May 15th, 9:00 AM - May 17th, 5:00 PM

Enumerative Induction

David Hitchcock
McMaster

ENUMERATIVE INDUCTION

David Hitchcock
Department of Philosophy
McMaster University

Abstract:
In an enumerative induction we project a property found in all the examined members of a class to a hitherto unexamined member of that class. I consider an unresolved disagreement between Stephen Thomas and John Nolt about how likely the conclusion of one example of such reasoning is, given the premisses. Reflection on their controversy, and on the example, indicates that many textbook commonplaces about enumerative inductions are false. In particular, uniformity of the examined members of a class does not necessarily make it highly likely that the next member will have the target property; this proposition may still be unlikely, or on the other hand be quite definite, depending on our background knowledge. Also, enumerative induction need not rely on any general assumption about the uniformity of nature or the resemblance of the future to the past.

***

Introduction

I propose to discuss the logic of enumerative induction. By an enumerative induction, I mean an inference of the following form:

> All examined Ms are B.
> Therefore, the next examined M will be B.

M is a variable ranging over kinds, or more generally over sets of individuals. B is a variable ranging over properties. Here is an example of such an inference:

> There were fifty marbles in this jar. The marbles were thoroughly stirred and mixed before sampling. The first forty-nine marbles that I removed, all of which were chosen at random, were all blue. Therefore, the next marble that I choose from the jar will also be blue. (Thomas 1997: 148)

Enumerative induction is a humble form of reasoning. All of us use it frequently, whenever we extrapolate from uniform past experience. Its logic *seems* obvious. It is not a deductively valid form of reasoning, we think, because the meaning of the premiss leaves open the possibility that the conclusion is false. But it can be inductively strong, as we say. By this we mean that, if the premiss cites enough instances of the kind, and we have some reason to think the examined instances are representative of the whole kind with respect to our target property, then the truth of the premisses makes it highly likely that the conclusion is true. Some reference may be made here to the need for a more sophisticated view of induction than that possessed by Russell's chicken (Russell, 1912), who confidently came up to the farmer every morning expecting to get its daily handout of grain, until one day the farmer cut its head off.

We may acknowledge that there is a problem of induction, first raised by Hume, a theoretical problem of justifying this and other sorts of inductive inference. Our inductive practices, we might say, assume that the future always resembles the past, or that nature is uniform, assumptions which we cannot prove to be true but for making which some reason can be given.

I shall argue, through reflection on the above example, that most of the above commonplaces are false.

1. The Thomas-Nolt Controversy

In an exchange some years ago in *Informal Logic*, Stephen Thomas and John Nolt gave wildly different answers to the question how likely it is that the conclusion of our example is true, assuming that the premisses are true. Thomas wrote as follows:

> If one calculates the probability that the remaining ball[1] is blue given that the first forty-nine drawn at random have been blue, the probability is well in excess of 80%. (Thomas 1984: 32)

Thomas made this claim as part of his case against John Nolt's proposal "to measure strength of reasoning by the proportion of conceivable worlds in which the conclusion is true among worlds in which the premises are true" (Nolt 1984). In reply, Nolt made the following comment about Thomas' claim of a probability well in excess of 80%:

> This claim seems plausible, but in fact there is no way to calculate such a probability from the information Thomas gives... if we do not know the number of blue balls in the urn initially and we make only the assumptions that Thomas gives, then no calculation will yield the probability that the remaining ball is blue. We can, without violating any mathematical law, assign that proposition any probability we like. (Nolt 1985: 56)

It is noteworthy that Thomas nowhere gives the method of calculation by which he gets a probability well in excess of 80%. On the other hand, Nolt gives no justification either for his striking claim that we can assign any probability we like, consistently with the premisses of the argument, to the proposition that the next marble that I choose from the jar will be blue. He says only:

> We intuitively feel that the probability of the 50th ball being blue ought to be high, given the premises; but this feeling is based on what Hume would have called our "habit" of extrapolating constant conjunction. It is based, in other words, on our implicit assumption of the uniformity or likely uniformity of the urn's contents, not on any mathematical principle... If we do not make such an assumption, however, then we have no basis for thinking that the colour of the first 49 balls is in any way relevant to the colour of the 50th. Since there are many possible colors other than blue, the inference is genuinely weak. (1985: 56)

Nolt defends the possible worlds test as helping us to notice the assumption.

