

Rässler, Susanne

Working Paper

Des Kartensammlers Dilemma

Diskussionspapier, No. 48/2003

Provided in Cooperation with:

Friedrich-Alexander-University Erlangen-Nuremberg, Chair of Statistics and Econometrics

Suggested Citation: Rässler, Susanne (2003) : Des Kartensammlers Dilemma, Diskussionspapier, No. 48/2003, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Statistik und Ökonometrie, Nürnberg

This Version is available at:

<https://hdl.handle.net/10419/29577>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Des Kartensammlers Dilemma

Susanne Rässler
Lehrstuhl für Statistik und Ökonometrie
Friedrich-Alexander-Universität Erlangen-Nürnberg
Lange Gasse 20
D-90403 Nürnberg
Email: susanne.raessler@wiso.uni-erlangen.de

Diskussionspapier 48 / 2003

Zusammenfassung

Käufer von Sammelbildern mögen sich bisweilen fragen, wieviele Käufe sie noch tätigen müssen, um bestimmte Bilder, die zu Gruppen in Tüten verpackt sind, zu erhalten. Dies ist eine gegenüber dem typischen "Coupon Collector's Problem" modifizierte Fragestellung, da es dort vornehmlich um die Anzahl der nötigen Käufe geht, um alle einzeln verpackte Bilder zu erhalten. In diesem Beitrag wird ein klassisches Modell aus der Stichprobentheorie angewandt, um die Wahrscheinlichkeiten für solche erfolgreichen Käufe zu berechnen. Neben der Annahme eines sog. fairen Spiels werden auch unterschiedliche Produktionswahrscheinlichkeiten erlaubt und ein Programmcode zur Berechnung der gesuchten Wahrscheinlichkeiten entwickelt.

Keyword: Sammelbilder, Auswahl mit unterschiedlichen Wahrscheinlichkeiten, geometrische Verteilung, Ziehen ohne Zurücklegen, Coupon Collector's Problem.

1 Einleitung

Der weltweit operierende Panini-Konzern¹ gibt zu bestimmten Themen Tüten mit Sammelbildern (Sticker) heraus.

Im Rahmen der aktuellen Bundesliga Spiele 2002/2003 sind dies Tüten von jeweils 7 unterschiedlichen Bildern aus einer Gesamtanzahl von 498 Bildern. Im Bundesgebiet soll es etwa 55000 Einzelhändler geben, die die bunten Tüten meist in direkter Kassennähe anbieten. Die Sticker werden millionenfach produziert und laut Panini zu einer “optimalen” Mischung zusammengestellt. Unter den Sammlern haben sich teilweise Tauschbörsen gebildet, das Internet gibt darüber reichhaltig Auskunft.

Hier wollen wir Fragen nachgehen, die von einigen Bildersammlern an uns gestellt wurden. Wieviele Tüten müssen im Durchschnitt gekauft werden, um bestimmte Bilder zu bekommen? Mit welcher Wahrscheinlichkeit ist noch eine bestimmte Anzahl an Tüten zu erwerben, um ein gesuchtes, fehlendes Bild der Sammlung in einer Tüte zu finden?

Dazu werden wir im folgenden Abschnitt erst einige grundlegende Begriffe der Kombinatorik einführen. Dann sollen zwei Varianten betrachtet werden. Zunächst wird im dritten Abschnitt angenommen, daß alle Bilder mit gleicher Wahrscheinlichkeit in die Tüten gepackt werden. Für den anderen Fall, daß es bestimmte Bilder gibt, die mit einer geringeren Wahrscheinlichkeit produziert werden, bieten wir im vierten Abschnitt ebenfalls einen Lösungsvorschlag an. Der fünfte Abschnitt gibt dann eine Zusammenfassung, der für diese Analyse implementierte Computercode kann aus dem Internet heruntergeladen werden.

