An Alternative Model of Probability Theory

by D Williams

Abstract: *An alternative model of probability is outlined and compared with the standard version. Difficulties in extending the Central Limit Theorem are resolved by adopting a new version of population mean developed using the new model and dx-less integrals. New types of sample means are proposed with one tested against the standard mean which appear to have faster convergence. Much remains to be done in this area.*

There are alternative models of geometry (euclidean, non-euclidean), analysis (standard, non-standard), set theory (ZF with/without AC) and so on. So why not probability theory?

An alternative model of probability theory (APT) can be constructed using functions of random numbers between 0 and 1 - (f(ran#) or f(r)) - instead of probability density functions (pr(x)). All the usual results follow … except there are a few minor differences for some distributions at some points.

Firstly we need something to play with.

From a given f(ran#) you can construct a pr(x) using the following transformations:

$$f(ran\#) \overset{replace}{\rightarrow} f(x) \overset{invert}{\rightarrow} cpr(x) = f^{-1}(x) \overset{d/dx}{\rightarrow} pr(x) = \frac{d}{dx} f^{-1}(x)$$

and vice-versa.

With APT, expectations are:

$$E(f(ran\#)) = \int_0^1 f(x)dx$$

Example:

$$E(ran\#^2) = \int_0^1 x^2 dx = \frac{1}{3}x^3\Big|_0^1 = 1/3$$

and

$$E(\frac{1}{2}*x^{-1/2}, 0 < x < 1) = \int_0^1 x*(1/2)*x^{-1/2}dx = \int_0^1 (1/2)*x^{1/2}dx = (1/3)x^3\Big|_0^1 = 1/3$$

For variances:

$$\sigma^2 = Vr(f(ran\#)) = \int_0^1 (f(x) - E(f(x)))^2 dx$$
$$= \int_0^1 (f(x))^2 dx - [E(f(x))]^2$$
$$= \int_0^1 (f(x))^2 dx - (\int_0^1 f(x)dx)^2$$

<u>Example</u>: consider pr(x)=x/2 (0<x<2)

With standard probability theory:

$$E(x) = \int_0^2 x(x/2)dx = x^3/6\Big|_0^2 = 4/3$$
$$E(x^2) = \int_0^2 x^2(x/2)dx = x^4/8\Big|_0^2 = 2$$
$$\therefore \sigma^2 = E(x^2) - [E(x)]^2 = 2 - (4/3)^2 = 2/9$$

With alternative probability theory:

$$pr(x) \Rightarrow f(x) = 2\sqrt{x} \ (for\ 0 < x < 1)$$
$$\sigma^2 = \int_0^1 (f(x))^2 dx - (\int_0^1 f(x)dx)^2$$
$$= \int_0^1 (2\sqrt{x})^2 dx - (\int_0^1 (2\sqrt{x})dx)^2$$
$$= 2x^2\Big|_0^1 - (4/3)x^{3/2}\Big|_0^1 = 2/9$$

There is also the Law of large Numbers (LLN):

<u>Old Prob Theory</u>	<u>New Prob Theory</u>
$\frac{S_n}{n} \to \mu \ as\ n \to \infty$ $for\ S_n = x_1 + x_2 + \ldots + x_n$ $with\ x_i = a\ random\ \text{variable}$	$\frac{\sum_1^n f(ran\#)}{n} \to \int_0^1 f(x)dx\ as\ n \to \infty$ $for\ suitable\ f(x)$

And the Central Limit Theorem (CLT):

<u>Old Prob Theory</u>	<u>New Prob Theory</u>
$\Pr\left(\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} < \beta\right) \to \Phi(\beta)\ as\ n \to \infty$	$\Pr\left(\frac{\frac{\sum f(r)}{n} - \int_0^1 f(x)dx}{\frac{\sigma}{\sqrt{n}}} < \beta\right) \to \Phi(\beta)\ as\ n \to \infty$

So far, so good.

However a minor problem arises if you consider a ***commonly used variant*** of the CLT, namely:

<u>Old Prob Theory</u>	<u>New Prob Theory</u>
$\Pr\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < \beta\right) \to \Phi(\beta)\ as\ n \to \infty$	$\Pr\left(\frac{\sum_1^n f(ran\#_i) - \int_0^1 f(x)}{\sigma\sqrt{n}} < \beta\right) \to \Phi(\beta)?\ as\ n \to \infty$

Here we are estimating the sum of n random variables not their average.

