Predicting With Uncertainty

Chris Harbron
MCO, Roche

PSI Conference : June 2025

Predicting With Uncertainty

Predictive Models

Personal Uncertainty Quantification

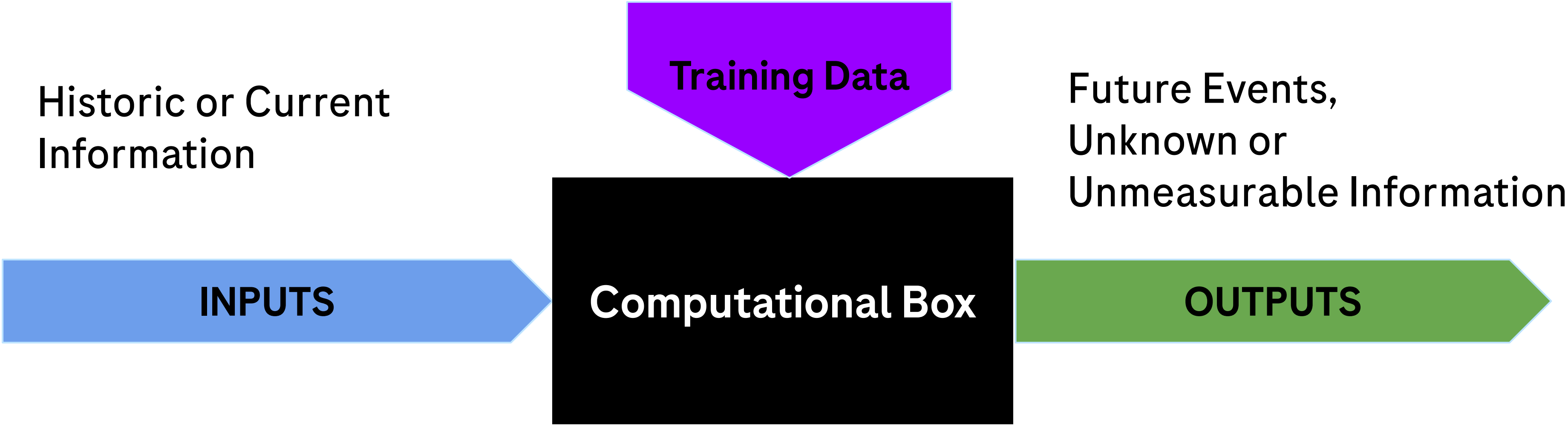
Uncertainty In Statistical Models

Uncertainty By Full Conformal Prediction

Uncertainty By Split Conformal Prediction

What Is A Predictive Model?

Historic or Current Information

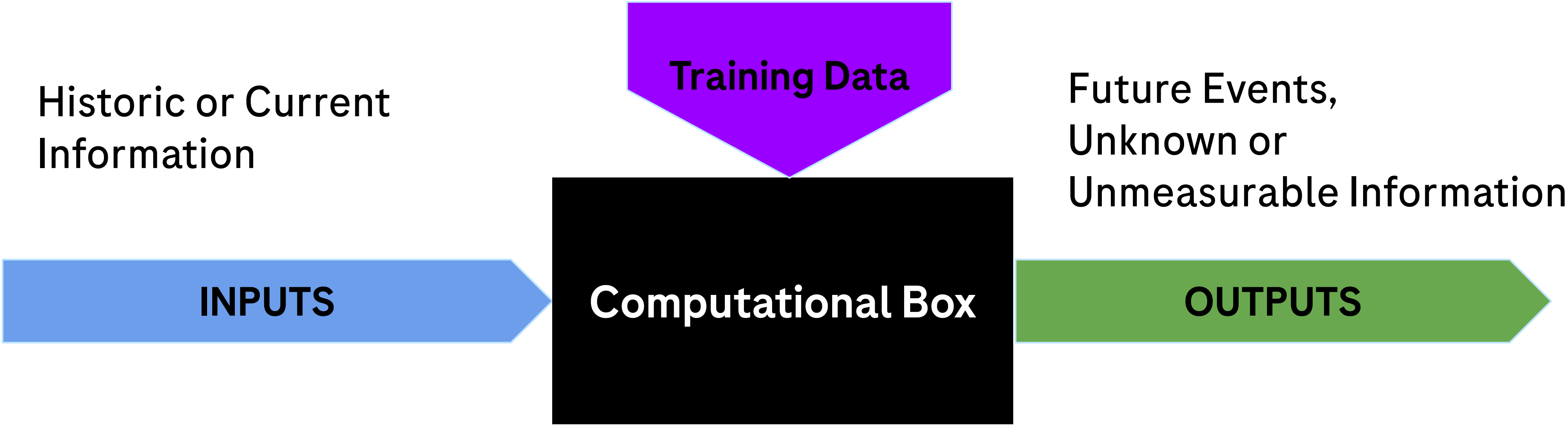


Future Events, Unknown or Unmeasurable Information

- Patient characteristics
- Medical history
- Diagnostic Images
- Omics

- Diagnosis
- Probability of response
- Risk score

What Is A Predictive Model?



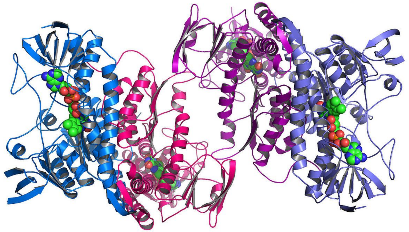
This computational box - the predictive model - could take many mathematical forms :

- Statistical modelling
- Machine learning
- Deep learners (AI)

Roche

Drivers For Development Of Predictive Models In HealthCare

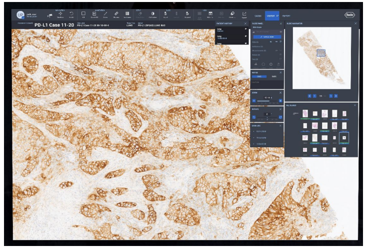
New Measurement Techniques



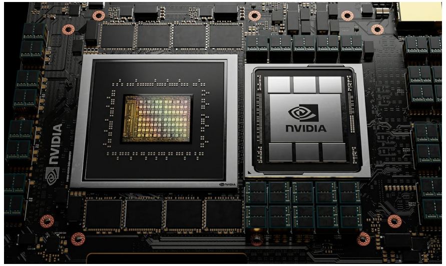
Increasing Availability Of Data At Scale

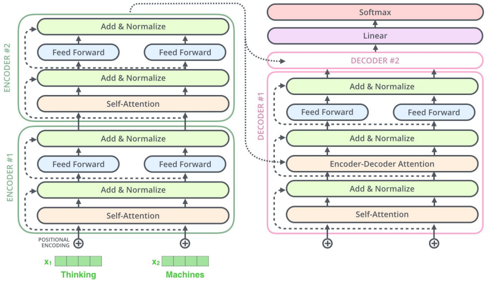
Genomics england

flatiron



Increased Computation



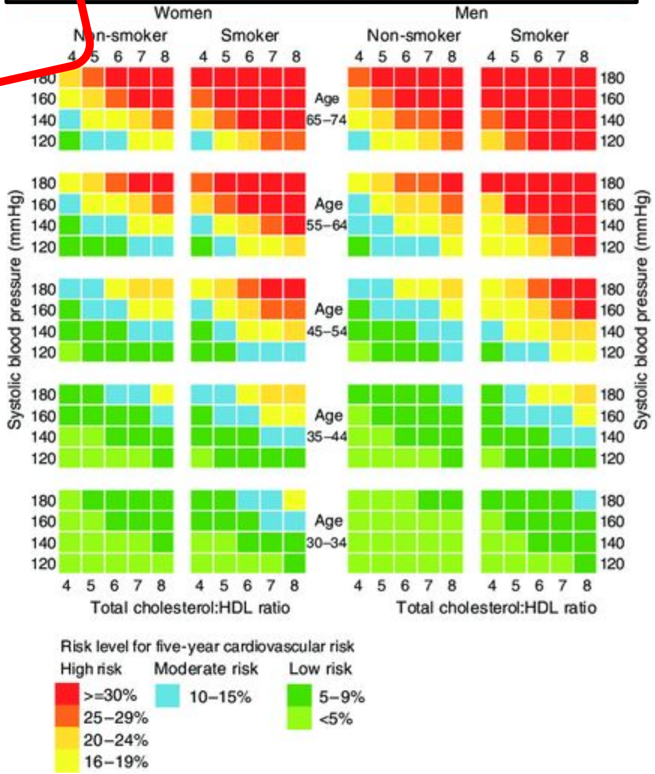


Examples Of Predictive Models In Clinical Use (1)

