PARAMETER ESTIMATION IN CREDIBILITY THEORY

FL. DE VYLDER

University of Louvain, Belgium

ABSTRACT.

The problem of distribution-free parameter estimation in recent credibility theory is discussed in the papers [1], [3] and [4] of the bibliography. Here, we consider a multiclass model with regression assumption. In that case, already treated by Ch. Hachemeister, [3], this author obtains an unsymmetrical matrix as an estimator of a covariance matrix. Of course, for practical use, this matrix is symmetrized in the obvious way. We show that this procedure can be avoided and that a lot of symmetrical unbiased estimators can be obtained at once.

By particularisations to the 1-rank model, we find the estimators given by Buhlmann and Straub, [1], [4].

In the multirank case, a generalization of the minimumvariance principle (minimization of the trace of the covariance matrix) leads to an optimal estimator of the mean regression vector. A noteworthy conclusion of our discussion is that there is no difference at all between the various credibility formulae (the inhomogenous formula, the homogeneous formula, the mean-free formula) if the mean regression vector is estimated optimally.

Finally we show that it must not be hoped to find, in a large set of unbiased estimators of the covariance matrix, one estimator furnishing, always, a semidefinite positive estimate

1. THE MULTICLASS MODEL WITH REGRESSION ASSUMPTION.

1.1. *Description of the model*

We consider the array of observable random variables

$$
\begin{array}{cccccc}
_1X_1 & _2X_1 & \ldots & _jX_1 & \ldots & _kX_1 \\
_1X_2 & _2X_2 & \ldots & _jX_2 & \ldots & _kX_2 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
_1X_s & _2X_s & \ldots & _jX_s & \ldots & _kX_s \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
_1X_t & _2X_t & \ldots & _jX_t & \ldots & _kX_t
\end{array}
$$

In the notation $_jX_s$ $(j = 1, 2, \ldots, k; s = 1, 2, \ldots, t)$, the left index j is the class index, the right index s is the year index. For example $_jX_s$ might be the claim rate of treaty j in year s in a reinsurer's portfolio, but other interpretations are possible. The column $_jX = (_jX_1, _jX_2, \ldots, _jX_t)'$, will be called the *class j*. To $_jX$ is associated the structure variable $_j\Theta$. We abbreviate:

$$
\Theta = (_1\Theta, _2\Theta, \ldots, _k\Theta).
$$

The numbers k (number of classes) and t (number of observation years) are fixed. As variable class-indices we use $i, j = 1, 2, \ldots, k$ and as variable time indices $r, s = 1, 2, \ldots, t$.

Before we specify the assumptions relating the observable and structure variables, we make some general remarks about the matrix notation used throughout the text. A $\frac{n}{m}$ matrix is one with m rows and n columns. Then $\frac{n}{m}$ is the *dimension* of the matrix. Rows, columns, scalars are particular matrices. The dimension $\frac{1}{1}$ is also denoted more simply by 1. Some relations are completed by the dimensions of the displayed matrices. The same symbol (without indices) is used for a matrix and for its elements (with indices). The inferior right index is the row-index. The superior right index is the column-index. *Matrix rules are applied to indices written on the right only.*

The following assumptions are made.

(i)   Independence of classes: $_1X, _2X, \ldots, _kX$ are independent.

(ii)   In each class, irrelevancy of other parameter values than that one of the given class: For each class-index j and function $f(.)$,

$$E(f(_jX)/\Theta) = E(f(_jX)/_j\Theta).$$

(iii) Independence of parameters: $_1\Theta, _2\Theta, \ldots, _k\Theta$ are independent.

(iv)   Equidistribution of the parameters $_1\Theta, _2\Theta, \ldots, _k\Theta$.

(v)   There exist functions $\mu_s(.)$ satisfying

$$E(_jX_s/_j\Theta) = \mu_s(_j\Theta).$$

(The assumption is in the fact that $\mu_s(.)$ does not depend on j.)

