Warum sich Ereignisse oft häufen:
Die Exponentialverteilung trifft man auch im Schwimmbad
von Wolfgang Riemer, Köln

Abstract

- Man kommt aus dem Schwimmbecken. Nur wenige Leute sind im Umkleideraum, aber „fast immer“ haben „fast alle“, die sich umziehen, ihre Schränke direkt neben dem Ihren.
- Das gleiche bei den Geburtstagen: Das Jahr hat 365 Tage, aber oft liegen die Geburtstage der Freunde so nahe beieinander, dass man gar nicht alle mit feiern kann.
- Man wartet schon seit Tagen auf wichtige Anrufe oder Mails - und dann kommen „alle auf einmal“

Unser Alltag ist voller Situationen, bei denen Dinge sich „scheinbar zufällig“ häufen. Dahinter steckt „System“: Wenn nämlich Ereignisse auf einer Zahlengeraden (zeitlich oder räumlich) „zufällig gleichmäßig“ verteilt sind, sind die Abstände zwischen ihnen exponential verteilt. Und das bedeutet, dass kurze Abstände viel häufiger auftreten als lange, dass sich die „Dinge“ also häufen..
Das Phänomen kennt man in einer diskreten Variante auch vom „Warten auf die 6“ beim Würfeln. Wenn man nämlich die gewürfelten Augenzahlen hintereinander schreibt, sind die Treffer auf der Zeitachse gleichmäßig verteilt. Die Abstände dazwischen sind die „Wartezeiten“ auf Erfolg, von denen man weiß dass sie geometrisch verteilt sind. Und die geometrische Verteilung ist die „diskrete Schwester“ der Exponentialverteilung, vgl. Abschnitt 7.
Der Artikel enthält vier Aufgabenstellungen, die man arbeitsteilig oder als Stationenlernen im Unterricht einsetzen kann. Dabei wird die Exponentialverteilung als Werkzeug verwendet, denn in vielen SII-Lehrplänen wird die Vernetzung von Analysis und Stochastik empfohlen und die Exponentialverteilung ist ein idealer Brückenschlag zwischen Integralrechnung und Stochastik . [1]
Der Autor „stolperte“ in die Thematik, als er im Rahmen eines Excel-Kurses eher zufällig eine Tabelle nach den Geburtstagen der mehr als 2000 Ehemaligen seiner Schule sortierte und die Differenzen zwischen den Geburtsdaten grafisch darstellen wollte.

1 Geburtstage

Fig. 1 zeigt einen Ausschnitt aus dem Geburtstagskalender für eine Jahrgangsstufe (60 Schüler) zusammen mit der Wartezeit (Abstand $X \geq 0$ in Tagen), die jeder einzelne Schüler bis zum nächsten „Stufengeburtstag“ warten muss[2]. Falls an einem Tag mehrere Schüler Geburtstag haben, beträgt für alle die Wartezeit jeweils 0 (zweite Zeile in Fig. 1). Fig. 2 erweckt den Anschein, als könne man das „Warten auf den nächsten Geburtstag“ näherungsweise durch eine Exponentialfunktion beschreiben. Das soll anhand der Aufgabenteile a) bis c) untersucht werden.

01.05	Sarah	1
02.05	Ann, Tim, Maika	0,0,0
03.05		
05.05	Jana	4
05.05		
06.05		
07.05		
08.05	Frank, Jim	0,0

Fig. 1 Geburtstagskalender (Ausschnitt). Angegeben ist für jede Person auch der Abstand bis zum nächsten Geburtstag.

