Auffällige Lottoreihen – „kurios“ oder einfach nur Zufall?

Gerd Riehl, Barsinghausen

Zusammenfassung: *Eine Pressemeldung über das „kuriose“ Ergebnis einer Lottoziehung gibt Anlass, die Wahrscheinlichkeit für das Auftreten bestimmter auffälliger Muster und für deren wiederholtes Vorkommen zu berechnen. Vergleicht man die so gewonnenen theoretischen Werte mit den empirischen Daten der bisher mehr als 5000 Ziehungen im Lotto „6 aus 49“, so findet man meist eine befriedigende Übereinstimmung – bis auf einen unwahrscheinlich krassen Ausreißer.*

1 Einleitung

Wenn die 6 Gewinnzahlen einer Lottoziehung im 7 × 7-Raster des Tippscheins ein auffälliges Muster bilden, berichtet über ein solches Ereignis auch gern die Tagespresse. Besonders sog. „Mehrlinge“, also mehrere direkt aufeinander folgende Gewinnzahlen, widersprechen der naiven Vorstellung von Zufälligkeit. Die folgende Pressemeldung (HAZ 2015, S. 8) über die Ausspielung vom 11. Februar 2015 ist ein typisches Beispiel, in dem auch die Fehlvorstellung deutlich wird, ein wiederholtes Auftreten desselben Vierlings sei sehr unwahrscheinlich.

Viererreihe wurde bereits gezogen

MÜNSTER. Lotto-Seltenheit im Doppelpack: Ausgerechnet die bei der vergangenen Lottoziehung ermittelten Gewinnzahlen in seltener Viererreihung (25-26-27-28) sind schon vor mehr als sechs Jahren einmal gezogen worden. „Ein solcher Vierling ist sehr selten“, sagte Westlotto-Sprecher Axel Weber am Donnerstag. In den bisherigen 5351 Lotto-Ziehungen seit 1955 habe es insgesamt 26 Vierlinge gegeben. Umso kurioser sei es, dass die identische Kombination bereits am 1. November 2008 unter den sechs Richtigen auftauchte. „So ist das mit Wahrscheinlichkeiten. Unsere Maschine hat schließlich kein Gedächtnis“, versicherte der Lotto-Sprecher. Reich macht die seltene Reihung der vier Zahlen allerdings allenfalls in Kombination mit weiteren Gewinnzahlen: Reihen sind bei den Spielern beliebt und so sinken die Gewinnquoten.

In den folgenden Abschnitten untersuchen wir, wie wahrscheinlich (oder unwahrscheinlich) ein Vierling in einer einzelnen Ziehung bzw. die Wiederholung eines bereits aufgetretenen Vierlings ist. Die danach zu erwartenden Anzahlen vergleichen wir mit den tatsächlich aufgetretenen Häufigkeiten.

2 Wie selten ist ein Vierling?

Wir bezeichnen einen Vierling (k; $k + 1$; $k + 2$; $k + 3$) mit V_k, wobei k die Werte 1 bis 46 annehmen kann. Es ist kaum möglich, die Wahrscheinlichkeit p_1 für das Auftreten eines Vierlings in einer einzelnen Ziehung genauer zu schätzen, als es der in der Zeitungsmeldung zitierte Lottosprecher mit der Formulierung „sehr selten“ getan hat. Um p_1 exakt zu berechnen, ermitteln wir die Anzahl aller möglichen Lottoreihen, die einen echten Vierling enthalten (der also nicht Bestandteil eines Fünflings oder Sechslings ist[1]). Dies erfordert eine Fallunterscheidung:

- Die „Randvierlinge“ V_1 und V_{46} dürfen nicht mit der 5 bzw. 45 kombiniert sein, sondern mit zwei der jeweils 44 Zahlen über 5 bzw. unter 45. Dafür gibt es je $\binom{44}{2} = 946$, zusammen also 1892 Möglichkeiten.
- Jeder der übrigen 44 Vierlinge darf nicht mit dem Vorgänger seiner kleinsten oder dem Nachfolger seiner größten Zahl (im Beispiel V_{25} des Zeitungsartikels also mit 24 oder 29) kombiniert sein, sondern mit zwei der restlichen 43 Zahlen. Für jeden der 44 Vierlinge gibt es $\binom{43}{2} = 903$, zusammen also 39732 Möglichkeiten.

