Prof. Sergio A. Alvarez
21 Campanella Way, room 569
Computer Science Department
Boston College
Chestnut Hill, MA 02467 USA

http://www.cs.bc.edu/∼alvarez/
alvarez@cs.bc.edu
voice: (617) 552-4333
fax: (617) 552-6790

CS244, Randomness and Computation Introduction

What is Randomness?

Situations requiring an impartial binary decision not based on any prior knowledge or preference are often settled by a coin toss. The outcome of such a toss is described as being "random". What does this mean? Here are some typical answers:

1. Both possible outcomes of the coin toss, heads and tails, are equally likely to occur.
2. The precise outcome, either heads or tails, cannot be predicted.

It may not be obvious at first sight, but the preceding two conditions are not equivalent. A coin that lands heads and tails equally often need not be unpredictable: an internal mechanism could simply vary the outcomes consistently between heads and tails on alternate tosses, for example. Conversely, even if it is impossible to predict the precise outcome of a toss for a particular coin, it isn't necessarily true that heads and tails are equally likely outcomes; the coin can slightly favor one of the two without becoming entirely predictable.

Repeatability. There is a subtle point here: if only a single toss is made, then the outcome could actually be completely determined in advance without this fact making it any easier for an observer to predict it. Even if the coin is not "rigged" artificially, one could appeal to some notion of "fate" to explain whatever outcome actually occurs. Such considerations are at odds with our ideas about randomness. Therefore, when describing an event as being random, we will assume that the act that leads to it can be repeated, at least in principle.

Random sequences

Assume that we repeat the act of coin tossing many times. The result is a long sequence of heads and tails. What is it about such a sequence that qualifies it as being random? As before, we invoke the properties of equal likelihood of heads and tails, and of unpredictability of individual outcomes. How can we tell whether a supposedly random sequence satisfies these properties? Equal likelihood of heads and tails translates to equal numbers of heads and tails. Well, almost. One should not expect the numbers of heads and tails to be exactly the same. In fact, if the numbers were always exactly the same, the sequence would be perfectly predictable: the next toss would always produce the outcome that has occurred least frequently so far! Instead, one expects the numbers of heads and tails to be *approximately* the same. Unpredictability demands that the outcome of a toss not be knowable in advance based on the outcomes of previous tosses. Thus, no clear patterns should be discernible in the elements of a random sequence.

Example. I used MATLAB to generate a near–random sequence of one million letters of the alphabet. In this case, there are 26 possible outcomes rather than two. Thus, equal likelihood corresponds to about 1 out of every 26 "tosses" resulting in any predetermined letter of the alphabet, or a total of about 38462 instances of each letter in a sequence of length one million. The results are very close to this. The actual numbers of a's, b's and c's obtained, for example, are 38459, 38182, and 38425. Numbers for other letters are similar. Therefore, it seems plausible that the generated sequence in fact exhibits equal likelihood for all outcomes – our first randomness requirement. What about unpredictability? We mentioned that no special patterns should be detectable in the sequence. I looked for the words "cat", "dog", and "rat" in the generated sequence. Shockingly, the first two of these words occur about 60 times as substrings, and the third occurs nearly 80 times!

Quantifying expectation

Is there more structure to the presumably random sequence in the above example than one should expect? Well, one should expect some chance occurrences of the words in question. The classical metaphor in this regard is a monkey typing random characters, with the possibility that it might write a literary masterpiece in this way if given enough time. This suggests that a theory of randomness should quantify the frequency with which various apparently structured outcomes are to be expected, even if the generating mechanism is assumed to be completely random. Indeed, characterizing such expected, or "average", behavior is a central point of probability theory. More on this later. Whether or not there are *any* detectable patterns in a given sequence is something that we will not be focusing on in this course. There is a related field that considers such questions further. The basic idea is to consider, for a given sequence, what the shortest program might be that prints the sequence. The shorter the program, the more structured (or predictable) the sequence. A truly random finite sequence will have a program that essentially reduces to a print statement with the target sequence as its argument. The longer the the random sequence, the longer the program needed to generate it. A truly random infinite sequence will not be printable by any (finite) program.

