TEACHING STATISTICS BASICS TO BIOLOGY AND HEALTH PROFESSIONALS THROUGH BAYESIAN IDEAS

J. Andrés Christen
CIMAT, Mexico
Romulo Lins
UNESP, Brazil
jac@cimat.mx

Bayesian statistics is usually viewed as an 'advanced' course; Bayesian methods are either seen as an optional extension or its founding ideas are seen as difficult to grasp. There are, however, two good reasons for changing this. First, the recent strong increase in interest in Bayesian methods, in several areas. Second, the fact that Bayesian methods are particularly relevant for the kinds of decisions health related professionals have to make. Here we will argue that the Bayesian approach makes a quite convenient platform for introducing key ideas of probability and statistics to biology and health professionals; we will also outline how such a course can be organized around everyday, easy to grasp, examples, exploring how those simple situations involve a complexity that requires a deeper discussion of the concepts and ideas underlying the modeling process that can lead to acceptable and adequate answers to them.

INTRODUCTION

In this presentation we describe and comment a short and introductory course on Bayesian Statistics. The course was presented as a 5 hour introduction to Bayesian statistics to 1st and 2nd year undergraduate students of Bioinformatics (Biology major plus information technology and statistics minors). We assumed basic knowledge of Probability theory (random variables) and recommended, but not as mandatory, a first course in classical statistical methods.

INITIAL QUESTIONS

We started by presenting some questions, without any prior explanation or discussion. These questions aimed at showing students that even in simple, real life everyday problems, there are complex probability issues that are of principal importance and that go well beyond the "favorable over total cases" definition of probability.

Question: 'What is the probability that if I toss a coin it lands on "heads"?'

Some students here will likely answer "0.5." However, the common thing is that only a minority will do, or even no one and the rest will remain silent expecting some sort of catch. Insisting a bit more on getting answers, students will recognize that a series of ambiguities will need to be solved before we can provide an answer to the question (which coin is used? how is it tossed? what is "heads"? etc.). It is quite relevant here that, even in this most simple and straightforward probability statement, assumptions must be made before a probability of heads may be agreed upon, usually that there are only two possible outcomes and that they have the same probability. That shows that *someone* must be making the assumptions and therefore probability must be conditioned to the person or *agent* making the statement. In this particular case, one can discuss the fact that prior to any experience with that coin, it is reasonable to assume that we have no reason to presume that one of the outcomes is more likely than the other.

Next question: "What is the probability that your lecturer has more than 50 pesos (around 5 USD) on her/him (pockets etc.), right now?"

Here, we show students that probability statements will possibly also depend on additional information each *agent* may have (who is the lecturer? possible income? etc.). In Mexico, students may survive on 50 pesos a day or less. A university lunch will cost possibly 30 to 40 pesos. Some people assign probabilities like 70% or 99% and other more conservative (maybe suspicious) say 20% or even less. It is easy to see that there is not *the* probability but *a* probability, conditional on the *agent* making the statement, meaning that the information available to the *agent* is constitutive part, in a sense, of that *a* probability.

Linking this with the coin situation, one can discuss, then, what if we begin with the 50-50 assumption, toss the coin and a head comes up: what is the probability that a second toss will also produce a heads? Let's further assume that the coin was chosen 'at random' by asking one of the students for a coin s/he had on the pocket. The 'wisdom of the streets' — meaning what people without specialized education on this would say — is quite likely that the probability is the same as before, ½. Interestingly, this answer coincides with that which would be given by a classical statistician, but for different reasons: while the layman would probably say that it is a regular coin picked at random, and with regular coins one always think of 50-50, the statistician would probably invoke the principle of insufficient reason — one throw is not sufficient to change my assumption. With a well motivated group of students this may lead to a quite sophisticated discussion on how many throws would be necessary to provide sufficient reason to make one change the 50-50 assumption.

Another quite challenging discussion could arise if the professor is lucky— as we have been on occasions — to have a student to be brave enough to put forward the thesis that in fact the probability of heads would *diminish* — even if for a tiny amount — once "in the long run heads and tails should equal, so there should be some sort of 'compensation' mechanism to guarantee that it happens." Do coins have memory (maybe of a quantum nature)? More generally, this may lead to a discussion of how strong animism is in everyday life (just consider a person punching a car and cursing it because it failed to start).

