

Transitioning from Discrete to Continuous Distribution *Mathematica* vs. Excel

—An Example

Haiduke Sarafian

The Pennsylvania State University, University College, York, USA

Email: has2@psu.edu

How to cite this paper: Sarafian, H. (2022) Transitioning from Discrete to Continuous Distribution *Mathematica* vs. Excel. *American Journal of Computational Mathematics*, 12, 25-32.

<https://doi.org/10.4236/ajcm.2022.121003>

Received: January 18, 2022

Accepted: March 5, 2022

Published: March 8, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Frequencies of the repeated integers of the first n digits of e.g. π utilizing commercial software are listed. The discrete distribution is utilized to evaluate its statistical moments. The distribution is fitted with a polynomial generating a continuous replica of the former. Its statistical moments are evaluated and compared to the former. The procedure clarifies the mechanism transiting from discrete to a continuous domain. Applying *Mathematica* the fitted polynomial is replaced with an interpolated function with controlled smoothing factor refining the quality of the fit and its corresponding moments. Knowledge learned assists in the understanding of the standard procedure calculating moments of e.g. Maxwell-Boltzmann continuous distribution in kinetic Theory of gases.

Keywords

Digits of π , Discrete Distribution, Continuous Distribution, Moments of Distribution, Excel, *Mathematica*

1. Introduction

Tabulating the statistical information such as distribution moments for either the discrete and/or continuous ensembles is the quantities of paramount interest when working with either abstract mathematical or collected data in natural sciences. It is somewhat trivial to evaluate the moments for a discrete data set, it is not obvious how to systematically transit from the discrete to the continuous situation.

One of the objectives of this report is that by way of example first to show how the moments are evaluated for a set of the discrete mathematical ensemble, and

then by applying the same conceptual method to extend the procedure for the continuous case.

To achieve this goal, we select one of the ~32,000 known constants in science e.g., the value of π . The shown procedure identically may be applied to any of the chosen constants. For an instant e , the Euler constant γ , the golden ratio ϕ , etc. Here in this report, we have chosen the π . We form an ensemble comprised of n digits of π ; naturally, this is a set of discrete integers. We then show how the statistical moments of the set are evaluated. Taking advantage of the commercially available software, e.g., Excel [1] we tally the data conducive to the needed distribution function. Having the distribution function on hand we evaluate the moments, such as the first, second, third, etc. Excel is an excellent numeric-based program with certain limitations. For instance, because it is a single-precision compiler it displays the digits of π up to 16 significant figures. As such it limits the number of elements of π_List . To circle this, one may use a commercially available scientific software e.g., *Mathematica* [2]. This allows extending the number of the digits of π_List literally to “infinite.” To transit from discrete to continuous and hence to evaluate the moments we form the extended π_List say with 50 elements. This list then is imported to Excel and is used as a basis to form the continuous distribution function by fitting it using a polynomial. Here again, Excel is limited to a maximum 6th order polynomial. For the sake of consistency when we utilize the *Mathematica*, we apply the same polynomial power; this results in the identical result. However, *Mathematica* has a useful option smoothening the quality of the fit. Utilizing this option, we perfect the fit. We include tables embodying the values of the calculated moments for all the scenarios.

This report is comprised of four sections. In addition to Section 1, introduction that outlines the motivation and goals, Section 2 is procedure; a description that embodies *Mathematica* codes, charts, tables as well as selected Excel’s charts. The interested reader may easily duplicate the steps and modify the codes adjusting to the need, for information c.f. [3] [4]. Section 3 is the conclusions and comments on what we learned.

