

# Statistics and Intuition for the Classroom

## KEYWORDS:

Teaching;  
Fractals;  
Randomness;  
Random walks;  
Quality control;  
Evolutionary biology.

*Sangit Chatterjee*

Northeastern University, Boston, Massachusetts, U.S.A.

*James Hawkes*

College of Charleston and Quant Systems  
Charleston, South Carolina, U.S.A.

## Summary

Examples and case studies are presented of statistical reasoning, thinking and intuition that may arise in perception of randomness and in particular for random walks. The relationship between art and science can be explored through various notions of the statistical concepts of randomness.

## ◆INTRODUCTION◆

*"I returned, and saw under the sun, that the race is not to the swift, nor the battle to the strong, neither yet bread to the wise, nor yet riches to men of understanding, nor yet favour to men of skill; but time and chance happeneth to them all."*  
Ecclesiastes

In this article we present a series of four examples tied together through a common theme of perceptions of randomness as manifested in examples of random processes. We have used these examples in our classrooms for illustrating statistical ideas, concepts, intuition and thinking. The aim is to present to the teachers of introductory and intermediate statistics courses some interesting material on random walks that they can use to illustrate important statistical principles. The presentation is from a teacher's perspective and the discussions and the lessons are therefore directed to elements that add to teaching.

## ◆PENSION FUND MANAGERS ◆

Our first example begins with the broad question of evaluating an employee's performance based on appropriate data. To make the task well defined and more approachable, we discuss the issue of evaluating pension fund managers. Pension fund managers are often evaluated on the returns they

derive by managing a portfolio of assets. Conventional wisdom might reward those who have had superior years and dismiss the poor performers. Is conventional wisdom reasonable? How can we evaluate twenty five managers' performance over a ten year period? Assuming the stock market behaves randomly, we can simulate a ten year performance of a fund manager by tossing a coin 10 times. The outcome head signifies a successful investment year (above average performance) and the outcome tail is an unsuccessful year, then a ten year history for a fund manager can be obtained by observing the outcomes of ten tosses of a coin.



Table 1. A summary of a ten year performance history of 25 portfolio managers.

Successful Years:	2	3	4	5	6	7
Frequency of Managers:	1	7	6	6	1	3

$$\bar{X}=4.48, S_x = 1.56$$

Let  $X$  represent the number of successful years for a fund manager (i.e. the number of heads in ten tosses). By repeating a ten toss sequence for each

of the 25 fund managers, we develop comparable histories. One of our classes generated the ten year history of 25 managers summarised in Table 1. There are two relatively unusual observations; one manager had only two good years and another had 8 good years. We see that 14 managers performed below the market average while 11 performed above the average.

What can we learn from the above simulation? If markets behave randomly, then a random choice of portfolio will put some managers' performances above the average and put some below the average. Whether a manager falls above or below the mean will have nothing to do with ability, market insight, or any other knowledge. Since the variation in performance is due to randomness, an insightful and lucky fund manager might appreciate this paraphrase of Mark Twain: "Let us be thankful for the unlucky fools. But for them the rest of us could not succeed." It is imperative to identify the cause of variation in fund performance; otherwise, our reward system becomes a roulette game rather than an evaluation/incentive system.

Assuming the stock market is completely random is not realistic. While there are non-random components in the market that can be exploited by a shrewd fund manager or an index arbitrageur, these components are minor in comparison to the extraordinary amount of apparent randomness in the market. Does the existence of small amounts of non-randomness affect our conclusions concerning the appropriateness of annual evaluations? If most of the variation in portfolio return is due to randomness, then it will be virtually impossible to detect a superior fund manager after only one year's performance. Perhaps after several years of observation it might be possible to detect a superior performance.

