The Continuum Hypothesis from A Continuous Uniform Distribution Analysis

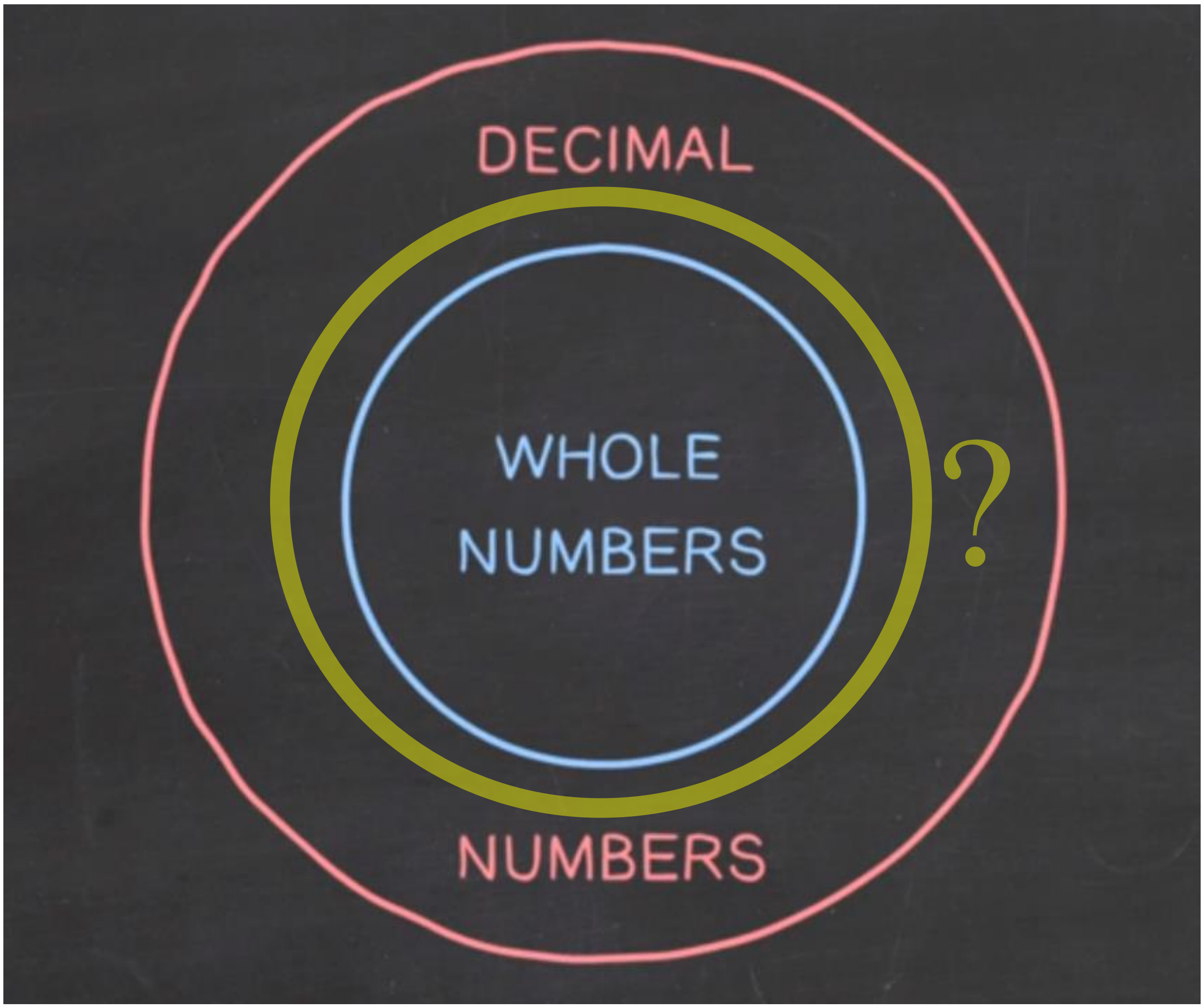
By

Fred Viole
OVVO Financial Systems
fred.viole@ovvofinancialsystems.com

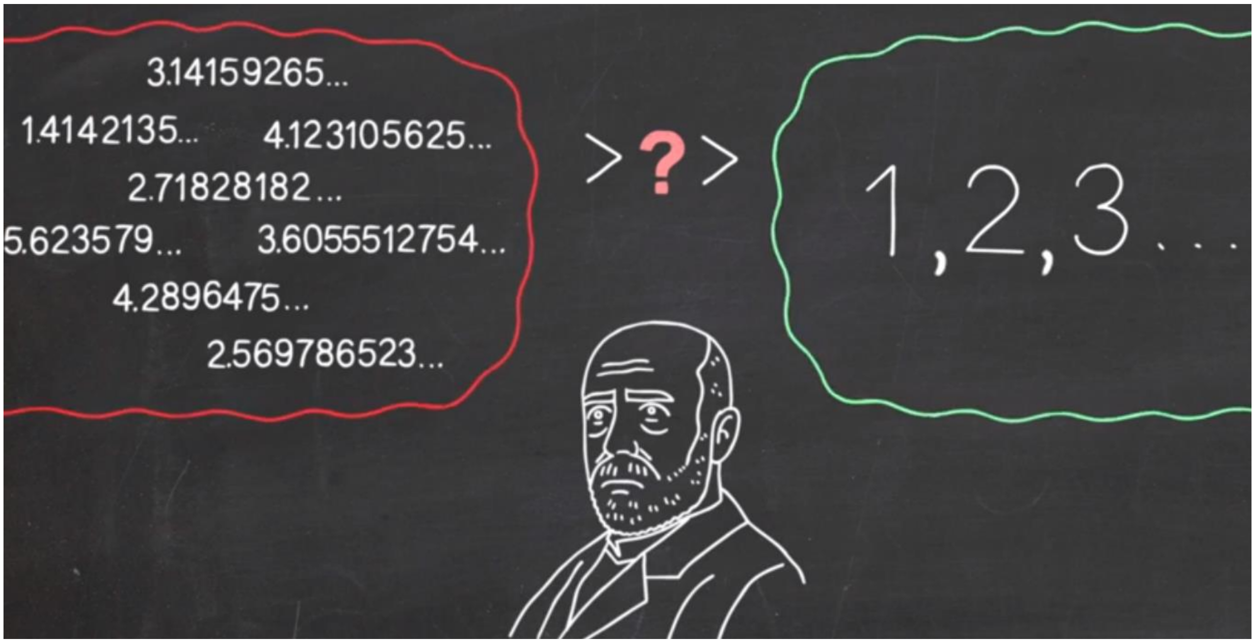
INTRODUCTION:

In a recent TED-Ed video by Dennis Wildfogel "How Big is Infinity?"[1], the continuum hypothesis is described and presented with graphical representation below.

> "We just pointed out that the set of decimal numbers, that is the real numbers, is a bigger infinity than the set of whole numbers. Cantor wondered whether there are infinities of different sizes between these two infinities. He didn't believe there were but couldn't prove it."



1 http://ed.ted.com/lessons/how-big-is-infinity



More formally stated,

