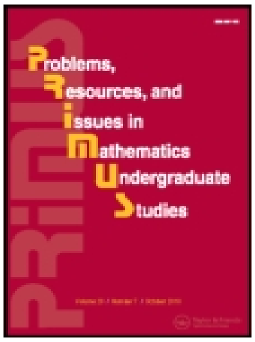


Problems, Resources, and Issues in Mathematics Undergraduate Studies

ISSN: 1051-1970 (Print) 1935-4053 (Online) Journal homepage: https://www.tandfonline.com/loi/upri20

# COMBINING DATA ANALYSIS AND PROBABILITY IN PRECALCULUS

Florence S. Gordon & Sheldon P. Gordon

**To cite this article:** Florence S. Gordon & Sheldon P. Gordon (1994) COMBINING DATA ANALYSIS AND PROBABILITY IN PRECALCULUS, Problems, Resources, and Issues in Mathematics Undergraduate Studies, 4:2, 109-116, DOI: 10.1080/10511979408965739

**To link to this article:** https://doi.org/10.1080/10511979408965739

Published online: 13 Aug 2007.

# COMBINING DATA ANALYSIS AND PROBABILITY IN PRECALCULUS

Florence S. Gordon[1] and Sheldon P. Gordon[2]

ADDRESS: (1) Department of Mathematics, New York Institute of Technology Old Westbury NY 11568 USA and (2) Department of Mathematics, Suffolk Community College Selden NY 11784USA.

ABSTRACT: In this article, we describe some details on an innovative alternative to standard precalculus courses. This course incorporates, among other things, ideas and applications of data analysis and probability in the process of preparing students for calculus. The article gives some examples of the types of applications and mathematical principles that are covered.

KEYWORDS: Precalculus reform, data analysis, curve fitting, probability, Monte Carlo simulations.

Over the years, the evolution of the mathematics curriculum has resulted in a fragmentation of the field of *mathematics* into small semester-long chunks known as courses. In the process, students see almost no interplay among different pieces of the mathematical mosaic; rather, they see mathematics as a long passageway with a series of doorways leading into seemingly unrelated chambers off at the sides.

The Math Modeling/PreCalculus Reform Project [1] is intended to develop an alternative to standard precalculus/college algebra and trigonometry courses based on the applicability of mathematics. In the process, we are attempting to broaden students' exposure to mathematics by presenting material from a wide variety of useful and interesting mathematical specialties in a unified way. Two of the major themes interwoven throughout the course are data analysis and probability.

Schultz and Rubenstein [5] make a strong case for integrating statistical ideas throughout the mathematics curriculum. One of the most effective ways for achieving this goal is through the use of random simulations of a variety of mathematical processes. Such simulations not only provide a vehicle for introducing statistical ideas into mathematics, but they also provide a very different and, in some ways, more practical and experimental perspective on the mathematics itself. They also provide the opportunity to make mathematics a more exciting and dynamic experience for the students. In a previous article [3], the present authors described a series of random simulations that can be introduced in calculus to bring a probabilistic perspective to many of the standard topics in the course. In [4], the authors describe the advantages of using random simulations in conjunction with presenting a treatment of radioactive decay to demonstrate the stochastic nature of the process.

In the present article, we will illustrate some specific examples of data analysis and probability that directly support the preparation of students for calculus. To begin, our firm belief is that probabilistic ideas, while a cornerstone of modern mathematics and an area that students must become accustomed to using routinely, should not be introduced in the formal settings that are all too usual. The standard marbles-in-urns problems do not provide insight; they promote confusion because too many of the results are so counterintuitive. With our approach, students become familiar with probability by seeing it applied to study phenomena about which they want to determine some information. The probabilistic results, via simulations, provide guidance about what happens in reality. Probability is introduced very early in the course via Monte Carlo simulations to estimate the value of $\pi$ and such simulations are used frequently at an informal level to provide information about the likelihood of events occurring. In the process, the students become comfortable with the idea of using probabilistic reasoning. At a later stage in the course, we present a more formal treatment of probability including an introduction to the binomial distribution (to motivate the binomial formula) and geometric probability.