2 Etwas Kombinatorik

Für die Packung einer jeden Tüte mit 7 Bildern bei insgesamt 498 verschiedenen Bildern läßt sich das Vorgehen wie folgt abbilden:

Wir ziehen Elemente aus einer Grundgesamtheit vom Umfang $N = 498$ ohne Zurücklegen, weil in jeder Tüte genau 7 verschiedene Bilder enthalten sind. Jede Tüte entspricht also einer Stichprobe aus dieser Grundgesamtheit vom Umfang $n = 7$. Da die Reihenfolge, in der die Bilder in die Tüten gepackt werden, zunächst nicht interessiert, handelt es sich um sogenannte ungeordnete Stichproben ohne Zurücklegen. Von diesen gibt es insgesamt

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \tag{1}$$

verschiedene Stichproben. Es bezeichne \mathcal{S} die Menge der möglichen Stichproben, dann

¹Siehe dazu die Website www.panini.de.

gilt für ungeordnete Stichproben des Umfanges n , $\mathcal{S} = \{s_1, s_2, \dots, s_{\binom{N}{n}}\}$.

Ist, unter bestimmten Umständen, die Reihenfolge noch von Interesse, dann können

$$n! \binom{N}{n} = \frac{N!}{(N-n)!} \quad (2)$$

verschiedene Stichproben betrachtet werden.² Für geordnete Stichproben des Umfanges n ist $\mathcal{S} = \{s_1, s_2, \dots, s_{n! \binom{N}{n}}\}$.

In der Tabelle 1 sind für ein einfaches Beispiel die möglichen Stichproben angegeben. Wir betrachten die Ziehung von Stichproben vom Umfang $n = 2$ aus einer Grundgesamtheit vom Umfang $N = 4$.

Tabelle 1: Stichprobenmuster bei $N = 4$ und $n = 2$

Unter Beachtung der Reihenfolge	Ohne Beachtung der Reihenfolge
(1,2)	(1,2)
(1,3)	(1,3)
(1,4)	(1,4)
(2,1)	–
(2,3)	(2,3)
(2,4)	(2,4)
(3,1)	–
(3,2)	–
(3,4)	(3,4)
(4,1)	–
(4,2)	–
(4,3)	–

Wir sehen in diesem einfachen Beispiel, daß es unter Beachtung der Reihenfolge genau $2! \binom{4}{2} = \frac{4!}{(4-2)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = 12$ verschiedene Stichproben gibt. Ohne Berücksichtigung der Reihenfolge sind es lediglich 6 verschiedene Stichproben.

3 Ein “fares” Spiel

Ist bei einem Würfel die Wahrscheinlichkeit, daß eine bestimmte Zahl fällt, für alle Zahlen gleich, spricht der Statistiker von einem fairen Würfel. Dies gilt ebenso für eine Münze.

²Für eine Einführung in die Kombinatorik siehe z.B. Casella und Berger (1990), S. 13ff.

Wir sprechen von einem fairen Euro, wenn die Wahrscheinlichkeiten bei einem Münzwurf für Kopf und Zahl gleich sind.

Wie sieht also die Situation des Kartensammlers bei gleichen “Füll”-Wahrscheinlichkeiten aus?

Tritt jedes Element mit der gleichen Wahrscheinlichkeit auf, treten auch die verschiedenen Stichproben mit gleicher Wahrscheinlichkeit auf, es ist

$$p(s) = \frac{1}{\binom{N}{n}}, s \in \mathcal{S}, \text{ ungeordnet, d.h. ohne Beachtung der Reihenfolge,} \quad (3)$$

$$p(s) = \frac{1}{n! \binom{N}{n}}, s \in \mathcal{S}, \text{ geordnet, d.h. mit Beachtung der Reihenfolge.} \quad (4)$$

Wird die Reihenfolge nicht beachtet, kann sich wiederum jedes Element nur in $\binom{N-1}{n-1}$ Stichproben befinden, da ein Element vorgegeben ist und die restlichen $n-1$ Elemente wiederum aus $N-1$ Elementen zufällig gezogen werden. Die Wahrscheinlichkeit π_i , eine bestimmte Einheit i für $i = 1, 2, \dots, N$ in einer ungeordneten Stichprobe vom Umfang n zu finden, läßt sich angeben mit

$$\pi_i = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}, \quad i = 1, 2, \dots, N, \quad (5)$$

vgl. auch Hájek (81), S. 51.

