Confidence distributions for predictive tail probabilities

Distribuzioni di confidenza per probabilità di previsione sulle code

Giovanni Fonseca, Federica Giummolè and Paolo Vidoni

Abstract In this short paper we propose the use of a calibration procedure in order to obtain predictive probabilities for a future random variable of interest. The new calibration method gives rise to a confidence distribution function which probabilities are close to the nominal ones to a high order of approximation. Moreover, the proposed predictive distribution can be easily obtained by means of a bootstrap simulation procedure. A simulation study is presented in order to assess the good properties of our proposal. The calibrated procedure is also applied to a series of real data related to sport records, with the aim of closely estimate the probability of future records.
Abstract *In questo lavoro proponiamo l'utilizzo di una procedura di calibrazione per determinare probabilità predittive per una variabile futura di interesse. Il metodo proposto fornisce distribuzioni di confidenza le cui probabilità si avvicinano a quelle vere con un buon ordine di approssimazione. Le distribuzioni predittive proposte si possono ottenere facilmente attraverso una procedura di bootstrap. Un primo studio di simulazione mostra le buone proprietà delle distribuzioni predittive ottenute. Il nuovo metodo viene anche applicato all'analisi di un insieme di dati reali riguardanti record sportivi, con lo scopo di stimare la probabilità di un nuovo record mondiale.*

Key words: athletic records, asymptotics, bootstrap, calibration, confidence distributions, generalised extreme value distribution, prediction.

Giovanni Fonseca
University of Udine, Via Tomadini 30/A, 33100 Udine (UD), Italy, e-mail: giovanni.fonseca@uniud.it

Federica Giummolè
Ca' Foscari University Venice, Via Torino 155, 30172 Mestre (VE), Italy, e-mail: giummole@unive.it

Paolo Vidoni
University of Udine, Via Tomadini 30/A, 33100 Udine (UD), Italy, e-mail: paolo.vidoni@uniud.it

1 Introduction

Consider the problem of predicting the value of a future or not yet observed random variable, using a sample generated by the same random mechanism. In the frequentist approach, prediction usually requires the specification of a suitable estimate for the (conditional) distribution of the interest random variable, based on the available data, which can be viewed as a confidence distribution ([4], [6]). In particular, this predictive distribution is considered for defining prediction intervals or more simply prediction quantiles, requiring that the associated coverage probability corresponds, exactly or approximately, to the prescribed target probability. Several papers have addressed this problem and, in particular, we mention the calibration approach introduced in [1], and the related bootstrap-based procedure proposed in [3]. In this paper we focus on the different, albeit related, problem of defining a predictive distribution giving well calibrated probabilities for the future random variable. The bootstrap calibration procedure, introduced for the quantiles, is applied in this dual framework, giving a new calibrated distribution in order to obtain predictive probabilities. This new proposal is briefly compared with the existing ones by considering an example involving normal distributed samples. Finally, a real data application, related to sport records and based on the GEV distribution, is presented.

2 Calibrated distributions for prediction probabilities

Let us define the notation and the general assumptions that we require for obtaining the result. Suppose that $\{Y_i\}_{i\geq 1}$ is a sequence of continuous random variables with probability distribution specified by the unknown d-dimensional parameter $\theta \in \Theta \subseteq \mathbf{R}^d$, $d \geq 1$; $Y = (Y_1, \ldots, Y_n)$, $n > 1$, 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.

As far as we know, modern literature has mainly focused on the problem of finding a predictive distribution which quantiles satisfy

$$E_Y\{G(\hat{Q}(\alpha;Y);\theta)\} = \alpha, \tag{1}$$

for all $\alpha \in (0,1)$, at least with a high approximation. In this work, we concentrate on the dual problem, that is finding a predictive distribution function $\hat{G}(z;y)$ such that, exactly or approximately,

$$E_Y\{Q(\hat{G}(z;Y);\theta)\} = z, \tag{2}$$

for every $z \in \mathbf{R}$. As it can be noted, instead of assessing the quantile function of Z we are trying to estimate the distribution function itself. In order to solve this problem, we simply apply the same procedure proposed by [3] to the quantile function $Q(\alpha;\theta)$ of Z instead of the distribution function itself. This easily lead to the definition of a new calibrated predictive distribution that may be useful for the calculation of probabilities for Z.

Consider the maximum likelihood estimator $\hat{\theta} = \hat{\theta}(Y)$ for θ, or an asymptotically equivalent alternative, and the estimative predictive distribution and quantile function, $G(z;\hat{\theta})$ and $Q(\alpha;\hat{\theta})$, respectively. 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, although its explicit expression is rarely available, it is well-known that it does not match the target value z even if, asymptotically, $A(z,\theta) = z + o(1)$, as $n \to +\infty$. It is easy to see that the function

$$Q_c(\alpha;\hat{\theta},\theta) = A\{Q(\alpha;\hat{\theta}),\theta\}, \tag{3}$$

which is obtained by substituting z with $Q(\alpha;\hat{\theta})$ in $A(z,\theta)$, is a proper quantile function, provided that $A(\cdot,\theta)$ is sufficiently smooth. Furthermore, the corresponding distribution function $G_c(z;\hat{\theta},\theta) = G\{A^{-1}(z,\theta);\hat{\theta}\}$ satisfies (2) 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 calibrated predictive quantile function (3) and the corresponding 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, $b = 1,\ldots,B$, be parametric bootstrap samples generated from the estimative distribution of the data and let $\hat{\theta}^b$, $b = 1,\ldots,B$, be the corresponding estimates. We can thus write

$$Q_c^{boot}(\alpha;\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B} Q\{G(z;\hat{\theta}^b);\hat{\theta}\}|_{z=Q(\alpha;\hat{\theta})}. \tag{4}$$

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

3 The Normal distribution: a simulation study

Let us first consider the case of prediction for a normally distributed random variable. If we use $\bar{Y} = \sum_i Y_i/n$ and $S = \sqrt{\sum_i (Y_i - \bar{Y})^2/(n-1)}$ as estimators for the unknown parameters, then $T = \sqrt{n/(n+1)}(Z - \bar{Y})/S$ is a pivotal quantity having a Student t distribution with $n-1$ degrees of freedom. Its quantiles satisfy (1) exactly. In spite of this, it could be also interesting to consider the calibrated procedure proposed in [3], which satisfies (1) approximately. Indeed, in some situations the sample mean and standard deviation may not be the most convenient estimators for the parameters and, thus, a pivotal quantity may not be easily available.

In the following we compare the estimative distribution function (Est), the exact distribution function obtained from the pivotal quantity (Piv), the quantile calibrated distribution function of [3] (Qcal) and our proposal, that we name probability calibrated distribution function (Pcal). Figure 1 represents an example of the different predictive distributions obtained from a particular sample y.

We have performed a simulation study in order to assess the properties of the different predictive distributions. Tables 1 and 2 show the results of a Monte Carlo simulation based on $M = 1000$ replications. The bootstrap procedure is based on $B = 500$ replications. The sample size is $n = 10$ and the true parameter values are $\mu = 0$ and $\sigma = 1$. We have compared the different predictive distributions on the basis of the corresponding coverage probability for $\alpha = 0.9, 0.95, 0.99$ (Table 1) and the mean quantiles of levels $\hat{G}(z; y)$ for $z = 1.5, 2, 2.5$ (Table 2). As expected, the pivotal and the quantile calibrated predictive distributions perform better with respect to criterion (1) whereas the probability calibrated predictive distribution outperforms the others with respect to criterion (2).

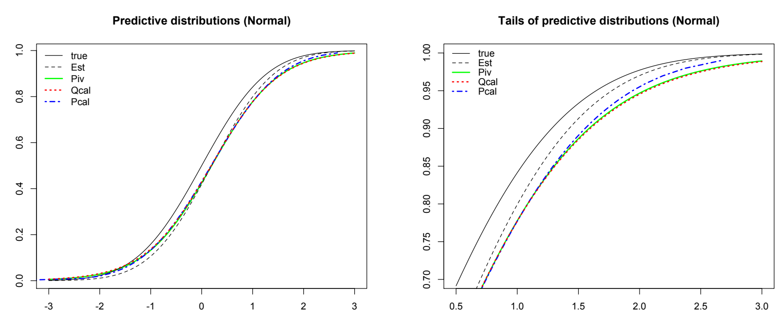

Fig. 1 Normal case: predictive distribution functions (left) and upper tails of predictive distribution functions (right).

Target	Piv	Est	Qcal	Pcal
$\alpha = 0.9$	0.902	0.876	0.902	0.895
$\alpha = 0.95$	0.946	0.919	0.945	0.936
$\alpha = 0.99$	0.990	0.972	0.990	0.982

Table 1 Normal model: coverage probabilities (standard errors are always smaller than 0.0025).

Target	Piv	Est	Qcal	Pcal
$z = 1.5$	1.40	1.63	1.40	1.47
$z = 2$	1.83	2.24	1.83	1.96
$z = 2.5$	2.16	2.75	2.16	2.30

Table 2 Normal model: mean quantiles of level $\hat{G}(z;y)$ (standard errors are always smaller than 0.025).

4 The GEV distribution: an application to athletic records

As a further example, we consider the case of prediction for the generalised extreme value (GEV) distribution, which is usually applied to model maxima of a process over certain time intervals; see for instance [2]. The GEV distribution has three parameters: location, scale and shape. It is important noticing that when the shape parameter is positive (Fréchet distribution) or equal to 0 (Gumbel distribution) the support of the distribution is not limited from above. We have collected annual records in the period 2001 to 2019 for female long jump from the web site of the World Athletics (formerly known as International Association of Athletics Federations (IAAF)) [5].

Using the proposed probability calibrated predictive distribution, we can properly compute probabilities related to the variable Z which represents the best performance in the year to come. In particular we can evaluate the probability of having a new world record in the next year as $\alpha_{WR} = P(Z > WR)$, where WR represents the present world record. This probability can also be used to evaluate the goodness of the world record: the smaller α_{WR} the better the world record. Moreover, from α_{WR} we can calculate the expected number of years for the next record, $T_{WR} = 1/\alpha_{WR}$.

In our example the estimate of the shape parameter of the GEV distribution is positive, thus the estimative GEV distribution function is a Fréchet distribution with no upper bound. Though, the confidence interval for the shape parameter includes 0 and hence, from an inferential point of view, the specification procedure indicates the Gumbel model as the obvious candidate. However, prediction can be quite affected by such a choice, as it can be seen from the results presented in Table 3. Figure 2 shows the estimative (solid), the quantile (dashed) and the probability (dotted) calibrated GEV (red) and Gumbel (black) distribution functions for women's long jump data. The bootstrap procedures are based on 1000 replications. The present world record (solid) is also represented.

The present world record, $WR = 7.52$ m, dates back to 1988 and is not included in the data. Using the GEV probability calibrated distribution instead of the Gumbel one, we take into account for the uncertainty related to the shape parameter

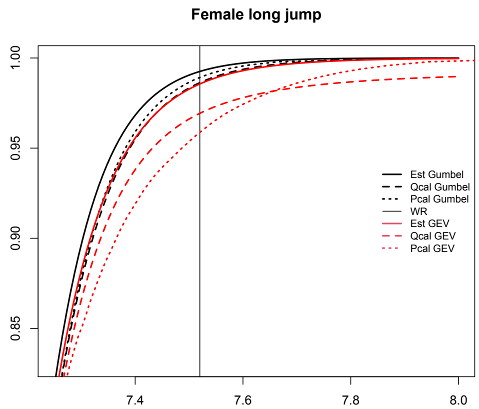

Fig. 2 Women's long jump: GEV and Gumbel predictive distribution functions.

estimation and we can properly assess the probability of improving the current world record: $\alpha_{WR} = P(Z > WR) = 0.041$. Notice also that both the GEV estimative and quantile calibrated predictive distributions wrongly estimate this probability to 0.014 and 0.031, respectively. The expected time for improving the current world record is about 24.4 years.

$WR = 7.52$	Est Gumbel	Qcal Gumbel	Pcal Gumbel	Est GEV	Qcal GEV	Pcal GEV
Probability	0.007	0.014	0.011	0.014	0.031	0.041
Expected time	134.1	73	90.9	69.5	32.5	24.4

Table 3 Probabilities of improving the current world record (WR) with corresponding mean waiting times.

Acknowledgements This research is partially supported by the Italian Ministry for University and Research under the PRIN2015 grant No. 2015EASZFS 003.

References

1. R. Beran. Calibrating prediction regions. *Journal of the American Statistical Association*, 85:715–723, 1990.
2. S. Coles. *An introduction to statistical modeling of extreme values*. Springer-Verlag, London, 2001.
3. G. Fonseca, F. Giummolè, and P. Vidoni. Calibrating predictive distributions. *Journal of Statistical Computation and Simulation*, 84:373–383, 2014.
4. N.L. Hjort and T. Schweder. Confidence distributions and related themes. *Journal of Statistical Planning and Inference*, 195:1–13, 2018.
5. World Athletics (I.A.A.F.). *https://www.worldathletics.org/*.
6. J. Shen, R. Liu, and M. Xie. Prediction with confidence: A general framework for prediction. *Journal of Statistical Planning and Inference*, 195:126–140, 2018.