Probability and Statistics in the Core Curriculum

David S. Moore[23]

J. Dossey (ed.), *Confronting the Core Curriculum*, Mathematical Association of America, 1997, 93–98

Introduction

Mathematics is distinguished from most other sciences by a lack of consensus on the content of an introductory overview for potential majors and other serious students. Chemists offer Chemistry 101–102, typically an introduction to the major branches of chemistry. We traditionally offer a year of calculus, followed in the second year by more calculus and perhaps some differential equations and linear algebra. This is hardly a balanced introduction to the nature and variety of mathematics. Attempts at reform have been common, and so has their failure. Table 1 shows one current reform proposal, the contents of a version of "Mathematics 101–102" developed with NSF funding (COMAP, 1997). There is *no* calculus, which is left for the second year. There is also no statistics, though probability does appear.

The absence of statistics in "Mathematics 101–102" should attract comment. After all, CUPM recommended in 1981 that "other mathematical sciences courses, such as computer science and applied probability and statistics, should be an integral part of the first two years of study." (See Steen 1989, page 5.) This suggestion has generally brought agreement in principle (though little action). I want to offer a partial disagreement in principle. I will argue that, whatever the merits of "Mathematics 101–102," its authors have done the right thing about probability and statistics. They have included the first and omitted the second. To my taste, they would have done well to exercise the same rigor with respect to computer science, in exchange for some continuous mathematics in the first college year. The point is this: a mathematics core ought to display to students the nature and variety of mathematics, including its applicability, but is not the place to develop the principles of related fields. Probability has an important place within mathematics. Statistics does not, and an attempt to include it will be disruptive.

[23]Department of Statistics, Purdue University, West Lafayette, IN 47907

MATH 101–102?

- CHANGE (sequences, difference equations, series)
- POSITION (vectors, analytic geometry)
- LINEAR ALGEBRA (matrices, eigenstuff, projections)
- COMBINATORICS
- GRAPHS AND ALGORITHMS
- ANALYSIS OF ALGORITHMS (time-complexity)
- LOGIC AND THE DESIGN OF "INTELLIGENT" DEVICES
- CHANCE
- MODERN ALGEBRA (groups, coding theory)

Table 1

Probability in a mathematics core

Probability has immediate attractions for a mathematics core. Chance phenomena are part of everyday experience and are important in the pure and applied sciences. Probability, the mathematical description of chance, is therefore especially attractive if *mathematical modeling* is one of the principles guiding the core curriculum. This is not true merely because probability models are interesting and have a wide field of application. Most areas of mathematics, when applied to modeling, describe deterministic behavior. It is intellectually stimulating to see how mathematics can also describe chance behavior. The chapter on "Some miscellaneous applications of simple probability" in Noble (1967) remains a good source of simple physical models. Areas such as learning (of rats, alas, not people), genetics, and transmission of rumors or disease are among the biological applications of probability modeling.

Moreover, probability tools that are both *discrete* and *elementary* are powerful enough to be interesting. Conditional probability and tree diagrams for multistage processes, though in simple settings they amount to little more than codified thinking about percentages, allow striking examples. Topics like finite Markov chains are only slightly further afield.

Example 1. ELISA tests are used to screen donated blood for the presence of HIV antibodies. When antibodies are present, ELISA is positive with probability about 0.997 and negative with probability 0.003. When the blood tested lacks HIV antibodies, ELISA gives a positive result with probability about 0.015 and a negative result with probability 0.985. (Because ELISA is designed to keep the AIDS virus out of blood supplies, the higher probability 0.015 of a false positive is acceptable in exchange for the low probability 0.003 of failing to detect contaminated blood. These probabilities depend on the expertise of the particular laboratory doing the test. The values given are based on a large national survey reported in Sloand et al. (1991).)

Now suppose that HIV screening is imposed on a large population, only 1% of which carry the antibody. Figure 1 displays the tree diagram of outcomes. We calculate easily that the probability that a person chosen at random from this population tests positive is 0.0249, and that the conditional probability that a person who tests positive has the antibody is 0.4016. That is, 60% of positive test results are false positives.

