

ON THE USE OF PARADOXES IN THE TEACHING OF PROBABILITY

Talma Leviatan
Tel Aviv University
Israel

Paradoxes have played an important role in the development of mathematics, as they brought about clarification of basic concepts and the introduction of new approaches. Probability theory offers a large variety of Paradoxes. Some of them are (nowadays) interesting mainly from a historical point of view, as the theory has already been adapted to resolve them. Others actually hide common misconceptions in a very subtle and tricky way. Introducing Paradoxes in class carries potential danger: it may result in a feeling of insecurity when the conflict between the mathematical solution and the intuition (or between two seemingly correct mathematical solutions) seems unresolvable. On the other hand, properly introduced, Paradoxes can play a very useful role in the classroom as they serve as leverage to fruitful discussions, and provoke deeper thinking about the (not always intuitive) probabilistic ideas.

INTRODUCTION

It is widely accepted that the study of Probability/Statistics is an essential part of one's education. Since we are constantly surrounded, both in everyday life and in our professional life by uncertainties, there is no way of understanding the world around us without being able to analyze such situations. Probability/Statistics supplies the mathematical framework and tools for this task. In this paper, I will discuss some of the problems one encounters while trying to teach these subjects, and offer methods to overcome the difficulties - methods developed during many years of teaching these subjects to university students of Exact-Sciences and Engineering. Those are top-level students; for whom this is not the first advanced math course. So, planned carefully, it is possible to teach in a single semester (3+1 hours a week) a full probability course of the level of, say, Ross (1998), or Pitman (1993). The course will have to include, of course, the introduction of standard problem-solving techniques, and all the common probabilistic models. Is this satisfying?

One problem is that we don't supply students with means to handle non-routine problems. It can hardly be expected that a student who solved, say, 70-80 problems during the semester to develop sophisticated general *strategies for solving probabilistic problems*. In addition to standard techniques and models, students also need a good "toolbox". We should incorporate into the regular curriculum sessions in which we discuss general probabilistic principles and introduce various, more sophisticated strategies specific to probability problem solving (see Leviatan (2000a), and also Leviatan (2000b) which is a problem-solving book accompanied by strategies).

But this is not enough. We are all aware of the fact that probability theory is infested with many *misconceptions, fallacies and pitfalls*. To quote Laplace: "The mind, like the sense of sight, has its illusions, and just as touch corrects those of the latter, so thought and calculations correct the former... One of the great advantages of probability calculus is that it teaches us to distrust our first impressions". Dealing with them should be an integral and routine part of any probability course. Regular textbooks neglect to do so. Also, since the high-level probability courses are fast paced, the decision usually is to deal with a misconception or a pitfall *only if* it arises spontaneously during class. The justification is that students are bright; so all those problems will be resolved naturally after a rigorous probability course. Unfortunately, experience actually shows that if one does not deal specifically with those misconceptions, they don't dissolve just by studying the subject formally. The problem here is that students tend to accept whatever they are taught in mathematics classes as truth (once they understand it, of course) without always noticing that it actually conflicts with their previous intuitions. Since their intuitions remain unchallenged, when those students encounter, later on, a real life situation in which they have to make a probabilistic judgment, the chances are that they will use their previous intuitions rather than their formal studies. My experience therefore is that it is best to confront all these difficulties and clear the mind of potential conflicts.

In order to accomplish this mission, I have developed the following three-tier method.

- Before teaching a new topic students are often asked (in class) to fill in a questionnaire with 5-10 *judgment problems* dealing with the new concepts. Their individual and their collective answers serve as a basis to the introduction of the new subject (see Leviatan, 1998).
- The discussion in class is accompanied by carefully selected *probabilistic puzzles* (which I will later on call “Quasi-Paradoxes”).
- Finally the discussion of a subject often ends up with the introduction of a famous Classical *Paradox* related to the subject.

Both the Quasi-Paradoxes and the Paradoxes play an important role in this scheme. In addition to their role in clarifying the new concepts, giving better insight to the theory, removing all sorts of potential conflicts between intuition and theory, and helping form new and correct intuitions, they introduce a new dimension to the course. Students cannot remain passive; they are forced to be sharp and witty, to constantly test themselves and to learn to expect the unexpected. My experience is that, carefully planned, all of this can be done within the framework of a standard course, and moreover that students enjoy it.

