Die Exponentialverteilung, Integralrechnung besucht die Stochastik – wie Tabellenkalkulation beim Modellieren helfen kann.

Wolfgang Riemer, Köln

Abstract:

Meist denkt man bei der Wahrscheinlichkeitsrechnung zunächst an Pfad- und Summenregel. Die folgenden Zeilen wollen dazu einladen, auch einmal direkt von der Analysis in die Stochastik einzutreten Der Einsatzbereich der Integralrechnung wird erheblich erweitert: Integrale erhalten die Bedeutung von Wahrscheinlichkeiten.

1 Integrale in der Wahrscheinlichkeitsrechnung: Wahrscheinlichkeitsdichte

Es regnet auf einen runden Gartentisch mit Radius 10 (dm). Fig. 1 vermittelt einen Eindruck von den auf dem Tischtuch gleichmäßig verteilten Tropfen. Wie sich die Abstände R der Regentropfen zum Tischmittelpunkt verteilen, zeigt Fig. 2. Große Abstände scheinen wahrscheinlicher als kleine.

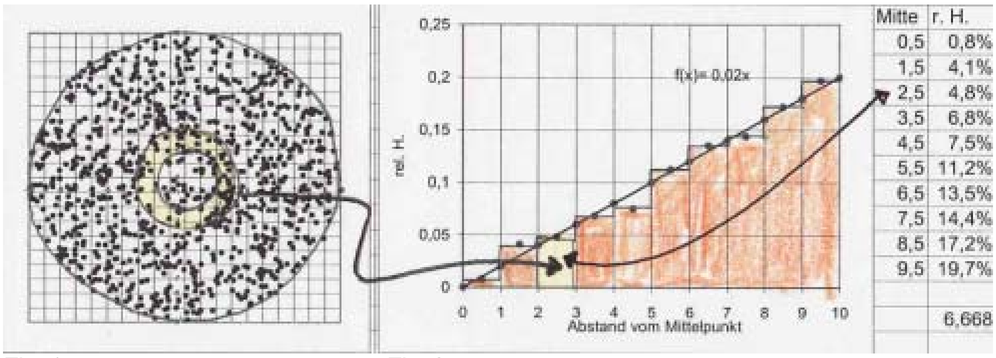


Fig. 1 Fig. 2

Das kann man auch anschaulich begründen: Wenn es gleichmäßig regnet, entspricht die Wahrscheinlichkeit, dass der Abstand R (das ist eine reellwertige Zufallsgröße) im Intervall [r-0,5;r+0,5] liegt, dem Flächenanteil des Kreisringes (mit Länge $\approx 2\pi r$ und Breite 1) an der gesamten Kreisfläche $r^2\pi = 100\pi$, es gilt $P(r-0{,}5 \le R \le r+0{,}5) \approx \frac{2\pi r \cdot 1}{10^2\pi} = \frac{1}{50} \cdot r$.

Diese Wahrscheinlichkeit nimmt proportional mit dem Radius r zu.

Im Säulendiagramm von Fig. 2 mit Säulenbreite 1 erkennt man, dass diese Wahrscheinlichkeit als Rechteckfläche zu dem Graphen der Funktion mit $f(x) = \frac{1}{50}x$ gedeutet werden kann. Die Wahrscheinlichkeit, dass der Abstand R in einem beliebigen Intervall [a;b] liegt, ergibt sich dann als Integral der Funktion f(x)=1/50 x. In der Tat gilt (genau, nicht nur gerundet)

$$P(a \le R \le b) = \frac{\mathit{Kreisringfläche}}{\mathit{Kreisfläche}} = \frac{b^2\pi - a^2\pi}{10^2\pi} = \frac{1}{100}(b^2 - a^2) = \int_a^b \frac{1}{50} x\,dx = \int_a^b f(x)\,dx$$

Man beachte: $f(r) = \frac{1}{50}r$ hat in Fig. 2 nicht die Bedeutung der „Wahrscheinlichkeit des Abstandes r“. Tatsächlich ist diese Wahrscheinlichkeit Null. Denn zum Abstand r gehört die Kreislinie mit Radius r, und die besitzt die Fläche Null.