Even though ELISA is quite accurate, most positives are false positives when the test is applied to a population in which true positives are rare. Similar results hold for drug screening and lie detectors. Gastwirth (1987) offers a sophisticated treatment.

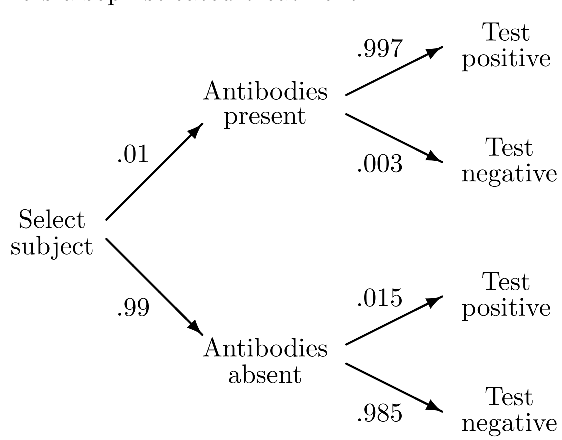

Figure 1

Probability also illustrates the *interconnections among subfields* that characterize contemporary mathematics. That is true at the advanced level, where probabilistic tools are important in areas such as number theory, PDEs, and harmonic analysis. But interconnections also appear in the first two years of college study, and are important to solving the "so much mathematics, so little time" problem. The use of combinatorics in calculating probabilities in symmetric settings is well known. Overemphasis on combinatorics has traditionally left students mystified by supposedly elementary probability, so I prefer to play down counting unless it will be studied and used elsewhere in the core. Calculations of continuous probabilities and expected values apply integration in ways that emphasize conceptual interpretations: probability is area under a curve, expected value is a weighted average of possible outcomes. A student who sees why expected value as an integral, $E[X] = \int xp(x)dx$, is the continuous analog of the discrete expected value as a sum, $E[X] = \sum x_i p(x_i)$ gains real insight into integration. Even discrete probability leads to the binomial theorem, geometric series, and other mathematical commonplaces. Finite Markov chains use the language and tools of matrix theory. And so on.

Probability also illustrates the *power of abstraction* in mathematics. The same rules describe all legitimate probability models, though the assignment of specific probabilities may vary greatly in nature and complexity. We can establish many facts once in general, than appeal to them in varied settings. If the use of an axiomatic approach is one of the themes of the core curriculum, the fact that all of general probability emanates from just three axioms is appealing. (I want to demonstrate the power of abstraction much more than I want to work from axioms. And in practice, we can deal axiomatically only with finite probability spaces, lest we meet the ugly fact that we can't assign probabilities to all sets of outcomes.)

Finally, probability lends itself to *computer simulation* as a tool for learning and for modeling. I believe that students ought to meet technology whenever "real" uses of the mathematics would employ technology. Simulation is one such setting. Simulations can demonstrate both the short-run unpredictability of random phenomena and the long-run regularity that probability describes. They allow study of problems that are simply stated but too hard for undergraduate analytical skills. (Probability abounds in such problems. Toss a balanced coin 10 times. What is the probability of getting a run of 3 or more consecutive heads?) Simulation can even help link probability to other topics in mathematics, through for example use of Monte Carlo methods to evaluate definite integrals.

There is room for differences of taste in selecting from this rich array of accessible material. My pedagogical taste runs to the concrete, and to developing mathematics in the context of applications. Simulation appeals to my taste for hands-on work and for technology. I would play down combinatorics and abstract general probability. Whatever our taste, interconnections among core topics are important both for efficiency and to illustrate the unity of mathematics. We should choose topics from probability (and from other subfields as well) with a view to the curriculum as a whole. That is, the first question to ask of any aspect of probability (or linear algebra, or calculus) is not how well it introduces the professional's view of that subfield, but how it contributes to an overview of mathematics in the context of selected aspects from other subfields.