For example, consider the question of the nature of the roots of a quadratic equation, $ax^2+bx+c=0$. The standard approach is to point out that the nature of the roots is determined by the sign of the discriminant, and apply the criterion to a dozen scattershot problems. A more interesting approach is to raise the question: how likely is it for a quadratic equation to have a pair of complex conjugate roots, or a double root? This can be answered fairly easily provided the coefficients are integers. We apply a simple computer program that uses ranges of integer values on the three coefficients, $a$, $b$ and $c$, as input and which then tests the resulting discriminant in each possible case. The students are asked to do some of this by hand to validate the results of the computer; of course, it also provides them with some skill development and reinforcement. But, this work is now done in the service of answering a question that interests them, not just for the sake of practicing a skill.

What if the coefficients are not integers? We talk about the *likely* effects of slight perturbations on the integer coefficients. We also apply a slightly more sophisticated version of the computer program described in the previous paragraph which generates any desired number of random sets of coefficients for either quadratic or cubic equation within any given set of ranges, analyzes the nature of the roots of each quadratic or cubic, and summarizes the results obtained. We have transformed the mathematics into an experimental activity with the computer providing a set of experimental data. We emphasize the fact that the results obtained represent only one single random sample; the percentages obtained are not precise, but rather represent a probabilistic assessment of the actual percentages of real and complex roots. We can demonstrate, through repeated runs of the program, that the results are reproducible in the sense that roughly the same percentages will occur on each run using the same set of parameters. For instance, with all coefficients positive and uniformly distributed, approximately 95% of all cubics have complex roots. (For students whose previous experience in algebra suggests that complex numbers are extremely rare, this comes as a shock. The unexpected frequency with which such roots occur also provides excellent motivation later in the course for the study of complex numbers.) We show some typical results for cubic equations $ax^3 + bx^2 + cx + d = 0$ with rational coefficients in the following table:

| Intervals for $a$, $b$, $c$, and $d$ | % of complex roots |
|---|---|
| all in [0,5] | 94.54% |
| all in [-3,3] | 78.43% |
| all in [-4,4] | 78.74% |
| all in [-5,5] | 78.93% |
| [0,4], [0,4], [0,4], [-4,0] | 88.40% |
| [0,4], [0,4], [-4,0], [0,4] | 74.8% |
| [0,4], [-4,0], [0,4], [0,4] | 88.4% |
| [0,4], [0,4], [-4,0], [-4,0] | 44% |
| [0,4], [-4,0], [-4,0], [0,4] | 44% |
| [0,4], [-4,0], [4,0], [-4,0] | 93.2% |

In addition, we point out to the students that the results of the probabilistic simulations with random coefficients are typically fairly close to the results with the same set of ranges for the coefficients in the integer case, so that the students see that the early discussion on perturbations makes sense. Consequently, the students see that Monte Carlo simulations can provide quite accurate information when one cannot get the answer directly.

Further, another feature of our course is the use of individualized computer investigations as described in [2]. For instance, each student is asked to use his or her social security number as the coefficients of an alternating polynomial and to determine the location of all real roots by a combination of graphical methods and the bisection method. It is fascinating to see how some of the students, entirely on their own, extend the ideas on the frequency of complex roots of quadratics and cubics to discuss the likelihood that an eighth degree polynomial will have real or complex zeros.

Let's consider another instance of probability in our course. Geometric probability provides an outstanding opportunity to introduce geometric ideas and develop some skills needed for calculus in settings that students find particularly interesting. Typically, if the random variable of interest is uniformly distributed, the probability of an event can be expressed as a ratio of two lengths, two time intervals, two areas or two volumes. For example:

> Suppose a 12" × 16" painting is mounted in a frame which is 2" wide and hangs on a wall. Someone is painting the ceiling, but doesn't bother to remove the picture. What is the probability that a drop of spattered paint lands on the picture (painting + frame), but doesn't hit the painting?

As an extension of this problem, consider:

> A painting whose area is 144 square inches is mounted in a 2" frame.
>
> a) Set up an expression in terms of the width $W$ of the painting for the probability that a random drop of paint that hits the picture (painting + frame) lands on the painting.
>
> b) Find the value of this probability for different widths $W$. Can you explain why the probability that the picture gets hit by the paint is small for some values of the width and larger for others?
>
> c) Plot the graph of the probability $P$ as a function of width $W$. From the graph, estimate the dimensions of the picture which has the greatest probability of being hit by a paint spatter.

