LA-5879-MS

Informal Report

UC-32 and UC-34

Reporting Date: February 1975

Issued: February 1975

A Monte Carlo Method for Problems of Hierarchical Inference

by

C. J. Everett

los alamos
scientific laboratory
of the University of California
LOS ALAMOS, NEW MEXICO 87544

UNITED STATES
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION
CONTRACT W-7405-ENG. 36

In the interest of prompt distribution, this report was not edited by the Technical Information staff.

Printed in the United States of America. Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22151
Price: Printed Copy $4.00 Microfiche $2.25

A MONTE CARLO METHOD FOR PROBLEMS OF HIERARCHICAL INFERENCE

by

C. J. Everett

ABSTRACT

A simple Monte Carlo method is presented for estimating the probability $p(S_1,\ldots,S_K|h)$ of the state vector $(S_1,\ldots,S_K)$ of the K terminal (data) nodes in a general inference tree (with an arbitrary number of nodes, each having a prescribed number of possible states) assuming the "hypothesis" node to be in state h. From this information, the probability $p(h|S_1,\ldots,S_K)$ of hypothesis h, assuming an observed state vector $(S_1,\ldots,S_K)$ at the terminal nodes, is then computed from Bayes' formula. The whole routine is easily coded, and a flow chart is included.

I. INTRODUCTION

If p(h,d) is a joint probability density for the pair of indices (h,d), then

$$p(h) \equiv \sum_d p(h,d) \tag{1}$$

is the corresponding marginal density for h, and

$$p(d|h) \equiv p(h,d)/p(h) \tag{2}$$

is the density for d, assuming h. Supposing that only p(h) and p(d|h) are known, then p(h|d), the probability of h, assuming d, is seen from Eqs. (1) and (2) to be

$$p(h|d) = p(h,d)\Big/\sum_h p(h,d)$$

$$= p(d|h)p(h)\Big/\sum_h p(d|h)p(h) \quad . \tag{3}$$

The last expression, in terms of the given densities, is Bayes' formula for p(h|d).

We are concerned here with its application to a "hierarchical inference tree," for which p(h) is given explicitly, and p(d|h) is expressible in terms of given transition probabilities, albeit in an extremely complicated way. In the following complete treatment of the problem, which is designed for computer coding, the role of Monte Carlo consists in the estimation of the probabilities p(d|h) by sampling methods, thus avoiding the difficult exact computation.

II. THE INFERENCE TREE

The problem and general method are best described by reference to a particular but sufficiently complicated example. Topologically, we are concerned with an "inference tree" such as that in Fig. 1. The N(=18) points n = 1,...,N are called "nodes" and are here numbered "level by level," advancing from left to right on each level ℓ, the levels proceeding downward from $\ell = 1$ to $\ell = L(=4)$, the lowermost. A table is stored giving the number $N(\ell)$ of the last node on each level, as in Table I.

TABLE I

LEVEL TERMINUS

ℓ	$N(\ell)$
1	4
2	11
3	16
4	18

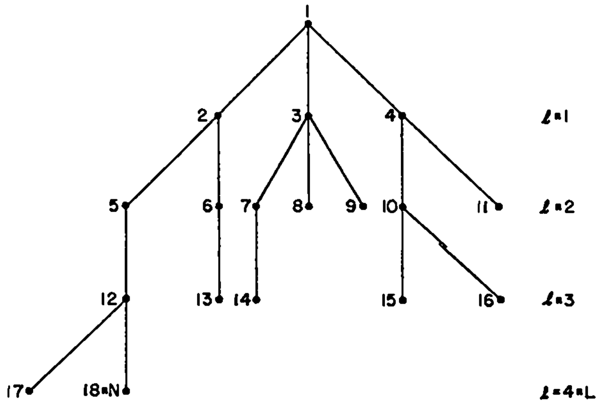

Fig. 1. An Inference Tree

The single topmost node, $n = 1$, is the "hypothesis node" and is the only one with no "ancestor." Each node $n \geq 2$ has exactly one ancestral node $m(n)$, and these are stored in a table such as Table II.

