NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY
51 HOVEY ROAD, PENSACOLA, FL 32508-1046

NAMRL-1406

SOME GENERAL QUANTITATIVE CONSIDERATIONS FOR THE STATISTICAL ANALYSIS OF ISOPERFORMANCE CURVES

D. J. Blower

ABSTRACT

In this paper, we present a recommended quantitative approach for analyzing the concept of isoperformance. The ideas outlined here rely upon the Bayesian version of *model evaluation.* We define models as hypotheses about the probabilities of subjects being categorized by a combination of predictor variables and criterion variables. From this foundation, a computational formula is derived whose value can be compared to a χ^2 distribution. For example, we are often interested in calculating the probability of a subject failing during some phase of flight training given that we have information on certain predictor variables. We would like to ascertain whether the extra information contained in such predictor variables is useful. If it is useful, then it enables us to predict the probability of failure for any given student. This ability to predict a change in the probability of failure, either in the upwards or downwards direction, is very helpful to managers and decision makers in the training community. In addition, these techniques can help answer the question of whether a candidate for flight training can "trade-off" a high score on one predictor variable for a low score on a different predictor variable. In particular, we would like to investigate the possibility of trading off different classes of predictor variables, say cognitive information processing variables and personality variables, and still achieve the same level of performance. The maximum entropy principle is used as a systematic disciplined approach to find parsimonious models.

Acknowledgments

I would like to acknowledge the continuing support of LT Hank Williams, Head of the Aviation Selection Division at NAMRL. My primary intellectual indebtedness for the ideas presented here can be traced to the monumental work by Edwin Thompson Jaynes on Bayesian statistics and maximum entropy. Professor Jaynes, Emeritus Professor of Physics at Washington University in St. Louis, Missouri, died on 30 April 1998 at the age of 75.

INTRODUCTION

In this paper, we present a recommended quantitative approach for analyzing the concept of isoperformance. An article in the journal *Human Factors* by Jones and Kennedy [1] prompted our current interest in applying isoperformance to selection and training issues.

The ideas outlined here rely upon the Bayesian version of *model evaluation.* We define models as hypotheses about the probabilities of subjects being categorized by a combination of predictor variables and criterion variables. From this foundation, a computational formula is derived whose value can be compared to a χ^2 distribution.

If this easily computed value falls into the upper 5% region of a χ^2 distribution with the appropriate degrees of freedom, then we reject the tentative model. On the other hand, if the value falls into the lower 95% region of the distribution, then we accept the model. Once a model is found that can be accepted, a few elementary rules from probability theory can be used to calculate the probability of events involved in isoperformance curves.

For example, we are often interested in calculating the probability of a subject failing during some phase of flight training given that we have information on certain predictor variables. Alternatively, one can focus on the positive and say that we are interested in the probability of a subject passing flight training. We would like to ascertain whether the extra information contained in such predictor variables is useful. If it is useful, then it enables us to predict the probability of failure (or passing) for any given student. This ability to predict a change in the probability of failure, either in the upwards or downwards direction, is very helpful to managers and decision makers in the training community.

In addition, these techniques can help answer the question of whether a candidate for flight training can "trade-off" a high score on one predictor variable for a low score on a different predictor variable. In particular, we would like to investigate the possibility of trading off different classes of predictor variables, say cognitive information processing variables and personality variables, and still achieve the same level of performance.

THE DATA BASE

The purpose of this paper is to provide the general quantitative foundations for analyzing isoperformance. From time to time, we shall employ fictitious data to illustrate the formulas. The analysis of actual data using these techniques will be presented in a subsequent report [2]. The fictitious data does, nonetheless, give some general idea of the actual data base we will be analyzing in the future for the isoperformance project.

As part of another project called the Pilot Prediction System (PPS), we have constructed a rather large and comprehensive data base consisting of various selection and training variables. A subset of this data base contains information on over a thousand Navy and Marine Corps candidates who entered pilot flight training from 1993 to early 1998.

Scores on the various subtests of the Aviation Selection Test Battery (ASTB) and all the grades from the academic ground school (API – Aviation Preflight Indoctrination) portion of training prior to actual flight training are part of this data base. We will concentrate on one of the subtests from the ASTB, the Pilot Biographical Inventory (PBI), and the final overall grade from API called the Navy Standard Score (NSS).

The raw score on the PBI is transformed into one of nine discrete categories so that PBI = 1, 2 $\cdots$ 9 with 1 being the lowest score and 9, the highest score. There are no candidates in the data base with a PBI = 1 so PBI will consist of eight categories. The API NSS is tranformed into one of six discrete categories, API = 1, 2 $\cdots$ 6 with, again, 1 representing low scores and 6 representing high scores from ground school. Thus, PBI and API represent the two predictor variables.

One criterion variable will be used in the subsequent analysis. This criterion variable simply records whether a candidate failed some later phase of flight training after API. The crux of the analysis then centers naturally upon the probability of failure given information about two predictor variables.

CONTINGENCY TABLES

As just mentioned, we will eventually analyze data from the PPS data base in our first assessment of isoperformance curves. The following schema will be used to set up the statistical derivations as detailed in later sections. Consider n cells that represent the n different ways that an event could happen. For us, these n cells represent the various combinations of categories for a given number of predictor variables and criterion variables.

For example, a subject in the data base is classified into one of eight categories on the PBI predictor variable, one of six categories on the API final grade predictor variable, and one of two categories to indicate success or failure in some phase of flight training. The total number of ways that a subject could be categorized given these three variables is n. Therefore, $n = 8 \times 6 \times 2 = 96$ different cells. The first cell would contain all those subjects with scores, PBI = 2, API = 1, ATTRITE = 0; the second cell all those subjects with scores PBI = 2, API = 2, ATTRITE = 0; the jth cell all those subjects with scores PBI = 6, API = 4, ATTRITE = 1; and the 96th and last cell all those subjects with scores PBI = 9, API = 6, ATTRITE = 1. These $n = 96$ cells can be arranged in any way that is convenient.