---

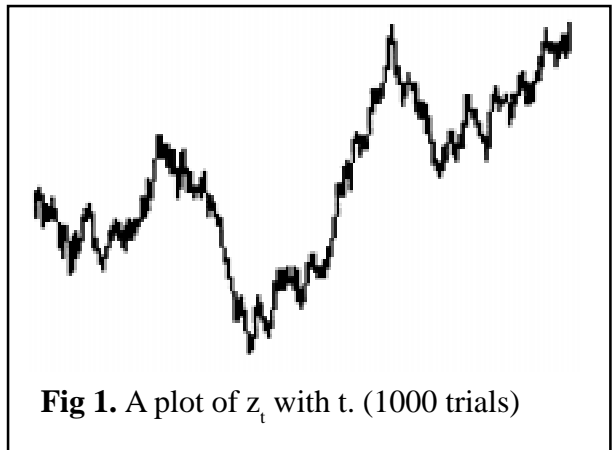
## ◆ STOCK MARKETS ◆ AND RANDOM WALKS

---

The discussion of the evaluation of pension fund managers and the behaviour of the movement of stock prices are closely aligned concepts. Our second example takes up for discussion the topic of stock price movement and an elementary simulation model associated with it. While there are many theories as to what makes stock prices move, a vast preponderance of the literature suggests that daily market movements are largely random. Unfortunately, those trying to model market fluctuations deterministically, even the simplest kind, will find the task daunting. A simple probability model

called a random walk is commonly used to describe the market behaviour of individual stocks as well as market indexes. To illustrate the random walk, suppose we have a bag of black and white marbles and let each student in the class draw a marble. If the marble is black we call it a (+1) and if white (-1). We designate each student's draw as  $x_t$ ,  $t = 1, 2, \dots, 100$ . We keep track of the sum of all the outcomes which we will denote as  $z_t = z_{t-1} + x_t$ . A plot of  $z_t$  against  $t$ , called a random walk, for 100 observations is given in Figure 1. The "patterns" that we see in Figure 1 look very much like the movement of a stock or stock market index over time.

Thinking about random walks can generate a number of interesting questions. Will the walk zig-zag across the  $t$ -axis frequently or will it sustain itself above or below the  $t$ -axis for prolonged periods? Does a random walk produce apparent patterns? We have observed that students' intuitions about these matters are generally at odds with the actual outcomes. We conclude by displaying several real economic time series which might be obtained from a financial newspaper such as the *Wall Street Journal*.



**Fig 1.** A plot of  $z_t$  with  $t$ . (1000 trials)

We then display data through computer experiments generated using MINITAB (Schaefer, 1989) or MATHEMATICA (Wolfram, 1991).

The following MINITAB commands create the plots for Figure 1:

```
MTB > random 100 c1;
SUBC> Bernoulli P1=0.5.
MTB> Letc2=2*c1-1
MTB > parsum c2 c3
MTB > tsplot c3
```

The MATHEMATICA commands to create a similar plot are:

```
ListPlot[
FoldList[ Plus, 0,
Table[ 2 * Random[Integer, {0, 1}]-1,
```

```
{1001]], PlotJoined->True]
```

The example illustrates a deep principle of uncertainty in that, though the individual outcomes are random, the aggregate output may indeed show “patterns”. However, such “patterns” are not persistent. The “patterns” we perceive are merely comfortable illusions caused by our own ignorance with uncertainty.

Random processes that wobble around some central value will display upward and downward “trends.” This is part of the predictable behaviour of randomness. Are there other predictable features to randomness? We mention stochastic processes as a discipline in probability, whose purpose is to discover features such as the notion of returns to origins, to further whet student appetites.

A graph like that given in Figure 1 is a discrete curve and its study is pervasive in nonlinear and dynamical systems of chaos and fractals. Such curves are self-similar everywhere. This theme is further elaborated in the next example.

