Unofficial Report SUF–PFY/91–02
Stockholm, December 20, 1991

PARABOLIC SPLINE FUNCTIONS USED IN RANDOM NUMBER GENERATION

by

Christian Walck

Particle Physics Group
Department of Physics
University of Stockholm

Contents

1 Introduction

It is well known that adding two uniform pseudorandom numbers yields a triangular distribution. Adding three we get a parabolic spline function, for four a cubic spline function etc. The parts are connected in so called *knots* in which the function values as well as the first derivative are continuous. Gradually we obtain a distribution closer and closer to the standard normal distribution, denoted $\mathcal{N}(x)$ below. Adding twelve random numbers is particularly convenient if we want to standardize the resulting variable (whose variance then become unity) and this is sometimes used as an approximative method for obtaining random numbers from a normal distribution. In this paper we analyze in detail the properties of the parabolic spline functions obtained by making linear combinations of three random numbers from a uniform distribution between zero and one.

2 Linear Combinations of Three Uniform Random Numbers

Write a linear combination of three uniform random numbers between zero and one as

$$a\xi_1 + b\xi_2 + c\xi_3 - d$$

where ξ_1, ξ_2 and ξ_3 are uniform random numbers between zero and one. Without loss of generality always use $d = (a + b + c)/2$ in order to get a distribution symmetric about zero and normalize the resulting function to unity. Also require that $0 \leq a \leq b < c$ since this will be sufficient to cover all possible cases. Because of the symmetry of the distribution only the function above $x = 0$ is regarded in the following.

3 Special Cases

First look at the special cases arising when not all of a, b and c are non-zero:

- $a = b = 0$ gives a uniform distribution between $-\frac{c}{2}$ and $\frac{c}{2}$.
- $a = 0$ and $b = c$ gives a triangular distribution between $-b$ and b with its maximum at $x = 0$ at a height of $\frac{1}{b}$.
- $a = 0$ and $b < c$ gives a trapezoidal distribution between $-\frac{b+c}{2}$ and $\frac{b+c}{2}$ having vertices at the points $(\pm\frac{c+b}{2}, 0)$ and $(\pm\frac{c-b}{2}, \frac{1}{c})$.

4 Parabolic Spline Functions

Turning to the more general case where a, b and c all are non-zero the equations for each interval $f_1, f_2, ..., f_n$ for $1 \leq n \leq 4$ are defined as

$$f_1(x) = a_0 + a_1 x + a_2 x^2$$

$$f_2(x) = b_0 + b_1 x + b_2 x^2$$

$$f_3(x) = c_0 + c_1 x + c_2 x^2$$

$$f_4(x) = d_0 + d_1 x + d_2 x^2$$

For negative x-values the function value is the same as for $|x|$. The first interval always start at $x_0 = 0$, the knot positions are denoted $x_1, x_2, .., x_{n-1}$ and the end-point x_n where n are the number of functions for $x > 0$ (*i.e.* the number of knots plus one). The function f_i is thus defined for the interval x_{i-1} to x_i .

5 Constraint Equations

The constraint equations used are

$$\begin{aligned}
f_1'(0) &= 0 \\
f_1(x_1) &= f_2(x_1) \\
f_1'(x_1) &= f_2'(x_1) \\
&\vdots \\
f_n(x_n) &= 0 \\
f_n'(x_n) &= 0 \\
\int_0^{x_1} f_1(x)dx + \ldots + \int_{x_{n-1}}^{x_n} f_n(x)dx &= \frac{1}{2}
\end{aligned}$$

Note, however, that this is the case for the true parabolic splines only, *i.e.* when a, b and c all are non-zero. *E.g.* in the trapezoidal case above the function value but *not* the first derivative is continuous at the knot.

6 Knot Positions

To find the positions of the knots write the expression as

$$\left(a\xi_1 - \frac{a}{2}\right) + \left(b\xi_2 + c\xi_3 - \frac{b+c}{2}\right).$$

The second part is distributed as a trapezoid symmetric around zero and with vertices at $x = \frac{c-b}{2}$ and $x = \frac{b+c}{2}$. The first part is a uniform random number between $-\frac{a}{2}$ and $+\frac{a}{2}$. Possible knots are thus at $\frac{c-b-a}{2}$, $\frac{c-b+a}{2}$, $\frac{b+c-a}{2}$ and the end-point is at $\frac{a+b+c}{2}$. Here one must watch up if $\frac{c-b-a}{2}$ is negative and in such a case regard the absolute value as a knot position. The maximum number of knots possible are three (on each side of $x = 0$).

7 Classification

The parabolic spline functions arising from linear combinations of three uniform random numbers are classified according to the number of knots. Also some subclasses are defined depending on the actual values of the linear coefficients.

- $a = b = c$. In this case there is only one genuine knot at $\frac{a}{2}$.
- $a = b$. Possible knot positions at $\frac{c-2a}{2}$, $|\frac{c-2a}{2}|$ and $\frac{c}{2}$. Besides the case $a = c$ which reduces to the previous case with $a = b = c$ we distinguish three possibilities:
  - $c = 2a$. In this case there is only one knot at $\frac{c}{2}$.
  - $c > 2a$. In this case there are two knots at $\frac{c}{2} - a$ and $\frac{c}{2}$.
  - $c < 2a$. In this case there are have two knots at $a - \frac{c}{2}$ and $\frac{c}{2}$.
- $b = c$. Possible knots at $\frac{a}{2}$ and $\frac{2c-a}{2}$. Except for the case where $c = a$, and thus $a = b = c$, there are two knots at $\frac{a}{2}$ and $\frac{2c-a}{2}$, respectively.
- a, b and c all different. Possible knots at $\frac{c-b-a}{2}$ or $|\frac{c-b-a}{2}|$, $\frac{c-b+a}{2}$ and $\frac{c+b-a}{2}$. This gives three knots unless $c - b - a = 0$ or $c - b + a = c + b - a$. The latter case implies that $a = b$ and is thus treated above. Here three cases are distinguished depending on if $c - a - b$ is negative, zero or positive (see table below).

To summarize the following classes are defined:

class	parameters	knots	x_1	x_2	x_3
Uni	$a = b = 0$	0			
Tri	$a = 0$ & $b = c$	0			
Tpz	$a = 0$ & $b < c$	1	$(c - b)/2$		
1a	$a = b = c$	1	$a/2$		
1b	$a = b$ & $c = 2a$	1	a		
2a	$b = c$	2	$a/2$	$b - a/2$	
2b	$c - b - a = 0$	2	a	b	
2c	$a = b$ & $c > 2a$	2	$c/2 - a$	$c/2$	
2d	$a = b$ & $c < 2a$	2	$a - c/2$	$c/2$	
3a	$c - b - a > 0$	3	$(c - b - a)/2$	$(c - b + a)/2$	$(c + b - a)/2$
3b	$c - b - a < 0$	3	$(a + b - c)/2$	$(c - b + a)/2$	$(c + b - a)/2$

Uni, Tri and Tpz stands for the uniform, triangular and trapezoidal case.