One traditional and convenient way of arranging these n cells is a two-dimensional table of rows and columns. In this arrangement, the n cells are called a cross-tabulation or contingency table. Using our previous example, the $n = 96$ cells could be displayed as two contingency tables each with eight rows for the eight categories of the PBI and six columns for the six categories of the API final grade. The first contingency table consists of all those subjects who failed some phase of flight training (ATTRITE = 0) while the second consists of all those subjects who passed all phases of flight training (ATTRITE = 1).

The symbol N will be used to indicate the total number of subjects allocated to the n cells. The number of subjects in the ith cell will be labeled N_i. Therefore,

$$\sum_{i=1}^{n} N_i = N.$$

Attached to each cell is a parameter, Q_i, that represents the probability for a subject to fall into the ith cell. The whole purpose of analyzing the contingency tables is to find values for the Q_i that are a good fit to the empirical frequency data in the PPS data base. Each separate consideration of a set of potential Q_i will be called a model and given the notation $\mathcal{M}_A, \mathcal{M}_B, \mathcal{M}_C \cdots$.

See Fig. 1 for a sketch of the salient points made in the above discussion. Two 8×6 contingency tables are shown. The table on the left consists of all the subjects in the data base who failed some phase of flight training, while the table on the right consists of those subjects who passed all phases of flight training. Each cell is numbered, starting with cell 1 and ending with cell 96. The actual number of subjects falling into cell 29 is N_{29}. The probability for a subject to be categorized into cell 16 is Q_{16}. The jth cell consists of the intersection ATTRITE = 1, PBI = 6, and API = 4. There are a total of

$$N = \sum_{i=1}^{n} N_i = 1,120$$

subjects in the data base who can be placed into one, and only one, of these 96 cells.

The models, $\mathcal{M}_A, \mathcal{M}_B, \mathcal{M}_C \cdots$, will embody various hypotheses regarding isoperformance curves. Such interesting hypotheses will concern independence or, the lack thereof, among the various Q_i. Other hypotheses to be investigated concern an increase of the Q_i with an increase in a variable score, and most especially, "tradeoffs" among certain of the Q_i. Addressing such hypotheses will allow us to accept or reject the idea that subjects can achieve equal probability of success in flight training by trading off high scores on one predictor variable with low scores on another predictor variable. We will always be guided by the principle of scientific parsimony, sometimes called "Occam's razor," to seek the simplest models that fit the data.

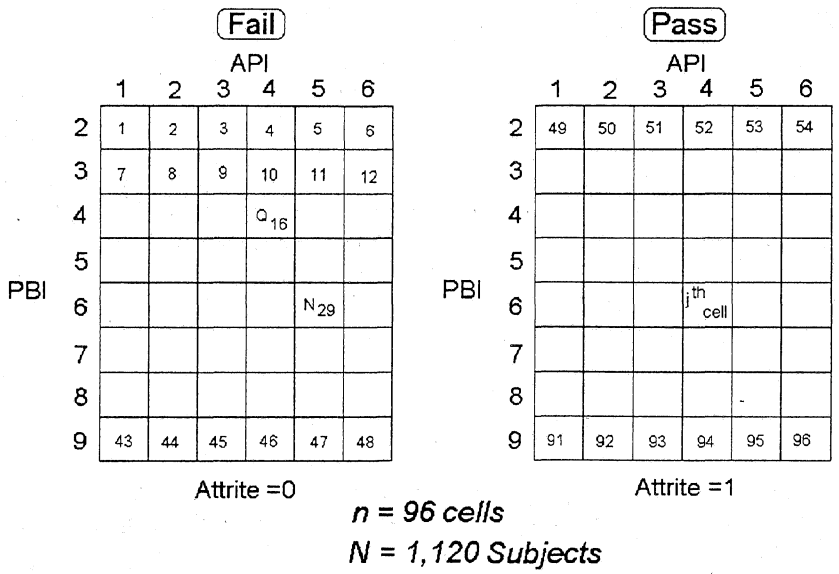

Figure 1: A sketch of a convenient arrangement of $n = 96$ cells into two 8×6 contingency tables.

THE BAYESIAN FORMALISM FOR MODEL EVALUATION

We first write Bayes's Formula for the posterior probability for any given model, say model $\mathcal{M}_A$. Then

$$P(\mathcal{M}_A|D,\mathcal{I}) = \frac{P(D|\mathcal{M}_A,\mathcal{I})\ P(\mathcal{M}_A|\mathcal{I})}{P(D|\mathcal{I})} \tag{1}$$

where D stands for the observed frequency data and $\mathcal{I}$ stands for all the background assumptions. The posterior probability for model $\mathcal{M}_A$ as conditioned on the truth of D and $\mathcal{I}$ is given on the left-hand side of Equation (1). The right hand side consists of the likelihood of the data conditioned on the truth of model $\mathcal{M}_A$ times the prior probability of model $\mathcal{M}_A$. The likelihood times prior component in the numerator is divided by the probability of the data. The denominator is the sum of all the terms that could appear in the numerator and thus is a sum over all possible models. In the future, we shall drop reference to the background assumptions, $\mathcal{I}$, to shorten the equations.