> *There is no set whose cardinality is strictly between that of the integers and that of the real numbers.*

Cantor gave two proofs that the cardinality of the set of integers is strictly smaller than that of the set of real numbers. His proofs, however, give no indication of the extent to which the cardinality of the integers is less than that of the real numbers. Cantor proposed the continuum hypothesis as a possible solution to this question.[2]

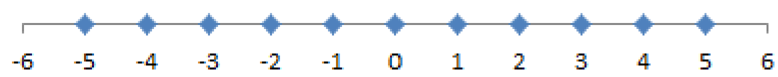
We take a different tact to proving the extent to which cardinality of the integers is less than that of the real numbers using the following assumptions:

- Set = Interval
- Rational Numbers = Discrete distribution
- Real Numbers = Continuous distribution

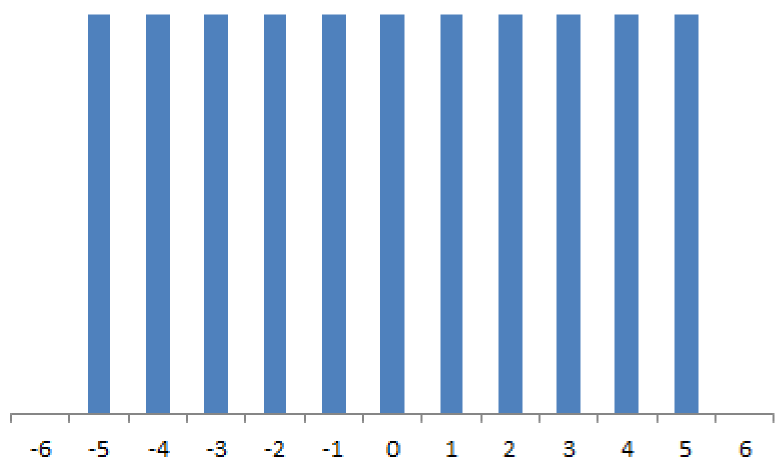
[2] http://en.wikipedia.org/wiki/Continuum_hypothesis

