Baby steps: a simple illustration of the Metropolis algorithm

Steve Simon

2021-03-19

### Abstract

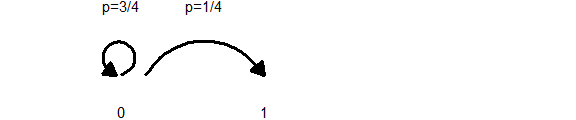
The Metropolis algorithm is a simple approach to generating random observations from a distribution where the density is known up to a proportionality constant. This algorithm is the easiest to understand and to implement of various methods used in Markov Chain Monte Carlo. In this paper, we show a simple example of the Metropolis algorithm to simulate the geometric distribution with =1/2, and describe the resulting simulation as analogous to a baby learning to walk.

### Introduction

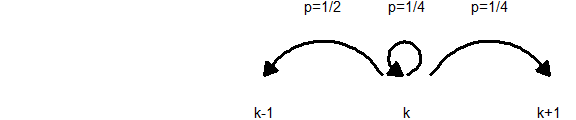
On March 24, 2017, the FiveThirtyEight website published a couple of mathematical puzzles, one which was contributed by this author.

“Your baby is learning to walk. The baby begins by holding onto a couch. Whenever she is next to the couch, there is a 25 percent chance that she will take a step forward and a 75 percent chance that she will stay clutching the couch. If the baby is one or more steps away from the couch, there’s a 25 percent chance that she will take a step forward, a 25 percent chance she’ll stay in place and a 50 percent chance she’ll take one step back toward the couch. In the long run, what percent of the time does the baby choose to clutch the couch?”

Figures 1 and 2 The transition probabilities are illustrated in Figure 1 and 2.



\*Figure 1. Baby walk transition probabilities at X = 0.



\*Figure 2. Baby walk transistion probabilities at X > 0.

Most statisticians will immediately recognize this as a discrete-time Markov chain, because the location of the baby at any time is a function only of the location at the previous time. The solu

You can solve this question easily with a computer simulation, but I derived the problem using the Metropolis algorithm to simulate a geometric distribution with p=1/2.

### The Metropolis algorithm

The Metropolis algorithm, first defined in Metropolis (1953) was originally developed to study the properties of substances composed of a finite number of interacting molecules. The method has proven useful for many physics problems such as simulated annealing (Kirkpatrick 1983).

It is also useful in simulating posterior distributions in a Bayesian analysis.

The posterior distribution is given by the following formula

if is continuous or

if is discrete.

The Metropolis algorithm is especially helpful if is high dimensional, because the evaluation of the denominator can difficult to do, either exactly or approximately. The Metropolis algorithm works even if the denominator is unknown because it relies on the ratio

and the denominators cancel out in this ratio.

Suppose you have a distribution, where the density p is known, but where it is difficult or impossible to directly generate a random sample from that distribution. Typically p is a multivariate distribution, but in the example I show below, it is a univariate distribution.

Select a starting point X0 that satisfies p(X0)>0. At time points t=1, 2, …, sample a provisional value X’, not from p (which is difficult or impossible), but from a jumping distribution J(X’|X0). Notice that this is a conditional distribution, which means that the spot that you jump to depends on the value of X0. For the Metropolis algorithm, this jumping distribution has to have a symmetry property J(a|b)=J(b|a).

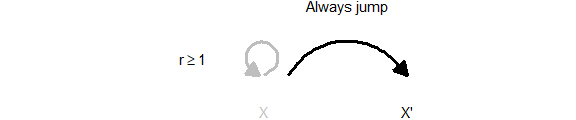
You will select a new value X1, which is either X’ or X0. The selection depends on the ratio r=p(X’)/p(X0). If r is greater than 1, you always select X’. If r is less than 1, you select X’ with probability r and X0 with probability 1-r.

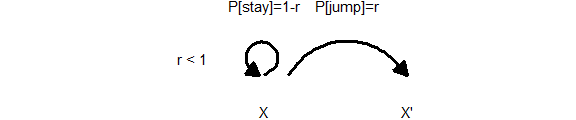
Now repeat this process using X1. Select a provisional value X’ from the jumping distribution J(X’|X1). Note that the jumping distribution is now conditional on a different value. Compute the ratio r=p(X’)/p(X1). Select X2 equal to either X1 or X’ depending on the ratio r. Continue this process to generate X3, X4, etc.

