zu übertragen. Für diese gezielte Übertragung muss das Gen, welches zu dem agronomischen Vorteil führt, bekannt sein. Das verantwortliche Gen kennt man jedoch nur in einigen Fällen. Schwierig wird es vor allem dann, wenn eine Merkmalsausprägung durch mehrere Gene bestimmt wird.

Dr. KIRSTEN SCHLÜTER, Leibniz-Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel, Olshausenstr. 62, 24098 Kiel, schlueter@ipn.uni-kiel.de, ist wissenschaftliche Angestellte in der Abt. Didaktik der Biologie. Einer ihrer Arbeitsschwerpunkte ist die Entwicklung und Evaluation von Unterrichtsmaterial im Bereich der Bio- und Gentechnik.

HANS HUMENBERGER

Additive Zahlzerlegungen und Lotto

Ausgehend von einer elementaren kombinatorischen Fragestellung soll ein relativ einfaches und ein relativ schwieriges Lotto-Problem untersucht werden: Verteilung und Erwartungswert der Länge des kleinsten bzw. größten »spacing« bei einer Lottoziehung. »spacings« sind die zwischen zwei benachbarten Lotto-Gewinnzahlen liegenden »unbenutzten Zahlenblöcke« bzw. »Löcher«. Bei beiden Aufgaben wird ein Computer-Algebra-System eingesetzt.

1 Additive Zahlzerlegungen

Problem: Auf wie viele Arten kann eine natürliche Zahl $n \geq 1$ in k natürliche Summanden größer oder gleich 1 zerlegt werden, wobei die Reihenfolge der Summanden eine Rolle spielen soll? ($10 = 3 + 3 + 4$ bzw. $10 = 4 + 3 + 3$ sollen also als zwei verschiedene Zerlegungen der 10 in $k = 3$ Summanden angesehen werden.)

Für kleine Werte von n und k findet man leicht die Anzahlen:

5 in $k = 3$ Summanden zerlegen ergibt sechs Möglichkeiten:

$$1+1+3, 1+3+1, 3+1+1, 2+2+1, 2+1+2, 1+2+2;$$

6 in $k = 2$ Summanden zerlegen ergibt fünf Möglichkeiten:

$$1+5, 5+1, 2+4, 4+2, 3+3.$$

Eine elementare Vorstellung, die zur allgemeinen Lösung führt, ist die folgende:

$$n = n_1 + n_2 + ... + n_k \text{ (alle } n_i \geq 1)$$

bedeutet einen Stab der Länge n (man denke dabei an einen Zollstock mit cm-Strichen) in k Teilabschnitte zu teilen; dazu müssen offensichtlich $k-1$ Teilungspunkte eingezeichnet (»Kerben geschlagen«) werden. Zur Auswahl für diese möglichen Teilungspunkte stehen alle »inneren« cm-Markierungen: 1, ..., $n - 1$, so dass klar ist, dass es dafür genau

$$\binom{n-1}{k-1}$$

Möglichkeiten gibt.

Eine weitere äquivalente Formulierung (neben Zerlegung einer natürlichen Zahl bzw. Teilung eines Stabes) ist: n ununterscheidbare Kugeln auf k Fächer so verteilen, dass in jedem Fach mindestens eine Kugel landet. Die Anzahl der Kugeln im Fach i entspricht dabei dem i-ten Summand.

Diese Vielfalt an Konkretisierungen (Modellen) ist im Unterricht besonders wichtig! Wir haben damit folgenden

Satz: *Es gibt* $\binom{n-1}{k-1}$ *Möglichkeiten, mit jeweiliger Beachtung der Reihenfolge*

- *einen Stab mit ganzzahliger Länge n in k Teilstücke mit ganzzahliger Länge zu zerlegen;*
- *eine natürliche Zahl n in k natürliche Summanden größer oder gleich l zu zerlegen;*
- *n ununterscheidbare Kugeln auf k Fächer so zu verteilen, dass in jedem Fach mindestens eine Kugel landet.*

