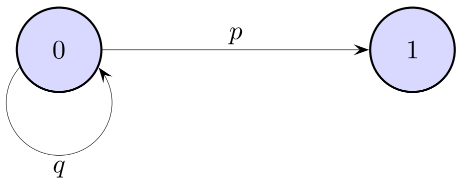
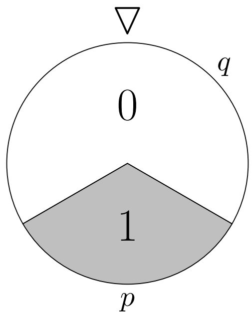
Vollständige Serie (Sammlerproblem)

Geometrische Verteilung

Geometrische Verteilung





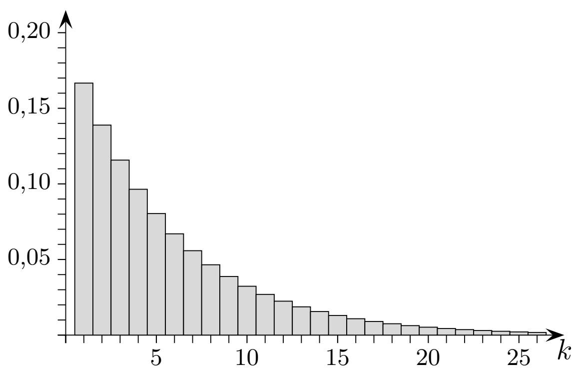
Ein Glücksrad wird solange gedreht bis zum 1. Mal eine 1 erscheint.
X sei die Anzahl der Übergänge bis zur Absorption.
Dem Graphen kann die Verteilung $p_k = q^{k-1}p,\ k = 1,\ 2,\ 3,\ \ldots$ unmittelbar abgelesen werden.

Wir ermitteln den Erwartungswert.

$$\begin{aligned}
E(X) &= 1p_1 + 2p_2 + 3p_3 + 4p_4 \ \ldots \\
&= 1p + 2pq + 3pq^2 + 4pq^3 \ldots \\
\\
&= p + pq + pq^2 + pq^3 \ \ldots \qquad (= P(X \geq 1) = P(\text{mindestens 1 Übergang}) = 1) \\
&\qquad + pq + pq^2 + pq^3 \ \ldots \qquad (= P(X \geq 2) = P(\text{mindestens 2 Übergänge}) = q,\ q \text{ ausklammern}) \\
&\qquad\qquad + pq^2 + pq^3 \ \ldots \qquad (= P(X \geq 3) = q^2,\ q^2 \text{ ausklammern}) \\
&\qquad\qquad\qquad + pq^3 \ \ldots \ = \ 1 + q + q^2 + q^3 + \ldots = \frac{1}{1-q} = \frac{1}{p}
\end{aligned}$$

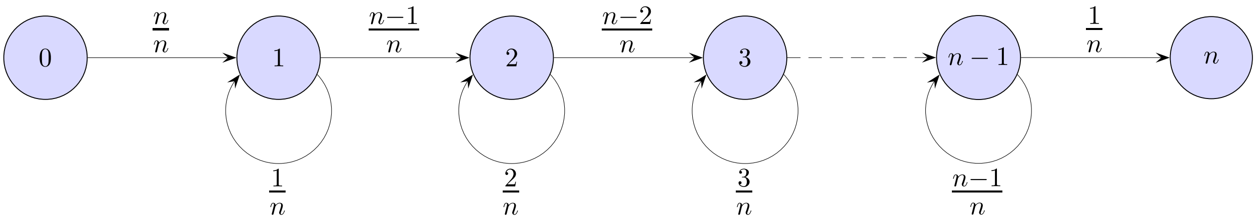
Das war auch zu erwarten.
Um z. B. eine 6 ($P = \frac{1}{6}$) zu würfeln, benötige ich im Schnitt 6 Würfe.

Warten auf eine 6: Wir würfeln solange, bis eine 6 erscheint.
X sei die Anzahl der benötigten Würfe, $P(X = k) = (1-p)^{k-1} \cdot p$
Z. B. ist die Wahrscheinlichkeit, mit höchstens 4 Würfen eine 6 zu erzielen, $P(X \leq 4) = 51{,}8\,\%$.



eleganter:
Sei m die mittlere Schrittzahl von 0 aus bis zur Absorption.
Dann gilt $m = p \cdot 1 + (1-p) \cdot (m+1)$, also $m = \frac{1}{p}$.

