Ann. Inst. Statist. Math.
36 (1984), Part A, 485-497

A DECISION-THEORETIC APPROACH TO SOME SCREENING PROBLEMS

ULRICH MENZEFRICKE

(Received Sept. 16, 1983)

Summary

Suppose an item is acceptable if its measurement on the variable of interest Y is $Y \leq u$. It may be expensive (or impossible) to measure Y, and a correlated variable X exists which is relatively inexpensive to measure and is used to screen items, i.e., to declare them acceptable if $X \leq w$. We examine two situations in both of which l acceptable items are needed. (i) Before use of the item, Y is measured directly to ensure acceptability: Should X be used for screening purposes before the Y measurement or not? (ii) Y cannot be measured directly before use, but screening is possible to determine the items that are to be used. We assume that X and Y have a bivariate normal distribution for which the parameters are known. Some comments are made about the case when the parameters are not known.

1. Introduction

We address ourselves to the decision-theoretic treatment of screening units or items for acceptance. This problem arises in various settings. An example from manufacturing is given in Owen, Li and Chou [6], where an automobile seat is attached to the frame by welding and it is desired that the weld hold even after a large stress is applied. It is possible to screen items by measuring X-ray penetration of the weld which is negatively correlated with the strength of the weld, rather than to measure the strength of the weld directly. Another example is the screening of applicants for employment, where ultimate performance of an individual is thought to be related to the score on an aptitude test.

A sampling-theoretic treatment of this problem is given, e.g., by Owen and his co-workers (Owen and Boddie [5], Owen, Li and Chou

Keywords: Bivariate normal probabilities, matric-variate t distribution, screening for acceptance.

[6], Owen and Su [7]), and by Madsen [4]. An item is acceptable if the variable of interest Y has $Y \leqq u$. In some situations the variable Y may be expensive or even impossible to measure before the item is used, and so it is difficult to ensure acceptability of the item. However, a variable X which is correlated with Y can be measured more easily, and this variable is to be used to screen items. Based on an X measurement items are deemed acceptable if $X \leqq w$. The variables X and Y are taken to have a bivariate normal distribution with means μ_x and μ_y, standard deviations σ_x and σ_y, and correlation p,

$$p(Z|\mu, \sigma)=(2\pi)^{-1}|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(Z-\mu)^t\Sigma^{-1}(Z-\mu)\right], \tag{1.1}$$

where $Z=(X, Y)^t$, $\mu=(\mu_x, \mu_y)^t$ and Σ is the covariance matrix. The following probabilities, computed from (1.1), will be used frequently:

$$\beta=\Pr(X \leqq w), \quad \gamma=\Pr(Y \leqq u) \quad \text{and} \quad \delta=\Pr(Y \leqq u|X \leqq w). \tag{1.2}$$

It is typically desirable to increase the marginal probability that an item is acceptable, γ, to a higher value, δ, after screening. In most, applications, the value of u is predetermined whereas that of w needs to be determined. When $\rho \geqq 0$, β increases whereas δ decreases in w. We assume throughout that $\rho \geqq 0$, if $\rho \leqq 0$ replace X by $-X$.

Owen and Boddie [5] and Owen and Su [7] treat the problem where β, the probability of acceptability before screening, is to be raised to δ after screening by choosing an appropriate value of w. They use a sampling theory approach to give solutions for various cases depending on which parameters are unknown. Owen, Li, and Chou [6] examine the problem where it is necessary that at least l out of n items are found acceptable with some specified probability. It is again necessary to choose an appropriate value for w. They give solutions for the cases when all parameters are known and when they are unknown. Because of the high number of parameters, it is not possible to provide extensive tables and the authors suggest that a computer algorithm be developed.

In this paper, we use a decision theoretic approach to problems similar to the one discussed by Owen, Li, and Chou [6]. In Section 2 we treat the case with all parameters known for which results are straightforward to obtain. In Section 3 the parameters are not known and have to be estimated. The computational problems become considerably more difficult than in Section 2. Some numerical examples are given in Section 4.

