A Non Bayesian Predictive Approach for Functional Calibration

Noslen Hernández[1], Rolando J. Biscay[2], and Isneri Talavera[1]

[1] Advanced Technologies Application Center, CENATAV, Cuba
[2] Universidad de Valparaíso, Chile
{nhernandez,italavera}@cenatav.co.cu, rolando.biscay@uv.cl

Abstract. A non Bayesian predictive approach for statistical calibration with functional data is introduced. This is based on extending to the functional calibration setting the definition of non Bayesian predictive probability density proposed by Harris (1989). The new method is elaborated in detail in case of Gaussian functional linear models. It is shown through numerical simulations that the introduced non Bayesian predictive estimator of the unknown parameter of interest in calibration (commonly, a substance concentration) has negligible bias and compares favorably with the classical estimator, particularly in extrapolation problems. A further advantage of the new approach, which is also briefly illustrated, is that it provides not only point estimates but also a predictive likelihood function that allows the researcher to explore the plausibility of any possible parameter value.

Keywords: statistical calibration, functional data analysis, regression.

1 Introduction

Statistical calibration plays a crucial role in many areas of technology such as pharmacology and chemometrics ([1–6]). In general, the calibration problem can be described as follows. A training sample $(z_i, x_i)_{1\leq i\leq n}$ of n independent observations in some space $\mathcal{Z}\times\mathcal{X}$ is available, which are generated according to conditional probability distributions $F\left(X/z_i\right)$ that belong to some specified statistical model $\{F_\theta\left(X/Z\right) : \theta\in\Theta\}$. The observations z_i may be non random (case of fixed design) or random (case of random design). Given a new observation X generated according to the distribution $F\left(X/z\right)$ with an unknown value z, it is desired to obtain an estimate or prediction $\widehat{z}$ of z.

There are several works that deal with this problem in the setting in which $z_1, ..., z_n$, z are observations of a real random variable and X is a random function. All of them are based on different approximations $\widehat{z}$ to the conditional mean $\widetilde{z} = E\left(Z/X\right)$ [7, 8]. This estimator $\widetilde{z}$ is optimal in the sense of minimizing the quadratic Bayesian risk $E\left(\widetilde{z}-Z\right)^2$. However, these approaches have two fundamental shortcomings: a) they focus on the case of random design; b) the estimator $\widetilde{z}$ has poor performance for extrapolation, i.e., when the unknown quantity Z is not generated by the same probabilistic mechanism as the previous data $z_1, ..., z_n$ and lies far away from this cluster of points.

L. Alvarez et al. (Eds.): CIARP 2012, LNCS 7441, pp. 781–788, 2012.

The calibration problem has also been studied in [9] for the setting in which Z and X are functions, the mean $\mu(z)$ of the statistical model $F_\theta(X/Z)$ is linear with respect to z, and the design is fixed. For this problem, a non Bayesian estimate estimator $\widehat{z}$ is introduced on the basis of a regularized inversion of the linear operator defined by the mean function μ. This generalizes to a functional framework the regularization of the so-called classical estimator for calibration proposed in [10] for the scalar linear model.

In the present work we are interested in the specific functional setting, common in chemometrics, in which the variable X is a function (e.g., a spectral curve), the variable $\mathbf{z}$ is a finite-dimensional vector (e.g., concentrations of some substances), and the statistical model is linear and Gaussian with respect to $\mathbf{z}$. In contrast to previous methods, we introduce an approach for statistical calibration based on a non Bayesian predictive framework. This extends to such functional setting the non Bayesian predictive approach for statistical calibration proposed in [11] for the scalar linear model.

More specifically, the non Bayesian predictive density introduced by Harris [12] is extended to the calibration setting just described, so providing, on the basis of the training sample, a non Bayesian *predictive density* $f_P(x;\mathbf{z})$, for a new observation x corresponding to the unknown $\mathbf{z}$. This allows one to define the non Bayesian *predictive likelihood* by

$$L(\mathbf{z}) = f_P(x;\mathbf{z}),$$

and the non Bayesian *predictive estimator* $\widehat{\mathbf{z}} = \arg\max_{\mathbf{z}} L(\mathbf{z})$.

