DS 3500: Advanced Programming with Data
Prof. Rachlin
Northeastern University

Simulation and Modeling

Predicting outcomes in the face of risk and uncertainty

Part 1: Extend the DRV Object Model

In class we developed a very basic object model for working with discrete random variables. In this homework, you will extend the object's functionality and apply your object to predicting the number of advanced civilizations in the galaxy capable of interstellar communication.

Here are the extensions you need to implement.

1. When plotting the distribution, users should have the *option* to overlay a cumulative distribution. (You will have seen examples of this the slide deck.). Also add an option for choosing whether the y-axis (probabilities) is log-scaled.

2. Add support for two continuous distributions: Uniform and Normal. You will approximate these continuous distributions with an underlying discrete random variable. Your constructor now needs to support additional options:

   type: discrete, uniform, or normal

   min/max: The upper and lower bound of a *uniform distribution.* N/A for other types.

   mean/stdev: The mean and standard deviation of a normal distribution. N/A for other types.

   bins: For uniform and normal distributions, how many bins used to approximate the continuous random variable. (You are approximating a continuous random variable with a discrete random variable so that we can reuse the methods of our DRV object.

   This is going to require some algorithmic thinking. Some numpy methods that might help you include **np.random.normal** and **np.linspace**

3. Implement methods to compute the expected value (E) and standard deviation (stdev) of a discrete random variable.

4. You may implement any other methods you deem helpful or necessary.

Part 2: Are we alone?

Are we alone in the universe? Or are there other intelligent alien civilizations within our galaxy? If so, how many? The Drake Equation, due to the American astronomer and astrophysicist Frank Drake, is an equation written in terms of probabilities. It is used to estimate how many actively communicating extraterrestrial civilizations there are in the Milky Way galaxy. As you will see, some of the terms in the equation can be estimated – others are purely guesswork – and YOU will be doing the guessing! The drake equation is:

$$N = R_* \cdot f_\mathrm{p} \cdot n_\mathrm{e} \cdot f_\mathrm{l} \cdot f_\mathrm{i} \cdot f_\mathrm{c} \cdot L$$

The outcome, N, is the number of civilizations in our galaxy with whom we might communicate. The other terms are:

Term	Meaning	Estimate, if possible
R_*	star formation rate	1.5 – 3.0
f_p	fraction of stars that have planets	Around 1.0
n_e	among stars having planets, how many of those planets can support life	1 – 5
f_1	fraction of life-supporting planets that actually develop life	???
f_i	fraction of planets with life that develop *intelligent* life	???
f_c	fraction of intelligent life-bearing planets that develop technology that releases detectable signals (such as radio waves) into space	???
L	the length of time (years) during which such civilizations release detectable signals into space. In other words, how many years do advanced civilizations last?	???

Use the Discrete Random Variable (DRV) class to estimate a distribution of values for N. Make use of any mixture of discrete, normal, and uniform distributions for your estimate of each of the key factors. For your probability distribution of each factor, e*ither cite a specific source or provide a reasonable justification based on your own intuition.*

What to Submit

Submit your code, documented estimates, and your final distribution plot. Also report your final distribution's Expected Value and Standard Deviation to a shared Google Sheet. (Link to be provided.)