2. Case where all parameters are known

In this section, the problem will be stated more generally than in the introduction. We assume that there is a vector X_p of r correlated variables to be used for screening and that the joint distribution of X_p and Y is $(r+1)$-variate normal with mean vector and covariance matrix

$$\mathrm{E}\begin{pmatrix} X_p \\ Y \end{pmatrix} = \begin{pmatrix} \mu_p \\ \mu_y \end{pmatrix} \quad \text{and} \quad V\begin{pmatrix} X_p \\ Y \end{pmatrix} = \begin{pmatrix} \Sigma_{pp} & \Sigma_{py} \\ \Sigma_{py}^t & \sigma_y^2 \end{pmatrix}, \tag{2.1}$$

where μ_p and Σ_{py} are $(r\times 1)$ vectors, Σ_{pp} is an $(r\times r)$ matrix, and μ_y and σ_y are scalars as defined in the introduction. As before an item is acceptable if $Y\leqq u$ and, without screening, a proportion $\gamma=\Pr(Y\leqq u)$ of such items is acceptable.

When using the correlated vector X_p for screening purposes, we need to choose an r-dimensional set A such that an item is deemed acceptable based on the correlated vector X_p, if $X_p\in A$. If we let $\beta=\Pr(X_p\in A)$ and $\delta=\Pr(Y\leqq u\,|\,X_p\in A)$, then it is desirable to choose A such that the conditional proportion of items actually acceptable, δ, be maximized for a given proportion of items deemed acceptable, β. It can be shown that under this condition the set A should be chosen such that

$$A=\{X_p\,|\,X=\Sigma_{py}^t\Sigma_{pp}^{-1}X_p\leqq w\},$$

where w is a constant to be chosen, i.e., the problem reduces to the bivariate one given in the introduction, where the correlated variable X is a linear combination of the correlated variables in X_p.

We then have that the parameters μ_x, σ_x and ρ, introduced in the previous section, relate to the parameters in (2.1) as follows,

$$\begin{aligned} &\mu_x=\Sigma_{py}^t\Sigma_{pp}^{-1}\mu_p, \\ &\sigma_x^2=\Sigma_{py}^t\Sigma_{pp}^{-1}\Sigma_{py} \quad \text{and} \\ &\rho=(\Sigma_{py}^t\Sigma_{pp}^{-1}\Sigma_{py})^{1/2}/\sigma_y=\sigma_x/\sigma_y\geqq 0. \end{aligned} \tag{2.2}$$

In the remainder of this section, we assume the parameters in (2.1) and thus μ_x, μ_y, σ_x, σ_y, and ρ to be known. This assumption may be appropriate if considerable information exists about the process generating X_p, and therefore X, and Y. Thus, the predictive distribution for a future observation $Z=(X, Y)^t$ is a bivariate normal distribution as in (1.1). The predictive probabilities β, γ, and δ computed from this distribution will form the basis for an evaluation of different screening procedures.

2.1. *Direct measurement of the performance variable is possible*

In this section we examine some models for which it is assumed that the performance variable Y can be measured directly before the item is to be used to ensure that the item is acceptable. It is to be determined if X should be measured first to screen for items which have a high probability of being acceptable on the performance variable. Let

$$c_1 = \text{cost of an } X\text{-measurement per unit, and}$$

$$c_2 = \text{cost of a } Y\text{-measurement per unit}.$$

We assume that l acceptable units are needed and that we sample until l such units are found (without loss of generality, we could take $l=1$). If Y is measured directly, action a_1, then the cost of measurement will be

$$c_2(l+f_y) ,$$

where f_y, the number of unacceptable items for which Y-measurements have been taken, has a negative binomial distribution with parameters l and γ. The expected cost of measuring Y directly is denoted by

$$\text{(2.3)} \qquad \mathrm{E}\, C(a_1) = \mathrm{E}_{f_y} \{c_2(l+f_y)\} = c_2 l/\gamma .$$

