THE DERIVE - NEWSLETTER #116

ISSN 1990-7079

THE BULLETIN OF THE DERIVE USER GROUP

+ CAS-TI

C o n t e n t s:

December 2019

Professor Michael de Villiers provided in his latest newsletter a couple of useful websites.

Articles (just recently released and older ones) written by Prof. de Villiers:

http://dynamicmathematicslearning.com/homepage4.html

Great News for Dynamic Geometry users:

Sketchpad is free:

You can download and install Sketchpad including its full functionality from

https://www.chartwellyorke.com/sketchpad/x24795.html

Unlock your license using License name and Authorization code as shown below:

License name: **Pythagoras Single User License 1863 KJUJAM**

Authorization code: **BDFBF968**

I tried, it works!!

Read and/or download a book published in the **UNESCO Digital Library:**

"The Influence of computers and informatics on mathematics and its teaching", (139 pages, pdf)

https://unesdoc.unesco.org/ark:/48223/pf0000093772
https://unesdoc.unesco.org/

Visit the Worldwide Center of Mathematics
(A collection of YouTube contributions, absolutely recommendable!!)

https://www.youtube.com/centerofmath

New Articles written by Prof. de Villiers:

Learn about the background of a Card Trick: Why does it work

http://dynamicmathematicslearning.com/math-explanation-math-trick-generalization-LTM24-dec18.pdf

Diagonal Division Ratios in a Quadrilateral

http://dynamicmathematicslearning.com/diagonal-division-ratios-quad.html

Find de Villier's previous newsletters at:

http://frink.machighway.com/~dynamicm/newsletter.html

Dear DUG Members,

Year 2019 comes to an end and I am very happy to have brought DNL#116 to an end in time. This newsletter presents the second parts of Simon Plouffe's and Helmut Heugl's talks given at ACA 2019 in Montreal.

I'd like to thank Prof. Plouffe for his great cooperation answering my questions and providing additional information during preparing his ppt-presentation for this issue. As you can read in Prof. Plouffe's final comment, we can expect a third part containing new and exciting results in his research field.

The information page shows some very useful links to interesting websites. There is a complete book to be downloaded, you can download and install Geometer's Sketchpad for free and I really liked the collection of YouTube-contributions offered in the Worldwide Center of Mathematics. I could imagine that students of our days – used to consume YouTube contributions – are enjoying the short lectures on nearly all fields of mathematics.

Guido Herweyers mentions in a short sentence (page 37) our travel to China. Yes, we were there for three exciting weeks. On the last page I show a small selection of our "few"- ~ 5500 – pictures.

Let's look ahead to another DUG year (it will be year 30 of its existence, which not so bad?).

Best regards and wishes to all of you

Noor and Josef

The *DERIVE-NEWSLETTER* is the Bulletin of the *DERIVE* & CAS-*TI User Group*. It is published at least four times a year with a content of 40 pages minimum. The goals of the *DNL* are to enable the exchange of experiences made with *DERIVE, TI*-CAS and other CAS as well to create a group to discuss the possibilities of new methodical and didactical manners in teaching mathematics.

Editor: Mag. Josef Böhm
D´Lust 1, A-3042 Würmla, Austria
Phone: ++43-(0)660 31 36 365
e-mail: nojo.boehm@pgv.at

Contributions:
Please send all contributions to the Editor. Non-English speakers are encouraged to write their contributions in English to reinforce the international touch of the *DNL*. It must be said, though, that non-English articles will be warmly welcomed nonetheless. Your contributions will be edited but not assessed. By submitting articles, the author gives his consent for reprinting it in the *DNL*. The more contributions you will send, the more lively and richer in contents the *DERIVE* & CAS-*TI Newsletter* will be.

Next issue: March 2020

Preview: Contributions waiting to be published

Some simulations of Random Experiments, J. Böhm, AUT, Lorenz Kopp, GER
Wonderful World of Pedal Curves, J. Böhm, AUT
Simulating a Graphing Calculator in *DERIVE*, J. Böhm, AUT
Cubics, Quartics – Interesting features, T. Koller & J. Böhm, AUT
Logos of Companies as an Inspiration for Math Teaching
Exciting Surfaces in the FAZ
BooleanPlots.mth, P. Schofield, UK
Old traditional examples for a CAS – What´s new? J. Böhm, AUT
Mandelbrot and Newton with *DERIVE*, Roman Hašek, CZ
Tutorials for the NSpireCAS, G. Herweyers, BEL
Dirac Algebra, Clifford Algebra, Vector-Matrix-Extension, D. R. Lunsford, USA
A New Approach to Taylor Series, D. Oertel, GER
Statistics of Shuffling Cards, Charge in a Magnetic Field, H. Ludwig, GER
Selected Lectures from TIME 2016
More Applications of TI-Innovator™ Hub and TI-Innovator™ Rover
Surfaces and their Duals, Cayley Symmetroid, J. Böhm, AUT
Affine Mappings –Treated Systematically, H. Nieder, GER
The Penney-Ante Game, MacDonald Phillips, USA
Hyper Operations for DERIVE, J Angres and others, GER
Investigations of Lottery Game Outcomes, W Pröpper, GER
A Traversing Bridge (Eine Rollbrücke), W. Alvermann, GER
DERIVE Bugs?, D. Welz, GER
Tweening & Morphing with TI-NspireCX-II-T, J. Böhm. AUT

Impressum:
Medieninhaber: *DERIVE* User Group, A-3042 Würmla, D´Lust 1, AUSTRIA
Richtung: Fachzeitschrift
Herausgeber: Mag. Josef Böhm

Lieber Josef,

im DERIVE-Newsletter 115 hat Helmut Heugl einen lesenswerten Aufsatz „Realizing the Concept of ‚Using Representations' Using CAS" veröffentlicht. Darin nennt er als Vorteile des Einsatzes digitaler Technologie beim Mathematik-Lernen ausschließlich die herausragende Eignung der digitalen Medien zur wichtigen und zentralen mathematischen Tätigkeit des Repräsentationswechsels. Dies kann in einem Aufsatz, der sich laut Überschrift auf Umsetzung des Konzeptes „Using Representations" beschränkt, natürlich nicht anders sein.

Leider ignoriert diese thematische Selbstbeschränkung den Blick auf eine zweite, ebenso wichtige und zentrale mathematische Tätigkeit, die durch den Einsatz digitaler Werkzeuge sehr erleichtert wird: die Mustererkennung. Gerade im Zusammenhang rekursiver Modelle erstellen digitale Werkzeuge große Datenmengen, denen ein Muster innewohnt und das auf der Basis dieser Datenmengen erkannt werden kann. Die Entdeckung dieses Musters sowie dessen explizite, formale Beschreibung sind ebenso zentrale und wichtige mathematische Tätigkeiten, wie beispielsweise der Repräsentationswechsel und die damit erreichbaren didaktischen Ziele.

Leider kenne ich kein praktisch erprobtes, wissenschaftlich begründetes Konzept zum Einsatz von CAS im Mathematikunterricht. Wenn dieses bereits vorliegt, müssten darin als Vorteile des Einsatzes von CAS bei der Vermittlung von Mathematik unbedingt die Stichwörter ‚Repräsentationswechsel' und ‚Mustererkennung' aufgegriffen werden. Darüber hinaus ist natürlich eine ‚kritische Bewertung auch von CAS-Ergebnissen' wichtiger Bestandteil eines Mathematikunterrichtes mit CAS-Einsatz.

All dies wurde im praktisch realisierten, CAS-einsetzenden Mathematikunterricht zu meiner aktiven Zeit (bis 2009) nicht berücksichtigt. Hat sich da inzwischen etwas Neues ergeben?

Herzliche Grüße

Roland

Dear Josef,

In DNL115 Helmut Heugl provided his worth reading article "Realizing the Concept of ‚Using Representations' Using CAS". He mentions as advantages of the use of digital technologies in math teaching solely their outstanding suitability for the important and central mathematical activity of changing the representation form. This cannot be other in an article which is according to its header restricted on realization of the concept "Using Representations".

Unfortunately, this thematic self-reduction disregards a second, equally important and central mathematical activity, which becomes very much easier by the use of digital tools: recognizing patterns. Particularly in connection with recursive models digital tools produce large amounts of data forming a pattern which can be detected based on these amounts of data. Recognizing this pattern and its explicit and formal description are as important and central mathematic activities as e.g. the change of representation form and the respective mathematical didactic goals.

Unfortunately, I don't know a tried and tested, scientifically justified concept of the use of CAS in mathematics teaching. If there is one, then the keywords "Change of Representation Form" and "Pattern Recognizing" must be included. Additionally a 'critical evaluation of CAS-results' is an important part of CAS-supported math teaching.

All this was not considered in practical realized CAS-supported math teaching in my active time (until 2009). I wonder, if things have changed since then?

Best regards

Roland

Hubert Langlotz, Germany

Dear Josef,

Helmut's problem has kept us busy in Thuringia for a long time. We take a pragmatic approach. (Problem described in Helmut Heugl's contribution in DNL#115)

As you can see in the attachment, which was written by Wilfried Zappe and me for Stark Verlag (Publisher), we try to help students to avoid this mistake providing recommendations as described in the attachment. The tns-file gives one more example what can happen when entering a list as function argument.

Hubert

How to avoid Dynamic Definitions:

Defining derivative functions with parameters you must avoid making the following mistake:

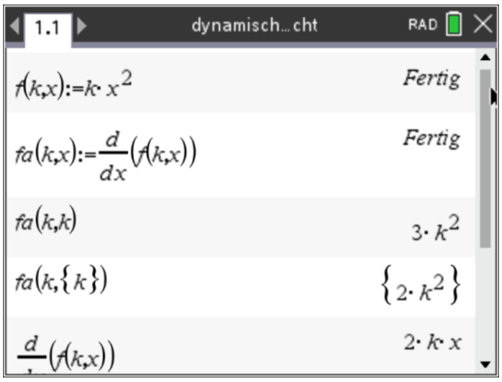
Example: the first derivative of $f(x) = k \cdot x^2$ is $f'(x) = 2k \cdot x$. Then the first derivative $f'(x = k) = 2k^2$.

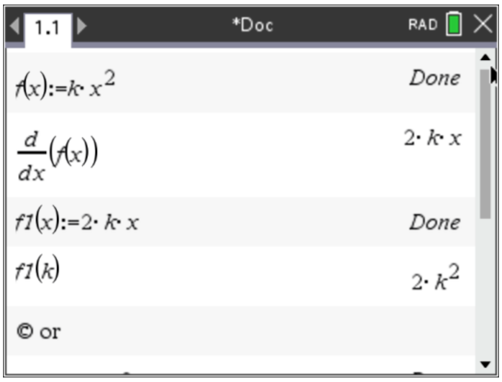
But CAS gives – performing the wrong procedure – $f'(x = k) = 3k^2$. See the first three expressions.

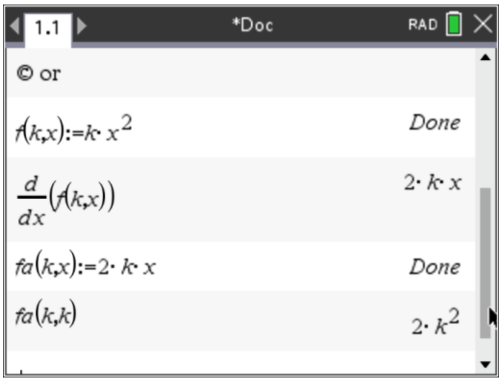
What's the reason?

CAS calculates the function value $fa(k,k) = k^3$ first and then differentiates wrt $k \longrightarrow 3k^2$.

Calculate such "dynamic definitions" by calculating and then storing the derivates as a new function as shown in the example below.

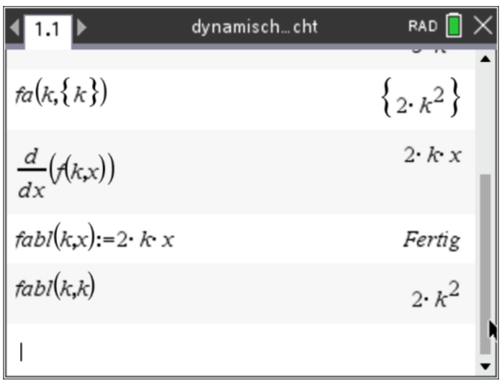






The right screen shows what happens if entering the second parameter as a list.

BTW: DERIVE behaves the same.



You can find a more extended article on this issue at:
https://education.ti.com/sites/DEUTSCHLAND/downloads/pdf/TI_Nachrichten_1_02final.pdf

π, the primes and the Lambert W-function (2)

Simon Plouffe [simon.plouffe@gmail.com]
Université de Nantes(IUT), Nantes, France

This was the last paragraph of part 1:

$$\pi(n) \approx \frac{n}{\ln(n)}$$

is the classic equation; we change it for (see Dusart thesis [2010] for details):

$$\pi(n) \approx \frac{n}{\ln(n) - 1}$$

If we compute the inverse of $\frac{n}{\ln(n)-1} = y$, then we receive (solve(%,n) with y renamed):

$$p_n \approx -n \cdot W_{-1}\left(-\frac{e}{n}\right).$$

With the n^{th} prime we have the formula $p_n \approx n\ln(n)$ and further $\pi(n) \approx \frac{n}{W_0(n)}$.

We proceed with part 2:

Now we will look at what the error looks for

$$p_n \approx -n \cdot W_{-1}\left(-\frac{e}{n}\right).$$

Then, one by one we eliminate different hypothesis about the difference between our calculated p_n and the real value.

- The size of what is left is comparable to one of the <straight lines> of $W_k(n)$.
- At first sight the value of the difference is a straight line (correlation is 0.9999). It is not!
- What is significant is the magnitude of the difference only.
- The $W_k(n)$ and logarithmic fit are indistinguishable.

