Monte Carlo Simulation.*

Greg Kochanski
http://kochanski.org/gpk

2005/01/25 10:27:49 UTC

1 Introduction

The idea behind Monte-Carlo simulations gained its name and its first major use in 1944 [Pllana, 2000], in the research work to develop the first atomic bomb. The scientists working on the Manhattan Project had intractably difficult equations to solve in order to calculate the probability with which a neutron from one fissioning Uranium[1] atom would cause another to fission. The equations were complicated because they had to mirror the complicated geometry of the actual bomb, and the answer had to be right because, if the first test failed, it would be months before there was enough Uranium for another attempt.

They solved the problem by realising that they could simulate the trajectories of individual neutrons, one at a time, using teams of humans implementing the calculation with mechanical calculators [Feynman, 1985, Man, 2004]. At each step, they could compute the probabilities that a neutron was absorbed, that it escaped from the bomb, or it started another fission reaction. They would pick random numbers, and, using the appropriate probabilities at each step, stop their simulated neutron or start new chains from the fission reaction.

The brilliant insight was that even though the trajectories of the simulated neutrons didn't match the trajectories of any real neutrons in detail, the statistical properties of a large ensemble of simulated trajectories would be identical to the statistics of the real trajectories. That mean they could compute reliable answers for the important question, which was the probability that a neutron would cause another fission reaction. All you had to do was simulate enough trajectories.

When Simulation is Valuable.: Q: In a free fall, how long would it take to reach the ground from a height of 1,000 feet? **A:** I have never performed this experiment.

2 Simple Example

2.1 Birthday Problem - Classical Approach

Simple examples of Monte-Carlo simulation are almost embarrassingly simple. Suppose we want to find out the probability that, out of a group of thirty people, two people share a birthday. It's a classic problem in probability, with a surprisingly large answer.

Classically, you approach it like this: Pick people (and their birthdays) randomly, one at a time. We will keep track of the probability that there are no shared birthdays.

- The first person can have any birthday, and there (obviously) can't be any sharing, so there is a 100% chance that the group has no shared birthdays so far.

[1] Or, Plutonium, of course.

Independence of Birthdays: In the Birthday problem, we assume that all the birthdays in the class are independent. This allows us to easily compute the probability that each person won't share birthdays with the previous people.
Independence of birthdays is actually quite a good assumption, since different people come from different families and there is no mechanism for correlating birthdays of different class members. It would not be so good if lots of twins took the same classes at the same university. It would fail miserably if primary schools started grouping students by half-year instead of by year.

- The second person has one chance of sharing with the first person, so there is a 364/365 chance of picking her birthday on a different day. The probability of having no shared birthdays is now 364/365.
- The third person has two chances to sharing with the first two people (i.e. two days are taken). so there is a 363/365 chance that he won't share with the previous two people. The probability of no shared birthdays is now $(364/365) \cdot (363/365)$.
- When we pick the fourth person, three days are taken so far, so there is a 362/365 chance that she won't share birthdays with anyone picked so far. The probability of no shared birthdays is now $(364/365) \cdot (363/365) \cdot (362/365)$.
- ...
- The thirtieth person has 29 chances of sharing a birthday with previous people, so there is a 336/365 chance of picking a person who doesn't share a birthday. The probability of having no shared birthdays at all is now $(364/365) \cdot (363/365) \cdot (362/365) \cdot \ldots \cdot (336/365)$.

The overall probability of having no shared birthdays is then 0.294, giving a 71% chance that at least one pair of people have overlapping birthdays. It's not too complex, as long as you see the trick of keeping track of the probability that there are shared birthdays, rather than trying to add up the probability that one or more people share. (Try it, it's a mess!) It also takes some thought to realise that the probabilities are conditioned properly, so that multiplying together all the various $P(N^{\text{th}}$ person doesn't overlap|first $N-1$ people don't overlap) factors.

2.2 Birthday Problem – Monte-Carlo Approach

The solution here is conceptually very simple:

1. Pick 30 random numbers in the range [1,365]. Each number represents one day of the year.
2. Check to see if any of the thirty are equal.
3. Go back to step 1 and repeat 10,000 times.
4. Report the fraction of trials that have matching days.

A computer program in Python to do this calculation is quite simple:

```
#!/usr/bin/env python

import random                 # Get a random number generator.

NTRIALS = 10000               # Enough trials to get an reasonably accurate answer.
NPEOPLE = 30                  # How many people in the group?

matches = 0                   # Keep track of how many trials have matching birthdays.
for trial in range(NTRIALS):                  # Do a bunch of trials...
        taken = {}                            # A place to keep track of which birthdays
                                              #       are already taken on this trial.
```

```
        for person in range(NPEOPLE):            # Put the people's birthdays down, one at a time...
                day = random.randint(0, 365)             # On a randomly chosen day.
                if day in taken:
                        matches += 1                     # A match!
                        break                            # No need to look for more than one.
                taken[day] = 1                           # Mark the day as taken.

print 'The fraction of trials that have matching birthdays is', float(matches)/NTRIALS
```

And the answer is:

```
The fraction of trials that have matching birthdays is 0.7129
```

3 Example in Class

- How many raisins do you add to a batch of dough to make M cookies to make sure (with probability P) that a random cookie has at least N raisins?
- How about that 99% of the cookies will have at least one raisin?
- How about that all the cookies will (with probability P) have at least one raisin?