It is shown that this new approach has a number of potential advantages: *i*) good performance for extrapolation; *ii*) negligible bias; *iii*) it can be applied to both random and fixed designs, *iv*) it offers not only a point estimate $\widehat{\mathbf{z}}$ but also a predictive likelihood function $l_P(z)$ that allows one to explore the likelihoods of all the possible values of the unknown $\mathbf{z}$; it permits to incorporate the information that some components of the vector $\mathbf{z}$ are known, when such information is available.

The rest of the paper is organized as follows. Section 2 presents the functional Non Bayesian Predictive estimator. Sections 3 illustrates its performance in a simulation study. Finally, some concluding remarks are given in Section 4.

2 Functional Non Bayesian Predictive Estimator

Let be given a sample of previous (training) data $(\mathbf{z}_i, x_i) = (z_{i1}, ..., z_{iq}, x_i)$ $(i = 1, ..., n)$ that follow the model:

$$x_i(t) = \beta_1(t) z_{i1} + ... + \beta_q(t) z_{iq} + e_i(t).$$

Here, $x_i \in \mathcal{X} = L_2(t, \mathbb{R})$ is a functional responses $(t \in [0,1])$; $\mathbf{z}_i \in \mathbb{R}^q$ is vector of covariates; $\boldsymbol{\beta}(t) = (\beta_1(t), ..., \beta_q(t))^\mathsf{T}$ is a vector of non random functions (coefficients); and $e_1, e_2, ...$ are independent zero-mean Gaussian functions in L_2

with covariance function σ_e. We will denote by $N_{L_2}(\mu, \sigma)$ the Gaussian distribution of a random function with mean μ and covariance function σ. Thus, the distribution of e_i is $P_e = N_{L_2}(0, \sigma_e)$.

This model can be written

$$\mathbf{x}\left(t\right) = \mathbf{Z}\boldsymbol{\beta}\left(t\right) + \mathbf{e}\left(t\right), \tag{1}$$

where $\mathbf{Z} = (Z_{ij})$ is the $n \times q$ design matrix, $\mathbf{x}\left(t\right) = \left(x_1\left(t\right), ..., x_n\left(t\right)\right)^{\mathsf{T}}$, and $\mathbf{e}\left(t\right) = \left(e_1\left(t\right), ..., e_n\left(t\right)\right)^{\mathsf{T}}$.

It is assumed that a new data x is available which follows model (1), that is

$$x(t) = \beta_1(t)z_1 + ... + \beta_q(t)z_q + e(t) = \boldsymbol{\beta}^{\mathsf{T}}\left(t\right)\mathbf{z} + e\left(t\right). \tag{2}$$

The problem of interest is to estimate the vector of variables $\mathbf{z}$ on the basis of the current observation x and the training data $\left(\mathbf{z}_i, x_i\right), i = 1, ..., n.$

Let $\widehat{\boldsymbol{\beta}}$ be the least squares estimate of $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}}\left(t\right) = \left(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathsf{T}}\mathbf{x}\left(t\right),$$

and $\widehat{\sigma}_e$ be the usual estimator of σ_e based on the residuals:

$$\widehat{\sigma}_e = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \widehat{x}_i\right) \otimes \left(x_i - \widehat{x}_i\right),$$

where $\widehat{x}_i = \mathbf{z}_i^{\mathsf{T}}\widehat{\boldsymbol{\beta}}$. Since the training vector of observations $\mathbf{x}$ is a Gaussian process with distribution $N_{L_2^n}\left(\boldsymbol{Z\beta}, \mathbf{I}\sigma_e\right)$, where $\mathbf{I}$ is the $n \times n$ identity matrix, $\widehat{\boldsymbol{\beta}}$ is a Gaussian process with distribution $P_{\widehat{\boldsymbol{\beta}}}\left(\cdot; \boldsymbol{\beta}, \sigma_e\right) = N_{L_2^q}\left(\boldsymbol{\beta}, \left(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\right)^{-1}\sigma_e\right)$. Define the probability distribution