Bemerkungen:

(1) Zur Lösung des Problems könnten auch »Kombinationen mit Wiederholung« herangezogen werden; obige Formulierung bedeutet ja: in jedes Fach eine Kugel geben und die restlichen $n - k$ Kugeln auf die k Fächer irgendwie[1] verteilen: Die entsprechende Formel (»Kombinationen mit Wiederholung«) ergibt

MNU 56/6 (1. 9. 2003) Seite 334–338 ISSN 0025-5866 © Bildungsverlag EINS – DÜMMLER · TROISDORF

$$\binom{k+(n-k)-1}{n-k} = \binom{n-1}{k-1}$$

Möglichkeiten. Hier sollen jedoch »Kombinationen mit Wiederholung« absichtlich vermieden werden, da sie in vielen Stochastikkursen in der Schule entweder gar nicht erwähnt oder zumindest nicht begründet werden.

(2) Obige elementare Sichtweise ist nicht nur dazu geeignet, »Kombinationen mit Wiederholung« bei diesem konkreten Problem zu vermeiden, sondern könnte umgekehrt dazu benutzt werden, die Formel für »Kombinationen mit Wiederholung« – bei Bedarf – allgemein zu begründen (neben den anderen »klassischen« Begründungsmöglichkeiten).

2 Zwei Lottoprobleme

Wir wollen die obige Erkenntnis nun auf zwei Lottoprobleme anwenden. Bei jeder Lotto-Ausspielung (z. B. deutsches Lotto: »6 aus 49«) werden 6 Zahlen aus $\{1, ..., 49\}$ gezogen (ohne Zurücklegen). Die Reihenfolge der Ziehungen spielt keine Rolle, im Anschluss an die Ziehung werden die sechs Gewinnzahlen ohnehin der Größe nach geordnet; der Index $i = 1, ..., 6$ einer Gewinnzahl t_i soll angeben, die wie vielte Zahl sie in der bereits geordneten Darstellung ist:

$1 \leq t_1 < t_2 < t_3 < t_4 < t_5 < t_6 \leq 49$ bzw.

$0 < t_1 < t_2 < t_3 < t_4 < t_5 < t_6 < 50$.

Die Abstände (Differenzen) $t_i - t_{i-1}$ paarweise aufeinander folgender Gewinnzahlen heißen in der englischen Literatur »spaces« oder »spacings«, wobei es dabei einige interessante Fragestellungen gibt, z. B. die Verteilung und den Erwartungswert des kleinsten bzw. größten »spacing« zu bestimmen.

Diese Abstände (vermindert um l) sind die Anzahl der zwischen t_{i-1} und t_i liegenden »Nicht-Gewinnzahlen« (Längen der in der Ziehung »unbenutzten Blöcke«): wenn z. B. $t_3 = 15$ und $t_4 = 20$ ist, so heißt dies ja, dass der Block der vier Zahlen $\{16,17,18,19\}$ unbenutzt blieb $(t_4 - t_3 = 5)$.

Vor t_1 und nach t_6 gibt es i. a. auch unbenutzte Blöcke, so dass es nahe liegt, die Randwerte $t_0 := 0$ und $t_7 := 50$ in die Betrachtung mit einzubeziehen[2], um auch die Länge der Randblöcke mittels Differenzen beschreiben zu können.

Die Abstände (Differenzen) D_i definieren wir dann in nahe liegender Weise:

$$D_i := t_i - t_{i-1} \text{ für } i = 1, \ldots, 7.$$

Damit ist die Analogie zum Obigen schon evident – in den drei obigen Kontexten ausgedrückt:

- Durch die Lottoziehung werden 6 Kerben (Markierungen) t_i $(i = 1, ..., 6)$ in einen Stab der Länge 50 geschlagen, d. h. der Stab wird dadurch in sieben Abschnitte zerlegt,
- die Zahl 50 wird in 7 Summanden zerlegt, die nichts anderes sind als die Differenzen[3] D_i, $(i = 1, ..., 7)$: $50 = D_1 + D_2 + ... + D_6 + D_7$.