Bei insgesamt $\binom{49}{6} = 13983816$ möglichen Tippreihen erhält man als gesuchte Wahrscheinlichkeit

$$p_1 = \frac{1892 + 39732}{13983816} = \frac{41624}{13983816} \approx 0{,}00298 \approx \frac{1}{336}.$$[2]

Im Folgenden bezeichnen wir die Zufallsvariable „Anzahl der Vierlinge in n_1 Lottoziehungen“ mit T; sie ist offenbar binomialverteilt mit den Parametern n_1 und p_1. In den $n_1 = 5356$ Lottoziehungen bis zum 28.2.2015 war $T = 25$ (Lottostatistik, 2015)[3], also deutlich mehr als der Erwartungswert

$$\mu_1 = n_1 \cdot p_1 \approx 15{,}94$$

der Trefferanzahl T. Mit der Standardabweichung

$$\sigma_1 = \sqrt{\mu_1 \cdot (1 - p_1)} \approx 3{,}99$$

ergibt sich $|25 - \mu_1| \approx 2{,}27\sigma_1$, eine Abweichung, die nur in 2,3 % aller Fälle überschritten wird.

Abbildung 1 veranschaulicht die zeitliche Abfolge aller Ausspielungen im Lotto „6 aus 49“. Die jeweils nach n Ziehungen erreichte Trefferanzahl t kann direkt mit ihrem Erwartungswert μ und dem üblicherweise als „Signifikanzgrenze“ verwendeten Wert $\mu + 2\sigma$ verglichen werden. Die Stufenlängen im Graphen von $n \to t$ sind ein Maß für die Zufallsvariable „Wartezeit von einem Auftreten eines Vierlings bis zum nächsten“, die wir mit W bezeichnen.

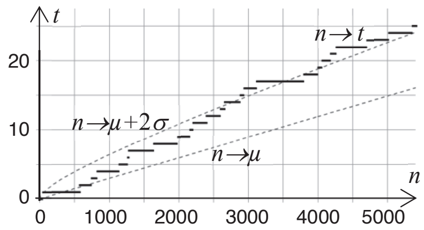

Abb. 1: Anzahl *t* der Lotto-Vierlinge nach *n* Ziehungen

Auffällig ist die kurze Wartezeit zwischen $t = 6$ und $t = 7$, wo $W = 2$ war.[4] Die Wahrscheinlichkeit, dass man nur eine oder zwei Ziehungen auf einen weiteren Vierling warten muss, ist mit

$$p_2 := P(W \le 2) = 1 - (1 - p_1)^2 \approx 0{,}00594$$

sehr klein. Allerdings relativiert sich dieses Urteil, wenn man die Zufallsvariable $K =$ „Anzahl kurzer Wartezeiten ($W < 3$) in 25 Vierlingsziehungen“ betrachtet. Es gilt nämlich

$$P(K = 0) = (1 - p_2)^{25} = [1(1 - p_1)^2]^{25}$$
$$= (1 - p_1)^{50} \approx 0{,}8615\,.$$

Das beobachtete Ereignis „$K > 0$“ war also immerhin mit Wahrscheinlichkeit 0,14 zu erwarten.

3 Wie kurios ist eine Wiederholung?

Wir prüfen nun die Aussage des Lottosprechers, es sei (wohl angesichts der Seltenheit eines Vierlings) „umso kurioser, dass die identische Kombination bereits am 1. November 2008 unter den sechs Richtigen auftauchte“.

