ON THE NON-ARBITRARY ASSIGNMENT OF PROBABILITY

William M. Briggs

Dept. Mathematics, Central Michigan University

Pearce Hall, Mt. Pleasant, MI 48859

email: mattstat@gmail.com

July 12, 2007

DRAFT

SUMMARY: How to assign numerical values for probabilities that do not seem artificial or arbitrary is a central question in Bayesian statistics. The case of assigning a probability on the truth of an proposition or event for which there is *no* evidence other than that the event is contingent, is contrasted with the assignment of a probability in the case where there is *definte* evidence that the event can happen in a finite set of ways. The truth of a proposition of this kind is frequently assigned a probability via arguments of ignorance, symmetry, randomness, the Principle of Indiffernce, the Principal Principal, non-informativeness, or by other methods. These concepts are all shown to be flawed or to be misleading. The *statistical syllogism* introduced by Williams in 1947 is shown to fix the problems that the other arguments have. An example in the context of model selection is given.

KEY WORDS: Finite exchangeability; Induction; Logical Probability; Model selection; Principle of Indifference; Principal Principle; Prior assignment.

1. Introduction

There are (at least) two central foundational problems in statistics: how to objectively justify probability models, and how to objectively assign probabilities to events and to the parameters of probability models. The goal of both of these operations is to insure that they are not arbitrary, or are not guided by the subjective whim of the user, and that they logically follow from the explicit evidence that is given or assumed to be known.

Concepts such as exchangeability, symmetry, and even direct appeals to physics or biology are sometimes given to posit a probability model (Diaconis, 1977; Diaconis and Freedman, 1980; Diaconis, 1988). However, frequently no justification other than habit—or ignorance of any alternative—is used to guide a user to select a particular model. Model correctness is not examined here. I take models as given, and instead look at the second question of probability assignment.

That problem is huge, so here only a small piece of it is taken in the context of logical probability: how to assign a probability value on the truth of an observation statement (or event) in two situations: when nothing is known about the event other than it could happen (definite knowledge of contingency), and when we know that the contingent event can happen in a certain finite number of ways. Questions like this are common in model selection, and are central to questions of probability interpretation. For a very good reason to be explained, formal examples will be withheld until after the central concepts are introduced.

First, it is useful to recall, what is often forgotten, that—both deductive and non-deductive—arguments of logic are nothing more than the study *between* statements, and only between the statements explicitly defined. How true is one thing given another? is the usual, and should be the only, question. The existence and characteristics of the statements themselves is left to other disciplines. Forgetting this distinction can lead, and has lead, to unnecessary arguments about the nature of logical probability. My attempt here is to clear up some of these controversies in the context of probability assignment for the truth of an elementary propostion.

Start with classical logic, mathematically describe in, e.g., (Schechter, 2005). So, suppose p is a premise and q a conclusion to the argument from p to q. We may write this argument in many ways: one of the most verbose—but clearest—is this:

$$\frac{p}{q} \tag{1}$$

which is to be read, "(the proposition) p (is true) therefore (the proposition) q (is true)" (the mathematically succinct way to write this is $p \Rightarrow q$). Logical probability makes statements about the conclusion of

(1) like this:

$$0 \leq \Pr(q|p) \leq 1. \tag{2}$$

(Incidentally, writing (1) like it is instead of $p \Rightarrow q$ makes it easier to see how (2) arises: we *could* write (2) as $\Pr(p \Rightarrow q)$ but this gives the feeling that we're asking about the probability of the *implication* "$\Rightarrow$" and not about the truth of q; the implication being false unless q is true, but see, e.g., Adams (1998) who writes it in in the alternative manner). Cox (1961), and like those in the logical probability tradition before him (de Laplace, 1996; Jeffreys, 1998; Keynes, 2004; Jaynes, 2003), states that if the limits 0 or 1 apply to the conclusion q of a given argument with premiss p, then q is, respectively, certainly false or certainly true. When the limits are reached, then the logical connective (between q and p) is said to be *deductive*. If the limits are not reached, then the logical connective is said to be *non-deductive*. Non-deductive arguments may be inductive, or they may be otherwise. The arguments from p to q are either valid if they are deductive, or invalid if they are not deductive. Invalid does *not* imply unreasonable; neither does deductive imply reasonable.

Here is a simple example of a deductive argument that is not reasonable (in the sense of relevance) adapted from from Schechter (2005): "If it is raining now, then red is a color. It is raining now. Therefore, red is a color." This is a valid argument in classical logic because the implication is always true (at least, when it's raining), as red is certainly a color. But the two concepts—color and rain—have no relevance to one another.

Recall these common definitions: *contingent* means not necessarily true or false, and an observation statement or *event* is some thing that can happen (is not necessarily false or impossible) in the given context (examples will be given below).