Wird hingegen die Reihenfolge beachtet, gibt es $(n-1)! \binom{N-1}{n-1} = 1 \cdot (N-1) \cdot (N-2) \cdots (N-n+1)$ günstige Stichproben, die ein bestimmtes Element i für $i = 1, 2, \dots, N$ enthalten. Die Wahrscheinlichkeit π_i , eine bestimmte Einheit i für $i = 1, 2, \dots, N$ in einer geordneten Stichprobe vom Umfang n zu finden, lautet daher

$$\pi_i = n \frac{(n-1)! \binom{N-1}{n-1}}{n! \binom{N}{n}} = \frac{n}{N}, \quad i = 1, 2, \dots, N. \quad (6)$$

Bei Auswahl ohne Zurücklegen mit gleichen Ziehungswahrscheinlichkeiten gibt es bezüglich der Wahrscheinlichkeiten, mit denen sich eine Einheit in einer Stichprobe befindet, keine Unterschiede zwischen geordneten und ungeordneten Stichproben.

Betrachten wir nun die Tüten. Wenn wir von einem gleichwahrscheinlichen Bestückungsalgorithmus ausgehen, beträgt die Wahrscheinlichkeit, zufällig eine Tüte zu kaufen, in der sich ein bestimmtes Bild befindet, gerade

$$\pi_i = \frac{n}{N} = \frac{7}{498} = 0.014056 \text{ für } i = 1, 2, \dots, 498 \text{ Bilder.} \quad (7)$$

Nun interessiert die Frage, wie lange man im Durchschnitt Tüten kaufen muß, bis ein bestimmtes Bild gefunden ist.³

Eine geeignete Modellierung hierfür bietet die geometrische Verteilung. Sie gibt an, wieviele Versuche (d.h. Tütenkäufe) mit einer gewissen “Erfolgswahrscheinlichkeit” (d.h., daß ein bestimmtes Bild enthalten ist) durchzuführen sind, bis endlich auch ein “Erfolg” eintritt.⁴

Es sei X eine Zufallsvariable, die die Mißerfolge zählt bis zum ersten Kaufakt einer Tüte, die einen Erfolg bringt, d.h. das gesuchte Bild enthält. Dann beschreibt p die Erfolgswahrscheinlichkeit, daß ein bestimmtes Bild in einer Tüte enthalten ist,

$$p = \pi_i = n/N, i = 1, 2, \dots, N.$$

Für die Wahrscheinlichkeit, daß der x -te Kauf endlich das gesuchte Bild enthält, ergibt sich

$$P(X = x) = p(1 - p)^{x-1} = \frac{n}{N} \left(1 - \frac{n}{N}\right)^{x-1} \quad \text{für } x = 1, 2, \dots \quad (8)$$

Für die durchschnittlich zu erwartenden Kaufakte gilt

$$E(X) = \frac{1}{p}. \quad (9)$$

Eine besondere Eigenschaft zeichnet die geometrische Verteilung aus, sie hat “kein Gedächtnis”. Dies bedeutet, daß die Wahrscheinlichkeit, nach bereits t_1 getätigten erfolglosen Käufen weitere erfolglose t_2 Käufe zu tun, gerade der Wahrscheinlichkeit entspricht, erfolglose t_2 Käufe zu tun, d.h.

$$P(X > t_1 + t_2 | X > t_1) = P(X > t_2). \quad (10)$$

Diese Eigenschaft entspricht auch unserem natürlichen Verständnis des Kartensammelns. Beim Kaufakt weiß der Verkäufer weder, welche Karten der Käufer bereits hat, noch welche Karten in den Tüten sind. Analog zur Situation beim Roulette erhöhen viele Fehlkäufe die Chance auf ein “gutes” Päckchen mit der gesuchten Karte nicht.

Schließlich können wir noch die Wahrscheinlichkeit angeben, mehr als k erfolglose Käufe zu tätigen:

$$P(X > k) = 1 - P(X \leq k) = 1 - \sum_{x=1}^k p(1 - p)^{x-1}. \quad (11)$$

³Beim Coupon Collector’s Problem betrachtet man die Anzahl der insgesamt nötigen Käufe von einzelnen Bildern, bis alle Bilder gesammelt sind. Die zusätzliche Komplexität der 7 unterschiedlichen Bilder in den Tüten wird dort nicht berücksichtigt, vgl. dazu Motwani und Raghavan (1995).

⁴Für weitere Einzelheiten und Beweise siehe hierzu beispielsweise Casella und Berger (1990), S. 98f.

