



UiO : **Department of Mathematics**
University of Oslo

Conformal Inference

A brief introduction

Anders Hjort

April 22, 2022

About this presentation

This presentation was part of a special syllabus PhD course the Department of Mathematics at the University of Oslo. The course syllabus was the following three seminal works on conformal inference:

- 1 Vladimir Vovk, Alex Gammerman, and Glenn Shafer. 2005.
Algorithmic Learning in a Random World (selected pages).
Springer Science+Business Media, Inc
- 2 Vladimir Vovk and Glenn Shafer. 2008.
A Tutorial on Conformal Prediction.
Journal of Machine Learning Research (9).
- 3 Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan Tibshirani and Larry Wasserman. 2018.
Distribution-Free Predictive Inference For Regression.
Journal of the American Statistical Association. 113.

About the presentation

The presentation will

- give a **practical and intuitive** introduction rather than go through very theoretical proofs
- focus primarily on **regression** rather than classification
- focus on **one-dimensional examples** where $y \in \mathbb{R}$ and $x \in \mathbb{R}$

Furthermore, the notation and terminology being used follows that of Lei et al. 2018 rather than that of Vovk et al. 2005.

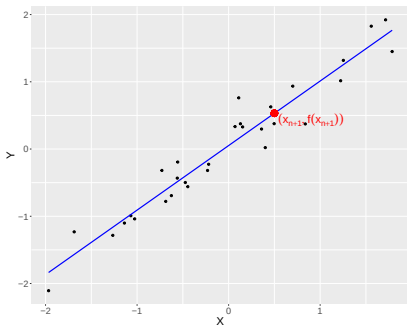
Table of contents

- 1 Introduction and intuition**
- 2 Full Conformal Inference**
- 3 Split Conformal Inference**
- 4 Normalized Conformal Inference**
- 5 Examples**
 - A linear example
 - A non-linear example
- 6 Conclusion**
- 7 Appendix**

Table of Contents

- 1 Introduction and intuition**
- 2 Full Conformal Inference
- 3 Split Conformal Inference
- 4 Normalized Conformal Inference
- 5 Examples
 - A linear example
 - A non-linear example
- 6 Conclusion
- 7 Appendix

Intuition



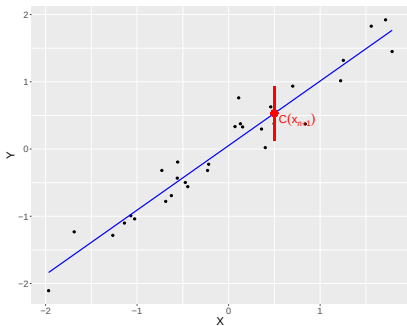
We have some training data points $(X_i, Y_i), i = 1, \dots, n$ and we have created a prediction model $\hat{Y} = f(X)$ that is trained on these. We use the model to obtain the prediction $\hat{Y}_{n+1} = f(X_{n+1})$.

How certain are we that $Y_{n+1} = \hat{Y}_{n+1}$? Instead of a point prediction, we want to create a **confidence region** $C(X_{n+1})$ such that

$$P(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha$$

for some $0 < \alpha < 1$.

Intuition



We have some training data points $(X_i, Y_i), i = 1, \dots, n$ and we have created a prediction model $\hat{Y} = f(X)$ that is trained on these. We use the model to obtain the prediction $\hat{Y}_{n+1} = f(X_{n+1})$.

How certain are we that $Y_{n+1} = \hat{Y}_{n+1}$? Instead of a point prediction, we want to create a **confidence region** $C(X_{n+1})$ such that

$$P(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha$$

for some $0 < \alpha < 1$.

Intuition

Satisfying $P(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha$ for some α is not hard. For instance, $C(X_{n+1}) = \mathbb{R}$ guarantees this.

Intuition

Satisfying $P(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha$ for some α is not hard. For instance, $C(X_{n+1}) = \mathbb{R}$ guarantees this.

We evaluate the confidence region primarily in two ways:

- **Empirical validity:** Are the confidence intervals valid at the desired $(1 - \alpha)$ level?
- **Efficiency:** How "tight" are the confidence regions?

A simple result about sample quantiles

- Let $U_1, U_2, \dots, U_n$ be i.i.d. samples of a random variable (we don't know from where).
- Let $U_{(1)}, U_{(2)}, \dots, U_{(n)}$ denote the order statistics of $U_1, U_2, \dots, U_n$ such that $U_{(1)}$ is the smallest value and $U_{(n)}$ is the largest value.