For this, we need a new type of integral – the dx-less integral ∫ f(x) – which, as the name suggests, is the standard integral *without the terminal dx.*

For example,

$$\int_0^1 (1+\ln(x))\mathrm{dx} = 0 \text{ but } \neq \int_0^1 (1+\ln(x)) = \ln(\sqrt{2})$$

For probabilities, it is necessary to use what I call the standard partition of the integral. That is,

$$\int_0^1 f(x) \overset{def}{=} \lim \{f(1/2), f(1/4)+f(3/4), f(1/6)+f(3/6)+f(5/6),...\}$$

$$\int_0^1 (1+\ln(x)) = \lim \{1+\ln(1/2), 1+\ln(1/4)+1+\ln(3/4),...\} = \ln(\sqrt{2})$$

This is necessary because the manner in which the limit of the partial sums is taken can determine the value of the dx-less integral, unlike with standard integrals.

Now, take f(x)=1+ln(x). Compare the variant of CLT under APT with the variant under standard PT with pr(x)= e^(x-1) for -∞<x <1, it's associated pdf.

$$For\ pr(x) = e^{(x-1)} \quad -\infty < x < 1$$
$$\mu = E(x) = 0\ and\ \sigma^2 = 1 \quad thus$$
$$\Pr\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < 0\right)$$
$$= \Pr\left(\frac{S_n - 0}{\sqrt{n}} < 0\right) = 1/2$$
$$but$$
$$for\ f(x) = 1 + \ln(ran\#)$$
$$\Pr\left(\frac{\sum_1^n 1 + \ln(ran\#) - \int_0^1 (1 + \ln(x))}{\sigma\sqrt{n}} < 0\right)$$
$$= \Pr\left(\frac{\sum_1^n 1 + \ln(ran\#) - \ln(\sqrt{2})}{\sigma\sqrt{n}} < 0\right) = 1/2$$

Notice that the centre of the normal curve has shifted from 0 in PT to ln(sqr(2))=0.346573...- a small but important difference.

This forces us to assume either:

i) there are 2 probability theories that can give different answers for some distributions, or
ii) one or the other probability theory requires amendment of some results for some distributions to achieve compatibility

Sadly, I don't know which it is, but I think (nothing more) that option ii) is more likely to be correct. And – don't scream – it is the standard probability theory that "probably" needs amendment.

That is, I'd like to conjecture that the CLT in standard PT should be rewritten as something like:

$$? \quad \Pr\left(\frac{\frac{S_n}{n} - \int_0^1 f(x)dx}{\frac{\sigma}{\sqrt{n}}} < \beta\right) \to \Phi(\beta)\ as\ n \to \infty \quad ?$$
$$for\ suitable\ f(x) \quad (i.e.\ continuous, non-decreasing,\ etc)$$

to conform with the APT version.

The variant form of the CLT naturally follows from this.

$$\Pr\left(\frac{\frac{S_n}{n} - \int_0^1 f(x)}{\frac{\sigma}{\sqrt{n}}} < \beta\right) \to \Phi(\beta) \quad as\ n \to \infty$$

$$\Rightarrow \Pr\left(\frac{S_n - \int_0^1 f(x)}{\sigma\sqrt{n}} < \beta\right) \to \Phi(\beta) \quad as\ n \to \infty$$

With the original CLT you need to add conditions and restrictions that make it less appealing.

My reason for suggesting this comes from playing with stochastic recursive equations like

$$P_{n+1} = e * ran\# * P_n\ with\ P_0 = (say)1000\ and\ e = 2.71828...$$

using spreadsheets and computer programs. Simulations suggest sqr(2)*P_0 more often appears as a better estimate of P_n for large n than P_0 (which is what you'd expect if the standard variant of CLT was true).

But please note I am **not** saying the CLT is wrong. It isn't. It is simply not capable of being adapted to the commonly used variant. The population mean is inadequate to the task. That is why you need dx-less integrals.