If we screen items first by determining whether $X \leqq w$ or not before making the Y-measurement, action a_2, the cost of measurement until l acceptable items are found is

$$c_1 f_x + (c_1+c_2)(l+f_y) ,$$

where, for a given f_y, f_x, the number of items which fail the screening test, has a negative binomial distribution with parameters $(l+f_y)$ and β, and, under the screening alternative, f_y has a negative binomial distribution with parameters l and δ. The expected cost of this screening alternative is

$$\text{(2.4)} \qquad \mathrm{E}\, C(a_2) = \mathrm{E}_{f_y} [\mathrm{E}_{f_x} \{c_1 f_x + (c_1+c_2)(l+f_y)\}] = c_1 \left[\frac{l}{\delta\beta}\right] + c_2 \left[\frac{l}{\delta}\right] .$$

Note that $\mathrm{E}\, C(a_2)$ depends on w, the cutoff value for the screening variable, and that an optimal value for w has to be found. The first term in brackets on the right-hand side of (2.4) represents the expected number of items checked on the screening variable X. This term is a monotonically decreasing function in w since $\delta\beta = \Pr(X \leqq w,\ Y \leqq u)$ increases in w. The second term in brackets represents the expected number of items which are also checked on the performance variable

Y. Since δ is a decreasing function in w this term is increasing in w.

It is easy to see that $\mathrm{E}\,C(a_2)\to\infty$ as $w\to-\infty$, and that $\mathrm{E}\,C(a_2)\to l(c_1+c_2)/\gamma$ as $w\to\infty$. Detailed numerical investigations suggest that $\mathrm{E}\,C(a_2)$ has a unique minimum for some value of w.

The higher c_2 is relative to c_1, i.e., the more costly it is to measure Y, the lower will be the optimal value of w. This optimal value for w can be used in (2.4) to compare the expected cost of the two alternatives. Screening is preferable if $\mathrm{E}\,C(a_2)\leqq\mathrm{E}\,C(a_1)$ or if $c_1/c_2\leqq\delta\beta/\gamma-\beta$ or if

$$\frac{c_1}{c_2}\leqq\Pr(X\leqq w\mid Y\leqq u)-\Pr(X\leqq w)\,, \tag{2.5}$$

where w is taken to be the optimal value of w obtained by minimizing $\mathrm{E}\,C(a_2)$. Note that (2.5) does not depend on l, the number of acceptable units needed. If X and Y are independent, i.e., if $\rho=0$, the right-hand side of (2.5) equals 0 and screening cannot be superior to measuring Y directly. The conditional probability $\Pr(X\leqq w\mid Y\leqq u)$ will be larger than $\Pr(X\leqq w)$ if ρ is positive, the difference increasing as ρ increases. Screening is then preferable if ρ is sufficiently high.

In many situations measuring Y will be considerably more expensive than measuring X, thus c_1/c_2 will be close to 0, and ρ need not be very high for screening to be superior to not screening. It can be seen from (2.4) and the discussion following it that, as $c_1/c_2\to0$, the optimal value of $w\to-\infty$ and only extremely few items pass the screening stage.

In some situation, e.g., when lifetime is the performance variable of interest, measurement of Y before use is clearly impossible and the model discussed in this section does not apply.

2.2. *Direct measurement of the performance variable is not possible*

We shall now investigate the situation where it is not possible to measure the variable Y before using a unit. We assume that l units acceptable on the Y variable (i.e., having $Y\leqq u$) are needed and any screening has to be carried out before any of the items are used. The model consists of choosing n items of which at least l have to be acceptable. If less than l are acceptable, a penalty will be incurred, which we assume not to depend on how many units short we are. Any unacceptable item among the n units leads to an additional cost. Let

c_1 = cost of an X-measurement per unit,

c_3 = cost of not having at least l acceptable units, and

c_4 = cost of using an unacceptable item (per unit).

