Literatur

Engel, Arthur (1987): Stochastik. Stuttgart: Klett.

Hefendehl-Hebecker, Lisa; Törner, Günter: Über Schwierigkeiten bei der Behandlung der Kombinatorik. In: DdM 4 (1984), 245–262

Henze, Norbert (2008): Stochastik für Einsteiger. Braunschweig/Wiesbaden: Vieweg.

Ineichen, Robert (1984): Stochastik. Eine Einführung in die elementare Statistik und Wahrscheinlichkeitsrechnung. Luzern: Raeber.

Krengel, Ulrich (2003): Einführung in die Wahrscheinlichkeitstheorie und Statistik. Braunschweig/Wiesbaden: Vieweg.

Kütting, Herbert; Sauer, Martin J. (2008): Elementare Stochastik. Heidelberg/Berlin: Spektrum Akademischer Verlag.

Müller, P. H. (1991): Lexikon der Stochastik. Berlin: Akademie Verlag.

Selter, Christoph; Spiegel, Hartmut (2004): Elemente der Kombinatorik. In: Müller, G.; Steinbring, H.; Wittmann, E. C. (Hg.): Arithmetik als Prozess. Seelze: Kallmeyer, 291–310.

Anschrift des Verfassers

Dr. rer. nat. Martin J. Sauer
Westfälische Wilhelms-Universität Münster
Fachbereich Mathematik und Informatik
Institut für Didaktik der Mathematik und der Informatik
Fliednerstraße 21
48149 Münster
mjsauer@math.uni-muenster.de

„Lottofieber in Deutschland" – stochastisch betrachtet

Gerd Riehl, Barsinghausen

***Zusammenfassung**: Ein Pressebericht vom Herbst 2007 über das Lottofieber in Deutschland gibt Anlass, einige Fragestellungen zum Lottospiel zu analysieren. Hypothesentests mit Daten aus dieser Zeit zeigen eindrucksvoll, dass selbst große prozentuale Abweichungen vom Erwartungswert nicht unbedingt signifikant sind. In diesen Fällen ist auch die Normalverteilung als Approximation der Binomialverteilung problematisch.*

1 Einleitung

Wahrscheinlichkeiten im Zusammenhang mit dem Lottospiel kann man nur in bestimmten Fällen mit dem schlichten Laplaceansatz berechnen. Dies ist insbesondere für alle Aufgaben erlaubt, die sich auf den Ziehungsvorgang selbst beziehen, den man als Urnenexperiment ohne Zurücklegen auffasst, wobei die 49 Kugeln als physikalisch nicht unterscheidbar angesehen werden.

Dagegen versagt dieser Ansatz, wenn für eine konkrete Kombination von sechs Zahlen und eine konkrete Teilnehmerzahl nach der zu erwartenden Anzahl der Gewinner in den einzelnen Rängen oder den zu erwartenden Gewinnquoten gefragt ist. Man kann lediglich mittlere erwartete Quoten berechnen (Abschnitt 2), von denen aber die tatsächlichen Quoten in der Regel überzufällig stark abweichen werden (Abschnitt 3).

In Abschnitt 4 behandeln wir mit dem Jackpot zusammenhängende Probleme. Den Abschluss bilden einige Anmerkungen zur Normalverteilung und zu den verwendeten Tests sowie Hinweise auf mögliche Schüleraktivitäten.

2 Hoch- oder Tiefquotenreihe?

Nachdem bei den elf Lottoveranstaltungen in der Zeit vom 24. Oktober bis Ende November 2007 kein Spieler einen Treffer im 1. Rang erzielt hatte und der sogenannte Jackpot dadurch auf über 30 Mio. Euro angewachsen war, las man überall in der Presse vom „Lottofieber in Deutschland". So berichtete die Hannoversche Allgemeine Zeitung über die Samstagziehung vom 1. Dezember am darauf folgenden Montag, also noch bevor die Auswertungen der Lottozentralen vorlagen, unter der Schlagzeile „Warten auf den Lotto-König". In diesem Artikel wurden auch Einschätzungen und Prognosen eines „Lottoexperten" zitiert (Abb. 1), die wir im Folgenden mathematisch analysieren wollen.

Dabei wenden wir uns zunächst dem Begriff der „Tiefquotenreihe“ zu.

Der Präsident des Verbandes der privaten Lottovermittler, Norman Faber, geht von mindestens einem Glückspilz aus. „Mit einer sehr hohen Wahrscheinlichkeit von rund 95 Prozent ist er geknackt“, sagte der Lottoexperte vom größten deutschen Anbieter für Tippgemeinschaften. Mit der 12, 21, 23, 29, 35 und 41 seien häufig getippte Zahlen gezogen worden. Es bestehe eine „hohe Wahrscheinlichkeit“, dass der Jackpot sogar mehrfach geknackt worden sei. Er geht deshalb von einer sogenannten Tiefquotenreihe in allen Gewinnrängen aus. „Das bedeutet, dass es in allen Gewinnklassen verhältnismäßig wenig Geld geben wird.“

Abb. 1: Zeitungsausschnitt vom 3.12.2007

Im Text der Abbildung 1 wird die Hypothese geäußert, die am 1.12. gezogenen Zahlen seien „häufig getippte Zahlen“. Bei einer Ziehung mit solchen beliebten Zahlen gibt es dann in den einzelnen Rängen auch viele Gewinner, auf die sich ein fester Anteil aller Einsätze verteilt, mit der Folge, dass die Quoten entsprechend niedrig sind.

