

Teaching Basic Probability and Statistics Using Computer Graphics

Henk Tijms - Amsterdam, The Netherlands

1. Introduction

In the physical sciences it has been long known that the most effective way to teach students is by having them "discover" the basic principles themselves. Unfortunately, probability and statistics have traditionally not been taught this way. With the advent of the personal computer, all that has changed. Microcomputers can do more than take the drudgery out of calculations. Their unique graphics and interactive capabilities can be used as well to teach basic concepts and ideas. Many of the basic ideas in probability and statistics seem exceedingly difficult for most students to grasp. It is extremely important to give students a sound intuitive feeling for basic concepts such as randomness and the normal curve before teaching them formal statistical theory. The micro makes it now possible to add a laboratory component to the teaching of probability and statistics. Key concepts such as the law of large numbers, random walks, and the central limit theorem can be made to come alive before one's eyes through computer graphics. Direct experience and actual experimentation is the best way the student can obtain a feeling for these basic concepts. It cannot be emphasised enough how important it is that students obtain at an early stage a feeling for probabilistic reasoning. People are not born with a natural feeling for probabilistic thinking, but it is of utmost importance in everyday life. Managers in industry realise more and more that without probabilistic thinking many of their problems cannot be adequately solved.

The micro is the ideal tool for students to develop a sound probabilistic intuition. Its most powerful feature - computer graphics - should be exploited. Most people find it much easier to grasp concepts in visual terms and through direct experiences than in terms of formulae. Of course, many authors have advocated the use of computer graphics to support the teaching of basic concepts in probability and statistics, but only a few have actually done the painful work of implementing their ideas in user-friendly software for the beginning student. The educational software

packages of Bowman and Robinson (1987) and Fisch and Griffeath (1988) should be mentioned. This paper discusses graphical software which has been developed in Kalvelagen and Tijms (1990), and is designed to introduce the beginning student in a motivating and coherent way to the following very basic concepts.

- (i) *The law of large numbers.* Computer animations of the experiment of coin and dice tossing is the best way to obtain insight in random fluctuations. Using an interactive simulation program it is possible to fight misconceptions that even short runs of the coin-tossing experiment should reflect the theoretical 50:50 ratio of heads to tails.
- (ii) *Random walks.* The graphical display of random walks provides the student with insight into random fluctuations. Showing the actual number of heads minus the expected number in the coin-tossing experiment is very instructive. Using this random walk, a natural link can be made with the central limit theorem.
- (iii) *The central limit theorem.* A visual demonstration of this very important concept in probability and statistics can be given. The student can interactively generate histograms for different sample sizes and see how fast the histogram approaches the normal curve as the sample size gets larger.

2. Law of large numbers

Many people expect that even short runs of the coin-tossing experiment should reflect the theoretical 50:50 ratio of heads to tails. This misconception is known as the gambler's fallacy since some gamblers feel that the probability of tails becomes larger after a run of heads. Mere exposure to the theory may not be sufficient to overcome such misconceptions. However, an interactive simulation program on the micro-computer is ideally suited to fight misconceptions such as the gambler's fallacy.

Computer animation of the experiment of coin and dice tossing is the best way to obtain a feeling for randomness. Within a few seconds students can simulate and repeat this statistical experiment on the micro. By looking at the graphical representation of the results, they see what randomness means. By doing the experiment of tossing a fair coin and observing the relative frequency at which heads appears, the student will see that this relative frequency may significantly differ from the value $1/2$ after a large number of tosses. As the number of tosses grows, the relative frequency will eventually approach the value $1/2$ according to the law of large numbers. To see before one's eyes this experimental demonstration of the law of large numbers is very instructive. Also, it is instructive to see that the relative frequency typically approaches the value of $1/2$ in a rather irregular way. To make the above concepts live, we developed in Kalvelagen and Tijms (1990) a software module that simulates the experiment of rolling a die. The program gives the user the option of using either a fair die or a loaded die. For the case of a loaded die, the user has to specify the probabilities of each of the six possible outcomes. By assigning positive probabilities to only two outcomes, the coin-tossing experiment is a special case of the die-tossing experiment. Once the data have been specified, a computer animation of the die-tossing experiment is given in real time. The computer output also shows the relative frequency of heads after n tosses for a number of intermediate values of n .