Wartezeit bis zu einem vollständigen Satz (Sammlerproblem)



Es geht um die Frage, wie lange es dauert, bis jedes der n möglichen Ergebnisse eines Zufallsversuchs mindestens einmal aufgetreten ist. Aus einer Urne mit n nummerierten Kugeln wird zufällig jeweils eine gezogen und wieder zurückgelegt. Dies wird solange wiederholt, bis jede Kugel mindestens einmal gezogen wurde. In der Grafik geben die Zustände 0 bis n die Anzahl der gesammelten Nummern an.
Sei X_i die Wartezeit, um vom Zustand $i-1$ nach i zu gelangen.
X_i ist geometrisch verteilt.
Die Übergangswahrscheinlichkeiten

$$p_i = \frac{n+1-i}{n}, \;\; i = 1, 2, 3, \ldots, n$$

verringern sich gemäß der abnehmenden Anzahl noch nicht gezogener Kugeln.
Im Zustand 2 z. B. muss eine von $n-2$ Kugeln gezogen werden, um in den Zustand 3 zu gelangen.
Im Schnitt werden hierzu $\frac{n}{n-2}$ Ziehungen (Kehrwert der Wahrscheinlichkeit, siehe 1. Seite) benötigt.

Die X_i sind unabhängige Zufallsvariable und die Summe

$$T_n = X_1 + X_2 + \ldots + X_n$$

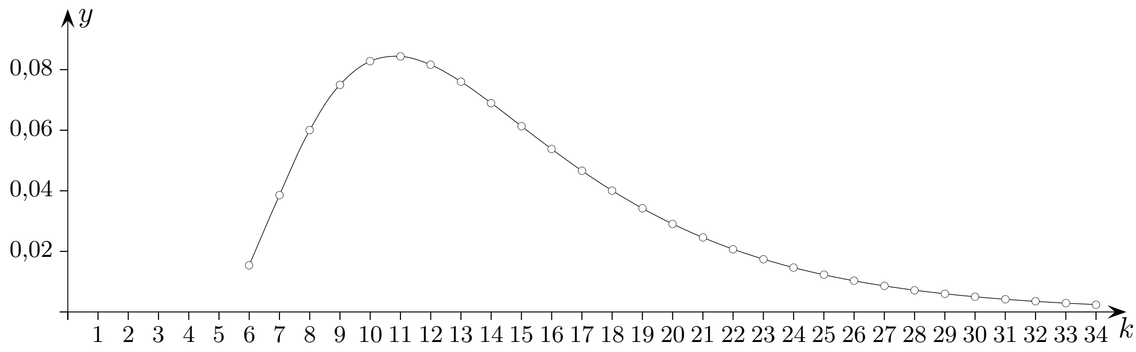
ist die Wartezeit für einen vollständigen Satz.

Der Erwartungswert errechnet sich somit:

$$\begin{aligned} E(T_n) &= \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + n \\ &= n\big(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}\big) \end{aligned}$$

Einen Würfel muss man im Schnitt 14,7 mal werfen, damit alle sechs Augenzahlen mindestens einmal auftreten.

Wahrscheinlichkeiten, eine vollständige Serie zu erhalten

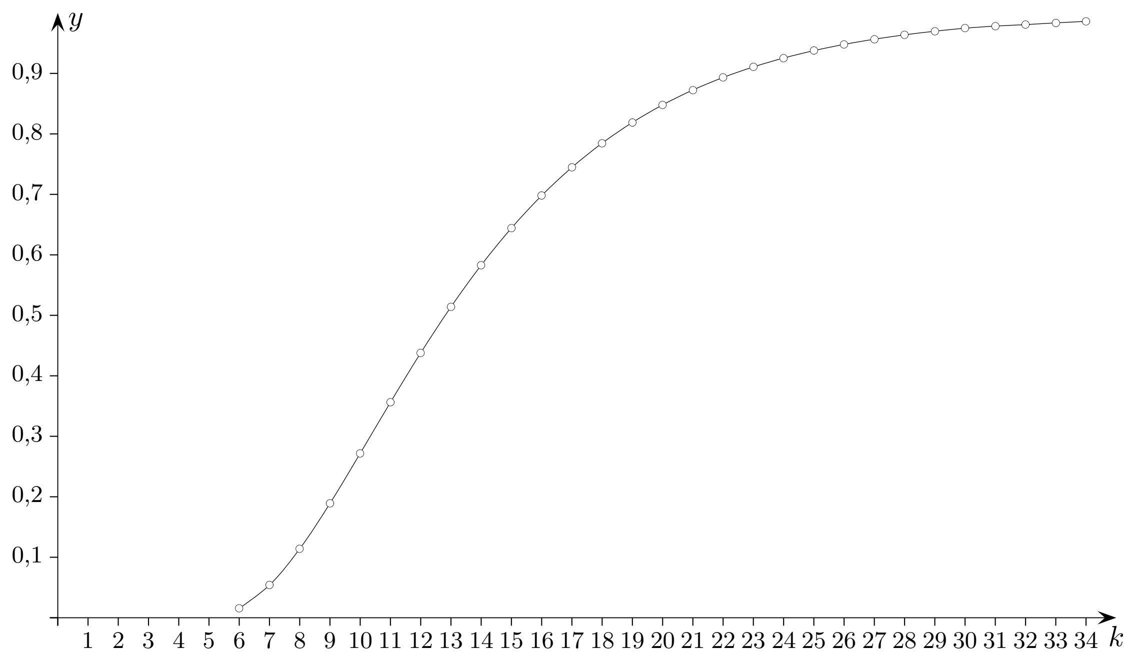


Die Wahrscheinlichkeiten, mit genau k Ziehungen eine vollständige Serie ($n = 6$) zu erhalten, sind hier grafisch dargestellt.