(vi)   There exist symmetrical definite positive $\frac{t}{t}$ matrices $_jv$ and a scalar function $\sigma^2(.)$ satisfying

$$\mathrm{COV}(_jX_s, _jX_r/_j\Theta) = \sigma^2(_j\Theta) \, _jv_r^s.$$

(vii) Regression assumption: For each j, the $\frac{1}{t}$ column $\mu(_j\Theta)$ of elements $\mu_s(_j\Theta)$, can be written as

$$\underset{\substack{t\\1}}{\mu(_j\Theta)} = \underset{\substack{g\ 1\\t\ g}}{y\beta(_j\Theta)},$$

where y is a known $\frac{g}{t}$ matrix and $\beta(.)$ a $\frac{1}{g}$ vector of elements $\beta_p(.)$. It is assumed moreover that y is of rank g and that $g < t$.

1.2. *Special assumptions*

Occasionally, (vi), (vii) will be specified in the following manner.

(BS): $_jv$ is a diagonal matrix with diagonal elements

$$1/_jp_1, \ 1/_jp_2, \ \ldots, \ 1/_jp_t.$$

(BS1): (BS) is true and moreover, y is the $_t^1$ column $y = (1, 1, \ldots, 1)'$. (Of course, then $\beta(\,.\,)$ is a scalar function.)

The case (BS) is introduced in Bühlmann and Straub, [1] and is further used in Hachemeister [4]. In [1], the number $_jp_s$ is the premium volume underlying treaty j in year s in a reinsurer's portfolio. In [4], each class is related to an American state and $_jp_s$ is a number of claims in state j in the observation period s.

Assumption (BS1) is a stationarity in time assumption, since then $\mu_s(.)$ does not depend on s.

In the sequel we assume (i) to (vii). The matrices $_jv$, y are supposed to be known. Assumptions (BS), (BS1) are mentioned explicitly if they are used.

1.3. *Summary of credibility theory results*

The following credibility approximations to the vector $\beta(_j\Theta)$ are known.

— The inhomogeneous approximation ([3], [5], [2])

$$_j\hat{B} = (1 - _jz) \ b + _jz \ _jB.$$

— The homogeneous approximation ([3], [2])

$$_j\tilde{B} = (1 - _jz) \ b \ S + _jz \ _jB.$$

— The homogeneous mean-free approximation ([5], [2])

$$_j\ddot{B} = (1 - _jz) \ B + _jz \ _jB.$$

In these formulae:

$$_jz = a \ y' \ _jd^{-1} \ y \ (1 + a \ y' \ _jd^{-1} \ y)^{-1},$$

$$_jB = (y' \ _jd^{-1} \ y)^{-1} \ y' \ _jd^{-1} \ _jX,$$

$$B = (\Sigma\ _iz)^{-1}\ \Sigma\ _iz\ _iB,$$
$$\begin{smallmatrix}1\\g\end{smallmatrix} = (\ \begin{smallmatrix}g\\g\end{smallmatrix})\qquad \begin{smallmatrix}g\\g\end{smallmatrix}\ \begin{smallmatrix}1\\g\end{smallmatrix}$$
$$S = \Sigma\ b'\ a^{-1}\ _iz\ _iB\ /\ \Sigma\ b'\ a^{-1}\ _iz\ b$$
$$1 = \ \begin{smallmatrix}g\\1\end{smallmatrix}\ \begin{smallmatrix}g\\g\end{smallmatrix}\quad \begin{smallmatrix}g\\g\end{smallmatrix}\ \begin{smallmatrix}1\\g\end{smallmatrix}\ /\ \begin{smallmatrix}g\\1\end{smallmatrix}\ \begin{smallmatrix}g\\g\end{smallmatrix}\quad \begin{smallmatrix}g\\g\end{smallmatrix}\ \begin{smallmatrix}1\\g\end{smallmatrix}$$

and further:

1) b is the $\begin{smallmatrix}1\\g\end{smallmatrix}$ vector of elements
$$b_p = E\ (\beta_p(_j\Theta)),\ \text{independent of } j.$$

2) $_jd$ is the $\begin{smallmatrix}t\\t\end{smallmatrix}$ matrix of elements
$$_jd_r^s = E\ \text{COV}\ (_jX_r,\ _jX_s\ /\ _j\Theta)\ = E(\sigma^2(_j\Theta))\ _jv_r^s = s^2\ _jv_r^s,$$

where
$$s^2 = E(\sigma^2(_j\Theta)),\ \text{independent of } j.$$

3) a is the $\begin{smallmatrix}g\\g\end{smallmatrix}$ matrix of elements
$$a_p^q = \text{COV}(\beta_p(_j\Theta),\ \beta_q(_j\Theta)),\ \text{independent of } j.$$

1.4. *Problem*

Our problem is to find unbiased estimators for b, s^2, a. For brevity, these quantities will be called, respectively, the *mean vector*, the *variance*, the *covariance matrix*.