Roche

1998

Framingham Cardiovascular Risk Score



2004

oncotype DX®

Breast Recurrence Score

PROLIFERATION	INVASION	HER2
Ki-67 STK15 Survivin Cyclin B1 MYBL2	Stromelysin 3 Cathepsin L2	GRB7 HER2

ESTROGEN	REFERENCE	OTHER
ER PR Bcl2 SCUBE2	Beta-actin GAPDH RPLPO GUS TFRC	GSTM1 CD68 BAG1

Recurrence Score= +0.47 x HER2 Group Score
-0.34 x ER Group Score
+1.04 x Proliferation group Score
+0.10 x Invasion Group Score
+0.05 x CD68
-0.08 x GSTM1
-0.07 x BAG1

Examples Of Predictive Models In Clinical Use (2)

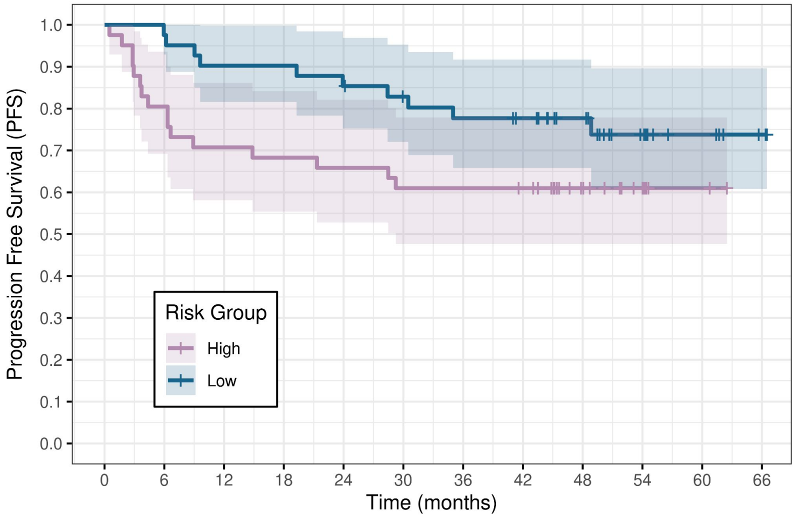
Roche

HemeProScore

Personalized Healthcare

2024

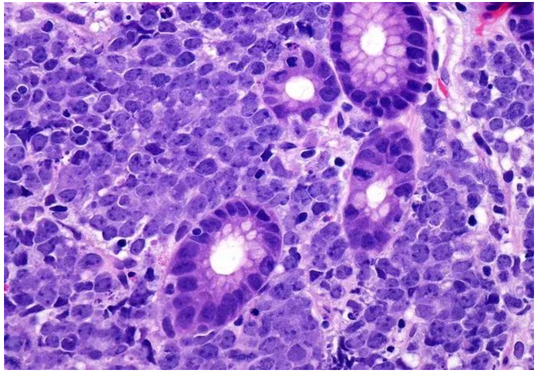
PFS in GOYA Test (F1-score)



Number at risk

Risk Group	0	6	12	18	24	30	36	42	48	54	60	66
High	41	33	29	28	27	25	25	24	14	8	2	0
Low	41	40	37	37	35	32	30	28	22	12	7	3

Time (months)



Personal Uncertainty Quantification

nature medicine

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature medicine › comment › article

Comment | Published: 11 October 2023

Clinical AI tools must convey predictive uncertainty for each individual patient

Christopher R. S. Banerji, Tapabrata Chakraborti, Chris Harbron & Ben D. MacArthur

nature machine intelligence

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature machine intelligence › perspectives › article

Perspective | Published: 23 April 2025

Personalized uncertainty quantification in artificial intelligence

Tapabrata Chakraborti, Christopher R. S. Banerji, Ariane Marandon, Vicky Hellon, Robin Mitra, Brieuc Lehmann, Leandra Bräuninger, Sarah McGough, Cagatay Turkay, Alejandro F. Frangi, Ginestra Bianconi, Weizi Li, Owen Rackham, Deepak Parashar, Chris Harbron & Ben MacArthur

Clinician collects clinical history and examines patient → Clinician inputs data into AI model, which predicts most likely diagnosis → Conformal prediction compares the distribution of AI-suggested diagnoses for the patient with other AI diagnoses for patients with similar demographics, and presents all diagnoses above a specified level of certainty → Clinician treats most likely condition or tests for further conditions