The collective point of view

We will be studying the theory and applications of probability. In probability, a *collective* viewpoint is taken. For instance, rather than considering individual "random sequences", we will be looking at *collections* of sequences generated by random means. To avoid future confusion, let me clarify that, in other contexts, a sequence may be considered to be a collection. This happens in certain situations in which single objects replace sequences as the basic objects of study. In such cases, we will focus on collections of single objects, and these collections may be viewed as sequences of single objects.

Archetypal example: the random walk. One day, you are extremely bored (hopefully not in CS244). You decide to go on a walk steered by chance. Being fundamentally safety-conscious, you limit the walk to a one-dimensional world: your own street. You have a coin in your pocket that you flip before taking each step. If the coin lands heads, you take a step to the right; if it lands tails, you take a step to the left. You manage to take 1000 steps in this way before losing patience. The results are shown in Fig. 1, where the vertical axis represents the distance from the starting point in steps. Overall, Fig. 1 displays a lot of erratic changes of direction, as one would expect. However, you actually manage to cover some ground, ending up nearly 20 steps to the right after 1000 steps. At your furthest, you make it a full 45 steps to the right of where you started! This seems noticeably one-sided.

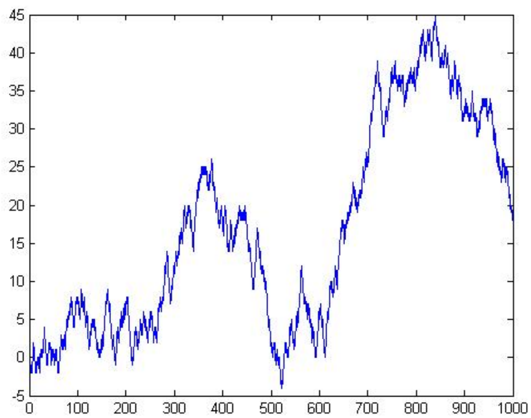

Figure 1: A 1000 step one-dimensional random walk.

The next day, boredom once again sets in. Encouraged by the somewhat surprising results of your initial excursion, you set out a second time, armed with your trusty steering coin. Later that night, you visualize your journey using the software from your navigation device, as shown in Fig. 2. This time, your random walk has taken you to the other end of the street, over 30 steps in all. So far, it seems that neither direction of travel is preferred by your coin. Furthermore, it is difficult to see any systematic trends in your random travels.

But, stubbornly, you persist. The results of your walk on the third day are shown in Fig. 3. This time, after short-lived excursions to both ends of the street, you have gotten absolutely nowhere! Not terribly exciting, but at least you don't have to face the awkward walk home after the last coin flip. The unsatisfying results of the day only serve to increase your determination. You *will* get to the bottom of these random walks, one way or another.

You continue with your daily walks through the first day of Spring, then Summer, and Fall, and, when the New Year comes, it finds you randomly stepping along your street in the snow once again. Each time, you view the day's results with anticipation, hoping for a

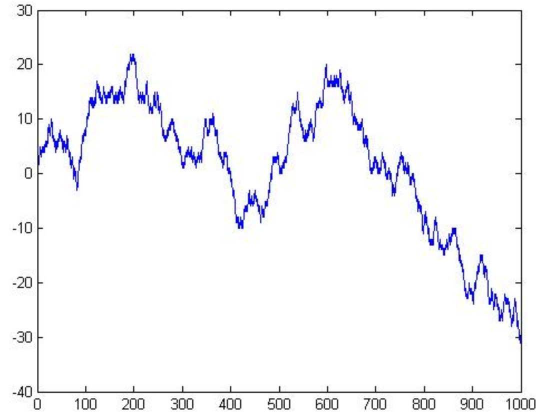

Figure 2: A second 1000 step one-dimensional random walk.