As is, the CLT should come with a warning saying something like:

"*Warning: Thou shalt NOT consider the sum of random variables using the CLT without first looking at the appropriate dx-less expression*"

Better still, it may be possible to replace μ with one of a sequence of "sub-means" possessing a smaller variance

$$\Pr\left(\frac{\frac{S_n}{n} - (\sum_{i=1}^{n} f(\frac{2i-1}{2n}))\frac{1}{n})}{\frac{\sigma^*}{\sqrt{n}}} < \beta\right) \to \Phi(\beta) \quad as\ n \to \infty$$

$$with\ \sigma^* \le \sigma \quad (for\ suitably\ configured\ f(x))$$

Giving a variant CLT of:

$$\Pr\left(\frac{S_n - (\sum_{i=1}^{n} f(\frac{2i-1}{2n}))}{\sigma^*\sqrt{n}} < \beta\right) \to \Phi(\beta) \quad as\ n \to \infty$$

APT has other features as well that commend it. It is easier to use w.r.t. spreadsheet and computer simulations and provides an alternative perspective that makes some problems easier to solve and suggests profitable areas to explore.

For instance, the term

$$\frac{(\sum_{i=1}^{n} f(\frac{2i-1}{2n}))}{n}$$

is just the Mid-point Rule approximation of the integral – so why not use other (better) approximations both for the "population mean" and sample mean to get "tighter" estimates?

For example, consider this new way of taking sample means.

Alternative type of sample mean

Take a sample of n data points ($\mathbf{a_i}$), order them ($\mathbf{a_1 \leq a_2 \leq ... \leq a_n}$), assign the following f(x) values ($\mathbf{f(1/2n)=a_1, f(3/2n)=a_2...f(2n-1/2n)=a_n}$) then estimate a curve passing through such points. Then estimate the area of the curve from 0 to 1 to give a different type of sample mean.

For samples of size 1 and 2 (where you approximate a constant and straight line respectively) you just get the regular sample mean.

But for 3 sample points things change.

For n=3, sample then order $a_1 \leq a_2 \leq a_3$ then assign $f(1/6)=a_1$, $f(3/6)=a_2$, $f(5/6)=a_3$

Now apply a quadratic curve ax^2+bx+c through the data points $(1/6,a_1)$, $(3/6,a_2)$, $(5/6,a_3)$ then solve the 3 simultaneous equations

