Note on Full Conformal Risk Control

Anastasios N. Angelopoulos

University of California, Berkeley

Abstract

We describe a full conformal version of the conformal risk control procedure of Angelopoulos et al. (2024). Like full conformal prediction, full conformal risk control allows the use of all data for both model training and calibration, and is thus more data-efficient than the original procedure, although it is also more computationally expensive.

1 Full conformal risk control

Conformal risk control has been previously described in a *split conformal* setting, wherein the risk of a pre-trained model is measured on an as-yet-unseen calibration dataset. The fact that the calibration dataset is unseen means that the loss incurred by the model on each calibration datapoint is exchangeable with the loss on the test point; this allows us to bound the risk on the new point. However, it requires a separate split of calibration data, which is not always possible, especially when the number of data points is small.

In this short note, we describe a *full conformal* version of conformal risk control, which allows the use of all data for both model training and calibration, just like the full conformal prediction procedure of Vovk et al. (2005). For a *cross-conformal* version of conformal risk control, the reader may also wish to reference the simultaneously developed approach of Cohen et al. (2024), which builds on the resampling procedures of Barber et al. (2021) and Vovk (2015). In some sense, these methods together complete the 'mosaic' of risk control procedures:

1. **Data-splitting.** The conformal risk control family of methods is introduced using data-splitting in Angelopoulos et al. (2024). This is analogous to the split-conformal procedure of Papadopoulos et al. (2002).

2. **Resampling.** Resampling-based procedures for conformal risk control are described in Cohen et al. (2024). This is analogous to the cross-conformal procedure of Vovk (2015).

3. **Full-conformal.** We describe the full-conformal risk control procedure in this note. This is analogous to the full-conformal prediction procedure of Vovk et al. (2005).

1.1 Setup and notation

Consider the following:

1. a feature space $\mathcal{X}$ and a label space $\mathcal{Y}$ with $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$;

2. an exchangeable set of feature-label pairs, $(X_1, Y_1), \ldots, (X_{n+1}, Y_{n+1})$, where the final label Y_{n+1} is unknown;

3. a sequence of functions $\mathcal{C}_\lambda : \mathcal{X} \times \mathcal{Z}^* \to \mathcal{H}$ for $\lambda \in [0, 1]$, and some third space $\mathcal{H}$, commonly taken to be $\mathcal{Y}$ or $2^{\mathcal{Y}}$;

4. a right-continuous loss function $\ell : \mathcal{Y} \times \mathcal{H} \to [0, B]$ satisfying, for all $(x, y) \in \mathcal{Z}$ and all $D \in \mathcal{Z}^*$,

$$\lambda_1 < \lambda_2 \implies \ell(y, \mathcal{C}_{\lambda_1}(x; D)) \geq \ell(y, \mathcal{C}_{\lambda_2}(x; D))$$

and

$$\ell(y, \mathcal{C}_0(x; D)) = B, \qquad \ell(y, \mathcal{C}_1(x; D)) = 0.$$

We furthermore assume that $\mathcal{C}$ is symmetric in its second argument, i.e., that for all $\sigma \in \Sigma$, where Σ is the set of all permutations of $[n+1]$, and all $x \in \mathcal{X}$, we have $\mathcal{C}(x; D) \overset{\text{a.s.}}{=} \mathcal{C}(x; D_\sigma)$. The goal is to control the risk at some level $\alpha \in [0, B]$, i.e., to form a set satisfying

$$\mathbb{E}[\ell(Y_{n+1}, \mathcal{C}(X_{n+1}))] \leq \alpha.$$

As in full conformal prediction, we will construct an augmented dataset based on a guess of Y_{n+1}. Define the augmented dataset as the ordered vector

$$D^y = ((X_1, Y_1), \dots, (X_{n+1}, y)).$$

When we guess the correct label, $D^{Y_{n+1}}$ is an exchangeable vector (i.e., one satisfying $D^{Y_{n+1}} \overset{\text{d}}{=} D_\sigma^{Y_{n+1}}$ for all $\sigma \in \Sigma$).

Finally, we will use the shorthand $L_i(\lambda; D) := \ell(Y_i, \mathcal{C}_\lambda(X_i; D))$ for $i \in [n]$ and $L_{n+1}^y(\lambda; D) := \ell(y, \mathcal{C}_\lambda(X_{n+1}; D))$ and

$$\hat{R}_n(\lambda; D) := \frac{1}{n}\sum_{i=1}^n L_i(\lambda; D) \qquad \hat{R}_{n+1}^y(\lambda; D) := \frac{1}{n+1}\sum_{i=1}^n L_i(\lambda; D) + \frac{1}{n+1} L_{n+1}^y(\lambda; D).$$

We also consider the vector of losses $L^y(\lambda) = (L_1(\lambda; D^y), \dots, L_{n+1}^y(\lambda; D^y))$. It is not hard to see that $L^{Y_{n+1}}(\lambda)$ is exchangeable for all λ.

