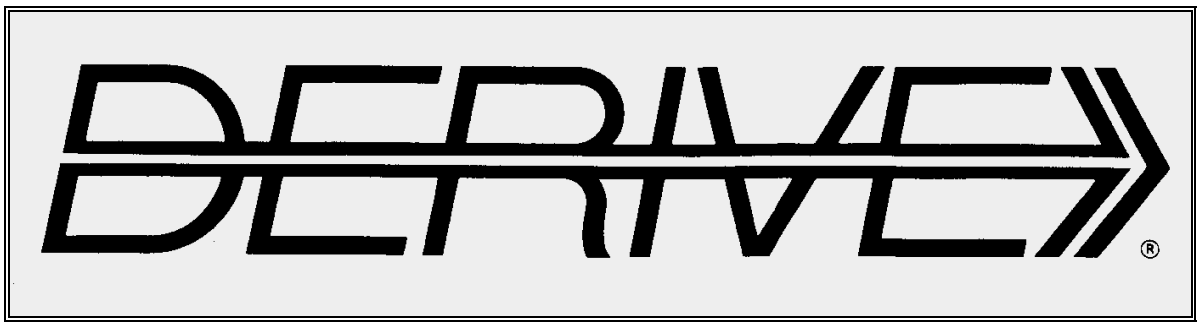
THE DERIVE - NEWSLETTER #88

ISSN 1990-7079

THE BULLETIN OF THE



USER GROUP

+ CAS-TI

Contents:

December 2012

There were some requests about DERIVE's compatibility with WINDOWS 8. I can appease all WIN 8 users – and all of you who intend to change to WIN 8. There are no problems, Josef.

DERIVE and WINDOWS 8

Our member Günter Schödl provided some information concerning DERIVE and WINDOWS 8:

Hello Josef!

Derive can be installed under Win 8 without any problems. Like under Win 7 the Help file cannot be accessed without a patch.

What you also can do is using Hyper-V under Win8 (it must be installed as a Windows component using appwiz.cpl, activate hardware virtualisation in the Bios of the PC), then you can install a virtual WinXP or Win 7 and run DERIVE.

The link for the patch (WinHlp32.exe) is

http://www.microsoft.com/de-de/download/details.aspx?id=35449

Then DERIVE will run as usual.

Greetings
Günter

Another valuable note from Günter:

There is a nasty Macro-error message appearing when calling the DERIVE Online-Help. You can find advice how to avoid this message at:

http://support.microsoft.com/kb/917607/de (German)

http://support.microsoft.com/kb/917607/en (English)

Latest news from TIME 2014

It is a great pleasure to announce the list of keynote speakers for TIME 2014:

Peter Baumgartner (Danube University Krems)
Regina Bruder (Technical University Darmstadt)
Bruno Buchberger (RISC Institute Linz)
Pavel Pech (University of South Bohemia, Budejovice)
Gilles Picard (École de technologie supérieure, Montréal)
Marlene Torres-Skoumal (VIS and International Baccalaureate Schools)

Liebe DUG-Mitglieder,

Ich musste (und wollte) bis heute auf einen schon lange versprochenen Beitrag warten. Daher die Verspätung mit DNL#88.

Es war meine feste Absicht, die letzte Folge von Piotr Trebisz' Schneckenhäusern zu bringen und war auch schon ziemlich fertig mit der Übersetzung als Statistik 4 von Guido Herweyers eine ungeheure Eigendynamik bekam. Einerseits konnte ich nicht widerstehen, die Simulationen mit DERIVE nachzuvollziehen und andererseits - wichtiger – gab es für mich noch offene Fragen, die Guido in dankenswerter Weise sehr umfassend beantwortete. Der Anhang zu den Funktionen von Zufallsvariablen ist für mich sehr wertvoll.

Fred Tydeman hat im letzten DNL beklagt, dass sich zu seinem in den DERIVE News vorgestellten Problem mit dieser speziellen Folge kein DERIVIAN gemeldet hat. Nun, in diesem DNL gibt es gleich drei z.T. sehr ausführliche Beiträge zu dieser Folge.

Es ist erfreulich, dass gelungen ist, Johann Wiesenbauer zu einer neuen Ausgabe seine Titbits zu ermuntern. In einem zweiteiligen Aufsatz beschreibt er eine DERIVE Implementierung des nach dem RSA-Algorithmus bekanntesten Public-Key Verschlüsselungssystems.

Bitte beachten Sie die wertvollen Hinweise zu DERIVE & WIN 8, die uns Günter Schödl zur Verfügung stellt. Er ist da immer sehr rasch voll informiert – und damit wir mit ihm, herzlichen Dank dafür lieber Günter.

Freuen Sie sich mit mir auf die nächsten Ausgaben mit u.a. Beiträgen über Primzahlen und Taylorreihen (D. Oertel), das Brüsseler Tor (E. van Lantschoot) und natürlich über die Schneckenhäuser, von denen Sie eine Kostprobe auf Seite 53 sehen können.

Es bleibt mir noch, Ihnen allen ein glückliches, erfolgreiches und gesundes Jahr 2013 zu wünschen.

Dear DUG Members,

I had (and I wanted) to wait for a long promised contribution until today. And this is the reason for the delay of DNL#88.

It was my intention to include the last part of Piotr Trebisz' snail house series and its translation was almost ready. But then Statistics 4 (Guido Herweyers) got an immense self dynamic. I could not resist reproducing the simulations using DERIVE and at the other hand – more important – there were some open questions for me which were kindly answered very comprehensive. In my opinion the appendix to the functions of random variables is very valuable.

In the last DNL Fred Tydeman complained that no DERIVIAN responded to his special sequence problem presented in the DERIVE News Group. Now, in this DNL he – and all our members – can find three very detailed contributions.

I am very happy that I was able to encourage Johann Wiesenbauer to a new Titbits-contribution. He describes in two parts a DERIVE implementation of the famous ElGamal encryption algorithm which is together with the RSA-encryption the best known public-key encryption method.

Please notify the valuable notes to DERIVE & WIN 8 which are provided by Günter Schödl. He is always fully informed about hard- and software news. Many thanks for your support, dear Günter.

Be looking forward to the next issues with among others contributions on prime numbers (D. Oertel), the "Brussel's Gate" (E. van Lantschoot) and on the snail shells. You can have a tasting on page 53.

For me remains wishing you a happy, successful and healthy year 2013.

Viele Grüße, kindest regards

Download all *DNL*-DERIVE- and TI-files from

http://www.austromath.at/dug/

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 3136365
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 2013

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
Tools for 3D-Problems, P. Lüke-Rosendahl, GER
Hill-Encryption, J. Böhm, AUT
Simulating a Graphing Calculator in *DERIVE*, J. Böhm, AUT
Do you know this? Cabri & CAS on PC and Handheld, W. Wegscheider, AUT
An Interesting Problem with a Triangle, Steiner Point, P. Lüke-Rosendahl, GER
Graphics World, Currency Change, P. Charland, CAN
Cubics, Quartics – Interesting features, T. Koller & J. Böhm, AUT
Logos of Companies as an Inspiration for Math Teaching
Exciting Surfaces in the FAZ / Pierre Charland´s Graphics Gallery
BooleanPlots.mth, P. Schofield, UK
Old traditional examples for a CAS – what's new? J. Böhm, AUT
Truth Tables on the TI, M. R. Phillips, USA
Where oh Where is It? (GPS with CAS), C. & P. Leinbach, USA
Embroidery Patterns, H. Ludwig, GER
Mandelbrot and Newton with *DERIVE*, Roman Hašek, CZK
Some Projects with Students, R. Schröder, GER
Structures in the Set of Prime Numbers, D. Oertel, GER
Dirac Algebra, Clifford Algebra, D. R. Lunsford, USA
Laplace Transforms, ODEs and CAS, G. Picard & Ch. Trottier, CAN
A New Approach to Taylor Series, D. Oertel, GER
Cesar Multiplication, G. Schödl, AUT
Henon & Co; Find your very own Strange Attractor, J. Böhm, AUT
Rational Hooks, J. Lechner, AUT
Mathematical Model for Snail Shells (4), P. Trebisz, GER
Simulation of Dynamic Systems with various Tools, J. Böhm, AUT
An APL-like SHAPE function in DERIVE 6, D. R. Lunsford, USA
Brussels Gate in Dendermonde, E. van Lantschoot, GER
Recursive Series of Numbers, Polynomials and Functions, D. Halprin, AUS
and others

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

Statistics with TI-Nspire 3.1/3.2 (Part 4)

Visualising and Simulating Dynamically with TI-Nspire

Guido Herweyers, KHBO Campus Oostende
guido.herweyers@khbo.be

Part 3: Discovering Probability Distributions

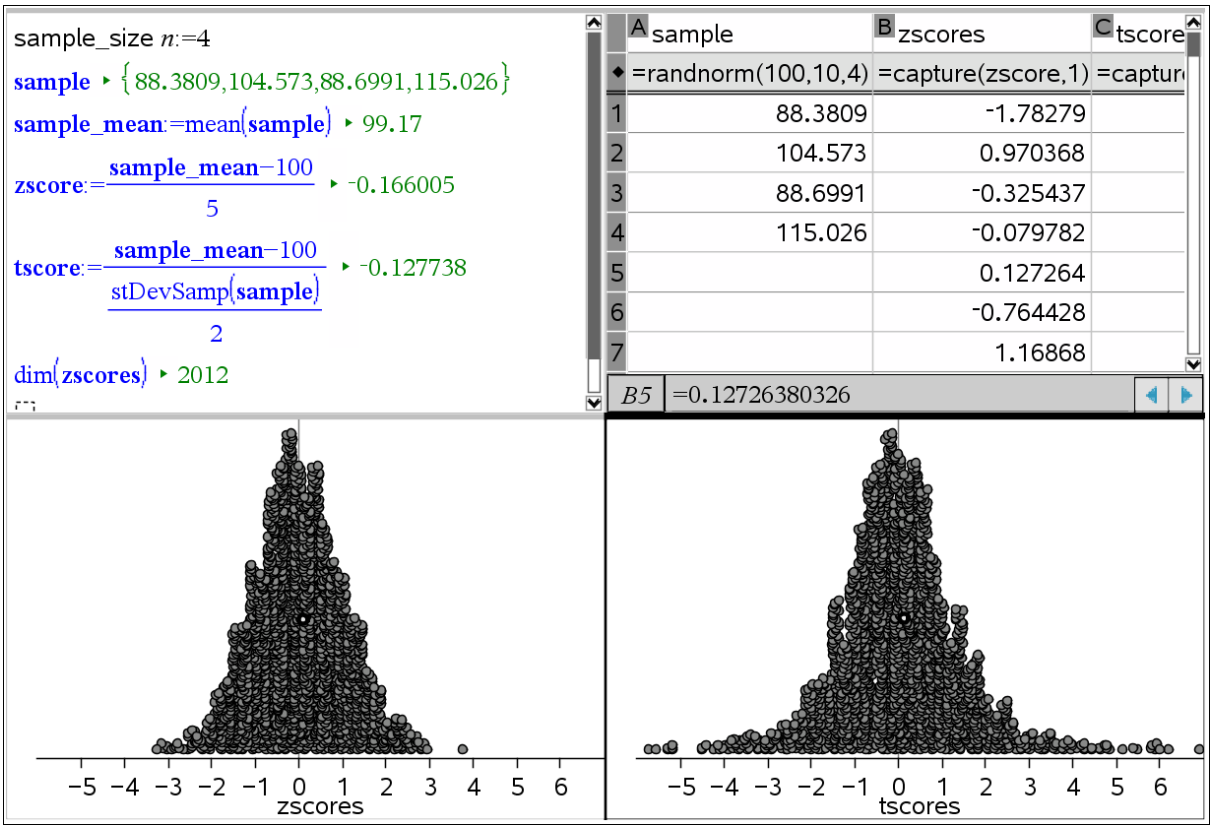
(1) z-scores versus t-scores

Set $X \sim N(\mu,\sigma)$ and take a sample of size n from this population. Then the mean of this sample is given by: $\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Standardisation gives $Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$.

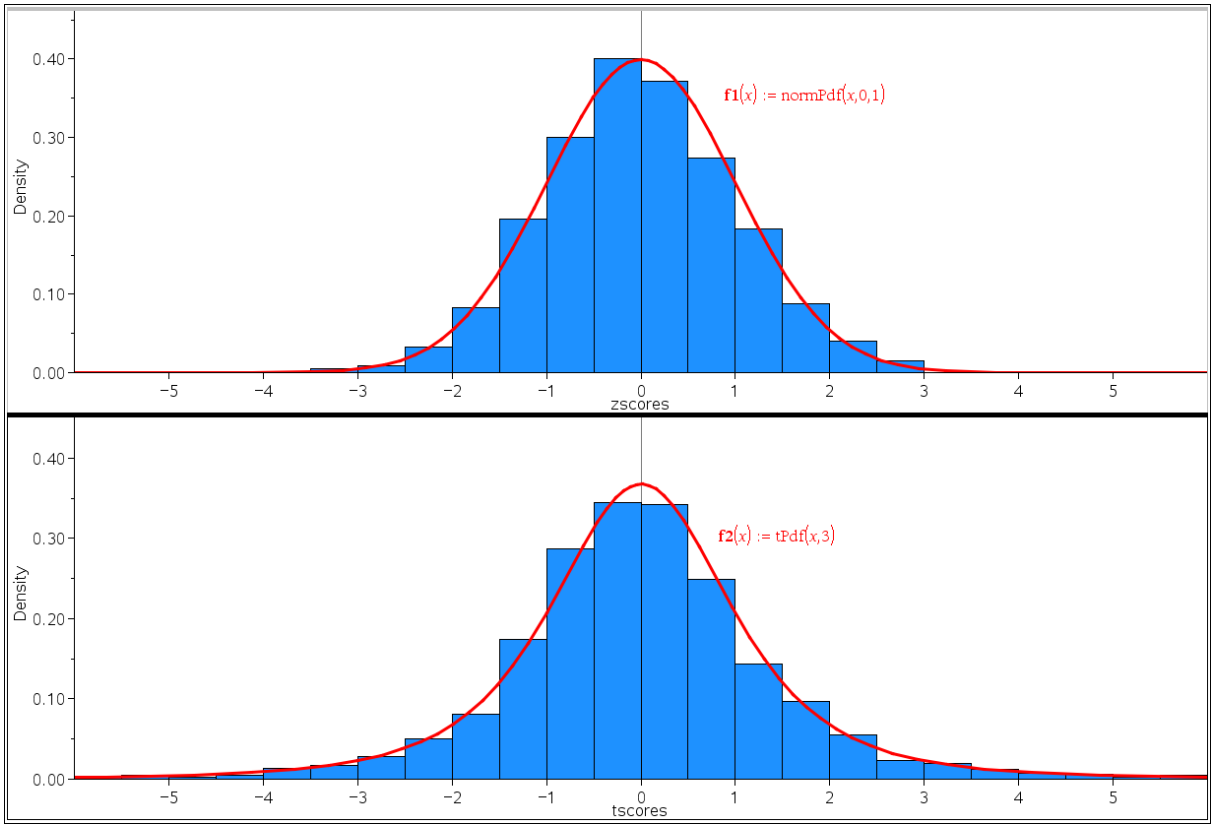
In case of an unknown standard deviation σ of the population σ will be estimated by the standard deviation of the sample s. Then we have a new random variable $T = \frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}}$.

We compare the "z-scores" $z = \frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ with the "t-scores" $t = \frac{\overline{x}-\mu}{\frac{s}{\sqrt{n}}}$ using their density diagrams

performing a simulation with the population mean $\mu = 100$, the population standard deviation $\sigma = 10$ and a sample size $n = 4$. (You can find hints how to reproduce the screen below in DNL#87 page 6.)



The t-scores show a larger dispersion than the z-scores



The t-scores are following a t-distribution with 3 degrees of freedom (3 = n–1). Its graph is also bell shaped, but it is **not** the normal distribution!

(Note: You can add the graphs of the distributions via **Data & Statistics** in the Documents Toolbox. Choose options **4: Analyze** > **4: Plot Function**.)

(2) Functions of random variables

Take any random number X in the interval [0, 1]. Then X (upper case!) is uniformly (or rectangular) distributed following the density function

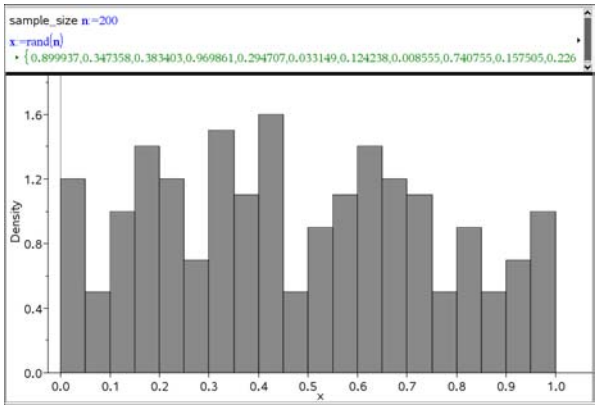
$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{else} \end{cases}.$$

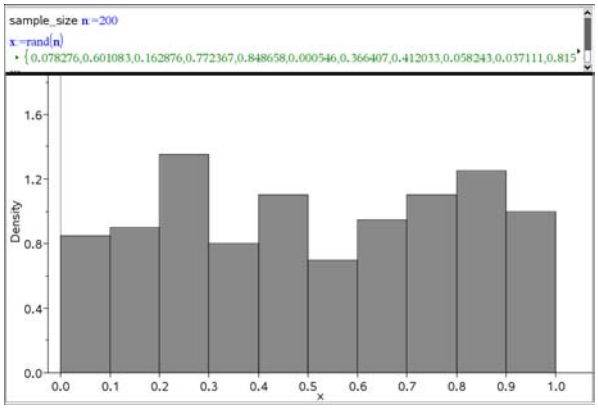
If a random number between 0 and 1 is generated then the random variable X gets a certain value x (lower case!) according to the probability mechanism of the density function f.

Generating many random numbers in the interval [0, 1] and presenting the respective density histogram we should observe that this histogram will become more and more stable and finally "converge" to the density function (probabilities are relative frequencies for many tries: law of large numbers).

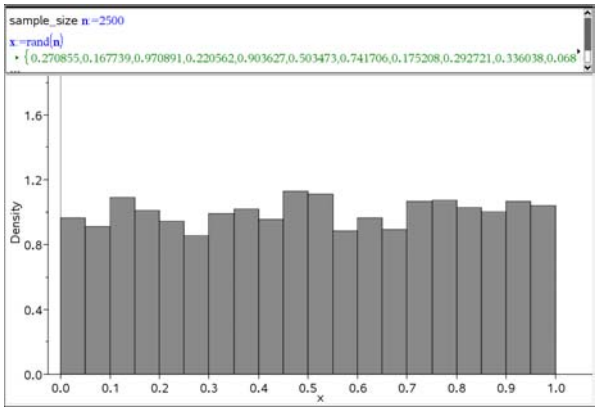
200 simulations (200 random numbers generated) will show big variation in the density histogram:.

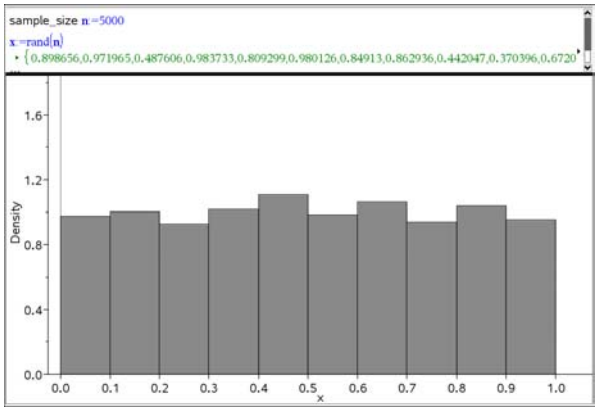
In the following we show two simulations with 200 random numbers uniformly distributed in [0, 1], the next one is the presentation of 2500 random numbers and finally we will generate 5000 random numbers in the interval.





For $n = 2500$ and then for $n = 5000$ variation becomes much smaller. The density histogram looks more and more like a rectangle (in case of not too small class widths).

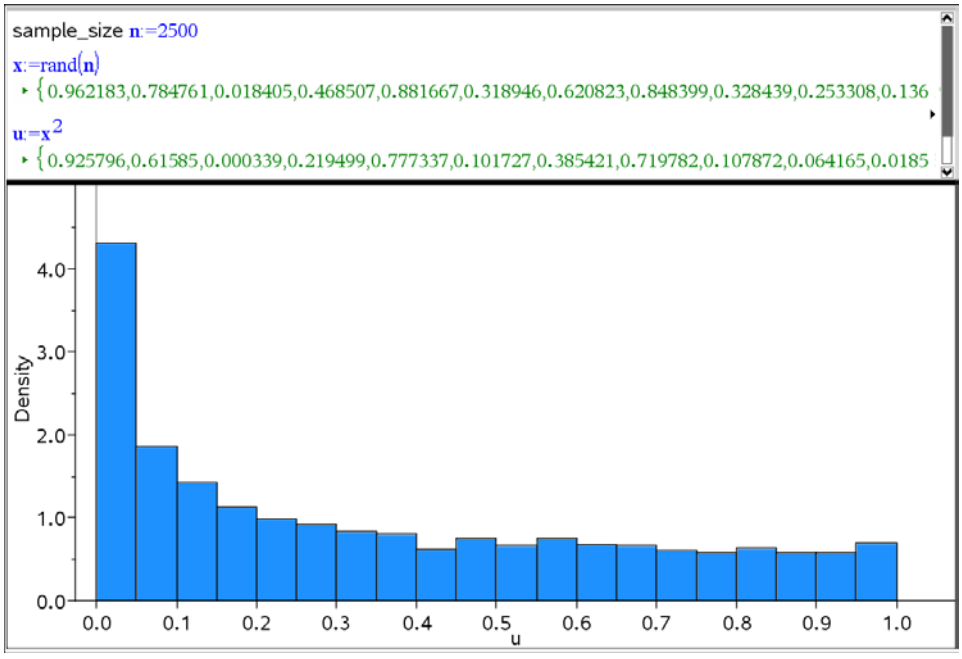




(Note: for $n = 10000$ we receive an error message: resource exhaustion.)

