Zur Berechnung der Wahrscheinlichkeit für das Vorliegen einer vollständigen Serie (Sammelbildproblem)

Heinz Althoff, Bielefeld

Zusammenfassung: Es wird gezeigt, mit welchen Ideen für das genannte Problem eine Rekursionsformel mit zwei Parametern, eine Rekursionsformel mit einem Parameter und eine explizite Formel entwickelt werden können.

In (Althoff 1999) beschreibe ich, wie man im Rahmen einer Unterrichtsreihe in einem Leistungskurs verschiedene stochastische Problemtypen (Geburtstagsproblem, Rosinenproblem, Sammelbilderproblem) mit dem *gleichen Laplace-Modell*, d. h. mit dem gleichen Stichprobenraum $S = Per_k^n(mW)$, untersuchen kann. Im Vergleich zum Geburtstagsproblem und zum Rosinenproblem erweist sich die Berechnung der Anzahl der *günstigen* Versuchsausfälle beim Sammelbilderproblem jedoch als wesentlich schwieriger. Ich möchte in diesem Aufsatz verdeutlichen, mit welchen typischen mathematischen Strategien zur Lösung des Sammelbilderproblems

- in (Strick 1994) eine Rekursionsformel entwickelt wird, die sich allerdings bei großen Zahlen als unhandlich erweist, da bei ihr eine Rekursion nach zwei Parametern erforderlich ist;
- durch Wahl eines anderen Modells (*hier:* einer anderen Zufallsgröße) eine Rekursionsformel gewonnen werden kann, in der die Rekursion nur noch nach einem Parameter erfolgt;
- wie man induktiv von dieser Rekursionsformel zu einer Vermutung für eine explizite Berechnungsformel gelangen kann;
- man aufgrund der Struktur dieser Formel erkennen kann, wie man sie auf eine allgemeinere (schon bekannte) Formel zurückführen und damit ihre Gültigkeit nachweisen kann.

Beim Sammelbilderproblem geht es um die Berechnung der Wahrscheinlichkeit $P(E)$ für das Ereignis

$E \triangleq$ *spätestens* das k-te gekaufte Sammelbild macht die Serie der n möglichen Sammelbildertypen vollständig.

Dabei kommt es im hier vorliegenden Laplace-Modell vor allem auf die Berechnung der Anzahl $|E|$ der für E günstigen Versuchsausfälle an. Im *Fächerbelegungsmodell*, das ich in (Althoff 1999) genauer erläutere, ist dann

$E \triangleq$ *spätestens* nach k Belegungsvorgängen (mit zugelassener Mehrfachbelegung eines Faches) sind alle n vorhandenen Fächer vollständig belegt.

Das Vorgehen in (Strick 1994) bei der Berechnung der Wahrscheinlichkeit für das Vorliegen einer vollständigen Serie kann man in der Kernidee nun folgendermaßen beschreiben:

Für $0 \le b \le k$ werden die Zufallsgrößen

$X_b \triangleq$ Anzahl der bei b Belegungsvorgängen belegten Fächer unter den n vorhandenen Fächern

definiert. Ferner wird mit

$A(f; b) = |\{X_b = f\}|$

die Anzahl der Möglichkeiten angegeben, bei b Belegungsvorgängen genau f der n möglichen Fächer vollständig zu belegen. Gesucht ist dann

$$A(n; k) \text{ bzw. } P(E) = \frac{|E|}{|Per_k^n(mW)|} = \frac{A(n;k)}{n^k}.$$

$A(n; k)$ ergibt sich mit Hilfe der *Rekursionsformel*

(1) $A(f; b) =$
$A(f; b-1) \cdot f + A(f-1; b-1) \cdot (n-(f-1))$ für $b \in \{1, 2, \ldots, k\}$ und $f \in \{2, 3, \ldots, n\}$

unter Beachtung der leicht einsehbaren *Randbedingungen*

(2) $A(f; b) = 0$ für $b \in \{0, 1, \ldots, f-1\}$ und
(3) $A(1; b) = n$ für $b \in \{1, 2, \ldots, k\}$.