The software for the die-tossing experiment is elementary, but it is very instructive. Students like to do the experiment themselves, and in this way they learn what randomness means. Much can sometimes be achieved by simple means!

3. Random walks

Let us consider the coin-tossing experiment using a fair coin. Many people erroneously believe that a run of heads should be followed by a run of tails so that heads and tails even out. In placing their bets many gamblers use some system that takes into account any imbalance between the past number of heads and tails. It is illusory to think that such a system can be of any help. The coin has no memory. Suppose a fair coin has been tossed 100 times and has landed heads 60 times. Then in the next 100 tosses the absolute difference between the number of heads and tails may still increase, while the relative difference decreases. For example, this occurs when heads appears 51 times in the next 100 tosses. The only thing one can be sure of is that the relative frequencies of heads and tails will eventually be equal, but there is not something like a law of averages for the absolute difference between heads and tails. In fact, the absolute difference between the number of heads and tails tends to increase as the number of tosses gets larger. This surprising fact can convincingly be shown by simulation and a graphical display of the results on the screen. For a simulation experiment consisting of 2000 tosses of a fair coin, Figure 1 displays the graph of the random walk giving the actual number of heads minus the expected number.

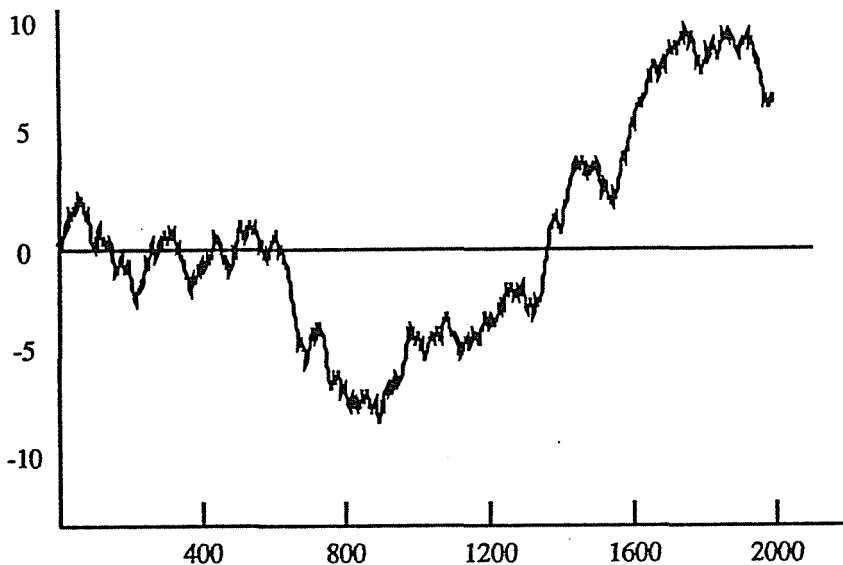


FIGURE 1

A random walk for the coin-tossing experiment

By trying different simulation runs of various lengths on the micro themselves, the students will learn by experience that realisations such as in Figure 1 are not an exception, but are typical for the coin-tossing experiment. The longer the run, the bigger and bigger the waves of the random walk and the relatively more rare the crossings of the x-axis. This finding is rather counterintuitive, but can be mathematically explained from the central limit theorem. However, before turning to this basic theorem, it is important that the student first learns by direct experience that a series of independent probability trials tends to behave better and better in the average sense but wilder and wilder in the absolute sense as the number of trials increases. This is a basic lesson in probability!