Man bezeichnet eine Funktionen f, aus der sich Wahrscheinlichkeiten durch Integration ergeben, als **Wahrscheinlichkeitsdichte**, denn es gilt

$$f(x) \approx \frac{1}{d}\int_{x}^{x+d} f(x)dx = \frac{\textit{Wahrscheinlichkeit}\ \text{des Intervalls } [x; x+d]}{\textit{Intervallbreite}\ \text{d}}$$

Und genau so wie „Autos je Stunde“ „Verkehrsdichte“ heißt, nennt man „Wahrscheinlichkeit je Länge" Wahrscheinlichkeitsdichte. Da die Wahrscheinlichkeit aller Ergebnisse zusammen 1 (100%) ergeben muss, gilt für jede Wahrscheinlichkeitsdichte $\int f(x)dx = 1$ in obigen Beispiel: $\int_0^{10} \frac{1}{50}x\,dx = 1$.

2 Mit Wahrscheinlichkeitsdichten „kontinuierliche“ Realität modellieren

Spannend wird Mathematik, wenn „reale“ (im Gegensatz zu simulierter) Realität in den Unterricht einzieht, wenn Daten gesammelt werden und zur Theoriebildung anregen oder als Prüfstein für zuvor aufgestellte „theoretische Modelle“ herangezogen, also mit Theorie in verglichen werden. So ist Datensammeln nicht nur Selbstzweck „weil halt Beschreibende Statistik im Lehrplan steht“. .Der Unterricht gewinnt „wissenschaftspropädeutische Tiefe“. Mit dieser Perspektive wurden Schüler mit folgenden Erkundungen beauftragt:

Arbeitsauftrag 1: „Wie lange leben Seifenblasen ?“

Vom obersten Stockwerk der Schule werden „Pustefix-Seifenblasen“ in die Freiheit „entlassen“. Stoppt die Lebensdauer, d. h. die Zeit (in Sekunden) bis zu ihrem Zerplatzen in der Luft. Wenn Seifenblasen keines „natürlichen Todes sterben“, weil sie an der Wand oder auf dem Boden zerplatzen, zählt die Messung nicht. Arbeitet in Zweierteams, wobei ihr abwechselnd pustet und protokolliert.
(Fig. 1 zeigt, was dabei herausgeommen ist)

Arbeitsauftrag 2: „Wie hoch sind Supermarkt-Rechnungen?

Zum Leidwesen ordnungsliebender Supermarkt-Filialleiter ist trotz der Papierkörbe der Fußboden im Bereich der Einkaufswagen von achtlos weggeworfenen Kassenzetteln übersäht. Sammelt über mehrere Tage diese Einkaufszettel und protokolliert die Höhe der Aufgedruckten Rechnungsbeträge.
Arbeitsteilige Einzelarbeit über eine Woche, wobei sich jeder einen anderen Supermarkt vornimmt, die Auswertung kann gemeinsam erfolgen. Fig. 4 zeigt ein Auswertungsbeispiel. Auch die Auswertung der Anzahl gekaufter Produkte ist möglich)

Arbeitsauftrag 3: „Wie lange dauern Telefongepräche?

Diesen Arbeitsauftrag kann übernehmen, wer Telefongespäch-Einzelabrechnungen bekommt oder solche Abrechnungen (unter Zusicherung strenger Anonymität) besprgen kann. Auf solchen Einzelabrechnungen werden stets die Gesprächsdauern vermerkt. Mühsamer ist es, nach jedem Gespräch die auf dem Display angezeigte Dauer zu protokollieren. Einzelarbeit nur für für besonders sorgfältige/engagierte Mitarbeiter!

Was in einem 4-Personen Privathaushalt mit Festnetz-Einzelabrechnung herauskam, zeigt Fig. 5.