Inductive arguments—which are arguments from contingent premisses which are premisses that are, or could have been, observed, to a contingent conclusion about something that has not been, and may not be able to be, observed—are, of course, central in probability. In an earlier paper (Briggs, 2006), I started with an example (due to Stove (1986), borrowing from Hume (2003)) of an inductive argument which everybody believes is reasonable. That was, ($p =$)Because all the many flames observed before have been hot, that ($q =$) *this* flame will be hot. Notice that no measure of reasonableness is given, no measure of how true the conclusion q is with respect to its premiss. We can give such a measure, and that we can do so is explained using the principles of logical probability (which I do not prove here; but see the references below).

The flames argument is inductive. *Not* all non-deductive arguments are inductive. Carnap (1950), the most widely known proponent of logical probability in the 20th century, unfortunately had the habit of calling all non-deductive inferences 'inductive', which, among other things, lead to a confusion about what logical probability is, and it is this confusion that is in part responsible for logical probability's current refugee status in statistics, Franklin (2001). This is fully described Stove (1973, 1986). In any case, I do not follow Carnap's terminology here, though I use deductive and non-deductive logic in what follows.

The main purpose of this article is to survey the most common arguments used in assigning probabilities to uncertain events where the event can happen in a finite number of (known) ways. These ways are usually assigned equal probability. The usual reasons given for equi-probable assignment are: ignorance, "no reason" or indifference, non-informativeness, symmetry, randomness, and some very well known mathematical arguments. All of these arguments, by no means mutually exclusive, will be shown to be flawed, or to be misleading, or to imply the necessity of subjectivity when it is not needed. Instead, an old argument, called the "statistical syllogism", will be re-introduced. The statistical syllogism avoids the problems inherent in the others, with the added benefit of clearly and completely delineating the information used in a given problem.

2. Ignorance

To stress: logical probability concerns itself with assigning probabilities to the conclusions of arguments with explicitly stated, and fixed, premisses. It is easy to assign probability when the argument is deductive: the probability being 0 or 1. But, of course, most arguments are not deductive: that is, they are non-deductive. *Not*, as is commonly assumed, all non-deductive arguments are inductive. For an example of a common, non-inductive (and non-deductive) argument, suppose we have *definite* knowledge, labelled e_c, that M is some non-contradictory contingent statement, proposition, or description of an event, and t any tautology. That is, we *know* that M is not necessarily true or false; we also do not know, we are ignorant, whether M will happen. The argument:

$$\frac{\begin{array}{c} t \\ e_c \end{array}}{M} \tag{3}$$

is not valid (and is read "t and e_c, therefore M"; or $t \wedge e_c \Rightarrow M$). Writing out details in this manner makes clear the tacit process of argumentation that is part of any prior probability assignment: all of our evidence is first amassed and then explicitly laid out *before* the probability assignment is made. The advantage of writing things in this extended manner is to very carefully bare what the arguments are actually saying.

The most common tautologies used in cases like this are $t =$"I am ignorant about M, but I know it can be true or false," or t="M will happen or it won't"; both of these ways of writing t implicitly attach the definite knowledge e_c that M is contingent, except that the first mistakenly adds "I am ignorant" since we know of M's contingency. Now, it is true that t; or the statement t is always true. A principle of logical probability gives:

$$0 < \Pr(M|t, e_c) < 1. \quad (4)$$

And that is the *best* we can ever do with only the definite knowledge that M is contingent (e.g., Keynes (2004)). This point, which has caused much confusion, is well worth reflecting upon, and which is amplified below. It follows from the well known logical fact that it is impossible to argue validly to a contingent conclusion given a necessarily true or tautologous premiss. This result, known since Aristotle, is not dependent on a particular t; any tautology or necessary truth will do.

Statements about the probability of M that lack evidence (other than t and e_c) frequently write (4) as "$0 < \Pr(M) < 1$": the missing evidence to the right of M, since it can be anything, is implied. This is usually harmless enough, but it can lead to troubles.

Now, the probability statement (4) represents the best (in the sense of most precise statement) that can be said in the face of no evidence, except for the *definite* evidence that we know M is contingent: (4) is, or should be, the probabilty assigned in the face of this *positive* knowledge. Since this probability is not definite, we *cannot* move towards definiteness using the rules of the probability calculus unless we learn something more about M.

Of course, the situation so far is *not* ignorance, since we have already specified that we *know* M is contingent. Suppose instead that somebody asked you, "What is the probability of M?" and refused to tell you anything about M: it may be contigent, it may be necessarily true, or M may even be complete gibberish. Then *no probability at all* can be assigned. If you do assign a probability it is because you are *adding* information that was not given to you, information you suppose that is true, but that may be false. The argument is changed and you cannot say your assignment is based on ignorance.

Some statisiticians—of the (subjective) Bayesian persuasion—would not like to settle for (4), which is a vague enough statement about M, and would insist that we find some concrete real number r such that $\Pr(M|t, e_c) = r$. To find this number, there is usually an appeal, to the utterers of (4), to announce some subjective opinion they might have about M, or even, if it can be believed, about how they would take bets with the Deity (or, for the secular, with Mother Nature) over the truth of M. This line was begun by Ramsey and de Finetti, and is summarized in e.g. (Press, 2003). I find this approach wholly unsatisfying. And so do those who still call themselves frequentists, and who still do so, at least in part, because of their distaste with this over reliance on subjectivity and insistence on resting probability upon a base of 'gambling' and 'betting.'

