*Black-box machine learning methods are now routinely used in high-risk settings, like medical diagnostics, which demand uncertainty quantification to avoid consequential model failures. Distribution-free uncertainty quantification (distribution-free UQ) is a user-friendly paradigm for creating statistically rigorous confidence intervals/sets for such predictions. Critically, the intervals/sets are valid without distributional assumptions or model assumptions, possessing explicit guarantees even with finitely many data points. Moreover, they adapt to the difficulty of the input; when the input example is difficult, the uncertainty intervals/sets are large, signaling that the model might be wrong. Without much work and without retraining, one can use distribution-free methods on any underlying algorithm, such as a neural network, to produce confidence sets guaranteed to contain the ground truth with a user-specified probability, such as 90%. Indeed, the methods are easy-to-understand and general, applying to many modern prediction problems arising in the fields of computer vision, natural language processing, deep reinforcement learning, and so on. This hands-on introduction is aimed at a reader interested in the practical implementation of distribution-free UQ who is not necessarily a statistician. We lead the reader through the practical theory and applications of distribution-free UQ, beginning with conformal prediction and culminating with distribution-free control of any risk, such as the false-discovery rate, false-positive rate of out-of-distribution detection, and so on. We will include many explanatory illustrations, examples, and code samples in Python, with PyTorch syntax. The goal is to provide the reader a working understanding of distribution-free UQ, allowing them to put confidence intervals on their algorithms, with one self-contained document.*

**How good is your prediction yˆ? If you are predicting the label y of a new instance, how confident are you that y = yˆ? In the case of regression where the label y is a number, how close do you think it is to yˆ?**

Every decision-making process powered by machine learning algorithms, it is crucial to support a prediction with a level of confidence.

*Indeed, as metrics provide us with a global level of confidence for machine learning algorithms, it is natural to look for similar local levels of confidence at the sample level.*

Bayesian models naturally give such local levels of confidence with posterior probabilities, and both aleatoric and epistemic uncertainties can be modeled. But it needs prior probabilities meaning distribution dependent.

Conformal prediction is a model agnostic (any model can incorporate conformal prediction)method to calculate uncertainty estimates for a supervised problem. The conformal prediction uses past experience to determine precise levels of confidence in predictions.

2 CP’s are there.

* Inductive conformal prediction (ICP). Model is created with training instances having labels and this model is used to predict test or unlabelled instances
* Transductive conformal prediction (TCP). It will have instances like semi-supervised (both labeled and unlabeled)meaning the model is created based on entire training and test data. When a new instance is provided again the entire model will be retrained. Computationally expensive.

For regression problems, conformal prediction changes the *point predictions to prediction intervals*. These *intervals surround the point estimate made by your model*, and the *size of the interval is directly tied to the degree of certainty that you want for your model.*

CP’s core mechanism is to find out non conformity score of calibration data and then report prediction intervals based on that score.Nonconformity Measure, *a function that gives some measure of dissimilarity between an example and a set of other examples. The higher the value of this function on an example, the more unlikely it is that this example belongs to the selected group of examples..intervals around predictions.*

*The underlying intuition is that inputs less similar to training data should lead to less certain estimates: this is captured by nonconformity scoring functions*

*Confidence Interval and Confidence Level*

When a confidence interval (CI) and confidence level (CL) are put together, the result is a statistically sound **spread of data**. For example, a result might be reported as “50% ± 6%, with a 95% confidence”. Let’s break apart the statistic into individual parts:

* The confidence interval: 50% ± 6% = 44% to 56%
* The confidence level: 95%

Confidence intervals are intrinsically connected to [confidence levels.](https://www.statisticshowto.com/confidence-level/) Confidence levels are expressed as a percentage (for example, a 90% confidence level). Should you repeat an experiment or survey with a 90% confidence level, we would expect that 90% of the time your results will match results you should get from a population. Confidence intervals are a range of results where you would expect the true value to appear. For example, you survey a group of children to see how many in-app purchases made a year. Your test is at the 99 percent confidence level and the result is a confidence interval of (250,300). That means you think they buy between 250 and 300 in-app items a year, and you’re confident that should the survey be repeated, 99% of the time the results will be the same.

