# Wartezeitprobleme[1]

Norbert Henze

Karlsruher Institut für Technologie

Institut für Stochastik

Kaiserstr. 89, 76133 Karlsruhe

Norbert.Henze@kit.edu

**Zusammenfassung:** Wartezeitprobleme treten immer dann auf, wenn zu jedem der Zeitpunkte 1,2,... ein Zufallsexperiment durchgeführt und hierbei auf das erstmalige Eintreten eines vorher definierten Bedienungskomplexes gewartet wird. Beispiele hierfür sind das Warten auf vorgeschriebene Ergebnis-Sequenzen in Bernoulli-Ketten oder *Kollisions- und Sammlerprobleme* beim sequentiellen zufälligen Verteilen von Teilchen auf Fächer. In dieser Arbeit werden verschiedene elementare Wartezeitprobleme behandelt und Hinweise auf höhere Gesichstpunkte gegeben. Der besondere Reiz dieser Probleme liegt in der Notwendigkeit einer mathematischen Modellierung einfach formulierbarer konkreter Fragestellungen sowie im Zusammenspiel von Kombinatorik und Analysis.

[1] Diese Arbeit ist Grundlage einer Lehrerfortbildungsveranstaltung an der Universität Bielefeld am 18.02.1998

# 1 Warten auf den ersten Treffer in einer Bernoulli–Kette: die geometrische Verteilung

Ein Zufallsexperiment besitze die möglichen Ausgänge *Treffer* und *Niete*, welche mit den Wahrscheinlichkeiten $p$ bzw. $1-p$ auftreten; dabei ist $0 < p < 1$ vorausgesetzt. Das Experiment werde in unabhängiger Folge wiederholt. Die Zufallsvariable $W_1$ beschreibe die *Anzahl der Nieten vor dem ersten Treffer*. $W_1$ besitzt die durch

$$P(W_1 = k) \;=\; (1-p)^k \cdot p \tag{1.1}$$

($k = 0, 1, 2, \ldots$) definierte *geometrische Verteilung mit Parameter* $p$, und wir schreiben hierfür kurz $W_1 \sim \mathcal{G}(p)$. Erwartungswert und Varianz von $W_1$ sind durch

$$E(W_1) \;=\; \frac{1}{p} - 1 \tag{1.2}$$

und

$$V(W_1) \;=\; \frac{1-p}{p^2} \tag{1.3}$$

gegeben. Bild 1.1 zeigt Stabdiagramme geometrischer Verteilungen für $p = 1/2$ und $p = 1/4$.

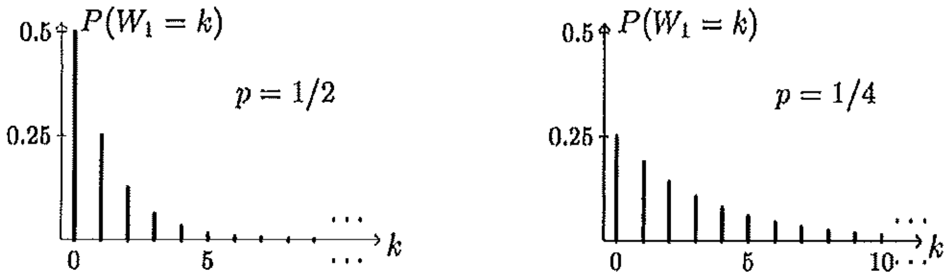


**Bild 1.1**: Stabdiagramme geometrischer Verteilungen

Wegen

$$\begin{aligned} P(W_1 = k + l \mid W_1 \geq k) \;&=\; \frac{P(W_1 = k + l, W_1 \geq k)}{P(W_1 \geq k)} \\ &=\; \frac{P(W_1 = k + l)}{P(W_1 \geq k)} \end{aligned}$$

$$
\begin{aligned}
&= \frac{(1-p)^{k+l} \cdot p}{(1-p)^k} \\
&= (1-p)^l \cdot p \\
&= P(W_1 = l)
\end{aligned}
$$

besitzt die geometrische Verteilung „kein Gedächtnis“: Werden $k$ Nieten beobachtet (Ereignis $\{W_1 \geq k\}$), so hat diese Information wahrscheinlichkeitstheoretisch keine Auswirkung auf den weiteren Verlauf des Wartezeitexperiments. Die bedingte Wahrscheinlichkeit, zu diesen $k$ noch weitere $l$ Nieten bis zum Auftreten des ersten Treffers zu beobachten, ist gleich der (nichtbedingten) Wahrscheinlichkeit, daß das Wartezeitexperiment mit dem Auftreten des ersten Treffers nach genau $l$ Nieten endet.

Die Eigenschaft $P(W_1 = k + l \mid W_1 \geq k) = P(W_1 = l)$ $(k, l \in \mathbb{N}_0)$ ist für die geometrische Verteilung in folgendem Sinn *charakteristisch*:

Ist $X$ eine $\mathbb{N}_0$–wertige Zufallsvariable mit $0 < P(X = 0) < 1$, welche die Eigenschaft

$$
P(X = k + l \mid X \geq k) = P(X = l) \qquad (k, l \in \mathbb{N}_0) \tag{1.4}
$$

der *Gedächtnislosigkeit* besitzt, so gibt es ein $p \in (0,1)$ mit $X \sim \mathcal{G}(p)$.

Zum Beweis schreiben wir kurz $p_j := P(X = j)$ und setzen $k = 1$ in (1.4). Es ergibt sich

$$
\frac{P(X = l + 1, X \geq 1)}{P(X \geq 1)} = P(X = l) \qquad (l \in \mathbb{N}_0)
$$

und somit wegen $P(X \geq 1) = 1 - p_0$ die Beziehung $p_{l+1} = p_l \cdot (1 - p_0)$ $(l \in \mathbb{N}_0)$. Wegen $0 < p_0 < 1$ folgt hieraus $p_l > 0$ für jedes $l \geq 0$, also

$$
\frac{p_{l+1}}{p_l} = 1 - p_0 \qquad (l \in \mathbb{N}_0).
$$

Wir erhalten somit

$$
p_j = \left( \prod_{l=0}^{j-1} \frac{p_{l+1}}{p_l} \right) \cdot p_0 = (1 - p_0)^j \cdot p_0 \qquad (j \in \mathbb{N}_0),
$$

so daß $X$ die geometrische Verteilung $\mathcal{G}(p_0)$ besitzt.

Für Knobelfüchse

- Konstruktion eines Wahrscheinlichkeitsraumes als Definitionsbereich für $W_1$.
- Herleitung der Beziehungen (1.2) und (1.3).
- Eine Urne enthält $r$ rote und $s$ schwarze Kugeln. Es werden rein zufällig nacheinander *ohne* Zurücklegen Kugeln gezogen, bis zum ersten Mal eine rote Kugel erscheint. Welche Verteilung besitzt die Anzahl der hierfür benötigten Züge?
- Ein echter Würfel wird solange geworfen, bis die erste Sechs auftritt. Wie groß ist die Wahrscheinlichkeit, vorher genau eine Vier zu werfen?
- Anja (A) und Bettina (B) drehen in unabhängiger Folge abwechselnd ein Glücksrad mit den Sektoren A und B. Das Glücksrad bleibe mit der Wahrscheinlichkeit $p$ (bzw. $1-p$) im Sektor A (bzw. B) stehen. Wie groß ist die Gewinnwahrscheinlichkeit für Anja, wenn sie beginnt? Für welchen Wert von $p$ besitzen beide Spielerinnen die gleiche Gewinnwahrscheinlichkeit?

# 2 Warten auf den $r$–ten Treffer: die negative Binomialverteilung

In Verallgemeinerung der bisher behandelten Fragestellung beschreibe die Zufallsvariable $W_r$ die *Anzahl der Nieten vor dem r–ten Treffer* in einer Bernoulli–Kette mit Trefferwahrscheinlichkeit $p$. Das Ereignis $\{W_r = k\}$ tritt genau dann ein, wenn der $(k+r)$–te Versuch den $r$–ten Treffer ergibt. Da jede spezielle Ergebnisfolge aus $r$ Treffern und $k$ Nieten die Wahrscheinlichkeit $p^r \cdot (1-p)^k$ besitzt und es $\binom{k+r-1}{k}$ Möglichkeiten der Verteilung von $k$ Nieten auf die $k+r-1$ Versuche vor dem (beim $(k+r)$–ten Mal auftretenden) $r$–ten Treffer gibt, gilt

$$P(W_r = k) \;=\; \binom{k+r-1}{k} \cdot p^r \cdot (1-p)^k, \qquad k \in \mathbb{N}_0. \tag{2.1}$$

Die durch (2.1) auf den nichtnegativen ganzen Zahlen definierte Verteilung heißt *negative Binomialverteilung mit Parametern* $r$ und $p$, kurz:

$$W_r \;\sim\; \mathcal{N}b(r,p).$$

Ihre Namensgebung verdankt die Verteilung $\mathcal{N}b(r,p)$ der Darstellung

$$P(W_r = k) = \binom{-r}{k} \cdot p^r \cdot (-(1-p))^k,$$

wobei der Binomialkoeffizient $\binom{z}{k}$ für $z \in \mathbb{R}$ und $k \in \mathbb{N}_0$ durch

$$\binom{z}{k} = \frac{z \cdot (z-1) \cdot \ldots \cdot (z-k+1)}{k!} \qquad (k \in \mathbb{N})$$

und $\binom{z}{0} = 1$ definiert ist.

Der Nachweis der Normierungseigenschaft $\sum_{k=0}^{\infty} P(W_r = k) = 1$ erfolgt am einfachsten über die Ableitungsformeln

$$\sum_{s=k}^{\infty} s^{\underline{k}} \cdot x^{s-k} = \frac{d^k}{dx^k}\left(\sum_{s=0}^{\infty} x^s\right) = \frac{d^k}{dx^k}\frac{1}{1-x} = \frac{k!}{(1-x)^{k+1}} \tag{2.2}$$

($|x| < 1$) für die geometrische Reihe. Hierbei steht allgemein

$$y^{\underline{m}} := y \cdot (y-1) \cdot (y-2) \cdot \ldots \cdot (y-m+1)$$

($y \in \mathbb{R}$, $m \in \mathbb{N}_0$, $y^{\underline{0}} := 1$) für die $m$–te *untere Faktorielle* von $y$. Unter Benutzung der Symmetriebeziehung $\binom{n}{j} = \binom{n}{n-j}$ für die Binomialkoeffizienten ergibt sich dann mit der Substitution $k = s - (r-1)$

$$\begin{aligned}
\sum_{k=0}^{\infty} P(W_r = k) &= p^r \cdot \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} \cdot (1-p)^k \\
&= \frac{p^r}{(r-1)!} \cdot \sum_{s=r-1}^{\infty} s^{\underline{r-1}} \cdot (1-p)^{s-(r-1)} \\
&= \frac{p^r}{(r-1)!} \cdot \frac{(r-1)!}{(1-(1-p))^r} \\
&= 1.
\end{aligned}$$

Tiefere Einsichten in die Struktur der negativen Binomialverteilung $\mathcal{N}b(r,p)$ erhält man durch Aufspaltung der Anzahl der Nieten vor dem $r$–ten Treffer in die Anzahl der Nieten vor dem ersten Treffer und die Anzahl der Nieten zwischen dem $(j-1)$–ten und dem $j$–ten Treffer ($j = 2,\ldots,r$). Aufgrund der Unabhängigkeit aller Versuche ist es intuitiv naheliegend, daß diese „Nieten–Anzahlen“ stochastisch unabhängige Zufallsvariablen mit gleicher geometrischer Verteilung $\mathcal{G}(p)$ sind, also $W_r = \sum_{i=1}^{r} X_i$ mit unabhängigen und je $\mathcal{G}(p)$ – verteilten Zufallsvariablen $X_1, \ldots, X_r$.