The K nodes without "progeny" (here n = 8, 9, 11, 13, 14, 15, 16, 17, 18) are called "terminal" or "data" nodes and are those subject to direct observation.

TABLE II

ANCESTRAL NODES

n	m(n)
1	0
2	1
3	1
4	1
5	2
6	2
7	3
8	3
9	3
10	4
11	4
12	5
13	6
14	7
15	10
16	10
17	12
18	12

III. NODAL STATES AND TRANSITION PROBABILITIES

Each node $n = 1,\ldots,N$ of the inference tree can exist in just one of a specified number $\sigma(n)$ of "states," $S(n) = 1,\ldots,\sigma(n)$. (No table of $\sigma(n)$ need be stored.)

It is supposed that the hypothesis node $n = 1$ has H (here 3) such states, $h = 1,\ldots,H$, and that the probability $p(h)$ of its existence in any state h is known. This information is stored in a table such as Table III.

It is further assumed that, given the state $i = 1,\ldots,\sigma(m)$ of the ancestral node $m = m(n)$ of any node $n \geq 2$, there is a known transition probability Π_{ij}^{mn} for node n to assume the j-th of its states, $j = 1,\ldots,\sigma(n)$. The corresponding cumulative probabilities are here stored in the form of a $\sigma(m) \times \sigma(n)$ matrix

$$\left(P_{ij}^{mn}\right)$$

(row index i, column index j), one such matrix for each linked pair of nodes $[m(n),n]$, $n = 2,\ldots,N$. (To simplify coding, these matrices may obviously be made square, and of identical size, if storage space permits.) In our example, we store Table IV. For instance, given that node $m = 10$ is in state $i = 3$, there is an 80% chance that node $n = 15$ will assume state $j = 2$.

IV. THE PRIOR DATA PROBABILITIES AND MONTE CARLO

Now suppose that node $n = 1$ is in a specified state h, and fix upon a particular "state vector" $[S(n_1),\ldots,S(n_K)]$ of the K terminal nodes $n = n_1,\ldots,n_K$. This connotes that, for each $\nu = 1,\ldots,K$, the terminal node n_ν is in its state $S(n_\nu)$ $[=1,\ldots,\sigma(n_\nu)]$. For the final application of Bayes' formula we require the probability

$$p[S(n_1),\ldots,S(n_K)|h] \qquad (4)$$

TABLE III

HYPOTHESIS PROBABILITIES

h	p(h)
1	0.3
2	0.2
3	0.5

TABLE IV

TRANSITION PROBABILITIES

1,2 p_{ij}

i \ j	1	2
1	0.3	1
2	0.4	1
3	0.5	1

1,3 p_{ij}

	1	2	3	4	5
1	0.2	0.4	0.6	0.8	1
2	0.1	0.3	0.5	0.7	1
3	0.3	0.4	0.7	0.9	1

1,4 p_{ij}

	1	2	3
1	0.2	0.7	1
2	0.3	0.6	1
3	0.5	0.8	1

2,5 p_{ij}

	1	2	3	4
1	0.5	0.7	0.8	1
2	0.2	0.3	0.7	1

2,6 p_{ij}

	1	2	3	4
1	0.3	0.4	0.9	1
2	0.4	0.5	0.8	1

3,7 p_{ij}

	1	2	3	4	5
1	0.2	0.3	0.4	0.5	1
2	0.4	0.5	0.7	0.9	1
3	0.1	0.4	0.6	0.8	1
4	0.5	0.6	0.7	0.9	1
5	0.6	0.7	0.8	0.9	1

3,8 p_{ij}

	1	2	3	4
1	0.3	0.5	0.8	1
2	0.4	0.6	0.9	1
3	0.2	0.6	0.7	1
4	0.1	0.4	0.8	1
5	0.5	0.7	0.9	1

