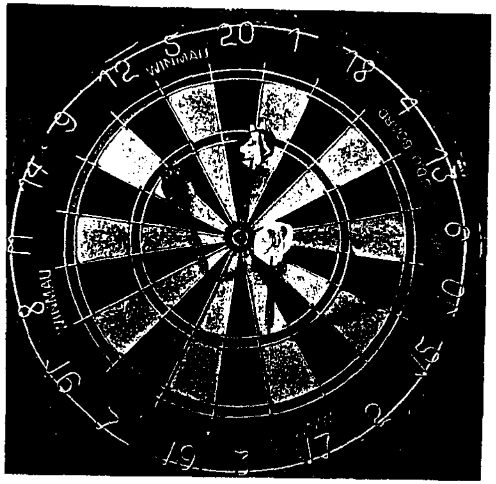
Es gibt $V_{mW} = n^r$ Möglichkeiten für die Zuteilung von Zahlen (aus n) auf die r Gruppenmitglieder. Es gibt $V_{oW} = \frac{n!}{(n-r)!}$ Möglichkeiten, daß alle r Mitglieder verschiedene Nummern haben. Daraus errechnen wir (es handelt sich um gleichwahrscheinliche Fälle) die Wahrscheinlichkeit, daß alle Mitglieder verschiedene Nummern haben zu

$$P(\text{verschiedene Nummern}) = \frac{n!}{(n-r)!\, n^r}.$$

Die Gegenwahrscheinlichkeit 1 - P(versch.) ist dann die Wahrscheinlichkeit, daß mindestens eine Nummer wenigstens doppelt vorkommt.

Anmerkung 2:

Beim angesprochenen englischen Pfeilwurfspiel werden Pfeile ("darts") auf eine Kreisscheibe ("clock-board") geworfen. Die Zielscheibe ist in 20 Sektoren aufgeteilt, die mit den Zahlen in dieser Reihenfolge belegt sind: 20, 1, 18, 4, 13, 6, 10, 15, 2, 17, 3, 19, 7, 16, 8, 11, 14, 9, 12, 5. Die 20 steht ganz oben, rechts folgt die niedrige Zahl 1 usw. Im Zentrum befinden sich ein Kreis ("bull's-eye") und ein Kreisring. Das Bullauge zählt 50 Punkte, der Treffer in den Ring bringt 25 Punkte.



Eine Anzahl von Kreisen teilt die 20 Sektoren nun so, daß ein äußerer Ring ("doubles", zweifach zählend) ein breiter (einfach zählend), wieder ein schmaler ("trebles", dreifach zählend) und nochmals ein einfach zählender breiter Ring entstehen. Welcher Wert für die einzelnen Sektoren gilt, gibt die Zahl am Äußeren der Scheibe an.
Die Scheibe ist in 1,70 m Höhe anzubringen, die Entfernung von der Abwurflinie muß 2,37 m betragen. Das Spiel beginnt mit einem Ausgangsstand von 501 Punkten, von denen man möglichst schnell auf exakt null Punkte herunterkommen muß.
Für das Beispiel im vorangegangenen Text setzt man also voraus, daß es nur die 20 Sektoren gibt, die die Nummern 1 bis 20 tragen, und jeder Wurf in einen Sektor trifft.

Anmerkung 3:

Dem Übersetzer teilte Prof. G. Fillbrunn eine problematische Stelle des Aufsatzes mit, die sich der Leser besonders ansehen **möge. Diese ist am Anfang der vierten Seite der Übersetzung.**
Das ursprüngliche Experiment besteht aus dem r-maligen Ziehen mit Zurücklegen aus einer Urne mit n Kugeln. Es wird dabei nach demjenigen r gefragt, für das man (mindestens) mit der Wahrscheinlichkeit 0,5 mindestens 2 gleiche Kugeln zieht. Dagegen wird an der hier kritisierten Stelle aus der Urne mit n Kugeln so lange gezogen, bis sich zwei gleiche Kugeln ergeben. Nach dem Wortlaut des Textes wird nach dem Erwartungswert μ der Anzahl der Ziehungen bis zum Abbruch gefragt.
Daß dies zwei verschiedene Probleme sind, zeigt das Beispiel für $n = 2$. Für $r = 2$ gilt $p = 0{,}5$. Dagegen erhält man $\mu = 2{,}5$ $(\neq r)$.
Was meinen Sie dazu?