The Bayesian approach actually compares the ratio of posterior probabilities for any two models, say model $\mathcal{M}_A$ and $\mathcal{M}_B$. This allows us to remove the complicated sum, $P(D)$, from further consideration:

$$\frac{P(\mathcal{M}_A|D)}{P(\mathcal{M}_B|D)} = \frac{P(D|\mathcal{M}_A)\ P(\mathcal{M}_A)}{P(D|\mathcal{M}_B)\ P(\mathcal{M}_B)}. \tag{2}$$

Another assumption is usually introduced at this point. The prior probability of all models is considered to be equal. No favor or bias is shown for a model when compared with any other model. Under this assumption, the ratio of posterior probabilities for any two models reduces to the ratio of their respective likelihoods under each given model:

$$\frac{P(\mathcal{M}_A|D)}{P(\mathcal{M}_B|D)} = \frac{P(D|\mathcal{M}_A)}{P(D|\mathcal{M}_B)}. \tag{3}$$

Now the question is, "How do we find the likelihood of the data given a particular model?" To answer this question, we again invoke Bayes's Formula, but this time at a lower level. We now write down Bayes's Formula for the posterior probability for any given contingency table based on the data and a given model. The notation F_j

is used for the jth contingency table:

$$P(F_j|D,\mathcal{M}_A) = \frac{P(D|\mathcal{M}_A,F_j)\ P(F_j|\mathcal{M}_A)}{P(D|\mathcal{M}_A)}. \tag{4}$$

The structure of Bayes's Formula is the same as that in Equation (1), but it is now expressed as the posterior probability for the jth contingency table as conditioned on the assumed truth of a given model. Observe that the denominator in Equation (4) is the very expression needed to solve Equation (3).

Since the term $P(D|\mathcal{M}_A)$ in the denominator of Equation (4) is the sum of all the terms that could appear in the numerator, it is written explicitly as

$$P(D|\mathcal{M}_A) = \sum_{i=1}^{K} P(D|\mathcal{M}_A,F_i)P(F_i|\mathcal{M}_A). \tag{5}$$

This is a sum over all K possible contingency tables that could arise from considering N subjects allocated to n cells. Equation (5) is also an axiom from probability theory and is given the name *marginalization.*

The final step among these strictly Bayesian manipulations is to determine $P(D|\mathcal{M}_A)$ and $P(D|\mathcal{M}_B)$. This turns out to be a relatively simple problem because we are dealing with noise-free data. We assume that we have been careful enough to correctly record the various categorical variables so that we do not have to account for any attached error in the frequency counts for these variables. The likelihood for the jth contingency table is therefore equal to 1 when the frequency data match the numbers in the contingency table and 0 for any contingency table where the data do not match the numbers in the table. Symbolically, this means

$$P(F_j|D,\mathcal{M}_A) = \frac{1 \times P(F_j|\mathcal{M}_A)}{[\ 1 \times P(F_j|\mathcal{M}_A)\ ] + [\ \sum_{i=1}^{K-1} 0 \times P(F_i|\mathcal{M}_A)\ ]}. \tag{6}$$

The denominator in Equation (6), the term we are seeking, therefore simplifies tremendously, reducing to

$$P(D|\mathcal{M}_A) = P(F_j|\mathcal{M}_A). \tag{7}$$

Likewise,

$$P(D|\mathcal{M}_B) = P(F_j|\mathcal{M}_B). \tag{8}$$

This completes the section on the Bayesian manipulations. The next section continues the derivation through to the point where we can write computer programs to analyze actual data.

FORMULA FOR COMPUTING THE ACCEPTANCE OR REJECTION OF ANY GIVEN MODEL

As the derivation for the actual formula used to compute whether to accept or reject a model is rather long and involved, we relegate the mathematical derivation to the Appendix. The interested reader may go there for all the details. Only the final formula is presented here as Equation (9).

$$2N\sum_{i=1}^{n} f_i \ln\left(\frac{f_i}{Q_i}\right) \sim \chi^2\ (\nu\ df). \tag{9}$$

As mentioned before, N is the total number of subjects allocated to the contingency tables. The observed relative frequency for the ith cell is given the notation f_i and is equal to N_i/N. Q_i refers to any model for assigning the probabilities that we might propose to test. A superscript will be attached to the Q_i to identify which model is being discussed so that Q_i^A will mean the probabilities for each of the n cells under Model A. Likewise, Q_i^B will mean the probabilities for each of the n cells under Model B, and so on. Equation (9) says that the number computed on the left-hand side, which must be positive, will be distributed according to the χ^2 distribution with ν degrees of freedom. We adopt the usual convention of rejecting any proposed model for the Q_i if the value

computed on the left-hand side falls into the upper 5% region of the χ^2 distribution with the appropriate degrees of freedom.

Numerical Examples

In this section, we present some simple numerical examples to illustrate the use of Equation (9). Consider the situation of $n = 8$ cells, conveniently arranged into two 2×2 contingency tables. The first table consists of those subjects who failed some phase of flight training, while the second table consists of those who passed all stages of flight training. The total number of subjects in the data base is $N = 100$. Figure 2 shows these two tables with the actual frequencies, N_i, filled in for all eight cells.

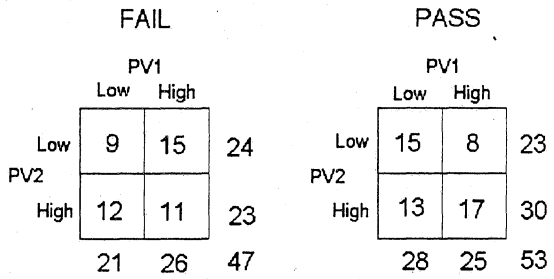

Figure 2: Two contingency tables showing fictitious data for 100 subjects. Each table shows the two predictor variables labeled $PV1$ and $PV2$ broken down into high and low scores. The table on the left shows the subjects who failed some phase of flight training while that on the right shows those who passed all phases of flight training.