2. FIXED-CLASS ESTIMATORS.

In this section we consider a fixed class $_jX$ and we make inferences based only on the variables in that class.

2.1. *Estimation of the mean vector*

2.1.1. *Theorem*

For the estimator
$$_j\hat{b} = (y'\ _jv^{-1}y)^{-1}\ y'_j v^{-1}\ _jX,\ (\begin{smallmatrix}1\\g\end{smallmatrix}), \tag{1}$$

we have
$$E(_j\hat{b}/_j\Theta) = \beta(_j\Theta), \tag{2}$$
$$E(_j\hat{b}) = b. \tag{3}$$

Demonstration. Follows from the fact that
$$E(_jX/_j\Theta) = \mu(_j\Theta) = y\beta(_j\Theta)$$
and the definition of b.

2.1.2. *Remark.* The arguments in favor of the estimator $_j\hat{b}$ are the same as those justifying the identically constructed estimator in multivariate regression theory. We shall not repeat them here, but we note however that such an estimator can be obtained as well by least-squares theory as under normal assumptions.

It is seen that $_j\hat{b}$ is $_jB$ defined in 1.3.

2.2. *Estimation of the variance*

2.2.1. *Lemma*

For any symmetrical $_t^t$ matrix r:

$$E(_jX'\ r\ _jX/_j\Theta) = \sigma^2(_j\Theta)tr(r\ _jv) + \mu'(_j\Theta)\ r\mu(_j\Theta).$$

Demonstration. We have, dropping everywhere the fixed class-index j,

$$E(X'rX/\Theta) = \sum_{rs} r_r^s E(X_r\ X_s/\Theta)$$

$$= \sum_{rs} r_r^s [E(X_r X_s/\Theta) - E(X_r/\Theta)\ E(X_s/\Theta)] + \sum_{rs} r_r^s E(X_r/\Theta)E(X_s/\Theta)$$

$$= \sum_{rs} r_r^s \text{COV}\ (X_r, X_s/\Theta) + \sum_{rs} \mu_r(\Theta)\ r_r^s\ \mu_s(\Theta)$$

$$= \sigma^2(\Theta)\ \sum_{rs} r_r^s v_r^s + \mu'(\Theta)\ r\mu(\Theta)$$

$$= \sigma^2(\Theta)\ tr(rv) + \mu'(\theta)\ r\mu(\Theta).$$

2.2.2. *Theorem*

For the estimator

$$_j\hat{s}^2 = \frac{1}{t-g}\ (_jX - y_j\hat{b})'\ _jv^{-1}\ (_jX - y_j\hat{b}), \qquad (4)$$

we have

$$E(_j\hat{s}^2/_j\Theta) = \sigma^2(_j\Theta), \qquad (5)$$

$$E(_j\hat{s}^2) = s^2 \qquad (6)$$

Demonstration

We drop everywhere the class-index j. Using (1), we have, after obvious simplifications:

$$(t-g)\ \hat{s}^2 = X'\ r\ X,$$

where

$$r = v^{-1} - v^{-1}y(y'v^{-1}y)^{-1}y'v^{-1}.$$

Therefore

$$tr(rv) = tr(1^\circ) - tr(v^{-1}y(y'v^{-1}y)^{-1}y')$$
$$= tr(1^\circ) - tr(y'v^{-1}y(y'v^{-1}y)^{-1})$$
$$= tr(1^\circ) - tr(1^{\circ\circ}) = t - g,$$

where 1°, $1^{\circ\circ}$ are respectively the ${}^t_t$ and the ${}^g_g$ unit matrix.

Also,

$$\mu'(\Theta)\, r\mu(\Theta) = \mu'(\Theta)\, ry\beta(\Theta) = 0,$$

since $ry = 0$. Therefore (5) follows from the lemma. Then (6) is evident from the definition of s^2.