4 A Linguistic Example: Optimality Theory

Let's try an artificial model with a certain amount of linguistic reality, to see how Monte-Carlo techniques might be applied to bigger problems.

We will assume H_0 from the "How could one do a statistical test on optimality theory?" chapter. We have three streams of events, **e**, **E**, and **f**. These events are observations the linguistic behaviour of a language learner, such as use of a particular phone in a particular context. We assume that at some time, that we will call $t = 0$, two constraints in the subject's interlanguage swap order, which simultaneously stops **e** and **E**, and starts **F**. What does it look like? This:

```
.  .  .  .  .  .  .  .  . E.  .EE.  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
.  .  .  .  .  . E.  .  .  E  .  .  .E .  .  .  .  .f .  .  .  .  .  . f.  .  .f
. e.E .  .  .E E  .  E  .  E  .  e  .  .  .  .  . f.  f  .  . f.  .f .  . f.  .f
.  .  .  .E .  .  .  .  E  .  .  .e .  .  .  .  f  .  .  . f.  .  f  .  .  ff .
.  .  . E.  .  .  .  .  .  .e .  .  .  .  . f.  .f . ffff.  .  .  .  .  .  . f.
.  .  .  .  .  .  .  .  e  . E.  . E.  .  .  .  .  . f.  .  .  f  .  .  .  .  .
.  .  .  .  .  .  .E .  . E. E.  .  .  .  .  .f .  f  . f.  .  ff .  .  .  .  .
.  .  .  .  E  .  .  .  .  .  .  .  .  .  .  .  . f.  .  .  . f.  .  .  . f.  .
.  .E .  .  .  .  .  .  . e.  .  .  .  .  . f.f .  .  .  . f.  .  .  .  . f. f.
.  .  .E .  .  .E .  .  .  .  .  .  .  .  .  .  .ff.  .  .ff.f .  .  .  f  ff .
.  .  . E.  .  .  .  .  .  .  .  .  .  .  .  .  .  . f. f.  .  f  .  f  .  .
.  . e.  .  .  . E.  .  .  E  . E.  .  .  .  .f .  . f.  .  .f .  .  .  .  .  .
.  .E .  .  . E.  .E .  .  .  .  .  .  .  .  . f. f.  .  .  .  .  .f .  .  .  .
.  .  .  .  .  .  .  .  .  .E .E .  .  .  .  .  f  .  .  .  .f . f.  .  .  .  .
.  . e.  .  .  .  .  .  . E.  .  . E.  .  .  .  . f.f .f .  .  .  .  .  . f. f.
. E.  .  .  .  .  .  .  .  .  .  .  .E .  . ff  .  .f .  .  .  .  .  f  .f .  f
.  .E .  .  .  .  .  .  .  EE .  .  .  .  .  .  .  .f .  .  .  .f f  f  f f.  .
.  .Ee.  .  .  .  .  E  .  . E.  .E .  .  .  .  .  .f .  .  f  .f .  .  .f f  f
.  .  .  .  .  .  .  .  .E .  .  .  .  .  f  .  .  .  .f . f.  .  . f.  .  f  .
.  .  .E .  .  .E .  .  .  .  .  .  .  .  .  .f .  .  . f.  .  .f .  .  .  .  .
```

Each line above is a simulation of a different language learner, and I have aligned the times where the constraints swap. I assume that **e** is seen on 1% of the observations, **E** on 5%, and **f** on 10% of the observations.

Looking at this, it's immediately obvious that H_0 is plausible. One finds (for some subjects) **e** and **E** events immediately to the left of $t = 0$, and (for some subjects) **f** events immediately to the right.

However, we can also learn how much power our statistical analysis will have. If we take just one line from that chart, corresponding to one subject, it looks like this:

```
.e EE .  .  .e .  .  . E.  E  E  .  .  . f.  .f .  .  f f. ff f.  .  .  .  .  f
```

We see a gap between the latest **E** and the earliest **f**. That's not due to a failure of H_0, but just random chance. It was possible that **E** might have happened just before $t = 0$, but it didn't. Likewise, **f** *might* have happened just after $t = 0$, but it didn't. Perhaps the subject didn't choose to say a word that contained the phone we were monitoring, or perhaps the subject just didn't say anything at all. For whatever reason, the first **f** event happens a few minutes after $t = 0$.

