

SOME APPLICATIONS OF UNIVARIATE CALIBRATION MODELS

Dr.JikceyIsaac*

Department of Mathematics & Statistics, Assumption College, Changanacherry, Kerala-686101, India.

Abstract

A statistical calibration problem is a kind of inverse prediction, or inverse regression. First of all we consider univariate calibration, then we extend this idea to multivariate case. Suppose we want to measure the amount of chemical, x , in a sample. Direct measurement is very difficult or expensive or time consuming. Hence we consider a related measurement, y , which is easy and inexpensive to obtain. Now we prepare N samples with known amounts of the chemical x_i , $i = 1, 2, \dots, N$. Measure the corresponding value of y , $i = 1, 2, \dots, N$. A model is fitted to these data and this model is used in a future study to estimate the true characteristic x using the less expensive or easily obtained measurement y . We deal with the situation where x is fixed (i.e., non-random).

Index Terms - calibration model; confidence region ; inverse regression.

1. Introduction

Let y_1, y_2, \dots, y_N are independent p variate normally distributed random vectors. We deal only with the situation where x is fixed (i.e. controlled)

Let $Y = (y_1, y_2, \dots, y_N)$ and $X = (x_1, x_2, \dots, x_N)$, Then Thus $y \sim N (AX, \Sigma)$

The columns of Y are independent multivariate normal random vectors with

$$E(Y) = AX \text{ and } \text{Cov} (\text{Vec} (Y)) = I_N \otimes \Sigma \quad (1)$$

where A is a $p \times q$ matrix of unknown parameters and the variance covariance matrix Σ may be partly known or completely unknown. The matrix X ($q \times N$) is also assumed to be of rank q . Now consider another $p \times 1$ normally distributed random vector y , corresponding to an unknown value θ of x and independent of Y in (1) and assume the same multivariate linear model (as in 1). Then we get

$$E(y) = A\theta \text{ and } \text{Cov} (y) = \Sigma \quad (2)$$

If $p = q = 1$ then we have the problem of univariate calibration. If p and/or q is more than one, the problem become multivariate calibration.

The confidence regions that are required in this context will be one of two types, depending on the particular application.

- (i) **Single use confidence regions:** If the calibration data, i.e, Y in (1) is used to construct a confidence region for θ_1 corresponding to a single future observation y_{N+1} .
- (ii) **Multiple use confidence regions:** If the calibration data is used to construct a sequence of confidence region, one at a time, for a sequence of unknown parameters $\theta_1, \theta_2, \theta_3, \dots$ we are in the set up of multiple use confidence regions. In other words, we have a sequence of independent future observations $y_{N+i} \sim N (A\theta_i, \Sigma)$, $i = 1, 2, \dots$, and as soon as we observe y_{N+i} , a confidence region is constructed for θ_i using y_{N+i} and the calibrated data Y . This is done a large number of times for $i = 1, 2, \dots$ and the θ_i 's could be different.

Some references on the construction of single use confidence regions are: Brown (1982), Fujikoshi and Nishii (1984), Oman and Wax (1984), Mathew and Zha (1996). The problem of constructing multiple use confidence regions addressed in Osborne (1991), Brown (1993) and Mathew and Zha (1997). The problem of deriving joint confidence regions is addressed in Thomas (1994) and Mathew and Sharma (1999). We consider univariate calibration through out this project.

2.Examples of Univariate Calibration

Example - 2.1

The estimation of blood alcohol concentration.

This deals with the calibration of breath estimates of blood alcohol concentration, based on the results of a laboratory test. Here, the y_j 's are breath estimates of blood alcohol concentrations and the x_j 's are the actual concentration of alcohol in the blood, obtained by a laboratory test. The relevant data are based on a study conducted at Acadiana Criminalistics laboratory, New Iberia, Louisiana. Here we estimate unknown blood alcohol concentration θ , after obtaining the corresponding breath estimate say y_0 using an appropriate fitted model between y_i 's and x_j 's. Such an estimation is preferable to the actual laboratory determination of the blood alcohol concentration, since it is much easier and faster to obtain the breath estimates.

