4. Stochastic Thinking

Wednesday, April 22, 2020 5:32 PM

Stochastic Thinking

Stochastic: Randomly determined; having a random probability distribution or pattern that may be analyzed statistically but may not be predicted precisely

There is a lot of uncertainty in our world.

"Whether or not the world is inherently unpredictable, the fact that we never have complete knowledge of the world suggests that we might as well treat it as inherently unpredictable."

Stochastic Process

An ongoing process where the next state might depend on both the previous state and some random element (example: dice roll)

```
import random

def rollDie():
    """returns a random int between 1 and 6"""
    return random.choice([1,2,3,4,5,6])

def testRoll(n = 10):
    result = ''
    for i in range(n):
        result = result + str(rollDie())
    print(result)
```

Three Basic Facts About Probability

1. Probabilities are always in the range 0 to 1. 0 if impossible, and 1 if guaranteed.
2. If the probability of an event occurring is p, the probability of it not occurring must be 1-p
3. When events are independent of each other, the probability of all of the events occurring is equal to a *product* of the probabilities of *each* of the events occurring.

Check first if you have independence before calculating with the assumption they're independent.

Die Simulation

A Simulation of Die Rolling

```
def runSim(goal, numTrials, txt):
    total = 0
    for i in range(numTrials):
        result = ''
        for j in range(len(goal)):
            result += str(rollDie())
        if result == goal:
            total += 1
    print('Actual probability of', txt, '=',
          round(1/(6**len(goal)), 8))
    estProbability = round(total/numTrials, 8)
    print('Estimated Probability of', txt, '=',
          round(estProbability, 8))

runSim('11111', 1000, '11111')
```

Structure
There'll be an outer loop, which is the number of trials.
    And then inside-- maybe it'll have a loop, or maybe it won't-- will be a single trial.
We'll sum up the results.
And then we'll divide by the number of trials

Output of Simulation

- Actual probability = 0.0001286
- Estimated Probability = 0.0
- Actual probability = 0.0001286
- Estimated Probability = 0.0

Random.choice is pseudo random: the algorithm begins with a seed, and with the same seed you get the same random set of numbers

Moral of the Story
The estimated probabilities seem to be wrong- it didn't calculate the actual probability… why not?

1. It takes a lot of trials before getting a good estimate
   of the frequency of occurrence of a rare event (check before believing an estimate)
2. One should not confuse the sample probability with the actual probability
3. There was really no need to do this by simulation, since there is a perfectly good closed form answer. We will see many examples where this is not true.

Then when are simulations useful?

Use of a Simulation

In a birthday simulation there's a closed solution for the probability of 2 people with the same birthday (assuming each day of the year has the same probability)

$$1-\frac{366!}{366^{N}*(366-N)!}$$

Approximating using a simulation can get you a similar answer.

But when you ask for the probability of 3 people with the same birthday, it become a MUCH harder mathematical question. But changing the simulation is extremely easy.

Summary

We use simulations to do probabilistic calculation because it's often way easier to do simulations

A description of computations that provide useful information about the possible behaviors of the system being modeled

- Descriptive, not prescriptive (it describes possible outcomes, but they don't tell you how to achieve possible outcomes)
- Only an approximation to reality

Why use models?

- To model systems that are mathematically intractable
- To extract intermediate results
- Play what if games by successively refining it