Hierzu ist zu bemerken, dass Wiederholungen nur dann auffallen, wenn sie wie hier bei einer besonderen Konstellation auftreten; zwei andere Viererkombinationen aus den Gewinnzahlen vom 11.2.2015 kamen ebenfalls schon einmal vor: (6-26-28-47) am 3.8.1968 und (6-27-28-47) am 20.5.2009. Dies blieb ebenso unerwähnt wie die Tatsache, dass sogar fünf der aktuellen Zahlen bereits am 1.11.2008 gezogen wurden, außer dem Vierling auch die 6[5].

Die Frage nach Wiederholung eines Vierlings stellt sich nur dann, wenn überhaupt wieder einmal einer gezogen wurde, was bisher 25-mal der Fall war. Die Zufallsvariable X gebe an, wie oft ein bestimmter (beliebig gewählter) Vierling V_k in der Folge von 25 Vierlingsziehungen aufgetreten ist. In einer einzelnen Vierlingsziehung ist dies mit der Wahrscheinlichkeit $p_3 \approx \frac{1}{46}$ der Fall, da es insgesamt 46 fast gleichwahrscheinliche Vierlinge gibt[6]. Die Verteilung von X kann man also gut durch die Binomialverteilung mit $n_3 = 25$ und $p_3 = \frac{1}{46}$ approximieren. Tabelle 1 zeigt die für unsere Aufgabe relevanten Werte dieser Verteilung.

X	0	1	2	> 2
$P(X = x)$	0,5773	0,3207	0,0855	0,0165

Tab. 1: Binomialverteilung $b_{25;\,1/46}$

Unter allen 46 möglichen Vierlingen sind demnach 26,55 zu erwarten, die gar nicht vorkommen, während 14,75 einmal und 4,7 mindestens zweimal in 25 Ziehungen zu erwarten sind.

Die tatsächlichen absoluten Häufigkeiten[7] weichen mit 28, 11 und 7 von den Erwartungswerten schon optisch nur wenig ab; ein χ^2-Test (bei dem die Klassen mit $X \ge 2$ zusammenzulegen sind) bestätigt diesen Eindruck mit dem bei $f = 2$ Freiheitsgraden unverdächtigen Wert von 2,16.

Dass bei der Ziehung am 11. Februar 2015 von einer „Kuriosität“ keine Rede sein kann, lässt sich auch ganz elementar begründen: Da nach den voraufgegangenen 24 Vierlingsziehungen bereits 12 der 46 möglichen Kombinationen genau einmal aufgetreten waren und 6 sogar zweimal, betrug die Wahrscheinlichkeit für eine Wiederholung beim 25. Auftreten eines Vierlings $\frac{18}{46} \approx 0{,}39$ (also kaum weniger als die Wahrscheinlichkeit, einen *neuen* Vierling zu ziehen). An dem Wert 0,39 ändert sich übrigens nichts vor der (noch ausstehenden) 26. Vierlingsziehung, da sich lediglich die Aufteilung der 18 bereits gezogenen Vierlinge von 12 + 6 auf 11 + 7 geändert hat.

4 Ein Ausreißer

Schon in Abbildung 1 war auffällig, dass die Trefferanzahl t für längere Zeiträume um mehr als 2σ über dem in n Ziehungen zu erwartenden Wert lag. Dies beruht vor allem auf der ungewöhnlichen Häufung von Vierlingen im Mittwochslotto: Statt der in 2257 Ziehungen zu erwartenden 6,7 Vierlinge traten 15 auf, (in 3099 Samstags-Ziehungen waren es nur 10 bei erwarteten 9,2). Wir untersuchen daher nun die beiden Ziehungsarten getrennt und wählen auch eine etwas andere graphische Darstellung zur Veranschaulichung der Daten: Statt der Trefferanzahl T nach n Ziehungen betrachten wir deren standardisierte Abweichung von ihrem Erwartungswert

$$Z := \frac{T - E(T)}{\sqrt{V(T)}}\,.$$

Dazu ein Beispiel: In den ersten 1009 Ziehungen im Mittwochslotto traten $t = 8$ Vierlinge auf, deutlich mehr als der Erwartungswert

$$E(T) = 1009 \cdot p_1 \approx 3{,}003\,.$$

Für unser Beispiel ergibt sich also

$$z \approx \frac{8 - 3{,}003}{\sqrt{3{,}003 \cdot (1 - p_1)}} \approx 2{,}887\,.$$

Der beobachtete Wert von T lag demnach zu diesem Zeitpunkt (Januar 1996) um fast drei Standardabweichungen über dem Erwartungswert. Die Graphen von $n \to z$ bis Ende Februar 2015, getrennt nach Samstags- und Mittwochslotto, zeigt Abbildung 2.

