FACULTY OF ENGINEERING AND ARCHITECTURE

13th Symposium on Conformal and Probabilistic Prediction with Applications

Conformal Predictive Systems under Covariate Shift

Jef Jonkers, Glenn Van Wallendael, Luc Duchateau, Sofie Van Hoecke

jef.jonkers@ugent.be

GHENT UNIVERSITY

IDLab INTERNET & DATA LAB

Overview

Introduction

GHENT UNIVERSITY

INTERNET & DATA LAB

Conformal predictive systems

What?

- Conformal predictive systems (CPSs) output conformal predictive distributions
  - p-values arranged into a non-parametric distribution
- CPS = smooth conformal transducer $\cap$ randomized predictive system (RPS)
- Can be used to calibrate a probabilistic regression model (deterministic prediction system)
- Probabilistically calibrated

Conformal predictive systems

What?

- Conformal predictive systems (CPSs) output conformal predictive distributions
  - p-values arranged into a non-parametric distribution
- CPS = smooth conformal transducer $\cap$ randomized predictive system (RPS)
- Can be used to calibrate a probabilistic regression model (deterministic prediction system)
- Probabilistically calibrated

Assumptions

- General IID model

Conformal predictive systems

What?

- Conformal predictive systems (CPSs) output conformal predictive distributions
  - p-values arranged into a non-parametric distribution
- CPS = smooth conformal transducer $\cap$ randomized predictive system (RPS)
- Can be used to calibrate a probabilistic regression model (deterministic prediction system)
- Probabilistically calibrated

Assumptions

- General IID model

Why?

- Decision-making
- Results into a central conformal predictor
- ...

IID assumption violated?

Consequences

- Losing validity guarantees

IID assumption violated?

Consequences

- Losing validity guarantees

When?

- Covariate shift
  - causal inference
  - time series
  - scientific discovery
  - transfer learning
- Label shift
  - time series
  - ...

IID assumption violated?

Consequences

- Losing validity guarantees

When?

- Covariate shift
  - causal inference
  - time series
  - scientific discovery
  - transfer learning
- Label shift
  - time series
  - ...

Remdy

- Conformal prediction (CP)
  - weighted CP (Tibshirani et al., 2019)
  - (Dt) Adaptive conformal inference (Gibbs and Candes, 2021; Gibbs and Candès, 2024)
  - ...
- CPS?

IID assumption violated?

Consequences

- Losing validity guarantees

When?

- Covariate shift
  - causal inference
  - time series
  - scientific discovery
  - transfer learning
- Label shift
  - time series
  - ...

Remdy

- Conformal prediction (CP)
  - weighted CP (Tibshirani et al., 2019)
  - (Dt) Adaptive conformal inference (Gibbs and Candes, 2021; Gibbs and Candès, 2024)
  - ...
- CPS?
  - **weighted CPS** (this work)
    - validity under covariate shift

Conformal Predictive System

GHENT UNIVERSITY

IDLab INTERNET & DATA LAB

Randomized predictive system

Definition (Randomized predictive system (RPS), Vovk et al., 2019)

A function $Q : \mathbf{Z}^{n+1} \times [0,1] \to [0,1]$ is an RPS:

R1.1 The function $Q(z_1, ..., z_n, (x_{n+1}, y), \tau)$ is monotonically increasing both in y and τ.

R1.2 For each $\tau, \tau' \in [0,1]$, $Q(z_1, ..., z_n, (x_{n+1}, y), \tau) \geq Q(z_1, ..., z_n, (x_{n+1}, y'), \tau')$ if $y > y'$,

R1.3 $\lim_{y \to -\infty} Q(z_1, ..., z_n, (x_{n+1}, y), 0) = 0, \qquad \lim_{y \to \infty} Q(z_1, ..., z_n, (x_{n+1}, y), 1) = 1$

R2 When $z_1 \sim P, ..., z_n \sim P, z_{n+1} \sim P$, and $\tau \sim \mathit{Uniform}(0,1)$, all assumed to be independent, the distribution of Q is uniform:

$$\forall \alpha \in [0,1] : \mathbb{P}\{Q(z_1, ..., z_n, z_{n+1}, \tau) \leq \alpha\} = \alpha$$