In this way one can get a good impression of the probability distribution of a random variable by simulation.

Variable X can be used to define new random variables, e.g. $U = X^2$. The probabiliy distribution (the density function) of U is approximated by the density historgram of a large number of generated random values $u = x^2$.

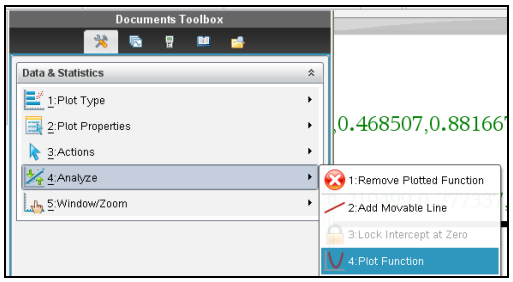


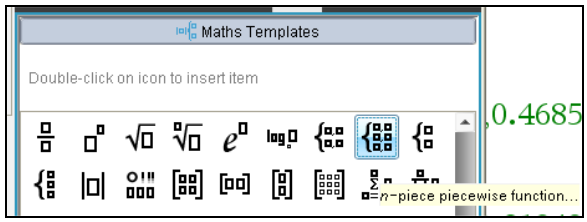
The density function of U is the first derivative of the distribution function F_U of U:
For $0 \le x \le 1$:

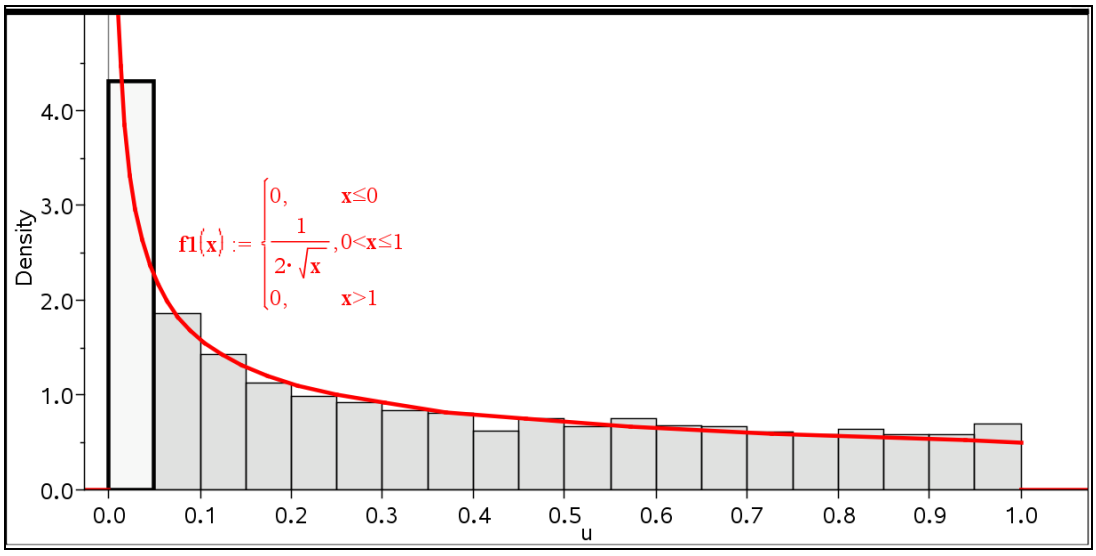
$$F_U(x) = P(U \le x) = P(X^2 \le x) = P(-\sqrt{x} \le X \le \sqrt{x}) = \int_0^{\sqrt{x}} 1 \cdot dt = \sqrt{x}$$

Hence, the density function of U is given by: $f_U(x) = F_U'(x) = \begin{cases} \dfrac{1}{2\sqrt{x}} & \text{for } 0 < x \le 1 \\ 0 & \text{else} \end{cases}$.

Add the graph of the density function using the Nspire-menus given below. The function can be edited by using the respective template for a piecewise defined function.







You are invited to double-check in the notes application that the (improper and defined) integral of the density function over the interval [0, 1] gives the expected result 1. (You can do this also in a calculator page.)

$$\int_0^1 \frac{1}{2\cdot\sqrt{x}}\,dx \blacktriangleright 1$$

(3) Functions of two random variables

Take any random number X and another one Y both from the interval [0, 1]. Assume that both are uniformly distributed in [0, 1].

Using these two variables new random variables can be defined, as for example:

$S = X + Y$, $P = X \cdot Y$, $M = \max(x, y)$, $K = \min(X, Y)$, etc.

What is the probability distribution of these newly generated variables?

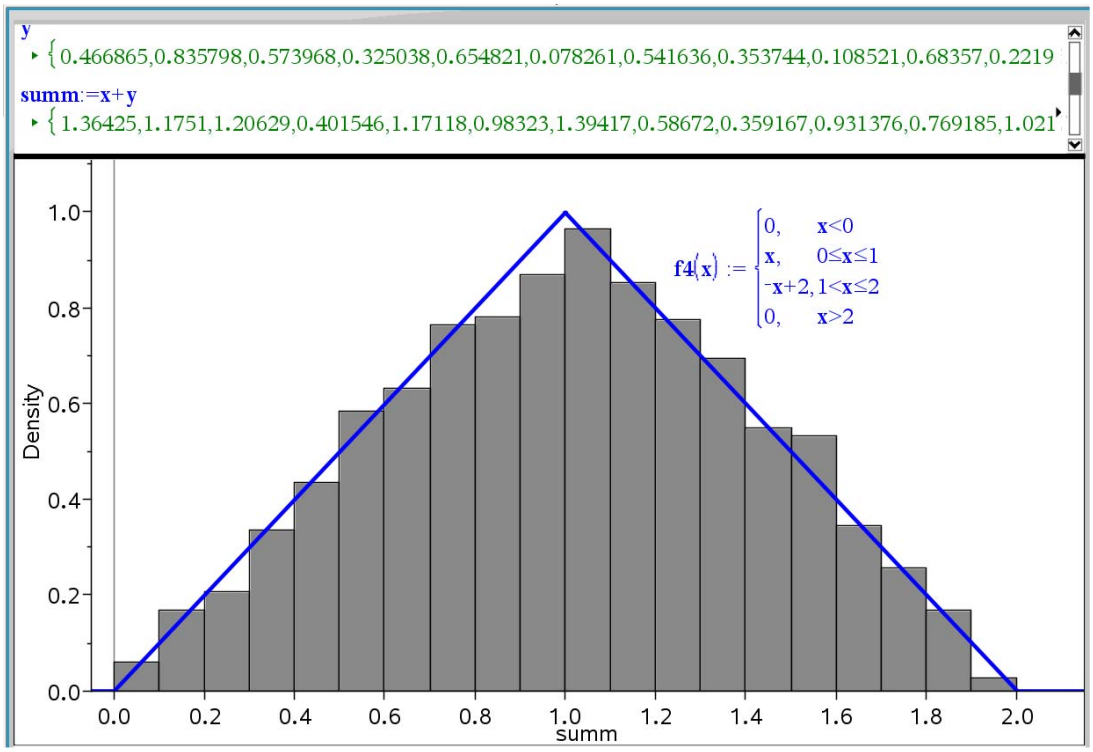
Let's experiment with random variable $S = X + Y$ first. Take two random numbers between 0 and 1 and form the concrete sum $s = x + y$. Then repeat this experiment very often.

Numbers x and y are created following the uniform distribution over [1, 0]. Then the sums are values between 0 and 2. The density diagram of the generated sums shall become closer and closer to the density function of S.

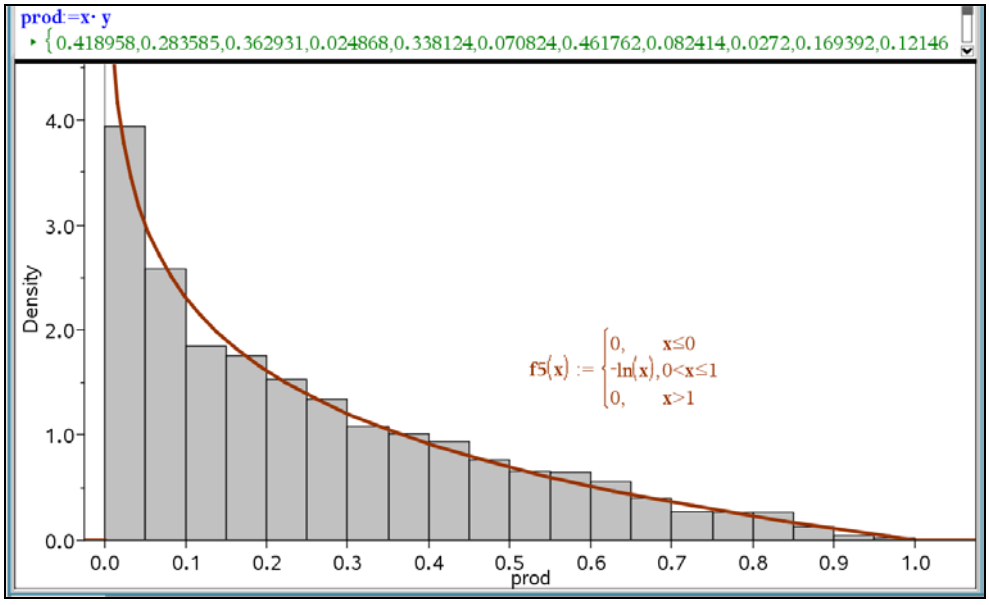
Do the same for variables P, M and K.

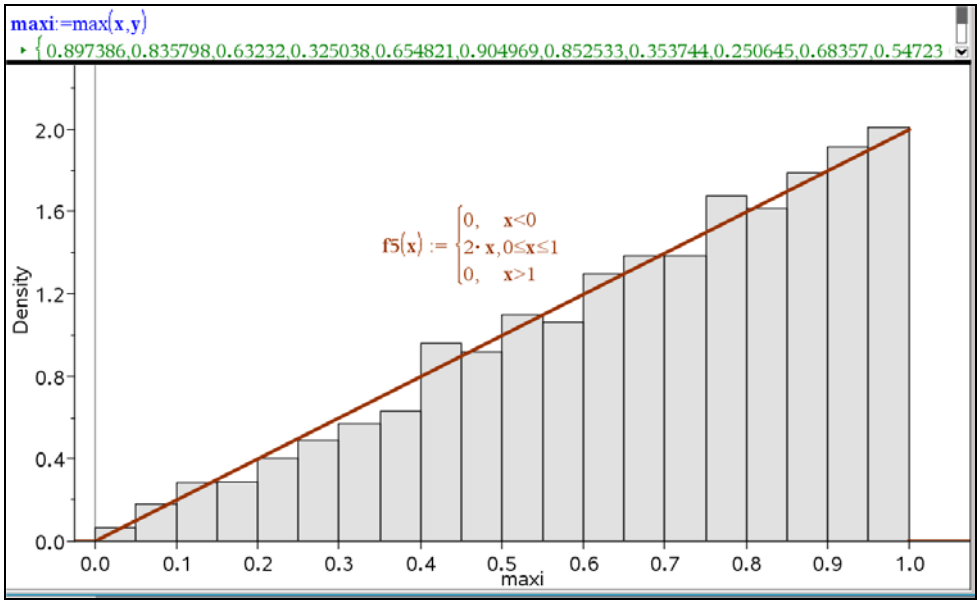
```
sample_size n:=3 ▸ 3
x:=rand(n):y:=rand(n)  (Two commands are separated by a colon. They are both executed when pressing ENTER.)
x ▸ {0.280027,0.588786,0.734613}
y ▸ {0.166901,0.023942,0.38266}
summ:=x+y ▸ {0.446928,0.612728,1.11727}
prod:=x· y ▸ {0.046737,0.014097,0.281107}
maxi:=max(x,y) ▸ {0.280027,0.588786,0.734613}
mini:=min(x,y) ▸ {0.166901,0.023942,0.38266}
```

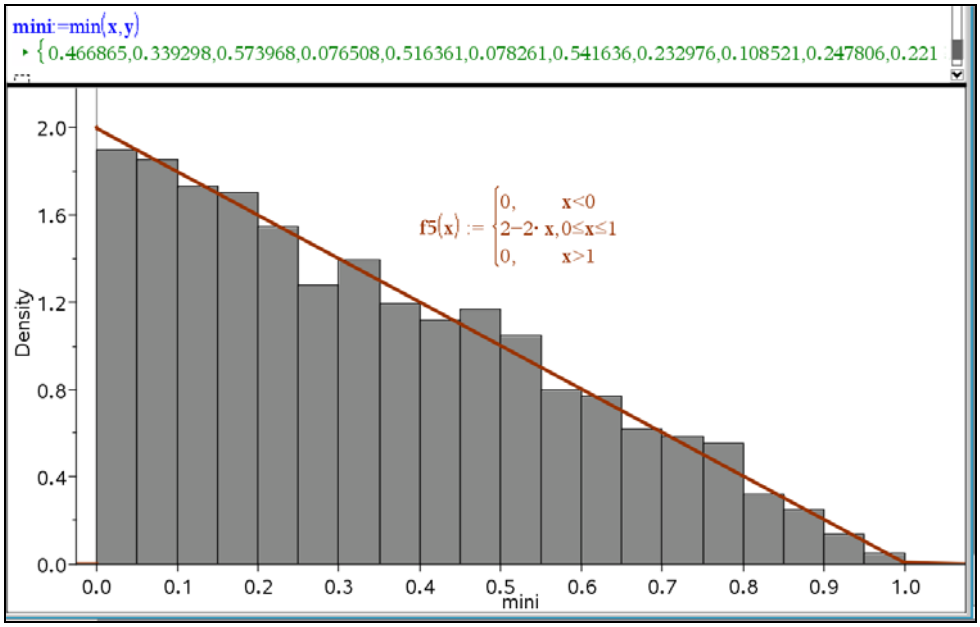
This is the density histogram for a sample size $n = 2500$ together with the respective probability density function.



[1] Information how to find the probability density function $f4(x)$ (pdf) is given later in this DNL.


prod:=x· y
{0.418958,0.283585,0.362931,0.024868,0.338124,0.070824,0.461762,0.082414,0.0272,0.169392,0.12146
f5(x) := 0, x≤0
-ln(x), 0<x≤1
0, x>1
Density
prod


maxi:=max(x,y)
{0.897386,0.835798,0.63232,0.325038,0.654821,0.904969,0.852533,0.353744,0.250645,0.68357,0.54723
f5(x) := 0, x<0
2· x, 0≤x≤1
0, x>1
Density
maxi


mini:=min(x,y)
{0.466865,0.339298,0.573968,0.076508,0.516361,0.078261,0.541636,0.232976,0.108521,0.247806,0.221
f5(x) := 0, x<0
2−2· x, 0≤x≤1
0, x>1
Density
mini

Exercise:

Take random variable $K = \min(X, Y)$ from the last example.

Simulate a sample of size 2500 and find the sample **mean** $\bar{x}$. This is an estimation of the theoretical expected value $E(K) = \mu = \int_0^1 x \cdot (2-2x)\,dx$.

Find also the sample standard deviation (**stdevsamp**), which is an estimation for the theoretical standard deviation $\sigma = \sqrt{\mathrm{Var}(K)} = \sqrt{\int_0^1 (x-\mu)^2 \cdot (2-2x)\,dx}$.

Calculate μ and σ. Check if sample mean and sample standard deviation in fact are lying close to μ and σ. Repeat the experiment for some samples of size 2500 or larger. Repeat the experiment for some samples of size 10. What is your conclusion?

Part 4:

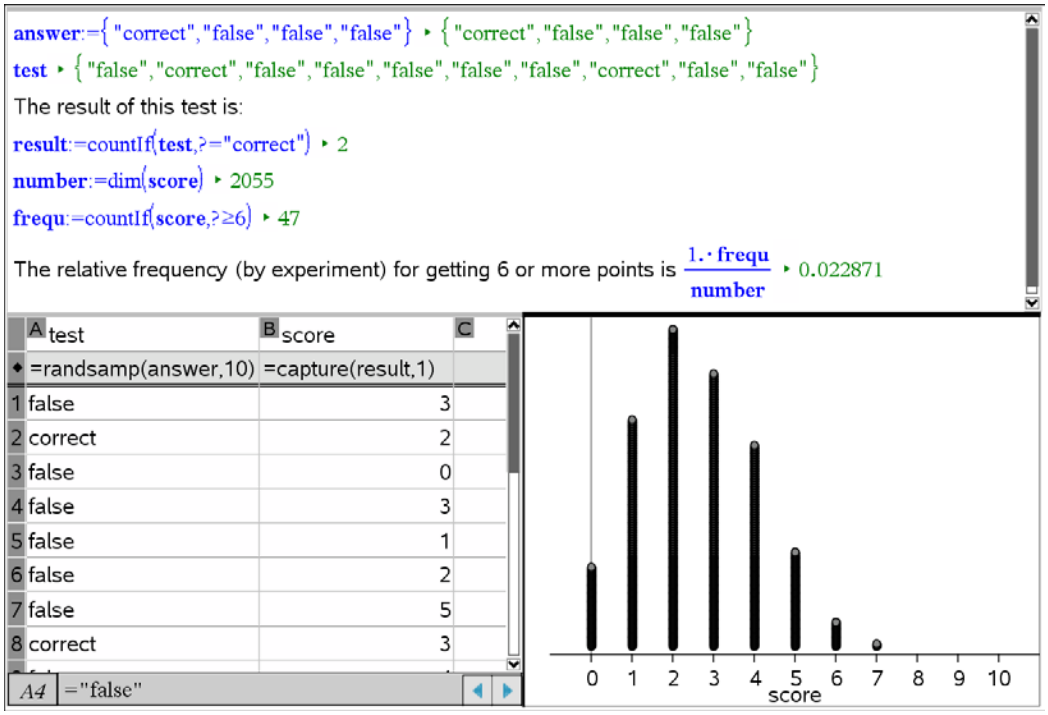
For the statistics background see [6], [7], [8], [9], [10], [11].

Example 1: Does the student say the truth?

A test consists of 10 multiple choice questions with four possible answers one of them being the correct one. A correct answer gives one point, a false one gains no point.

A student reaches 6 points of the 10 possible ones and he assures that he had made a random choice for each question (because he didn't study the subject). Can you believe him?

By simulating this test and keeping the scores by automatically data storing (capture) one can investigate how often a score of 6 or higher is appearing.



The experimental frequency of gaining a score above 5 is according to a simulation of 2055 tests only ~2.3%. It seems to be sure that the student does not tell the truth.

Let X the correct answers when guessing at each problem. Then X follows a binomial distribution with $n = 10$ and success probability $p = 0.25$.

Let's check the experimental frequency table by comparing it with the theoretical probability distribution:

	A k	B theoretical	C by_experiment	D	E
◆	=seq(n,n,0,10)	=binompdf(10,1/4)	=seq(approx(countif		
1	0	0.056314	0.06472		
2	1	0.187712	0.185888		
3	2	0.281568	0.260341		
4	3	0.250282	0.223358		
5	4	0.145998	0.164964		
6	5	0.058399	0.077859		
7	6	0.016222	0.020438		
8	7	0.00309	0.002433		
9	8	0.000386	0.		
10	9	0.000029	0.		
11	10	9.53674E-7	0.		

C **by_experiment**:=seq$\left(\text{approx}\left(\frac{\text{countif}(\textbf{score},?=n)}{\dim(\textbf{score})}\right),n,0,10\right)$

The hypothesis test is performed as follows:

The null hypothesis is set: H_0: $p = 0.25$ (the student guessed)

The alternative hypothesis is: H_1: $p > 0.25$ (the student studied and did not guess with each question having the same success probability)

The test variable is the number of correct answers X with $X \sim B(10,0.25)$ – assuming that H_0 is true.

The observed value of the test variable is $x = 6$. The exceeding probability or the p-value is $P(X \geq 6) = P(6 \leq X \leq 10) =$ binomcdf(10,0.25,6,10) = 1.97%.

$\text{binomCdf}\left(10,\frac{1}{4},6,10\right)$	0.019728
$1-\text{binomCdf}\left(10,\frac{1}{4},5\right)$	0.019728

This probability is very small, thus H_0 is rejected which means that we can be (almost) sure that the student does not tell the truth (and he did study).

Example 2: Is this die a correct one?