Die Formel (1) erhält man durch folgende Überlegungen:

Wenn *nach* dem b-ten Belegungsvorgang genau f von n möglichen Fächern belegt sein sollen, müssen nach dem $(b-1)$-ten Belegungsvorgang entweder schon alle f Fächer belegt gewesen sein (dafür gab es $A(f; b-1)$ Möglichkeiten) oder erst $f-1$ Fächer belegt gewesen sein (dafür gab es $A(f-1; b-1)$ Möglichkeiten); für den b-ten Belegungsvorgang gibt es dann im ersten Fall f Möglichkeiten (da jedes der f schon belegten Fächer in Frage kommt) und im zweiten Fall $n-(f-1)$ Möglichkeiten (da jetzt jedes noch nicht belegte Fach in Frage kommt).

Führt man nun die Berechnung von $A(n; k)$ konkret durch, so zeigt sich schon bei kleinen Zahlen für n und k (Beispiel: $E_1 \triangleq$ nach 10 Würfen mit einem L-Würfel sind alle 6 Augenzahlen wenigstens einmal vorgekommen), daß ein unverhältnismäßig großer Schreib- und Rechenaufwand erforderlich ist. Bei größeren Zahlen (Beispiel: $E_2 \triangleq$ von den 2000 Schülern einer Schule hat an *jedem* Tag des Jahres *wenigstens einer* Geburtstag) führt der Aufwand auch bei Benutzung eines Computers zu langen Rechenzeiten und eventuell sogar zu Speicherplatzproblemen. Hauptursache dafür ist, daß bei Benutzung der Rekursionsformel (1) eine Rekursion bezüglich beider Parameter *b* und *f* durchzuführen ist.

Hauptaugenmerk bei der Herleitung einer besser handhabbaren Rekursionsformel muß also sein, daß in ihr nur noch eine Rekursion bezüglich eines der beiden Parameter nötig ist und der Wert des anderen Parameters konstant gehalten werden kann. Dies kann man dadurch erreichen, daß man mit Zufallsgrößen arbeitet, die nicht von einem vorgegebenen Wert von *b*, sondern von einem vorgegebenen Wert von *f* ausgehen:

Von *n* vorhandenen Fächern sind *f* Fächer zum Belegen zugelassen. Dann werden für $1 \le f \le n$ die Zufallsgrößen

$Y_f \mathrel{\hat=}$ Anzahl der zur vollständigen Belegung der *f* zum Belegen zugelassenen Fächer erforderlichen Belegungsvorgänge

definiert. Ferner wird mit

(4) $B(f; k) = |\{Y_f \le k\}|$

die Anzahl der Möglichkeiten angegeben, *genau* die *f* zur Belegung zugelassenen Fächer mit *höchstens k* Belegungsvorgängen vollständig zu belegen. Gesucht ist dann $B(n; k)$ und $P(E) = P(Y_n \le k) = \dfrac{B(n;k)}{n^k}$.

Für $k < n$ ist natürlich $B(n; k) = 0$.

Für $k \ge n$ betrachte ich zunächst den Fall $n = 1$. Aus der Definition (4) ergibt sich dafür $B(1; k) = 1$.

Um $B(n; k)$ für $n \ge 2$ rekursiv durch $B(f; k)$ mit $1 \le f < n$ ausdrücken zu können, schreibe ich $B(n; k)$ als Differenz aus der Anzahl n^k aller *möglichen* Versuchsausfälle und der Anzahl der *ungünstigen* Versuchsausfälle, d. h. der Anzahl der Möglichkeiten, mit *k* Belegungsvorgängen höchstens $n - 1$ Fächer vollständig zu belegen. Diese Anzahl der ungünstigen Versuchsausfälle ergibt sich als $\sum_{f=1}^{n-1} C(f;k)$, wobei $C(f; k)$ für $1 \le f \le n-1$ die Anzahl der Möglichkeiten ausdrückt, mit *k* Belegungsvorgängen *genau f* zur Belegung zugelassene Fächer vollständig zu belegen. Da man zunächst $\binom{n}{f}$ Möglichkeiten hat, aus den *n* Fächern *f* zur vollständigen Belegung auszuwählen, und danach $B(f; k)$ Möglichkeiten, mit höchstens *k* Belegungsvorgängen *diese f* ausgewählten Fächer vollständig zu belegen, ergibt sich $C(f, k) = \binom{n}{f} \cdot B(f; k)$ und damit die neue Rekursionsformel

(5) $B(n; k) = n^k - \sum_{f=1}^{n-1} \binom{n}{f} B(f;k)$

mit der Randbedingung $B(1; k) = 1$.

