Wie lange dauert es (im Mittel), bis ...?

Teilnehmer:

Valentin Bonjer	Käthe-Kollwitz-Gymnasium
Thomas Dittmar	Heinrich-Hertz-Gymnasium
Henriette Kirsten	Heinrich-Hertz-Gymnasium
Max Lindner	Käthe-Kollwitz-Gymnasium
Anton Pusch	Käthe-Kollwitz-Gymnasium
Fabian Schiemann	Andreas-Gymnasium
Maximilian Stepper	Herder-Gymnasium
Alexeij Ward	Herder-Gymnasium
Alma Wettig	Andreas-Gymnasium

mit tatkräftiger Unterstützung durch:

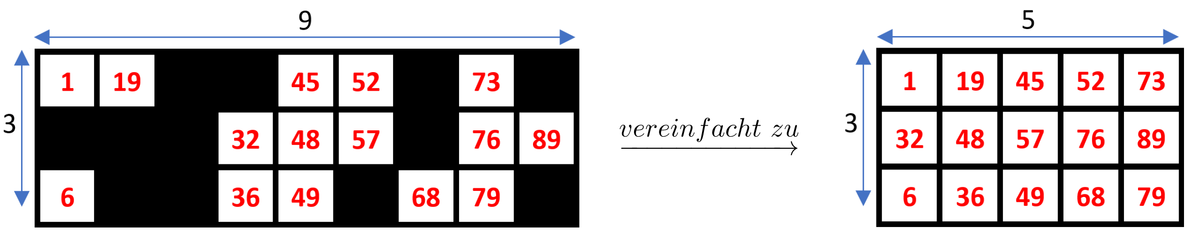
Frank Schiller	Käthe-Kollwitz-Gymnasium

Gruppenleiter:

Elke Warmuth	Humboldt-Universität zu Berlin

1. BINGO im Spiegel der Mathematik

Seit 1929 begeistert uns das mathematisch höchst interessante Glücksspiel BINGO. Zunächst füllen alle Spieler einen BINGO-Schein aus. In der europäischen Variante besteht dieser Schein aus $3 \cdot 9$ Feldern mit paarweise verschiedenen Zahlen von 1 bis 90. In jeder Zeile sind fünf gefüllte und vier leere Kästchen:



Regeln

Nacheinander werden zufällig unabhängig alle 90 Kugeln ohne Zurücklegen mit gleicher Wahrscheinlichkeit gezogen. Es gibt die Gewinnvarianten 1-, 2-, 3-er BINGO. Wir betrachten hier aber nur das *3-er BINGO*. Wer zuerst alle 15 Zahlen auf seinem BINGO-Schein abhaken kann, gewinnt.

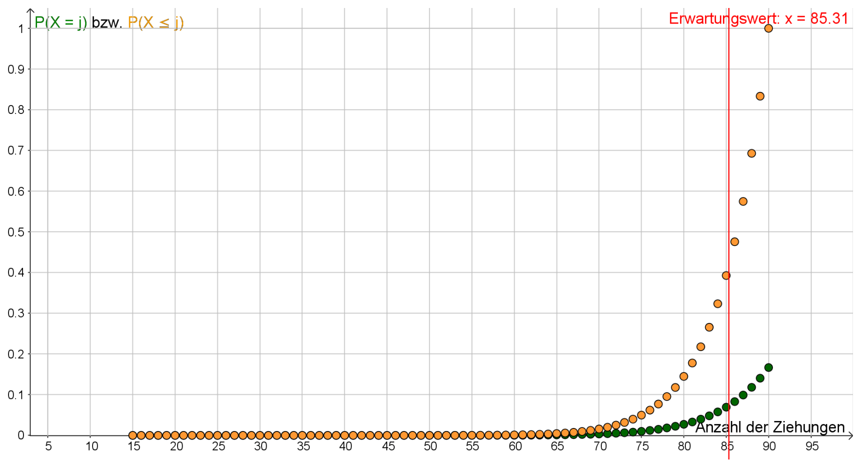
Übersetzung in die Stochastik

Es sei X die Anzahl der Ziehungen bis zum 3-er Bingo für einen festen Spieler. Es gilt $15 \leq X \leq 90$, da der Bingoschein 15 Zahlen hat und es 90 Kugeln gibt. Hierbei gilt:

$$I) \quad P(X \leq j) = \frac{\binom{75}{j-15}}{\binom{90}{j}} = f(j).$$

Wir färben die Kugeln in rote (die Kugeln mit den Zahlen auf dem Schein) und schwarze (die restlichen). Unter den j Kugeln befinden sich alle 15 roten und $j - 15$ der 75 schwarzen Kugeln. Um die Gewinnwahrscheinlichkeit für die j-te Ziehung zu betrachten, subtrahieren wir alle vorhergehenden Ziehungen:

$$II) \quad P(X = j) = f(j) - f(j-1) = \frac{\binom{75}{j-15}}{\binom{90}{j}} - \frac{\binom{75}{j-16}}{\binom{90}{j-1}}.$$



Entwicklung der Wahrscheinlichkeit für *3-er BINGO*

Mit steigendem (j) steigt auch die Gewinnwahrscheinlichkeit für die j-te Ziehung.

Hätten Sie gedacht, dass erst bei der letzten Kugel die Wahrscheinlichkeit für einen 3-er BINGO am größten $\left(\frac{1}{6}\right)$ *ist?*

1.1. Uminterpretation als BINGO-Lotto-Modell

Wir stellen das Zufallsexperiment als 90-Tupel dar:

$$\underbrace{\underbrace{c_1\, c_2\, c_3\, c_4\, c_5\, c_6\, c_7\, c_8\, c_9\, c_{10} \cdots}_{j-1 \text{ Kugeln mit 14 roten}}\, c_j}_{j \text{ Kugeln mit 15 roten}} \cdots c_{90}$$

15 richtige rote Kugeln
75 falsche schwarze Kugeln
c_j: letzte Zahl des Scheins, damit man *3-er BINGO* erzielt

Hierbei lassen sich rote bzw. schwarze Kugeln unterscheiden. Der Zahlenwert der Kugel ist irrelevant. Vor der Ziehung von c_j befinden sich $j-1$ Kugeln. Die Positionen der ersten 14 roten Kugeln sind aus einer Menge von $j-1$ Elementen ausgewählt. Für das Tupel gibt es $\binom{90}{15}$ Möglichkeiten. Äquivalent zu II) gilt:

$$III) \quad P(X=j) = \frac{\binom{j-1}{14}}{\binom{90}{15}} = f(j) - f(j-1).$$

Dies lässt sich verallgemeinern zu:

$$IV) \quad P(X=j) = \frac{\binom{j-1}{r-1}}{\binom{s}{r}} \qquad r\text{: Zahlen auf dem Schein} \qquad s\text{: Kugeln insgesamt}$$

1.2. Erwartungswert

Des Weiteren interessiert uns, wie viele Ziehungen ein einzelner Spieler im Mittel abwarten muss, bis dieser Spieler *3-er BINGO* rufen kann.

$$\begin{aligned}
E(X) &= \sum_{j=r}^{s} j \cdot P(X=j) && IV) \text{ einsetzen} \\
&= \sum_{j=r}^{s} j \cdot \frac{\binom{j-1}{r-1}}{\binom{s}{r}} && \binom{s}{r}^{-1} \text{ ausklammern} \\
&= \frac{1}{\binom{s}{r}} \cdot \sum_{j=r}^{s} j \cdot \binom{j-1}{r-1} && \binom{j-1}{r-1} \text{ mit } r \text{ erweitern} \\
&= \frac{1}{\binom{s}{r}} \cdot \sum_{j=r}^{s} r \cdot \binom{j}{r} && r \text{ ausklammern} \\
&= \frac{r}{\binom{s}{r}} \cdot \sum_{j=r}^{s} \cdot \binom{j}{r}.
\end{aligned}$$

Summation über alle Wahrscheinlichkeiten und Indexverschiebung ergibt[1]:

$$\begin{aligned}
E(X) &= \frac{r}{\binom{s}{r}} \cdot \binom{s+1}{r+1} && \text{kürzen} \\
&= \frac{r \cdot (s+1)}{r+1}.
\end{aligned}$$

[1]vgl. N. Henze, H. Humenberger: Stochastische Überraschungen beim Spiel BINGO. In: Stochastik in der Schule 31 (2011) 3, S. 8

2. Das Sammlerproblem

Seit jeher übt das Komplettieren einer Sammlung eine Faszination sowohl auf Sammler als auch auf Mathematiker aus. Dabei interessiert meist, wie viel Zeit dieser Vorgang in Anspruch nimmt.