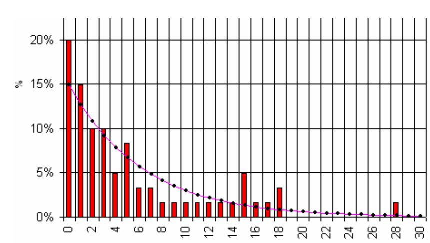

Fig. 2 Verteilung der Abstände (Jahrgangsstufe mit 60 Schülern)

a) Bestimmen Sie den Mittelwert $\bar{x}$ der relativen Häufigkeitsverteilung von Fig. 2, d. h. die mittlere Anzahl $\bar{x}$ von Tagen bis zum nächsten Geburtstag. Tipp: $\bar{x} = 0 \cdot h_0 + 1 \cdot h_1 + ... + 28 \cdot h_{28}$

Dabei ist h_i die relative Häufigkeit der Warezeit i

b) Da die Exponentialverteilung mit der Dichtefunktion f mit $f(x) = \lambda \cdot e^{-\lambda x}$ den Erwartungswert $\mu = \frac{1}{\lambda}$ besitzt, wird man für die Dichte den Parameter $\lambda = \frac{1}{\mu} \approx \frac{1}{\bar{x}}$ wählen. Vergleichen Sie die relativen Häufigkeiten h_0 , h_1,… aus Fig. 2 mit den Wahrscheinlichkeiten der Form

$$P(0 \le X < 1) = \int_0^1 \lambda \cdot e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^1, \quad P(1 \le X < 2) = \int_1^2 \lambda \cdot e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_1^2) \ldots$$

c) Erstellen Sie selber einen Geburtstagskalender wie in Fig. 1 (etwa für Ihre Jahrgangsstufe) und simulieren Sie 10 Geburtstagskalender mit 60 Geburtstagen oder so vielen Geburtstagen, wie ihr Schüler in Ihrer Stufe sind. Fassen Sie die Daten aus den 10 Simulationen zusammen und erstellen Sie ein „gemeinsames“ Säulendiagramm wie in Fig. 2 und vergleichen Sie die relativen Häufigkeiten wieder mit den Wahrscheinlichkeiten aus b). Der Abstand des letzten Dezember-Gebutstages zum ersten Januar-Geburtstag zählt mit.

2 Schwimmbad

Im Umkleideraum stehen n=100 Schränke, von denen k=20 zufällig belegt sind. Nach dem Schwimmen wollen Sie sich umziehen. Wie groß der Abstand zum nächsten belegten Schrank? Auch der Abstand des letzten Schrankes zum ersten soll mitzählen, so dass man bei 20 Belegungen auch 20 Abstände bekommt.

a) Simulieren Sie diese Situation mehrfach mit Zetteln oder Zufallszahlen.

b) Untersuchen Sie, ob die Abstände k≥1 zum jeweils nächsten Schrank näherungsweise exponential verteilt sind.

Hinweis: der Mittelwert der Abstände ist zwischen den 20 belegten Schränken ist stets $\bar{x} = \frac{n}{k} = \frac{20}{100} = 5$. Vergleichen Sie die relativen Häufigkeiten h_1, h_2,… der Abstände 1, 2, …mit den Wahrscheinlichkeiten der Form $P(0 \le X \le 1) = \int_0^1 \lambda \cdot e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^1$,

$P(1 < X \le 2) = \int_1^2 \lambda \cdot e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_1^2)$,… wobei man $\lambda = \frac{k}{n} = \frac{1}{\bar{x}}$ wählt.

c) Zusatz: Variieren Sie n und k.

3 Zufallsdezimalzahlen

Wenn man 100 Zufallszahlen aus dem Intervall [0;1], der Größe nach sortiert, erhält man 99 Abstände (Differenzen). Wie sieht die zugehörige Verteilung aus?

a) Erzeugen Sie eine Liste (L1) aus 100 solcher Zufallszahlen, sortieren Sie diese. Bilden Sie hieraus die Liste (L2) der Abstände aus je zwei aufeinander folgenden Zahlen und stellen Sie die Häufigkeitsverteilung der Abstände graphisch dar. Bilden Sie dazu Klassen der Breite 0,03. Kommentieren Sie.