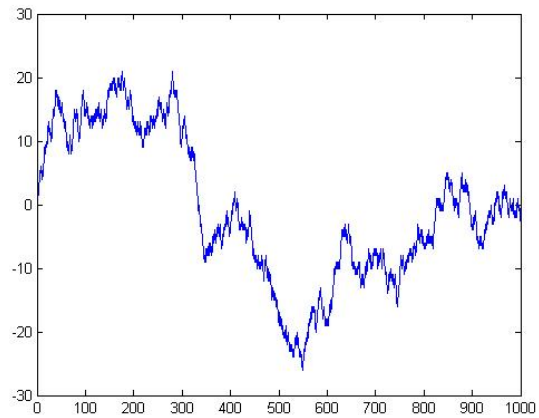

Figure 3: Day 3 random walk.

pattern. Each time, it eludes you. As you have done each day before, you halfheartedly file the results in your night table drawer and call it another day. Three years pass. That night, after the inevitable emptiness repeats itself, you are unable to contain your frustration any longer. You empty your drawer, throwing its contents on the floor in a fit of rage. Graphs pile on top of graphs, each barely visible outline still taunting you through the thin sheets of paper above it. Then, suddenly, you notice something in this chance superposition of diagrams: if the axes are aligned exactly in all of the graphs, the jumble of all of the graphs appears to suggest a coherent shape. Could it be that this behavior is not entirely random? See Fig. 4.

Certain features are apparent in Fig. 4. The underlying shape is symmetrical with respect to the horizontal axis: neither left nor right is a preferred direction in your walks. You knew this, and yet it is now as clear as day in this one composite diagram. But also, the locations in the diagram that are covered by the collection of graphs appear to fall within a relatively

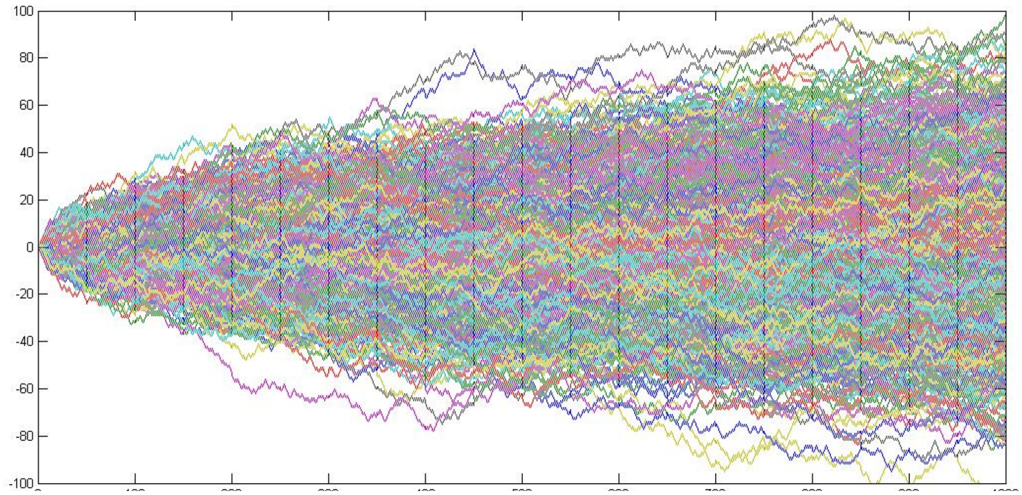

Figure 4: A collection of 1000 superimposed 1000 step one-dimensional random walks.

well-defined region. This region is narrow at the left, for small "time" (number of steps taken), and gets gradually larger toward the right, as time increases. In other words, there is a tendency to wander away from the starting point as the walk progresses, in a somewhat *predictable* fashion. Yes, some individual graphs deviate from this description, but, overall, the *collection* of graphs has this tantalizingly nonrandom appearance.

Charting a course

The considerations above suggest some long-term themes for our study of randomness. We will aim to understand random variation from a collective point of view. This will only be possible by considering random "experiments", such as a coin toss, or a random walk, that can be repeated many times. We will seek to quantify the random variations that occur in collectives resulting from many repetitions of such random experiments. Individual random results will remain unpredictable, but we will retain the hope that, if the number of repetitions is very large, there may be some determinism in the collective behavior of such repetitions. Let the random walk be an internal guide to this exploration. We have seen glimpses of structure underneath the surface, but many questions remain. How do we quantify the "average" distance traveled? Is there a well-defined typical deviation from this average consistent with the shape seen in Fig. 4? How wide is the covered region at the far right end? Exactly what shape is this? Is this shape stable, or merely a fluke of a particularly lucky three years' worth of walks? With these questions in mind, let us begin.