Arbeitsauftrag 4: „Unter welchen Hausnummer wohnst du?
In dieser Arbeitsgruppe durchsucht ihr eure privaten Adressbücher, oder Seiten eines Telefonbuchs in eurem Heimatort - und ihr notiert so viele Hausnummern wie ihr in einer Stunde in den Computer eintippen könnt. Noch eleganter ist es, in einer Telefonbuch-CD alle Hausnummern in ausgewählten Orten herauszukopieren. Arbeitsform: arbeitsteilige Kleingruppen, wobei man Telefonbuchseiten bzw. Orte der Adress-CD aufteilt. Fig. 6 zeigt beispielsweise die Ergebnisse einer bundesweiten Stichprobe mit 1000 Adressen.

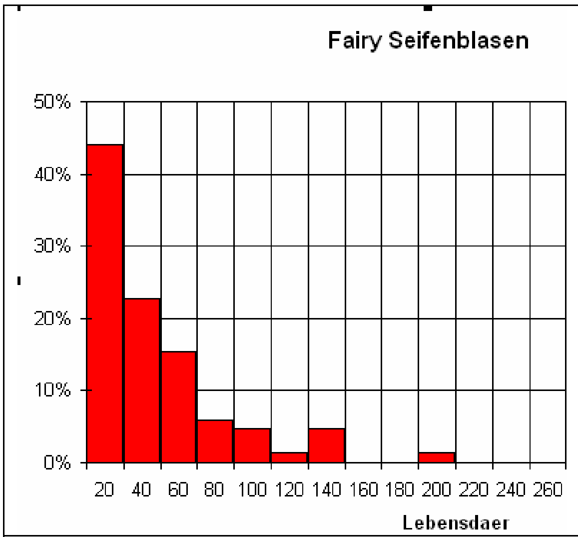


Fig. 3 Lebensdauer von Seifenblasen

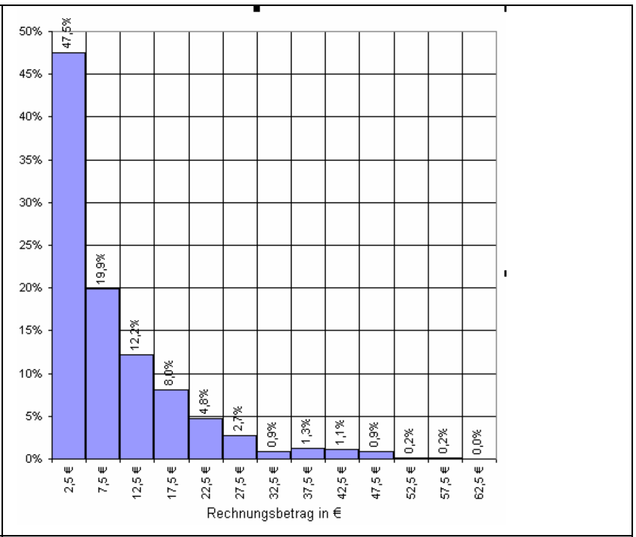


Fig. 4 Supermarkt-Rechnungsbeträge

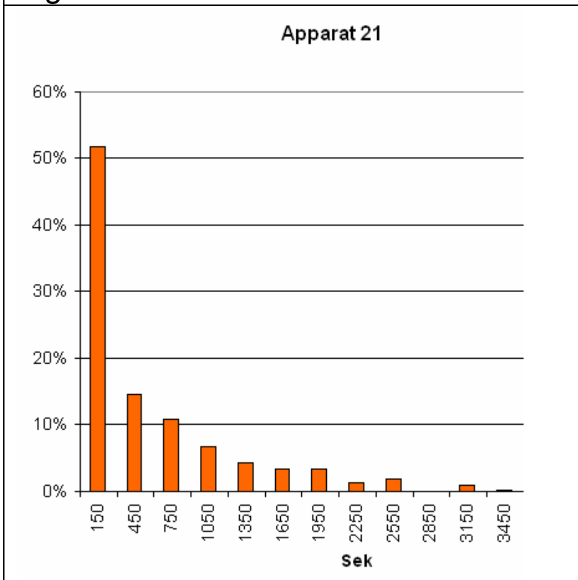


Fig. 5 Telefon-Gesprächsdauer

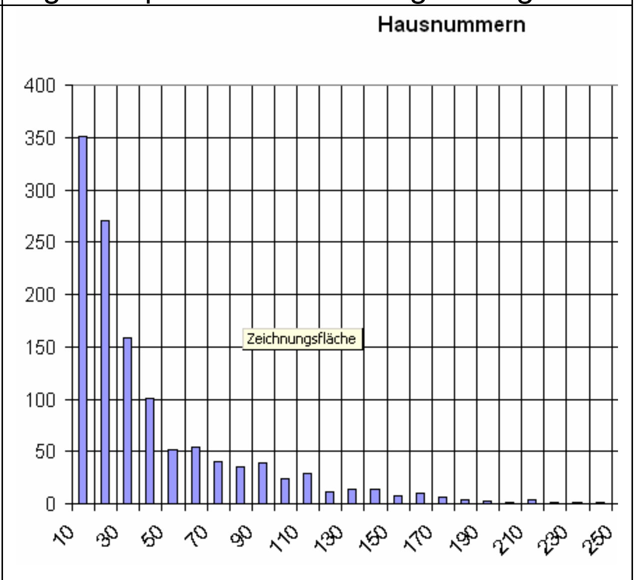


Fig. 6 Hausnummern

Die Diagramme sehen alle so aus, als hätten sie etwas mit einer „fallenden Exponentialfunktion“ zu tun

3 Ein Modell finden

Fallende Exponentialfunktion lassen sich beschreiben durch

$f_a(x) = k \cdot a^{\cdot x}$ mit 0<a<1 bzw. $f_\lambda(x) = k \cdot e^{-\lambda \cdot x}$ mit λ>0.

Wie sind die Parameter zu wählen, wenn f die „zugehörige“ Wahrscheinlichkeitsdichte sein soll?