Not all Bayesians would insist that you must say how you'd bet for or against M. Some try to find r by an argument like the following: "Well, M can be true, or it may be false. So it must be that $\Pr(M) = \frac{1}{2}$." No, it musn't. The first sentence to this argument is just t, and nothing has been gained. The step from the conclusion to the probability statement is therefore arbitrary (as many have felt before; e.g. (Fisher, 1973)).

The argument can be modified, by inserting some additional evidence: say, $e_\circ =$"M is equally like to be true or false", which I hope you will agree is the same as saying $e_\circ =$"$\Pr(M|t) = \frac{1}{2}$." The argument is then:

$$\begin{array}{l} M \text{ is true or it is false} \\ \Pr(M|t) = \frac{1}{2} \\ \hline \Pr(M|t) = \frac{1}{2}\ . \end{array} \tag{5}$$

This is a curiously dogmatic argument; nevertheless, it *is* a valid one; however, the (major) premiss is the same as the conclusion, which isn't wrong, but it is begging the question. This is usually and loosely called a fallacy, but the conclusion *does* follow from assuming the premisses are true, therefore the argument *is* valid: it is just of no use. (A helpful way to read this argument is to say "p is true, therefore p is true." Attaching the tautology t, or any other tautology or necessary truth, changes nothing; it is then "$p\&t$ is true, therefore $p\&t$ is true.")

There is still the matter of assigning a probability statement to the conclusion of (5), which is:

$$\Pr\left(\text{``}\Pr(M|t) = \frac{1}{2})\text{''} \mid e_\circ, t\right) = 1, \tag{6}$$

a statement which is cruicial to understand: it just says that the conclusion deductively follows from the premisses.

The argument (5) is usually recognized for what it is, and instead, in their search for an r, people will more likely say "Well, M can be true, or it may be false, *and I have no reason to think that it is false or that it is true. I am indifferent.* So it must be that $\Pr(M) = \frac{1}{2}$." This kind of argument is sometimes called the "Principle of Indifference," advanced by Laplace and Keynes (2004) and criticized in e.g. Howson and Urbach (1993). It is the "indifference" or "no reason" clause that is the start of troubles.

3. No Reason & indifference

The minor premiss in (5), "Both [possibilities for M] are equally likely" is evidently itself a conclusion from the premiss, "I have no reason to think that M is false or that it is true," or "I am indifferent about M." Now, this argument, in its many forms, has lead a happy life. It, or a version of it, shows up in discussion of priors frequently, and also, of course, in discussions about model selection, e.g. (Bernardo and Smith, 2000). But it is an argument that should not have had the attention it did. For we can rewrite it like this :

> I do not know—I am *ignorant*; I have no reason to know—whether M is true or false, but it can only be true or false.
>
> ---
>
> M (7)

The conclusion to (7) is usually assigned probability $\Pr(M) = \frac{1}{2}$. This argument, I hope you can see, is not valid: the conclusion certainly does not follow from the premiss, and the probability statement is arbitrary. Here's why. This argument *is* valid:

> M is true or it is false
>
> ---
>
> I do not know—I am *ignorant*; I have no reason to know—whether M is true or false, but it can only be true or false. (8)

It should now be obvious that the conclusion is nothing more than a restatement of the initial tautology! To be explicit: saying you do not know anything about M, in English, means you *know nothing*, and therefore cannot assign any probability, not even the bounds of (4). But if you are saying you do not know whether it is true or false, this is the same as saying that you *know* that it can be true or false, that is, $t \wedge e_c$. So, despite our repeated insistence of "ignorance," we are back to (4), which is to say, right where we started. It should, therefore, be—but it is not—astounding that people have instead come to the probability statement "$\Pr(M) = \frac{1}{2}$" for the conclusion of (7).

This leaves "indifference", which isn't exactly wrong, but it has unnecessary connotations of subjectivity, and, for some, a certain implication that the probabilities are equal (and so begs the question). The subjectivity is implied in the sense that we are *setting* the probabilities by our will, or that, somehow, our opinions matter as to what the probabilities are (see Franklin (2001) for a discussion of how Neyman used a similar trick applied to confidence interval interpretation).

4. Symmetry

Up to this point, I have been very careful not to give an example for M, some concrete, real-world thing upon which to fix the idea in your mind. This was on purpose. Because it is difficult, if not nearly impossible, especially if you are a working statistician, to avoid adding hidden premisses to (3) or (7) once you have such an example in mind, and then to criticize the conclusion that (4) is indeed valid. To emphasize: (4) is the correct statement to make given that the *only* definite evidence for M is t and that, e_c, M is contingent.