---

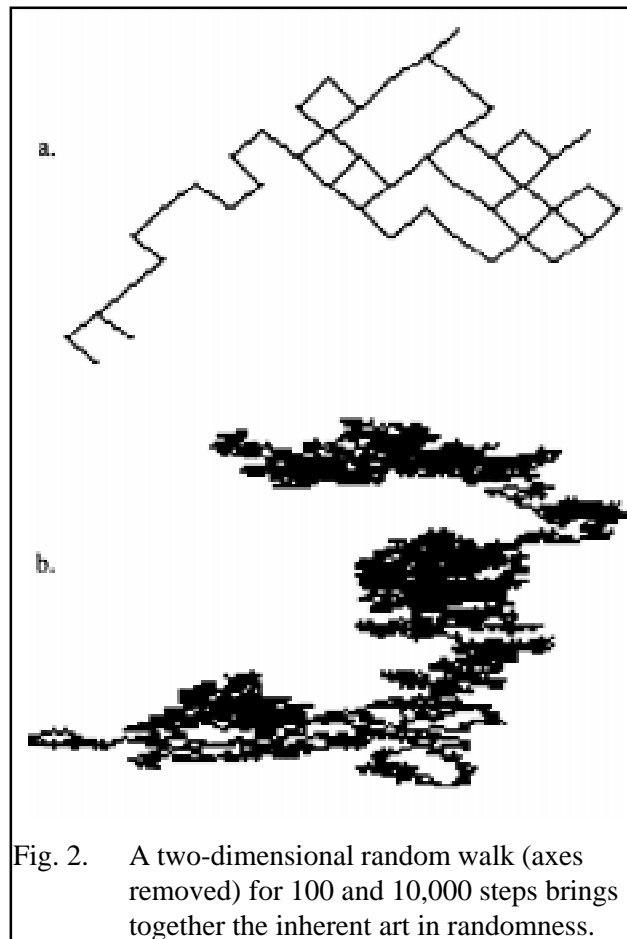
## ◆TWO-DIMENSIONAL ◆ RANDOM WALK

---

For our third example, we extend our discussions and experiments to higher dimensional random walks. The motivation may still be finance related, e.g. observing a stock and a bond index displaying random up or down behaviour at a given time, or in physical sciences where migration of molecules from motion due to thermal heat may be of interest. For such cases the outcome of each trial is represented by an x and y coordinate and we are interested in the walk produced by joining successive outcomes. Figures 2a and 2b represent outcomes for 100 and 100000 trials produced through MATHEMATICA

```
ListPlot[FoldList[Plus, { 0,0} ,Table[{  
2*Random[Integer,{0,1}]-  
1,2*Random[Integer,{0,1}]-1},{100000}]],  
PlotJoined->True, PlotStyle-> {  
RGBColor[1,0,0]}, Axes->False]
```

Random walks in three-dimensional space are used by biologists to study movement of protein molecules in aqueous medium, effectiveness of drug delivery systems within the body and other such systems. A very clear account of applications of random walks in various dimensions with biological applications involving diffusion and diffusion with drift can be found in Berg (1993). Raup (1991) gives many interesting applications of



random walks with reflecting and absorbing barriers through models in palaeontology. For example, if a genus has a stock of 10 species and adds or loses a species with equal probability every million years, a small time in the scale of geological time, how are we to understand the statistical characteristics of this genus becoming extinct? Simulation experiments with coin tosses produce interesting facts about the many possibilities of history that the genus may have including the stark fact that it dies with certainty, no matter how large a stock of species the genus has to begin with!

A two dimensional random walk produces a fractal which promotes a discussion of art and science and the place of randomness in them. Typically, we discuss the work of Mandelbrot (1982) and various applications of non-linear mathematics (Gleick, 1987, Chatterjee and Yilmaz, 1992) and share with students the beautiful books of Mandelbrot (1982), Pietgen and Richter (1988), Pickover (1990) and others. This interaction with students after an experiment with relevant materials has helped us to get students interested in the course material, see the complexity involved with simple randomness and relate the study of statistics to real world events.

## ◆ APPARENT ORDER AND DISORDER ◆

Our final example does not explicitly relate to random walks but continues the theme of human perceptions of order and disorder. Randomness and order, according to most students in an introductory statistics course are obvious and easy to discern. This module illustrates such a dichotomy is not so obvious and in fact may be a truly difficult and counter-intuitive task. Our earlier examples demonstrate this counter-intuitive fact that patterns can emerge from an apparently disorderly process.

The night sky looks random to an astronomer, yet it is a well known fact that space is mostly vacuous with clumped matter such as galaxies. So, to an earthly observer, the sky looks random yet it is anything but random. Note the pattern we see in the sky is due to relative position with regard to the actual constellations.

Figure 3a plots points generated by the function  $\sin(100x)$ ,  $\{x, 1, 100\}$  and in Figure 3b the same points are joined. We ask students about the randomness of Figure 3a before showing them Figure 3b. The students invariably see randomness in Figure 3a. Figure 3c shows us a set of random points generated by MATHEMATICA for comparison purpose.

Finally, Figures 3d and 3e give plots with 500 points generated from the same function. Students usually have little difficulty in seeing order in Figure 3d.

Professor Stephen Jay Gould, the great evolutionary biologist, gives an excellent illustration of apparent order and disorder in his description of the glow of glow-worm grotto of Waitomo cave in New Zealand (1991). He also provides computer simulations which are reproduced in Figure 4a and 4b. Figure 4b is produced by insertion of dots at random while Figure 4a is produced by a fixed law (ordered by fields of inhibition around each dot). Gould describes these two pictures as:

*the curious psychological effect. Most of us would see order in the strings and clumps of the figure on the right, and would interpret the figure on the left, with its lack of apparent pattern, as random. In fact the opposite is true, and our ordinary conceptions are faulty."*

The class discussion of these concepts leaves the students with plenty to think about: order, randomness and the delicate boundary between the two and the human perception of all three.

