Warum sich Ereignisse oft häufen

Die Exponentialverteilung im Schwimmbad und bei Geburtstagen – Ein Lernzirkel mit Stationen zum Experimentieren und zum Nachdenken

Sie kommen nach dem Schwimmen in den Umkleideraum. Es sind nur wenige Menschen da, 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. Oder Sie warten 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. Daher treten kurze Abstände viel häufiger auf als lange. Die „Dinge" häufen sich also. 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.

Stationen zum Experimentieren

Die **Stationen 1** bis **3** stellen drei „Abstands"-Probleme vor (im einzelnen: Geburtstage, Schränke im Schwimmbad, Zufalls-Dezimalzahlen). Es wird jeweils die Verteilung der Abstände zum nächsten Geburtstag, zum nächsten belegten Schrank oder zur nächsten Zufallszahl (Differenz) untersucht. Stellt man die Häufigkeitsverteilung grafisch dar, scheint es einen exponentiellen Zusammenhang zu geben. Das Bild in Station 1 zum Beispiel erweckt den Anschein, als könne man das „Warten auf den nächsten Geburtstag" näherungsweise durch eine Exponentialfunktion beschreiben.

Es ist klar, dass die Abstände zwischen Geburtstagen oder belegten Schwimmbadschränken nicht *exakt* durch eine Exponentialverteilung beschrieben werden. 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 der drei „Abstands"-Beispiele durch Exponentialverteilungen gut angenähert werden. (Und es 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.)

Die Exponentialverteilung

Eine positive *reellwertige* Zufallsgröße $X \geq 0$ 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

$$P(a \leq X \leq b) = \int_a^b \lambda \cdot e^{-\lambda x} dx = [-e^{-\lambda x}]_a^b.$$

Insbesondere gilt $P(X \leq t) = 1 - e^{-\lambda t} = 1 - (e^{-\lambda})^t$. Dabei hat X den Erwartungswert $\mu = 1/\lambda$.

Die Geometrische Verteilung

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

Wenn man eine exponentialverteilte (reellwertige) 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 $t \in \mathbb{Z}$ sind nämlich die Bedingungen $Z \leq t$ und $X \leq t$ gleichwertig und es gilt:

$$P(Z \leq t) = P(X \leq t) = 1 - e^{-\lambda t} = 1 - \left(\frac{1}{e^{\lambda}}\right)^t = 1 - q^t.$$

Es ist daher nicht verwunderlich, dass man zur Approximation statt der Exponential- auch die geometrische Verteilung nutzen kann. Man braucht dann (genau wie beim Warten auf den nächsten Treffer in einer Bernoullikette) 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.

Stationen zum Nachdenken

In **Stationen 4** wird das (in den Stationen 1 bis 3 beobachtete „näherungsweise") Auftreten der Exponentialverteilung analytisch begründet. Die vertiefenden **Stationen 5** und **6** beschäftigen sich mit den exakten Verteilungen beim Geburtstags- und Schwimmbadproblem. In beiden Fällen werden von der Struktur her k Kugeln auf n Fächer verteilt – und die Abstände zwischen belegten Fächer (mit der Pfadregel bzw. mit der Kombinatorik) untersucht. Wir arbeiten exemplarisch mit $k = 60$ und $n = 365$.

Beim Geburtstagsproblem dürfen die Fächer mehrfach belegt sein (was zum Abstand 0 führt), beim Schwimmbadproblem nicht. Dort hat der minimale Abstand den Wert 1. Darüber hinaus wird die Brücke zur Exponential- und zur geometrischen Verteilung geschlagen. Mit Excel lassen sich die Unterschiede zwischen exakten und Näherungs-Verteilungen visualisieren (vgl. S. 60).

Mit einem Lotto-Problem in Station 6 wird die Schwimmbad-Thematik auf eine weitere alltägliche Situation übertragen

Wolfgang Riemer,
Köln

Anmerkung

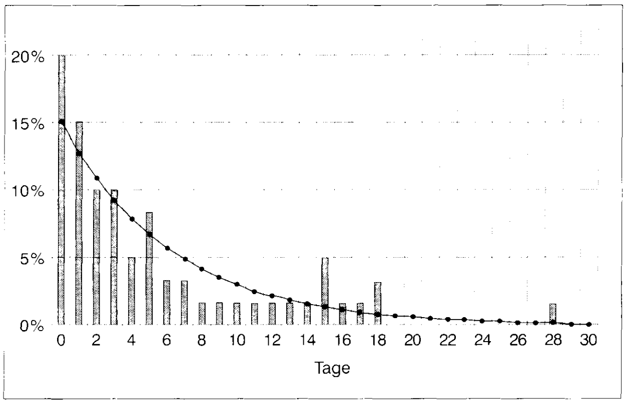
Die hier verwendeten Dateien kann man herunterladen unter www.riemer-koeln.de

