ABSTRACT. A major difficulty for currently existing theories of inductive inference involves the question of what to do when novel, unknown, or previously unsuspected phenomena occur. In this paper one particular instance of this difficulty is considered, the so-called sampling of species problem.

The classical probabilistic theories of inductive inference due to Laplace, Johnson, de Finetti, and Carnap adopt a model of simple enumerative induction in which there are a prespecified number of types or species which may be observed. But, realistically, this is often not the case. In 1838 the English mathematician Augustus De Morgan proposed a modification of the Laplacian model to accommodate situations where the possible types or species to be observed are not assumed to be known in advance; but he did not advance a justification for his solution.

In this paper a general philosophical approach to such problems is suggested, drawing on work of the English mathematician J. F. C. Kingman. It then emerges that the solution advanced by De Morgan has a very deep, if not totally unexpected, justification. The key idea is that although 'exchangeable' random sequences are the right objects to consider when all possible outcome-types are known in advance, exchangeable random partitions are the right objects to consider when they are not. The result turns out to be very satisfying. The classical theory has several basic elements: a representation theorem for the general exchangeable sequence (the de Finetti representation theorem), a distinguished class of sequences (those employing Dirichlet priors), and a corresponding rule of succession (the continuum of inductive methods). The new theory has parallel basic elements: a representation theorem for the general exchangeable random partition (the Kingman representation theorem), a distinguished class of random partitions (the Poisson-Dirichlet process), and a rule of succession which corresponds to De Morgan's rule.

1. INTRODUCTION

An important question rarely discussed in accounts of inductive inference is what to do when the utterly unexpected occurs, an outcome for which no slot has been provided. Alternatively – since we know this will happen on occasion – how can we coherently incorporate such new information into the body of our old beliefs? The very attempt to do so seems paradoxical within the framework of Bayesian inference, a theory of consistency between old and new information.

This is not the problem of observing the ‘impossible’, that is, an event whose possibility we have considered but whose probability we judge to be 0. Rather, the problem arises when we observe an event
whose existence we did not even previously suspect; this is the so-called problem of 'unanticipated knowledge'. This is a very different problem from the one just mentioned: it is not that we judge such events impossible – indeed, after the fact we may view them as quite plausible – it is just that beforehand we did not even consider their possibility. On the surface there would appear to be no way of incorporating such new information into our system of beliefs, other than starting over from scratch and completely reassessing our subjective probabilities. Coherence of old and new makes no sense here; there are no old beliefs for the new to cohere with.

A special instance of this phenomenon is the so-called sampling of species problem. Imagine that we are in a new terrain, and observe the different species present. Based on our past experience, we may anticipate seeing certain old friends – black crows, for example – but stumbling across a giant panda may be a complete surprise. And, yet, all such information will be grist to our mill: if the region is found rich in the variety of species present, the chance of seeing a particular species again may be judged small, while if there are only a few present, the chances of another sighting will be judged quite high. The unanticipated has its uses.

Thus, the problem arises: How can the theory of inductive inference deal with the potential existence of unanticipated knowledge, and, how can such knowledge be rationally incorporated into the corpus of our previous beliefs? How can we predict the occurrence of something we neither know, nor even suspect, exists? Subjective probability and Bayesian inference, despite their many impressive successes, would seem at a loss to handle such a problem given their structure and content. Nevertheless, in 1838 the English mathematician Augustus De Morgan proposed a method for dealing with precisely this difficulty. This paper describes De Morgan's proposal and sets it within the context of other attempts to explain induction in probabilistic terms.

The organization of the paper is as follows. The second section gives some historical background and briefly describes De Morgan's rule. As will be seen, although the statement of the rule is unambiguous, its justification – at least, as described by De Morgan – is unclear, and our goal will be to understand why De Morgan's rule makes sense. We begin this task by briefly reviewing, in the third section of the paper, the classical analysis of the inductive process in probabilistic terms. This is very well-known material, and our goal here is simply to set