3,9 p_{ij}

	1	2
1	0.5	1
2	0.4	1
3	0.3	1
4	0.2	1
5	0.1	1

4,10 p_{ij}

	1	2	3
1	0.4	0.8	1
2	0.5	0.7	1
3	0.3	0.5	1

4,11 p_{ij}

	1	2
1	0.3	1
2	0.5	1
3	0.7	1

5,12 p_{ij}

	1	2
1	0.5	1
2	0.2	1
3	0.6	1
4	0.8	1

6,13 p_{ij}

	1	2
1	0.4	1
2	0.3	1
3	0.5	1
4	0.2	1

7,14 p_{ij}

	1	2
1	0.1	1
2	0.3	1
3	0.2	1
4	0.4	1
5	0.6	1

10,15 p_{ij}

	1	2
1	0.5	1
2	0.7	1
3	0.2	1

10,16 p_{ij}

	1	2	3
1	0.2	0.8	1
2	0.4	0.7	1
3	0.3	0.5	1

12,17 p_{ij}

	1	2
1	0.4	1
2	0.6	1

12,18 p_{ij}

	1	2	3	4
1	0.2	0.6	0.8	1
2	0.3	0.7	0.9	1

of the terminal state vector $[S(n_1),\ldots,S(n_K)]$, assuming n = 1 to be in state h. Even for relatively simple trees, an analytical approach, although well defined, its far from easy (Ref. 1). In principle, we should consider separately each possible sequence of transitions $i \to j$ throughout the tree which result in the stipulated state vector, multiplying all the corresponding transition probabilities Π_{ij}^{mn} for all pairs (m,n) of linked nodes, and then summing all such products for all possible transition sequences.

In contrast, the estimation of the probability (4) by sampling is transparently simple, since we may regard a terminal state vector as the result of a kind of shower, or cascade descending from n = 1, and governed by the given transition probabilities. Thus, assuming n = 1 to be in the particular state h, one "throws" successively for the state S(n) of each node n = 2,...,N(1) on level $\ell = 1$, using the stored transition probability tables p_{hj}^{1n}. Having determined all states on level 1, the states of all nodes on level 2 may then be thrown for in similar fashion, and so down through the lowest level L. At this point, the states S(n) of all nodes n = 1,...,N have been assigned, and the results recorded in a temporary storage table as exemplified by Table V.

TABLE V

CASCADE STATES

n	S(n)
1	h
2	S(2)
.	.
.	.
.	.
* 8	S(8)
* 9	S(9)
10	S(10)
* 11	S(11)
12	S(12)
* 13	S(13)
* 14	S(14)
* 15	S(15)
* 16	S(16)
* 17	S(17)
* 18	S(18)

In particular, a state vector $[S(n_1),\ldots,S(n_K)]$ has been determined for the K terminal nodes $(n_1,\ldots,n_K)$ (starred in Table V). There are

$$D \equiv \sigma(n_1)\cdots\sigma(n_K)$$

such vectors in all, and they are easily enumerated by a single index, d = 1,2,...,D, as shown in the Appendix. Having computed the index d of the state vector resulting from the above cascade, a 1 is tallied in a storage location N(h,d). The latter may be visualized in a tabular form, such as Table VI for our example. Repetition of the process for a sufficiently large number C of cascades yields all the possible terminal state vectors d = 1,...,D, each with approximately its correct frequency N(h,d). Norming by C then gives the Monte Carlo approximation to the probability (4), namely

$$p(d|h) \cong N(h,d)/C$$

and the result is stored back in location N(h,d).

The routine is followed as indicated for each of the states h = 1,...,H of n = 1. At this point, Table VI contains an entire listing of (approximate) probabilities (4), namely

$$N(h,d) \sim p(d|h)$$

h = 1,2,...,H, d = 1,2,...,D, and the first part of the problem is complete.

V. THE POSTERIOR PROBABILITY OF HYPOTHESIS h

The ultimate question in hierarchical inference asks for the probability p(h|d) of a certain hypothesis (state h of node n = 1), assuming a particular data vector, of index d, to be observed at the terminal nodes $n_1,\ldots,n_K$. As noted in Sec. I, this is readily computed as

TABLE VI

INITIALLY, FREQUENCIES N(h,d)

h \ d	1	2	···	3072 = D
1				
2				
3				

$$p(h|d) = p(d|h)\ p(h) \Big/ \sum_h p(d|h)\ p(h)$$

which in our notation at this point is approximately

$$N(h,d)\ p(h) \Big/ \sum_h N(h,d)\ p(h) \quad . \qquad (5)$$

The details of machine computation, which are routine, are indicated in the flow chart of Fig. 2. A storage block must be reserved for the D sums

$$\Delta(d) \equiv \sum_h N(h,d)\ p(h)$$

as indicated in Table VII.