To validly arrive at an r, new evidence about M must be added. These additional premisses have to be of a certain concrete character themselves. They cannot be anything like "M can be true or false" or any other restatement of t. They cannot contain the probability statement of the conclusion, as in "M is equally likely true or false." Nor can they measure some form of 'ignorance,' because, at the most, that is nothing different than "M can be true or false." The best—in the sense of being the most precise—probability statement that can be made given these arguments is (4). So if we are to find an r what can these additional premisses be?

Before I tell you, let me first fill in the blank about M, and give you a real example. When I do, unless you are a highly unusual person, you will almost certainly *instantly* think, "Of course the probability of M is a $\frac{1}{2}$! What is the problem!"

Let M represent the fact that I see a head when next I flip this coin.

Are you with the majority who insist that the probability of M must be $\frac{1}{2}$? Before you answer, notice that the 'coin flip' M is *entirely* different from any other M' where all you know is that M' is contingent. For example, if instead of a coin flip, suppose M represented the outcome of an experiment where you to open a box and examine some object inside and note whether you can see an 'H'. Now all you know is that M is contingent and can be true or false. Based *solely* on the information you have, you do not know any other possibilities. You do *not* know that an 'H' or some other letter or object might appear. You do not know, even, whether a snake may jump out of the box. If you imply that because the question asked something about an 'H', that the result must be 'H' or some other letter, probably a 'T', then you are *adding* evidence that you were *not* given.

Back to the coin flip. Why is the probability of M $\frac{1}{2}$? Symmetry, perhaps? As in, "It can fall head or tail and there is no reason to prefer—I am indifferent—to head over tail"? But isn't that the same as ignorance, that is, the same as the tautology and knowledge of contingency? It is. Because substitute 'be true' for 'fall head' and 'be false' for 'fall tail' and you are right back at the tautology. Or symmetry as in, "Heads and tails are equally likely because I have no reason to think otherwise"? Again, "no reason to think otherwise" or "Heads and tails are equally likely" or "indifference" are begging the question or can be misleading.

The anticlimatic answer for assigning probability to a definite M is the statistical syllogism, as defined by Williams (1947) in the coin flip example:

Just 1 out of 2 of the possible sides are Heads

M is an side

M is a head (9)

This inductive argument is, of course, invalid. But we can now justify saying $\Pr(M|e_s) = \frac{1}{2}$, where e_s is the evidence of the two premisses (and which together imply e_c). Adding arguments to e_s about symmetry, or 'fair' coins, or ignorance does not change the probability of the conclusion, because these arguments are all equivalent to adding t or "$P(M|t) = 1/2$" to the list of premisses. A 'fair' coin, after all, carries with it the assumption that the probability *is* 1/2: which is begging the question.

Symmetry has often been used, and objected to, as a principle to assign probability, e.g. Strevens (1998); Bartha and Johns (2001); Hájek (2007). Arguments based on symmetry tend to be misleading because the examples are always chosen in such a way that they are "physically balanaced" or physically symmetric, which gives rise to a certain confusion. For example, Strevens (1998) imagines that one side is painted red on a dodecahedral die and asks the probability (in a 'fair' roll) of seeing the red side. He assigns 1/12 because of (physical) symmetry. Hájek (2007)—and many, many other authors, invoking something about a "priviledged partition"—then argue this assignment is indeed correct under physical symmetry (one partition of the outcome). But (in another partitioning) they say that you either see the red or you don't, so that under this view, the probability is 1/2. Both probability assignments can't be right, so logical probability itself must be flawed! Well, the "either see red or not" is the tautology, which is very different information than physical symmetry: these two different pieces of information should certainly give different probability assessments, so it is to logical probability's credit—and not its detriment—that it does so. (And we have already seen that under the "see red or not" partition, the answer is (4) and not 1/2. Also, all priviledged partition arguments have a distinct subjective quality about them: why choose any partition not based on the statistical syllogism unless you are intent on creating difficulties where they do not exist?)

Again, Streven's die is physically symmetric (though there is also, in the literature, considerable and legitimate worry about how we would know that the device is perfectly symmetric etc. etc.), and so the probability assignment seems natural. And it is this conicidence of physical symmetry and the correct (as I will argue) probability assigment that has lead to all kinds of confusion about "physical" probability. This confusion would disappear were we to consider the problem under the light of the statistical syllogism (clearly, much more on this subject can be said, but this is not the goal of this paper).

To clarify that physical symmetry is not needed, consider this example: suppose I have an n-sided object, one side of which is painted red: what's the probability of red? My object may—or may *not*—be physically symmetric. It may be some amorphous blob, no two sides having the same surface area. It may be physically symmetric down to the quark. But you are *not* entitled to say it is physically symmetric without additional evidence. Just as equally, you do not have any evidence that my object is physically *a*symmetric. And so, you can only appeal to the statistical syllogism.

Another example. Suppose there are 10 men in a room and just 9 of these 10 are Schmenges. M is a man in the room. The conclusion "M is a Schmenge" by the statistical syllogism has probability $\frac{9}{10}$. Note that you rarely hear the term "fairness" applied to situtations like this as you do with coins and dice(is there such a thing as a 'fair' room full of Schmenges?).