8 Parabolic Coefficients

In the case of one knot the constraint equations are sufficient to solve for the unknown coefficients a_0, a_1, a_2, b_0, b_1 and b_2. In these cases one may simplify and write:

$$f^{1a}(x) = \begin{cases} \frac{1}{a}\left(\frac{3}{4} - \left(\frac{x}{a}\right)^2\right) & \text{if } 0 \leq |x| < \frac{a}{2} \\ \frac{1}{2a}\left(\frac{3}{2} - \frac{|x|}{a}\right)^2 & \text{if } \frac{a}{2} \leq |x| \leq \frac{3a}{2} \\ 0 & \text{if } \frac{3a}{2} \leq |x| \end{cases}$$

$$f^{1b}(x) = \begin{cases} \frac{1}{a}\left(\frac{1}{2} - \left(\frac{x}{2a}\right)^2\right) & \text{if } 0 \leq |x| < a \\ \frac{1}{a}\left(1 - \frac{|x|}{2a}\right)^2 & \text{if } a \leq |x| < 2a \\ 0 & \text{if } 2a \leq |x| \end{cases}$$

In the cases with more than one knot we have determined the coefficients empirically.

	Uni	Tri	Tpz	1a	1b	2a	2b
a_0	$\frac{1}{c}$	$\frac{1}{b}$	$\frac{1}{c}$	$\frac{3}{4a}$	$\frac{1}{2a}$	$\frac{4b-a}{4b^2}$	$\frac{1}{c}$
a_1	0	$-\frac{1}{b^2}$	0	0	0	0	0
a_2	0	0	0	$-\frac{1}{a^3}$	$-\frac{1}{4a^3}$	$-\frac{1}{ab^2}$	$-\frac{1}{2abc}$
b_0			$\frac{b+c}{2bc}$	$\frac{9}{8a}$	$\frac{1}{a}$	$\frac{1}{b}$	$\frac{b+c}{2bc}$
b_1			$-\frac{1}{bc}$	$-\frac{3}{2a^2}$	$-\frac{1}{a^2}$	$-\frac{1}{b^2}$	$-\frac{1}{bc}$
b_2			0	$\frac{1}{2a^3}$	$\frac{1}{4a^3}$	0	0
c_0						$\frac{(a+2b)^2}{8ab^2}$	$\frac{c}{2ab}$
c_1						$-\frac{a+2b}{2ab^2}$	$-\frac{1}{ab}$
c_2						$\frac{1}{2ab^2}$	$\frac{1}{2abc}$

	2c	2d	3a	3b
a_0	$\frac{1}{c}$	$\frac{4a-c}{4a^2}$	$\frac{1}{c}$	$\frac{4ab-(c-a-b)^2}{4abc}$
a_1	0	0	0	0
a_2	0	$-\frac{1}{a^2c}$	0	$-\frac{1}{abc}$
b_0	$\frac{4a^2+4ac-c^2}{8a^2c}$	$\longleftarrow$	$\frac{8ab-(c-b-a)^2}{8abc}$	$\longleftarrow$
b_1	$\frac{c-2a}{2a^2c}$	$\longleftarrow$	$\frac{c-b-a}{2abc}$	$\longleftarrow$
b_2	$-\frac{1}{2a^2c}$	$\longleftarrow$	$-\frac{1}{2abc}$	$\longleftarrow$
c_0	$\frac{(2a+c)^2}{8a^2c}$	$\longleftarrow$	$\frac{b+c}{2bc}$	$\longleftarrow$
c_1	$-\frac{2a+c}{2a^2c}$	$\longleftarrow$	$-\frac{1}{bc}$	$\longleftarrow$
c_2	$\frac{1}{2a^2c}$	$\longleftarrow$	0	$\longleftarrow$
d_0			$\frac{(a+b+c)^2}{8abc}$	$\longleftarrow$
d_1			$-\frac{a+b+c}{2abc}$	$\longleftarrow$
d_2			$\frac{1}{2abc}$	$\longleftarrow$

Among all these parabolic functions one, in fact, distinguishes only five distinct types if written in their most general form:

$$f_1^{\mathrm{UNI}}(x) = f_1^{\mathrm{TPZ}}(x) = f_1^{2c}(x) = f_1^{3a}(x) = \frac{1}{c}$$

$$f_1^{\mathrm{TRI}}(x) = f_2^{\mathrm{TPZ}}(x) = f_2^{2a}(x) = f_2^{2b}(x) = f_3^{3a}(x) = f_3^{3b}(x) = \frac{1}{bc}\left(\frac{b+c}{2} - x\right)$$

$$f_1^{1a}(x) = f_1^{2a}(x) = f_1^{2d}(x) = f_1^{3b}(x) = \frac{1}{c} - \frac{1}{abc}\left(\left(\frac{c-a-b}{2}\right)^2 + x^2\right)$$

$$f_1^{1b}(x) = f_1^{2b}(x) = f_2^{2c}(x) = f_2^{2d}(x) = f_2^{3a}(x) = f_2^{3b}(x) = \frac{1}{c} - \frac{1}{2abc}\left(x - \frac{c-a-b}{2}\right)^2$$

$$f_2^{1a}(x) = f_2^{1b}(x) = f_3^{2a}(x) = f_3^{2b}(x) = f_3^{2c}(x) = f_3^{2d}(x) = f_4^{3a}(x) = f_4^{3b}(x) = \frac{1}{2abc}\left(\frac{a+b+c}{2} - x\right)^2$$

formulæ which sometimes simplify if we insert the specific relations between a, b and c. Note, however, that the ranges in which a certain function is valid vary in different cases due to the actual knot positions. The last equation implies that all true parabolic spline functions tail off in the same way with the function value as well as the first derivative becoming zero at the end-point. The expressions are written for $x > 0$, use $|x|$ to obtain function values for $x < 0$.

9 Moments

The mean value of the parabolic spline functions are zero and thus algebraic and central moments coincide. Moreover, due to the symmetry of the distributions, all moments off odd order vanish. Even order moments for the different classes are given by

$$\begin{aligned}
\mu_{2\ell}^{\mathrm{UNI}} &= \frac{1}{2\ell+1}\cdot\left(\frac{c}{2}\right)^{2\ell} \\
\mu_{2\ell}^{\mathrm{TRI}} &= \frac{2b^{2\ell}}{(2\ell+1)(2\ell+2)} \\
\mu_{2\ell}^{\mathrm{TPZ}} &= \frac{(c+b)^{2\ell}-(c-b)^{2\ell}}{bc2^{2\ell+1}(2\ell+1)(2\ell+2)} \\
\mu_{2\ell}^{1a} &= \frac{3}{4}\cdot\frac{3^{2\ell+2}-1}{(2\ell+1)(2\ell+2)(2\ell+3)}\left(\frac{a}{2}\right)^{2\ell} \\
\mu_{2\ell}^{1b} &= \frac{2a^{2\ell}(2^{2\ell+2}-1)}{(2\ell+1)(2\ell+2)(2\ell+3)} \\
\mu_{2\ell}^{2a} &= \frac{(2b+a)^{2\ell+3}-(2b-a)^{2\ell+3}-2a^{2\ell+3}}{ab^2 2^{2\ell+2}(2\ell+1)(2\ell+2)(2\ell+3)} \\
\mu_{2\ell}^{2b} &= \frac{2}{ab(a+b)}\cdot\frac{(a+b)^{2\ell+3}-b^{2\ell+3}-a^{2\ell+3}}{(2\ell+1)(2\ell+2)(2\ell+3)} \\
\mu_{2\ell}^{2c} &= \frac{2}{a^2c2^{2\ell+3}}\cdot\frac{(2a+c)^{2\ell+3}-2c^{2\ell+3}-(c-2a)^{2\ell+3}}{(2\ell+1)(2\ell+2)(2\ell+3)} \\
\mu_{2\ell}^{2d} &= \frac{2}{a^2c2^{2\ell+3}}\cdot\frac{(2a+c)^{2\ell+3}-2c^{2\ell+3}-(2a-c)^{2\ell+3}}{(2\ell+1)(2\ell+2)(2\ell+3)} \\
\mu_{2\ell}^{3a} &= \frac{2}{abc2^{2\ell+3}}\cdot\frac{(c+b+a)^{2\ell+3}-(c+b-a)^{2\ell+3}-(c-b+a)^{2\ell+3}-(c-b-a)^{2\ell+3}}{(2\ell+1)(2\ell+2)(2\ell+3)} \\
\mu_{2\ell}^{3b} &= \frac{2}{abc2^{2\ell+3}}\cdot\frac{(c+b+a)^{2\ell+3}-(c+b-a)^{2\ell+3}-(c-b+a)^{2\ell+3}-(a+b-c)^{2\ell+3}}{(2\ell+1)(2\ell+2)(2\ell+3)}
\end{aligned}$$