2.3. *Relation for the covariance matrix*

2.3.1. *Remark*

The covariance matrix a cannot be estimated from observations in one class. However, the following relation (8) is the first step in the construction of unbiased estimators for a. Observe that, as is indicated, the relations (7), (8) are ${}^g_g$ matrix relations.

2.3.2. *Theorem*

$$E((_j\hat{b} - b)\, (_j\hat{b} - b)'/_j\Theta) = \sigma^2(_j\Theta)\, (y'\,_j v^{-1}y)^{-1} +$$
$$+ (\beta(_j\Theta) - b)\, (\beta(_j\Theta) - b)', \; {\binom{g}{g}}, \qquad (7)$$

$$E((_j\hat{b} - b)\, (_j\hat{b} - b)') = s^2(y'\,_j v^{-1}y)^{-1} + a, \; {\binom{g}{g}}. \qquad (8)$$

Demonstration

We drop everywhere j. First we prove:

$$E(XX'/\Theta) = \sigma^2(\Theta)\, v + y\beta(\Theta)\beta'(\Theta)y'. \qquad (9)$$

Indeed, the ${}^s_r$ element of the first member of (9) is

$$E(X_r X_s/\Theta) = \text{COV}(X_r, X_s/\Theta) + E(X_r/\Theta)\, E(X_s/\Theta)$$
$$= \sigma^2(\Theta)v_r^s + \mu_r(\Theta)\, \mu_s(\Theta).$$

So we have (9) since the last expression is the ${}^s_r$ element of the matrix

$$\sigma^2(\Theta)\, v + \mu(\Theta)\mu'(\Theta) = \sigma^2(\Theta)\, v + y\beta(\Theta)\beta'(\Theta)\, y'.$$

By (1): $\widehat{bb}' = (y'v^{-1}y)^{-1}y'v^{-1}XX'\,v^{-1}y(y'v^{-1}y)^{-1}.$

By an application of $E(./\Theta)$, using (9):

$$E(\widehat{bb}'/\Theta) = \sigma^2(\Theta)\, (y'v^{-1}y)^{-1} + \beta(\Theta)\, \beta'(\Theta). \qquad (10)$$

From (2) and the relation

$$(\hat{b} - b)(\hat{b} - b)' = \hat{b}\hat{b}' - b\hat{b}' - \hat{b}b' + bb'$$

it follows that

$$E(\,(\hat{b} - b)(\hat{b} - b)'/\Theta) = E(\hat{b}\hat{b}'/\Theta) - b\beta'(\Theta) - \beta(\Theta)\,b' + bb'.$$

Combining this relation with (10), we have (7). Then (8) follows.

3. GLOBAL ESTIMATORS.

Here we use the statistical material of all the classes.

3.1. *Estimation of the mean vector*

3.1.1. *Theorem*

Whatever be the $\frac{g}{g}$ matrices $_j\pi$ satisfying $\Sigma\,_j\pi = \mathrm{I}$, the vector

$$\hat{b} = \Sigma\,_j\pi\,_j\hat{b} \tag{11}$$
$$\underset{g}{1} = \underset{g}{0}\,\underset{g}{1}$$

is an unbiased estimator of b.

Demonstration

Use (3).

3.1.2. *Natural estimator*

In the (BS) case, the estimator

$$\hat{b} = \Sigma\,_jp\,_j\hat{b}, \tag{12}$$

where the scalars $_jp$ are defined by

$$_jp = \Sigma\,_jp_s/\Sigma\,_tp_s, \tag{13}$$

will be called the *natural estimator* of b. The natural estimator is a particular estimator (11) obtained by taking for $_j\pi$ the diagonal matrix with each diagonal element equal to $_jp$. The numbers $_jp$ will be called the *natural weights*. The matrices $_j\pi$ in (11) can be considered as generalized weights.

The natural estimator $\hat{b}$ is used (at least implicitly) in Buhlmann and Straub, [1], in the (BS1) case.

3.2. *Estimation of the variance*

3.2.1. *Theorem*

Whatever be the scalar weights ${}_j\rho$ satisfying $\sum_j {}_j\rho = 1$,

$$\mathcal{S}^2 = \sum_j {}_j\rho \; {}_j\mathcal{S}^2 \tag{14}$$

is an unbiased estimator of s^2.