Sind umgekehrt $X_1, X_2, \ldots, X_r$ unabhängige Zufallsvariablen mit derselben geometrischen Verteilung $\mathcal{G}(p)$, so gilt:

$$X_1 + X_2 + \ldots + X_r \sim \mathcal{N}b(r,p). \tag{2.3}$$

Zum Beweis von (2.3) beachten wir, daß das Ereignis $\{X_1 + \ldots + X_r = k\}$ die Vereinigung der für verschiedene $r$–Tupel $(i_1, \ldots, i_r) \in \mathbb{N}_0^r$ mit $i_1 + \ldots + i_r = k$ unvereinbaren Ereignisse $\{X_1 = i_1, X_2 = i_2, \ldots, X_r = i_r\}$ darstellt. Da jedes dieser Ereignisse dieselbe Wahrscheinlichkeit

$$\begin{aligned} P(X_1 = i_1, X_2 = i_2, \ldots, X_r = i_r) &= \prod_{l=1}^{r} P(X_l = i_l) \\ &= \prod_{l=1}^{r} \left((1-p)^{i_l} \cdot p\right) \\ &= (1-p)^k \cdot p^r \end{aligned}$$

besitzt, folgt

$$\begin{aligned} P(X_1 + \ldots + X_r = k) &= \sum_{\substack{(i_1,\ldots,i_r) \in \mathbb{N}_0^r, \\ i_1+\ldots+i_r=k}} P(X_1 = i_1, \ldots, X_r = i_r) \\ &= (1-p)^k \cdot p^r \cdot \left|\{(i_1, \ldots, i_r) \in \mathbb{N}_0^r : i_1 + \ldots + i_r = k\}\right| \end{aligned}$$

und somit wegen

$$\left|\{(i_1, \ldots, i_r) \in \mathbb{N}_0^r : i_1 + \ldots + i_r = k\}\right| = \binom{k+r-1}{k} \tag{2.4}$$

die Behauptung.

Als Folgerung aus (2.3) und (1.2), (1.3) erhält man

$$E(W_r) = r \cdot \left(\frac{1}{p} - 1\right), \qquad V(W_r) = r \cdot \frac{1-p}{p^2} \tag{2.5}$$

für den Erwartungswert und die Varianz einer Zufallsvariablen $W_r$ mit Verteilung $\mathcal{N}b(r,p)$.

Für Knobelfüchse

- Weisen Sie Gleichung (2.4) nach.
- Konstruieren Sie einen Wahrscheinlichkeitsraum als Definitionsbereich für $W_r$ und $X_1, \ldots, X_r$.
- Welche Verteilung besitzt (in der oben beschriebenen Situation) die Zufallsvariable $X_1$ unter Bedingung $X_1 + X_2 = k$?

# 3 Warten auf den ersten Doppeltreffer und die Fibonacci–Zahlen

Wir betrachten jetzt speziell eine Bernoulli–Kette mit Trefferwahrscheinlichkeit 1/2, also z.B. den Wurf einer echten Münze, und fragen nach der Verteilung der zufälligen Anzahl $D_2$ von Versuchen, die nötig sind, bis *erstmalig direkt hintereinander 2 Treffer* auftreten. Im Gegensatz zu früher sollen hier die beiden letzten, jeweils einen Treffer ergebenden Versuche mitgezählt werden. Symbolisieren wir einen Treffer mit 1 und eine Niete mit 0, so nimmt $D_2$ z.B. für die Ergebnisfolge 0 1 0 0 0 1 0 1 1 den Wert 9 an.

Bezeichnet allgemein $\{a_1\, a_2 \ldots a_l\}$ das Ereignis, daß die Bernoulli–Kette mit der Ergebnisfolge $a_1, \ldots, a_l$ $(a_j \in \{0,1\})$ beginnt, so gilt offenbar

$$P(D_2 = 2) \;=\; P(\{11\}) \;=\; \frac{1}{4}\,, \tag{3.1}$$

$$P(D_2 = 3) \;=\; P(\{011\}) \;=\; \frac{1}{8}\,, \tag{3.2}$$

sowie

$$P(D_2 = 4) \;=\; P(\{0011\}) + P(\{1011\}) \;=\; \frac{2}{16} \;=\; \frac{1}{8}.$$

Zur Herleitung eines allgemeinen Ausdrucks für $P(D_2 = n)$ zerlegen wir im Fall $n \geq 4$ das Ereignis $\{D_2 = n\}$ nach den drei disjunkten Möglichkeiten $\{0\}$, $\{10\}$ und $\{11\}$ für den Beginn der Bernoulli–Kette. Es gilt

$$P(D_2 = n \mid \{0\}) \;=\; P(D_2 = n-1),$$

da das Ereignis $\{D_2 = n\}$ bei einer Niete im ersten Versuch nach genau $n-1$ der *weiteren* Versuche eintritt und da die ab dem 2. Versuch beginnende Wartezeit wieder verteilt ist wie die ab dem 1. Versuch beginnende. In gleicher Weise erhalten wir

$$P(D_2 = n \mid \{10\}) \;=\; P(D_2 = n-2).$$

Wegen $P(D_2 = n \mid \{11\}) = 0$ (im Falle eines Doppeltreffers in den beiden ersten Versuchen kann das Ereignis $D_2 = n$ für $n \geq 4$ nicht eintreten!) liefert die Formel von der totalen Wahrscheinlichkeit die Rekursionsbeziehung

$$P(D_2 = n) \;=\; P(D_2 = n-1) \cdot P(\{0\}) + P(D_2 = n-2) \cdot P(\{10\}) \qquad (n \geq 4)$$

oder kürzer

$$q_n \;=\; \frac{1}{2} \cdot q_{n-1} + \frac{1}{4} \cdot q_{n-2} \qquad (n \geq 4), \tag{3.3}$$

wobei $q_n := P(D_2 = n)$ gesetzt wurde. Mit den Anfangsbedingungen $q_2 = 1/4$, $q_3 = 1/8$ und (3.3) ist die Folge $(q_n)_{n\geq 2}$ festgelegt.

Der Ansatz zur Lösung der Differenzengleichung (3.3) ist

$$q_n \;=\; \gamma \cdot x^n \tag{3.4}$$

mit geeigneten Größen $\gamma$ und $x$. Einsetzen von (3.4) in (3.3) liefert nach Division durch $x^{n-2}$ und $\gamma$ die quadratische Gleichung $x^2 - x/2 - 1/4 = 0$, welche die beiden Lösungen

$$x_1 \;=\; \frac{1+\sqrt{5}}{4}\,, \qquad x_2 \;=\; \frac{1-\sqrt{5}}{4} \tag{3.5}$$

besitzt. Nach Konstruktion genügt die durch

$$q_n \;:=\; \gamma_1 \cdot x_1^n + \gamma_2 \cdot x_2^n \tag{3.6}$$

definierte Folge $(q_n)_{n\geq 1}$ der Rekursionsformel (3.3). Die Konstanten $\gamma_1, \gamma_2$ bestimmen sich aus den Anfangsbedingungen

$$\frac{1}{4} = \gamma_1 \cdot x_1^2 + \gamma_2 \cdot x_2^2\,, \qquad \frac{1}{8} = \gamma_1 \cdot x_1^3 + \gamma_2 \cdot x_2^3$$

zu

$$\gamma_1 \;=\; \frac{2}{\sqrt{5}\cdot(\sqrt{5}+1)}\,, \qquad \gamma_2 \;=\; \frac{2}{\sqrt{5}\cdot(\sqrt{5}-1)}\,. \tag{3.7}$$

Insgesamt folgt durch Einsetzen von (3.5), (3.7) in (3.6) und Zusammenfassen das Endergebnis

$$P(D_2 = n) \;=\; q_n \;=\; \frac{1}{\sqrt{5}} \cdot \left( \left( \frac{1+\sqrt{5}}{2} \right)^{n-1} - \left( \frac{1-\sqrt{5}}{2} \right)^{n-1} \right) \cdot \frac{1}{2^n} \tag{3.8}$$

($n \geq 2$), welches auch kompakter als

$$q_n \;=\; \frac{f_{n-2}}{2^n}\,, \qquad n \geq 2,$$

geschrieben werden kann. Hierbei ist $(f_n)_{n \geq 0}$ die durch die Anfangsbedingungen $f_0 = f_1 = 1$ und die Rekursionsformel $f_n = f_{n-1} + f_{n-2}$ ($n \geq 2$) definierte Folge der *Fibonacci-Zahlen* (siehe z.B. Heuser, 1991, S.377). Bild 3.1 zeigt das Stabdiagramm der Verteilung von $D_2$.

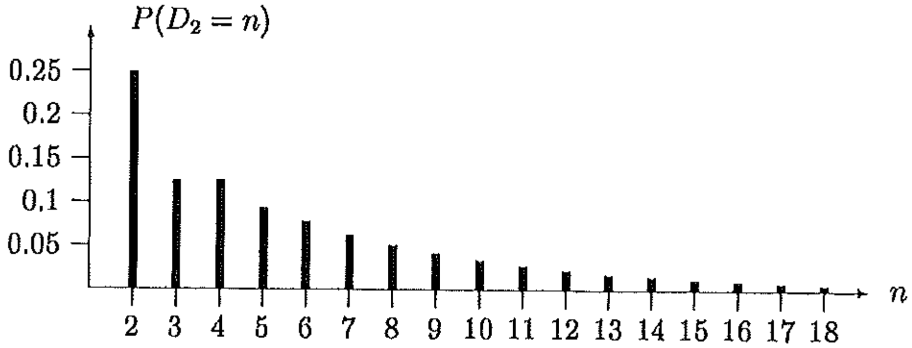


**Bild 3.1:** Verteilung der Versuchsanzahl bis zum ersten Doppeltreffer in einer Bernoulli-Kette mit Trefferwahrscheinlichkeit 1/2

Den Erwartungswert von $D_2$ erhält man entweder mit (3.8) über die Darstellungsformel $E(D_2) = \sum_{n=2}^{\infty} n \cdot q_n$ ((2.2) benutzen!) oder einfacher über die häufig als *Mittelwertsregel* bezeichnete und ein Analogon der Formel von der totalen Wahrscheinlichkeit darstellenden Beziehung

$$\begin{aligned} E(D_2) \;&=\; E(D_2\,|\,0) \cdot P(\{0\}) \;+\; E(D_2\,|\,10) \cdot P(\{10\}) \\ &\quad+\; E(D_2\,|\,11) \cdot P(\{11\}). \end{aligned} \tag{3.9}$$

Hier steht allgemein $E(D_2\,|\,a_1 \ldots a_l)$ für den Erwartungswert von $D_2$ unter der Bedingung, daß die Bernoulli-Kette mit der Ergebnisfolge $a_1 \ldots a_l$ beginnt. Setzen wir kurz $\mu := E(D_2)$, so geht (3.9) wegen

$$E(D_2\,|\,0) \;=\; 1 + \mu, \qquad E(D_2\,|\,10) \;=\; 2 + \mu, \qquad E(D_2\,|\,11) \;=\; 2$$

in

$$\mu = (1+\mu)\cdot\frac{1}{2} + (2+\mu)\cdot\frac{1}{4} + 2\cdot\frac{1}{4}$$

über, und es folgt $E(D_2) = 6$.