for $\ell = 1, 2, \ldots$. Since we are dealing with polynomials this is the result of straightforward but somewhat tedious calculations. The class 3 form is the most general from which the others could be obtained by inserting the proper relations between the coefficients.

10 Standard Normal Distribution

To describe some connections in which the linear combination of three random numbers are useful we take the case where we want to make random numbers from the standard normal distribution. The aim is to cover as large part as possible of the area under the standard normal distribution

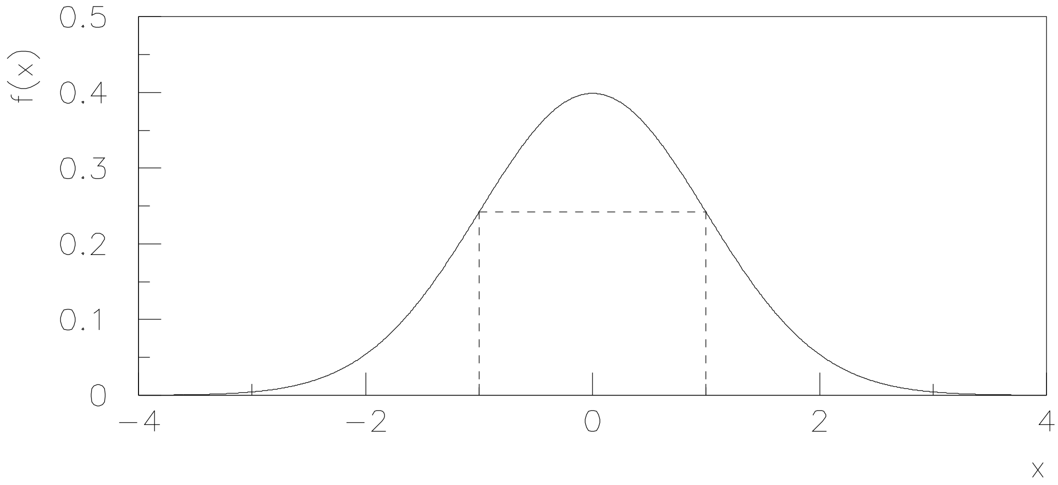
$$\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

with a parabolic spline function from which random number generation is fairly fast. The remaining parts may then be dealt with using somewhat more tedious reject-accept and tail techniques. Starting from the simplest case such methods are described in the following subsections.

10.1 Uniform Distribution

The largest (maximizing the area) uniform distribution which may be inscribed under the standard normal curve has parameters $a = b = 0$ and $c = 2$ as is shown in the figure below. It covers

48.39 % of the area under the curve.



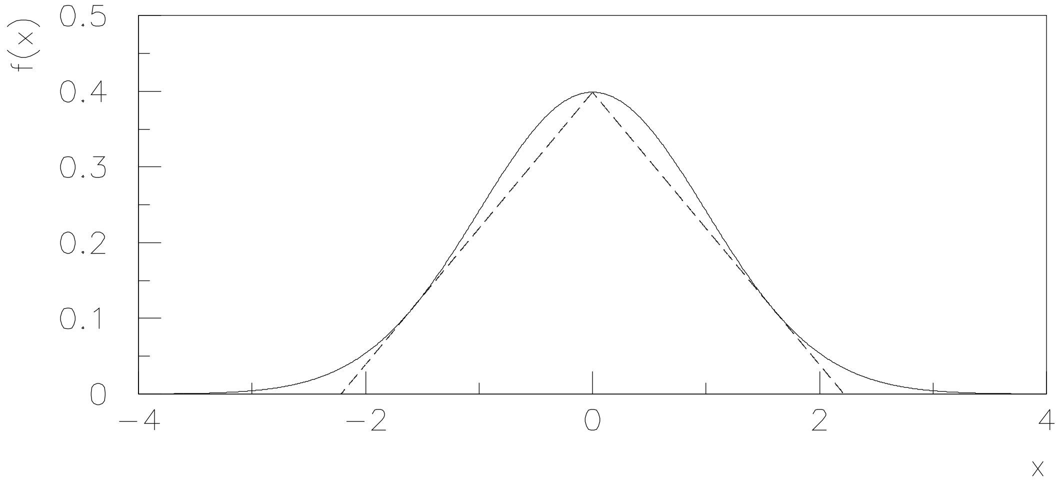
This is easily calculated by maximizing the area

$$\mathcal{Y}(c) = c\mathcal{N}(\frac{c}{2}) = \frac{c}{\sqrt{2\pi}}e^{-c^2/8} \Rightarrow \frac{d\mathcal{Y}(c)}{dc} = \frac{1}{\sqrt{2\pi}}e^{-c^2/8}\left(1 - \frac{c^2}{4}\right)$$

from which we see, by putting the derivative to zero, that c=2 gives the maximum area $\mathcal{Y}(2) = \sqrt{2/\pi e}$.