A simple result about sample quantiles

- Let $U_1, U_2, \dots, U_n$ be i.i.d. samples of a random variable (we don't know from where).
- Let $U_{(1)}, U_{(2)}, \dots, U_{(n)}$ denote the order statistics of $U_1, U_2, \dots, U_n$ such that $U_{(1)}$ is the smallest value and $U_{(n)}$ is the largest value.
- Define now $\hat{q}_{1-\alpha}$ to be the $(1 - \alpha)$ th quantile of the observations;
$$\hat{q}_{1-\alpha} = U_{(\lceil (n+1)(1-\alpha) \rceil)}$$

A simple result about sample quantiles

- Let $U_1, U_2, \dots, U_n$ be i.i.d. samples of a random variable (we don't know from where).
- Let $U_{(1)}, U_{(2)}, \dots, U_{(n)}$ denote the order statistics of $U_1, U_2, \dots, U_n$ such that $U_{(1)}$ is the smallest value and $U_{(n)}$ is the largest value.
- Define now $\hat{q}_{1-\alpha}$ to be the $(1 - \alpha)$ th quantile of the observations;
$$\hat{q}_{1-\alpha} = U_{(\lceil (n+1)(1-\alpha) \rceil)}$$

A quick result:

We draw a new U_{n+1} from the same distribution as $U_1, U_2, \dots, U_n$. Now,

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

A simple result about sample quantiles

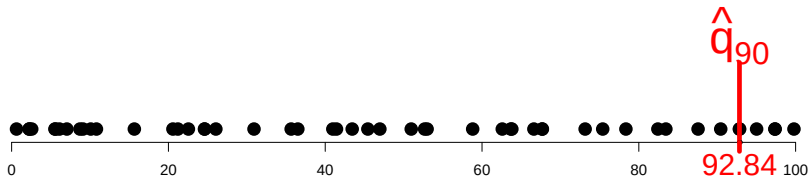


Figure: $U_1, \dots, U_{100}$ marked on a number line with $\hat{q}_{90} = 92.84$ in red. If a new data point U_{n+1} is drawn from the same distribution, we know that that $P(U_{n+1} \leq 92.84) \geq 0.9$.

Table of Contents

- 1 Introduction and intuition
- 2 Full Conformal Inference**
- 3 Split Conformal Inference
- 4 Normalized Conformal Inference
- 5 Examples
 - A linear example
 - A non-linear example
- 6 Conclusion
- 7 Appendix

Full CI algorithm

For any given *proposed value* y we check if it should be included in the confidence set.

- Create an augmented data set $(X_1, Y_1), \dots, (X_n, Y_n), (X_{n+1}, y)$;

Full CI algorithm

For any given *proposed value* y we check if it should be included in the confidence set.

- Create an augmented data set $(X_1, Y_1), \dots, (X_n, Y_n), (X_{n+1}, y)$;
- Train a model on this augmented data set, denoted $f_y(x)$;

Full CI algorithm

For any given *proposed value* y we check if it should be included in the confidence set.

- Create an augmented data set $(X_1, Y_1), \dots, (X_n, Y_n), (X_{n+1}, y)$;
- Train a model on this augmented data set, denoted $f_y(x)$;
- Calculate the residuals $R_{y,i} = |Y_i - f_y(X_i)|$ for all observations in the training set and $R_{y,n+1} = |y - f_y(X_{n+1})|$ for the new instance X_{n+1} ;

Full CI algorithm

For any given *proposed value* y we check if it should be included in the confidence set.

- Create an augmented data set $(X_1, Y_1), \dots, (X_n, Y_n), (X_{n+1}, y)$;
- Train a model on this augmented data set, denoted $f_y(x)$;
- Calculate the residuals $R_{y,i} = |Y_i - f_y(X_i)|$ for all observations in the training set and $R_{y,n+1} = |y - f_y(X_{n+1})|$ for the new instance X_{n+1} ;
- Rank $R_{y,n+1}$ among the residuals $R_{y,1}, \dots, R_{y,n}$. The rank is defined as

$$\pi(y) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbb{1}\{R_{y,i} \leq R_{y,n+1}\}$$

Full CI algorithm

For any given *proposed value* y we check if it should be included in the confidence set.

- Create an augmented data set $(X_1, Y_1), \dots, (X_n, Y_n), (X_{n+1}, y)$;
- Train a model on this augmented data set, denoted $f_y(x)$;
- Calculate the residuals $R_{y,i} = |Y_i - f_y(X_i)|$ for all observations in the training set and $R_{y,n+1} = |y - f_y(X_{n+1})|$ for the new instance X_{n+1} ;
- Rank $R_{y,n+1}$ among the residuals $R_{y,1}, \dots, R_{y,n}$. The rank is defined as

$$\pi(y) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbb{1}\{R_{y,i} \leq R_{y,n+1}\}$$

- Include y in $C(X_{n+1})$ if and only if

$$\pi(y) \leq \frac{\lceil (1 - \alpha)(n + 1) \rceil}{n + 1}.$$

Full CI algorithm (continued)

If we repeat this over all $y \in \mathbb{R}$ we can construct

$$C(X_{n+1}) = \left\{ y \in \mathbb{R} : \pi(y) \leq \frac{\lceil (1 - \alpha)(n + 1) \rceil}{n + 1} \right\} \quad (1)$$

with the guarantee that

$$P(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha.$$

Example

Let $\alpha = 0.1$ and $n = 99$. Then (1) is equivalent to

$$C(X_{n+1}) = \left\{ y \in \mathbb{R} : \pi(y) \leq 0.9 \right\},$$

or simply put: "Include all the y values that yield a residual $R_{y,n+1}$ that is smaller than the 90th percentile of the other residuals $R_{y,i}$."