Für Knobelfüchse

- Welche Verteilung besitzt die Wartezeit bis zum ersten Doppeltreffer in einer Bernoulli-Kette mit Trefferwahrscheinlichkeit $p$?
- Wie groß ist der Erwartungswert der Wartezeit bis zum ersten Auftreten der Sequenz 111 (Dreifachtreffer) in einer Bernoulli-Kette mit Trefferwahrscheinlichkeit 1/2 ?
- Welche Verteilung besitzt die Wartezeit bis zum ersten Auftreten der Sequenz 10?

# 4 Warten auf die erste Kollision: die erste Gewinnreihenwiederholung im Lotto

Gegeben seien $n$ von 1 bis $n$ numerierte Fächer. Es wird zu jedem der *Zeitpunkte* $1, 2, \ldots$ ein Fach rein zufällig ausgewählt (Gleichverteilungsannahme!) und mit einem Teilchen besetzt, wobei diese Besetzungen stochastisch unabhängig voneinander erfolgen sollen. Wir interessieren uns für den zufälligen Zeitpunkt, zu dem erstmalig ein Teilchen in ein bereits besetztes Fach gelangt, betrachten folglich die Zufallsvariable

$K_n$ := Zeitpunkt der ersten *Kollision* beim sukzessiven rein zufälligen Besetzen von $n$ Fächern.

Offenbar nimmt $K_n$ die Werte $2, 3, \ldots, n+1$ an, und es gilt

$$P(K_n \geq k+1) = \frac{n\cdot(n-1)\cdot(n-2)\cdot\ \ldots\ \cdot(n-k+1)}{n^k} \tag{4.1}$$

für jedes $k = 1, 2, \ldots, n+1$. Um (4.1) einzusehen, beachte man, daß das Ereignis $\{K_n \geq k+1\}$ gleichbedeutend damit ist, daß bei der rein zufälligen Verteilung von

$k$ Teilchen auf $n$ Fächer alle Teilchen in verschiedene Fächer gelangen. Bei Annahme eines Laplace-Modells über dem Ergebnisraum aller $k$-Permutationen der Zahlen $1, 2, \ldots, n$ geben Zähler und Nenner in (4.1) gerade die Anzahl der günstigen Fälle ($k$-Permutationen ohne Wiederholung) bzw. die Anzahl aller möglichen Fälle ($k$-Permutationen mit Wiederholung) an.

Aus (4.1) folgt durch Komplement-Bildung

$$P(K_n \leq k) \;=\; 1 - \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right) \tag{4.2}$$

($k = 2, 3, \ldots, n+1$; $P(K_n \leq 1) = 0$).

Tabelle 4.1 enthält die Wahrscheinlichkeiten $P(K_n \leq k)$ in Abhängigkeit von $k$ für $n = 13\,983\,816 = \binom{49}{6}$. Dieser Fall, bei dem jeder möglichen Gewinnreihe (ohne Zusatzzahl) im Lotto 6 aus 49 ein „Fach" entspricht, hat am 29.6.1995 in der Presse Furore gemacht, als die erste Gewinnreihenwiederholung nach 3016 Ausspielungen als „sensationell früh" empfunden wurde (siehe Henze, 1997, Kapitel 10).

| $k$ | $P(K_n \leq k)$ | $k$ | $P(K_n \leq k)$ | $k$ | $P(K_n \leq k)$ |
|---|---|---|---|---|---|
| 500 | 0.0089 | 4500 | 0.5152 | 8500 | 0.9245 |
| 1000 | 0.0351 | 5000 | 0.5909 | 9000 | 0.9448 |
| 1500 | 0.0773 | 5500 | 0.6609 | 9500 | 0.9603 |
| 2000 | 0.1332 | 6000 | 0.7240 | 10000 | 0.9720 |
| 2500 | 0.2002 | 6500 | 0.7792 | 10500 | 0.9806 |
| 3000 | 0.2751 | 7000 | 0.8266 | 11000 | 0.9868 |
| 3500 | 0.3546 | 7500 | 0.8662 | 11500 | 0.9912 |
| 4000 | 0.4356 | 8000 | 0.8986 | 12000 | 0.9942 |

**Tabelle 4.1:** Wahrscheinlichkeiten für die erste Gewinnreihenwiederholung im Lotto 6 aus 49 ($n = 13\,983\,816$) nach höchstens $k$ Ausspielungen

Es mag überraschend erscheinen, daß bei fast 14 Millionen möglichen Tippreihen eine Kollision nach höchstens 3016 Ziehungen keinesfalls unwahrscheinlich ist (die exakte Wahrscheinlichkeit $P(K_n \leq 3016)$ berechnet sich mit (4.2) zu $0.2775\ldots$). Insbesondere kann schon darauf gewettet werden, daß die erste Gewinnreihenwiederholung nach höchstens 4500 Ausspielungen erfolgt. Der Grund hierfür ist, daß wir auf *irgendeine* und nicht auf eine *bestimmte* Kollision warten.

Um diesen wichtigen Unterschied zu verdeutlichen, beschreibe nun wieder allgemein die Zufallsvariable $T_j$ die Anzahl der Versuche, bis *Fach Nr.* $j$ zum *zweiten* Mal besetzt wird, also eine Kollision im $j$-ten Fach stattfindet. Interessieren wir uns nur für dieses spezielle Fach, so kann dessen Besetzung bzw. Nichtbesetzung zu jedem Zeitpunkt als *Treffer* bzw. *Niete* gedeutet werden. Wir befinden uns also in der Situation einer Bernoulli-Kette mit Trefferwahrscheinlichkeit $p = 1/n$. Da $T_j - 2$ die Anzahl der Nieten vor dem zweiten Treffer zählt, besitzt $T_j - 2$ die negative Binomialverteilung $\mathcal{N}b(2, 1/n)$ (vgl. Kapitel 2). Insbesondere gilt $E(T_j) = 2 + 2 \cdot (n-1) = 2 \cdot n$ (vgl. (2.5)), was nur bedeutet, daß wir im Lotto-Beispiel auf den zweiten Sechser mit einer *bestimmten* Zahlenreihe im Durchschnitt fast 28 Millionen Ausspielungen warten müssen.

Daß wir uns nicht für eine bestimmte, sondern für *irgendeine* Kollision interessieren, drückt sich in der Gleichung

$$K_n = \min(T_1, T_2, \ldots, T_n)$$

aus. Die Zeit bis zur ersten Kollision ist also ein *Minimum* von $n$ (nicht stochastisch unabhängigen) Zufallsvariablen, wobei $T_j - 2$ für jedes $j$ die negative Binomialverteilung $\mathcal{N}b(2, 1/n)$ besitzt.

Wie im folgenden gezeigt wird, ist der Zeitpunkt der ersten Kollision bei der rein zufälligen sukzessiven Besetzung von $n$ Fächern *von der Größenordnung* $\sqrt{n}$.

Satz 4.1: Für jede positive reelle Zahl $t$ gilt die Grenzwertaussage

$$\lim_{n\to\infty} P\left(K_n \le \sqrt{n} \cdot t\right) = 1 - e^{-\frac{t^2}{2}} .$$

Beweis: Zu vorgegebenem $t > 0$ existiert für jede genügend große Zahl $n$ eine natürliche Zahl $k_n$ mit

$$2 \le k_n \le \sqrt{n} \cdot t \le k_n + 1 \le n + 1 \tag{4.3}$$

(warum?), und es folgt

$$P(K_n \le k_n) \le P(K_n \le \sqrt{n} \cdot t) \le P(K_n \le k_n + 1). \tag{4.4}$$

Unter Verwendung der Ungleichung

$$\log x \;\leq\; x - 1 \qquad (x > 0) \tag{4.5}$$

für die Logarithmus-Funktion (Skizze!) und der Summenformel $\sum_{j=1}^{m} j = \frac{m(m+1)}{2}$ erhalten wir aus (4.2) die Abschätzung

$$\begin{aligned} P(K_n \leq k_n) &= 1 - \prod_{j=1}^{k_n-1}\left(1 - \frac{j}{n}\right) = 1 - \exp\left(\sum_{j=1}^{k_n-1}\log\left(1 - \frac{j}{n}\right)\right) \\ &\geq 1 - \exp\left(-\frac{1}{2}\cdot\frac{k_n\cdot(k_n-1)}{n}\right). \end{aligned}$$

Völlig analog liefert die durch Ersetzen von $x$ durch $1/x$ in (4.5) entstehende Ungleichung $\log x \geq 1 - 1/x$ die Abschätzung

$$\begin{aligned} P(K_n \leq k_n + 1) &\leq 1 - \exp\left(-\sum_{j=1}^{k_n}\frac{j}{n-j}\right) \\ &\leq 1 - \exp\left(-\frac{1}{2}\cdot\frac{k_n\cdot(k_n+1)}{n-k_n}\right) \end{aligned}$$

(man beachte, daß $n - j \geq n - k_n$ für $j \in \{1,\ldots,k_n\}$!). Da (4.3) die Grenzwertaussagen

$$\lim_{n\to\infty}\frac{k_n\cdot(k_n-1)}{n} = \lim_{n\to\infty}\frac{k_n\cdot(k_n+1)}{n-k_n} = t^2$$

nach sich zieht, konvergieren beide Schranken in (4.4) gegen $1 - \exp(-t^2/2)$, womit der Beweis erbracht ist.

Setzen wir in Satz 4.1 speziell $t = \sqrt{2\cdot\log 2}$, so folgt für großes $n$

$$P\left(K_n \leq \sqrt{n\cdot 2\cdot\log 2}\right) \approx \frac{1}{2} \tag{4.6}$$

und somit speziell $P(K_n \leq 4403) \approx \frac{1}{2}$ im Fall $n = 13\,983\,816$. Der Beweis von Satz 4.1 zeigt aber auch, daß die Wahrscheinlichkeit $P(K_n \leq k)$ durch

$$1 - \exp\left(-\frac{k\cdot(k-1)}{2\cdot n}\right) \leq P(K_n \leq k) \leq 1 - \exp\left(-\frac{k\cdot(k-1)}{2\cdot(n-k+1)}\right) \tag{4.7}$$

nach unten und oben abgeschätzt werden kann.