10.2 Triangular Distribution

The largest triangular distribution which may be inscribed has parameters $a = 0$ and $b = c = 2.216$ and covers 88.41 % of the area. This is shown in the figure below



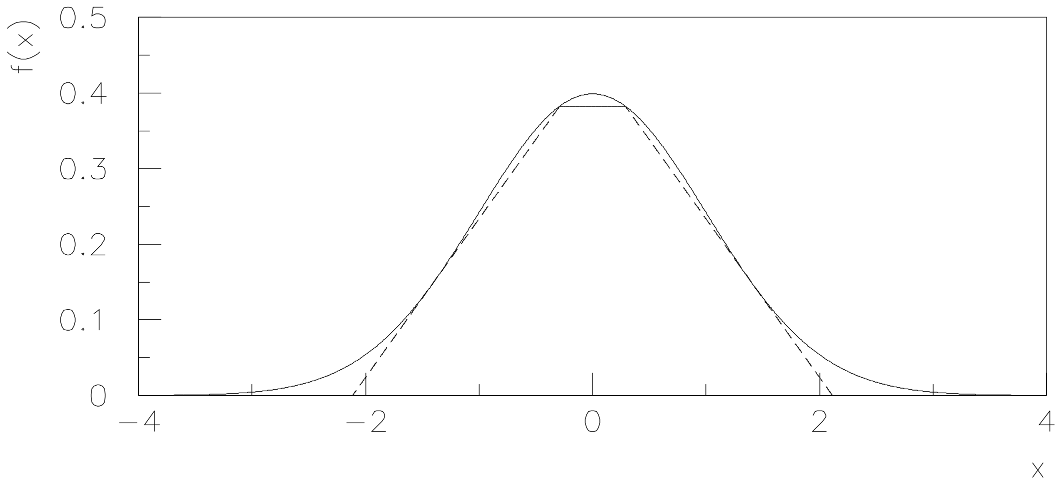
To calculate this we regard the point, x, at which the triangle slope touches the standard normal curve. In this point the function values as well as the first derivative of the curve and the triangle are the same. From this we may calculate that $b = c = x + \frac{1}{x}$. The maximum area is obtained when the triangle peak is at the peak of the curve. This gives the equation

$$\mathcal{N}(0) = \mathcal{N}(x) - x\frac{d\mathcal{N}(x)}{dx} \quad \Rightarrow \quad e^{-x^2/2}\left(x^2+1\right) - 1 = 0$$

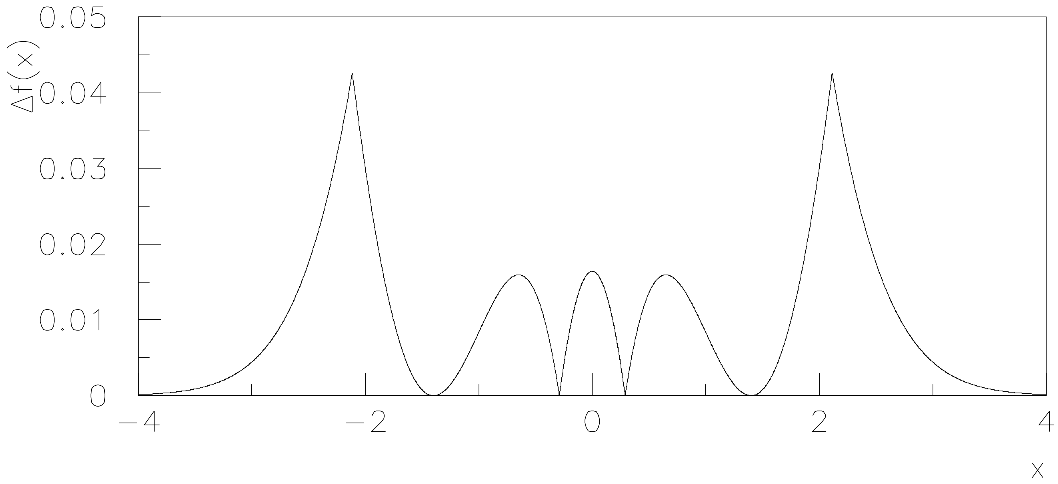
which solved *e.g.* by the Newton-Raphson method gives $x = 1.5852$ implying $c = b = 2.21604$.

10.3 Trapezoidal Distribution

The largest trapezoid which may be inscribed has parameters $a = 0$, $b = 1.8243$ and $c = 2.404$ and covers 91.95 % of the area. It is shown in the figure below



This method is used in one of our random number routines (RNNRMT) and is in fact a quite fast method to obtain such random numbers [1, 2]. The remaining parts under the standard normal curve, shown in the figure below, are treated by common reject-accept and tail techniques.



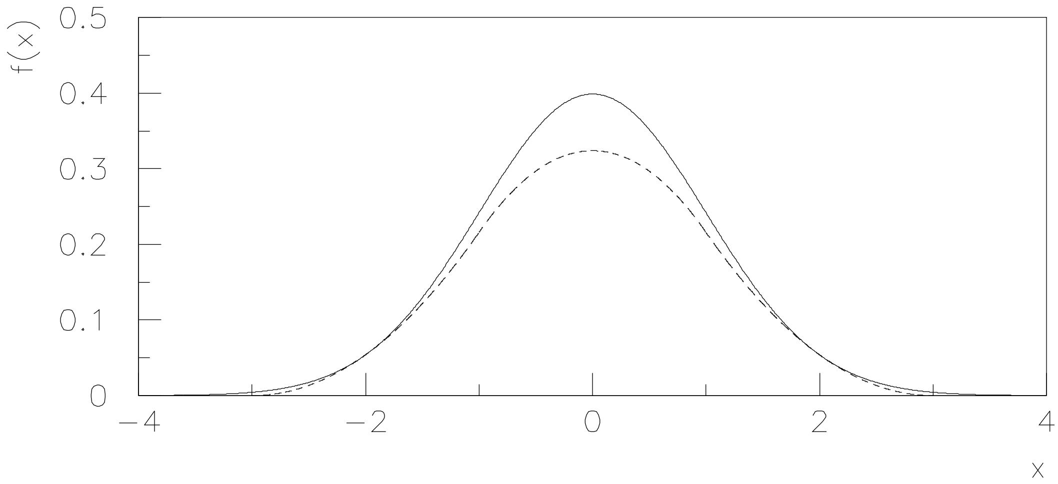
The trapezoidal case we treat in a similar manner as for the triangular distribution. Regarding the touching point, x, between the curve and the slope of the trapezoid we get the end-point of the trapezoid at $x_2 = x + \frac{1}{x}$. To find in which point, x_1, the slope of the trapezoid intersects the standard normal curve we solve the equation

$$\mathcal{N}(x_1) = \mathcal{N}(x) + (x_1 - x)\frac{d\mathcal{N}(x)}{dx} \Rightarrow e^{-x_1^2/2} + (x(x_1 - x) - 1)e^{-x^2/2} = 0$$

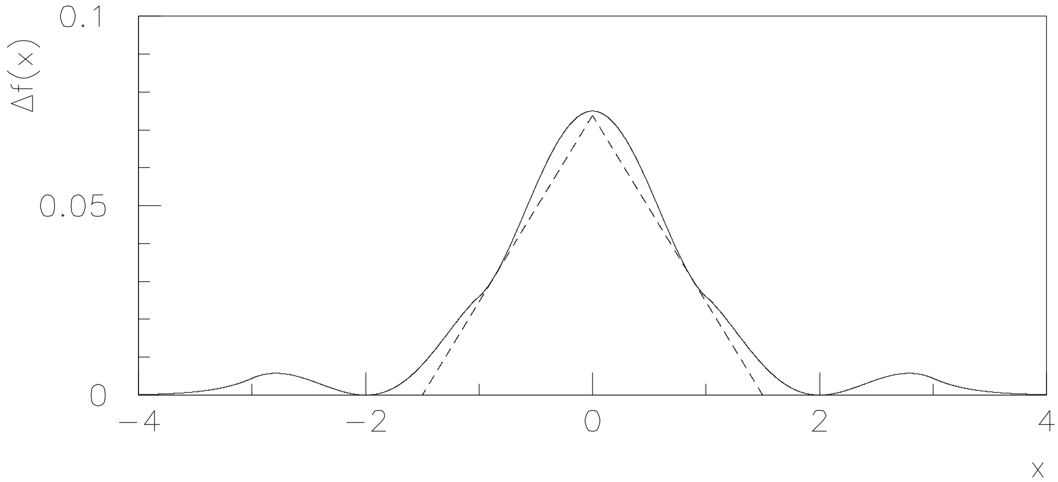
using the Newton-Raphson method. We then maximize the area of the trapezoid area as a function of x and find $x = 1.3995$, $x_1 = 0.2897$ and $x_2 = 2.1141$. This corresponds to $b = x_2 - x_1 = 1.82438$ and $c = x_1 + x_2 = 2.40371$. Note that for $x < 1$ our method is senseless since the trapezoid slope then would intersect the curve. The trapezoid would then not be inscribed. Also for $x > 1.5852$ the trapezoid collapses to a triangle.