A Simple Example

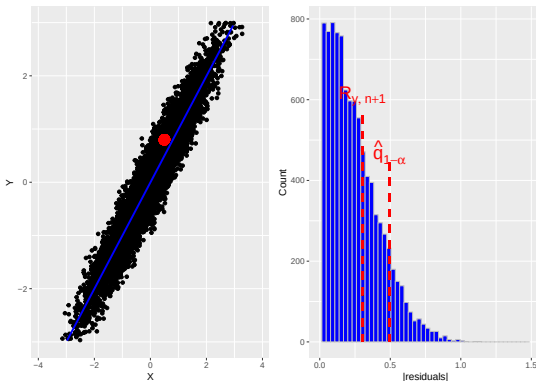


Figure: A data set of $(X_i, Y_i), i = 1, \dots, 10\,000$. We seek to create a confidence region for $C(X_{n+1} = 0.5)$ and propose $y = 0.8$ as a trial value. Since $R_{y, n+1} < \hat{q}_{1-\alpha}$ we accept $y = 0.8$ in $C(X_{n+1})$.

A Simple Example II

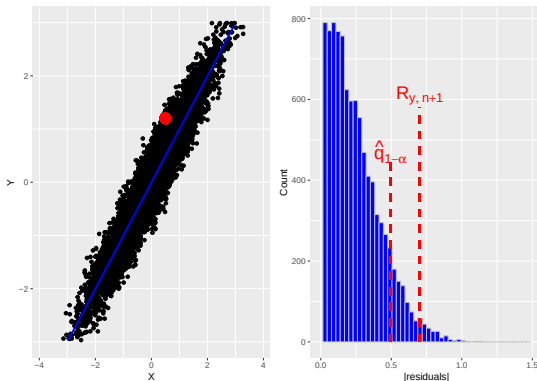


Figure: The same setting, but with a new trial value $y = 1.2$. This time we reject the value.

Why does it work?

The ranking $\pi(Y_{n+1})$ is uniformly distributed over $\{\frac{1}{n+1}, \frac{2}{n+1}, \dots, 1\}$ regardless of what kind of distribution the underlying data is drawn from. This means that

$$P\left(\pi(Y_{n+1}) \leq \frac{\lceil (1 - \alpha)(n + 1) \rceil}{n + 1}\right) \geq 1 - \alpha. \quad (2)$$

Why does it work?

The ranking $\pi(Y_{n+1})$ is uniformly distributed over $\{\frac{1}{n+1}, \frac{2}{n+1}, \dots, 1\}$ regardless of what kind of distribution the underlying data is drawn from. This means that

$$P\left(\pi(Y_{n+1}) \leq \frac{\lceil (1 - \alpha)(n + 1) \rceil}{n + 1}\right) \geq 1 - \alpha. \quad (2)$$

This can be interpreted as saying $1 - \pi(Y_{n+1})$ provides a p -value for the hypothesis test $H_0 : Y_{n+1} = y$.

Assumptions

We assume [exchangeable](#) data. Shafer and Vovk 2007 introduce it as follows:

Exchangeability (informally)

Consider variables $Z_1, \dots, Z_N$. Suppose that for any collection of N values, the $N!$ different orderings are equally likely. Then we say that $Z_1, \dots, Z_N$ are *exchangeable*.

Assumptions

We assume **exchangeable** data. Shafer and Vovk 2007 introduce it as follows:

Exchangeability (informally)

Consider variables $Z_1, \dots, Z_N$. Suppose that for any collection of N values, the $N!$ different orderings are equally likely. Then we say that $Z_1, \dots, Z_N$ are *exchangeable*.

We furthermore assume that regression algorithm used to produce $f(\cdot)$ treats the data points **symmetrically**.

Summary so far

In a nutshell:

For any new example (X_{n+1}, y) we calculate its similarity (conformity) with the observed data $(X_1, Y_1), \dots, (X_n, Y_n)$. We include the new example in the confidence region if the conformity is sufficient enough (depending on α).

Summary so far

In a nutshell:

For any new example (X_{n+1}, y) we calculate its similarity (conformity) with the observed data $(X_1, Y_1), \dots, (X_n, Y_n)$. We include the new example in the confidence region if the conformity is sufficient enough (depending on α).

A problem:

The procedure is extremely computationally heavy. We need to check a large number of y trial values to construct $C(X_{n+1})$.