Experiment: Geburtstage

Die **Tabelle** zeigt einen Ausschnitt aus dem Geburtstagskalender für eine Jahrgangsstufe (60 Schüler) zusammen mit der Wartezeit (Abstand $X \geq 0$ in Tagen), die jedes Geburtstagskind bis zum nächsten „Stufengeburtstag" warten muss. Falls an einem Tag mehrere Schüler Geburtstag haben, beträgt für alle die Wartezeit jeweils 0. Das **Bild** erweckt den Anschein, als könne man das „Warten auf den nächsten Geburtstag" näherungsweise durch eine Exponentialfunktion beschreiben.


© Fotolia.com

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



→ **AUFGABEN**

1. Bestimmen Sie den Mittelwert $\bar{x}$ der relativen Häufigkeitsverteilung im Bild oben, d. h. die mittlere Anzahl $\bar{x}$ von Tagen bis zum nächsten Geburtstag.
   Tipp: Es gilt $\bar{x} = 0 \cdot h_0 + 1 \cdot h_1 + \ldots + 28 \cdot h_{28}$. Dabei ist h_i die relative Häufigkeit der Warezeit i.

2. 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 $h_0, h_1, \ldots$ aus dem Bild mit den Wahrscheinlichkeiten der Form

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

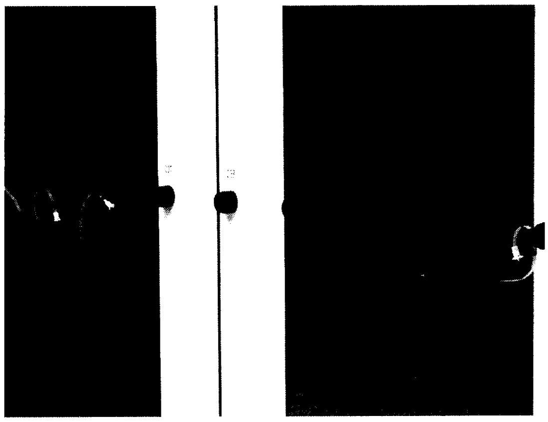
3. Erstellen Sie selber einen Geburtstagskalender (etwa für Ihre Jahrgangsstufe oder den unten abgebildeten Mathe-Kurs) und simulieren Sie 10 Geburtstagskalender. Fassen Sie die Daten aus den 10 Simulationen zusammen und erstellen Sie ein „gemeinsames" Säulendiagramm wie im Bild und vergleichen Sie die relativen Häufigkeiten wieder mit den Wahrscheinlichkeiten aus Aufgabe 2. Der Abstand des letzten Dezember-Gebutstages zum ersten Januar-Geburtstag zählt mit.


Foto: W. Riemer

STATION 2

Experiment: Schwimmbad

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



→ AUFGABEN

1. Simulieren Sie diese Situation mehrfach mit Zetteln oder Zufallszahlen.

2. Untersuchen Sie, ob die Abstände $k \geq 1$ zum jeweils nächsten Schrank näherungsweise exponential verteilt sind.
   Hinweis: Der Mittelwert der Abstände ist zwischen den zwanzig belegten Schränken ist stets $\bar{x} = \frac{n}{k} = \frac{20}{100} = 5$.
   Vergleichen Sie die relativen Häufigkeiten $h_1, h_2, \ldots$ der Abstände 1, 2, … mit den Wahrscheinlichkeiten der Form

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

   wobei man $\lambda = \frac{k}{n} = \frac{1}{\bar{x}}$ wählt.

3. Zusatzaufgabe: Verändern Sie n und k.

STATION 3

Experiment: 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?

→ AUFGABEN

1. Erzeugen Sie eine Liste (L1) aus 100 solcher Zufallszahlen und 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. Was stellen Sie fest?

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

3. 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.

Mappe1

	A
1	=Zufallszahl()
2	0,63211424
3	0,26944457
4	0,59558935
5	0,42063761
6	0,66712016
7	0,1670165
8	0,92819087
9	0,01517211
10	0,01405633
11	0,98277363
12	0,96238977
13	0,61290336
14	0,79777541
15	0,36290233
16	0,42714205
17	0,28553277
18	0,66890268
19	0,61410421
20	0,43537746

STATION 4

Theorie: Exponentialverteilung

Satz: Wenn auf einer Zahlengeraden Zufallszahlen gleichmäßig so verteilt sind, 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 $(1 - x)^{100}$.