With problems such as this, we set the stage for calculus in a variety of important ways.

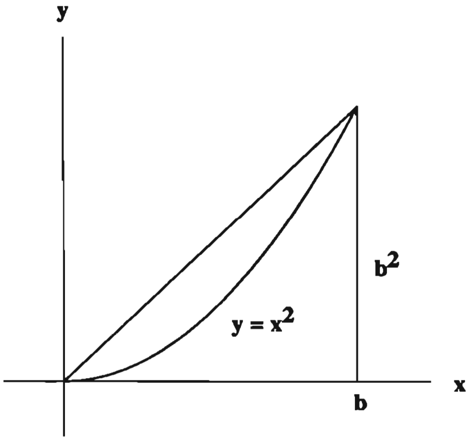


Figure 1. Area under a portion of the parabola $y = x^2$.

We ask questions which spark student interest. We emphasize the need to convert words into mathematics. We reinforce graphical ideas. We provide, in the probabilistic setting, teasers for the notions of calculus such as optimization in this case.

Geometric probability problems often require finding the ratio of two areas in the plane. While we could restrict our attention to simple regions that involve only (we hope) known facts from geometry, we decided that that would impose too great a limitation on what we could discuss. Instead, we treat the question of estimating the area of a region in the plane determined by a function using Monte Carlo simulations. Embed the desired region in an appropriate rectangle, generate large numbers of random points in the rectangle and keep track of what percentage of those points land in the desired region. The result, when the number of points is reasonably large, is usually quite a good approximation to the true area. One of the applications that these ideas allow us to address is the famous Buffon needle problem (see, for instance, [6]) which requires knowing the area under one arch of the sine curve.

The following is another interesting application of these ideas. Suppose we want to estimate the area under a portion of the parabola $y = x^2$ from 0 to any number $b$. As shown in Figure 1, the desired area $A$, when compared to the area of the circumscribing triangle, satisfies:

$$A < \frac{1}{2} b \cdot b^2 = \frac{1}{2} b^3.$$

We now take some specific values for $b$, say $b = 1, 2, 3, 4$ and 5 and apply a Monte Carlo simulation to estimate the area of each of the corresponding regions. Using $n = 2000$ random points in each circumscribing rectangle, we obtain the following sets of estimates on one set of runs with a computer:

| b | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Area | .33123 | 2.69264 | 8.99911 | 21.42751 | 41.38991 |

We can now apply some ideas on data analysis and nonlinear curve fitting that have already been introduced earlier in the course to determine the best fit to this set of data. The pattern clearly is growing faster than linear (look at the successive differences of the terms), but more slowly than exponential (look at the successive ratios of the terms). This suggests that the best fit may be a power function. Using the statistical routines of the TI-81, for example, we quickly find that the best power function fit to this set of data is

$$A = .333197 \cdot b^{3.0006}$$

with a correlation coefficient of $r = .999992$. This suggests that the exact formula for the area under the curve is likely to be $A = \frac{1}{3}x^3$.

Running a few more sets of random simulations can verify this conclusion. However, a useful device is to have each of the students conduct the same experiment in a computer lab by generating his or her own sets of data and performing the associated statistical analysis individually. The results of each student's calculations on the values of the parameters can then be tabulated on the board. This demonstrates several things. First, the students see the consistency in the results. Second, they see how the mean tends to average out the variations to produce a more accurate answer.

We note that comparable "derivations" can be performed with the area under any power function curve. A particularly effective second case (it is actually a problem given by the first author on a final exam in the course in Spring, 1993) is to treat $y = \sqrt{x}$ since the associated region can be approximated well with both a lower estimate $A > \frac{2}{3}x^{3/2}$ using an inscribed triangle and an upper estimate $A < 1 \cdot x^{3/2}$ using a circumscribed rectangle, as shown in Figure 2. The correct formula is, of course, $A = \frac{2}{3}x^{3/2}$.

