Mathematical Properties of Continuous Ranked Probability Score Forecasting

Forecast Verification and Data Assimilation Workshop

Romain Pic[1], Clément Dombry[1], Philippe Naveau[2] and Maxime Taillardat[3]

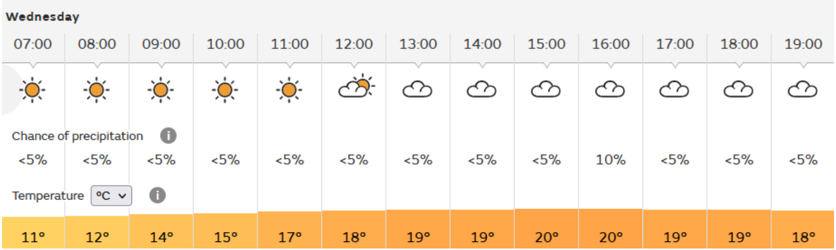
[1]Laboratoire de Mathématiques de Besançon, Université de Bourgogne Franche-Comté

[2]Laboratoire des Sciences du Climat et de l'Environnement, Université de Versailles Saint-Quentin

[3]Centre National de Recherches Météorologiques, Météo France

Table of Contents

Met Office


Wednesday
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
Chance of precipitation
<5%
<5%
<5%
<5%
<5%
<5%
<5%
<5%
<5%
10%
<5%
<5%
<5%
Temperature
°C
11°
12°
14°
15°
17°
18°
19°
19°
20°
20°
19°
19°
18°

Uncertainty in Forecasting

Ernest Cooke (MWR, 1906)

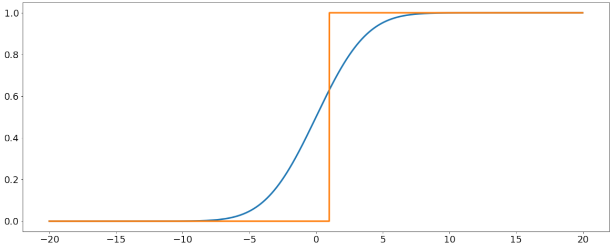
All those whose duty it is to issue regular daily forecasts know that there are times when they feel **very confident** and other times when they are **doubtful** as to coming weather. It seems to me that the condition of confidence or otherwise forms a **very important part of the prediction**.

- Various approaches :
  - Point forecasting (e.g. mean, quantile...)
  - Ensemble forecasting
  - **Distribution forecasting** : cumulative distribution function, density, quantile function, copula...
- Evaluation :
  - for point forecasting : squared error for the mean, pinball loss for the quantile.
  - for probabilistic forecasting : ?
- **Question :**
  How can we compare a distribution and an observation? $\rightarrow$ **Scoring Rules** [Gneiting and Katzfuss, 2014]

Continuous Ranked Probability Score

- Continuous Ranked Probability Score (CRPS) : [Matheson and Winkler, 1976]

$$\mathrm{CRPS}(F, y) = \int_{\mathbb{R}} (F(z) - \mathbb{1}_{y \leq z})^2 \mathrm{d}z$$



- Expected CRPS : $\overline{\mathrm{CRPS}}(F, G) = \mathbb{E}_{Y \sim G}\left[CRPS(F, Y)\right]$

$$\overline{\mathrm{CRPS}}(F, G) - \overline{\mathrm{CRPS}}(G, G) = \int_{\mathbb{R}} (F(z) - G(z))^2 \mathrm{d}z$$

$$\overline{\mathrm{CRPS}}(F, G) \geq \overline{\mathrm{CRPS}}(G, G) \text{ (strictly proper)}$$

Table of Contents

Distributional Regression

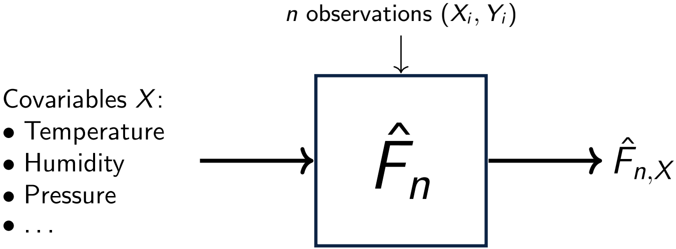
- $Y \in \mathbb{R}$ variable of interest, $X \in \mathbb{R}^d$ covariables with $(X, Y) \sim P$.
- Forecaster's goal : estimate the conditional distribution of Y given X, noted F_X^*.

Covariables X:
- Temperature
- Humidity
- Pressure
- ...

$\longrightarrow$ Model $\longrightarrow \hat{F}_X$

- Verification with the CRPS

- In practice, the model can be fitted on a **training sample** $(X_i, Y_i)_{1 \leq i \leq n}$ assumed i.i.d. following P.



Main Questions

$$R_P(\hat{F}_n) = \mathbb{E}_{(X_i,Y_i)\sim P}\mathbb{E}_{(X,Y)\sim P}\left[\mathrm{CRPS}(\hat{F}_{n,X}, Y)\right]$$
$$R_P(F^*) = \mathbb{E}_{(X,Y)\sim P}\left[\mathrm{CRPS}(F^*_X, Y)\right]$$

- Since the CRPS is strictly proper, $R_P(\hat{F}_n) \geq R_P(F^*)$ with equality if $\hat{F}_n = F^*$.

Questions :

- Consistency $R_P(\hat{F}_n) \to R_P(F^*)$ for large sample sizes ?
- **Best achievable rate of convergence?**

In point regression : need to restrain to a class of distributions to obtain non-trivial results on the rate of convergence.

- What definition of convergence do we choose?
- Minimization of the maximal error on a class of distributions. (minimax error)

Definition

A sequence of positive numbers $(\boldsymbol{a_n})$ is called an **optimal minimax rate of convergence** on the class $\mathcal{D}$ if

- (lower bound) any algorithm $\hat{F}_n$ satisfies $\sup_{P \in \mathcal{D}}(R_P(\hat{F}_n) - R_P(F^*)) \geq \epsilon a_n$ for $\epsilon > 0$ and n large enough.
- (upper bound) there exists an algorithm $\hat{F}_n$ satisfying $\sup_{P \in \mathcal{D}}(R_P(\hat{F}_n) - R_P(F^*)) < \epsilon^{-1} a_n$ for $\epsilon > 0$ and n large enough.

Class of Distributions

Class of distributions defined by the following conditions :

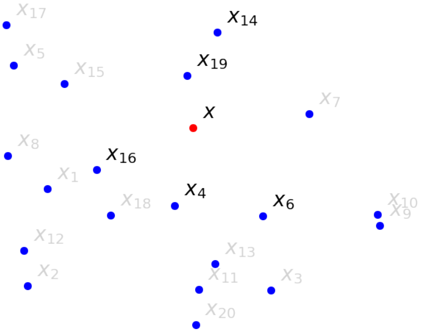
i) Covariables condition : $X \in [0,1]^d$ P_X-a.s.;
$\rightarrow$ More generally a compact.

ii) Regression condition : For all $x \in [0,1]^d$, $\int_{\mathbb{R}} F_x^*(z)(1 - F_x^*(z))\mathrm{d}z \leq M$;
$\rightarrow$ The dispersion of $Y|X = x$ remains bounded for all $x \in [0,1]^d$.

iii) CRPS condition : $\|F_{x'}^* - F_x^*\|_{L^2} \leq C\|x' - x\|^h$ for all $x, x' \in [0,1]^d$.
Using knowledge from previous observations at $X = x_i$ to extrapolate the value at $X = x$ $\rightarrow$ Need regularity of F^*.

• (lower bound) Use a subclass with a binary response to obtain a **lower minimax rate of convergence** : $\boldsymbol{a_n = n^{-\frac{2h}{2h+d}}}$. [Györfi et al., 2002]

k-Nearest Neighbors



$$\hat{F}_{n,x}(z) = \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{1}_{y_{i:n}(x) \leq z}, \quad {}_{i:n}(x) \text{ index of the } i\text{-th nearest neighbor of } x.$$

- **Remark :** k-NN are a type of **Analog Method**. [Delle Monache et al., 2013]

Main Results

- (lower bound) Use a subclass with a binary response to obtain a **lower minimax rate of convergence** : $\boldsymbol{a_n = n^{-\frac{2h}{2h+d}}}$. [Györfi et al., 2002]

Proposition

Assume $P \in \mathcal{D}^{(h,C,M)}$ and let $\hat{F}_n$ be the k-NN model. Then, for $d \geq 2$,

$$R_P(\hat{F}_n) - R_P(F^*) \leq {c_d}^h C^2 \left(\frac{k_n}{n}\right)^{2h/d} + \frac{M}{k_n}$$

where $c_d = \frac{2^{3+\frac{2}{d}}(1+\sqrt{d})^2}{V_d^{2/d}}$ and V_d is the volume of the unit ball in $\mathbb{R}^d$.

Theorem

For $d \geq 2$, the optimal minimax rate of convergence on the class $\mathcal{D}^{(h,C,M)}$ is $\boldsymbol{a_n = n^{-\frac{2h}{2h+d}}}$. Moreover, the k-NN algorithm reaches the optimal rate of convergence for

$$k_n = \left(\frac{Md}{2hC^2c_d^h}\right)^{\frac{d}{2h+d}} n^{\frac{2h}{2h+d}}.$$

Conclusion

- Optimal minimax rate of convergence for distributional regression in any $d \geq 2$.
- What happens in $d = 1$? **Kernel methods** reach this optimal minimax rate of convergence in any d.
- Not only methods based on the minimization of the CRPS, also methods using the CRPS for verification.
- Upper bound on the convergence rate for k-NN (and kernel methods) at fixed n.
- Extension to usual weighted CRPSs.
- Perspectives :
  - Study other algorithms : Random Forests (e.g. QRF [Taillardat et al., 2016]).
  - Study other definitions of convergence : other distances.
  - Adapt other classical results to the distributional regression framework.

Preprint : Mathematical Properties of Continuous Ranked Probability Score Forecasting, Pic et al. (`https://arxiv.org/abs/2205.04360`)

References I

Delle Monache, Luca et al. (2013). "Probabilistic weather prediction with analog ensemble". In: *Monthly Weather Review* 141, pp. 3498–3516. DOI: https://doi.org/10.1175/MWR-D-12-00281.1.

Gneiting, Tilmann and Matthiass Katzfuss (2014). "Probabilistic Forecasting". In: *Annual Review of Statistics and its Applications*. DOI: 10.1146/annurev-statistics-062713-085831.

Györfi, Lászlò et al. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics. Springer.

Matheson, James E. and Robert L. Winkler (1976). "Scoring Rules for Continuous Probability Distributions". In: *Management Science* 22 (10). DOI: 10.2307/2629907.

Taillardat, Maxime et al. (2016). "Calibrated Ensemble Forecasts using Quantile Regression Forests and Ensemble Model Output Statistics.". In: *Monthly Weather Review* 144 (6). DOI: 10.1175/MWR-D-15-0260.1.