Man könnte auch sagen:

- 50 ununterscheidbare Kugeln werden so in sieben Fächer verteilt, dass in jedem Fach mindestens eine Kugel ist (die Anzahl der Kugeln im Fach i entspricht dem Summand bzw. der Differenz D_i).

Unter den Differenzen (Summanden, »spacings«) D_i gibt es natürlich eine kleinste und eine größte – mehrere Differenzen D_i können gleichzeitig dieses Minimum (Maximum) annehmen: z. B. in $50 = 12 + 8 + 1 + 12 + 1 + 7 + 9$ tritt sowohl das Minimum 1 als auch das Maximum 12 zweifach auf (diese Differenzen würden übrigens folgenden Gewinnzahlen entsprechen: $t_1 = 12$, $t_2 = 20$, $t_3 = 21$, $t_4 = 33$, $t_5 = 34$, $t_6 = 41$).

Wir definieren:

$m := \min D_i$ und $M := \max D_i$ für $i = 1, ..., 7$,

wobei m und M (kleinste bzw. größte Differenz) von der konkreten Ziehung abhängige Zufallsgrößen sind. Im Folgenden wollen wir deren Verteilungen bzw. Erwartungswerte berechnen.

2.1 Verteilung und Erwartungswert der kleinsten Differenz m

Der größtmögliche Wert von m ist 7, denn wenn alle $D_i \geq 8$ wären, so wäre $\sum D_i \geq 56 > 50$ - also schon zu viel, d. h. $m \in \{1, ..., 7\}$.

Bei wie vielen der insgesamt $\binom{49}{6}$ möglichen Lotto-Ziehungen hat die kleinste Differenz m die Werte 1, ..., 7? Dies entspricht: wie viele additive Zerlegungen der Zahl 50 in 7 natürliche Summanden gibt es mit kleinstem Summand $m = 1, ..., 7$? Wir wollen also die Werte

$h(j) :=$ Anzahl aller Lotto-Ziehungsergebnisse mit. $m = j$ $(j = 1, ...,7)$ bestimmen.

Die Verteilung der Zufallsgröße m ist dann durch

$$P(m = j) = \frac{h(j)}{\binom{49}{6}} \text{ mit } j = 1, ..., 7$$

und deren Erwartungswert durch

$$\mathrm{E}(m) = \sum_{j=1}^{7} \frac{j h(j)}{\binom{49}{6}} \text{ gegeben.}$$

Wir erhalten die Werte $h(j)$, indem wir zuerst Überlegungen zu

$f(j) :=$ Anzahl aller Lotto-Ziehungsergebnisse mit $m \geq j$ $(j = 1, ...,7)$

anstellen; dann ist nämlich $h(j) = f(j) - f(j + 1)$ und wir können Verteilung und Erwartungswert berechnen.

[1] Ohne Zusatzbedingung, dass in jedem Fach mindestens eine Kugel landen muss.

[2] Diese sind natürlich keine Gewinnzahlen!

[3] In diesem Sinn entsprechen hier den Summanden also Differenzen.