Similar experiments could be conducted with other classes of functions, but the results, unfortunately, do not work out quite as simply. For instance, consider the function $f(x) = \cos x$. One sample series of Monte Carlo simulations for the area under the curve for $x$ corresponding to 15°, 30°, ..., 90° produced the following set of values where $x$ is measured in radians:

| x | .2618 | .5236 | .7854 | 1.0472 | 1.3090 | 1.5708 |
|---|---|---|---|---|---|---|
| Area | .259 | .499 | .703 | .859 | .974 | .991 |

An examination of these entries might suggest that they are approximately equal to the corresponding values of the sine function, at least while x remains relatively small. Repeated runs of the experiment, and especially averaging out the results of these runs, will further confirm this suspicion. However, there are no simple data analysis procedures available in our course for fitting trigonometric functions to data.

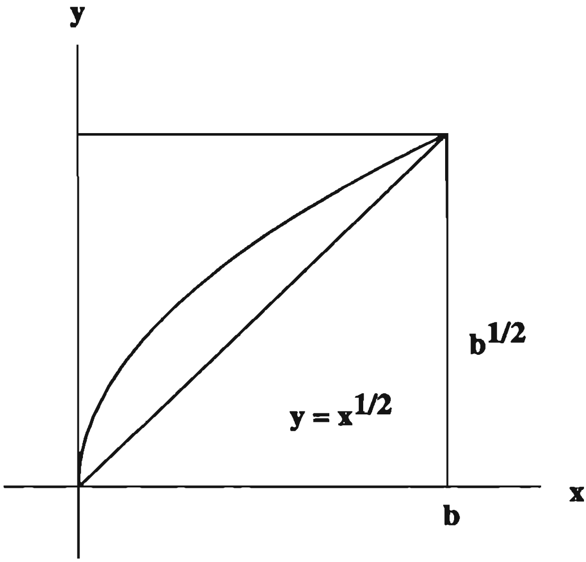


Figure 2. Area under a portion of the parabola $y = \sqrt{x} = x^{1/2}$.

The situation with exponential functions is, if anything, worse. The area under the curve $f(x) = b^x$ is given by

$$A = \frac{(b^x - 1)}{\ln b}$$

for any base $b$. As a consequence, a pure exponential function is not the most appropriate candidate to fit the data that would be produced - it gives a somewhat misleading answer. For instance, with $10^x$, the corresponding best fit among exponential functions typically turns out to be approximately $10.24^x/\ln 10$. The authors would be very happy to hear from any interested readers who can suggest a simple motivation that leads to the appropriate answer without knowledge of calculus.

## REFERENCES

1. Fusaro, B. A. and S. P. Gordon. 1992. Filling the Tank: Models for PreCalculus Reform. *UME Trends*. 4(August): 8.

2. Gordon, Sheldon P. 1992. Individualized Computer Investigations for Calculus. *College Math Journal*. 23: 426-429.

3. Gordon, Sheldon P. and Florence S. Gordon 1990. Applications of Monte Carlo Methods in Calculus. *The AMATYC Review*. 11: 9-14.

4. Gordon, Sheldon P. and Florence S. Gordon. Random Simulations of Radioactive Decay. *PRIMUS*. 3(3): 323-330.

5. Schultz, James E. and Rheta N. Rubenstein. 1990. Integrating Statistics into a Course on Functions. *Mathematics Teacher*. 83: 612-616.

6. Uspensky, J. V. 1937. *Introduction to Mathematical Probability*. New York: McGraw-Hill.

## ACKNOWLEDGEMENT

The work described in this article was supported in part by the National Science Foundation under grant #USE-91-50440 to the Math Modeling/PreCalculus Reform Project and grant #USE-89-53923 to the Harvard Calculus Reform project, as well as under a grant from New York Institute of Technology's NYIT/AAUP Faculty Research program.

## BIOGRAPHICAL SKETCHES

Shelly Gordon is professor of mathematics at Suffolk Community College. He is project director of the Math Modeling/PreCalculus Reform Project and is a member of the Harvard Calculus Reform Project working group. He is co-editor of the recent MAA Notes volume, *Statistics for the Twenty First Century*, and is particularly interested in the implications and applications of technology throughout the undergraduate mathematics curriculum.

Flo Gordon is professor of mathematics at New York Institute of Technology. She is a member of the Math Modeling/PreCalculus Reform Project and is responsible for developing the materials on data analysis and probability applications for the project. She is co-editor of *Statistics for the Twenty First Century* and co-author of the textbook, *Contemporary Statistics: A Computer Approach.*