We will consider two alternatives. (i) a_1: Pick n_1 units and use them, and (ii) a_2: Screen units first until n_2 are available for use. It is then easy to show that the expected costs for the two alternatives are

$$\mathrm{E}\,C(a_1)=c_3\sum_{i=0}^{l-1}\begin{bmatrix}n_1\\ i\end{bmatrix}\gamma^i(1-\gamma)^{n_1-i}+c_4n_1(1-\gamma)\ , \tag{2.6}$$

and

$$\mathrm{E}\,C(a_2)=c_1n_2/\beta+c_3\sum_{i=0}^{l-1}\begin{bmatrix}n_2\\ i\end{bmatrix}\delta^i(1-\delta)^{n_2-i}+c_4n_2(1-\delta)\ . \tag{2.7}$$

The decision variable for the first alternative is n_1. If we let $g(n_1)=\mathrm{E}\,C(a_1)$, then

$$g(n_1+1)-g(n_1)=c_4(1-\gamma)+c_3\left\{\sum_{i=0}^{l-1}\begin{bmatrix}n_1+1\\ i\end{bmatrix}\gamma^i(1-\gamma)^{n_1+1-i}-\sum_{i=0}^{l-1}\begin{bmatrix}n_1\\ i\end{bmatrix}\gamma^i(1-\gamma)^{n_1-i}\right\}\ . \tag{2.8}$$

Hald ([3], p. 191) shows that the expression in braces is: (i) decreasing, (ii) concave if $n_1\leqq(l-1)/\gamma$, and (iii) convex if $n_1\geqq(l-1)/\gamma$. Thus, the form of (2.8) implies that $\mathrm{E}\,C(a_1)$ has at most two local minima. If there are two such minima, one of them is at $n_1=l$. Note that as $c_4/c_3\to0$, the optimal value of $n_1\to\infty$.

In the second alternative there are two decision variables, n_2 and w. As a function of n_2, both the first and third terms in (2.7) increase in n_2 whereas the second term decreases. As a function of w, the first term in (2.7) decreases in w since β increases in w and the remaining two terms increase in w since δ decreases in w. Given that there is a positive relationship between X and Y, $c_1\to0$ implies that screening is free and that $w\to-\infty$, i.e., exceptionally few items pass the screening test all of which having conditional probability of being acceptable $\delta\to1$ so that $n_1\to l$ and no bad items are used.

Screening is desirable if $\mathrm{E}\,C(a_1)\geqq\mathrm{E}\,C(a_2)$ or if

$$\frac{c_1}{c_3}\frac{n_2}{\beta}+\frac{c_4}{c_3}[n_2(1-\delta)-n_1(1-\gamma)]\leqq\sum_{i=0}^{l-1}\left[\begin{bmatrix}n_2\\ i\end{bmatrix}\delta^i(1-\delta)^{n_2-i}-\begin{bmatrix}n_1\\ i\end{bmatrix}\gamma^i(1-\gamma)^{n_1-i}\right]\ . \tag{2.9}$$

The second term on the left-hand side of (2.9) will typically be negative, since $\delta\geqq\gamma$ and $n_1\geqq n_2$. The left-hand side represents the expected cost change due to using the screening alternative and to using less unacceptable when using the screening alternative as a fraction of the cost of not having enough acceptable units, and the right-hand side represents the reduction due to screening in the probability of not having enough acceptable units.

It does not appear possible to carry out the minimization of $\mathrm{E}\,C(a_1)$ and $\mathrm{E}\,C(a_2)$ in closed form. Some simplification is possible if l, the number of acceptable units needed, equals 1. In this case, $\mathrm{E}\,C(a_1)=c_3(1-\gamma)^{n_1}+c_4n_1(1-\gamma)$ and minimization with respect to n_1 leads to choosing one of the two integer values of n_1 closest to

$$n_1=\ln\left[\frac{-c_4(1-\gamma)}{c_3\ln(1-\gamma)}\right]\Big/\ln(1-\gamma)\ .$$

In a similar way, minimization over n_2 can be carried out for $\mathrm{E}\,C(a_2)$ given a value of w, but minimization over w has to be done numerically.

3. Case where all parameters are unknown

We will now examine the more realistic situation where the parameters of the bivariate distribution in (1.1) are unknown. The discussion, however, will not be as complete as that in Section 2 because there are severe computational problems in this case, and some approximations are needed.