Um nun beurteilen zu können, ob dies in unserem Fall so ist, vergleichen wir die tatsächlichen Quoten der einzelnen Ränge mit den „mittleren Quoten“, die wir unter der Annahme berechnen, alle Tipps seien unabhängig voneinander durch einen Zufallsprozess zustande gekommen, bei dem jede der 49 Zahlen dieselbe Wahrscheinlichkeit besitzt, gezogen zu werden (Nullhypothese H_0).

In einem ersten Schritt ermitteln wir die erwartete Anzahl an Gewinnern bei der fiktiven Anzahl von $N^* = 10 \cdot \binom{49}{6} = 139\,838\,160$ abgegebenen Tipps (diese Größenordnung hat N, wenn der Jackpot gut gefüllt und das Lottofieber folglich hoch ist). Unter H_0 tritt dann jede der 13 983 816 möglichen Kombinationen durchschnittlich zehnmal auf.

Insbesondere hat die Anzahl der Spieler mit den 6 Gewinnzahlen den Erwartungswert 10, davon 9 im 2. Rang. Den 1. Rang, bei dem die Superzahl ins Spiel kommt und eventuell ein Jackpot ausgezahlt wird, lassen wir in diesem und dem folgenden Abschnitt zunächst unbeachtet. Diesen Sonderfall werden wir in Abschnitt 4 ausführlich behandeln.

Tabelle 1 zeigt die Definition der Ränge sowie für jeden Rang die erwartete Anzahl an Gewinnern und die Anteile am Ausschüttungsbetrag.

Rang	Treffer	erw. Anzahl	Anteil
1	6 mit Superzahl	1	10 %
2	6 ohne Superzahl	9	8 %
3	5 mit Zusatzzahl	60	5 %
4	5 ohne Zusatzzahl	2 520	13 %
5	4 mit Zusatzzahl	6 300	2 %
6	4 ohne Zusatzzahl	129 150	10 %
7	3 mit Zusatzzahl	172 200	8 %
8	3 ohne Zusatzzahl	2 296 000	44 %

Tabelle 1: Gewinnränge beim deutschen Lotto

Erläuterung zur Berechnung der Anzahlen am Fall „4 Treffer“: Es gibt dann drei Arten von Zahlen, die 6 eigentlichen Gewinnzahlen (von denen 4 angekreuzt sein müssen), eine Zusatzzahl (die für den 5., aber nicht den 6. Rang getippt sein muss) und 42 weitere Zahlen, von denen eine (5. Rang) bzw. zwei (6. Rang) gewählt sein müssen. Für die Zahl der möglichen Kombinationen erhält man

$$k_5 = \binom{6}{4} \cdot \binom{42}{1} = 630 \text{ bzw. } k_6 = \binom{6}{4} \cdot \binom{42}{2} = 12\,915$$

und für die erwartete Zahl der Gewinner $10 \cdot k_5$ bzw. $10 \cdot k_6$ (vgl. 3. Spalte von Tabelle 1).

Weicht die Zahl N der Tipps für eine Ausspielung von dem fiktiven Wert N^* ab, so ändern sich zwar die Erwartungswerte für die Anzahl der Gewinner in den einzelnen Rängen, nicht aber diejenigen für die Quoten, da sich die Einsätze und damit der Ausschüttungsbetrag im selben Verhältnis ändern wie die Teilnehmerzahl.

Es genügt also, im zweiten Schritt die erwarteten Quoten für den Fall von N^* Tipps zu bestimmen. Bei einem Einsatz von 0,75 € je Spiel und einer Ausschüttung von 50 % der Einsätze stehen dann 52 439 310 € für Gewinne zur Verfügung. Wie sich dieser Betrag auf die vom Jackpot unabhängigen Ränge 2 bis 8 verteilt und welche mittleren Quoten sich je Rang ergeben, ist aus dem linken Teil von Tabelle 2 ersichtlich. Zum Vergleich stehen rechts die tatsächlichen Quoten von drei Ziehungen.

Offenbar bilden die Zahlen vom 1. 12. 2007 (im Widerspruch zur Prognose des Experten Faber) eine Hochquotenreihe. Dagegen ergab die Ziehung am 5. 12. 2007 ein typisches Beispiel für eine Tiefquotenreihe. Dass man in beiden Fällen die Abweichungen von den mittleren Quoten nicht als zufallsbedingt ansehen kann, zeigen wir im folgenden Abschnitt.