Smooth conformal transducer

Definition (Smooth conformal transducer, Vovk et al., 2022)

The conformal transducer determined by a conformity measure A is defined as,

$$Q(z_1, ..., z_n, (x_{n+1}, y), \tau) := \sum_{i=1}^{n+1} [R_i^y < R_{n+1}^y] \frac{1}{n+1} + \sum_{i=1}^{n+1} [R_i^y = R_{n+1}^y] \frac{\tau}{n+1}$$

where $(z_1, ..., z_n)$ is the training sequence, $\tau \in [0, 1]$, x_{n+1} is a test object, and for each label y the corresponding conformity score R_i^y is defined as

$$\begin{aligned} R_i^y &:= A(z_1, ..., z_{i-1}, z_{i+1}, ..., z_n, (x_{n+1}, y), z_i), \qquad i = 1, ..., n \\ R_{n+1}^y &:= A(z_1, ..., z_n, (x_{n+1}, y)). \end{aligned}$$

CPS = conformal transducer $\cap$ RPS

- For smooth conformal transducers: $Q(z_1, ..., z_n, z_{n+1}, \tau) \sim Uniform(0, 1)$
  - coincides with R2 of RPS
  - under IID assumption
- To satisfies R1.1 and R1.3, the conformal transducer can only be defined by certain conformity measures A
  - monotonic conformity measures (e.g., $\alpha_i := y_i - \hat{y}_i$)

Weighted Conformal Predictive System

Problem setting: covariate shift

A covariate shift is a distributional shift where the test object (x_{n+1}, y_{n+1}) is differently distributed, i.e. $x_{n+1} \sim \tilde{P}_X$, than the training data $z_i = (x_i, y_i), i = 1, ..., n$ where $x_i \sim P_X$, thus $\tilde{P}_X \neq P_X$. However, the relationship between inputs and labels remains fixed.

$$
\begin{aligned}
(x_i, y_i) &\overset{\text{iid}}{\sim} P = P_X \times P_{Y|X}, \quad i = 1, ..., n \\
(x_{n+1}, y_{n+1}) &\sim \tilde{P} = \tilde{P}_X \times P_{Y|X}
\end{aligned}
$$

Weighted conformal predictive system

- Provides probabilistic calibrated predictive distributions under covariate shift
- Builds on the same idea as weighted CP Tibshirani et al., 2019.
  - Weight each conformity score by a probability proportionally to the likelihood ratio $w(x_i) = \frac{d\tilde{P}_X(x_i)}{dP_X(x_i)}$
- Weighted CPS = Smooth weighted conformal transducer $\cap$ Weighted RPS

Smooth weighted conformal transducer

Definition (Smooth weighted conformal transducer)

The weighted conformal transducer determined by conformity measure A and likelihood ratio $w(x_i) = \frac{d\tilde{P}_X(x_i)}{dP_X(x_i)}$ is defined as,

$$Q(z_1, ..., z_n, \frac{d\tilde{P}}{dP}, (x_{n+1}, y), \tau) := \sum_{i=1}^{n+1} [R_i^y < R_{n+1}^y] p_i^w(x) + \sum_{i=1}^{n+1} [R_i^y = R_{n+1}^y] p_i^w(x) \tau$$

where

$$p_i^w(x) = \frac{w(x_i)}{\sum_{j=1}^n w(x_j) + w(x)}, \qquad p_{n+1}^w(x) = \frac{w(x)}{\sum_{j=1}^n w(x_j) + w(x)}.$$

Weighted randomized predictive system

Definition (Weighted randomized predictive system (WRPS))

A function $Q : \mathbf{Z}^{n+1} \times [0,1] \to [0,1]$ is an WRPS:

R1.1 The function $Q(z_1, ..., z_n, \frac{d\tilde{P}}{dP}, (x_{n+1}, y), \tau)$ is monotonically increasing both in y and τ.