Although Nolt had the last word, Thomas continues to maintain that the premisses of such an argument give strong support to its conclusion. About a similar example, he writes the following in the 1997 edition of his text:

> The large proportion of marbles examined and the fact that the marbles were thoroughly stirred before sampling and were chosen at random, all contribute to the strength of this reasoning. Yet the reasons do not make the truth of the conclusion totally certain. It remains possible that one (or even both) of the two marbles still in the bag is not made of clear glass. [In the example under discussion,

> 48 marbles selected at random without replacement from a bag of 50 marbles were clear glass.—DH] Although we can imagine that the reasons are true while the conclusion is false, this situation is unlikely. Consequently, the step from these reasons to the conclusion is rated as *strong* [italics in original—DH]. Unlikely as it may be, the logical possibility that a remaining marble is not clear glass (despite the fact that the first 48 drawn at random were clear glass) makes this step of reasoning less than 100 per cent certain, the highest possible degree of strength. (Thomas, 1997: 131)

My reason for discussing Thomas' example, then, is that there is an unresolved disagreement about how likely the conclusion is, given the premisses, a disagreement whose resolution may shed light on our theorizing generally about enumerative induction. The example may of course seem absurd. Who cares, someone might say, how likely the premisses make the conclusion? All one has to do is pull out the last marble and find out what colour it is. But picking a marble from a jar—or more exactly a ball from an urn—has been used since Laplace (1952/1819) as a metaphor for conducting an experiment or making a systematic observation. When we make extrapolations from experiments or observations, there is typically no limit to the number of further instances which might be observed. In such cases we cannot pull out all the remaining instances and look at them.

2. Improbability of true premisses with a false conclusion

Thomas appeals to its being unlikely "that the reasons are true while the conclusion is false" (1997:131); in other words, he regards as unlikely the following situation: 49 marbles selected at random without replacement from this jar containing 50 marbles are blue, while the non-selected marble is not blue. But it is unclear what sort of likelihood we are being asked to estimate when we contemplate such a situation, and this unclarity perhaps explains how Thomas came up with his probability "well in excess of 80%" that the conclusion is true given that the reasons are true (1984: 32). It sounds as if we are being asked to estimate the conditional probability that all 49 marbles selected at random without replacement are blue, given that the jar contains 49 blue marbles and one non-blue marble. And this probability is 1/50, or .02.[2]

Since the probability of a set of mutually exclusive and jointly exhaustive events sums to 1, then we can work out by subtraction that the probability of a blue marble being chosen last is .98. And we might then be inclined to take this as the likelihood that the conclusion of our argument is true, given that the reasons are true. But this would be a mistake. For the argument corresponding to the probability of .98 is the following argument:

> 49 marbles are selected at random without replacement from a jar containing 49 blue marbles and one non-blue marble.
> Therefore, the marble left in the jar is blue.

And that is clearly a different argument from the one whose inference we are evaluating, since it includes in the premiss the assumption that one of the original 50 marbles is not blue and it does not include the information that the 49 examined marbles are blue (information with which the premiss and the conclusion are in fact jointly incompatible).

What has gone wrong? In construing as we did the likelihood that the reasons are true while the conclusion is false, we have treated the problem as one of estimating the probability of a given outcome of a not yet completed stochastic (indeterministic) process. But our problem does not involve any such stochastic processes. We are supposing that the 49 marbles have already been selected, and that they are all blue. There is no indeterminacy about their colour, or about the colour of the remaining marble in the jar. It is true that, prior to their selection, if 49 marbles in the jar were blue and one was non-blue, the result we observed was very unlikely, with a probability of only .02. It is also true that, if all 50 marbles in the jar were blue prior to the selection of the 49, then the result we observed was inevitable, with a probability of 1.[3]

But it is still unclear what these two conditional probabilities have to do with the probability we are seeking to determine.

