PREDICTION USING POOLED REGRESSIONS

Ivor Francis, Cornell University and Samprit Chatterjee, New York University

Introduction

An applicant for admission to college is usually required to take a series of tests in support of his application. In many instances the applicant's scores are then weighted and added to form one composite score, and this score will determine the success or failure of his application. The reasoning behind this procedure is that this composite score is presumed to relate to, and therefore can be used to predict, ability to succeed in college.

A psychologist might argue, however, that potential applicants form not one population, but rather they should be regarded as comprising several populations each having distinct psychological and cultural properties. This being the case it would be better, as far as uniform accuracy of prediction is concerned, to use a different set of weights for different populations rather than the same weights for everybody.

We shall suppose that it is not possible to measure directly those psychological and cultural factors which would place a candidate into one of these populations. The question is how should we estimate his ability to succeed? Several possibilities suggest themselves. Firstly we can use the same weights for everybody, as we are supposing is currently done. But if we knew the best weights for each population perhaps we could use the observed scores for an individual to first classify him into one of the populations and then use the weights that are best for that population. Alternatively, if we know the probability π_i that the new individual came from the i-th population, $i = 1, \ldots, k$, perhaps we should weight the separate population estimates with these probabilities.

The problem, then, is the following: from each of k populations we have independent sample observations on the variables $(y, x_j, j = 1, \ldots, p)$, the size of the sample in the i^{th} population being n_i, $i = 1, \ldots, k$. In addition we have observations on (x_j) for an individual known to have come from one of the k populations, but from which one is not known. We wish to estimate the value of y for this new individual.

First we shall present a maximum likelihood solution and then a discussion of methods of pooling the regressions of the separate populations.

The Maximum Likelihood Approach

Suppose we have k populations, with n_i observations from the i-th population, where $\Sigma n_i = n$. Let the observations on (x_j) for the new observation be denoted by $\underline{X}^*$, a $(p \times 1)$ vector. The Y-value corresponding to $\underline{X}^*$ will be denoted by Y^*. Assume y is a scalar random variable and $\underline{x}$ is a $(p \times 1)$ vector random variable with the following distribution in the i-th population

$$y \sim N(\underline{x}^t \beta_i, \sigma^2)$$

$$\underline{x} \sim N(\underline{\mu}_i, \underline{\Sigma}), \quad i = 1, \ldots, k$$

where $\underline{\beta}_i$ and $\underline{\mu}_i$ are $(p \times 1)$ parameter vectors, and $\underline{\Sigma}$ is a $(p \times p)$ matrix, $\underline{\Sigma}$ being assumed the same in all populations.

Let $\underline{Y}_i$ denote the $(n_i \times 1)$ vector of observations on y in the i-th population, and let $\underline{X}_i$ denote the $(p \times n_i)$ matrix of observations on $\underline{x}$ in the i-th population.

In the manner of Hartley and Rao (1968) we introduce an indicator vector $\underline{\theta}$, where

$$\underline{\theta} = (\theta_1, \ldots, \theta_k) \quad \text{such that } \Sigma\, \theta_i = 1$$

where $\theta_i = 1$ if the new individual came from the i-th population and $\theta_i = 0$ otherwise.

The likelihood function for $\underline{\beta}_i$, $\underline{\mu}_i$, $\underline{\Sigma}$, σ^2, $\underline{\theta}$, and Y^* is

$$L(\underline{\beta}_i, \underline{\mu}_i, \underline{\Sigma}, \sigma^2, \underline{\theta}, Y^*) = L_1 \times L_2 \times L_3 \times L_4$$

where

$$L_1 = \frac{C_1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2} \sum_1^k (\underline{Y}_i - \underline{X}_i^t \underline{\beta}_i)^t (\underline{Y}_i - \underline{X}_i^t \underline{\beta}_i)\right]$$

$$L_2 = \frac{C_2}{\sigma} \exp\left[-\frac{1}{2\sigma^2} (Y^* - \sum_i^k \theta_i \underline{X}^{*t} \underline{\beta}_i)^2\right]$$

$$L_3 = \frac{C_3}{|\underline{\Sigma}|^{n/2}}$$

$$\cdot \exp\left[-\frac{1}{2} \sum_{i=1}^k \sum_{j=1}^{n_i} (\underline{X}_{ij} - \underline{\mu}_i)^t \underline{\Sigma}^{-1} (\underline{X}_{ij} - \underline{\mu}_i)\right]$$

$$L_4 = \frac{C_4}{|\underline{\Sigma}|^{1/2}} \exp\left[\frac{1}{2} \sum_{i=1}^k \theta_i (\underline{X}^* - \underline{\mu}_i)^t \underline{\Sigma}^{-1} (\underline{X}^* - \underline{\mu}_i)\right]$$