Curriculum planners considering material from probability might seek inspiration from Snell (1988). Though aimed at upperclass students, the book is rich, concrete, makes heavy use of computing, and offers nice historical remarks. It is, however, a mathematician's book with little attention to modeling.

The trouble with probability

The trouble with probability is that it is conceptually the hardest subject in elementary mathematics. Psychologists, beginning with Tversky and Kahneman, have suggested that our intuition of chance profoundly contradicts the laws of probability that describe actual random behavior. They have also demonstrated that incorrect concepts remain firmly embedded in students who can correctly solve formal probability problems. See e.g. Tversky and Kahneman (1983) and the collection by Kapadia and Borovcnik (1991). Garfield and Ahlgren (1988) conclude a review by stating that "teaching a conceptual grasp of probability still appears to be a very difficult task, fraught with ambiguity and illusion."

We run the risk—no, we face the near certainty—that students will learn a formalism not accompanied by a substantial understanding of the behavior that the mathematics describes. Probability is the count of favorable outcomes divided by the count of all outcomes. Probability is area under a curve and can be found by integration. The record suggests that we are unlikely to move most students beyond that level of understanding.

One root of the trouble with probability is lack of experience with the long-term regularity that the mathematics purports to describe. Chance variation is familiar, but chance appears haphazard because we very rarely see the large number of similar trials needed for the emergence of regular patterns. It is not accidental that games of chance, which impose a structure of repeated independent trials, were the historical setting for Pascal and Fermat and have been a staple of teaching ever since. Simulation allows learners to gain some experience with long-run chance behavior. We ought to mix simulation and model building with the mathematics that so strongly appeals to us. We ought to note specific instances (such as the prevalence of runs and other "nonrandom" behavior in short sequences of random trials) in which our intuition fails. But we should also be aware in advance that, given the limited time available in a core curriculum for extended experience with chance behavior, a conceptual understanding of probability will elude many of our students.

The trouble with statistics

The trouble with statistics is that it is not mathematics. It is a discipline that (like economics or physics) makes heavy and essential use of mathematics but has its own subject matter. Many engineers and scientists will find a knowledge of statistics useful and will wish to study the subject. Most mathematics students should study some statistics as a quantitative tool that complements their mathematical training. Selected applications from statistics (or economics, or physics) can add richness to the mathematics core curriculum. But it is unfair to both mathematics and statistics to attempt a substantial treatment of a separate discipline in the mathematics core.

That bald statement reflects the self-understanding of most statisticians. It may surprise some mathematicians, who regard statistics as a (somewhat trivial) field of mathematics. Probabilists, specialists in the field of mathematics most applied in statistics, often know better—note David Aldous's (1994) saying that he "is interested in the applications of probability to all scientific fields *except statistics*." Let me outline the facts behind the position. Moore (1988) is a more polemical statement of the case.

Statistics is a methodological discipline, the science of inference from empirical data. Under the influence of computing, statistics research and (more slowly) instruction have in recent years returned to their roots in data and scientific inference. Here is the statistician's view of statistics. For more detail, see the essays in Hoaglin and Moore (1992).

Data analysis, the examination of data for interesting patterns and striking deviations from those patterns, is one of the main foci of contemporary statistics. Data analysis uses both an ample kit of clever tools and a clear strategy for exploring data, but it has no mathematical theory. Graphical display, usually automated and made interactive via software, is always the starting point. Numerical summaries and (sometimes) compact mathematical models follow. Data analysis is specific and concrete. As George Cobb likes to say, "In mathematics, context obscures structure. In data analysis, context provides meaning." In mathematics, abstraction often gets to the heart of the matter. In data analysis, abstraction strips away the details of a particular data set, and so hides the matters of greatest interest.

