Probabilistic Information Processing Systems: Evaluation with Conditionally Dependent Data[1]

PATRICIA ANN DOMAS AND CAMERON R. PETERSON[2]

University of Michigan

Previous research on Probabilistic Information Processing (PIP) systems has used data that are conditionally independent. In the real world, data are frequently conditionally dependent, that is, given an hypothesis, the occurrence of one datum influences the likelihood of occurrence of a second datum. If the use of a PIP system is desired when the data are known to be conditionally dependent, then it is necessary to know if PIP is an appropriate system for the processing of dependent data. One experiment compared PIP with a second system, POP. PIP operators gave more optimal estimates when the data were conditionally independent; however, POP estimators gave more optimal estimates when the data were conditionally dependent. A second experiment attempted, unsuccessfully, to produce a modified PIP system that would give optimal estimates for both conditionally dependent and conditionally independent data. A revised technology of diagnostic information processing based on conditionally dependent data was proposed.

Observation of events provides man with evidence about prevailing and future states of the environment. In diagnosis, he revises his certainty about hypothesized states on the basis of evidence. Bayes's theorem is a formally optimal rule for revision of one's relative certainty about hypotheses in light of new evidence. When intuitive inferences are compared with optimal inferences from Bayes's theorem, it turns out that people extract less certainty from the data than does Bayes's theorem; that is, people are conservative (Phillips & Edwards, 1966).

Edwards (1966) has designed an information processing system that circumvents such conservatism. In this Probabilistic Information Processing (PIP) system, the task of diagnostic information processing is

77

decomposed into two subtasks. In the first task, man assesses the diagnostic impact of a current datum D_n, on a pair of hypotheses H_a and H_b, as expressed in the likelihood ratio, $L_n = P(D_n|H_a)/P(D_n|H_b)$. The second task consists of the aggregation of these separate likelihood ratios across all data. Likelihood ratios and prior odds may be combined by the odds-likelihood ratio form of Bayes's theorem, $\Omega_n = L_n\Omega_{n-1}$. The Ω_n is the posterior odds in favor of H_a in light of the data up to and including the current datum D_n. The Ω_{n-1} is the odds after receiving the previous $n - 1$ data.

In a large simulation-type experiment, Edwards, Phillips, Hays, and Goodman (1968) compared a PIP system, in which subjects estimated likelihood ratios, to a POP system in which subjects estimated Ω_n, the posterior odds in favor of each hypothesis. When estimating posterior odds, POP subjects essentially performed both the assessment and aggregation tasks. Information that led posterior odds estimators to arrive at odds of $4\frac{1}{2}:1$ led PIP operators to more extreme estimates of $99:1$. Wheeler and Edwards (personal communication) performed an experiment using a normal data generating process, and again POP was conservative with respect to PIP. When both POP and PIP were compared with the veridical model, POP was conservative, and PIP was nearly optimal. Thus, these results argue in favor of using PIP systems.

Previous research comparing PIP and POP procedures, however, has used data that are conditionally independent of each other. Formally, independence obtaines between two items of data, D_n and D_{n-1}, in light of some hypothesis, H_i, when $P(D_n,D_{n-1}|H_i) = P(D_n|H_i)P(D_{n-1}|H_i)$. This relation is violated whenever $P(D_n|H_i)$ changes as a function of D_{n-1}, the datum that occurs on observation $n - 1$; that is, whenever D_n and D_{n-1} are conditionally dependent.

Data dependency is encountered in various forms. Schum (1966) presented subjects with scenarios of data classes whose joint conditional probabilities exhibited conditional dependence; that is, the dependency displayed magnified the likelihood ratios. Conditional dependence, however, may have the effect of either enhancing the likelihood ratio or reducing it. The present study considers a different form of dependence, redundancy, which occurs when the nth datum, D_n, may repeat information from D_{n-1}; that is, the likelihood ratios are reduced. When D_n is the same as D_{n-1}, they are completely dependent or redundant, and no new information is obtained from D_n. As an example of data redundancy, consider weather forecasting. When a forecaster attempts to predict the probability that it will rain tomorrow, he has several sources of data. Across all the weather charts available to him, such measures as temperature and pressure are repeated many times. A report of high

pressure on a current chart should mean less to the forecaster if he has previously incorporated that information into his probability estimate.