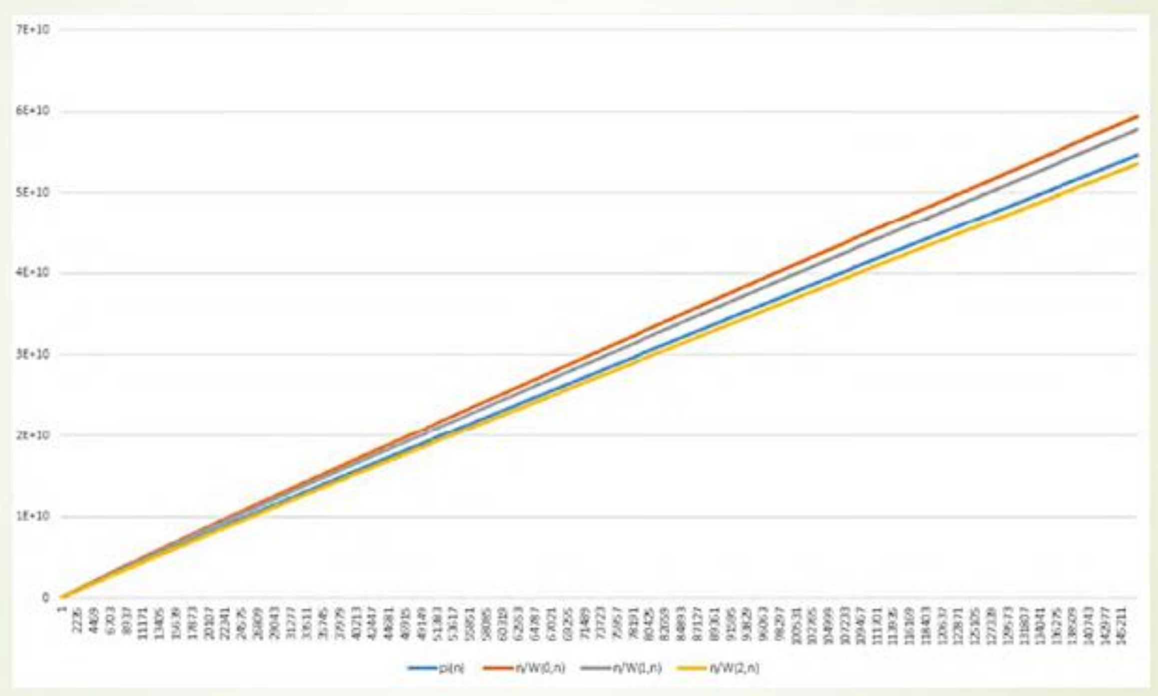
For this comparison we need to consider the extent of the known tables of p_n.

My own table is up to 20 000 billions. The known long-range table is only up to 10^{24}.
(powers of 10 only)

The table for $\pi(n)$ is up to 27 000 billions and the long range goes up to 10^{27}.

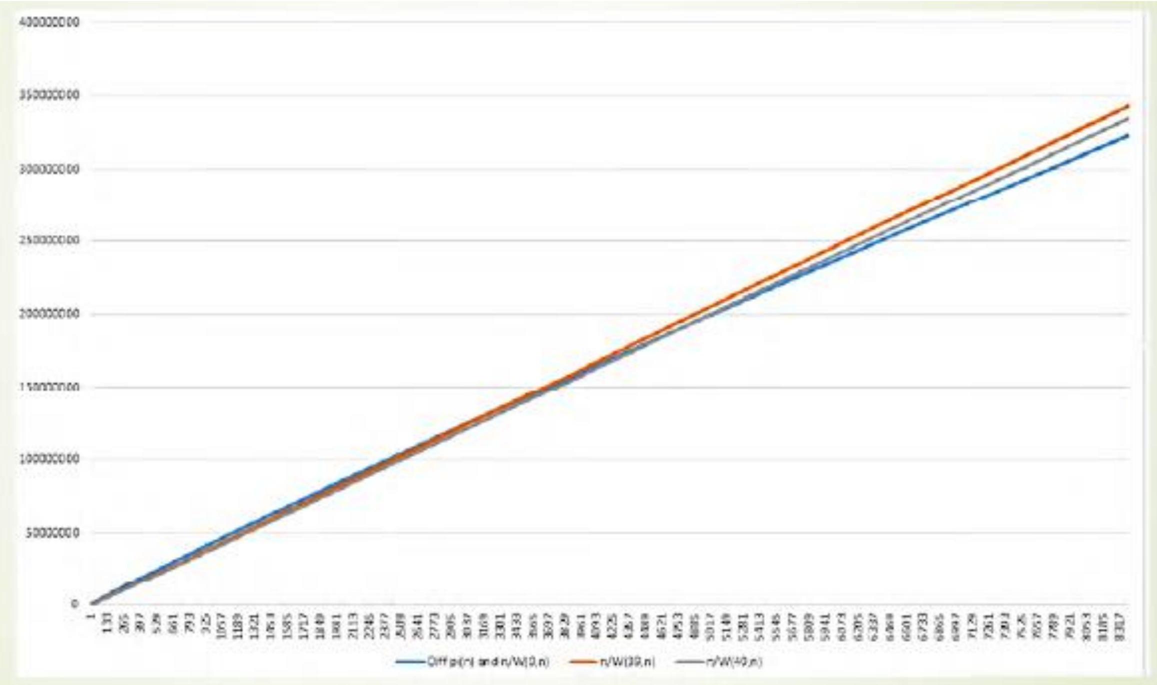
The punctual very large known primes are useless for this study for a very simple reason:
We don't know the rank of these primes.

$$\frac{n}{W_k(n)} \text{ for large values of } n$$



$\pi(n)$ is the straight line in blue between the values of $\dfrac{n}{|W_k(n)|}$.

Since the values lie in between, it is natural to use differences to realize we still have <u>the same</u> thing but smaller.



$$\frac{n}{W_{39}(n)} < \pi(n) - \frac{n}{W_0(n)} < \frac{n}{W_{40}(n)}$$

This is strange.

The known behaviour of p_n and $\pi(n)$ is not exactly predictable, the evaluation of the Riemann with $li(x)$ (integral logarithm []) not simple.

The last computations of p_n and $\pi(n)$ were ‘difficult’ (Months of computer time).

From numerical evidence then

$$p_n \approx -n\, W_{-1}\left(-\frac{e}{n}\right) - \frac{n}{W_0(n)}.$$

The surprise is that: what is left (again) is something that resembles exactly what we had in the first place (!). The ‘curve’ is still a ‘straight line’ but the magnitude is smaller.

The only plausible explanation is that we have here the *Matryoshka Principle*: Russian Puppets.

$$p_n \approx n \ln n$$

$$-n\, W_{-1}\left(-\frac{e}{n}\right) - \frac{n}{W_0(n)}$$

is more precise.

Tested at $n = 10^{24}$: $p_{10^{24}} \approx 58308642550474983476717666$

The real value being: 58310039994836584070534263

Now if we continue with this *matryoshka principle*, what is the next term!

For $p_{10^{24}}$ by using a bisection method to find the next term in the form of

$$\frac{n}{W_k(n)}.$$

The next terms are 114 and 96606.

$W_k(n)$ can be approximated with formula 4.10 in *Corless et al*.

$$W_k(n) \approx \ln(n) + 2\pi k\, i + \ln(\ln(n) + 2\pi k\, i)$$

$$p_{10^{24}} = -10^{24} W_{-1}\left(-\frac{e}{10^{24}}\right) - \frac{10^{24}}{W_0(10^{24})} + \frac{n}{W_{114}(10^{24})} + \frac{n}{W_{96606}(10^{24})} + \ldots$$

The value is 58310039994824799949493554

Compared to: 58310039994836584070534263 (12 exact digits).

With 3 terms: 6 exact digits.

In 1994, *B. Salvy* published a paper to dig out an algorithm to get dozens of terms in the Cipolla formula:

$$p_n \sim n\left(\ln(n) + \ln(\ln(n)) - 1 + \frac{\ln(\ln(n)) - 2}{\ln(n)} - \frac{\ln(\ln(n))^2 - 6\ln(\ln(n)) + 11}{2\ln(n)^2} + \cdots\right)$$

The formula is quite similar to the asymptotic expansion of $W_0(n)$:

$$W_0(n) \approx L_1 - L_2 + \frac{L_2}{L_1} + \frac{L_2(-2+L_2)}{2L_1^2} + \frac{L_2(6-9L_2+2L_2^2)}{6L_1^3} + \frac{L_2(-12+36L_2-22L_2^2+36L_2^3)}{12L_1^4} + \ldots$$

Here $L_1 = \ln(n)$ and $L_2 = \ln(\ln(n))$.

[*] $li(x) = \int_0^x \frac{dt}{\ln t}$

In principle, with enough terms with $n >> 1$ it should do the thing.

Not exactly, with 72 terms of *Cipolla-Salvy* formula we get 12 digits exact, too. There is a limit to it.

The expansion in Lambert functions is much simpler.

For the moment, the only clue I have about the 4th term is that it is proportional to the $[\log_2(n) + 1]$, that is the log of n in base 2.

For the info, here is the 10th term of the *Cipolly-Salvy* expansion:

```
k*ln(k)*(1+(ln(ln(k))-1)/ln(k)+(ln(ln(k))-2)/ln(k)^2+(-1/2*
ln(ln(k))^2+3*ln(ln(k))-11/2)/ln(k)^3+(1/3*ln(ln(k))^3-7/2*
ln(ln(k))^2+14*ln(ln(k))-131/6)/ln(k)^4+(-1/4*ln(ln(k))^4+
23/6*ln(ln(k))^3-49/2*ln(ln(k))^2+159/2*ln(ln(k))-1333/12)/
ln(k)^5+(1/5*ln(ln(k))^5-49/12*ln(ln(k))^4+73/2*ln(ln(k))^3
-367/2*ln(ln(k))^2+3143/6*ln(ln(k))-13589/20)/ln(k)^6+(-1/6
*ln(ln(k))^6+257/60*ln(ln(k))^5-1193/24*ln(ln(k))^4+1027/3*
ln(ln(k))^3-17917/12*ln(ln(k))^2+47053/12*ln(ln(k))-193223/
40)/ln(k)^7+(1/7*ln(ln(k))^7-89/20*ln(ln(k))^6+959/15*ln(ln
(k))^5-13517/24*ln(ln(k))^4+6657/2*ln(ln(k))^3-39769/3*ln(
ln(k))^2+493568/15*ln(ln(k))-32832199/840)/ln(k)^8+(-1/8*ln
(ln(k))^8+643/140*ln(ln(k))^7-14227/180*ln(ln(k))^6+34097/
40*ln(ln(k))^5-76657/12*ln(ln(k))^4+616679/18*ln(ln(k))^3-\
642111/5*ln(ln(k))^2+36780743/120*ln(ln(k))-893591051/2520)
/ln(k)^9+(1/9*ln(ln(k))^9-1321/280*ln(ln(k))^8+119603/1260*
ln(ln(k))^7-218809/180*ln(ln(k))^6+1328803/120*ln(ln(k))^5-\
2696687/36*ln(ln(k))^4+33904723/90*ln(ln(k))^3-40633409/30*
ln(ln(k))^2+7921124011/2520*ln(ln(k))-2995314311/840)/ln(k)
^10)
```

But, let's go back to $\pi(n)$. We had $\pi(n) \approx \dfrac{n}{W_0(n)}$.

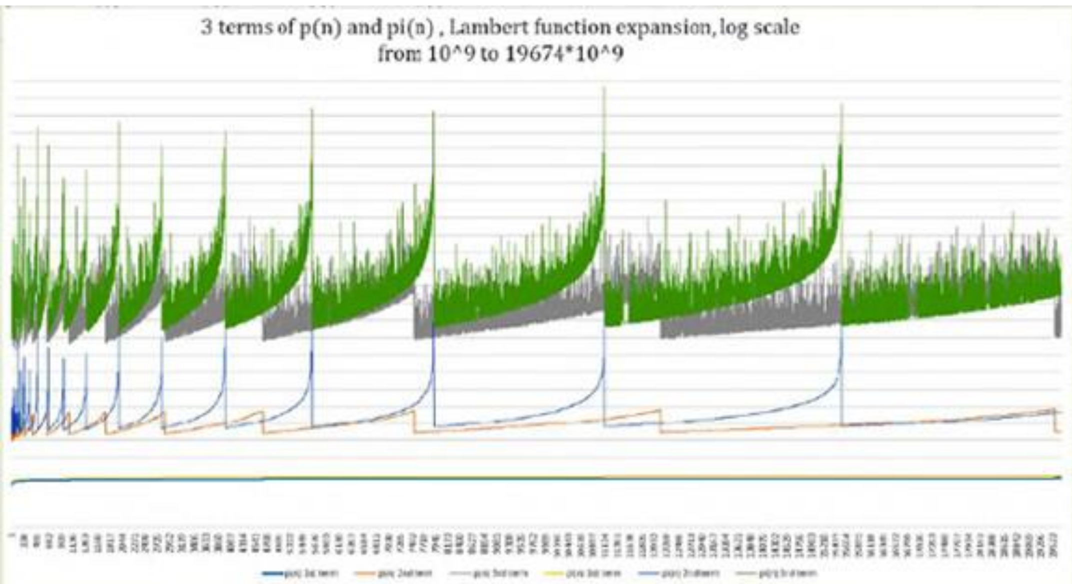
We apply the same scheme let's say for 10^7 for $\pi(n)$ and p_n:

$$\frac{p_{10^7}}{10^7} = -W_{-1}\left(-\frac{e}{10^7}\right) - \frac{1}{W_0(10^7)} + \frac{1}{W_{22}(10^7)} + \frac{1}{W_{763}(10^7)} + \frac{1}{W_{5323546}(10^7)} - \ldots$$

$$\frac{\pi\left(10^7\right)}{10^7} = \frac{1}{W_0(10^7)} - \frac{1}{W_{22}(10^7)} + \frac{1}{W_{640}(10^7)} + \frac{1}{W_{2174463}(10^7)} - \ldots$$

It is similar, how similar is it?

This is the graph of the coefficients of $\pi(n)$ and p_n in the Lambert expansion, every billion from 10^9 to $19674 \cdot 10^9$, log scale.



This is where we can apply the *duck principle*:

> If an animal has a beak lie a duck, feathers like a duck, the color of a duck, quacks like a duck and has two feet like a duck, … then it's a duck.

With this Lambert function expansion, the two quantities $\pi(n)$ and p_n are the same (except for the first term of p_n).

Recently, a certain *André LeClair* and *Guilherme França* (2014) had a formula for the n^{th} zero of Riemann's Zeta function. It follows the same idea. If $N(n)$ is the number of non-trivial zeros (considering only the imaginary part) then

$$N(n) \approx \frac{n}{2\pi}\ln\left(\frac{n}{2\pi}\right) - \frac{n}{2\pi} + \frac{11}{8}.$$

By inverting (functionally) the formula we obtain a formula for the n^{th} zero:

$$\sigma(n) \approx \frac{(8n-11)\pi}{4W\left(\frac{8n-11}{8e}\right)}.$$

The formula is spectacular in precision.

$$\sigma(1) \approx -\frac{3}{4}\frac{\pi}{W\left(-\frac{3}{8e}\right)}$$

is 14.5213469… when the real value is 14.13472514.