To summarize, our evidence e_s for M, in some concrete situation like in these examples, is that M is contingent (e_s encompasses e_c) and that we know or assume that it can happen in any of m out of n different ways: this justifies our saying $\Pr(M|e_s) = \frac{m}{n}$ (the tautology t is still on the right hand side with e_s, but supressed for ease of notation).

I have yet to find anybody who disagress with the probability assignments implied by the statistical syllogism (except for the common mistaken "priveledged partition" arguments like those in Hájek (2007)). Equally compelling, there is no argument against the statistical syllogism (as described fully in (Stove, 1973, 1986)). Even more importantly, and perhaps surprising to some, is that the statistical syllogism is itself *derived* from uniform probability across the individual events that make up the "sample space": see a complete discussion in Stove (1986) pp. 92-97, who credits Carnap (1950) with the first proof of this. In the case of the Schmenges, this is men in the room, or: $\Pr(\text{man 1 Schmenge}|e_s) = dots \Pr(\text{man 10 Schmenge}|e_s)$. That is to say, if you are convinced of the probability assigned implied by the statistical syllogism, you must admit the equi-probability of the underlying events.

It is true that the statistical syllogism gives the same results as the traditional arguments of 'ignorance', 'fairness', or symmetry give, but it does not carry the same baggage. The other arguments, while they contain the necessary information that M is contingent and the sufficient information that M has $n = 2$ or $n = 10$ etc. states, also carry extra hidden assumptions, information that is not explicit and that can cause consternation and disagreement, because not everybody would necessarily put the same value on these hidden assumptions. There is no hidden information to the statistical syllogism. Except maybe something having to do with "randomness."

5. Whither randomness?

An objection to the statistical syllogism might have something to do with "randomness", and how it is invoked to select, or to "sample", say, men from a room. This may seem fair line of inquiry because of practical interest in the conclusion M. I may want to act like a subjective Bayesian and bet, say, on the chance that the man I grab is a Schmenge, or there might be other reasons why I want to accurately assess the probability of M. But arguments about randomness are, just as are arguments about ignorance, irrelevant (Campbell and Franklin, 2004).

If you were to grab a man out of the room randomly: how can you be sure that the probability that he is a Schmenge is $\frac{9}{10}$? Suppose you were to "sample" the men by opening the door and grabbing the nearest man and noting whether or not he is a Schmenge. Or perhaps that doesn't sound "random" enough to you. Instead, you order the men inside to polka madly, to run about and bounce off the walls and to not stop; then you reach in a grab one. This sampling procedure becomes an additional premise, so that we have:

(e_{s_1}) Just 9 out of the 10 men are Schmenges

(e_{s_2}) M is a man in the room

(e_r) Men are arranged in the room randomly

The man M that I grab will be a Schmenge (10)

Here, I take "randomly" to mean, as it can only mean, that "I have no idea—I am ignorant—of how the men are arranged". To show this, first suppose that *all* we know is that there are men in a room, but *nothing* else. That is, our only evidence is e_r, which is just another way of saying, "There are men in the room, and I have no idea who they are or how they are arranged." Tacit in this is the idea that there may be some Schmenges in the room, which, of course, means that there may *not* be any. That is, e_r is equivalent to, "M may be true or it may be false". This is our old friend, the tautology t, which we have already seen adds nothing to the argument that would allow us to assign a definite probability to the conclusion.

Now, if you *did* know something about the way the men were arranged in the room, then this is evidence you *must* include in the list of premisses, where it would quite legitimately and naturally change the probability of M. But just saying your evidence is "random", or your experiment was "randomly" sampled, adds nothing. This should not be surprising, as Bayesians have long known that randomness is not a concept that is needed in experiments such as patient assignment in, e.g., clinical trials Howson and Urbach (1993).

It should also not be necessary to say that we do not need to assume anything about infinite "trials" of men in rooms to arrive at the probability of M. Some (objective) Bayesians try this kind of argument in an attempt justify their priors by invoking something called the *Principal Principle*, which states

> that if the objective, physical probability of a random event (in the sense of its limiting relative-frequency in an infinite sequence of trials) were known to be r and if no other relevant information were available, then the appropriate subjective degree of belief that the event will occur on any particular trial would also be r: (Howson and Urbach, 1993, p. 240).

Ignoring the fact that we can never *know* what happens after an infinite amount of time, and so can *know* r, or that we cannot imagine an infinite number of rooms filled with Schmenges, but pretending that we can, the Principal Principle says "$\Pr(M|\Pr(M) = r) = r$" (it adds the premiss "$\Pr(M) = r$" which is taken to be the 'objective' or physical probability of M), but which we can now see is just begging the question.

6. Mathematical attempts

The discussion to this point has been, of course, philosophical, an approach that will certainly induce discomfort in some because of its seeming lack of rigor. So to stiffen the discussion, here are two, of many, well known mathematical approaches to the problem of justifying equi-probable assignment. The results from both agree with the results from the statistical syllogism; nevertheless, I think both arguments fail, in the sense that they are circular, because they both assume the statistical syllogism or equi-probability in their proofs.