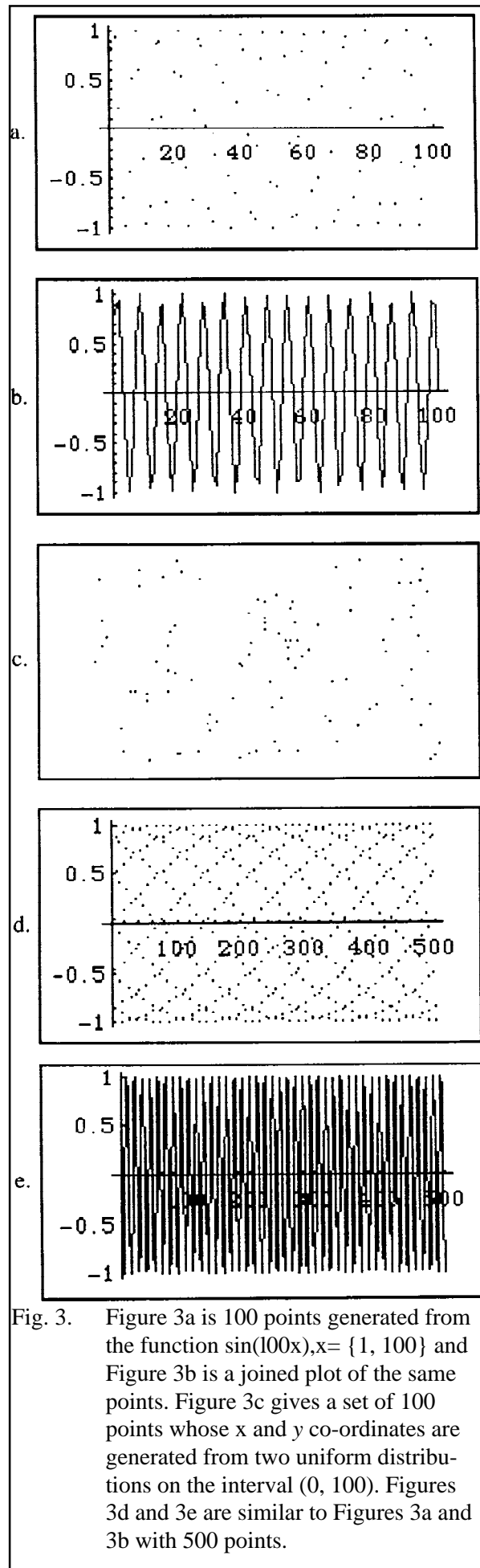
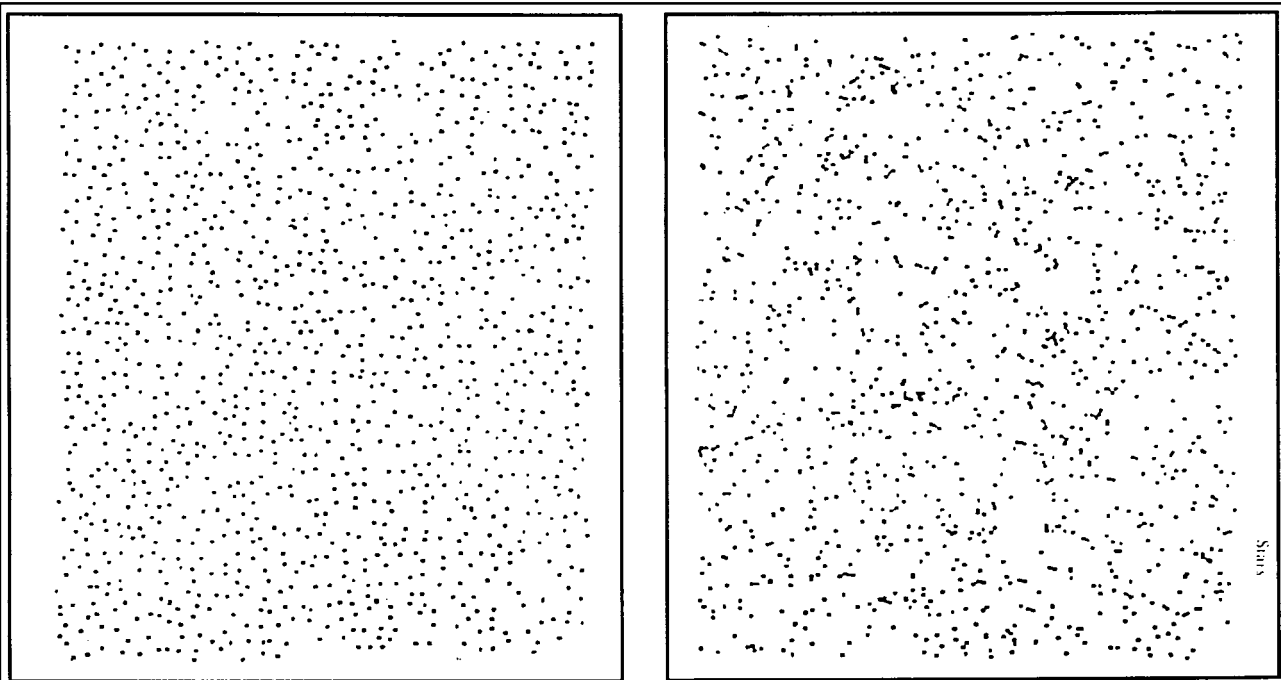