If the use of a PIP system is desired in a situation known to involve conditionally dependent data, then it is necessary to know how a PIP system handles these interdependencies. A PIP system probably improves performance because it decomposes the diagnostic task. When data are independent a PIP operator need consider only one datum at a time, and estimate L_n for each datum in isolation from all other data. The degree of decomposition in PIP, however, is not possible if the data are conditionally dependent. The operator cannot consider D_n independently of D_{n-1}, since L_n changes as a function of which datum occurred on observation $n - 1$.

The present study addressed the question of conditionally dependent data: Which response mode, PIP or POP, elicits the more nearly optimal inferences when the data exhibit a redundant form of conditional dependence.

<center>EXPERIMENT 1: PIP VS POP</center>

Method

Design. The experiment required a diagnostic task involving dependent data. Four different degrees of data dependency were used in which the probabilities of a redundant datum were 0, $\frac{1}{6}$, $\frac{3}{6}$, and $\frac{5}{6}$. A nonrepeated datum (i.e., $D_n \neq D_{n-1}$) will be denoted by $\bar{R}$; R denotes a repeated datum (i.e., $D_n = D_{n-1}$). The likelihood ratio for, $\bar{R}$ at any sequential location, is given by

$$L_n = \frac{P(\bar{R}|H_a)}{P(\bar{R}|H_b)} = \frac{P(D_n|H_a)}{P(D_n|H_b)}. \tag{1}$$

The likelihood ratio for R is given by

$$L_n = \frac{P(R|H_a)}{P(R|H_b)} = \frac{P(D_n|H_a) + k}{P(D_n|H_b) + k}, \tag{2}$$

where $k = p/(1 - p)$, and p is the probability of repetition.

Two different response modes, PIP and POP, were combined factorially with the four levels of data redundancy. In the PIP mode, subjects estimated a conditional likelihood ratio, $P(D_n|H_a, D_{n-1}, D_{n-2}, \ldots, D_1)/P(D_n|H_b, D_{n-1}, D_{n-2}, \ldots, D_1)$, for each datum, D_i, $i = 1, \ldots, n$. In the POP mode subjects estimated posterior odds for the entire sequence of data; these quantities are of the form, $P(H_a|D_1, D_2, \ldots, D_n)/P(H_b|D_1, D_2, \ldots, D_n) = \Omega_n$.

Subjects. Twenty-two University of Michigan men undergraduates served individually as subjects. They were paid at the rate of $1.50/hr.

Apparatus. The random devices generating data samples were two bookbags, one containing 80% red chips and 20% blue chips, the other having the reverse proportions. The numerical proportions of each bag were displayed to the subject. Five dice were used to determine the probabilities of repeating a datum. These dice had, respectively, 0, 1, 3, 5, and 6 black sides, representing probabilities of repetition, $p = 0$, $\frac{1}{6}$, $\frac{3}{6}$, $\frac{5}{6}$, and 1. The remaining sides on the dice were white. The number of black sides on each die was listed for the subject, while the numerical values of p were not displayed. A third display cumulatively recorded the sequence of chip colors for the subject.

Procedure. In the first part of the experiment, the experimenter trained a subject on one of the two response modes. Both bookbags and their contents were shown to the subject and their proportions displayed. The experimenter selected one of the bags by flipping a concealed coin and then sampled chips sequentially and with replacement. After each chip was drawn, the subject responded verbally with either a likelihood ratio or an odds estimate appropriate to the assigned response mode. The experimenter explained that after the draw of each chip, one die would be rolled. The subject could observe the roll but not the outcome. If the die outcome was a white side, the experimenter would replace the chip and resample from the previously chosen bookbag; but if the outcome were black, the experimenter would redisplay the same chip shown on the previous draw. Outward appearances were such that the subject could not distinguish between the processes of resampling and redisplaying. Initially the subject practiced with the all white and all black die, so that he knew for certain whether the experimenter was resampling a new chip or redisplaying the same chip. Before beginning the experimental trials, the experimenter ran through a practice trial drawing three chips and rolling the die with $p = \frac{1}{2}$ twice, once between the first and second, and once between the second and third draws. The subject was instructed to use the probabilistic information concerning data redundancy, provided by the die, in making his estimates for sequences of data.

In the experimental trials, the experimenter selected a bookbag before each new trial. Each subject was instructed that there would be eight trials with three draws per trial. The subject was also told on each trial which of four dice having $p = 0$, $\frac{1}{6}$, $\frac{3}{6}$, or $\frac{5}{6}$, was being used, and that the bookbag and die would always be the same during a given trial. The order of presentation of the dice was predetermined but the sequences of chip draws were not controlled. All subjects saw the four dice in the same order of presentation.