R1.2 For each $\tau, \tau' \in [0,1]$, $Q(z_1, ..., z_n, \frac{d\tilde{P}}{dP}, (x_{n+1}, y), \tau) \geq Q(z_1, ..., z_n, \frac{d\tilde{P}}{dP}, (x_{n+1}, y'), \tau')$ if $y > y'$,

R1.3 $\lim_{y \to -\infty} Q(z_1, ..., z_n, \frac{d\tilde{P}}{dP}, (x_{n+1}, y), 0) = 0$,
$\lim_{y \to \infty} Q(z_1, ..., z_n, \frac{d\tilde{P}}{dP}, (x_{n+1}, y), 1) = 1$,

R2 When $z_1 \sim P, ..., z_n \sim P$, $z_{n+1} \sim \tilde{P}$, and $\tau \sim Uniform(0,1)$, all assumed to be independent, the distribution of Q is uniform:

$$\forall \alpha \in [0,1] : \mathbb{P}\{Q(z_1, ..., z_n, \frac{d\tilde{P}}{dP}, z_{n+1}, \tau) \leq \alpha\} = \alpha$$

WCPS = weighted conformal transducer ∩ WRPS

- To satisfies R1.1 and R1.3, the weighted conformal transducer can only be defined by certain conformity measures A.
  - monotonic conformity measures (e.g., $\alpha_i := y_i - \hat{y}_i$)
  - proof similar to CPS
- For smooth weighted conformal transducers: $Q(z_1, ..., z_n, \frac{d\tilde{P}}{dP}, z_{n+1}, \tau) \sim Uniform(0, 1)$
  - coincides with R2 of WRPS
  - under covariate shift assumption and knowing $w(x_i) = \frac{d\tilde{P}_X(x_i)}{dP_X(x_i)}$
  - proof not provided yet

Weighted split conformal predictive system

We can also extend the computationally efficient split conformal predictive system by introducing a smooth weighted split conformal transducer.

Definition (Smooth weighted split conformal transducer)

The weighted split conformal transducer determined by conformity measure A and likelihood ratio $w(x_i) = \frac{d\tilde{P}_X(x_i)}{dP_X(x_i)}$ is defined as,

$$Q(z_1, ..., z_n, \frac{d\tilde{P}}{dP}, (x_{n+1}, y), \tau) := \sum_{i=m+1}^{n} [R_i < R^y] p_i^w(x) + \sum_{i=m+1}^{n} [R_i = R^y] p_i^w(x)\tau + p_{n+1}^w(x)\tau$$

where

$$p_i^w(x) = \frac{w(x_i)}{\sum_{j=1}^{n} w(x_j) + w(x)}, \qquad p_{n+1}^w(x) = \frac{w(x)}{\sum_{j=1}^{n} w(x_j) + w(x)}.$$

Experiments

GHENT UNIVERSITY

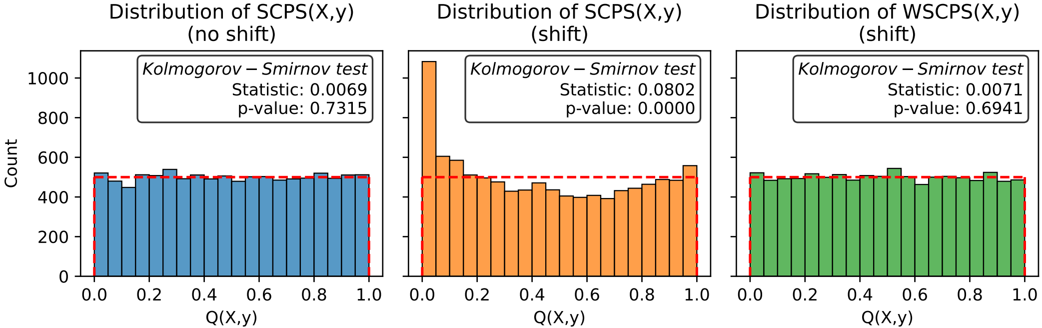
IDLab INTERNET & DATA LAB

Setup

From Kang and Schafer, 2007, also used in Yang et al. (2022), where each observation i is generated in the following way:

- $(x_{i1}, x_{i2}, x_{i3}, x_{i4})^T$ is independently distributed as $N(0, I_4)$ where I_4 represents the 4×4 identity matrix.
- $y_i = 210 + 27.4x_{i1} + 13.7x_{i2} + 13.7x_{i3} + 13.7x_{i4} + \varepsilon_i, \quad \text{where} \quad \varepsilon_i \sim N(0, 1)$
- $w(x) = \exp(-x_{i1} + 0.5x_{i2} - 0.25x_{i3} - 0.1x_{i4})$, which represents the likelihood ratio of the covariate distributions of the shifted test set D_{test} and training set D_{train}.

