Probabilistic prediction: aims and solutions

Giovanni Fonseca[1], Federica Giummolè[2], Paolo Vidoni[1]

[1] Dept. of Economics and Statistics, University of Udine, Udine, Italy
[2] Dept. of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice, Italy

E-mail for correspondence: giovanni.fonseca@uniud.it

Abstract: The goodness of a predictive distribution depends on the aim of the prediction. This presentation intends to shed light on properties of predictive distributions in use nowadays. We also propose a new predictive distribution that may be useful to obtain calibrated predictions for the probabilities of a future random variable of interest. This predictive distribution can be easily computed by a simple bootstrap procedure. In order to compare the different predictive distributions, some simulation studies are also presented.

Keywords: Bootstrap, Calibration, Prediction.

1 Introduction

Let us define the notation and the general assumptions that we will use in the sequel. Suppose that $\{Y_i\}_{i\geq 1}$ is a sequence of continuous random variables with probability distribution depending on an unknown d-dimensional parameter $\theta \in \Theta \subseteq \mathbf{R}^d$, $d \geq 1$; $Y = (Y_1, \cdots, Y_n)$, $n > 1$, is observable, while $Z = Y_{n+1}$ is a future or not yet available observation. For simplicity, we consider the case of Y and Z being independent random variables and we indicate with $G(z;\theta)$ and $Q(\alpha;\theta)$ the distribution function and the quantile function of Z, respectively. Given the observed sample $y = (y_1, \ldots, y_n)$, we look for a predictive distribution $\hat{G}(z;y)$, with corresponding quantile function $\hat{Q}(\alpha;y)$, that fullfills some good requirements for prediction.
There are different desirable properties that a predictive distribution should possess. Here we consider only two of the most important:

(A) calibrated quantile function: $E_Y[G\{\hat{Q}(\alpha;Y);\theta\}] = \alpha$, $\forall \alpha \in (0,1)$

(B) calibrated distribution function: $E_Y[Q\{\hat{G}(z;Y);\theta\}] = z$, $\forall z \in \mathbf{R}$.

Unfortunately, these properties cannot be satisfied at the same time. They regard different aspects of a predictive distribution and depend on the target of the prediction itself. A quantile function is calibrated if, in mean, it coincides with the inverse of the true distribution function. This last property can also be expressed in terms of coverage probabilities, since

$$E_Y[G\{\hat{Q}(\alpha;Y);\theta\}] = P_{Z,Y}\{Z \leq \hat{Q}(\alpha;Y)\}.$$

Similarly, a predictive distribution function is calibrated if, in mean, it coincides with the inverse of the true quantile function. While from a theoretical point of view the knowledge of the distribution function coincides with that of the quantile function for continuous random variables, this is not true when we talk about predictive distributions. Thus, a predictive distribution may be good for estimating quantiles but not as good for estimating probabilities and the converse is also true.

In the sequel we always consider the maximum likelihood estimator (mle) $\hat{\theta} = \hat{\theta}(Y)$ for θ, or an asymptotically equivalent alternative. The estimative predictive distribution and quantile functions, $G(z;\hat{\theta})$ and $Q(\alpha;\hat{\theta})$ respectively, usually satisfy properties (A) and (B) with an error term of order $O(n^{-1})$, as the sample size $n \to +\infty$, see e.g. Barndorff-Nielsen and Cox (1996). It is well known that this error term could be substantial, in particular for small sample sizes.

2 Calibrated quantile functions

Modern literature has largely focused on the problem of prediction limits, that is the problem of finding a predictive distribution which quantiles satisfy property (A) with a high approximation. This requirement is usually met when a pivotal quantity for prediction is available. Unfortunately in many situations of practical interest, a pivot is not known. Furthermore, even in the case of the normal distribution, sometimes the unknown parameters are estimated using ad hoc estimators whose exact distribution is unknown. As a consequence, the distribution of a quantity such as $(Z - \hat{\mu})/\hat{\sigma}$ is not known. Thus, it becomes of interest in the applications to find alternative approximate solutions.