Für unsere Tüten bedeutet dies, daß im Durchschnitt $1/p = N/n = 498/7 = 71.1428$ Käufe notwendig sind, um eine bestimmte Karte zu erhalten. Die Wahrscheinlichkeit, mehr als 100 erfolglose Käufe zu tätigen, liegt beispielsweise noch bei

$$P(X > 100) = 1 - \sum_{x=1}^{100} 0.014056(1 - 0.014056)^{x-1} = 1 - 0.7606315 = 0.2393685$$

also bei etwa 24%.

Erst nach 210 Käufen wird man mit einer Wahrscheinlichkeit von etwa 95% eine Tüte kaufen, die das gesuchte Bild enthält. Voraussetzung dafür war, daß die Bilder mit gleichen Wahrscheinlichkeiten in die Tüten kommen. Wie verändert sich nun diese Situation, wenn die Tüten mit ungleichen Wahrscheinlichkeiten bestückt werden? Dies soll der folgende Abschnitt klären.

4 Ein Lösungsansatz

Wir betrachten nun die Ziehung von geordneten Stichproben ohne Zurücklegen vom gleichen Umfang n aus einer Grundgesamtheit vom Umfang N . Die Menge \mathcal{S} der möglichen Stichproben ist $\mathcal{S} = \{s_1, s_2, \dots, s_{n! \binom{N}{n}}\}$. Um die Wahrscheinlichkeiten bestimmen zu können, mit denen bestimmte Elemente in den Stichproben auftreten, ist es sinnvoll, sog. Auswahl- und Einschlußwahrscheinlichkeiten zu definieren.

4.1 Auswahl- und Einschlußwahrscheinlichkeiten

Die Berechnung der Auswahl- und Einschlußwahrscheinlichkeiten kann sich als schwierig erweisen, wenn die Einheiten ohne Zurücklegen für eine Stichprobe vom Umfang $n > 1$ gezogen werden. Es sei z_i die Auswahlwahrscheinlichkeit, die i -te Einheit, $i = 1, 2, \dots, N$, im ersten Zug in eine Stichprobe vom Umfang n aufzunehmen. Es gilt $\sum_{i=1}^N z_i = 1$.

Ferner beschreibe wieder

$$\pi_i = \sum_{\substack{s \in \mathcal{S} \\ i \in s}} p(s) \quad , i = 1, 2, \dots, N, \quad (12)$$

die Einschlußwahrscheinlichkeit, mit der sich die i -te Einheit in einer Stichprobe vom Umfang n befindet, wobei über alle Stichproben summiert wird, die die i -te Einheit enthalten.

Im weiteren seien die Wahrscheinlichkeiten $p(s) > 0$ für alle Stichproben $s \in \mathcal{S}$ vom Umfang n . Ebenso soll $z_i, \pi_i > 0$ für $i = 1, 2, \dots, N$ und alle Stichprobenumfänge $n = 1, 2, \dots, N$ gelten. Da die Einschlußwahrscheinlichkeiten von n abhängen, werden diese im folgenden mit $\pi_i(n)$ bezeichnet.

Bei Auswahl ohne Zurücklegen sind Auswahlwahrscheinlichkeiten und Einschlußwahrscheinlichkeiten im allgemeinen ungleich. Eine einmal gewählte Einheit kann nicht ein zweites Mal gezogen werden, so daß jede der Auswahlseinheiten höchstens einmal in der resultierenden Stichprobe vertreten ist. Dies bedingt, daß die vorherbestimmten Auswahlwahrscheinlichkeiten nicht aufrechterhalten werden können und im Verlauf der Auswahlprozedur monoton wachsen. Die für den nachfolgenden Auswahlvorgang verbleibende Grundgesamtheit schrumpft in dem Maße, wie Einheiten bereits ausgewählt und damit aus der noch verfügbaren Grundgesamtheit ausgeschieden sind.