Probabilistic calibration under covariate shift

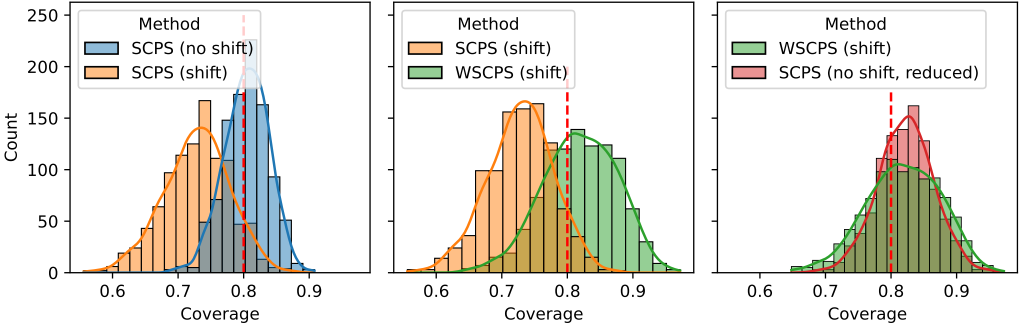


Intervals are still possible ...

$$C_{n+1}(x_{n+1}) := \left\{ y \in \mathbb{R} | \quad Q(z_1, ..., z_n, \frac{d\tilde{P}}{dP}, (x_{n+1}, y), \tau) \in [\alpha, 1 - \alpha] \right\}$$

Intervals are still possible ...

$$C_{n+1}(x_{n+1}) := \left\{ y \in \mathbb{R} \middle| \quad Q(z_1, ..., z_n, \frac{d\tilde{P}}{dP}, (x_{n+1}, y), \tau) \in [\alpha, 1-\alpha] \right\}$$



crepes-weighted

GHENT UNIVERSITY

INTERNET & DATA LAB

crepes-weighted: crepes made weighted

- Conformal prediction (regression, Tibshirani et al., 2019)
- Conformal prediction (classification)
- Conformal predictive systems (this work)
- See: https://github.com/predict-idlab/crepes-weighted

SCAN ME

crepes-weighted: crepes made weighted

```python
from sklearn.ensemble import RandomForestRegressor
from crepes_weighted import WrapRegressor

rf_wcps = WrapRegressor(RandomForestRegressor(n_estimators=100,
    random_state=17))
rf_wcps.fit(X_train_prop, y_train_prop)

```

crepes-weighted: crepes made weighted

```
from sklearn.ensemble import RandomForestRegressor
from crepes_weighted import WrapRegressor

rf_wcps = WrapRegressor(RandomForestRegressor(n_estimators=100,
    random_state=17))
rf_wcps.fit(X_train_prop, y_train_prop)

rf_wcps.calibrate(X_cal, y_cal, likelihood_ratios=
    shifted_likelihood_cal, cps=True)

```

crepes-weighted: crepes made weighted

```
from sklearn.ensemble import RandomForestRegressor
from crepes_weighted import WrapRegressor

rf_wcps = WrapRegressor(RandomForestRegressor(n_estimators=100,
    random_state=17))
rf_wcps.fit(X_train_prop, y_train_prop)

rf_wcps.calibrate(X_cal, y_cal, likelihood_ratios=
    shifted_likelihood_cal, cps=True)

int_wcps = rf_wcps.predict_int(X_test[idx_shift], y=y_test[idx_shift],
    likelihood_ratios=shifted_likelihood_test[idx_shift])
```