Die Werte $f(j)$ sind leicht zu erhalten, denn $m \geq j$ bedeutet ja: $D_i \geq j$ für alle $i = 1, ..., 7$, d. h. jeder Summand beträgt mindestens j:
$D_1 \geq j \wedge D_2 \geq j \wedge D_3 \geq j \wedge D_4 \geq j \wedge D_5 \geq j \wedge D_6 \geq j \wedge D_7 \geq j$.
Wie viele Möglichkeiten gibt es dafür?
Wir wollen das Problem mit obigem Zerlegungssatz einer Zahl in Summanden lösen, bei dem jeder Summand ja mindestens 1 beträgt, wobei hier in diesem Zusammenhang m. E. die Vorstellung mit den Kugeln und Fächern besonders nahe liegt:
Wir denken uns zunächst in jedem der 7 Fächer $j - 1$ Kugeln. Wenn wir nun die »restlichen« $50 - 7(j - 1) = 57 - 7j$ Kugeln noch auf diese Fächer in unserer speziellen Weise aufteilen (in jedes Fach kommt mindestens eine weitere Kugel), so befinden sich in jedem Fach – genau wie verlangt – mindestens j Kugeln.
Wir brauchen also nur zu zählen, wie viele Möglichkeiten es gibt, $57 - 7j$ Kugeln auf 7 Fächer aufzuteilen, so dass in jedem Fach mindestens eine Kugel landet; laut oben gibt es dafür genau $\binom{56-7j}{6}$ Möglichkeiten, d. h.

$$f(j) = \binom{56-7j}{6} \quad \text{mit } j = 1, ..., 7.$$

Man beachte: für $j = 8$ erhält man $f(8) = 0$, was in $h(7) = f(7) - f(8)$ Bedeutung hat. Mit einem Computer-Algebra-System können nun die Werte $h(j) = f(j) - f(j + 1)$ für $j = 1, ..., 7$ berechnet werden, deren Summe natürlich

$$\binom{49}{6} = 13\,983\,816$$

ergeben muss (die zugehörigen Werte $P(m = j)$ sind in Tabelle 1 mit zwei signifikanten Ziffern angegeben.
Nun kann auch der Erwartungswert mit berechnet werden:

$$\mathrm{E}(m) \approx 1{,}52,$$

ein auf dem ersten Blick vielleicht erstaunlich kleiner Wert. Beachtet man aber, dass z. B. die Wahrscheinlichkeit $P(m \leq 2) \approx 0{,}884$ bzw. $P(m \leq 3) \approx 0{,}973$ beträgt, d. h. dass mehr als 97 % aller möglichen Lotto-Ziehungsergebnisse kleinste Differenz $m = 1,2$ oder 3 haben, so ist die Größe des obigen Erwartungswertes klarer!
Bemerkung: Hätte man kein CAS verwendet, so könnte der Term zur Berechnung des Erwartungswertes wegen $h(j) = f(j) - f(j + 1)$ noch etwas vereinfacht werden (»teleskopartige« Summe):

$$\mathrm{E}(m) = \frac{1}{\binom{49}{6}} \sum_{j=1}^{7} j \cdot [f(j) - f(j+1)]$$

$$= \frac{f(1) - f(2) + 2 \cdot [f(2) - f(3)] + 3 \cdot [f(3) - f(4)] + ... + 7 \cdot [f(7) - \overset{0}{\overbrace{f(8)}}]}{\binom{49}{6}}$$

$$= \frac{f(1) + f(2) + f(3) + ... + f(7)}{\binom{49}{6}}$$

$$= \frac{\binom{49}{6} + \binom{42}{6} + \binom{35}{6} + \binom{28}{6} + \binom{21}{6} + \binom{14}{6} + \binom{7}{6}}{\binom{49}{6}}$$

j	$h(j)$	$P(m = j) = \frac{h(j)}{13983816}$
1	8 738 030	0,62
2	3 622 626	0,26
3	1 246 420	0,089
4	322 476	0,023
5	51 261	0,0037
6	2 996	0,00021
7	7	0,00000050