We assume that a training set of k bivariate observations $z_i^t=(x_i, y_i)$, $i=1,\cdots,k$, is available. The sample observations can be summarized by the mean vector $\bar{z}$ and the matrix v, where

$$\text{(3.1)}\qquad \bar{z}=\begin{bmatrix}\bar{x}\\ \bar{y}\end{bmatrix}\qquad\text{and}\qquad v=\sum_{i=1}^{k}(z_i-\bar{z})(z_i-\bar{z})^t\ .$$

Letting

$$\text{(3.2)}\qquad \mu=\begin{bmatrix}\mu_x\\ \mu_y\end{bmatrix}\qquad\text{and}\qquad \Sigma=\begin{bmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{bmatrix}\ ,$$

we can combine a prior distribution for μ and Σ^{-1}, here taken to be diffuse and proportional to $|\Sigma^{-1}|^{-3/2}$, with the likelihood arising from the training set to get the posterior distribution for μ and Σ^{-1},

$$\text{(3.3)}\qquad p(\mu,\Sigma^{-1})\propto|k\Sigma^{-1}|^{1/2}\exp\left[-\frac{k}{2}(\mu-\bar{z})^t\Sigma^{-1}(\mu-\bar{z})\right]$$
$$\times|\Sigma^{-1}|^{(k-1-3)/2}\exp\left[-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}v)\right]\ .$$

Using this posterior distribution we now examine the two cases already discussed in Section 2.

3.1. *Direct measurement of the performance variable is possible*

Equations (2.3) and (2.4) give the expected costs of measuring Y directly and of using the screening variable before measuring Y,

$$\text{(3.4)} \qquad \mathrm{E}\, C(a_1 \mid \mu, \Sigma^{-1}) = c_2 l/\gamma \ ,$$

and

$$\text{(3.5)} \qquad \mathrm{E}\, C(a_2 \mid \mu, \Sigma^{-1}) = c_2 l/\delta + c_1 l/(\delta\beta) \ ,$$

where our notation explicitly shows the dependence of the conditional expected cost figures on the unknown parameters μ and Σ^{-1}. To obtain (unconditional) expected cost figures, μ and Σ^{-1} need to be integrated out of (3.4) and (3.5) leading to

$$\text{(3.6)} \qquad \mathrm{E}\, C(a_1) = c_2 l\, \mathrm{E}\,[\gamma^{-1}] \ ,$$

and

$$\text{(3.7)} \qquad \mathrm{E}\, C(a_2) = c_2 l\, \mathrm{E}\,[\delta^{-1}] + c_1 l\, \mathrm{E}\,[(\delta\beta)^{-1}] \ ,$$

where the expectations on the right-hand sides are with respect to μ and Σ^{-1} whose density is given in (3.3). I am not aware of closed-form expressions for $\mathrm{E}\,[\gamma^{-1}]$, $\mathrm{E}\,[\delta^{-1}]$, and $\mathrm{E}\,[(\delta\beta)^{-1}]$ so that these expectations have to be found using numerical integration techniques. The integration is fairly straightforward for $\mathrm{E}\,[\gamma^{-1}]$ since only two parameters, μ_y and σ_y, are involved, but when finding $\mathrm{E}\,[\delta^{-1}]$ and $\mathrm{E}\,[(\delta\beta)^{-1}]$ all five parameters are involved leading to considerable computational difficulties in the evaluation of $\mathrm{E}\, C(a_2)$ in (3.7). An additional numerical difficulty arises with respect to $\mathrm{E}\, C(a_2)$ in that it depends on w, the cutoff value for the screening variable for which an optimal value has to be found.

3.2. *Direct measurement of the performance variable is not possible*

In this case further numerical difficulties arise. Given μ and Σ^{-1}, future observations $Z_i = (X_i, Y_i)^t$ are independent and the conditional predictive distribution for $Z_1, \cdots, Z_n$ is the product of n bivariate normal distributions

$$\text{(3.8)} \qquad p(\underset{\sim}{Z} \mid \mu, \Sigma^{-1}) = \prod_{i=1}^{n} (2\pi)^{-1} |\Sigma^{-1}|^{1/2} \exp\left[-\frac{1}{2}(Z_i - \mu)^t \Sigma^{-1} (Z_i - \mu)\right] \ ,$$

where $\underset{\sim}{Z} = (Z_1, \cdots, Z_n)^t$, and the symbol "$\sim$" is used under a quantity representing n future observables.