Rg.	%	Ausschüttung	Gewinner	mittlere Quote	28. 11. 2007	1. 12. 2007	5. 12. 2007
2	8	4 195 144,80 €	9	466 127,20 €	854 821,40 €	606 248,90 €	263 865,10 €
3	5	2 621 965,50 €	60	43 699,43 €	71 235,10 €	54 129,30 €	34 632,20 €
4	13	6 817 110,30 €	2 520	2 705,20 €	3 926,70 €	3 573,20 €	1 853,10 €
5	2	1 048 786,20 €	6 300	166,47 €	218,20 €	207,10 €	128,10 €
6	10	5 243 931,00 €	129 150	40,60 €	45,90 €	47,60 €	34,10 €
7	8	4 195 144,80 €	172 200	24,36 €	28,40 €	28,90 €	18,10 €
8	44	23 073 296,40 €	2 296 000	10,05 €	10,00 €	11,30 €	9,20 €

Tabelle 2: Bestimmung der mittleren Quoten und Vergleich mit den realen Daten dreier Ausspielungen

3 Analyse der Abweichungen von den Erwartungswerten

Wie bereits erwähnt, ist die für einen bestimmten Rang unter H_0 erwartete Quote von der Anzahl N der abgegebenen Tipps unabhängig; für die zu erwartenden *Abweichungen* vom Erwartungswert gilt dies jedoch nicht mehr. Wir betrachten daher nun die Zufallsgrößen G_r der *Gewinner im Rang r*, deren Verteilung leichter zugänglich ist.

Wir beginnen wieder mit dem Fall $N = N^*$ und bezeichnen die zugehörigen Zufallsgrößen mit G_r^*. Für einen festen Rang r ($2 \leq r \leq 8$) ist G_r^* binomialverteilt mit den Parametern N^* und p_{0r}. Dabei ergibt sich p_{0r}, indem man μ_r^*, den Erwartungswert von G_r^* (Spalte 4 von Tabelle 2), durch N^* dividiert. Anschließend ergibt sich die Varianz von G_r^* zu $V(G_r^*) = \mu_r^* \cdot (1 - p_{0r})$ und die Standardabweichung σ_r^* daraus durch Radizieren. Tabelle 3 zeigt die Ergebnisse.

r	$\mu_r^* = E(G_r^*)$	p_{0r}	$V(G_r^*)$	σ_r^*
2	9	$6{,}436 \cdot 10^{-8}$	9	3,00
3	60	$4{,}291 \cdot 10^{-7}$	60	7,75
4	2 520	$1{,}802 \cdot 10^{-5}$	2 520	50,2
5	6 300	$4{,}505 \cdot 10^{-5}$	6 300	79,4
6	129 150	0,000924	129 031	359,2
7	172 200	0,001231	171 988	414,7
8	2 296 000	0,016419	2 258 302	1 502,8

*Tabelle 3: Varianz und Standardabweichung von Gr**

Verallgemeinern wir nun die Ergebnisse auf den Fall einer beliebigen Teilnehmerzahl N, so sind zwar die p_{0r} unabhängig von N, Erwartungswerte und Varianzen jedoch proportional zu N.

Bei einer Ausspielung mit $N = c \cdot N^*$ abgegebenen Tipps ist also $\mu_r = c \cdot \mu_r^*$ und $V(G_r) = c \cdot V(G_r^*)$, während für die Standardabweichung $\sigma_r = \sqrt{c} \cdot \sigma_r^*$ gilt.

Beispiel 1: Eine Hochquotenreihe

Zur Ziehung am 1. 12. 2007 wurden 202 082 967 Tipps abgegeben ($c = 1{,}445$; $\sqrt{c} = 1{,}202$). Tabelle 4 zeigt links die damit berechneten Werte für μ_r und σ_r, daneben die tatsächlich eingetretenen Besetzungszahlen g_r der Zufallsgrößen G_r und rechts deren Abweichungen von den μ_r in Prozent und in Vielfachen von σ_r (gemäß $g_r - \mu_r = z_r \cdot \sigma_r$).

r	$\mu_r = E(G_r)$	σ_r	g_r	Abweichung	
	unter H_0			%	z_r
2	13,01	3,61	10	-23,1	-0,83
3	86,71	9,31	70	-19,3	-1,79
4	3 642	60,3	2 757	-24,3	-14,7
5	9 104	95,4	7 317	-19,6	-18,7
6	186 637	432	159 055	-14,8	-63,9
7	248 850	499	209 481	-15,8	-79,0
8	3 317 996	1 807	2 949 045	-11,1	-204

Tabelle 4: Daten zur Ziehung am 1. 12. 2007

Wir hatten bereits anhand der Daten von Tabelle 2 vermutet, es liege hier eine Hochquotenreihe vor, was äquivalent damit ist, dass die G_r kleinere Werte annehmen als unter H_0 erwartet. Im Sinne der Beurteilenden Statistik bedeutet dies: Wir haben die Gegenhypothese H_1: $p_r < p_{0r}$ zu testen. Die aus der Tabelle ersichtlichen prozentualen Abweichungen sind alle negativ und von ähnlicher Größenordnung, sprechen also für H_1. Obwohl die Zahl der Gewinner im 2. Rang um mehr als 20 % niedriger ist als erwartet, kann die Nullhypothese hier nicht verworfen werden; Tabelle 5 zeigt links die mit EXCEL berechnete Binomialverteilung von G_2 und die zugehörige Verteilungsfunktion.