$a(1/6)^2+b(1/6)+c=a_1$
$a(3/6)^2+b(3/6)+c=a_2$
$a(5/6)^2+b(5/6)+c=a_3$

This gives the quadratic equation

$ax^2+bx+c=(9/2)(a_3-2a_2-a_1)x^2+3(3a_2-a_3-2a_1)x+(1/8)(15a_1-10a_2+3a_3)$

Now take the integral of this to give the alternative sample mean of

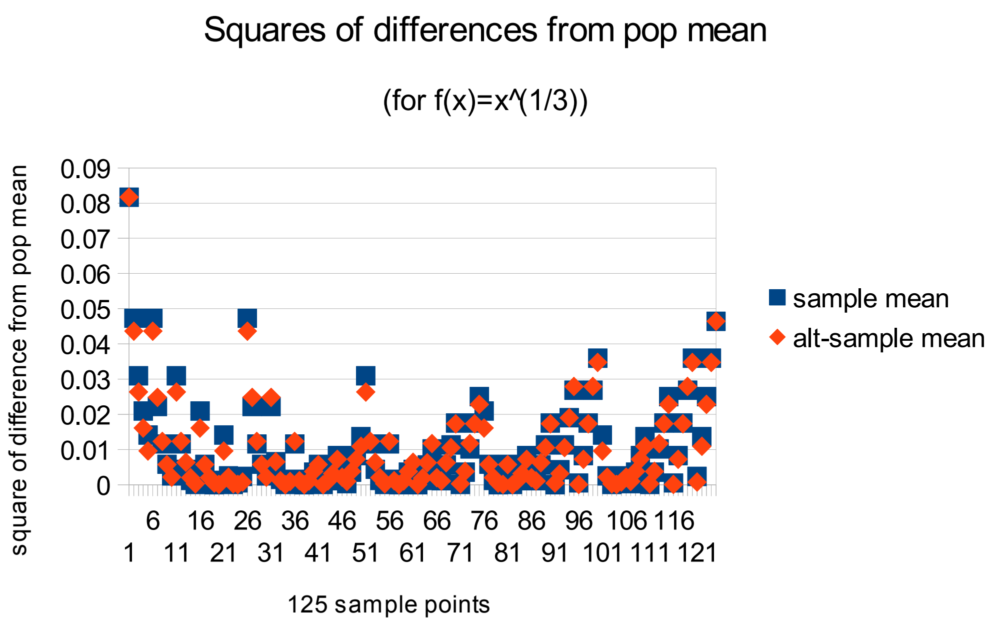
$(3/8)a_1+(1/4)a_2+(3/8)a_3$

To test this I used a spreadsheet with 125 points evenly spread between 0 and 1 with f(x)=x^(1/3) equivalent to $pr(x)=3x^2$ for $0 \leq x \leq 1$. I then compared the standard sample mean with the new one.

Calculating the sum of the squares of the differences from the population mean (here ¾) gave the following values:

Sample mean: 1.29946
Alternative sample mean: 1.24116

Thus, in this example, the values *suggest* that the new sample mean is better at estimating the population mean than the standard sample mean.



<u>Graph</u>: *result of spreadsheet test using 125 evenly spaced points*

In general, for n data points, you need to look at the matrix equation

$$\begin{bmatrix} a \\ b \\ c \\ \ldots \\ k \end{bmatrix} = \begin{bmatrix} \left(\frac{1}{2n}\right)^{n-1} & \left(\frac{1}{2n}\right)^{n-2} & \ldots & \left(\frac{1}{2n}\right) & 1 \\ \left(\frac{3}{2n}\right)^{n-1} & \left(\frac{3}{2n}\right)^{n-2} & \ldots & \left(\frac{3}{2n}\right) & 1 \\ \left(\frac{5}{2n}\right)^{n-1} & \left(\frac{5}{2n}\right)^{n-2} & \ldots & \left(\frac{5}{2n}\right) & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \left(\frac{2n-1}{2n}\right)^{n-1} & \left(\frac{2n-1}{2n}\right)^{n-2} & \ldots & \left(\frac{2n-1}{2n}\right) & 1 \end{bmatrix}^{-1} * \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \ldots \\ a_n \end{bmatrix}$$

Then solve the integral equation of the (n-1)th degree polynomial using the coefficients of a,b,c,...,k.

Whether this difference applies in general or just for some distributions remains to be determined. Also the use of other types of approximate numerical integration needs looking at.

<u>Ways Forward</u>

It would be interesting to know if APT has consequences in the “real world”, in particular quantum mechanics. Can wave functions be physically constructed that display the discrepancy shown above? Could an experiment be devised to determine which probability theory quantum mechanics uses under what circumstances?

APT permits the use of dx-less integrals which have many unusual properties that deserve closer attention. *The problem with standard calculus is that it can erase some of the fine textures by "ignoring smaller terms"*. Dx-less integrals may allow some of these fine textures back.

The question of whether there are other types of probability theory also suggests itself.

A different metric in an alternative (non-euclidean) geometry gives us relativity. What could a different probability theory give us?

D Williams
everythingflows@hotmail.com

Dx-less Integrals

Dx-less integrals are just standard integrals ***without the terminal dx.***

Surprisingly,

i) Such integrals occasionally converge.
ii) Of those that do converge, the result ***can*** differ from the standard integral.
iii) These dx-less integrals have application in estimating the long-term modes of certain stochastic recursive equations (most especially the standard P→(1+b-d)*P of population dynamics).

