**ConvolutionsofDistributions.docx -**  CSC148, 5/8/18

**#1 Introduction**

As we know, a valid simulation depends on correct representation of system “processes” such as arrivals, services, etc.

So far server/service studies the distribution of arrivals was assumed to follow one of the common pmfs seen in systems:

Constant/deterministic, uniform, Poisson/exponential, triangular, empirical, and so on.

However, many system processes involve COMBINATIONS of distributions such as the above - - for both discrete and continuous cases.

We first look at one of the simplest discrete simulations, and see that something as simple as iterating (i.e., adding) it to itself creates a much more complicated distribution.

**#2 Example of a convolution of Uniform discrete distributions**

Given the simplicity of the uniform distribution on interval [a,b] (refer to it as Du),

it might seem that convolution of Du with itself would have a simple pmf … perhaps another linear function.

But this is not true !

**Example**: Random variable x1 is the result of one toss of a fair die, with Prob(x1=j) = 1/6, for 1 <= j <= 6.

This is a discrete uniform distribution on [1,6].

The simplest convolution involving x1 is x1 with itself: For notation clarity, let x2 be a distribution identical to x1.

Then, the result (total count of dots) of two independent and identically-distributed (iid) die tosses is modeled by the distribution of x1 + x2, the “sum” of distributions x1 and x2.

But, complications are evident:

1) **pmf of x1 + x2** has domain [2,12, not [1,6]. Result 1 is impossible, and 12 is the maximum possible result.

Enumeration of the 11 discrete points in the pmf (**call it f**) of x1 + x2 is shown in table below.

2) **f** is not linear; instead, f is the simplest form of a spline (piecewise polynomial)

pmf of f G = graph of pmf of f

6/36 x ( max Prob(x1+x2 = j) )

5/36 x x

4/36 x x

3/36 x x

2/36 x x

1/36 x x

0 1 2 3 4 5 6 7 8 9 10 11 12

|  |  |
| --- | --- |
| **(x1 + x2) result** | **Prob** |
| 1 | 0 |
| 2 | 1/36 |
| 3 | 2/36 |
| 4 | 3/36 |
| 5 | 4/36 |
| 6 | 5/36 |
| 7 | 6/36 |
| 8 | 5/36 |
| 9 | 4/36 |
| 10 | 3/36 |
| 11 | 2/36 |
| 12 | 1/36 |

f is a discrete symmetric 11-point triangular distribution with domain [2,12] (forming the “base of the triangle”) and

max Prob of (1/6) at domain point 7 (corresponding to the “height” of this isosceles triangle).

To check that G is in fact a probability distribution, calculate the area “AofG” under G in interval [1,7].

Do this by imagining G as a continuous straight line to use calculus for finding AofG:

∫ (1 to 7) (1/36)x – (1/36) < = The slop-intercept form of a straight line equation

= (1/36)(x2/2) – (x/36) | evaluated at limits (1 and 7) = ( (x2/72) – (x/36) ) | evaluated at limits (1 and 7) = 36/72 = 1/2.

By the symmetry of G, the entire AofG value over [1,12] is 2 \* (1/2) = 1, thus f is a probability function.

Therefore, convolutions of even the same simple distribution with itself, are not simple, because the domain changes and the general nature of the pmf becomes increasingly complex for iterated convolutions (x1 + x2) + x3, etc.

For n convolutions of Unif[0,1], the pmf shifts more and more right, and approaches a normal distribution.

**#3 A graph of Unif(0,1) convolution Unif(0,1)**

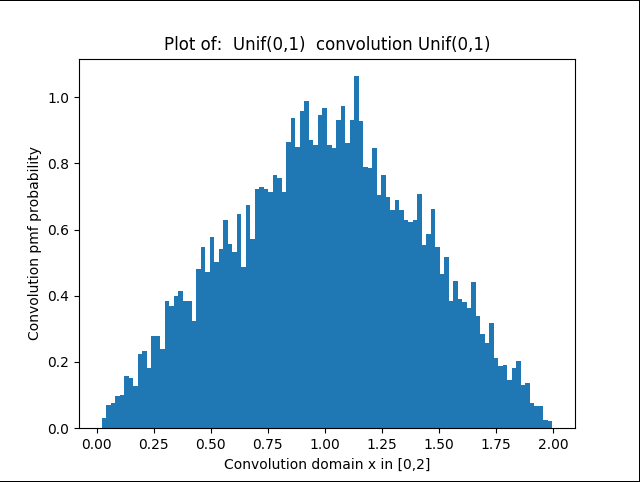
Some integral calculus determines the pmf of the case of a CONTINUOUS Uniform distribution on (0,1) with itself.