Demonstration

Use (6).

3.2.2. *Natural and unweighted estimators*

In the (BS) case, the estimator

$$\mathcal{S}^2 = \sum_j {}_jp \; {}_j\mathcal{S}^2 \tag{15}$$

will be called the *natural estimator of s^2*.

In the general case, the estimator

$$\mathcal{S}^2 = \frac{1}{k} \sum_j {}_j\mathcal{S}^2 \tag{16}$$

will be called the *unweighted estimator of s^2*.

The unweighted estimator is considered in Bühlmann and Straub, [1] in the (BS1) case and also in Hachemeister, [3] in the more general (BS) case.

3.3. *Estimation of the covariance matrix*

3.3.1. *Theorem*

Let ${}_{ij}k$ be weights satisfying ${}_{ij}k = {}_{ji}k$, $\sum_{ij} {}_{ij}k = 1$ and set ${}_{i\Sigma}k = \Sigma_{ij}k$. Let $\mathcal{S}^2$ be an unbiased estimator of s^2. Then the ${}^\theta_\theta$ matrix $\hat{a}$ defined by the relation

$$\sum_{ij} {}_{ij}k \left({}_i\widehat{b} - {}_j\widehat{b} \right) \left({}_i\widehat{b} - {}_j\widehat{b} \right)' =$$

$$2(1 - \sum_i {}_{ii}k) \, \hat{a} + 2 \, \mathcal{S}^2 \sum_i ({}_{i\Sigma}k - {}_{ii}k) \, (y'_i \, v^{-1}y)^{-1}, \; \binom{\theta}{\theta}, \tag{17}$$

is an unbiased estimator of a.

Demonstration

$$\left({}_i\widehat{b} - {}_j\widehat{b} \right) \left({}_i\widehat{b} - {}_j\widehat{b} \right)' = \left(({}_i\widehat{b} - b) - ({}_j\widehat{b} - b) \right) \left(({}_i\widehat{b} - b) - ({}_j\widehat{b} - b) \right)' =$$

$$({}_i\widehat{b} - b) ({}_i\widehat{b} - b)' + ({}_j\widehat{b} - b) ({}_j\widehat{b} - b)' - ({}_i\widehat{b} - b) ({}_j\widehat{b} - b)' -$$

$$({}_j\widehat{b} - b) ({}_i\widehat{b} - b)'.$$

Therefore, by the assumption of independence of classes, by (3), (8), writing momentarily $_i w$ for $s^2 \, (y' \, _i v^{-1} y)^{-1}$:

$$E(\Sigma_{ij} \, _{ij}k \, (_i\hat{b} - _j\hat{b}) \, (_i\hat{b} - _j\hat{b})') =$$

$$\Sigma_{ij} \, _{ij}k(_iw + a) + \Sigma_{ij} \, _{ij}k \, (_jw + a) - 2 \, \Sigma_{ij} \, _{ij}k \, \delta_{ij} \, (_iw + a) =$$

$$2 \, a \, (\mathrm{I} - \Sigma_i \, _{ii}k) + 2 \, \Sigma_i \, (_i\Sigma k - _{ii}k)_i w.$$

From this the theorem is clear.

3.3.2. *Natural estimator*

In the (BS) case, let $_{ij}k = \, _ip \, _jp$. Then $\hat{a}$ defined by (17) will be called the *natural estimator of a*, for the given $\hat{s}^2$, even if the latter estimator is not the natural one.

If $\hat{b}$ is the natural estimator (12) of b, then

$$\Sigma_{ij} \, _ip \, _jp \, (_i\hat{b} - _j\hat{b}) \, (_i\hat{b} - _j\hat{b})' =$$

$$\Sigma_{ij} \, _ip \, _jp \, (_i\hat{b} \, _i\hat{b}' + _j\hat{b} \, _j\hat{b}' - _i\hat{b} \, _j\hat{b}' - _j\hat{b} \, _i\hat{b}') =$$

$$2 \, \Sigma_i \, _ip \, _i\hat{b} \, _i\hat{b}' - 2 \, \hat{b} \, \hat{b}' = 2 \, \Sigma_i \, _ip \, (_i\hat{b} - \hat{b}) \, (_i\hat{b} - \hat{b})'.$$

