PRINCIPLES AND STRATEGIES IN TEACHING PROBABILITY

Talma Leviatan
Department of Mathematics, Beit Berl Teachers' College, Israel
talmale@post.tau.ac.il

We propose to teach tertiary probability focusing on general probabilistic principles that lead to general probabilistic problem-solving strategies. The method enhances theoretical understanding of general models and at the same time, supplies students with a toolbox for handling the extremely wide range of possible applications of probability theory. According to experience, the method is particularly effective when applied to prospective teachers, as it helps in building up self-confidence both as students and as future teachers, a confidence that may later be transferred to their students.

INTRODUCTION

Designing a Probability course is a dual challenge: Probabilistic ideas are unfamiliar and often non-intuitive for students, resulting in many misconceptions in students' understanding and application of the theory (see Kahneman, Slovic & Tverski, 1982; Leviatan 1998, 2002). In addition, students find it confusing that a wide range of possible applications follows each new topic and that no two problems seem to be alike.

Years of debate about these issues have resulted in the formulization of some general *probabilistic principles* that lead to a reasonable number of general problem solving strategies. Once a well-defined collection of strategies is compiled, a probability course can then be designed around a structured strategies "discovery" process (preferably by the students themselves, in the classroom setting). The strategies are used repetitively to solve problems of increasing complexity, with combination of strategies sometimes used to handle a single more complex problem.

In section 1 we present some of the theoretical principles, in section 2 we describe some resulting strategies, and finally we offer directions for future research.

SECTION 1: PROBABILISTIC PRINCIPLES

It is advisable to start probability courses (in particular ones designed for prospective teachers) with a historical account, presenting the various historical approaches to probability theory. The most important models presented in such an introduction are:

- Pascal & Fermat's model (applicable only to finite symmetric sample spaces): The probability that an event A will occur in a given experiment, is defined as $P(A) = \dfrac{\text{number of favorable outcomes } (\#A)}{\text{total number of outcomes of the experiment } (\#\Omega)}$
- Von Mises' (more general) model, which is based on many independent repetitions of an experiment: The probability of an event A is defined via the *frequency* of the repetitions that resulted with A, relative to the total number of repetitions.
- A theoretical discussion of the *characterizing properties* of probabilities of events, under both models, leads to the final historical step: The *axiomatic approach*.

The conclusion of such an account is that the general theoretical framework of a tertiary Probability course should be axiomatic, while in problem solving one naturally resorts to the above first two applicational approaches.

In this section we present some general probabilistic principles that lie behind successful problem solving and lead to general problem solving strategies. Identifying such principles makes the presentation of examples in class more transparent and demystifies the ways problem solving is approached.

- *The principle* of *multi-stage experiment*: to better understand a complex probabilistic experiment, *equivalently* describe it as if performed in stages where each stage is a simple probabilistic experiment.

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

[Note: $\#\Omega = \#\Omega_1 \times \cdots \times \#\Omega_k$ where $\#\Omega_i$ - the number of outcomes of stage *i* experiment. Also, computing $\#A_i$, where A_i - an event of stage *i*, is typically a simple task.]

Examples: An experiment of tossing 2 identical coins at once is probabilistically equivalent to two consecutive tossing of a single coin.

An experiment of drawing *n* balls together out of an urn is probabilistically equivalent to *n* draws, without replacement, one ball at a time (see an example below).

The same goes for *n* balls placed in compartments (occupancy problems), arrangement of *n* objects in a row (permutation problems), etc.

- *The multi-stage symmetry principle* (a special case of a more general symmetry principle): If the outcomes of each stage are equally likely then the outcomes of the whole multi-stage experiment are equally likely.

  [A very useful observation because identifying symmetry at each stage is usually straightforward.]
- *The principle of a multi-stage events*: A is called a multi-stage event if it can be expressed as $A = A_1 \times \cdots \times A_k$ where A_i is an event of stage *i* experiment, in which case $\#A = \#A_1 \times \cdots \times \#A_k$.

  Using complement/union operations, it is often possible to express a given event in terms of a finite number of simple multistage events.

  Example (can easily be generalized to the Binomial model): B - Exactly *3* vowels (a,e,i,o,u) in a *k* letter code selected at random. B is not a multistage event but we can express $B = B^{1,2,3} \cup B^{1,3,4} \cup \cdots$ as a union of $\binom{k}{3}$ disjoint multistage stage events: $B^{i,j,m}$ - vowels only in the *i,j,m* places.

