Reprinted by permission from IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS
Vol. SMC-4, No. 1, January 1974, pp. 116-117

The Best Two Independent Measurements Are Not the Two Best

THOMAS M. COVER

***Abstract*—Consider an item that belongs to one of two classes, $\theta = 0$ or $\theta = 1$, with equal probability. Suppose also that there are two measurement experiments E_1 and E_2 that can be performed, and suppose that the outcomes are independent (given θ). Let E_i' denote an independent performance of experiment E_i. Let $P_e(E)$ denote the probability of error resulting from the performance of experiment E. Elashoff [1] gives an example of three experiments E_1,E_2,E_3 such that $P_e(E_1) < P_e(E_2) < P_e(E_3)$, but $P_e(E_1,E_3) < P_e(E_1,E_2)$. Toussaint [2] exhibits binary valued experiments satisfying $P_e(E_1) < P_e(E_2) < P_e(E_3)$, such that $P_e(E_2,E_3) < P_e(E_1,E_3) < P_e(E_1,E_2)$. We shall give an example of binary valued experiments E_1 and E_2 such that $P_e(E_1) < P_e(E_2)$, but $P_e(E_2,E_2') < P_e(E_1,E_2) < P_e(E_1,E_1')$. Thus if one observation is allowed, E_1 is the best experiment. If two observations are allowed, then two independent copies of the "worst" experiment E_2 are preferred. This is true despite the conditional independence of the observations.**

I. Introduction

In the classification or hypothesis testing problem, it is well known that the most informative k element subset of n measurements is not necessarily the union of the k individually most informative measurements. An easy example can be generated by allowing statistical dependence among the measurements, thereby making it redundant to use the two measurements with individually lowest probabilities of error.

If one goes no further than this example, one might be led to believe that the difficulties in finding the best k measurements arise solely from dependence among the measurements. Elashoff [1] and Toussaint [2] have shown that the essence of the problem is retained even when all the measurements are (conditionally) independent. This correspondence provides another such example.

II. Family of Examples

Consider the following example:

$$\begin{array}{lll} & E_1 & E_2 \\ \theta = 0 & X = \begin{cases} 1, & p_0 \\ 0, & 1 - p_0 \end{cases} & Y = \begin{cases} 1, & r_0 \\ 0, & 1 - r_0 \end{cases} \\ \theta = 1 & X = \begin{cases} 1, & p_1 \\ 0, & 1 - p_1 \end{cases} & Y = \begin{cases} 1, & r_1 \\ 0, & 1 - r_1 \end{cases} \end{array}$$

where $\Pr\{\theta = 0\} = \Pr\{\theta = 1\} = \frac{1}{2}$. Also, let (E_i,E_i') denote two independent repetitions of E_i. Thus, for example, for (E_1,E_1'), $\Pr\{(X,X') = (1,0) \mid \theta\} = p_\theta(1 - p_\theta)$.

The Bayes probability of error is given for a discrete random variable X by

$$P_e(E) = \sum_x \min\{\Pr\{\theta = 0\}P_0(x), \Pr\{\theta = 1\}P_1(x)\}.$$

Thus, for example,

$$\begin{aligned} P_e(E_1) &= \tfrac{1}{2}\min\{1 - p_0, 1 - p_1\} + \tfrac{1}{2}\min\{p_0,p_1\} \\ &= \tfrac{1}{2}[1 - |p_0 - p_1|]. \end{aligned}$$

Choose

$$p_0 = 0.96, p_1 = 0.04, r_0 = 0.9, r_1 = 0.$$

We then have

$$\begin{aligned} P_e(E_1) &= 0.04 \\ &< P_e(E_2) = 0.05 \end{aligned}$$

and

$$\begin{aligned} P_e(E_2,E_2') &= 0.005 \\ &< P_e(E_1,E_2) = 0.022 \\ &< P_e(E_1,E_1') = 0.040. \end{aligned}$$

Interpreting this, we see, for example, that two observations of the best experiment E_1 yield no decrease in the probability of error 0.04, while two observations of E_2 decrease the probability of error 0.05 by a factor of 10.

III. Comments

The main idea of this example is that experiment E_2 has a moderate probability $\frac{1}{2}(r_0 + r_1)$ of yielding an observation ($Y = 1$) of very high information (conditional probability of error approximately zero). Thus two independent observations of E_2 will yield such an observation with much higher probability. Experiment E_1, on the other hand, has no events of outstandingly high information. In fact, since we have chosen $p_0 = 1 - p_1$ in E_1, two experiments are no better than one.

The set of values, r_0,r_1,p_0,p_1 for which the desired ordering on experiments is induced is fairly large. In particular it is not necessary to choose $r_1 = 0$, nor is it necessary that the prior probabilities be equal.

In general, whenever the divergence measure J yields a different ordering on E_1 and E_2 than does the probability of error P_e, then n independent repetitions of the high P_e experiment will yield lower P_e than n independent repetitions of the lower P_e experiment, for a suitably large value of n.

Acknowledgment

B. Boyle, Massachusetts Institute of Technology, E. Persoon, Purdue University, and Chandrasekaran, Ohio State, contributed references and discussions.

References

[1] J. D. Elashoff, R. M. Elashoff, and G. E. Goldman, "On the choice of variables in classification problems with dichotomous variables," *Biometrika*, vol. 54, pp. 668–670, 1967.
[2] G. T. Toussaint, "Note on optimal selection of independent binary-valued features for pattern recognition," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-17, p. 618, Sept. 1971.

Manuscript received June 5, 1973; revised July 30, 1973. This work was supported by the Air Force Office of Scientific Research under Contract F44620-69-C-0101.
The author is with the Departments of Electrical Engineering and Statistics, Stanford University, Stanford, Calif. 94305.