If the quotients obtained in (5) are stored back at the locations N(h,d), Table VI will contain the entire set of approximations

$$p(h|d) \sim N(h,d)$$

at the end of the problem. Printing the complete result is obviously impractical in an involved case such as our example, and is of course limited to the information actually required. Various sum checks can easily be included, as indicated in Fig. 2.

Aside from the permanent storage already mentioned, one must include one additional table such as Table VIII.

TABLE VII

BAYESIAN DENOMINATORS

d	1	2	···	3072 = D
Δ(d)				

TABLE VIII

ADDITIONAL CONSTANTS

H	3
C	10^5
L	4
D	3072

VI. A FLOW CHART

In the following chart, which is designed to handle the general problem as just described, only the formula for the terminal state index d (cf. the Appendix) and the print routine are subject to change. The letter r below refers to the next random number of the random number generator. It is understood that transfers to previous entries are executed on Y (yes) or N (no), with the indicated indices advanced by 1, whereas the opposite decision always leads to the next step. The printed sum-check parameters $T_1,\ldots,T_h$ and U should all be unity. Moreover, if any column of the final Table VI is printed, its sum should also be unity.

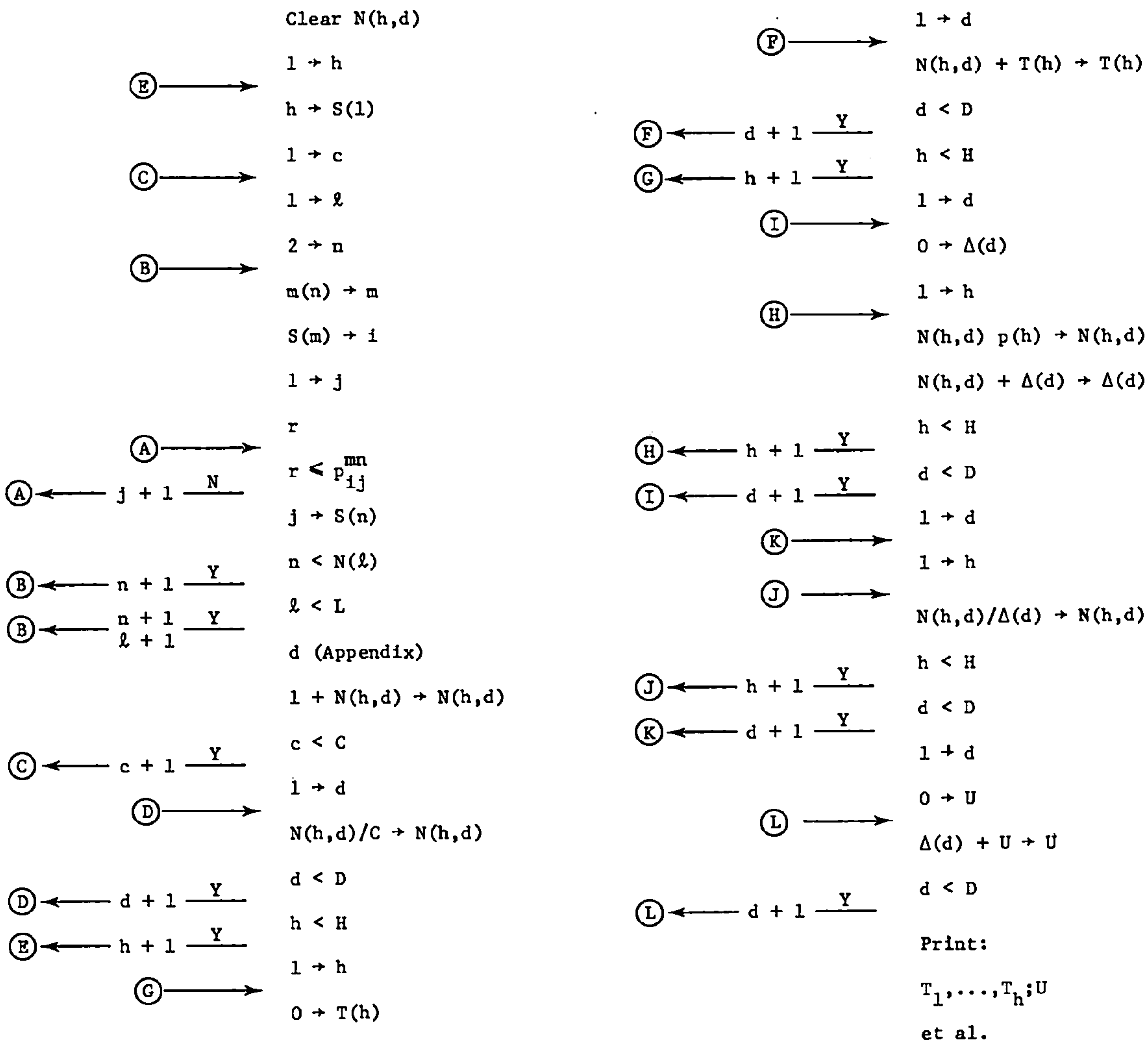

Fig. 2. Flow Chart

APPENDIX

ENUMERATION OF STATE VECTORS

For machine purposes, we require an enumeration of the set of all

$$D = \sigma(n_1) \cdots \sigma(n_K)$$

terminal state vectors $[S(n_1),...,S(n_K)]$ by means of a single index

$$d = 1, 2, \cdots, D \quad .$$