  [$\#\Omega = 26 \times \cdots \times 26 = 26^k$ (equally likely), $\#B^{i,j,m} = 5^3(26-5)^{k-3}$.]

The next two more subtle principles are very useful and should always be kept in mind.

- *The focus principle*: It is often possible to focus on certain aspects of the experiment and disregard irrelevant aspects.

  Example: The probability that the *l*'th letter in a *k* digit code is a vowel is 5/26. All other letters in the code are irrelevant.
- *The principle of ordered/non-ordered sampling*: when sampling *k* elements without replacements out of a population, one often has the *probabilistic freedom* to choose between ordered and non-ordered sampling models.

  Example (can easily be generalized to the *Hypergeometric model*): *k*=3 balls are drawn without replacement out of an urn containing *n*=2 black (B) balls and *m*=4 white (W) balls. A – Exactly 2 white balls drawn. The ordered approach (see the first three principles) is to think equivalently of 3 balls drawn without replacement out of an urn containing 6 distinguishable balls B_1, B_2, W_1, W_2, W_3. A is not a multistage event, but it can be expressed as a union of 3 such events, yielding $P(A) = 3 \times {}^{2\times4\times3}\!/_{6\times5\times4}$. Note that this approach becomes "messy" for large values of *k,n,m*. Applying the non-ordered approach $\#\Omega = \binom{6}{3} = \frac{6\times5\times4}{3\times2\times1}, \#A = \binom{2}{1}\cdot\binom{4}{2} = \frac{2\times4\times3}{2\times1}$, yielding exactly the same result as in the ordered model approach.

A typical probability course offers many opportunities to apply these principles both while solving problems and while formulating general models.

SECTION 2: PROBLEM SOLVING STRATEGIES

We argue that the most efficient approach to problem solving is to formulate problem solving strategies. The probabilistic strategies offered here are easy to disseminate and their names are used as buzz words throughout the course. For a more detailed list of strategies, arranged in four layers, and more examples, see Leviatan, (2005)

- *First layer strategies* are based (directly or indirectly) on the axioms of probability. These seemingly simple strategies will prove to be very effective throughout the entire course.
  - *The additive strategy*: Decompose the event A into a union of simpler *disjoint events,* preferably *equally likely*, and *add* their individual probabilities.
  - *The complement strategy*: Check if the complementary event, $\overline{A}$, is simpler.
  - *The decomposition strategy*: Decompose a complicated event A into a finite union of *n* simpler events $A_1,\dots,A_n$ - not necessarily disjoint, and use the inclusion-exclusion formula $P(A) = S_1 - S_1 + S_1 - \dots \pm S_n$ (see e.g. Ross, 2009, and final example here).
- *Second layer strategies* are strongly related to the general principles of probability described in section 1:
  - *The Equivalent problem strategy*: Try to formulate a simpler equivalent problem.
    Example: The event A- Bob and Bill celebrate their birthdays on the same day, is equivalent to A'- Bob celebrates his birthday on, say, Jan 1. Thus P(A)=1/365.
    [The multistage principle is an important special case of this universal strategy.]
  - *The Symmetry strategy*: Check first for symmetries in the probability space.
    Example: A train arrives at the station every 60 minutes. Bob, Dan and Ron, each arrive at a random time. A - Bob waits longer than Dan. To compute P(A), one can avoid uniform continuous probability spaces. It suffices to observe the *symmetry* between the two persons to deduce P(A)=1/2. Similarly P(they arrive according to alphabetical order) = $1/3!=1/6$.
  - *The focus strategy*: When dealing with a complicated experiment focus on relevant elements of the experiment, ignore the rest.
    Example: *n*=9 knights sit randomly around a table. A - exactly $k = 0,1,\dots,4$ persons sit between King Arthur and Lancelot. To compute P(A) ignore all other 7 knights and focus on Arthur and Lancelot seated by a 9-seat table. Let King Arthur choose any seat, now Lancelot chooses a seat at random, thus $P(A) = 2/8$.

We will not present here 0th layer *graphical strategies*, or 3rd layer more advanced strategies. Note that often a solution of a more complex problem requires the use of several strategies, one after the other. We conclude with two such examples.