Es wird also beim ersten Zug die i -te Einheit mit einer Auswahlwahrscheinlichkeit von z_i entnommen. Die Einschlußwahrscheinlichkeit, mit der sich die i -te Einheit in einer Stichprobe vom Umfang $n = 1$ befindet, ist dann $\pi_i(1) = z_i$. Beim zweiten Zug wird eine der verbleibenden Einheiten mit einer relativen Auswahlwahrscheinlichkeit gezogen, d.h. die j -te Einheit wird mit einer Wahrscheinlichkeit von $z_j/(1 - z_i)$ entnommen. Beim nächsten, dem dritten Zug wird dann die k -te Einheit mit einer Wahrscheinlichkeit von $z_k/(1 - z_i - z_j)$ ausgewählt. Weitere Züge werden analog durchgeführt.

Die Einschlußwahrscheinlichkeit $\pi_i(2)$, die i -te Einheit in einer Stichprobe vom Umfang $n = 2$ zu finden, berechnet sich aus der Wahrscheinlichkeit z_i , diese i -te Einheit im ersten Zug zu ziehen und aus der bedingten Wahrscheinlichkeit $\delta_i(2)$, die i -te Einheit erst im zweiten Zug zu ziehen. Um die nachfolgenden Formulierungen zu verallgemeinern, werden alle Einheiten i , die in einem k -ten Zug gezogen werden können, mit i_k bezeichnet. Die Wahrscheinlichkeit z_{i_k} hingegen gibt immer die Auswahlwahrscheinlichkeit an, die Einheit i_k im ersten Zug zu ziehen. Es ist also

$$\delta_i(2) = \sum_{\substack{i_1=1 \\ i_1 \neq i}}^N z_{i_1} \frac{z_i}{1 - z_{i_1}} = \sum_{\substack{i_1=1 \\ i_1 \neq i}}^N z_i \frac{z_{i_1}}{1 - z_{i_1}} \quad , i = 1, 2, \dots, N,$$

die Wahrscheinlichkeit, daß die i -te Einheit beim zweiten Zug in die Stichprobe kommt. Somit ergibt sich für

$$\pi_i(2) = z_i + \sum_{\substack{i_1=1 \\ i_1 \neq i}}^N z_i \frac{z_{i_1}}{1 - z_{i_1}} \quad , i = 1, 2, \dots, N. \quad (13)$$

Analoges gilt für eine Stichprobe vom Umfang $n = 3$.

$$\delta_i(3) = z_i \sum_{\substack{i_1=1 \\ i_1 \neq i}}^N \left(\frac{z_{i_1}}{1 - z_{i_1}} \sum_{\substack{i_2=1 \\ i_2 \neq i, i_1}}^N \frac{z_{i_2}}{1 - z_{i_1} - z_{i_2}} \right) \quad , i = 1, 2, \dots, N,$$

ist die Wahrscheinlichkeit, daß die i -te Einheit im dritten Zug in die Stichprobe kommt.

Für die Einschlußwahrscheinlichkeit gilt dann

$$\begin{aligned}\pi_i(3) &= z_i + \sum_{\substack{i_1=1 \\ i_1 \neq i}}^N z_i \frac{z_{i_1}}{1 - z_{i_1}} + \sum_{\substack{i_1=1 \\ i_1 \neq i}}^N \left(z_i \frac{z_{i_1}}{1 - z_{i_1}} \sum_{\substack{i_2=1 \\ i_2 \neq i, i_1}}^N \frac{z_{i_2}}{1 - z_{i_1} - z_{i_2}} \right) \\ &= \pi_i(2) + \delta_i(3) \quad , i = 1, 2, \dots, N.\end{aligned}\tag{14}$$

Auf diese Weise läßt sich die Einschlußwahrscheinlichkeit $\pi_i(n)$, mit der sich die i -te Einheit in einer Stichprobe des Umfangs n befindet, allgemein formulieren und in ein rekursives Computerprogramm umsetzen.

Es ist

$$\pi_i(n) = \pi_i(n-1) + \delta_i(n)\tag{15}$$

für $n = 2, 3, \dots, N$ und $i = 1, 2, \dots, N$ mit $\pi_{i_1}(1) := z_{i_1}$ und

$$\delta_i(n) = z_i \sum_{\substack{i_1=1 \\ i_1 \neq i}}^N \left(\frac{z_{i_1}}{1 - z_{i_1}} \sum_{\substack{i_2=1 \\ i_2 \neq i, i_1}}^N \left(\frac{z_{i_2}}{1 - z_{i_1} - z_{i_2}} \cdots \sum_{\substack{i_{n-1}=1 \\ i_{n-1} \neq i, i_1, \dots, i_{n-2}}}^N \frac{z_{i_{n-1}}}{1 - z_{i_1} - \cdots - z_{i_{n-1}}} \right) \right)$$

für $n = 2, 3, \dots, N$ und $\delta_i(1) := z_i$.