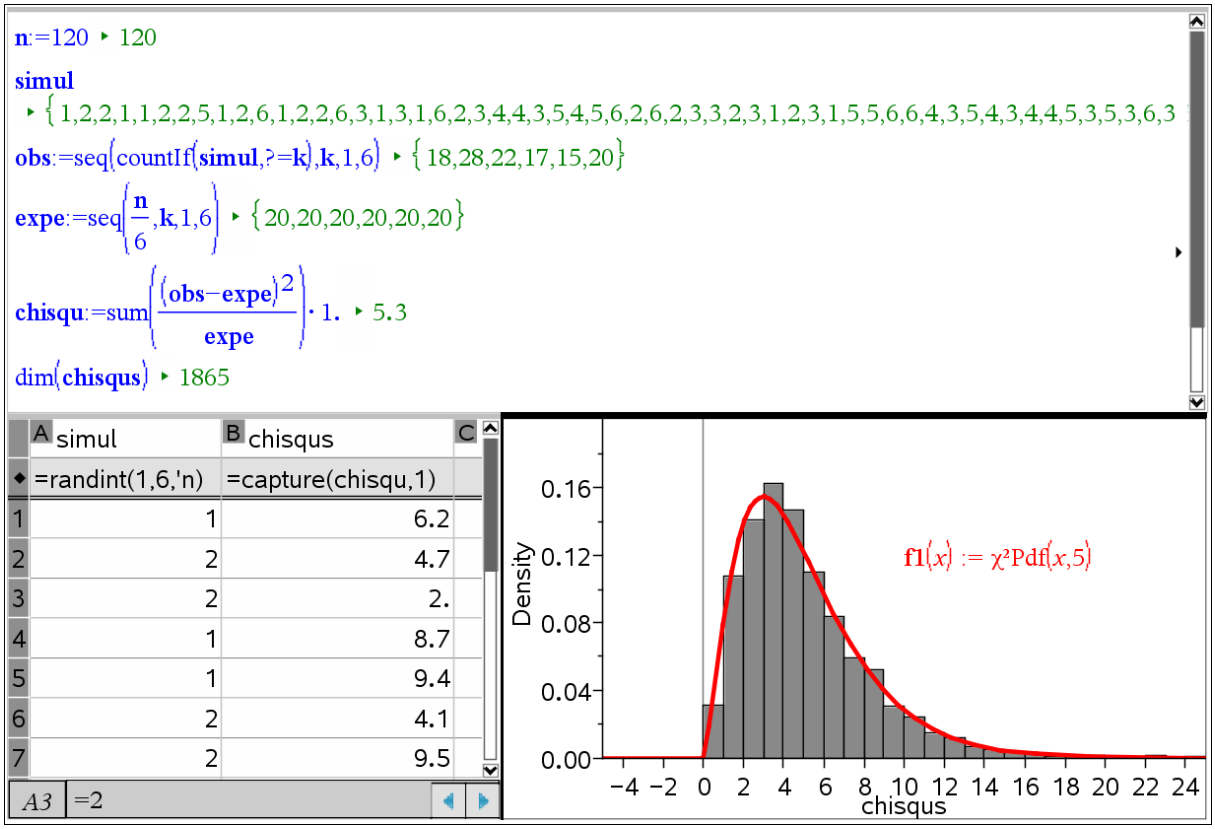
When rolling a correct die 120 times one can expect that the numbers 1 through 6 are appearing approximately 20 times each. Which deviations of this expectation are possible? A measure for the deviation of the expected values is the

$$\text{Chi-Square value: } \chi^2 = \sum \frac{(\text{observed value} - \text{expected value})^2}{\text{expected value}} = \sum_i \frac{(O_i - e_i)^2}{e_i}.$$

If this random variable turns out to be small then this die should be correct. But if one finds a large χ^2-value then it is possible that the true probability distribution does not correspond with the expected discrete uniform distribution.

For getting an impression of the distribution of the random variable χ^2 it will be sufficient to perform a large number of simulations in order to create a density diagram of the observed χ^2-values which will become stable on the long run.

The distribution is right skewed. As you can learn from the graphic representation the χ^2-distribution with 5 degrees of freedom forms a good model for the probability distribution.



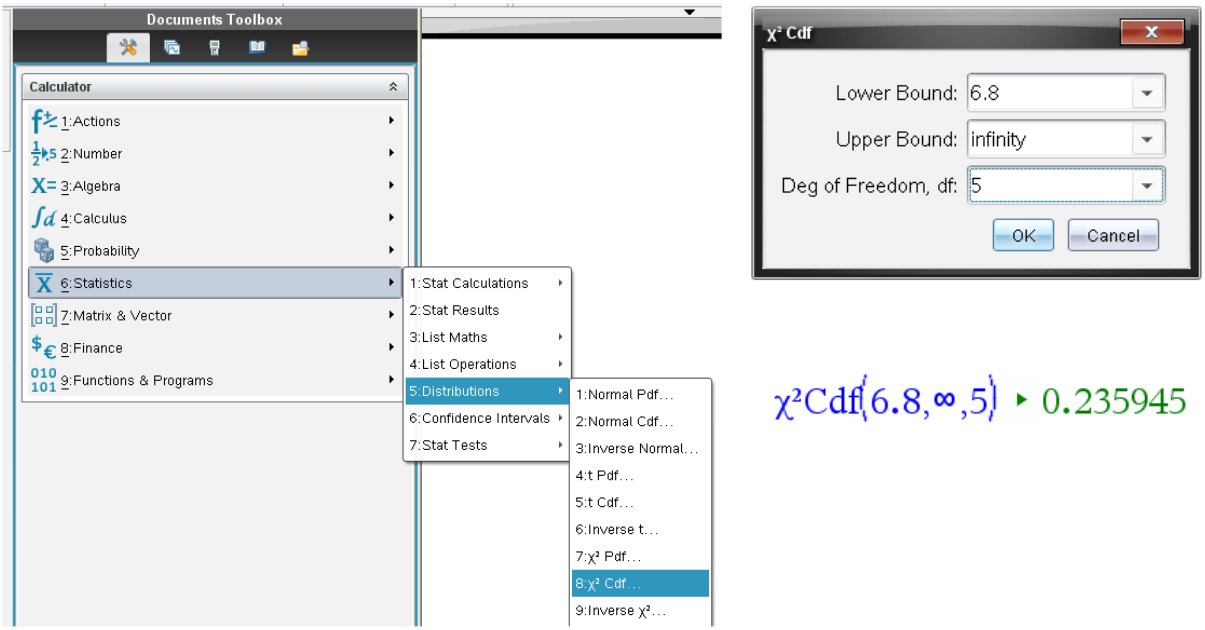
Roll a die 120 times in order to check if it is a correct one. Assume that you will get the following results:

numbers	1	2	3	4	5	6
observed	17	12	23	18	25	15
expected	20	20	20	20	20	20

The respective χ^2-value can be calculated as:

$$\chi^2 = \frac{(17-20)^2}{20} - \frac{(12-20)^2}{20} - \frac{(23-20)^2}{20} - \frac{(18-20)^2}{20} - \frac{(25-20)^2}{20} - \frac{(15-20)^2}{20} = 6.8$$

Can we expect this result? The *p*-value is (we will ask TI-NspireCAS) …



… 0.236. Thus, the chance that the value of the χ^2-statistics is at least as great as the observed value of 6.8 is 23.6% and this is not extra ordinary.

The null hypothesis can not be rejected by this "goodness-of-fit test":

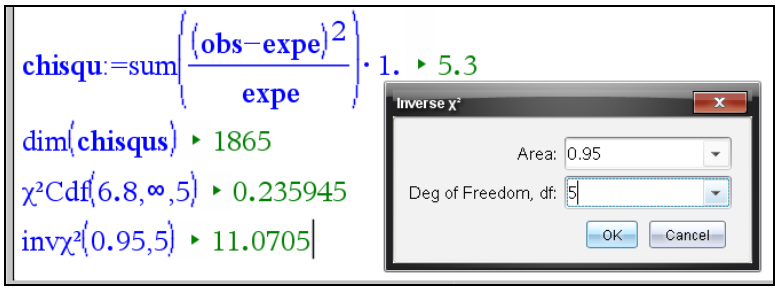
H_0: the die follows a uniform discrete distribution

versus

H_1: the rolled numbers are not uniformly distributed.

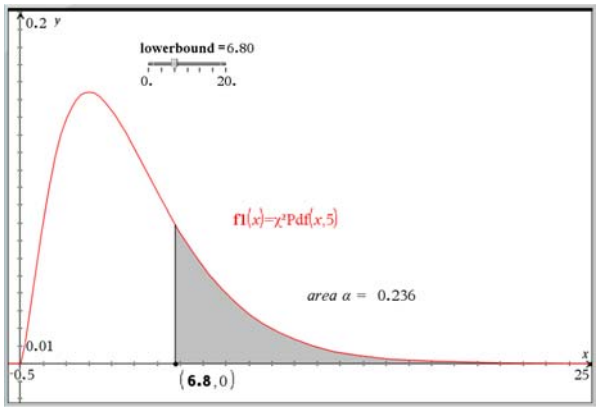
The question is: which observed χ^2-value will lead us to a rejection of the null hypothesis at a significance level of $\alpha = 5\%$?

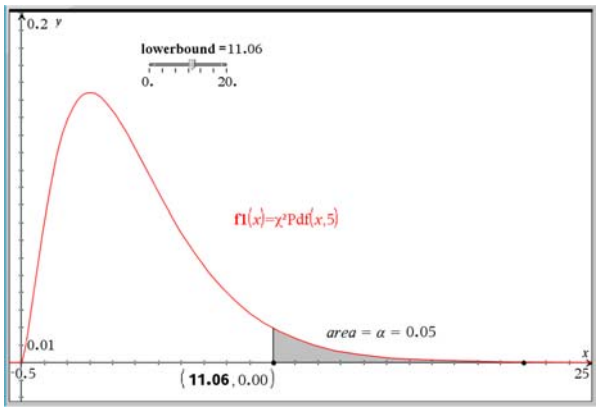
This critical value is obtained by using the inverse of the χ^2 distribution function.



The critical value is 11.07.

The critical value can also be found using a slider:





Example 3: Testing a proportion

A sweets manufacturer affirms that 30% of a special sort of candies is yellow. You want to check this. So you buy a bag of 50 sweets and find out that 23 of them are yellow – this is 46%!

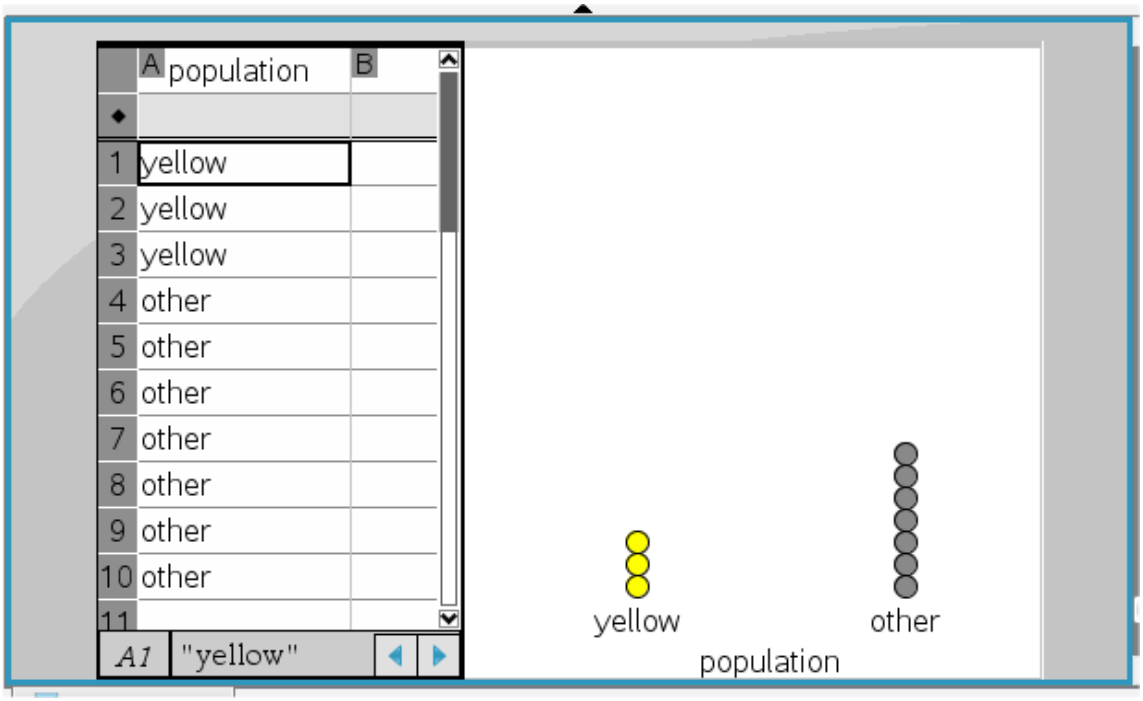
Is this fact sufficient for concluding that the manufacturer does not tell the truth?

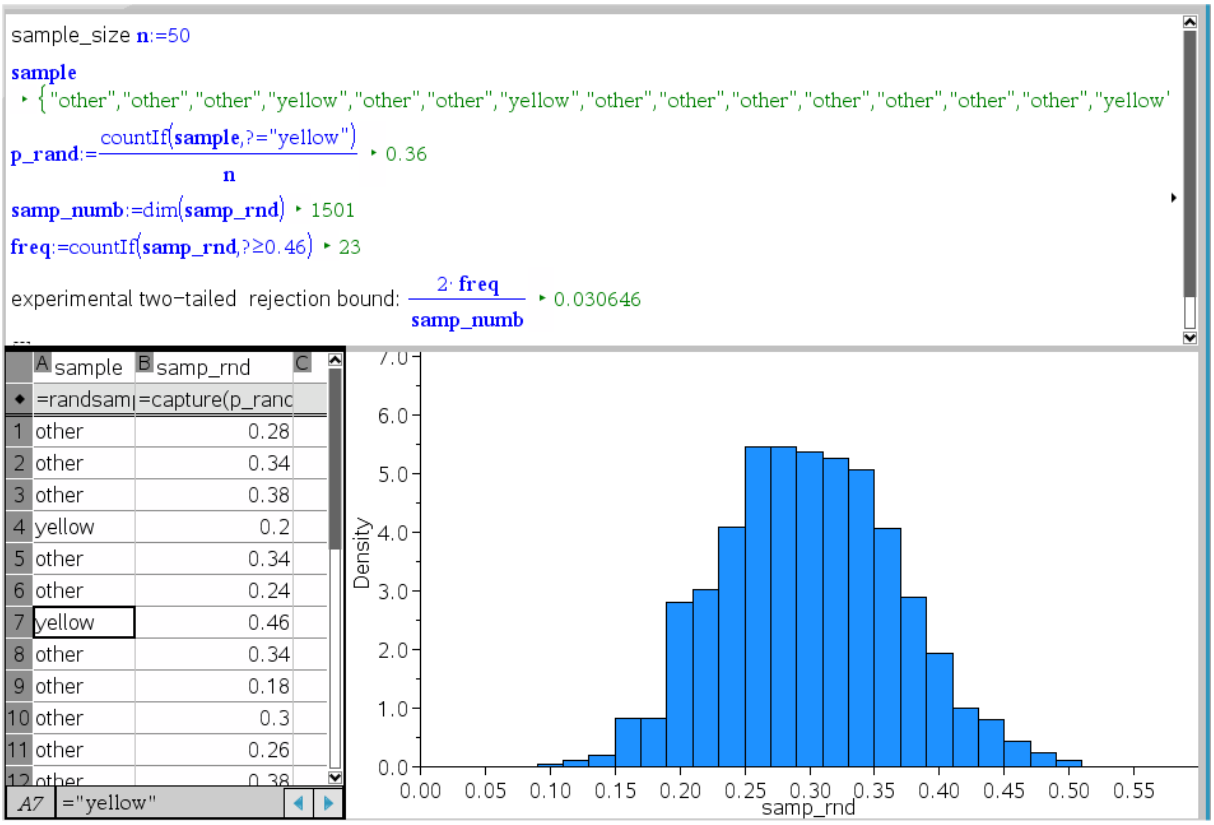
Let p the proportion of the population.

The hypothesis test is H_0: $p = 0.3$ versus H_1: $p \neq 0.3$ (a two-tailed test).

The simulation:

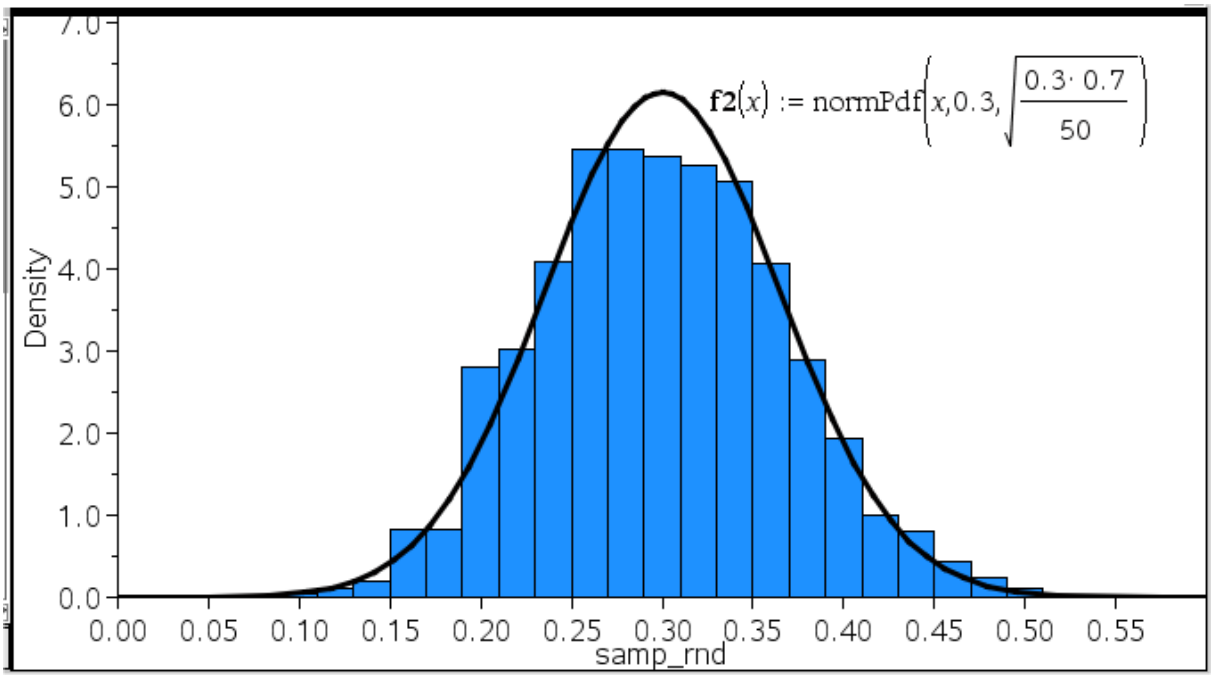
Assume that H_0 is true. Take a sample of 50 candies out from a population containing 30% yellow sweets.
Investigate how much the sample proportion $\hat{p}$ can diverge from the (fixed) population proportion p.





The experiment shows that the two tailed exceeding probability is 3.06% (having drawn 1501 samples). The conclusion is that we can reject H_0 with a 5% significance level.

Although the probability distribution of the sample proportions is a discrete one – which will become clear taking a class width less 0.02 – it can be approximated by a normal distribution with mean $\mu = 0.3$ and standard deviation $\sigma = \sqrt{\frac{0.3 \cdot 0.7}{50}} \approx 0.65$.



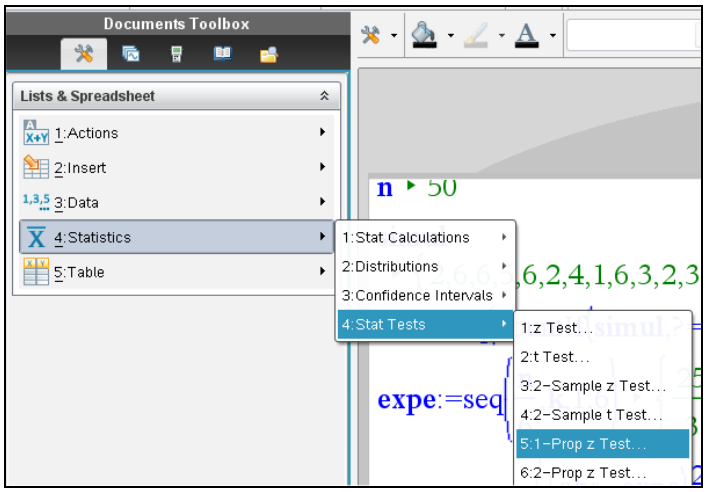
Taking the normal distribution as probability model the two tailed exceeding probability is given by $2 \cdot P(\hat{P} \geq 0.46) \approx 1.4\%$ (without continuity correction!)

