ASTR/PHSX 315

Modeling Using Stochastic Methods

Graham W. Wilson

February 7, 2022

1. Modeling random processes introduction
2. Random numbers - the key ingredient to probabilistic modeling
3. Probability and statistics refresher

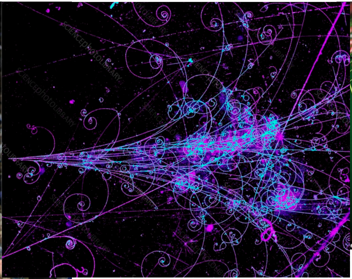
Introduction to Modeling Random Processes

Many processses, including many physical processes, can be modeled using **stochastic** methods based on **probability**.

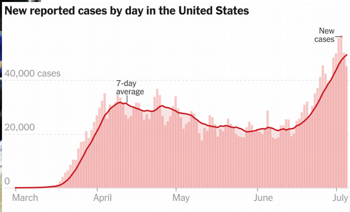
- Clearly **games** of chance like poker, roulette, blackjack.
- But also lots of **sports**, that have elements of chance to them, like basketball, baseball, football, golf, soccer, tennis.
- Particularly true for fields that are undeniably **probabilistic** in the sense that they are underpinned by the ideas of quantum mechanics, like particle physics. Governs probabilities for interaction and **decay** and subsequent interactions of particles with the detector.
- But it is also very useful for more everyday fields, where uncertainties and variations are facts of life. For example **elections** and generally anything with intrinsic uncertainties, such as experimental measurements.
- Even in **epidemiology**, some models are stochastic and take into account estimates for the probability distributions of quantities such as the number of infections per case, the incubation time, and infectiousness vs time.

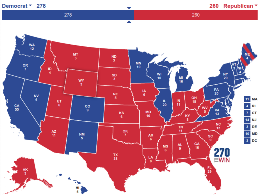
The key tool is the use of **random numbers**. When you master their use, you are empowered to model almost anything!

Examples




etty Images


New reported cases by day in the United States
New cases
40,000 cases
7-day average
20,000
0
March
April
May
June
July


Democrat 278
260 Republican
278
260
270 to WIN

Random Numbers

Many computational applications rely on the availability of computer-generated sequences of **random numbers** (sometimes referred to as pseudo-random).

The standard implementation results in **uniform** random numbers that are real numbers in the range [0.0, 1.0) with a uniform (ie. flat) probability distribution.

For every small interval, dx, in this range the probability of obtaining a random number from it is identical:

$$p(x) = 1$$

and $p(x)$ is a **probability density function** with the correct normalization since:

$$\int_0^1 p(x) dx = \int_0^1 dx = 1$$

Uniform.py example

```python
# Uniform.py
import random

# Illustrate standard uniform random number generator

# Initialize the random number generator using specified seed
SEED = 202
random.seed(SEED)

NINSTANCES = 10           # Number of random numbers to generate

# Generate uniform random numbers
for i in range(NINSTANCES):
    u = random.uniform(0.0,1.0) # standard uniform random numbers
                                # in range [0.0,1.0)
    print('uniform random number',i,'u =',u)
```

Probability and Statistics Reminder

A reminder of some of the concepts and terminology underlying probability.

There are **discrete** probability distributions. A good example is a dice, where each discrete outcome has a probability of $\frac{1}{6}$. Another example is the outcomes per state in a US presidential election. The sum of the probabilities of all possible outcomes is 1.

There are also **continuous** probability distributions like the uniform distribution we discussed, the exponential distribution associated with decay processes, and the normal distribution (or Gaussian in physics-land) or bell-shaped curve in lay-person language. The integral of the **probability density function**, $p(x)$, over the allowed domain is 1. Three of the basic distributions are:

$$\text{Uniform, } p(u) = \frac{1}{(b-a)}; \;\; u \in [a, b]$$

$$\text{Exponential, } p(t;\tau) = \frac{1}{\tau} \exp\left(-t/\tau\right); \;\; t \in [0, \infty)$$

$$\text{Normal/Gaussian, } p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\, exp[-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2]; \;\; x \in (-\infty, \infty)$$

$$\text{Standard}(\mu = 0, \sigma = 1)\text{ Gaussian, } p(z) = \frac{1}{\sqrt{2\pi}}\, exp(-z^2/2); \;\; z \in (-\infty, \infty)$$

Probability 2

Given the probability density function, $p(x)$, abbreviated as pdf, we can calculate the standard distribution quantities like the mean, usually denoted, μ, defined as the expectation value of x,

$$\mu = \langle x \rangle = \mathbb{E}(x) = \int x\, p(x)\, dx$$

the second moment, the expectation value of x^2,

$$\langle x^2 \rangle = \mathbb{E}(x^2) = \int x^2\, p(x)\, dx$$

and the variance defined as

$$\mathbb{V}(x) = \mathbb{E}((x-\mu)^2) = \langle x^2 \rangle - \mu^2 = \sigma^2$$

where the standard deviation, σ, is defined as

$$\sigma = \sqrt{\mathbb{V}(x)}$$

(correct for distributions – for samples there are some small adjustments). When using random numbers we should check that these statistics behave as we expect.