So, we expect to see a gap, even if H_0 is true. Well, what about another hypothesis, H_1? Assume that H_1 says that **f** events start 3 minutes later than the end of **E**. Will we be able to reject that? No, not on the basis of a single subject. It's inside the gap.

We might be able to reject variants of H_1 that prescribe a large spacing between the end of **E** and the beginning of **f**, but the power of any statistical test will clearly depend on the spacing that H_1 prescribes. Clearly, we will never be able to rule out all variant of H_1. Some will be so close as to be indistinguishable to H_0.

This leads us to the normal procedure for a hypothesis test. We can't prove H_0 by rejecting all variants of H_1. Invariably, some variants of H_1 are too close to reject. Instead, we ask if we can reject H_0. If we aren't able to reject it statistically, we (tentatively) accept it, and assume that it is true.

Here's an example where we *could* reject H_0:

```
.e EE .  .e .  . E.  E  E  .  .  .  .  .  .  .  .  . f.  .f .  f f. ff f.  .  f
```

In this example, the gap is too big. Off to the left, we tend to get an **E** every 2 dots or so, and over to the right, we tend to get a **f** every 2 dots or so. Consequently, we might expect a four-dot gap between the last **E** and the first **f**. Instead, the gap is 9 dots, much bigger than we might expect. We'll later do the statistics to prove that this would be a very improbable result, if H_0 is true, consequently H_0 probably wouldn't be true.

Finally, there is the added complexity that in a real experiment, no omniscient genii is around to align the times when the constraint ordering swaps. In real life, the data would look more like this:

```
.  .  .  .  .  .  .  .  .  . E.  .E .  .  .  . f  .f .  .  .  . f  .  . f  .
.  .  .  .  .  .  E.  .  .  .  .  .  . E  EE .  .  .  .  .  .  .  .  . f. f.  .
.  .  E  .  .  .  .  .  .  .  .  .  .  . E  .e .  .  . f  .  . f.  .  .  .  .
.  . E.  .  .  .  .  .  .  .  .f .f f  .  . f.  .  . f  .  .  . f  .  f  .f
.  .  . .E .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  . E.  .  .  .  .
.  .  .  .  .  .  .  .  . E.E .  .  .  .  . E  .  .  .  .  .  .  .  . E.  . ff
.  .  .  .  .  .  .  .  .  .  .e .  .  .  .  .  . EE .  .  .  .  .  .  .f .  f
.  .  . E.  .E .  .  .  .  .E .  .  .  .  .  .  .  .  .  .  .  .  .  E  .  .
.  .  .  .  .  .  .  .E .  .  . E  .  .  .  .  .  .  .  .  f  .  .f .  .  .  .
.  . .E .  .  .  .E .  . E  .  .  .  .  .  .  .  .  .  . E.  .  .  .  .f .f
.  .  .  .  .  .  .  .  .  . f. f.  f f.  f  f  f  .  .  .  .f f  .f .  .  .  .
.  .  .  .  .  .  ffff .  .  .  .  . f  .  . f  .  . f.  .  .f .  . f.  ff .
.  .  .  .  .  . E.  .  .E .  E  .  .  .  .  .  .  .  f  . f.f . f.  .f .  .  .
.  . .E .  .  .E .  .  .  .  .  .  e.  .  . E  .  .  .  .  .  .  .  .  .  .  .
.  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  f  . f.  . f.  .f .  .  .  .
.  .  .  e  .  .  .  .  .  .  .  .E .  .  .  .  . E  .  .  .  . f.  .f .  .  .
.  .  .  .  .  .  E  .  . .E .  .  .  .  f.  .  .  .  . f.  . f.  f f.  .  .
.  .E .  .  .  .  E  .  . .E .  .  .E .  .  .  .  e. E.  .  .  .  .  .  . f.  .
.  . E.  .  .e .  .  f f.  f  .  .  .  .  .  .  .  .  .f .  .  .f .  .  . f.  .
.  . E.  .E E E.  .  .  .  .  .  .  .  .  E  .  .  .  .  .  .  .  .  .  .  .  .
```

Here, H_0 is true, individually, for each line, but the subjects don't learn the language at the same rate, and consequently don't all make the constraint interchange at the same time.

References

Richard Feynman. *Surely You're Joking Mr. Feynman!* Bantam, 1985.

History – Los Alamos – Oversight Committee Formed. The Manhattan Project Heritage Preservation Association, http://www.childrenofthemanhattanproject.org/HISTORY/H-06c12.htm, 2004. URL `http://www.childrenofthemanhattanproject.org/HISTORY/H-06c12.htm`.

Sabri Pllana. History of Monte Carlo method. http://www.geocities.com/CollegePark/Quad/2435/index.html, August 2000. URL `http://www.geocities.com/CollegePark/Quad/2435/index.html`.