<u>OUR METHOD:</u>

Number lines do not offer an area consideration as the area of each point equals zero.[3]



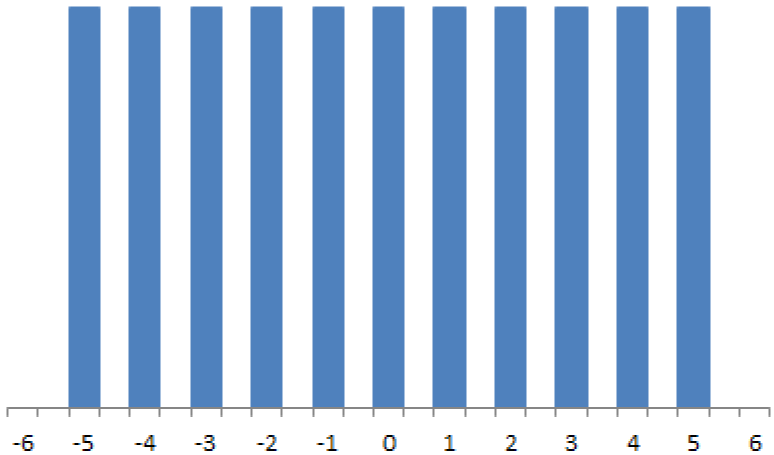
However, a vertical shift to a number line as is the case with a uniform distribution, does lend itself to area analysis.



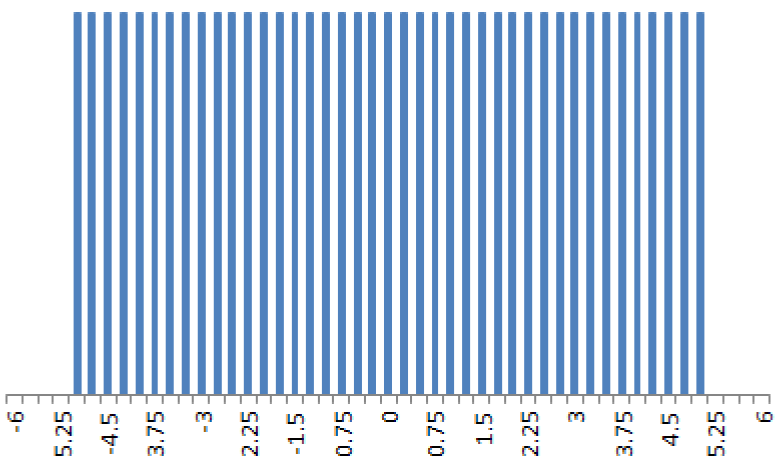
The uniform distribution can also be one of two varieties: discrete and continuous. It is in this comparison we find Cantor's observation to hold, namely the area of the continuous uniform distribution is larger than that of the discrete uniform distribution. The images below will illustrate.

[3] We know from calculus that $\int_a^b(x)dx = F(b) - F(a)$ and if $F(b) = F(a)$, the integral of a point equals zero.

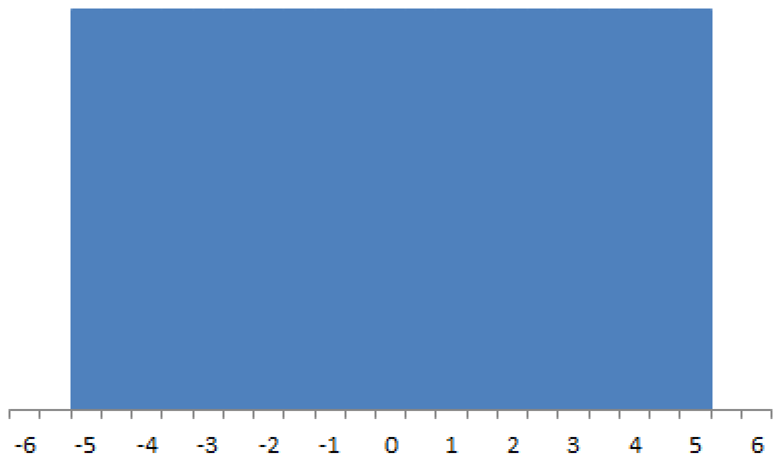
A discrete uniform distribution (note the gaps between the bars),



There are an infinite amount of points (rational numbers) we could add to this interval, however, there are then an infinite amount of areas (real numbers) between those bars.