So the natural estimator $\hat{a}$ results from the relation

$$\Sigma_i \, _ip \, (_i\hat{b} - \hat{b}) \, (_i\hat{b} - \hat{b})' =$$

$$(\mathrm{I} - \Sigma_i \, _ip^2) \, \hat{a} + \hat{s}^2 \, \Sigma_i \, _ip(\mathrm{I} - _ip) \, (y' \, _iv^{-1}y)^{-1}, \, \binom{g}{g}, \qquad (18)$$

where $\hat{b}$ is the natural estimator (12) of b.

3.4. *The (BS1) case*

3.4.1. *Notations*

Here we consider the (BS1) case and use the notations

$$_j p_\Sigma = \Sigma_\bullet \, _j p_s, \; _\Sigma p_\Sigma = \Sigma_j \, _j p_\Sigma = \Sigma_{j\bullet} \, _j p_s.$$

Then the natural weights are $_j p = \, _j p_\Sigma / _\Sigma p_\Sigma$.

We use the abbreviations

$$_j X_E = \sum_\bullet \frac{_j p_s}{_j p_\Sigma} \, _j X_s, \; _E X_E = \sum_j \frac{_j p_\Sigma}{_\Sigma p_\Sigma} \, _j X_E = \sum_{j\bullet} \frac{_j p_s}{_\Sigma p_\Sigma} \, _j X_b.$$

3.4.2. *Estimation of the mean*

Now $_j\hat{b}$, $\hat{b}$ are scalars denoted by $_j\hat{m}$, $\hat{m}$. By particularisation

of the general results we have $_j\hat{m} = {}_jX_E$ and the natural mean equals $\hat{m} = {}_E\bar{X}_E$.

3.4.3. Estimation of the variance

The j-th class variance estimator is, by particularisation of (4):

$$_j\hat{s}^2 = \frac{1}{t-1} \Sigma\ _jp_s\ (_jX_s - {}_jX_E)^2. \tag{19}$$

The unweighted estimator is

$$\hat{s}^2 = \frac{1}{k(t-1)} \Sigma\ _jp_s\ (_jX_s - {}_jX_E)^2. \tag{20}$$

This is the estimator considered in Bühlmann and Straub, [1].

3.4.4. Estimation of a

The natural estimator $\hat{a}$, a scalar in this case, results from the relation

$$(1 - \Sigma\ _jp^2)\ \hat{a} = \Sigma\ _jp({}_jX_E - {}_EX_E)^2 - (k-1)\ \hat{s}^2/{}_\Sigma p_\Sigma, \tag{21}$$

obtained from (18). The Bühlmann and Straub, [1] estimator $\hat{\hat{a}}$ results from the relation

$$(1 - \Sigma_j\ _jp^2)\ \hat{\hat{a}} = \sum_j \frac{_jp_s}{{}_\Sigma p_\Sigma}\ (_jX_s - {}_E\bar{X}_E)^2 - \frac{kt-1}{{}_\Sigma p_\Sigma}\ \hat{s}^2. \tag{22}$$

By the identity

$$\Sigma\ _jp_s\ (_jX_s - {}_E\bar{X}_E)^2 = \Sigma\ _jp_s({}_jX_s - {}_jX_E)^2 + \Sigma\ _jp_\Sigma\ (_jX_E - {}_E\bar{X}_E)^2, \quad 23$$

it is seen that $\hat{\hat{a}} = \hat{a}$ if $\hat{s}^2$ is the unweighted estimator (20).

4. Optimal Estimation of the Mean Regression Vector

4.1. Optimal estimator

An estimator $\hat{e}$ in a set E of vector estimators shall be called *optimal* in E if the trace of the covariance matrix of $\hat{e}$ is minimal, in comparison with the traces defined similarly for the other elements in E. If E is a set of scalar estimators, the principle invoked is that of minimum-variance.

We leave the question of an optimal $\hat{a}$ or $\hat{s}^2$ unsettled. We consider, here, the case of an optimal $\hat{b}$ given by (11). We prove that the optimal sequence $(_1\pi, {}_2\pi, \ldots, {}_k\pi)$ is the sequence $(_1z, {}_2z, \ldots, {}_kz)$ of credibility matrices (see 1.3), except for the constant pre-factor $(\Sigma_iz)^{-1}$.