To obtain the (unconditional) predictive distribution $p(\underset{\sim}{Z})$ for $\underset{\sim}{Z}$, μ and Σ^{-1} need to be integrated out of (3.8) using (3.3),

$$p(\underset{\sim}{Z})=\int\int p(\underset{\sim}{Z}\mid \mu, \Sigma^{-1})p(\mu, \Sigma^{-1})d\mu d\Sigma^{-1} . \tag{3.9}$$

This last step introduces dependence among the n future observations $Z_1, \cdots, Z_n$. After some manipulations, (3.9) reduces to

$$p(\underset{\sim}{Z})\propto|I_2+v^{-1}(\underset{\sim}{Z}-\bar{z}^t 1_n)^t A^{-1}(\underset{\sim}{Z}-\bar{z}^t 1_n)|^{-(\nu+n+2-1)/2} , \tag{3.10}$$

where $\nu=k-2$, I_2 is the the (2×2) identity matrix, $\underset{\sim}{Z}$ is of order $(n\times 2)$, $\bar{z}$ is given in (3.1), 1_n is an $(n\times 1)$ vector of ones, and A is an $(n\times n)$ matrix,

$$A=\frac{1}{k}\begin{bmatrix} k+1 & 1 \cdots & 1 \\ 1 & k+1 \cdots & 1 \\ 1 & 1 \cdots & k+1 \end{bmatrix} . \tag{3.11}$$

The distribution in (3.10) is a matric-variate t distribution with ν degrees of freedom which is discussed, e.g., by Dickey [2], and Box and Tiao ([1], § 8.4). Using results given in Box and Tiao, the marginal predictive distribution for n future values of the performance variable, $\underset{\sim}{Y}=(Y_1, \cdots, Y_n)$, is a multivariate t distribution,

$$p(\underset{\sim}{Y})\propto[1+v_y^{-1}(\underset{\sim}{Y}-\bar{y}1_n)^t A^{-1}(\underset{\sim}{Y}-\bar{y}1_n)]^{-(\nu+n)/2} , \tag{3.12}$$

where $\bar{y}$ is given in (3.1), $v_y^{-1}=\sum_{i=1}^{k}(y_i-\bar{y})^2$, and $\underset{\sim}{Y}$ is an $(n\times 1)$ vector. A similar result holds for the correlated variable, X.

In order to derive results similar to the ones in Section 2.2, probabilities need to be computed from (3.10) and (3.12). To illustrate the kind of probability calculations needed, consider the model leading to equation (2.6). One of the probabilities needed there is the probability that i units out of n_1 have $Y\leqq u$, in particular that the first i units have $Y\leqq u$ and the remaining (n_1-i) units have $Y\geqq u$,

$$p=\Pr\left[\left[\bigcap_{j=1}^{i}(Y_j\leqq u)\right]\left[\bigcap_{j=i+1}^{n_1}(Y_j\geqq u)\right]\right] . \tag{3.13}$$

When all parameters are known, there is predictive independence among the n_1 future observations, and $p=\gamma^i(1-\gamma)^{n_1-i}$, where $\gamma=\Pr(Y_1\leqq u)$, say. When the parameters are unknown, the n_1 future observations are predictively dependent and computation of p from (3.12) is considerably more difficult than in the case when all parameters are known. At this point, it seems that the computations needed for an expected cost figure comparable to that in (2.6) are prohibitive, and an approximate solution has to be used.