10.4 Parabolic Spline $a = b = c = 2$

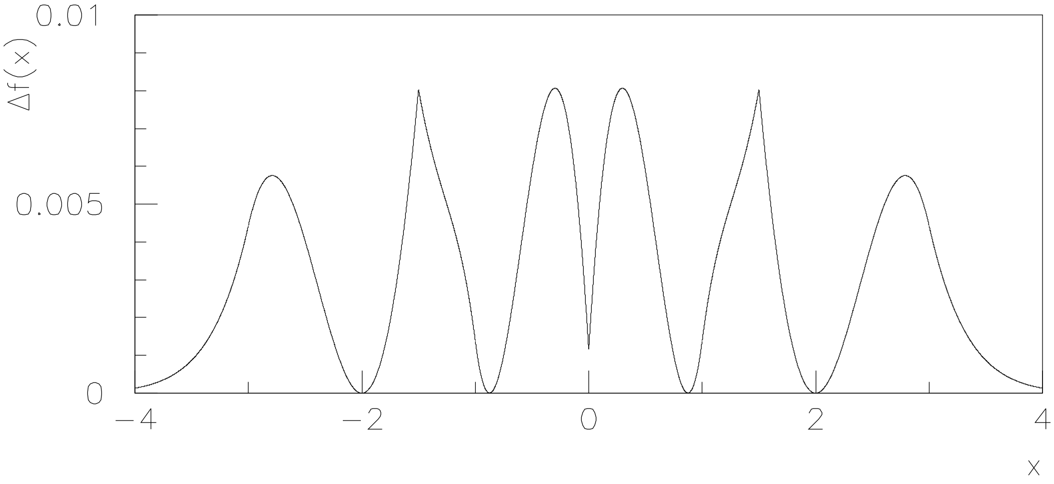
Using $a = b = c = 2$ we note that such a distribution is used in a method by George Marsaglia which we have coded in the routine **RNNRM6** [3, 4, 2]. This is shown in the figure below



Indeed, this method goes even further in inscribing a triangular distribution, with parameters $a = 0$ and $b = c = \frac{3}{2}$, under the curve for the difference between the standard normal curve and the parabolic spline function.

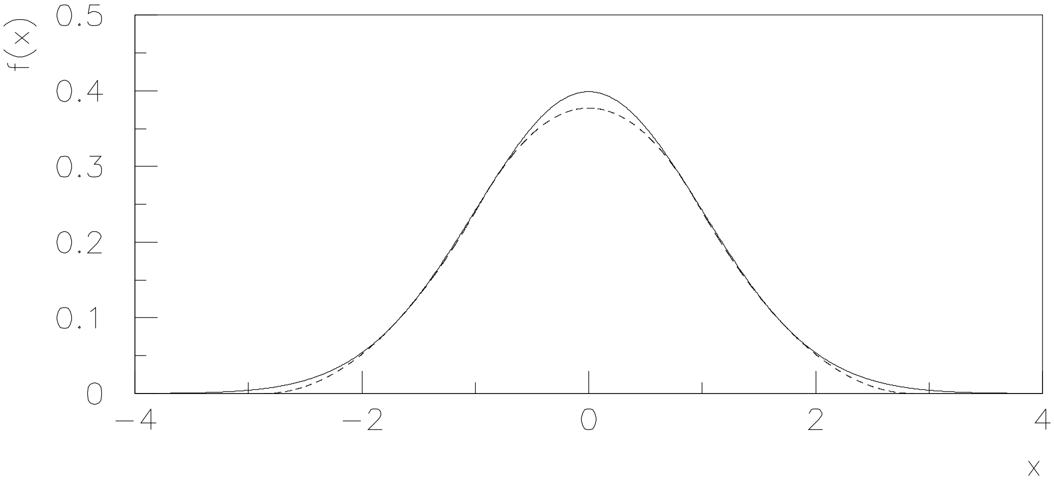


The area covered by the parabolic spline function is $16e^{-2}/\sqrt{2\pi} = 86.39\%$ and another 11.08 % is covered by the triangular distribution. The remaining area, shown in the figure below, is treated by common tail and reject-accept techniques.

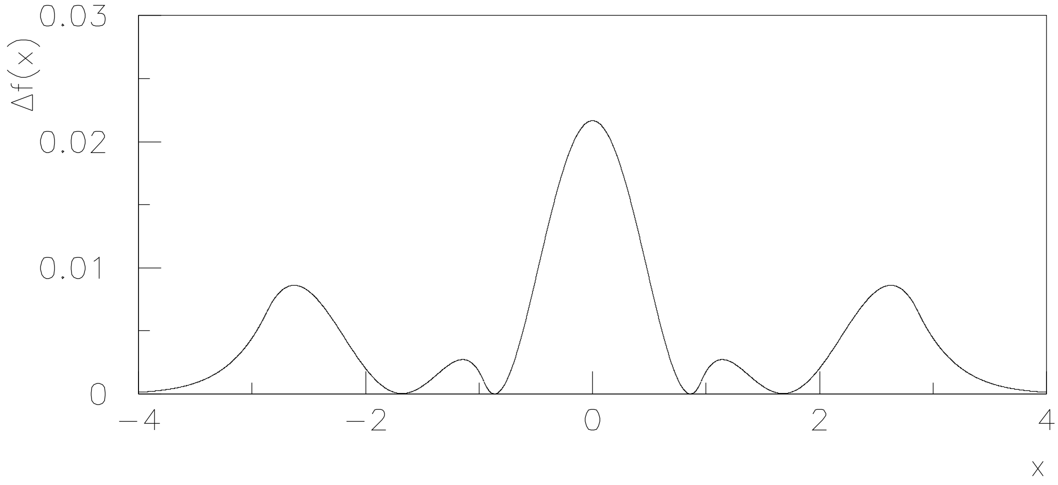


10.5 Parabolic Spline $a = b = c$

Marsaglia uses $a = b = c = 2$ a convenient choice of a parabolic spline with only one knot (class 1a). Trying to maximize the area for the case $a = b = c$ we find that for $a = b = c = 1.913$ we get a distribution which covers 96.23 % of the standard normal curve. The normal curve with such a parabolic spline inscribed is shown in the figure below