4.2. *Lemma*

Let $_1m, _2m, \ldots, _km$ be definite positive symmetrical $_g^g$ matrices. Let $_1x, _2x, \ldots, _kx$ be variable $_g^g$ matrices. Then the minimum of the trace

$$tr(\Sigma_i \, _ix \, _im \, _ix'), \tag{24}$$

subject to the constraint $\Sigma_i \, _ix = \mathrm{I}$, is reached for

$$_jx = (\Sigma_i \, _im^{-1})^{-1} \, _jm^{-1}, \; (j = \mathrm{I}, 2, \ldots, k). \tag{25}$$

Demonstration

If $_ix$ is fixed and if y is an arbitrary $_g^g$ vector, we have

$$y' \, (_ix \, _im \, _ix') \; y = (y' \, _ix)_im(y' \, _ix)' \geq 0$$

since $_im$ is definite positive. Therefore $_ix \, _im \, _ix'$ is semidefinite positive and has a nonnegative trace. Thus, (24) is ≥ 0. It is a quadratic form in the kg^2 variables $_ix_p^q$. If we eliminate g^2 variables by the constraints $\Sigma_i x = \mathrm{I}$, we have a quadratic polynomial in $(k - \mathrm{I})g^2$ independent variables that is never negative. Such a function is minimum for finite values of the variables. (See, for example, the lemma 2.5 in De Vylder, [2]). Now we shall apply Lagrange's method and we shall find a unique extremum. From the preceding discussion it follows that this extremum must be the minimum.

We introduce the g^2 Lagrange multipliers λ_α^β corresponding to the constraints

$$\Sigma_i \, _ix_\alpha^\beta = \delta_\alpha^\beta.$$

We minimize

$$L = tr \, (\Sigma_i \, _ix \, _im \, _ix') - 2 \sum_{i\alpha\beta} \lambda_\alpha^\beta \, _ix_\alpha^\beta$$

$$= \sum_{i\alpha\beta\gamma} \, _ix_\alpha^\beta \, _ix_\alpha^\gamma \, _im_\gamma^\beta - 2 \sum_{i\alpha\beta} \lambda_\alpha^\beta \, _ix_\alpha^\beta.$$

We must have

$$0 = \frac{\mathrm{I}}{2} \frac{\partial L}{\partial_j x_p^q} = \sum_\gamma \, _jx_p^\gamma \, _jm_\gamma^q - \lambda_p^q,$$

or, in matrix form,

$$_jx \, _jm = \lambda.$$

Then, successively,

$$_jx = \lambda_j m^{-1}, \; \mathrm{I} = \Sigma_j \, _jx = \lambda\Sigma_j \, _jm^{-1}, \; \lambda = (\Sigma_j \, _jm^{-1})^{-1},$$

and (25). Note that the existence of the inverse matrix of $\Sigma_j m^{-1}$ results from the assumptions.

4.3. Lemma

The covariance matrix of $\hat{b}$, given by (11), is

$$\Sigma_i \pi \ (s^2 \ (y'_i v^{-1} y)^{-1} + a)_i \pi', \ \binom{g}{g}. \tag{26}$$

Demonstration

From (8) and from the independence of classes:

$$E((\Sigma_i \pi \ _i\hat{b} - b) \ (\Sigma_j \pi \ _j\hat{b} - b)') =$$
$$E(\Sigma_i \pi \ (_i\hat{b} - b) \ (\Sigma_j \pi (_j\hat{b} - b))') =$$
$$\Sigma_{i\,j} \ _i\pi \ E((_i\hat{b} - b) \ (_j\hat{b} - b)') \ _j\pi' \ \delta_{ij} =$$
$$\Sigma_i \ _i\pi \ (s^2(y' \ _i v^{-1} y)^{-1} + a) \ _i\pi'.$$

4.4. Theorem

The optimal estimator $\hat{b}$ in the class of estimators (11) is

$$\hat{b} = \Sigma_j \ (\Sigma_i \ _i z)^{-1} \ _j z \ _j b, \tag{27}$$

where the $_j z$ are the credibility matrices defined in 1.3.