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

$$y_0 = \beta_0 + \beta_1 \theta + e_0, i = 1, 2, \dots, n. \quad (3)$$

$$e_i \sim N(0, \sigma^2), e_0 \sim N(0, \sigma^2)$$

Legal limit of safe driving is $\theta \leq 0.10\%$.

The problem is to estimate the unknown blood alcohol concentration θ using breath estimate y_0 .

Example -2.2

The estimation of weight of rubber sheet

The major crop of the farmers of Kerala is rubber. This is one of the main income yielding crops in Kerala. The usual method of making rubber sheets is through the collection of rubber latex through farmer societies. This way of collection reduces the hardship of the farmers, since they need not bother about the different stages of processing of rubber sheets. But one problem remains, that is, related to the immediate calculation of the weight of the sheet from 1kg rubber latex. It is very difficult to calculate the weight of the sheet from 1kg rubber latex and it consumes much time. This is an area where we would be able to make use of statistical calibration theory which may be of interest to farmers in the locality.

For this we can proceed as follows.

Take samples of rubber latex from each farmer. Number them respectively, and then prepare dry rubber sheet of each sample and weigh them. Let Y be the weight of rubber latex and X be the weight of rubber sheet. We can fit a suitable relationship between X and Y using available data of the form

$Y = a_0 + a_1X$. Using this model we can estimate unknown value θ of x for a particular value of y .

Hence through the application of calibration theory we are in a position to measure the weight of each rubber sheet from the rubber latex at the earliest and again it helps the farmer and buyer to fix the price on the spot accurately. Thus this theory would be a milestone in the development of rubber plantation and business in Kerala. Similar methods can be applied for grading of rubber sheets as RMS-1, RMS-2, RMS-3, RMS-4 etc. the same method can be used for grading of spices like cardamom, determining the quality of tea leaves etc.

2.3 Estimation of Enzyme Concentration in Blood Plasma

An example of data from a controlled calibration experiment is given in Aitchison and Dunsmore (1975). Enzyme concentration in blood plasma can be determined by a long and costly laboratory method where as an auto analyser method is quick and cheap. Here to calibrate the auto analyser, nine plasma samples selected to cover the range of enzyme concentrations have each been divided into four aliquots, one aliquot being assigned to the laboratory method and the other three to separate analyser determinations.

2.4 Estimation of the Age of a Dead Hair

A biological study reported in Lieftinck Koeijers (1988). The problem is to estimate the age of a dead hair. It was noted that the age is related to several criteria, like body weight, eye lens weight and length of the hind foot, and furthermore, a linear relationship exists between the various criteria and age (on a log scale). The linear model parameters can be estimated based on a calibration experiment involving several hares of known ages. The model can be used to estimate the unknown age of a dead hare on the basis of the vector of measured criteria.

3. Analysis of Univariate Calibration

Here we consider the models (3.1). $e_i \sim N(0, \sigma^2)$, $e_0 \sim N(0, \sigma^2)$, y_i 's are independent. y_i , $i = 1, 2, \dots, n$ are called calibration data.

i) Estimation of θ

\bar{y} = average of the y_i 's

\bar{x} = average of the x_i 's

$$\left. \begin{aligned} S_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2 \\ S_{yy} &= \sum_{i=1}^n (y_i - \bar{y})^2 \\ S_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \end{aligned} \right\} \quad (4)$$

$$\left. \begin{aligned} \hat{\beta}_1 &= \frac{S_{xy}}{S_{xx}} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\sigma}^2 &= \frac{1}{n-2} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \end{aligned} \right\} \quad (5)$$