2. Procedure

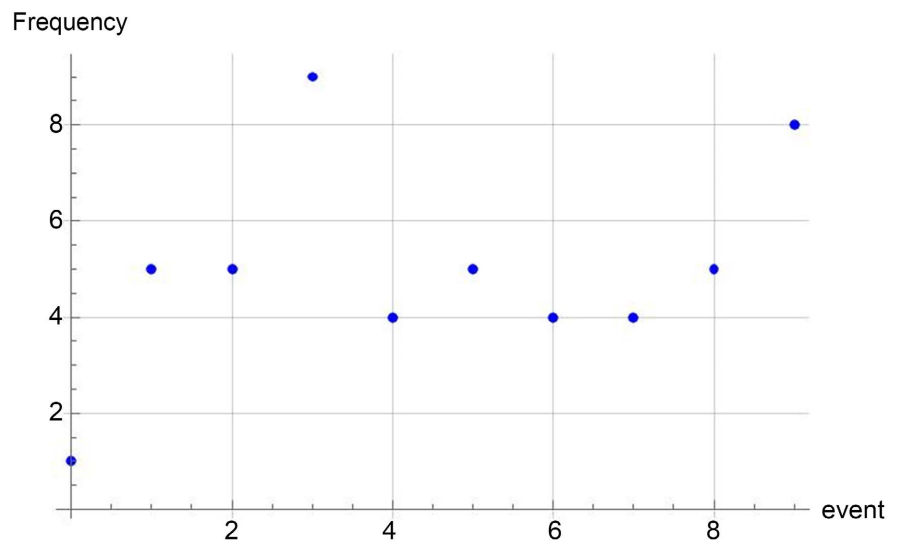
For the sake of efficiency, we begin with *Mathematica*, as such first we form the π_List , a list of digits of π . Nmax defines the number of desired significant digits, e.g., 50. Shown program is crafted such that with this input parameter one single keystroke runs the entire program with the needed output.

```
Nmax=50;
pi=First[RealDigits[N[ $\pi$ ,Nmax]]];
Next, we tabulate the tallied digits, (see Table 1)
table=TableForm[Tally[pi]/.{p_,q_}->{q,p},TableHeadings->{Automatic,
{"Frequency","digit/Event"}}]
```

By defining a few auxiliary components, we display the Frequency vs. the Range. This is shown in Figure 1.

Table 1. Frequencies of the first 50 digits of Pi vs. the digits.

Frequency	Digits
9	3
5	1
4	4
5	5
8	9
5	2
4	6
5	8
4	7
1	0

**Figure 1.** Display of frequencies vs. ranges *i.e.*, the event.

```
lengthtallypi=Length[Tally[pi]];
data50=Transpose[{Table[{table[[1,n,2]],table[[1,n,1]]},{n,1,lengthtallypi}]];
listplotdata50=ListPlot[data50,AxesOrigin->{0,0},PlotStyle->Blue,AxesLabel->{
"event", "Frequency"},GridLines->Automatic]
```

To check the integrity of the normalized distribution we form,

```
Frequencies=Table[table[[1,n,1]],{n,1,lengthtallypi}];
```

```
events=Table[table[[1,n,2]],{n,1,lengthtallypi}];
```

```
frequencies=N[1/Nmax Frequencies];
```

```
Apply[Plus,frequencies]
```

1.

This shows the output of this step is a 1. Assuring the distribution is normalized to unity.

Note also, Apply is a standard *Mathematica* command. In this particular-code, it applies the Plus function *adding* the elements of the listed frequencies.

Next, we evaluate the mean and the RMS of the distribution and compare them to the output of *Mathematica*'s library functions,

```
{discreteaverage,mean}={Apply[Plus,frequencies events],N[Mean[pi]]};
{discreteRMS,RMS}={sqrt{Apply[Plus,frequencies
events^2]],N[RootMeanSquare[pi]]};
{4.939,4.94}
{5.662,5.662}
```

As shown, these are identical. These steps ensure the accuracy of our program conducive to laying the bases for transiting to the continuous scenario. We also evaluate the 3rd moment making the point that evaluating the nth order moment is no challenge.

```
discretethirdmoment=cuberoor{Apply[Plus,frequenciesevents^3]};
```

The summary of the calculation is shown in **Table 2**.

To transit to the continuous domain and be compatible with the capabilities of Excel we consider a 5th order polynomial for the model. Note that by trial and error the 5th and the 6th order proven to be indistinguishable.

```
model=a+b x+c x^2+d x^3+e x^4+f x^5;
Flatten[data50,{2}][[1]];
fit1=FindFit[Flatten[data50,{2}][[1]],model,{a,b,c,d,e,f},x];
fit1x=model/.fit1
Show[{listplotdata50,ListPlot[Table[{x,fit1x},{x,0,9}],PlotStyle->Red]]}
1.03357 +4.47364 x-0.845862 x^2-0.0632867 x^3+0.02162 x^4-0.00102564 x^5
```

Noting the numeric coefficients of the fitted polynomials utilized *Mathematica* and Excel are the same.

Figure 2 and **Figure 3** show the fitted polynomials have the correct trend, however, they aren't to satisfaction. Taking these polynomials as their face value we evaluate their corresponding statistical moments.

To do so and fulfill one of our aimed objectives *i.e.*, transiting from discrete to continuous we take the following steps: since normalized discrete distribution is

subject to $\frac{1}{N_{\max}} \sum_{i=1} F_i$, where F_i is the number of events we replace $\frac{1}{N_{\max}} \rightarrow \frac{\Delta x}{9}$ and $F_i \rightarrow F(x)$ where $F(x)$ is the fitted polynomial. With these substitutions, the normalization condition reads,

$$\frac{1}{N_{\max}} \sum_{i=1} F_i \rightarrow \frac{1}{9} \int_0^9 F(x) dx, \quad (1)$$

And because the fitted polynomial or any fitted modeled function is not normalized by multiplying the model function with a constant we enforce the normalization so,

$$\text{normalized Distribution} = \alpha = 9 \left(\int_0^9 \text{fit1} x dx \right)^{-1};$$

Table 2. First three moments associated with the first 50 digits of Pi.

Average	4.94
RMS	5.66
3 rd moment	6.17

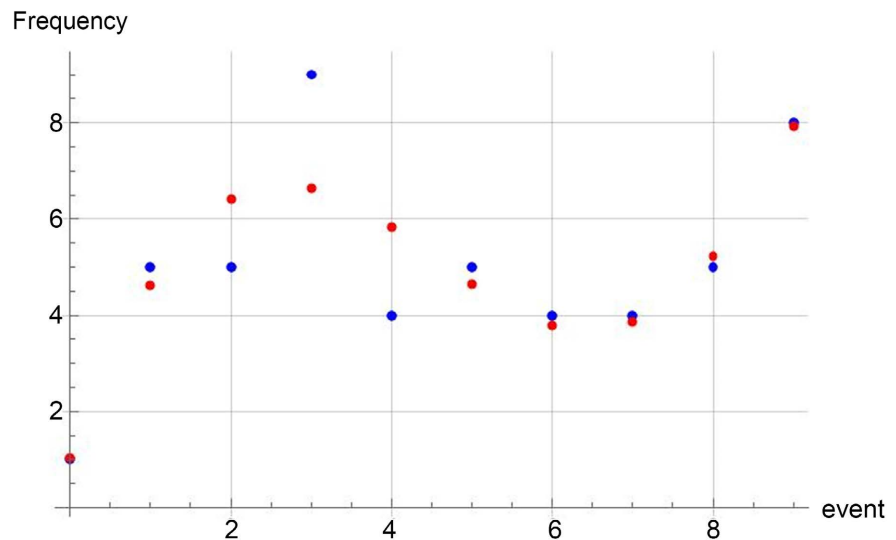


Figure 2. The blue dots are the data shown in **Figure 1**, the red dots are the values of the Frequencies applied to the polynomial, *i.e.*, our model.

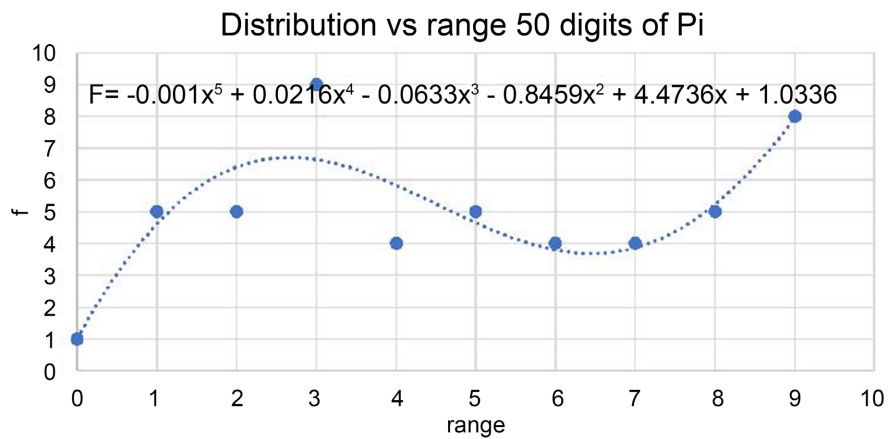


Figure 3. Excel fitted curve with the explicitly employed polynomial.

Now by applying the normalized distribution we evaluate the moments,

$$\text{continuous Average} = 1/9 \int_0^9 x (\alpha \text{fit} 1x) dx ;$$

$$\text{rms Continuous} = \sqrt{\int_0^9 x^2 \frac{1}{9} (\alpha \text{fit} 1x) dx} ;$$

$$\text{third Moment} = \sqrt[3]{\int_0^9 x^3 \frac{1}{9} (\alpha \text{fit} 1x) dx} ;$$

The summary of the output is tabulated in **Table 3**.

The tabulated values of **Table 2** and **Table 3** reveal the differences between the discrete and its corresponding continuous distributions.

As we pointed out **Figure 2** and/or **Figure 3** are not quite to our satisfaction. If we were using Excel as an ultimate tool this would have been our best fit with its accompanied evaluated moments given in **Table 3**. However, *Mathematica* offers a procedure improving the fit quality. With *Mathematica* the discrete limited data are shown with the blue dots in **Figure 1** may be interpolated generating unlimited implicit data making the fit procedure much satisfactory. After trial and error, we noted Interpolation Order of degree 4 works the best. **Figure 4** is the result.

As shown interpolated function fits exactly the data, dots shown on the right plate are doubly overlapped dots. Meaning the smoothened data is exactly overlaps with discrete ones. The green continuous curve is the 5th order polynomial that is smoothened by *Mathematica*. Having such a perfect fit *i.e.*, continuous distribution, we calculate its moments. Beforehand the green curve ought to be normalized,

$$\beta = N \left[9 \left(\int_0^9 \text{interpolate}[x] dx \right)^{-1} \right];$$

And the moments are,

```
averageInterpolatedx=1/9 NIntegrate[x (β interpolate[x]),{x,0,9}];
rmsInterpolatedx=sqrt{1/9 NIntegrate[x^2 β interpolate[x],{x,0,9}]};
thirdmomentInterpolatedx=cuberoor{1/9 NIntegrate[x^3 β interpolate[x],{x,0,9}}];
```

Table 3. Average, RMS and the third moment of the continuous distribution of the first 50 digits of Pi.

Average	4.56
RMS	5.23
3 rd moment	5.72

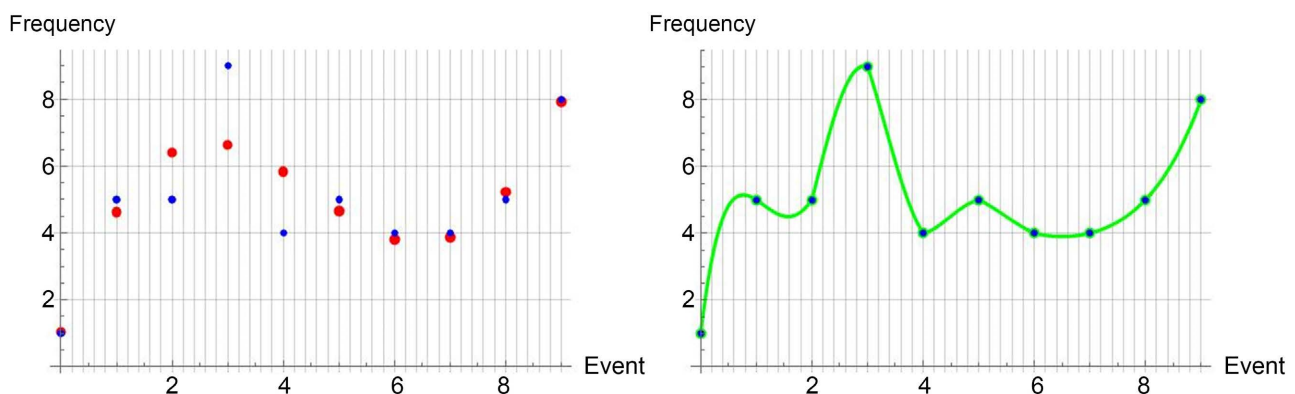


Figure 4. The left plate is the same as **Figure 2**, the right plate is the Interpolated fit including the smoothening factor.

Table 4. Summary of the first three moments associated with the three scenarios. Description of each case is embedded in the text.

	Average	RMS	3 rd moment
Discrete	4.94	5.66	6.17
Polynomial	4.56	5.23	5.72
Interpolated	4.47	5.16	5.66

Table 4 is the calculated moments for three scenarios in this report.

Table 4 embodies the values of the body of the work presented in this report. It compares the values of various moments of discrete, continuous, and improved continuous distributions. To form an opinion about the quality of these moments and the associated quality of distribution functions one needs to have **Figure 4** in mind.

3. Conclusions

We set two aims crafting this investigation. For practical purposes by a way of example we show how the common knowledge evaluating statistical moments of a discrete distribution is extended by evaluating similar moments for a continuous distribution. Steps shown in this report transiting between these two sets of distributions fill the missing gap that is overlooked in the literature. Our approach identified the shortcoming of a useful program such as Excel justifying looking for a replacement such as the powerful scientific program *Mathematica*. For a manageable list of a discrete list of integer numbers, we set the list length to 50, as mentioned *Mathematica* contrary to Excel can extend the list length literally to infinite. Specific of the given example is applied for π , identical steps may be taken utilizing e.g., the e , Euler gamma γ , etc. and combinations amongst e.g., π^e , $e^{\sqrt{\pi}}$ so there is no limitation replicating examples.

The lesson learned is that with a solid understanding of steps needed to calculate the moments of a continuous distribution, calculation conducive to the moments for distributions encounter in physical science, e.g., speed distribution given by Maxwell-Boltzmann [5], or probability distribution for quantum system easily may be duplicated as well.

Acknowledgements

The author acknowledges the John T. and Page S. Smith Professorship funds for completing and publishing this work.

The author appreciates the referee's comment about the reference [6]. While read proofing the final version of the article [6] is added to the list.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] MicroSoft®, Excel. <http://www.microsoft.com>
- [2] *Mathematica*® V13.0. <http://Wolfram.com>
- [3] Wolfram, S. (2003) *Mathematica Book*. 5th Edition, Cambridge University Press, New York, NY.
- [4] Sarafian, H. (2019) *Mathematica Graphics Examples*. 2nd Edition, Scientific Research Publishing, Wuhan.
- [5] Longair, M.S. (1984) *Theoretical Concepts in Physics*. Cambridge University Press, New York, NY.
- [6] Lin, M. (2018) Techniques of Discrete Function Transfers into Continuous Function in Practice. *Engineering*, **10**, 680-687.
<https://doi.org/10.4236/eng.2018.1010049>