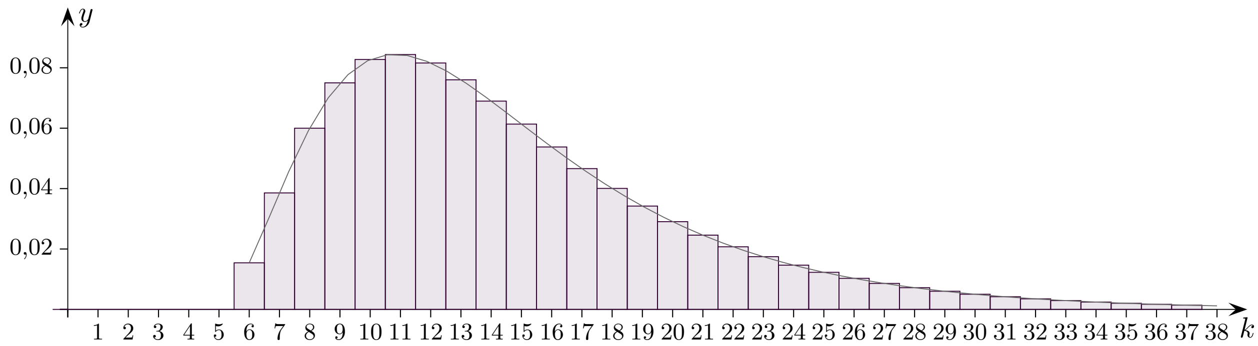
Ich habe die Wahrscheinlichkeiten aus der Taylorentwicklung der erzeugenden Funktion von T_6 ermittelt. Alternativ liefert die Anwendung der Siebformel:

$$P(T_n = k) = \sum_{r=1}^{n-1} (-1)^{r-1} \cdot \binom{n}{r} \cdot \left(1 - \frac{r}{n}\right)^{k-1} \cdot \frac{r}{n}$$

In der unteren Grafik sind die Wahrscheinlichkeiten dargestellt, mit höchstens k Ziehungen eine vollständige Serie zu erhalten (kumulierte Wahrscheinlichkeiten).

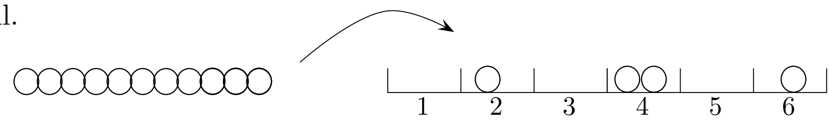


$P(T_6 = k)$ GeoGebra



Anwendung der Siebformel, um $P(T_n = k)$ zu ermitteln

Wir wechseln das Modell.



$m = 15$ Kugeln sollen auf $n = 6$ (z. B.) Fächer nacheinander zufällig verteilt werden.
Gegenüber dem Kugel-Fächer-Modell sind die Bezeichnungen m und n vertauscht.

Eine vollständige Serie liegt vor, wenn sich in jedem Fach mindestens eine Kugel befindet.
Sei ω eine Versuchsreihe der Länge m:

$$\omega = (\,2,\ 4,\ 2,\ 1,\ 3,\ 1,\ 3,\ 6,\ 4,\ 6,\ 3,\ 3,\ 5,\ 1,\ 2\,)$$

Für jedes Fach j wird eine Zufallsvariable W_j definiert, die die Anzahl der Versuche (die Wartezeit) angibt, bis das Fach j erstmalig eine Kugel erhält.

$$W_1(\omega) = 4$$

$$W_2(\omega) = 1$$

$$W_3(\omega) = 5$$

$$\dots$$

$$W_n(\omega) = 8$$

Die Zufallsvariable der Wartezeit auf eine vollständige Serie lässt sich in der Form

$$T_n = \max(W_1, W_2, \dots, W_n)$$

ausdrücken.

Den Schlüssel zur Bestimmung der Verteilung von T_n bildet die Gleichheit der Ereignisse:

$$\{T_n > k\} = \bigcup_{j=1}^{n} \{W_j > k\}$$

Die Menge $\{T_n > k\}$ beinhaltet alle Versuchsreihen ω, so dass nach k Versuchen noch keine vollständige Serie vorliegt. Für diese ω gibt es jeweils ein Fach, in dem nach k Versuchen noch keine Kugel liegt, d. h. $W_j(\omega) > k$, $j \in \{1, 2, \dots, n\}$.