Our probability is in fact some sort of epistemic probability, the degree of certainty that it is reasonable for us to attach to a proposition given certain information. An initial approximation to understanding this kind of probability is to think of it in terms of the betting odds one should be willing to give that the proposition is true. Given that the first 49 marbles selected from the jar are blue, what would be fair odds if one person bet another person a dollar that the 50th marble will also be blue? The basic approach to understanding this sort of probability was worked out by Frank P. Ramsey in a paper written in 1926 and published posthumously in 1931 (Ramsey, 1990/1931). For a recent moderate defence and elaboration of the approach, see Kaplan (1996).

Note that this is not probability in any frequency sense, as when we say that the probability is .25 that a playing card in a standard deck is a heart, meaning that one in every four cards in the deck is a heart. We are dealing with the probability that a particular proposition about a particular marble is true.[4]

Once we recognize that we are dealing with an epistemic probability, we can see that the word "while" is misleading. We are not dealing with the probability of a conjunction, or the complement of the probability of a conjunction. The epistemic probability of interest is not the probability that the reasons are true and the conclusion true. That would involve some sort of assessment of how likely it is that the first 49 marbles picked at random without replacement from this jar are blue. But we have no background information on the basis of which to calculate such a probability. We are simply given that it is so. Nor is the probability of interest the complement of the probability that the reasons are true and the conclusion false, for similar reasons.

Rather, what we are trying to work out is the conditional probability that the conclusion is true, given that the premisses are true. Equivalently, it is the complement of the conditional probability that the conclusion is false, given that the premisses are true.[5]

We might be tempted to reason that our conclusion is probably true, since the evidence is otherwise so unlikely to occur. Our reasoning would have the following form:

> The evidence rules out all but two possibilities.
> One of these possibilities when combined with the evidence means that what we observed was bound to occur before we started.
> The other when combined with the evidence means that what we observed was very unlikely before we started.
> Therefore, the first possibility is probably actual.

If we consider the parallel inductive generalization, we could consider our evidence as settling a decision between two general hypotheses: that all 50 marbles are blue and that 49 were blue and one non-blue. Our reasoning would have the following form:

> The evidence rules out all but two hypotheses.

On one hypothesis, the evidence was bound to occur.
On the other hypothesis, the probability of the evidence was .02.
Therefore, the first hypothesis is probably correct.

This is an argument of the form:

E.
If E, then H or H'.
If H, then E.
If H', then very probably not E.
Therefore, probably H.

(H is a variable ranging over hypotheses, E a variable ranging over evidence.) This form of argument looks legitimate. But we shall have occasion to question its validity.)

3. Bayes' theorem

We have seen that the conditional probability we seek is not the probability that a certain result will occur given a certain hypothesis, but the probability that a hypothesis is true given a certain result. From now on I shall call this the *posterior probability*. For such posterior probabilities, we must have recourse to Bayes' theorem, which enables us to calculate the posterior probability if we have three pieces of information. First, we need the *prior probability* of the hypothesis, that is, the probability that the hypothesis is true, given our background knowledge independently of the new evidence. (I shall call this "P(H/K)", where P() is the probability function, H is the hypothesis and K is our background knowledge apart from the new evidence.) Second, we need the *posterior likelihood*, the likelihood of the evidence on the assumption that the hypothesis is true, again assuming the same background knowledge which we have independently of the new evidence. (I shall call this "P(E/H & K)", where E is the new evidence.) Third, we need the so-called *prior likelihood* of the evidence, that is, the likelihood that the evidence is true on the assumption of our background knowledge, without assuming the truth of the hypothesis under investigation. (I shall call this "P(E/K)".) Bayes' theorem tells us that the posterior probability of a hypothesis on new evidence is its prior probability multiplied by the ratio of the posterior likelihood of the evidence to its prior likelihood:

$$P(H/E \,\&\, K) = P\,(H/K) \times P(E/H \,\&\, K) \div P(E/K)$$

Bayes' theorem rests on a definition of a conditional probability P(A/B) as the result of dividing the probability that both A and B obtain by the probability that B obtains, provided that this latter probability is not zero. Thus, for example, the probability that a card in a deck of playing cards is a heart, given that it is red, is the probability that a card is both a heart and red (13/52) divided by the probability that it is red (26/52), or 1/2. If one replaces the conditional probabilities in Bayes' theorem according to this definition, one sees that the theorem is correct, provided that the hypothesis under investigation has a non-zero prior probability, and assuming that probabilities can be assigned to the component propositions.

