

# Positive evidence for non-arbitrary assignments of probability

William M Briggs

*mattstat@gmail.com; Dept. Mathematics, 214 Pearce Hall, Central Michigan University, Mt. Pleasant, MI 48859*

**Abstract.** 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 to the truth of a proposition or event for which there is *no* evidence other than that the event is contingent, is contrasted with the assignment of probability in the case where there is *definite* 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 Indifference, 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.

**Keywords:** Induction; Logical Probability; Principle of Indifference; Probability assignment.

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. 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.

So, suppose  $p$  is a premise and  $q$  a conclusion to the argument from  $p$  to  $q$ : which is an argument that states, “(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 truth of the conclusion  $q$  like this:

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

Cox [1961], and like those in the logical probability tradition before him [2, 3, 4, 5], 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 argument from  $p$  to  $q$  is invalid or *non-deductive* and  $q$  only probable.

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 premises which are premises 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 [6], I examine induction in statistics and probability.

This article surveys 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 equiprobable 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.

## IGNORANCE

To stress: logical probability concerns itself with assigning probabilities to the conclusions of arguments with explicitly stated, and fixed, premises. It is easy to assign probability when the argument is deductive: the probability being 0 or 1. 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:  $t \wedge e_c \Rightarrow M$  is not valid (and is read “ $t$  and  $e_c$ , therefore  $M$  is true”). 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:

$$0 < \Pr(M \text{ (is true)} | t, e_c) < 1. \quad (2)$$

And that is the *best* we can ever do with only the definite knowledge that  $M$  is contingent (e.g., Keynes [4]). 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 (like  $M$ ) 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.

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 contingent, 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 (Bayesian) statisticians would not like to settle for (2), 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 (2), to announce some subjective opinion they might have about  $M$ , or even about how they would take bets over the truth of  $M$ . 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 ( $M$  is true) to the probability statement is therefore arbitrary (as many have felt before; e.g. [7]).

The argument can be modified, by inserting some additional evidence: say,  $e_o = "M$  is equally like to be true or false", so that we have " $e_o$ , therefore  $\Pr(M|t) = \frac{1}{2}$ ." This argument is dogmatic; nevertheless, it *is* valid; however, the premiss  $e_o$  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 premises are true, therefore the argument *is* valid: it is just of no use.

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 [4] and criticized in e.g. [8]. It is the "indifference" or "no reason" clause that is the start of troubles.

## NO REASON & INDIFFERENCE

The minor premiss in "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. [9]. 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 (therefore  $M$  is true)." The implicit conclusion is usually assigned probability  $\Pr(M) = \frac{1}{2}$ . The argument, I hope you can see, is not valid and the probability statement is arbitrary. Here's why. This argument *is* valid: " $M$  is true or it is false; therefore, 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." It should now be obvious that this 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 (2). 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, you *know*  $t \wedge e_c$ . So, despite our repeated insistence of "ignorance," we are back to the bounds of eq. (2), which is to say, right where we started.

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 [10] for a discussion of how Neyman used a similar trick applied

to interpreting classical confidence intervals).

## SYMMETRY

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] for the coin flip posed in the familiar form of a syllogistic argument (common example of a syllogism: “All men are mortal, Socrates is a man, etc.”):

Just 1 out of 2 of the possible sides are Heads

$M$  is a side

---

$M$  is a head

(3)

This argument, is, of course, invalid in the sense that the conclusion ( $M$  is a head) is not entailed by the premises. But we can assign the probability  $\Pr(M|e_s) = \frac{1}{2}$  that the conclusion is true, where  $e_s$  is the evidence of the two premises ( $e_s$  implies  $e_c$ ). This probability assignment, made explicit in the form of the statistical syllogism, is *derived* from assuming uniform probability across the individual events that make up the “sample space”: see a complete discussion in Stove [15] pp. 92-97, who credits Carnap [1950] with the first proof of this. In this case, this sides of a coin, or:  $\Pr(H|e_s) = \Pr(T|e_s)$ . 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.

Symmetry has often been used, and objected to, as a principle to assign probability, e.g. [12, 13, 14]. Arguments based on symmetry tend to be misleading because the examples are always chosen in such a way that they are “physically balanced” or uniformly

symmetric, which gives rise to a certain confusion. For example, Strevens [12] 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 [14]—and many, many other authors, invoking something about a "privileged partition"—and argue that Strevens's 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 (just  $t \wedge e_c$ ), the probability assignment is eq. (2) and not  $1/2$ . Also, all privileged 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, Strevens's die is physically symmetric, a very strong assumption that 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 *asymmetric*. And so, you can only appeal to the statistical syllogism ("Just 1 side of  $n$  etc.").

## WHITHER RANDOMNESS?

Suppose there are 10 men in a room and just 9 of these 10 are Schmenges.  $M$  is a man in the room. The probability that the conclusion " $M$  is a Schmenge" is true by the statistical syllogism is  $\frac{9}{10}$ . But 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,  $e_r =$  "Men are arranged in the room randomly."