> Another one: "What is the probability that it rains tomorrow in *X* (some known but distant place)? What is the probability that has rained yesterday in *X*?

Here we explore the aspects examined before about agents and assumptions and introduce one final and crucial one: students *feel* that a probability statement *could* be made for both questions, nevertheless for the second we are talking about something in the past, an event that already occurred and should dispense a probability statement, requiring instead only a 'yes' or a 'no.' Here we can argue that the important point is that we do not *know* — even if as yet — whether it rained in *X* or not, bringing uncertainty and the need of using a probability statement for the event. Also, it is possible to discuss that although in this case as with the coin we are talking about the same thing, probability, the coin can be tossed as many times as we want but the 'rain in *X* at such and such day' is a one-off event; that might help our students to fully appreciate the point above: that probability statements are about events — past or future — about which we do not have sufficient information to make assertive statements and that, again, brings in the fact that we are talking about *someone* having or not sufficient information.

AXIOMATICS

One topic that is only taught in advance courses of Bayesian statistics is the axiomatic foundations. However, we introduce the axioms (as given by DeGroot, 1970) without giving any proof of the basic theorems (like transitivity etc.); one can nicely justify the axioms (rules of the game) for preferences, since students have been engaged — through the initial questions — in making probability (and thus preference) statements about complex events (rain in the past, coin tossing, etc.). The advantage of introducing the axioms is discussing the "auxiliary event" (axiom 5 in DeGroot), and its direct link to the definition of probability as comparison to a standard event (we avoid explaining the more precise but difficult definition as a preference relation between two lotteries). Using this, we establish a definition of probability that goes well beyond the much more usual, and quite restricted, frequentist definition.

SOME MATHEMATICS:

The next part of the course involves trying to go into more technical calculations. Given the axioms, a basic principle is established for Bayesian inference:

"uncertainty is measured with probabilities"

In that respect, Bayes' theorem is a means for "updating probabilities with evidence (data)." The most basic form of Bayes' theorem is proved and an interpretation of its parts is given (prior and posterior probabilities, model, and normalization constant). A Bernoulli trials example is explained in which the success probability is fixed to values 0.2, 0.5 and 0.7 (representing the degree of some enzyme reproduction in a colony of cells, and these values are known from some previous theory). Having 20 trials and 12 successes, the posterior distribution is obtained using a table.

A second example is explained using Bernoulli trials in general. Using a beta prior the posterior is shown to be beta also. Students are then introduced to the *R* programming language and they use a simple function to calculate the corresponding beta posterior distribution for any beta prior and data. We show how the prior and data affects the posterior and that with a lot of data the effect of the prior diminishes. This is a simple, hands on, introduction to the important results of Bayesian asymptotics, that otherwise will be too technical to cover.

The expected outcome of 'mathematising' the discussion is that students will have a stronger feeling that incorporating an agent's knowledge and assumptions into further probability statements is not simply a matter of 'pure intuition' or common-sense; the mathematical tools offer a *negotiated* way that allow several people to account for that updating in an acceptable way. More in general — and we think this should be included in the discussion — this is a characteristic of any mathematical modeling, be it using statistical concepts and tools or otherwise, in any field of knowledge.

THE GENDER PROBABILITY PROBLEM

A couple has had 5 pregnancies resulting in boys, and they wish to know the probability that with their 6th pregnancy they will — finally! — have a girl.

Apart from any "technical" considerations one can explore the possible human and cultural interests in asking such a question. In some cultures it is considered normal that fathers wish to have sons (so they can, for instance, play football with them), while mothers wish to have daughters (perhaps to keep them company while the 'boys' play football). In other cultures having a daughter means the parents will have to save money for the dowry if they want to get her married. Generally speaking, that simple, naïve, question may trigger a quite interesting — and always due — conversation about gender questions in relation to culture and society.

To the question itself, all sorts of reactions come up. *Although* they are Biology students, most of them will say that "it is still 0.5" (what have they done with all that impressive information about the first five pregnancies?). However some will cast doubt and may argue that "given the evidence the probability should be less than ½." Even more interesting, some 1 or 2 students say that "the probability should be higher now" arguing some sort of "compensation" given the large amount of boys in the family.