a) Zunächst muss wegen $1 = \int_0^\infty k \cdot e^{-\lambda \cdot t} dt = \left[-\frac{k}{\lambda} e^{-\lambda \cdot t} \right]_o^x = \frac{k}{\lambda}$ gelten k=λ also $f_\lambda(x) = \lambda \cdot e^{-\lambda \cdot x}$.

b) Dann muss λ so „justiert“ werden, dass - z. B. im Seifenblasenbeispiel die Intervallwahrscheinlichkeiten

$$P(0 \le X \le 10) = \int_0^{10} \lambda \cdot e^{-\lambda \cdot t} dt = [-e^{-\lambda x}]_0^{10} = 1 - e^{-10\lambda}, \quad P(10 \le X \le 20) = \int_{10}^{20} \lambda \cdot e^{-\lambda \cdot t} dt = e^{-10\lambda} - e^{-20\lambda} \ \ldots$$

u. s. w. optimal zu den gemessenen relativen Häufigkeiten 45%, 22% der Lebensdauerintervalle passen. Das erledigt man (ohne Theoriekenntnisse) durch ausprobieren - am besten in Teilgruppen, wobei jede Gruppe einen anderen Parameterwert λ durchrechnet. Oder man verwendet das Kalkulationsblatt zur Visualisierung der Exponentialverteilung aus Fig. 7.

	A	B	C
1			
2			0,03
3	rechte Klassen Grenze	Klassen Mitte	Lambda^ Wahrschk.v
4			
5	20	10	45,12%
6	40	30	24,76%
7	60	50	13,59%
8	80	70	7,46%
9	100	90	4,09%
10	120	110	2,25%
11	140	130	1,23%
12	160	150	0,68%
13	180	170	0,37%
14	200	190	0,20%