Abstrahiert kann man schon das wiederholte Rollen eines Würfels als einen solchen Sachverhalt ansehen. Ziel ist es hierbei jede Augenzahl mindestens einmal zu erwürfeln. Dabei ist ein Wurf ein Erfolg, wenn wir eine noch nicht erzielte Zahl würfeln. Mit der Wahrscheinlichkeit $p_1 = 1-(1-1)\cdot p$ ist es im ersten Versuch sicher, einen Erfolg zu erzielen, im zweiten ist die Wahrscheinlichkeit dann $p_2 = 1-(2-1)\cdot p$. Wenn es schon $i-1$ Erfolge gab, ist die Wahrscheinlichkeit für den i-ten Erfolg im nächsten Wurf $p_i = 1-(i-1)p$. Hierbei ist $p = \frac{1}{6}$.

Es sei X_i die Anzahl der Würfe nach der $i-1$-ten bis zur i-ten verschiedenen Augenzahl und X die Anzahl der Würfe bis zur vollständigen Serie. Dann ist $X = X_1 + X_2 + \cdots + X_6$. Für den Erwartungswert von X_i haben wir gezeigt:

$$E(X_i) = \frac{1}{p_i}.$$

Da wir den Erwartungswert der Anzahl der Versuche X wissen wollen, muss man die Erwartungswerte der einzelnen Würfe dafür, dass man eine noch nicht geworfene Zahl erzielt, aufsummieren:

$$E(X) = \sum_{i=1}^{6} \frac{1}{1-(i-1)\cdot p} = \frac{1}{\frac{6}{6}} + \frac{1}{\frac{5}{6}} + \frac{1}{\frac{4}{6}} + \frac{1}{\frac{3}{6}} + \frac{1}{\frac{2}{6}} + \frac{1}{\frac{1}{6}} = 14,7.$$

Daran erkennt man, dass man nach durchschnittlich 15 Würfen also alle sechs Augenzahlen erwürfelt hat.
Diese Formel kann man verallgemeinern. Mit n als Gesamtanzahl der verschiedenen Elemente unserer Reihe und p als Wahrscheinlichkeit jedes einzelnen Elements (mit $p = \frac{1}{n}$, d. h., dass wir im Folgenden immer von einem Laplace-Experiment ausgehen) sieht die Formel wie folgt aus:

$$E(X) = \sum_{i=1}^{n} \frac{1}{1-(i-1)\cdot p} = \frac{1}{\frac{n}{n}} + \frac{1}{\frac{n-1}{n}} + \cdots + \frac{1}{\frac{1}{n}} = n\left[\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{2} + \frac{1}{1}\right].$$

Diese Formel funktioniert allerdings nur, wenn $s = 1$ ist. Im allgemeinen Fall werden bei jeder Ziehung s verschiedene Elemente auf gut Glück aus einer n-elementigen Menge gezogen.

Für ein beliebiges s sei X_n die Anzahl der Versuche bis zur vollständigen Serie. Wir benötigen folgende Formeln:

1.) für die Einzelwahrscheinlichkeit, dass beim k-ten Versuch die Serie komplettiert ist:

$$P(X_n = k) = \sum_{r=1}^{n-s} (-1)^{r-1} \binom{n}{r} q_r^{k-1}(1-q_r),\ k > a-1,$$

2.) für den Erwartungswert:

$$E(X_n) = \sum_{r=1}^{n-s} (-1)^{r-1} \binom{n}{r} \frac{q_r^{a-1}(q_r - a(q_r-1))}{1-q_r}.$$

Hierbei ist a die Mindestanzahl an Versuchen, die es braucht, bis die Serie komplettiert ist ($\frac{n}{s}$ aufgerundet liefert a).