Simulations

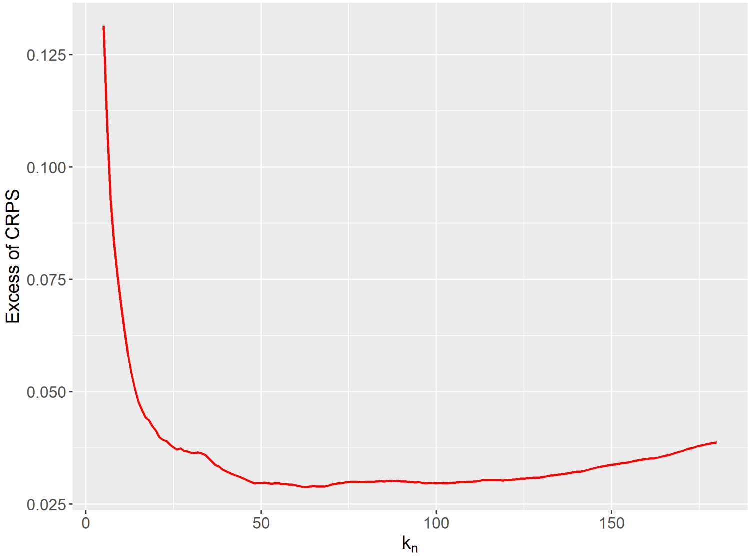
- $X_1, X_2 \sim \mathcal{U}([0,1])$
- $Y = X_1 + X_2 + \sigma\epsilon$ with $\epsilon \sim \mathcal{N}(0,1)$
- $Y|X \sim \mathcal{N}(X_1 + X_2, \sigma^2)$
- Checking the conditions :
  - **i)** $\boldsymbol{X \in [0,1]^d}$ $\boldsymbol{P_X}$**-a.s.;** ✓
  - **ii)** **For all** $\boldsymbol{x \in [0,1]^d}$**,** $\boldsymbol{\int_{\mathbb{R}} F_x^*(z)(1 - F_x^*(z))\mathrm{d}z \leq M}$**;**
    $\rightarrow M = \frac{\sigma}{\sqrt{\pi}}$ ✓
  - **iii)** $\boldsymbol{\|F_{x'}^* - F_x^*\|_{L^2} \leq C\|x' - x\|^h}$ **for all** $\boldsymbol{x, x' \in [0,1]^d}$.
    $\rightarrow$ Hard to get optimal values for C and h but $h = 1$ works. ✓
- k-NN :

$$\hat{F}_{n,x}(z) = \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{1}_{y_{i:n}(x) \leq z}$$

$$\overline{\mathrm{CRPS}}(F_{n,x}, F_x^*) - \overline{\mathrm{CRPS}}(F_x^*, F_x^*) = \int_{\mathbb{R}} \left(\frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{1}_{y_{i:n}(x) \leq z} - \Phi\left(\frac{z - (x_1 + x_2)}{\sigma} \right) \right)^2 \mathrm{d}z$$

CRPS vs. k_n

Parameters : $\sigma = 1$ and $n = 200$.



Scaling of k_n with n, $\sigma = 2$

$$k_n \propto n^{\frac{2h}{2h+d}}$$

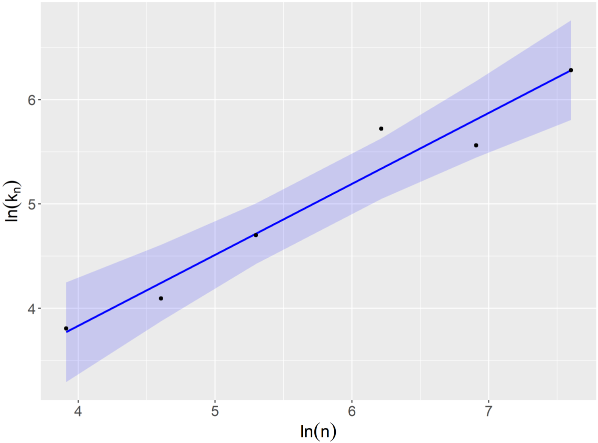


Figure: Equation : $y = 1.1 + 0.68x,\ R^2 = 0.952$

Scaling of k_n with σ, $n = 200$

$$k_n \propto M^{\frac{d}{2h+d}} \propto \sigma^{\frac{d}{2h+d}}$$

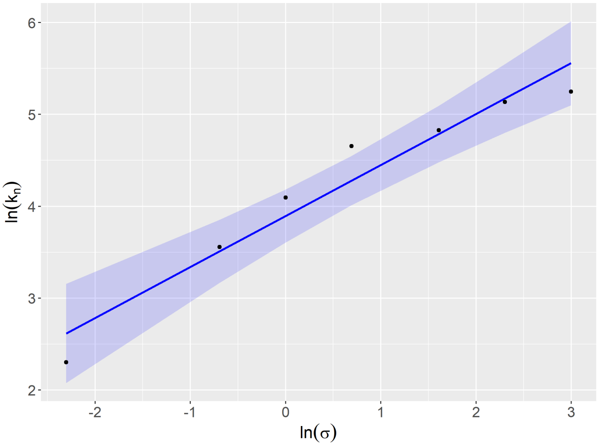


Figure: Equation : $y = 3.9 + 0.56x,\ R^2 = 0.942$