Die Siebformel besagt nun in diesem Fall:

$$P\Big(\bigcup_{j=1}^{n} \{W_j > k\}\Big) = \sum_{r=1}^{n} (-1)^{r-1} \cdot S_r \qquad \text{mit} \quad S_r := \sum_{1 \le i_1 < \ldots < i_r \le n} P(\{W_{i_1} > k\} \cap \ldots \cap \{W_{i_r} > k\})$$

Die rechts stehenden Wahrscheinlichkeiten lassen sich leicht angeben, z. B. gilt:

$$P(\{W_1 > k\} \cap \{W_2 > k\} \cap \{W_3 > k\}) = \left(\frac{n-3}{n}\right)^k$$

Das Ereignis $\{W_1 > k\} \cap \{W_2 > k\} \cap \{W_3 > k\}$ tritt ein, wenn in den ersten k Versuchen keines der Fächer 1, 2, 3 besetzt wird, d. h. wenn bei jedem der ersten k Versuche jeweils die Nummern 4, 5, 6 ausgewählt werden und dies geschieht mit der angegebenen Wahrscheinlichkeit.

Anwendung der Siebformel, um $P(T_n = k)$ zu ermitteln

Die Wahrscheinlichkeiten in

$$S_r := \sum_{1\leq i_1<\ldots<i_r\leq n} P(\{W_{i_1} > k\} \cap \ldots \cap \{W_{i_r} > k\})$$

sind daher alle identisch $\left(\frac{n-r}{n}\right)^k$.

Mit der speziellen Siebformel

$$P\Big(\bigcup_{j=1}^{n} A_j\Big) = \sum_{r=1}^{n} (-1)^{r-1} \cdot \binom{n}{r} \cdot P(A_1 \cap A_2 \cap \ldots \cap A_r)$$

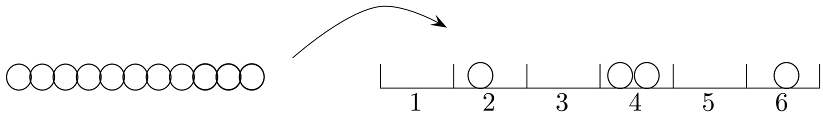
erhalten wir:

$$P(\{T_n > k\}) = \sum_{r=1}^{n} (-1)^{r-1} \cdot \binom{n}{r} \cdot \left(\frac{n-r}{n}\right)^k$$

Nun ist nur noch $P(T_n > k-1) = P(T_n > k) + P(T_n = k)$ (Schreibweise wie üblich vereinfacht)
umzustellen, das Bisherige zweimal anzuwenden,
auszuklammern und das Ersehnte liegt vor:

$$P(T_n = k) = \sum_{r=1}^{n-1} (-1)^{r-1} \cdot \binom{n}{r} \cdot \left(1 - \frac{r}{n}\right)^{k-1} \cdot \frac{r}{n}$$

Anwendung der Siebformel, leicht vereinfacht



m Kugeln sollen auf $n = 6$ (z. B.) Fächer nacheinander zufällig verteilt werden.

Wie groß ist die Wahrscheinlichkeit, dass mindestens ein Fach leer bleibt?

Wir betrachten alle Versuchsreihen der Länge m:

$$\omega = (\,2,\ 4,\ 2,\ 1,\ 3,\ 1,\ 3,\ 6,\ 4,\ 6,\ 3,\ 3,\ 4,\ 1,\ 2\,)$$

Für jedes Fach j definieren wir das Ereignis: A_j = Fach A_j bleibt leer. Dann ist

$$P(A_i) = (\tfrac{n-1}{n})^m$$

$$P(A_i \cap A_j) = (\tfrac{n-2}{n})^m$$

$$P(A_i \cap A_j \cap A_k) = (\tfrac{n-3}{n})^m$$

$$\ldots$$

$$P(A_i \cap A_j \cap A_k \cap \ldots \cap A_n) = 0$$

Mit der Siebformel

$$P(A_1 \cup A_2 \cup \,\ldots\, \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \,\ldots$$
$$(-1)^{n+1} \cdot P(A_1 \cap A_2 \cap \,\ldots\, \cap A_n)$$

erhalten wir:

$$P_m = \sum_{r=1}^{n} (-1)^{r-1} \cdot \binom{n}{r} \cdot (\tfrac{n-r}{n})^m$$

Z. B. für $n = 4$ Fächer

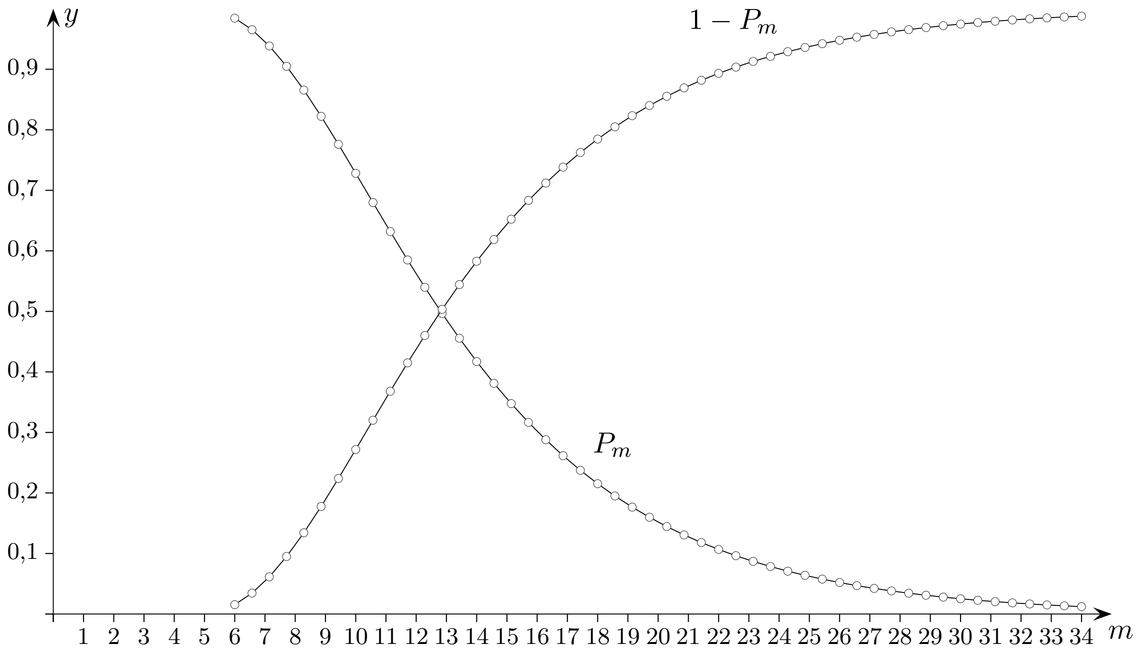
$$P_m = P(A_1 \cup A_2 \cup A_3 \cup A_4) = 4 \cdot (\tfrac{3}{4})^m - 6 \cdot (\tfrac{2}{4})^m + 4 \cdot (\tfrac{1}{4})^m$$

und für $n = 6$

$$P_m = P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6) = 6 \cdot (\tfrac{5}{6})^m - 15 \cdot (\tfrac{4}{6})^m + 20 \cdot (\tfrac{3}{6})^m - 15 \cdot (\tfrac{2}{6})^m + 6 \cdot (\tfrac{1}{6})^m$$

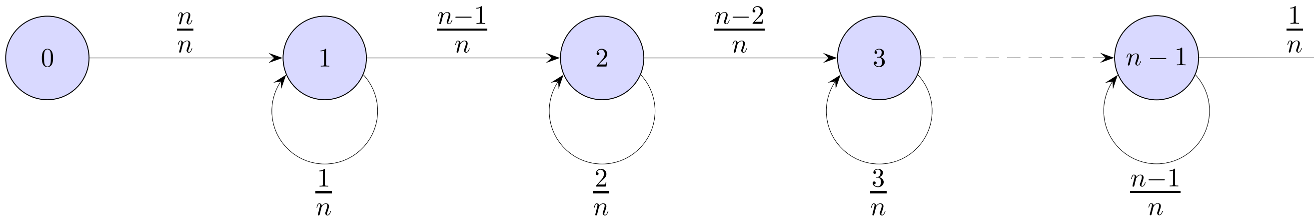
Anwendung der Siebformel, leicht vereinfacht

$n = 6$



Die Wahrscheinlichkeiten P_m streben für wachsendes m natürlich gegen 0.
$1 - P_m$ ist die Wahrscheinlichkeit, mit m Kugeln jedes Fach mindestens einmal zu belegen, das ist die Wahrscheinlichkeit für eine vollständige Serie.

Wahrscheinlichkeiten für eine vollständige Serie schrittweise berechnen



Die Wahrscheinlichkeiten, um nach k Versuchen in einen bestimmten Zustand i zu gelangen, können bei Übergangsdiagrammen dieser Art auf einfache Weise ermittelt werden.

Mit dem 1. Versuch gelangt man in den Zustand 1 (1). (Wahrscheinlichkeiten in Klammern)

Mit dem 2. Versuch bleibt man im Zustand 1 ($\frac{1}{n}$) oder gelangt in den Zustand 2 ($\frac{n-1}{n}$).

Mit dem 3. Versuch sind die Zustände 1, 2 und 3 erreichbar.

Zustand 1 nur, wenn für $k = 2$ der Zustand 1 schon vorlag,
Zustand 2 nur, wenn für $k = 2$ die Zustände 1 oder 2 vorlagen,
Zustand 3 nur, wenn für $k = 2$ der Zustand 2 vorlag.

Mit dem 4. Versuch sind die Zustände 1, 2, 3 und 4 erreichbar.

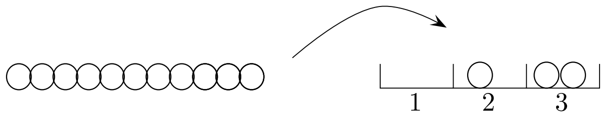
Zustand 1 nur, wenn für $k = 3$ der Zustand 1 schon vorlag,
Zustand 2 nur, wenn für $k = 3$ die Zustände 1 oder 2 vorlagen,
Zustand 3 nur, wenn für $k = 3$ die Zustände 2 oder 3 vorlagen,
...