Um auf diese Formeln zu kommen überlegen wir, dass X_n die maximale Wartedauer aller zu ziehenden Elemente ist. Der Wertebereich geht dabei von a bis ∞.

Es sei $k > 0$ fest. Mit A_j bezeichnen wir das Ereignis, dass die Wartedauer auf das Element j größer als k ist.
Das Ereignis $\{X_n > k\}$ tritt genau dann ein, wenn mindestens eins der Ereignisse A_j eintritt. Also gilt:

$$P(X_n > k) = P\left(\bigcup_{j=1}^{n} A_j\right).$$

Wir wenden die Formel des Ein- und Ausschließens an und erhalten:

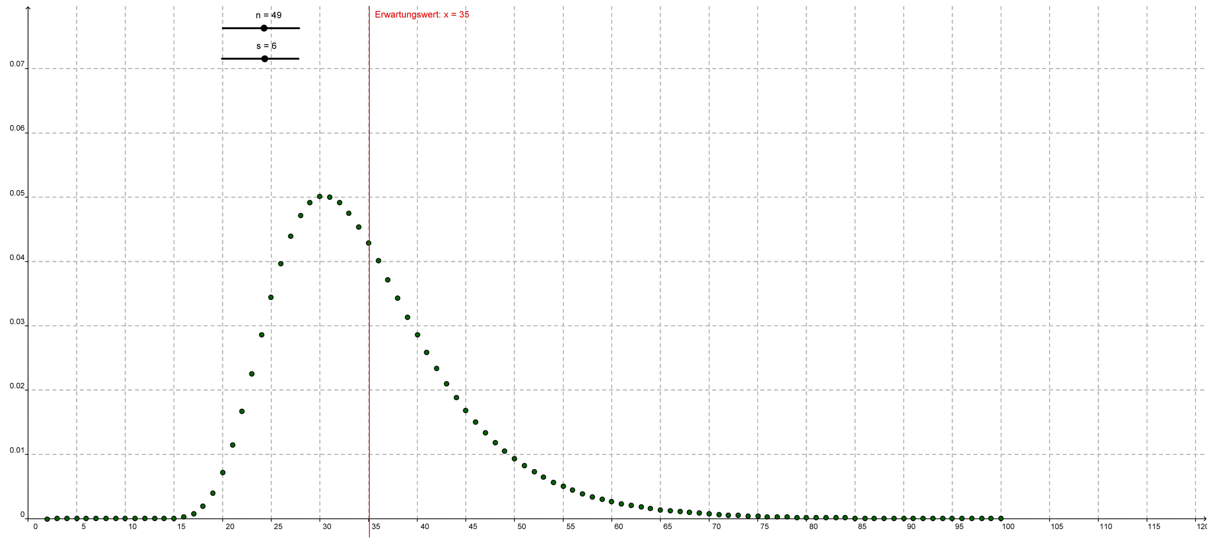
$$P(X_n > k) = \sum_{r=1}^{n} (-1)^{r-1} \binom{n}{r} q_r^k.$$

Hierbei ist $q_r = \frac{\binom{n-r}{s}}{\binom{n}{s}}$ und gibt die Wahrscheinlichkeit an, dass die zu ziehenden s Elemente aus der betrachteten $(n-r)$-elementigen Menge genommen werden.

Wegen $P(X_n > k-1) = P(X_n > k) + P(X_n = k)$ ergibt sich nun Formel 1.) für die Einzelwahrscheinlichkeit. Um auf den Erwartungswert zu kommen nutzen wir die Ableitung, wie im folgenden Beispiel dargestellt, und nach Umstellen erhalten wir Formel 2.).

$$\text{Beispiel}: \sum_{k=1}^{\infty} k x^{k-1} = \frac{d}{dx}\left(\sum_{k=1}^{\infty} x^k\right) = \frac{d}{dx}\left(\frac{x}{1-x}\right) = \frac{1}{(1-x)^2}, |x| < 1.$$

Wir sind jetzt mit der Formel 2.) im Stande zu ermitteln, wann im Mittel alle Lottozahlen mindestens ein Mal gezogen wurden. Folgende Grafik veranschaulicht dies:



Hier erkennt man, dass nach durchschnittlich 35 Ziehungen alle Lottozahlen mindestens ein Mal gezogen wurden. Die Punkte geben die Einzelwahrscheinlichkeiten zur Komplettierung der Serie für jede Ziehung an.
Diese Formeln kann man auf alle Sachverhalte dieser Art übertragen.