crepes-weighted: crepes made weighted

```
from sklearn.ensemble import RandomForestRegressor
from crepes_weighted import WrapRegressor

rf_wcps = WrapRegressor(RandomForestRegressor(n_estimators=100, random_state=17))
rf_wcps.fit(X_train_prop, y_train_prop)

rf_wcps.calibrate(X_cal, y_cal, likelihood_ratios=shifted_likelihood_cal, cps=True)

int_wcps = rf_wcps.predict_int(X_test[idx_shift], y=y_test[idx_shift], likelihood_ratios=shifted_likelihood_test[idx_shift])
dist_wcps = rf_wcps.predict_cps(X_test, y=y_test, likelihood_ratios=shifted_likelihood_test, return_cpds=True)
```

crepes-weighted: crepes made weighted

```
from sklearn.ensemble import RandomForestRegressor
from crepes_weighted import WrapRegressor

rf_wcps = WrapRegressor(RandomForestRegressor(n_estimators=100,
    random_state=17))
rf_wcps.fit(X_train_prop, y_train_prop)

rf_wcps.calibrate(X_cal, y_cal, likelihood_ratios=
    shifted_likelihood_cal, cps=True)

int_wcps = rf_wcps.predict_int(X_test[idx_shift], y=y_test[idx_shift],
    likelihood_ratios=shifted_likelihood_test[idx_shift])
dist_wcps = rf_wcps.predict_cps(X_test, y=y_test, likelihood_ratios=
    shifted_likelihood_test, return_cpds=True)
p_values_wcps = rf_wcps.predict_cps(X_test, y=y_test, likelihood_ratios
    =shifted_likelihood_test)
```

crepes-weighted: crepes made weighted

```python
from sklearn.ensemble import RandomForestRegressor
from crepes_weighted import WrapRegressor

rf_wcps = WrapRegressor(RandomForestRegressor(n_estimators=100, random_state=17))
rf_wcps.fit(X_train_prop, y_train_prop)

rf_wcps.calibrate(X_cal, y_cal, likelihood_ratios=shifted_likelihood_cal, cps=True)

int_wcps = rf_wcps.predict_int(X_test[idx_shift], y=y_test[idx_shift], likelihood_ratios=shifted_likelihood_test[idx_shift])
dist_wcps = rf_wcps.predict_cps(X_test, y=y_test, likelihood_ratios=shifted_likelihood_test, return_cpds=True)
p_values_wcps = rf_wcps.predict_cps(X_test, y=y_test, likelihood_ratios=shifted_likelihood_test)
```

Conclusion

Conclusion

- Successfully extended CPS to handle covariate shift
  - by introducing WCPS/WSCPS
- Practical implementation available via open-source Python package

Future work

- Formal proofs of theoretical conjectures
- Application to counterfactual inference
- Making CPS more conditionally probabilistically calibrated
- Weighted full CPS to `crepes-weighted`
- Dealing with prior (likelihood ratio)
  - imprecise probabilities, ...

References I

Gibbs, Isaac and Emmanuel Candes (2021). "Adaptive Conformal Inference Under Distribution Shift". In: *Advances in Neural Information Processing Systems*. Vol. 34, pp. 1660–1672.

Gibbs, Isaac and Emmanuel J. Candès (2024). "Conformal Inference for Online Prediction with Arbitrary Distribution Shifts". In: *Journal of Machine Learning Research* 25.162, pp. 1–36.

Kang, Joseph D. Y. and Joseph L. Schafer (2007). "Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data". In: *Statistical Science* 22.4, pp. 523–539.

Tibshirani, Ryan J et al. (2019). "Conformal Prediction Under Covariate Shift". In: *Advances in Neural Information Processing Systems*. Vol. 32.

Vovk, Vladimir, Alexander Gammerman, and Glenn Shafer (2022). *Algorithmic Learning in a Random World*.

Vovk, Vladimir et al. (2019). "Nonparametric predictive distributions based on conformal prediction". In: *Machine Language* 108.3, pp. 445–474.

References II

Yang, Yachong, Arun Kumar Kuchibhotla, and Eric Tchetgen Tchetgen (2022). *Doubly Robust Calibration of Prediction Sets under Covariate Shift*. arXiv:2203.01761 [math, stat].