So precise that they could evaluate precise values of $\sigma(n)$ with $n = 10^{1000}$ by using an additional Newton like interpolation.

We have here a quantum leap compared to previous models.

Again, if we go back to the classic known equations: $\sigma(n) \approx \dfrac{2\pi n}{\ln(n)}$

and $N(n)$ (Riemann) is $N(n) \approx \dfrac{n}{2\pi}\ln\left(\dfrac{n}{2\pi e}\right)$.

And now by solving for n in each case we get

$$\frac{\sigma(n)}{2\pi n} \approx \frac{1}{W\left(\frac{n}{e}\right)} \qquad \frac{N(n)}{2\pi n} \approx -W_{-1}\left(\frac{-2\pi}{n}\right)$$

$$\frac{p_n}{n} \approx -W_{-1}\left(-\frac{e}{n}\right) - \frac{1}{W_0(n)} + \frac{1}{W_{\{2\ln(n)\}+1}(n)}$$

If we collect the four formulas we found, dividing by either n or $2\pi n$ we get

$\dfrac{\pi(n)}{n} \approx \dfrac{1}{W_0(n)}$	$\dfrac{p_n}{n} \approx -W_{-1}\left(-\dfrac{e}{n}\right)$
$\dfrac{\sigma(n)}{2\pi n} \approx \dfrac{1}{W\left(\frac{n}{e}\right)}$	$\dfrac{N(n)}{2\pi n} \approx -W_{-1}\left(\dfrac{-2\pi}{n}\right)$

$$\pi(n) \cong n\left(\frac{1}{W_0(n)} - \frac{1}{W_k(n)}\cdots\right) \qquad p_n \cong n\left(-W_{-1}\left(-\frac{e}{n}\right) - \frac{1}{W_0(n)}\cdots\right)$$

$$N(n) = 2\pi n\left(-W_{-1}\left(\frac{-2\pi}{n}\right)\cdots\right) \qquad \sigma_n = 2\pi n\left(\frac{1}{W_0\left(\frac{n}{e}\right)} + \frac{1}{W_1(n)}\cdots\right)$$

From there, two possible directions:

If the Euler principle applies then we should have a sum and a product on each side.

Or, the expression with primes needs to be completed with an expression using the zeros of the Zeta function (1/2 + *i t*), then it has to match with the equivalent expression with the primes on the other two equations. (*i* is a counter and *t* is the real part of the zeros)

I leave this question as an exercise to be completed …

Here is a model we can try for $\pi(n)$ and p_n:

$$\frac{\pi(n)}{n} \approx \frac{1}{W_0(n)} - \sum_{k=1}^{\infty} \frac{1}{W_k(n) + e^{kf(n)}}$$

$$p_n \approx -W_{-1}\left(-\frac{e}{n}\right) - \frac{1}{W_{0(n)}} - \sum_{k=1}^{\infty} \frac{1}{W_k(n) + e^{kg(n)}}$$

Here $f(n)$ and $g(n)$ are found using a logarithmic fit. For $f(n)$, at n = 4635000018752, the formula gives 146388867645773 *exactly*.

For $g(n)$, the formula is quite similar.

Comparing the two:

$f(n)$ = 5.1407131338852538860618655508885 + 0.048089483129908800105508959416636 ln(n)

$g(n)$ = 5.1425259035418911897661770856362 + 0.048391979782557290855113082298999 ln(n)

Is this just a coincidence?

Let's try another model:

$$\frac{\pi(n)}{n} \approx \frac{1}{W_0(n)} - \sum_{k=1}^{\infty} \frac{1}{W_k(n) + k^{f(n)}}$$

$$p_n \approx -W_{-1}\left(-\frac{e}{n}\right) - \frac{1}{W_{0(n)}} - \sum_{k=1}^{\infty} \frac{1}{W_k(n) + k^{g(n)}}$$

$f(n)$ = 4.8493349460 + 0.0557287326 ln(n)

$g(n)$ = 4.8604042576 + 0.068197411 ln(n)

I think that there is a possibility that one could find eventually an exact formula for $\pi(n)$, p_n, $N(n)$ and $\sigma(n)$.

And now, something completely different:

This is *Viète* formula (1593):

$$\frac{2}{\pi}=\frac{\sqrt{2}}{2}\cdot\frac{\sqrt{2+\sqrt{2}}}{2}\cdot\frac{\sqrt{2+\sqrt{2+\sqrt{2}}}}{2}+\frac{\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2}}}}}{2}\cdot\ \ldots$$

As far as mathematic is concerned, nobody knows what is the binary expansion of $\frac{2}{\pi}$. This is just a bunch of zeros and ones at random. Perhaps we will never find the pattern in it.

But there is another approach to the problem.

If we are looking at individual bits of this number, we do not see anything.

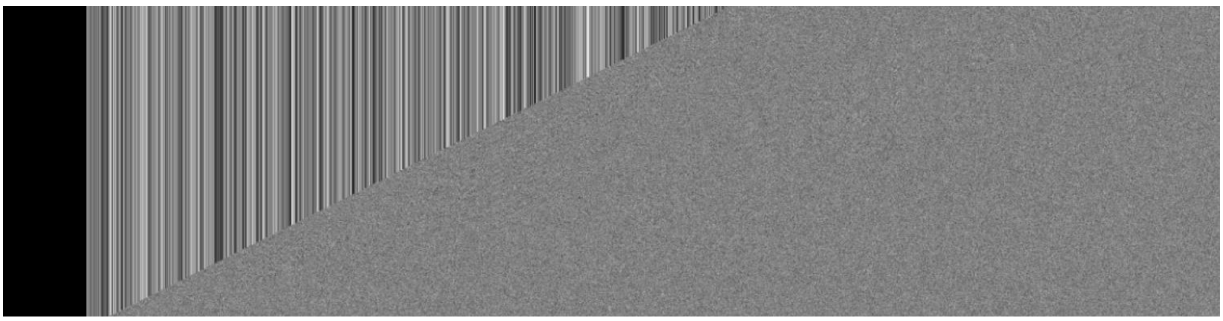
But let's consider this instead:
We align all the partial products in Viète expansion and look at the bits as a whole, everything is computed in binary.

The first line is $\frac{\sqrt{2}}{2}$.

The second line is $\frac{\sqrt{2}}{2}\cdot\frac{\sqrt{2+\sqrt{2}}}{2}$.

The third line is $\frac{\sqrt{2}}{2}\cdot\frac{\sqrt{2+\sqrt{2}}}{2}\cdot\frac{\sqrt{2+\sqrt{2+\sqrt{2}}}}{2}$.

With 99715 bits x 25000 terms gives this:



An image of 2.5 billion pixels.

The long vertical lines at the left are the slow convergence of the product to $\frac{2}{\pi}$.

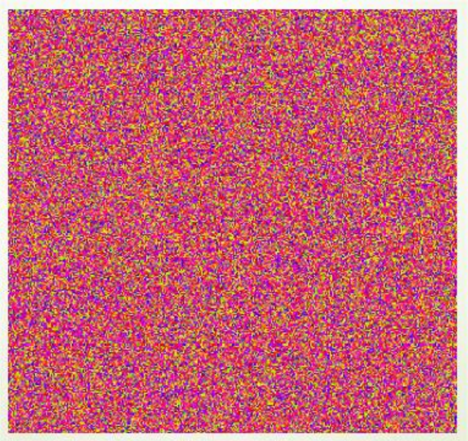
We don't see much of a pattern here.

As a comparison, this is an image of the first 100 million decimal digits of π, colorized with 10 colors, blue, green, red, yellow, …

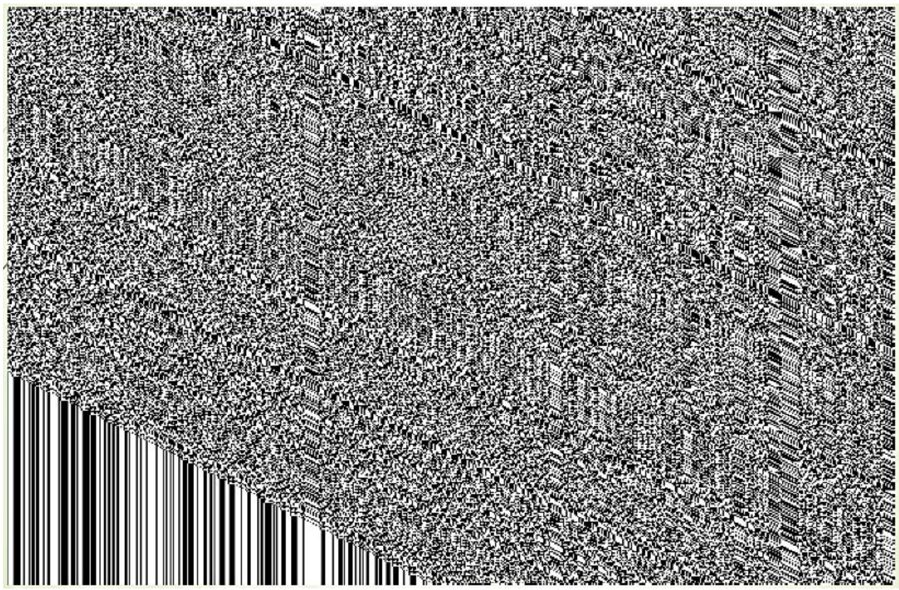
The image is 10000 x 10000.

If we zoom in: this is pretty much random data.

One experiment was done with the first 1000 billion digits of π: 10000 images of 10000 x 10000 (see web page).



If we zoom on Viète image, we see this:



I have no explanation for it.

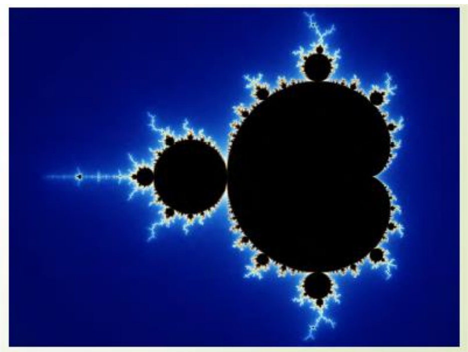
The successive square roots are producing the effect. The thing we can say that:

From one term to the other: bits are not random and the pattern is quite persistent.

Are there any other algebraic curiosities like that?

When experimenting with square roots and square roots of square roots, I wanted to know if a persistent pattern occurs in the Mandelbrot set.

As we know, there are similarities. I wanted to know what if we approach one limit point? If the limit point is algebraic, are there any patterns in these algebraic numbers? (in binary perhaps?).



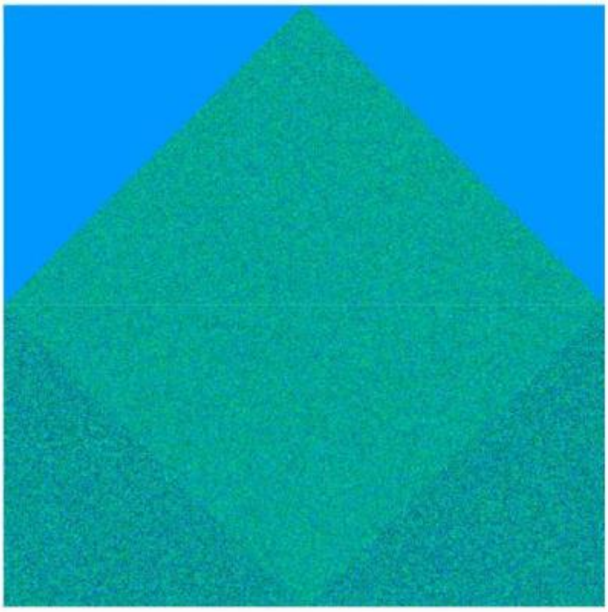
Well, I have found one formula:

If $f(n) = 1 + \frac{\sqrt[4]{16^{4n}+1}}{16^n+1}$, then the binary expansion of *f* has a very, very long and persistent pattern.

When n = 4096 then at position 1342238724 there are 4118 successive bits of this number that are all ‘0’.

(estimated) at n = 1 000 000 the persistent pattern goes up to the 80 000 015 000 004th position. At n 1 000 000 000 the position is at 8.0 x 10^{18}.