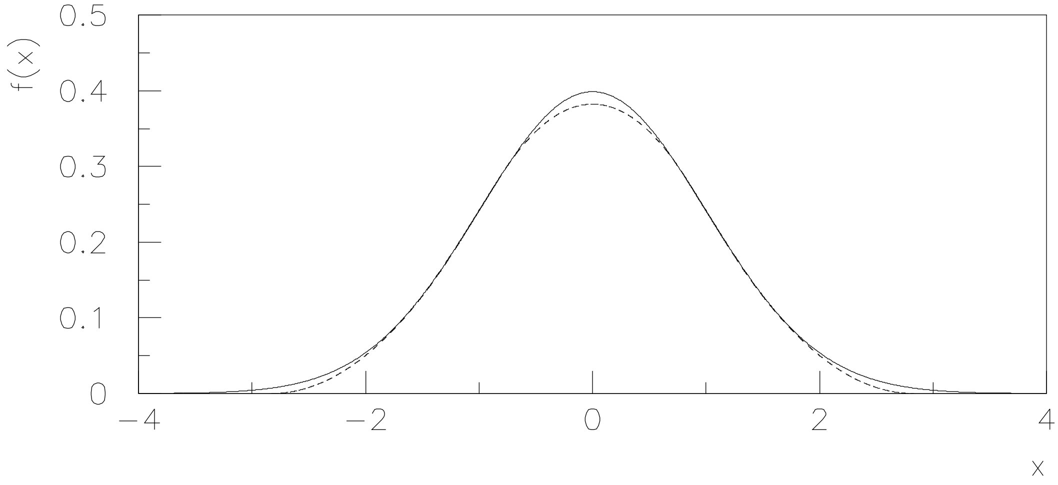


The difference between the standard normal curve and the inscribed parabolic spline function is shown below

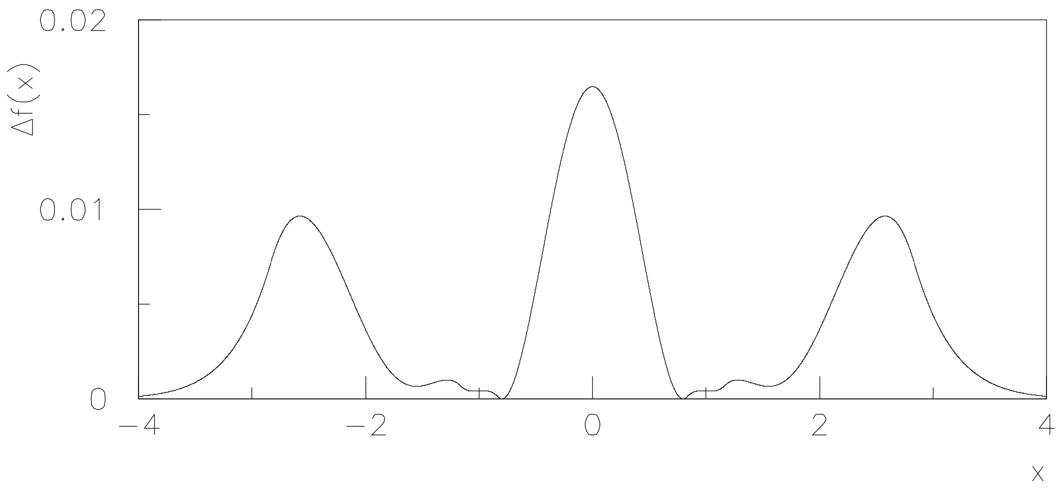


10.6 Parabolic Spline

Leaving a, b and c all free we find a maximum very close to $a = 5/3$, $b = 2$ and $c = 2.015$. In this case we cover 96.74 % of the total area. This particular parabolic spline function belongs to class 3b above. The three knots are at 0.826, 0.841 and 1.174 and the end-point at 2.841. The second part, f_2, is thus used in a very small region between 0.826 and 0.841. By making $b = c = 2$ we still cover 96.62 % of the area with a function from class 2a having one knot less which might be somewhat more convenient. The last case is plotted in the figure below



The difference between the standard normal curve and the inscribed parabolic spline function is shown below



11 Higher Order Spline Functions

Taking the average of N independent uniform variables (from zero to one) we obtain a spline function of the form [5]

$$f_k(\overline{x}) = \frac{N^N}{(N-1)!}\sum_{i=0}^{k}(-1)^i\binom{N}{i}\left(\overline{x}-\frac{i}{N}\right)^{N-1} \quad \text{for} \quad \frac{k}{N} \le \overline{x} \le \frac{k+1}{N} \quad \text{where} \quad k=0,1,\ldots,N-1$$

If we want the distribution in the shifted sum $x = N\overline{x} - \frac{N}{2}$ a simple change of variables yields

$$f_k(x) = \frac{N^{N-1}}{(N-1)!}\sum_{i=0}^{k}(-1)^i\binom{N}{i}\left(\frac{x}{N}+\frac{1}{2}-\frac{i}{N}\right)^{N-1} \quad \text{for} \quad k-\frac{N}{2} \le x \le k+1-\frac{N}{2}$$

where $k = 0, 1, \ldots, N-1$. This function consists of N polynomial arcs of degree $N-1$ which are joined at $N-1$ points (knots) in which the first $N-2$ derivatives are continuous.

This is easily seen since the difference between the ℓ:th derivatives at the knots is given by

$$\frac{d^\ell f_{k+1}(x)}{dx^\ell} - \frac{d^\ell f_k(x)}{dx^\ell} = \frac{N^{N-1-\ell}}{(N-1-\ell)!}(-1)^{k+1}\binom{N}{k+1}\left(\frac{x}{N}+\frac{1}{2}-\frac{k+1}{N}\right)^{N-1-\ell}$$

for $k = 0, 1, \ldots N-2$, which is zero at the knot at $x = k+1-\frac{N}{2}$ for all values of ℓ less or equal to $N-2$. At the end-points the function value as well as the first $N-2$ derivatives are all zero.

According to the Central Limit Theorem the sum, or average, of independent variables from *any* distribution with finite variance approaches a normal distribution as N grows. In our case the standardized variable

$$z_N = \frac{\sum\limits_{i=1}^{N} x_i - \frac{N}{2}}{\sqrt{\frac{N}{12}}}$$

with x_i independent uniform variables between zero and one approaches a standard normal distribution. The distribution in z_N is

$$f_k(z_N) = \frac{\sqrt{\frac{N}{12}}}{(N-1)!}\sum_{i=0}^{k}(-1)^i\binom{N}{i}\left(\sqrt{\frac{N}{12}}z_N+\frac{N}{2}-i\right)^{N-1} \quad \text{for} \quad k-\frac{N}{2} \le \sqrt{\frac{N}{12}}z_N \le k+1-\frac{N}{2}$$