Both of these arguments start with the definite knowledge e that M can be decomposed into a finite number of possibilities (like our Schmeges) $M_1, M_2, \ldots, M_n$, $n < \infty$. This, again, already carries with it the knowledge that M is contingent.

First permutation argument (logical probability) Jaynes (2003): Introduce evidence e which states that either M_1 or M_2 or etc. M_n

can be true, but that only one of them can be true. In the case where M is a coin flip, the result can be either M_1="head" or M_2="tail". Thus, $\Pr(M_1 \vee M_2 \vee \cdots \vee M_n|e) = \sum_{i=1}^{n} \Pr(M_i|e)$. We want to assign the probabilities $\Pr(M_i|e)$ for $i = 1 \ldots n$. The set of possibilities is $M = \{M_1, M_2, M_3, \ldots M_n\}$. Let π be a permutation on the set $\{1, 2\}$. Let $M' = \{M_{\pi(1)}, M_{\pi(2)}, M_3, \ldots M_n\}$. That is, the set M and M' are the same except the first two indexes have been swapped in M'. The evidence e is fixed. Therefore, it must be that $\Pr(M_1|e)_M = \Pr(M_{\pi(2)}|e)_{M'}$ and $\Pr(M_2|e)_M = \Pr(M_{\pi(1)}|e)_{M'}$. Jaynes then makes a crucial step, which is to add evidence to e which states that the evidence is "indifferent" to M_1 and M_2, i.e.

> if it [the evidence] says something about one, it says the same thing about the other, and so it contains nothing that would give [us] *any reason* to prefer one over the other. (p. 39, emphasis mine)

Accepting this for the moment, e then says that our state of knowledge about M or M' is equivalent, including the order of the indexes. Thus, (note the change in indexes) $\Pr(M_1|e)_M = \Pr(M_{\pi(1)}|e)_{M'}$, $\Pr(M_2|e)_M = \Pr(M_{\pi(2)}|e)_{M'}$ *and* $\Pr(M_j|e)_M = \Pr(M_j|e)_{M'}, j = 3, \ldots, n$. Which implies $\Pr(M_1|e)_M = \Pr(M_2|e)_M$: that is to say, equi-probable prior assignment.

This argument is fine if what Jaynes says in the quotation holds. But we can see in it the presence of two tell-tale phrases, our old friends, "indifferent" and "no reason", which are used, and are needed, to justify the final step. This is just begging the question all over again, for how else could the evidence e be "indifferent"? That is, Jaynes has assumed the statistical syllogism as part of the evidence e, which is what he set out to prove.

Second permutation argument (finite exchangeability) Diaconis (1977): This argument is more mathematically complicated and was originally used to justify a use of de Finetti's representation theorem for finite sequences. Recall what this famous original theorem does: it gives an infinite sequence of exchangeable 0-1 variables a formal (induced) representation as a probability model with a *unique* measure of the probability model's parameters. The key, of course, is that the sequence *must* be infinite. Diaconis, after showing that some finite exchangeable sequences fail to be represented as probability models with unique measures, goes on to offer a proof for certain other finite exchangeable sequences that do.

I follow Diaconis (1977) as closely as possible, almost copying the theorem as it stands but using my notation; interested readers should consult the original if they are interested in the details, particularly since the original uses graphical notions which I do not elaborate here. Let $\mathcal{P}_n$ represent all probabilities on $M = \prod_{i=1}^{n} M_i$ where $M_i = \{0, 1\}, \forall i$, where M is a finite ($n < \infty$) sequence of 0-1 variables. $\mathcal{P}_n$ may be thought of as the probability models on M: it may be written in coordinate form by $p = (p_0, p_1, \ldots, p_{2^n-1})$ where p_j represents the outcome j where $0 \leq j < 2^n$ is the binary expansion of j written with n binary digits. Diaconis gives the example if $n = 3, j = 1$ refers to the point 001. Let $M(m, n)$ be the set of j with exactly m ones. The number of elements in $M(m, n)$ is $\binom{n}{m}$: this much is true—the number of elements in $M(m, n)$ *is* $\binom{n}{m}$—regardless of what the actual probabilities of any outcomes are.

Now, let $\mathcal{E}_n$ be the exchangeable measures in $\mathcal{P}_n$: $\mathcal{E}_n$ will take the place of the measure on $\mathcal{P}_n$'s 'parameters'. The theorem is stated thus: $\mathcal{E}_n$ has $n+1$ points $e_0, e_1, \ldots, e_n$, where e_m is the measure putting mass $1/\binom{n}{m}$ at each of the coordinates $j \in M(m, n)$ and mass 0 at the other coordinates. (Uniqueness of each point in $\mathcal{E}_n$ is also covered, but not of interest here.) How is this theorem proved?

e_n represents the measure of drawing n balls without replacement from an urn with n balls, m of which are marked 1, and $n - m$ marked 0, so each e_n is exchangeable. If e_n can be written as a proper mixture of other exchangeable points, has the form $e_n = pg_1 + (1 - p)g_0$, where $0 < p < 1$: also, g_1, g_0 must assign 0 probability to the outcomes which e_n assigns 0 probability. But because of exchangeability of the coordinates $j \in M(m, n)$ g_1 and g_0 must be equal. And because the probability for any $j \in M(m,n)$ must sum to 1—and here is the big assumption used in the proof—*the mass of each coordinate is* $1/\binom{n}{m}$.