This was generated using a *different random number stream* for random variate values for each uniform distribution.

(See a way to generate iid RNSs in python: section#6)

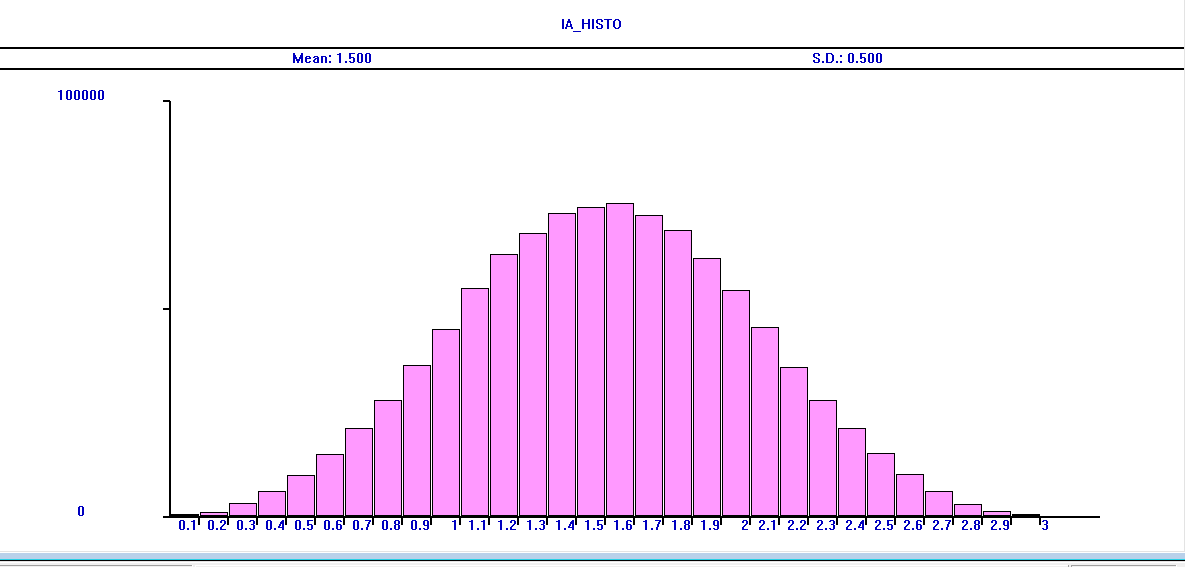
The domain stretches to [0,2], and the pmf function is simply x for 0 < x <= 1, and is (2-x) for 1 < x <=2.

For the DISCRETE case, the pmf graph shown was computed in python3 using calls to the numpy (Numerical processing library) module. It distinctly illustrates the triangular distribution of this convolution’s pmf (100,000 sample values).



**#4 Three convolutions of Unif(0,1)**

One million arrivals: distribution clearly show a non-linear (quadratic) pmf for this distribution (gpssW)



**#5 The Irwin-Hall distribution – wiki summary**

Reference: https://en.wikipedia.org/wiki/Irwin%E2%80%93Hall\_distribution

The first part of this article is as follows (diagrams also in first part of this article):

“In [probability](https://en.wikipedia.org/wiki/Probability) and [statistics](https://en.wikipedia.org/wiki/Statistics), the **Irwin–Hall distribution**, named after [Joseph Oscar Irwin](https://en.wikipedia.org/wiki/Joseph_Oscar_Irwin) and [Philip Hall](https://en.wikipedia.org/wiki/Philip_Hall), is a [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution)for a [random variable](https://en.wikipedia.org/wiki/Random_variable) defined as the sum of a number of [independent](https://en.wikipedia.org/wiki/Statistically_independent) random variables, each having a [uniform distribution](https://en.wikipedia.org/wiki/Uniform_distribution_(continuous)).[[1]](https://en.wikipedia.org/wiki/Irwin%E2%80%93Hall_distribution#cite_note-1) For this reason it is also known as the **uniform sum distribution**.”