Die Berechnung wird übersichtlich in einer Tabelle (hier für $n = 6$) angeordnet.

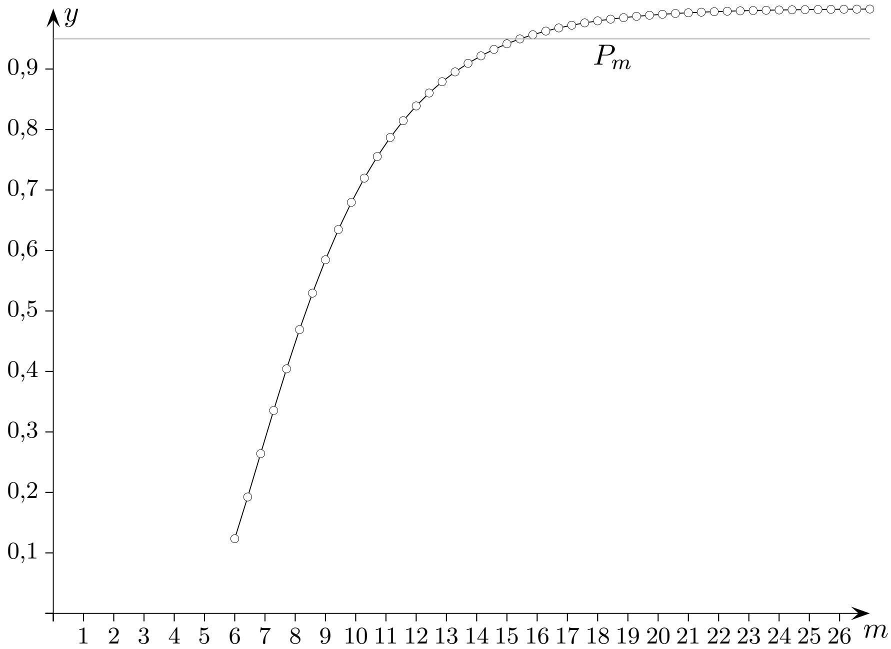
Versuche k \ Zustände i	1	2	3	4	5	6
1	1	0	0	0	0	0
2	$\frac{1}{6}$	$\frac{5}{6}$	0	0	0	0
3	$(\frac{1}{6})^2$	$\frac{1}{6}\cdot\frac{5}{6}+\frac{5}{6}\cdot\frac{2}{6}$	$\frac{5}{6}\cdot\frac{4}{6}$	0	0	0
4	$(\frac{1}{6})^3$	$\frac{35}{216}$	$\frac{120}{216}$	$\frac{5}{6}\cdot\frac{4}{6}\cdot\frac{3}{6}$	0	0
5	$(\frac{1}{6})^4$		$\frac{125}{324}$		$\frac{5}{54}$	0
		$\frac{155}{7776}$	$i-1$	i		
$k+1$			$\frac{n-i+1}{n}$	$\frac{i}{n}$		

An den Pfeilen stehen die Wahrscheinlichkeiten, im $(k+1)$ten Schritt vom Zustand $i-1$ nach i zu wechseln oder den Zustand i beizubehalten.

in jedem Brötchen mindestens 2 Rosinen



m Kugeln sollen auf $n = 3$ Fächer nacheinander zufällig verteilt werden.
Wie viele Kugeln sind mindestens erforderlich, damit mit (mindestens) 95 %-iger Wahrscheinlichkeit sich in jedem Fach mindestens 2 Kugeln befinden?

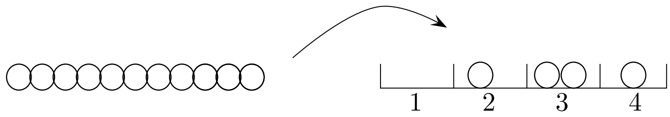


Die Wahrscheinlichkeiten P_m streben für wachsendes m natürlich gegen 1.
Die Antwort lautet: 16 Kugeln, $P_{16} = 95{,}9\,\%$