The subject marked his estimates on a log scale from 1:1 to 1,000,000:1. The subject selected the hypothesis he considered more likely, and indicated his likelihood ratio or odds in favor of that hypothesis. Estimates of the form X:1 were required. He recorded all responses in an individual response booklet, with one response per page.

RESULTS

Both conditionally dependent and independent data could occur during a given experimental trial. Estimates based on each type of data were collected within a level of dependency and analyzed separately. Results are based on the data for 22 subjects, 11 in each of the two response-mode conditions. In the PIP mode, conditional likelihood ratios were equal to the subjects' actual estimates. In the POP condition, conditional likelihood ratios were inferred from subjects' posterior odds estimates using the odds-likelihood ratio form of Bayes's theorem.

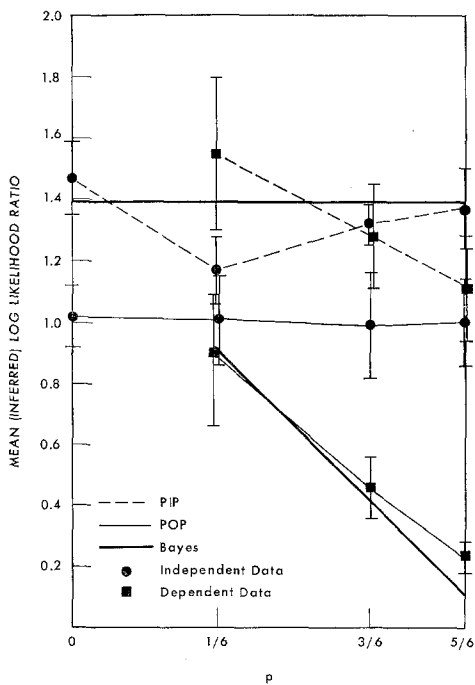Figure 1 shows subjects' mean inferred log likelihood ratios (ILLRs)

Fɪɢ. 1. PIP vs POP: Mean ILLRs as a function of p. Vertical bars indicate SEM.

as a function of p. The heavier lines are optimal estimates calculated from Eqs. (1) and (2). The circles and squares represent estimates based on independent or dependent data, respectively. In the PIP condition, mean ILLRs are connected by a broken line. An unbroken line connects mean ILLRs in the POP condition.

Figure 1 provided answers to two questions. First, the results of previous research were confirmed: PIP was more optimal than POP when the data were conditionally independent. Second, when the data were conditionally dependent, PIP estimates were again extreme with respect to POP, but suboptimally so. PIP operators gave conditional likelihood ratio estimates based on conditionally dependent data that were not significantly different from estimates based on conditionally independent data. On the other hand, as the probability of repetition increased, POP estimators appropriately decreased the diagnostic impact of the dependent data. At all levels of dependency, POP was more optimal than PIP.

EXPERIMENT 2: CONDITIONAL ODDS

The diagnostic impact of new information in a datum is expressed in its conditional likelihood ratio. Experiment 1 indicated that subjects had difficulty understanding and estimating conditional likelihood ratios. Thus, at the suggestion of Kurt Snapper (personal communication), a modified system was designed to make the meaning of the conditional likelihood ratio more apparent. It seemed plausible that the use of odds may have made subjects more sensitive to conditional dependence. Therefore, the modified system required both that subjects respond in odds and that the machine be allowed to aggregate. This was accomplished with the response mode of conditional odds; conditional in the sense that subjects were instructed to assume that all previous data had occurred and that all previous data had led to odds of 1:1, that is $\Omega_{n-1} = 1:1$. Thus, conditional odds were formally identical to corresponding conditional likelihood ratios.

Method

Design. Four different degrees of data redundancy, $p = 0$, $\frac{1}{6}$, $\frac{3}{6}$, and $\frac{5}{6}$, were combined with the conditional odds response mode.

Subjects. Eleven University of Michigan men undergraduates served individually as subjects. They were paid at the rate of \$1.50/hr.