Taking logs:

$$\log L = P + Q + \text{constant}$$

where

$$P = -\frac{1}{2\sigma^2}\left[\sum_1^k (\underline{Y}_i - \underline{X}_i^t \underline{\beta}_i)^t (\underline{Y}_i - \underline{X}_i^t \underline{\beta}_i)\right.$$

$$\left. + (Y^* - \sum_1^k \theta_i \underline{X}^{*t} \underline{\beta}_i)^2\right] - \frac{n+1}{2} \log \sigma^2$$

$$Q = -\frac{n+1}{2}\log|\underline{\Sigma}|$$

$$-\frac{1}{2}\Big[\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\underline{X}_{ij}-\underline{\mu}_i)^t\,\underline{\Sigma}^{-1}(\underline{X}_{ij}-\underline{\mu}_i)$$

$$+\sum_{i=1}^{k}\theta_i(\underline{X}^*-\underline{\mu}_i)^t\,\underline{\Sigma}^{-1}(\underline{X}^*-\underline{\mu}_i)\Big]$$

In the next two sections we find the conditional maximum likelihood estimates (MLE) given $\theta_s = 1$. Given $\theta_s = 1$ denote the conditional maximum likelihood estimates of the parameters by $\hat{\underline{\beta}}_i$, $\hat{\underline{\mu}}_i$, $\hat{\underline{\Sigma}}$, $\hat{\sigma}^2$, $\hat{Y}^*$.

Conditional MLE of Y^*, σ^2 and $\underline{\beta}_i$

$\frac{\partial \log L}{\partial \underline{\beta}_s} = 0$ gives

$$(\underline{X}_s\underline{X}_s^t + \underline{X}^*\underline{X}^{*t})\,\hat{\underline{\beta}}_s = \underline{X}_s\underline{Y}_s + \underline{X}^*\hat{Y}^*. \qquad (1)$$

$\frac{\partial \log L}{\partial \underline{\beta}_i} = 0$ for $i \neq s$ gives

$$\underline{X}_i\underline{X}_i^t\hat{\underline{\beta}}_i = \underline{X}_i\underline{Y}_i. \qquad (2)$$

$\frac{\partial \log L}{\partial \sigma^2} = 0$ gives

$$\hat{\sigma}^2 = \frac{1}{n+1}\Big[\Sigma(\underline{Y}_i-\underline{X}_i^t\hat{\underline{\beta}}_i)^t(\underline{Y}_i-\underline{X}_i^t\hat{\underline{\beta}}_i) + (\hat{Y}^*-\underline{X}^{*t}\hat{\underline{\beta}}_s)^2\Big]. \qquad (3)$$

$\frac{\partial \log L}{\partial Y^*} = 0$ gives

$$\hat{Y}^* = \underline{X}^{*t}\hat{\underline{\beta}}_s. \qquad (4)$$

Substituting (4) in (1) we get

$$\underline{X}_s\underline{X}_s^t\hat{\underline{\beta}}_s = \underline{X}_s\underline{Y}_s.$$

Thus $\hat{\underline{\beta}}_i = (\underline{X}_i\underline{X}_i^t)^{-1}(\underline{X}_i\underline{Y}_i)$ for all $i = 1, \ldots, k$. (5)

Substituting (4) in (3) we get

$$\hat{\sigma}^2 = \frac{1}{n+1}\sum_i(\underline{Y}_i-\underline{X}_i^t\hat{\underline{\beta}}_i)^t(\underline{Y}_i-\underline{X}_i^t\hat{\underline{\beta}}_i). \qquad (6)$$

Conditional MLE of $\underline{\mu}_i$ and $\underline{\Sigma}$

In the expression for the likelihood above only the terms L_3 and L_4 contain $\underline{\mu}_i$ and $\underline{\Sigma}$. Thus we can find the MLE's of $\underline{\mu}_i$ and $\underline{\Sigma}$ by maximizing $L_3 \times L_4$. But since $\theta_s = 1$ we simply have a situation where there are n_i observations from $N(\underline{\mu}_i, \underline{\Sigma})$, $i \neq s$, and $n_s + 1$ from $N(\underline{\mu}_s, \underline{\Sigma})$.