Here we quickly recall the procedure used in Fonseca et al. (2014), since we will follow the same steps in the next section. The starting point is the coverage probability associated to the estimative quantile function $Q(\alpha;\hat{\theta})$:

$$P_{Y,Z}\{Z \leq Q(\alpha;\hat{\theta});\theta\} = E_Y[G\{Q(\alpha;\hat{\theta});\theta\}] = C(\alpha,\theta).$$

Although an explicit expression of this coverage probability is rarely available, it is well-known that it does not match the target value α. Fonseca et al. (2014) noticed that the function $G_c(z;\hat{\theta},\theta) = C\{G(z;\hat{\theta}),\theta\}$, obtained by substituting α with $G(z;\hat{\theta})$ in $C(\alpha,\theta)$, is a proper predictive distribution function, whose associated quantile function is calibrated, giving coverage probability equal to the target nominal value α, for all $\alpha \in (0,1)$. A suitable parametric bootstrap estimator for $G_c(z;\hat{\theta},\theta)$ may be readily defined as

$$G_c^{boot}(z;\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B} G\{Q(\alpha;\hat{\theta}^b);\hat{\theta}\}|_{\alpha=G(z;\hat{\theta})},$$

where $\hat{\theta}^b$, $b=1,\dots,B$, are estimates obtained with B bootstrap samples from $G(z;\hat{\theta})$. The corresponding α-quantile defines, for each $\alpha \in (0,1)$, a prediction limit having coverage probability equal to the target α, with an error term which depends on the efficiency of the bootstrap simulation procedure.

3 Calibrated distribution functions

In this section we address the dual problem, looking for predictive distributions that satisfy property (B). We use exactly the same ideas proposed by Fonseca et al. (2014) and recalled in the previous section, applied to the distribution function instead of the quantile function. The result is a new predictive distribution that may be useful for predicting probabilities for the interest variable Z, instead of quantiles.
The estimative distribution function is not well calibrated in the sense of property (B). Infact, the mean of quantiles of level equal to $G(z;\hat{\theta})$ is

$$E_Y[Q\{G(z;\hat{\theta});\theta\}] = A(z,\theta)$$

and it does not match the target value z. Instead, the function

$$Q_c(\alpha;\hat{\theta},\theta) = A\{Q(\alpha;\hat{\theta}),\theta\}, \qquad (1)$$

obtained by substituting z with $Q(\alpha;\hat{\theta})$ in $A(z,\theta)$, is a proper predictive quantile function whose distribution function $G_c(z;\hat{\theta},\theta) = G\{A^{-1}(z,\theta);\hat{\theta}\}$ satisfies property (B) for every $z \in \mathbf{R}$. Indeed,

$$\begin{aligned} E_Y[Q\{G_c(z;\hat{\theta},\theta);\theta\}] &= E_Y[Q\{G(A^{-1}(z,\theta);\hat{\theta});\theta\}] \\ &= A\{A^{-1}(z,\theta),\theta\} = z. \end{aligned}$$

The predictive quantile function (1) and the corresponding calibrated predictive distribution are not useful in practice, since they depend on the unknown parameter θ. However, a suitable parametric bootstrap estimator for $Q_c(\alpha;\hat{\theta},\theta)$ may be readily defined. Let y^b, $b=1,\dots,B$, be parametric bootstrap samples generated from the estimative distribution of the data and let $\hat{\theta}^b$, $b=1,\dots,B$, be the corresponding estimates. We can thus write

$$Q_c^{boot}(\alpha;\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B} Q\{G(z;\hat{\theta}^b);\hat{\theta}\}|_{z=Q(\alpha;\hat{\theta})}.$$

The corresponding distribution function allows to predict the target probability $P(Z \leq z_0)$, for each $z_0 \in \mathbf{R}$, with an error term which depends on the efficiency of the bootstrap simulation procedure. Indeed, the estimate is the value α_0 such that $Q_c^{boot}(\alpha_0; \hat{\theta}) = z_0$.