Bei unserem einseitigen Test ergibt sich daraus $k < 7$ als Verwerfungsbereich auf dem 5%-Niveau bzw. $k < 5$ auf dem 1%-Niveau.

Diese Situation ändert sich grundlegend, wenn μ_r und damit σ_r größer werden. Schon für den dritten Rang reicht der beobachtete Wert $g_3 = 70$ (trotz kleinerer prozentualer Abweichung als bei $r = 2$) aus, um H_0 auf dem 5%-Niveau zu verwerfen. Um H_0 auch auf dem 1%-Niveau ablehnen zu können, müsste allerdings $G_3 < 66$ sein, wie man dem rechten Teil von Tabelle 5 entnimmt.

k	$P(G_2=k)$	$P(G_2 \leq k)$	k	$P(G_3=k)$	$P(G_3 \leq k)$
0	$2{,}25 \cdot 10^{-6}$	$2{,}25 \cdot 10^{-6}$	63	…	0,0047
1	$2{,}92 \cdot 10^{-5}$	$3{,}15 \cdot 10^{-5}$	64	0,0019	0,0066
2	0,0002	0,0002	65	0,0025	***0,0091***
3	0,0008	0,0010	66	0,0033	0,0124
4	0,0027	***0,0037***	67	0,0043	0,0167
5	0,0070	0,0107	68	0,0055	0,0221
6	0,0151	***0,0258***	69	0,0069	0,0290
7	0,0281	0,0539	70	0,0085	0,0375
8	0,0456	0,0995	71	0,0104	***0,0478***
9	0,0659	0,1654	72	0,0125	0,0603
10	0,0857	0,2512	73	0,0148	0,0752

Tabelle 5: Verwerfungsbereiche für r = 2 und r = 3 bei der Veranstaltung am 1. 12. 2007

Bei allen weiteren Rängen liegen hochsignifikante Abweichungen von der Nullhypothese vor, wie schon ein Blick auf die Werte von z_r in der letzten Spalte von Tabelle 4 lehrt. Man kann daher getrost H_0 ablehnen und die Gegenhypothese annehmen.

Beispiel 2: Eine Tiefquotenreihe

Zu ähnlichen Ergebnissen, nur mit umgekehrten Vorzeichen, führt die Analyse der Daten für die folgende Ziehung (mit N = 184 705 570; c = 1,321; $\sqrt{c} = 1{,}149$). In den Tabellen 6 und 7 sind die zu Tabelle 4 und 5 analogen Daten wiedergegeben.

Um die Hypothese „Tiefquotenreihe" zu prüfen, testen wir wiederum einseitig, aber hier lautet H_1: $p_r > p_{0r}$. Den Werten der Verteilungsfunktionen von G_2 und G_3 in Tabelle 7 bzw. deren Gegenwahrscheinlichkeiten entnimmt man $G_2 > 18$ (oder $G_2 \geq 19$) und $G_3 > 94$ auf dem 5%-Niveau sowie $G_2 > 21$ und $G_3 > 101$ auf dem 1%-Niveau als Verwerfungsbereiche.

Die Grenzen des 1%-Niveaus werden in diesem Beispiel nur knapp verfehlt, sodass man H_0 mit gutem Grund bereits anhand der Daten für den 2. und 3. Rang verwerfen kann. Erst recht gilt dies für die Ränge 4 bis 8; hier reicht zum Nachweis der Signifikanz wegen der hohen Werte von z_r die Tschebyschew-Ungleichung.

r	$\mu_r = E(G_r)$	σ_r	g_r	Abweichung	
	unter H_0			%	z_r
2	11,9	3,45	21	+76,7	+2,64
3	79,3	8,90	100	+26,2	+2,33
4	3 329	57,7	4 859	+46,0	+26,5
5	8 321	91,2	10 808	+29,9	+27,3
6	170 588	413	202 556	+18,7	+77,4
7	227 451	477	305 561	+34,3	+164
8	3 032 677	1 727	3 303 393	+8,9	+157

Tabelle 6: Daten zur Ziehung am 5. 12. 2007

k	$P(G_2 \leq k)$	$P(G_2 > k)$	k	$P(G_3 \leq k)$	$P(G_3 > k)$
12	0,5888	0,4112	92	0,9289	0,0711
13	0,6933	0,3067	93	0,9423	0,0577
14	0,7821	0,2179	94	0,9535	***0,0465***
15	0,8524	0,1476	95	0,9629	0,0371
16	0,9047	0,0953	96	0,9707	0,0293
17	0,9412	0,0588	97	0,9770	0,0230
18	0,9654	***0,0346***	98	0,9822	0,0178
19	0,9805	0,0195	99	0,9863	0,0137
20	0,9894	0,0106	100	0,9895	0,0105
21	0,9945	***0,0055***	101	0,9921	***0,0079***
22	0,9973	0,0027	102	0,9941	0,0059