There are two predictor variables, $PV1$ and $PV2$, with two levels for each predictor variable called "Low" and "High." The left-hand side of Equation (9) says to compute

$$2N\sum_{i=1}^{8} f_i \ln\left(\frac{f_i}{Q_i^A}\right) = (2 \times 100) \times$$

$$\left[\left(\frac{9}{100}\right)\ln\left(\frac{9/100}{Q_1^A}\right) + \left(\frac{15}{100}\right)\ln\left(\frac{15/100}{Q_2^A}\right) + \cdots \left(\frac{17}{100}\right)\ln\left(\frac{17/100}{Q_8^A}\right)\right].$$

What is model $\mathcal{M}_A$ so that we can substitute values for the Q_i^A? It is up to us to choose whatever hypothesis we are interested in investigating. For starters, let's pick the simplest hypothesis we can think of, that is, that all eight Q_i are equal. Now we can fill in the values for Q_i^A based on this hypothesis:

$$\begin{aligned} 2N\sum_{i=1}^{n} f_i \ln\left(\frac{f_i}{Q_i^A}\right) &= (2 \times 100) \times \\ &\quad \left[.09\ln\left(\frac{.09}{.125}\right) + .15\ln\left(\frac{.15}{.125}\right) + \cdots .17\ln\left(\frac{.17}{.125}\right)\right] \\ &= 200 \times (-.02957 + .027348 + \cdots + .052272) \\ &= 200 \times .02784 \\ &= 5.57. \end{aligned}$$

This value of 5.57 is compared to a χ^2 distribution with $\nu = 7$ df. The critical value that cuts off the upper 5% of this χ^2 distribution is 14.07. Therefore, 5.57 fits comfortably within this distribution and does not fall into the

rejection region. We cannot reject model $\mathcal{M}_A$ that says that the probability for a subject to fall into any of the eight categories is the same.

The probability of failure is $Q_1 + Q_2 + Q_3 + Q_4 = .50$, which is the same as the probability of passing $Q_5 + Q_6 + Q_7 + Q_8 = .50$. There is no effect due to the predictor variables either. There is the same probability of .25 of being categorized in the low or high level of either predictor variables no matter whether you pass or fail.

Now consider a second example as shown in Fig. 3. N remains at 100 subjects. The value computed by

FAIL

PV2 \ PV1	Low	High	
Low	5	7	12
High	4	6	10
	9	13	22

PASS

PV2 \ PV1	Low	High	
Low	19	19	38
High	22	18	40
	41	37	78

Figure 3: Two different contingency tables showing fictitious data for 100 subjects. Each table shows the two predictor variables labeled $PV1$ and $PV2$ broken down into high and low scores. The table on the left shows the subjects who failed some phase of flight training while that on the right shows those who passed all phases of flight training.

Equation (9) for this second example is 34.62, which falls into the upper 5% region of the χ^2 distribution with $\nu = 7$ df. Therefore, for these data, we must reject model $\mathcal{M}_A$ that says all eight $Q_i = .125$.

What alternative model might fit the data better than model $\mathcal{M}_A$? A casual inspection of Fig. 3 will reveal that the number of attritions is much less than the number of graduates. Let model $\mathcal{M}_B$ posit that the ratio of passing to failing is 4:1 so that

$$Q_{Pass} = Q_5 + Q_6 + Q_7 + Q_8 = .80$$

and

$$Q_{Fail} = Q_1 + Q_2 + Q_3 + Q_4 = .20.$$

Otherwise, there are no further constraints on the Q_i. Within the pass and fail groups we want the Q_i to be evenly spread out. This foreshadows the idea of maximum entropy to be introduced later in the report. The specification of model $\mathcal{M}_B$ is shown below in Table 1, along with the previous $\mathcal{M}_A$ and a new model, $\mathcal{M}_C$, to be discussed shortly.

The value computed by Equation (9) for model $\mathcal{M}_B$ is 1.62. The degrees of freedom must be adjusted downwards by 1 since we have introduced a new constraint. The critical value of the χ^2 distribution for $\nu = 6$ df is 12.59, so we are well within the region where we would accept model $\mathcal{M}_B$. The data do not allow us to reject the hypothesis that $P(\text{Pass}) = .80$ and $P(\text{Fail}) = .20$. However, by accepting model $\mathcal{M}_B$, we still do not see any effects due to either of the predictor variables.

For a third and final numerical example, extending the insights from the first two examples, please refer to Fig 4. For these data, model $\mathcal{M}_A$ has a value of 145.70, so it is clearly rejected. The revised thinking incorporated into model $\mathcal{M}_B$ is not much better at 115.47 and it too must be rejected.

We have to search for another plausible model that fits these new data. We will retain the hypothesis that $Q_{Pass} = .80$ and $Q_{Fail} = .20$ from model $\mathcal{M}_B$. Within each of the two groups there appears to be a strong effect due to the predictor variables, $PV1$ and $PV2$. If we now attribute a strong theoretical impact for low $PV1$ and $PV2$ scores to predict failure and high $PV1$ and $PV2$ scores to predict success, then a model like model $\mathcal{M}_C$ as shown in Table 1 might work.

Table 1: The specification of the eight Q_i values for models $\mathcal{M}_A$, $\mathcal{M}_B$, and $\mathcal{M}_C$.

Cell	$\mathcal{M}_A$ Q_i	$\mathcal{M}_B$ Q_i	$\mathcal{M}_C$ Q_i
1	.125	.05	.14
2	.125	.05	.02
3	.125	.05	.02
4	.125	.05	.02
5	.125	.20	.08
6	.125	.20	.08
7	.125	.20	.08
8	.125	.20	.56
Sums	1.00	1.00	1.00

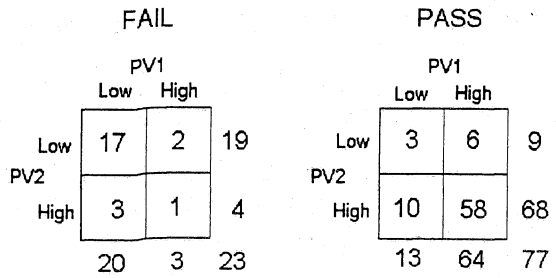

Figure 4: The final two contingency tables showing fictitious data for 100 subjects. Each table shows the two predictor variables labeled $PV1$ and $PV2$ broken down into high and low scores. The table on the left shows the subjects who failed some phase of flight training while that on the right shows those who passed all phases of flight training.