Table of Contents

- 1 Introduction and intuition
- 2 Full Conformal Inference
- 3 Split Conformal Inference**
- 4 Normalized Conformal Inference
- 5 Examples
 - A linear example
 - A non-linear example
- 6 Conclusion
- 7 Appendix

Split conformal inference

Instead of the full conformal inference procedure, consider the following:

- Split the original training data set into two: a *training data set* $\mathcal{I}_1$ and a *calibration data set* $\mathcal{I}_2$

Split conformal inference

Instead of the full conformal inference procedure, consider the following:

- Split the original training data set into two: a *training data set* $\mathcal{I}_1$ and a *calibration data set* $\mathcal{I}_2$
- Train a predictive model $f(x)$ on $\mathcal{I}_1$

Split conformal inference

Instead of the full conformal inference procedure, consider the following:

- Split the original training data set into two: a *training data set* $\mathcal{I}_1$ and a *calibration data set* $\mathcal{I}_2$
- Train a predictive model $f(x)$ on $\mathcal{I}_1$
- Apply $f(x)$ to the calibration set and obtain a set of residuals $R_i = |f(X_i) - Y_i| \forall i \in \mathcal{I}_2$

Split conformal inference

Instead of the full conformal inference procedure, consider the following:

- Split the original training data set into two: a *training data set* $\mathcal{I}_1$ and a *calibration data set* $\mathcal{I}_2$
- Train a predictive model $f(x)$ on $\mathcal{I}_1$
- Apply $f(x)$ to the calibration set and obtain a set of residuals $R_i = |f(X_i) - Y_i| \forall i \in \mathcal{I}_2$
- Let $\hat{q}_{1-\alpha}$ be the $(1 - \alpha)$ th percentile of R_i

Split conformal inference

Instead of the full conformal inference procedure, consider the following:

- Split the original training data set into two: a *training data set* $\mathcal{I}_1$ and a *calibration data set* $\mathcal{I}_2$
- Train a predictive model $f(x)$ on $\mathcal{I}_1$
- Apply $f(x)$ to the calibration set and obtain a set of residuals $R_i = |f(X_i) - Y_i| \forall i \in \mathcal{I}_2$
- Let $\hat{q}_{1-\alpha}$ be the $(1 - \alpha)$ th percentile of R_i
- We want to predict Y_{n+1} from a test set $\mathcal{I}_3$. We create the confidence region

$$C_{split}(X_{n+1}) = [f(X_{n+1}) \pm \hat{q}_{1-\alpha}]$$

Split conformal inference

As long as the training set and the calibration set are exchangeable, we can still guarantee that

$$P(Y_{n+1} \in C_{split}(X_{n+1})) \geq 1 - \alpha. \quad (3)$$

Note that (with both methods) we only get **marginal coverage**. It would be more useful to get guaranteed **conditional coverage**, i.e.

$$P(Y_{n+1} \in C(X_{n+1}) | X_{n+1}) \geq 1 - \alpha.$$

Unfortunately, Lei and Wasserman 2014 showed that finite sample conditional validity is impossible.¹ This distinction is very important!

¹Technically, they proved that the only time this is possible is in the trivial case where the size of the interval goes towards ∞ .

A Simple Example

We generate $N = 10\,000$ observations from

$$X \sim \mathcal{N}(0, 1), \quad \varepsilon \sim \mathcal{N}(0, e^{0.6 \cdot X}) \quad Y = X + \varepsilon.$$

Randomly sample 2/3 of the observations for the training set and 1/3 to the test set. The training set is split equally into $\mathcal{I}_1$ and $\mathcal{I}_2$. We want to get confidence region that are valid at $\alpha = 0.1$.

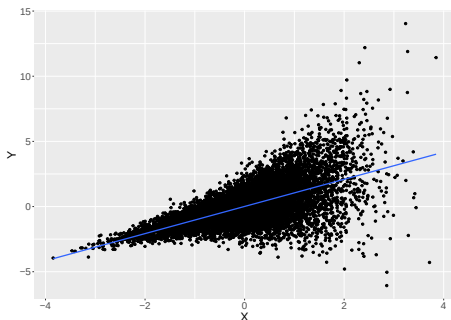
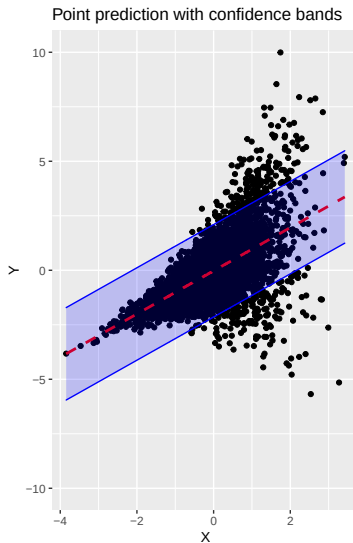
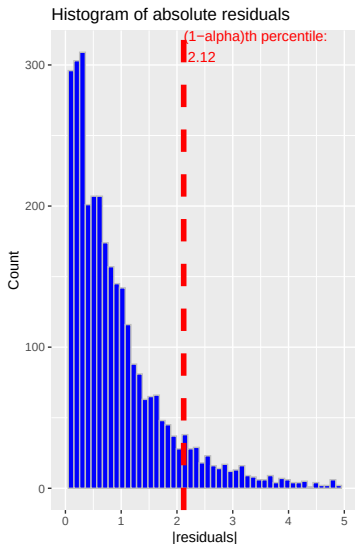


Figure: The generated data set.

A Simple Example



Summary so far

In a nutshell:

Split conformal inference solves the computational issues with full conformal inference by separating the fitting and the ranking. As long as we have exchangeable data we still get valid confidence regions.