$$\mu_P\left(\cdot; \mathbf{z}, \boldsymbol{\beta}, \sigma_e\right) = \int P_X\left(\cdot; \mathbf{z}, \boldsymbol{\gamma}, \sigma_e\right) P_{\widehat{\boldsymbol{\beta}}}\left(d\boldsymbol{\gamma}; \boldsymbol{\beta}, \sigma_e\right),$$

where $P_X\left(\cdot; \mathbf{z}, \boldsymbol{\beta}, \sigma_e\right) = N_{L_2}\left(\mathbf{z}^{\mathsf{T}}\boldsymbol{\beta}, \sigma_e\right)$ is the distribution of the observation X according to the model. Thus, $\mu_P\left(\cdot; \mathbf{z}, \boldsymbol{\beta}, \sigma_e\right) = N_{L_2}\left(\mathbf{z}^{\mathsf{T}}\boldsymbol{\beta}, \left(1 + \mathbf{z}^{\mathsf{T}}\left(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\right)^{-1}\mathbf{z}\right)\sigma_e\right)$. Under mild conditions, this measure has a density $g\left(\cdot; \mathbf{z}, \boldsymbol{\beta}, \sigma_e\right)$ with respect to the measure P_e, which is given by

$$g\left(\cdot; \mathbf{z}, \boldsymbol{\beta}, \sigma_e\right) = \int f_X\left(\cdot; \mathbf{z}, \boldsymbol{\gamma}, \sigma_e\right) P_{\widehat{\boldsymbol{\beta}}}\left(d\boldsymbol{\gamma}; \boldsymbol{\beta}, \sigma_e\right),$$

where $f_X\left(\cdot; \mathbf{z}, \boldsymbol{\gamma}, \sigma_e\right) = dP_X\left(\cdot; \mathbf{z}, \boldsymbol{\gamma}, \sigma_e\right)/P_e$ is the density of $P_X\left(\cdot; \mathbf{z}, \boldsymbol{\gamma}, \sigma_e\right)$ with respect to P_e.

Let φ_l, λ_l be, respectively, the eigenfunctions and eigenvalues of the covariance function σ_e. From known results on equivalence of Gaussian measures [13] it follows that

$$g\left(\cdot; \mathbf{z}, \boldsymbol{\beta}, \sigma_e\right) = \frac{1}{\sqrt{1 + \mathbf{z}^{\mathsf{T}}\left(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\right)^{-1}\mathbf{z}}}\exp\left\{\frac{1}{1 + \mathbf{z}^{\mathsf{T}}\left(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\right)^{-1}\mathbf{z}}\sum_{l=1}^{\infty}\frac{\mathbf{z}^{\mathsf{T}}\boldsymbol{\beta}^l}{\lambda_l}\left(x^l - \frac{\mathbf{z}^{\mathsf{T}}\boldsymbol{\beta}^l}{2}\right)\right\},$$

where

$$\begin{aligned} \boldsymbol{\beta}^l &= \left(\langle \boldsymbol{\beta}_j, \varphi_l \rangle\right)_{1 \le j \le q}, \\ x^l &= \langle x, \varphi_l \rangle . \end{aligned}$$

Here, $\langle \cdot, \cdot \rangle$ denotes the inner product in $L_2\left([0,T]\right)$.