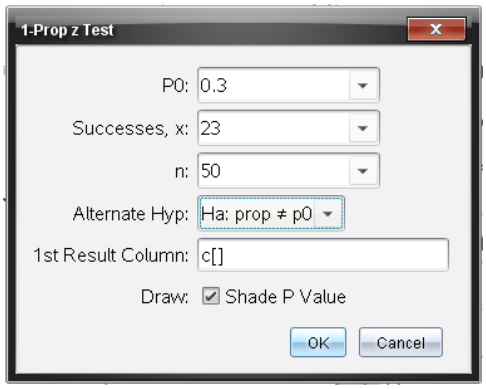
$$\sqrt{\frac{0.3 \cdot 0.7}{50}} \qquad 0.064807$$

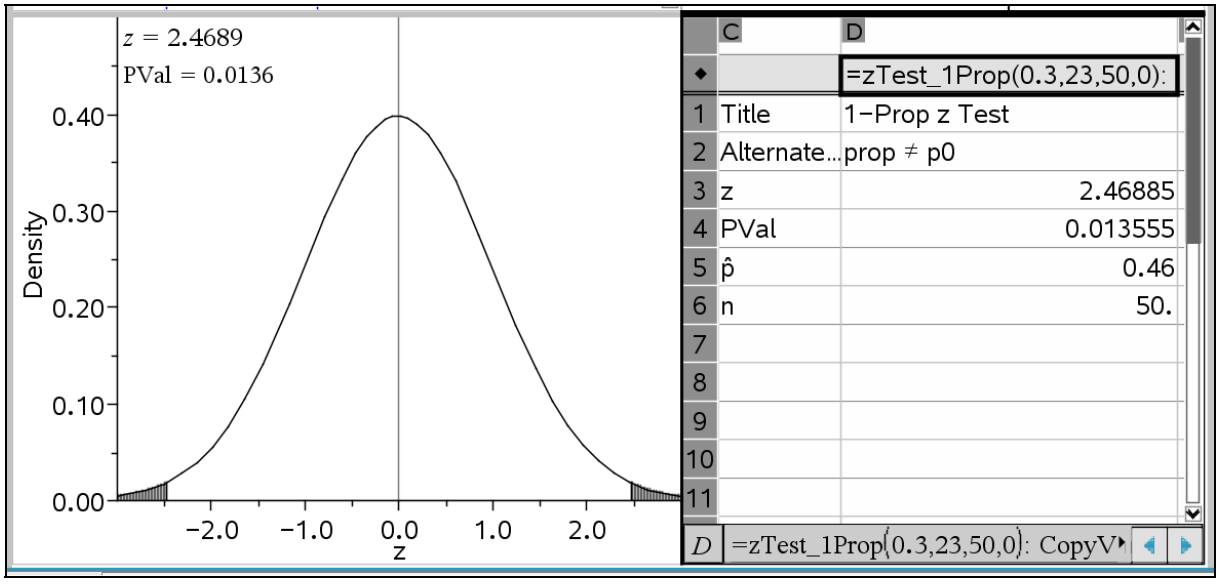
$$2 \cdot \text{normCdf}\left(0.46, 1, 0.3, \sqrt{\frac{0.3 \cdot 0.7}{50}}\right) \qquad 0.013555$$

The prop z Test with 1 proportion works with the standard normal distributed test variable $Z = \dfrac{\hat{P} - 0.3}{\sqrt{\dfrac{0.3 \cdot 0.7}{50}}}$. The observed value is $z = \dfrac{0.46 - 0.3}{\sqrt{\dfrac{0.3 \cdot 0.7}{50}}} \approx 2.47.$

Then the exceeding probability or the *p*-value is $2 \cdot P(Z \geq 2.47) \approx 1.4\%$.







References:

Websites:

[1] NIS (National Institute for Statistics in Belgium): http://statbel.fgov.be/nl

[2] Kerncijfers 2009, Belgium in a European perspective http://economie.fgov.be/nl/modules/pressrelease/statistieken/generale/world_statistics_day.jsp

[3] Information about TI-Nspire: www.education.ti.com

[4] Life guards: www.redderaanzee.wobra.be

[5] Roulette, Rules and Chances: www.casino-gids.be/artikels/spelregels/beginners/roulette.php

[6] Guido Herweyers, *Cahier 8: Betrouwbaarheidsintervallen en testen van hypothesen*, avaiable at www.t3vlaanderen.be

Books

[7] J. Beirlant, G. Dierckx, M. Hubert, *Statistiek en Wetenschap*, Acco, Leuven, 2005.

[8] M. H.Degroot,M. J. Schervish, *Probability and Statistics*, Pearson International Edition, 2010.

[9] D. S. Moore, G. P. McCabe, *Statistiek in de Praktijk*, Academic Service, Schoonhoven, 2006.

[10] D. S. Yates, D. S. Moore, G. P. McCabe, *The practice of Statistics, TI-83 Graphing Calculator Enhanced*, W. H. Freeman and Company, New York, 1999.

[11] R. E. Walpole, R. H. Myers, S. L. Myers, K. Ye, *Probability and Statistics for Engineers and Scientists*, Pearson International Edition, 2011.

Comment of the Editor:

Translating and working through Guido's *Cahier* was really a pleasure. I learned a lot about statistics and a lot about TI-Nspire's latest version.

Two open questions remained:

(1) How to find the probability density functions for functions of random variables (Paragraph 3)? In my books I found the convolution integral for the sum $X + Y$ but that was all. I asked Guido and he promptly sent the required information, thanks for this. You can inform in an extra contribution in this DNL.

(2) I wondered if it were possible to perform Guido's experiments and simulations with DERIVE, too with not too much efforts. As you will see on the following pages, it worked. I used the opportunity to write a few short functions/programs to implement some Nspire-functionalties for DERIVE like `randSamp` in *DNL#87* (in most cases earlier I did the other way round: DERIVE tools for TI-Nspire). The functions for plotting the diagrams have been developed in an earlier Statistics-Tool-contribution.

Following Guido's *Cahier* I start comparing the distribution of z-scores and t-scores.

I implement `randnorm(mean,standard deviation,sample size)` for DERIVE. The counting loop is necessary for avoiding repeated samples.

```
      randnorm(x_, s_, ss, dummy, i) :=
        Prog
          dummy := RANDOM(0)
          i := 0
#1:       Loop
            i :+ 1
            If i = 1000 exit
          VECTOR(RANDOM_NORMAL(s_, x_), k, ss)

#2:   samples(x_, s_, ss, n) := VECTOR(randnorm(x_, s_, ss), k, n)

#3:   samples(100, 10, 4, 10)

      [ 111.1  113.9  83.86  130.5 ]
      [                            ]
      [ 83.49  98.68  92.46  82.44 ]
      [                            ]
      [ 96.36  97.88  108.4  91.59 ]
      [                            ]
      [ 99.87  102.8  94.33  103.5 ]
      [                            ]
      [ 87.54  94.62   102   94.75 ]
#4:
      [ 105.3  85.1   104.6   109  ]
      [                            ]
      [ 121.2  101.6  78.25  107.1 ]
      [                            ]
      [ 107.8  93.31  114.4  113.1 ]
      [                            ]
      [ 99.32  95.44  107.6  95.06 ]
      [                            ]
      [ 109.4  116.6  91.84  103.4 ]
```

`z_t_scores(mean,stdev,sample size,number of simulations)` produces two lists of corresponding scores (according to the columns presented on page 3).

```
      z_t_scores(x_, s_, ss, n, j, sd, smp, sm) :=
        Prog
          j := 1
          zscores := []
          tscores := []
          Loop
            If j > n exit
#5:         smp := APPROX(randnorm(x_, s_, ss))
            sm := AVERAGE(smp)
            sd := STDEV(smp)
            tscores := APPEND(tscores, [(sm - x_)/(sd/√ss)])
            zscores := APPEND(zscores, [(sm - x_)/(s_/√ss)])
            j :+ 1
          RETURN "zscores and tscores stored"
```

The scores are stored in two lists generated of 1500 simulations. Comparing max and min values shows the different dispersion. We cannot perform it dynamically like with Nspire.

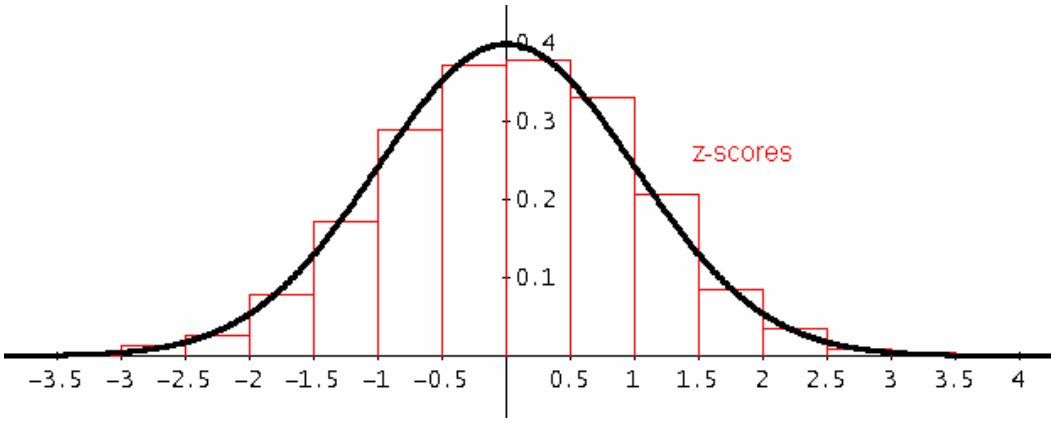
```
#6:   z_t_scores(100, 10, 4, 1500)

#7:   zscores and tscores stored

#8:   [MIN(zscores), MAX(zscores)] = [-2.95, 3.814]

#9:   [MIN(tscores), MAX(tscores)] = [-16.09, 15.25]
```

Like in most cases the graphs are much more impressive. (`normpdf(x,μ,σ)` is not implemented in DERIVE. `normpdf(x)` is the N(0,1) distribution).

```
#12:  HISTDENS(zscores, -4, 4, 16)
```

$$\#13:\quad \text{normpdf}(x,\ \mu := 0,\ \sigma := 1) := \frac{1}{\sigma\cdot\sqrt{(2\cdot\pi)}}\cdot \text{EXP}\left(-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

```
#14:  TABLE(normpdf(x), x, -5, 5, 0.01)
```

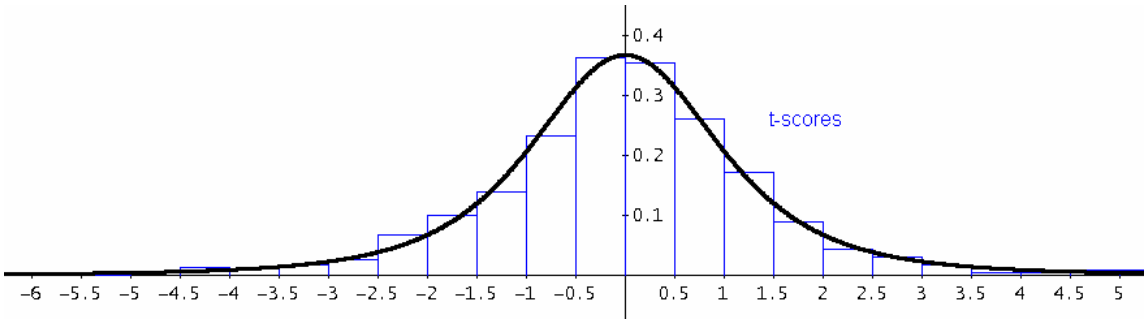


The `TABLE`-function is used to obtain a thick Gaussian bell shaped curve. For finding the pdf of the Student t-distribution I define the respective function `tpdf(x,degrees of freedom)`. In DERIVE the cdf (the distribution function is implemented as `STUDENT(x,n)`.)

```
#15:  HISTDENS(tscores, -17, 17, 68)
```

$$\#16:\quad \text{tpdf}(x,\ n) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{(n\cdot\pi)}\cdot\Gamma\left(\frac{n}{2}\right)}\cdot\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}$$

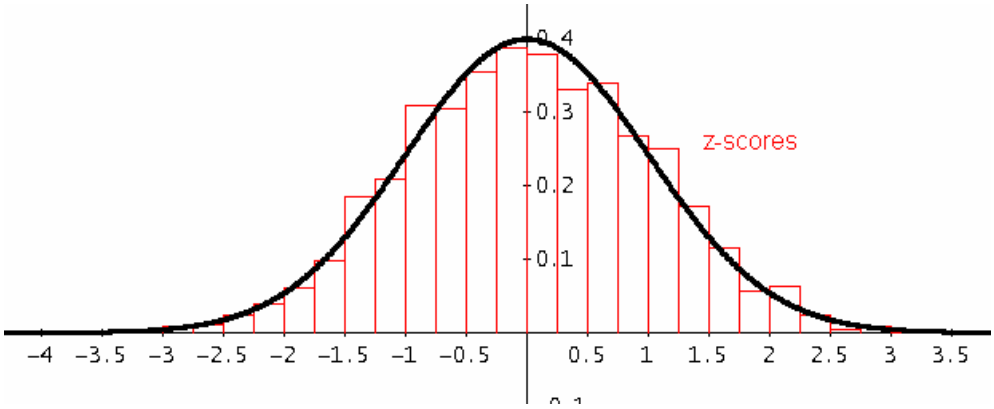
```
#17:  TABLE(tpdf(x, 3), x, -20, 20, 0.01)
```



Let's have a second example: 1500 simulations, sample size = 10, mean = 50, stdev =2

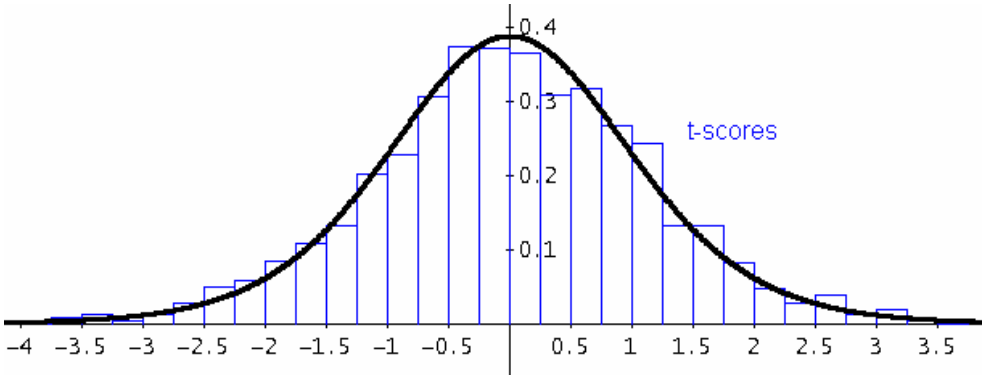
```
#18:  z_t_scores(50, 2, 10, 1500)

#19:  zscores and tscores stored

#20:  [MIN(zscores), MAX(zscores)] = [-3.119, 2.901]

#21:  [MIN(tscores), MAX(tscores)] = [-3.747, 4.261]
```

```
#22:  HISTDENS(zscores, -4, 4, 32)
#23:  TABLE(normpdf(x), x, -5, 5, 0.01)
```



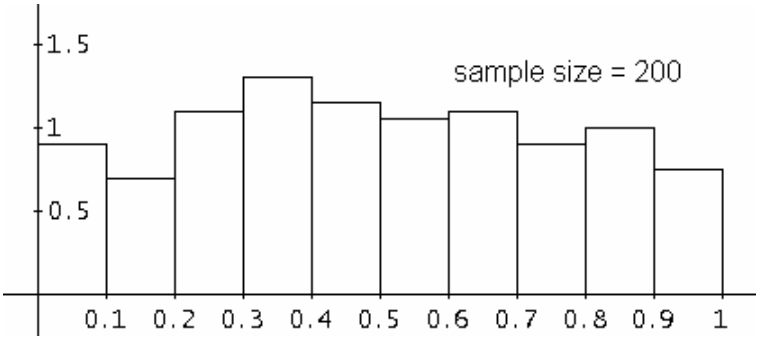
The intervals have a width of 0.25. Both pdfs fit pretty well.

```
#24:  HISTDENS(tscores, -4, 5, 36)
#25:  TABLE(tpdf(x, 9), x, -5, 5, 0.01)
```



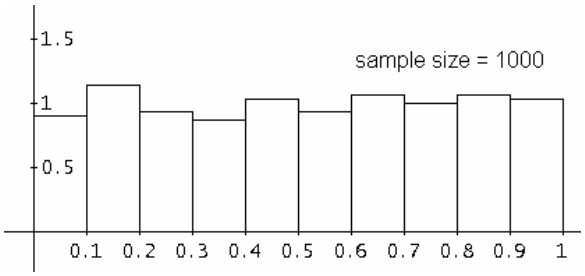
I proceed with the pdfs of one random variable. `rand(n)` generates a sequence of n uniformly distributed random numbers from [0,1] (with n > 1). For n = 1 take `RANDOM(1)`.

```
#26:  rand(n) := VECTOR(RANDOM(1), k, n)
#27:  rand(5) = [0.6787, 0.8265, 0.4101, 0.5258, 0.4188]
#28:  HISTDENS(rand(200), 0, 1, 10)
```

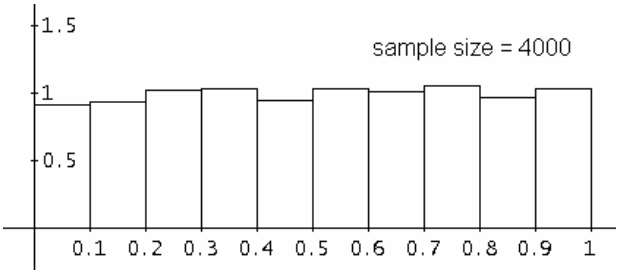


I start with sample size 200 and increase up to a sample size of 10 000 which wouldn't be so easy done with Nspire!

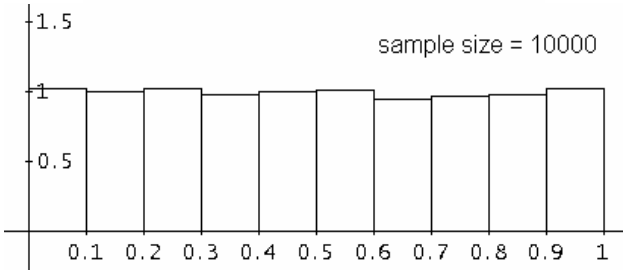
HISTDENS(rand(1000), 0, 1, 10)



HISTDENS(rand(4000), 0, 1, 10)

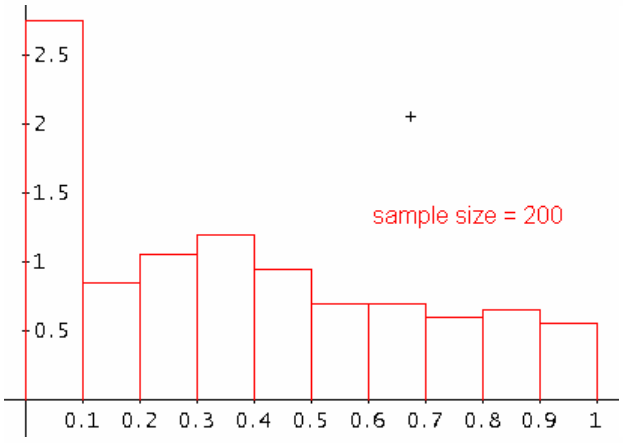


HISTDENS(rand(10000), 0, 1, 10)

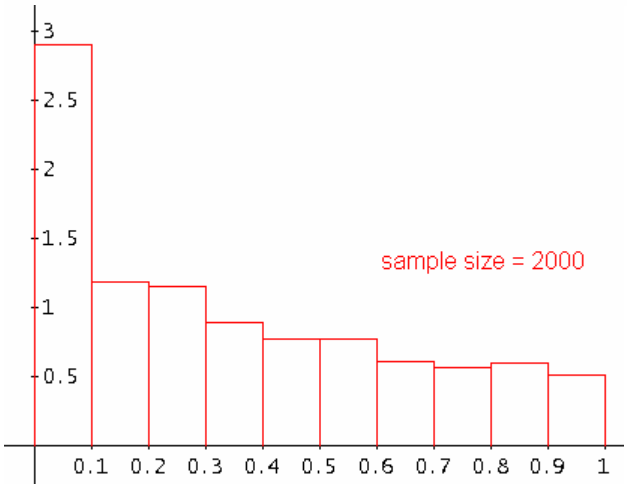


Now we generate 200 random variables x and study the distribution of the squares.

HISTDENS(VECTOR(rand(1)^2, k, 200), 0, 1, 10)



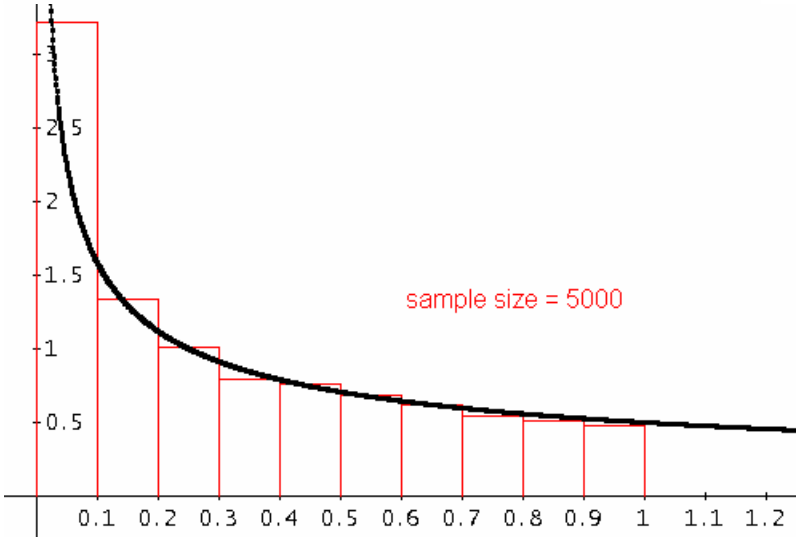
HISTDENS(VECTOR(rand(1)^2, k, 2000), 0, 1, 10)



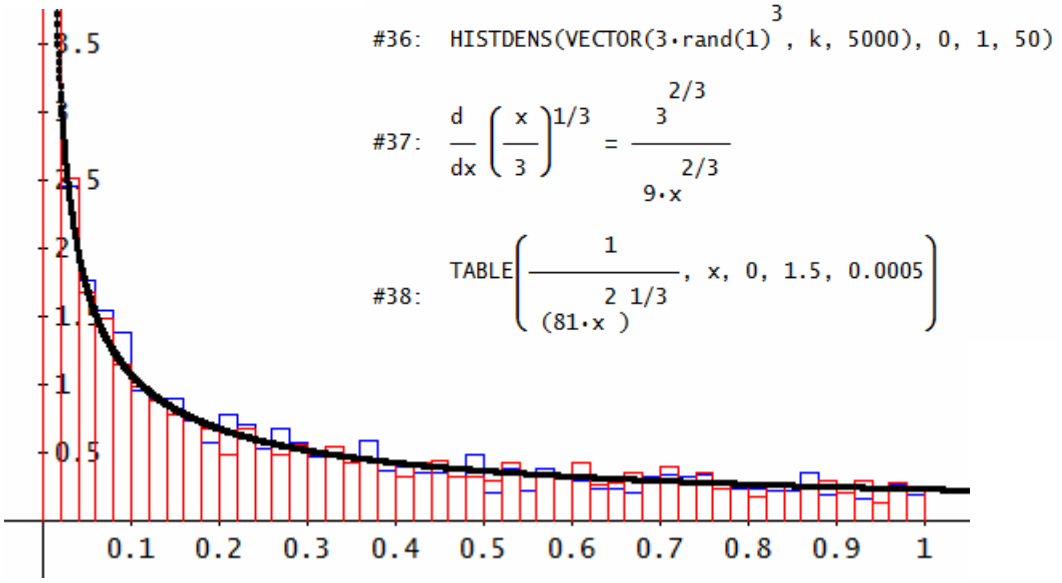
Can you guess the form of the probability density function?