Migraine
Most likely diagnosis

→

Migraine	Ischemic stroke	Hemorrhagic stroke	Cluster headache
98%	<1%	<1%	<1%

→ Treat migraine

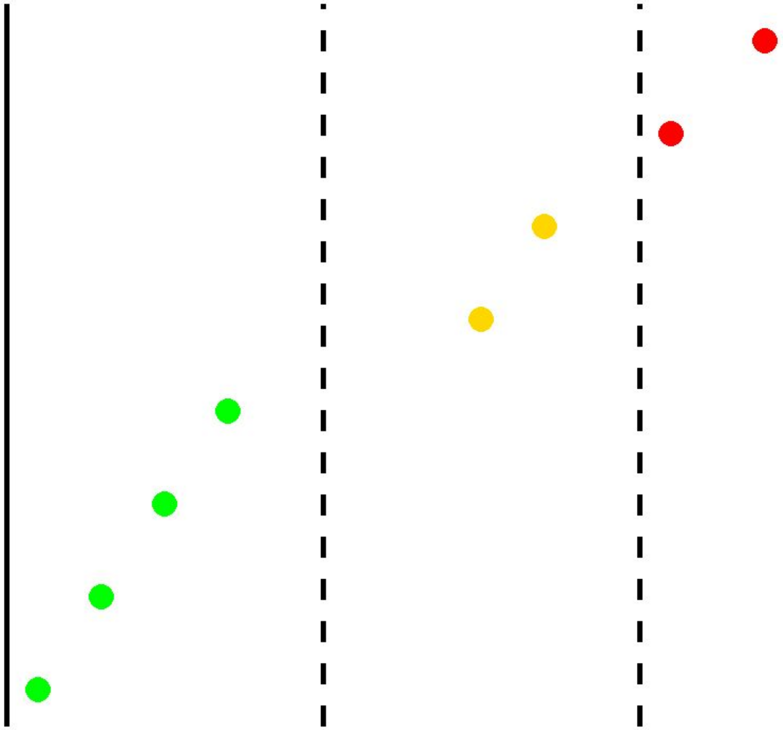
Migraine
Most likely diagnosis

→

Migraine	Ischemic stroke	Hemorrhagic stroke	Cluster headache
50%	20%	5%	<1%

→ Investigate cerebral vascular pathology

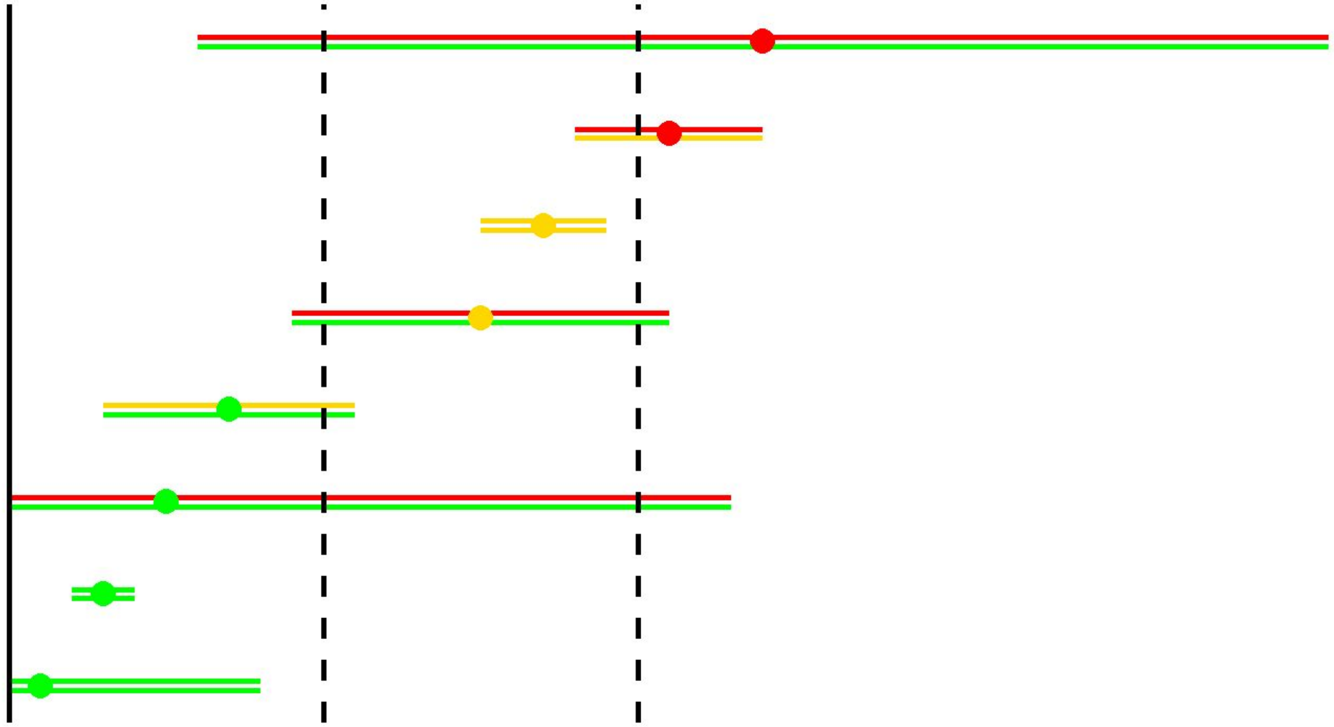
Personal Uncertainty Quantification With Risk Scores



Low Risk | Medium Risk | High Risk

Leading To Different Treatment Regimens

Personal Uncertainty Quantification With Risk Scores



Low Risk | Medium Risk | High Risk

Leading To Different Treatment Regimens

Uncertainty Quantification : An Expectation From Regulators

Roche

Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products

Guidance for Industry and Other Interested Parties

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Tala Fakhouri, 301-837-7407; (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010; or (CDRH) Digital Health Center of Excellence, digitalhealth@fda.hhs.gov.

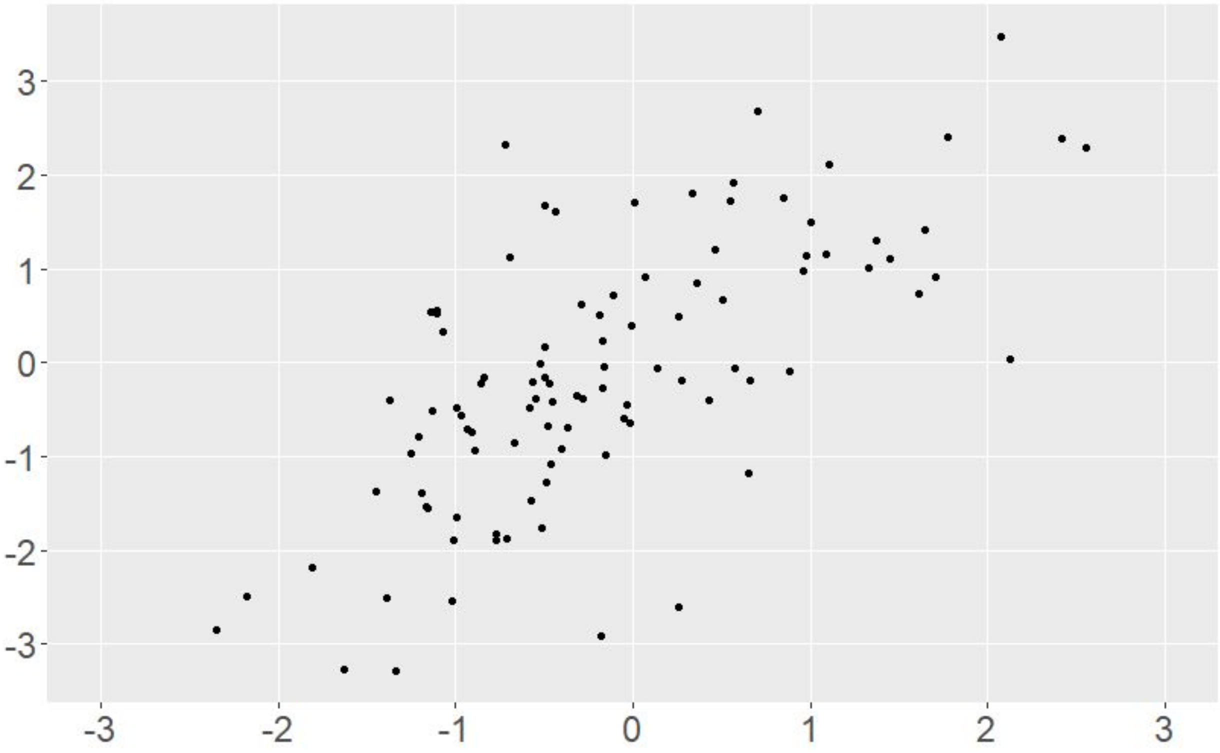
U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Center for Veterinary Medicine (CVM)
Oncology Center of Excellence (OCE)
Office of Combination Products (OCP)
Office of Inspections and Investigations (OII)

January 2025
Artificial Intelligence

59502407dft.docx

Specify the process by which the uncertainty and confidence level of model predictions were estimated. If relevant, include any other descriptions or metrics that quantify confidence or uncertainty. Information regarding the uncertainty of model output is important because it helps interpret model outputs. Repeatability and/or