Clearly, the intution that gave rise to these particular masses asserted in the proof came from the *fact* that the number of elements in $M(m,n)$ is $\binom{n}{m}$. However, other masses work too, as long as they sum to one and assign a probability of 0 to the other coordinates not in $M(m,n)$. For example, for $j \in M(m,n)$ assign $1/2m$ for the first m coordinates and $1/(2(\binom{n}{m} - m))$ to the remaining $\binom{n}{m} - m$ coordinates.

The reason that the $1/\binom{n}{m}$ mass was chosen is understandable, but there was no explicit reason for it (other than having the probabilities sum to 1) and the desire for symmetry and the equi-probable assignment. So again, the statistical syllogism/equi-probability is assumed.

7. Examples

Suppose you are considering M_1 and M_2 as the only competing models for some situation. Then, using the statistical syllogism and the logical probability assignments it implies as above, $\Pr(M_1 \vee M_2|e_s) = \Pr(M_1|e_s) + \Pr(M_2|e_s) = 1$ and $\Pr(M_1|e_s) = \Pr(M_2|e_s) = \frac{1}{2}$. This is the justification for starting with equal probability in model selection. After x is observed, then it is easy (in principle) to calculate $\Pr(M_1|x, e_s)$ and $\Pr(M_2|x, e_s)$.

It is no surprise that this is the same point reached by appealing to the Principle of Indifference (or even the Principle of Maximum Entropy for a finite number of model choices; Jaynes (2003)). The statistical syllogism gives the same answers as the Principle of Indifference, but not by the same route and, again, without the hidden assumptions or metaphysical baggage. The built-in question-begging of that principle is gone, and there is no appeal to subjectivity, which many find so distasteful.

Arguments against objectively assigned probabilities often centers on evidence external to that used by the authors. The best known example is Laplace's Rule of Succession used for finding the probability of seeing the sun rise tomorrow (read Jaynes (2003, chap. 18) for a fascinating look at this oft-cited topic: what follows here is a simplification and a change of what Laplace actually argued, he was trying to find the value of a continuous parameter; this changes here to the probability of a proposition). That is, certain evidence e is given for the truth of M, and a probability is then logically assigned to M. But the critic has in mind evidence e', which may contain e but also has arguments different than e, which would naturally lead to a different probability assignment. The probably assignment induced by e is then critized as "absurd" and the principles of logical probability which gave rise to it is rejected. This is overwhelmingly true for Laplace's example.

Along this same line, it is often heard that one must select priors, either on models or parameters, *before* seeing the data, lest the data somehow modify your pure 'prior' thoughts. This view is false, at least in the strict logical sense, because whether you apply the statistical syllogism before or after seeing your data it is irrelevant to the probability you assign. This probability assignment is based *only*, for example in the case of model selection, on the argument $M_1 \vee M_2$ is an outcome etc. The probability assignment "$\Pr(M|e_s) = 1/2$" is true no matter when in time it was made.

8. Conclusion

Logical probability is a much neglected subject in the statistical community. The only book in many years to appear on the subject is Jaynes (2003) (other books in the maximum entropy (MAXENT) traditional have been published, but these are not of the same scope as Jaynes). The Bayesian revolution from the later part of the 20th century, remarkable in many ways, mainly eschewed logical probability and fixed on the idea that probabilities are subjective.

It is this focus on subjectivity which has made statisticians comfortable using words like "ignorance", "fair" (though that term predates the revolution), "no reason", and especially "gamble", "indifferent","betting" and so on when they assign probabilities. These terms *feel* or are directly subjective; they are words to put your beliefs behind.

And once you have brought in belief, you make it difficult to discover the hidden assumptions behind your belief. But words like "ignorance" etc. are misleading, as was shown, in assigning probabilities, and they should be eliminated from the discussion.

In conversation, I have had it pointed out that the same results as the statistical syllogism can be had by appealing the the Principle of Maximum (information) Entopy. I agree with this. However, the apparatus of MAXENT is certainly not needed; and it is not clear that the assumptions etc. of that system are simpler than those of the statistical syllogism. The uniform probability assumption over events that is used to derive the statistical syllogism is just true; but is it true that the probability assignment should also maximize entropy? Maybe. But if you are trying to convince somebody of the correctness of logical probability, it should be clear that you introduce MAXENT at such an early stage, you are then asking a lot more from your audience.