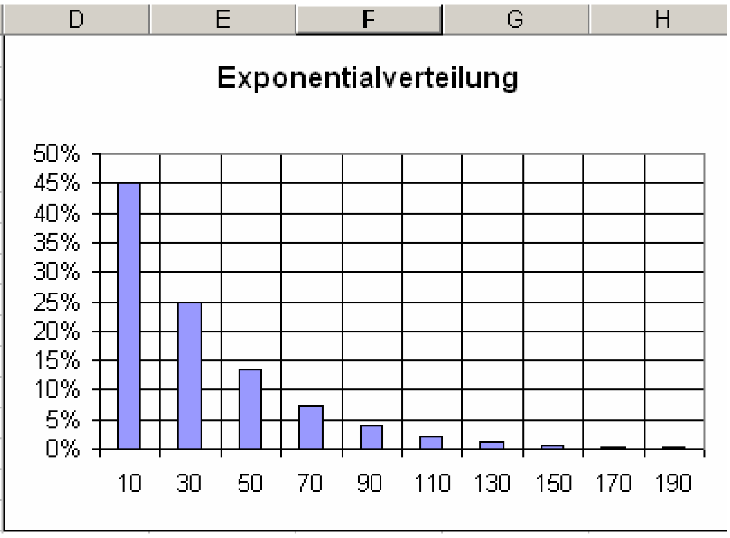
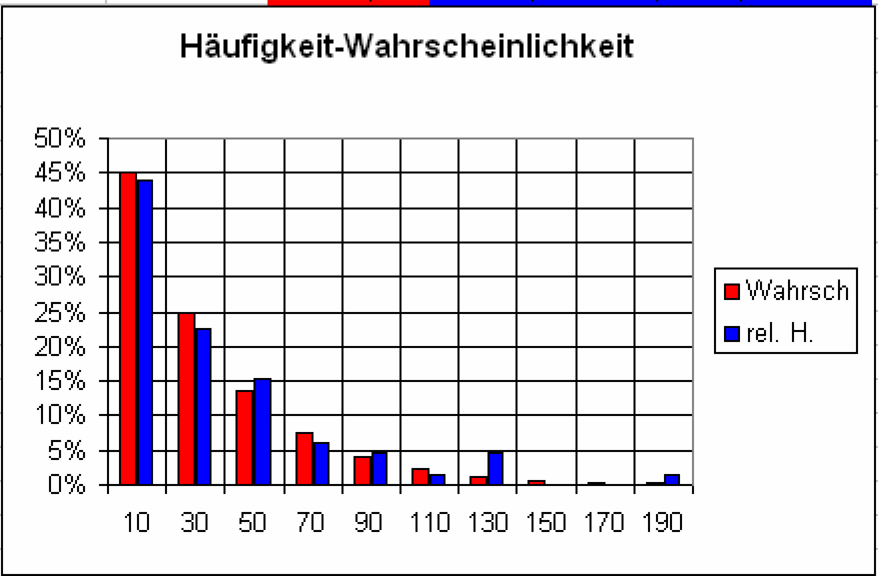


Fig. 7 Kalkulationsblatt Exponentialverteilung mit Säulendiagramm, hier für $\lambda = 0{,}03$. In Zelle C6 der Befehl =EXP(-C$2*A5)-EXP(-C$2*A$6) oder gleichwertig =EXPONVERT(A6,C$2,WAHR)- EXPONVERT(A5,C$2,WAHR),der nach unten kopiert werden kann.

Wenn man die (z. B. durch einen λ-Schieberegler) veränderbaren Wahrscheinlichkeiten und die „gemessenen“ relativen Häufigkeiten nebeneinander stellt (Fig. 8), gelingt die Anpassung besonders eindrucksvoll. Wichtig ist, dass durch das Hilsmittel „Tabelllenkalkulation die Wahrscheinlichkeitsdichten ihren scheinbare Unhandlichkeit verlieren.

Ergebnis: Das Modell der Exponentialverteilung passt hervorragend zu den gemessenen Lebensdauern. Die Lebensdauern von Seifenblasen sind exponentialverteilt. Das ist für die Lebensdauer von Industrieprodukten wie Glühbirnen, Kühlschränken, Autos bekannt, für Seifenblasen hatte das wohl bisher noch niemand zuvor untersucht, ebenso wenig wie für Aldi-Kassenzettel oder Hausnummern.

	A	B	C	D	E	F	G
1				37,33 <- Mittelwert			$\bar{x}$
2		Lambda ->	0,03	0,03 <- 1/Mittelwert			$1/\bar{x}$
3	rechte Klassen Grenze	Klassen Mitte	Wahrsch.	rel. H	Daniel	Marc	Natalie
4							
5	20	10	45,12%	44,00%	24	3,64	11
6	40	30	24,76%	22,67%	25	11,12	6
7	60	50	13,59%	15,33%	102	59,39	3
8	80	70	7,46%	6,00%	15	5,58	21
9	100	90	4,09%	4,67%	140	12,69	18
10	120	110	2,25%	1,33%	88	46,21	26
11	140	130	1,23%	4,67%	8	37,18	124
12	160	150	0,68%	0,00%	197	6,59	18
13	180	170	0,37%	0,00%	137	19,11	42
14	200	190	0,20%	1,33%	180,2	43,24	23



((Spalten Daniel Marc Natalie weglassen))

Fig. 8

Wenn man in Fig. 7 das Augenmerk auf den optimalen Parameter λ der Exponentialverteilung und den Mittelwert $\bar{x}$ der experimentellen Daten (hier die mittlere Lebensdauer der Seifenblasen) legt, entdeckt man (hier wie auch in den anderen Beispielen), dass Wahrscheinlichkeiten und relative Häufigkeiten besonders gut zusammenpassen, wenn gilt $\lambda = \frac{1}{\bar{x}}$. Das kann kein Zufall sein…

4 Mit der Wahrscheinlichkeitsdichte den Mittelwert vorhersagen
(mit dem Mittelwert den Modellparameter finden)
Um den Zusammenhang zwischen der Wahrscheinlichkeitsdichte und dem Mittelwert der zugehörigen experimentellen Daten zu klären, greifen wir auf das Eingangsbeispiel mit den Regentropfen zurück.
Der mittlere Abstand $\bar{r} = \frac{1}{1000}(r_1 + r_2 + \ldots + r_{1000})$ der 1000 Regentropfen vom Ursprung beträgt in Fig. 1. 6,625 (dm). Diesen **Mittelwert** kann man vorab mit Hilfe der Wahrscheinlichkeitsdichte (bis auf Zufallsschwankungen) durch das Integral

$$\mu = \int_0^{10} x \cdot f(x)dx = \int_0^{10} x \cdot \frac{1}{50} x dx = \left[\frac{1}{150} x^3\right]_0^{10} = 6\frac{2}{3}$$ **vorhersagen**.

Deswegen heißt $\mu = \int_0^{10} x \cdot f(x)dx$ **erwarteter Mittelwert** oder kurz **Erwartungswert**.

Begründung: Die Regentropfen, die z. B. im dritten Kreisring 2<R<3 (mit Mittellinie bei $m_3 = 2{,}5$ und relativer Häufigkeit h_3=4,5%) liegen, liefern zum Mittelwert $\bar{r}$ den Beitrag $\approx m_3 \cdot h_3 = 2{,}5 \cdot 0{,}045$. Entsprechendes gilt für die anderen Kreisringe. Man erkennt, dass man den Mittelwert über die Häufigkeitsverteilung näherungsweise berechnen kann als: $\bar{r} \approx m_1 \cdot h_1 + \ldots + m_{10} \cdot h_{10} = 0{,}5 \cdot 0{,}008 + \ldots + 9{,}5 \cdot 0{,}197$. Diese Summe kann man als eine Rechtecknäherung zum Integral $\int_0^{10} x \cdot f(x)dx$ deuten. Man erkennt allgemein:

> Bei allen Zufallsgrößen X mit Wahrscheinlichkeitsdichte f liegt der Mittelwert $\bar{x}$ einer Stichprobe in der Nähe des Erwartungswertes $\mu = \int_0^{10} x \cdot f(x)dx$. Die Abweichungen sind zufallsbedingt und werden umso kleiner, je größer der Stichprobenumfang ist.

Und nun zurück zur Exponentialverteilung: Der Erwartungswert (also der erwartete Mittelwert) entpuppt sich als $\mu = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \left[-\frac{\lambda x + 1}{\lambda} \cdot e^{-\lambda x}\right]_0^{\infty} = \frac{1}{\lambda}$.

Damit ist die Beobachtung begründet, dass bei exponentialverteilten Zufallsgrößen der Kehrwert des Mittelwertes in der Nähe des Parameters λ liegen muss. Man kann also tatsächlich λ aus den Daten schätzen, und muss nicht mehr mit Excel „probieren".

Notiz zum Weiterdenken:
Bei exponentialverteilten Zufallsgrößen ist die Standardabweichung
$\sigma = \sqrt{\int_0^{\infty} (x-\mu)^2 \cdot f(x)dx} = \sqrt{\int_0^{\infty} (x-\mu)^2 \cdot \lambda \cdot e^{-\lambda x} dx} = \frac{1}{\lambda}$ genau so groß wie der Erwartungswert. Wenn Mittelwert und empirische Standardabweichung bei experimentellen Daten nahe

beieinander liegen, ist das ein weiteres Indiz für das Vorliegen einer Exponentialverteilung.

5 Simulationen mit dem GTR
oder: wie Integrale auch noch in die Wahrscheinlichkeitsrechnung kommen können

Wem das Experimentieren im Mathematikunterricht „suspekt“ oder die Überlegung zum Erwartungswert zu mühsam zu zeitaufwändig ist, der kann

a) sich der Thematik „Integrale in der Stochastik“ auch über Simulationen mit dem GTR nähern - oder
b) die Tatsache dass der Erwartungswert μ den Mittelwert $\overline{x}$ (die Standardabweichung σ die empirische Standardabweichung s) vorhersagt, auch wie folgt absichern:

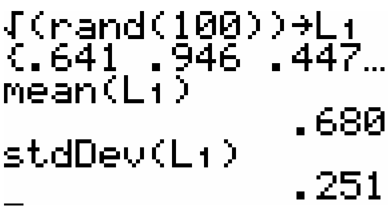
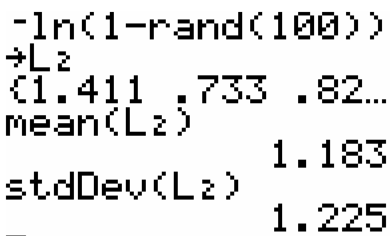
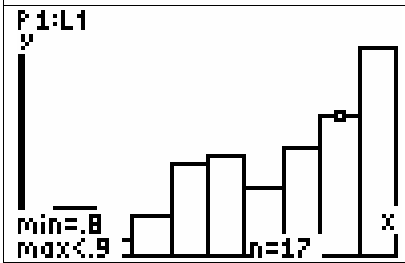
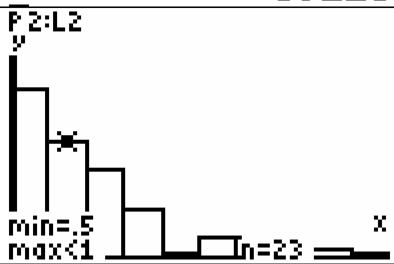
Jeder Taschenrechner erzeugt gleichverteilte Zufallszahlen X über „rand“ zwischen 0 und 1. Die Wahrscheinlichkeit, dass X zwischen a und b liegt ist $P(a \le X \le b) = b - a = \int_a^b 1dx$.

Diese Zufallszahlen haben also eine Wahrscheinlichkeitsverteilung mit der Dichte f(x)=1 über [0;1]. Wie sieht es aus, wenn man auf die Zufallszahlen eine Funktion anwendet, also aus den im Intevall (0;1) gleichverteilten Zahlen $x_1, x_2, ..., x_n$ die Zahlen

(1) Y mit $y_k = \sqrt{x_k}$, wobei gilt $0<y_k<1$

(2) Z mit $z_k = -\ln(1 - x_k)$ wobei gilt $0<z_k$

Wie eine Liste aus 100 solcher Zufallszahlen samt Mittelwert und Standardabweichung mit dem TI 82 erzeugt wird - und wie man einen ersten Eindruck von der Verteilung erhält, zeigen Fig. 9 und 10..