The first 270 million bits of $1 + \frac{\sqrt{2 \cdot 4^{8192} + 2\sqrt{16^{8192}+1}}}{2^{16386}}$:

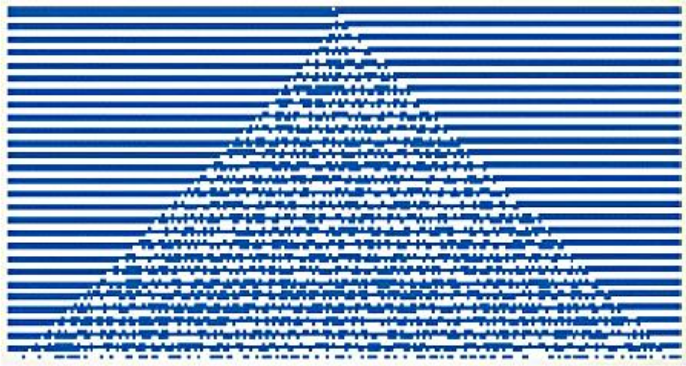


In false color: The bits were colorized to enhance the contrast

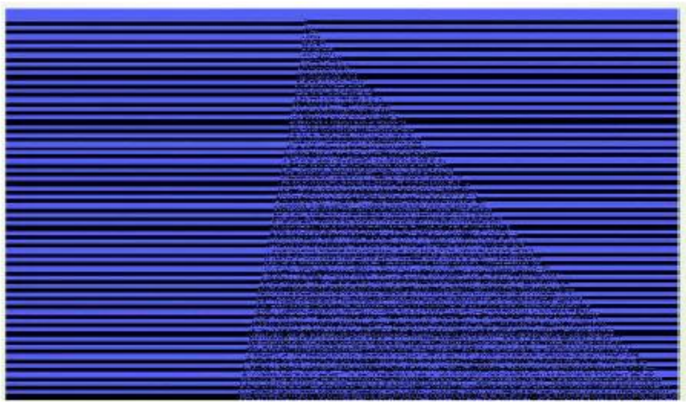
This is the binary expansion of *f*(100) (same as above):

```
f(100)=1.0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000010
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000111111111111111111111111111111
1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
1111111111111111111111111111111111111101100000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000010100111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
1111111111111111111111111111111111111111111111111111111111111111111111110010100110000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000100101111110111111111111111111111111111111111111111111111111
1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
1111111000110100101111000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000001011010101011111001111111111
11111111111111111111111111111111111111111111111111111111111…
```

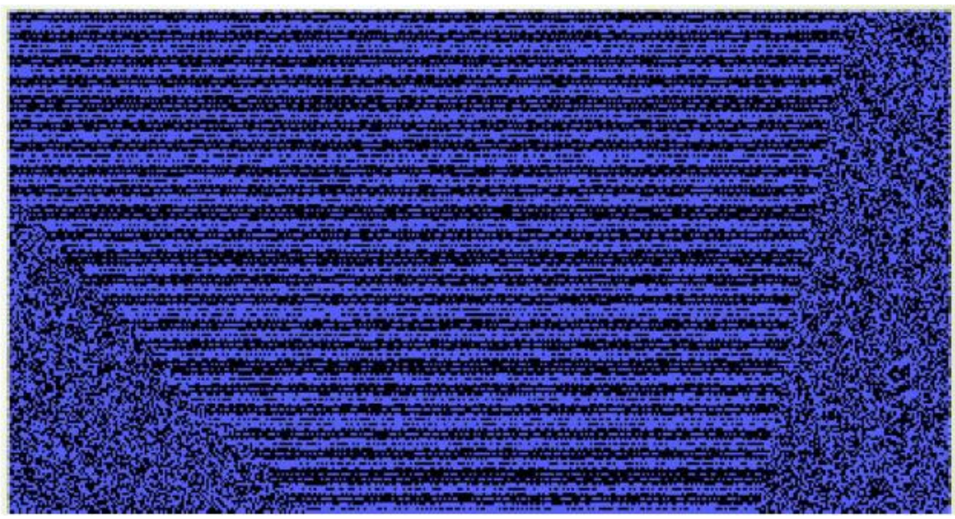
When we zoom in:



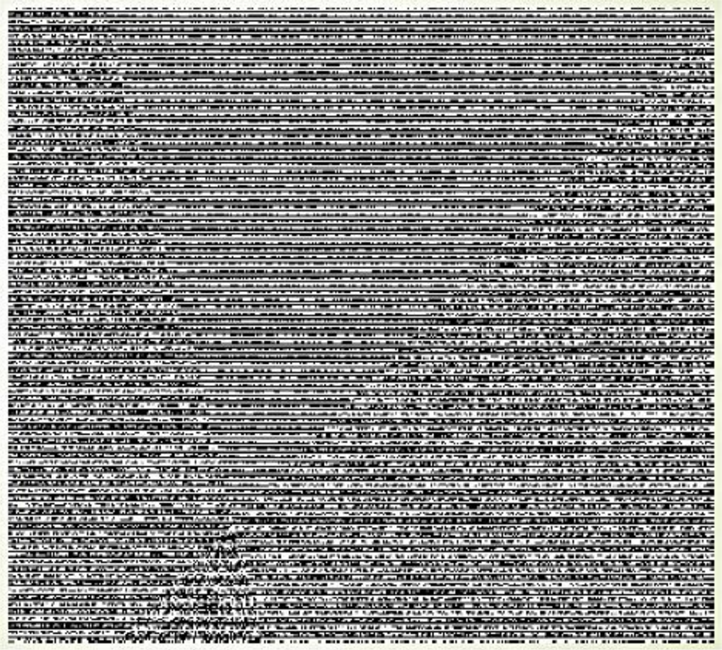
The first 446 million bits of the number $1+\frac{\sqrt{2\cdot 32^{16383}+3\sqrt{4^{16383}+1}}}{2\cdot 16^{32766}}$



At another point (depending on the choice of width):



Zoom at the bit level (from an image of 332 m pixels).



The data

- In all, 74 TB of mathematical data, mostly numbers,
- 41 TB of primes, from 2 to 805994098476893 (2602 billion entries)
- 5.6 TB of ζ zeros, 103 billion zeros
- OEIS tables (and extended tables)
- The Inverter, 41 digits (small version with 11.3 billion entries), 64 digits, 17.2 billion entries
- Inverter 41: 1.008 TB, http://plouffe.fr/ip/
- Inverter 64: 2.15 TB
- High resolution images: 1206000, 1.773 TB

Errors found …

- Prime[8200000000] and Prime[93000000000] just hangs in *Mathematica*.
- The tables of zeros at http://www.lmfdb.org/zeros/zeta do contain errors when the decimal expansion finishes by 00. 11 errors where found.

The index and the corresponding zeros do not match.

In both cases they are working on the problem.

Thank you for your attention

Merci de votre attention

References

[1] M. Cipolla, *La determinazione assintotica dell' nimo numero primo*, Rend. Accad. Sci. Fis-Mat. Napoli (3) 8 (1902), 132-166

[2] P. Dusart, *The k-th prime is greater than k (log k + log log k – 1) for k_2*, Math. Comp.68 (1999) 411-415

[3] B. Salvy, Fast computation of some asymptotic functional inverses, J. Symbolic Comput. 17 (1994), 227-236

[4] Visser, Matt, Primes and the Lambert function: https://www.mdpi.com/2227-7390/6/4/56/htm

[5] LeClair, André: Franca, Guilherme, Transcendental equations satisfied by the individual zeros of Riemann ζ, Dirichlet and modular *L*-functions: https://arxiv.org/pdf/1502.06003.pdf

[6] Corless, R.M.; Gonnet, G.H.; Hare, D.E.G.; Jeffrey, D.J.; Knuth, D.E. (1969). "On the Lambert W function" (Postscript). Advances in Computational Mathematics 5: 329-359. https://cs.uwaterloo.ca/research/tr/1993/03/W.pdf

[7] The Lambert function: http://www.orcca.on.ca/LambertW/ (nice poster!)

[8] The Lambert W function: https://dlmf.nist.gov/4.13

[9] Hayes, Brian: American Scientist, 2005 https://www.americanscientist.org/sites/americanscientist.org/files/2005216151419_306.pdf

[10] OEIS, Online Encyclopaedia of Integer Sequences: http://oeis.org

[11] Encyclopaedia of Integer Sequences, Neil J.A. Sloane, Simon Plouffe, 1995 , Academic Press.

[12] Lambert W function: https://en.wikipedia.org/wiki/Lambert_W_function

[13] Tomás Oliveira e Silva, Extended tables of π(n):. http://sweet.ua.pt/tos/primes.html

[14] LMFDB site, section about zeta zeros: https://www.lmfdb.org/zeros/zeta/

[15] Jean-Paul Delahaye, *Merveilleux nombres premiers*, Bélin Pour La Science 2011 (2ème édition).

Last mail from Simon Plouffe:

Well, I have good news and bad news.
The bad news is that you will be perhaps obliged to write a third section because I have found more.

The <approximation> I had for p(n) and pi(n) are quite precise now, so precise that I have a formula for p(n) more precise than the one of Cipolla-Salvy and also for pi(n) which is more precise than the one of Riemann. I have both at the same time. The next finding is that the formula for both results is <the same>, yes the same. I have found a way to have p(n) and pi(n) at the same time using the same formula, one being the inverse of each other. There is a way to get the precise value of both using fsolve of Maple in a particular way. The trick is that in both situations the solution is the solution to a transcendental equation. My solution is good from 2 to 7.35 x 10 ^ 15. I think I can push to 10^16.

I am writing the article on this subject right now. It will be on ArXiv and or Vixra too.
Best regards and merry Xmas.

Applications of Computer Algebra – ACA 2019
Montreal, July 16 – 20, 2019

Realizing the Concept of “Multiple Representations” by using CAS (2)

Helmut Heugl

Example 3: Upper and lower sums
Source Gertrud Aumayr

At first, we look at strictly monotonic increasing functions.

Given is the function f1: $f1(x) = 0.2 \cdot x^2 + 1$

The idea of this prefabricated applet:

Graphic representation: After definition of a slider for n the rectangles of the upper and the lower sums are drawn in the interval [0,6]

CAS representation in the notes application: Using the sum-operator, the product sums for the upper and lower sum are calculated. For comparison also the definite integral is calculated as a black box.

Task:

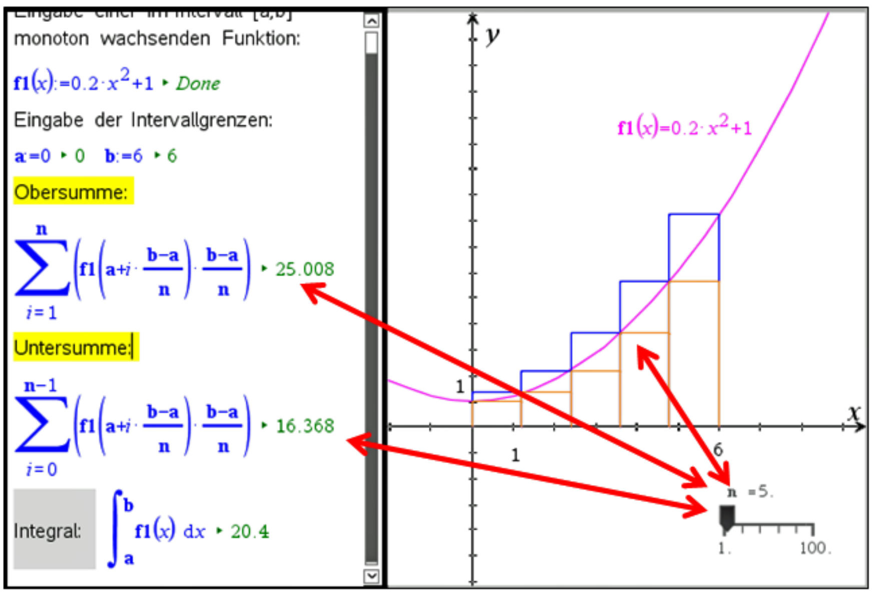
Observe the development of upper- and lower-sums for growing n. Look at the product sums calculated by CAS as well as the graphic representation.

Solution:

Modern digital media offers the strategy of

Dynamic linking:

Graphical manipulations as well as algebraic manipulations can simultaneously be transferred in other modes of representations. Consequences of manipulations in the algebraic window can directly be observed in the graphic window and vice versa.

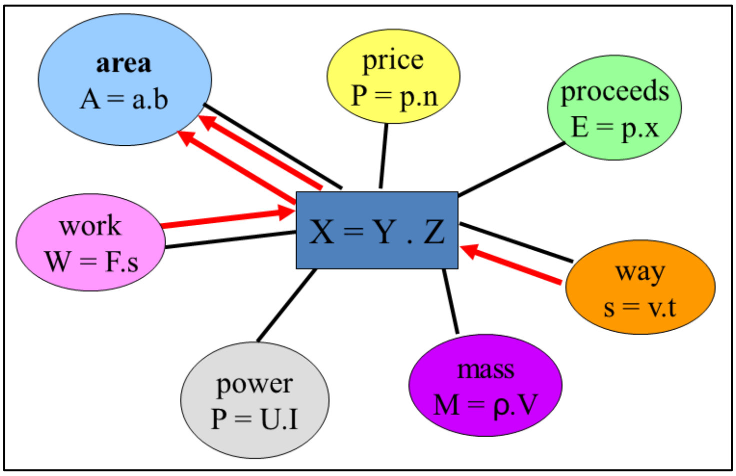


Dynamic linking:

Modifying the number n of the rectangles with the slider in the graphic window initiates the equivalent change of the product sum in the algebraic window.

Didactical comment:

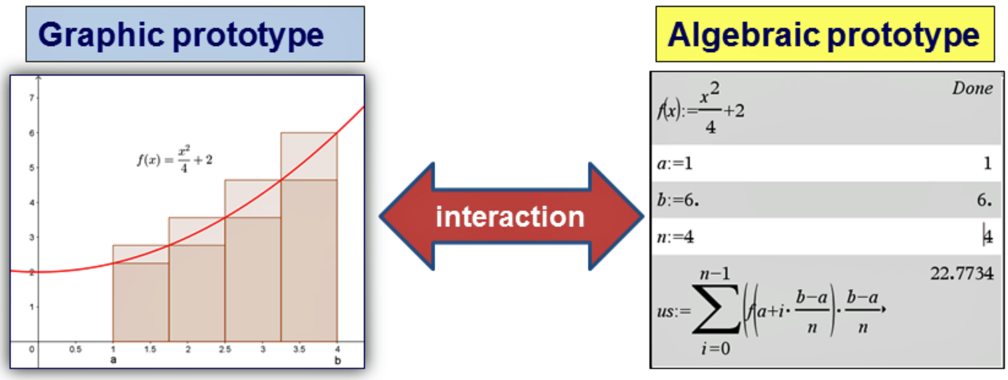
The development of the concept of the definite integral normally starts with the interpretation as an area. But after having generated first characteristics of the concept of a definite integral we should better forget this interpretation because the power of integrals are the many several interpretations in many other contexts.



The power of Integral ⬄ the relationship of many applications to the area because of a common mathematical model

The abstract interpretation of the definite integral is the limit of sums of products

Depictive representation | Descriptive representation



For this important cognitive step of abstraction, the interaction between these modes of representation is very important

Example 4: Ship collision?

The HMS Serena and the HMS Magic are sailing on the Atlantic Ocean: Their progress is being monitored by radar tracking equipment which is geographically oriented (W-E/N-S). As they come onto the observer´s rectangular screen, HMS Serena is at a point 900 mm from the bottom left corner of the screen along the lower edge. HMS Magic is at a point 100 mm above the lower left corner along the left edge. One minute later their positions have changed as follows:

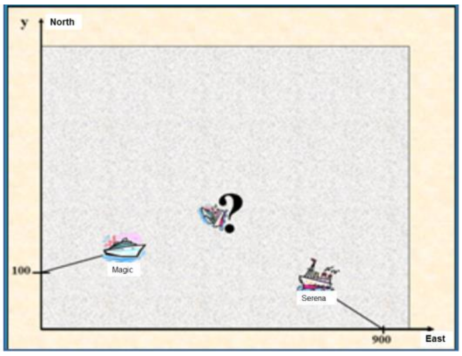
- HMS Serena has moved to a location on the screen that is 3 mm west and 2 mm north of the previous location and
- HMS Magic has moved 4 mm east and 1 mm north.