Tab. 1. Die Verteilung von m bei »6 aus 49«

2.2 Verteilung und Erwartungswert der größten Differenz M

Dieses Problem kann zwar prinzipiell gleich angegangen werden, ist aber letztlich doch um einiges schwieriger; dies wird an zwei Stellen besonders deutlich werden.
Zunächst findet man wieder leicht, dass die größte Differenz $M \in \{8, ..., 44\}$ sein muss, denn wenn alle $D_i \leq 7$ wären, so wäre $\sum D_i \leq 49 < 50$ (zu klein!); andererseits kann kein $D_i \geq 45$ sein, denn $45 + 1 + 1 + 1 + 1 + 1 + 1 = 51 > 50$ (zu groß!).
Bei wie vielen der insgesamt $\binom{49}{6}$ möglichen Lotto-Ziehungen (additive Zerlegungen der Zahl 50) nimmt M die Werte 8, ..., 44 an?
Es sei
$H(j) := $ Anzahl aller Lotto-Ziehungsergebnisse mit $M = j \quad (j = 8, ..., 44)$.
Die Verteilung der Zufallsgröße M ist dann durch

$$P(M = j) = \frac{H(j)}{\binom{49}{6}} \quad \text{mit } j = 8, ..., 44$$

und deren Erwartungswert durch

$$\mathrm{E}(M) = \sum_{j=8}^{44} \frac{jH(j)}{\binom{49}{6}} \quad \text{gegeben.}$$

Wir wenden zunächst wieder jenen Trick an, der sich auch oben bewährt hat: Wenn wir
$F(j) := $ Anzahl aller Lotto-Ziehungsergebnisse mit $M \geq j \quad (j = 8, ..., 44)$
gefunden haben, dann ist analog $H(j) = F(j) - F(j + 1)$ und unser Problem gelöst.
Diese Anzahlen $F(j)$ sind nun aber nicht mehr so leicht wie obige $f(j)$ zu erhalten. Außerdem sind es nun viel mehr mögliche Werte für j, so dass ein konkreter Überblick über die Werte und die konkrete Berechnung des Erwartungswertes ohne CAS nicht mehr sinnvoll ist.
$M \geq j$ bedeutet, dass mindestens ein $D_i \geq j \ (i = 1, ..., 7)$ ist d. h., mindestens ein Summand beträgt mindestens j (es dürfen nicht alle $D_i < j$ sein):

$$D_1 \geq j \vee D_2 \geq j \vee D_3 \geq j \vee D_4 \geq j \vee D_5 \geq j \vee D_6 \geq j \vee D_7 \geq j.$$

Wie viele Möglichkeiten gibt es dafür? Wir haben die Mächtigkeit einer Vereinigungsmenge zu zählen (mit

$\{D_i \ge j\}$ ist die Menge aller Lotto-Ziehungsergebnisse mit $D_i \ge j$ gemeint:

$$\{M \ge j\} = \{D_1 \ge j\} \cup \{D_2 \ge j\} \cup \{D_3 \ge j\} \cup \{D_4 \ge j\} \cup \{D_5 \ge j\} \cup \{D_6 \ge j\} \cup \{D_7 \ge j\},$$

was natürlich an die Ein-Ausschaltformel[4] erinnert:

$$F(j) = |\{M \ge j\}| = \binom{7}{1}|\{D_1 \ge j\}| - \binom{7}{2}|\{D_1, D_2 \ge j\}| + \binom{7}{3}|\{D_1, D_2, D_3 \ge j\}| - \binom{7}{4}|\{D_1, D_2, D_3, D_4 \ge j\}| + \binom{7}{5}|\{D_1, \ldots, D_5 \ge j\}| - \binom{7}{6}|\{D_1, \ldots, D_6 \ge j\}| + \binom{7}{7}|\{D_1, \ldots, D_7 \ge j\}|,$$

wobei wir diese Werte $F(j)$ für alle möglichen $j = 8, \ldots, 44$ brauchen [für die Berechnung der Werte $H(j) = F(j) - F(j+1)$]. Wenn wir nun die Anzahlen

$Z(j, i) := |\{D_1, \ldots, D_i \ge j\}|$ für alle $j = 8, \ldots, 44$ bzw. $i = 1, \ldots, 7$ zählen ~~wollen~~ können, so ist unser Problem gelöst, denn wir haben dann:

$$F(j) = \sum_{i=1}^{7} (-1)^{i=1}\binom{7}{i} Z(j, i) \quad \text{mit } j = 8, \ldots, 44. \qquad (1)$$