#34: HISTDENS(VECTOR(rand(1)^2, k, 5000), 0, 1, 10)

#35: TABLE$\left(\frac{1}{2\cdot\sqrt{x}}, x, 0, 2, 0.001\right)$



What about $U = 3X^3$?



I superimposed two 5000 simulations runs and then added the respective pdf. I am quite sure that you can easily follow how to derive the density function (compare with page 6).

The density function of U is the first derivative of the distribution function F_U of U:
For $0 \le x \le 1$:

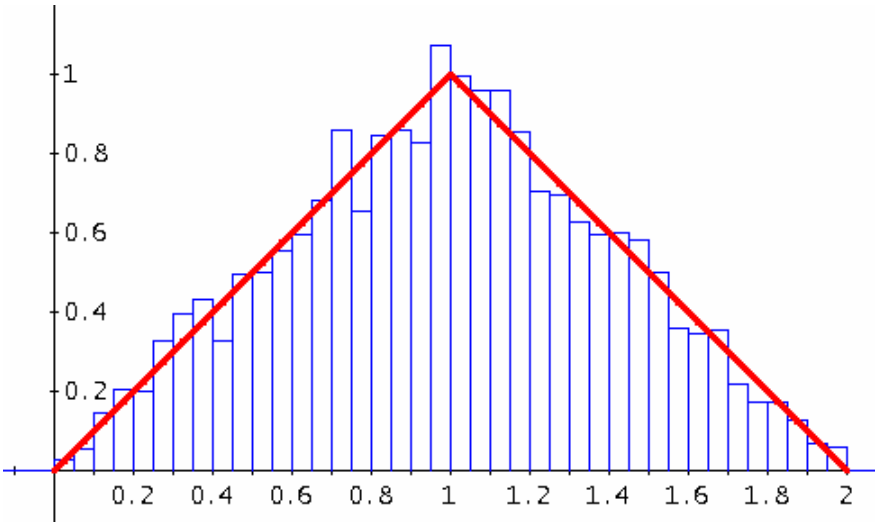
$$F_U(x) = P(U \le x) = P(3X^3 \le x) = P(X \le \sqrt[3]{\frac{x}{3}}) = \int_0^{\sqrt[3]{\frac{x}{3}}} 1 \cdot dt = \sqrt[3]{\frac{x}{3}}$$

Hence, the density function of U is given by: $f_U(x) = F_U'(x) = \begin{cases} \frac{1}{\sqrt[3]{81x^2}} & \text{for } 0 < x \le 1 \\ 0 & \text{else} \end{cases}$. Does it fit?

Functions of two Random Variables

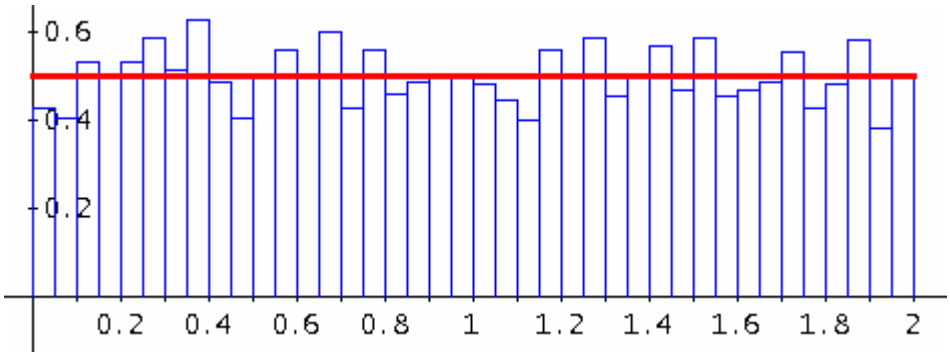
Distribution of $X + Y$ with 3000 simulations performed:

```
#39:  HISTDENS(VECTOR(rand(1) + rand(1), k, 3000), 0, 2, 40)

      f(x) :=
        If x < 0
          0
          If x ≤ 1
#40:        x
            If x ≤ 2
              -x + 2
              0

#41:  TABLE(f(x), x, 0, 2, 0.001)
```



Compare with the pdf of $U = 2X$.

```
#42:  HISTDENS(VECTOR(2·rand(1), k, 3000), 0, 2, 40)

#43:  TABLE(1/2, x, 0, 2, 0.001)
```



It should be no problem to find the pdf $f(u) = 0.5$.

Part 4

Example 1:

I need my DERIVE function `randsamp(population list, sample size)`, s = 0 by default.

```
      randsamp(l, n, s := 0, dummy, i, samp) :=
        Prog
          dummy := RANDOM(0)
          i := 1
          samp := []
          Loop
#44:        If i > n
              RETURN samp
            k := RANDOM(DIM(l)) + 1
            samp := APPEND(samp, [l↓k])
            If s = 1
              l := DELETE(l, k)
            i :+ 1

#45:  ans := [1, 0, 0, 0]
```

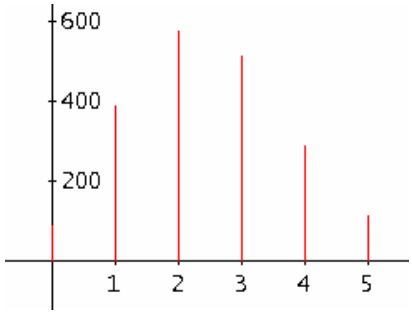
```
      results(list, k, n, t, i_, j_) :=
        Prog
          i_ := 0
          res := []
          Loop
            If i_ > n
              RETURN "results stored in res"
#46:        t := Σ(randsamp(list, k))
            res := APPEND(res, [t])
            j_ := 0
            Loop
              j_ :+ 1
              If j_ = 500 exit
            i_ :+ 1
```

```
#47:  results(ans, 10, 2000) = results stored in res
```

This gives the number of correct answers simulating 2000 tests consisting of 10 multiple choice questions each.

`FRETAB(list)` and `FREQDIAG(list)` as well are functions provided in my statistics tools utility file. If you load the `stat_4.dfw` file all functions and programs needed are available and you can see them via Author > Function Definition.

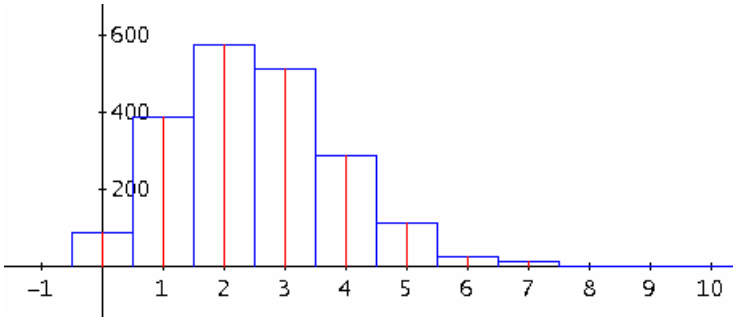
```
#48:  FRETAB(res)

#49:  [ 0   1    2    3    4    5    6    7 ]
      [                                      ]
      [ 88  389  574  511  288  112  27   12 ]

#50:  FREQDIAG(res)
```



We have 88 tests with no single correct answer and there at the other side 12 tests with 7 correct answers. There are no tests with 8 or more correct answers – with randomly chosen answers!!

`HISTO(list, start, end, number of equal wide classes)` plots the respective histogram,

```
#51:  HISTO(res, -0.5, 10.5, 11)
```



Example finished:

```
#52:  DIM(SELECT(k ≥ 6, k, res)) = 39
```

$$\#53:\quad \frac{39}{2000} = 0.0195$$

$$\#54:\quad 1 - \text{BINOMIAL\_DISTRIBUTION}\left(5,\ 10,\ \frac{1}{4}\right) = 0.01972$$

```
#55:  results(ans, 10, 5000) = results stored in res
```

$$\#56:\quad \frac{\text{DIM(SELECT(k} \geq 6, \text{k, res))}}{5000} = 0.0188$$

Example 2:

`randint(a,b,n)` generates *n* integer random numbers *x* with $a \leq x \leq b$. `simul(number)` generates *number* die rolls.

```
#57:  randint(a_, b_, n_) := VECTOR(RANDOM(b_ - a_ + 1) + 1, k, n_)

#58:  randint(1, 6, 20)

#59:  simul(n_) := randint(1, 6, n_)
```

Make a first test with 120 experiments:

```
#60:  test := simul(120)

#61:  test := [2, 3, 3, 2, 2, 3, 3, 2, 2, 3, 2, 5, 4, 2,
```

$$\#62:\quad \text{FRETAB(test)} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 9 & 16 & 27 & 21 & 25 & 22 \end{bmatrix}$$

```
      obs_ := (FRETAB(test))
#63:                        2

#64:  exp_ := [20, 20, 20, 20, 20, 20]
```

I start performing the χ^2-test (same variable names are used as in Guido's paper):

$$\#65:\quad \text{chisqu(obs, exp)} := \sum_{i=1}^{\text{DIM(obs)}} \frac{(\text{obs}_i - \text{exp}_i)^2}{\text{exp}_i}$$

```
      chisqus(n, i, sim, o) :=
        Prog
          i := 0
          chiqs := []
          Loop
#66:        If i = n
              RETURN "chisquares stored in chiqs"
            sim := simul(120)
            o := VECTOR(FREQ(a, sim), a, 6)
            chiqs := APPEND(chiqs, [chisqu(o, exp_)])
            i :+ 1

#67:  chisqus(2000)
```

```
#68:  chisquares stored in chiqs

#69:  [MIN(chiqs), MAX(chiqs)] = [0.1, 21.3]

#70:  HISTDENS(chiqs, 0, 22, 22)
```

Before plotting the density diagram I inform myself in the DERIVE Online Help if the χ^2-pdf is available or not? It is not. What I find is `CHI_SQUARE(x,n)` which is the χ^2-cdf – the distribution function with variable *x* and *n* degrees of freedom.

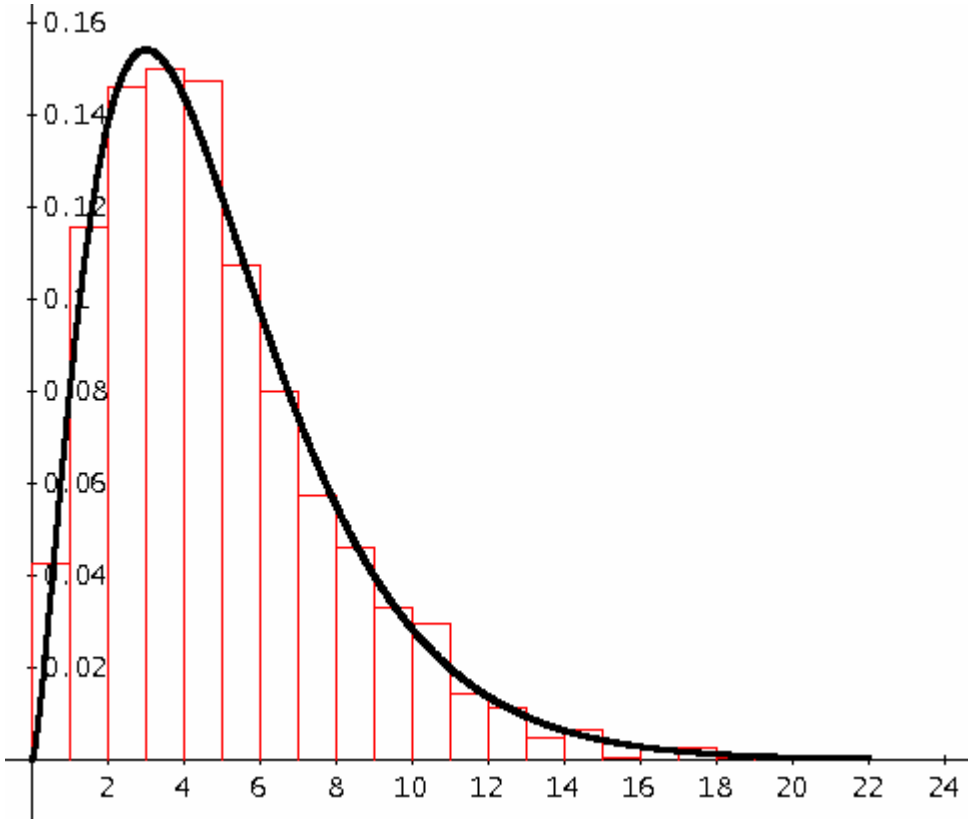
I define my own χ^2-pdf as follows and proceed according Guido's guide line.

#71: $$\chi2\_pdf(x,\ df) := \frac{1}{2^{df/2}\cdot\Gamma\left(\frac{df}{2}\right)}\cdot x^{df/2\ -\ 1}\cdot e^{-\ x/2}$$

```
#72:  χ2_pdf(x, 5)
```

#73: $$\frac{\sqrt{2}\cdot e^{-\ x/2}\cdot x^{3/2}}{6\cdot\sqrt{\pi}}$$

```
#74:  TABLE(χ2_pdf(x, 5), x, 0, 22, 0.01)
```



The result is a really nice and convincing plot, isn't it?

We could do without defining the pdf remembering that the pdf is the derivative of the cdf (probability distribution function):

Please compare #73 from above and #77 on the next page!

$$\#75:\quad \frac{d}{dx}\,CHI\_SQUARE(x,\ 5)$$

$$\#76:\quad \frac{\sqrt{2}\cdot e^{-x/2}\cdot(x^2+3\cdot x+3)}{6\cdot\sqrt{\pi}\cdot\sqrt{x}}-\frac{\sqrt{2}\cdot e^{-x/2}\cdot(x+1)}{2\cdot\sqrt{\pi}\cdot\sqrt{x}}$$

$$\#77:\quad \frac{\sqrt{2}\cdot e^{-x/2}\cdot x^{3/2}}{6\cdot\sqrt{\pi}}$$

Calculating the critical value is an easy job now:

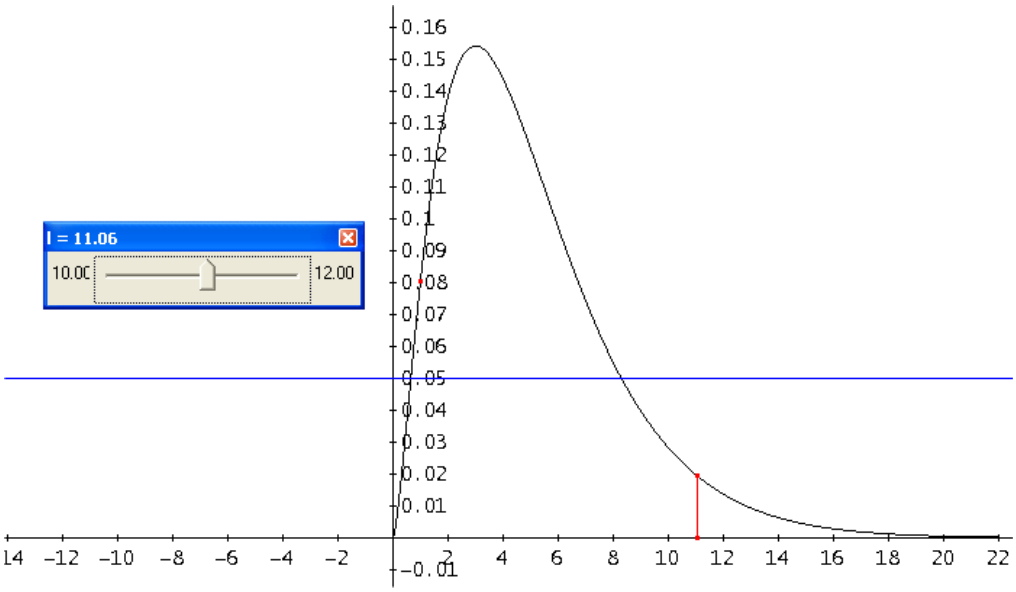
```
#78:  1 - CHI_SQUARE(6.8, 5) = 0.2358

#79:  NSOLVE(CHI_SQUARE(x, 5) = 0.95, x) = (x = 11.07)

#80:  NSOLUTIONS(1 - CHI_SQUARE(x, 5) = 0.05, x) = [11.07]
```

Working with a slider is very attracting with TI-Nspire. With little fantasy we can do a similar animation with DERIVE, too:

We plot the χ^2-pdf and the (red) segment – expression #81 – with its point (l,0) on the x-axis moveable. #82 is the area under the pdf for $x \geq l$.

$$\#81:\quad \begin{bmatrix} l & 0 \\ l & 1.5\cdot l^{0.5}\cdot e^{-0.1329\cdot l} \end{bmatrix}$$

```
#82:  1 - CHI_SQUARE(l, 5)
```



The value of the area is given by the distance between the blue line and the x-axis. Moving the slider moves the segment which moves the blue line. Try to move the blue line to a distance of α = 5% and then read off the value of l. This is it.

Note: Finding the pdf of the t-distribution as derivative of `STUDENT(x,n)` is possible but not so easy.

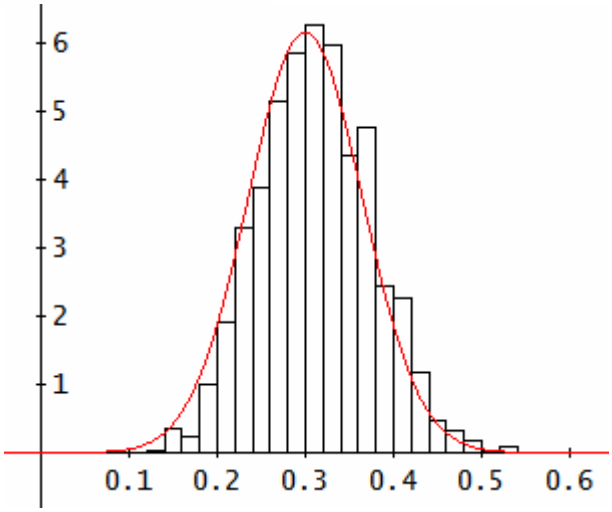
Example 3:

I believe that the following expressions are self explanatory.

```
#83:  population := [yellow, yellow, yellow, other, other, other, other, other, other, other]

#84:  sample := randsamp(population, 50)

                    DIM(SELECT(c = yellow, c, sample))
#85:  p_rand := ----------------------------------
                              DIM(sample)

#86:  samp_rnd(n) := VECTOR(p_rand, k, n)

       2·DIM(SELECT(v ≥ 0.46, v, samp_rnd(2000)))
#87:   ------------------------------------------ = 0.023
                          2000
```

We add the plots:

```
#88:  HISTDENS(samp_rnd(2000), 0, 1, 50)

#89:  normpdf(x, 0.3, √(0.3·0.7/50))
```



And this is the remaining calculation done with DERIVE:

$$2\cdot\left(\text{NORMAL}\left(1,\ 0.3,\ \sqrt{\frac{0.3\cdot0.7}{50}}\right)-\text{NORMAL}\left(0.46,\ 0.3,\ \sqrt{\frac{0.3\cdot0.7}{50}}\right)\right)=\text{ERF}\left(\frac{5\cdot\sqrt{21}}{3}\right)-\text{ERF}\left(\frac{8\cdot\sqrt{21}}{21}\right)$$

$$2\cdot\left(\text{NORMAL}\left(1,\ 0.3,\ \sqrt{\frac{0.3\cdot0.7}{50}}\right)-\text{NORMAL}\left(0.46,\ 0.3,\ \sqrt{\frac{0.3\cdot0.7}{50}}\right)\right)=0.01355$$

$$\frac{0.46-0.3}{\sqrt{\frac{0.3\cdot0.7}{50}}}=2.468$$

$$2\cdot(1-\text{NORMAL}(2.468))=0.01358$$

Probability Density Functions (pdf) for Combined Random Variables

Guido Herweyers, Belgium

Let X and Y be independent random variables with a common uniform distribution on the interval [0,1], having p.d.f. $f(x)=1$ for $0\le x\le 1$ and 0 otherwise.

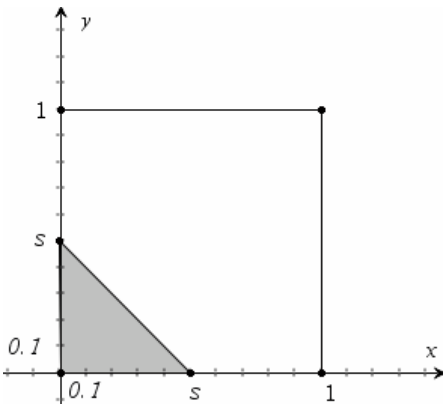
Then the joint p.d.f. of X and Y is $g(x,y)=f(x)f(y)=1$ for $0\le x\le 1$, $0\le y\le 1$ and 0 otherwise.