Note that $g\left(\cdot; \mathbf{z}, \boldsymbol{\beta}, \sigma_e\right)$ takes into consideration the uncertainty about $\boldsymbol{\beta}$ in determining the true density $f_X\left(\cdot; \mathbf{z}, \boldsymbol{\beta}, \sigma_e\right)$ through the estimate $\widehat{\boldsymbol{\beta}}$. On this basis it is possible, by extending to the functional setting the approach initiated in [12], to define a non Bayesian *predictive density* for the new observation x corresponding to $\mathbf{z}$ as the following empirical version of $g\left(\cdot; \mathbf{z}, \boldsymbol{\beta}, \sigma_e\right)$:

$$f_P\left(x; \mathbf{z}\right) = \frac{1}{\sqrt{1 + \mathbf{z}^\intercal \left(\mathbf{Z}^\intercal \mathbf{Z}\right)^{-1} \mathbf{z}}} \exp\left\{ \frac{1}{1 + \mathbf{z}^\intercal \left(\mathbf{Z}^\intercal \mathbf{Z}\right)^{-1} \mathbf{z}} \sum_{l=1}^{m} \frac{\mathbf{z}^\intercal \widehat{\boldsymbol{\beta}}^l}{\widehat{\lambda}_l} \left(x^l - \frac{\mathbf{z}^\intercal \widehat{\boldsymbol{\beta}}^l}{2} \right) \right\},$$

where $m = m\left(n\right)$ is a specified integer such that $m \to \infty$ as $n \to \infty$; $\widehat{\varphi}_k$, $\widehat{\lambda}_k$ are, respectively, the eigenfunctions and eigenvalues of the covariance function $\widehat{\sigma}_e$; and

$$\begin{aligned} \widehat{\boldsymbol{\beta}}^l &= \left(\left\langle \widehat{\boldsymbol{\beta}}_j, \widehat{\varphi}_l \right\rangle\right)_{1 \le j \le q}, \\ x^l &= \langle x, \widehat{\varphi}_l \rangle . \end{aligned}$$

We also define the non Bayesian *predictive likelihood* function by

$$L\left(\mathbf{z}\right) = f_P\left(x; \mathbf{z}\right),$$

and the non Bayesian *predictive estimator* $\widehat{\mathbf{z}}$ of $\mathbf{z}$ by

$$\widehat{\mathbf{z}}_P = \arg\max_{\mathbf{z}} L\left(\mathbf{z}\right).$$

In cases in which some components of the vector $\mathbf{z}$ are known, they are not considered in this maximization.

3 A Simulation Study

The feasibility and the performance of the introduced functional calibration method are here explored through a simulation study. For simplicity, we consider models with only one covariate z (i.e., $q = 1$), so they have the specific form:

$$x(t) = \beta(t) z + e(t).$$

The covariate values in the training data were generated following a normal distribution $z_i \sim N(15, 1.5)$, $i = 1, ..., n$. The size of the training sample was set to $n = 300$.

The Gaussian error process $e(t)$ was simulated with the covariance function

$$\sigma_e(s,t) = \sum_{i=1}^{200} \lambda_i \phi_i(s)\phi_i(t),$$

where $(\phi_i(t))_i$ is the trigonometric basis on $L_2([0,1])$ (i.e., $\phi_{2k-1} = \sqrt{2}\cos(2\pi kt)$, $\phi_{2k} = \sqrt{2}\sin(2\pi kt)$), and the eigenvalues were set to $\lambda_i = 0.06/(i^{1.01})$.

The coefficient functions were specified to be of the form $\beta(t) = Cg(t)\sin(2\pi t)$, where $g(t)$ is the density function $N(0.5, 0.05)$ (i.e., a peak) and C is a constant.

Different simulation settings were considered according to the signal (β) to noise (σ_e) ratio by varying the constant C in the mean function: "good", "moderate" and "bad" scenarios correspond to $C = 2.5$, $C = 1$ and $C = 0.2$, respectively. The means functions for different scenarios are shown in Figure 1.