1.2 The full conformal risk control procedure

Define

$$\hat{\lambda}^y = \hat{\lambda}(D^y) = \inf\left\{\lambda : \hat{R}_{n+1}^y(\lambda; D^y) \leq \alpha\right\}$$

and the intermediate sets

$$\mathcal{C}_{\hat{\lambda}^y}^y(x) = \mathcal{C}_{\hat{\lambda}^y}(x; D^y).$$

Then we combine to get the final prediction set as

$$\mathcal{C}(x) = \bigcup_{y \in \mathcal{Y}} \mathcal{C}_{\hat{\lambda}^y}^y(x) = \mathcal{C}_{\sup_{y \in \mathcal{Y}} \hat{\lambda}^y}(x).$$

This prediction set has the desired guarantee:

Theorem 1 (Validity of full conformal risk control for monotone losses)**.** *In the above setting, we have*

$$\mathbb{E}[\ell(Y_{n+1}, \mathcal{C}(X_{n+1}))] \leq \alpha.$$

Proof. We have that $D^{Y_{n+1}}$ is exchangeable by definition. The first step is to prove that $L^{Y_{n+1}}$ is exchangeable. Beginning with the definition of $L^{Y_{n+1}}$, we have that for any $\sigma \in \Sigma$,

$$\begin{aligned}
L^{Y_{n+1}}(\lambda) &= (L_1(\lambda; D^{Y_{n+1}}), \dots, L_{n+1}^{Y_{n+1}}(\lambda; D^{Y_{n+1}})) \\
&= (\ell(Y_1, \mathcal{C}_\lambda(X_1; D^{Y_{n+1}})), \dots, \ell(Y_{n+1}, \mathcal{C}_\lambda(X_{n+1}; D^{Y_{n+1}}))) \\
&\overset{\text{d}}{=} (\ell(Y_{\sigma(1)}, \mathcal{C}_\lambda(X_{\sigma(1)}; D_\sigma^{Y_{n+1}})), \dots, \ell(Y_{\sigma(n+1)}, \mathcal{C}_\lambda(X_{\sigma(n+1)}; D_\sigma^{Y_{n+1}}))) \\
&= (\ell(Y_{\sigma(1)}, \mathcal{C}_\lambda(X_{\sigma(1)}; D^{Y_{n+1}})), \dots, \ell(Y_{\sigma(n+1)}, \mathcal{C}_\lambda(X_{\sigma(n+1)}; D^{Y_{n+1}}))) \\
&= (L_{\sigma(1)}(\lambda; D^{Y_{n+1}}), \dots, L_{\sigma(n+1)}(\lambda; D^{Y_{n+1}})).
\end{aligned}$$

Thus, $L^{Y_{n+1}}(\lambda)$ is a vector of exchangeable random functions.

Now we proceed to the main result.

$$
\begin{aligned}
\mathbb{E}\left[\ell(Y_{n+1}, \mathcal{C}(X_{n+1}))\right] &\le \mathbb{E}\left[\ell(Y_{n+1}, \mathcal{C}_{\hat{\lambda}(D^{Y_{n+1}})}(X_{n+1}))\right] \\
&= \mathbb{E}\left[\frac{\sum_{\sigma\in\Sigma} \ell(Y_{\sigma(n+1)}, \mathcal{C}_{\hat{\lambda}(D_{\sigma}^{Y_{n+1}})}(X_{\sigma(n+1)}))}{|\Sigma|}\right] \\
&= \mathbb{E}\left[\frac{\sum_{\sigma\in\Sigma} \ell(Y_{\sigma(n+1)}, \mathcal{C}_{\hat{\lambda}(D^{Y_{n+1}})}(X_{\sigma(n+1)}))}{|\Sigma|}\right] \\
&= \mathbb{E}\left[\frac{1}{n+1}\sum_{i=1}^{n+1} \ell(Y_i, \mathcal{C}_{\hat{\lambda}(D^{Y_{n+1}})}(X_i)\right] \\
&= \mathbb{E}\left[\hat{R}_{n+1}^{Y_{n+1}}(\hat{\lambda}^{Y_{n+1}}; D^{Y_{n+1}})\right] \le \alpha
\end{aligned}
$$

□

Acknowledgements

I would like to thank Vincent Blot, Nicolas Brunel, and Tijana Zrnić for helpful discussions, and my advisors Jitendra Malik and Michael I. Jordan for their guidance and support.

References

Anastasios N. Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. In *Proceedings of the International Conference on Learning Representations*, 2024.

Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. Predictive inference with the jackknife+. *Annals of Statistics*, 49(1):486–507, 2021.

Kfir M. Cohen, Sangwoo Park, Osvaldo Simeone, and Shlomo Shamai. Cross-validation conformal risk control. *arXiv preprint arXiv:2401.11974*, 2024.

Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. *European Conference on Machine Learning*, 2002.

Vladimir Vovk. Cross-conformal predictors. *Annals of Mathematics and Artificial Intelligence*, 74(1–2):9–28, 2015.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.