Die Anzahlen $B(f, k)$ kann man natürlich ebenfalls mit (5) auf die $B(f^*, k)$ mit $f^* < f$ zurückführen, wenn man *n* durch *f* und *f* durch f^* ersetzt.

Da in (5) die Rekursion wie angestrebt nur noch nach dem Parameter *f* erfolgt (*k* ist jetzt konstant), wird nicht nur im Vergleich zur Formel (1) der Rechenaufwand wesentlich reduziert, sondern auch die Hoffnung erweckt, durch Betrachtung einiger konkreter Beispiele für *n* vielleicht sogar eine *explizite* Formel für $B(n; k)$ und damit natürlich auch für $P(E)$ zu entdecken.

Für $n = 1$ wurde oben schon $B(1; k) = 1$ hergeleitet.

Für $n = 2$ gilt:

$$B(2; k) = 2^k - \sum_{f=1}^{1} \binom{2}{f} B(f;k) = 2^k - \binom{2}{1} B(1; k) = 2^k - 2$$

Für $n = 3$ gilt:

$$B(3; k) = 3^k - \sum_{f=1}^{2} \binom{3}{f} B(f;k)$$

$$= 3^k - [\binom{3}{1} B(1; k) + \binom{3}{2} B(2; k)]$$

$$= 3^k - 3 \cdot 1 - 3 \cdot (2^k - 2) = 3^k - 3 \cdot 2^k + 3 =$$

$$\binom{3}{0} \cdot 3^k - \binom{3}{1} \cdot 2^k + \binom{3}{2} \cdot 1^k = \sum_{i=0}^{3-1} (-1)^i \cdot \binom{3}{i} \cdot (3-i)^k$$

Entsprechend erhält man für $n = 4$ (der Leser möge sich durch eine ausführliche Rechnung davon überzeugen!)

$$B(3; k) = \sum_{i=0}^{4-1} (-1)^i \cdot \binom{4}{i} \cdot (4-i)^k$$

Das führt (für $k \ge$ n) zur *Vermutung*

(6) $B(n; k) = \sum_{i=0}^{n-1} (-1)^i \cdot \binom{n}{i} \cdot (n-i)^k$

und der sich daraus ergebenden Formel

(7) $P(E) = \dfrac{\sum_{i=0}^{n-1} (-1)^i \cdot \binom{n}{i} \cdot (n-i)^k}{n^k} = \sum_{i=0}^{n-1} (-1)^i \cdot \binom{n}{i} \cdot \left(1 - \dfrac{i}{n}\right)^k$.

Den Term (7) zur Berechnung von $P(E)$ kann man ohne große Mühe in den Taschenrechner TI-92 (oder in einen Computer) eingeben. Für die oben erwähnten Beispiele berechnet der TI-92 dann $P(E_1) = 0{,}271812$ in Sekundenschnelle und $P(E_2) = 0{,}216119$ in wenigen Minuten.

Die Formel (6) findet man übrigens schon als (10) auf Seite 81 in (Fricke 1984), allerdings ist der dort mit vollständiger Induktion durchgeführte Beweis der Formel nicht ganz einfach zu durchschauen. Ein Beweisversuch des Autors auf einem etwas anderen

Weg führte zwar ebenfalls zum Ziel, bedurfte aber großer Anstrengungen; deshalb wird an dieser Stelle auf einen Abdruck dieses relativ umfangreichen Beweises verzichtet.

Stattdessen möchte ich noch auf eine Idee in (Treiber 1988) hinweisen. Dort wird verdeutlicht, wie man die Formel (6) auf das allgemeinere *Zählprinzip des Ein- und Ausschließens* zurückführen kann.

Man kann nun die Formel des Ein- und Ausschließens mit vollständiger Induktion beweisen (Treiber verzichtet darauf, man findet den Beweis aber z. B. in (Henze 1997) auf S. 72) oder sie wieder auf ein allgemeineres Zählprinzip, den binomischen Lehrsatz zurückführen, wie es z. B. in (Danckwerts/Vogel/Bovermann 1985) auf den Seiten 128/129 und 152 auch für LK-Schüler verständlich dargestellt ist.