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" [17]. 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.

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  $r$ : [8, 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.

## MATHEMATICAL ATTEMPTS

The following arguments start with the definite knowledge  $e$  that  $M$  is contingent and can be decomposed into a finite number of possibilities (like coin flips or rooms of Schmenges)  $M_1, M_2, \dots, M_n, n < \infty$ .

**First permutation argument** (logical probability) [5]: 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,  $\Pr(M_1 \vee M_2 \vee \dots \vee M_n|e) = \sum_{i=1}^n \Pr(M_i|e)$ . We want to assign the probabilities  $\Pr(M_i|e)$  for  $i = 1 \dots n$ . The set of possibilities is  $M = \{M_1, M_2, M_3, \dots M_n\}$ . Let  $\pi$  be a permutation on the set  $\{1, 2\}$ . Let  $M' = \{M_{\pi(1)}, M_{\pi(2)}, M_3, \dots M_n\}$ . 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  $\Pr(M_j|e)_M = \Pr(M_j|e)_{M'}, j = 3, \dots, n$ . Which implies  $\Pr(M_1|e)_M = \Pr(M_2|e)_M$ : that is to say, equiprobable 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 uniform probability (and thus, the statistical syllogism) as part of the evidence  $e$ , which is what he set out to prove.

**Second permutation argument** (finite exchangeability) [18]: Space does not permit a detailed examination of this argument; details will appear in a future paper.

## DISCUSSION

Suppose you are considering  $M_1$  and  $M_2$  as the only competing models for some situation. Then, using the statistical syllogism ("Just  $M_1$  or  $M_2$  etc.") 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 so  $\Pr(M_1 | e_s) = \Pr(M_2 | e_s) = \frac{1}{2}$ . 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 [5]. 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.

In conversation, I have had it pointed out that the same results as the statistical syllogism can be had by appealing to the Principle of Maximum (information) Entropy (MAXENT), or via other complex mathematical arguments. I agree with this. However, the additional complex apparatus of MAXENT, with its own set of axioms and assumptions, is certainly not needed. The uniform probability assumption over events that is used to derive the statistical syllogism is just true; but is it also true that the probability assignment *should* maximize entropy? Maybe, but what do we mean by "should"? If you are trying to convince somebody of the correctness of logical probability, it should be clear that if you introduce MAXENT at 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. This paper is certainly not a complete answer to the question of how to assign probabilities in an objective way in all situations. 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 infinite, as in the case of parameters in models like the normal distribution. [19] 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—and mysticism—that 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 this 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 (such as contingency and know number of possible outcomes). People, I believe, would be more inclined to try to understand Bayesian methods (and be made aware of the multitude of shortcomings of classical probability) if we who promote them are more careful—and justifiably positive—in our language.

## ACKNOWLEDGMENTS

I am heavily indebted to an anonymous referee who helped to clarify the arguments used in this paper.

## REFERENCES

1. R. T. Cox, *Algebra of Probable Inference*, Johns Hopkins University Press, Baltimore, 1961.
2. M. de Laplace, *A Philosophical Essay on Probabilities*, Dover, Mineola, NY, 1996.
3. H. Jeffreys, *Theory of Probability*, Oxford University Press, Oxford, 1998.
4. J. M. Keynes, *A Treatise on Probability*, Dover Phoenix Editions, Mineola, NY, 2004.
5. E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, 2003.
6. W. M. Briggs, [arxiv.org/pdf/math.GM/0610859](https://arxiv.org/pdf/math.GM/0610859) (2006).
7. R. Fisher, *Collected Papers of R.A. Fisher*, University of Adelaide, Adelaide, 1973, vol. 2, chap. The logic of inductive inference, pp. 271–315.
8. C. Howson, and P. Urbach, *Scientific Reasoning: the Bayesian Approach*, Open Court, Chicago, 1993, second edn.
9. J. M. Bernardo, and A. F. M. Smith, *Bayesian Theory*, Wiley, New York, 2000.
10. J. Franklin, *Erkenntnis* **55**, 277–305 (2001).
11. D. Williams, *The Ground of Induction*, Russell & Russell, New York, 1947.
12. M. Strevens, *Noûs* **3:22**, 231–246 (1998).
13. P. Bartha, and R. Johns, *Philosophy of Science* **68**, S109–S122 (2001).
14. A. Hájek, “A philosopher’s guide to probability,” in *Uncertainty: Multi-disciplinary Perspectives on Risk*, Earthscan, 2007.
15. D. Stove, *The Rationality of Induction*, Clarendon, Oxford, 1986.
16. R. Carnap, *Logical Foundations of Probability*, Chicago University Press, Chicago, 1950.
17. S. Campbell, and J. Franklin, *Synthese* **138**, 79–99 (2004).
18. P. Diaconis, *Synthese* **36**, 271–281 (1977).
19. I. H. Jermyn, *Annals of Statistics* **33**, 583–605 (2005).