b) Bei 100 sortierten Zahlen im Intervall [0;1] ist der Erwartungswert des Abstandes μ=0,01. Ermitteln Sie Mittelwert $\bar{x}$ der Differenzen in Ihrer Liste L2.

c) Ist das Ergebnis mit der Annahme einer Exponentialverteilung vereinbar? Berechnen Sie - unter Annahme einer Exponentialverteilung - die Wahrscheinlichkeit, dass der Abstand zweier aufeinander folgender Zahlen mehr als 0,01 beträgt. Vergleichen Sie mit ihrem experimentellen Ergebnis.

Für Einstellung der Klassenbreite 0,3 nutzen Sie die Einstellung Xscl im Window Menü

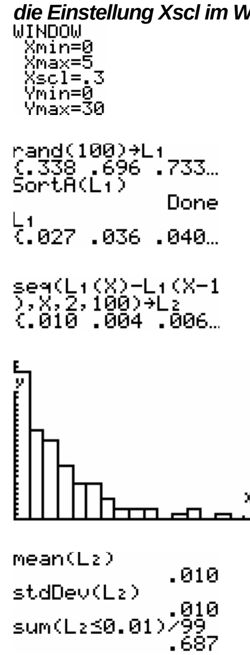

Fig. 3 GTR-Befehle

4 Theorie: Exponentialverteilung

Satz: Wenn auf einer Zahlengeraden Zufallszahlen gleichmäßig so verteilt, dass im Mittel auf eine Einheit n Zahlen kommen, dann sind die Abstände zwischen benachbarten Zahlen exponentialverteilt mit der Dichte $f(x)=\lambda\cdot e^{-\lambda x}$ wobei für den Parameter gilt $\lambda = n$ bzw. $\mu=\frac{1}{n}$.

Dieser Satz soll am Beispiel n=100 in drei Schritten begründet werden.

Schritt 1: Lassen Sie in Gedanken 100 Zufallszahlen zufällig auf das Intervall [0;1] „fallen“. Nehmen Sie in Gedanken eine beliebige Stelle des Intervalls heraus. Begründen Sie: Die Wahrscheinlichkeit, dass im Intervall [a;a+x] keine Zufallszahl liegt, ist $(1-x)^{100}$.

Schritt 2: Nun lässt man 200 Zufallszahlen auf das Intervall [0;2] „fallen“. Begründen Sie: Die Wahrscheinlichkeit, dass im Intervall [a;a+x] keine Zufallszahl liegt, ist nun $(1-\frac{x}{2})^{2\cdot 100}$ … und wenn man n*100 Zufallszahlen auf das Intervall [0;n] fallen lässt, liegen mit Wahrscheinlichkeit $(1-\frac{x}{n})^{n\cdot 100}\approx e^{-100x}$ keine Zahlen in [a;a+x].

Schritt 3: [3] Folgern Sie hieraus: Wenn eine Zahlengerade gleichmäßig von Zufallszahlen mit der Dichte „100 Zahlen je Einheit“ bevölkert ist, dann ist der Abstand zwischen zwei aufeinander folgenden Zahlen exponentialverteilt mit der Dichte $f(x)=\lambda\cdot e^{-\lambda x}$ wobei für den Parameter gilt $\lambda=100$ bzw. $\mu=\frac{1}{100}$. Zeigen Sie dazu: $P(0\le X\le x)=1-e^{-100x}$