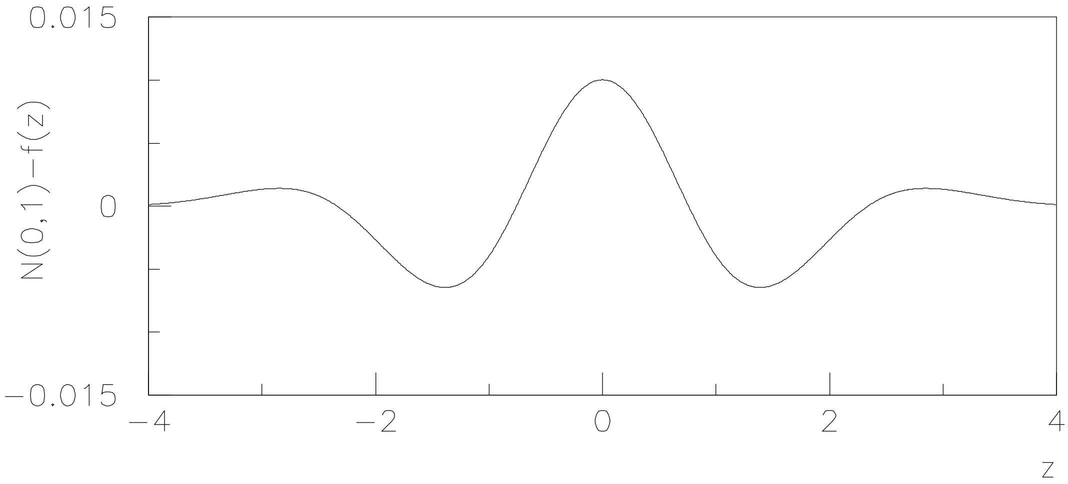
where $k = 0, 1, \ldots, N-1$.

In particular the choice $N = 12$ is convenient since the denominator becomes unity in this case. This is often used as a simple generator of standard normal random numbers. Note, however, that such a generator is neither fast nor exact and much better methods exist [2].

11.1 Sum of Six Uniform Random Numbers

In the previous section we described a method where we made use of the sum of three uniform variables in which case $z_3 = 2\sum_{i=1}^{3} x_i - 3$ corresponding to a parabolic spline with $a = b = c = 2$.

In the figure below we show the difference between the standard normal curve and the spline function resulting from standardizing the sum of six independent variables from a uniform distribution from zero to one.



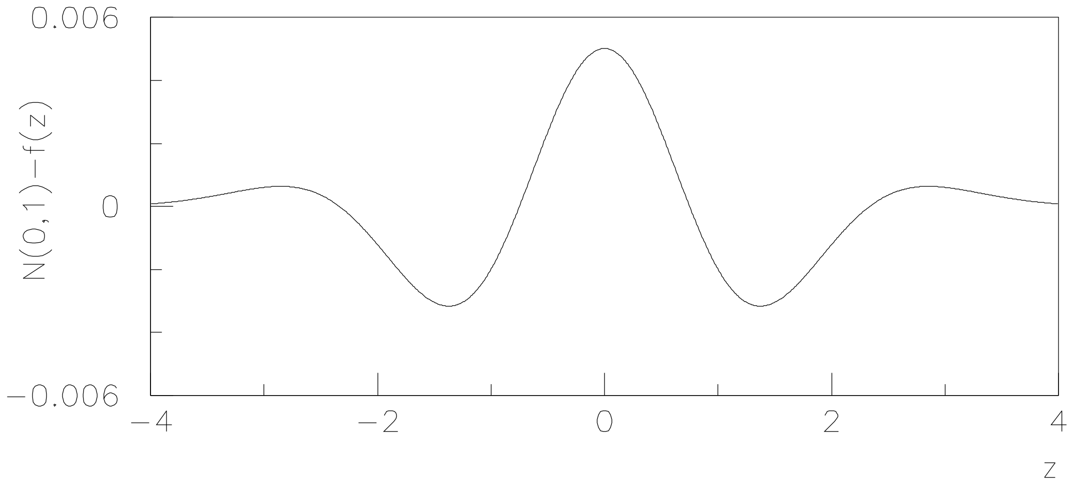
In this case the spline function is given by

$$f(x) = \begin{cases} \frac{11}{20} - \frac{x^2}{2} + \frac{x^4}{4} - \frac{|x|^5}{12} & \text{if} \quad 0 \le |x| \le 1 \\ \frac{17}{40} + \frac{5|x|}{8} - \frac{7x^2}{4} + \frac{5|x|^3}{4} - \frac{3x^4}{8} + \frac{|x|^5}{24} & \text{if} \quad 1 \le |x| \le 2 \\ \frac{81}{40} - \frac{27|x|}{8} + \frac{9x^2}{4} - \frac{3|x|^3}{4} + \frac{x^4}{8} - \frac{|x|^5}{120} & \text{if} \quad 2 \le |x| \le 3 \end{cases}$$

and the maximum deviation from the normal curve is 0.010.

11.2 Sum of Twelve Uniform Random Numbers

If we instead add the classical twelve uniform variables the maximum difference between the standard normal curve and the resulting spline function, shown in the figure below, is 0.005.



As seen in the figures the spline functions are below the standard normal curve centrally, the maximum deviation being at $z = 0$. This is mainly compensated for in the region $1 \leq z_N \leq 2$ where the spline functions are too high. As the number of variables added grow we gradually approach closer and closer to the standard normal curve. Although the spline function seems to be broader centrally the standard normal curve has longer tails. Indeed, the spline function stops at $\pm\frac{N}{2}$ while the standard normal curve in principle extends to $\pm\infty$.

11.3 Polynomial Coefficients

In order to evaluate the polynomial coefficient in $f_k(x)$ we make the following algebra

$$
\begin{aligned}
f_k(x) &= \frac{N^{N-1}}{(N-1)!}\sum_{i=0}^{k}(-1)^i\binom{N}{i}\left(\frac{x}{N}+\frac{1}{2}-\frac{i}{N}\right)^{N-1} = \\
&= \frac{N^{N-1}}{(N-1)!(2N)^{N-1}}\sum_{i=0}^{k}(-1)^i\binom{N}{i}\sum_{j=0}^{N-1}\binom{N-1}{j}(2x)^j(N-2i)^{N-1-j} = \\
&= \frac{1}{(N-1)!2^{N-1}}\sum_{j=0}^{N-1}x^j\binom{N-1}{j}2^j\sum_{i=0}^{k}(-1)^i\binom{N}{i}(N-2i)^{N-1-j} = \\
&= \frac{1}{(N-1)!2^{N-1}}\sum_{j=0}^{N-1}a_jx^j = \sum_{j=0}^{N-1}\alpha_jx^j \quad \text{with} \\
a_j &= \binom{N-1}{j}2^j\sum_{i=0}^{k}(-1)^i\binom{N}{i}(N-2i)^{N-1-j} \quad \text{and} \\
\alpha_j &= \frac{a_j}{(N-1)!2^{N-1}}
\end{aligned}
$$

In this way the coefficients a_j are all integers which is convenient for the calculations (*cf* below for calculations of moments).

11.4 Moments

By the symmetry of the spline function all odd moments vanish and by the construction of z_N the second central moment, the variance, is unity. Higher central moments of even order of the spline functions slowly approaches the asymptotic values from below. For the spline function with $N = 12$ we have $\mu_4 = \frac{29}{10}$, $\mu_6 = \frac{569}{42} \approx 13.55$, $\mu_8 = \frac{2569}{30} \approx 85.63$, $\mu_{10} = \frac{14797}{22} \approx 672.6$ and $\mu_{12} = \frac{567885}{91} \approx 6240.5$.