The '50-50' approach resembles the street wisdom about the coin. The 'less than 0.5 for a girl' approach *could* resemble the 'sufficient reason' argument for the coin. And the 'compensation' argument *could* resemble the similar one for the coin.

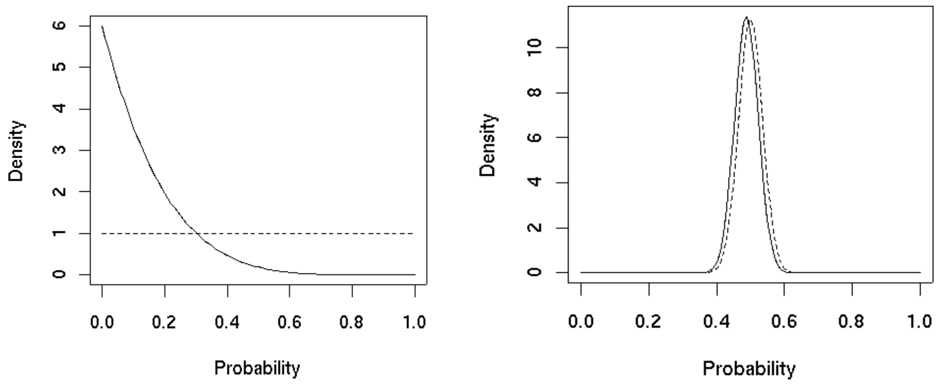
The trouble is, although there are seemingly similar approaches in the two cases, the factors conditioning the judgments in each case are quite dissimilar. However in the first case the reasoning might be the same — "independent events," so to speak, as the birth of two children ('individual humans') seem to be as independent as two tosses of a coin, and there are only two possibilities (0 or 1), in the other two they may not be.

Would five consecutive heads outcomes warrant that the coin was unfair? Not quite likely. So why would five consecutive boys warrant the couple of parents was biased towards boys? Maybe the fact that not many people have five children — and those who do are not likely to have, say, 20 — could be a factor, considering that one can easily produce hundreds of coin tosses, so in the gender situation five 'tosses' become a quite reasonable sample. In a similar way, although coins do not plausibly have a memory (going against animism), human bodies/beings might have such a memory (through emotions, wishes, souls and other factors that most of us are definitely not willing to associate coins with).

Again, the technical approach to this whole discussion might help students to gain consciousness of the role of a statistical approach to phenomena to which there is an intrinsic associated uncertainty. If the same theory can be used to study both what to say about five consecutive boy births and five heads coin tosses, then one can gain some appreciation of the role of such theory and tools, as well as gaining insight into, in this case, the fact that human couples can indeed be 'biased' towards boys or girls. In the early 90's a doctor went on national television in the UK to say that after five boys the couple has "of course" a greater chance of having a girl, as a kind of compensation towards 50-50. Amazingly, in that same occasion he discussed techniques for increasing the probability of a given gender outcome, techniques that were clearly based on the fact that a combination of factors involving the parents' physiology can indeed bias the 'selection' process (e.g., acidity of vaginal fluids). In other words, he implicitly refused to take those assumptions into account and stayed firmly with the 50-50 view, *to the extent of adopting the 'compensation theory.'* The interest of biology and health professionals in statistics is related to decision making, and that in itself recommends the Bayesian approach.

As much as after five heads, heads should be expected, after five boys a boy should be expected — no matter how little we judge the balance to lean to one side, particularly because a new pregnancy has certainly much more impact in a couple's life than a new coin toss.

Theory may be used to approach this problem. One may start asking, for instance, if the gender yielding each pregnancy is independent from pregnancy to pregnancy, given the unknown probability of male/female. Students are invited to give arguments against (conditional) independence in this case. With this a model for Bernoulli trials is deduced and, assuming a beta prior, a beta posterior is obtained. The former *R* code may be used to obtain the corresponding posterior.