Some Simple Simulated Data



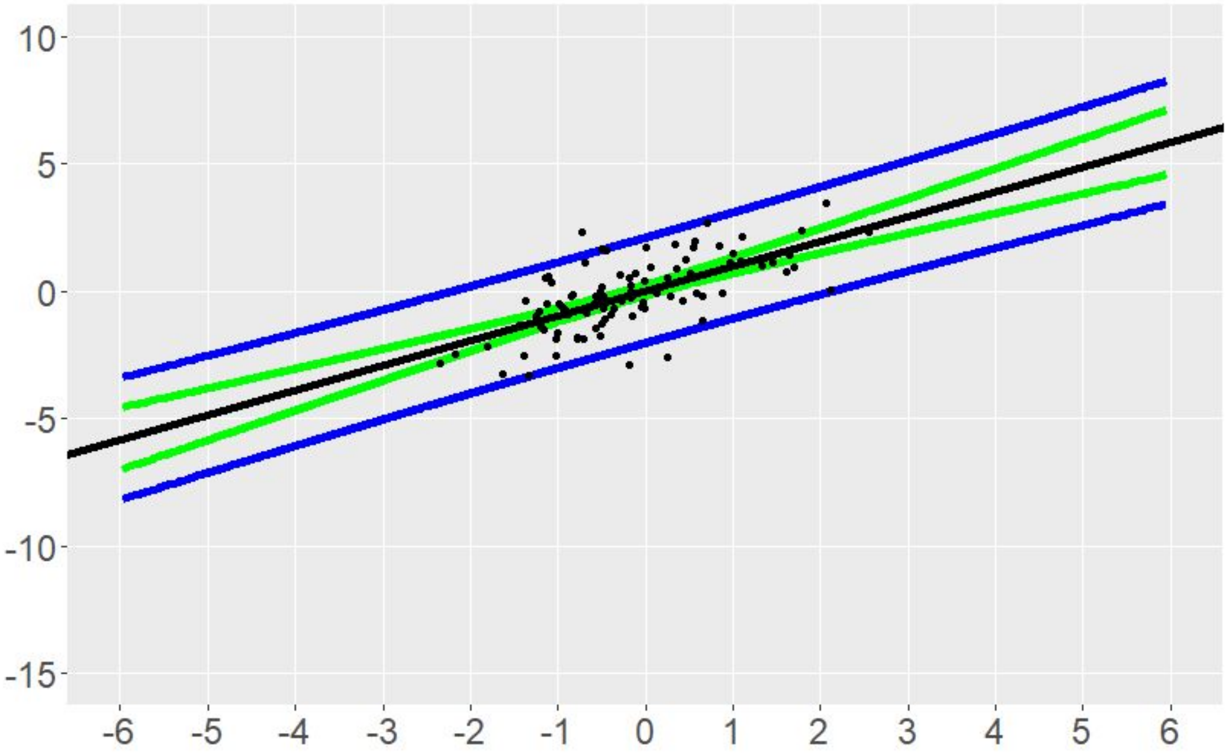
- n=100

- X ~ N(0,1)
- Y = X + N(0,1)

-> Correlation of 0.7

Starting With A Well Understood Model

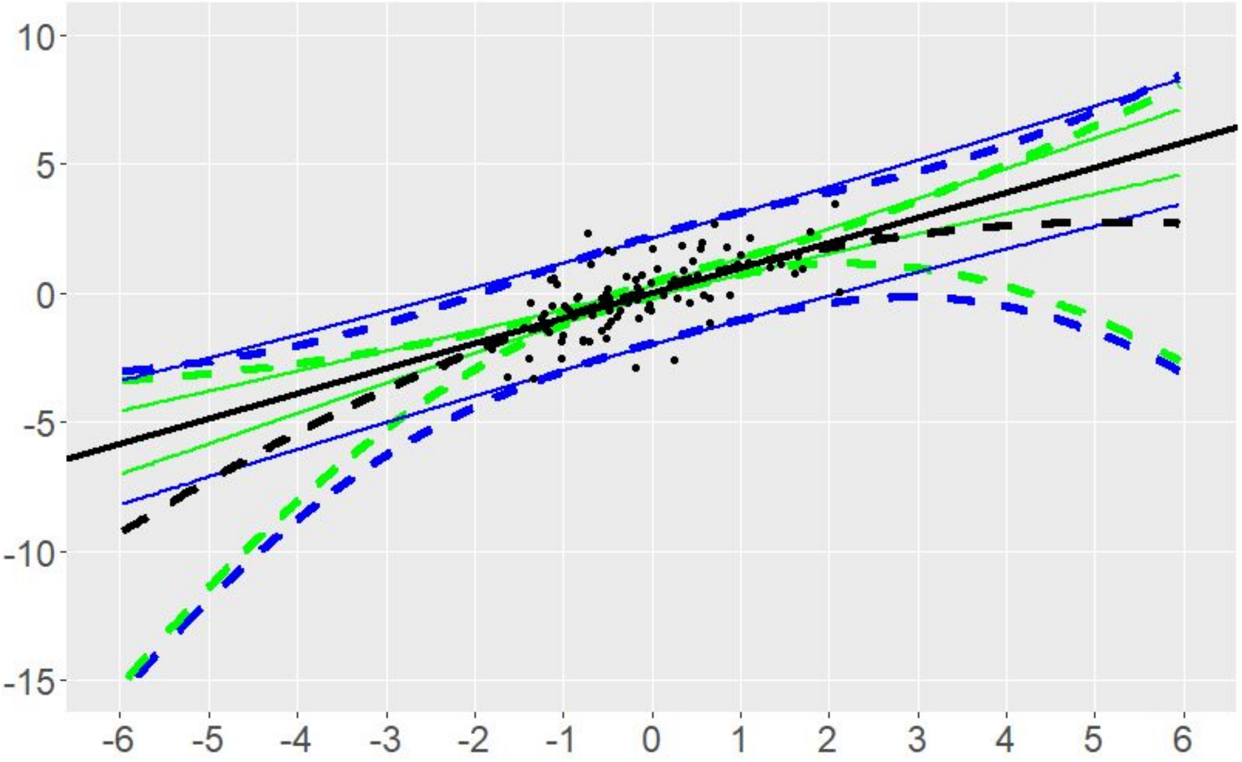
Linear Regression



- Two types of interval
- **Confidence interval** width varies
- **Prediction Interval** approximately constant width
- Prediction error dominates parameter uncertainty

Starting With A Well Understood Model

Quadratic Regression With Same Data (Simulated data is linear)



- Similar intervals in centre of data
- Much wider intervals when extrapolating
- When extrapolating, model uncertainty dominates

Starting With A Well Understood Model

Some Maths

$$\sigma_{Pred} = \sqrt{\sigma^2_{Conf} + \sigma^2_{Error}}$$

- Prediction error is a combination of two sources
- If one is much larger, it will dominate

$$\sigma^2_{Conf} = x_{new}(X^T X)^{-1} x^T_{new} \times \sigma^2_{Error}$$

The modelling error is a multiple of the observation error, depending upon the design, sample size and new data point

