Über das Sammeln vollständiger Figurenserien

Ingmar Rubin, Berlin

aus Newgroup de.sci.mathematik

Vorgeschichte

Vor einiger Zeit fand ich den folgenden Beitrag in der Newsgroup de.sci.mathematik:

Liebe Gruppe,
diesen Sommer hatten wir ein Ferienhaus in der Bretagne gemietet. Es waren wundervolle zwei Wochen mit erstaunlich unbretonischem Wetter, die mich so richtig abschalten ließen, so daß ich den Kopf mal wieder für etwas Mathematik frei hatte. Und schon hielt mich ein Problem in Bann, das ich bis heute nicht gelöst habe.
Wir kauften unsere Sachen meistens im Supermarkt. Habt Ihr schon mal eine Weinabteilung in einem französischen Supermarkt gesehen? Nein? Solltet Ihr aber. Aber zur Sache: in diesem Supermarkt gab es einen Automaten mit Figuren zum neuen Heffalump-Film. (Für Nichteingeweihte bzw. Kinderlose: das ist das neueste Winnie-Puuh-Machwerk von Disney.) Es gab sechs verschiedene Figuren, wenn man ein 2-Euro-Stück reinwarf, kam eine raus. Irgendeine. Meine Tochter wollte aber das Ferkel.
Nun gut, wir warfen zwei Euro rein und bekamen den Esel. Auch ganz nett. Am nächsten Tag kam Tigger. Am übernächsten auch. Und nun beging ich den Fehler, meiner Tochter zu erklären, daß, wenn man vielleicht zehnmal zieht, die Chance bestimmt schon recht hoch ist, auch mal das Ferkel zu kriegen. Ab dann mußten wir nämlich täglich in diesen Supermarkt. War nicht so schlimm, habe ich Euch schon erzählt, wie groß die Weinabteilung ... Ach so.
Und dann wollte ich's wissen: wie hoch ist die Wahrscheinlichkeit bei angenommener Gleichverteilung und beliebig vielen Figuren im Automaten, mit zehn Zügen alle sechs zu haben? Kann nicht so schwer sein, dachte ich, nahm Papier und Bleistift und fing an. Schnell hatte ich raus, daß es bei sechs Zügen $1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6 = 720$ Möglichkeiten gibt, die Figuren zu bekommen und $6^6 = 46656$ Möglichkeiten insgesamt. Das ist eine Chance von $1,5\%$, konnte ich meine Tochter trösten. Dann wollte ich weiterrechnen. Ich weiß schon nicht mehr, ob und wie ich es schaffte, die Zahl für sieben Züge auszurechnen, die Zettel habe ich weggeworfen. Ich glaube, ich hatte etwa 20% für neun Züge raus, was uns nicht gerade ermutigte,

und tatsächlich hatten wir zum Schluß, glaube ich, fünfzehn Figuren, aber kein Ferkel. Das haben wir dann für fünf Euro im Supermarkt gekauft, sah zwar nicht ganz so aus wie das aus dem Automaten, aber immerhin.
Zu Hause am Computer dachte ich 'jetzt will ich's wissen' und schrieb ein kleines Primitivprogramm, das z. B. bei allen Neunerkombinationen der Zahlen 1 bis 6 diejenigen zählt, bei denen eben alle sechs Zahlen vorkommen, und ließ es laufen.
Das Ergebnis:

Züge	Kombinationen mit 1-6	Kombinationen gesamt	Prozent
6	720	46656	1,5
7	15120	279936	5,4
8	191520	1679616	11,4
9	1905120	10077696	18,9
10	16435440	60466176	27,2
11	129230640	362797056	35,6
12	953029440	2176782336	43,8
13	6711344640	13060694016	51,4

Kein Wunder also. Aber wie rechnet man sowas? Anders ausgedrückt: welche Formel ergibt die Wahrscheinlichkeit, daß bei einem Wurf mit n Würfeln die Zahlen 1 bis 6 erscheinen? Das kann doch nicht so schwer sein. Die Gesamtkombinationen sind natürlich die Sechserpotenzen, aber bei den anderen Zahlen verliere ich mich andauernd in irgendwelchen Fakultäten und Permutationen. Vielleicht kommt noch ein Programmierfehler dazu, und die Zahlen stimmen sowieso nicht.
Hilft mir jemand?
Viele Grüße Steffen

Eine äquivalente Aufgabenstellung

Eine äuivalente Aufgabenstellung zum oben geschilderten Problem fand ich einige Wochen später in der gleichen Newsgroup:

Hallo Forum!
Ich habe ein Problem aus der Wahrscheinlichkeitsrechnung, das ich nicht sicher gelöst bekomme. Es lautet: Wie groß ist die Wahrscheinlichkeit, daß bei 10-maligem Werfen eines normalen Würfels mit 6 Seiten, die Zahlen 1 bis 6 jeweils mindestens einmal auftauchen.
Meine eigene Überlegung lautet:

$$p = \frac{6! \cdot 6^4}{6^{10}} \tag{1}$$

Ich würde mich über das richtige Ergebnis freuen - Benedikt Kaiser (benedikt.kaiser@planet-interkom.de)

Lösungsansatz aus der Kombinatorik

von Horst Kraemer (horst.kraemer@snafu.de)

Der Ansatz $6! \cdot 6^4$ berechnet nur einen kleinen Teil der gesuchten Wurffolgen - z.B. alle Folgen, die an den Stellen 1-6 alle Zahlen aus $1 \ldots 6$ genau einmal enthalten und an den Stellen $7 \ldots 10$ beliebige Werte. Dies sind aber viel zu wenig. Die korrekte Loesung ist leider nicht ganz so einfach.
Die Formel für die Anzahl der Zahlenfolgen der Länge 10, die alle Elemente aus $1, 2, 3, 4, 5, 6$ mindestens einmal enthalten (mathematisch die Anzahl der surjektiven Abbildungen einer 10-Menge in eine 6-Menge), ist

$$\begin{aligned} N = \mathrm{Sur}(10,6) = {} & 6^{10} \\ & -5^{10} \cdot C(6,5) \\ & +4^{10} \cdot C(6,4) \\ & -3^{10} \cdot C(6,3) \\ & +2^{10} \cdot C(6,2) \\ & -1^{10} \cdot C(6,1) \\ & = 16.435.440 \end{aligned}$$

wobei $C(n, k)$ jeweils der Binomialkoeffizient

$$C(n,k) = \frac{n!}{n! \cdot (k-n)!} \tag{2}$$

ist (Anzahl der k-Kombinationen aus n Elementen). Eine einfachere Formel gibt es für diese Zahl leider nicht. Höchstens die kürzere Schreibweise:

$$\mathrm{Sur}(n,k) = \sum_{i=1}^{k} (-1)^{k-i} \cdot C(k,i) \cdot i^n \tag{3}$$

Die Wahrscheinlichkeit beträgt dann für n Würfe:

$$p(n,k) = \frac{\mathrm{Sur}(n,k)}{k^n} = \frac{\sum_{i=1}^{k} (-1)^{k-i} \cdot C(k,i) \cdot i^n}{k^n} \tag{4}$$

Die Wahrscheinlichkeit beim Würfel mit $k = 6$ und 10 Würfen beträgt:

$$p(10,6) = \frac{\mathrm{Sur}(10,6)}{6^{10}} \approx 0.271812 \tag{5}$$

Das Ergebnis deckt sich mit der Computersimulation aus der ersten Aufgabenstellung. Die strenge Herleitung der Formel mit vollständiger Induktion ist einfach und kurz, aber leider etwas unintuitiv.

Verteilungsfunktion und Wahrscheinlichkeitsdichte

Die Formel (4) gestattet uns die Verteilungsfunktion graphisch darzustellen. Wir berechnen dazu die Wahscheinlichkeitswerte für das Intervall $1 \leq n \leq 40$ und verbinden die Punkte zu einer durchgehenden Kurve. Die Ableitung der Vertei-

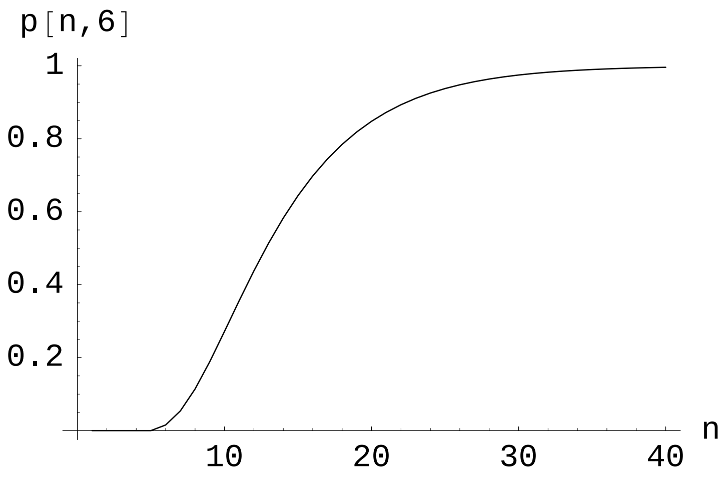

Abbildung 1: Verteilungsfunktion für eine Serie aus 6 Figuren und $1 \leq n \leq 40$

lungsfunktion ergibt die Wahrscheinlichkeitsdichte (bei der Normalverteilung die bekannte Glockenkurve). Wir bilden die fortlaufenden Differenzen zwischen den oben berechneten Werten und erhalten die Graphik in Abbildung 2 (alle Punkte sind zu einem durchgehenden Kurvenzug verbunden). Aus der Wahrscheinlich-

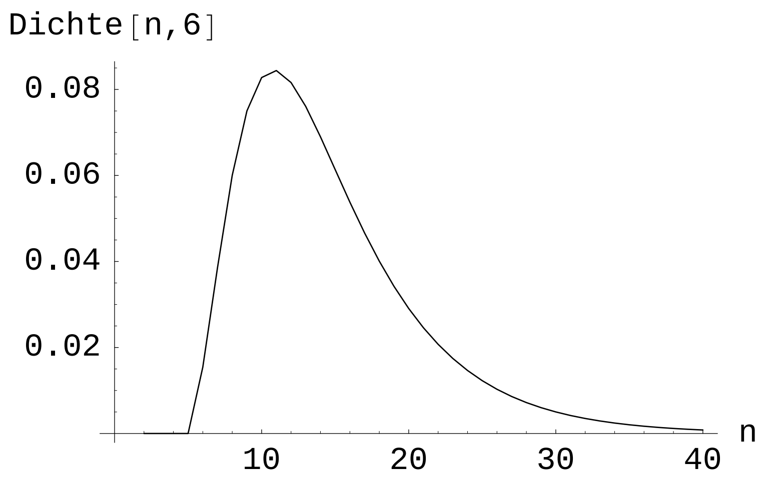

Abbildung 2: Wahrscheinlichkeitsdichte für eine Serie aus 6 Figuren

keitsdichte können wir den Erwartungswert berechen, d.h. die mittlere Anzahl

Ziehungen nach der wir die Serie vollständig haben ($k = 6$, $n = 60$):

$$EW[k] = \sum_{i=1}^{n} (i+1) \cdot [p(i+1,6) - p(i,6)] \approx 14.693 \tag{6}$$

Wahrscheinlichkeitsberechnung mittels Markowketten

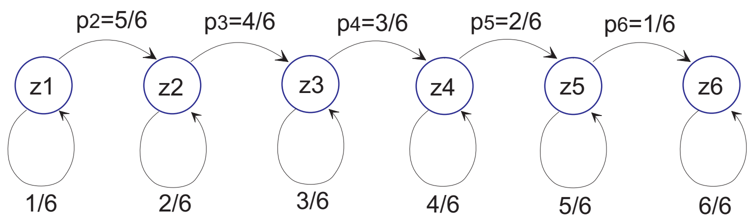

Abbildung 3: Markowkette für eine Serie aus 6 Figuren

Wir können das Sammeln der Figurenserie mit Hilfe einer *Markowkette* modellieren (Abbildung 3). Die Zustände $z_1 \dots z_6$ bezeichnen die Anzahl der Figuren aus der Serie, die wir bereits besitzen. Immer wenn wir beim Ziehen aus der Vorratsbox eine neue Figur aus der Serie erhalten, wandern wir in der Zustandskette um eine Position nach rechts.
Wir starten bei z_1, d.h. wir haben die erste Figur bereits gezogen. Bei der folgenden Ziehung sind 5 von 6 Figuren für uns günstig, d.h. die Übergangswahrscheinlichkeit von z_1 nach z_2 beträgt 5/6. In einem von 6 Fällen ziehen wir die gleiche Figur und verharren im Zustand z_1. Die Gesamtheit der Übergangswahrscheinlichkeiten wird in einer Übergangsmatrix $\mathbf{P}$ zusammengefasst:

$$\mathbf{P} = \begin{bmatrix} 1/6 & 0 & 0 & 0 & 0 & 0 \\ 5/6 & 2/6 & 0 & 0 & 0 & 0 \\ 0 & 4/6 & 3/6 & 0 & 0 & 0 \\ 0 & 0 & 3/6 & 4/6 & 0 & 0 \\ 0 & 0 & 0 & 2/6 & 5/6 & 0 \\ 0 & 0 & 0 & 0 & 1/6 & 6/6 \end{bmatrix} \tag{7}$$

Die Wahrscheinlichkeiten mit der wir uns im Zustand $z_1 \dots z_6$ befinden, fassen wir in den Vektoren $p_i[z_1, z_2, z_3, z_4, z_5, z_6]$ zusammen. Der Startzustand wird durch den Vektor $p_1[1, 0, 0, 0, 0, 0]$ repräsentiert, d.h. wir haben die erste Figur bereits gezogen und befinden uns mit Sicherheit bei z_1. Jeden weiteren Vektor erhalten wir aus der rekursiven Gleichung:

$$p_{i+1} = p_i \cdot \mathbf{P},\ i = 1 \dots n-1 \quad \rightarrow \quad p_n = p_1 \cdot \mathbf{P}^{(n-1)} \tag{8}$$

Der erste Spaltenvektor in $P^{(n-1)}$ entspricht nun gerade dem Vektor p_n). Durch fortlaufendes potenzieren der Matrix P erhalten wir Schritt für Schritt die Wahrscheinlichkeitsverteilung für die Zustände $z_1 \dots z_6$. Für eine Serie mit $k = 6$ erhalten wir nach der 6.Ziehung:

$$\mathbf{P}^{(6-1)} = \begin{pmatrix} 0.000128601 & 0. & 0. & 0. & 0. & 0. \\ 0.0199331 & 0.00411523 & 0. & 0. & 0. & 0. \\ 0.231481 & 0.108539 & 0.03125 & 0. & 0. & 0. \\ 0.501543 & 0.439815 & 0.301312 & 0.131687 & 0. & 0. \\ 0.231481 & 0.385802 & 0.509259 & 0.540381 & 0.401878 & 0. \\ 0.0154321 & 0.0617284 & 0.158179 & 0.327932 & 0.598122 & 1. \end{pmatrix}$$

Der erste Spaltenvektor zeigt für den Zustand z_6 eine Wahrscheinlichkeit von $p_6[6] = 0.0154321$ an. Diese Ergebnis hatten wir auch mittels Kombinatorik im vorangehenden Abschnitt berechnet. Nach $n = 10$ Ziehungen sieht die Matrix $\mathbf{P}^{(10-1)}$ wie folgt aus:

$$\mathbf{P}^{(10-1)} = \begin{pmatrix} 9.9229 \times 10^{-8} & 0. & 0. & 0. & 0. & 0. \\ 0.00025353 & 0.0000508053 & 0. & 0. & 0. & 0. \\ 0.0185161 & 0.00760928 & 0.00195313 & 0. & 0. & 0. \\ 0.203052 & 0.132941 & 0.0721775 & 0.0260123 & 0. & 0. \\ 0.506366 & 0.486314 & 0.431206 & 0.335589 & 0.193807 & 0. \\ 0.271812 & 0.373085 & 0.494664 & 0.638399 & 0.806193 & 1. \end{pmatrix}$$

Die Wahrscheinlichkeit die Serie nach 10 Ziehungen vollständig zu haben beträgt $p_{10}[6] = 0.271812$.

Mittlere Wartezeit auf eine vollständige Serie

Wir wollen jetzt die mittlere Anzahl an Zügen berechnen, bis wir ein Serie aus k Elementen vollständig erhalten. Mit anderen Worten geht es um die Bestimmung des *Erwartungswertes*. Dazu sei die folgende Computersimulation vorangestellt.
In ARIBAS (`http://www.mathematik.uni-muenchen.de/~forster/sw/aribas.html`) schreiben wir die Funktion $serie(k, n)$:

```
function serie(k, n :integer) : real;

var figur : array[k+1] of integer;
    i,j,p,su    : integer;
    y           : real;
begin
 for i:=1 to k do figur[i]:=0; end;
 su:=0;
 for i:=1 to n do
    for j:=1 to k do figur[j]:=0; end;
    j:=0;
    while sum(figur) < k do
      inc(j);
      p:= random(k) + 1;
```

```
        figur[p]:=1;
      end;
      su:= su + j;
   end;
   y:= su/n;
   return y;
 end;
```

Das Programm liefert uns den Erwartungswert für eine Serie aus k Figuren, wobei das Ziehen einer kompletten Serie $n-$ mal wiederholt wird. Wir wählen n möglichst hoch ($n = 10^6$), um einen plausiblen Mittelwert zu erhalten. Die folgende Tabelle zeigt die simulierten Erwartungswerte von $k = 2 \ldots 12$. Die Folge der Erwartungswerte ist monoton steigend, wie man aus der ersten Differenz erkennt.

k	EW	$EW[k] - EW[k-1]$
2	3.0	0
3	5.5	2.5
4	8.34	2.84
5	11.41	3.07
6	14.71	3.30
7	18.153	3.442
8	21.686	3.53
9	25.47	3.78
10	29.2795	3.809
11	33.2175	3.938
12	37.2495	4.032

Berechnung des Erwartungswertes mit Markowketten

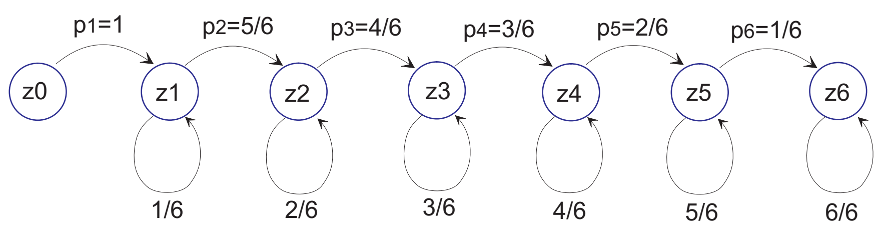

Abbildung 4: Markowkette für eine Serie aus 6 Figuren

Wir können den Erwartungswert für das Sammeln einer Serie der Länge s aus der Markowkette berechnen. Startpunkt ist der Zustand z_0, d.h. wir besitzen noch keine Figur. Von z_0 nach z_1 beträgt Übergangswahrscheinlichkeit $p_1 = 1$, d.h. es ist sicher, daß wir bei der ersten Ziehung eine neue Figur erhalten. Der Erwartungswert von z_0 nach z_1 ist $EW_1 = 1/p_1 = 1$, d.h. wir werden immer in

einem Zug den Zustand p_1 erreichen.
Beim Übergang von z_1 nach z_2 sind 5 Figuren für uns nützlich. In einem von 6 Fällen werden wir im Zustand z_1 verharren. Für den Erwartungswert gilt (also im Mittel nach vieviel Zügen schaffen wir den Übergang z_1 nach z_2):

$$p_2 = p(z_1 \to z_2) = \frac{5}{6}, \qquad EW_2 = \frac{1}{p_2} = \frac{6}{5} \tag{9}$$

Analog berechnen wir die Erwartungswerte für die weiteren Übergänge:

$$p_3 = p(z_2 \to z_3) = \frac{4}{6}, \qquad EW_3 = \frac{1}{p_3} = \frac{6}{4} \tag{10}$$

$$p_4 = p(z_3 \to z_4) = \frac{3}{6}, \qquad EW_4 = \frac{1}{p_4} = \frac{6}{3} \tag{11}$$

$$p_5 = p(z_4 \to z_5) = \frac{2}{6}, \qquad EW_5 = \frac{1}{p_5} = \frac{6}{2} \tag{12}$$

Nach dem wir 5 Figuren besitzen, beträgt die Wahrscheinlichkeit nur noch 1/6 die letzte, fehlende Figur der Serie zu ziehen. Im Mittel werden wir also 6 mal ziehen müssen, um die gewünschte Figur zu erhalten:

$$p_6 = p(z_5 \to z_6) = \frac{1}{6}, \qquad EW_6 = \frac{1}{p_6} = 6 \tag{13}$$

Die Summe aller Übergangserwartungswerte gibt uns die mittlere Anzahl an Ziehungen, bis wir die Serie vollständig haben:

$$EW = \sum_{i=1}^{6} EW_i = 1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + 6 = 6 \cdot \frac{147}{60} = 14.7 \tag{14}$$

Allgemein erhalten wir den Erwartungswert für eine Serie aus s Elementen zu:

$$EW(s) = 1 + \frac{s}{s-1} + \frac{s}{s-2} + \cdots + \frac{s}{2} + s \tag{15}$$

$$EW(s) = s \cdot \left(\frac{1}{s} + \frac{1}{s-1} + \frac{1}{s-2} + \cdots + \frac{1}{2} + 1 \right) \tag{16}$$

$$EW(s) = s \cdot \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{s} \right) = s \cdot \sum_{i=1}^{s} \frac{1}{i} = s \cdot \mathrm{Harm}(s) \tag{17}$$

Hinweis: $\mathrm{Harm}(s)$ bezeichnet hier die Partialsumme der *harmonischen Reihe*. Sie ist in vielen Mathematikprogrammen fertig als Funktion enthalten, so z.B. in Mathematica als HarmonicNumber[i].