When dealing with the case where all parameters are known, the predictive distribution is (i) based on the normal distribution and (ii) characterized by independence of future observations. When all parameters are unknown, the predictive distribution is (i) based on the t-distribution, and (ii) characterized by dependence of future observations. The computational difficulties mentioned in the last paragraph are due to the dependence among future observations. As an approximation we propose to ignore this dependence and take probabilities for future observations to be independent and based on a bivariate t-distribution that would result if there were to be only one future observation Z_1,

$$p(Z_1) \propto \left[1+(Z_1-\bar{z})^t\left[\frac{k+1}{k}v\right]^{-1}(Z_1-\bar{z})\right]^{-(\nu+2)/2}. \tag{3.14}$$

When this approximation is used, the probabilities β, γ, and δ given in (1.2) are computed from (3.14) and, with this change, the results in both Sections 2.1 and 2.2 can be used as an approximation.

This approximation may not be adequate if the size of the training set, k, is small. In this case the correlations between X_i and X_j, and Y_i and Y_j, $i \neq j$, which equal $(k+1)^{-1}$, are appreciable.

4. Numerical illustrations

4.1. *Direct measurement of the performance variable is possible*

In this section I will first present some numerical results for the case where all parameters are known. Without loss of generality, I take the means and variances of the bivariate distribution in (1.1) to be 0 and 1, respectively. In Table 1 are given some results for the cases where an item is acceptable if $Y \leqq u_1 = -1$ and $Y \leqq u_2 = 1$, and $l = 1$, i.e., one acceptable item is needed. Table 1 associates with selected values of the correlation ρ between X and Y values of c_1/c_2 such that there is indifference between the screening and no-screening alternatives. In addition, the optimal value of w is given which is needed for the screening alternative, and so are $\beta = \Pr(X \leqq w)$ and $\delta = \Pr(Y \leqq$

Table 1. Indifference values of c_1/c_2 corresponding to selected values of ρ ($u = 1$ or -1)

	$u_1=-1$				$u_2=1$			
ρ	c_1/c_2	w	β	δ	c_1/c_2	w	β	δ
.6	.38	−.33	.37	.32	.072	.33	.63	.94
.7	.45	−.41	.34	.33	.086	.41	.66	.95
.8	.54	−.50	.31	.44	.10	.51	.69	.97
.9	.66	−.62	.27	.54	.12	.63	.73	.98
.95	.71	−.72	.24	.63	.13	.73	.77	.99
.99	.78	−.87	.19	.81	.15	.87	.81	.995

$u|X\leqq w)$. For example, when $u=-1$ and $\rho=.9$ the indifference cost ratio c_1/c_2 is .66. Furthermore, the optimal value of $w=-.62$ at this indifference point so that a proportion $\beta=.27$ of items is deemed acceptable on the basis of the screening variable X and the probability of an item being acceptable is increased from $\gamma=\Pr(Y\leqq u=-1)=.16$ to $\delta=.54$. If at $\rho=.9$ the actual cost ratio $c_1/c_2\leqq.66$, the screening alternative is optimal. The optimal value of w, of course, depends on the value of c_1/c_2.

Let us next examine an example where the parameters have to be estimated. Suppose the training set consists of $k=15$ observations, that the sample mean vector $\bar{z}=0$ and that the sample covariance matrix v/k has variances 1 and correlation $r=.8$. An item is acceptable if $Y\leqq u=-1$, and $l=1$. If furthermore, the cost values are $c_1=5.7$ and $c_2=100$, then the expected cost under the no-screening alternative a_1 is 669 and that under the screening alternative a_2 is 257 with optimal value $w=-1.10$. Thus, the screening alternative is preferred in this case. If $k\to\infty$, i.e., all parameters are known, and all other values are unchanged, then the expected cost under a_1 is 630 and that under a_2 is 217 with optimal value $w=-1.17$ so that again the screening alternative is preferred.

4.2. *Direct measurement of the performance variable is not possible*

In this section we illustrate the results of Section 2.2. Let us reconsider a problem given in Owen, Li, and Chou [6]. The performance variable Y represents temperature inside an oven that is used to bake corn chips. The correlated variable X is temperature of the outside metal, and it is much easier to measure than Y. It is known that (X, Y) has a normal distribution with parameters $\mu_x=80°$, $\sigma_x=2°$, $\mu_y=204°$, $\sigma_y=5°$, and $\rho=.9$. An oven run is acceptable if $Y\geqq 200°$. Since direct measurement of Y is difficult, it may be advisable to deem oven runs to be acceptable if the correlated variable X exceeds a quantity x_w. It is necessary that at least $l=9$ oven runs are acceptable on the basis of Y.