PARADOXES AND QUASI-PARADOXES

We begin by clarifying the general notion of a Paradox. The following definitions are taken from the dictionary (Cassel Concise Dictionary): 1. A statement, view or doctrine contrary to received opinion. 2. An assertion seemingly absurd but really correct. 3. A self-contradictory or essentially false and absurd statement. 4. A phenomenon at variance with normal ideas of what is probable, natural, or possible. In other words we are *not* necessarily talking about the Logical notion of a Contradiction, which is a statement that is both True and False. As we all know, mathematics cannot contain contradictions. Any mathematical theory that resulted in Logical Paradoxes (Set Theory for example) had to be modified in order to rid itself of those Paradoxes. So, what do we call a Paradox in Probability? I will not attempt to give an exact definition here. Classically the title “Paradox” was attached to problems that could be solved by two (sometimes more) different probabilistic approaches, both seeming to follow routine procedures, but eventually leading to contradictory answers. We will give here some classical examples, show briefly how they can be resolved, and indicate how to use them in class.

Lately, as the circle of people involved (in one capacity or another) in teaching probability has largely expanded, there has been inflated use of the term “Paradox” in more popular probabilistic literature. Much in accordance with the above general dictionary definition, any seemingly self-contradictory statements or any result in probability that is contrary to common sense is now popularly called a paradox. Most of these (along with many other examples) are familiar to any probabilist. Many of them prove to be very useful examples in class, but since they play a different role in the process of learning and understanding the theory, I would like to make a distinction between the two categories, so to refer to them as Quasi-Paradoxes (QPs). Actually, there is a lot in common between those two categories. The purpose of both is “to arrest attention and provoke fresh thoughts” (Britannica). The first reaction in class to both is usually that of amusement mixed with surprise and confusion; and the underlying meaning of both can be revealed only by careful scrutiny. These properties make them desirable tools in the process of learning.

But the two categories play different roles in class. The introduction of QPs carries no potential danger, so in view of their usefulness it is advisable, in my opinion, to include them in probability courses of all levels. Though each such QP is important at a certain stage of the course, they will usually attract our attention only for a short time. They will serve their immediate purpose (usually to clarify a notion) and will soon be forgotten. In an advanced probability course, typically a 5-10 minute discussion in class will satisfactorily clarify the notion. “Classical” paradoxes are made of different stuff; they are more perplexing, they cannot be resolved in a 10- minute discussion, and they usually lead to heated arguments even after they are “resolved” in class. There is something irresistible about them, which lingers long after class is over. Their introduction to a class of not-so-sophisticated students carries the potential danger that they can lead to unnecessary confusion. But, on the other hand, their (careful) introduction in the course eventually leads to lines of inquiry that might otherwise have been neglected.

QUASI-PARADOXES - THEIR PLACE AND ROLE IN CLASS

In this category I include any result that is contrary to common sense and intuition. Probability theory is full of such results. The idea of introducing them is not to undermine students' self-confidence, but rather to sharpen and reformulate their underdeveloped intuitions. But since their introduction in class is part of a bigger scheme, let me sketch the whole picture.

Before starting a new subject, I often offer several teasers. One way is to include a well-selected (guided) problem in the previous week's assignments. For instance, let us consider the first topic of the course - the various definitions of probability: Classical (Pascal & Fermat), Frequency (Von Mises), Axiomatic (Kolmogorov). Their introduction in class should be accompanied by many discrete and continuous examples in which the probability of events can be obtained from the relevant definition. Within this subject, before introducing the use of tree diagrams as a way of describing the sample space, and finding all the "favorable outcomes", one can assign the following famous problem of Pascal: Two evenly matched players are playing a series of games. The first to accumulate 4 victories gets \$100. How should they divide the cash if the games were interrupted after player A had accumulated 2 victories, and player B, 3? (*The common answer here is according to a ratio of 2:3, but actually the ratio should be 1:3. The mistake is not a result of bad intuition, but is rather a natural guess resulting from lack of relevant thinking tools- to be discussed soon after.*)

In class, I often start any topic by distributing a short *questionnaire* with relevant "judgment problems". For example, the first such questionnaire may include the following problems:

- According to statistics, during World War II, a pilot of the Royal Air Force had a 5% chance of being downed in any given mission. A pilot has been assigned to 10 missions. Would you say that his survival chances are smaller, equal, or perhaps bigger than 50%? (*The essay from which this information was taken has claimed a 50% risk, a common pitfall.*)
- Two honest dice are tossed. Which event is more likely: the sum of the numbers is 9, or the sum is 10? (*A common mistake is to assume equal probabilities. Why?*)
- In a family of 4 kids, which event is more likely: (i) two of each sex, (ii) 3 of one sex and 1 of the other sex? (*A common answer is (i), because it looks like a "more typical" event.*)