Das Paradoxon der frühen ersten Kollision ist in anderem Gewand als *Geburtstagsproblem* wohlbekannt (siehe z.B. Engel, 1987, S.39, Riehl, 1997, oder Strick, 1991). Beim

Geburtstagsproblem ist nach der Wahrscheinlichkeit gefragt, daß unter $k$ rein zufällig ausgewählten Personen mindestens zwei an demselben Tag Geburtstag haben. Deuten wir die 365 Tage des Jahres (Schaltjahre seien unberücksichtigt) als Fächer und die Personen als Teilchen, so entspricht das Feststellen der Geburtstage dem rein zufälligen Besetzen der 365 Fächer mit $k$ Teilchen. Hierbei wird zwar die unrealistische Annahme einer Gleichverteilung der Geburtstage über alle 365 Tage gemacht; es kann aber gezeigt werden, daß Abweichungen von dieser Annahme die Wahrscheinlichkeit für einen Mehrfachgeburtstag nur vergrößern (siehe z.B. Joag-Dev u. Proschan, 1992, oder Munford, 1977). Da beim Geburtstagsproblem $P(K_{365} \leq 23) = 0.507\ldots > 1/2$ gilt (vgl. (4.2)), kann getrost darauf gewettet werden, daß unter 23 (oder mehr) Personen mindestens zwei an demselben Tag Geburtstag haben. Wegen $\sqrt{365 \cdot 2 \cdot \log 2} = 22.49\ldots$ ist dabei die Approximation (4.6) schon für $n = 365$ sehr gut.

Da der Erwartungswert einer $\mathbb{N}_0$-wertigen Zufallsvariablen $X$ durch

$$\begin{aligned} E(X) &= \sum_{j=1}^{\infty} j \cdot P(X=j) = \sum_{j=1}^{\infty} \left( \sum_{k=1}^{j} 1 \right) P(X=j) \\ &= \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} P(X=j) = \sum_{k=1}^{\infty} P(X \geq k) \end{aligned}$$

gegeben ist und da $K_n$ die Werte $2, 3, \ldots, n+1$ annimmt, folgt nach (4.1)

$$E(K_n) = \sum_{k=0}^{n} P(K_n \geq k+1) = \sum_{k=0}^{n} \frac{n^{\underline{k}}}{n^k} \tag{4.8}$$

und somit speziell $E(K_{365}) = 24.616\ldots$ . Der rechts stehende Ausdruck ist für große Werte von $n$ unhandlich. Deshalb liegt es nahe, nach einer Näherungsformel für $E(K_n)$ zu suchen. Blom, Holst und Sandell (1994, S. 190) geben hierfür den Ausdruck

$$E(K_n) \approx \sqrt{\frac{\pi n}{2}} + \frac{2}{3} \tag{4.9}$$

an, was für den Fall $n = 365$ den Wert $24.611\ldots$ liefert. Die von Engel (1987, S. 63) ohne Beweis angegebene Entwicklung für $E(K_n)$ ist fehlerhaft; der Term $-\frac{1}{3}$ dort muß durch $+\frac{2}{3}$ ersetzt werden.

Der Schlüssel zur Gewinnung von Näherungsformeln für (4.8) ist die Gleichung

$$k! = \int_0^{\infty} e^{-t} \cdot t^k \, dt \quad , \qquad k \in \mathbb{N}_0,$$

welche zusammen mit dem binomischen Lehrsatz und der Substitution $u = t/\sqrt{n}$ zur Integraldarstellung

$$\begin{aligned}
\sum_{k=0}^{n} \frac{n^{\underline{k}}}{n^k} &= \sum_{k=0}^{n} \frac{n^{\underline{k}}}{k!} \cdot \frac{1}{n^k} \cdot \int_0^\infty e^{-t} \cdot t^k \, dt \\
&= \int_0^\infty e^{-t} \cdot \sum_{k=0}^{n} \binom{n}{k} \cdot \left(\frac{t}{n}\right)^k dt \\
&= \int_0^\infty e^{-t} \cdot \left(1 + \frac{t}{n}\right)^n dt \\
&= \sqrt{n} \cdot \int_0^\infty \left(e^{-u/\sqrt{n}} \cdot \left(1 + \frac{u}{\sqrt{n}}\right)\right)^n du \qquad (4.10)
\end{aligned}$$

führt. Wegen

$$e^{-u/\sqrt{n}} = 1 - \frac{u}{\sqrt{n}} + \frac{u^2}{2n} + o\left(\frac{1}{n}\right)$$

erhält man

$$e^{-u/\sqrt{n}} \cdot \left(1 + \frac{u}{\sqrt{n}}\right) = 1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right),$$

so daß der in (4.10) auftretende Integrand gegen $\exp(-u^2/2)$ konvergiert. Überlegt man sich noch, daß dieser Grenzübergang mit der Integration in (4.10) vertauscht werden darf (siehe Klamkin und Newman, 1967), so folgt die grobe Approximation

$$E(K_n) \approx \sqrt{n} \cdot \int_0^\infty \exp\left(-\frac{u^2}{2}\right) du = \sqrt{\frac{n\pi}{2}}.$$

Die bessere Näherungsformel (4.9) ergibt sich durch eine genauere Betrachtung des in (4.10) stehenden Integranden.

# 5 Warten auf den ersten Dreifachgeburtstag

Eine interessante Variante des Geburtstagsproblems entsteht, wenn nach der Wahrscheinlichkeit dafür gefragt wird, daß in einer Gruppe von $k$ Personen mindestens *drei* an demselben Tag Geburtstag haben (siehe z.B. Schrage, 1990, 1992 oder Holst, 1995). Wir behandeln dieses Problem wie in Kapitel 4 in einem Teilchen/Fächer-Modell mit $n$ Fächern, wobei die Teilchen nacheinander rein zufällig und unabhängig voneinander in die Fächer fallen. Von Interesse ist jetzt die Zufallsvariable

$$K_{n,3} \; := \; \text{Zeitpunkt der ersten } \textit{Dreifach-Kollision} \text{ beim sukzessiven rein zufälligen Besetzen von } n \text{ Fächern.}$$

$K_{n,3}$ gibt den Zeitpunkt an, zu dem erstmalig ein Teilchen in ein bereits mit *zwei* Teilchen besetztes Fach gelangt. Die eingangs beschriebene Wahrscheinlichkeit läßt sich dann mit Hilfe von $K_{n,3}$ in der Form $P(K_{365,3} \leq k)$ ausdrücken.

Modellieren wir die Verteilung von $k$ Teilchen auf die $n$ Fächer als $k$–Permutation $(i_1, \ldots, i_k)$ mit der Deutung von $i_l$ als Nummer des Faches für das $l$–te Teilchen, und setzen wir $1\{\cdot\} = 1$ bzw. $1\{\cdot\} = 0$, falls die in $\{\cdot\}$ stehende Aussage wahr bzw. falsch ist, so gibt

$$h_j \; := \; \sum_{m=1}^{k} 1\{i_m = j\}$$

die Anzahl der Teilchen im $j$–ten Fach an. Da das Ereignis $\{K_{n,3} \geq k+1\}$ genau dann eintritt, wenn jedes $h_j$ kleiner oder gleich 2 ist, erhalten wir

$$P(K_{n,3} \geq k+1) \; = \; \frac{\left|\left\{(i_1, \ldots, i_k) : 1 \leq i_1, \ldots, i_k \leq n, \; \max\limits_{1 \leq j \leq n} h_j \leq 2\right\}\right|}{n^k} \, .$$

Dieser geschlossene Ausdruck führt auf das Problem der Bestimmung der Anzahl aller Verteilungsmöglichkeiten von $k$ Teilchen auf $n$ Fächer mit der Nebenbedingung, daß jedes Fach höchstens zwei Teilchen enthält. Hier bietet sich an, das Ereignis $\{K_{n,3} \geq k+1\}$ nach der Anzahl $l$ der mit jeweils *genau* 2 Teilchen besetzen Fächer zu zerlegen. Hierbei läuft $l$ von 0 bis $[k/2]$ $(= \max\{m \in \mathbb{Z} : m \leq k/2\})$. Bezeichnet $w(k, l)$ die Wahrscheinlichkeit, daß bei der rein zufälligen Verteilung von $k$ Teilchen auf $n$ Fächer genau $l$ Fächer je zwei Teilchen enthalten und die restlichen $k - 2 \cdot l$ Teilchen allein in ihrem jeweiligen Fach liegen, so gilt mit der Konvention, ein Produkt über die leere Menge gleich 1 zu setzen

$$w(k, l) \; = \; \frac{1}{n^k \cdot l!} \cdot \prod_{\nu=0}^{l-1} \binom{k - 2 \cdot \nu}{2} \cdot n^{\underline{l}} \cdot (n - l)^{\underline{k-2 \cdot l}} \tag{5.1}$$

und folglich wegen $n^{\underline{l}} \cdot (n-l)^{\underline{k-2\cdot l}} = n^{\underline{k-l}}$

$$P(K_{n,3} \geq k+1) \; = \; \sum_{l=0}^{[k/2]} w(k, l) \; = \; \frac{1}{n^k} \cdot \sum_{l=0}^{[k/2]} \frac{1}{l!} \cdot \prod_{\nu=0}^{l-1} \binom{k - 2\nu}{2} \cdot n^{\underline{k-l}}$$

(vgl. Schrage, 1990, für den Spezialfall $n = 365$). Formel (5.1) ergibt sich, indem zunächst 2 der $k$ Teilchen, danach 2 der restlichen $k - 2$ Teilchen ... und schließlich 2 der übriggebliebenen $k - 2 \cdot (l - 1)$ Teilchen für die doppelt besetzten Fächer ausgewählt werden. Diese $l$ Teilchenpaare können dann auf $n^{\underline{l}}$ Weisen ihren Fächern zugeordnet werden. Danach werden die übrigen $k - 2 \cdot l$ Teilchen auf die freien $n - l$ Fächer verteilt, was auf $(n - l)^{\underline{k-2\cdot l}}$ Weisen möglich ist. Der Faktor $1/l!$ tritt auf, weil es nicht auf die Reihenfolge der $l$ Auswahlen von je zwei Teilchen ankommt.

Zur numerischen Berechnung von $P(K_{n,3} \geq k + 1)$ ist die für $k = 0, 1, \ldots, n$ geltende Rekursionsformel

$$w(k, l+1) = \frac{1}{(l+1) \cdot (n - k + l + 1)} \cdot \binom{k - 2 \cdot l}{2} \cdot w(k, l)$$

($l = 0, 1, 2, \ldots$) hilfreich. In Tabelle 5.1 sind die Wahrscheinlichkeiten $P(K_{n,3} \leq k)$ für den Fall $n = 365$ und verschiedene Werte von $k$ aufgeführt.

| $k$ | $P(K_{n,3} \leq k)$ | $k$ | $P(K_{n,3} \leq k)$ | $k$ | $P(K_{n,3} \leq k)$ |
|---|---|---|---|---|---|
| 10 | 0.0009 | 80 | 0.4182 | 130 | 0.8911 |
| 20 | 0.0082 | 87 | 0.4995 | 140 | 0.9357 |
| 30 | 0.0285 | 88 | 0.5111 | 150 | 0.9648 |
| 40 | 0.0669 | 90 | 0.5342 | 160 | 0.9822 |
| 50 | 0.1264 | 100 | 0.6459 | 170 | 0.9917 |
| 60 | 0.2072 | 110 | 0.7455 | 180 | 0.9965 |
| 70 | 0.3065 | 120 | 0.8280 | 200 | 0.9995 |

**Tabelle 5.1:** Wahrscheinlichkeiten für mindestens einen Dreifachgeburtstag ($n = 365$) bei $k$ Personen

Es zeigt sich, daß man schon bei 88 oder mehr Personen darauf wetten kann, daß mindestens 3 dieser Personen an demselben Tag des Jahres Geburtstag haben. Ein auf 100 000 wiederholten Simulationen der Wartezeit $K_{365,3}$ basierendes Histogramm ist in Bild 5.1 veranschaulicht.