Designs for data production through sample surveys and experiments have long been a staple of statistics in practice. Their elementary principles are core content in statistics instruction, and their detailed elaborations provide employment for professional statisticians. Although one central idea—the deliberate use of chance selection in producing data—provides a basis for probability analysis, data production like data analysis is not an inherently mathematical subject.

Formal inference is the area of statistics that does have a mathematical theory, based on probability. In fact, inference has several competing theories. The domain of applicability of formal inference is more restricted than that of data analysis. How restricted is disputed. Because statistical inference is a formalization of inductive inference from data to an underlying population or process, it is full of conceptual difficulties and heated debates. The debates concern not the correctness of the mathematics, but the nature and scope of inferential reasoning. Statistical inference is based on mathematical models, but now places heavy emphasis on *diagnostics*, methods that allow data to criticize and even falsify models. The result in practice is a dialog between data and model that reflects the empirical spirit of data analysis. Here is a very brief example of the inadequacy of a mathematics-based approach to formal inference even when diagnostics and philosophy are left aside.

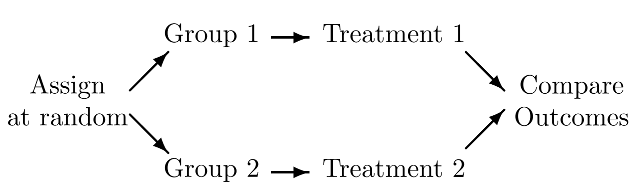

Figure 2

Example 2. A standard setting for elementary inference is the two-sample problem: two independent sets of observations are drawn from populations assumed to be normally distributed. We wish to compare (say) the mean responses μ_1 and μ_2 in the populations. The mathematical model is

$$X_1, X_2, \ldots, X_n \quad \text{iid} \quad N(\mu_1, \sigma_1)$$

$$Y_1, Y_2, \ldots, Y_m \quad \text{iid} \quad N(\mu_2, \sigma_2)$$

Formal inference is based on this model. But the model is radically incomplete. The model, and the formal inference, is the same for two independent samples from two populations and for data from a randomized comparative experiment. Yet the experiment (Figure 2) is intended to allow cause-and-effect conclusions, while an observational study cannot give convincing evidence of causation. The distinction between observation and experiment, and the reasoning of randomized comparative experiments, are among the most important topics in basic statistics. They are inherently statistical, with little mathematical content, and are out of place in a mathematics curriculum.

The American Statistical Association and the MAA have formed a joint committee to discuss the curriculum in elementary statistics. The recommendations of that group reflect the view of statistics just presented. Here are some excerpts (Cobb (1991).

> Almost any course in statistics can be improved by more emphasis on data and concepts, at the expense of less theory and fewer recipes. To the maximum extent feasible, calculations and graphics should be automated.
>
> Any introductory course should take as its main goal helping students to learn the basics of statistical thinking. [These include] the need for data, the importance of data production, the omnipresence of variability, the quantification and explanation of variability.

Data analysis, statistical graphics, data production, and even the somewhat arcane reasoning behind "statistical significance" are mismatched with the mathematical content needed by potential math majors. Yet "statistics" that ignores these topics isn't a responsible introduction to statistics. Statistics in a mathematics core curriculum is an oxymoron.

I should add at once that although mathematics can prosper without statistics, the converse fails. Bullock's

(1994) claim that "Many statisticians now insist that their subject is something quite apart from mathematics, so that statistics courses do not require any preparation in mathematics." draws a clearly false implication. Although the place of statistics in mathematics instruction may be marginal, the place of mathematics in statistics instruction remains central.

Applications of Mathematics in Statistics

The decision not to teach statistics for its own sake does not rule out applying mathematics to statistical problems. Consider, for example, the topic of *prediction.*

Example 3. Knowing which of several groups something belongs to can help predict its properties if the groups differ in the property we want to predict. For example, knowing that a hot dog is a "meat hot dog" or a "poultry hot dog" in the government classification helps predict how many calories the hot dog has. Given data on many brands of meat and poultry hot dogs, we find that the mean calorie count is about 160 for meat and 125 for poultry. With no other information, we might use the group mean as a prediction for an individual hot dog.