3. Muster in Bernoulli-Ketten

Gesucht wird der Erwartungswert der Wartezeit X auf ein beliebiges gegebenes Muster in einer Bernoulli-Kette. Muster meint hierbei eine bestimmte, vorher festgelegte Abfolge von Erfolgen und Misserfolgen.

Beschrieben wird der Vorgang durch *Markow-Ketten erster Ordnung.* Eine homogene Markow-Kette mit dem Zustandsraum $S = \{1, 2, \dots\}$ ist ein Zufallsprozess in diskreter Zeit, der zu jedem der Zeitpunkte $1, 2, \dots$ seinen momentanen Zustand wechseln kann. Mit der Übergangswahrscheinlichkeit $p_{ij}, i, j \in E$ geht er vom Zustand i in den Zustand j über. Diese Wahrscheinlichkeit hängt nicht von den vorherigen Zuständen, sondern nur vom aktuellen Zustand ab (Markow-Eigenschaft). Markow-Ketten können sogenannte absorbierende Zustände besitzen, das sind Zustände, die nicht mehr verlassen werden können. Dieser absorbierende Zustand ist bei uns das Erreichen des festgelegten Musters.

Wir nutzen hierzu die

$$\textit{zweite Mittelwertsregel}: E_i(X) = 1 + \sum_{j=1}^{n} p_{ij} E_j(X).$$

Bei der zweiten Mittelwertsregel geht man davon aus, dass der Zustand i nicht der absorbierende Zustand ist, man braucht also mindestens noch einen Schritt zum gewünschten Muster. Die Summe gibt – gewichtet nach der Wahrscheinlichkeit der nächsten Zustände – den (bedingten) Erwartungswert von X ausgehend vom Zustand i an.

Zur praktischen Demonstration sollen die Erwartungswerte der Wartezeit auf folgende drei Muster eines idealen Münzwurfes ermittelt und miteinander verglichen werden: 101010; 110011; 111111.

Am Beispiel des Musters 101 soll die Berechnung von $E(X)$ über ein lineares Gleichungssystem verdeutlicht werden. Der Graph der Markow-Kette für dieses Beispiel sieht folgendermaßen aus

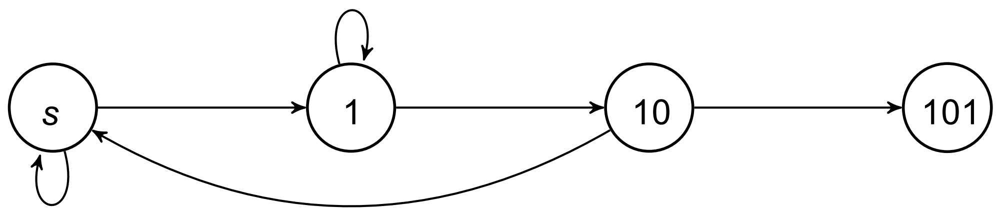


Figure 1: Markow-Kette Erstellt von Jonas Wanke

Hierbei ist s der „Start"-Zustand. In diesem Zustand befindet sich der Prozess zu Beginn und wenn nach 10 in der Bernoulli-Kette eine 0 geworfen wird, weil diese das Muster „zerstört". Die weiteren Zustände sind 1, 10 und 101. Da es sich um eine ideale Münze handelt, sind alle Übergangswahrscheinlichkeiten gleich $\frac{1}{2}$. Nach der zweiten Mittelwertsregel gilt mit der abkürzenden Bezeichnung $E_s(X) = E(X)$