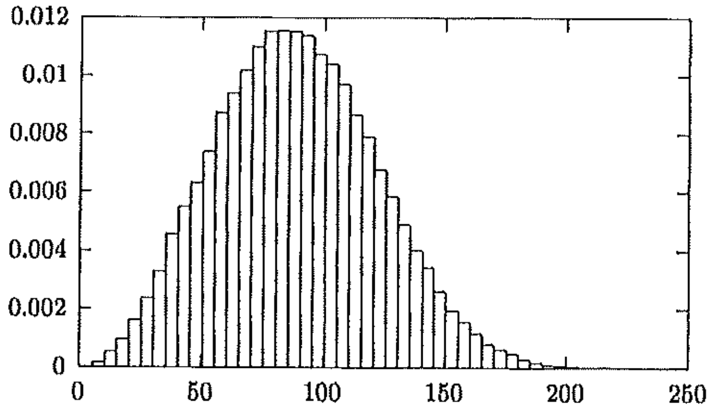


**Bild 5.1:** Histogramm der Verteilung der Dreifach-Kollisionszeit $K_{365,3}$ (100 000 Simulationen)

Bezeichnet in Verallgemeinerung von $K_{n,3}$ die Zufallsvariable $K_{n,c}$ den *Zeitpunkt der ersten c–fachen Kollision* beim rein zufälligen sukzessiven Besetzen von $n$ Fächern, so gelten die Integraldarstellung

$$E(K_{n,c}) = \int_0^\infty e^{-t} \cdot \left( \sum_{j=0}^{c-1} \frac{1}{j!} \left( \frac{t}{n} \right)^j \right)^n dt \tag{5.2}$$

(Holst, 1995; Klamkin und Newman, 1967) sowie die asymptotische Gleichheit

$$E(K_{n,c}) \sim \sqrt[c]{c!} \cdot \Gamma\left(1 + \frac{1}{c}\right) \cdot n^{1-1/c} \quad \text{bei } n \to \infty \tag{5.3}$$

(Klamkin und Newman, 1967). Hier bedeutet das Symbol „$\sim$", daß der Quotient aus linker und rechter Seite in (5.3) beim Grenzübergang $n \to \infty$ gegen 1 konvergiert. $\Gamma(x)$ steht für die durch

$$\Gamma(x) := \int_0^\infty e^{-t} \cdot t^{x-1}\, dx \qquad (x > 0)$$

definierte *Gamma–Funktion.*

Holst (1995) gibt ohne Beweis für den zur Diskussion stehenden Fall $c = 3$ die im Vergleich zu (5.3) genauere Approximation

$$E(K_{n,3}) \approx \frac{6^{1/3} \cdot \Gamma(1/3)}{3} \cdot n^{2/3} + \frac{6^{2/3} \cdot \Gamma(2/3)}{6} \cdot n^{1/3} + \frac{21}{40}$$

für großes $n$ an. Diese liefert im Fall $n = 365$ den Näherungswert $E(K_{365,3}) \approx 88.725$.

Mit Hilfe einer Verallgemeinerung der von Klamkin und Newman entwickelten Integraldarstellung (5.2) leitet Dwass (1969) den Grenzwertsatz

$$\lim_{n\to\infty} P(K_{n,c} \le n^{1-1/c} \cdot t) \;=\; 1 - \exp\left(-\frac{t^c}{c!}\right) \tag{5.4}$$

($t > 0$) her. Da $K_{n,c}$ für $c = 2$ die in Kapitel 4 untersuchte Zeit $K_n$ bis zur ersten (Zweifach–)Kollision ist, stellt diese Aussage eine Verallgemeinerung von Satz 4.1 dar. Ein einfacher Beweis für (5.4) findet sich in Henze (1998b).

# 6 Warten auf eine vollständige Serie: das Sammlerproblem

Zu einem Sammelalbum gehören $n$ Bilder, die in Packungen zu je $s$ Bildern verkauft werden. Wir nehmen an, daß für jede Packung alle $\binom{n}{s}$ möglichen Bilder–Auswahlen gleichwahrscheinlich sind. Ferner seien die Bilder–Auswahlen für verschiedene Packungen stochastisch unabhängig voneinander. Hier stellen sich in natürlicher Weise die folgenden Fragen:

1. Wie viele Packungen müssen "im Mittel" gekauft werden, bis jedes der $n$ Bilder mindestens einmal vorhanden ist?

2. Wie groß ist die Wahrscheinlichkeit, nach dem Kauf von $k$ Packungen erstmals alle Bilder zu besitzen?

Diese und ähnliche Fragen sind klassische Probleme der Wahrscheinlichkeitsrechnung, welche in der Literatur als *Sammlerproblem*, *Coupon-Collector-Problem* oder im speziellen Fall $s = 1$ als *Problem der vollständigen Serie* bekannt sind (siehe z.B. Henze, 1999, Kapitel 24 oder Stadje, 1989, und die dort angegebene Literatur).

Für eine Formulierung des Sammlerproblems in einem abstrakten Teilchen/Fächer–Modell werden die möglichen Bilder durch $n$ von 1 bis $n$ numerierte Fächer ersetzt. Dem Kauf einer Packung mit $s$ verschiedenen Bildern entspricht dann die rein zufällige

Besetzung von $s$ verschiedenen Fächern mit je einem Teilchen — im folgenden *Besetzungsvorgang* genannt. Von Interesse ist hier die Zufallsvariable

$$V_{n,s} \quad = \quad \text{Anzahl der Besetzungsvorgänge, bis (zum ersten Mal)} \atop \text{jedes Fach mindestens ein Teilchen enthält.}$$

Die eingangs gestellten Fragen 1. und 2. führen dann offenbar auf die Angabe des Erwartungswertes von $V_{n,s}$ bzw. auf die Wahrscheinlichkeit $P(V_{n,s} = k)$. Bevor wir uns diesen Problemen zuwenden, seien einige beliebte Einkleidungen des Sammlerproblems vorgestellt:

a) Werfen eines Würfels, bis jede Augenzahl mindestens einmal aufgetreten ist ($n = 6,\ s = 1$),

b) Drehen eines Roulette–Rades, bis jede Zahl mindestens einmal aufgetreten ist ($n = 37,\ s = 1$),

c) Anzahl von Ausspielungen im Zahlenlotto 6 aus 49, bis jede Zahl mindestens einmal Gewinnzahl gewesen ist ($n = 49,\ s = 6$).

Das in c) beschriebene Sammler–Experiment wird von den staatlichen Lottogesellschaften in jährlich neu startenden *Treffer–Bilanzen* laufend dokumentiert. Tabelle 6.1 zeigt die Treffer–Bilanz des Samstagslottos 6 aus 49 nach den ersten 11 Ausspielungen des Jahres 1996.

**Treffer–Bilanz 1996**
**Nach 11 Ausspielungen**

| | | | | | | |
|---|---|---|---|---|---|---|
| **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| 1 | – | 5 | 2 | 1 | 1 | 1 |
| **8** | **9** | **10** | **11** | **12** | **13** | **14** |
| 1 | 2 | 1 | – | 1 | 1 | 2 |
| **15** | **16** | **17** | **18** | **19** | **20** | **21** |
| 1 | – | 2 | 1 | 2 | 4 | 1 |
| **22** | **23** | **24** | **25** | **26** | **27** | **28** |
| 1 | 1 | 1 | 3 | 2 | – | – |
| **29** | **30** | **31** | **32** | **33** | **34** | **35** |
| 2 | 1 | 1 | 1 | 1 | 2 | 2 |
| **36** | **37** | **38** | **39** | **40** | **41** | **42** |
| 3 | 1 | 3 | – | 1 | 1 | 2 |
| **43** | **44** | **45** | **46** | **47** | **48** | **49** |
| 3 | – | – | – | 2 | 1 | 1 |

Tabelle 6.1: Treffer–Bilanz 1996 nach 11 Ausspielungen

## 6.1 Der Spezialfall $s = 1$

Im speziellen Fall $s = 1$ ist eine Modellierung der Wartezeit $V_{n,1}$ als *Summe stochastisch unabhängiger Wartezeiten* möglich. Hierzu bezeichnen wir einen Versuch als „Treffer", wenn er zur Besetzung eines noch freien Faches führt. Damit ist der erste Versuch immer ein Treffer. Da nach dem Erzielen des $j$-ten Treffers jeder der weiteren Versuche mit Wahrscheinlichkeit $(n-j)/n$ den nächsten Treffer ergibt ($j = 1, \ldots, n-1$) und da alle Versuche unabhängig voneinander ablaufen, besitzt $V_{n,1}$ die Darstellung

$$V_{n,1} \;=\; 1 \;+\; Y_1 \;+\; Y_2 \;+\; \ldots \;+\; Y_{n-1} \tag{6.1}$$

mit unabhängigen Zufallsvariablen $Y_1, \ldots, Y_{n-1}$, wobei $Y_j - 1$ die geometrische Verteilung $\mathcal{G}((n-j)/n)$ besitzt und anschaulich für die Anzahl der *Fehlversuche zwischen dem $j$-ten und dem $(j+1)$-ten Treffer* steht. Wegen

$$E(Y_j) \;=\; \frac{n}{n-j}, \qquad V(Y_j) \;=\; \frac{n \cdot j}{(n-j)^2}$$

(vgl. (1.2) und (1.3)) ergeben sich aus der Darstellung (6.1) Erwartungswert und Varianz von $V_{n,1}$ zu

$$E(V_{n,1}) \;=\; 1 + \sum_{j=1}^{n-1} E(Y_j) \;=\; n \cdot \left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}\right),$$

$$V(V_{n,1}) \;=\; \sum_{j=1}^{n-1} V(Y_j) \;=\; n^2 \cdot \left(\sum_{j=1}^{n-1} \frac{1}{j^2} \;-\; \frac{1}{n} \cdot \sum_{j=1}^{n-1} \frac{1}{j}\right).$$

Im speziellen Fall $n = 6$ (Warten auf jede Augenzahl beim Würfelwurf) erhält man $E(V_{6,1}) = 14.7$. Unter Benutzung der Beziehungen

$$\lim_{n \to \infty} \left(\sum_{j=1}^{n} \frac{1}{j} \;-\; \log n\right) \;=\; C \;=\; 0.57721\ldots \quad (\textit{Euler} - \textit{Mascheroni} - \textit{Konstante})$$

(siehe z.B. Heuser, 1991, S.185) und

$$\lim_{n \to \infty} \sum_{j=1}^{n} \frac{1}{j^2} \;=\; \frac{\pi^2}{6}$$

(Heuser, 1993, S.150) folgen für großes $n$ die Approximationen

$$E(V_{n,1}) \;\approx\; n \cdot (\log n \;+\; C), \qquad V(V_{n,1}) \;\approx\; n^2 \cdot \frac{\pi^2}{6}.$$

## 6.2 Der allgemeine Fall $s \geq 1$

Im allgemeinen Fall führen wir geeignete Zufallsvariablen $W_1, W_2, \ldots, W_n$ ein. $W_j$ beschreibt die Anzahl der Besetzungsvorgänge, bis das $j$-te Fach mindestens ein Teilchen enthält. Hiermit läßt sich $V_{n,s}$ als *maximale Wartezeit* in der Form

$$V_{n,s} = \max(W_1, W_2, \ldots, W_n)$$

ausdrücken. Offenbar nimmt $V_{n,s}$ die möglichen Werte $a, a+1, a+2, \ldots$ an, wobei $a$ durch

$$a := \min\left\{m \in \mathbb{N} : \frac{n}{s} \leq m\right\} \tag{6.2}$$

gegeben ist.