Die Zählung von $Z(j, 1) = \{D_1 \ge j\}|$ gelingt relativ einfach für alle möglichen $j = 8, \ldots, 44$: Wir geben ins 1. Fach zunächst $j - 1$ Kugeln und in die anderen keine Kugel. Die restlichen $50 - (j - 1) = 51 - j$ Kugeln können wir nun auf die gewohnte Weise auf die 7 Fächer verteilen (in jedem Fach landet mindestens eine weitere Kugel), wofür es

$$\binom{50-j}{6} = Z(j, 1) \quad \text{mit } j = 8, \ldots, 44$$

Möglichkeiten gibt. Die so definierten Werte $Z(j, 1)$ können in der Berechnung der Werte $F(j)$ nach Gleichung (1) für alle $j = 8, \ldots, 44$ herangezogen werden (für das größtmögliche $j = 44$ ergibt sich

$Z(44, 1) = \binom{6}{6} = 1$).

Die Zählung von $Z\{j, 2) = |\{D_1, D_2 \ge j\}|$ ist formal genau so einfach:
Wir geben ins 1. und ins 2. Fach zunächst jeweils $j - 1$ Kugeln und in die anderen keine Kugel. Für die restlichen $50 - 2(j - 1) = 52 - 2j$ Kugeln haben wir nun

$$\binom{(52) - 2j}{6} = Z(j, 2) \qquad (2)$$

Möglichkeiten, sie auf die 7 Fächer zu verteilen (in jedem Fach soll wieder mindestens eine weitere Kugel landen):
$Z(j, 2)$ ist für $j = 8, \ldots, 44(?)$ zu berechnen. Die Schwierigkeit ist durch das Fragezeichen bei 44 angedeutet: Natürlich können nicht in zwei Fächern gleichzeitig 44 Kugeln sein, wenn nur insgesamt 50 Kugeln zu verteilen sind!
Bei $Z(j, 2)$ sind nur die Werte $j = 8, \ldots, 22$ inhaltlich sinnvoll: wenn in zwei Fächern je 23 Kugeln wären, so hätten wir $23 + 23 + 1 + 1 + 1 + 1 + 1 = 51 > 50$! Auch an der Formel (2) ist zu erkennen, dass sich bei $j = 23$ der Wert 0 ergibt: $Z(23,2) = \binom{5}{6} = 0$ – ein »sinnvoller« Wert, denn für $j > 22$ gibt es eben keine Möglichkeit mehr, dass in zwei Fächern mindestens j Kugeln sind. Dasselbe (der sinnvolle Wert 0) ergibt sich auch für $j = 24$ und $j = 25$.
Nun kommt aber eine kleine »Misere«: Für $j = 26, \ldots, 44$ ergibt sich für $Z(j, 2)$ (aufgrund der Definition von Binomialkoeffizienten) leider nicht der einzig richtige Wert 0, sondern z. B.:

$$Z(26, 2) = \binom{-1}{6} = \frac{(-1)\cdot(-2)\cdot(-3)\cdot(-4)\cdot(-5)\cdot(-6)}{6\cdot5\cdot4\cdot3\cdot2\cdot1} = 1 \ne 0$$

$$Z(30, 2) = \binom{-9}{6} = \frac{(-9)\cdot(-10)\cdot(-11)\cdot(-12)\cdot(-13)\cdot(-14)}{6\cdot5\cdot4\cdot3\cdot2\cdot1} = 3003 \ne 0$$

D. h. die Werte $Z(j, 2)$ wären bei der Berechnung der $F(j)$ nach (1) nicht mehr für alle $j = 8, \ldots, 44$ verwendbar, sondern nur noch für $j = 8, \ldots, 22$ (25) – Fallunterscheidungen je nach j wären nötig!
Um in jedem Fall bei »zu großem j« den einzig sinnvollen Wert 0 für $Z(j, 2)$ – und allgemein für $Z(j, i)$ – zu erreichen, führen wir einen »nichtnegativen Binomialkoeffizienten« (vgl. [1]) ein , wodurch wir von weiteren (lästigen) Fallunterscheidungen verschont bleiben[5].