5 heuristische Synthese (zu 1 bis 4)
Es ist klar, dass das die Abstände zwischen Geburtstagen oder belegten Schwimmbadschränken nicht exakt durch eine Exponentialverteilung beschrieben werden, denn erstens sind exponentialverteilte Zufallsgrößen nicht ganzzahlig und zweitens können sie prinzipiell beliebig große Werte annehmen. Wenn man aber hinreichend viele diskrete Zufallszahlen auf einem beschränkten Intervall verstreut, dessen diskrete Gitterpunkte nahe beieinander liegen, spielen die Unterschiede zwischen dem unendlichen kontinuierlichen und der endlichen diskreten Sicht nur eine geringe Rolle. Das erklärt die „universelle Bedeutung“ der Exponentialverteilung und macht plausibel, dass die Ergebnisse aus 1 bis 3 durch Exponentialverteilungen gut angenähert werden…. und erklärt sogar die oft angeführte Behauptung, dass die Dauer von Telefongesprächen in großen Büros exponential verteilt sein soll. Man stelle sich dazu ein Sekretariat vor, in dem kontinuierlich Arbeiten - gleichmäßig über den Tag verteilt - zu erledigen sind und die (exponentialverteilten) Freiräume dazwischen für Telefonate genutzt werden.

6 exakte Lösungen
Der Vollständigkeit halber sollen auch die exakten Lösungen des Geburtstags- und des Schwimmbadproblems mit Pfadregel und Kombinatorik in Form einer „Zusatzaufgabe“ entwickelt werden. In beiden Fällen werden von der Struktur her k Kugeln auf n Fächer verteilt. Dann werden die Abstände zwischen den belegten Fächern untersucht. Beim Geburtstagsproblem dürfen die Fächer mehrfach belegt sein (das führt jeweils zum Abstand 0), beim Schwimmbadproblem nicht. Dort hat der minimale Abstand den Wert 1.
Wir rechnen mit k=60 und n=365.

6.1 Geburtstage (Mehrfachbelegung)
Greifen Sie einen der 60 Schüler (wir nennen ihn „Tim“) heraus. Begründen Sie: Die Wahrscheinlichkeit, dass „Tims Wartezeit auf den nächsten Geburtstag“ den Wert

a) 0 hat, ist $P(X=0)=1-\left(\frac{364}{365}\right)^{59}\approx 0{,}0765$

(„Trefferwahrsceinlichkeit“ p)

b) 1 hat, ist $P(X=1)=\left(\frac{364}{365}\right)^{59}-\left(\frac{363}{365}\right)^{59}$

c) hat, ist $P(X=d)=\left(\frac{365-d}{365}\right)^{59}-\left(\frac{365-d-1}{365}\right)^{59}$

d) 364 hat, ist $P(X=365)=\left(\frac{1}{365}\right)^{59}$

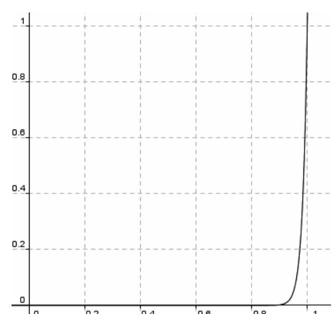

Fig. 4 $f(x)=x^{59}$

Tipp zu b) Begründe $\left(\frac{364}{365}\right)^{59}$ ist die Wahrscheinlichkeit $P(X\geq 1)$

e) (optional) Fig. 4 zeigt den Graphen zu $f(x)=x^{59}$. Erläutern Sie wie man an diesem Graphen die Wahrscheinlichkeiten aus a) bis d) ablesen könnte. Wie würden sich diese Wahrscheinlichkeiten ändern, wenn man statt 60 Schülern 30 (100) in der Stufe hätte?.

f) Erläutern Sie die folgende Termumformung und begründen Sie, dass die Abstände zwischen den Geburtstagen näherungsweise exponentialverteilt sind mit dem Parameter $\lambda = \frac{59}{365}$

$$P(X \geq d) = \left(\frac{365-d}{365}\right)^{59} = \left(1 - \frac{\frac{59 \cdot d}{365}}{59}\right)^{59} \approx e^{-\frac{59}{365} \cdot d} = e^{-\lambda \cdot d}$$, d=0, 1, 2,…

g) Vergleichen Sie die exakten Wahrscheinlichkeiten P(X=d) mit den Näherungswerten $P(d-1 \leq X < d) = \int_{d-1}^{d} \lambda \cdot e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_{d-1}^{d}$, die die Exponentialverteilung mit Parameter $\lambda = \frac{59}{365}$ liefern würde.