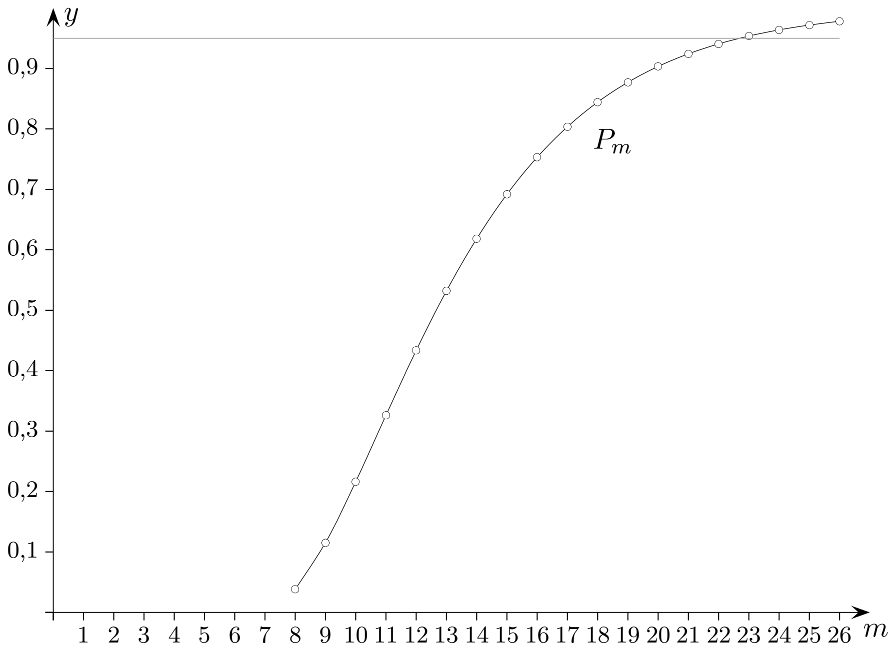
$$P_m = \frac{\sum\limits_{k=2}^{m-4} \left[\binom{m}{k} \sum\limits_{i=2}^{m-k-2} \binom{m-k}{i} \right]}{3^m}$$

Um die günstigen Möglichkeiten zu ermitteln, werden die Möglichkeiten für das 1. Fach (Anzahl $\sum\limits_{k=2}^{m-4} \binom{m}{k}$) mit den Möglichkeiten für das 2. Fach (Anzahl $\sum\limits_{i=2}^{m-k-2} \binom{m-k}{i}$) kombiniert. Die restlichen Kugeln landen im 3. Fach. Die obere Summationsgrenze $m-4$ stellt sicher, dass für die übrigen beiden Fächer mindestens 4 Kugeln übrig bleiben.

in jedem Brötchen mindestens 2 Rosinen



m Kugeln sollen auf $n = 4$ Fächer nacheinander zufällig verteilt werden.
Wie viele Kugeln sind mindestens erforderlich, damit mit (mindestens) 95 %-iger Wahrscheinlichkeit sich in jedem Fach mindestens 2 Kugeln befinden?



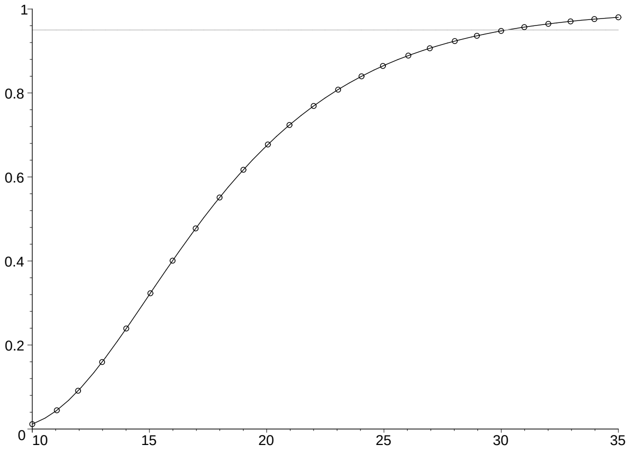
Die Antwort lautet: 23 Kugeln, $P_{23} = 95{,}4\,\%$

$$P_m = \frac{\sum\limits_{k=2}^{m-6} \Big[\binom{m}{k} \sum\limits_{i=2}^{m-k-4} \Big[\binom{m-k}{i} \sum\limits_{j=2}^{m-k-i-2} \binom{m-k-i}{j} \Big]\Big]}{4^m} = \frac{\sum\limits_{k=2}^{m-6} \Big[\sum\limits_{i=2}^{m-k-4} \Big[\sum\limits_{j=2}^{m-k-i-2} \binom{m}{k}\binom{m-k}{i}\binom{m-k-i}{j} \Big]\Big]}{4^m}$$

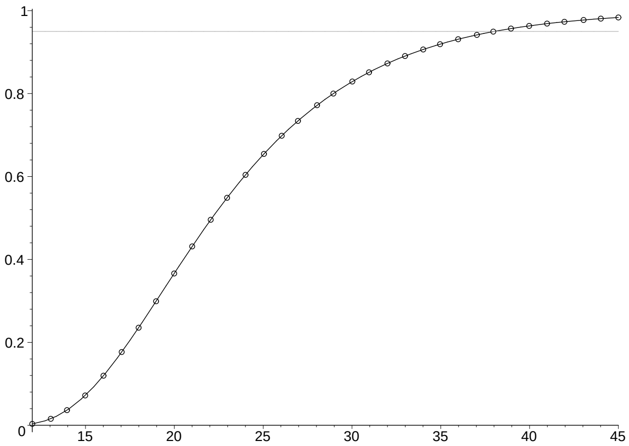
Mit dieser Zählmethode sind der Berechnung (Maple) leider Grenzen gesetzt.
Mir ist jedoch nichts Einfacheres eingefallen.
Immerhin gibt es für 23 Kugeln 70 368 744 177 664 mögliche Verteilungen, von denen 67 112 688 842 496 die geforderte Bedingung erfüllen.