We examine in detail how all of the constraints are satisfied by this model. First of all, $\sum Q_i$ must equal 1. The second constraint is that $Q_1 + Q_2 + Q_3 + Q_4 = .20$ and $Q_5 + Q_6 + Q_7 + Q_8 = .80$. Thirdly, low $PV1$ and $PV2$ scores are equal for the fail group, $Q_1 + Q_3 = Q_1 + Q_2$, and high $PV1$ and $PV2$ scores are equal for the pass group, $Q_6 + Q_8 = Q_7 + Q_8$. Notice that the rule for keeping as many Q_i equal as possible is followed and that the ratio of 4:1 is followed as well as we move from the fail group to the pass group.

Equation (9) produces a value of 6.84 for model $\mathcal{M}_C$. The degrees of freedom must be reduced by one again to account for the added constraint. The critical value demarcating the 95% and 5% regions of the χ^2 for $\nu = 5$ *df* equals 11.07. This is a model we can accept. Table 2 summarizes the three models examined and their status for the data as given in Fig. 4.

Table 2: Summary of the three models examined for the data in Fig. 4.

Model	χ^2	df	Status
$\mathcal{M}_A$	145.70	7	Rejected
$\mathcal{M}_B$	115.40	6	Rejected
$\mathcal{M}_C$	6.84	5	Accepted

CALCULATING THE PROBABILITY OF EVENTS

Once we have found the most conservative model that can be accepted, it is a relatively simple matter to find the probability for any event of interest. For example, it is usually of interest to calculate the probability for attrition as a function of the predictor variables. A particular case can be expressed symbolically as

$$P(\text{Fail}|PV1 = \text{low and } PV2 = \text{high})$$

which is read as the probability of failing given that a subject scored low on predictor variable one and scored high on predictor variable two.

Probability theory provides a well-known solution for this situation. Abstractly, the probability of event A conditioned on the truth of event B is written

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{10}$$

where $P(A \cap B)$ refers to the joint occurrence of events A and B. If the event A can be broken down into K mutually exclusive and exhaustive events, then the probability of the jth category of A is written as[1]

$$P(A_j|B) = \frac{P(A_j \cap B)}{\sum_{i=1}^{K} P(A_i \cap B)}. \tag{11}$$

In the case that concerns us, A is the event of success in flight training and it is broken down into just $K = 2$ categories, Pass or Fail. These two categories are mutually exclusive and exhaustive. That is to say, a given subject must be in one of these two categories and given that he or she is in one of the two categories, she or he cannot be in the other category. The conditioning information B is the score on the predictor variable.

Equation (11) can now be rewritten simply as

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(A_1 \cap B) + P(A_2 \cap B)}. \tag{12}$$

If we let event A_1 stand for fail, event A_2 for pass, and event B for a low score on predictor variable one, then Equation (12) becomes

$$P(\text{Fail}|PV1 = \text{low}) = \frac{P(\text{Fail and } PV1 = \text{low})}{P(\text{Fail and } PV1 = \text{low}) + P(\text{Pass and } PV1 = \text{low})}. \tag{13}$$

Our acceptable model will then provide us with the probabilities to substitute into Equation (13).

Let us return to the first numerical example as depicted in Fig. 2. The numerator in Equation (13) is the intersection of the Fail cells with $PV1$ =low, which is $Q_1 + Q_3 = .25$. At this point, we need find only the second term in the denominator. This is the intersection of the Pass and $PV1 =$ low cells, which is $Q_5 + Q_7 = .25$. Substituting these probabilities into Equation (13) yields

$$\begin{aligned} P(\text{Fail}|PV1 = \text{low}) &= \frac{.25}{.25 + .25} \\ &= .50. \end{aligned} \tag{14}$$

However, this probability is the same as P(Fail) not conditioned on any information, i.e.,

$$Q_1 + Q_2 + Q_3 + Q_4 = .50.$$

[1] This is Bayes's Formula written in another way.

The extra information contained in the $PV1$ score was of no help whatsoever. It is irrelevant information.

The same tactic just outlined can be employed when conditioning on any number of predictor variables. Say that we are interested in the information provided by the scores on both $PV1$ and $PV2$:

$$P(\text{Fail}|PV1 = \text{low and } PV2 = \text{low})$$

This is equal to

$$\frac{P(\text{Fail and } PV1 = \text{low and } PV2 = \text{low})}{P(\text{Fail and } PV1 = \text{low and } PV2 = \text{low}) + P(\text{Pass and } PV1 = \text{low and } PV2 = \text{low})}.$$

The cell that is the intersection in the numerator is $Q_1 = .125$, and the cell that is the intersection of the second term in the denominator is $Q_5 = .125$. Therefore,

$$\begin{aligned} P(\text{Fail}|PV1 = \text{low and } PV2 = \text{low}) &= \frac{.125}{.125 + .125} \\ &= .50. \end{aligned} \tag{15}$$

So, once again, the information from both predictor variables was completely irrelevant or useless. Conditioning on this extra information did not change the probability of failing from what we knew when we did not have this information, that is, $P(\text{Fail}) = .50$.

What about the second numerical example as illustrated in Fig. 3? Does the extra information from the predictor variables help here? The same formula applies so all we have to do is plug in the correct values for the probabilities. In the second numerical example, model $\mathcal{M}_B$ was found to be an acceptable model with the values $Q_1 = .05$ and $Q_5 = .20$

$$\begin{aligned} P(\text{Fail}|PV1 = \text{low and } PV2 = \text{low}) &= \frac{.05}{.05 + .20} \\ &= .20. \end{aligned} \tag{16}$$

However, $P(\text{Fail}) = Q_1 + Q_2 + Q_3 + Q_4 = .20$ as well. Here also the predictor variables are providing no useful information with regard to the probability of failing.