Task:

a) Will the two ships collide if they maintain their speeds and remain on their respective course? If so, when?

b) If not, how close do they actually come to each other?

Note: 1 mm ≙ 100 m



Solution:

$$serena(t):=\begin{bmatrix}900.\\0\end{bmatrix}+t\cdot\begin{bmatrix}-3\\2\end{bmatrix} \quad \text{Done}$$

$$magic(s):=\begin{bmatrix}0\\100.\end{bmatrix}+s\cdot\begin{bmatrix}4\\1\end{bmatrix} \quad \text{Done}$$

$$\text{solve}(serena(t)=magic(s),\{t,s\}) \quad s=136.36364 \text{ and } t=118.18182$$

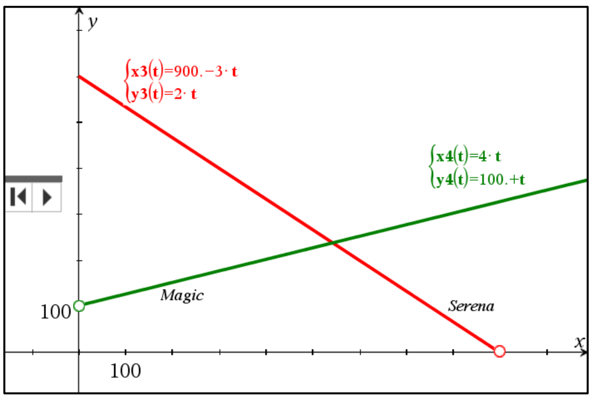
$$serena(118.18182) \quad \begin{bmatrix}545.45454\\236.36364\end{bmatrix}$$

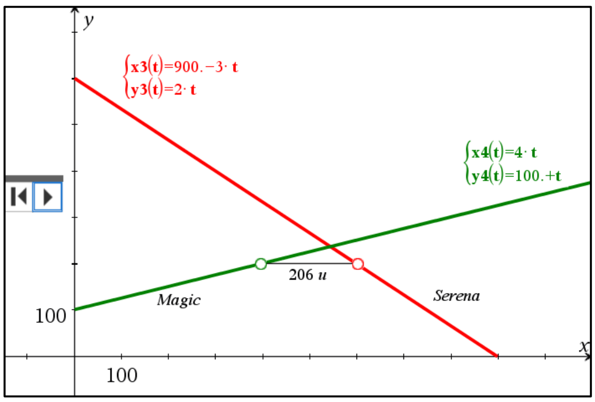
$$magic(136.36364) \quad \begin{bmatrix}545.45456\\236.36364\end{bmatrix}$$

When calculating in the algebraic window the intersection point of the paths of the two ships a possible conclusion is: At this point the ships will collide.

But students have to ask: Are they at this point at the same time? And: "What is the physical interpretation of the parameters?" The answer to the second question is: Parameters s and t represent the time (in minutes).

Now a shift to a depictive representation of the problem is helpful. Modern tools like TI Nspire CAS cannot only draw graphs of the paths they can also simulate the movement of the two ships with respect to the time t and show their distance.



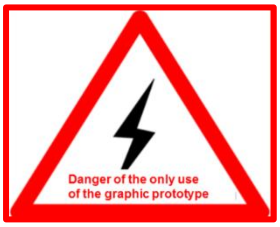


The simulation shows: The ships are not at the intersection point at the same time.

The exact calculation of the velocity of the two ships and their minimal distance needs a return to a descriptive interpretation in the algebraic window:

norm$(serena(1)-serena(0))$	3.6055513
norm$(magic(1)-magic(0))$	4.1231056
$3.6055513 \cdot 6$	21.633308
$v_{Serena}=3.6055513 \cdot 6$	$v_{Serena}=21.633308$
$v_{Magic}=4.1231056 \cdot 6$	$v_{Magic}=24.738634$

norm$(serena(t)-magic(t))$	$5\cdot\sqrt{2\cdot(t^2-256.\cdot t+16400.)}$
$dist(t):=$norm$(serena(t)-magic(t))$	*Done*
$dist1(t):=\frac{d}{dt}(dist(t))$	*Done*
solve$(dist1(t)=0,t)$	$t=128.$
$dist(128)$	28.284271
$28.284271\cdot 100$	2828.4271



The only creation of a depictive representation such as drawing is not sufficient for successful cognitive problem solving even the drawing is correct. Sometimes the specific perceptual structure or other perceptual attributes can obscure the relevant structural attributes.

Example 5: What is $\sqrt{2}$?

Source: Applet "Integrator" developed by H.-J. Elschenbroich, H. Langlotz and G. Aumayr

To find an answer a definition of irrational numbers formulated by R. Fischer is helpful:

> *The step to irrational numbers takes place by declaring the possibility of an arbitrary approach as a number*

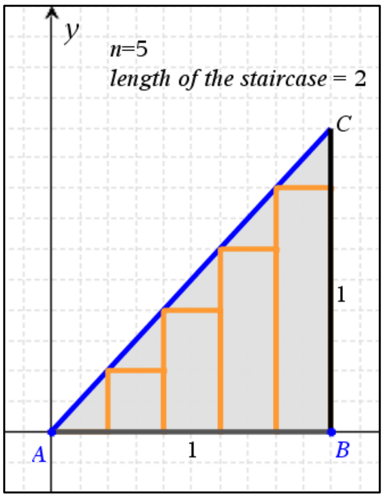
We can simulate this arbitrary approach by using the following depictive representation:

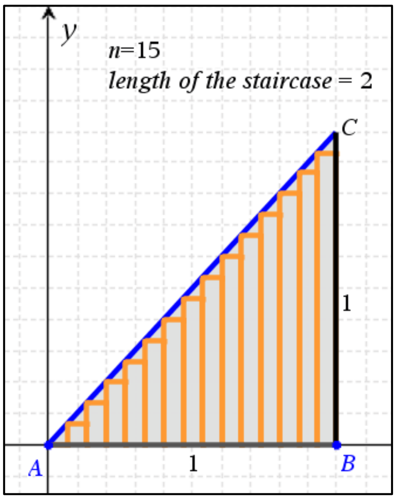
Task: What is $\sqrt{2}$?

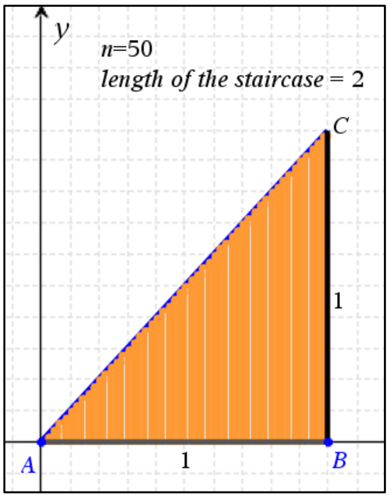
We draw the graph of f(x) = x in the interval [0,1]. By using this applet, we can draw the lower sum with respect to the number n. Now a staircase can be seen.

The central question is: "What is the length of the staircase?"

With a slider the number n can be enlarged. For growing n we again ask: "What is the length of the staircase?"







Graphic representation shows:

- For any n the length of the stair is 2
- For arbitrary large n the staircase approaches more and more to the diagonal of the square with side length 1.

Conclusion:

- We have a constant sequence with elements 2 ⇨ the limit is 2.
- The graphic representation shows: The staircase with constant length 2 “converges” to the diagonal with length $\sqrt{2}$.
- **We have shown by visualization $\sqrt{2} = 2$**

Conclusion 3.2: The Interaction between graphic representations and CAS is an important contribution for building effective mathematical mental models which are the corner-stones of problem solving.

3.3 Interaction between graphic representation, CAS and data represented in interactive tables

Solving more complex problems often needs the interaction between mor than two modes of representation.

Example 6: **A computer virus**

Within two hours nearly all PCs in a company with 4500 networked PC-working stations are attacked by a virus. After a complete cleanup the network administrator can reconstruct the time flow:

Time (minutes)	0	20	40	60	80	100	120
Number of infected PCs	15	99	598	2346	4024	4435	4492

Task:

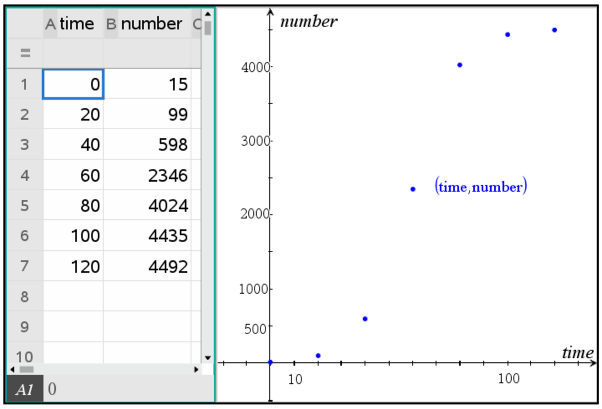
(1) What function could describe the growth process at the best?

(2) After how many minutes a maximum rate of change can be observed?

Solution:

Step 1: Entering the given data in the “Lists&Spreadsheet“ application

Step 2: Drawing the graph of the scatter plot



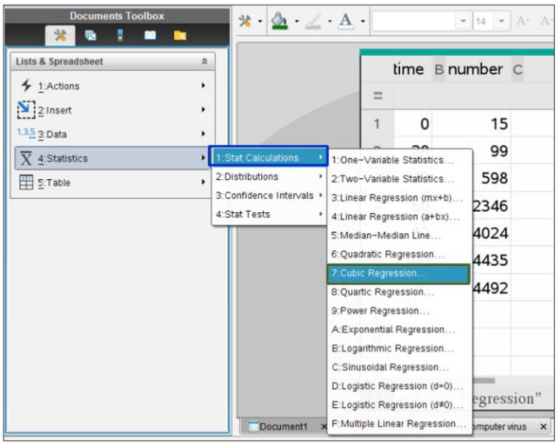
The descriptive representation (1) of the table allows no conclusion about the function type.

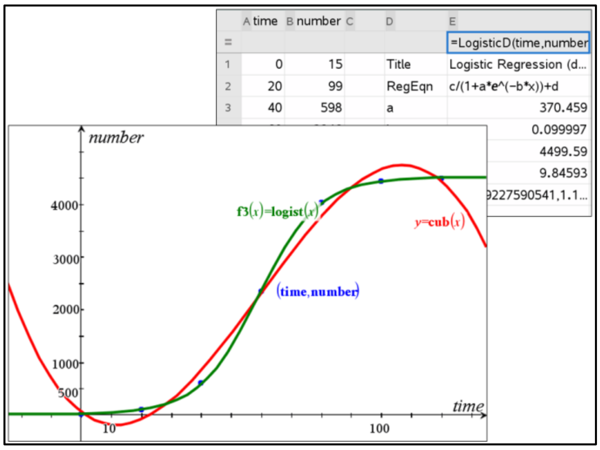
⇨ the depictive representation (2) of the graph of scatter plot is necessary

Now assumptions about the function type are conceivable, e.g.

- a polynomial function with grade 3 or
- a logistic function.

Step 3: Looking for the equations of the suitable regression functions

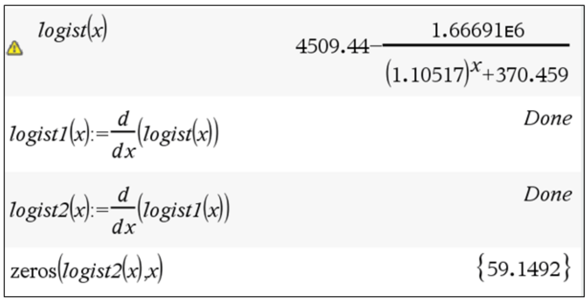




Tools like TI Nspire offer the descriptive representations (3) of the function equation together with parameters like R^2 the coefficient of determination. The depictive representation (4) of the regression functions together with the given parameters lead to the solution of the problem.

Task (2):
After how many minutes a maximum rate of change can be observed?

Now the descriptive representation of CAS is necessary.



Conclusion3.3: A solution of such a problem is only thinkable with the interaction of various descriptive and depictive representations. If algebraic operations are necessary then the descriptive representation mode of CAS is important.

3.4 Interaction between recursive prototypes, graphic representations and CAS

The only new content in our mathematics curriculum caused by the use of technology is the representation of functional dependencies by recursive models. Thereby a large area of growth processes becomes accessible for mathematics education which before only could be solved with complex differential equations. Also, for the development of the concept of irrational numbers the use of recursive models is very helpful.

But the only use of the depictive representation of the result of the simulation process can only lead to assumptions. For ensured conclusions about characteristics of the functions like convergence etc. descriptive representations which are possible by using CAS are necessary.

Example 7: A recursive model for the approximation of irrational numbers
[Schweiger, 2013]

Given are two recursive models for the approximation of $\sqrt[k]{a}$

Model 1: $x_{n+1} = \frac{1}{2} \cdot \left(x_n + \frac{a}{x_n^{\,k-1}} \right)$ and Model 2: $x_{n+1} = \frac{1}{k} \cdot \left((k-1) \cdot x_n + \frac{a}{x_n^{\,k-1}} \right)$

Task: Investigate the convergence of the two models of $\sqrt[k]{a}$

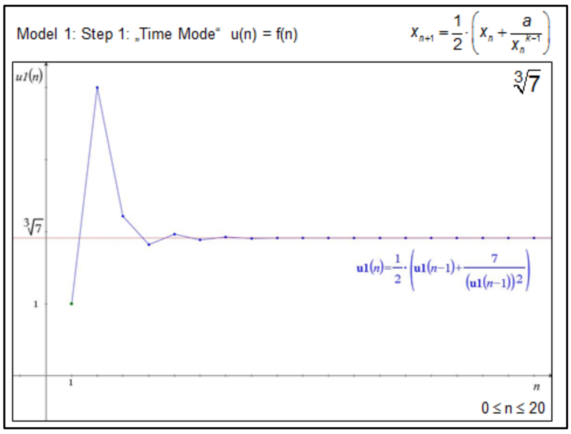
The learning process proceeds in two phases:

- **Phase 1** (experimental phase): Draw the graphs of the sequence in the "time mode" and the "web mode". Is a convergence observable?
  In this experimental phase depictive representations of the graphs lead to conjectures.
- **Phase 2** (exactifying phase): Calculate the fixed points of the sequences and investigate the character of the fixed point by using the fixed-point theorem.
  A secured answer is only possible with a proof by symbolic operations ⇨ the interaction between the graphic representation and the algebraic representation offered by CAS is necessary.