In order to evaluate the moments we make the following algebra

$$
\begin{aligned}
\mu_{2\ell}(z_N) &= \left(\frac{12}{N}\right)^{\ell} \sum_{k=0}^{N-1} \int_{k-\frac{N}{2}}^{k+1-\frac{N}{2}} x^{2\ell} f_k(x)dx = \\
&= \left(\frac{12}{N}\right)^{\ell} \sum_{k=0}^{N-1} \frac{1}{(N-1)!} \int_{k-\frac{N}{2}}^{k+1-\frac{N}{2}} \sum_{i=0}^{k} (-1)^i \binom{N}{i} \left(x + \frac{N}{2} - i\right)^{N-1} x^{2\ell} dx = \\
&= \left(\frac{12}{N}\right)^{\ell} \frac{1}{(N-1)!} \sum_{k=0}^{N-1} \sum_{i=0}^{k} (-1)^i \binom{N}{i} \sum_{j=0}^{N-1} \binom{N-1}{j} \left(\frac{N}{2} - i\right)^{N-1-j} \int_{k-\frac{N}{2}}^{k+1-\frac{N}{2}} x^{j+2\ell} dx = \\
&= \left(\frac{12}{N}\right)^{\ell} \frac{1}{(N-1)!} \sum_{k=0}^{N-1} \sum_{i=0}^{k} (-1)^i \binom{N}{i} \sum_{j=0}^{N-1} \binom{N-1}{j} \\
&\quad \left(\frac{N}{2} - i\right)^{N-1-j} \left[\frac{\left(k+1-\frac{N}{2}\right)^{2\ell+j+1} - \left(k - \frac{N}{2}\right)^{2\ell+j+1}}{2\ell+j+1}\right] = \\
&= \left(\frac{12}{N}\right)^{\ell} \frac{1}{(N-1)!} \sum_{k=0}^{N-1} \sum_{i=0}^{k} \\
&\quad \sum_{j=0}^{N-1} \frac{(-1)^i \binom{N}{i} \binom{N-1}{j} (N-2i)^{N-1-j} \left[(2k+2-N)^{2\ell+j+1} - (2k+N)^{2\ell+j+1}\right]}{2^{N-1-j}(2\ell+j+1)2^{2\ell+j+1}} = \\
&= \left(\frac{3}{N}\right)^{\ell} \frac{1}{k(N-1)!2^N} \sum_{k=0}^{N-1} \sum_{i=0}^{k} \sum_{j=0}^{N-1} k_j (-1)^i \binom{N}{i} \binom{N-1}{j} \\
&\quad (N-2i)^{N-1-j} \left[(2k+2-N)^{2\ell+j+1} - (2k+N)^{2\ell+j+1}\right] \quad \text{where} \\
k &= (2\ell+1)(2\ell+2)\cdots(2\ell+N) = \frac{(2\ell+N)!}{(2\ell)!} \quad \text{and} \\
k_j &= \frac{k}{2\ell+j+1}
\end{aligned}
$$

Rewritten in this way the $\frac{N^2(N+1)}{2}$ terms summed are all integers. This is convenient if we want to make calculation for large N for two reasons

- The terms have alternating signs and for $N > 40$ even using quadruple real precision we get severe round of errors already for μ_{10}.
- Using rational numbers with big integers involves operations where *e.g.* the cancellation of common factors in the nominator and the denominator is tedious.

The formula above, however, enables us to make calculations for larger values of N utilizing a home-made library for big integer operations. The remaining limitations is that the number of terms summed grow fast with N (se above) and that the integers dealt with becomes increasingly large.

In the table below we show these central moments for different values of N as compared to the asymptotic values for the standard normal distribution which are given by $\mu_{2n} = (2n)!/2^n \cdot n! = (2n-1)!!$.

N	μ_4	μ_6	μ_8	μ_{10}	μ_{12}
1	1.80	3.86	9.0	22	56
2	2.40	7.71	28.8	118	513
3	2.60	9.76	44.7	232	1314
4	2.70	10.93	55.8	331	2187
5	2.76	11.67	63.6	411	2993
6	2.80	12.19	69.3	473	3695
8	2.85	12.86	77.1	565	4807
10	2.88	13.27	82.1	627	5624
12	2.90	13.55	85.6	673	6240
16	2.93	13.90	90.2	733	7102
20	2.94	14.12	93.0	772	7671
24	2.95	14.26	94.9	799	8074
30	2.96	14.41	96.9	826	8496
36	2.97	14.51	98.2	845	8788
48	2.98	14.63	99.9	869	9167
60	2.98	14.70	100.9	884	9402
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
∞	3.00	15.00	105.0	945	10395

In looking at the exact values of the moments given as rational numbers it was found empirically that

$$\mu_{2\ell}(z_N) = \frac{1}{(2\ell+1)!N^{\ell-1}} \sum_{i=0}^{\ell-1} a_i N^i$$

where a_i are integer coefficients. In particular

$$\begin{aligned}
\mu_4 &= \frac{15N-6}{5N} = 3 - \frac{1.2}{N} \\
\mu_6 &= \frac{1}{7N^2}\left(105N^2 - 126N + 48\right) = 15 - \frac{18}{N} + \frac{48}{7N^2} \\
\mu_8 &= \frac{1}{45N^3}\left(4725N^3 - 11340N^2 + 10908N - 3888\right) = 105 - \frac{252}{N} + \frac{1212}{5N^2} - \frac{432}{5N^2} \\
\mu_{10} &= \frac{27}{11N^4}\left(385N^4 - 1540N^3 + 2684N^2 - 2288N + 768\right) = \\
&= 945 - \frac{3780}{N} + \frac{6588}{N^2} - \frac{5616}{N^3} + \frac{20736}{11N^4} \\
\mu_{12} &= 10395 - \frac{62730}{N} + \frac{169884}{N^2} - \frac{1261656}{5N^3} + \frac{6911136}{35N^4} - \frac{28657152}{455N^5} \\
\mu_{14} &= 135135 - \frac{1135134}{N} + \frac{4432428}{N^2} - \frac{49945896}{5N^3} + \frac{66729312}{5N^4} - \frac{48874752}{5N^5} + \frac{2985984}{N^6} \\
&\vdots \\
\mu_{2\ell} &= \mu_{2\ell}^{\text{ideal}} - \mathcal{O}\left(\frac{1}{N}\right) = (2\ell-1)!! - \mathcal{O}\left(\frac{1}{N}\right)
\end{aligned}$$

12 Summary

We have investigated, in some detail, the distribution arising from

- a linear combination of three independent uniform variables, and
- adding N independent uniform variables and standardizing the result.

Applications where random numbers from the standard normal distribution are generated are described.

References

[1] J.H.Ahrens and U.Dieter. *Computer Methods for Sampling from the Exponential and Normal Distributions.* Comm. ACM **15** (1972) 873–882.

[2] Ch. Walck. *Random Number Generation.* USIP Report 87-15, University of Stockholm, December 1987.

[3] Birger Jansson. *Random Number Generators.* Victor Pettersons Bokindustri AB, Stockholm (1966).

[4] G.Marsaglia and T.A.Bray. *A Convenient Method for Generating Normal Variables.* SIAM review **6** (1964) 260–264.

[5] W.T.Eadie, D.Drijard, F.E.James, M.Roos and B.Sadoulet. *Statistical Methods in Experimental Physics.* North-Holland Publishing Company, Amsterdam-London (1971).