For the moment, we denote $\sigma(i_\nu)$ by σ_ν and $S(i_\nu)$ by S_ν. If the states of terminal node n_ν were numbered

$$S'_\nu = 0, 1, \cdots, \sigma_\nu - 1$$

then the function

$$S'_1 + \sigma_1 S'_2 + \sigma_1\sigma_2 S'_3 + \cdots + (\sigma_1 \cdots \sigma_{K-1}) S'_K$$

would run over all the integers $0,1,\ldots,(\sigma_1\cdots\sigma_K)-1$. Since in our notation S_ν runs over the integers $1,2,\ldots,\sigma_\nu$, it is clear that a suitable formula for the index $d = 1,2,\ldots,\sigma_1\cdots\sigma_K = D$ is

$$d = S_1 + \sigma_1 S_2 + \sigma_1\sigma_2 S_3 + \cdots + (\sigma_1\cdots\sigma_{K-1})S_K$$

$$- [1 + \sigma_1 + \sigma_1\sigma_2 + \cdots + (\sigma_1\cdots\sigma_{K-1})] + 1$$

or, returning to our original notation,

$$d = S(n_1) + \sigma(n_1)S(n_2) + \sigma(n_1)\sigma(n_2)S(n_3)$$

$$+ \cdots + [\sigma(n_1)\cdots\sigma(n_{K-1})]S(n_K) - \Delta$$

where

$$\Delta = \sigma(n_1) + \sigma(n_1)\sigma(n_2) + \cdots + [\sigma(n_1)\cdots\sigma(n_{K-1})] \quad .$$

Thus, in our example, the terminal nodes n_ν with their numbers $\sigma(n_\nu)$ of states are (cf. Table IV)

	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
	8	9	11	13	14	15	16	17	18
$\sigma(n_\nu)$:	4	2	2	2	2	2	3	2	4

In this case our formula for d is

$$d = S(8) + 4S(9) + 8S(11) + 16S(13) + 32S(14)$$

$$+ 64S(15) + 128S_{16} + 384S_{17} + 768S_{18} - 1404$$

where d runs over the integers $1,2,\ldots,3072 = D$. The flow chart of Fig. 2 is quite general except for the step at which d is computed. This step is tailored to the particular inference tree and specially coded.

REFERENCE

1. C. W. Kelly III, S. Barclay, "A general Bayesian Model for Hierarchical Inference," Organizational Behavior and Human Performance, Dec. 1973.