Tabelle 7: Verwerfungsbereiche für r = 2 und r = 3 bei der Veranstaltung am 5. 12. 2007

Beispiel 3: Eine „unbeliebte" Lottozahl

Zum Abschluss dieses Abschnitts behandeln wir noch ein Beispiel, das sich in die Klassifizierung als Hoch- oder Tiefquotenreihe nicht einzufügen scheint. Bei der Ziehung am 28. 11. 2007, zu der 113 976 187 Tipps abgegeben wurden (c = 0,815; $\sqrt{c} = 0{,}903$), ergaben sich die Daten in Tabelle 8.

r	$\mu_r = E(G_r)$	σ_r	g_r	Abweichung	
	unter H_0			%	z_r
2	7,34	2,71	4	-45,5	-1,23
3	48,90	6,99	30	-38,7	-2,70
4	2 054	45,3	1 415	-31,1	-14,1
5	5 135	71,7	3 916	-23,7	-17,0
6	105 265	324	93 114	-11,5	-37,5
7	140 353	374	120 045	-14,5	-54,2
8	1 871 373	1 357	1 877 933	+0,4	+4,84

Tabelle 8: Daten zur Ziehung am 28. 11. 2007

Auffällig ist hier die Vorzeichenumkehr bei der Abweichung zwischen g_8 und μ_8; im 8. Rang liegt die Quote *unter* der mittleren Quote (vgl. Tab. 2), während die oberen Ränge das typische Aussehen einer Hochquotenreihe haben. Allerdings erkennt man auch schon in Tabelle 4 und 6 eine solche „gegenläufige Tendenz", auch dort sind im 8. Rang die Beträge der prozentualen Abweichungen deutlich kleiner als in den oberen Rängen.

Diese zunächst überraschende Erscheinung findet eine einfache Erklärung darin, dass Abweichungen von den erwarteten Teilnehmerzahlen in den oberen (oder allen) Gewinnrängen in der einen Richtung – hier nach unten – spätestens bei den Verlierern (mit 2, 1 oder 0 Gewinnzahlen) durch Abweichungen in der andere Richtung – in diesem Fall nach oben – kompensiert werden müssen.

Dieser Effekt ist in unseren Beispielen bisher nicht deutlich geworden, da wir in den Tabellen nur die *Gewinner* aufgeführt haben. So macht es übrigens auch der Deutsche Lotto-Block, der die rund 98 % der Spieler, die leer ausgehen, aus verständlichen Gründen nicht erwähnt.

An den Daten der Tabelle 8 ist noch etwas anderes auffällig: Von je zwei Rängen mit gleicher Anzahl richtig getippter Gewinnzahlen weist der höhere (*mit* Zusatzzahl) gegenüber dem niedrigeren (*ohne* Zusatzzahl) die betragsmäßig größere prozentuale Abweichung auf. Dies lässt vermuten, dass mit der 22 als Zusatzzahl eine weniger beliebte und daher seltener getippte Lottozahl gezogen wurde.

Um diese Hypothese am Beispiel „4 Richtige" zu testen, formulieren wir H_0 in der Form „Die 97 030 Gewinner haben jede der 43 Nichtgewinnzahlen mit der Wahrscheinlichkeit $p_0 = \frac{1}{43}$ getippt". Die Anzahl der Gewinner im 5. Rang ist dann binomialverteilt mit n = 97 030 und $p = 2 \cdot p_0$ (da jeder der n Spieler 2 der 43 Nichtgewinnzahlen angekreuzt hat, ist dies die Wahrscheinlichkeit, dass eine davon die 22 ist). Wenn H_0 gilt, wären also 4 513 Gewinner im 5. Rang zu erwarten; der eingetretene Wert 3 916 ist bei $\sigma = 65{,}6$ um $9{,}1\sigma$ – also hochsignifikant – kleiner als erwartet.

Auch hier kann man H_0 nicht verwerfen, wenn man nur die oberen Ränge berücksichtigt: Im 3. Rang wären bei n = 1 445 Spielern mit 5 Richtigen und $p = 1 \cdot p_0$ (bei hier nur einer Nichtgewinnzahl) 33,6 Gewinner zu erwarten. Der eingetretene Wert 30 liegt um weniger als eine Standardabweichung darunter ($\sigma = 5{,}73$).

Noch deutlicher als bei 4 Richtigen sprechen auch hier die Zahlen für 3 Richtige gegen H_0. Es ist n = 1 997 978 und $p = 3 \cdot p_0$, sodass im 7. Rang 139 394 Gewinner zu erwarten waren, während die tatsächliche Anzahl von 120 045 (bei $\sigma = 360$) um mehr als 50σ darunter lag. Man kann daher die 22 mit großer Sicherheit als eine „unbeliebte" Zahl ansehen, die unterdurchschnittlich oft getippt wird.