Thus the conditional MLE's of $\underline{\mu}_i$ and $\underline{\Sigma}$ can be obtained from standard multivariate theory (e.g. Anderson, (1958) p. 248). These are

$$\hat{\underline{\Sigma}} = \frac{1}{n+1}\underline{A} = \frac{1}{n+1}\sum_{i=1}^{k}\underline{A}_i$$

where

$$\underline{A}_i = \sum_{j=1}^{n_i}(\underline{X}_{ij}-\bar{\underline{X}}_i)(\underline{X}_{ij}-\bar{\underline{X}}_i)^t, \text{ for } i \neq s$$

and

$$\underline{A}_s = \sum_{j=1}^{n_s}(\underline{X}_{sj}-\bar{\underline{X}}_s)(\underline{X}_{sj}-\bar{\underline{X}}_s)^t + (\underline{X}^*-\bar{\underline{X}}_s)(\underline{X}^*-\bar{\underline{X}}_s)^t$$

and finally

$$\hat{\underline{\mu}}_i = \bar{\underline{X}}_i.$$

Estimating θ

Substituting these conditional MLE's into the expression for the likelihood above we see that max $(L_1 \times L_2)$ is independent of s. The maximum value of $L_3 \times L_4$ is

$$\frac{C}{|\hat{\underline{\Sigma}}|^{\frac{n+1}{2}}}$$

where $\hat{\underline{\Sigma}}$ is a function of s, and so we will denote it by $\hat{\underline{\Sigma}}^{(s)}$. Thus the MLE of s is that value which minimizes $|\hat{\underline{\Sigma}}^{(s)}|$ or equivalently $|\underline{A}^{(s)}|$.

Discussion

1. The regression is estimated from $n\ (= \Sigma n_i)$ observations on which y and $\underline{x}$ are available.
2. Classification is based on all the $n + 1$ observations on which $\underline{x}$ is available.
3. The procedure operates as follows:
   a) Compute
   $$|\underline{A}^{(s)}| = \Big|\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\underline{X}_{ij}-\bar{\underline{X}}_i)(\underline{X}_{ij}-\bar{\underline{X}}_i)^t + (\underline{X}^*-\bar{\underline{X}}_s)(\underline{X}^*-\bar{\underline{X}}_s)^t\Big|$$
   for all $s = 1, \ldots, k$.
   b) If r is the value of s for which $|\underline{A}^{(s)}|$ is minimum, assign $\underline{X}^*$ to the r-th population.

c) Estimate Y^* by

$$\hat{Y}^* = \underline{X}^{*t}\,\hat{\beta}_r$$

where

$$\hat{\beta} = (\underline{X}_r\underline{X}_r^t)^{-1}(\underline{X}_r\underline{Y}_r)$$

Pooled Regressions

We assume the usual regression model in each population, and we let the true regression in the i-th population be denoted by R_i, the estimated (least squares) regression by $\hat{R}_i$, the variance of the dependent variable by Π_i^2, and the standard error of estimate by σ_i. Thus

$$y_{ij} = R_i + \epsilon_{ij} \qquad j = 1, \ldots, n_i$$

where $\quad \epsilon_{ij} \sim N(0, \Pi_i^2)$

For the special case of simple linear regression,

$$R_i = \alpha_i + \beta_i x$$

$$\hat{R}_i = \hat{\alpha}_i + \hat{\beta}_i x$$

and
$$\sigma_i^2 = \Pi_i^2\left(\frac{1}{n_i} + \frac{(x - \bar{x}_i)^2}{\sum_j (x_{ij} + \bar{x}_i)^2}\right)$$

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the usual least squares estimates.

The population of origin of the new individual is unknown, but we suppose that the probability he came from the i-th population is π_i, $i = 1, \ldots, k$.

We propose to investigate several methods of pooling the regressions. Which method is best will depend on the criterion used to evaluate the predicting ability of these methods. We propose to use the mean square error (MSE) since this is attractive in itself, and can be looked at as comprising the variance of the estimator plus the square of the bias. Thus minimizing the MSE is controlling the size of both these parameters.