4 The normal distribution

Let $Y_1, \ldots, Y_n, Z$ be independent and normally distributed with mean μ and standard deviation σ, both unknown. In this context the pivotal quantity $T = \sqrt{n/(n+1)}(Z - \bar{Y})/S$ is useful for prediction, with $\bar{Y}$ and S the sample mean and sample standard deviation, respectively. Its distribution is Student t with $n-1$ degrees of freedom. The quantile function $\sqrt{(n+1)/n}\, Q_t\,(\alpha; n-1)\, S + \bar{Y}$, satisfies property (A). Hence, in this case, the calibrated quantile function presented in section 2 replicates the distribution obtained from the pivot. However, as shown in the following simulation study, the pivotal distribution is not the best choice for prediction of probabilities.
The following tables show the results of Monte Carlo simulations based on $M = 10000$ replications and $B = 500$ bootstrap replications for the computation of the calibrated distributions. The sample size is $n = 10, 25$ and the true parameter values are $\mu = 0$ and $\sigma = 1$. We have compared the estimative distribution with the mle, the predictive distribution obtained from the pivotal quantity and the two bootstrap calibrated predictive distributions on the basis of the corresponding coverage probability for $\alpha = 0.5, 0.9, 0.95, 0.99, 0.999$ (Table 1) and the mean quantiles of levels $\hat{G}(z; y)$ for $z = 0, 1.5, 2, 2.5, 3.5$ (Table 2). The best performances are written in bold face, clearly showing how the aim of the prediction should influence on the choice of the predictive distribution.

TABLE 1. Coverage probabilities. Standard errors smaller than 0.001.

	Target	estim.	pivotal	qu. calib.	pr. calib.
n=10	$\alpha = 0.5$	**0.500**	**0.500**	**0.500**	**0.500**
	$\alpha = 0.9$	0.861	**0.900**	0.899	0.892
	$\alpha = 0.95$	0.914	**0.950**	0.949	0.939
	$\alpha = 0.99$	0.967	**0.990**	**0.990**	0.981
	$\alpha = 0.999$	0.989	**0.999**	**0.999**	0.995
n=25	$\alpha = 0.5$	**0.500**	**0.500**	**0.500**	**0.500**
	$\alpha = 0.9$	0.885	**0.900**	**0.900**	0.897
	$\alpha = 0.95$	0.936	**0.950**	**0.950**	0.946
	$\alpha = 0.99$	0.983	**0.990**	**0.990**	0.987
	$\alpha = 0.999$	0.997	**0.999**	**0.999**	0.998

TABLE 2. Mean quantiles of level $\hat{G}(z; y)$. Standard errors smaller than 0.001

	Target	estim.	pivotal	qu. calib.	pr. calib.
n=10	$z = 0$	-0.001	-0.001	**0.000**	**0.000**
	$z = 1.5$	1.734	1.411	1.411	**1.504**
	$z = 2$	2.312	1.803	1.804	**2.004**
	$z = 2.5$	2.889	2.151	2.153	**2.505**
	$z = 3.5$	4.044	2.732	2.741	**3.498**
n=25	$z = 0$	**0.000**	**0.000**	**0.000**	**0.000**
	$z = 1.5$	1.581	1.465	1.465	**1.500**
	$z = 2$	2.108	1.920	1.920	**2.000**
	$z = 2.5$	2.635	2.350	2.350	**2.500**
	$z = 3.5$	3.689	3.130	3.133	**3.500**

References

Barndorff-Nielsen, O.E. and Cox, D.R. (1996). Prediction and asymptotics. *Bernoulli*, **2**, 319 – 340.

Fonseca, G., Giummolè F. and Vidoni, P. (2014). Calibrating predictive distributions. *Journal of Statistical Computation and Simulation*, **84**, 373 – 383.