The diagram clearly shows that as n of convolutions grows, the pmf approaches a normal distribution and the domain slowly widens, as well.

Note that the horizontal scaling in the pmfs below are narrower than used in the n=3 case obtained with gpssW

(that whose wider horizontal scale makes the axis marks clearer)

|  |
| --- |
| **Irwin–Hall distribution** |
| Probability density function  [Probability mass function for the distribution](https://en.wikipedia.org/wiki/File:Irwin-hall-pdf.svg) |
| Cumulative distribution function  [Cumulative distribution function for the distribution](https://en.wikipedia.org/wiki/File:Irwin-hall-cdf.svg) |

Here are some of the full (non-constant) pmf function definitions for small n:

n=2: x, 0 <= x <= 1, (2 – x) , 1 <= x <= 2

n=3: (1/2)x\*\*2, 0 <= x <=1, (1/2)(-2x\*\*2 + 6x - 3), 1 <= x <= 2, (1/2)(x\*\*2 -6x + 9), 2 <= x <= 3

n=4: (1/6)x\*\*3, 0 <= x <= 1, (1/6) (-3x\*\*3 + 12x\*\*2 - 12x + 4), 1 <= x <= 2, (1/6) (3x\*\*3 -24x\*\*2 + 60x – 44), 2 <= x <= 3,

and (1/6) (-x\*\*3 + 12 x\*\*2 -48x + 64), 3 <= x <= 4

**#6 . Convolution of two iid exponentially (exp) distributed random variables**

This example uses two RNS (random number stream) generators from the built-in python library –

import random

import matplotlib.pyplot as plt

:

RNS1, RNS2 = random.Random(), random.Random() # Use instantiable built-in class to get RNSs for iids

# Use a python “list comprehension” to get lots of points in pmf of the convolution

convolutionSamples = [ RNS1.expovariate(1) + RNS2.expovariate(1) for k in range (10000)]

# Plot the pfm

fig, axis = plt.subplots()

axis.hist(convolutionSamples, bins=100, normed=True)

axis.set\_title("Plot of: exp distr, mean 1 convolution exp distr, mean 1")

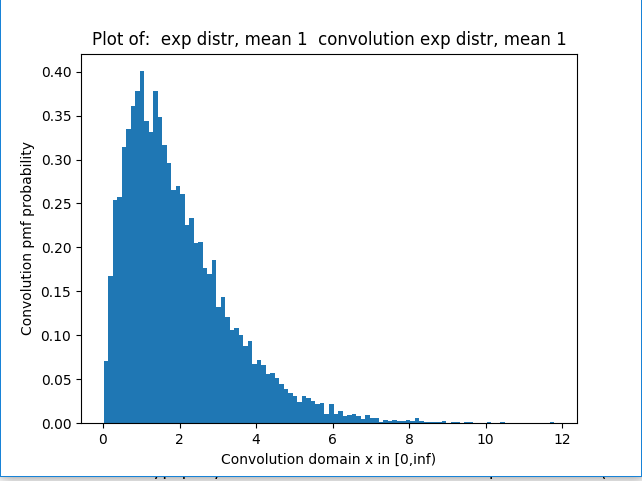
axis.set\_xlabel("Convolution domain x in [0,inf)")

axis.set\_ylabel("Convolution pmf probability")

# Save convolution pmf for Demos

fig.savefig("UnifConvUnif\_01.png")

plt.show()



This convolution result is in the family of gamma distributions

Unlike an exp distribution, a gamma distribution leans away from the y-axis in the part of the pmf with highest probability.

Convolutions of exp distributions are very common in natural as well as man-made systems = >

there are many applications of the gamma distribution.

In reliability engineering, the above gamma distribution can *model the expected lifetime of a system having a main component and an identical backup component* in case the main fails catastrophically …

There are MANY textbooks/papers/tutorials on the convolution concept in physics, engineering etc.

Many physical phenomena, such as signal mixing in electrical circuits, are understood and represented using

convolutions.