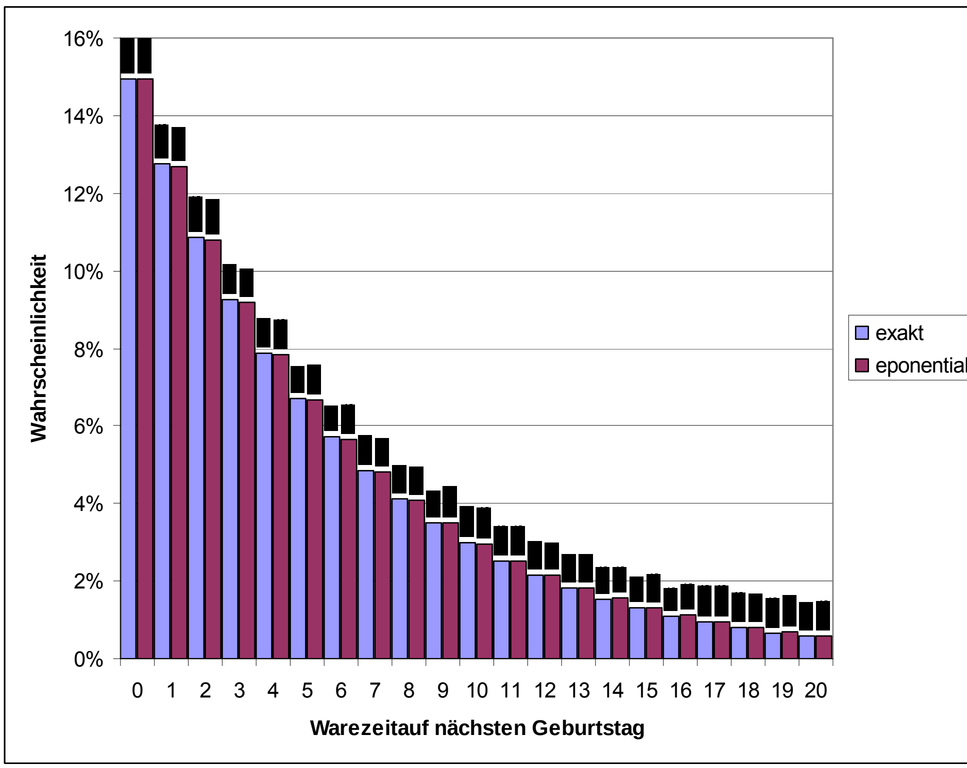

Fig. 5 Lösung zu g

6.2 Schränke (keine Mehrfachbelegung)

Greifen Sie „Tims“ Schrank heraus. Begründen Sie: Für die Wahrscheinlichkeit des Abstandes X zum nächsten belegten Schrank gilt:

a) $P(X \geq 2) = \left(\dfrac{\binom{364-1}{59}}{\binom{364}{59}} \right) = \dfrac{363 \cdot 362 \cdot \ldots \cdot 305}{364 \cdot 363 \cdot \ldots \cdot 306} = \dfrac{305}{364}$,

Insbesondere gilt $P(X = 1) = \dfrac{59}{364}$ („Trefferwahrscheinlichkeit“ p)

b) $P(X \geq 3) = \left(\dfrac{\binom{364-2}{59}}{\binom{364}{59}} \right) = \dfrac{362 \cdot 361 \cdot \ldots \cdot 304}{364 \cdot 363 \cdot \ldots \cdot 306} = \dfrac{305}{364} \cdot \dfrac{304}{363}$