Den Schlüssel zur Bestimmung der Verteilung von $V_{n,s}$ bildet die Gleichung

$$\{V_{n,s} > k\} = \bigcup_{j=1}^{n} \{W_j > k\}, \quad k \geq a-1. \tag{6.3}$$

Diese drückt nur aus, daß die maximale Wartezeit $V_{n,s}$ genau dann größer als $k$ ist, wenn mindestens eine der Wartezeiten $W_j$ größer als $k$ ist. Schreiben wir kurz $A_j$ für das Ereignis $\{W_j > k\}$, so liegt wegen $P(V_{n,s} > k) = P(\cup_{j=1}^{n} A_j)$ die Anwendung der Formel des Ein- und Ausschließens

$$P\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{r=1}^{n} (-1)^{r-1} \cdot S_r \tag{6.4}$$

(siehe z.B. Henze, 1999, Kapitel 11) nahe. Dabei ist

$$S_r := \sum_{1 \leq i_1 < \ldots < i_r \leq n} P(A_{i_1} \cap \ldots \cap A_{i_r}) \tag{6.5}$$

(Summation über alle $r$-elementigen Teilmengen $\{i_1, \ldots, i_r\}$ von $\{1, \ldots, n\}$). Um die Wahrscheinlichkeit $P(\cup_{j=1}^{n} A_j)$ zu berechnen, benötigen wir also für jedes $r = 1, \ldots, n$ und jede Wahl von $i_1, \ldots, i_r$ mit $1 \leq i_1 < \ldots < i_r \leq n$ die Wahrscheinlichkeit $P(A_{i_1} \cap \ldots \cap A_{i_r})$.

Offenbar tritt das Ereignis $A_{i_1} \cap \ldots \cap A_{i_r}$ genau dann ein, wenn in den ersten $k$ Versuchen keines der Fächer mit den Nummern $i_1, \ldots, i_r$ besetzt wird, d.h. wenn bei jedem der ersten $k$ Besetzungsvorgänge jeweils $s$ Fächer aus der $(n-r)$-elementigen Nummern-Menge $\{1,2,\ldots,n\}\setminus\{i_1,\ldots,i_r\}$ ausgewählt werden. Die Wahrscheinlichkeit dafür, daß dies bei *einem* Versuch geschieht, ist durch

$$q_r := \frac{\binom{n-r}{s}}{\binom{n}{s}}, \quad n-r \geq s, \tag{6.6}$$

gegeben (Laplace-Modell). Aufgrund der Unabhängigkeit von Ereignissen, welche sich auf verschiedene Besetzungsvorgänge beziehen, gilt dann

$$P(A_{i_1} \cap \ldots \cap A_{i_r}) = \begin{cases} q_r^k, & \text{falls} \quad r \leq n-s, \\ 0, & \text{falls} \quad r > n-s, \end{cases}$$

so daß die in (6.5) definierte Summe $S_r$ für $r \leq n-s$ die Gestalt $S_r = \binom{n}{r} q_r^k$ annimmt. Nach (6.4) und (6.3) folgt

$$P(V_{n,s} > k) = \sum_{r=1}^{n-s} (-1)^{r-1} \cdot \binom{n}{r} \cdot q_r^k, \quad k \geq a-1, \tag{6.7}$$

mit $a$ wie in (6.2). Wegen $P(V_{n,s} > k-1) = P(V_{n,s} > k) + P(V_{n,s} = k)$ ergibt sich nun die Verteilung von $V_{n,k}$ durch Differenzbildung in (6.7), und wir erhalten das folgende Resultat.

**Satz 6.1:** Die Anzahl $V_{n,s}$ der zur Besetzung aller Fächer nötigen Versuche im Sammlerproblem mit $n$ Fächern besitzt die Verteilung

$$P(V_{n,s} = k) = \sum_{r=1}^{n-s} (-1)^{r-1} \cdot \binom{n}{r} \cdot q_r^{k-1} \cdot (1-q_r), \quad k \geq a,$$

und den Erwartungswert

$$E(V_{n,s}) = \sum_{r=1}^{n-s} (-1)^{r-1} \cdot \binom{n}{r} \cdot \frac{q_r^{a-1} \cdot (q_r - a \cdot (q_r - 1))}{1-q_r} \tag{6.8}$$

mit $a$ wie in (6.2).

Dabei ergibt sich (6.8) durch direkte Rechnung aus der Darstellungsformel $E(V_{n,s}) = \sum_{k=a}^{\infty} k \cdot P(V_{n,s} = k)$ unter Beachtung von (2.2) sowie

$$\sum_{k=1}^{a-1} k \cdot x^{k-1} = \frac{d}{dx}\left(\frac{x^a - 1}{x-1}\right) = \frac{a \cdot x^{a-1} \cdot (x-1) - (x^a - 1)}{(x-1)^2}, \quad |x| < 1.$$

Die Verteilung von $V_{n,s}$ ist für den Fall $n = 6, s = 1$ (Wartezeit, bis beim Würfeln jede Augenzahl aufgetreten ist) in Bild 6.1 veranschaulicht. Deutlich erkennbar ist eine für stochastische Extremwertprobleme typische Asymmetrie ($V_{n,s}$ ist ein *Maximum* von Zufallsvariablen!).

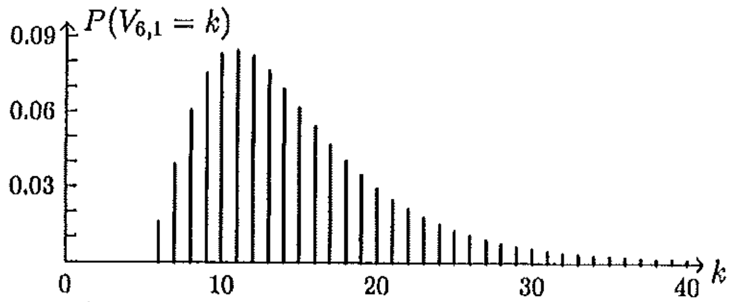


**Bild 6.1**: Verteilung der Wartezeit beim Sammlerproblem mit $n = 6, s = 1$

In gleicher Weise zeigt Bild 6.2 die Verteilung von $V_{n,s}$ für den Fall $n = 49$, $s = 6$ (Anzahl der Ausspielungen im Lotto 6 aus 49, bis jede Zahl als Gewinnzahl aufgetreten ist).

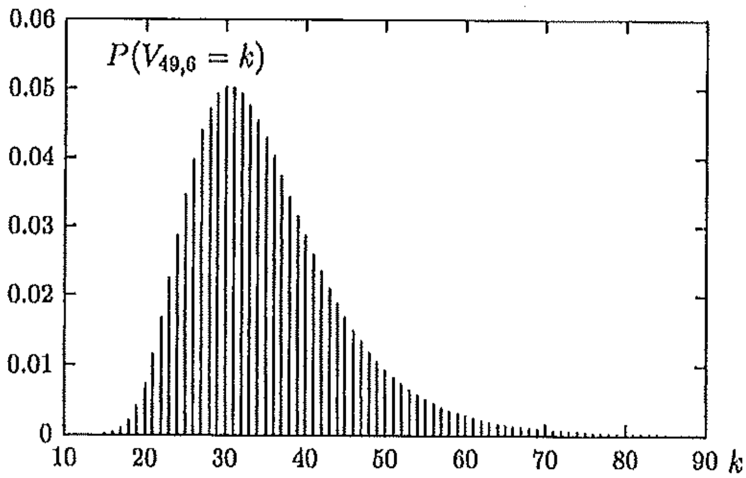


**Bild 6.2**: Verteilung der Wartezeit beim Sammlerproblem mit $n = 49$, $s = 6$

Komplementbildung in (6.7) liefert hier den Wert $P(V_{49,6} \leq 52) = 0.946\ldots$. Somit kann getrost darauf gewettet werden, daß im Laufe eines Jahres jede Lottozahl mindestens einmal Gewinnzahl ist. Nebenbei sei bemerkt, daß dieser Fall in genau 38 der ersten 40 Jahre (= 95%!) der deutschen Lottogeschichte eintrat. Der nach (6.8) berechnete Erwartungswert von $V_{49,6}$ ist $35.08\ldots$, was bedeutet, daß beim Lotto im Mittel nach etwa 35 Wochen jede Zahl mindestens einmal gezogen worden ist.

## 6.3 Ein Grenzwertsatz

Im folgenden soll die Grenzverteilung von $V_{n,1}$ beim Grenzübergang $n \to \infty$ bestimmt werden. Obwohl $V_{n,1}$ nach (6.1) eine Summe unabhängiger Zufallsvariablen ist, gelangt der zentrale Grenzwertsatz (Approximation durch eine Normalverteilung) hier nicht zur Anwendung.