There are several possible QPs to *accompany* this subject. For example:

- The gambler fallacy: In a casino, a roulette wheel stopped on a red number in the last 7 games. Would you rather bet on red or on black in the next game?
- Is it a favorable policy to wait until there is a long run of reds and then bet on black?
- A small hospital had 10 births this week; a larger hospital had 100 births this week. Which event is more probable: more than 6 boys in the small hospital, or more than 60 boys in the larger hospital? (*A well known fallacy, due to Kahneman and Tversky.*)
- The subway puzzle (Mosteller, 1965): A young innocent guy goes every evening, at some random time, to the nearby subway station. He has agreed with his girlfriend that if the northbound train comes first he will take it and thus spend the evening with her; but if the southbound train comes first he will take it and visit his mother. Over the long run, the girlfriend observed that he actually visited her on only about 1/3 of the evenings. As there are exactly the same number of trains going in each direction she accused him of cheating on her. Can you help prove his innocence?

In the rest of this section, we describe some other useful QPs. As we will see, many of them are connected to *conditional probabilities and independence*. Indeed, those are notorious subjects from the point of view of misconceptions and fallacies. With so much talk about the era of information, it seems like a paradox in itself that so many people lack the intuition of how to make use of additional information in their computations. Another source of problems is the improper use of *symmetries*. People feel free to assume symmetry by default (and often neglect to see real symmetries that require some insight to discover...).

Lewis Carrol's urn: An urn contains a single ball and you know that it is either black or white with equal probabilities. You add a white ball to the urn, and then you take out one ball, at random. The ball selected turns out to be white. What is the probability that the ball remaining in the urn is white too? (*It is not 1/2! Why? Compute and explain.*)

Mrs. Cohen has a son: Mrs. Cohen gave birth to (non-identical) twins. You know she has a son because you remember that there was a birth-party (circumcision) soon after birth. You meet Mrs. Cohen with one baby (unfortunately dressed in yellow). What is the probability that it is a boy? (*Many students will say 1/2 because they don't know how to make use of the extra information!*) Another nice variation of this puzzle, that leads to a different direction of inquiry, is to compare the probability of E- both are boys, given A - at least one is a son (as before), B - the elder is a son. Surprisingly, the two are not equal.

(Both are positively correlated with E, but the second has stronger correlation!) This example also leads to a discussion of how exactly the new information (a son) has been acquired.

The two aces puzzle (Related to the previous one): Why is the probability of having an additional ace in a hand of bridge different when you know A- the hand has an ace, or B- the hand has an ace of hearts. (*Same remarks as above.*)

A winner and a gentleman (non-transitive dice game): The game involves 4 specially numbered dice A,B,C,D: A has four 4s and two 0s, B has only 3s, C has four 2s and two 6s, D has three 1s and three 5s. In each game one player chooses a die as he pleases and then the second player can choose a die out of the remaining three. The dice are tossed, the player who gets a higher number wins. You play a series of such games. The surprise is that you can afford being a gentleman and let the other player always pick his die first, and yet make the games favorable to you! *The reason is that die A beats die B, which in turn beats C, which beats die D, which (completing the circle) beats die A. As a result you can guarantee 2/3 chance of winning each game. How?*

The Run puzzle: Two players toss a fair coin repeatedly. Player A bets on a run of 2 consecutive heads (HH), player B bets on a run of tail followed by head (TH). The player whose run appears first, wins the game. Even though the chance of getting HH is equal to that of TH, the game is impressively favorable to player B (his chances of winning are 3 times larger than that of A). The same will be true if each player has his own coin and they toss them simultaneously.

The Elevator and the Paranoid (or is he?): A busy office building has one elevator (that spends on average the same amount of time on each floor). Mr. X's office is on a high floor and he often has to wait for the elevator to take him up to the top floor to meet the big boss. Mr. X strongly believes that the first elevator to stop at his floor is more often the one going down than the one going up. Is he paranoid or is there some more rational explanation? (see Gardner, 1981)

Which value is most likely? A drunken person returns home and wants to get in. There are 10 keys on his key ring and only one fits. He pulls a key at random and tries to open the door. If it doesn't fit, the key returns to the ring and he pulls a key (out of 10) again and again until the door opens. Under these circumstances, the number of trials required is 1,2,3,... Which of these values is most likely? *Students are always surprised to find that the most likely value (i.e. the mode) here is 1 (no matter how many keys are involved!). This is a good starting point for discussing the difference between the various measures of location (the mean here is 10 and the median is 7).*