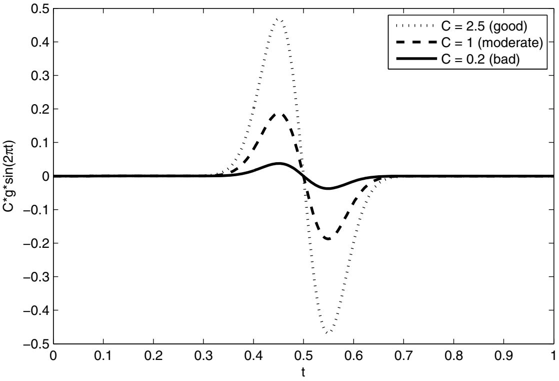

Fig. 1. Mean curves for the good, moderate and bad scenarios

Our interest is in methods with good performance for extrapolation, thus we focus comparisons with the classical estimator

$$\hat{z}_C = \frac{\left\langle x, \widehat{\beta} \right\rangle}{\left\| \widehat{\beta} \right\|^2}.$$

The Mean Square Error (MSE) $E(z - \widehat{z})^2$ is used as comparison criteria, which is computed on the basis of $B = 3000$ sample repetitions.

Figure 2 shows the MSE curves for both estimators and all the settings. It can be observed that, in all simulation settings, both estimators perform worse as the covariate value z goes away from the center of the calibration range. In the good setting the difference between the estimators is not so noticeable. As the setting becomes worse, differences between the estimators become remarkable. Inside the calibration range, the performance of both estimators is quite similar, but

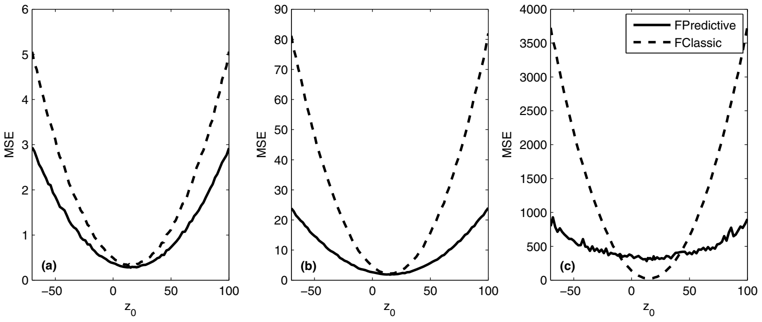

Fig. 2. Curves of RMSE as function of z for the estimators $\hat{z}_C$ and $\hat{z}_P$ in scenarios good(a), moderate (b) and bad (c)

outside such range (i.e., for extrapolation) the non Bayesian predictive estimator shows much less MSE than the classical estimator.

Figure 3 shows plots of mean values of the estimates (predictions) versus true values of the covariate in the bad scenarios for both estimators. It can be appreciated that the classical estimator is highly biased while the introduced estimator has negligible bias.

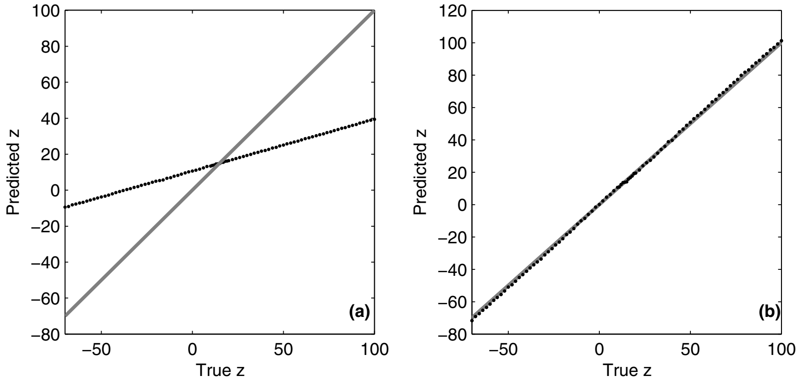

Fig. 3. Points represent predicted z values versus true z values in the bad setting for the (a) Classic estimator and the (b) Non Bayesian Predictive estimator. The diagonal line is also shown, for reference.

