those in which the treasure is followed by a single block length of thread, and those in which the treasure is left isolated from the thread. The first type is three times as frequent as the second. Given that the type is of the first kind, we know where the treasure must be. And if it is not, then a uniform ("semi-prior") distribution among the three remaining positions gives a perfectly reasonable answer.

REFERENCE


Comment

G. E. P. BOX and G. C. TIAO*

Stone himself has suggested that the first problem he poses is academic and we agree with this. We do, however, feel that the treasure hunt in Flatland needs further analysis.

After the woman and soldier left the Bayesian statistician (who was distantly related to the soldier by marriage)\(^2\) at intersection O, he considered the problem as follows: The data y will consist of the direction in which the thread is pointing from the endpoint \((N, S, E, W)\). The parameter of interest \(\theta\) concerns the position of the treasure relative to the endpoint \((N, S, E, W)\). There is also a nuisance parameter \(\phi\), crucial to the problem, which determines whether the thread has been pulled back along its original path. The state of the parameter \(\phi\) can be expressed by letting \(\phi = 0\) if the thread has not been pulled back, and \(\phi = 1\) if it has.

The Bayesian statistician recognizes this as a problem of the type mentioned, for example, by Fisher [2, p. 19] and referred to on [1, p. 12] of our book, where the conduct of the experiment determines the prior probabilities exactly. Specifically, because of the manner in which the expedition has been conducted, he knows that, a priori

i. \(p(\theta = N) = p(\theta = S) = p(\theta = E) = p(\theta = W) = \frac{1}{4}\),

ii. \(p(\phi = 0) = \frac{1}{4}\) and \(p(\phi = 1) = \frac{1}{4}\),

iii. \(\theta\) and \(\phi\) are independent.

After the Bayesian statistician was allowed to follow the thread to its endpoint, he observes \(y\) and applies Bayes theorem in the usual way so that

\[ p(\theta, \phi | y) \propto p(y | \theta, \phi) p(\theta)p(\phi). \]

Suppose he found that \(y = E\). The required components would then be as in the table.

| \(\theta\) | \(\phi\) | \(p(y = E | \theta, \phi)\) | \(p(\theta)p(\phi)\) |
| --- | --- | --- | --- |
| E | 0 | 1 | 3/16 |
| E | 1 | 0 | 1/16 |
| N | 0 | 0 | 1/16 |
| N | 1 | 1/3 | 1/16 |
| W | 0 | 0 | 1/16 |
| W | 1 | 1/3 | 1/16 |
| S | 0 | 0 | 1/16 |
| S | 1 | 1/3 | 1/16 |

It readily follows that

\( p(\theta = E | y = E) = \frac{3}{4} \) and \( p(\theta \neq E | y = E) = \frac{1}{4} \)

The example is interesting in that it illustrates the importance of correctly formulating a statistical problem. We do not recommend improper priors in our book.

REFERENCES


Comment

D. A. S. FRASER*

Stone continues in his diligent search for flaws in the Bayesian theory of statistical inference. In the present paper he considers two examples in which Bayesian strong inconsistency can occur with a flat prior.

* D.A.S. Fraser is professor, Department of Mathematics, University of Toronto, Toronto 5, Canada.

Only a relatively few statisticians believe that a single theory of inference can be the answer for all of statistics. The committed Bayesians, however, are prominent among such believers. There are, of course, substantial arguments against the Bayesian theory as a single theory of inference; a summary of these arguments may be
found in Fraser [5, 9]. These arguments are not primarily concerned with the Bayesian method as a tool in the statistician’s tool bag; rather, they are concerned with the catholic claim for Bayesian theory and with the meaning and consequences of the theory in scientific contexts. The two examples considered by Stone are somewhat remote from Bayesian theory in a scientific context; indeed, the theory will not rise or fall on the basis of the examples. Nevertheless the examples are extremely interesting; they are concerned with implications of the theory, and they are presented with the attractive flair that we expect from Stone.

Each of the examples has a rather obvious transformation group. Accordingly each has an analogous structural model. A structural model has (i) a distribution on an initial error or variation space, and (ii) a class of transformations from the initial space to an isomorphic response space; the class of transformations has usually been taken to be a group. Dempster [3] and Beran [1] have examined a more general model without the restrictions to the group property. The necessity (for simple probability calculations) of the group property was presented in [4] and some hazards of neglecting the group property were described [6] in the commentary on the McGilchrist [10] paper.

The first example is somewhat artificial from an applied statistical viewpoint but it is extremely interesting from a transformation group viewpoint. The example involves the free group on two generators, and a product of two group elements can contain a partial memory of the two components. Consideration of this characteristic raises questions whether the restrictions to the group property should be strengthened to include, say, amenability. The example deserves further attention.

Stone has overlooked, it seems, the most important aspect of the first example. The example is a very clear and beautiful example against the likelihood principle; the Bayesian, strong inconsistency, is a rather pale property by comparison. Clearly, the example deserves to be remembered as a counter example to the likelihood principle. Of course, not many statisticians believe in the likelihood principle, but a direct and powerful example against it has long been needed.

The simple minded structural version of the first example has the form $Y = \theta Z$, where $\theta = \alpha_1 \cdots \alpha_n$ belongs to the free group $G$ on two generators and $Z$ has probability $\frac{1}{2}$ at each of $a, a^{-1}, b, b^{-1}$, where $a$ and $b$ are the two generators. This differs from the usual examples of structural models in several ways: (i) the extreme high dimensional characteristics of the free group $G$, (ii) the almost trivial dimensionality of the distribution for $Z$, (iii) the property that a product of group elements contains a certain trace of what elements produced the product. The example is extremely interesting and it argues that the conditions determined in [4] should be examined in detail.

A structural model as just described is a distortion of the example in its applied context—the model has a transformation operating on error rather than error applied to the transformation (as in the example). A more realistic model can be presented as $Y = \theta Z$, where $Z$ has probability $\frac{1}{2}$ at each of $h_1(\alpha_n), h_2(\alpha_n), h_3^{-1}(\alpha_n), h_4^{-1}(\alpha_n)$ and

$$h_1(a') = a^{\text{sym}(r)} \quad h_2(a') = b^{\text{sym}(r)} \quad h_3(b') = b^{\text{sym}(r)} \quad h_2(b') = a^{\text{sym}(r)}$$

In the applied context the randomizations applied to the woman by the soldier have only one reference direction—the street that the soldier and woman have just exited from; the present model incorporates this reference. This is not a structural model as generally used, but it does allow probability statements. If $Z \subseteq \{h_1(\alpha_n), h_2(\alpha_n), h_3^{-1}(\alpha_n)\}$, then the actual value of $Z$ can be identified from the observed $Y$. Thus, designating $Z$ as the last element in $Y$ is correct with probability $\frac{1}{2}$. And accordingly, the transformation type model has the needed optimum properties.

The second example considered by Stone is more sharply focused on the Bayesian theory of inference. Some structural aspects of this example were raised in [2], but they had been examined earlier in [7].

In this second example Stone suggests that “some flat connected with group decomposition” may be needed from Fraser [8]. This reference, however, is concerned with inadequacies of the ancillary concept in classical statistics and does not bear on group factorization and the structural model.

REFERENCES