Expectation Estimator (EE)

The simplest pooled estimator is

$$EE = \Sigma\, \pi_i \hat{R}_i \tag{7}$$

The expected value of this is $\Sigma\, \pi_i R_i$. For convenience we shall denote this by $\bar{R}$. Thus

$$E(EE) = \Sigma\, \pi_i R_i = \bar{R} \tag{8}$$

The expected MSE in predicting the mean value of y is

$$MSE(EE) = \sum_j \pi_j\left[\sum_i \pi_i^2\sigma_i^2 + (\bar{R} - R_j)^2\right]$$

$$= \Sigma\, \pi_i (R_i - \bar{R})^2 + \Sigma\, \pi_i^2\sigma_i^2 \tag{9}$$

$$= \Sigma\, \pi_i R_i^2 - \bar{R}^2 + \Sigma\, \pi_i^2\sigma_i^2$$

A Weighted Expectation Estimator (WEE)

Since we are using minimum MSE as our criterion of goodness we are not restricted to using estimates whose expected bias is zero. Thus we consider

$$WEE = a\, \Sigma\, \pi_i \hat{R}_i \tag{10}$$

where a is a constant to be determined. To find the value of a which minimizes the MSE we set equal to zero the derivative of

$$MSE = a^2 \sum_i \pi_i^2\sigma_i^2 + \sum_j \pi_j (a\,\bar{R} - R_j)^2$$

Solving for a we get

$$a = \frac{\bar{R}^2}{\Sigma\, \pi_i^2\sigma_i^2 + \bar{R}^2} \tag{11}$$

For this optimum value of a the expected value and the MSE of this estimate are

$$E(WEE) = \frac{\bar{R}^2}{\Sigma\, \pi_i^2\sigma_i^2 + \bar{R}^2}\cdot \bar{R} \tag{12}$$

$$MSE(WEE) = \Sigma\, \pi_i (R_i - \bar{R})^2 + \frac{\bar{R}^2}{\Sigma\, \pi_i^2\sigma_i^2 + \bar{R}^2}\cdot \Sigma\, \pi_i^2\sigma_i^2 \tag{13}$$

Optimal Weighted Estimator (OWE)

The two estimators considered so far do not take into consideration the possibility that there may be different sample sizes and therefore differing amounts of information available in the different populations. So let us consider an estimate

$$OWE = \Sigma\, a_i \hat{R}_i \tag{14}$$

where the a_i, $i = 1, \ldots, k$ are constants chosen to minimize the expected MSE:

$$MSE = \sum_i a_i^2 \pi_i^2 + \sum_j \pi_j \left(\sum_i a_i R_i - R_j\right)^2$$

Differentiating this with respect to a_i, $i = 1, \ldots, k$, and setting the derivatives equal to zero we solve for the a_i:

$$a_i = \frac{\frac{R_i}{\sigma_i^2}\bar{R}}{1 + \Sigma \frac{R_i^2}{\sigma_i^2}} \qquad (15)$$

The estimator, the expected value, and the MSE using these optimal a_i's are

$$OWE = \frac{\bar{R}\,\Sigma \frac{R_i \hat{R}_i}{\sigma_i^2}}{1 + \Sigma \frac{R_i^2}{\sigma_i^2}} \qquad (16)$$

$$E(OWE) = \frac{\Sigma \frac{R_i^2}{\sigma_i^2}}{1 + \Sigma \frac{R_i^2}{\sigma_i^2}} \cdot \bar{R} \qquad (17)$$

$$MSE(OWE) = \Sigma\,\pi_i(R_i - \bar{R})^2 + \frac{\bar{R}^2}{1 + \Sigma \frac{R_i^2}{\sigma_i^2}} \qquad (18)$$

Constrained Weighted Estimator (CWE)

For those who would like the coefficients to sum to unity we offer:

$$CWE = \Sigma\, a_i \hat{R}_i \quad \text{where } \Sigma\, a_i = 1. \qquad (19)$$

To find the optimum a_i we differentiate with respect to a_i, $i = 1, \ldots, k$

$$\varphi = \Sigma\, a_i^2 \sigma_i^2 + \sum_j \pi_j (\sum_i a_i R_i - R_j)^2 + \lambda(\sum_i a_i - 1)$$

and set the derivatives equal to zero. To solve these equations we first multiply the i-th equation by $\frac{1}{\sigma_i^2}$ and add over i. Then multiply the i-th equation by $\frac{R_i}{\sigma_i^2}$ and add over i. These two equations can be solved for $\Sigma\, a_i R_i$ and λ, and thus we can solve for a_i:

$$a_i = \frac{1}{D}\left[\frac{1}{\sigma_i^2}(1 + \Sigma \frac{R_i^2}{\sigma_i^2} - \Sigma \frac{R_i}{\sigma_i^2}\bar{R}) + \frac{R_i}{\sigma_i^2}(\Sigma \frac{1}{\sigma_i^2}\bar{R} - \Sigma \frac{R_i}{\sigma_i^2})\right] \qquad (20)$$

where $D = (1 + \Sigma \frac{R_i^2}{\sigma_i^2})\,\Sigma \frac{1}{\sigma_i^2} - (\Sigma \frac{R_i}{\sigma_i^2})^2$ which is positive by Schwarz' inequality.

Thus the expected value and the MSE of the CWE are

$$E(CWE) = \bar{R} + \frac{1}{D}(\Sigma \frac{R_i}{\sigma_i^2} - \bar{R}\,\Sigma \frac{1}{\sigma_i^2}) \qquad (21)$$

$$MSE(CWE) = \Sigma\,\pi_i(R_i - \bar{R})^2 + \frac{\bar{R}^2}{1 + \frac{R_i^2}{\sigma_i^2}} + \frac{(1 + \Sigma \frac{R_i^2}{\sigma_i^2} - \bar{R}\,\Sigma \frac{R_i}{\sigma_i^2})^2}{D(1 + \Sigma \frac{R_i^2}{\sigma_i^2})} \qquad (22)$$

Lumped Estimator (LE)

Up to this point we have had to estimate the regression in each population, and then make our prediction using some set of weights. An alternative method is to simply lump all observations together as if they were from one large population, estimate the regression in that large population, and use this regression for prediction. We shall call this the Lumped Estimator (LE).

One situation where this would be a desirable alternative is where the degrees of freedom in the separate populations are rather small for the separate estimations.

Comparing the ability to predict of the LE with that of earlier weighted estimators becomes very involved. We present here a comparison for the case of simple linear regression, and we make one or two simplifying assumptions. We shall use the following notation:

$$LE = \hat{\alpha} + \hat{\beta}\,x \qquad (23)$$

$$\hat{\beta} = \frac{\Sigma\,\Sigma\,(y_{ij} - \bar{y})(x_{ij} - \bar{x})}{D_x}$$

$$D_x = \Sigma\,\Sigma\,(x_{ij} - \bar{x})^2 = n(W + B) \qquad (24)$$

$$W = \frac{1}{n}\Sigma\,\Sigma\,(x_{ij} - \bar{x}_i)^2$$

$$B = \Sigma\, p_i(\bar{x}_i - \bar{x})^2$$

$$p_i = \frac{n_i}{n} \qquad (25)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}$$

$$\bar{y} = \Sigma\, p_i \bar{y}_i$$

$$\bar{x} = \Sigma\, p_i \bar{x}_i$$

Now it is easily shown that

$$\hat{\beta} = \frac{\Sigma\, D_i \hat{\beta}_i + \Sigma\, n_i(\bar{x}_i - \bar{x})\bar{y}_i}{D_x}$$

where

$$D_i = \Sigma(x_{ij} - \bar{x}_i)^2$$

For simplification we shall assume that it is approximately true that

$$\frac{D_i}{n_i} = W \quad \text{for all } i = 1, \ldots, k.$$

Note that this is assuming that the "within-sample" variance of the x_{ij} is approximately the same in all populations. In this case

$$\hat{\beta} = \frac{W}{W+B} \Sigma p_i\hat{\beta}_i + \frac{B}{W+B} \hat{\beta}^*$$

where $$\hat{\beta}^* = \frac{\Sigma p_i(\bar{x}_i - \bar{x})\bar{y}_i}{\Sigma p_i(\bar{x}_i - \bar{x})^2} \tag{26}$$

Note that $\hat{\beta}^*$ is the slope of a weighted regression line fitted to the centroids of the k samples.

Thus the lumped estimator is

$$\begin{aligned} LE &= \bar{y} + (x - \bar{x}) \Sigma p_i\hat{\beta}_i \\ &\quad + \frac{B}{W+B}(x - \bar{x})(\hat{\beta}^* - \Sigma p_i\hat{\beta}_i) \\ &= \Sigma p_i\hat{R}_i + \Sigma p_i\hat{\beta}_i(\bar{x}_i - \bar{x}) \\ &\quad + \frac{B}{W+B}(x - \bar{x})(\hat{\beta}^* - \Sigma p_i\hat{\beta}_i) \end{aligned} \tag{27}$$

The expected value is

$$\begin{aligned} E(LE) &= \Sigma p_iR_i + \Sigma p_i\beta_i(\bar{x}_i - \bar{x}) \\ &\quad + \frac{B}{W+B}(x - \bar{x})(\beta^* - \bar{\beta}) \end{aligned} \tag{28}$$

where $$\bar{\beta} = \Sigma p_i\beta_i \tag{29}$$

Note that β^* may be regarded as the slope of the regression <u>between</u> populations, and $\bar{\beta}$ an average slope <u>within</u> populations.

Now LE is the usual least squares estimator. Let us assume for the sake of simplification that the variance of y in the i-th population is a constant, i.e.

$$\eta_i = \eta, \quad \text{for all } i = 1, \ldots, k. \tag{30}$$

Thus

$$\text{var}(LE) = \frac{\eta^2}{n}\left(1 + \frac{(x - \bar{x})^2}{W+B}\right) \tag{31}$$

Note that if the η_i are not assumed equal the variance of LE can be found from an alternative form of LE, namely

$$\begin{aligned} LE = \Sigma p_i\Big\{ &\left[1 + \frac{(x - \bar{x})(\bar{x}_i - \bar{x})}{W+B}\right]\bar{y}_i \\ &+ \frac{W}{W+B}(x - \bar{x})\hat{\beta}_i\Big\}. \end{aligned}$$

<u>The MSE of the Lumped Estimator</u>

If, as earlier, we let π_i denote the probability that the individual comes from the i-th population we can write down the expected MSE:

$$MSE(LE) = \text{var}(LE) + \Sigma \pi_i(R_i - E(LE))^2$$

where var(LE) and E(LE) are given in equations (31) and (28).

There now arises the problem of comparing this MSE with the MSE of earlier pooled estimators. This is quite difficult, especially if we can assume nothing about the relationship between p_i and π_i. Even if the total sample is random, a complete analysis would discuss the variance of p_i as and estimate of π_i. If the sample is large p_i would approximate π_i. But anyway, in order that we can get a feel for the relative goodness of LE as opposed to EE, the expectation estimator, we shall assume that it is approximately true that

$$p_i = \pi_i \tag{32}$$

In this case we can show that

$$\begin{aligned} MSE(LE) &= \frac{\eta^2}{n}\left(1 + \frac{(x - \bar{x})^2}{W+B}\right) + \Sigma \pi_i(R_i - \bar{R})^2 \\ &\quad + \Big[\Sigma \pi_i\beta_i(\bar{x}_i - \bar{x}) \\ &\quad + \frac{B}{W+B}(x - \bar{x})(\beta^* - \bar{\beta})\Big]^2 \end{aligned} \tag{33}$$

For comparing this with MSE(EE) in equation (9) we note first that σ_i^2 can be written:

$$\sigma_i^2 = \frac{\eta_i^2}{n_i}\left(1 + \frac{(x - \bar{x}_i)^2}{W}\right),$$

and since we are assuming $\eta_i = \eta$ for all i, and that $n_i = n\,\pi_i$,

$$\begin{aligned} \Sigma \pi_i^2\sigma_i^2 &= \frac{\eta^2}{n} \Sigma \pi_i\left(1 + \frac{(x - \bar{x}_i)^2}{W}\right) \\ &= \frac{\eta^2}{n}\left(1 + \frac{(x - \bar{x})^2}{W} + \frac{B}{W}\right) \end{aligned}$$

