1 Conformalized Quantile Regression

This chapter introduces *Conformalized Quantile Regression (CQR)* as the first of two main post-processing procedures that we implemented in the `postforecasts` package.

Section 1.1 explains the original Conformalized Quantile Regression algorithm as proposed in Romano, Patterson, and Candès (2019). There, we highlight potential limitations of the traditional implementation that could potentially be fixed by more flexible modifications, which are discussed in Section 1.2 and Section 1.3.

1.1 Traditional CQR

All derivations in this sections are taken from the original paper (Romano, Patterson, and Candès 2019). The authors motivate Conformalized Quantile Regression by stating two criteria that an ideal procedure for generating prediction intervals should satisfy:

- It should provide valid coverage in finite samples without making strong distributional assumptions
- The resulting intervals should be as short as possible at each point in the input space

According to the authors, CQR performs well on both criteria while being *distribution-free* and adaptive to *heteroscedasticity*.

1.1.1 Statistical Validity

The algorithm that CQR is build upon is statistically supported by the following Theorem:

Theorem 1.1. *If* $(X_i, Y_i), i = 1, \ldots, n+1$ *are exchangeable, then the* $(1-\alpha) \cdot 100\%$ *prediction interval* $C(X_{n+1})$ *constructed by the CQR algorithm satisfies*

$$P\left(Y_{n+1} \in C(X_{n+1})\right) \geq 1 - \alpha.$$

Moreover, if the conformity scores E_i *are almost surely distinct, then the prediction interval is nearly perfectly calibrated:*

$$P\left(Y_{n+1} \in C(X_{n+1})\right) \leq 1 - \alpha + \frac{1}{|I_2| + 1},$$

where I_2 *denotes the calibration set.*

Thus, the first statement of Theorem 1.1 provides a coverage *guarantee* in the sense that the adjusted prediction interval is *lower-bounded* by the desired coverage level. The second statement adds an *upper-bound* to the coverage probability which gets tighter with increasing sample size and asymptotically converges to the desired coverage level $1 - \alpha$ such that lower bound and upper bound are asymptotically identical.

1.1.2 Algorithm

The CQR algorithm is best described as a multi-step procedure.

Step 1:
Split the data into a training and validation (here called *calibration*) set, indexed by I_1 and I_2, respectively.

Step 2:
For a given quantile α and a given quantile regression algorithm $\mathcal{A}$, calculate lower and upper interval bounds on the training set:

$$\{\hat{q}_{\alpha,low},\ \hat{q}_{\alpha,high}\} \leftarrow \mathcal{A}\left(\{(X_i, Y_i) : i \in I_1\}\right).$$

Step 3:
Compute *conformity scores* on the calibration set:

$$E_i := \max\left\{\hat{q}_{\alpha,low}(X_i) - Y_i,\ Y_i - \hat{q}_{\alpha,high}(X_i)\right\} \quad \forall\, i \in I_2$$

For each i, the corresponding score E_i is *positive* if Y_i is *outside* the interval $[\hat{q}_{\alpha,low}(X_i),\ \hat{q}_{\alpha,high}(X_i)]$ and *negative* if Y_i is *inside* the interval.

Step 4:
Compute the *margin* $Q_{1-\alpha}(E, I_2)$ given by the $(1-\alpha)(1+\frac{1}{1+|I_2|})$-th empirical quantile of the scores E_i in the calibration set. For small sample sizes and small quantiles α the quantile above can be greater than 1 in which case it is simply set to 1 such that the maximum value of the score vector is selected.

Step 5:
On the basis of the original prediction interval bounds $\hat{q}_{\alpha,low}(X_i)$ and $\hat{q}_{\alpha,high}(X_i)$, the new *post-processed* prediction interval for Y_i is given by

$$C(X_{n+1}) = \left[\hat{q}_{\alpha,low}(X_i) - Q_{1-\alpha}(E, I_2),\ \hat{q}_{\alpha,high}(X_i) + Q_{1-\alpha}(E, I_2)\right].$$

Note that the *same* margin $Q_{1-\alpha}(E, I_2)$ is subtracted from the original lower quantile prediction and added to the original upper quantile prediction. This limitation is adressed in Section 1.2.

1.2 Asymmetric CQR

As noted in Section 1.1 this section suggests a first extension to the original algorithm. Instead of limiting ourselves to choosing the *same* margin $Q_{1-\alpha}(E, I_2)$ for adjusting the original lower and upper quantile predictions, we allow for individual and, thus, generally different margins $Q_{1-\alpha,low}(E, I_2)$ and $Q_{1-\alpha,high}(E, I_2)$ such that the post-processed prediction interval is given by

$$C(X_{n+1}) = \left[\hat{q}_{\alpha,low}(X_i) - Q_{1-\alpha,low}(E, I_2),\ \hat{q}_{\alpha,high}(X_i) + Q_{1-\alpha,high}(E, I_2)\right].$$

This asymmetric version additionally requires a change in the computation of the conformity scores. Instead of considering the elementwise maximum of the differences between observed values Y_i and original bounds, we simply compute two separate score vectors:

$$\begin{aligned} E_{i,low} &:= \hat{q}_{\alpha,low}(X_i) - Y_i \quad \forall\, i \in I_2 \\ E_{i,high} &:= Y_i - \hat{q}_{\alpha,high}(X_i) \quad \forall\, i \in I_2 \end{aligned}$$

1.3 Multiplicative CQR

On top of the asymmetric CQR version introduced in Section 1.2, we can extend the CQR algorithm further. So far, the adjustments to the original prediction interval were always chosen in *additive* form. It may be useful to leverage the *magnitude* of the original bounds more explicitly by using *relative* or *multiplicative* adjustments.

Hence, we again compute separate margins $Q_{1-\alpha,low}(E, I_2)$ and $Q_{1-\alpha,high}(E, I_2)$ which are now *multiplied* with the existing forecasts. The post-processed prediction interval is then given by

$$C(X_{n+1}) = \left[\hat{q}_{\alpha,low}(X_i) \cdot Q_{1-\alpha,low}(E, I_2),\ \hat{q}_{\alpha,high}(X_i) \cdot Q_{1-\alpha,high}(E, I_2)\right].$$

Just like the asymmetric version, the computation of the score vectors is changed accordingly to respect the new multiplicative relationship:

$$\begin{aligned} E_{i,low} &:= \frac{Y_i}{\hat{q}_{\alpha,low}(X_i)} \quad \forall\, i \in I_2 \\ E_{i,high} &:= \frac{Y_i}{\hat{q}_{\alpha,high}(X_i)} \quad \forall\, i \in I_2, \end{aligned}$$

where we have to exclude original predictions with the value 0. Since in our application of Covid-19 Cases and Deaths all values are non-negative, we threshold the scores at zero such that $E_{i,low}$ equals 0 whenever $\hat{q}_{\alpha,low}(X_i) \leq 0$.

Romano, Yaniv, Evan Patterson, and Emmanuel J. Candès. 2019. "Conformalized Quantile Regression." *arXiv:1905.03222 [Stat]*, May. http://arxiv.org/abs/1905.03222.