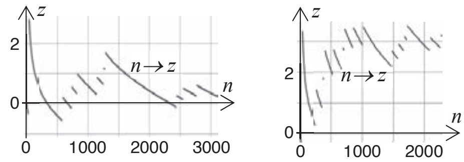

Abb. 2: Abweichung zwischen der Anzahl aufgetretener und erwarteter Vierlinge (in Vielfachen von σ) links: Samstagslotto rechts: Mittwochslotto

Typisch sind die Sprünge bei jedem neuen Vierling, so z. B. in der 1010. Ziehung beim Mittwochslotto. $E(T)$ und $V(T)$ ändern sich nur minimal gegenüber dem obigen Beispiel (3,006 statt 3,003 bzw. 2,997 statt 2,994), da aber der Minuend im Zähler von Z nun von 8 auf 9 springt, wächst auch z sprunghaft von 2,887 auf 3,466 an.

Während beim Samstagslotto die Abweichung vom Erwartungswert ab $n = 58$ durchgehend unter 2σ liegt, meist sogar unter 1σ, beträgt sie beim Mittwochslotto – anders als zu erwarten – ab $n = 531$ stets mehr als 2σ, über längere Zeiträume sogar mehr als 3σ.

Das Auftreten vieler Vierlinge bedeutet auch, dass es viele kurze und wenig lange Wartezeiten gegeben hat. Um dies genauer zu untersuchen, fassen wir die möglichen Werte der Wartezeiten zu 5 Klassen zusammen, die jeweils etwa die Wahrscheinlichkeit 0,2 haben. Die Klassengrenzen sind also dorthin zu legen, wo der Wert der Verteilungsfunktion F von W möglichst nahe bei einem Vielfachen von 0,2 liegt.[8] Abbildung 3 veranschaulicht den Sachverhalt für die erste Klassengrenze (das 0,2-Quantil).

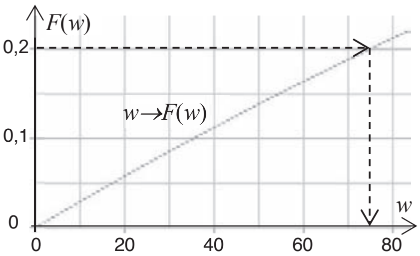

Abb. 3: Bestimmung der ersten Klassengrenze

Die zugehörige Berechnung liefert

$$\begin{aligned} F(w) \approx 0{,}2 &\Leftrightarrow 1 - F(w) \approx 0{,}8 \\ &\Leftrightarrow (1 - p_1)^w \approx 0{,}8 \\ &\Leftrightarrow w \approx 74{,}85\,. \end{aligned}$$

Wegen $w \in \mathbb{N}$ wird dieser Wert auf 75 gerundet. Wie Abbildung 2 (links) zeigt, fielen beim Samstagslotto die Wartezeiten auf V_1 und V_7 ($w_1 = 35$ und $w_7 = 2$) in die Klasse der kürzesten Wartezeiten.

Entsprechend findet man als weitere Klassengrenzen 171, 307 und 540, wobei in die zweite Klasse $w_3 = 154$, $w_4 = 87$, $w_6 = 107$ und $w_9 = 105$ fallen, in die mittlere $w_{10} = 215$, in die vierte $w_5 = 322$ und in die letzte $w_2 = 548$ und $w_8 = 1194$.