1) Distribution of $S=X+Y$

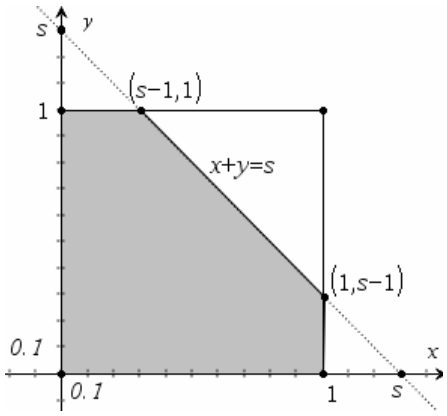
Let $S=X+Y$, then the c.d.f. of S is $H(s)=\mathrm{P}(S\le s)=\mathrm{P}(X+Y\le s)=\iint_A g(x,y)\,dxdy$,

where $A=\{(x,y)\,|\,x+y\le s\}$.

Let R be the square $R=[0,1]\times[0,1]$, then $H(s)=\begin{cases} 0 & , \quad s\le 0 \\ \iint_{A\cap R} dxdy=\text{area}(A\cap R) & , \quad 0<s\le 2 \\ 1 & , \quad s>2 \end{cases}$.



area $H(s)=\dfrac{s^2}{2}, 0<s\le 1$



area $H(s)=1-\dfrac{(2-s)^2}{2}, 1\le s\le 2$

The p.d.f. of S is $h(s)=\dfrac{dH(s)}{ds}=\begin{cases} 0, & s<0 \\ s, & 0\le s\le 1 \\ 2-s, & 1<s\le 2 \\ 0, & s>2 \end{cases}$

Remark:

The distribution of $S=X+Y$, for independent random variables X and Y, is the convolution of the distributions of X and Y:

$$(f*g)(s)=h(s)=\int_{-\infty}^{\infty} f(s-x)\cdot g(x)\,dx$$

f and g are the densities of X and Y respectively.

This is done as follows:

$X \sim U(0,1)$, $Y \sim U(0,1)$ and $S = X + Y$ with $0 \le w \le 2$

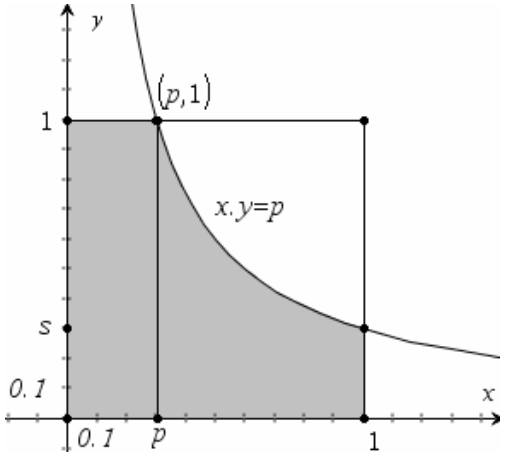
$$h(s) = \begin{cases} 0 \quad , s < 0 \\ \int_0^w f(x)g(s-x)dx = \int_0^w dx = w, \quad 0 \le w \le 1 \\ \int_{w-1}^1 f(x)g(s-x)dx = \int_{w-1}^1 dx = 2-w, \quad 1 < w \le 2 \\ 0 \quad , s > 2 \end{cases}$$

2) Distribution of $S = X \cdot Y$

Let $S = X \cdot Y$, then the c.d.f. of S is $H(s) = \mathrm{P}(S \le s) = \mathrm{P}(X \cdot Y \le s) = \iint_A g(x,y)\,dxdy$,

where $A = \{(x,y) | x \cdot y \le s\}$.

Let R be the square $R = [0,1] \times [0,1]$, then $H(s) = \begin{cases} 0, & s \le 0 \\ \iint_{A \cap R} dx\,dy = \text{area}(A \cap R), & 0 < s \le 1 \\ 1, & s > 1 \end{cases}$.



area $H(s) = s + \int_s^1 \frac{s}{x} dx = s - s\ln(s), 0 < s \le 1$

The p.d.f. of S is $h(s) = \frac{dH(s)}{ds} = \begin{cases} 0, & s < 0 \\ -\ln(s), & 0 \le s \le 1 \\ 0, & s > 1 \end{cases}$

The value $h(0)$can be assigned arbitrarily, e.g $h(0) = 0$.

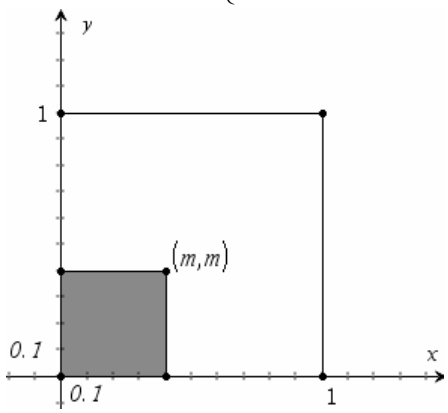
Remark: the p.d.f. of S is not bounded on the interval [0,1] !

3) Distribution of $M = \max(X,Y)$

Let $M = \max(X,Y)$, then the c.d.f. of M is $H(m) = \mathrm{P}(M \le m) = \mathrm{P}(\max(X,Y) \le m) = \iint_A g(x,y)\,dx\,dy$,

where $A = \{(x,y) \mid \max(x,y) \le m\} = \{(x,y) \mid x \le m \text{ and } y \le m\}$.

Let R be the square $R = [0,1]\times[0,1]$, then $H(m) = \begin{cases} 0, & m \le 0 \\ \iint_{A\cap R} dx\,dy = \text{area}\left(A \cap R\right), & 0 < m \le 1 \\ 1, & m > 1 \end{cases}$.



area $H(m) = m^2, 0 < m \le 1$

The p.d.f. of M is $h(m) = \dfrac{dH(m)}{dm} = \begin{cases} 0, & m < 0 \\ 2m, & 0 \le m < 1 \\ 0, & p > 1 \end{cases}$

The value $h(1)$ can be assigned arbitrarily, e.g. $h(1) = 2$.

4) Distribution of $K = \min(X,Y)$

Let $K = \min(X,Y)$, then the c.d.f. of K is $H(k) = \mathrm{P}(K \le k) = \mathrm{P}(\min(X,Y) \le k) = \iint_A g(x,y)\,dx\,dy$,

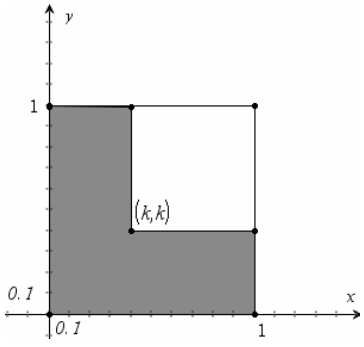
where $A = \{(x,y) \mid \min(x,y) \le k\} = \{(x,y) \mid x \le k \text{ and } y \le k\}$.

Let R be the square $R = [0,1]\times[0,1]$, then $H(k) = \begin{cases} 0, & k \le 0 \\ \iint_{A\cap R} dx\,dy = \text{area}\left(A \cap R\right), & 0 < k \le 1 \\ 1, & k > 1 \end{cases}$.

The p.d.f. of K is

$$h(k) = \frac{dH(k)}{dk} = \begin{cases} 0, & k < 0 \\ 2-2k, & 0 < k \le 1 \\ 0, & k > 1 \end{cases}$$

The value $h(0)$ can be assigned arbitrarily, e.g. $h(0) = 2$.



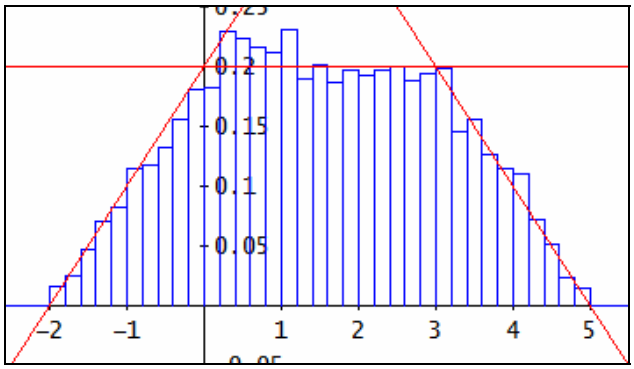
area $H(k) = 1-(1-k)^2 = 2k-k^2, 0 < k \le 1$

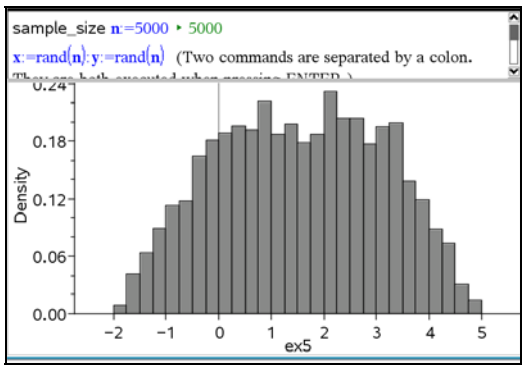
I must admit that this was completely new for me and I – nasty me – sent more problems to Guido and asked for some reference literature because the only facts which I could find were the notes about applying convolution for $X + Y$. They are given on page 29.
I performed some simulations using TI-NspireCAS and DERIVE as well and wanted to find the formulae for the respective density functions. My problems were:

5) Distribution of $S = 5X - 2Y$

```
HISTDENS(VECTOR(5·RANDOM(1) - 2·RANDOM(1), k, 5000), -3, 6, 45)
```

$$\left[\frac{x}{10} + 0.2,\ 0.2,\ 0.5 - \frac{x}{10}\right]$$





Inspecting the density diagrams I had the impression of a trapezoidal distribution and I tried to find the boundary lines as you can see left above.

And this is how Guido treated this distribution:

Distribution of $S = 5X - 2Y$

The reader can verify that the p.d.f. of S is $h(s) = \dfrac{dH(s)}{ds} = \begin{cases} 0, & s \le -2 \\ 0.1s + 0.2, & -2 < s \le 0 \\ 0.2, & 0 < s \le 3 \\ -0.1s + 0.5, & 3 < s \le 5 \\ 0, & s > 5 \end{cases}$

Remark: Let $S = X - Y$, X has a uniform distribution on the interval [0,5] and Y a uniform distribution on the interval [0,2], then we expect that S has the same distribution as problem (6).

Indeed, the joint p.d.f. of X and Y is $g(x,y) = \frac{1}{5} \cdot \frac{1}{2} = 0.1$ for $0 \le x \le 5$, $0 \le y \le 2$ and 0 otherwise.

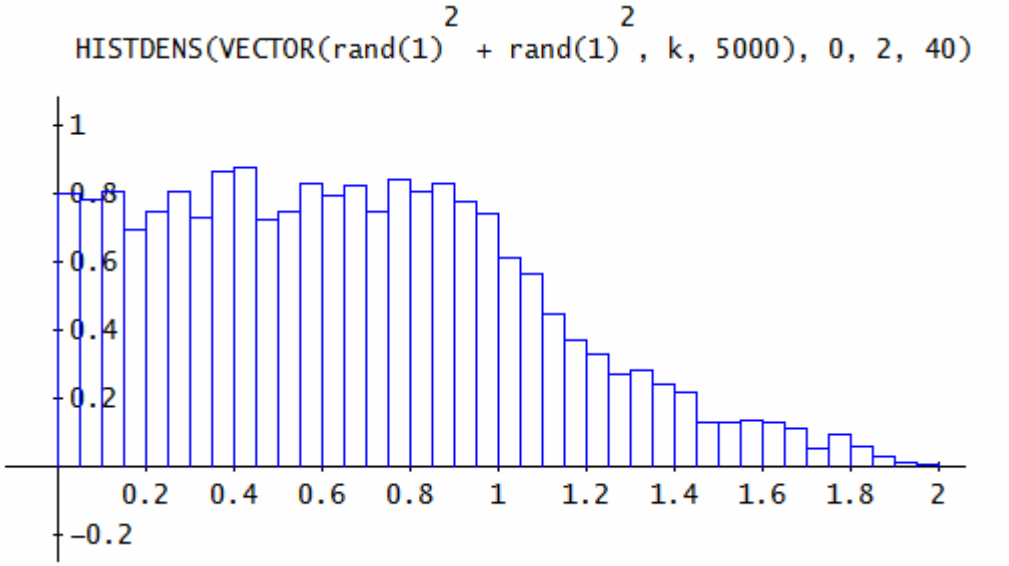
The p.d.f. of S is (see M. H.Degroot, M. J. Schervish, *Probability and Statistics*, Fourth Edition, Pearson International Edition, 2010, page 178):

$$h(s) = \int_{-\infty}^{\infty} g(s+y, y)dy = \begin{cases} 0, & s \le -2 \\ \int_{-s}^{2} 0.1dy = 0.2 + 0.1s, & -2 < s \le 0 \\ \int_{0}^{2} 0.1dy = 0.2, & 0 < s \le 3 \\ \int_{0}^{5-s} 0.1dy = 0.5 - 0.1s, & 3 < s \le 5 \\ 0, & s > 5 \end{cases}$$

This is exactly the trapezium from above! My conjecture is confirmed.

For my next problem I was not able to find the density function by inspection only.

6) Distribution of $S = X^2 + Y^2$



All what I could imagine was that the first part ($0 \le x \le 1$) of the density function seems to be a horizontal line ($y \approx 0.8$?), and what's the second part??

Here is Guido's answer:

Distribution of $S = X^2 + Y^2$

Let $S = X^2 + Y^2$, then the c.d.f. of S is $H(s) = \mathrm{P}(S \le s) = \mathrm{P}\left(X^2 + Y^2 \le s\right) = \iint_A g(x,y)\,dxdy$,

where $A = \left\{(x,y) \middle| x^2 + y^2 \le s\right\}$.

Let R be the square $R = [0,1] \times [0,1]$, then

$$H(s) = \begin{cases} 0 \quad , & s \le 0 \\ \iint_{A \cap R} dxdy = \text{area}\left(A \cap R\right), & 0 < s \le 2 \\ 1 \quad , & s > 2 \end{cases} = \begin{cases} 0, & s \le 0 \\ \dfrac{\pi s}{4}, & 0 < s \le 1 \\ 1 - \displaystyle\int_{\sqrt{s-1}}^{1} \left(1 - \sqrt{s - x^2}\right) dx, & 1 < s \le 2 \\ 1, & s > 2 \end{cases}$$

$$\text{The p.d.f. of } S \text{ is } \quad h(s) = \frac{dH(s)}{ds} = \begin{cases} 0, & s < 0 \\ \dfrac{\pi}{4}, & 0 < s \le 1 \\ \dfrac{\pi}{4} - \arctan\left(\sqrt{s-1}\right), & 1 < s \le 2 \\ 0, & s > 2 \end{cases}$$

The value $h(0)$ can be assigned arbitrarily, e.g. $h(0) = 0$.

DERIVE gives another output for the integral. But plotting both functions shows the identity for both expressions.

$$H(s) := 1 - \int_{\sqrt{(s-1)}}^{1} (1 - \sqrt{(s - x^2)})\, dx$$

$$H(s) := -\frac{s \cdot \mathrm{ATAN}(\sqrt{(s-1)})}{2} + \frac{s \cdot \mathrm{ASIN}\left(\frac{1}{\sqrt{s}}\right)}{2} + \sqrt{(s-1)}$$

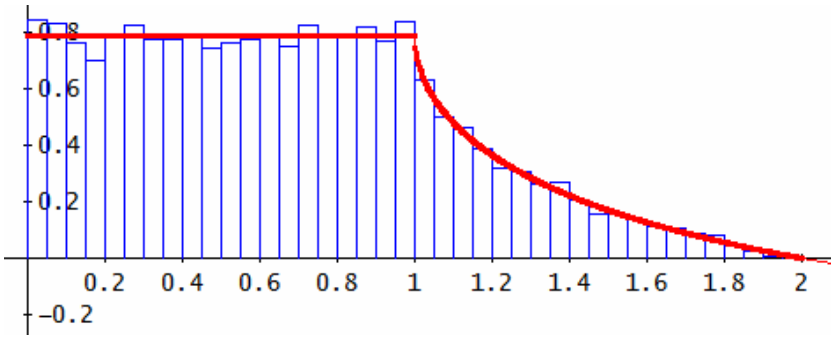
It needs defining the domain for *s* then DERIVE confirms Guido's result.

$$\frac{d}{ds} H(s) = \frac{\mathrm{ASIN}\left(\frac{1}{\sqrt{s}}\right)}{2} - \frac{\mathrm{ATAN}(\sqrt{(s-1)})}{2}$$

$$s :\in \mathrm{Real}\ (1, 2]$$

The plot below is really convincing (density diagram based on 10 000 values $X^2 + Y^2$).

$$\frac{\mathrm{ASIN}\left(\frac{1}{\sqrt{s}}\right)}{2} - \frac{\mathrm{ATAN}(\sqrt{(s-1)})}{2} = \frac{\pi}{4} - \mathrm{ATAN}(\sqrt{(s-1)})$$

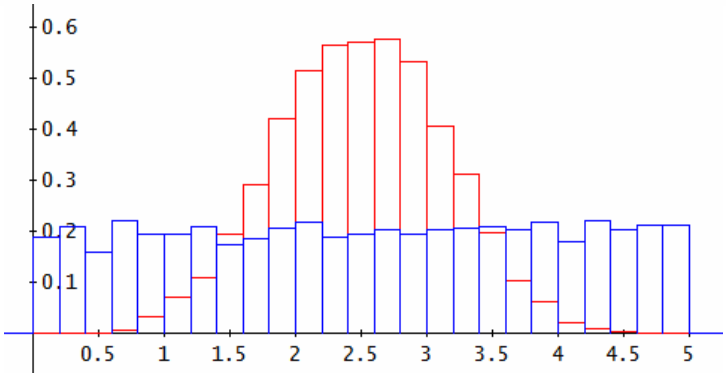


Guido' comment: It is interesting to mention that X^2+Y^2 follows a χ^2-distribution with two degrees of freedom, if both X and Y are independent standard normal distributed. (Show this by simulation!!)

The last example is left for the reader.

Let's compare the density functions of $V = 5X$ and $U = X + X + X + X + X$!

```
HISTDENS(VECTOR(rand(1) + rand(1) + rand(1) + rand(1) + rand(1), k, 5000), -3, 6, 45)
HISTDENS(VECTOR(5·rand(1), k, 5000), -3, 6, 45)
```



Which one is the blue one, and which one is the red one (*U* or *V*)?
What makes the difference?
What are the density functions?

Von: Robert SETIF [robert.setif@gmail.com]
An: Josef Böhm
Betreff: variance of list with weights in Derive

Dear Josef,

It seems to me that there is no command for variance of lists of numbers with their weights.
For instance [61,64,67,70,73] with [5,18,42,27,8] which is 5 times 61, 18 times 64,...
Matlab and XCAS find 8.5275, but Maple_16 finds 12.0207.
???
No command in Mathematica_6, nor in MuPad, nor in Scilab and nor perhaps in XMaxima.
Best regards.

Dear Robert,
Please have a look, this is my DERVE work sheet dealing with your question:

#1: `[numbs := [61, 64, 67, 70, 73], freq := [5, 18, 42, 27, 8]]`

#2: $$vari(n, f) := \frac{\sum_{i=1}^{DIM(n)} n_i \cdot f_i}{\Sigma(f)} - \left(\frac{n \cdot f}{\Sigma(f)}\right)^2$$

#3: `vari(numbs, freq) = 8.5275`

#4: `This meets the XCAS/Matlab-result.`

I repeat with a shorter list given in two ways:

#5: `[n1 := [2, 3, 5], f1 := [3, 1, 4]]`

#6: `list := [2, 2, 2, 3, 5, 5, 5, 5]`

#7: `vari(n1, f1) = 1.984375`

DERIVE's variance and standard deviation are the statistics (sample) values = n-1 weighted.
`vari1` is another form of `vari`

#9: $STDEV(list)^2 = 2.267857142$

#10: $$vari1(n, f, av) := \frac{\sum_{i=1}^{DIM(n)} \left(n_i - \frac{n \cdot f}{\Sigma(f)}\right)^2 \cdot f_i}{\Sigma(f)}$$

#11: `vari1(numbs, freq) = 8.5275`

`vari2` is the DERIVE way.

#13: $$vari2(n, f) := \frac{\sum_{i=1}^{DIM(n)} \left(n_i - \frac{n \cdot f}{\Sigma(f)}\right)^2 \cdot f_i}{\Sigma(f) - 1}$$

#14: `vari2(n1, f1) = 2.267857142`

#15: `VARIANCE(list) = 2.267857142`

#16: `vari2(numbs, freq) = 8.613636363`

But this does also not look like the Maple result.

And this is the third way to achieve the same result:

$$\#19:\quad vari3(n,\ f) := \frac{1}{\Sigma(f)} \cdot \sum_{i=1}^{DIM(n)} n_i \cdot f_i^2 - \left(\frac{n \cdot f}{\Sigma(f)}\right)^2$$

```
#20:  vari3(n1, f1) = 1.984375

#21:  vari3(numbs, freq) = 8.5275
```

So you can be quite sure that 8.5275 is the correct result. I have no idea where the Maple-result comes from.

To be on the safe side, I repeated the calculation with the TI-Nspire statistics tool which confirms the earlier results.

	A vals	B freq	C	D	E	F	G
◆					=OneVar('vals,'freq):		
1	61	5		Title	One-Variable Stati...		
2	64	18		$\bar{x}$	67.45		
3	67	42		Σx	6745.		
4	70	27		Σx^2	455803.		
5	73	8		sx := $s_{n-...}$	2.9349	8.61364	
6				σx := $\sigma_{nX...}$	2.92019	8.5275	
7				n	100.		
8				MinX	61.		
9				Q_1X	67.		
10				MedianX...	67.		
11				Q_3X	70.		
12				MaxX	73.		