4 Der Jackpot

Wir haben in den bisherigen Untersuchungen nur die Gewinnränge 2 bis 8 berücksichtigt. In diesem Abschnitt behandeln wir nun den 1. Rang, der sich von den anderen in zwei Punkten unterscheidet:

Ein Spieler mit 6 Richtigen gewinnt im 1. Rang, wenn die Endziffer seiner Spielscheinnummer mit der in einem Laplaceexperiment ausgelosten Superzahl übereinstimmt; sonst gewinnt er im 2. Rang. Also ist die bedingte Wahrscheinlichkeit P (Superzahl richtig|6 Richtige) = 0,1, woraus man die in Tabelle 1 ohne Herleitung angegebenen Erwartungswerte $\mu_1^* = 1$ und $\mu_2^* = 9$ erhält. Als zweite Besonderheit kommt beim 1. Rang der sogenannte Jackpot ins Spiel: Wenn kein Teilnehmer im 1. Rang gewinnt, wird der darauf entfallende Anteil der Ausschüttungssumme dem 1. Rang der folgenden Veranstaltung zugeschlagen.

So hatten sich im Jackpot bis Ende November 2007 in 11 Ziehungen ohne Gewinner im 1. Rang fast 30,9 Mio. Euro angesammelt und nach der Ziehung am 1. 12. knapp 38,5 Mio. Euro. Diese verteilten sich dann, nachdem der Jackpot am 5. 12. „geknackt" worden war, zusammen mit den regulären 6,9 Mio. Euro dieser Ausspielung auf drei Gewinner im 1. Rang, von denen jeder über 15,1 Mio. Euro erhielt.

Wir untersuchen nun die Prognosen hinsichtlich des Jackpots im Zeitungsartikel der Abbildung 1. Als der Lottoexperte Faber die Aussagen machte, wusste er noch nicht, dass zwar 10 Spieler alle 6 Gewinnzahlen richtig getippt hatten, aber keiner von ihnen einen Schein mit der Superzahl 0 besaß. Wir bezeichnen mit R die Zufallsgröße *„Anzahl der Spieler mit 6 Richtigen"*, mit S die *„Anzahl der Spieler mit 6 Richtigen und Superzahl"*. Am 1. 12. waren also die Ereignisse $R = 10$ und $S = 0$ eingetreten.

In unserem Beispiel ist S offenbar binomialverteilt mit $n = 10$ und $p = 0{,}1$; damit ist $P(S = 0) = 0.9^{10}$, also rund 35 %. Die Wahrscheinlichkeit, dass der Jackpot in dieser Situation (bei $R = 10$) geknackt wird, ist $P(S > 0) = 1 - P(S = 0) \approx 65$ %, also deutlich kleiner als die behaupteten „rund 95 Prozent".

Auch Fabers Aussage, es bestehe eine „hohe Wahrscheinlichkeit" für ein mehrfaches Knacken des Jackpots erweist sich als falsch: Diese ist nur $P(S > 1) = P(S > 0) - P(S = 1) \approx 0{,}65 - 0{,}39 = 0{,}26$.

Die Hauptursache der Fehleinschätzungen liegt in der irrtümlichen Annahme einer Tiefquotenreihe. Daran wird sehr schön erkennbar, dass es hier um *bedingte* Wahrscheinlichkeiten geht. Durch eine genauere Notation kann man dies verdeutlichen: $P(S > 0 \mid R = 10) \approx 0{,}65$ und $P(S > 1 \mid R = 10) \approx 0{,}26$.

Allerdings wären die Mutmaßungen von Abb. 1 realistischer gewesen, wenn tatsächlich am 1. 12. eine Tiefquotenreihe aufgetreten wäre.

Eine solche lag dann, wie oben gezeigt, am 5. 12. vor. Bei dieser Ziehung war $g_1 = 3$ und $g_2 = 21$, S also binomialverteilt mit $n = 24$ und $p = 0{,}1$. Die Wahrscheinlichkeit, den Jackpot zu knacken, ist nun $P(S > 0 \mid R = 24) = 1 - 0{,}9^{24} \approx 0{,}92$ und wegen $P(S = 1 \mid R = 24) = 24 \cdot 0{,}1 \cdot 0{,}9^{23} \approx 0{,}21$ ergibt sich als Wahrscheinlichkeit für mehrfaches Knacken des Jackpots $P(S > 1 \mid R = 24) \approx 0{,}92 - 0{,}21 = 0{,}71$. Diese ist nun tatsächlich recht hoch, während erstere immer noch unter 95 % liegt.