Below are some examples of dx-less integrals (notice the difference!):

Dx-less Integral	**Standard Integral**
$\int_0^1 (3-2*\ln(2)+2*\ln(x)+\ln(2-x)) = \ln(2)$	$\int_0^1 (3-2*\ln(2)+2*\ln(x)+\ln(2-x))dx = 0$
$\int_0^1 (1+\ln(x)) = \ln(\sqrt{2})$	$\int_0^1 (1+\ln(x))dx = 0$
$\int_0^1 (\frac{\pi^2}{12}+\frac{\ln(x)}{1+x}) = \ln(\sqrt{2})$	$\int_0^1 (\frac{\pi^2}{12}+\frac{\ln(x)}{1+x})dx = 0$
$\int_0^1 (\frac{-\pi^2}{12}+\frac{\ln(1+x)}{x}) = 0$	$\int_0^1 (\frac{-\pi^2}{12}+\frac{\ln(1+x)}{x})dx = 0$
$\int_0^1 (\gamma+\ln\|\ln(x)\|) = \ln(\sqrt{2}), \quad \gamma = 0.5772157...$	$\int_0^1 (\gamma+\ln\|\ln(x)\|)dx = 0$
$\int_0^1 (\ln(2)+\ln(\sin(\pi*x/2))) = \ln(\sqrt{2})$	$\int_0^1 (\ln(2)+\ln(\sin(\pi*x/2)))dx = 0$
$\int_0^1 (\frac{-\pi}{2}+\frac{\ln(1+\cos(\pi*x))}{\cos(\pi*x)}) = \ln(\frac{1}{2})$	$\int_0^1 (\frac{-\pi}{2}+\frac{\ln(1+\cos(\pi*x))}{\cos(\pi*x)})dx = 0$
$\int_0^1 (\frac{\pi^2}{8}+\frac{\ln(x)}{(1-x^2)}) = \ln(\sqrt{2})$	$\int_0^1 (\frac{\pi^2}{8}+\frac{\ln(x)}{(1-x^2)})dx = 0$

Notice the difference between dx-less and standard in ***<u>some</u>*** (not all) of the above.

Below are some calculations based on finite sum approximations (ie: mid-point rule estimates using N subintervals of equal width) that seem to support the above.

$$\int_0^1 (\frac{-\pi}{2}+\frac{\ln(1+\cos(\pi*x))}{\cos(\pi*x)}) = \ln(\frac{1}{2})$$

N (=number of subintervals of width delta x)	Finite Sum Approximation
10	-0.690922805
100	-0.6931245588
200	-0.6931412817
1000	-0.6931361702

$$\int_0^1 \left(\frac{\pi^2}{8} + \frac{\ln(x)}{(1-x^2)}\right) = \ln(\sqrt{2})$$

N (=number of subintervals of width delta x)	Finite Sum Approximation
10	0.3442650586
100	0.3463629748
1000	0.34655228572

(note: ln(sqr(2)) = 0.3465735903….)

$$\int_0^1 (\ln(2) + \ln(\sin(\pi * x/2))) = \ln(\sqrt{2})$$

N (=number of subintervals of width delta x)	Finite Sum Approximation
10	0.346573
100	0.34657359
200	0.346573589

$$\int_0^1 (\gamma + \ln|\ln(x)|) = \ln(\sqrt{2}), \quad \gamma = 0.5772157...$$

N (=number of subintervals of width delta x)	Finite Sum Approximation
10	0.2702834762
100	0.2952827875
500	0.3051016614
2000	0.3110572407

$$\int_0^1 (3 - 2*\ln(2) + 2*\ln(x) + \ln(2-x)) = \ln(2)$$

N (=number of subintervals of width delta x)	Finite Sum Approximation
10	0.6868999
100	0.69252218
1000	0.6930847566

(note: ln(2)= 0.6931471806…)

$$\int_0^1 (1 + \ln(x)) = \ln(\sqrt{2})$$

N (=number of subintervals of width delta x)	Finite Sum Approximation
10	0.3424093466
100	0.3461569271
1000	0.3465319952

What is needed is a test to determine which dx-less integrals i) converge to the same value as the normal integral ii) converge to a different value or iii) diverge. Any suggestions?

All comments to: everythingflows@hotmail.com

Here are some partial product approximations of some dx-less ***product*** integrals. To obtain the corresponding dx-less integrals just take the ln of the product to obtain the partial sum approximations.

<u>Some finite product approximations of certain f(x)</u>

The simple BASIC program

```
label start
e=exp(1)
input n
p=1
d=2*n
for i=1 to n
x=(2*i-1)/d
p=e*x*p
next i
print p
goto start
```

can be used to calculate partial **product** approximations for f(x)=e*x, giving the output (for various input n) of:

n=	p=
1	1.3591409...
10	1.40873667...
100	1.4136244...
1000	1.4141544...
10,000	1.414207669...