Fig. 3. Figure 3a is 100 points generated from the function  $\sin(100x)$ ,  $x = \{1, 100\}$  and Figure 3b is a joined plot of the same points. Figure 3c gives a set of 100 points whose  $x$  and  $y$  co-ordinates are generated from two uniform distributions on the interval  $(0, 100)$ . Figures 3d and 3e are similar to Figures 3a and 3b with 500 points.



a.

b.

Fig. 4. Figure 4b is a random collection of points generated from two independent uniform distributions for their x and y co-ordinates respectively. Figure 4a is a plot of the same number of points generated through a fixed law, ordered by fields of inhibition around each dot (from a random start). Most students see a pattern in Figure 4b and randomness in Figure 4a while in reality the exact opposite is true.

## ◆ CONCLUSIONS ◆

We have presented four illustrations of randomness that are somewhat counter-intuitive and that students embrace with enthusiasm. The examples are connected through the common theme of perceptions of random walk in various situations. Other examples include understanding the control of an industrial process through Deming's (1989) funnel experiment, observation of the nature of variation of a stable process through Deming's bead experiment (Deming 1989), examples of Poisson processes from events in daily life, and Simpson's paradox. A technical report describing these experiments can be requested from the first author. Much of statistics appeals to the intuition. However, there are a number of important ideas that are counter-intuitive. We feel that most elementary texts do not address these counter-intuitive notions in the way that they should. Our purpose in writing this article was to provide some interesting problems that elementary students could understand and so appreciate the subtleties of statistics.

### References

- Berg, H. C. (1993). *Random Walks in Biology*. Princeton University Press, Princeton, New Jersey.
- Chatterjee, S. and Yilmaz, M. R. (1992), Chaos, Fractal and Statistics (with comments). *Statistical Science*, 7, 49-121
- Deming, W. Edwards. (1989). *Out of the Crisis*. MIT Press, Cambridge, Massachusetts.
- Gleick, J. (1987). *Chaos: Making of a New Science*. Viking, New York.
- Gould, S. J. (1991). Glow, Big Glowworm in *Bully for the Brontosaurus*. W. W. Norton and Company, New York.
- Hawkes, J. (1993). *Discovering Statistics and Adventure in Problem Solving*. Quant Publishing, Charleston, S. C.
- Mandelbrot, B. (1982). *The Fractal Geometry of Nature*. W. H. Freeman, San Francisco.
- Piatgen, H. O. and Richter, P. H. (1988). *The Beauty of Fractals*. Springer, New York
- Pickover, C. A. (1990). *Computers, pattern, chaos and Beauty*. St. Martin's Press, New York
- Raup, D. R. (1991). *Extinction*. W. W. Norton and Company, New York.
- Schaefer, R. L. and Farber, E. (1989). *The Student Edition of Minitab*. Addison-Wesley, Reading, MA.
- Tanur, J. M., Mosteller, F. M., Lehman, E. L., Link, R. F., Pieters, R. S. and Rising, G. R. (1989). *Statistics: A Guide to the Unknown* Wordsworth & Brooks/Cole, Pacific Grove, California.
- Wolfram, S. (1991). *MATHEMATICA, A System for Doing Mathematics*. Addison-Wesley, MA.