F6 $=e6^2$

Best regards
Josef

Giuseppe Ornaghi [g.ornaghi2@tin.it]

Is it possible, and how, to define in Derive LI function (logarithmic integral) for a complex number?
Thank you very much,
Giuseppe

Fred Tydeman's problem – A non recursive definition of a special sequence

In DNL#87 Fred complained that he only received answers from *MATHEMATICA* and Maple Users. Fortunately DUG members also tackled the problem:

This is Benno Grabinger's answer:

The number of fix point free permutations of n elements is given by

$$a(1) = 0$$
$$a(n) = n \cdot a(n-1) + (-1)^n$$

It is easy to derive the explicit formula from the recursive definition:

$$a(1) = 0$$
$$a(2) = 2\,a(1) + 1 \qquad \left| : 2! \right.$$
$$a(3) = 3\,a(2) - 1 \qquad \left| : 3! \right.$$
$$a(4) = 4\,a(3) + 1 \qquad \left| : 4! \right.$$
$$\ldots$$
$$a(n) = n\,a(n-1) + (-1)^n \quad \left| : n! \right.$$

This gives:

$$a(1) = 0$$
$$\frac{a(2)}{2!} = \frac{a(1)}{1!} + \frac{1}{2!}$$
$$\frac{a(3)}{3!} = \frac{a(2)}{2!} - \frac{1}{3!}$$
$$\frac{a(4)}{4!} = \frac{a(3)}{3!} + \frac{1}{4!}$$
$$\ldots$$
$$\frac{a(n)}{n!} = \frac{a(n-1)}{(n-1)!} + \frac{(-1)^n}{n!}$$

adding all equations leads to:

$$\frac{a(n)}{n!} = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} \mp \ldots \frac{(-1)^n}{n!}$$
$$a(n) = n! \cdot \left(\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} \mp \ldots \frac{(-1)^n}{n!} \right)$$

Hence, the desired explicit formula is $a(n) = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}$

Substituting in the power series $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ for x = -1 then

$e^{-1} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \approx \left(1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} \mp \ldots\right)$ and we obtain an estimation $a(n) \approx n! \cdot e^{-1} = \frac{n!}{e}$.

On a problem by Fred Tydeman

Stefan Welke, stefanwelke@web.de

In DNL #87 Fred Tydeman asked for a non recursive definition of the recursively defined sequence $f(0) := 1$ and $f(n) := n \cdot f(n-1) + (-1)^n$ for $n > 0$, which is closely related to the factorial function, see the remark at the end. The computation should be done with *DERIVE*. This definition is easily seen to be equivalent to the following two step recursion, which has some advantages:

$$f(0) := 1, f(1) := 0, \text{ and } f(n) := (n-1)\left(f(n-1) + f(n-2)\right) \text{ for } n > 1$$

Note, that we start here with $n = 0$ in contrast to Fred. Now we turn this definition into an iteration by the observation that the two step recursion above ist equivalent to the following matrix equation:

(0.1) $$\begin{pmatrix} 0 & 1 \\ n-1 & n-1 \end{pmatrix} \begin{pmatrix} f(n-2) \\ f(n-1) \end{pmatrix} = \begin{pmatrix} f(n-1) \\ f(n) \end{pmatrix} \text{ for } n > 1$$

We set $M_k := \begin{pmatrix} 0 & 1 \\ k & k \end{pmatrix}$ and get by iteration:

(0.2) $$\begin{pmatrix} f(n) \\ f(n+1) \end{pmatrix} = \left(\prod_{k=1}^{n} M_k\right) \begin{pmatrix} f(0) \\ f(1) \end{pmatrix} \text{ for } n > 0$$

Here the product is meant as $\prod_{k=1}^{n} M_k = M_n \cdot M_{n-1} \cdot \ldots \cdot M_1$, because the matrices M_k do not commute for different values of k. The equation (1.2) gives a non recursive definition for consecutive pairs of elements of Fred's sequence for arbitrary initial values $f(0)$ and $f(1)$.

In Fred's case we have $f(1) = 1 \cdot f(0) - 1 = f(0) - 1$, so we arrive at

(0.3) $$\begin{pmatrix} f(n) \\ f(n+1) \end{pmatrix} = \left(\prod_{k=1}^{n} M_k\right) \begin{pmatrix} f(0) \\ f(0) - 1 \end{pmatrix} \text{ for } n > 0.$$

A straightforward implementation in *DERIVE* is the function g, which works with matrix multiplication, the initial value for $n = 0$ is p:

```
      g(n, p) :=
        If n = 0
           p
           Prog
             k := 1
#1:          m := [0, 1; 1, 1]
             Loop
               k := k + 1
               If k > n
                  RETURN (m·[p; p - 1])↓1↓1
               m := [0, 1; k, k]·m

#2:    VECTOR(g(q, 1), q, 0, 7)

#3:    [1, 0, 1, 2, 9, 44, 265, 1854]

#4:    VECTOR(g(q, 0), q, 0, 10)

#5:    [0, -1, -1, -4, -15, -76, -455, -3186, -25487, -229384, -2293839]
```

An even better and faster implementation looks at two consecutive vectors and avoids actually the matrix multiplications:

$$\begin{pmatrix} f(n-1) \\ f(n) \end{pmatrix} \xrightarrow{M_n} \begin{pmatrix} f(n) \\ f(n+1) \end{pmatrix} = \begin{pmatrix} f(n) \\ n \cdot \big(f(n-1) + f(n)\big) \end{pmatrix} \text{ for } n > 0$$

This reads as a *DERIVE* function:

```
      f(n, p) :=
        If n = 0
           p
           Prog
             k := 0
#6:              v := [p, p - 1]
             Loop
               k := k + 1
               If k > n - 1
                  RETURN v↓2
               v := [v↓2, k·(v↓1 + v↓2)]

#7:   VECTOR(f(r, 1), r, 0, 15)

#8:   [1, 0, 1, 2, 9, 44, 265, 1854, 14833, 133496, 1334961, 14684570,
       176214841, 2290792932, 32071101049, 481066515734]
```

And in 0.030 seconds on my quad-core computer:

```
#9:   f(1000,3)
```

```
#10:
      952804520525856627451321481964974025676536975193112726642291666949765841919494112
365407404364993677146919049839307459512519926419927453074574888827562417477590532604895 1
281860582566832862823502815886112938051389882429141427164275206137517645533585921717984 3
004791771295911722384985969830043498984343083514845834353249704255909472461206205514006 1
097831598680176879109618344263421749219892298120873202823473157047533562867507379099425 0
887572764911473124345410524490869390286269805766447401777629524335098010182422631739034 6
357200269808845957844818774080122606947707087914061754814072685246092520391551998447377 1
939954583713258096064456354502839996333102312180849140277184137900974521225414384051030 0
669478992352368881480274742408754262738955814638725307561628915444478942010422157126550 8
731062050399047230353710533855390240863321167327839680702138615462267010158994297986404 8
544883606449467780351560110934490459872477148389292131841530623862195540795317190368427 4
856477429287550985508727408729102519666011729370784837259365709304219813891738039411007 9
927192103523138126901856878860378118796617358613463970339851614180556730310881073258440 7
747006380736585460051537169742524286536808426295600840584462696249514617011747360259762 1
086388240532962543018556325757658743166249459848918633671234441451904364629857158439949 9
558541548476805548942647342430301013609786062106870289744583896747209702066339127259765 8
230405837634231343099177291486596559761846493730624102632236504460171280664818018227015 6
783314834503006464489120970138266103888049158588307434191149589760944536180284603461323 4
291029680893652091565517048531740435087346565001509233811258163379078370287356879628808 0
812603499053348247371860775197088001242818191898069418940670287532472843163627226947982 5
452268333445743704891641645303971861501737627168298175871004180154865682727683725178965 6
082233421496761272807120740798670107857775938734019940281572580053417844967282746125727 7
891400775884939368951044606366466181244232690845010090562144183259619738341738341114123 5
083624337200301843917385969251322626024222728825821828119317425685411190398618821938507 4
369555037863254523627491039857104093177364590599853537342924579359374969972326638729649 6
652209651478601790550497378446752904610013640443569004619151256075796299966226876637994 6
759926667722228779295636645389126431722393268539068601119113908474985451813508750353980 6
686862195997387003647310820647089080512559176603565166026316625607185906652349440493287 3
989243033885387310363168734739688132495065765084286985470381074859852651721482664917019 2
279447500448155506860 01
```

`f(10000,3)` needs 0.631 seconds.

We conclude with a different approach, which gives a closed form expression for Fred's sequence, and which unveals a relation to the factorial and to the exponential function. At first we give a generalisation of Fred's definition:

$$For\ a, f(0) \in \mathbb{C}\ \ let\ \ f(n) := n \cdot f(n-1) + a^n \ \text{ for } n > 0\,.$$

Now consider the quotient $\frac{f(n)}{n!}$. We get by the definition of f:

$$\frac{f(n)}{n!} = \frac{n \cdot f(n-1) + a^n}{n \cdot (n-1)!} = \frac{f(n-1)}{(n-1)!} + \frac{a^n}{n!} \tag{0.4}$$

Repeated application of (1.4) gives:

$$\frac{f(n)}{n!} = \frac{f(0)}{0!} + \sum_{k=1}^{n} \frac{a^k}{k!} = f(0) - 1 + \sum_{k=0}^{n} \frac{a^k}{k!} \tag{0.5}$$

The first immediate consequence is a closed form representation for $f(n)$:

$$f(n) := \left(f(0) - 1 + \sum_{k=0}^{n} \frac{a^k}{k!} \right) \cdot n! \tag{0.6}$$

The second consequence is the convergence of the quotient sequence:

$$\lim_{n \to \infty} \frac{f(n)}{n!} = f(0) - 1 + e^a \tag{0.7}$$

An equivalent result is the asymptotic behaviour of f :

(0.8) $$f(n) \sim \left(e^a + f(0) - 1\right) \cdot n! \text{ for large } n$$

Finally we consider the derive function `h(n)` which computes the values of f according to formula (1.6) with $a = -1$ and $f(0) = 1$ as Fred's initially posed sequence.

#11: $h(n) := \sum\left(\frac{(-1)^k}{k!}, k, 0, n\right) \cdot n!$

```
#12:    VECTOR(h(q), q, 0, 7)

#13:    [1, 0, 1, 2, 9, 44, 265, 1854]
```

Although the function definition for `h` looks much simpler than the previous definitions for `g` and `f`, this implementation has the disadvantage that computation time explodes. `h(1000)` needs 2.82 seconds and `h(5000)` needs 518.8 seconds. (1.6) is a very pleasant formula in the eyes of a mathematician, but it is not suitable for programming purposes.

And there was a third reaction from Australia sent by our DUG member David Halprin:

Josef

In December 2011, I sent you this letter with attachment BOEHM.PDF in the hope that you would find space in a DNL. The purpose was to elicit reader feedback as well as a challenge to see assorted ingenuities.

I know you have been very busy.

BTW I spent many hours of many weeks in a fruitless attempt to solve the Fred Tydeman recursive problem with true mathematical reasoning. I read that some people came up with a Mathematica or Maple `solution' but that evades the kernel of the problem.

The first definition explicitly defines the series without recourse to recursive methodology

The second definition, albeit recursive is insufficient, since it is incomplete.

I spent my time trying to find an expression that defines the n^{th} term as efficiently as Tydeman's first definition, but with its basis in a recursively defined series, as one gets with Fibonacci and Lucas types of series.

All these types of series can be treated under the one generalisation, called a General Admixture Series, (G.A.S.) for which I wrote a very detailed paper back in 1989 covering Tribonacci, Quatracci, Pentacci and much higher, for which all could be defined with any of the 5 methods below:-

1) Each term is the sum of the previous 2,3,4,5,6, terms etc..

2) An equation, which determines the $(n+1)^{th}$ term.

3) The sum to the $(n+1)^{th}$ term.

4) The limiting value of the ratio of two successive terms.

5) The generating function.

viz. from 2) above

(a^n - b^n)/(a - b) for Fibonacci-type series, or (a^n + b^n) for Lucas-type series.

However, I had to give up, sadly. I hope that someone among the readership will achieve a real mathematical solution.

Herzlichst

David H

First of all I'd like to apologize for my not publishing David's paper from 2011. It is following. David promised to submit an updated version of his Tribonacci-Quattracci… paper mentioned above. He wrote:

Josef
Coincidentally I spent a solid 8 hours today re-attempting the Tydeman series, still thinking I had cracked it, but alas and alack, not to be as yet. I was constantly referring to my methodology in my paper, attached RECURSIV.PDF. Please put me in the queue for publishing it, but NOT this copy of today, since it needs some additions, for which I have made some notes within. However please read it and enjoy. You will see that my approach is completely different from Stefan Welke and Benno Grabinger, whose interesting papers you sent me; thanks very much.

My paper ends with quite a challenge for the readers. Maybe my hoped-for answer is in the too-hard basket for today's mathematics???? I would certainly welcome your comments.
Herzlichst
David H

DILEMMA AND/OR PARADOXON

David Halprin

I have been attending math. seminars at Melbourne University for years. Most of them are under the auspices of MUMS (Melbourne University Mathematical Society", whose members are an admixture of students, ex-students and interested parties.

URL: http://www.ms.unimelb.edu.au/~mums/seminars/pastseminars.html

These daytime seminars are slotted in between formal lectures and are allowed one hour. Recently, there was some time to spare, so the lecturer presented a question to the attendees for them to come up, one at a time, to chalk their solution(s) on the green board. Belatedly, I realised that I should copy down all these solutions, but I only managed to copy five; the last two I missed out and could not duplicate, so I am requesting the readers to submit their opinions of what the last two solutions were.

(What were the answers of student 6) and 7)? You will find the solution of this problem in the next DNL – provided by David, Josef.)

(If you are not familiar with $D^{1/2}$ which is the 0.5 derivative then read David's contribution "*On the lighter Side of Operational Calculus*" in DNL#35. Josef)

QUESTION

Solve for $y = f(\theta)$

$$y = f(\theta),\quad Dy = \frac{dy}{d\theta} = f'(\theta),\quad D^2 y = \frac{d^2 y}{d\theta^2} = -y$$

1) First-Year Student

$$y = \sin\theta,\quad Dy = \cos\theta,\quad D^2 y = -\sin\theta \qquad \textbf{Q.E.D.}$$

2) First-Year Student

$$y = \cos\theta,\quad Dy = -\sin\theta,\quad D^2 y = -\cos\theta \qquad \textbf{Q.E.D.}$$

3) Second-Year Student
With respect, you two guys are both correct, but one needs to combine both solutions under one function.

$$y = \mathrm{c}i\mathrm{s}\theta = \cos\theta + i\cdot\sin\theta = e^{i\cdot\theta}$$
$$Dy = i\cdot e^{i\cdot\theta},\quad D^2 y = -e^{i\cdot\theta} \qquad \textbf{Q.E.D.}$$

4) Second-Year Student
With respect, y'all missed out on another avenue of calculus venturing

Let $y = \mathrm{sun}\theta = D^{1/2}\sin\theta,\quad D^{1/2}\mathrm{sun}\theta = D^{1/2}\cdot D^{1/2}\mathrm{sin}\theta = D\sin\theta = \cos\theta$

Let $D^{1/2}\cos\theta = \mathrm{cus}\theta,\quad D^{1/2}\mathrm{cus}\theta = D^{1/2}\cdot D^{1/2}\mathrm{cos}\theta = D\cos\theta = -\sin\theta$

$$D\mathrm{sun}\theta = D^{1/2}\cdot D^{1/2}\mathrm{sun}\theta = \; D^{1/2}\mathrm{cos}\theta = \mathrm{cus}\theta$$
$$D^2\mathrm{sun}\theta = D\mathrm{cus}\theta = D^{1/2}\cdot D^{1/2}\mathrm{cus}\theta = -D^{1/2}\mathrm{sin}\theta = -\mathrm{sun}\theta \qquad \textbf{Q.E.D.}$$

5) Second-Year Student
With great respect, what about

Let $y = \mathrm{cus}\theta = D^{1/2}\cos\theta,\quad D^{1/2}\mathrm{cus}\theta = D^{1/2}\cdot D^{1/2}\mathrm{cos}\theta = D\cos\theta = -\sin\theta$

Let $D^{1/2}\sin\theta = \mathrm{sun}\theta,\quad D^{1/2}\mathrm{sub}\theta = D^{1/2}\cdot D^{1/2}\mathrm{sin}\theta = D\sin\theta = \cos\theta$

$$D\mathrm{cus}\theta = D^{1/2}\cdot D^{1/2}\mathrm{cus}\theta = \; -D^{1/2}\mathrm{sin}\theta = -\mathrm{sun}\theta$$
$$D^2\mathrm{cus}\theta = -D\mathrm{sun}\theta = -D^{1/2}\cdot D^{1/2}\mathrm{sun}\theta = -D^{1/2}\mathrm{cos}\theta = -\mathrm{cus}\theta \qquad \textbf{Q.E.D.}$$

6) Third-Year Student
With respect, you two guys are both correct, but one needs to combine both solutions under one function.

7) Third-Year Student
With greatest respect, the combined solution of the third solver and the sixth solver need to be combined as one umbral solution, don't you think?

David Halprin

Titbits (39)- Emulating the ElGamal Cryptosystem (Part 1)

by Johann Wiesenbauer, Vienna

Well, there has been quite a break since my last column here due to the fact that I was very busy lately. Nevertheless, I'm ready now to continue now with one of my favourite topics dealing with the use of elliptic curves in public-key cryptography. To be more precise, I would like to make a setup in DERIVE for the ElGamal cryptosystem, by far the most important competitor of RSA. Most of the routines, which I introduce here, I developed for my talk at DES-TIME-conference in Malaga in 2010. If you missed my contribution in the proceedings of that conference, this article might make up for it to some extent.

Let's start with some general remarks on the ElGamal cryptosystem. As you might already know, it's based on the discrete logarithm problem or DLP for short. In its most general form what you need is a cyclic group G along with a generator g such that for any $h \in G$ the equation $g^x = h$ is usually impossible to solve in a reasonable time. Needless to say that the group G must be rather big to fulfil this condition, otherwise you could find x simply by trial and error. The "classical" choice for G is the multiplicative group of the residue class ring mod p for some prime p with say at least 1024 bits, but nowadays cyclic groups consisting of points on an elliptic curve over some finite field are very popular, too. In fact, we are considering here the latter case, where the finite field will always be a residue class ring mod p again for some prime p with at least 160 bits. It turns out then that group G itself has about the same order of magnitude, hence it can be much smaller than in the first case, which is one of the main reasons of the popularity of ECC (=Elliptic Curve Cryptography) in any environment where only a small amount of resources is available, like on smartcards or handhelds.

This said I'll focus more on the mathematics behind ElGamal now rather than on other cryptographic aspects. For a start, let me remind you of same facts concerning the arithmetic on elliptic curves over a residue class mod p, where p is a - usually "big" - prime number (cf. my Titbits 30 if you need to brush up on this topic). Basically, we are dealing with an algebraic curve of the form

$$y^2 = x^3 + ax + b \quad (a,b \in \mathbf{Z}/p\mathbf{Z})$$

along with the point O at infinity, where the discriminant $4a^3+27b^2$ of the polynomial on the right-hand side does not vanish mod p, i.e., this polynomial hasn't got multiple roots in the residue class ring $\mathbf{Z}/p\mathbf{Z}$. Under these conditions the finite set $E_{a,b}(\mathbf{Z}/p\mathbf{Z})$ of points on this form an abelian group, if the addition of two points U and V on the curve and the n-th additive power of U is defined in a way that can be seen by looking at the following DERIVE-routines:

```
add(u, v, a, p, k_) :=
  Prog
    If u = [p, p] ∨ v = [p, p]
       RETURN u + v - [p, p]
    If u↓1 = v↓1 ∧ MOD(u↓2 + v↓2, p) = 0
       RETURN [p, p]
    If u = v
       k_ := MOD((3·u↓1^2 + a)·INVERSE_MOD(2·u↓2, p), p)
       k_ := MOD((v↓2 - u↓2)·INVERSE_MOD(v↓1 - u↓1, p), p)
    a := MOD(k_^2 - u↓1 - v↓1, p)
    [a, MOD(k_·(u↓1 - a) - u↓2, p)]

mult(u, n, a, p, b_, u_) :=
  Prog
    b_ := [p, p]
    u_ := [u↓1, p - u↓2]
    If n < 0
       RETURN mult(u_, -n, a, p)
    Loop
      If n = 0
         RETURN b_
      If MOD(n, 4) = 1
         [n :- 1, b_ := add(b_, u, a, p)]
      If MOD(n, 4) = 3
         [n :+ 1, b_ := add(b_, u_, a, p)]
      u := add(u, u, a, p)
      u_ := [u↓1, p - u↓2]
      n :/ 2
```

Basically, `mult(u,n,a,p)` makes use of the so-called NAF-representation of n, which uses the "digits" -1,0,1 and is unique provided that never two nonzero digits are adjacent (hence the notation NAF, by the way, meaning "non adjacent form"). If n is even then the least significant digit is always 0, otherwise it is 1 or -1, depending on whether $n \equiv 1 \bmod 4$ or $n \equiv -1 \bmod 4$. By removing this last digit and responding to its value in a similar way as in the original "Square and multiply"- method, except for using inverse points whenever -1 occurs, one can proceed in an iterative way as can be seen above until the stop condition n = 0 becomes true. Actually, it is supposed to be a little bit faster than the normal right-to-left exponentiation based on the binary representation of n, as the number of nonzero digits is on average only about one third of the total number of digits, but it is used here mainly for didactic reasons. What we take advantage here is the stunningly easy way to compute inverses in our group, after all, we only have to change the sign of the y-coordinate for this.