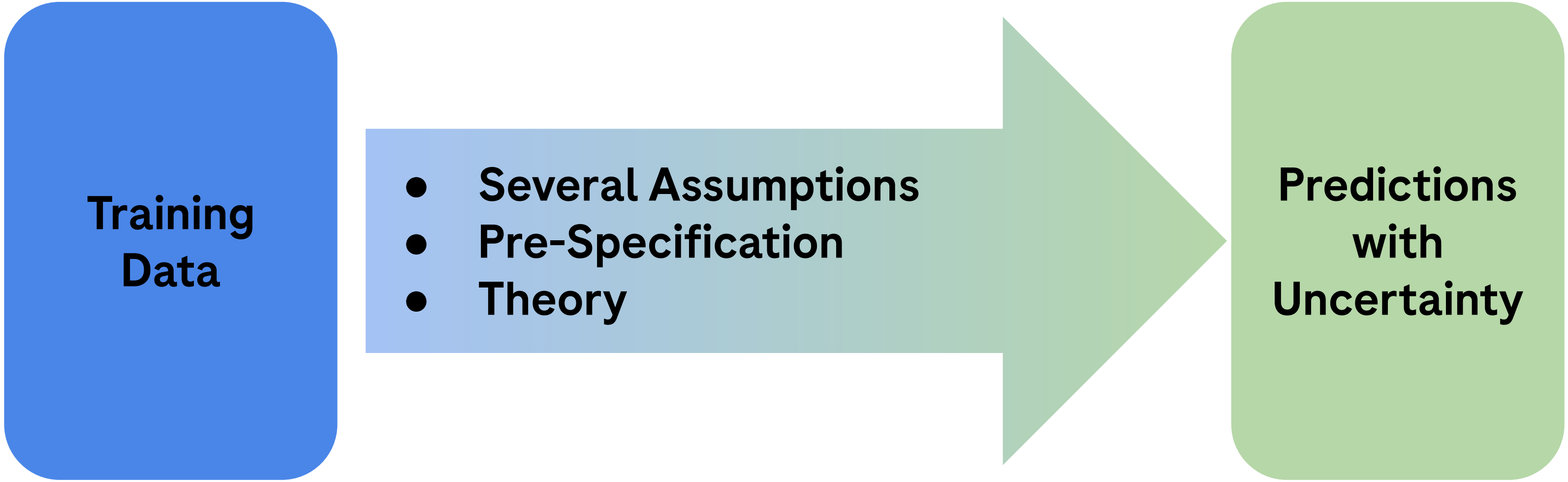
$$\sigma^2_{Conf} \approx \left[\frac{1}{n} + \frac{x^2_{new}}{\sum x^2_i} + \underbrace{\frac{x^4_{new}}{\sum x^4_i}}_{\text{Only for Quadratic}} \right] \times \sigma^2_{Error}$$

In our situation this only becomes important when

$$|x_{new}| >>> max(|x_i|)$$

i.e. extrapolation

Linear Regression



With Many Modern ML/AI Methods, No Such Theory Exists

- Random Forest
- XGBoost
- Lasso
- Elastic Net
- kNN
- Neural Net
- Deep Learner
- GBM
- SVM

Full Conformal Prediction

1998

Learning by Transduction

A. Gammerman, V. Vovk, V. Vapnik
Department of Computer Science
Royal Holloway, University of London
Egham, Surrey TW20 0EX, UK
{alex,vovk,vladimir}@dcs.rhbnc.ac.uk

2012

v2 [cs.LG] 24 Sep 2012

Conditional validity of inductive conformal predictors

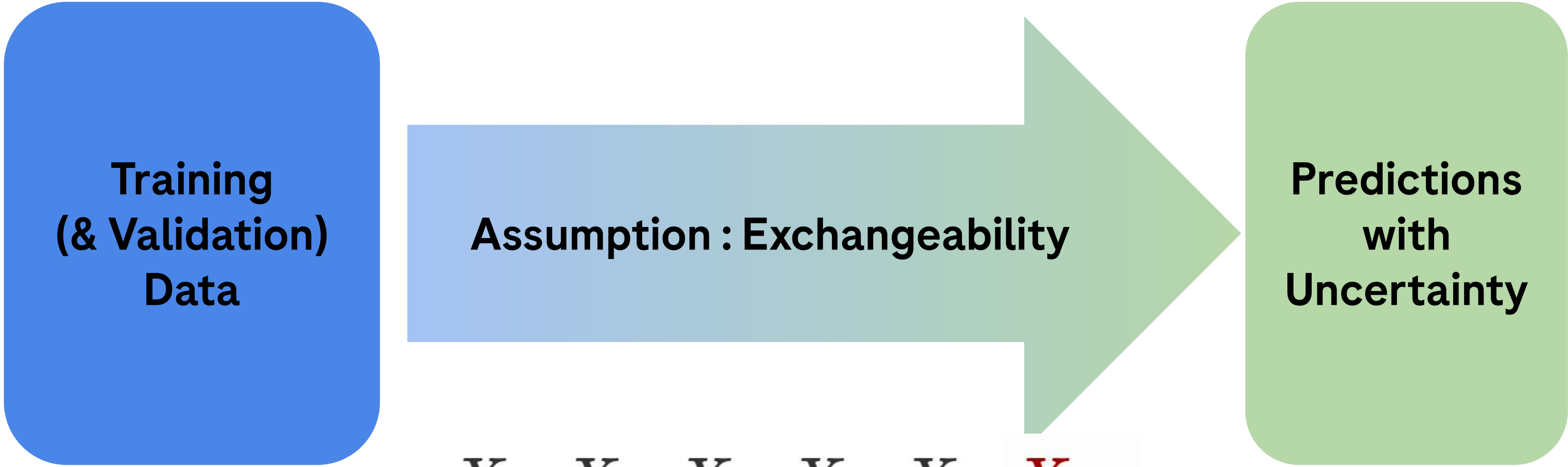
Vladimir Vovk
v.vovk@rhul.ac.uk
http://vovk.net

August 10, 2018

Abstract

Conformal predictors are set predictors that are automatically valid in the sense of having coverage probability equal to or exceeding a given confidence level. Inductive conformal predictors are a computationally efficient version of conformal predictors satisfying the same property of validity. However, inductive conformal predictors have been only known to control unconditional coverage probability. This paper explores various versions of conditional validity and various ways to achieve them using inductive conformal predictors and their modifications.