The justification of applying Bayes' theorem to the calculation of our epistemic probability is that this epistemic probability satisfies the axioms of the probability calculus (the Kolmogoroff axioms). For a moderate justification of this claim, see Kaplan (1996).

4. First solution: assumption of independence

Now let us attempt to apply Bayes' theorem to our problem.

1) *Prior probability of the hypothesis*: What is the probability that the 50th marble to be drawn from the jar is blue, given our background knowledge but not given the evidence of the colour of the first 49 marbles drawn? We have been told nothing about the way the marbles were put in the jar, who put them there, for example, or how they were selected. But we know that a marble can be any colour at all, or a combination of colours. So we might suppose that the prior probability, independently of our new evidence, that the 50th marble to be drawn is blue is the same as the probability that a randomly selected marble from the set of all marbles now existing on earth is blue. Let us call this probability "p".

2) *Prior likelihood of the evidence*: Making no assumption of an inclination on the part of whoever filled the jar to prefer uniformity to diversity, I would put the prior likelihood of the first 49 marbles being blue at the product of the likelihood of any particular marble's being blue, in other words at p to the 49th power (p^{49}). That is, I am assuming that the colour of any marble in the jar is independent of the colour of the other marbles in the jar.

3) *Posterior likelihood of the evidence*: Assuming our background knowledge, what is the likelihood, given that the 50th ball to be drawn is blue, that the first 49 will also be blue? Not knowing anything about how the marbles were selected to be put in the jar, I find it difficult to say. I have no particular reason for thinking that marbles in a single jar are likely to be of uniform colour. It is quite possible to put in a jar marbles of a great variety of colours. So I would personally be inclined to think that the colour of the 50th marble picked from the jar would not change the likelihood of any particular sequence of colours drawn in the first 49 drawings. In other words, I would put the posterior likelihood also at p to the 49th power (p^{49}).

Putting these numbers in the formula, one sees immediately that the probability of the 50th ball being blue, given that the first 49 drawn at random without replacement are blue, is p. That is, it is exactly the same as the prior probability. The evidence of drawing 49 balls in a row which are blue has made no difference at all to the probability that the 50th one is blue!

If this result is correct, then Nolt is correct in maintaining that we can make the posterior probability anything we like, since we can cook up details of the situation which will make the value of p whatever we like. (If p is zero, then Bayes' theorem does not apply. But new evidence cannot change the probability of something which has a zero probability on the basis of background knowledge.)

This result is so counter-intuitive that it demands explanation. Has there not been some mistake in the calculation? Does Bayes' theorem really apply to such situations?

The explanation, I believe, is that, in filling out the background knowledge for calculating the prior likelihood and posterior likelihood, I have assumed that the colour of each marble in the jar is independent of the colour of every other marble. I have treated the drawing of a marble from the jar in the same way as one treats the flipping of a fair coin or the turning of a fair roulette wheel. Just as a long run of heads does not change the odds that a fair coin will come up heads the next time, so a long run of blue marbles drawn from the jar does not change the odds that the next marble drawn from the jar will be blue. To think otherwise is to subscribe to the gambler's superstitious and fallacious belief that there are such things as streaks, that once something starts coming up a certain way several times in succession in a chance setup, it is more likely than usual to keep coming up that way.

5. Second solution: assumption of uniformity

Nolt claims that the intuitive belief that drawing 49 blue marbles in succession from a jar containing 50 marbles increases the likelihood that the 50th marble will be blue depends on an assumption about the uniformity in colour, or likely uniformity in colour, of the jar's contents.

Suppose, then, we build into our background knowledge some assumption that the marbles in the jar are, or are likely to be, uniform in colour.