Da die Klassen praktisch gleichwahrscheinlich sind, ist der Erwartungswert für die Klassenhäufigkeiten bei den 10 Vierlingsziehungen an Samstagen stets 2 und bei den 15 Mittwochsziehungen jeweils 3. In den Abbildungen 4 und 5 sind diese den tatsächlich aufgetretenen absoluten Häufigkeiten der fünf Wartezeitkategorien gegenüber gestellt.

Beim Samstagslotto ist eine geringfügige Verschiebung in Richtung kleiner Wartezeiten zu erkennen, dem entspricht auch die durchschnittliche Wartezeit von 276,9 (etwas weniger als der Erwartungswert 336 der geometrischen Verteilung mit $p_1 = \frac{1}{336}$).

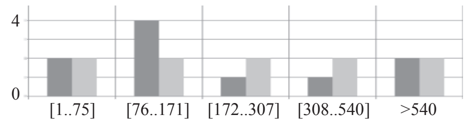

Abb. 4: Geringe Abweichung von einer Gleichverteilung der Wartezeitklassen beim Samstagslotto

Eine wesentlich deutlichere Asymmetrie zeigt sich beim Mittwochslotto, wo die längste Wartezeit nur 464 betrug und die mittlere Wartezeit mit 150,4 Ziehungen kleiner als die Hälfte des Erwartungswerts 336 war.

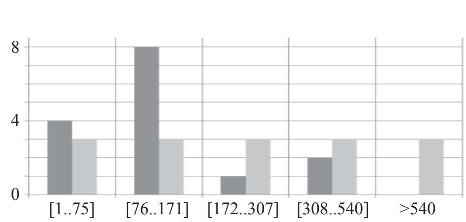

Abb. 5: Starke Abweichung von einer Gleichverteilung der Wartezeitklassen beim Mittwochslotto

Überprüft man die Häufigkeitsverteilungen mithilfe des χ^2-Tests auf die Abweichung von der erwarteten Gleichverteilung, so erhält man im ersten Fall den bei $f = 4$ Freiheitsgraden „normalen“ Wert 3, im zweiten dagegen 13,33, was eine hochsignifikante Abweichung bedeutet ($\alpha < 0{,}01$).

5 Schlussbemerkungen

Wir haben in den Abschnitten 2 und 3 einige naive Vorstellungen zum Thema Zufälligkeit mit Mitteln der Stochastik hinterfragt und widerlegt. Diese Vorgehensweise darf aber nicht dazu verführen, signifikante Ergebnisse wie das am Ende von Abschnitt 4 gewonnene absolut zu setzen. Ohne eine begründete Gegenhypothese zur Nullhypothese „Gleichverteilung" hat die festgestellte Signifikanz keinen Wert; es besteht überhaupt keine Veranlassung, bei der Ziehung der Lottozahlen eine Bevorzugung von Vierlingen anzunehmen.

Im Übrigen stehen der uns aufgefallenen Erscheinung in der Gesamtheit aller Ziehungen mit einem der fünf Mehrlingstypen in beiden Lottoversionen neun gegenüber, die keine signifikante Abweichung aufweisen. Dass unter den zehn Fällen einer mit der sehr kleinen Wahrscheinlichkeit 0,01 ist, hat immerhin eine Wahrscheinlichkeit von fast 0,1 und kann daher als „rein zufällig" beurteilt werden.

Gerade die letzten Gedanken lassen die hier behandelten Fragestellungen wertvoll für den Unterricht erscheinen. Darüber hinaus sei auf die Verwendung des Binomialmodells in verschiedenen Zusammenhängen hingewiesen: T = „Anzahl der Vierlinge in n_1 Lottoziehungen", K = „Anzahl kurzer Wartezeiten in n_2 Vierlingsziehungen" und X = „Anzahl eines bestimmen Vierlings in n_3 Vierlingsziehungen" sind jeweils binomialverteilt mit den oben genannten Trefferwahrscheinlichkeiten p_1, p_2 bzw. p_3 (mit der in Anmerkung 6 genannten Einschränkung).

