Review for
"Distribution-Free Predictive Inference for Regression"

Jiaxin Hu

This paper is a textbook-like paper that comprehensively investigates the conformal prediction inference under the classical i.i.d. or exchangability setting. The basic idea behind the conformal inference is a simple fact about the sample quantiles: Let $U_1, \ldots, U_n$ be and i.i.d. (exchangeable) samples and U_{n+1} be an i.i.d. copy, then we have

$$\mathbb{P}(U_{n+1} \leq \hat{q}_{1-\alpha}) \geq 1 - \alpha, \tag{1}$$

where $\alpha \in (0, 1)$ is the given miscoverage level, and $\hat{q}_{1-\alpha}$ is the $(1 - \alpha)$-th upper quantile of $U_1, \ldots, U_n$. The inequality (1) fails immediately if the exchangeability of U_k is not satisfied. Hence, the exchangability or i.i.dness is an essential assumption for the classical conformal inference, and the extension to the non-exchangeable case is thus important since exchangeability or i.i.dness are hardly satisfied in some applications.

Starting from (1) reviews the conformal methods with important variants: (1) generic conformal prediction set with full data; (2) split conformal prediction set which separates the full data to training and testing sets; (3) multiple splits that is equal to the intersection of multiple $1 - \alpha/N$ prediction sets, where N is the number of splits; (4) Jackknife prediction sets using leave-one-out residuals. Take-aways are below:

1. The original conformal method (1) is suppose to possess the best statistical properties while the computational cost to obtain (1) is expensive due to the repetitive model fitting with new predictor (X_{n+1}, y) with different y.
2. The split approach (2) dramatically decrease the computational issue, whereas, the coverage performance relies more on the i.i.dness and the fitted model may not achieve the best fit, since only half of the data is used to train and the other half develops the confidence interval.
3. Multiple split (3) faces the "Bonferroni-intersection trade off", and the Bonferroni effect dominates in this problem and thereof usually outputs a wider interval than regular split.
4. Jackknife (4) uses more traning data than split approach (2) and (3). However, the coverage of (4) is not guaranteed with finite samples.

This paper also studies the statistical accuracy of original conformal sets, in terms of the interval length compared with the (super-)oracle intervals with true quantiles. Unlike coverage property, the length accuracy depends on the consistency of the estimator and requires more regularity assumption on the sample distribution. The extra distribution assumption requires the independence between the noise and the mean, which excludes Bernoulli, Poisson, and other mean-variance dependent data.

Interesting extensions of in-sample and locally weighted conformal inference are also discussed. The in-sample conformal set constructs the confidence interval of the estimated response in training data, rather than the new prediction. The weighted conformal prediction address the correlation between the residual and the predictor. Correction is proposed to eliminate the mean effects in the residual, which somehow solve the mean-variance dependence issue in some data.

Last, this paper provides a conformal-based variable selection method. The importance of the variable is described by the leave-one-covariate-out difference, $\Delta_j = |\text{residual (no variable } j)| - |\text{residual (full)}|$. We select the important variables with Δ_j whose lower bound of the conformal sets is above 0.

Random thoughts:

1. In Wednesday's statistics seminar (speaker: Kristin Linn, title: Estimation and Evaluation of Individualized Treatment Rules Following Multiple Imputation), the precision medicine analysis also want to evaluate the uncertainty of the treatment response when the patient take a recommended treatment. Kristin used cross-validation and the bootstrap resampling to construct the confidence interval of the response. This application is close to the in-sample conformal inference in Lei's paper and avoids the model assumptions of the regression. We may also use this conformal techniques in other personalized studies.

2. In Section 3, Lei's paper indicates the prediction sets have different property when the estimator has different consistency and stability conditions. Particularly, the perturb-one sensitivity is very similar with the algorithm stability defined in Barber's black box testing paper. Barber's paper indicates the perturb-one sensitivity is a very strong stability property, and Lei's paper consider an alternative sampling stability. We have extend the testing idea with this alternative stability.

3. The model-free variable selection with conformal prediction sets can be applied to many biostatistical studies. For example, we may select critical genes/SNPs/traits by consider the lower bound of the conformal sets. The advantage of conformal sets is that no model assumption is needed. Simulations in Lei's paper also indicate conformal sets tends to give a more accurate intervals based on least squared estimator when the linear regression assumptions are violated. Another potential application is: whether we can use conformal set to find the relatively important features/genes/entities? For example, in precision medicine, we would suggest top 10% patients with high risk to accept the intervention. How to determine the 10% based on the conformal results? These may be interesting future questions.

References