c) $P(X \geq d) = \left(\dfrac{\binom{364-d+1}{59}}{\binom{364}{59}} \right) = \dfrac{305}{364} \cdot \dfrac{304}{363} \cdot \ldots \cdot \dfrac{307-d}{366-d}$

d) $P(X \geq 306) = \left(\dfrac{\binom{59}{59}}{\binom{364}{59}} \right) = \dfrac{59 \cdot 58 \cdot \ldots \cdot 1}{364 \cdot 363 \cdot \ldots \cdot 306} = \dfrac{305}{364} \cdot \dfrac{304}{363} \cdot \ldots \cdot \dfrac{1}{60}$

e) Berechnen Sie daraus mit Hilfe einer Tabellenkalkulation die Wahrscheinlichkeiten P(X=d) und vergleichen Sie mit den Wahrscheinlichkeiten $P(d-1 < X \leq d) = \int\limits_{d-1}^{d} \lambda \cdot e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_{d-1}^{d}$, die die Exponentialverteilung mit Parameter $\lambda = \dfrac{60}{365}$ (also mit dem gleichen Erwartungswert) liefern würde.

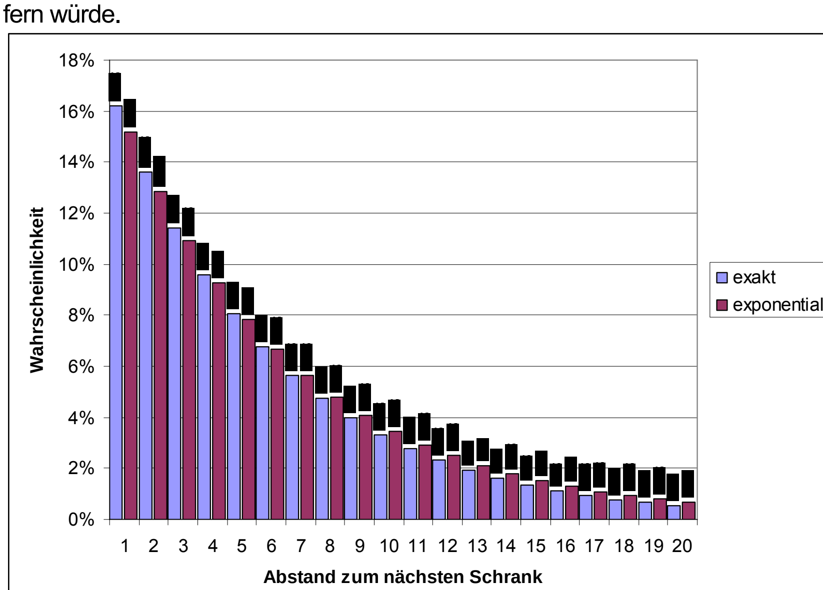

Fig. 6 Lösung zu e

7 Die Exponentialverteilung und ihre diskrete Schwester, die geometrische Verteilung

7.1 Erinnerung

(1) Eine positive **reellwertige** Zufallsgröße X≥0 heißt **exponentialverteilt** mit dem Parameter λ (der Wahrscheinlichkeitsdichte $f_\lambda(x) = \lambda \cdot e^{-\lambda x}$), wenn sich die Wahrscheinlichkeit, dass X zwischen a und b liegt, berechnen lässt durch $P(a \le X \le b) = \int_a^b \lambda \cdot e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_a^b$ insbesondere gilt $P(X \le t) = 1 - e^{-\lambda t} = 1 - (e^{-\lambda})^t$ .X hat den Erwartungswert μ=1/ λ .

(2) Eine **ganzzahlige** Zufallsgröße Z ist **geometrisch verteilt** mit dem Parameter q, wenn gilt $P(Z = t) = p \cdot q^{t-1}$. Insbesondere gilt $P(Z \le t) = 1 - q^t$.
Z hat den Erwartungswert μ=1/p, wobei gilt p=1-q. p heißt „Trefferwahrscheinlichkeit".