Sampling from pdfs - aka generating random variates

We want algorithms/recipes/functions for generating for example, u, t, x, z values according to the following 4 distributions. This is called sampling from these probability distributions leading to "random variates" or "random deviates".

$$\text{Uniform, } p(u) = \frac{1}{(b-a)}; \;\; u \in [a, b]$$

$$\text{Exponential, } p(t;\tau) = \frac{1}{\tau} \exp\left(-t/\tau\right); \;\; t \in [0, \infty)$$

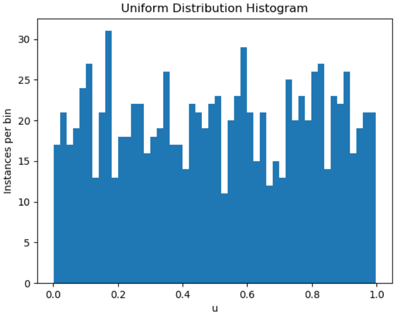
$$\text{Normal/Gaussian, } p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\, exp[-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2]; \;\; x \in (-\infty, \infty)$$

$$\text{Standard}(\mu = 0, \sigma = 1)\text{ Gaussian, } p(z) = \frac{1}{\sqrt{2\pi}}\, exp(-z^2/2); \;\; z \in (-\infty, \infty)$$

Often these are readily available. Sometimes we may need to learn more general techniques for producing our own, especially for less common distributions.
i) **testrandom.py** implements a variable transformation method: if $u \sim U(0, 1)$ then $t = -\tau \log(1-u)$ is distributed as $p(t;\tau)$.
ii) Given $z \sim N(0, 1)$ (standard normal), then $x = \mu + \sigma z$ is normal with mean, μ, and standard deviation, σ.

Uniform Distribution on [0,1)

(See UniformPlot.py for code)



Here we have 1000 random numbers drawn from the uniform distribution. Sample (mean, rms) are (0.504, 0.291) Error on the mean, $\sigma/\sqrt{N} = 0.009$.

What is the expected mean? Right, 0.5.
What is the expected rms? One can show that this is $1/\sqrt{12} = 0.289$.
So these 1000 random numbers are statistically consistent with expectations.
That is as much and as little as we can expect.
With uniform random numbers like this we can then produce random numbers from other distributions.

Recipe for generating a Gaussian distribution

One of the simplest (but not exact) ways to generate a Gaussian is to make use of a technique that puts the **Central Limit Theorem** (CLT) into practice.

Let's say we have 12 uniform random numbers from the previous exercise on [0,1).

We construct,

$$z = (\sum_{i=1}^{12} u_i) - 6$$

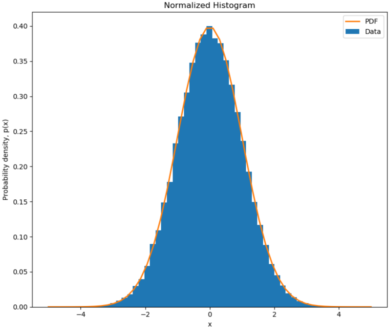
Then we can show that

$$\mathbb{E}(z) = \mu = 0, \quad \mathbb{V}(z) = \sigma^2 = (\sum_{i=1}^{12} \frac{1}{12}) = 1$$

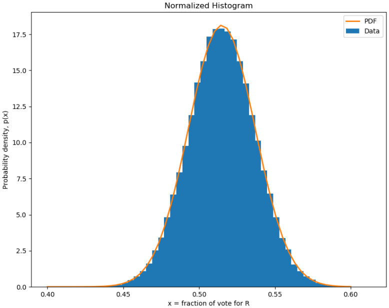
is a very good approximation to a standard normal/Gaussian.

The CLT says something like, "when independent and identically distributed random variables are added, their properly normalized sum tends toward a normal distribution as their number grows even if the original variables are not themselves normally distributed" (Wikipedia)

This is one of the reasons why the normal distribution is so common and useful



Standard Normal. 68.4% within $\pm 1\sigma$. 95% within $\pm 1.96\sigma$.



Example with mean, μ, of 51.5% and standard deviation, σ, of 2.2%.

What is the probability to measure $x < 0.50$? Can you find this in a standard table? Hint, calculate, the z-value from $z = (x - \mu)/\sigma$.

Models and Assumptions

- When using models, it is very important to state the model assumptions.
- A good place to start is to appreciate that all models are usually in principle "wrong". But that is not a good reason to discount them.
- It can just be that there are uncertainties on the correct parameters
- Or more fundamentally that the parametrization is wrong, maybe neglects important additional phenomena
- But that does not make them at all useless.
- Best guesses informed by data, and adhering to underlying scientific principles, are important for decision makers.

Two very important but difficult things to model are climate change, and the COVID-19 pandemic. They are particularly difficult, in that society has a role in their eventual trajectories, and the problems are multi-faceted.
I was interested already in epidemiological modeling. In 2006 when teaching the graduate computational physics class, one of the potential class projects was on modeling a global pandemic (inspired by the Avian Flu (H5N1) which had a human mortality rate of 53%) and risked spreading around the world if human-to-human transmission got established.

COVID-19 Pandemic Modeling

Often a lot of insight can be garnered from simplified models that while not pretending to give a complete picture, encapsulate some of the fundamentals, like energy conservation, probability conservation, causality, etc.

- Imperial College Model
  Look for example at the influential study by Imperial College, London researchers re the COVID-19 pandemic from March 2020. It predicted rather well the evolution, led to the lockdowns, bought valuable time, but also pointed out presciently that lockdowns were likely socially unsustainable.
- IHME Models (Institute of Health Metrics and Evaluation)
  This is a "hybrid" model encompassing empirical data and underlying epidemiological modeling. Relied on by many for medium-term predictions for hospital capacity etc.
  Global COVID-19 projections
- My TakeAway
  This is a difficult task, given real-world phenomena involved like evolution and social behavior, and stuff like test availability, etc, and is not as amenable to the first-principles scientific approach. At the peak of the omicron wave, they estimate that only 13% of actual infections are reported as cases.