A problem:

The confidence regions will have the same width for every part of the feature space. Not very informative.

Table of Contents

- 1 Introduction and intuition
- 2 Full Conformal Inference
- 3 Split Conformal Inference
- 4 Normalized Conformal Inference**
- 5 Examples
 - A linear example
 - A non-linear example
- 6 Conclusion
- 7 Appendix

Normalized Conformal Inference

In order to introduce confidence regions that are of different size, we apply a more general *non-conformity measure*:

$$R_i = \frac{|Y_i - f(X_i)|}{\hat{\rho}(X_i)}, \quad (4)$$

where $\hat{\rho}(X_i)$ is sometimes referred to as the **difficulty estimation function**.²

²Lei et al. 2018 note that $\hat{\rho}(x_i)$ is an estimate of the conditional mean absolute deviation of $(Y - f(X))|X = x$.

Normalized Conformal Inference

In order to introduce confidence regions that are of different size, we apply a more general *non-conformity measure*:

$$R_i = \frac{|Y_i - f(X_i)|}{\hat{\rho}(X_i)}, \quad (4)$$

where $\hat{\rho}(X_i)$ is sometimes referred to as the **difficulty estimation function**.²

We create confidence regions

$$C_{\text{normalized}}(X_{n+1}) = \left[\underbrace{f(X_{n+1})}_{\text{point prediction}} \pm \underbrace{\hat{q}_{1-\alpha}}_{\text{a fixed value}} \cdot \underbrace{\hat{\rho}(X_{n+1})}_{\text{difficulty est.}} \right],$$

where $\hat{q}_{1-\alpha}$ is once again the $(1 - \alpha)$ quantile of R_i from the calibration set .

²Lei et al. 2018 note that $\hat{\rho}(x_i)$ is an estimate of the conditional mean absolute deviation of $(Y - f(X))|X = x$.

Normalized conformal inference

How can we find a suitable function for $\hat{p}(x_i)$? Note that

- Learn $\hat{p}(X_i)$ by fitting e.g. a polynomial to $|Y_i - f(X_i)|$;
- Use some kind of model-specific difficulty function, like proposed by Boström et al. 2017 for random forest models;
- Domain knowledge;
- $\hat{p}(X_i) = \hat{Y}_i$ or $\hat{p}(X_i) = \exp(\gamma \cdot \hat{Y}_i)$ are proposed by Lim and Bellotti 2021 for the application of Automated Valuation Models.

Note: The difficulty function should be learned on the training data $\mathcal{I}_1$, not the calibration data $\mathcal{I}_2$!

Note: The validity guarantees from split conformal inference carries over, since the result about sample quantiles holds also for normalized residuals!

Table of Contents

- 1 Introduction and intuition
- 2 Full Conformal Inference
- 3 Split Conformal Inference
- 4 Normalized Conformal Inference
- 5 **Examples**
 - A linear example
 - A non-linear example
- 6 Conclusion
- 7 Appendix

A Simple Example (Revisited)

We generate $N = 10\,000$ observations from

$$X \sim \mathcal{N}(0, 1), \quad \varepsilon \sim \mathcal{N}(0, e^{0.6 \cdot X}) \quad Y = X + \varepsilon.$$

Split equally between training set, calibration set and test set (1/3 in each). We want to get confidence region that are valid at $\alpha = 0.1$.

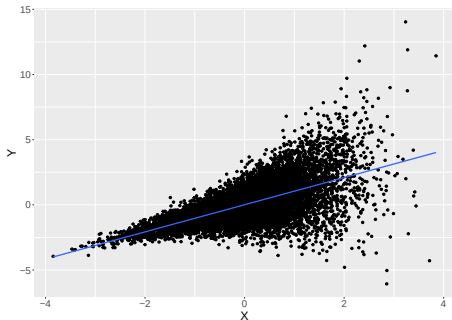
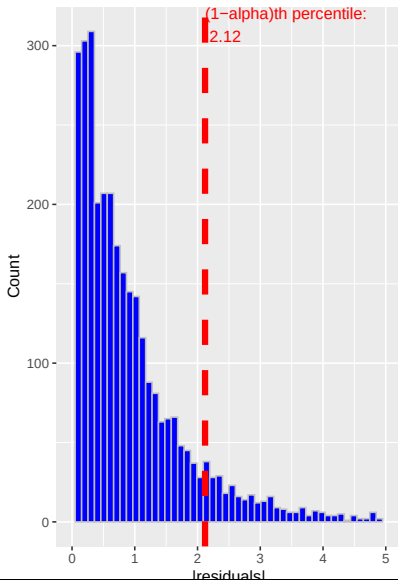


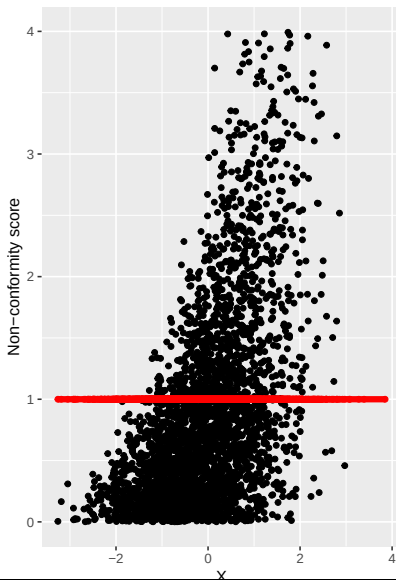
Figure: The generated data set.