**Satz 6.2**: Für jede reelle Zahl $x$ gilt

$$\lim_{n\to\infty} P(V_{n,1} \le n \cdot (x + \log n)) = \exp\left(-e^{-x}\right).$$

Beweis: Zu gegebenem $x \in \mathbb{R}$ sei $n \in \mathbb{N}$ so groß gewählt, daß die Ungleichung $x + \log n \ge 1$ erfüllt ist. Setzen wir $k_n := [n \cdot (x + \log n)]$, wobei allgemein $[y]$ die größte ganze Zahl kleiner oder gleich $y$ bezeichnet, so gilt aufgrund der Ganzzahligkeit von $V_{n,1}$ sowie aufgrund von (6.7)

$$\begin{aligned} P(V_{n,1} > n \cdot (x + \log n)) &= P(V_{n,1} > k_n) \\ &= \sum_{r=1}^{n-1} (-1)^{r-1} \cdot \binom{n}{r} \cdot q_r^{k_n}. \end{aligned}$$

Wir werden im folgenden entscheidend von der Tatsache Gebrauch machen, daß die in der Formel des Ein- und Ausschließens (6.4) bei Abbruch der alternierende Summe entstehenden Partialsummen abwechselnd obere und untere Schranken für die linke Seite liefern (siehe z.B. Feller, 1968, S. 110); es gilt also

$$P(V_{n,1} > n \cdot (x + \log n)) \le \sum_{r=1}^{2l-1} (-1)^{r-1} \cdot \binom{n}{r} \cdot q_r^{k_n}, \tag{6.9}$$

$$P(V_{n,1} > n \cdot (x + \log n)) \ge \sum_{r=1}^{2l} (-1)^{r-1} \cdot \binom{n}{r} \cdot q_r^{k_n} \tag{6.10}$$

für jedes *feste* $l = 1, 2, \ldots$ . Wegen

$$\lim_{n\to\infty} \binom{n}{r} \cdot q_r^{k_n} = \frac{e^{-xr}}{r!} \tag{6.11}$$

(Beweis folgt) liefern (6.9) und (6.10) für jedes feste $l = 1, 2, \ldots$ die Abschätzungen

$$\limsup_{n\to\infty} P\left(V_{n,1} > n\cdot(x+\log n)\right) \;\leq\; -\sum_{r=1}^{2l-1} \frac{(-e^{-x})^r}{r!}\,,$$

$$\liminf_{n\to\infty} P\left(V_{n,1} > n\cdot(x+\log n)\right) \;\geq\; -\sum_{r=1}^{2l} \frac{(-e^{-x})^r}{r!}\,.$$

Läßt man hier $l$ gegen Unendlich streben, so folgt die Existenz des Grenzwertes und zugleich dessen Wert: Es gilt

$$\begin{aligned}\lim_{n\to\infty} P\left(V_{n,1} > n\cdot(x+\log n)\right) &= -\sum_{r=1}^{\infty} \frac{(-e^{-x})^r}{r!}\\ &= -\left(\exp\left(-e^{-x}\right) - 1\right),\end{aligned}$$

was zu zeigen war. Für den noch fehlenden Nachweis von (6.11) setzen wir $q_r = (n-r)/n$ in die linke Seite von (6.11) ein und erhalten

$$\binom{n}{r}\cdot q_r^{k_n} \;=\; \frac{1}{r!}\cdot\frac{n^{\underline{r}}}{n^r}\cdot n^r\cdot\left(1-\frac{r}{n}\right)^{n(x+\log n)}\cdot\left(1-\frac{r}{n}\right)^{\varepsilon_n}$$

mit $\varepsilon_n = [n\cdot(x+\log n)] - n\cdot(x+\log n)$, also $-1 < \varepsilon_n \leq 0$. (6.11) folgt dann aus

$$\lim_{n\to\infty}\frac{n^{\underline{r}}}{n^r} \;=\; 1 \;=\; \lim_{n\to\infty}\left(1-\frac{r}{n}\right)^{\varepsilon_n}$$

und

$$\begin{aligned}&\lim_{n\to\infty} n^r\cdot\left(1-\frac{r}{n}\right)^{n(x+\log n)}\\ &= \lim_{n\to\infty}\exp\left(r\cdot\log n + n\cdot(x+\log n)\cdot\log\left(1-\frac{r}{n}\right)\right)\\ &= \exp(-r\cdot x).\end{aligned}$$

Hierbei ergibt sich das letzte Gleichheitszeichen aus den Ungleichungen

$$-\frac{r}{n-r} \;\leq\; \log\left(1-\frac{r}{n}\right) \;\leq\; -\frac{r}{n}$$

(vgl. die Logarithmus-Ungleichungen auf Seite 13). In einer anderen Formulierung besagt Satz 6.2, daß die Limesverteilung der Zufallsvariablen $V_{n,1}/n - \log n$ beim Grenzübergang $n \to \infty$ die *Gumbelsche Extremwertverteilung* ist. Die Gumbelsche

Extremwertverteilung bildet eine der drei klassischen Limesverteilungen für Maxima von unabhängigen und identisch verteilten Zufallsvariablen (siehe z.B. Galambos, 1978, S. 71). Sie besitzt die Verteilungsfunktion $G(x) = \exp(-e^{-x})$ und die in Bild 6.3 veranschaulichte Dichte $g(x) = G'(x) = e^{-x} \cdot \exp(-e^{-x})$. Ihr Erwartungswert ist die Euler–Mascheroni–Konstante $C = 0.57721\ldots$, und ihre Varianz ist $\pi^2/6$.

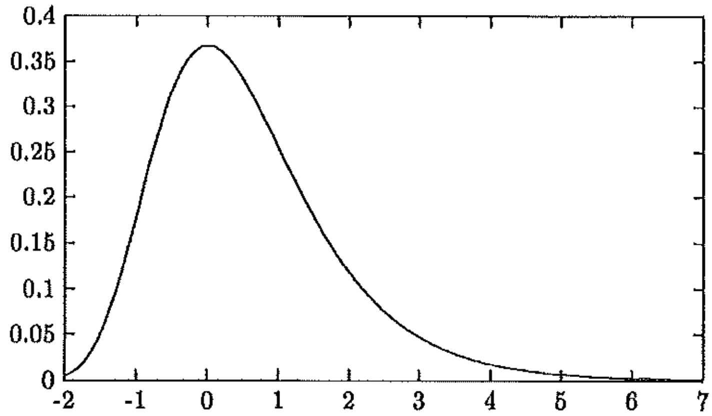


**Bild 6.3:** Dichte der Gumbelschen Extremwertverteilung

Aus Satz 6.2 folgt das auf den ersten Blick überraschende Resultat

$$\lim_{n\to\infty} P(V_{n,1} > M \cdot n) \;=\; 1 \tag{6.12}$$

für jede (noch so große) positive Zahl $M$. Hat man also sogar 1000 mal so viele Bilder erworben, wie Plätze im Album vorhanden sind, so ist es bei einer genügend großen Anzahl von möglichen Bildern praktisch sicher, daß die zufällig zusammengestellten $1000 \cdot n$ Bilder zum Erzielen einer vollständigen Serie nicht ausreichen.

Zum Nachweis von (6.12) wählen wir zu beliebig vorgegebenen Zahlen $M > 0$ und $\varepsilon$ mit $0 < \varepsilon < 1$ eine reelle Zahl $x$ derart, daß die Ungleichung

$$1 - \exp\left(-e^{-x}\right) \;\geq\; 1 - \varepsilon \tag{6.13}$$

erfüllt ist. Für genügend großes $n$ gilt dann $n \cdot (x + \log n) \geq M \cdot n$ und somit

$$P\left(V_{n,1} > n \cdot (x + \log n)\right) \;\leq\; P\left(V_{n,1} > M \cdot n\right), \tag{6.14}$$

da aus dem Ereignis $\{V_{n,1} > n \cdot (x + \log n)\}$ das Ereignis $\{V_{n,1} > M \cdot n\}$ folgt. Aus (6.13), (6.14) und Satz 6.2 ergibt sich dann

$$\begin{aligned} 1-\varepsilon &\leq 1-\exp\left(-e^{-x}\right) \\ &= \lim_{n\to\infty} P\left(V_{n,1} > n\cdot(x+\log n)\right) \leq \liminf_{n\to\infty} P\left(V_{n,1} > M\cdot n\right) \\ &\leq 1 \end{aligned}$$

und somit die Behauptung, da $\varepsilon$ beliebig klein gewählt werden kann.

## 6.4 Das Sammlerproblem mit ungleichen Wahrscheinlichkeiten

Wir greifen jetzt wieder den Spezialfall $s = 1$ auf, betrachten also die Besetzung mit jeweils einem Teilchen, lösen uns jedoch von der Annahme einer Gleichverteilung über alle $n$ Fächer. Es werde angenommen, daß die Teilchen unabhängig voneinander mit der Wahrscheinlichkeit $p_j$ ins $j$-te Fach gelangen, wobei $p_j > 0$ und $p_1 + \ldots + p_n = 1$ vorausgesetzt ist. Die Zufallsvariable $V_{n,1}$ bezeichne wie früher die Mindestanzahl der zum Erreichen einer vollständigen Serie benötigten Teilchen.

Auch in dieser allgemeineren Situation läßt sich die Verteilung von $V_{n,1}$ mit Hilfe der Formel des Ein- und Ausschließens bestimmen. Bezeichnet hierzu $A_j$ das Ereignis, daß das $j$-te Fach nach $k$ Besetzungsvorgängen kein Teilchen enthält, so gilt

$$P(V_{n,1} > k) = P\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{r=1}^{n}(-1)^{r-1}\cdot S_r \tag{6.15}$$

mit $S_r$ wie in (6.5). Wegen der Unabhängigkeit der einzelnen Besetzungsvorgänge gilt weiter

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_r}) = (1 - p_{i_1} - p_{i_2} - \ldots - p_{i_r})^k$$

für jedes $r$ mit $1 \leq r \leq n-1$ und jede Wahl von $i_1, \ldots, i_r$ mit $1 \leq i_1 < \ldots < i_r \leq n$. Unter Beachtung von $S_n = P(A_1 \cap \ldots \cap A_n) = 0$ liefert dann (6.15) das Resultat

$$P(V_{n,1} > k) = \sum_{r=1}^{n-1}(-1)^{r-1} \sum_{1\leq i_1<\ldots<i_r\leq n} (1 - p_{i_1} - \ldots - p_{i_r})^k \tag{6.16}$$

$(k = 1, 2, \ldots)$, woraus die Verteilung von $V_{n,1}$ durch Differenzbildung gemäß $P(V_{n,1} = k) = P(V_{n,1} > k-1) - P(V_{n,1} > k)$ erhalten werden kann. Für den Erwartungswert von $V_{n,1}$ ergibt sich mit (6.16) (geometrische Reihe, vgl. die auf Seite 14 hergeleitete Formel)

$$\begin{aligned}
E(V_{n,1}) &= 1 + \sum_{k=1}^{\infty} P(V_{n,1} > k) \\
&= 1 + \sum_{r=1}^{n-1}(-1)^{r-1} \sum_{1\le i_1<\ldots<i_r\le n} \frac{1-p_{i_1}-\ldots-p_{i_r}}{p_{i_1}+\ldots+p_{i_r}} \\
&= 1 + \sum_{r=1}^{n-1}(-1)^{r-1} \sum_{1\le i_1<\ldots<i_r\le n} \frac{1}{p_{i_1}+\ldots+p_{i_r}} - \sum_{r=1}^{n-1}(-1)^r \cdot \binom{n}{r} \\
&= \sum_{r=1}^{n-1}(-1)^{r-1} \sum_{1\le i_1<\ldots<i_r\le n} \frac{1}{p_{i_1}+\ldots+p_{i_r}} - (-1)^n.
\end{aligned}$$

Eine kompakte Darstellung von $E(V_{n,1})$ ist

$$E(V_{n,1}) = \int_0^{\infty}\left(1 - \prod_{j=1}^{n}\left(1 - e^{-p_j t}\right)\right) dt$$

(Boneh und Hofri, 1997, S. 48). Boneh und Hofri benutzen diese Formel, um zu beweisen, daß der Erwartungswert von $V_{n,1}$ in Abhängigkeit der Wahrscheinlichkeiten $p_1,\ldots,p_n$ für den Fall gleichwahrscheinlicher Fächer (d.h. $p_1 = \ldots = p_n = 1/n$) minimal wird. Dieses plausible Ergebnis war lange Zeit ein offenes Problem.