Full Conformal Prediction



$$X_1, X_2, X_3, X_4, X_5, X_6$$

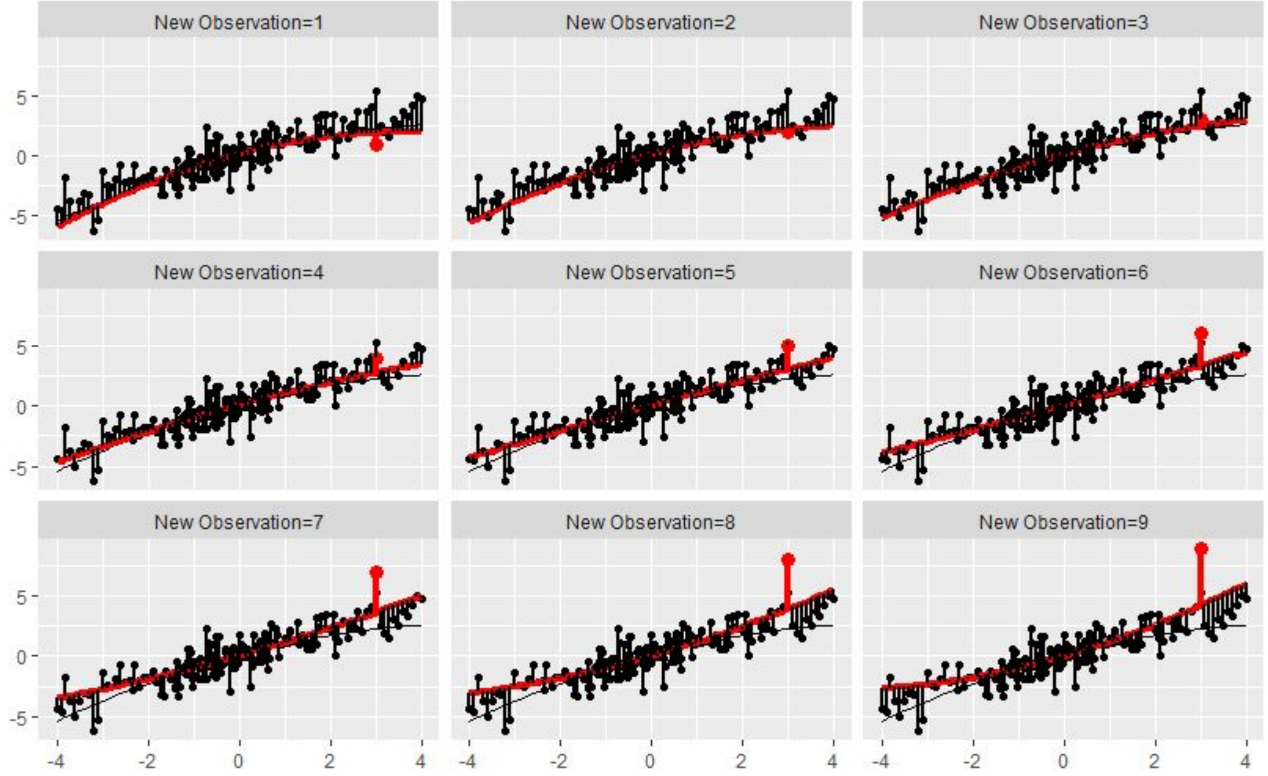
and

$X_4, X_3, X_6\ X_1, X_2, X_5$ **or any other permutation**

have the same joint distribution

Full Conformal Prediction

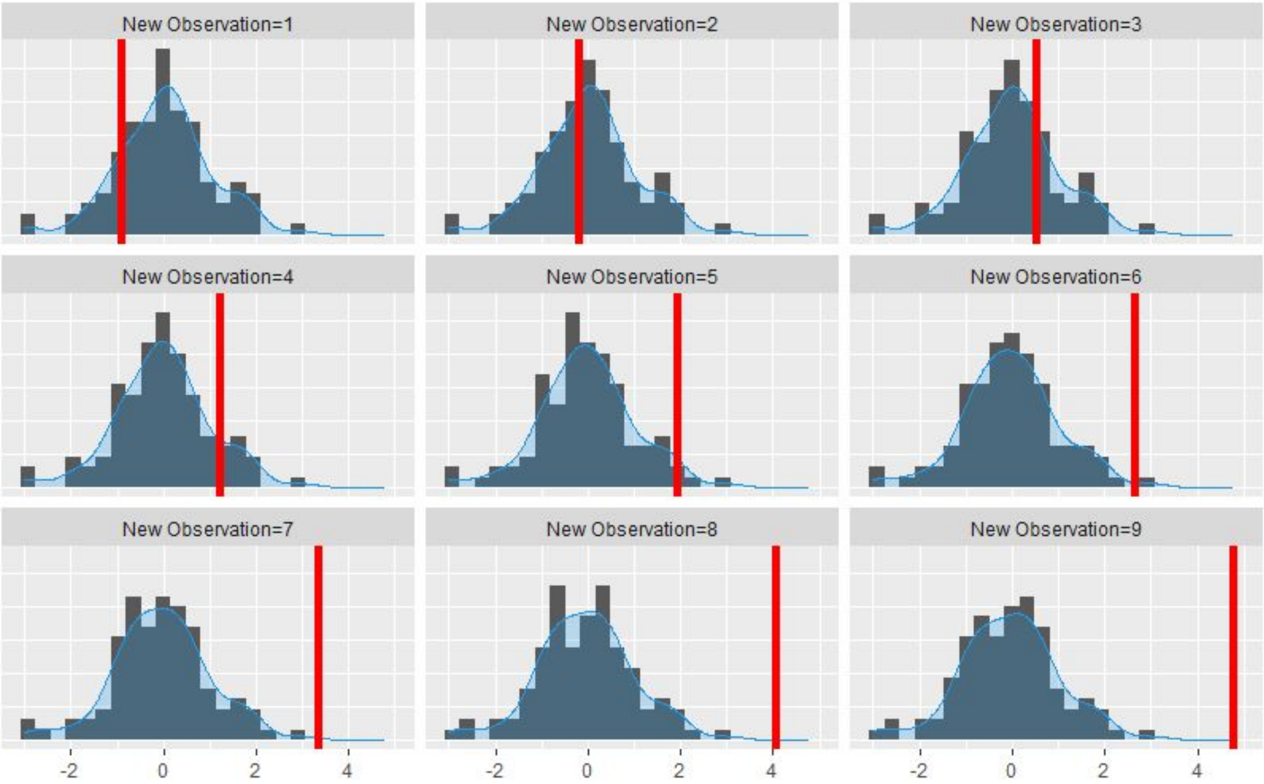
Quadratic Regression



- For a new data point to be predicted (here at x=3)
- Consider a range of potential values that could be observed
- For each potential value, assume it was observed and refit the model
- Collect the residuals

Full Conformal Prediction

Quadratic Regression

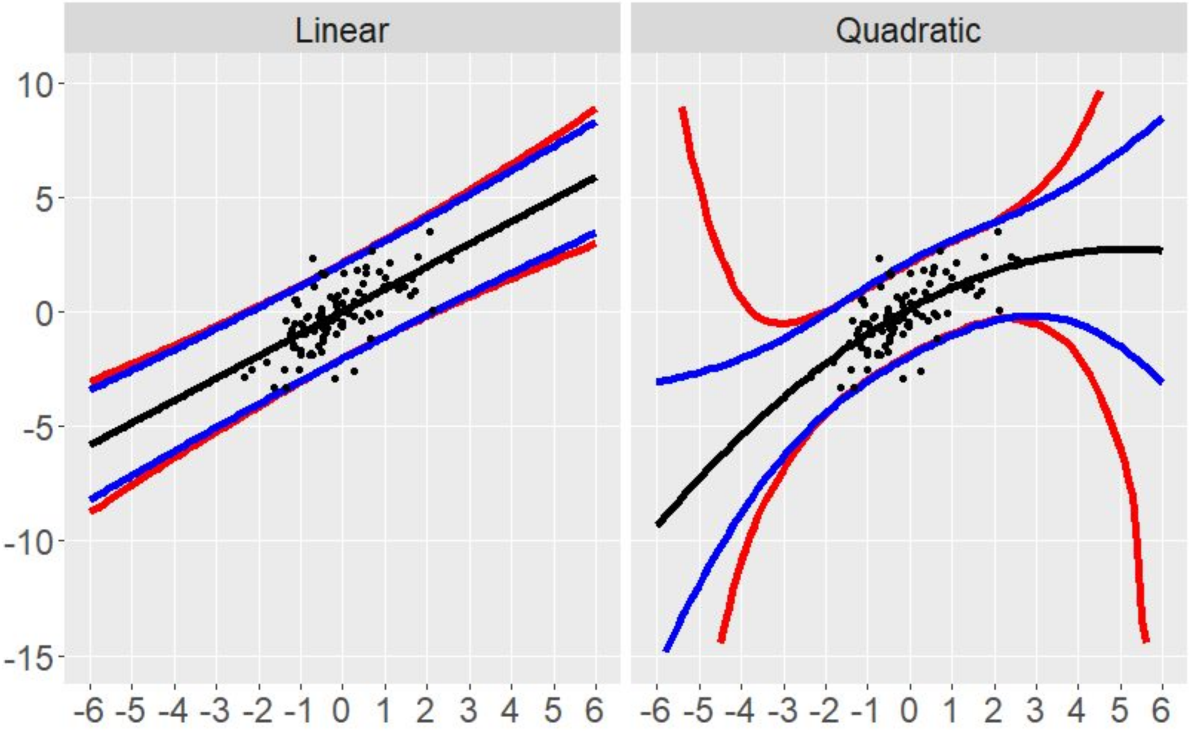


Compare the residual at the new point with the distribution of residuals

Prediction Interval = Those assumed observations where the residual lies within the distribution