(I used the free Small Basic program available on the web for these calculations)

For f(x)=(e/4)*(x+1), replace “p=e*x*p” in the above program with “p=(e/4)*(x+1)*p” to get output of:

n=	p=
1	1.019355...
10	1.002083381...
100	1.000208352...
1000	1.0000208335...
10,000	1.0000020833...

For f(x)=(4e/27)*(x+2), make the appropriate swap to get output of:

n=	p=
1	1.006771...
10	1.0006944...
100	1.00006944...
1000	1.000006944...
10,000	1.0000006944...

For f(x)=(16/27)*(x+2)/(x+1), make the appropriate swap to get output of:

n=	p=
1	0.987654...
10	0.99861397...
100	0.99986112...
1000	0.999986111...
10,000	0.9999986111...

For

$$\frac{\prod_{0}^{1/3}(x+1)}{\sqrt{\prod_{1/3}^{2/3}(x+2/3)\prod_{2/3}^{1}(x+1/3)}} = 0.866...?$$

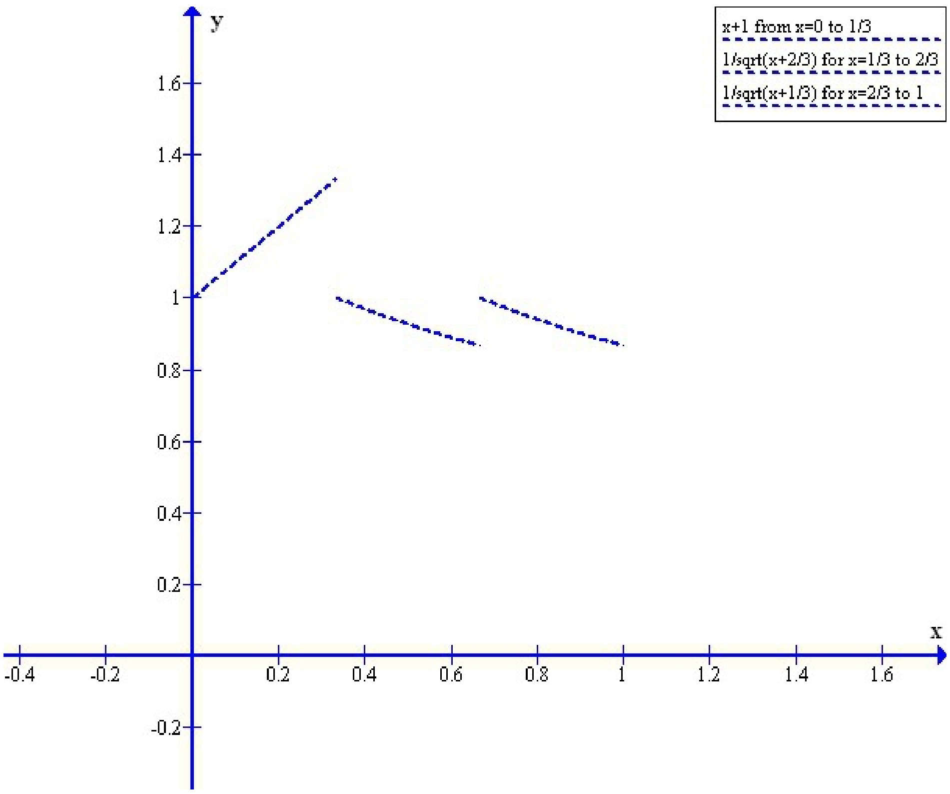
Use

```
label start
input n
p=1
d=2*n
for i=1 to n
x=(2*i-1)/d
```

```
if x<1/3 then p=(x+1)*p
if x>1/3 and x<2/3 then p=(1/sqr(x+2/3))*p
if x>2/3 then p=(1/sqr(x+1/3))*p
next i
print p
goto start
```