Damit der Jackpot mit einer Wahrscheinlichkeit von mindestens 95 % geknackt wird, müsste die Zahl k der Spieler mit 6 Richtigen so hoch sein, dass $P(S = 0 \mid R = k) = 0{,}9^k < 0{,}05$ wird; es muss also $k \cdot \lg 0{,}9 < \lg 0{,}05 \Leftrightarrow k > \frac{\lg 0{,}05}{\lg 0{,}9} \approx 28{,}4$ sein, d. h., mindestens 29 Spieler müssten 6 Richtige haben.

5 Schlussbemerkungen

Bei den Hypothesentests der Abschnitte 3 und 4 konnten wir Binomialverteilungen benutzen, weil jeweils nur *ein* bestimmtes Ereignis als „Treffer" beobachtet wurde (in 3: Gewinn in einem festen Rang; in 4: Superzahl richtig). Dabei haben wir unter H_0 die Wahrscheinlichkeit für das Eintreten von „Treffer" als konstant angenommen und als Zufallsgröße die *Anzahl der Treffer* betrachtet.

Zur Approximation der Binomialverteilung

Nun ist den Schülern zwar das Binomialmodell im Zusammenhang mit Hypothesentests vertraut, die Berechnung von Verteilungen wie in Abschnitt 3 ist aber sehr mühsam, wenn als Hilfsmitteln nur einfache Taschenrechner zur Verfügung stehen. Üblicherweise arbeitet man in dieser Situation mit der Normalverteilung als Approximationen für die Binomialverteilung, wobei als Faustregel für deren Anwendbarkeit die Bedingung $\sigma^2 > 9$ gilt.

Obwohl diese Voraussetzung am 1. 12. bei G_2 mit $\sigma^2 \approx 13$ klar erfüllt war (vgl. Tab. 4), weichen hier Normal- und Binomialverteilung an den Grenzen der Verwerfungsbereiche stark voneinander ab.

Die kumulierte Binomialverteilung $B_{n;p}$ wird durch die Normalverteilung gut approximiert, wenn die Normalverteilungskurve die Stufen des Graphen der Treppenfunktion etwa in der Mitte schneidet. Dann gilt $B_{n;p}(k) \approx \Phi\left(\frac{k + 0{,}5 - \mu}{\sigma}\right)$. Dies ist am besten nahe $\mu - \sigma \approx 9{,}4$ und $\mu + \sigma \approx 16{,}6$ der Fall; innerhalb der 1σ-Umgebung von μ sind die Werte der Normalverteilung zu klein (Abbildung 2).

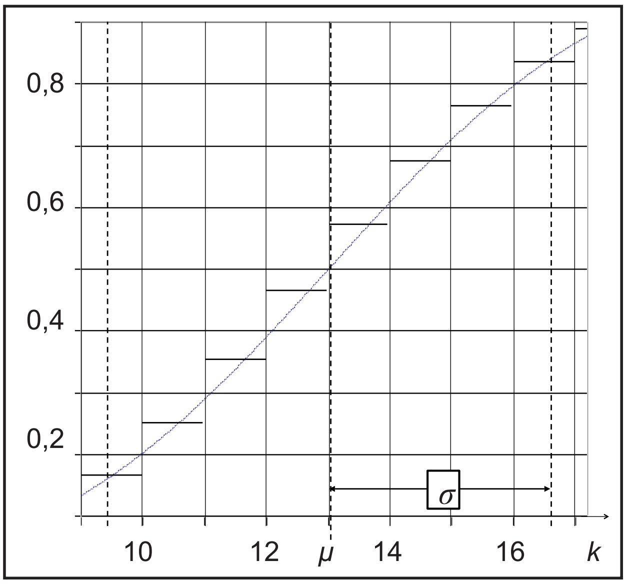

Abb. 2: Normalverteilung zu klein für |k – μ| < σ

Außerhalb der 1σ-Umgebung von μ wächst der relative Fehler bei der Approximation durch die Normalverteilung rasch an, wie die Abbildungen 3 und 4 veranschaulichen.

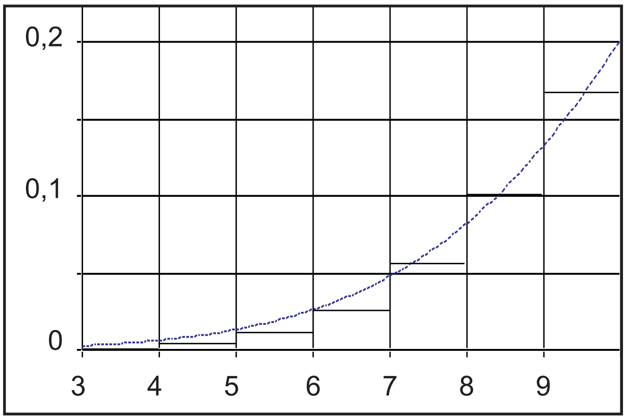

Abb. 3: Schlechte Approximation für k « μ (die Werte der Normalverteilung sind zu groß)