Later on, we will also need the order of an elliptic curve, i.e., the number of its points. There are two essentially different ways of achieving this goal. The first one is to choose the coefficients a and b of the elliptic curve essentially at random and compute its order thereafter using Schoof's algorithm or its more advanced version, the SEA-algorithm due to Schoof, Elkies and Atkins. We use here the second approach though, which makes use of a special class of elliptic curves, the so-called CM-curves. For this we need a set of discriminants for which the corresponding imaginary quadratic fields have a "small" class number. Below is a complete list of all those values belonging to the class numbers 1 or 2:

```
Δ := [-3, -4, -7, -8, -11, -19, -43, -67, -163, -15, -20, -24, -35, -40,
  -51, -52, -88, -91, -115, -123, -148, -187, -232, -235, -267, -403, -427]
```

Next, we must choose for our given prime p the first number D from this set such that the equation

$4p = x^2 + |D|y^2$

has got an solution (x,y) with integers x and y. (In the unlikely case that no such D is available, we must switch to a different prime p.) In order to solve the Diophantine equation above for given values of p and D, we introduce another routine due to Cornaccia-Smith:

```
CS(p, D, a_, b_, c_, r_) :=
  Prog
    If JACOBI(D, p) < 1
       RETURN false
    b_ := SQUARE_ROOT(D, p)
    If ODD?(b_ - D)
       b_ := p - b_
    a_ := 2•p
    c_ := FLOOR(2•√p)
    Loop
      If b_ ≤ c_ exit
      r_ := MOD(a_, b_)
      a_ := b_
      b_ := r_
    a_ := 4•p - b_^2
    If MOD(a_, ABS(D)) > 0
       RETURN false
    c_ := √(a_/ABS(D))
    If ¬ INTEGER?(c_)
       RETURN false
    [±b_, ±c_]

[CS(23, -3), CS(23, -4), CS(23, -7)]

                          [false, false, [±8, ±2]]
```

I won't go into details as to how and why this routine works, but if you have a closer look at it, you will see, that it starts with an even square root of D mod p, (which might fail to exist though and leads to an abortion then) and contains elements of the Euclidean Algorithm thereafter which is carried out to certain point starting with 2p and the square root of D mod p above. The outcome might be used to construct the solutions (x,y) of the equation above, but there are still more chances of a failure, as you can see by inspecting the routine more closely. Even though, in view of the rather large list Δ the chances are good that we come up in the end with a handful of suitable values of D.

Now we must distinguish between three basic cases, namely

I. D = -3

II. D = -4

III. D < -4

We will assume the rather small value p = 37, so we might be able to check all computations manually. Furthermore, for this special value of p all 3 cases can be demonstrated.

```
p := 37

SELECT(VECTOR?(CS(p, D)), D, Δ)

                        [-3, -4, -7, -11, -67, -123]
```

Case I: D=-3

In the first place, there are six optional choices for the elliptic curves, which are all of the form

$$y^2 = x^3 - g^k,\ k=0,1,2,3,4,5$$

for some quadratic nonresidue mod p. Moreover, in the case $g \equiv 1 \bmod 3$, which is true in this example, g must not be a cube mod p. For example, g = 2 is an appropriate value for p = 13 here due to

```
                                                (p - 1)/3
g := ITERATE(IF(JACOBI(g_, p) = -1 ∧ MOD(g_           , p) ≠ 1, g_, g_ + 1), g_, 2)

                                        g := 2
```

For each of these curves E it is true that #E is one of the six numbers $p + 1 \pm u$, $p + 1 \pm (u \pm 3v)/2$, where the signs of u and v may be chosen independently here. In Order to find for each curve the correct value of n one must test the condition $n \cdot U = 0$ for some randomly chosen point $U \neq O$ on the elliptic curve. The following routine will do exactly this for us.

```
NOP1(p, k := 0, g_ := 1, n_, t_, x_ := -1, u_, v_, w_) :=
  Prog
    w_ := CS(p, -3)
    If ¬ VECTOR?(w_)
       RETURN ?
    Loop
      g_ :+ 1
      If JACOBI(g_, p) = -1 exit
    a := 0
    b := MODS(- g_^k, p)
    u_ := ABS(FIRST(w_))
    v_ := ABS(FIRST(REST(w_)))
    t_ := {u_, -u_, (u_ + 3·v_)/2, (u_ - 3·v_)/2}
    t_ := t_ ∪ {- (u_ + 3·v_)/2, - (u_ - 3·v_)/2}
    Loop
      Loop
        x_ :+ 1
        If JACOBI(x_^3 + b, p) = 1 exit
      If x_ ≥ p
         RETURN MAP_LIST(p + 1 + s_, s_, t_)
      w_ := [x_, SQUARE_ROOT(x_^3 + b, p)]
      t_ := SELECT(mult(w_, p + 1 + s_, 0, p) = [p, p], s_, t_)
      If DIM(t_) = 1
         RETURN p + 1 + FIRST(t_)
```

We check this routine by computing the number of points on E: $y^2=x^3-2$ both by using it and a simple brute force computation:

$$\text{NOP}(a, b, p) := p + 1 + \sum_{x=0}^{p-1} \text{JACOBI}(x^3 + a\cdot x + b, p)$$

```
VECTOR([y^2 = x^3 + MODS(- 2^k, p), n := NOP1(p, k), n = NOP(a, b, p)], k, 0, 5)'
```

$$\begin{bmatrix} y^2 = x^3 - 1 & y^2 = x^3 - 2 & y^2 = x^3 - 4 & y^2 = x^3 - 8 & y^2 = x^3 - 16 & y^2 = x^3 + 5 \\ n := 48 & n := 49 & n := 39 & n := 28 & n := 27 & n := 37 \\ \text{true} & \text{true} & \text{true} & \text{true} & \text{true} & \text{true} \end{bmatrix}$$

Case II: D=-4.

Here the corresponding CM-curves are of the form

$y^2=x^3 - g^k x$, k=0,1,2,3

where g is again some quadratic nonresidue mod p, for example g = 2 for p = 13 as above.

This time, for each of these curves E the number #E of its points is given by one of the four numbers p + 1 ± u, p + 1 ± 2v, where (u,v) is again any solution of the Diophantine equation #4 above. The correct value among these can be found out in an analogous way as before using the following routine:

```
NOP2(p, k := 0, g_ := 1, n_, t_, x_ := -1, u_, v_, w_) :=
  Prog
    w_ := CS(p, -4)
    If ¬ VECTOR?(w_)
       RETURN ?
    Loop
      g_ :+ 1
      If JACOBI(g_, p) = -1 exit
    a := MODS(- g_^k, p)
    b := 0
    u_ := ABS(FIRST(w_))
    v_ := ABS(FIRST(REST(w_)))
    t_ := {u_, -u_, 2·v_, - 2·v_}
    x_ := 1
    Loop
      Loop
        x_ :+ 1
        If JACOBI(x_^3 + a·x_, p) = 1 exit
      If x_ ≥ p
         RETURN MAP_LIST(p + 1 + s_, s_, t_)
      w_ := [x_, SQUARE_ROOT(x_^3 + a·x_, p)]
      t_ := SELECT(mult(w_, p + 1 + s_, a, p) = [p, p], s_, t_)
      If DIM(t_) = 1
         RETURN p + 1 + FIRST(t_)
```

$$\text{VECTOR}\left(\left[y^2 = x^3 + \text{MODS}(-2^k, p)\cdot x,\ n := \text{NOP2}(p, k),\ n = \text{NOP}(a, b, p)\right], k, 0, 3\right)$$

$$\begin{bmatrix} y^2 = x^3 - x & y^2 = x^3 - 2\cdot x & y^2 = x^3 - 4\cdot x & y^2 = x^3 - 8\cdot x \\ n := 40 & n := 26 & n := 36 & n := 50 \\ \text{true} & \text{true} & \text{true} & \text{true} \end{bmatrix}$$

<u>Case III: D<-4.</u>

Here the corresponding CM-curves are of the form

$y^2=x^3-3rs^3x+2rs^5$ or $y^2=x^3-3rs^3g^2x+2rs^5g^3$

where g is again some quadratic nonresidue mod p, e.g. g = 2 for p = 13 as above.

But what about those mysterious numbers r and s occurring in these formulas, where do they come from? Well, telling you the whole story (just in case, you don't know it yet!) would take a lot of space here and lead us too far away. If you are interested in it (as well as in many other topics here I didn't have the time to dwell on) , I refer you to the wonderful book "Prime Numbers – A Computational Perspective" by R. Crandall and C. Pomerance, which is sort of a "bible" when it comes to computational number theory.

Here comes the solution:

rlist :=

D	r
-7	[125, 0, 0]
-8	[125, 0, 0]
-11	[512, 0, 0]
-19	[512, 0, 0]
-43	[512000, 0, 0]
-67	[85184000, 0, 0]
-163	[151931373056000, 0, 0]
-15	[1225, -2080, 5]
-20	[108250, 29835, 5]
-24	[1757, -494, 2]
-35	[-1126400, -1589760, 5]
-40	[54175, -1020, 5]
-51	[75520, -7936, 17]
-52	[1778750, 5125, 13]
-88	[181713125, -44250, 2]
-91	[74752, -36352, 13]
-115	[269593600, -89157120, 5]
-123	[1025058304000, -1248832000, 41]
-148	[499833128054750, 356500625, 37]
-187	[91878880000, -1074017568000, 17]
-232	[1728371226151263375, -11276414500, 29]
-235	[7574816832000, -190341944320, 5]
-267	[3632253349307716000000, -12320504793376000, 89]
-403	[16416107434811840000, -4799513373120384000, 13]
-427	[564510997315289728000, -5784785611102784000, 61]

, slist :=

D	s
-7	189
-8	98
-11	539
-19	513
-43	512001
-67	85184001
-163	151931373056001
-15	5929
-20	174724
-24	1058
-35	2428447
-40	51894
-51	108241
-52	1797228
-88	181650546
-91	205821
-115	468954981
-123	1033054730449
-148	499835296563372
-187	4520166756633
-232	1728371165425912854
-235	8000434358469
-267	3632369580717474122449
-403	337209989988725140 77
-427	609691617259594724421

```
rslookup(D, p) :=
  Prog
    r := rlist↓POSITION(D, rlist COL 1)↓2
    r := r↓1 + r↓2·SQUARE_ROOT(r↓3, p)
    s := slist↓POSITION(D, slist COL 1)↓2
```

To simplify matters, as for those values of r and s, I'll simply look them up in the tables on the previous page, rather than compute them from scratch. Needless to say that these tables can also be found in the book quoted above though in a slightly different form, In particular, a triple [a,b,c] in the table named `rlist` corresponds to the term a + b·c mod p, where again the square root must be computed mod p.

For each of these curves E it is true that n = #E is one of the two numbers p + 1 ± u, where (u,v) is again a solution of our standard Diophantine equation #4 above. The eventually correct value of n can be found in an analogous way as before using the following routine:

```
NOP3(p, D, k := 0, g_ := 1, n_, t_, x_ := -1, u_, v_, w_) :=
  Prog
    w_ := CS(p, D)
    If ¬ VECTOR?(w_)
       RETURN ?
    rslookup(D, p)
    Loop
      g_ :+ 1
      If JACOBI(g_, p) = -1 exit
    a := MODS(- 3·r·s^3·g_^(2·k), p)
    b := MODS(2·r·s^5·g_^(3·k), p)
    u_ := ABS(FIRST(w_))
    v_ := ABS(FIRST(REST(w_)))
    t_ := {u_, -u_}
    Loop
      Loop
        x_ :+ 1
        If JACOBI(x_^3 + a·x_ + b, p) = 1 exit
      If x_ ≥ p
         RETURN MAP_LIST(p + 1 + s_, s_, t_)
      w_ := [x_, SQUARE_ROOT(x_^3 + a·x_ + b, p)]
      t_ := SELECT(mult(w_, p + 1 + s_, a, p) = [p, p], s_, t_)
      If DIM(t_) = 1
         RETURN p + 1 + FIRST(t_)
```

```
REVERSE(VECTOR([n := NOP3(p, -7, k), n = NOP(a, b, p), y^2 = x^3 + a·x + b], k, 0, 1)')
```

$$\begin{bmatrix} y^2 = x^3 + 13\cdot x - 3 & y^2 = x^3 + 15\cdot x + 13 \\ \text{true} & \text{true} \\ n := 32 & n := 44 \end{bmatrix}$$

Ok, one last routine and we are finally finished with our setup. It deals with the fact that not every elliptic curve is suitable for our purposes, but what we need is an elliptic curve of an appropriate order that is "almost" a prime, i.e., that is divisible only by small factors. The following routine will show if this condition is

sufficiently fulfilled by removing small factors from the order n of the curve up to a certain bound s. If it is a "good" curve, the remaining part should be a prime even if s is rather small.

```
trialdiv(n, s := 10^5, d_ := 4, s_ := [], t_ := 2, d_ := 4) :=
  Loop
    Loop
      If MOD(n, t_) > 0 exit
      s_ := ADJOIN(t_, s_)
      n :/ t_
    t_ :+ d_^IF(t_ ≥ 5)
    d_ := 6 - d_
    If t_^2 > n
       If n > 1
          RETURN REVERSE(ADJOIN(n, s_))
          RETURN REVERSE(s_)
    If t_ > s
       Prog
         d_ := APPEND(IF(PRIME?(n), "P", "C"), STRING(DIM(n)))
         RETURN REVERSE(ADJOIN(d_, s_))
```

Now, what do you think: Will our routines above pass the acid test when using primes of realistic order in a cryptographic environment, say with 160 bits and more? Let's try it out.

```
                  159           159
p := NEXT_PRIME(2    + RANDOM(2   ))

p := 979041373284628512978859729772372544317253699343

SELECT(VECTOR?(CS(p, D)), D, Δ) = [-3, -19, -163, -24, -123]
```

Ok, we have in total 6+2+2+2+2=10 elliptic curves a tour disposal, let's check now if there is also one with a „big" prime factor. To be honest, there is also a lot of „garbage", but two or three are really good (see the accompanying DERIVE-File for a complete list!), the best one being the last below

```
                                                    2    3                  ]
VECTOR(REVERSE([n := NOP3(p, -19, k), trialdiv(n), y  = x  + a·x + b]), k, 0, 1)

[   2    3
 y  = x  - 207368750592·x + 36382017816364032       [19, 467, C45]       n :=

  2    3
 y  = x  - 1866318755328·x + 982314481041828864  [3, 3, 5, 11, 457, C43]  n :=

  979041373284628512978859508298659993048973669403 ]

  979041373284628512978859951246085095585533729285 ]
```

```
...
...
VECTOR(REVERSE([n := NOP3(p, -123, k), trialdiv(n), y^2 = x^3 + a·x + b]), k, 0, 1)

[

]

  2    3                                                                    ~
 y  = x  + 10812091809269166011513553763102577591592406849 3·x + 43197587274344143962 67~
                                                                           ~
   2    3                                                                  ~
  y  = x  - 59531104504035719426398910931405610739370829 06·x - 85147915342623285823889~

  565719777286824050204662 41  [11, 64763, P43]  n :=

  313869796106871491803609       [3, P48]      n :=

  979041373284628512978857852967500664853268633179

  979041373284628512978861606577244423781238765509 ]
```

As a matter of fact, I'll use exactly this one in part 2 of this treatise on the ElGamal dealing with all the goodies it has to offer. Hope to see you again here!

Invitation to visit Europe's Cultural Capital 2013 - Marseille

Hello,
I hope that you are keeping well .
Maybe you will remember Bernard Eggger who was in Time Montreal in 2004 ? He is now president of the French association of math's teachers for Provence Region (Aix-Marseille)
And as Marseille is Europe's capital for culture in 2013 , he decided to held the Association Annual Meeting there .
I am also involved in the preparation of this meeting, not so much scientifically, but on the organisation level.
It is not exactly an international meeting because all the conferences and workshops will be in French ,yet for this "capital" year it is planned to have a kind of Science Fair (called Souk) which will be more specifically open to teachers from abroad, if they which to present Posters, or whatever , even not in French. It is open to all of course, although we suppose that it will be more particularly of interest for teachers from Mediterranean countries.
As it is the first (and probably last) time that I am busy myself with this kind of venture ,
I thought that I would let you know .

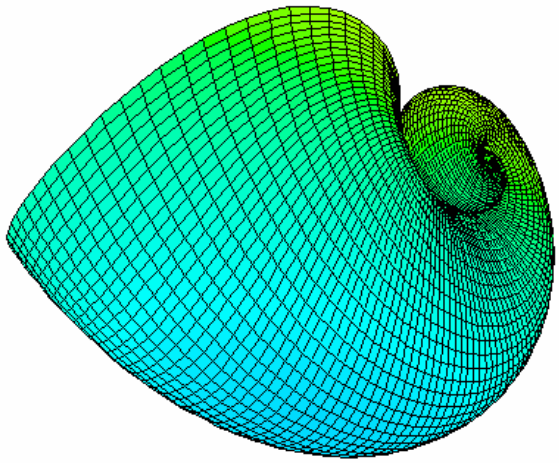
So you may just visit the site :

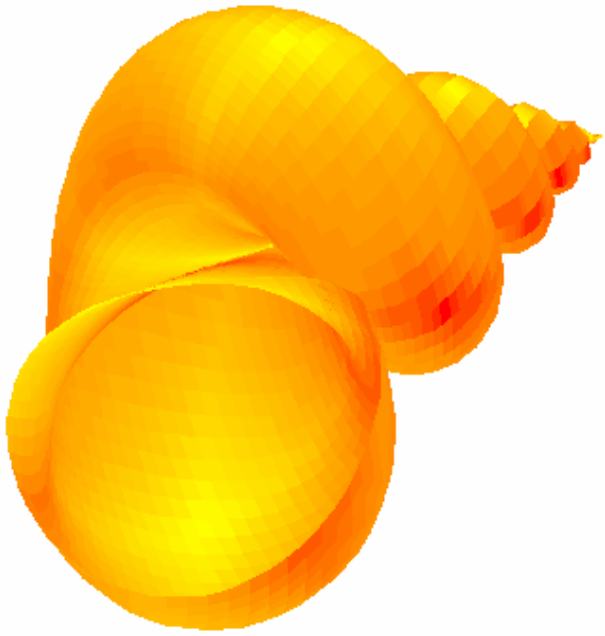
« Les mathématiques au carrefour des cultures de la Méditerranée »

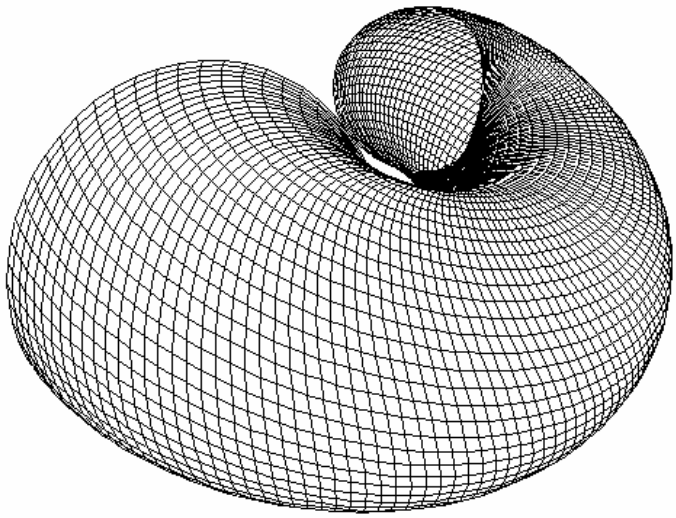
http://www.jnmarseille2013.fr

With best regards
Mit freundlichen Grüßen
Marie-Laure Laurent

I prepared Piotr Trebisz's 4th part of his "Snail Shell" series. Guido Herweyers' Statistics 4 got its self dynamic and needed more space than expected. So I must leave the Conical Shells for the next DNL. See some "mouthwatering" plots.







Impressions from Laos, Vietnam and Cambodia

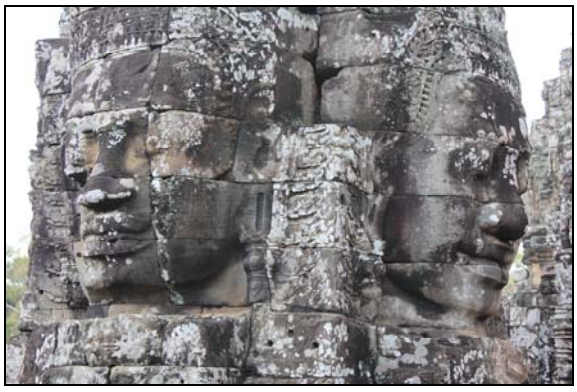
Sunset over the Mekong, Luang Prabang, Laos

Vientiane, Laos

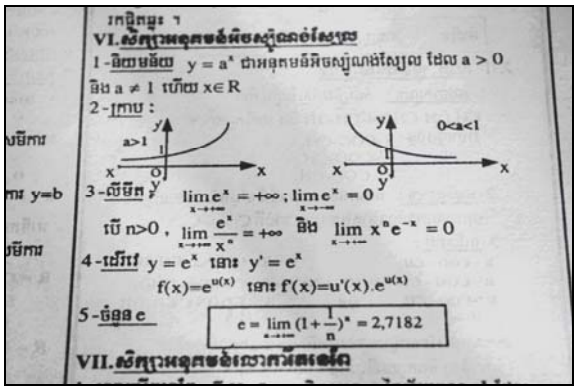
One Pillar Pagoda, Hanoi

Students in the Temple of Literature, Hanoi

Halong Bay



Bayon, Angkor Wat, Cambodia



Calculus in Cambodia