Residuals 1: No difficulty function

Histogram of absolute residuals

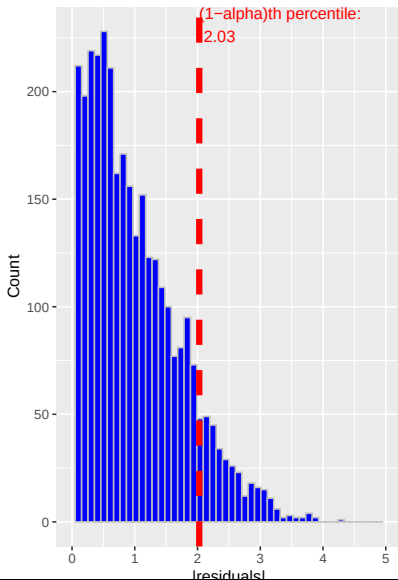


Residuals and difficulty function

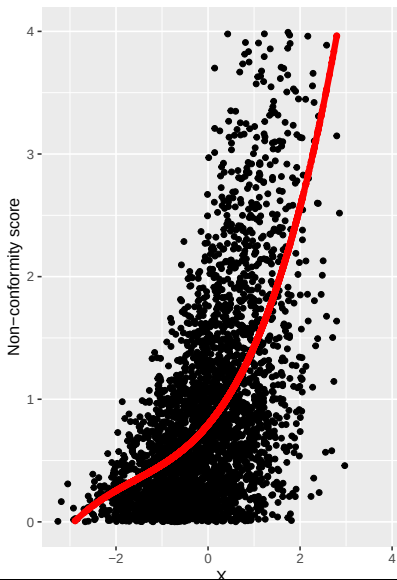


Residuals 2: Polynomial difficulty function

Histogram of absolute residuals

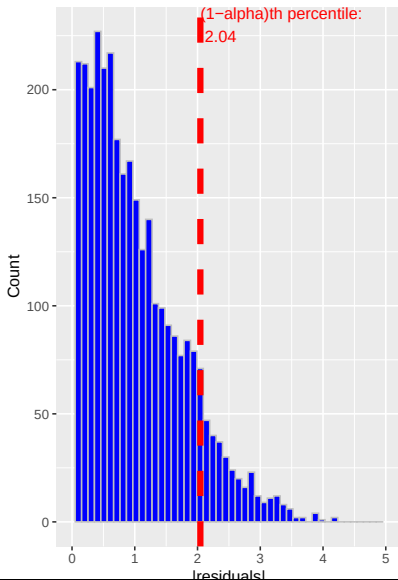


Residuals and difficulty function

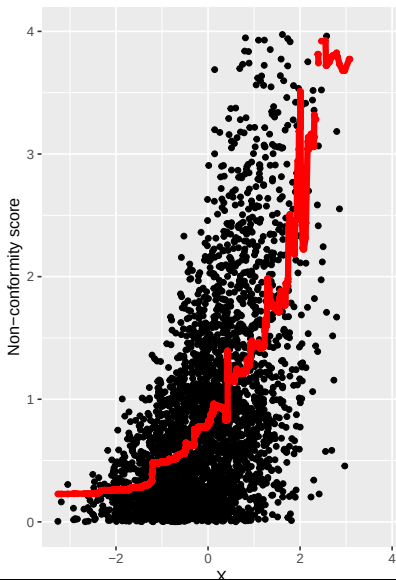


Residuals 3: Random forest difficulty

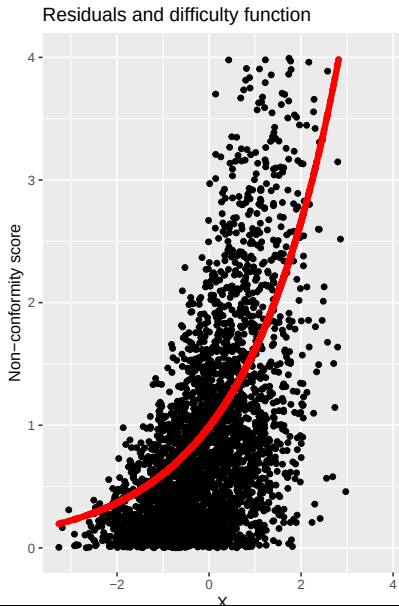
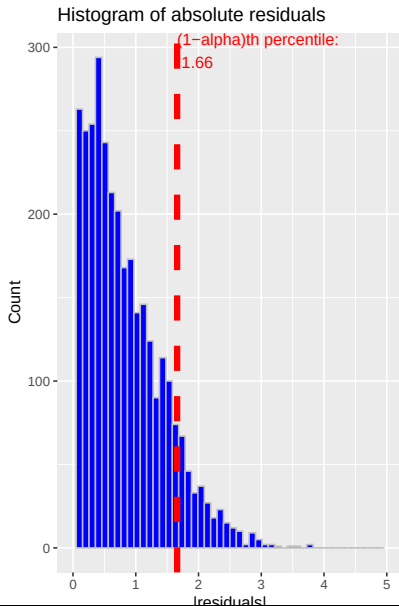
Histogram of absolute residuals