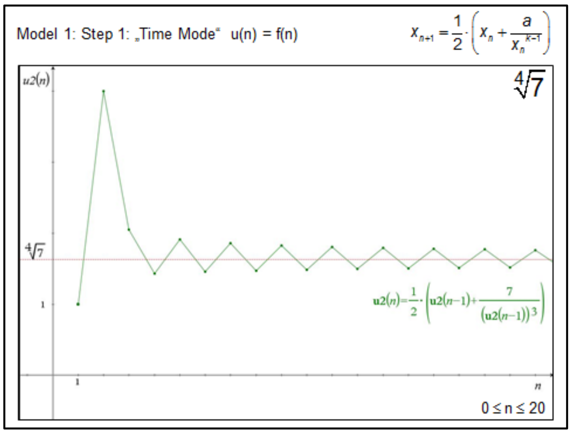
Phase 1 (experimental phase): Investigating the graphs, coming to suppositions:
Students can use two representation modes of the graph:

- Step 1: They use the "Time Mode" u(n) = f(n)
- Step 2: They use the "Web Mode" u(n) = g(u(n-1))

For the experimental phase I have chosen: a = 7 and k = 3, 4, 5

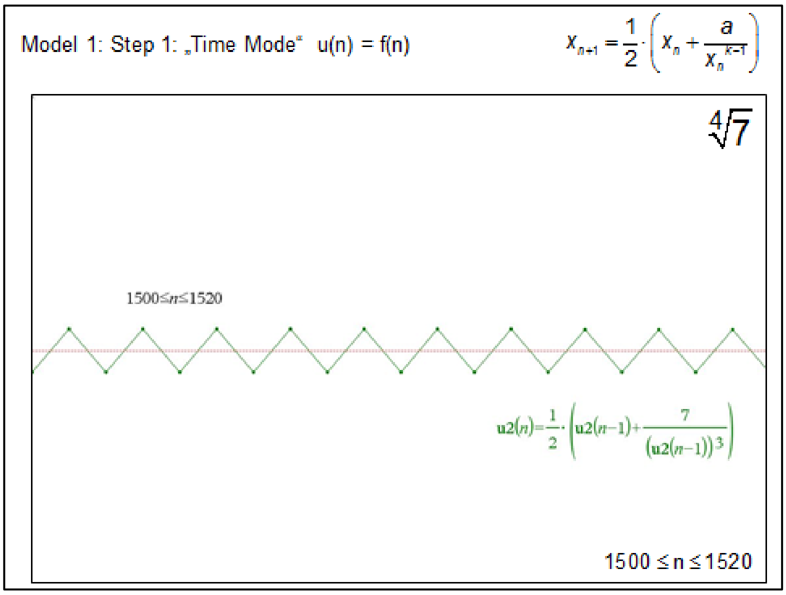


k = 3:
Model 1 seems to be convergent



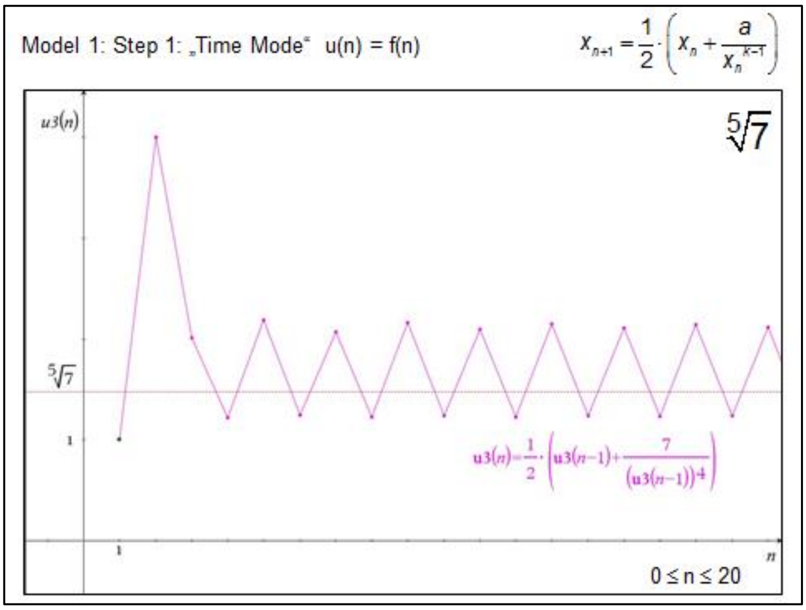
k = 4:

Observing the interval [1,20] we cannot be sure if model 1 is convergent. But by changing the window variables we look at other regions



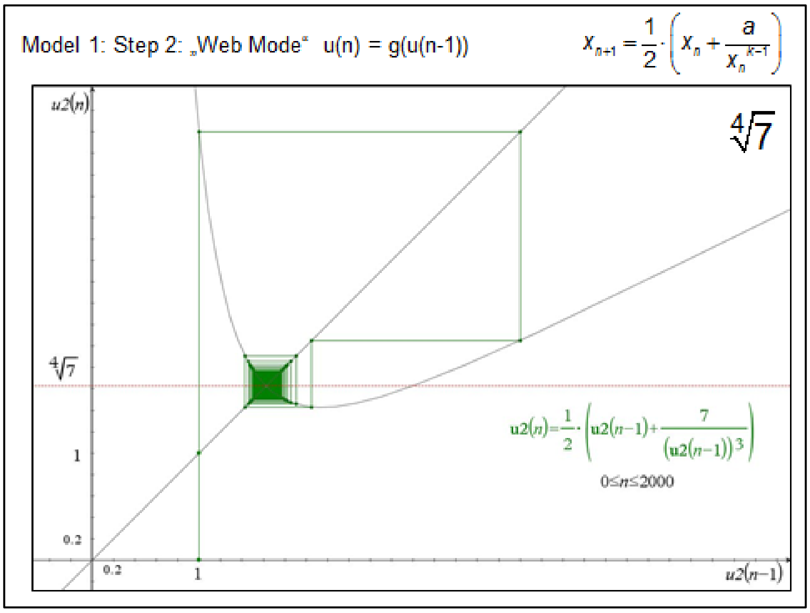
k = 4:

Let us look at the interval [1500, 1520]. We still cannot decide the convergence of model 1.



k = 5:

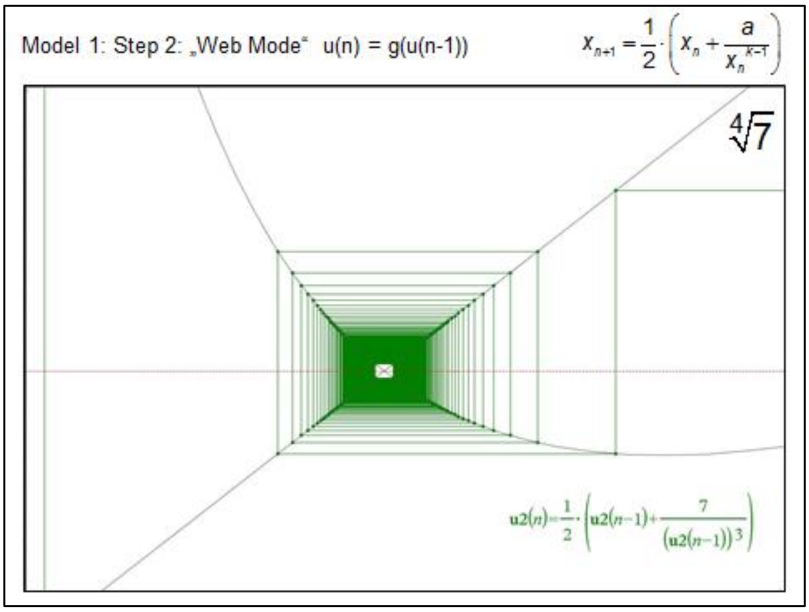
It is improbable that model 1 converges



Let us look at the **"web-mode"**:

k = 4:

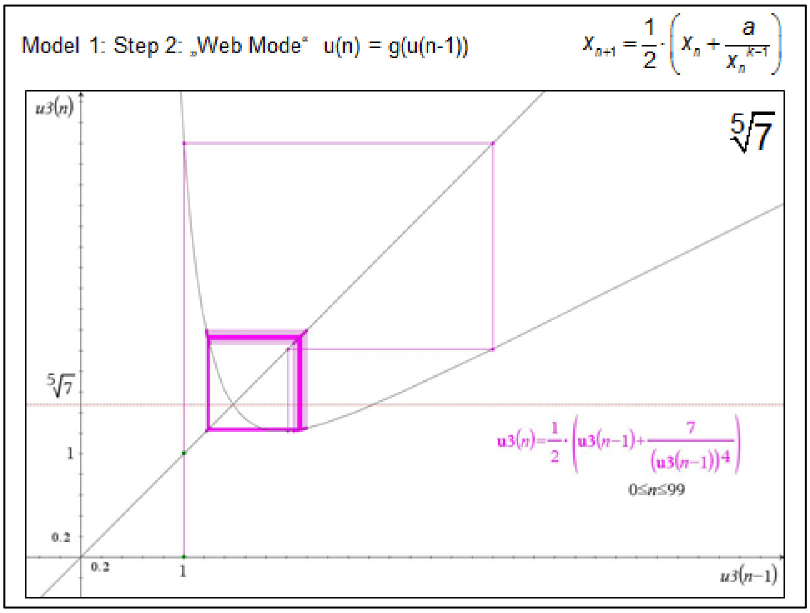
Model 1 seems to be convergent



k = 4:

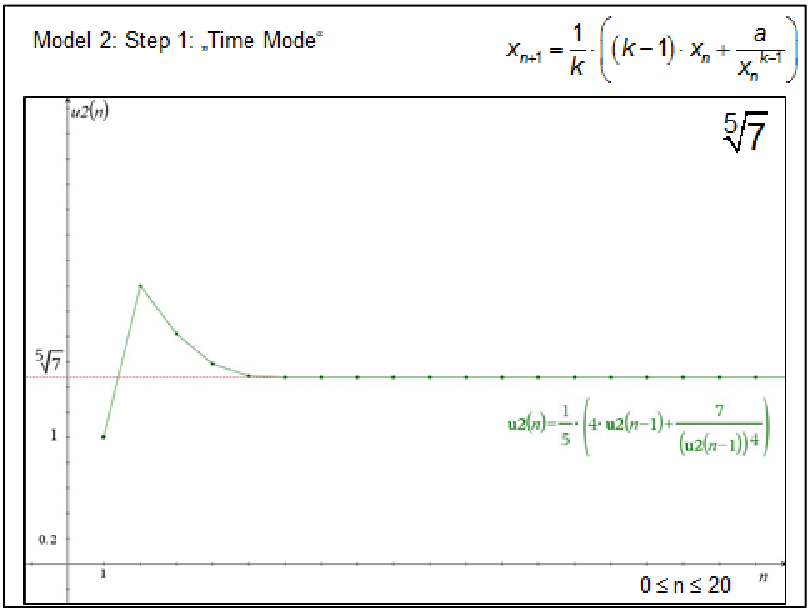
When being critical we zoom in and now we see that it will probably not be convergent.

But we have still no sureness. We have only visualized the interval [0,2000].



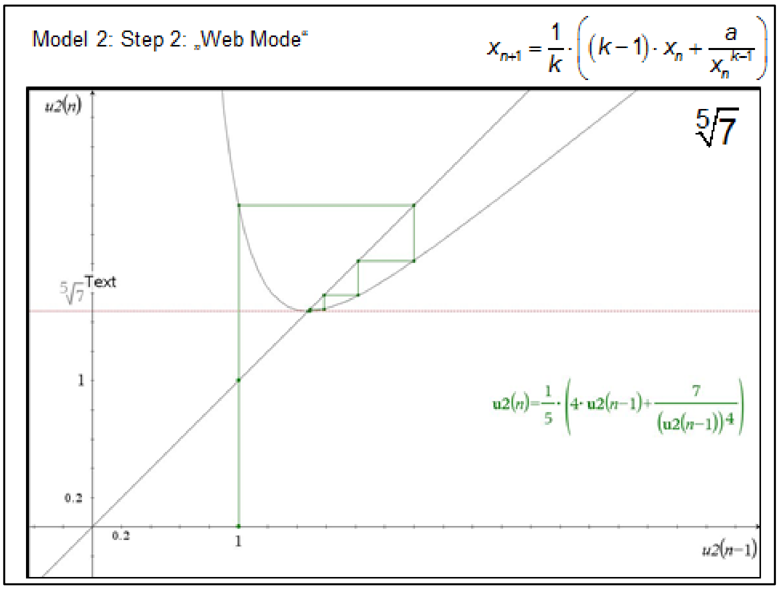
k = 5:

The graph in the web mode strengthens the supposition that model 1 is not convergent.



k = 5:

Model 2 seems to be convergent also for k =5.



k = 5:

The web graph strengthens the supposition.

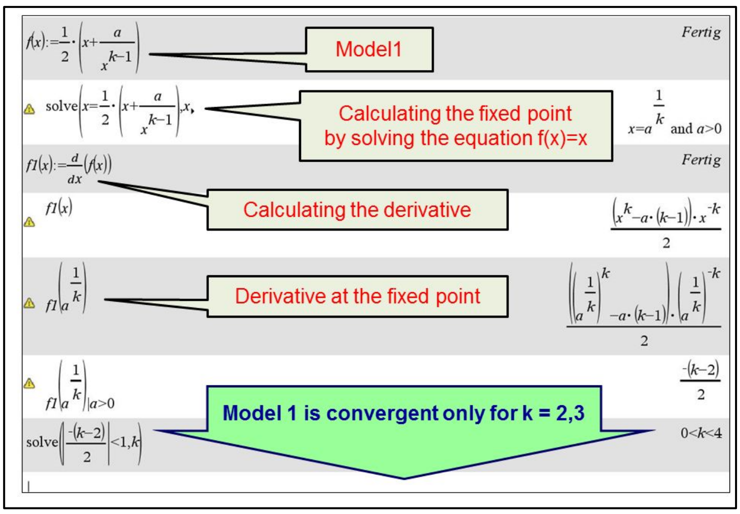
Phase 2 (exactifying phase): Proving the convergence of the given sequences

I think this example clearly shows the possibilities but also the limits of the use of technology in the experimental learning phase or we can say often an experimental phase is only possible by technology. But we cannot be content with these results, in an exactifying phase the power of CAS is necessary to get certainty.