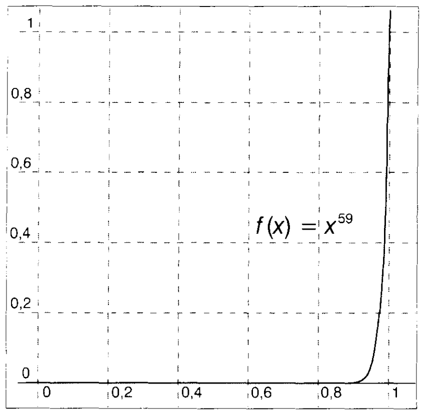
Schritt 2: Nun lässt man 200 Zufallszahlen auf das Intervall [0; 2] „fallen". Begründen Sie:
Die Wahrscheinlichkeit, dass im Intervall $[a; a + x]$ keine Zufallszahl liegt, ist nun $\left(1 - \frac{x}{2}\right)^{2 \cdot 100}$ …
und wenn man $n \cdot 100$ Zufallszahlen auf das Intervall $[0; n]$ fallen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

Schritt 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: $P(0 \leq X \leq x) = 1 - e^{-100x}$

Zur Erinnerung an e^x: Erklären Sie folgende Aussage durch ein Zahlenbeispiel mit einem Taschenrechner: 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 den Faktor x, sondern um den Faktor e^x anwachsen. $\left(1 + \frac{x}{n}\right)^n \xrightarrow[n \to \infty]{} e^x$.

STATION 5

Theorie: Geburtstage (Mehrfachbelegung)



Greifen Sie einen der 60 Schüler (wir nennen ihn „Tim") heraus. Begründen Sie folgende Aussagen:

1. Die Wahrscheinlichkeit, dass „Tims Wartezeit auf den nächsten Geburtstag"
   a. den Wert 0 hat, ist $P(X = 0) = 1 - \left(\frac{364}{365}\right)^{59} \approx 0{,}0765$. („Trefferwahrscheinlichkeit" p)
   b. den Wert 1 hat, ist $P(X = 1) = \left(\frac{364}{365}\right)^{59} - \left(\frac{363}{365}\right)^{59}$.
   c. den Wert d hat, ist $P(X = d) = \left(\frac{365 - d}{365}\right)^{59} - \left(\frac{365 - d - 1}{365}\right)^{59}$.
   d. den Wert 364 hat, ist $P(X = 365) = \left(\frac{1}{365}\right)^{59}$.
   e. Erläutern Sie, wie man an dem Graphen $f(x) = x^{59}$ die Wahrscheinlichkeiten aus Teil a) bis d) ablesen könnte. Wie würden sich diese Wahrscheinlichkeiten ändern, wenn man statt 60 Schülern 30 (100) in der Stufe hätte?

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

2. Erläutern Sie die folgende Termumformung und begründen Sie, dass die Abstände zwischen den Geburtstagen näherungsweise exponentialverteilt sind mit dem Parameter $\lambda = \frac{59}{365}$.
   $P(X \geq d) = \left(\frac{365 - d}{365}\right)^{59} = \left(1 - \frac{\frac{59 \cdot d}{365}}{59}\right)^{59} \approx e^{-\frac{59}{365}d} = e^{-\lambda \cdot d}$, $d = 0, 1, 2, \ldots$

4. Vergleichen Sie die exakten Wahrscheinlichkeiten $P(X = d)$ mit den Näherungswerten $P(d - 1 \leq X < d) = \int_{d-1}^{d} \lambda \cdot e^{-\lambda \cdot x}\, dx = \left[-e^{-\lambda x}\right]_{d-1}^{d}$, $d = 0, 1, 2, \ldots$, die die Exponentialverteilung mit Parameter $\lambda = \frac{59}{365}$ liefern würde.

6 Theorie: Schränke (keine Mehrfachbelegung)

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

a. $P(X \geq 2) = \frac{\binom{365-1}{59}}{\binom{364}{59}} = \frac{363 \cdot 362 \cdot \ldots \cdot 305}{364 \cdot 363 \cdot \ldots \cdot 306} = \frac{305}{364}$, **b.** $P(X \geq 3) = \frac{\binom{364-2}{59}}{\binom{364}{59}} = \frac{362 \cdot 361 \cdot \ldots \cdot 304}{364 \cdot 363 \cdot \ldots \cdot 306} = \frac{305}{364} \cdot \frac{304}{363}$

c. $P(X \geq d) = \frac{\binom{365-d+1}{59}}{\binom{364}{59}} = \frac{305}{364} \cdot \frac{304}{363} \cdot \ldots \cdot \frac{307-d}{366-d}$ **d.** $P(X \geq 306) = \frac{\binom{59}{59}}{\binom{364}{59}} = \frac{59 \cdot 58 \cdot \ldots \cdot 1}{364 \cdot 363 \cdot \ldots \cdot 306} = \frac{305}{364} \cdot \frac{304}{363} \cdot \ldots \cdot \frac{1}{60}$