(3) Wenn man eine exponentialverteilte (rellwertige) Zufallsgröße X mit dem Parameter λ aufrundet, entsteht eine ganzzahlige Zufallsgröße Z, die geometrisch verteilt ist mit dem Parameter $q = 1/e^\lambda$ (bzw. $\lambda = \ln(1/q)$). Für $t \in Z$ sind nämlich die Bedingungen $Z \le t$ und $X \le t$ gleichwertig und es gilt $P(Z \le t) = P(X \le t) = 1 - e^{-\lambda t} = 1 - (\frac{1}{e^\lambda})^t = 1 - q^t$.

Es ist daher nicht verwunderlich, dass man zu Approximation statt der Exponential- auch die geometrische Verteilung nutzen kann. Man braucht dann (genau wie beim Warten auf den nächsten Treffer in einer Bernouliikette) eine geeignete Trefferwahrscheinlichkeit p, die sich aber (im Gegensatz zur Bernoullikette) bei jedem Warteschritt auf den nächsten Geburtstag oder belegten Schrank (ein wenig) ändert, was aber für die relevanten Abstände d kaum eine Rolle spielt.

7.2 Schwimmbadproblem

Hier verwendet „Tim“ für die geometrische Näherung als Trefferwahrscheinlichkeit (die Wahrscheinlichkeit, dass der Schrank rechts neben seinem eigenen belegt ist), p=59/364, (vgl. 6.2). Dann gilt $P(X=d) \approx q^{d-1} \cdot p$, d=1, 2, 3,…

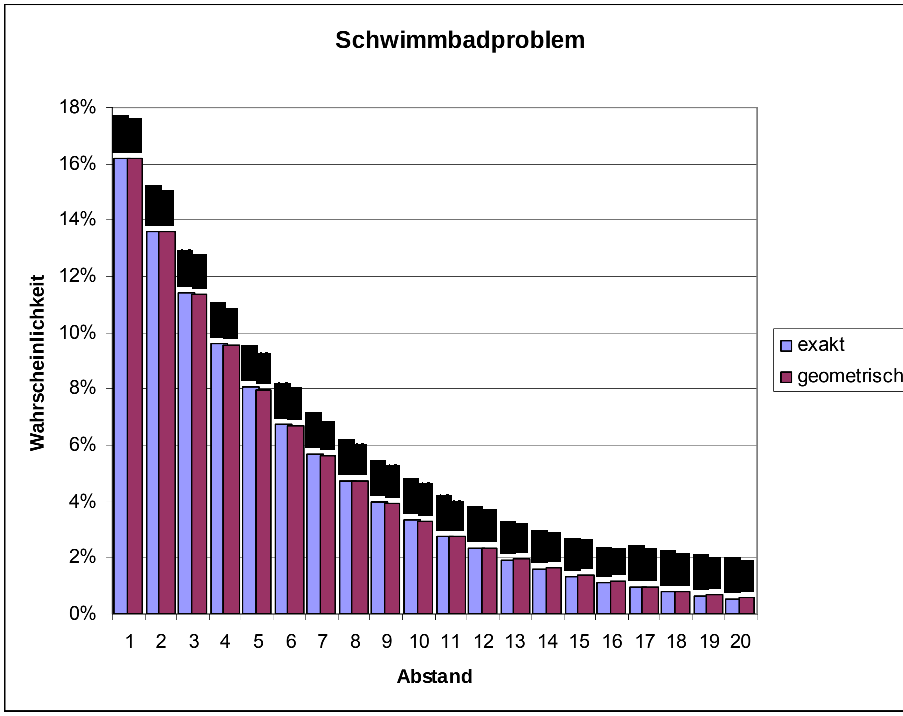

Fig. 7 Exakte Wahrscheinlichkeiten und geometrische Approximation im Vergleich

7.3 Geburtstagsproblem

Für die Wahrscheinlichkeit, dass Tim keinen Geburtstagszwilling hat, ist $q = \left(\frac{364}{365}\right)^{59}$. Die Trefferwahrscheinlichkeit (dass er einen Geburtstagszwilling hat, also d=0 gilt,) ist dann $p = 1 - \left(\frac{364}{365}\right)^{59}$