Once the students have learned by experimentation that the random walk tends to exhibit bigger and bigger fluctuations as the number of tosses grows, they may be ready to learn more about the mathematical explanation of this behaviour. At this point a natural link can be made with the central limit theorem. For the moment, taking the central limit theorem for granted, how do we explain mathematically the behaviour exhibited in Figure 1? Defining the random variable $X_i = 1$ if the i^{th} toss gives heads and $X_i = 0$ otherwise, the actual number of heads minus the expected number can be represented as $X_1 + \dots + X_n - n\mu$, where $\mu = 1/2$ denotes the mean of the X_i 's. Since the normal distribution has about 68% of its mass within one standard deviation of the mean, a simple consequence of the central limit theorem is that

$$P\{|X_1 + \dots + X_n - n\mu| > \sigma\sqrt{n}\} \approx 0.32 \quad \text{for } n \text{ large enough,}$$

where $\sigma = 1/2$ denotes the standard deviation of the X_i 's. In words, the probability that after n tosses the random walk $X_1 + \dots + X_n - n\mu$ will take on a value larger than $\sigma\sqrt{n}$ is about 32% when n is large. This gives the mathematical explanation that the absolute difference between the number of heads and tails tends to increase as the number of tosses grows. This result is not contradictory to the fact that the difference between the relative frequencies of heads and tails tends to zero, since the absolute difference between the number of heads and tails roughly tends to increase proportionally to the square root of the number of tosses. This important but subtle point is best understood through experimental studies using computer graphics.

4. Central limit theorem

The central limit theorem is the most important result in probability and statistics. Every beginning student should know about it. Using the random walk discussed above, the student's interest could be aroused in this theorem, but how to explain it? The central limit theorem is extremely difficult to prove. Moreover, the proof will not substantially help the student to understand the working of the theorem. It will not give the student a clear insight into how large n should be before the sum $X_1 + \dots + X_n$ of n independent and identically distributed random variables X_1, \dots, X_n is approximately normally distributed. An intuitive feeling for the central limit theorem is best obtained through experimental studies based on computer graphics.

A possible way to demonstrate graphically the central limit theorem is to use computer simulation. In an interactive way many samples may be drawn from a given

probability distribution and a graphical display of the probability histogram of the sum may be given for different sizes. It will then be seen that the probability histogram gets close to the bell-shaped form of the normal density as the sample size increases. The drawback of the simulation approach is that for any fixed sample size n , many observations of the sum are needed before the simulated probability density $X_1 + \dots + X_n$ is sufficiently close to the true density. How many observations are needed is often not clear. Hence the simulation approach has the drawback that the law of large numbers interferes with the central limit theorem. To avoid this complication, we advocate an analytical approach in combination with computer graphics. A restriction to discrete probability distributions allows us to compute analytically the true probability density of the sum $X_1 + \dots + X_n$ for any fixed value of n . How to do this can be easily explained to the beginning student (cf. Kalvelagen and Tijms, 1990). An illustrative case is obtained by using the good old die, and taking

$$X_k = \text{the number of points shown at the } k^{\text{th}} \text{ rolling of a die.}$$

The sum $X_1 + \dots + X_n$ then represents the total score when the die is rolled n times.

The software module developed in Kalvelagen and Tijms (1990) gives the student the option of using a fair die or a loaded die. For a loaded die, the student has to specify for $j = 1, \dots, 6$ the probability $p(j)$ of getting j points at any given throw of the die. For example, by choosing $p(1) = p$ and $p(2) = 1-p$ for some $0 < p < 1$, the student has the option to verify experimentally that the binomial probability density tends to the normal curve as the number of trials gets large.

Once the underlying probability distribution, $\{p(j)\}$, of the die and the number, n , of throws have been specified, the computer program calculates the probability density of the sum $X_1 + \dots + X_n$ and displays the graph of this density on the screen. A glance at the screen is sufficient to see whether this density has the bell-shaped form of the normal curve or not. Since the probabilities $p(j)$ of the die can be varied, students can discover themselves that how large n should be before the probability density of the sum $X_1 + \dots + X_n$ is close to the normal curve, depends very much on the degree of asymmetry in the underlying probability distribution of the die. It is quite instructive to find out that the more asymmetric the underlying distribution, the larger n should be. It is really fun to learn about the central limit theorem in this experimental way.

References

- Bowman, A W and Robinson, D R (1987) *Introduction to Statistics*. Adam Hilger, Bristol, England.
- Fisch, B and Griffeath, D (1988) *Graphical Aids for Stochastic Processes*. Wadsworth and Brooks/Cole, California.
- Kalvelagen, E and Tijms, H C (1990) *Exploring Operations Research and Statistics in the Micro Lab*. Prentice-Hall, Englewood Cliffs, New Jersey.