If β_0 and β_1 are known,

$$\hat{\theta} = \frac{y_0 - \beta_0}{\beta_1} \quad (6)$$

Since β_0 and β_1 are unknown, replace by the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

$$\begin{aligned} \hat{\theta} &= \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} \\ &= \bar{x} + \frac{S_{xx}}{S_{yy}} (y_0 - \bar{y}) \end{aligned} \quad (7)$$

ii) Confidence intervals for θ

Not possible to obtain confidence intervals based on the distribution of $\hat{\theta}$; the distributions depend on β_1 . Consider,

$$y_0 - \hat{\beta}_0 - \hat{\beta}_1 \theta \sim N \left[0, \sigma^2 \left(1 + \frac{1}{n} + \frac{(\theta - \bar{x})^2}{S_{xx}} \right) \right] \quad (8)$$

Define

$$T(\theta) = \frac{y_0 - \hat{\beta}_0 - \hat{\beta}_1 \theta}{\left[\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(\theta - \bar{x})^2}{S_{xx}} \right) \right]^{1/2}} \quad (9)$$

$T(\theta) \sim$ student's t distribution with d.f. $n-2$

$$\{\theta : |T(\theta)| \leq t_{\alpha/2}\} \quad (10)$$

This represents 100 $(1 - \alpha)$ % confidence region for θ . (Fieller's theorem)

Remark

This need not be an interval. Is an interval if $\frac{\sqrt{S_{xx}} |\hat{\beta}_1|}{\hat{\sigma}} > t_{\alpha/2}$, i.e.,

if $H_0 : \beta_1 = 0$ is rejected.

4. Conclusion

The earliest articles on calibration are due to Eisenhart (1939) and Fieller (1954), dealing with univariate calibration. Since then, numerous articles have appeared dealing with point and confidence set estimation in univariate calibration. An excellent review of there appears in the article by Osborne (1993) and in the book by Brown (1994). An important unsolved problem is the minimization of difficulties in the computational work associated with joint and multiple use confidence regions. Bayesian methods can also be developed for constructing more efficient confidence regions. This is another area which needs further research. Some such works are currently under investigation. However, the developments in multivariate calibration is fairly recent and some problems related in this area are not yet to be solved.

REFERENCES

1. Brown, P.J. (1982). Multivariate calibration (with discussion). *J. Roy. Statist. Soc., Ser. B*, 44, 287-321.
2. Brown, P.J. (1993). *Measurement, regression and calibration*. Oxford University Press, New York.
3. Fujikoshi, Y. and Nishii, R. (1984). On distribution of a statistic in multivariate inverse regression analysis. *Hiroshima Mathematics Journal*, 14, 215-225.
4. Jikcey Isaac and Jose K.K. (2006). A study on certain improved calibration methods and their use in biological and industrial research. *Journal of Science and Society*.
5. Jose K.K. and Jikcey Isaac (2007). Approximate single use confidence Regions in Multivariate Controlled Calibration. *Jour. Stat. Planning and Inference*.
6. Jose K.K. and Jikcey Isaac (2007b). Estimation of Foetal age by ultra sound measurements using multivariate calibration theory. *STARS Int. Journal* 5, 2, 12-20.
7. Lieftinck – Koeijers, C.A.J. (1988). Multivariate calibration: A generalization of the classical estimator. *Journal of Multivariate Analysis*, 25, 31-44.
8. Mathew, T. and Sharma, M.K. (1999). Joint confidence region in the multivariate calibration problem. *The Annals of Statistics*, Preprint.
9. Mathew, T. and Zha, W. (1996). Conservative confidence regions in multivariate calibration. *Annals of Statistics*, 24, 707-725.
10. Mathew, T. and Zha, W. (1997). Multiple Use Confidence regions in Multivariate Calibration. *Journal of American Statistical Association*, Vol. 92, 1141-1150.
11. Oman, S.D. and Wax, Y. (1984). Estimating fetal age by ultrasound measurements: An example of multivariate calibration. *Biometrics*, 40, 947-960.

12. Osborne, C. (1991). Statistical calibration: a review. *International Statistical Review*, 59, 309-336.
13. Thomas, E.V. (1991). Errors – in – variables estimation in multivariate calibration. *Technometrics*, 33, 405-413.
14. William G. Hunter and Warren F. Lamboy, (1981). A Bayesian analysis of the linear calibration problem. *Technometrics*, Vol. 23, 323-350.