Definition: $$\binom{n}{m}^+ := \begin{cases} \binom{n}{m} & \text{für } n \ge 0 \\ 0 & \text{für } n < 0 \end{cases}$$

Auch und gerade mit CAS lassen sich solche Definitionen durchführen und in den Berechnungen weiter verwenden!
In der Berechnung der Werte $Z(j, 2)$ (und auch für die anderen $i = 1, \ldots, 7$ in $Z(j, i)$)können wir dadurch »beruhigt« alle $j = 8, \ldots, 44$ einbeziehen und brauchen uns über dieses Problem keine Gedanken mehr zu machen:

$$Z(j, 2) := \binom{51 - 2j}{6}^+ \quad \text{für } j = 8, \ldots, 44.$$

Der allgemeine Fall $Z(j, i)$: Wir geben in die ersten i-Fächer zunächst jeweils $j - 1$ Kugeln und in die anderen keine Kugel. Für die restlichen $50 - i(j - 1)$ Kugeln haben wir nun

$$\binom{49 - i(j-1)}{6}$$

Möglichkeiten, sie auf die 7 Fächer zu verteilen (in jedem Fach soll wieder mindestens eine weitere Kugel landen); die Werte $Z(j, i)$ sind durch die oben definierten

---

4 Auch »Prinzip der Inklusion-Exklusion« genannt. Dies ist der erste Punkt, der doch um einiges schwieriger ist als beim obigen Pendant *m*. Dort hatten wir die Möglichkeit eine Durchschnittsmenge zu zählen, was leichter ist.

5 Dies ist der zweite Punkt, der doch um einiges schwieriger ist als bei der kleinsten Differenz *m*.

nichtnegativen Binomialkoeffizienten für alle möglichen i, j einheitlich angeb- bzw. definierbar:

$$Z(j, i) := \binom{49 - i(j-1)}{6}^{+} \quad j = 8, ..., 44; \quad i = 1, ..., 7$$

Damit ist $F(j)$ nach (1) nun leicht anzugeben, und zwar einheitlich für alle möglichen $j = 8, ..., 44$[6]:

$$F(j) := \sum_{i=1}^{7} (-1)^{i-1}\binom{7}{i}\underbrace{\binom{49 - i(j-1)}{6}^{+}}_{Z(j,\,i)} \quad j = 8, ..., 44.$$

Die Verteilung von M geben wir – der Übersichtlichkeit halber – jetzt nicht mehr für alle Werte $j = 8, ..., 44$ an, sondern nur für eine kleine Auswahl; mit CAS sind die entsprechenden Werte nach obigen Definitionen auf Knopfdruck (mit zwei signifikanten Ziffern) zu erhalten (Tab. 2).

Zur Kontrolle kann dienen, dass natürlich

$$\sum_{i=8}^{44} H(j) = \binom{49}{6} = 13983816 \text{ sein muss.}$$

Für den Erwartungswert ergibt sich (man beachte $F(45) = 0$):

$$E(M) = \sum_{j=8}^{44} \frac{j \cdot H(j)}{\binom{49}{6}} = \binom{49}{6}^{-1} \sum_{j=8}^{44} j \cdot [F(j) - F(j+1)]$$

$$= \binom{49}{6}^{-1} [8F(8) + F(9) + F(10) + ... + F(10)] \approx 17,7.$$

Ohne CAS-Einsatz ist das Problem der größten Differenz M nicht sinnvoll behandelbar: erstens sind es zu viele mögliche j-Werte und zweitens ist die »Umsetzung« der Definition von $\binom{m}{n}^{+}$ auch nur mit CAS möglich (ohne konkrete Berechnungen, z. B. die Werte $H(j)$ und die Kontrolle

$$\sum_{j=8}^{44} H(j) = \binom{49}{6}, \text{bliebe } \binom{m}{n}^{+}$$

in gewisser Weise nur ein »totes« Symbol).