The last to leave puzzle: It is customary to assume that the processes of customers entering a service station, and of those finishing their service are both Poisson processes. The problem: You enter the bank and see the two cashiers each busy serving a customer. The cashier that finishes first will serve you. What is the probability that you will be the last to leave the bank? *It seems that the probability should be close to 1. Not so, from the lack of memory property of exponential dist. one can deduce (no computation required!) that the probability is actually 1/2.*

The Cluster puzzle: Consider a Poisson process over some area (e.g., the flying-bomb hits of London during WWII). A Poisson distribution results from uniformity assumptions of the rate. Yet, the Poisson model itself actually predicts the existence of clusters at random locations! (In London, even though the rate was 0.93 hits per unit area there were several areas with 4 or more hits). *This example can lead to discussion of evidence, say, in cancer clusters.*

CLASSICAL PARADOXES AND HOW TO USE THEM (VERY CAREFULLY...)

I will describe here briefly, some of the "Classical Paradoxes". It seems that paradoxes are to the experts what QPs are to laymen. The conflicts they cause are much deeper and they

pose a real challenge even to professional mathematicians; but their investigation always leads us back to the basics of probability theory.

Often I offer a paradox at the end of class (after finishing a subject) and postpone the discussion until the next time. The bright students always “catch the bite” and become hooked on the problem. Sometimes the discussions are not held in class but informally outside. One should be very careful about using such paradoxes, as the discussion may add to the confusion of students who are not yet sure about the new concepts.

The Roulette Paradox: A balanced roulette wheel is spun. The angle at which it stops is recorded. What is the probability that the angle will be exactly π ? It follows from the countable-additivity axiom of probability (actually from continuity) that this probability should be zero! But the same is true for any given angle. It seems that the roulette will not be able to come to rest. The paradox is that, as we know, the roulette will stop at some point a , which originally had probability zero. How possible? *This paradox is a good introduction to continuous sample space problems, in which this situation appears routinely. In such problems, one should also get used to the fact that, unlike the discrete case, in continuous problems the probability of an event is not computed from the probabilities of the simple events composing it.*

The Prisoners' Paradox: Captain Hook imprisons Wendy, John and Michael. Tonight one of them (chosen randomly) will be executed; the other two will be released. Wendy appeals to a guard to tell her the name of one brother to be released. Her argument is that no harm will be done because she already knows that at least one of her brothers will be released, so her chances of being the one executed ($1/3$) will not change. The guard refuses, saying that given the new information, there will be only 2 candidates, so her chances will increase to $1/2$. Who is right?

Monty Hall Paradox: The more famous brother of the previous paradox (see for example Bar Hillel & Falk, 1982). Wrong convictions here are vehement. So much so, that computerized simulations exist on the web for this game. Those are helpful tools, but they cannot replace the mathematical resolution. *It is advisable to introduce this paradox in conjunction with the previous one. They both involve the right use of conditional probabilities, a discussion of symmetry and false symmetry, and also inquiries about the mechanism behind the guard's way of selecting the name he will announce (or Monty's way of selecting the door) in the situation where there are two possibilities for that.*

Simpson's Paradox: There are many real life examples to this insufficiently known Paradox. I present here real data taken from two local newspapers in my neighborhood. One newspaper reported that school α scored better (96.3% passed the test) than school β (93.2% passed) on a language comprehension test given to all 4th graders in the country. The other newspaper gave a more detailed report: among the group of new immigrants, school β scored better (50% compared to 30%) but also among the rest of the students, school β scored better (98% compared to 97%). It seems as if we have two conflicting reports. Could they both be correct? *To settle the paradox one has to use the total probability formula, and to get another bit of information: the percentage of new immigrants in each school. It turned out that school α had only 1% while school β (which was near an absorption center) had 10%. This settled the conflict.*

The Unlucky Investor: An energetic boy got a piggy bank for his birthday. He decided that from then on he will number every bill he gets from his grandparents (1,2,...) and put it all in his bank. During the first half year he got 2 bills, but at the end of this period he pulled out 1 bill (chosen at random). In the next $1/4$ year he got 2 more bills, but at the end of this period he pulled 1 bill chosen at random from the 3 bills in his bank. In the next $1/8$ year he repeated the same routine etc. (each period is half the length of the previous period). What is the probability that any of the bills he got during this year, will remain in his bank after a full year of the above activity? Paradoxically the probability is 0, even though it is clear that he only spent half of his money. *Can we offer the boy good financial advice without making him cut his expenses?*