The rationale for this approach is that you should look randomly in a variety of directions. If the new location is more likely than your current location, always sample the new location. If the new location is little less likely, then it makes sense most of the time to jump to the new location, but once in a while you will be better off staying at the current location. If the new location is much less likely, you will only rarely jump to that part of the distribution.

With any reasonable jumping function, you can show that this algorithm visits all the regions of the distribution and with probabilities that are appropriately large or small.

Figure x illustrates the decision rule on jumping or not.





### Applying the Metropolis algorithm to the Geometric distribution

The Geometric distribution is a discrete disctribution with support on the non-negative integers. The probabilies for the hypergeometric distribution are computed as

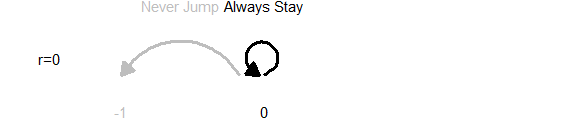
For , the formula simplifies to

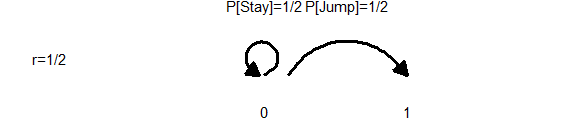
To simulate the Metropolis algorithm, specify a jumping distribution

Define a jumping distribution with probability 1/2 at X’=X-1 and at X’=X+1. This is a simple jumping distribution that moves one unit to the left or right with equal probability.

Note that the ratio of two consecutive probabilities is

At X=0, a jump to the left represents jumping to a value (-1) with probability 0. A jump to the right represents jumping to a value which is half as likely.

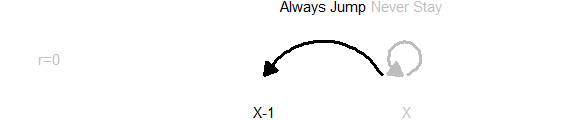




Combine these two cases to compute the transition probabilities 3/4 for staying at 0 (continue to clutch the couch) and 1/4 for taking the first step away from the couch.

Once away from the couch, the probabilities change. A step to the left represents a doubling of the probability, so a step to the left will always be taken.

A step to the right represents a halving of the probability, so a step to the right will be taken half of the time and the baby stays in place half of the time.





Combine these to get transtion probabilities of 1/2 for going one step to the left, 1/4 for staying in place, and 1/4 for going one step to the right.

### Discussion

John Kruschke motivates the Metropolis algorithm in Chapter 7 of his delightful book on Bayesian data analysis (Kruschke 2014). He describes an island hopping politician who wants to spend time on each island in proportion to its population. He can hop to the east or west, or stay on the current island, and uses the mechanics of the Metropolis algorithm to choose when to hop islands and when to stay put. The baby-walk example is similar in spirit, but provides a concrete distribution to evaluate.

Jim Albert and Jingchen Hu propose an arbitrary discrete distribution with five values and shows how students can simulate from this distribution using coin flips.

### Bibliography

Oliver Roeder. Will The Baby Walk Away? Will The Troll Kill The Dwarves? FiveThirtyEight (2017, Mar 24). <https://fivethirtyeight.com/features/will-the-baby-walk-away-will-the-troll-kill-the-dwarves/>. Accessed 2021-04-02.

S. Kirkpatrick; C. D. Gelatt; M. P. Vecchi Science, New Series, Vol. 220, No. 4598. (May 13, 1983), pp. 671-680.

Kruschke, John (2014). Doing Bayesian Data Analysis. A Tutorial with R, JAGS, and Stan. 2nd edition. Academic Press.

Metropolis, Nicholas; Rosenbluth, Arianna W.; Rosenbluth, Marshall N.; Teller, Augusta H.; Teller, Edward (1953). “Equation of State Calculations by Fast Computing Machines”. J. Chem. Phys. 21 (6): 1087. <doi:10.1063/1.1699114>.