$$E(X) = 1 + \frac{1}{2}E_1(X) + \frac{1}{2}E(X) \quad (1)$$

$$E_1(X) = 1 + \frac{1}{2}E_1(X) + \frac{1}{2}E_{10}(X) \quad (2)$$

$$E_{10}(X) = 1 + \frac{1}{2}E(X) + \frac{1}{2}E_{101}(X) \quad (3)$$

Es gilt $E_{101}(X) = 0$, da 101 der absorbierende Zustand ist. Wir setzen $E_{10}(X)$ in (2) ein und erhalten:

$$\frac{1}{2}E_0(X) = \frac{3}{2} + \frac{1}{4}E(X).$$

Nach Umformung ergibt sich das folgende lineare Gleichungssystem in $E(X)$ und $E_1(X)$

$$\begin{aligned} E(X) - E_1(X) &= 2 \\ -E(X) + 2E_1(X) &= 6, \end{aligned}$$

das für $E(X)$ die eindeutige Lösung $E(X) = 10$ hat.

Analog – aber umfangreicher – lassen sich für das oben genannte Beispiel die einzelnen Erwartungswerte ausrechnen.

$101010 : E(X) = 84$,

$110011 : E(X) = 70$,

$111111 : E(X) = 126$.

Anmerkung: Alle Ergebnisse entstanden unter der Annahme eines idealen Münzwurfes, bei dem Kopf und Zahl bzw. Erfolg und Misserfolg gleichwahrscheinlich sind.

Rekurrente Ereignisse

Um die mittlere Wartezeit auf andere Typen von Mustern zu bestimmen, haben wir uns zunächst mit rekurrenten Ereignissen in einem Zufallsprozess beschäftigt.

Definition. Die Ereignisse $H_1, H_2, \dots$ heißen rekurrent, wenn für $i, j \in \mathbb{N}$ mit $i > j$ gilt, dass

$$P(H_i | \overline{H_1} \cap \dots \cap \overline{H_{j-1}} \cap H_j) = P(H_{i-j}).$$

Beispiel Random Walk. Beim Random Walk wird ein Partikel im eindimensionalen Raum betrachtet, das sich bei jedem Zeitschritt vom Punkt i aus mit der Wahrscheinlichkeit $\frac{1}{2}$ entweder zum Punkt $i + 1$ oder zum Punkt $i - 1$ bewegt.
Für $k = 1, 2, \dots$ bezeichnen wir mit H_k das Ereignis, dass das Partikel im k-ten Schritt den Punkt 0 erreicht. Die oben genannte Formel heißt für den Random Walk Folgendes: Sei $i = 8$ und $j = 6$. Die Bedingung auf der linken Seite bedeutet, dass das Partikel erstmalig beim sechsten Schritt wieder den Punkt 0 erreicht. Da die Ereignisse $H_1, H_2, \dots$ rekurrrent sind, ist die Wahrscheinlichkeit, bei Schritt 8 wieder den Nullpunkt zu erreichen, unabhängig von den Schritten **vor** dem ersten Erreichen des Nullpunktes. Wir können den Prozess bei Schritt 6 als wieder neu gestartet annehmen. Die Wahrscheinlichkeit, dass das Partikel den Nullpunkt erreicht, ist also bei Schritt 8 und 2 gleich, wenn es bei Schritt 6 den Punkt 0 erreicht hatte.

Um rekurrente Ereignisse beschreiben zu können, gibt es erzeugende Funktionen *H(s)* für die Auftrittswahrscheinlichkeiten und *T(s)* für die Wartezeit. Zwischen diesen zwei erzeugenden Funktionen besteht folgender Zusammenhang:

$$H(s) = \frac{1}{1 - T(s)}.$$

Runs

Runs sind ein spezielle Art von Mustern in Bernoulli-Ketten, bei denen r Erfolge (oder auch Misserfolge) aufeinanderfolgen. Zur Berechnung des Erwartungswertes der Wartezeit nutzt uns die oben für das Muster 101 erläuterte Methode nichts, da wir $E(X)$ allgemein für beliebige r berechnen wollen.

Mit Hilfe von erzeugenden Funktionen für rekurrente Ereignisse lässt sich die Formel des Erwartungswertes bestimmen als

$$E(X) = \frac{1 - p^r}{qp^r},$$

wobei p die Wahrscheinlichkeit eines Erfolges, q die eines Misserfolges und r die Länge des Runs bezeichnet.