In the third example, we finally do observe an influence on the probability of failing by knowing the scores on the predictor variables. Refer back to Table 1 where the values of $Q_1 = .14$ and $Q_5 = .08$ for model $\mathcal{M}_C$ are listed. In this case,

$$\begin{aligned} P(\text{Fail}|PV1 = \text{low and } PV2 = \text{low}) &= \frac{.14}{.14 + .08} \\ &= .64. \end{aligned} \tag{17}$$

Knowing that a subject scored low on both $PV1$ and $PV2$ raised the probability of failing from .20 to .64. The scores on these predictor variables are valuable information that permit us to materially change our assessment of failing.

HOW TO FIND MODELS CONSISTING OF PRESCRIBED INFORMATION

We have nearly completed the quantitative overview for the analysis that we intend to carry out for isoperformance curves. One item still remains to be discussed, however. How does one manage to assign values to the Q_i and thus arrive at plausible models?

In the numerical examples that were presented earlier, this task was not so difficult. The assigned values could be intuited without too much difficulty. But we really require some disciplined, systematic method for assigning the Q_i that doesn't depend upon someone's intuitive insight. In this final section, we provide such a method for assigning the Q_i; a method that has a number of attractive features. The method is called the *Maximum Entropy Principle* (MEP), and it permits us to systematically generate only the models with the known information that we have consciously inserted and to avoid models with hidden assumptions about information we do not wish to insert.

The mathematical derivation behind the MEP will not be presented in this report. A lengthy and thorough tutorial on this subject is available in Volume II of my textbook [3]. The treatment in my book is based entirely upon the seminal work on the MEP by Edwin T. Jaynes. Instead, we present here only the formulas that show how one assigns values to the Q_i.

Equation (18) below presents the simplest form of the MEP where only one piece of information has been inserted into a model:

$$Q_i = \frac{e^{\lambda_1 A_1(x_i)}}{\sum_{i=1}^{n} e^{\lambda_1 A_1(x_i)}} \tag{18}$$

where λ_1 is a constant value called a Lagrange multiplier. The value for λ_1 can be determined through numerical methods. We shall use only a very simple trial and error technique to find λ_1. $A_1(x_i)$ is the notation for a *constraint* on the n values of the Q_i. As the argument x_i indicates, there exists a separate value for each of the Q_i we are trying to assign. The denominator in Equation (18) consists of the sum over all n possible cells, that is, the sum of each possible term that could appear in the numerator. These remarks about the MEP formula will be clarified by the numerical examples to follow.

As the first, and easiest example of the MEP, consider the case where $\lambda_1 = 0$. This is the case where we are inserting no information in the form of a constraint about the model. Actually this is not quite true. There is one piece of information that is universally present in the MEP. This is the constraint that all n Q_i must sum to 1. This constraint is universally present because all probability distributions must sum (or integrate) to 1. The essence of the MEP is that the assignment of the Q_i must have maximum entropy subject to the constraints imposed. When the only constraint is that the sum of the Q_i must sum to 1, the distribution with maximum entropy is found by applying Equation (18) with $\lambda_1 = 0$:

$$Q_i = \frac{e^{\lambda_1 A_1(x_i)}}{\sum_{i=1}^{n} e^{\lambda_1 A_1(x_i)}} \tag{19}$$

$$= \frac{e^{0 \times A_1(x_i)}}{\sum_{i=1}^{n} e^{0 \times A_1(x_i)}} \tag{20}$$

$$e^{0 \times A_1(x_i)} = 1 \tag{21}$$

$$\sum_{i=1}^{n} e^{0 \times A_1(x_i)} = n \tag{22}$$

$$Q_i = \frac{1}{n}. \tag{23}$$

This is exactly model $\mathcal{M}_A$ that we assigned intuitively in the earlier numerical examples.

Let us see how we could arrive at model $\mathcal{M}_B$ using the MEP. We now introduce a constraint as one piece of information that we wish to insert into the assignment. That constraint is that $P(\text{Fail}) = .20$. Perhaps the easiest way of writing this constraint is to place a 1 in the first four cells and a 0 in the last four cells as the values for $A_1(x_i)$. Table 3 shows the subsequent computation of the Q_i using Equation (18).

Table 3: The MEP assignment of the Q_i for one constraint. This corresponds to model $\mathcal{M}_B$.

Cell	$A_1(x_i)$	$\exp(\lambda_1 A_1(x_i))$	Q_i
1	1	.25	.05
2	1	.25	.05
3	1	.25	.05
4	1	.25	.05
5	0	1.00	.20
6	0	1.00	.20
7	0	1.00	.20
8	0	1.00	.20
	$\lambda_1 = -1.3863$	5.00	1.00

Table 4: The MEP assignment of the Q_i for two constraints. This corresponds to model $\mathcal{M}_C$.

Cell	$A_1(x_i)$	$A_2(x_i)$	$\exp(\lambda_1 A_1(x_i) + \lambda_2 A_2(x_i))$	Q_i
1	1	1	1.75	.14
2	1	0	.25	.02
3	1	0	.25	.02
4	1	0	.25	.02
5	0	0	1.00	.08
6	0	0	1.00	.08
7	0	0	1.00	.08
8	0	1	7.00	.56
	$\lambda_1 = -1.3863$	$\lambda_2 = 1.94593$	12.50	1.00

The MEP formalism can be extended straightforwardly to more than one constraint. An additional Lagrange multiplier, λ_2, and constraint function, $A_2(x_i)$, are placed into Equation (18). This results in

$$Q_i = \frac{e^{\lambda_1 A_1(x_i) + \lambda_2 A_2(x_i)}}{\sum_{i=1}^{n} e^{\lambda_1 A_1(x_i) + \lambda_2 A_2(x_i)}}. \tag{24}$$

We can use Equation (24) to find models with two pieces of information inserted, and we can be sure that only these two pieces are involved. An example of such a model with two constraints was model $\mathcal{M}_C$. In this model, we entertained the hypothesis that a predictor variable was associated with success in training in addition to a given value for the probability of failing. Table 4 shows the numerical computations needed to assign the Q_i values for this model ensuing from Equation (24).