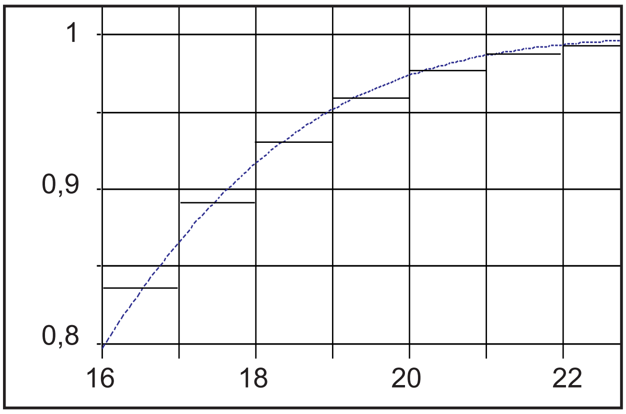

Abb. 4: Schlechte Approximation für k » μ (die Werte der Normalverteilung sind zu groß)

Wenn man versucht, Verwerfungsbereiche mit der Normalverteilung zu bestimmen, können Fehler auftreten, denn die Grenzen liegen, abhängig vom Signifikanzniveau, etwa 2σ vom Erwartungswert entfernt in Gebieten schlechter Approximation.

So ergibt sich beim rechtsseitigen Test auf dem 1%-Niveau als Verwerfungsbereich $G > 22$ [aus $B_{n;p}(21) = 0{,}9859$ und $B_{n;p}(22) = 0{,}9923$], während die Normalverteilung mit $\Phi\left(\frac{21{,}5-\mu}{\sigma}\right) = 0{,}9907$ den Bereich $G > 21$ liefert. Dieser besitzt aber mit $P(G>21) = 1 - B_{n;p}(21) = 0{,}0141$ eine deutlich zu große Irrtumswahrscheinlichkeit.

Auch beim zweiseitigen Test auf dem 1%-Niveau ergibt sich mit der Normalverteilung ein falscher Annahmebereich ($4 \le G_2 \le 22$ statt $5 \le G_2 \le 23$). Dieses Beispiel zeigt, dass bei der Approximation durch die symmetrische Normalverteilung die *Schiefe* der Binomialverteilung verloren geht.

In Fällen wie den hier betrachteten erhält man übrigens ausgezeichnete Näherungswerte mit der auch für einfache Taschenrechner zugänglichen Poisson-Verteilung.

Binomialtest oder χ^2-Test?

Es mag etwas mühselig erscheinen, dass wir jeden Rang einzeln mit einer neuen Binomialverteilung getestet haben, und in der Tat wäre es einfacher, als Modell eine Multinomialverteilung zu wählen mit den Grundwahrscheinlichkeiten p_{0r} (Tabelle 3) sowie p_{01} für den 1. Rang und p_{09} für die Verlierer. Die gegebene Häufigkeitsverteilung könnte man dann mit dem χ^2-Test auf signifikante Abweichung vom Modell testen.

Abgesehen davon, dass dieser Test wegen seiner Komplexität und des Begriffs der Freiheitsgrade im Schulunterricht selten behandelt wird, spricht auch ein innermathematischer Grund dagegen. In unseren Beispielen ging es stets um *einseitige* Tests; mit dem χ^2-Test überprüfen wir aber eine ganze Verteilung, in der Abweichungen in einer Richtung für bestimmte Merkmalswerte durch entgegengesetzt gerichtete bei anderen Werten kompensiert werden, ein Phänomen, das uns schon bei dem Beispiel 3 in Abschnitt 3 begegnet ist.

Schüleraktivitäten

Mit dem Kenntnisstand von Tabelle 3 können die Schülerinnen und Schüler die Daten der Tabellen 4, 6 und 8 selbstständig erarbeiten und dabei die in Abschnitt 3 behandelten Auffälligkeiten entdecken und analysieren.

Die auch für Schüler interessante Frage, wie man Tiefquotenreihen beim Tippen vermeidet, lässt sich an Beispielen erörtern, die man mit Anzahl der Gewinner und Quoten im Internet findet [1]. Dabei muss man darauf achten, nicht nur einzelne häufig getippte Zahlen zu vermeiden, sondern auch bestimmte Zahlenkombinationen.

So ergeben die Reihen 4, 6, 12, 18, 24, 30, (36) vom 15. 2. 2003 und 9, 13, 23, 27, 38, 40, (29) vom 4. 10. 1997 charakteristische Muster auf dem 7×7-Feld eines Tippscheins; aus den Zahlen 1, 4, 17, 31, 33, 49, (19) vom 21. 1. 1978 lassen sich ungewöhnlich viele Geburtsdaten bilden. In allen drei Fällen wurden die Ausschüttungsbeträge für zwei benachbarte Ränge zusammengelegt, damit nicht auf den höheren Rang eine geringere Quote entfiel als auf den niedrigeren.

Website

[1] www.lotto.de/lotto_6aus49_archiv.html (21.3.2008)

Anschrift des Verfassers:

Gerd Riehl
Obere Mark 6
30890 Barsinghausen
Elfriede.Riehl@t-online.de