Thus from equation (9)

$$\begin{aligned} MSE(EE) &= \frac{\eta^2}{n}\left(1 + \frac{(x - \bar{x})^2}{W} + \frac{B}{W}\right) \\ &\quad + \Sigma \pi_i(R_i - \bar{R})^2. \end{aligned} \tag{34}$$

Comparison of the Pooled Estimators

It can be seen from equations (9), (13), (18), and (22) under what circumstances the various pooled estimators would be superior, and by how much. It is clear that OWE is best, but it must be remembered that the optimal weights are functions of the parameters, and therefore have to be estimated. Estimating these weights will obviously increase the MSE of these estimators. Since it seems reasonable that any improvement, if any, would be due to a large extent to an improved estimate of the variance, an investigation of the MSE of the estimated OWE should estimate the R_i's and the σ_i's.

One observation on MSE(OWE). We note that this is small whenever at least one coefficient of variation $\frac{R_i}{\sigma_i}$ is large at the value of x we are estimating for. A small σ_i provides some evidence that the new individual came from the i-th population. And from equation (15), if $\frac{R_i}{\sigma_i^2}$ is large then the weight for the i-th is large.

In an unpublished paper Francis (1969) considers the related problem of estimating the regression in a particular population when sample observations are also available from other populations which are considered somewhat similar. There again, a comparison of the pooled estimators will have to investigate the effect of estimating the optimal weights.

Comparison of EE and LE

By comparing equations (33) and (34) we can say something about EE versus LE. For example, suppose

$$\Sigma\ \pi_i \beta_i (\bar{x}_i - \bar{x}) = 0$$

This would be approximately true, for instance, if all the β_i are roughly equal, or alternatively if all the $\bar{x}_i$ were roughly equal. (Figures 1 and 2 display the concentration ellipses for two examples for which k = 3, where all the β_i are roughly equal, and where the within-sample variances of x would be approximately all equal. The slope of the line AB is $\bar{\beta}$ and the slope of CD is β^*.) In this case the difference between (33) and (34) is

$$\mathrm{MSE(LE)} - \mathrm{MSE(EE)} = \frac{\sigma^2}{n} \left\{ z^2\left[\left(1 + \frac{B}{W}\right)^{-1} - 1\right] - \frac{B}{W}\right\} + \frac{B^2 W}{(W + B)^2}\, z^2(\beta^* - \bar{\beta})^2 \tag{35}$$

where $$z^2 = \frac{(x - \bar{x})^2}{W} .$$

The following observations can be made:

i) Very small B could imply a very large β^*: EE is better (see Figure 2).

ii) When B approaches W in size and when $(\beta^* - \bar{\beta})^2 \gg 0$, then EE is better, except when $x \doteq \bar{x}$ and then:

iii) When B approaches W in size, and $x \doteq \bar{x}$ then LE is best.

iv) When B is extremely large we would surely modify our π_i's, given the x observation on our new individual, to eliminate most populations, unless the original samples were not random. (See section on Posterior Probabilities below.)

Posterior Probabilities

We have been assuming that π_i is known for all i. But as was suggested in observation (iv) above, if these are only prior probabilities, and if the original observations on (x_i) were a random sample from all populations, then these sample observations, together with the observations on (x_i) for the new individual, should be used to modify the π_i's. We shall not pursue this topic in this paper.

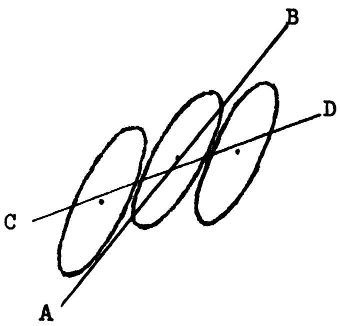

Figure 1

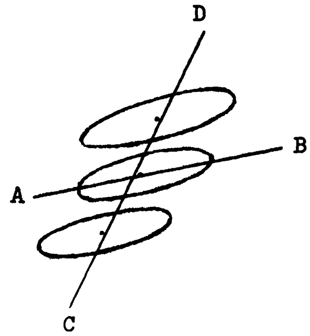

Figure 2

References

Anderson, T. W., Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, Inc., 1958.

Francis, I. S., "On Pooling Regressions," Unpublished, 1969.

Hartley, H. O. and Rao, J. N. K., "Classification and Estimation in Analysis of Variance Problems," Review of the International Statistical Institute, Vol. 36: 2, 1968, pp. 141-147.