First consider the assumption that the marbles are all of one colour. In this case, the premisses of our argument make the conclusion quite certain. The probability that the 50th marble to be selected will be blue is 1, given that the first 49 are blue. But notice that it does not take observation of 49 marbles to get this high probability. One is enough, if we assume that all the marbles in the jar are the same colour. And intuitively this seems not quite right. For when would we be in a position to know in advance that all the marbles in a jar are the same colour, and on what basis?

We would be in such a position (or close to it) if we had a justified theory from which we could deduce uniformity of colour. We have such theories in real life when we perform experiments or observations on instances of natural kinds in order to determine the value of some variable which theoretical considerations indicate is uniform for that kind. For example, we might be measuring the solubility of (pure) sulphur in (pure) water. Ignoring for the moment measurement error and effects of trace impurities, we can say that one measurement will give us the answer we need.

Notice that in such a situation the conclusion follows definitely, as we might say, from the premiss, not probabilistically. So the common statement that all inductive enumerations are probabilistic is incorrect. In cases where we have a good theoretical basis for expecting uniformity, the conclusion follows non-probabilistically. This does not mean that our acceptance of the conclusion is infallible, because we may be mistaken about a premiss or about our background "knowledge". But the inference is not probabilistic.

6. Third solution: assumption of likely uniformity

A marble in a jar, however, is not like a pure sample of sulphur combined with a pure sample of water, at least with respect to its colour. Let us suppose that our background knowledge tells us that there is a 50% probability that the marbles in this jar are of uniform colour. This new information makes no difference to the prior probability that the last marble chosen is blue; it remains at p. But the prior likelihood that the first 49 marbles selected are blue is now much higher. If we assume that the alternative to complete uniformity of colour is random variation, then the prior likelihood is $(p \times .5) + (p^{49} \times .5)$. The posterior likelihood, on the other hand, is $.5 + (p^{49} \times .5)$, since the observation of one blue marble makes it necessary that, if the marbles are of uniform colour, they are all blue. The posterior probability of the hypothesis, given this new piece of background information, is $p \times [.5 + (p^{49} \times .5)] \div [p \times .5) + (p^{49} \times .5)]$, which simplifies to $(1 + p^{49}) \div (1 + p^{48})$, or practically 1, if we assume that p is considerably less than 1.

Changing the likelihood that the marbles are uniform in colour does not affect the posterior probability of the hypothesis very much, provided the alternative is random variation among the colours marbles are known to have. Even if the likelihood of uniformity in colour is as low as 10^{-10}, the posterior probability of the hypothesis turns out to be $(1 + [10^{10} p^{49}] - p^{49}) \div (1 + [10^{10} p^{48}] - p^{48})$, which is still practically one, assuming that p is a small fraction of 1. (That is, assuming that far fewer than half the marbles on earth right now are blue.)

Even entertaining the alternative that the marbles in the jar are of two colours, with the relative proportion of these colours determined by chance, does not reduce the posterior probability of our hypothesis much below 1. Suppose our background knowledge, prior to drawing the first 49 marbles, makes two colours and one colour equally likely. Then the prior likelihood of the evidence is $(.5 \times p) + (.5 \times p \times .5^{48})$, and the posterior likelihood is $(.5 \times 1) + (.5 \times 0)$. So the posterior probability of the hypothesis is $1 \div (1 + .5^{48})$, which is still practically one.

So, if we assume as part of our prior background knowledge that the marbles in the jar are likely to be uniform in colour, we do indeed end up with the intuitive result that drawing 49 blue marbles in a row from the jar makes it almost certain that the last marble will be blue.

7. Fourth solution: no assumption of (likely) uniformity

But Nolt's claim is still not vindicated. We need to show that the intuitive result *requires* an assumption of uniformity, or likely uniformity, of the colour of the jar's contents. What if we made no such assumption in advance?

