Conformal prediction

Maximilian Kasy

Department of Economics, University of Oxford

Winter 2025

Outline

- A general procedure for constructing confidence sets for predictions.
- Proof of validity under minimal assumptions.
- Special cases: Constructions using
  - Classification.
  - Confidence bands.
  - Quantile regressions.
- Unconditional and conditional coverage.

Takeaways for this part of class

- Any method for point prediction can be converted into a method for uncertainty quantification.
- Size control only relies on exchangeability of test observations with calibration data.
- Better predictors yield tighter confidence sets.
- Important caveat: Coverage is unconditional. Conditional coverage requires modification.

General procedure

- Take some scoring function $s(x, y)$.
- Fix a calibration sample of observations $i = 1, \ldots, n$, and let

$$s_i = s(X_i, Y_i).$$

- Let $\tilde{\alpha} = 1 - \frac{\lceil (n+1)(1-\alpha) \rceil}{n}$ (rounded down significance level).
- Let $\hat{q}$ be the $1 - \alpha$ quantile of $\{s_1, \ldots, s_n\}$.
- For a new observation X_{test}, let

$$\mathcal{C}(X_{test}) = \{y : \ s(X_{test}, y) \leq \hat{q}\}.$$

Coverage

- **Theorem**:
  - Suppose $\{(X_1, Y_1), \ldots, (X_n, Y_n), (X_{test}, Y_{test})\}$ are i.i.d.
  - Then
$$P(Y_{test} \in \mathcal{C}(X_{test})) \geq 1 - \alpha.$$
- **Proof**:
  - Let $s_{(i)}$ be the ith order statistic of $\{s_1, \ldots, s_n\}$, and $s_{test} = s(X_{test}, Y_{test})$.
  - By exchangeability (i.i.d.), $P(s_{test} \leq s_{(i)}) = \frac{i}{n+1}$.
  - Therefore
$$P(s_{test} \leq s_{\lceil (n+1)(1-\alpha) \rceil}) = \frac{\lceil (n+1)(1-\alpha) \rceil}{n+1} \geq 1 - \alpha.$$

Scores for classification

- Suppose $P(Y = y|X = x) \approx \hat{f}(y|x)$.
- (E.g. multinomial logit, or neural net classifier).
- One possible score:

$$s(x,y) = 1 - \hat{f}(y|x).$$

- Problem with this score: Small $\mathcal{C}$ when uncertainty is high.
- Better: Let $j(y,x)$ be the rank of $\hat{f}(y|x)$ across y. Define

$$s(x,y) = \sum_{y'} \mathbf{1}(j(y|x) \geq j(y'|x)) \cdot \hat{f}(y|x)$$

Score from point prediction

- Let $\hat{f}(x)$ be a point-predictor of y given x.
- E.g. Lasso, random forest, neural net, ...
- Define

$$s(x,y) = |y - \hat{f}(x)|.$$

- Yields confidence band of constant width.

Score from uncertainty estimate

- Taking into account conditional uncertainty: Let additionally $\hat{\sigma}(x)$ be a predictor of $|y - \hat{f}(x)|$. E.g. estimate of conditional standard deviation.
- Define

$$s(x, y) = \frac{|y - \hat{f}(x)|}{\hat{\sigma}(x)}.$$

- Yields confidence bands with width proportional to $\hat{\sigma}(x)$.

Score from quantile regressions

- Fit two separate models for conditional quantiles $t_{\frac{\alpha}{2}}(x)$ and $t_{1-\frac{\alpha}{2}}(x)$ of Y given X.
- Loss function for estimation of qth quantile $t_q(x)$:

$$l(t,y) = q \cdot \max(t-y,0) + (1-q) \cdot \max(y-t,0).$$

- Define

$$s(x,y) = \max\left(t_{\frac{\alpha}{2}}(x) - y, y - t_{1-\frac{\alpha}{2}}(x)\right).$$

- Resulting confidence band:

$$\mathcal{C}(x) = \left[t_{\frac{\alpha}{2}}(x) - \hat{q}, t_{1-\frac{\alpha}{2}}(x) + \hat{q}\right].$$

Unconditional versus conditional coverage

- The theorem guarantees unconditional coverage: $P(Y_{test} \in \mathcal{C}(X_{test})) \geq 1 - \alpha$.
- It does *not* guarantee conditional coverage: $P(Y_{test} \in \mathcal{C}(X_{test}) | X_{test}) \geq 1 - \alpha$.
- We can modify conformal prediction to hold conditional on groups:
  - Let $g = g(x) \in \{0, \ldots, G\}$.
  - Use conditional empirical quantiles $\hat{q}^g$ for construction of $\mathcal{C}(X)$.

Full conformal prediction

- Having a calibration sample not used for estimation is wasteful.
- Full conformal prediction avoids this.
- Computationally demanding: Requires re-fitting model for every possible value y.
- Size control again holds under exchangeability.

Full conformal prediction, continued

- For each possible outcome y for a test observation:
  - Fit a model $\hat{f}^y$ on training data and the hypothetical observation (X_{test}, y).
  - Calculate the quantile $\hat{q}^y$ of the scores $s(X_i, Y_i, \hat{f}^y)$.
- Define

$$\mathcal{C}(X_{test}) = \left\{ y : \; s(X_{test}, y, \hat{f}^y) \leq \hat{q}^y \right\}.$$

- This uses the same data for estimation and calibration.

References

Angelopoulos, A. N. and Bates, S. (2021). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. arXiv preprint arXiv:2107.07511.