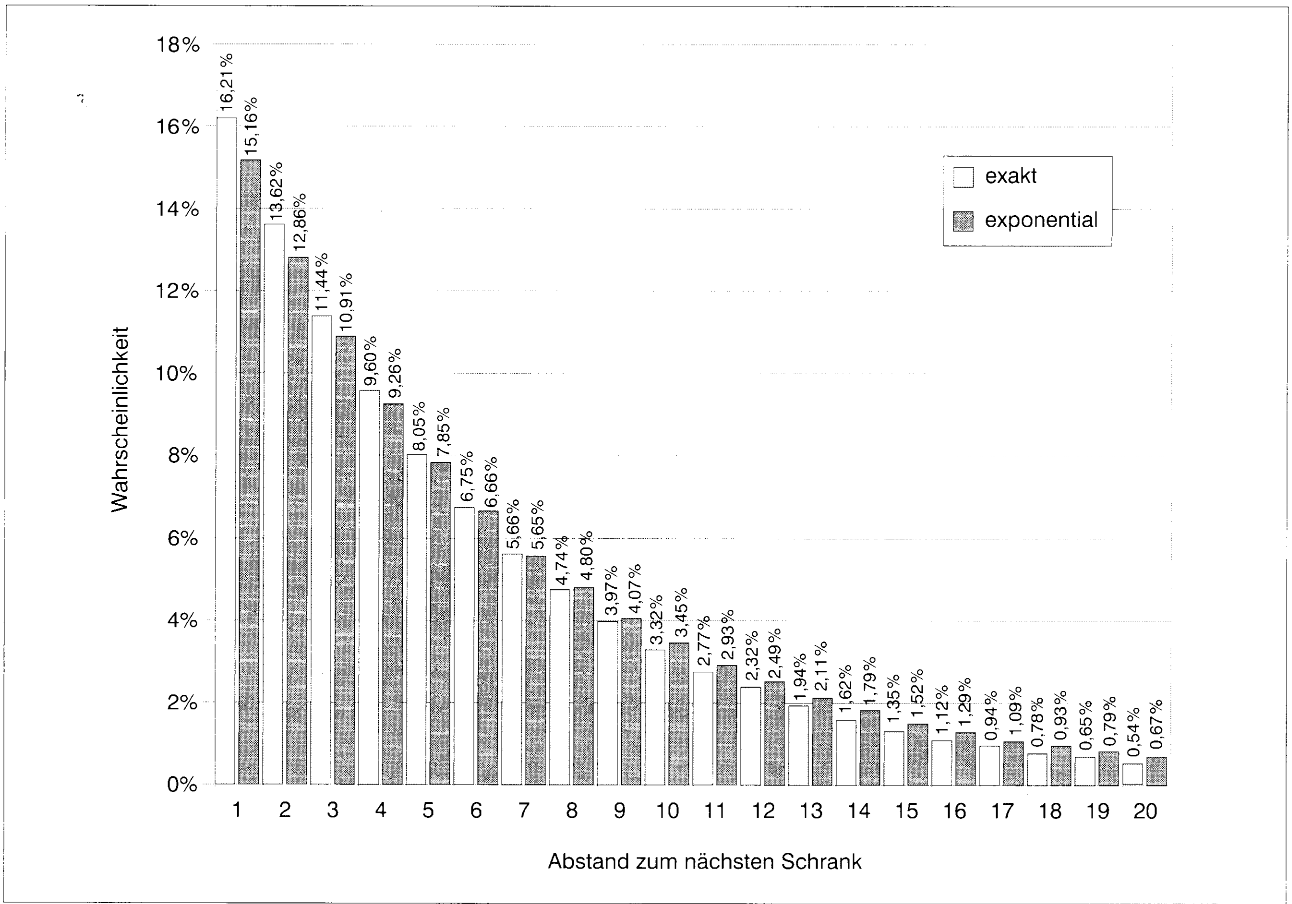
Insbesondere gilt $P(X = 1) = \frac{59}{364}$ („Trefferwahrscheinlichkeit“ p)

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

3. Zum Weiterdenken und Weitersuchen...
Im Netz gibt es unzählige Lottostatistiken. Auf die Idee, die Abstände zwischen den Lottozahlen auszuwerten, kam bisher noch niemand. Untersuchen Sie, was dabei herauskommen müsste und vergleichen Sie Ihre Erwartungen mit einer eigenen Datenerhebung.

Lottozahlen

Samstag	21.03.2009	10	13	21	23	34	38
Mittwoch	18.03.2009	4	17	29	37	42	47
Samstag	14.03.2009	5	8	26	44	45	49



Vergleich der exakten Lösung mit der Näherung