Murphy's Law Paradox: You have to choose one of two sealed identical envelopes. You know that one of them contains double the amount of money of the other. After you choose an envelope before you open it, you can change your mind and switch. What would you do to increase your profit? As always, Murphy's Law ensures that whatever you decide to do, it will turn out that the other possibility was better. Indeed, denote the amount of money in the envelope you choose by x , the other envelope will contain an amount of $x/2$ with probability $1/2$, or $2x$ with

probability $1/2$. So switching will give you on average $5x/4$, which is larger than x . It seems clear that you should switch. The Paradox here is that after switching and before opening, the same argument will convince you to switch back to the original envelope...*To solve the paradox for a specific value of x , one has to be careful in writing down the probability space, the (delicate) mistake lies there. If, on the other hand x is considered to have a random integer value, one should discuss the probability distribution of x . It cannot be just be a randomly chosen integer (why?). This paradox can lead also to discussions of prior distribution and the Bayesian approach in probability.*

Bertrand's Paradox: A cord is chosen at random on a given circle. You are asked to compute the probability that its length exceeds the length of an equilateral triangle inscribed in the circle. Bertrand offers 2 perfectly correct solutions, resulting in 2 different answers: $1/2$, $1/3$. *The discussion leads to inquiries about the meaning of the expression "at random" and how does one achieves randomness in real life experiments.*

The Waiting Time Paradox: If busses run exactly every 20 minutes, than a traveler arriving at a bus stop at a random time, will have to wait 10 minutes on average. The average amount of time that passed from the previous bus is also 10 minutes. Now consider a lonely road where cars pass according to Poisson process (a reasonable assumption) with rate of 3 per hour. A hitchhiker arrives there at a random time. By assumption, the average time between two consecutive cars is 20 minutes. The alarming news is that our hitchhiker will have to wait 20 minutes on average for the next car to pass. Moreover the average time from the previous car is also 20 minute! *What happened to the additive property of expectation? The solution lies in the fact that according to the above procedure, longer periods have larger probabilities of being selected.*

The St. Petersburg Paradox: The St. Petersburg game is played by flipping a fair coin until it comes up tails. The total number of flips required, n , determines the payoff, which equals $\$2^n$. The game has an infinite number of possible outcomes. The expected payoff here is $\$2(1/2) + 2^2(1/2)^2 + 2^3(1/2)^3 + \dots = \$1 + 1 + 1 + \dots$ which is infinite. The paradox here is that the payoff (which is finite with probability 1) will *always* be less than the expected payoff! Note also that intuition tells you not to offer more than, say, $\$20$ to play this game. *This paradox can lead us to discussion of the meaning of an expected value as a fair participation fee in a single game and in a series of games. Historically, the St. Petersburg Paradox played an important role in the development of decision theory and the concept of utility function of money.* We will not discuss here some more recent paradoxes, such as *Parrondo's Paradox* that describes two simple unfavorable coin-tossing games A , B , such that over a sequence of games in each of which the player chooses randomly between A and B , the game becomes favorable.

In closing let me just remark that all these paradoxes and QPs are useful tools to serve a purpose. They are not the purpose of the course; they cannot be allowed to replace the *theory*. One should remember that our main goal in teaching a probability course is not to entertain, nor even to sharpen the mind, but to master a new and challenging subject in mathematics, a subject that has a lot of real life *applications* that should be discussed too. Lately I have come across more and more populist "probability courses" that are devoted mainly to "critical reading of newspaper essays" and to "Paradoxes and Surprises in Probability". In other words the "main course" (which is indeed not easy to digest) is skipped and only the "dessert" is served. It may be sweet, but definitely is not a nutritious diet. Let me just say that they are not "my cup of teach".

REFERENCES

- Bar Hillel, M., & Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition*, 11.
- Gardner, M. (1981). *Aha! Gotcha, paradoxes to puzzle and delight*. W.H. Freeman and Co.
- Leviatan, T. (1998). Misconceptions, fallacies, and pitfalls in the teaching of probability. *ICM*.
- Leviatan, T. (2000a). *Probability and statistics for sciences*. Amihai Press (In Hebrew).
- Leviatan, T. (2000b). Strategies and Principles of Problem solving in Probability. *ICME-9*.
- Mosteller, F. (1965). *Fifty challenging problems in probability*. Addison-Westely.
- Pitman, J. (1993). *Probability*. Springer Verlag.
- Ross, Sh. (1998). *A first course in probability* (5th edn.). London: Macmillan.