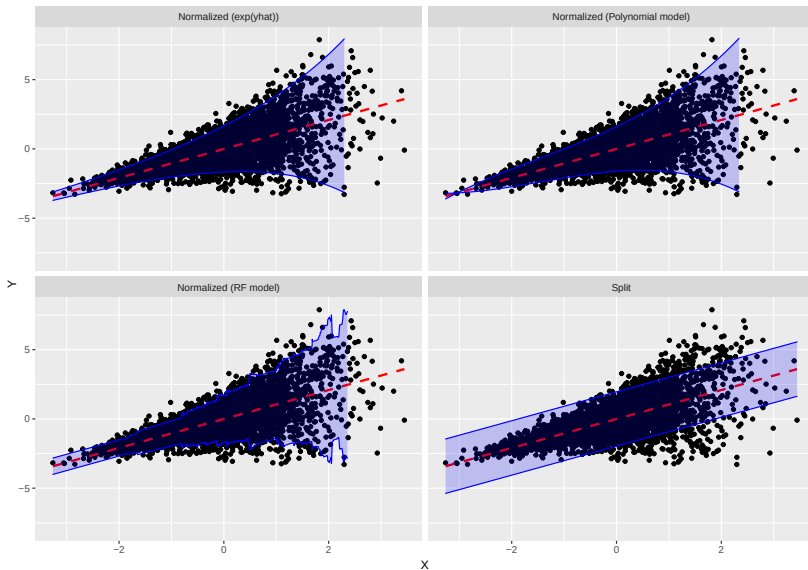
Residuals and difficulty function



Residuals 4: $\exp(0.6 \cdot \hat{Y})$ as difficulty



Results on test set



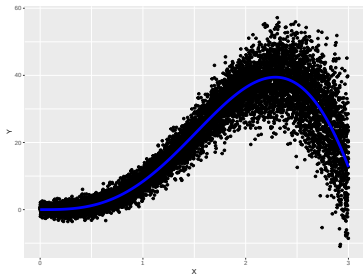
Which method is best?

Difficulty estimation method	Empirical coverage	Mean Interval Length
None	0.896	4.24
Polynomial	0.891	3.82
Random forest	0.892	3.88
$\exp(0.6 \cdot \hat{Y})$	0.891	3.69

Table: Evaluation of the different methods for creating confidence regions with split conformal inference. The test set used has $N_{test} = 3\,333$.

Non-linear example

Generate $N = 10\,000$ observations from $X \sim U(0, 3)$, $\varepsilon \sim \mathcal{N}(0, e^{0.8 \cdot X})$, $Y = 10X^2 \cdot \sin X + \varepsilon$. Split equally between training set, calibration set and test set (1/3 in each). We want to get confidence region that are valid at $\alpha = 0.1$.



Underlying prediction model $f(X)$ is a quite simple random forest with 100 trees of depth 4. Try two different difficulty functions: $\hat{\rho} = 1$ (no difficulty) and $\hat{\rho}(X)$ fitted by a third degree polynomial.

No difficulty function

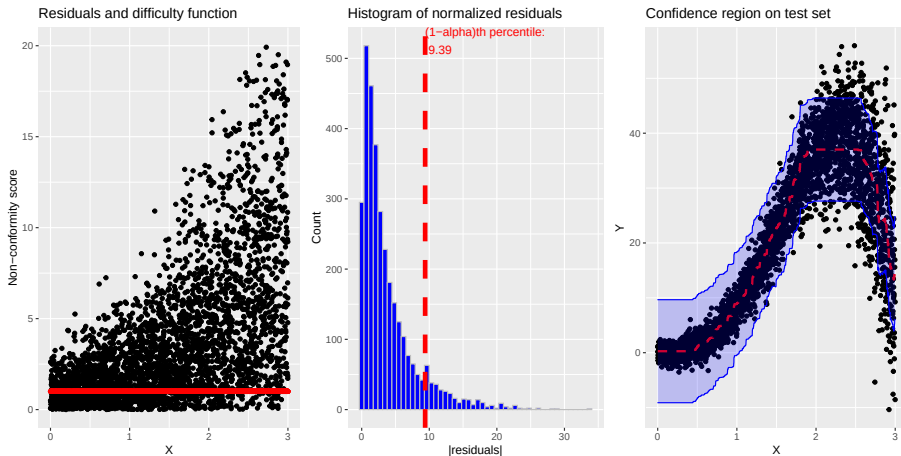


Figure: **Left:** Absolute residuals as a function of X with $\hat{p}(X) = 1$ in red. **Middle:** Histogram of residuals with $\hat{q}_{90}$ highlighted. **Right:** Point prediction with confidence regions on a test set.