All three constraints are satisfied by this assignment to the Q_i. The universal constraint that all Q_i sum to 1 is satisfied. The constraint inserted by model $\mathcal{M}_B$ that the probability of failing is equal to .20 is satisfied. The additional constraint inserted by model $\mathcal{M}_C$ that low scores on both predictor variables lead to a higher probability of failing and that high scores on both predictor variables lead to a higher probability of passing is satisfied. The MEP also tells us the correct degrees of freedom for the χ^2 test. It is $\nu = n -$ number of constraints, which is $\nu = 5$ for model $\mathcal{M}_C$.

It is important to emphasize that *this information and only this information* has been inserted into the

assignment. This means that the information entropy of the assignment given to the Q_i in Table 4 is the maximum possible entropy given the constraints. There are other assignments to the Q_i that satisfy all three constraints, but they possess an entropy that is less than the MEP assignment.

SUMMARY

We have shown that a well-known formula from information theory can be derived from a Bayesian model evaluation approach to contingency tables. The value computed by this function of the cross-entropy is compared to a χ^2 distribution to judge whether a proposed model is acceptable. Such models refer to probabilities for a flight candidate being placed in a particular cell of a contingency table. Each cell represents an intersection of some number of predictor variables and a criterion variable. Simple numerical examples illustrating this concept were presented in this report. A follow-on report [2] will use the techniques developed here to analyze isoperformance issues. In this practical application, PBI and API scores are used as the predictor variables and attrition in any phase of flight training is employed as the criterion variable.

REFERENCES

1. Jones, M.B. and Kennedy, R.S. Isoperformance Curves in Applied Psychology. *Human Factors,* 38(1), 167–182, 1996.
2. Blower, D.J. Statistical Analysis of Isoperformance Issues in Navy Flight Training. *NAMRL Technical Report.* In review. 1999
3. Blower, D. J. *An Introduction to Scientific Inference. Volume II: The Maximum Entropy Principle.* Published privately by the author, 1999.

Appendix
Derivation of Information Entropy Formula from Bayesian Model Evaluation

At the stage where we left the derivation in the earlier section of this paper, we had to find the prior probability for any contingency table based on the truth of some given model. As we mentioned earlier, each model assigns some definite value to each of the Q_i values, n in number. Each Q_i, remember, assigns a probability for a subject to be categorized into the ith cell of the contingency table. The prior probability for the numbers appearing in any contingency table is based on the multinomial formula. The prior probability for any contingency table based on model $\mathcal{M}_A$ is therefore,

$$P(F_j|\mathcal{M}_A) = W(F_j)\, Q_1^{A^{N_1}} Q_2^{A^{N_2}} \cdots Q_n^{A^{N_n}}. \tag{25}$$

In the same manner, the prior probability for the same contingency table based on a different model, model $\mathcal{M}_B$, is

$$P(F_j|\mathcal{M}_B) = W(F_j)\, Q_1^{B^{N_1}} Q_2^{B^{N_2}} \cdots Q_n^{B^{N_n}}. \tag{26}$$

The symbol $W(F_j)$ refers to the multiplicity factor; the number of ways that each contingency table could be formed without regard to the order that subjects are placed into the cells.

We can now form the ratio of posterior probabilities for the two models as

$$\frac{P(\mathcal{M}_A|D)}{P(\mathcal{M}_B|D)} = \frac{P(D|\mathcal{M}_A)P(\mathcal{M}_A)}{P(D|\mathcal{M}_B)P(\mathcal{M}_B)}. \tag{27}$$

Because we are assuming that the prior probabilities of the two models are equal, we can write the ratio of posterior probabilities as the ratio of likelihoods:

$$\frac{P(\mathcal{M}_A|D)}{P(\mathcal{M}_B|D)} = \frac{P(D|\mathcal{M}_A)}{P(D|\mathcal{M}_B)} \tag{28}$$

$$= \frac{P(F_j|\mathcal{M}_A)}{P(F_j|\mathcal{M}_B)} \tag{29}$$

$$= \frac{W(F_j)\, Q_1^{A^{N_1}} Q_2^{A^{N_2}} \cdots Q_n^{A^{N_n}}}{W(F_j)\, Q_1^{B^{N_1}} Q_2^{B^{N_2}} \cdots Q_n^{B^{N_n}}}. \tag{30}$$

The multiplicity factor cancels in this ratio so that

$$\frac{P(\mathcal{M}_A|D)}{P(\mathcal{M}_B|D)} = \frac{Q_1^{A^{N_1}} Q_2^{A^{N_2}} \cdots Q_n^{A^{N_n}}}{Q_1^{B^{N_1}} Q_2^{B^{N_2}} \cdots Q_n^{B^{N_n}}}. \tag{31}$$

At this juncture, we bring in a classical theorem from non-Bayesian statistics, the asymptotic property of the likelihood ratio test [1]. This theorem states that a quantity, $-2\ln\lambda$, where λ is a ratio of likelihoods as in Equation (31), will be distributed according to the chi-square (χ^2) distribution as $N \rightarrow \infty$[2]. Jaynes's similar Entropy Concentration Theorem [2] can also be invoked. This kind of transformation carried out on the posterior probabilities of the two models will then be distributed as a χ^2 distribution,

$$-2\ln\left[\frac{P(\mathcal{M}_A|D)}{P(\mathcal{M}_B|D)}\right] \sim \chi^2\ (\nu\ df). \tag{32}$$

[2] this use of λ is to be distinguished from its use as the Lagrange multiplier.

