Pramāṇa - J. Phys., Vol. 32, No. 1, January 1989, pp. 95–97. © Printed in India.

On the consistent inference of probability distributions

Y TIKOCHINSKY
Racah Institute of Physics. The Hebrew University of Jerusalem, Jerusalem 91904, Israel

MS received 29 August 1988

Abstract. We respond to criticism (Karbelkar 1986) concerning our approach to consistent inference of probabilities for reproducible experiments (Tikochinsky *et al* 1984).

Keywords. Inductive inference; maximum entropy method.

PACS Nos 02·50; 03·65; 05·20

In a recent article published in this journal, Karbelkar (1986) criticized our approach (Tikochinsky *et al* 1984a) to a consistent determination of an algorithm for inferring probability distributions. He also claimed to have shown that the result, namely, the maximum entropy algorithm (Jaynes 1957a, b), is not the only one to satisfy our consistency requirement. The purpose of this comment is to restate and clarify our aim and method, apparently, missed by Karbelkar.

We asked ourselves a simple question. Suppose we want to write a computer program for inferring a probability distribution $\{p_i\}$ given the following pieces of input: (1) The number of states s, (2) the degeneracies g_i, $i=1,\ldots,s$, (3) the number of constraints $(m+1)$, where $m+1<s$, (4) the matrix of "dynamical variables" A_{ri}, $r=0,\ldots,m$, $i=1,\ldots,s$, $A_{0i}=1$, and (5) the $(m+1)$ values $\langle A_r\rangle$ of the constraints $\langle A_r\rangle=\sum_{i=1}^{s}p_iA_{ri}$. For example, we could employ as our algorithm a maximum principle inferring $\{p_i\}$ by maximizing an additive functional $S[p]$ of the form

$$S[p]=\sum_{i=1}^{s}g_ih(p_i/g_i) \tag{1}$$

(with $h(x)$ a convex function) subject to the constraints.

Now run the program for the following two problems:

(a) $s=n, g_i, m+1, A_{ri}, \langle A_r\rangle$ to infer $\{p_i\}$,

(b) $s=l, g_{\mathrm{N}}, m+1, B_{r\mathrm{N}}, \langle B_r\rangle$ to infer $\{P_{\mathrm{N}}\}$.

Here N is an index $\mathrm{N}=1,\ldots,l$. In the special case where

$$\mathrm{N}=(N_1,N_2,\ldots,N_n),\quad \sum_{i=1}^{n}N_i=N\ (n \text{ an integer}),$$

$$g_{\mathrm{N}}=N!/(N_1!N_2!\ldots N_n!),$$

$$B_{r\mathrm{N}}=\sum_{i=1}^{n}N_iA_{ri},\quad \langle B_r\rangle=N\langle A_r\rangle,\quad\text{and}\quad l=\binom{N+n-1}{n-1}$$

(with A_{ri} and $\langle A_r \rangle$ the same as in problem (a)), we *could* regard problem (b) as a representation of N independent repetitions of an experiment presented by problem (a). Of course, the computer program grinds the result $\{P_N\}$ regardless of our interpretation. We want, however, the result $\{P_N\}$, in this particular case, to satisfy $P_N = g_N p_1^{N_1} \cdots p_n^{N_n}$, where $\{p_i\}$ are the results of problem (a). Hence our consistency condition. It turns out that this consistency condition, together with the requirement that all data should be treated uniformly in the same way, in fact, suffices to determine the algorithm for the computer program (Tikochinsky *et al* 1984a, b). Moreover, if we decide to choose as our algorithm a maximum principle of the form (1), our consistency condition determines the functional $S[p]$, (up to a monotonic function of S) as the Shannon entropy $S = -\sum p_i \log(p_i/g_i)$ (see Tikochinsky *et al* 1984c).

Karbelkar's Appendix A addresses a completely different problem: What are the additive functionals $S[p]$ of the form (1) such that $\max_p S[p]$ for problem (a) yields the same result $\{p_i\}$ as $\max_{P_N} S[P_N]$, *where P_N is confined to the multinomial form*

$$P_N = g_N p_1^{N_1} \cdots p_n^{N_n}$$

and satisfies

$$\langle B_r \rangle = \sum_{i=1}^{n} \langle N_i \rangle A_{ri} = N \langle A_r \rangle, \quad r = 0, \ldots, m,$$

with $\langle A_r \rangle = \sum_i p_i A_{ri}$. Karbelkar then shows that $S[p] = \sum_i g_i (p_i/g_i)^k$, for some real k, as well as the more familiar $S[p] = -\sum p_i \log(p_i/g_i)$ are such functionals. Here, the assumption that P_N represents probabilities in N independent repetitions of the experiment presented in problem (a), is explicitly used. But this is precisely what we wanted to avoid. We set out to find an algorithm which will treat uniformly all problems of the type (a), regardless of whether they arise from a simple or a super experiment, and yet will give consistent results for the special case of N independent repetitions of a simple experiment.

I am indebted to Dr N Lind for drawing my attention to Karbelkar's work.

References

Jaynes E T 1957a *Phys. Rev.* **106** 620
Jaynes E T 1957b *Phys. Rev.* **108** 171
Karbelkar S N 1986 *Pramana – J. Phys.* **26** 301
Tikochinsky Y, Tishby N Z and Levine R D 1984a *Phys. Rev. Lett.* **52** 1357
Tikochinsky Y, Tishby N Z and Levine R D 1984b *Phys. Rev.* **A30** 2638
Tikochinsky Y, Tishby N Z and Levine R D 1984c *Phys. Rev. Lett.* **55** 337

S N Karbelkar responds

To begin with, we did not claim that the maximum entropy algorithm is not the only one to satisfy TTL (Tikochinsky *et al* 1984) consistency. What was argued was that the TTL consistency conditions are by no means logically compelling. In particular, it was pointed out that the information in the outcomes of a simple and a superexperiment need not be the same in general. More specifically, two superexperiments may yield the same information though in one case the repetitions (of the simple experiment) are independent while in the other case not so. It is only in the case of independent repetitions that the information in a simple and a superexperiment is the same. When, in reality, there are cases where the information in a simple and a superexperiment (we have correlated repetitions in mind) is not the same it is *incorrect* to demand on the part of an inference algorithm to induce the same probabilities for the experiment and the superexperiment. But this is precisely the TTL consistency requirement. So our main objection was (and continues to be) to the TTL consistency requirement. The very far-reaching claim of TTL to have derived the maximum entropy method from consistency conditions that must be satisfied by *any* algorithm is refuted once we show that their statement of the consistency is not logically binding. When a requirement (desire?) is not logically compelling it becomes a part of a modelling hypothesis: no less subjective (contrary to the TTL claim) than Shannon's entropy or similar considerations.

However, we do not stop at merely pointing out that the TTL consistency requirement is not compelling, but go on to restate the consistency requirement relevant to the problem. We argued that the information in a simple and a superexperiment is equivalent if and only if the repetitions forming the superexperiment are known to be independent. We therefore use this additional information explicitly in the form of constraints. This restatement of consistency condition turns out (Appendix A, Karbelkar 1986) to be much less restrictive than the very stringent *but* subjective TTL consistency.

We stand by our criticism of the TTL work.

I am indebted to Dr Lind for bringing my paper to the notice of Dr Tikochinsky.

References

Karbelkar S N 1986 *Pramana – J. Phys.* **26** 301

Tikochinsky Y, Tishby N Z and Levine R D 1984 *Phys. Rev. Lett.* **52** 1357

S N KARBELKAR

Joint Astronomy Programme
Physics Department
Indian Institute of Science
Bangalore 560012, India