Suppose that, according to our background knowledge, it is highly unlikely before we select any marbles from the jar that they are all of the same colour. Let the probability that this is so be q. Suppose also that our background knowledge makes it equally unlikely before we select any marbles from the jar that they are of only two colours. Then we have the following values:

$$P(H/K) = p$$
$$P(E/H \& K) = q + (q \times 0) = q$$
$$P(E/K) = (p \times q) + (p \times q \times .5^{48}) = p \times q \times (1 + .5^{48})$$

Thus the posterior probability, P(H/E & K), is $1 \div (1 + .5^{48})$, which is again very close to 1. Hence the intuitive belief that the observed result makes it highly likely that the last marble in the jar is blue does not necessarily depend on a prior assumption that the contents of the jar are uniform, or likely to be uniform, in colour. In our example, we can set q as low as we please.

Conclusion

The extent to which the premisses of an enumerative induction make its conclusion probable thus varies with our background knowledge about the kind and attribute with respect to which the inference is being made. Without assuming some such background knowledge, we cannot calculate a definite probability. But we can assign the conclusion a high probability, given the premisses, even if our background knowledge does not include any assumption that nature is in general uniform or that the future resembles the past. Nor need it include an assumption that this particular kind is uniform, or likely to be uniform, with respect to this attribute. On the other hand, our background knowledge may be of such a sort that the conclusion of the inductive enumeration follows quite definitely from its premisses, not merely probabilistically.

Notes

1. In adapting his textbook example, Thomas changed the marbles to balls.

2. The easy way to calculate this probability is to calculate the probability that the non-blue ball will be chosen last. Since the selection is random, there is an equal chance of the non-blue ball being chosen on each draw, namely one out of 50; therefore, this is the probability that the non-blue ball will be drawn last. The same result follows if, using the multiplication rule for calculating the probability of the conjunction of independent events, we multiply together the probability of choosing a blue ball on the first draw (49/50), the probability of choosing a blue ball on the second draw if a blue ball has been chosen on the first draw (48/49), and so on.

3. These calculations assume a number of facts about marbles in general and about this jar in particular which are not explicitly stated in the premisses of the original argument:

1) The number of marbles in a jar changes only by removing or adding marbles to the jar. (Contrast the number of drops of water in a glass, or the number of rabbits in a cage, or the number of mothballs in an open jar.)
2) Marbles do not spontaneously change colour on their own. (Contrast chameleons.)
3) No marbles left or entered the jar after it had 50 marbles except for the 49 I removed.
4) In particular, I did not replace any of the 49 marbles I removed.

4. Some theorists want to give our sort of probability a frequency interpretation. For some cogent arguments against this approach, see Hacking (1965).

5. Such conditional probabilities are not the same as the probability of a material conditional. To take a simple example from playing cards, the conditional probability that a card is a queen, given that it is a face card, is .33. But the probability of the material conditional that a card is a queen if it is a face card is the probability of the disjunction that a card is a queen or not a face card, which is .85.

Bibliography

Hacking, Ian. *Logic of Statistical Inference*. Cambridge: Cambridge University Press, 1965.

Kaplan, Mark. *Decision Theory as Philosophy*. Cambridge: Cambridge University Press, 1996.

Laplace, Pierre Simon, Marquis de. *Philosophical Essay on Probabilities*. Translated from the sixth French edition by Frederick Wilson Truscott and Frederick Lincoln Emory, with an introductory note by E. T. Bell. New York: Dover, 1952. (First French edition published in 1819.)

Nolt, John. "More on induction and possible worlds: replies to Thomas and Kahane." *Informal Logic* 7, (1985) 1: 55-60.

Ramsey, F. P. *Philosophical Papers*. Cambridge: Cambridge University Press, 1990. (Chapter 4, "Truth and probability", with four postscripts, was first published in 1931.)

Russell, Bertrand. *The Problems of Philosophy*. London: Williams and Norgate, 1912. Chapter VI, "Induction".

Thomas, Stephen N. "Degrees of validity and ratios of conceivable worlds." *Informal Logic* 6, (1984) 3: 31-34.

Thomas, Stephen N. *Practical Reasoning in Natural Language*, 4th edition. Upper Saddle River, NJ: Prentice-Hall, 1997. (Previous editions published in 1973, 1981 and 1986.)

View Comentary by R. Girle

View Index of Papers and Commentaries

Return to Main Menu