At first, we need **some theoretical prerequisites**:

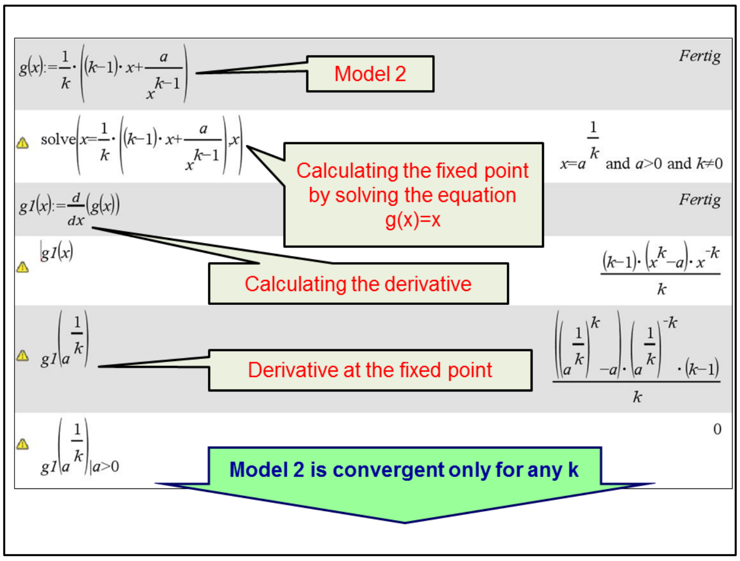
- **A fixed point x*** of a function f is an element of the function's domain that is mapped to itself, that is f(x*) = x*.
- A fixed point is **an attractive fixed point x*** of a sequence given by a difference equation $x_n = f(x_{n-1})$, if f converges to x*, that is $\lim_{n\to\infty} f(x_n) = x^*$.
- **The fixed point theorem:** A fixed point x* of a difference equation $x_n = f(x_{n-1})$ *(f is continuous and differentiable) is an attractive fixed point, if* $|f'(x^*)|<1$ *and is distractive, if* $|f'(x^*)|>1$.

The task of phase 2 is: Calculate the fixed points and investigate the character of the fixed point by using the fixed point theorem.



Without knowing the condition that the radicand a must be positive students will not get suitable results.

The CAS tool is critical: When looking at the solution of the equation f(x) = x we will find the condition which the CAS tool expects: **a > 0**.



Results:

Model 1 is convergent only for k = 2,3

Model 2 is convergent only for any k

Conclusion 3.4 By simulating technology offers a depictive representation of the recursive model which can be used for verbal descriptions of the functional dependence or for conjectures. But exact results or proofs of conjectures need the close interaction with symbolic representations offered by CAS.

Summary

I tried to express the relevance of following thesis:

- The principle of multiple representations, the combined use of different representations, is a key strategy for teaching and learning mathematics.
- Computer based multimedia learning differs from traditional learning environments because several modes of representations are available simultaneously. This fact and the possibility of a direct exchange of information between the several representations ("dynamic linking") encourages the development of a stable mental model.
- CAS are the best modes of representations to express the symbolic language of mathematics because these symbolic representations allow not only the depiction of symbolic objects but also the performance of mathematical operations.
- **In a computer-based learning environment CAS are indispensable for realizing the principle of "Multiple Representations".**

Dr. Helmut Heugl

ACDCA (Austrian Center for Didactics of Computer Algenra)

Pädagogische Hochschule Niederösterreich
Regionales Fachdidaktikzentrum - Mathematik und Informatik

hheugl@aon.at

Business in the Public Pool

Sebastian Rauh, Germany, joined the DUG on 28 September 2019, 08:36. On just the same day I received his mail (see below) at 16:55. This was the start of a very intense exchange of mails and files. Finally, we involved our member Guido Herweyers from Belgium. I translated the mails and the notes-pages in the Nspire-files, Josef

Hallo Herr Böhm,

im letzten Jahrgang des Abiturs in NRW wurde eine Aufgabe gestellt, die in meinen Augen durch die SuS nicht lösbar war. Auch die angebotene Lösung war, meines Erachtens nach, nicht richtig. Ich konnte die Aufgabe allerdings selbst nicht so lösen, wie ich das wollte und habe deshalb eine Simulation geschrieben.

Nach 3 Durchläufen (a ca 15min) scheint sich zu bestätigen, dass die Musterlösung nicht stimmt.

Abitur NRW 2019 CAS LK (erhöhtes Anforderungsniveau); M_19_c_L_HT_B5_GG

Original:

Für ein Schwimmbad besitzen 2000 Personen eine Jahreskarte. Für einen bestimmten Tag beschreibt die Zufallsgröße X die Anzahl der Jahreskartenbesitzer, die das Schwimmbad besuchen. Vereinfachend soll davon ausgegangen werden, dass X binomialverteilt ist. Dabei beträgt die Wahrscheinlichkeit dafür, dass ein zufällig ausgesuchter Jahreskartenbesitzer an diesem Tag das Schwimmbad besucht 10%.

b) Auf dem Gelände des Schwimmbades wird ein Kiosk betrieben. Der Besitzer nimmt vereinfachend an, dass jeder Gast 4 €, 12 € oder gar kein Geld ausgibt. Die Wahrscheinlichkeit, dass ein Gast 4 € ausgibt, betrage 50%, die Wahrscheinlichkeit, dass ein Gast 12 € ausgibt, betrage 30%.

(2) *Ermitteln Sie die Wahrscheinlichkeit, dass der Besitzer des Kiosks an dem betrachteten Tag erwartete Einnahmen von den Jahreskartenbesitzern hat, die mindestens 1000 € betragen.*

Lösungsweg laut Korrekturanweisung:

Sei n die Anzahl der benötigten Jahreskartenbesitzer:

Aus $0{,}5\cdot n\cdot 4 + 0{,}3\cdot n\cdot 12 = 1000$ folgt $n = \dfrac{1000}{0{,}5\cdot 4 + 0{,}3\cdot 12} = \dfrac{1000}{5{,}6} \approx 178{,}6$

Der TR liefert $P_{2000;0,1}(X \geq 179) = 0{,}947$.

Die gesuchte Wahrscheinlichkeit beträgt daher ca. 95%.

Translation (JB)

Hello Mr. Böhm,

in the last final exam, a problem was posed which in my opinion cannot be solved by the students. Even the offered expected sample solution is – in my opinion – not correct.

But I couldn’t solve the problem as I had wished. So, I wrote a simulation program.

Three runs of the simulation (~ 15 min.) seem to confirm that the sample solution must be wrong.

Final exam NRW 2019 CAS LK (increased level); M_19_c_L_HT_B5_GG

2000 persons are annual ticket holders for a public pool. Random variable X describes the number of ticket holders who attend the pool on a certain day. Simplifying we can assume that X is binomial distributed. The probability that a random chosen ticket holder is attending the pool on this day is 10%.

b) There is a kiosk on the premises of the pool. Its owner simplifying assumes that every guest spends 4 €, 12 € or nothing at all. The probability that a guest spends 4 € is 50%, that he spends 12 € is 30%.

   (2) *Find the probability that the owner of the kiosk can expect revenues of at least 1000 € from the annual ticket holders.*

Solution according to the official correction instruction:

Let n the number of the ticket holders needed:

From $0{,}5\cdot n\cdot 4 + 0{,}3\cdot n\cdot 12 = 1000$ follows $n = \dfrac{1000}{0{,}5\cdot 4 + 0{,}3\cdot 12} = \dfrac{1000}{5{,}6} \approx 178{,}6$

The calculator gives $P_{2000;0,1}(X \geq 179) = 0{,}947$.

Hence, the wanted probability is approximately 95%.

Despite the fact, that the modeling is extremely inappropriate, the solution is wrong imho.

If 179 ticket holders are in the pool, then the probability to make revenues of at least 1000 € not 1. It should be approximated by a multinomial distribution because the numbers might become too large.

$\mathbf{eink}(10) \blacktriangleright 28 \qquad 3\cdot\sqrt{2000\cdot 0.1\cdot(1-0.1)} \blacktriangleright 40.2492$

pk1, pk2, pk3 is the probability, that with exact 159,160,...241 guests at least 1000 € are earnt (each run is poduced by 80*1000 single simulations).

pk1

▸ {0.034,0.023,0.036,0.053,0.061,0.074,0.078,0.109,0.122,0.156,0.169,0.184,0.219,0.252,0.282

pk2

▸ {0.024,0.033,0.038,0.038,0.06,0.065,0.093,0.106,0.112,0.143,0.165,0.191,0.222,0.239,0.301,(

pk3

▸ {0.021,0.025,0.031,0.04,0.061,0.079,0.08,0.124,0.125,0.144,0.184,0.177,0.226,0.275,0.3,0.31

This is the probability that 159,160,..241 guests show up.

bin:=seq(binomPdf(2000,0.1,k),k,160,240)

▸ {0.000288,0.000365,0.000461,0.000577,0.000719,0.000888,0.001091,0.001331,0.001614,0.00

Hence, the probability that 159 people appear and spend at least 1000 €:

pk1[1]·**bin**[1] ▸ 0.00001

The probability that on one day 1000 € are spent by ticket owners is:

p1:=sum(**pk1**·**bin**) ▸ 0.874639 $\quad$ *p2*:=sum(**pk2**·**bin**) ▸ 0.873168 $\quad$ **p3**:=sum(**pk3**·**bin**) ▸ 0.874942

mean({**p1**,**p2**,**p3**}) ▸ 0.87425

File: pool_bad.tns (English and German)

eink(n) simulates the purchases of n guests spending 4€, 12€ or 0€ with the given probabilities.

Dear Mr. Rauh,

I was busy with your end examination problem yesterday evening after having received your mail. At first sight I found the result of the provided solution feasible, because I assumed an expected value of 100*4+60*12 = 1120, which is considerably above 1000. Hence, the result could be correct.

Today morning I made a simulation, but a little bit different as you did:

eink2(100)	0.98
eink2(100)	0.99
eink2(1000)	0.94
eink2(1000)	0.951
eink2(1000)	0.953
eink2(5000)	0.95
eink2(5000)	0.9478
eink2(10000)	0.9523
eink2(10000)	0.9466
eink2(20000)	0.94905

```
"eink2" stored successfully
Define eink2(n)=
Func
Local z,ums,l_ums,i,anz
anz:=0
l_ums:={}
For i,1,n
z:=randBin(2000,0.1)
ums:=z· 0.5· 4+z· 0.3· 12
©l_ums:=augment(l_ums,{ums})
If ums≥1000: anz:=anz+1
EndFor
Return 1.· anz/n
EndFunc
```

File: pool_bad_1.tns (English and German)

The number of the ticket holders is binomial distributed, I can get this number as a random variable (*z*). This gives the distribution of the purchases and then of the total revenues (*ums*). I don't use another probability distribution.

Using function **eink2(n)** it is very easy to simulate a large number of days. The result fits to the sample solution (95%).

I hope that I could help you,
best regards
Josef

Hello, Mr. Böhm

Thanks for inspecting the problem. Now I have understood, what my problem really is:

Your interpretation of the problem is:

"Find the probability that the kiosk owner can expect at least 1000 € from the annual ticket holders." as question for the expected value. This is provided in the sample solution and is probably so interpreted by the school authority, too.

My interpretation is different. See an example:

x people are in the pool and buy ice cream for 1 € with a probability of 50%. 10 € should be earnt. Buy/Buy not is binomial distributed, too.

According to the solution 20 people should be in the pool.

But, using binomial distribution leads to the prob that earning 10 € or more having 20 visitors is only 59%. 95% probability can be achieved only with 14 visitors or more.

Now, the problem becomes more complicated:

The number of guests is again binomial distributed with expected value 20, if e.g. the prob to visit the bath among 100 persons is 20%. Now we have a two-dimensional binomial distribution because even for 19 visitors the probability to earn 10 € is not 0 and for 21 visitors the prob is not 1.

And because this is not complicated enough there are also ticket holders who spend 2 €, e.g. p(0 €) = 0.2, p(1 €) = 0.5 and p(2 €) = 0.3.

This results strictly speaking in a two-dimensional multinomial distribution. As I cannot calculate this, I performed a simulation.

Dear Mr. Rauh,

and that's the nasty (and interesting as well) property of stats and prob theory problems. Many of them are not so clear and unique as the old dice or counters-box-problems.

Sometimes one interprets the problem too complicated or even in another way as intended by the poser of the task – and even he/she is sometimes blind in his/her work.

When I was a teacher it could have happened that a text of a task was absolutely clear for me – but not for the students (and often they were right!).

I must admit that probably I had not found the solution how it was given in the sample solution.

But I found out something other which is interesting:

I refined my model, by assuming that the 50% and 30% part of the ticket holders is also binomial distributed. Additionally, the average earnings are presented.

I receive very accurate the expected value of 1120 € (same like in my first model), probability is significantly smaller.

Maybe that the reason is that the probabilities are not independent. (10% of all visitors are ticket holders, and then the 50%/30% …?)

The 2 € clients did not appear in the original problem?

Regards
Josef

Again Josef, two days later:

Dear colleague,

I could not resist revising the pool-simulation once more. Using a second parameter (s) you can perform both simulations.

In case 1 (s = 1) we have a smaller dispersion for the 50% and 30% parts because only one distribution has to be considered.

Case 2 is different.

I have no idea how to use this approach to find the exact solution for the problem by a calculation.