Die gezeigten Zusammenhänge zwischen Auswahl- und Einschlußwahrscheinlichkeiten verdeutlichen, wie mit steigendem Stichprobenumfang n die Berechnung der $\pi_i(n)$ immer aufwendiger wird. Weiteres ist ausführlich in Rässler (1996) erläutert.

4.2 Ein einfaches Beispiel

Betrachten wir nochmal eine kleine, endliche Grundgesamtheit von $N = 4$ unterschiedlichen (und identifizierbaren) Einheiten, aus der eine Stichprobe vom Umfang $n = 2$ ohne Zurücklegen gezogen wird. Die Auswahlwahrscheinlichkeiten werden nach dem oft zitierten Beispiel von Yates und Grundy⁵ gewählt, mit $z_1 = 0.1, z_2 = 0.2, z_3 = 0.3$ und $z_4 = 0.4$. Die Menge \mathcal{S} aller Stichproben enthält $n! \binom{N}{n} = 12$ mögliche Stichproben. Die Wahrscheinlichkeit für eine bestimmte geordnete Stichprobe $s = (i, j)$ beträgt

$$p(s) = p(i, j) = z_i \frac{z_j}{1 - z_i} \quad \text{für } i, j = 1, 2, 3, 4, i \neq j.$$

⁵Siehe dazu Cochran (1977), S. 259.

Die Tabelle 2 zeigt die Verteilung dieses Stichprobenmusters:

Tabelle 2: Stichprobenmuster für $N = 4$ und $n = 2$

s	$p(s)$	s	$p(s)$
(1,2)	0.0222	(3,1)	0.0429
(1,3)	0.0333	(3,2)	0.0857
(1,4)	0.0444	(3,4)	0.1714
(2,1)	0.0250	(4,1)	0.0667
(2,3)	0.0750	(4,2)	0.1333
(2,4)	0.1000	(4,3)	0.2000

Die Einschlußwahrscheinlichkeiten $\pi(2)$ werden dann wie folgt berechnet:

$$\begin{aligned}
 \pi_1 &= \sum_{\substack{s \in \mathcal{S} \\ 1 \in s}} p(s) \\
 &= 0.0222 + 0.0333 + 0.0444 + 0.0250 + 0.0429 + 0.0667 \\
 &= 0.2345 \text{ bzw.} \\
 &= 0.1 + 0.1 \frac{0.2}{1 - 0.2} + 0.1 \frac{0.3}{1 - 0.3} + 0.1 \frac{0.4}{1 - 0.4}, \\
 \pi_2 &= 0.4413, \\
 \pi_3 &= 0.6083, \\
 \pi_4 &= 0.7159.
 \end{aligned}$$

4.3 Wieviele noch?

Zur Bestimmung der Erfolgswahrscheinlichkeit, eine bestimmte Karte in einer Tüte, d.h. Stichprobe, vom Umfang $n = 7$ zu finden, ist die Berechnung der Einschlußwahrscheinlichkeit $\pi_j(7)$ zu implementieren bei beliebig gewählten Auswahlwahrscheinlichkeiten. Für diese Berechnungen existieren keine fertigen Softwarelösungen, so daß eine eigene Routine implementiert werden muß.⁶

⁶Wir wollen schlußendlich eine Java-Lösung anbieten, so daß jeder interessierte Sammler selber an den "Stellschrauben" drehen und die Erfahrungen aus seiner Sammelleidenschaft verwenden kann.