Giving output of

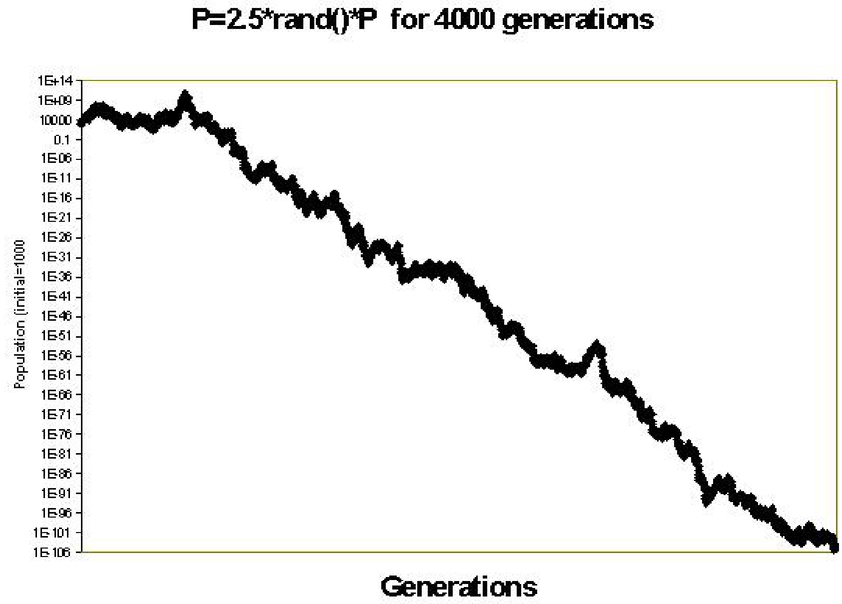
n=	p=
1	0.925320...
10	0.8726...
100	0.86668...
1000	0.86609155...
10,000	0.8660320...



Graph of function above

And so on (but what the hell is 0.866...?).

It is also relatively easy to make approximations using **spreadsheets** which can then be graphed. The ***rand()*** function can also be used to make stochastic products and series as well. Comparisons can then be made between the stochastic output and the dx-less estimators.



Graph of Recursive Product via spreadsheet

Here are some simple dx-less product integrals to start playing with:

$$\prod_0^1 \frac{f(x)}{f(1-x)} = 1$$

$$\prod_0^1 e^{\alpha(x-1/2)} = 1 \; for \; \alpha \in \mathbb{R}$$

$$\prod_0^{1/2} f(x) \prod_{1/2}^1 1/f(1-x) = 1$$

$$\prod_0^{1/3} f(x) \sqrt{\prod_{1/3}^{2/3} 1/f(x-1/3) \prod_{1/3}^{2/3} 1/f(1-x)} = 1$$

$$\prod_0^1 e * (\frac{a^a}{(a+1)^{a+1}})(x+a) = 1 \; for \; a \in \mathbb{R}_{>0}$$

... and so on

And some more general dx-less product integrals (none proved, just significant numerical evidence for):

$$\prod_0^1 (\sqrt{e} * x)^x = 1$$

$$\prod_0^1 e^{-\pi^2/12} * (1+x)^{1/x} = 1$$

$$\prod_0^1 2 * (\sin(\pi x/2)) = \sqrt{2}$$

$$\prod_0^1 e^{\gamma} \ln(1/x) = \sqrt{2} \quad \textit{where } \gamma = 0.5772...$$

$$\prod_0^1 \sqrt{2} * (\sin(\pi x))^x = \sqrt{2}$$

$$\prod_0^1 e^{-\pi/2} * (1+\cos(\pi x))^{1/(\cos(\pi x))} = 1/2$$

$$\prod_0^1 x\left[(2-x)\wedge(12/\pi^2(1-x)\right] = \sqrt{2}$$

$$\prod_0^1 (\frac{e^{\sqrt{2}}}{1+\sqrt{2}})x(x+\sqrt{x^2+1}) = \sqrt{2}$$

$$\prod_0^1 2\sin(\pi x) = 2$$

$$\prod_0^1 (x + 0.54331...) = 1$$

$$\prod_0^1 (2x + 0.17696...) = 1$$

$$\prod_0^1 \frac{2}{3} * e \wedge (\frac{x^2 - x}{\ln(x)}) = 1$$

$$\prod_0^1 x * e \wedge (\sin^2(2\pi x)) = \sqrt{2}$$

– where all the above products are over Q(odd/even) – that is, all rationals with odd numerators and even denominators.

There is so much to do.