Apparatus and procedure. The apparatus and procedure were the same as in Expt. 1. In responding, however, subjects estimated conditional odds.
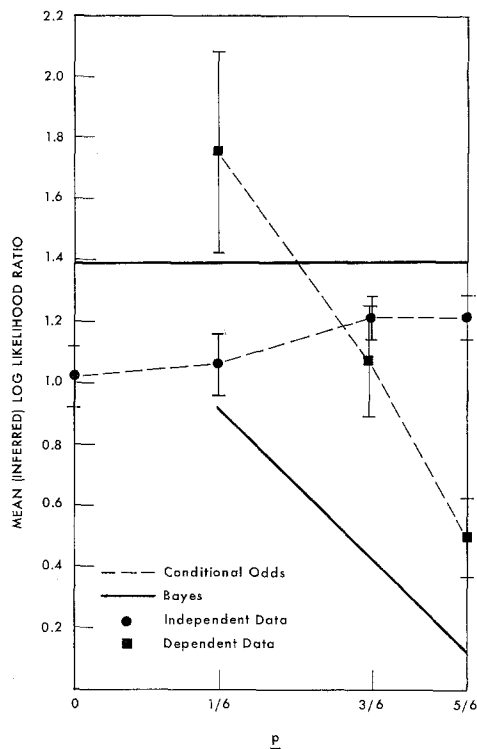
FIG. 2. Conditional odds: Mean ILLRs as a function of p. Vertical bars indicate SEM.

RESULTS

All data for the 11 subjects were included in the analysis. The subjects' actual estimates of conditional odds were inferred as conditional likelihood ratios. Figure 2 displays mean ILLRs as a function of p. Optimal estimates calculated from Eqs. (1) and (2) are connected by the heavier lines. A broken line connects mean ILLRs in the conditional odds mode. Estimates based on independent data are represented by circles, dependent data estimates are represented by squares.

Using the modified response mode, subjects were able to discriminate between independent and dependent data. Conditional odds estimators gave estimates based on dependent data that were significantly different $(p < .05)$ from estimates based on independent data. The subjects decreased the diagnostic impact of the dependent datum as p increased. However, their increased sensitivity to the degree of dependency occurred at the cost of suboptimality. Estimates based on conditionally

independent data were conservative, while estimates based on conditionally dependent data were excessive.

DISCUSSION

Experiment 1 indicated that the decomposition concept of PIP resulted in insensitivity to the degree of dependency between data. In Expt. 2, modification of the response mode produced an output that was sensitive to the degree of dependency, but at a cost in terms of optimality, that is, conservatism or radicalism.

Since neither PIP nor conditional odds response modes handled data interdependencies satisfactorily, the results of Expts. 1 and 2 discourage the complete decomposition of the information processing task. The decomposition concept of PIP and conditional odds consists of dividing diagnostic information into two subtasks: human assessment and human aggregation. Zlotnick (1968) has proposed an intermediate form of information processing involving both human aggregation and the concept of decomposition: The task is partially decomposed into independent "packages" of dependent data; man aggregates evidence within a "package" of conditionally dependent data. Machines then aggregate evidence across "packages" that are independent of each other. The present results, as well as Schum's (1966), suggest that Zlotnick's proposal will be an attractive compromise. That is, the information processing system should be sensitive to dependencies within data packages, and it should reduce conservatism due to aggregation across packages.

Gustafson (1969) has provided further empirical support for this proposal. The comparison of a PIP system to three other forecasting techniques showed that the detrimental effects of conditional dependencies can be removed by subjects classification of data into conditionally independent complexes.

REFERENCES

EDWARDS, W. Non-conservative probabilistic information processing systems. ESD-TDR-66-404, University of Michigan, Institute of Science and Technology, Report 5893-22-F, December 1966.

EDWARDS, W., PHILLIPS, L. D., HAYS, W., & GOODMAN, B. C. Probabilistic information processing systems: Design and evaluation. *IEEE Transactions on Systems Science and Cybernetics,* 1968, Vol. SSC-4, 248–265.

GUSTAFSON, D. H. Evaluation of probabilistic information processing in medical decision making. *Organizational Behavior and Human Performance,* 1969, **4,** 20–34.

PHILLIPS, L. D., & EDWARDS, W. Conservatism in a simple probability inference task. *Journal of Experimental Psychology,* 1966, **72,** 346–354.

SCHUM, D. A. Inferences on the basis of conditionally nonindependent data. *Journal of Experimental Psychology,* 1966, **72,** 401–409.

SNAPPER, K. J. Personal communication.
WHEELER, G., & EDWARDS, W. Personal communication.
ZLOTNICK, J. A theorem for prediction. *Foreign Service Journal,* 1968, **45**, 20.