Eine konkrete Schüleraufgabe im Zusammenhang mit m und M könnte lauten: Löse das entsprechende Problem für ein anderes Lotto-System, z. B. österreichisches Lotto »6 aus 45«. Dazu müssen die nötigen Überlegungen mit anderen konkreten Zahlenwerten nochmals durchgeführt werden (wieder mit CAS).

Nach zwei »Durchgängen« mit konkreten Zahlenwerten sollte auch der allgemeine Fall eines »r aus n-Lotto« (Zerlegung der Zahl $n + 1$ in $r + 1$ Summanden) keine großen Schwierigkeiten mehr bereiten: Dafür müssen wiederum dieselben Überlegungen (nicht inhaltlich, sondern nur formal etwas schwieriger) durchgeführt und in das CAS »implementiert« werden, so dass die interessierenden Werte auf Knopfdruck erhalten werden können.

[6] Hier könnte die Summe – ohne Wertveränderung – nur bis $i = 6$ genommen werden, da $Z(j, 7) = 0$ für alle $j \geq 8$ ist: Bei 50 Kugeln können nicht in allen 7 Fächern mindestens 8 Kugeln sein!

j	$H(j)$	$P(M = j) = \frac{H(j)}{13983816}$
8	924	0,000066
9	24 738	0,0018
⋮	⋮	⋮
15	1 378 265	0,099 [hier ist $P(M = j)$ maximal!]
⋮	⋮	⋮
30	81 396	0,0058
⋮	⋮	⋮
40	882	0,000063
⋮	⋮	⋮
43	42	0,0000030
44	7	0,00000050

Tab. 2. Die Verteilung von M bei »6 aus 49«

Hier noch die allgemeinen Formeln:
Kleinste Differenz m: ([x] bezeichnet die Gaußklammer: größte ganze Zahl ≤ x)

$$f_{r,n}(j) = \binom{n - (r+1)(j-1)}{r}, \; j = 1, ..., \left[\frac{n+1}{r+1}\right];$$

$$h_{r,n}(j) = f_{r,n}(j) - f_{r,n}(j+1), \; j = 1, ..., \left[\frac{n+1}{r+1}\right];$$

$$E(m_{r,n}) = \binom{n}{r}^{-1} \sum_{j=1}^{\left[\frac{n+1}{r+1}\right]} j \cdot h_{r,n}(j) = ...$$

$$= \binom{n}{r}^{-1} \sum_{j=1}^{\left[\frac{n+1}{r+1}\right]} \underbrace{\binom{n - (r+1)(j-1)}{r}}_{f_{r,n}(j)}$$

Größte Differenz M:

$$F_{r,n}(j) = \sum_{j=1}^{r+1} (-1)^{i-1}\binom{r+1}{i}\binom{n - i(j-1)}{r}^{+},$$

$$j = \left[\frac{n+1}{r+1}\right] + 1, ..., n - (r-1);$$

$$H_{r,n}(j) = F_{r,n}(j) - F_{r,n}(j+1);$$

$$E(M_{r,n}) = \binom{n}{r}^{-1} \sum_{j=\left[\frac{n+1}{r+1}\right]+1}^{n-(r-1)} j \cdot H_{r,n}(j).$$

Literatur

[1] N. HENZE: The Distribution of Spaces in Lottery Tickets. – Fibonacci Quarterly **33** (1995) 426–431.

Dr. HANS HUMENBERGER, IEEM, FB Mathematik, Universität Dortmund, 44 221 Dortmund, Hans.Humenberger@math.uni-dortmund.de war zehn Jahre lang in Wien jeweils zur Hälfte an der Universität und an einem Gymnasium beschäftigt und ist seit 2000 Dozent für Didaktik der Mathematik an der Universität Dortmund.