in jedem Brötchen mindestens 2 Rosinen

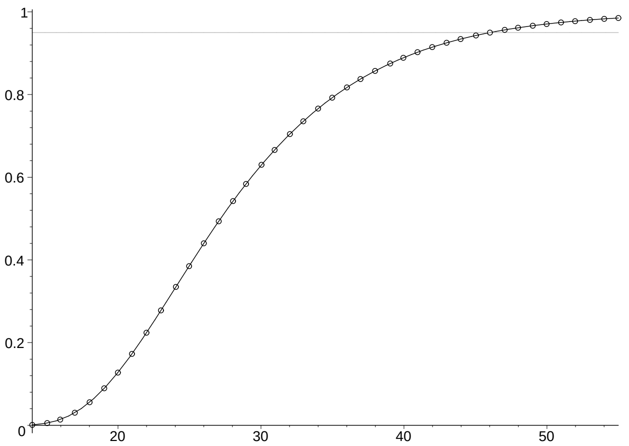
$n = 5$ Fächer



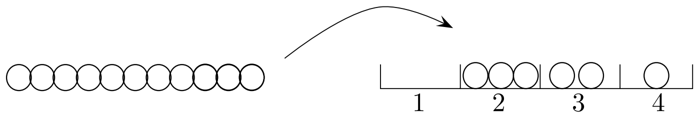
$n = 6$ Fächer



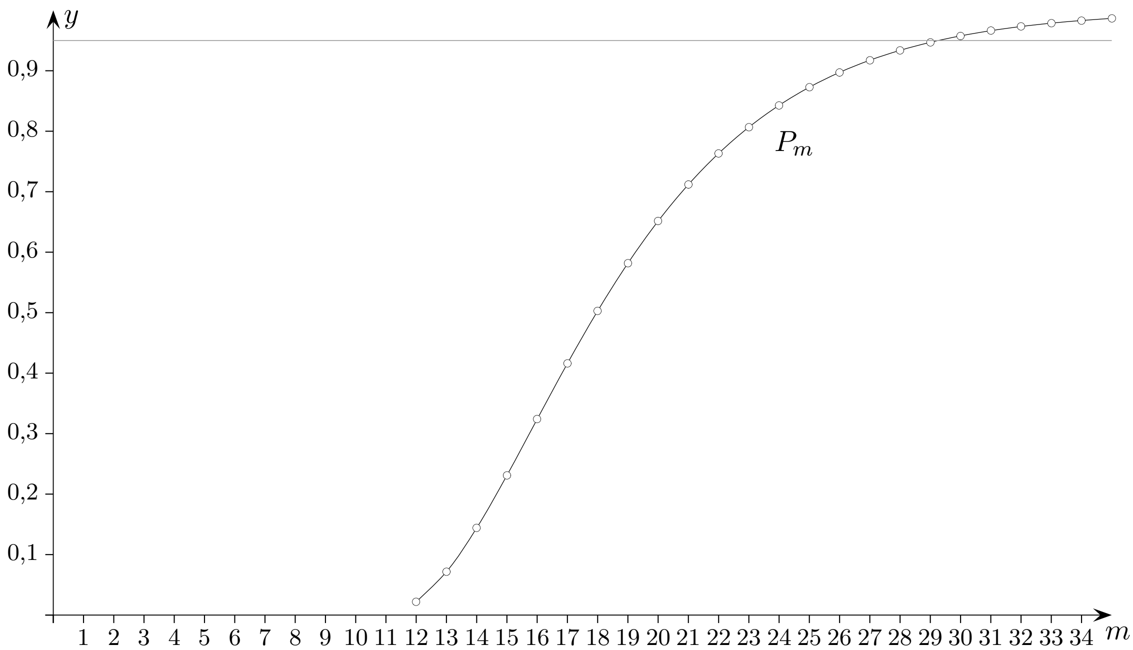
$n = 7$ Fächer



in jedem Brötchen mindestens 3 Rosinen



m Kugeln sollen auf $n = 4$ Fächer nacheinander zufällig verteilt werden.
Wie viele Kugeln sind mindestens erforderlich, damit mit (mindestens) 95 %-iger Wahrscheinlichkeit sich in jedem Fach mindestens 3 Kugeln befinden?



Die Antwort lautet: 30 Kugeln, $P_{30} = 95{,}7\,\%$

$$P_m = \frac{\sum\limits_{k=3}^{m-9} \Big[\binom{m}{k} \sum\limits_{i=3}^{m-k-6} \Big[\binom{m-k}{i} \sum\limits_{j=3}^{m-k-i-3} \binom{m-k-i}{j} \Big]\Big]}{4^m} = \frac{\sum\limits_{k=3}^{m-9} \Big[\sum\limits_{i=3}^{m-k-6} \Big[\sum\limits_{j=3}^{m-k-i-3} \binom{m}{k}\binom{m-k}{i}\binom{m-k-i}{j} \Big]\Big]}{4^m}$$