Let us take the cost of an X-measurement to be $c_1=1$ without loss of generality. The cost of having less than l acceptable bake runs is $c_3=10$, and each item erroneously deemed acceptable leads to a cost of $c_4=1.2$. Other cost relations are possible, and each new situation will require a careful analysis of these cost relations. We have to determine now if it is advisable to screen oven runs or not. We will first examine the screening alternative for which expected cost are given in (2.7). It is necessary to find the optimal cutoff point for the correlated variable X and the optimal number of oven runs to be deemed acceptable to minimize expected cost $\mathrm{E}\,C(a_2)$.

The problem as set out in the first paragraph of this section is not yet in the form required for the results in Section 2.2, where we can assume without loss of generality that the two variables X and Y are standardized. Denote these standardized variables by X_s and Y_s. Then, in the notation of Section 2.2, we have that $\gamma=\Pr(Y_s\leqq u=.8)=.788$. Given γ, the cost values, and $\rho=.9$, a simple computer program will find the values of n_2 and w, the cutoff point for the correlated variable, minimizing the expected cost of (2.7). The optimal values are $n_2=10$ and $w=.71$ so that items are deemed acceptable if $X_s\leqq .71$.

In terms of the oven run example, oven runs will be deemed acceptable if the temperature of outside metal $X\geqq \mu_x-(.71)\sigma_x=78.6°$. This means the long run probability that $Y\geqq 200°$ is increased from $\gamma=.788$ before screening to $\delta=.946$ after screening. Given the cost values $c_1=1$, $c_3=10$, and $c_4=1.2$, it is optimal to accept or reject as many oven runs as are necessary until 10 runs are deemed acceptable on the basis of $X\geqq 78.6°$. The probability that at least 9 out of these 10 oven runs are acceptable on the basis of Y is .901 after screening, and .341 before screening. These results are comparable to the ones given in Owen et al. [6] where a non-decision theoretic approach is taken in which the cutoff point for the correlated variable X is to be found, and "we wish the probability to be .90 that at least 9 of the next 10 accepted bake runs equal or exceed the required 200°, and we would like to achieve this probability by accepting or rejecting as many oven runs as needed based on X".

The expected cost for this optimal screening procedure is $\mathrm{E}\,C(a_2)=(1)(10)/(.761)+(10)(1-.901)+(1.2)(10)(1-.946)=14.8$. Without screening, the expected cost of the corresponding procedure, i.e., to accept the first 10 oven runs, is $\mathrm{E}\,C(a_1)=(10)(1-.341)+(1.2)(10)(1-.788)=9.1$ so that no-screening would be preferable to screening. If the no-screening alternative is selected, expected cost can be reduced further by minimizing (2.4) over n_1, the optimal result being that the first $n_1=15$ oven runs should be accepted without screening so that $\mathrm{E}\,C(a_1)=(10)(1-.975)+(1.2)(15)(1-.788)=4.1$.

UNIVERSITY OF TORONTO

REFERENCES

[1] Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, Reading, Mass., Addison-Wesley.

[2] Dickey, J. M. (1967). Matric-variate generalization of the multivariate t-distribution and the inverted multivariate t-distribution, *Ann. Math. Statist.*, 38, 511-518.

[3] Hald, A. (1981). *Statistical Theory of Sampling Inspection by Attributes*, Academic Press, London.

[4] Madsen, R. W. (1982). A selection procedure using a screening variate, *Technometrics*,

20, 301-306.
[5] Owen, D. B. and Boddie, J. W. (1976). A screening method for increasing acceptable product with some parameters unknown, *Technometrics*, **18**, 195-199.
[6] Owen, D. B., Li, L. and Chou, Y-M. (1981). Prediction intervals for screening using a measured correlated variable, *Technometrics*, **23**, 165-170.
[7] Owen, D. B. and Su, Y-L.H. (1977). Screening based on normal variables, *Technometrics*, **19**, 65-68.