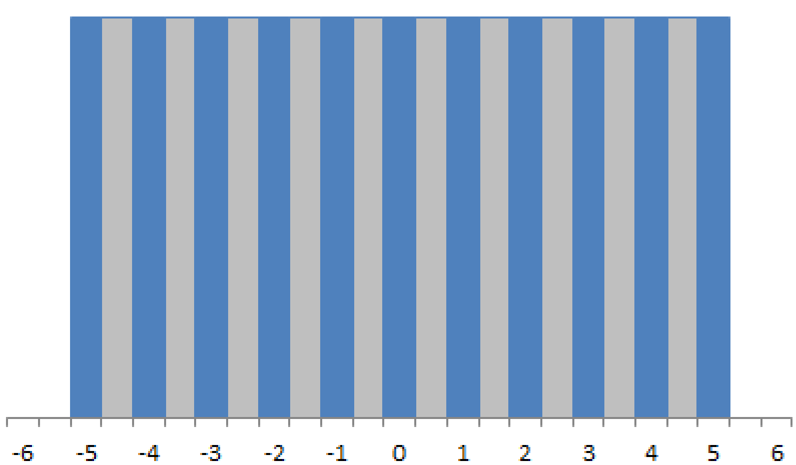
A truly continuous uniform distribution (note the lack of gaps),



Unfortunately, truly continuous distributions cannot be constructed, since every analysis is that of the discrete variety.

> "*All actual sample spaces are discrete, and all observable random variables have discrete distributions. The continuous distribution is a mathematical construction, suitable for mathematical treatment, but not practically observable*." E.J.G. Pitman (1979).

However, Viole and Nawrocki [2012] propose a method to capture the "area between the bins" **for any distribution family, uniform included**. The "area between the bins" (regardless of number of discrete observations [rational numbers]) in grey,



CUMULATIVE DISTRIBUTION FUNCTIONS:

One method to measure these both the discrete and continuous areas is through cumulative distribution function ratios. When comparing cumulative distribution functions, we need to distinguish between discrete and continuous. For the uniform distribution in the IMSL Fortran Numerical Stat library, the CDF is represented by:

$$\mathrm{F}(x|A,B) = \begin{cases} 0, & if\ x < A \\ \dfrac{x-A}{B-A}, & if\ A \le x \le B \\ 1, & if\ x > B \end{cases}$$

Viole and Nawrocki [2012] demonstrate that this is identical to a *discrete* degree zero lower partial moment represented by the equation:

$$LPM(n,h,x) = \frac{1}{T}\left[\sum_{t=1}^{T} max\{0, h - x_t\}^n\right] \qquad (1)$$

And the partial moment CDF is represented by

$$\frac{LPM(0,h,X)}{[LPM(0,h,X) + UPM(0,l,X)]} - \frac{\varepsilon}{2} \qquad (2)^4$$

where x_t represents the observation *x* at time *t*, *n* is the degree of the LPM, *q* is the degree of the UPM, *h* is the target for computing below target returns, ε is the point probability, and *l* is the target for computing above target returns.[5]

For a Uniform Distribution $\mu = 10.00045$ consisting of 5 million observations with 300 iteration seeds, the discrete vs. continuous CDF comparison has a consistent relationship. The continuous area (real numbers included) > the discrete area (rational numbers) as represented in table 1 from Viole and Nawrocki [2012].

---

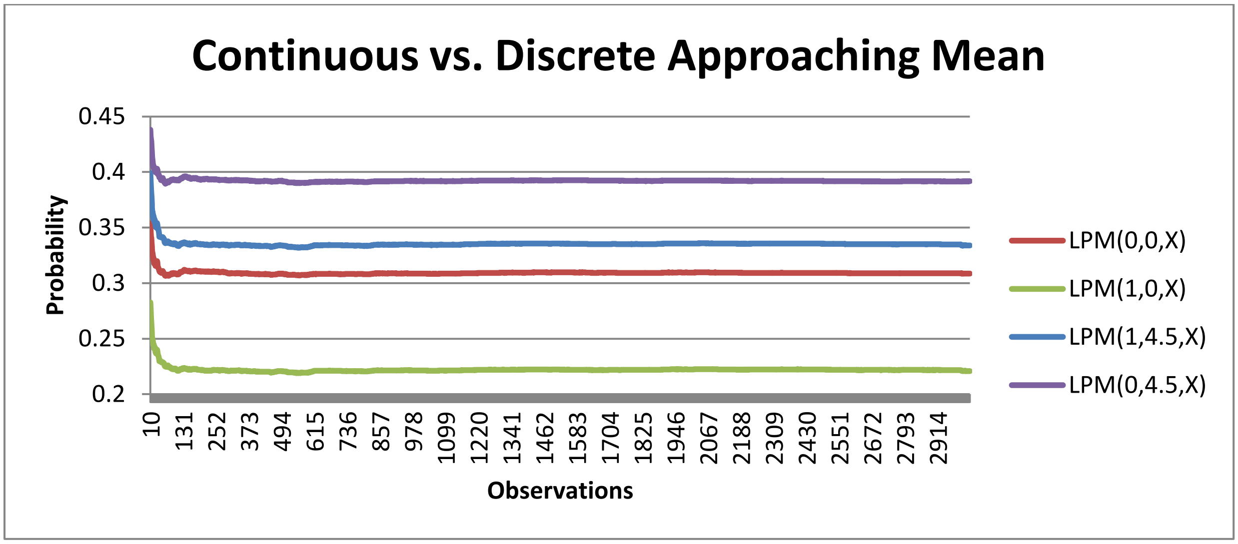
[4] It is important to note that $LPM(0,x,X)$ is a probability measure and will yield a result from 0 to 1. Thus, the ratio of $LPM(0,x,X)$ to the entire distribution ($LPM_{ratio}(0,x,X)$) is equal to the probability measure itself, $LPM(0,x,X)$.

[5] Equations 1 and 2 will generate a 0 for degree 0 instances of 0 results.

Uniform Distribution Probabilities - 5 Million Draws 300 Iteration Seeds		
UNDF(X ≤ 0.00) = .4	LPM(0, 0, X) = .4	LPM(1, 0, X) = .3077
UNDF(X ≤ 4.50) = .445	LPM(0, 4.5, X) = .445	LPM(1, 4.5, X) = .3913
UNDF(X ≤ Mean) = .5	LPM(0, μ, X) = .5	LPM(1, μ, X) = .5
UNDF(X ≤ 13.5) = .535	LPM(0, 13.5, X) = .535	LPM(1, 13.5, X) = .5697

Table 1. Uniform distribution results illustrate convergence of $LPM(0, x, X)$ to UNDF and consistent relationship between $LPM(0, x, X)$ and $LPM_{ratio}(1, x, X)$ above and below the mean target.

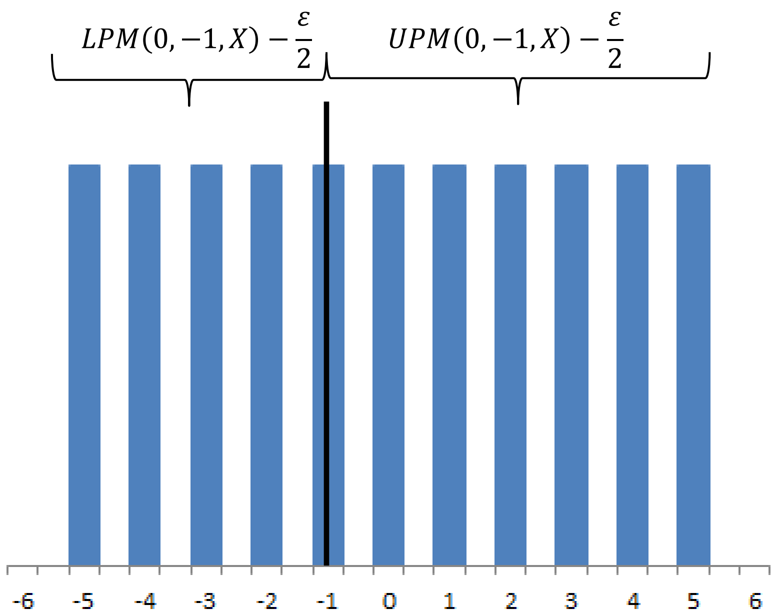
The figure below illustrates the consistent continuous vs. discrete CDF relationship, when 5 million observations (rational numbers) were added to a normal distribution analysis.



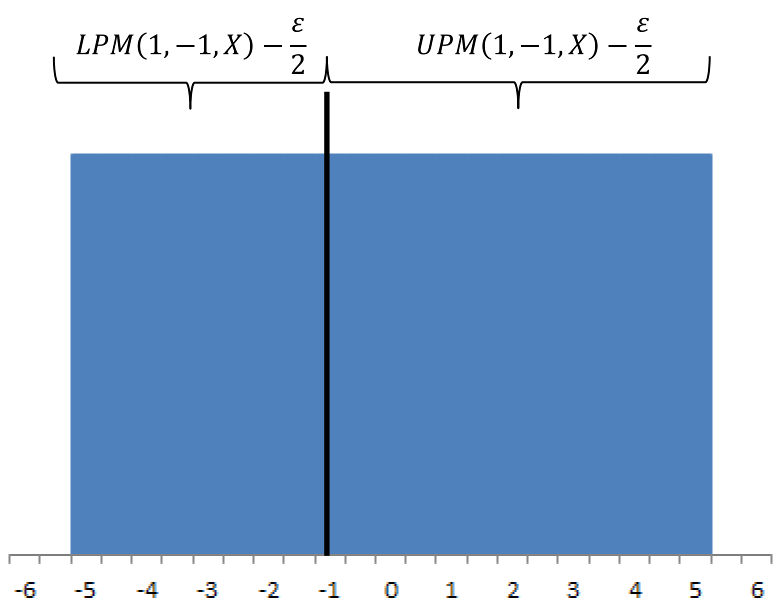
Continuous estimate converges towards discrete estimate as the target approaches sample mean (as *h* is increased from 0 to 4.5). The LPM *n*=0, *h=0* is denoted as LPM(0,0,X), LPM *n*=1, *h*=0) is denoted by LPM(1,0,X), LPM *n*=1, *h*=4.5 is denoted as LPM(1,4.5,X) and the LPM *n*=0, *h*=4.5 is LPM(0,4.5,X).

The discrete uniform distribution is measured with a degree zero partial moment while the continuous uniform distribution is measured with a degree one partial moment to capture the continuous area. This "area between the bins" contains all of the real numbers (whole numbers, fractions, decimals and irrational numbers). The unaccounted for continuous area is the reason when comparing an interval [a,b] to the entire distribution, the differences are exacerbated the smaller the interval per the figure above.

Interval [-5,-1] for the discrete uniform distribution where point probability $\frac{\varepsilon}{2}$ is pronounced,



Interval [-5,-1] for the continuous uniform distribution,



$LPM(0,-1,X)-\frac{\varepsilon}{2} > \frac{LPM(1,-1,X)}{[LPM(1,-1,X)+\boldsymbol{UPM(1,-1,X)}]}-\frac{\varepsilon}{2}$ for all targets < μ due to the greater amount of area between the bins in the right hand side of the distribution (UPM), overcompensating the denominator of the partial moment ratio. The inverse is true for all targets > μ.

CONCLUSION:

The question then turns to whether there exists a smaller set of numbers that lie within the "area between the bins". The answer is no. ***Whenever the degree is greater than zero, these numbers (rational and real) between the bins are automatically factored, turning the analysis to the continuous variety by weighting the deviations.*** Larger and smaller deviations than degree one will yield taller and shorter continuous areas respectively, when compared to the figure above.

So given the same interval, nothing exists outside of the continuous or discrete distribution per our area analysis of a vertically shifted number line, however weighted; thus supporting Cantor's views that no such set whose cardinality is strictly between that of the integers and that of the real numbers exists.

REFERENCES:

Viole, F. and Nawrocki, D. [2012a]. "Deriving Cumulative Distribution Functions & Probability Density Functions Using Partial Moments." Available at SSRN: http://ssrn.com/abstract=2148482