<table>
<tr><td>Zufallsgröße $Y = \sqrt{rand}$
</td><td>Zufallsgröße $Z = -\ln(1 - rand)$
</td></tr>
<tr><td></td><td></td></tr>
<tr><td>Fig. 9: 17 von 100 Y-Werte liegen zwischen 0,8 und 0,9.</td><td>Fig. 10: 23 von 100 Z-Werten liegen zwischen 0,5 und 1
.</td></tr>
</table>

Die relativen Häufigkeiten scheinen in Fig. 9 linear anzusteigen und in Fig. 10 , exponentiell zu fallen. Beides kann man wir folgt begründen:
(1) $Y = \sqrt{X}$ liegt immer dann in [a;b], wenn X in [a²;b²] liegt. Damit ist

$P(a \le X \le b) = b^2 - a^2 = [x^2]_a^b = \int_a^b 2 \cdot x dx$. Die Wahrscheinlichkeitsdichte zu Y ist also linear:

f(x)=2x und die Wahrscheinlichkeit, dass Y zwischen 0,9 und 1 liegt, beträgt .
$\int_{0,8}^{0,9} 2 \cdot x dx = 17\%$, was in Fig. 9 mit der relativen Häufigkeit „17 von 100“ (zufällig genau)

übereinstimmt. Auch der Erwartungswert $\mu = \int_0^1 x \cdot 2x dx = \frac{2}{3}$ stimmt mit dem Mittelwert

0,680, die Standardabweichung $\sigma = \sqrt{(x - \frac{2}{3})^2 \cdot 2x dx} = 0{,}237$ mit der empirischen

Standardabweichung 0,251 gut überein.

(2) $Z = -\ln(X-1)$ liegt immer dann in [a;b], wenn X zwischen $1 - e^{-a}$ und $1 - e^{-b}$ liegt. Die

Wahrscheinlichkeit dafür ist ($P(a \le Z \le b) = (1 - e^{-b})$-($1 - e^{-a}$)= $e^{-a} - e^{-b} = \int_a^b -e^{-x} dx$ Die

Wahrscheinlichkeitsdichte zu Z ist also mit f(x)=e^{-x} tatsächlich exponentiell. Es handelt sich um die Exponentialverteilung mit Parameter λ=1 (also μ=1 und σ=1).
Die Wahrscheinlichkeit, dass Z zwischen 0,5 und 1 liegt, beträgt also $\int_{0,5}^{1} -e^{-x} dx = 23{,}9\%$,

was mit der relativen Häufigkeit „23 von 100“ gut übereinstimmt. Auch der Mittelwert (1,183) und die empirische Standardabweichung (1,225) liegen in der Nähe der theoretischen Werte.

Anmerkung:
Für die Kassenzettelrechnungen und die Hausnummer findet man eine Begründung für das Vorliegen einer Exponentialverteilung in [6] und [7].

Literatur:
Ein komplettes Kurskonzept des Autors zum Einstieg in die Stochastik im Anschluss an die Analysis, in dem auch die Normalverteilung eine zentrale Rolle spielt, erscheint in
[1] Lambacher-Schweizer Baden-Württemberg, Kursstufe, 2008 Klett Verlag.

Seifenblasenexperiment::
[2] Lambacher-Schweizer NRW Klasse 6 (734462), S. 165, Klett Verlag:
[3] W. Riemer, Wie schnell platzen Träume? PM 171, Okt. 2007

[Kassenzettelexperiment
[4] Lambacher-Schweizer NRW Klasse 6 (734462), S. 165 Klett Verlag:

[5] Eichler, W. Riemer, Auf Spurensuche im Supermarkt: Daten, die auf der Erde liegen. Erscheint in: Stochastik in der Schule 2008
[6] Lambacher-Schweizer Stochastik (73243) Klett-Verlag S. 154

Hausnummern-Experiment
[7] Lambacher-Schweizer Stochastik (73243) S. 160f
[8] W. Riemer: Hausnummern: MNU 53/8 S. 491-493
[9] Die verwendeten Excel-Blätter und die Datensätze sind zu finden unter www.riemer-koeln.de.

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

Der Autor unterrichtet am Heinrich-Mann-Gymnasium in Köln und ist Seminarleiter im Studienseminar Köln