*In both the regression and classification versions of the problem, the size of the output is controlled by the user. This parameter affecting the output is known as the significance.*

For eg: if you set the significance to 0.1, your sets and intervals will contain the actual class or the true value 90% of the time. This property means that the true value of a test instance will fall within your interval 90% of the time, or the true value of the test instance will be in the prediction

**Procedure to calculate CP(ICP):**

1. The data is split into three sets training, calibration, and test.
2. Train a model with training data set.
3. Predict the unseen calibration data set using this trained model.
4. Then non-conformity scores are calculated on the calibration set. Identify a prediction probability rule(quantile) using the unseen calibration data set scores.
5. Next, from these calibration scores and a user-specified significance level, conformal prediction intervals are constructed for new examples(test instances)
6. The intervals are calculated roughly by taking the nth percentile of the scores based on the user significance. Since the non-conformity scores for regression are the difference between the predicted calibration and true calibration, the conformity scores are the errors.

**Two notions help quantify uncertainty for conformal prediction at the sample level.**

* Credibility measures how likely the sample is to come from the train set, this is given by the minimal significance level such that the conformal region is empty.
* Confidence estimates how certain the model is that the prediction is a singleton.

**Short Comings and Concerns:**

1. Conformal prediction is defined under an exchangeability assumption. This assumption means that *conformal prediction assumes that the data you use to train the data, validate and test the data are exchangeable* (The data you select for training your model you could have put into your test set. Or consider the data you used for calibration. Again, this data could have been used for training or testing.)

*For the data to be exchangeable, the entire dataset must come from the same distribution. While this aspect is a given for many datasets and future instances, data distributions change over time for other datasets. Conformal prediction can still provide* *some uncertainty estimation in these cases, but the theoretical guarantees are no longer assured.*

1. Uncertainty estimation has a trade-off. For example, suppose you want to have 90% guarantees that your prediction interval contains the true value. In that case, you are going to have larger intervals. But if you only want to be 50% certain, then your intervals will be lesser.More confidence larger intervals,less confidence smaller intervals

Questions:

1. How do you are defining that these models are having uncertainties? Once we confirm that model is having uncertainty, then we can say that we are using conformal predictions to overcome those uncertainties.

Needs:(25th min)

1.Uncertainties from Deep DB.(No need to run conformal predictions on this uncertainties.We can do it if we want to just to compare uncertainties reported by conformal predictions to uncertainties estimated by Deep DB)

—--------------------- 3 csv’s—------------------------------

2.Predictions of job-light from MSCN and uncertainities from Bayesian

3.Predictions of job-light from Neuro Card and uncertanties from Bayesian

4.Predictions and uncertainties from Deep DB(uncertainties not from Bayesian but from Deep DB itself)

References:

<https://blog.dataiku.com/measuring-models-uncertainty-conformal-prediction>

<https://towardsdatascience.com/how-to-add-uncertainty-estimation-to-your-models-with-conformal-prediction-a5acdb86ea05>

<https://www.quora.com/What-is-the-difference-between-inductive-and-transductive-learning> - the difference between inductive and transductive

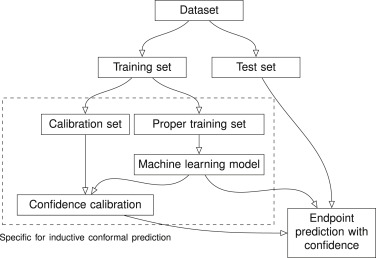
<https://link.springer.com/article/10.1007/s10994-014-5453-0>

<https://www.sciencedirect.com/science/article/pii/S089360801100150X?casa_token=Kk3lQafpWz8AAAAA:KL01O-9CcIytjGhQMrShQc3YCeDVUh0PDtWdtslF5WChkTCbU57hmohdoA5x4b8-XMFx9UaI4cI> – anaspaulous paper

<https://uq360.readthedocs.io/en/latest/regression_metrics.html>

https://colab.research.google.com/drive/1uRsj9GvF-MJrnEYvBczqSBDnJpTmFWAX?usp=sharing

<https://arxiv.org/abs/2010.09107> <https://github.com/hamrel-cxu/EnbPI>



# Predicting With Confidence: Using Conformal Prediction in Drug Discovery

<https://link.springer.com/article/10.1007/s10994-014-5453-0> - algorithm and various conformity scores

<https://www.diva-portal.org/smash/get/diva2:1288708/FULLTEXT01.pdf> - code inspiration

<https://arxiv.org/ftp/arxiv/papers/1908/1908.03569.pdf> - good read to understand statistically

background-

uc technqs in details--bysn,cnfrml,models 3--in detail

problem statement--to be asked

research qstn--to be asked

arch design--5 arch and explanation

exp set up-- all the exp we have done(hyperparam tuning)

ensemble

result-q errors,intervals, all diagrams/graphs