Interesting use of elementary mathematics arises from looking for *optimal* predictions. The mean is optimal if our criterion is to minimize the sum of the squares of the errors made. The median is optimal if we seek to minimize the sum of the absolute errors. The midrange is optimal if we wish to minimize the maximum error. There is no simple rule for the point that minimizes the median of the absolute or squared errors. This simple setting leads to: the idea of optimization by stated criteria; the fact that the optimal result can vary with the criterion; the fact that the solution may not be unique (the median often isn't) and may not have a simple expression; and of course the technique needed to minimize the criterion functions. Can students show by counterexample that the median, which minimizes the mean absolute error, does *not* minimize the median absolute error? Can they show by example for $n = 3$ that the least median of squares solution is radically unstable?

Example 4. Now suppose that we have more information on which to base a prediction. We have data on an explanatory variable x as well as on the response y we wish to predict. For example, we may want to use the height x from which a rubber ball is dropped to predict its rebound height y.

Plot the data. The graph shows an approximate straight line relationship—not perfect due to measurement error and other factors. If we draw a line through the data, we can use the fitted line to predict y from x. What line shall we draw? Ask the students to discuss criteria. Distances from points to a line are usually measured perpendicular to the line. But in this setting it is usual to use vertical deviations because we are predicting y. The least squares criterion (minimize the sum of the squared vertical deviations) leads by elementary calculus to recipes for the slope and intercept of the optimal line. Formulating the problem requires more thought than solving it. If we prefer to minimize the sum of the absolute vertical deviations, on the other hand, there is no closed-form solution. If we attempt to minimize the median of the absolute deviations, there is no simple recipe for the solution and the computations rapidly become infeasible.

These examples require little background in statistics; even the goal of prediction can be removed if the instructor wishes. They are also chosen to avoid probability, the branch of mathematics most often applied to statistics. For the purposes of a core mathematics curriculum, good applications in another discipline must be comprehensible without much grounding in that discipline. This is as true of applications to statistics as to physics or economics.

References

[1] Aldous, David (1994), "Triangulating the circle, at random," *American Mathematical Monthly*, 101, 223–233. The remark appears in the biographical note accompanying the paper.

[2] Bullock, James O. (1994), "Literacy in the language of mathematics," *American Mathematical Monthly*, 101, 735–743.

[3] Cobb, George (1991), "Teaching statistics: more data, less lecturing," *Amstat News*, December 1991.

[4] COMAP (1997), *Principles and Practice of Mathematics*, New York: Springer.

[5] Garfield, J. and Ahlgren, A. (1988), "Difficulties in learning basic concepts in probability and statistics: Implications for research," *Journal for Research in Mathematics Education*, 19, 44–63.

[6] Gastwirth, J. L. (1987), "The statistical precision of medical screening procedures," *Statistical Science*, 2, 213–222.

[7] Hoaglin, David C. and Moore, David S. (eds.) (1992), *Perspectives on Contemporary Statistics*, MAA Notes 21, Washington: MAA.

[8] Kapadia, R. and Borovcnik, M. (eds.) (1991), *Chance Encounters: Probability in Education*, Dordrecht: Kluwer.

[9] Moore, David S. (1988), "Should mathematicians teach statistics" (with discussion), *College Mathematics Journal*, **19**, 3–7.

[10] Noble, Ben (1967), *Applications of Undergraduate Mathematics in Engineering*, New York: Macmillan.

[11] Sloand, E. M. et al. (1991), "HIV testing: state of the art," *Journal of the American Medical Association*, 266, 2861–2866.

[12] Snell, J. Laurie (1988), *Introduction to Probability*, New York: Random House.

[13] Steen, Lynn Arthur, (ed.) 1989, *Reshaping College Mathematics*, Washington: MAA.

[14] Tversky, A. and Kahneman, D. (1983), "Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment," *Psychological Review*, 90, 293–315.