I attempted to cast light on a few common hidden assumptions in the simplest possible situations. It is certainly not a complete answer to the question of how to assign probabilities in an objective way in all models. The statistical syllogism can clearly be applied to assign priors on probability model parameters when those parameters can take a finite number of values or states. The class of probability models which contain such parameters may or may not be very large, but it is at least not empty, though it of course does not contain the most frequently used probability models, such as those, say, from the exponential family. I make no attempt in this paper to justify, or modify, the use of the statistical syllogism in the case where the number of outcomes is countably or uncountably infinte, as in the case of parameters in models like the normal distribution. Jermyn (2005) is a good starting place for these topics.

But, however simple, the statistical syllogism clearly works and does not suffer from the same flaws as earlier arguments—arguments which may have given the same answers sometimes, but come loaded with hidden assumptions, assumptions which have been barriers to acceptance of Bayesian methods. Too, the statistical syllogism is completely objective and it eliminates any hint of "randomness" and "chance" and the complexity these terms imply. To this, much of this paper may seem like quibbling. After all, the results using the statistical syllogism agree with those (at least in these examples) that would be had appealing to "no reason" etc. But the impression of agreement is false. For one, people who would insist, for example, that all probability calculations cannot begin before a properly defined measure space has been carefully laid out, should not quail from a demand for the preciseness of language used in describing such models. More importantly, the terms "no reason" etc. are all improperly *defensive* and are negative. Using them with respect to assigning probabilities naturally creates a certain suspicion in those who hear them that something funny is going on.

The terms also over-emphasize, and even use when they should not, subjectivity. With the statistical syllogism, these problems disappear. For one: there is no subjectivity; the probability assignment follows logically from the information given. And the statistical syllogism emphasizes the *definite, positive* knowledge that exists. People, I believe, would be more inclined to to try to understand Bayesian methods (and the multitude of shortcomings of classical probability) if we who promote them are more careful—and justifiably positive—in our language.

Several attempts at mathematically assigning priors were shown to be begging the question. Other mathematical attempts at assigning equi-probable priors, such as those by e.g. Kerns and Székely (2007), which use signed measures, may be be useful, but since signed measures imply "negative probability," it is not clear that these attempts belong to applied (real-life) statistics.

References

Adams, E. W. (1998). *A Primer of Probability Logic*. CSLI Publications, Leland Stanford Junior University.

Bartha, P. and Johns, R. (2001). Probability and symmetry. *Philosophy of Science*, 68:S109–S122.

Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian Theory*. Wiley, New York.

Briggs, W. M. (2006). Broccoli reduces the risk of splenetic fever! the use of induction and falsifiability in statistics and model selection. *arxiv.org/pdf/math.GM/0610859*.

Campbell, S. and Franklin, J. (2004). Randomness and induction. *Synthese*, 138:79–99.

Carnap, R. (1950). *Logical Foundations of Probability*. Chicago University Press, Chicago.

Cox, R. T. (1961). *Algebra of Probable Inference*. Johns Hopkins University Press, Baltimore.

de Laplace, M. (1996). *A Philosophical Essay on Probabilities*. Dover, Mineola, NY.

Diaconis, P. (1977). Finite forms of de finetti's theorem on exchangeability. *Synthese*, 36:271–281.

Diaconis, P. (1988). *Recent progress on de Finetti's notions of exchangeability*, volume Bayesian Statistics 3, pages 111–125. Oxford

Univeristy Press.

Diaconis, P. and Freedman, D. (1980). Finite exchangeable sequences. *Annals of Probability*, 8(4):745–764.

Fisher, R. (1973). *Collected Papers of R.A. Fisher*, volume 2, chapter The logic of inductive inference, pages 271–315. University of Adelaide, Adelaide.

Franklin, J. (2001). Resurrecting logical probability. *Erkenntnis*, 55:277–305.

Hájek, A. (2007). A philosopher's guide to probability. In *Uncertainty: Multi-disciplinary Perspectives on Risk*. Earthscan.

Howson, C. and Urbach, P. (1993). *Scientific Reasoning: the Bayesian Approach*. Open Court, Chicago, second edition.

Hume, D. (2003). *A Treatise of Human Nature*. Oxford Univeristy Press, Oxford, corrected edition.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge.

Jeffreys, H. (1998). *Theory of Probability*. Oxford University Press, Oxford.

Jermyn, I. H. (2005). Invariant bayesian estimation on manifolds. *Annals of Statistics*, 33(2):583–605.

Kerns, G. J. and Székely, G. J. (2007). De finetti's theorem for finite exchangeable sequences. In review *Annals of Probability*.

Keynes, J. M. (2004). *A Treatise on Probability*. Dover Phoenix Editions, Mineola, NY.

Press, S. J. (2003). *Subjective and Objective Bayesian Statistics*. Wiley, New York, second edition.

Schechter, E. (2005). *Classical and Nonclassical Logic*. Princeton University Press, Princeton.

Stove, D. (1973). *Probability and Hume's Inductive Scepticism*. Clarendon, Oxford.

Stove, D. (1986). *The Rationality of Induction*. Clarendon, Oxford.

Strevens, M. (1998). Inferring probabilities from symmetries. *Noûs*, 3:22:231–246.

Williams, D. (1947). *The Ground of Induction*. Russell & Russell, New York.