(a) (b)

Figure 1: Prior (dotted) and posterior beta distributions for 5 failures in 5 Bernoulli trials, (a) with a noninformative uniform prior and (b) with a beta (100, 100) informative prior, representing prior information about the gender probability in any given pregnancy

A flat (uniform) prior is first displayed for discussion, see Figure 1(a). In this case, the probability of girl seems rather low (definitely less that 0.5). However, it is not rare that some students ask "why did we use that flat prior? Isn't it what we now *a priori*, that is, before any pregnancy?" This is especially interesting since we are talking about students that have been exposed to Bayesian ideas for only 2 hours! Indeed, we rapidly agree, even for the people that felt the posterior should be less than 0.5, that the prior should be quite peaked around 0.5. A rather different posterior is obtained, concentrated around 0.5 and just slightly shifted to 0 (see Figure 1(b)), coinciding with the general intuition that the probability will need to remain 0.5 basically as before. That is, our probability of a girl in the 6th pregnancy remains, for all practical means, 50-50. But to reach this conclusion we needed to use additional information beyond the mere data set ($X = \{0, 0, 0, 0, 0\}$): our *a priori* information about the probability of having a girl in any given pregnancy. And in this case the amount of information is rather large and a quite peaked prior distribution around 0.5 most be agreed up on, as in Figure 1(b), as opposed to the rather neutral (and inappropriate for the pregnancy example) nature of the prior in Figure 1(a).

But again, and this is not a minor issue, the slight change here has to be assessed against the actual situation, involving a possible pregnancy and not a mere coin toss. This is the kind of awareness that health professionals would benefit from, an awareness of the fact that the *consequences* of a decision must taken into account, and that leads us to our next topic.

HYPOTHESIS TESTING

Finally, and very important for Biology students, is a brief discussion on a Bayesian approach to hypothesis testing. Following the basic principle established earlier: "uncertainty is measured with probabilities," one comes to the idea that what is needed is to calculate the posterior probability of each hypothesis. A final example is given, in which an experimental drug is tested on 20 patients with 15 successes. The standard drug is known to have a 50% efficacy. The hypothesis is established as:

$$H_1: p < 0.5 \text{ vs. } H_2: p >= 0.5$$

Using a uniform prior, the beta posterior is obtained and the probability of H_1: $p < 0.5$ (efficacy of the experimental drug is less than 50%) is calculated as 0.0133. Is the experimental drug better?

Students tend to view 0.0133 as small (1.3%) and many tend to implicitly decide in favor of the experimental drug. But, as for now, we are only able to say that "it is highly probable that the experimental drug has a higher efficacy than the standard drug, given the current evidence." However, making a decision will necessarily need to take into account the consequences of our decisions. Students then are presented with two fictional choices:

1) You are the winner of a worldwide raffle and are invited to go to space for 1 week, orbiting the Earth, completely for free. You are warned, however, that there is a 1% probability that the rocket blows at some point of the trip and you die. Will you take the trip?
   Some students (risk prone) immediately say "yes," while others call them "crazy" saying a cutting "no." After some discussion we go on to the next question:
2) You are invited to enter a room for 10 min, next door, for an experiment the lecturer is conducting on human behavior. The room has no windows, nothing inside and you are warned that there is a 1% probability that the room explodes while you are inside and you die. Will you enter the room? Of course, no one wants to enter the room, finding it a ridiculous idea.

Nevertheless it is a 1% probability of dying in both examples. Without studying decision theory, we suggest to students that decisions are not only based on probabilities but, crucially, on the *consequences* to be faced or enjoyed, and "small" or "large" probabilities are only a matter of the context we are dealing with.

DISCUSSION

With this brief description of a short course in Bayesian statistics, we show that Bayesian ideas and sophisticated probability concepts may be taught to Biology students in a non-technical way, and that the usual concept that Bayesian statistics is an advanced course is simply misleading. Moreover, Bayesian statistics may well be a first course in statistics, once having some basic concepts of calculus and probability theory, which commonly undergraduate Biology students take in their first year. Of course, more material can be studied, like normal sampling, nuisance parameters and marginalization. This only takes basic knowledge of Calculus. Moreover, it is quite straightforward to establish the difference between probability and decision without embarking on any technical discussion on decision theory. For biology students and health related professionals, learning the important concepts of probability and developing an intuition based on them is probably the central aim of such part of their statistical education, They can then apply and use that intuition and those insights in their everyday practice, rather than trying to dominate the technical aspects, which could adequately — whenever necessary — be dealt with by (or at least consulted with) a statistician.

The same general suggestion applies to mathematics education, that is, it is rather different to think with concepts than to be able to go into the technical side of it. It remains an open question to investigation to assess the extent to which a technical mastery of statistics or mathematics fosters or hinders 'thinking with the concepts.'

REFERENCES

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.