Abschließend sei auf ergänzende Untersuchungen hingewiesen, die mit dem Einsatz von Simulationen möglich werden. Die von uns benutzte Annahme der Binomialverteilung für X ist dann überflüssig, die unterschiedlichen Wahrscheinlichkeiten für innere und Randvierlinge sind in einer Simulation ohne Probleme zu realisieren. Noch wertvoller ist, dass statt der relativ wenigen realen Daten umfangreiches Material erzeugt werden kann.

Anmerkungen

1 In Lottostatistik (2015) werden die beiden bisher aufgetretenen Fünflinge auch je zweimal bei den Vierlingen mitgezählt (die am 30. Juli 2014 gezogenen Reihe 9-10-11-12-13 also einmal als V_9 und einmal als V_{10}). Wir zählen diese unechten Vierlinge nicht mit.

2 Auf dasselbe Ergebnis führt eine andere Herleitung, bei der man die Fälle „Vierling mit Zwilling" und „Vierling mit 2 isolierten Zahlen" unterscheidet (Riehl 2013, S. 27, Tab. 3, Zeile 7 und 9). Bemerkenswerterweise gibt es auch hier in dem einen Fall 1892 und im anderen 39732 Möglichkeiten.

3 Die Anzahl der bisher aufgetretenen Vierlinge ist nicht 26, wie in HAZ (2015) angegeben, sondern 25, davon 10 im Samstags- und 15 im Mittwochslotto.

4 In Abbildung 1 ist noch eine weitere recht kurze Stufe zu erkennen (zwischen $t = 10$ und $t = 11$); dort war $W = 23$, also erheblich größer als zwischen $t = 6$ und $t = 7$; der Unterschied wird in der groben Graphik aber kaum erkennbar.

5 Im Internet findet man eine komfortable Möglichkeit für die Suche nach früher bereits gezogenen Zahlenkombinationen (Lottozahlensuche 2015); dort sind die Gewinnzahlen für das Mittwochslotto allerdings erst vom Jahr 2000 an erfasst.

6 Wie aus der Berechnung von p_1 in Abschnitt 2 hervorgeht, haben die beiden Randvierlinge V_1 und V_{46} mit $\frac{946}{41624} \approx 0{,}02273$ eine etwas größere Wahrscheinlichkeit als die übrigen mit $\frac{903}{41624} \approx 0{,}02168$. Da die beiden Werte nur wenig von $\frac{1}{46} \approx 0{,}02174$ abweichen, kann man X als angenähert binomialverteilt ansehen.

7 In 15 Mittwochs-Ziehungen kamen V_4, V_8, V_{10}, V_{11}, V_{13}, V_{14}, V_{25}, V_{30} und V_{46} je einmal, V_{18}, V_{26} und V_{40} je zweimal vor; in 10 Samstags-Ziehungen traten V_{10}, V_{22}, V_{25}, V_{33}, V_{35}, V_{36}, V_{38} und V_{46} je einmal auf, V_{16} zweimal (Lottostatistik 2015, ohne unechte Vierlinge, vgl. Anm. 1). V_{10}, V_{25} und V_{46} traten sowohl im Samstags- als auch im Mittwochslotto auf, so dass zusammen 7 Vierlinge doppelt vorkamen.

8 Da F – wie in Abbildung 3 dargestellt – eine Treppenfunktion ist, werden die Werte $k \cdot 0{,}2$ im Allgemeinen nicht angenommen. Dadurch weichen die Wahrscheinlichkeiten der Klassen geringfügig von 0,2 ab (0,2003; 0,1990; 0,2002; 0,2005; 0,1999).

Literatur

HAZ (2015): Hannoversche Allgemeine Zeitung Nr. 37 vom 13. Februar 2015.

Lottostatistik (2015): www.mylottoy.net [→ Deutschland → Statistiken → fortlaufend] (Zugriff: 2.3.2015).

Lottozahlen-Suche (2015): www.lottozahlenonline.de [→ Suche nach Lottozahlen] (Zugriff: 2.3.2015).

Riehl, G. (2013): Eine neue Modellierung für benachbarte Zahlen beim Lotto. *Stochastik in der Schule* 33 (3), 23–27.

Anschrift des Verfassers

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