Polynomial difficulty function

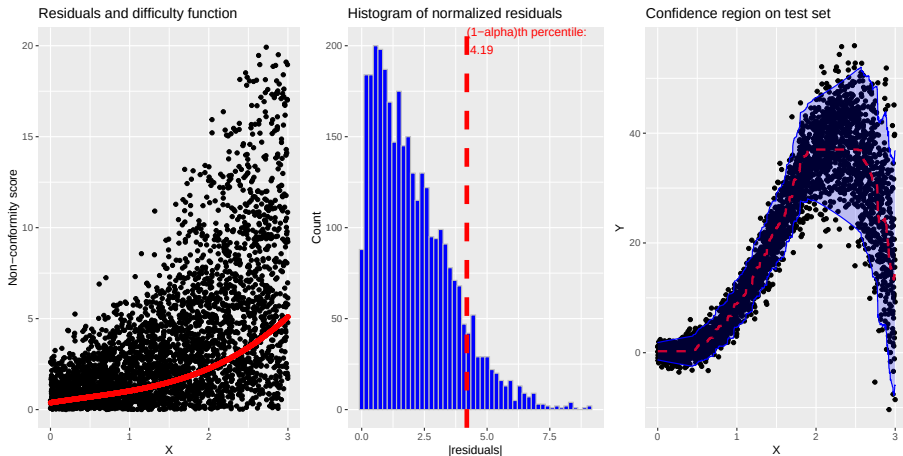


Figure: **Left:** Absolute residuals as a function of X with third degree polynomial $\hat{p}(X)$ in red. **Middle:** Histogram of residuals with $\hat{q}_{90}$ highlighted. **Right:** Point prediction with confidence regions on a test set.

Which method is best?

Difficulty estimation method	Empirical coverage	Mean Interval Length
None	0.92	18.8
Polynomial	0.943	16.0

Table: Evaluation of the different methods for creating confidence regions with split conformal inference. The test set used has $N_{test} = 3\,333$.

Table of Contents

- 1 Introduction and intuition
- 2 Full Conformal Inference
- 3 Split Conformal Inference
- 4 Normalized Conformal Inference
- 5 Examples
 - A linear example
 - A non-linear example
- 6 Conclusion**
- 7 Appendix

Conclusion

- Conformal inference is a framework that allows us to create confidence regions around predictions **regardless of the underlying algorithm**
- Full conformal inference was introduced first by Vovk et al. 2005, but the focus in most modern applications and research is on **split conformal inference**.
- **Normalized conformal inference** makes it possible to create confidence regions of different size for different parts of the feature space, but great effort and care must be put into crafting a suitable difficulty function
- As long as the data in the training set, calibration set and test set are **exchangeability**, the confidence regions are guaranteed to be valid
- It must be stressed that we get **marginal coverage**, not conditional coverage (unfortunately)

Acknowledgements and disclaimer

- **Disclaimer:** I am far from an expert on this topic. This is my attempt at summarizing the works of Vovk et al. 2005, Shafer and Vovk 2007 and Lei et al. 2018. Many of the phrases in this presentation is therefore copied directly from one or many of these works
- The presentation is also somewhat inspired by other presentations on the topic, for instance Ryan Tibshirani's very insightful presentation titled "Advances and Challenges in Conformal Inference " ³
- Thanks also to Ryan Tibshirani and Henrik Boström for responding to my questions about conformal inference

³<https://www.stat.cmu.edu/~ryantibs/talks/conformal-2019.pdf>

Table of Contents

- 1 Introduction and intuition
- 2 Full Conformal Inference
- 3 Split Conformal Inference
- 4 Normalized Conformal Inference
- 5 Examples
 - A linear example
 - A non-linear example
- 6 Conclusion
- 7 Appendix**

References I

-  Angelopoulos, Anastasios N. and Stephen Bates (2021). ‘A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification’. In: *CoRR* abs/2107.07511.
-  Boström, Henrik, Henrik Linusson, Tuve Löfström and Ulf Johansson (2017). ‘Accelerating difficulty estimation for conformal regression forests’. In: *Annals of Mathematics and Artificial Intelligence* 81.
-  Lei, Jing, Max G’Sell, Alessandro Rinaldo, Ryan J. Tibshirani and Larry Wasserman (2018). ‘Distribution-Free Predictive Inference for Regression’. In: *Journal of the American Statistical Association* 113.523, pp. 1094–1111.
-  Lei, Jing and Larry Wasserman (2014). ‘Distribution-free prediction bands for non-parametric regression’. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.

References II

-  Lim, Zhe and Anthony Bellotti (2021). ‘Normalized nonconformity measures for automated valuation models’. In: *Expert Systems with Applications* 180, pp. 115–165.
-  Shafer, Glenn and Vladimir Vovk (July 2007). ‘A tutorial on conformal prediction’. In: *Journal of Machine Learning Research* 9.
-  Vovk, Vladimir, Alex Gammerman and Glenn Shafer (2005). *Algorithmic Learning in a Random World*. Springer-Verlag, Heidelberg.

UiO : Department of Mathematics

University of Oslo



Anders Hjort



Conformal Inference

A brief introduction