Damit hat man die Näherung $P(X = d) \approx q^d \cdot p$, d=0, 1, 2,

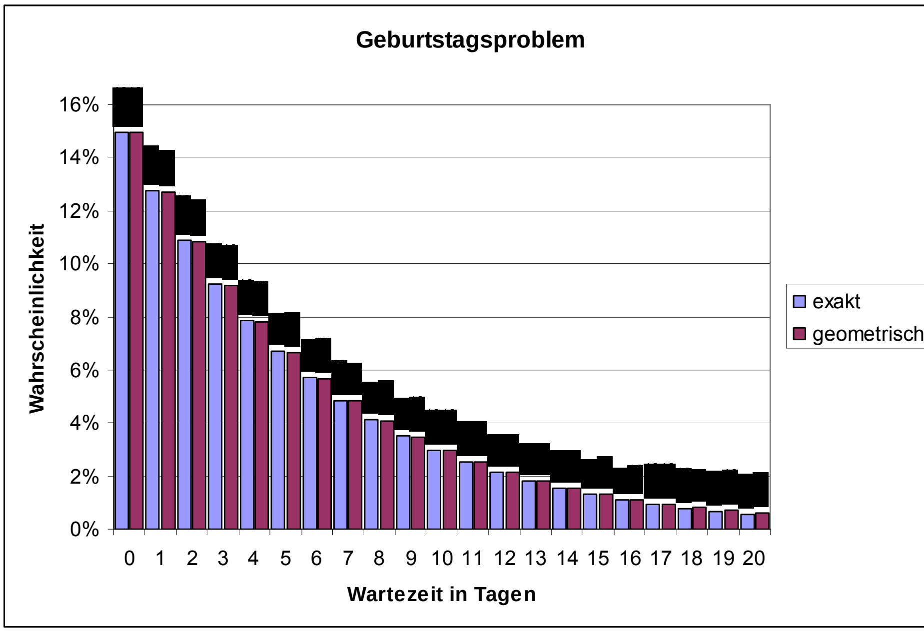

Fig. 8 Exakte Wahrscheinlichkeiten und geometrische Approximation im Vergleich

Der Autor unterrichtet Mathematik, Physik und Informatik am Heinreich-Mann-Gymnasium in Köln und ist Fach- und Hauptseminarleiter am Studienseminar für Lehrämter an Schulen, Köln

Dr. Wolfgang Riemer
August-Bebel-Str. 80
50259 Pulheim
w.riemer@arcor.de

---

[1] Wer sich an deren Eigenschaften erinnern möchte oder sich für dem Zusammenhang zwischen stetiger Exponentialverteilung und diskrteter geometrischer Verteilung interessiert ist oder bei Erarbeitung des Themas auf die stetige Exponentialverteilung ganz verzichten möchte, sei auf den letzten Abschnitt 7 verwiesen.

[2] Wenn man nicht von den 60 Schülern ausgeht, sondern von den 365 Tage des Jahres und die Abstände zwischen den belegten Tagen untersucht, ergibt sich eine Zufallsgröße, deren Werte nicht 0 annehmen können.

[3] Zur Erinnerung an e^x : Erklären Sie folgende Aussage durch ein Zahlenbeispiel mit einem TR: Wenn man ein Kapital mit dem Prozentsatz p (dem Wachstumsfaktor x=1+p) statt nur einmal im Jahr stetig („sekündlich") verzinsen würde, dann würde es in einem Jahr nicht um x sondern um den Faktor e^x anwachsen. $(1+\frac{x}{n})^n \xrightarrow[n\to\infty]{} e^x$.