Im Spezialfall von 3 Fächern ergibt sich durch Komplementbildung in (6.16)

$$P(V_{3,1} \le k) = 1 - (1-p_1)^k - (1-p_2)^k - (1-p_3)^k + p_1^k + p_2^k + p_3^k$$

($k = 1, 2, 3, \ldots$) sowie

$$E(V_{3,1}) = \frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} - \frac{1}{1-p_1} - \frac{1}{1-p_2} - \frac{1}{1-p_3} + 1.$$

| $k$ | $P(V_{3,1} \leq k)$ $p_1 = p_2 = p_3 = \frac{1}{3}$ | $P(V_{3,1} \leq k)$ $p_1 = \frac{1}{2},\ p_2 = \frac{1}{3},\ p_3 = \frac{1}{6}$ |
|---|---|---|
| 3 | 0.2222 | 0.1667 |
| 4 | 0.4444 | 0.3333 |
| 5 | 0.6173 | 0.4707 |
| 6 | 0.7407 | 0.5787 |
| 7 | 0.8258 | 0.6629 |
| 8 | 0.8834 | 0.7286 |
| 9 | 0.9221 | 0.7802 |
| 10 | 0.9480 | 0.8212 |
| 15 | 0.9931 | 0.9328 |
| 20 | 0.9991 | 0.9736 |

**Tabelle 6.2:** Wahrscheinlichkeit für eine vollständige Serie nach höchstens $k$ Versuchen bei drei Fächern

Tabelle 6.2 zeigt die Wahrscheinlichkeiten $P(V_{n,1} \leq k)$ für den gleichwahrscheinlichen Fall $p_1 = p_2 = p_3 = 1/3$ sowie den Fall $p_1 = 1/2$, $p_2 = 1/3$, $p_3 = 1/6$. Beide Situationen können als Wartezeitprobleme mit einem Würfel aufgefaßt (und simuliert) werden. Dabei entsprechen im gleichwahrscheinlichen Fall je 2 der 6 möglichen Ergebnisse einem Fach. Im zweiten Fall werden etwa die Ergebnisse 1,2 oder 3 dem ersten Fach, die Ergebnisse 4 oder 5 dem zweiten Fach und das Ergebnis 6 dem dritten Fach zugeordnet. Es ist deutlich zu erkennen, daß die Wartezeit bis zum Erreichen einer vollständigen Serie bei ungleichen Wahrscheinlichkeiten für die einzelnen Fächer im Vergleich zum gleichwahrscheinlichen Fall *stochastisch größer* wird. Dies wird auch an den Erwartungswerten (7.3 gegenüber 5.5 im gleichwahrscheinlichen Fall) deutlich.

Im Spezialfall $n = 2$ läßt sich der Erwartungswert von $V_{n,1}$ auch sehr einfach über die Zerlegung nach dem Ergebnis des ersten Versuchs herleiten. Es gilt

$$E(V_{2,1}) = p_1 \cdot E(V_{2,1}|1) + p_2 \cdot E(V_{2,1}|2). \tag{6.17}$$

Hierbei ist $E(V_{2,1}|j)$ der Erwartungswert von $V_{2,1}$ unter der Bedingung, daß das erste Teilchen ins $j$–te Fach fällt. Da wir nach der Besetzung von Fach 1 im ersten Versuch zum Erreichen einer vollständigen Serie auf einen mit Wahrscheinlichkeit $p_2$ eintretenden Treffer in unabhängigen Experimenten warten, folgt

$$E(V_{2,1}|1) \;=\; 1+\frac{1}{p_2}$$

(vgl. Abschnitt 1) und analog $E(V_{2,1}|2) = 1 + 1/p_1$. Einsetzen in (6.17) liefert dann

$$E(V_{2,1}) \;=\; 1+\frac{p_1}{p_2}+\frac{p_2}{p_1}\,.$$

Da die Funktion $x \longmapsto 1+x+1/x$ im Bereich $x > 0$ ihr Minimum für $x = 1$ annimmt, ergibt sich hier mit elementaren Mitteln, daß die durchschnittliche Wartezeit zum Erreichen einer vollständigen Serie bei 2 Fächern im Fall gleichwahrscheinlicher Fächer am kürzesten ist.

Für Knobelfüchse

- Teilchen fallen unabhängig voneinander und rein zufällig in jeweils eines von $2 \cdot k$ Fächern. Die Zufallsvariable $H_{1,k}$ bezeichne die Anzahl der Teilchen, die benötigt werden, damit *die Hälfte* aller Fächer mit jeweils mindestens einem Teilchen besetzt ist. $H_{2,k}$ sei die Anzahl der Teilchen, welche danach noch zum Erreichen einer vollständigen Serie (jedes Fach enthält mindestens eine Kugel) benötigt werden. Welchen Erwartungswert besitzen $H_{1,k}$ und $H_{2,k}$? Was kann über das Verhältnis dieser Erwartungswerte für $k \to \infty$ ausgesagt werden?

- Ein Tourist, welcher 6 Städte besuchen will, startet seine Rundreise in einer beliebigen dieser Städte und wählt anschließend rein zufällig eine der restlichen fünf Städte für seine nächste Reise aus. Er möchte bereits besichtigte Städte nicht noch einmal besuchen; sein Gedächtnis ist jedoch so schlecht, daß er sich stets nur noch an seinen letzten Aufenthaltsort erinnern kann (den er als nächstes Reiseziel ausschließt). Wie viele Städtereisen muß der Tourist bis zum Ende seiner Besichtigungstour im Durchschnitt unternehmen, wenn er sein jeweils nächstes Reiseziel rein zufällig auswählt?

# 7 Weitere Wartezeitprobleme

Einige interessante Varianten und Verallgemeinerungen der betrachteten Kollisions- und Sammlerprobleme sind:

a) Eine Schulklasse bestehe aus $n$ Jungen und $n$ Mädchen. Wie groß muß $n$ mindestens sein, damit mit der Mindestwahrscheinlichkeit $1/2$ mindestens ein Mehrfachgeburtstag innerhalb der Menge der Jungen oder innerhalb der Menge der Mädchen auftritt? (Crilly und Nandy, 1987)

b) Wie lange dauert es, Jacks Haus zu bauen? (ein Sammlerproblem mit Nebenbedingungen, da man manche Teile des Hauses wie etwa den Schornstein erst dann sammeln kann, wenn man vorher andere Teile (das Dach) erhalten hat) (Riddiough und McColl, 1997)

c) In der Situation von Abschnitt 6.1 sei $T_{n,r}$ die Anzahl der Teilchen, die man benötigt, um $r$ *vollständige Serien* zu erzielen. Es gilt

$$E(T_{n,r}) \;=\; n \cdot \log n + (r-1) \cdot \log\log n + n \cdot K_r + o(n)$$

mit $K_r = 0.57721\ldots - \log(r-1)!$ (Newman und Shepp, 1960) sowie

$$\lim_{n\to\infty} P(T_{n,r} \le n \cdot (\log n + (r-1)\log\log n + x)) \;=\; \exp\left(-\frac{e^{-x}}{(r-1)!}\right)$$

($x \in \mathbb{R}$; Erdös und Rényi, 1961).

d) Am 26.11.1693 schrieb Newton an Pepys: „Ich nehme an, daß A mit einem Würfel und B mit zweien wirft, der erstere, bis er eine Sechs wirft, der letztere genausooft um zwei Sechsen, und ich fand, daß A im Vorteil ist [...]. Denn hier ist A deswegen im Vorteil, weil er als erster wirft.“ A ist tatsächlich im Vorteil, aber mitnichten, weil er als erster wirft! (Haller, 1997; Henze, 1998a)

# Literatur

**Blom, G., Holst, L. und Sandell, D.** (1994). Problems and Snapshots from the World of Probability. Springer–Verlag. New York.

**Boneh, A. und Hofri, M.** (1997). The Coupon-Collector Problem Revisited— A Survey of Engineering Problems and Computational Methods. Communications in Statistics – Stoch. Models 13, 39–66.

**Crilly, A. und Nandy, S.** (1987). The birthday problem for boys and girls. Mathem. Gazette 71, 19–22.

**Dwass, M.** (1969). More Birthday Surprises. Journ. Combinatorial Theory 7, 258–261.

**Engel, A.** (1987). Stochastik. Ernst Klett Schulbuchverlag. Stuttgart.

**Erdös, P. und Rényi, A.** (1961). On a Classical Problem of Probability Theory. Magy. Tud. Akad. Mat. Kutató Int. Közl 6, 215–220.

**Feller, W.** (1968). An Introduction to Probability Theory and Its Applications. Vol. 1. 3. Auflage. Verlag Wiley. New York.

**Galambos, J.** (1978). The Asymptotic Theory of Extreme Order Statistics. Verlag Wiley. New York.

**Haller, R.** (1997). Zog Pepys falsche Schlüsse? Und hat Newton recht? Stochastik in der Schule 17, 49–54.

**Henze, N.** (1999). Stochastik für Einsteiger, 2. Auflage. Verlag Vieweg, Braunschweig, Wiesbaden.

**Henze, N.** (1998a). Die Auflösung eines Wartezeit-Paradoxons oder Newton hatte nur teilweise recht! Stochastik in der Schule 19, 1999, 30–37.

**Henze, N.** (1998b). A Poisson Limit Theorem for a Generalized Birthday Problem. Statist. & Probab. Letters 39, 1998, 333–336.

**Heuser, H.** (1991). Lehrbuch der Analysis. Teil 1. 9. Auflage. Verlag B.G. Teubner. Stuttgart.

**Heuser, H.** (1993). Lehrbuch der Analysis. Teil 2. 8. Auflage. Verlag B.G. Teubner. Stuttgart.

**Holst, L.** (1995). The General Birthday Problem. Random Structures and Algorithms 6, 201–208.

**Joag–Dev, K. und Proschan, F.** (1992). Birthday Problem with Unlike Probabilities. American Mathem. Monthly 99, 10–12.

**Klamkin, M.S. und Newman, D.J.** (1967). Extensions of the Birthday Surprise. Journ. Combinatorial Theory 3, 279–282.

**Munford, A.G.** (1977). A Note on the Uniformity Assumption in the Birthday Problem. The American Statistician 31, 119.

**Newman, D.J. und Shepp, L.** (1960). The Double Dixie Cup Problem. Amer. Mathem. Monthly 67, 58–61.

**Riddiough, R. und McColl, J.** (1997). Wie lange dauert es, Jacks Haus zu bauen? Stochastik in der Schule 17, 49–54.

**Riehl, G.** (1997). Das Geburtstagsproblem und die Intuition. Praxis der Mathematik 39, 177–181.

**Schrage, G.** (1990). Ein Geburtstagsproblem. Mathem. Semesterber. 37, 251–257.

**Schrage, G.** (1992). Ein Geburtstagsproblem. Stochastik in der Schule 12, 30–36.

**Stadje, W.** (1989). Die Wartezeitverteilung für ein verallgemeinertes Sammlerproblem. Didaktik der Mathematik 4, 313–320.

**Strick, H.K.** (1991). Überlegungen zum Geburtstagsproblem. In: Mathematik lehren und lernen. Festschrift für Heinz Griesel. Postel, H., Kirsch, A., Blum, W. (Hrsg.), S. 207–216. Verlag Schroedel, Hannover.