Comparing Statistical And Conformal Intervals

Statistical , Full Conformal,



For linear regression Conformal Intervals practically identical to statistical intervals

For quadratic regression agreement within region of data

Conformal gives larger intervals when extrapolating - where parameter uncertainty starts to dominate

Full Conformal Prediction

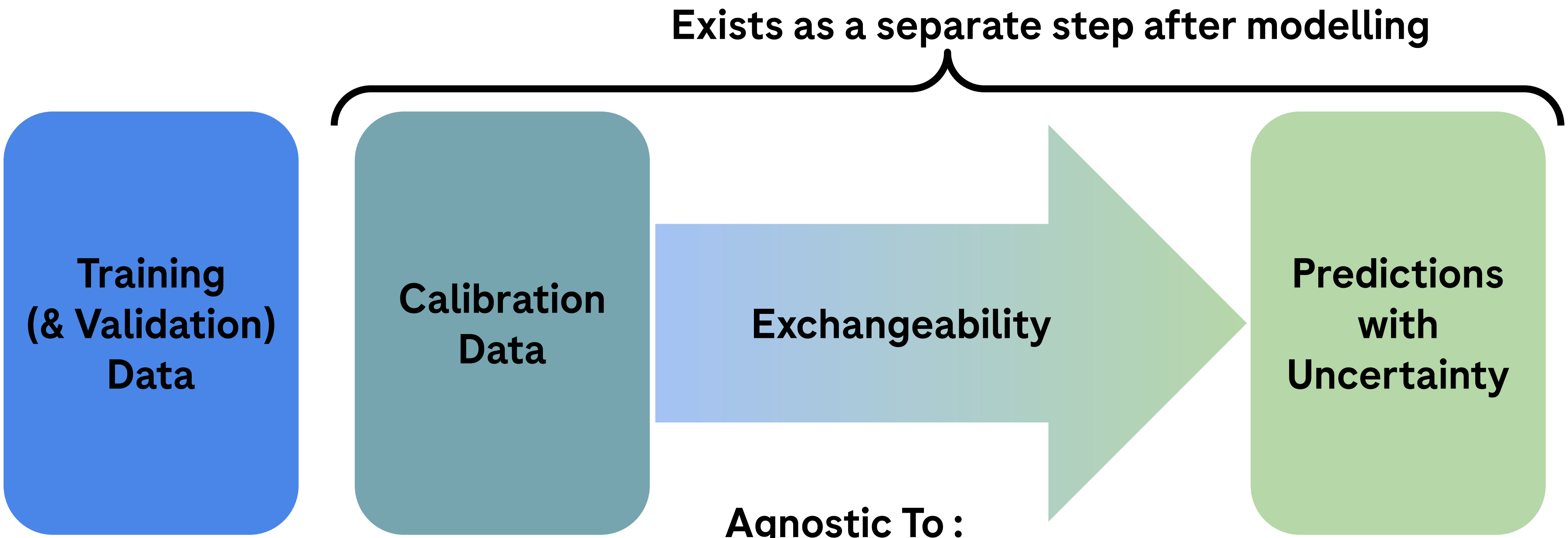
Some Comments

- Agnostic To :
  - Model
  - Distributions
  - Heterogeneity

- Generalisable - here using residuals, but can use any similarity score

- Computationally heavy, so reliant on either :
  - The model being computationally light to repeatedly fit
  - Some clever maths being available as a short-cut
  - Being able to make some useful approximations

Split Conformal Prediction

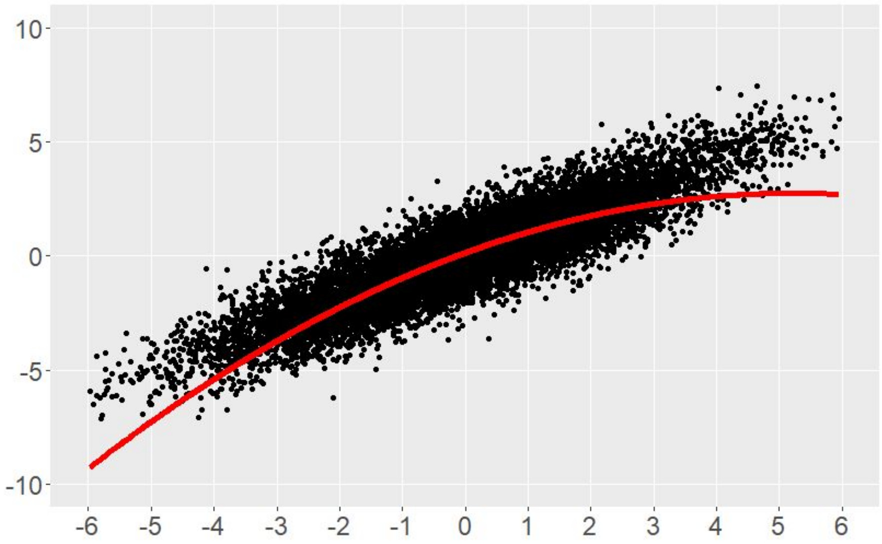


Agnostic To :