Demonstration

From the definition of $_j z$ follows the relation

$$(y' \ _j d^{-1} y) + a = \ _j z^{-1} \ a.$$

Then, since

$$s^2 (y' \ _j v^{-1} y)^{-1} = (y' \ _j d^{-1} y)^{-1},$$

the theorem follows from the lemma's.

4.5. Corollary

If b is estimated optimally, there is no difference between the credibility approximations $_j\hat{B}$, $_j\tilde{B}$, $_j\ddot{B}$ to $\beta(_j\Theta)$ given in 1.3.

4.6. Remarks

It seems that we are in a circular situation if we try to use the optimal $\hat{b}$, since this $\hat{b}$ depends on a and that $\hat{b}$ is needed in, for example, the natural estimator $\hat{a}$ of a.

However, this anomaly is only apparent, since the first member of (18) can be written without $\hat{b}$. In other words, in (18) $\hat{b}$ *must* be the

natural estimator and not the optimal one. It is not excluded, however, that (18) can be optimized in some way by a method of successive approximations, using successively improved $\hat{b}$'s and, eventually, redefinitions of the numbers $_jp_s$.

5. NON-NEGATIVITY CONSIDERATIONS

The covariance matrix a is semidefinite positive. In particular, in the (BS1) case the number a is non-negative. It is known that the estimator $\hat{a}$ can provide negative values. In such cases, Bühlmann and Straub, [1] estimate a by o.

A similar method can be used if a is a matrix. For example, suppose that $a°$ is an estimate of a and that $a°$ is not semidefinite positive. Then make $a°$ diagonal by an orthogonal transformation. Replace the negative diagonal elements (i.e. the negative characteristic values) by o and apply the inverse orthogonal transformation.

If all diagonal elements of $a°$ are positive and if $a°$ is not semi-definite positive the following method can also be used. Multiply all non-diagonal elements of $a°$ by the same number x. Then if x decreases from 1 to o the matrix becomes necessarily semidefinite positive. Keep the largest possible x.

Of course, a justification of these methods is difficult to find. Moreover, the estimators redefined in such a way are no longer unbiased. But it must be kept in mind that it is preferable to have an estimate that might be bad, than no estimate at all. And also that the application of credibility formulae with wrong parameters introduces unfairnesses in the different classes, but that these counterbalance each other, at least if b is estimated correctly.

Finally, let we go back to the general formula (17) and let we consider the following question. Is a reasonable general choice of the weights $_{ij}k$ and $_jρ$ (in S^2) possible in such a way that the resulting $\hat{a}$ always is semidefinite positive? The answer is negative. Indeed, let we consider the (BS1) case with each $_jp_s = 1$. Then our general hypothetical rule for fixing the weights must lead to equal weights $_jρ$ since we start from a symmetrical situation. For the same reason, we must have

$$_{ij}k = α \ (i \neq j), \ _{ii}k = β$$

for some $α$ and $β$. Since we must have $Σ_{ij}k = 1$, there is one in-dependent parameter, say $α$, left. But an inspection of (17) shows that this parameter simplifies in that relation. So we may take $α =$

β. Then $\hat{a}$ is the natural estimator, given by (21). The particular case $k = 2$, $t = 2$ shows that $\hat{a} < 0$ for the values

$$_1X_1 = 1, \ _2X_1 = 1,$$

$$_1X_2 = 0, \ _2X_2 = 0.$$

BIBLIOGRAPHY

[1] BUHLMANN, H., and STRAUB, E., (1970) Glaubwurdigkeit fur Schaden-satze, Mitteilungen der Vereinigung Schweizerischer Versicherungs-mathematiker.

[2] DE VYLDER, FL., (1976) Geometrical Credibility, Scandinavian Actuarial Journal.

[3] HACHEMEISTER, CH., Credibility for regression models with application to trend

[4] STRAUB, E., Credibility in practice

[5] TAYLOR, G. C., Abstract credibility

The papers [3], [4] were presented to the Actuarial Research Conference on Credibility Theory, Berkeley, 1974. They are published in the book "Credibility" edited by P. M Kahn, Academic Press, 1975.

The paper [5] will be published imminently.