Best regards

Josef

```
badegast(1000,1)        {0.939,1115.39}
badegast(1000,2)        {0.859,1123.04}
badegast(100,1)         {0.94,1115.41}
badegast(100,2)         {0.88,1131.2}
badegast(10000,1)       {0.9483,1119.9}
badegast(10000,2)       {0.8669,1121.06}
© mit Standardabweichungen:
badegast(500,1)
                        {6.36321,3.81793}
                        {0.96,1121.74}
badegast(500,2)
                        {9.61266,7.54916}
                        {0.878,1124.31}
```

```
badegast                                           6/18
Define badegast(n,s)=
Func
Local z,ums,s,s_ums,i,anz,z2,z3,l1,l2
l1:={□}:l2:={□}
anz:=0
s_ums:=0
For i,1,n
z:=randBin(2000,0.1)
z2:={0.5·z,randBin(z,0.5)}:z3:={0.3·z,randBin(z,0.3)}
© Disp z: Disp z2:Disp z3
l1:=augment(l1,{z2[s]}):l2:=augment(l2,{z3[s]})
ums:=z2[s]·4+z3[s]·12
© Disp {z,z2[s],z3[s],ums}
s_ums:=s_ums+ums
If ums≥1000
 anz:=anz+1
EndFor
Disp 1.·{√(varSamp(l1)),√(varSamp(l2))}
Return {1.·anz/n,1.·s_ums/n}
©Return l_ums
EndFunc
```

File: bad_2.tns (German) and pool_2.tns (English) (badegast = guest of the pool/bath)

Hello Mr. Böhm,

Yes, you are right, probability theory problems are those problems which require strongest understanding of the text. It's what I enjoy the most right now.

What concerns your simulation:

Your probability for earning more than 1000 EURO is approximatively 0.866, my simulation gives 0.874. This is amazingly close.

It’s clear for me that the expected value is above, but I had not expected to hit is so exactly.

(At the other hand, the last calculation runs over 10000 days – i.e. ~ 50 years (5 days week).

But dispersion is much greater, why?

I suppose that the deviation is caused by the addition of the binomial distributions for 4 EURO/12 EURO. Intuitively I had expected a greater probability to overcome 1000 EURO in this case.

(I will keep in mind your trick multiplying by 1.)

I added the 2 EURO clients to clarify my problem. I am sorry if this caused some confusion.

Best regards

Sebastian Rauh

Dear colleague Rauh,

did you find out something new in connection with the pool problem?

If you agree to pass on the problem, then I'd ask a colleague (and DUG member) – very experienced in stats and prob theory – for an explanation of the difference of the two simulations and how to find the result of the given problem directly.

Regards

Then I forwarded the problem together with our findings so far, our friend Guido Herweyers from Belgium. Some days later I received his answer.

Dear Josef,

Find attached my comment to the exam problem.

It was really a challenge ☺.

Swimming Bath Problem

Number of annual ticket owners is binomial distributed with n=2000, p=10%.
X is the respective random number.

The owner of the kiosk expects: 50% of the clients spend 4 EURO, 30% 12 EURO.
The sample solution is correct: the question concerns the expected earnings and not the actual revenue!

P(expected revenue ≥ 1000) = P(0.5 **4 + 0.3*X*12)≥1000)
= P(5.6 X ≥ 1000)
= P(X179)
= binomCdf(2000,0.1,179,2000) ▸ 0.947289

expected value: E(5.6X)=5.6 E(X) = 5.6· 2000· 0.1 ▸ 1120.

variance: Var(5.6X)= 5.6^2*Var(X)= $(5.6)^2$· 2000· 0.1· 0.9 ▸ 5644.8

standard deviation: $\sqrt{5664.8}$ ▸ 75.2649

Simulation of sample solution (= your simulation with parameter 1) confirmed:
badegast1(1000) ▸ {0.952,1117.89,73.7757}

variance: $var(5.6X)= 5.6^2 \cdot var(X)= (5.6)^2 \cdot 2000 \cdot 0.1 \cdot 0.9$ ▸ 5644.8

standard deviation: $\sqrt{5664.8}$ ▸ 75.2649

Simulation of sample solution (= your simulation with parameter 1) confirmed:

badegast1(1000) ▸ {0.952,1117.89,73.7757}

2. Parameter 2: the 4 EURO clients are also binomial distributed with z=number of the present annual ticket holders, p = 0.5; the 12 EURO clients with z and p = 0.3

Other question: P(Revenue ≥ 1000) =?

X1 is the number of the 4 EURO clients, *X2* is the number of the 12 Euro clients,

U is the revenue

$U=4*X1+12*X2$

In your simulation (parameter 2) the random variables *X1* and *X2* are independent, but it can happen that a guest spends 4+12 = 16!

X1 and *X2* are dependent.

We solve the more difficult problem P(revenue ≥ 1000) =? by simulation and calculate the exact probability.

1) Simulation

Here we have a multinomial distribution (https://en.wikipedia.org/wiki/Multinomial_distribution) with n=2000 and three possible outcomes. Every ticket holder spends:

4 € with probability $0.1 \cdot 0.5$ ▸ 0.05

12 € with probability $0.1 \cdot 0.3$ ▸ 0.03

0 € with probability $0.1 \cdot 0.2+0.9 \cdot 1$ ▸ 0.92

rand() gives a random number between 0 and 1:

0≤rand()≤0.05 : 4€

0.05<rand()≤0.08 : 12€

0.08<rand()≤1 : 0€

X1 is the number of the 4 EURO clients, *X2* is the number of the 12 Euro clients,

X1 and *X2* are dependent binomial distributed random variables.

U is the revenue

$U=4 \cdot X1+12 \cdot X2$

expected value: $E(U)=4 \cdot E(X1)+12 \cdot E(X2)=4 \cdot 2000 \cdot 0.05+12 \cdot 2000 \cdot 0.03$ ▸ 1120

$U=4\cdot X1+12\cdot X2$

expected value: E(U)=4·E(X1)+12·E(X2)= $4\cdot 2000\cdot 0.05+12\cdot 2000\cdot 0.03$ ▸ 1120.

variance: Var(U)=4^2·Var(X1)+12^2·Var(X2)+2·4·12·Cov(X1,X2)

$=16\cdot 2000\cdot 0.05\cdot 0.95+144\cdot 2000\cdot 0.03\cdot 0.97-96\cdot 2000\cdot 0.05\cdot 0.03$ ▸ 9612.8

standard deviation: $\sqrt{9612.8}$ ▸ 98.0449

2) Exact probability:

$P(\text{revenue} < 1000) =$

$$\sum_{i=0}^{83}\left(\sum_{j=0}^{249-3\cdot i}\left(\text{nCr}(2000,i)\cdot \text{nCr}(2000-i,j)\cdot (0.05)^{j}\cdot (0.03)^{i}\cdot (0.92)^{2000-i-j}\right)\right) \blacktriangleright 0.104925$$

i is the number of the 12 Euro guests, j is the number of the 4 Euro Gäste

$83\cdot 12$ ▸ 996 : maximum value for i= 83 (j=0)

$249\cdot 4$ ▸ 996 : maximum value for j=249 (i=0)

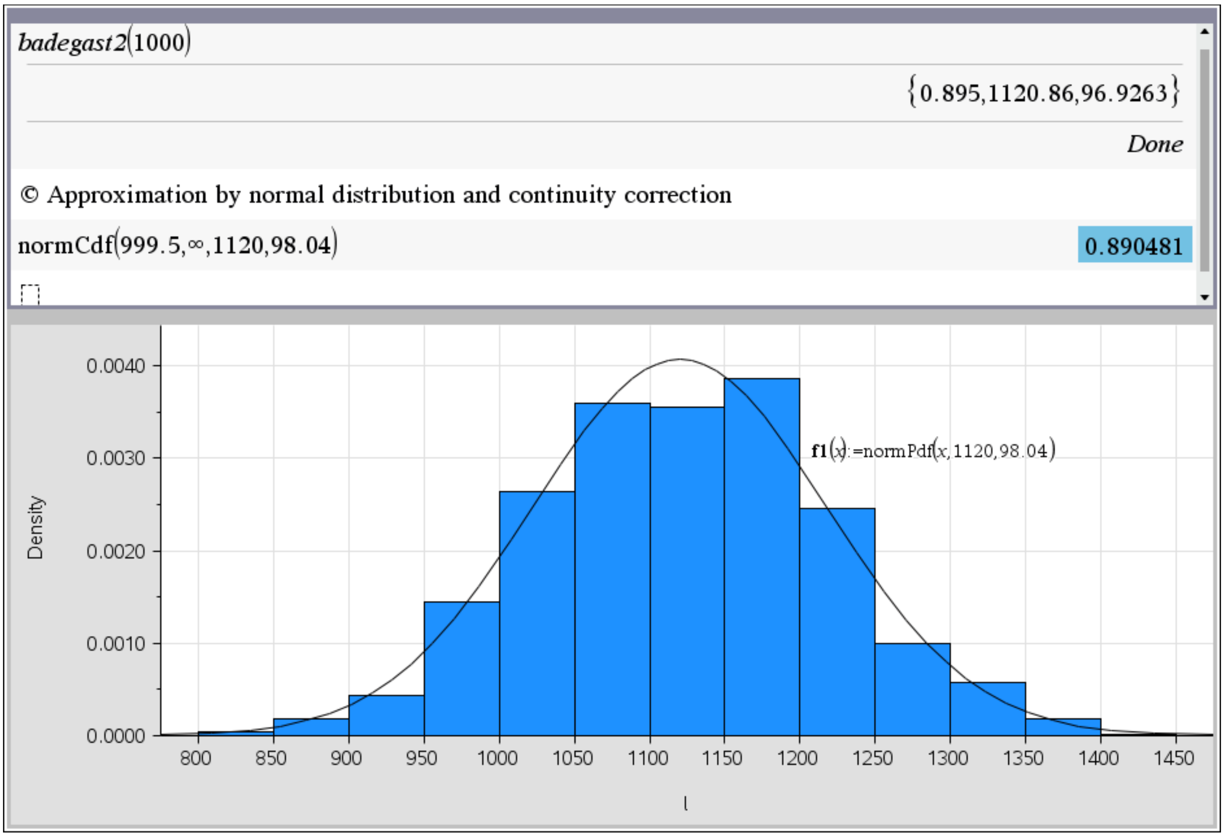
E.g. if i=10 then j can be maximum 249−3.10=219 → $10\cdot 12+219\cdot 4$ ▸ 996

$P(\text{revenue} \geq 1000) = 1- P(\text{revenue} < 1000) =1-0.104925$ ▸ 0.895075

Simulation confirms the calculation!

"badegast2" stored successfully

```
Define badegast2(n)=
Prgm
Local sim,ums,s,s_ums,i,anz,x1,x2
l:={ }
© l is a global variable (works only in a program, not in a function!)
anz:=0
s_ums:=0
For i,1,n
sim:=rand(2000)
x1:=countIf(sim,0≤?≤0.05)
x2:=countIf(sim,0.05<?≤0.08)
ums:=4· x1+12· x2
l:=augment(l,{ums})
s_ums:=s_ums+ums
If ums≥1000
 anz:=anz+1
EndFor
Disp {1.·anz/n,1.·s_ums/n,1.·stDevSamp(l)}
EndPrgm
```



File: bad_3.tns (German) and pool_3.tns (English)

Dear Guido,

many thanks that you accepted the challenge.

Sebastian Rauh – he provided the problem – has also suspected that it must be the multinomial distribution.

I had also the idea that the two events (spending 4€ or 12€) should be dependent, but I was unable to use this information. (Shame on me!)

Once more many thanks. This will make a very nice contribution in our next newsletter, which might be useful for teaching in the classroom.

What great a solution! Excellent idea to apply the inverse probability and really clever: the double sum, especially using the 3i. Reading this, then it seems tob e very simple …

You should absolutely publish this. It is really beautiful mathematics (born from an extremely silly task).

Best regards

Sebastian Rauh

But this is not the end of the story, Josef

Thanks for Sebastian Rauh's positive comment.

Find attached calculation and simulation for independent random variables X1 and X2. Unfortunately, this **wrong** solution gives almost the same results as the correct one.

Have save trip to and from China and best regards

Guido

Calculation of the probability with independent binomial ditributed random variables X1 und X2, Revenue U=4X1+12X2.

variance: Var(U)=16Var(X1)+144Var(X2)=$16\cdot 2000\cdot 0.05\cdot 0.95+144\cdot 2000\cdot 0.03\cdot 0.97 \blacktriangleright 9900.8$

standard deviation: $\sqrt{9900.8} \blacktriangleright 99.5028$

Because of independency: P(X2=i und X1=j) = P(X2=i).P(X1=j)

$$P(U<1000)=\sum_{i=0}^{83}\left(\sum_{j=0}^{249-3\cdot i}\left(\mathrm{nCr}(2000,i)\cdot(0.03)^i\cdot(0.97)^{2000-i}\cdot \mathrm{nCr}(2000,j)\cdot(0.05)^j\cdot(0.95)^{2000-j}\right)\right) \blacktriangleright 0.108384$$

P(U≥1000)= 1− P(U<1000) = $1-0.108384 \blacktriangleright 0.891616$

Probability with dependent variables X1 and X2 (correct solution) is

P(U≥1000) = 0.8951 (see swim_bath_2.tns)

Unfortunately, the difference to the other solution (independent X1 and X2), which is the wrong solution is only 0.35%, the standard deviations 99.50 (independent) versus 98.04 (dependent) are nearly the equal, too.

One cannot detect the difference by simulation.

badegast(5000)	{0.8928,1120.94,99.876}

badegast 1/14

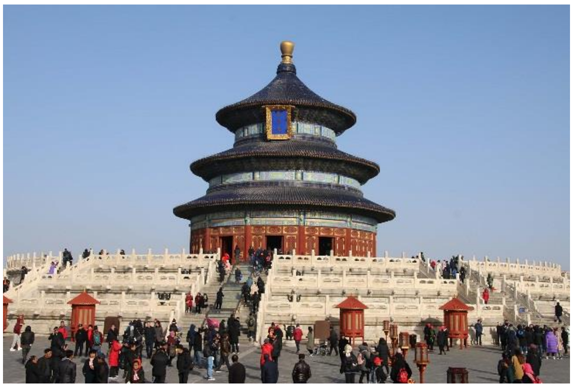
```
Define badegast(n)=
Func
Local ums,s_ums,i,anz,x1,x2,l
l:={}
anz:=0
s_ums:=0
For i,1,n
x1:=randBin(2000,0.05)
x2:=randBin(2000,0.03)
ums:=4·x1+12·x2
l:=augment(l,{ums})
s_ums:=s_ums+ums
If ums≥1000
 anz:=anz+1
EndFor
Return 1.·{anz/n,s_ums/n,stDevSamp(l)}
EndFunc
```

File: bad_4.tns (German) and pool_4.tns (English)

Stone Wood, Kunming

Jade Dragon Snow Mountain, Lijiang

Rice Terraces, Yuanyang



Heaven Temple, Beijing

Lama Temple, Beijing

Panda Station, Chengdu

Great Wall, Jian Shan Ling

Great Wall, Jian Shan Ling