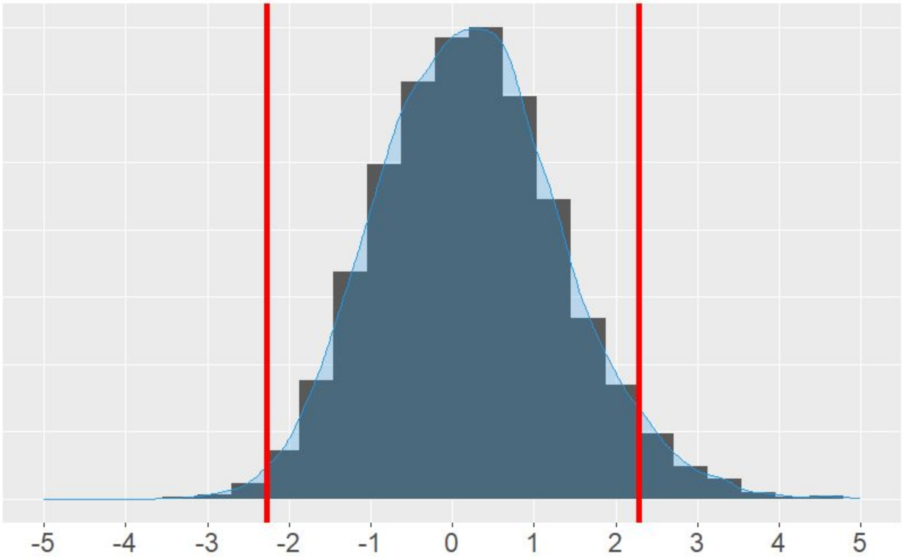
- Model
- Distributions
- Heterogeneity

Generalisable

Split Conformal Prediction
Quadratic Regression



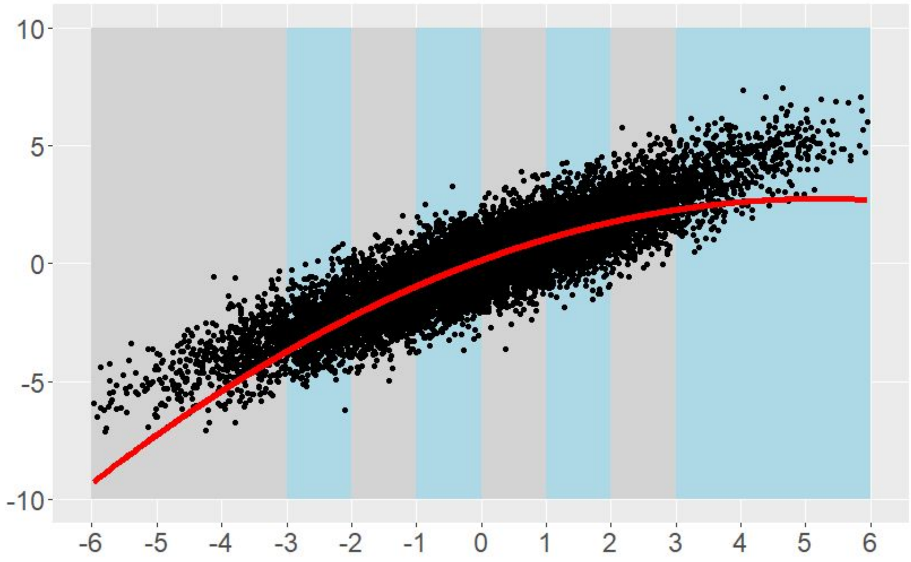
Prediction from the smaller training set, shown against a much larger calibration set

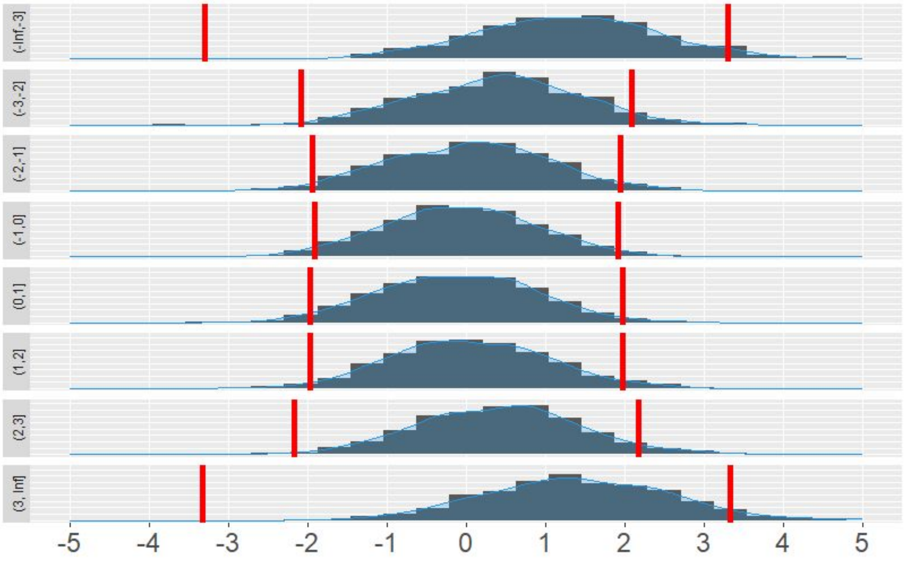


Take appropriate quantiles of the residuals in the calibration set

Split Conformal Prediction

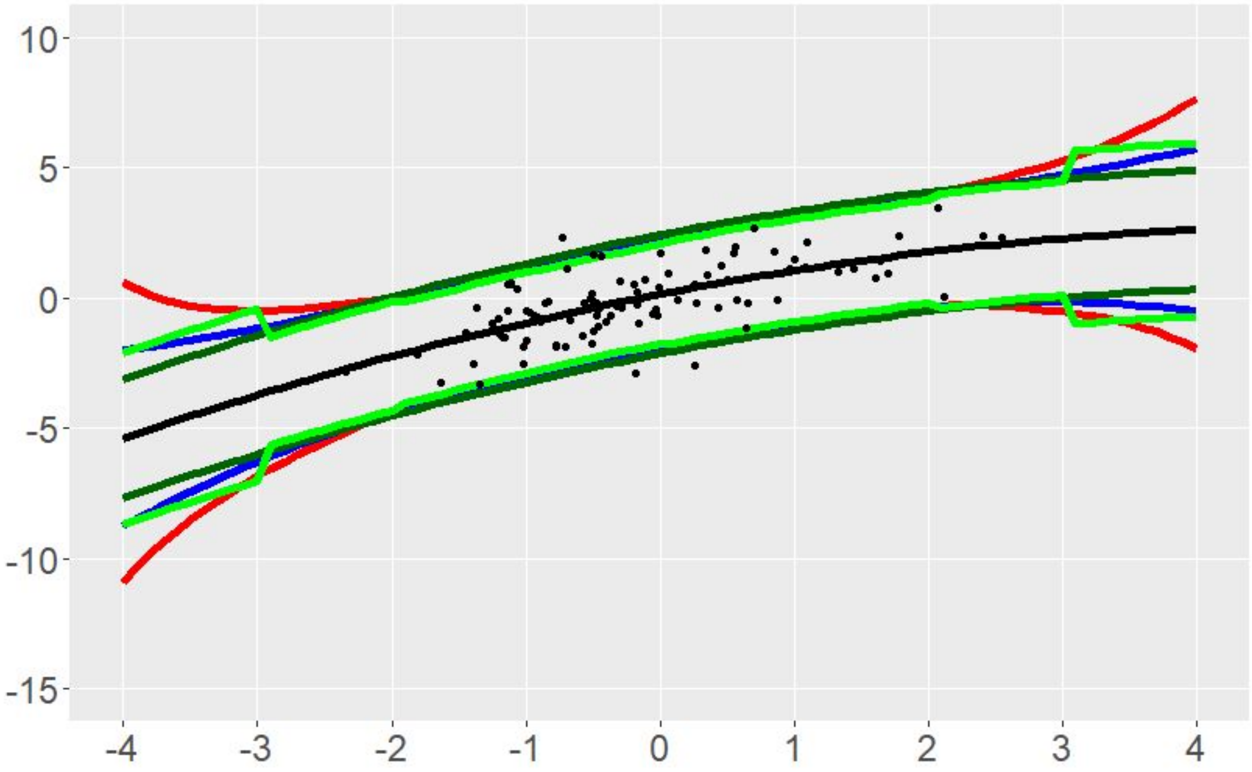
Quadratic Regression





- In this scenario we stratified by x value
- Could stratify by region, gender, race, missingness patterns etc.
- Could stratify either by expected heterogenity and/or expected variation in model fit

Split Conformal Prediction
Quadratic Regression



Statistical, **Full Conformal**, **Split Conformal** **Stratified Split Conformal**

More sophisticated applications available to allow for heterogeneity in data

Modelling Sources Of Error

Statistical	Model two sources of error independently and combine
Full Conformal	Two sources of error contribute to the prediction uncertainty - typical residual size and how ‘flexible’ the model is at that point
Split Conformal	Assumes an “average” error combining both sources and uses stratification to allow for differences -> marginal coverage Fine when residual error dominates, less so when the model itself is uncertain

Concluding Comments

- Whilst forming the basis for statistics, uncertainty in general is a new concept for much of the Computer Science ML/AI World

- Here was a simple example to make a link to statistical modelling.
  In most high dimensional ML/AI methodologies, no rigorous uncertainty methodology exists

- Split conformal intervals provide a marginal coverage
    - Would ideally like the coverage to be as close to conditional as possible
        - Steps such as stratifying take us in that direction

- Uncertainty with Generative AI/Large Language Models much harder to define problem - but work starting as way of addressing hallucinations

Doing now what patients need next