Es gilt:

$$\begin{aligned}
\pi_j(\mathbb{T}) &= z_j + z_j \sum_{\substack{i_1=1 \\ i_1 \neq j}}^N \frac{z_{i_1}}{1 - z_{i_1}} + z_j \sum_{\substack{i_1=1 \\ i_1 \neq j}}^N \left(\frac{z_{i_1}}{1 - z_{i_1}} \sum_{\substack{i_2=1 \\ i_2 \neq j, i_1}}^N \frac{z_{i_2}}{1 - z_{i_1} - z_{i_2}} \right) \\
&+ z_j \sum_{\substack{i_1=1 \\ i_1 \neq j}}^N \left(\frac{z_{i_1}}{1 - z_{i_1}} \sum_{\substack{i_2=1 \\ i_2 \neq j, i_1}}^N \left(\frac{z_{i_2}}{1 - z_{i_1} - z_{i_2}} \sum_{\substack{i_3=1 \\ i_3 \neq j, i_1, i_2}}^N \frac{z_{i_3}}{1 - z_{i_1} - z_{i_2} - z_{i_3}} \right) \right) \\
&+ z_j \sum_{\substack{i_1=1 \\ i_1 \neq j}}^N \left(\frac{z_{i_1}}{1 - z_{i_1}} \sum_{\substack{i_2=1 \\ i_2 \neq j, i_1}}^N \left(\frac{z_{i_2}}{1 - z_{i_1} - z_{i_2}} \sum_{\substack{i_3=1 \\ i_3 \neq j, i_1, i_2}}^N \left(\frac{z_{i_3}}{1 - z_{i_1} - z_{i_2} - z_{i_3}} \sum_{\substack{i_4=1 \\ i_4 \neq j, i_1, i_2, i_3}}^N \frac{z_{i_4}}{1 - z_{i_1} - z_{i_2} - z_{i_3} - z_{i_4}} \right) \right) \right) \\
&+ z_j \sum_{\substack{i_1=1 \\ i_1 \neq j}}^N \left(\frac{z_{i_1}}{1 - z_{i_1}} \sum_{\substack{i_2=1 \\ i_2 \neq j, i_1}}^N \left(\frac{z_{i_2}}{1 - z_{i_1} - z_{i_2}} \sum_{\substack{i_3=1 \\ i_3 \neq j, i_1, i_2}}^N \left(\frac{z_{i_3}}{1 - z_{i_1} - z_{i_2} - z_{i_3}} \sum_{\substack{i_4=1 \\ i_4 \neq j, i_1, i_2, i_3}}^N \left(\frac{z_{i_4}}{1 - z_{i_1} - z_{i_2} - z_{i_3} - z_{i_4}} \right) \right) \right) \right) \\
&+ z_j \sum_{\substack{i_5=1 \\ i_5 \neq j, i_1, i_2, i_3, i_4}}^N \frac{z_{i_5}}{1 - z_{i_1} - z_{i_2} - z_{i_3} - z_{i_4} - z_{i_5}} \\
&+ z_j \sum_{\substack{i_1=1 \\ i_1 \neq j}}^N \left(\frac{z_{i_1}}{1 - z_{i_1}} \sum_{\substack{i_2=1 \\ i_2 \neq j, i_1}}^N \left(\frac{z_{i_2}}{1 - z_{i_1} - z_{i_2}} \sum_{\substack{i_3=1 \\ i_3 \neq j, i_1, i_2}}^N \left(\frac{z_{i_3}}{1 - z_{i_1} - z_{i_2} - z_{i_3}} \sum_{\substack{i_4=1 \\ i_4 \neq j, i_1, i_2, i_3}}^N \left(\frac{z_{i_4}}{1 - z_{i_1} - z_{i_2} - z_{i_3} - z_{i_4}} \right) \right) \right) \right) \\
&+ z_j \sum_{\substack{i_5=1 \\ i_5 \neq j, i_1, i_2, i_3, i_4}}^N \left(\frac{z_{i_5}}{1 - z_{i_1} - z_{i_2} - z_{i_3} - z_{i_4} - z_{i_5}} \sum_{\substack{i_6=1 \\ i_6 \neq j, i_1, i_2, i_3, i_4, i_5}}^N \frac{z_{i_6}}{1 - z_{i_1} - z_{i_2} - z_{i_3} - z_{i_4} - z_{i_5} - z_{i_6}} \right)
\end{aligned}$$

Unterstellen wir der Einfachheit halber, daß eines der 498 Bilder, es sei die Nummer j , etwa die Ziehungswahrscheinlichkeit $z_j = 0.5 \frac{1}{N} = 0.5/498 = 0.001$ (wir verwenden hier die Rundung 0.001) hat, die anderen wiederum mit gleichen Wahrscheinlichkeiten zu je $z_i = (1 - 0.001)/497 = 0.00201$, für $i = 1, 2, \dots, N, i \neq j$ produziert werden.

Aufgrund der ineinander geschachtelten Summen, muß der Programmcode extra optimiert werden. Mit der an sich relativ leistungsstarken Statistik-Software S-PLUS, ist zwar die Programmierung der Einschlußwahrscheinlichkeiten einfach, jedoch aus Speichergründen keine Berechnung über $N = 10$ hinaus mehr möglich. Die rekursive Lösung erlaubt zwar große N Werte, liefert jedoch kein Ergebnis mehr für Werte über $n = 5$. Wir entwickeln daher eine schnelle C++ Lösung.

Für die Einschlußwahrscheinlichkeiten finden wir schließlich folgendes Ergebnis:

$$\pi_j(7) = 0.007022622 \tag{16}$$

$$\pi_i(7) = 0.01407163 \quad i = 1, 2, \dots, N, i \neq j, \tag{17}$$

Damit läßt sich eine geometrische Verteilung formulieren für die Anzahl der Mißerfolge bis zum ersten Treffer, d.h. der ersten Tüte mit der gesuchten Karte j , die Trefferwahrscheinlichkeit ist gerade $\pi_j(7)$.

Gemäß (9) sind durchschnittlich $1/p = 1/0.007022622 = 142.3970 \approx 142$ Käufe zu tätigen, bis endlich das gesuchte Bild j enthalten ist.

Die Wahrscheinlichkeit, mehr als 200 erfolglose Käufe nach dem Bild j zu tätigen, beträgt

$$P(X > 200) = 1 - \sum_{x=1}^{200} 0.007022622(1 - 0.007022622)^{x-1} = 0.242555$$

also wiederum etwa 24%.

Um allerdings mit einer 95% Wahrscheinlichkeit, genau dieses Bild zu erhalten, müssen nach (11) 425 Käufe getätigt werden. Um eine Sicherheit von 99% für den Einkauf dieses fehlenden Bildes zu garantieren, müssen etwa 653 Kaufakte getätigt werden.

5 Zusammenfassung

Unter Verwendung der Theorie für Stichprobenziehungen mit unterschiedlichen Wahrscheinlichkeiten ohne Zurücklegen, gelingt es, die Wahrscheinlichkeit dafür zu berechnen, daß sich ein bestimmtes Element aus der Grundgesamtheit in einer Stichprobe befindet. Übertragen auf die Sammelbilder können so die Wahrscheinlichkeiten bestimmt werden, mit denen ein besonderes, dem Sammler fehlendes Bild in einer zufällig gekauften Tüte steckt. Unter Annahme gleicher Bestückungswahrscheinlichkeiten sind noch 211 Tüten zu kaufen, um mit einer Wahrscheinlichkeit von 95% ein fehlendes Bild in einer sieben-Bilder-Tüte zu finden.

Geht man davon aus, daß ein bestimmtes Bild beispielsweise nur ungefähr halb so oft wie alle anderen in die Tüten gepackt wird, erhöht sich diese Anzahl auf ca. das Doppelte, genauer auf 425 Kaufakte.

Danksagung

Unser besonderer Dank geht an Hendrik Rässler, Vorstand der raessler automation & consulting AG in Gründung, der eine optimierte und schnelle C++ Routine zur Berechnung dieser Einschlußwahrscheinlichkeiten erstellte. Die Routine kann aus dem Internet unter www.missing-data.de/downloads/kadi heruntergeladen werden.

Literaturverzeichnis

- Casella, G., Berger, R.L. (1990). *Statistical Inference*. Duxbury, Belmont, California. New York.
- Cochran, W.G. (1977). *Sampling Techniques*. John Wiley and Sons, New York.
- Hájek, J. (1981). *Sampling From a Finite Population*. Marcel Dekker, New York.
- Motwani, R., Raghavan, P. (1995). *Randomized Algorithms*. University Press, Cambridge.
- Rässler, S. (1996). *Stichprobenverfahren bei sukzessiver Auswahl mit unterschiedlichen Wahrscheinlichkeiten im Wirksamkeitsvergleich*. Vandenhoeck & Ruprecht, Göttingen.