Example: 3 girls and 5 boys form four pairs at random. A- three mixed-sex pairs. The first two principles lead to $\#\Omega = \binom{8}{2}\cdot\binom{6}{2}\cdot\binom{4}{2}\cdot\binom{2}{2}$ (a 4-stage experiment). Note that this approach introduces an element of *order* between pairs which has to be taken into account also in the numerator (!), thus $\#A = \binom{5}{2}\cdot 4\cdot 3\cdot 3\cdot 2\cdot 2$, and $P(A) = \dfrac{\binom{5}{2}\cdot 4\cdot 3\cdot 3\cdot 2\cdot 2}{\binom{8}{2}\cdot\binom{6}{2}\cdot\binom{4}{2}\cdot\binom{2}{2}} = \dfrac{4}{7}$. A much simpler approach is to adopt the *focus strategy*: focus on the problem of 3 girls choosing their partners at random. A 3-stages approach yields $P(A) = \dfrac{5\cdot 4\cdot 3}{7\cdot 5\cdot 3} = \dfrac{4}{7}$.

Note that originally there was no element of order between pairs in the experiment, to use the multi-stage approach we introduced an element of order (see *order/ non-order principle*) which necessarily effects both denominator and numerator!

Example (The matching problem): Don Juan writes *n* letters to his current *n* girl friends. Absentmindedly, he puts the letters at random into the envelopes. What is the probability of A- Don Juan ends up with no girl friend? First use *the complement strategy*: $\overline{A}$ - at least one girl friend left. Then use the *decomposition* strategy to express $\overline{A} = A_1 \cup ... \cup A_n$ where A_i - letter number *i* gets into the right envelope. Using the *symmetry* strategy, note that $P(A_1) = \cdots = P(A_n)$, $P(A_i \cap A_j) = P(A_1 \cap A_2)$, $P(A_i \cap A_j \cap A_k) = P(A_1 \cap A_2 \cap A_3)$ etc. Using the *focus* strategy, note that $P(A_1) = \frac{1}{n}$, $P(A_1 \cap A_2) = \frac{1 \cdot 1}{n \cdot (n-1)}$, $P(A_1 \cap A_2 \cap A_3) = \frac{1 \cdot 1 \cdot 1}{n \cdot (n-1) \cdot (n-2)} \ldots$

Thus $S_1 = \sum_{i=1}^{n} P(A_i) = \binom{n}{1} \times \frac{1}{n} = \frac{1}{1!}$, $S_2 = \sum_{i<j} P(A_i \cap A_j) = \binom{n}{2} \times \frac{1}{n(n-1)} = \frac{1}{2!}$, $S_3 = \binom{n}{3} \times \frac{1}{n(n-1)(n-2)} = \frac{1}{3!}$, …

Using the inclusion-exclusion formula, $P(\overline{A}) = S_1 - S_2 + \cdots \pm S_n = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \cdots \pm \frac{1}{n!}$,

yielding finally: $P(A) = 1 - P(\overline{A}) = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots \pm \frac{1}{n!}$. [Note $P(A) \to 1/e$ as $n \to \infty$.]

CONCLUSIONS

- To be able to solve *non-routine problems* students should be supplied with a proper *toolbox*. A printed list of effective strategies developed and demonstrated throughout the course seems to do the trick.
- The strategies should be intimately connected to the basic mathematical concepts (rules and principles) of that subject, thus enhancing understanding of the new theoretical ideas.
- The strategies should appear as simple as possible. To solve a more complex problem, a combination of several strategies may be required.
- It is beneficial to expose the general principles that lie behind the strategies.

During the course we compile a list of all probabilistic principles and strategies (about 20) observed throughout the course, each exemplified by a simple "generic example". The principles and strategies, constantly discussed, supply a good starting point for approaching a new problem and for developing new models. It will be interesting to make a more systematic research on the effectiveness of the combined approach suggested here.

REFERENCES

Kahneman, D., Slovic, P., & Tverski, A. (1982) *Judgment under Uncertainty: Heuristics and Biases.* Cambridge University Press, New York.

Leviatan, T. (1998). Misconceptions, fallacies, and pitfalls in the teaching of probability, *Proceedings of ICM.*

Leviatan, T. (2002). On the use of paradoxes in the teaching of probability. *Proceedings of ICOTS-6.*

Leviatan, T. (2005). Teaching mathematics via problem solving: The case of probability. *Proceedings of EARCOME3.*

Ross, S. (2009). *A first course in Probability,* 8th edition. International edition publisher: Prentice Hall.