Literatur:

Althoff, H. (1999): Das Lösen stochastischer Probleme mit Hilfe eines Fächerbelegungsmodells – ein Bericht über eine Unterrichtsreihe in einem Leistungskurs 13. - In: Berichte aus dem Seminar für Didaktik der Mathematik WS 97/98 und SS 98, Universität Bielefeld

Danckwerts, R./Vogel, D./Bovermann, K. (1985): Elementare Methoden der Kombinatorik. - Verlag Teubner Stuttgart.

Fricke, A. (1984): Das stochastische Problem der vollständigen Serie. - In: Der Mathematikunterricht 30, Heft 1.

Henze, N. (1997): Stochastik für Einsteiger. - Verlag Vieweg Braunschweig/Wiesbaden

Strick, H. K. (1994): Modellieren als Aufgabe des Stochastikunterrichts. - In: Praxis der Mathematik 36, Heft 2.

Treiber, D. (1988): Zur Wartezeit auf eine vollständige Serie. - In: Didaktik der Mathematik 16, Heft 3.

Bedingte Erwartungswerte

Hans Humenberger, Wien

Zusammenfassung: Das Thema *Bedingte Erwartungswerte* ist in der didaktischen Literatur nicht oft zu finden. Es ist daher kein Wunder, daß sie in vielen Stochastikkursen für Schüler (und bisweilen vielleicht auch für Studenten) fehlen. Dies ist u.E. insbesondere dann „schade", wenn in einem Stochastikkurs sowohl bedingte Wahrscheinlichkeiten als auch Erwartungswerte besprochen werden. Die dann noch fehlende Symbiose zum bedingten Erwartungswert ist nicht mehr mit viel zusätzlichem Aufwand verbunden, öffnet aber eine Reihe von Einsatz- bzw. Anwendungsmöglichkeiten (siehe z.B. KILIAN 1987 oder HUMENBERGER 1998abc).

Nach einer kurzen Replik von bedingten Wahrscheinlichkeiten und Erwartungswerten bei diskreten Zufallsvariablen wird ein intuitiver Zugang zu bedingten Erwartungswerten (in Analogie zu bedingten Wahrscheinlichkeiten) kurz dargestellt. Die (etwas abstraktere) Definition eines *bedingten Erwartungswertes* ist jedoch mit dieser der Intuition entsprechenden Sichtweise verträglich. Es folgen in dieser Arbeit dann einige kurze Beispiele, die das „Potential" von bedingten Erwartungswerten illustrieren sollen.

1 Bedingte Erwartungswerte

1.1 Bedingte Wahrscheinlichkeiten und Erwartungswerte

Bedingte Erwartungswerte stehen zu normalen Erwartungswerten im selben Verhältnis wie bedingte Wahrscheinlichkeiten zu normalen Wahrscheinlichkeiten. Deshalb sei kurz an bedingte Wahrscheinlichkeiten erinnert. Meist geschieht der Zugang zu diesen intuitiv anhand eines Beispiels wie dem folgenden. Jemand hat aus einer Personengruppe, die sich aus Damen – Herren bzw. Brillenträgern – Nicht-Brillenträgern (wie in Tab. 1 zu lesen) zusammensetzt, eine Person zufällig ausgewählt.

	D	H
B	2	3
$\neg B$	5	4

Tab. 1: Damen (D) – Herren (H), Brillenträger (B) – Nicht-Brillenträger ($\neg B$)

Man erhält aus dieser Tabelle sofort für die einzelnen Wahrscheinlichkeiten:

$$P(D) = \frac{7}{14} = \frac{1}{2}, \quad P(H) = \frac{7}{14} = \frac{1}{2},$$

$$P(B) = \frac{5}{14}, \quad P(\neg B) = \frac{9}{14}.$$

Jemand verrät einem, daß es sich bei der gewählten Person um eine Dame handelt. Wie groß ist dann die Wahrscheinlichkeit, daß diese Dame eine Brillenträgerin ist?

Auch wenn jemand noch nichts von *bedingten Wahrscheinlichkeiten* gehört hat, so wird er argumentieren: Aufgrund der Information, die ich habe, kommen nur mehr noch die 7 Damen in Frage („Anzahl der möglichen Fälle"), wobei 2 von diesen („günstige Fälle") Brillenträgerinnen sind. Der