An attractive feature of the non Bayesian predictive approach for functional calibration is that it provides not only point estimates of the covariate z but also predictive likelihoods of all possible values of this parameter. This allows one to complement the point estimate $\hat{z}_P$ with a likelihood-based appreciation of the location of z by plotting the relative predictive likelihood curve. As an illustration, Figure 4 shows the relative predictive likelihood curves for two samples generated from the good and the bad setting. It can be observed that predictive

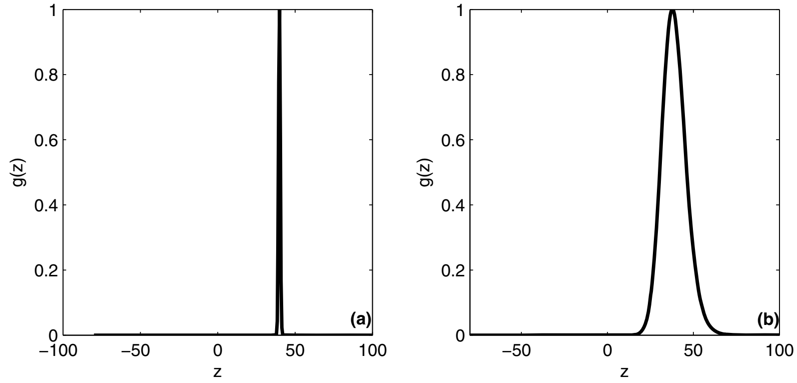

Fig. 4. Relative predictive likelihood functions for samples from the a) good and b) bad scenarios corresponding to the true value $z = 40$ of the covariate

likelihood curves show more precision in the assessment of z (i.e., they are more spiky) for good settings than for bad settings.

4 Conclusions

This papers introduces a non Bayesian predictive calibration approach for functional data. Its rationality comes from taking into consideration model uncertainty by averaging with respect to the distribution of the parameter estimator, a device that shows to have a regularizing effect in the resulting solution to the calibration (inverse) problem. By construction, the introduced method can be applied to both random and fixed designs. It shows negligible bias, and much better performance for extrapolation than the classical estimator. It also has the following advantageous feature that is lacking in previous Bayesian and non Bayesian approaches to calibration: it provides not only a point estimate but also a non Bayesian predictive likelihood function that can be used to assess the plausibility of any possible value of the covariate to be predicted. Finally, it is worth of note that the introduced approach can be extended to a wide variety of statistical models. In case of very complex models it might be implemented by means of a bootstrap approximation to the predictive density.

References

1. Osborne, C.: Statistical calibration: A review. Int. Stat. Rev. 59, 309–336 (1991)
2. Martens, H., Naes, T.: Multivariate calibration. Wiley, Chichester (1989)
3. Brown, P.J.: Measurement, Regression, and Calibration. Clarendon Press, Oxford (1993)
4. Massart, D.L., Vandeginste, B.G.M., Buydens, L., Jong, S.D., Lewi, P.J., Smeyers-Verbeke, J.: Handbook of Chemometrics and Qualimetrics: Part B. Elsevier Science B. V., The Netherlands (1997)

5. Lavine, B.K., Workman, J.: Fundamental reviews: Chemometrics. Anal. Chem. 74, 2763–2770 (2002)
6. Walters, F.H., Rizzuto, G.T.: The calibration problem in statistics and its application to chemistry. Anal. Lett. 21, 2069–2076 (1988)
7. Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis: Theory and Practice (Springer Series in Statistics). Springer, New York (2006)
8. Ramsay, J., Silverman, B.: Functional Data Analysis. Springer, Berlin (1997)
9. Cuevas, A., Febrero, M., Fraiman, R.: Linear functional regression: the case of fixed design and functional response. The Canadian Joumal of Statistics 30, 285–300 (2002)
10. Hagwood, C.: The calibration problem as an ill-posed inverse problem. J. Stat. Plan. Infer. 31, 179–185 (1992)
11. Hernández, N., Biscay, R.J., Talavera, I.: A non-bayesian predictive approach for statistical calibration. Journal of Statistical Computation and Simulation 82, 529–545 (2012)
12. Harris, I.R.: Predictive fit for natural exponential families. Biometrika 76, 675–684 (1989)
13. Grenander, U.: Abstract Inference. Wiley, New York (1981)