Substituting the right hand side of Equation (31) as the likelihood ratio, the transformation yields,

$$-2\ln\left[\frac{Q_1^{A^{N_1}}Q_2^{A^{N_2}}\cdots Q_n^{A^{N_n}}}{Q_1^{B^{N_1}}Q_2^{B^{N_2}}\cdots Q_n^{B^{N_n}}}\right] = -2\left[N_1\ln\left(\frac{Q_1^A}{Q_1^B}\right)+N_2\ln\left(\frac{Q_2^A}{Q_2^B}\right)+\cdots N_n\ln\left(\frac{Q_n^A}{Q_n^B}\right)\right] \quad (33)$$

$$= -2\sum_{i=1}^{n} N_i\ln\left(\frac{Q_i^A}{Q_i^B}\right) \quad (34)$$

$$-2\sum_{i=1}^{n} N_i\ln\left(\frac{Q_i^A}{Q_i^B}\right) = 2\sum_{i=1}^{n} N_i\ln\left(\frac{Q_i^B}{Q_i^A}\right) \quad (35)$$

$$2\sum_{i=1}^{n} N_i\ln\left(\frac{Q_i^B}{Q_i^A}\right) \sim \chi^2\ (\nu\ df). \quad (36)$$

In Equation (35), we made use of the following identity

$$\ln\left(\frac{x}{y}\right) = \ln x - \ln y$$

$$-(\ln x - \ln y) = \ln y - \ln x$$

$$= \ln\left(\frac{y}{x}\right).$$

If we let Q_i^B stand for the very best model as a benchmark reference, then the Q_i for model $\mathcal{M}_B$ will be exactly the same as the observed frequencies, N_i/N. Equation (36) now looks like this after making this choice for model $\mathcal{M}_B$,

$$2\sum_{i=1}^{n} N_i\ln\left(\frac{Q_i^B}{Q_i^A}\right) = 2\sum_{i=1}^{n} N_i\ln\left(\frac{N_i/N}{Q_i^A}\right). \quad (37)$$

Now we want to get Equation (37) into a form that expressly shows the frequencies, f_i. Multiply and divide the right-hand side of Equation (37) by N to achieve

$$2N\times\sum_{i=1}^{n}\frac{N_i}{N}\ln\left(\frac{N_i/N}{Q_i^A}\right) = 2N\sum_{i=1}^{n} f_i\ln\left(\frac{f_i}{Q_i^A}\right). \quad (38)$$

The summation term in Equation (38) is well-known in information theory as *cross-entropy.* We will give it the notation $H(f,\mathcal{M}_A)$ to indicate that it is the information cross-entropy of the actual frequencies with some model for the Q_i, here Model A,

$$H(f,\mathcal{M}_A) \equiv \sum_{i=1}^{n} f_i\ln\frac{f_i}{Q_i^A}. \quad (39)$$

As our final statement, then, we see that any model can be accepted or rejected on the basis of its information cross-entropy and where it falls in relation to a χ^2 distribution

$$2NH(f,\mathcal{M}) \sim \chi^2\ (\nu\ df). \quad (40)$$

We will adopt the usual criterion for rejection of a proposed model on the basis of whether $2NH(f,\mathcal{M})$ falls into the upper 5% region of the χ^2 distribution.

REFERENCES

1. Hoel, P.G. *Introduction to Mathematical Statistics.* 4th edition. John Wiley & Sons, 1971.
2. Jaynes, E.T. Concentration of Distributions at Entropy Maxima. *Papers on Probability, Statistics and Statistical Physics.* ed. by R. D. Rosenkrantz, Kluwer Academic Publishers, 1989.

Reviewed and approved 10/1/99

R. R. STANNY, Ph.D.
Technical Director

This research was sponsored by the Office of Naval Research under work unit 6.2B994 03330/0126-7903.

The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1 OCT 1999	

4. TITLE AND SUBTITLE
Some General Quantitative Considerations for the Statistical Analysis of Isoperformance Curves

5. FUNDING NUMBERS
B994 03330/0126
Work Unit 7903

6. AUTHOR(S)
D. J. Blower

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Aerospace Medical Research Laboratory
51 Hovey Road
Pensacola Fl 32508-1046

8. PERFORMING ORGANIZATION REPORT NUMBER
NAMRL-1406

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217-5660

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Approved for public release; distribution unlimited.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

In this paper we present a recommended quantitative approach for analyzing the concept of isoperformance. The ideas outlined here rely upon the Bayesian version of *model evaluation*. We define models as hypotheses about the probabilities of subjects being categorized by a combination of predictor variables and criterion variables. From this foundation, a computational formula is derived whose value can be compared to a χ^2 distribution. For example, we are we are often interested in calculating the probability of a subject failing during some phase of flight training given that we have information on certain predictor variables. We would like to ascertain whether the extra information contained in such predictor variables is useful. If it is useful, then it enables us to predict the probability of failure for any given student. This ability to predict a change in the probability of failure, either in the upwards or downwards direction, is very helpful to managers and decision makers in the training community. In addition, these techniques can help answer the question of whether a candidate for flight training can "trade-off" a high score on one predictor variable for a low score on a different predictor variable. In particular, we would like to investigate the possibility of trading off different classes of predictor variable. In particular, we would like to investigate the possibility of trading off different classes of predictor variables, say cognition information processing variables and personality variables, and still achieve the same level of performance. The maximum entropy principle is used as a systematic disciplined approach to find parsimonious models.

14. SUBJECT TERMS
Isoperformance, Bayesian model evaluation, Information theory, Maximum entropy

15. NUMBER OF PAGES
26

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR