Isotonic Distributional Regression (IDR): A powerful nonparametric calibration technique

Johanna F. Ziegel

University of Bern

Workshop: Predictability, dynamics and applications research using the TIGGE and S2S ensembles

2 – 5 April 2019

ECMWF, Reading

Joint work with Alexander Henzi and Tilmann Gneiting

Introduction

Goal:
Provide calibrated probabilistic predictions for a real-valued quantity Y (e.g. cumulated precipitation amount) based on an ensemble of predictions $X = (X^{(1)}, \ldots, X^{(d)})$.

Requirement:
Sufficient training data available: $(X_1, Y_1), \ldots, (X_n, Y_n)$

Characteristics of IDR:

- Generic (non-parametric) method providing a competitive benchmark for prediction (with respect to CRPS)
- Leads to calibrated probabilistic predictions (flat PIT histogram)
- (Almost) No tuning parameters
- May be outperformed by carefully tuned parametric postprocessing methods

Introduction

Goal:
Provide calibrated probabilistic predictions for a real-valued quantity Y (e.g. cumulated precipitation amount) based on an ensemble of predictions $X = (X^{(1)}, \dots, X^{(d)})$.

Requirement:
Sufficient training data available: $(X_1, Y_1), \dots, (X_n, Y_n)$

Characteristics of IDR:

- Generic (non-parametric) method providing a competitive benchmark for prediction (with respect to CRPS)
- Leads to calibrated probabilistic predictions (flat PIT histogram)
- (Almost) No tuning parameters
- May be outperformed by carefully tuned parametric postprocessing methods

Introduction

Goal:
Provide calibrated probabilistic predictions for a real-valued quantity Y (e.g. cumulated precipitation amount) based on an ensemble of predictions $X = (X^{(1)}, \ldots, X^{(d)})$.

Requirement:
Sufficient training data available: $(X_1, Y_1), \ldots, (X_n, Y_n)$

Characteristics of IDR:

- Generic (non-parametric) method providing a competitive benchmark for prediction (with respect to CRPS)
- Leads to calibrated probabilistic predictions (flat PIT histogram)
- (Almost) No tuning parameters
- May be outperformed by carefully tuned parametric postprocessing methods

Fundamental assumption of IDR

"If the predictions increase we expect an increase of the outcomes."

Making this intuition precise

"If the predictions increase. . . "

Partial order on the covariates:
$x = (x_1, \dots, x_d), x' = (x'_1, \dots, x'_d) \in \mathbb{R}^d$

$$x \leq_p x' \quad \text{if} \quad x_1 \leq x'_1,\ \dots,\ x_d \leq x'_d.$$

". . . we expect an increase of the outcomes."

Stochastic order on predictive distributions: F, G cdfs

$$F \preceq G \quad \text{if} \quad F(z) \geq G(z) \text{ for all } z \in \mathbb{R}.$$

Equivalent:

$$F \preceq G \quad \text{if} \quad F^{-1}(\alpha) \leq G^{-1}(\alpha) \text{ for all } \alpha \in (0,1).$$

Making this intuition precise

"If the predictions increase..."

Partial order on the covariates:
$x = (x_1, \ldots, x_d), x' = (x'_1, \ldots, x'_d) \in \mathbb{R}^d$

$$x \leq_p x' \quad \text{if} \quad x_1 \leq x'_1, \ \ldots, \ x_d \leq x'_d.$$

"...we expect an increase of the outcomes."

Stochastic order on predictive distributions: F, G cdfs

$$F \preceq G \quad \text{if} \quad F(z) \geq G(z) \text{ for all } z \in \mathbb{R}.$$

Equivalent:

$$F \preceq G \quad \text{if} \quad F^{-1}(\alpha) \leq G^{-1}(\alpha) \text{ for all } \alpha \in (0,1).$$

Isotonic distributional regression (IDR)

Estimate the cdf-valued function $X \mapsto F_X$ with

$$F_X = \mathcal{L}(Y|X)$$

under the assumption that F_X is isotone, that is,

$$X \leq_p X' \implies F_X \preceq F_{X'}.$$

Minimization problem: Define $\hat{F}_X$ to be the isotone cdf-valued G_X minimizing

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{CRPS}(G_{X_i}, Y_i).$$

Result: There exists a unique minimizer $\hat{F}_X$ which we call the IDR.

Isotonic distributional regression (IDR)

Estimate the cdf-valued function $X \mapsto F_X$ with

$$F_X = \mathcal{L}(Y|X)$$

under the assumption that F_X is isotone, that is,

$$X \leq_p X' \quad \Longrightarrow \quad F_X \preceq F_{X'}.$$

Minimization problem: Define $\hat{F}_X$ to be the isotone cdf-valued G_X minimizing

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{CRPS}(G_{X_i}, Y_i).$$

Result: There exists a unique minimizer $\hat{F}_X$ which we call the IDR.

Isotonic distributional regression (IDR)

Estimate the cdf-valued function $X \mapsto F_X$ with

$$F_X = \mathcal{L}(Y|X)$$

under the assumption that F_X is isotone, that is,

$$X \leq_p X' \quad \Longrightarrow \quad F_X \preceq F_{X'}.$$

Minimization problem: Define $\hat{F}_X$ to be the isotone cdf-valued G_X minimizing

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{CRPS}(G_{X_i}, Y_i).$$

Result: There exists a unique minimizer $\hat{F}_X$ which we call the IDR.

Constructing the IDR

Let $z \in \mathbb{R}$. Minimizing

$$\sum_{\ell=1}^{n} (g_z(X_\ell) - \mathbb{1}\{Y_\ell > z\})^2$$

over all increasing functions $g_z : \mathbb{R}^d \to \mathbb{R}$ has a unique optimal solution that can be computed by solving a quadratic programming problem.

- $\hat{F}_X : z \mapsto 1 - \hat{g}_z(X)$ is a valid cdf
- $X \mapsto \hat{F}_X$ is the IDR

Sidenote:

Closed form of the optimal solution for a total order ($d = 1$)

$$\hat{g}_z(X_\ell) = \min_{j \geq \ell} \max_{i \leq j} \frac{1}{(j - i + 1)} \sum_{t=i}^{j} \mathbb{1}\{Y_t > z\}.$$

Constructing the IDR

Let $z \in \mathbb{R}$. Minimizing

$$\sum_{\ell=1}^{n} (g_z(X_\ell) - \mathbb{1}\{Y_\ell > z\})^2$$

over all increasing functions $g_z : \mathbb{R}^d \to \mathbb{R}$ has a unique optimal solution that can be computed by solving a quadratic programming problem.

- $\hat{F}_X : z \mapsto 1 - \hat{g}_z(X)$ is a valid cdf
- $X \mapsto \hat{F}_X$ is the IDR

Sidenote:

Closed form of the optimal solution for a total order ($d = 1$)

$$\hat{g}_z(X_\ell) = \min_{j \geq \ell} \max_{i \leq j} \frac{1}{(j - i + 1)} \sum_{t=i}^{j} \mathbb{1}\{Y_t > z\}.$$

Constructing the IDR

Let $z \in \mathbb{R}$. Minimizing

$$\sum_{\ell=1}^{n} (g_z(X_\ell) - \mathbb{1}\{Y_\ell > z\})^2$$

over all increasing functions $g_z : \mathbb{R}^d \to \mathbb{R}$ has a unique optimal solution that can be computed by solving a quadratic programming problem.

- $\hat{F}_X : z \mapsto 1 - \hat{g}_z(X)$ is a valid cdf
- $X \mapsto \hat{F}_X$ is the IDR

Sidenote:
Closed form of the optimal solution for a total order ($d = 1$)

$$\hat{g}_z(X_\ell) = \min_{j \geq \ell} \max_{i \leq j} \frac{1}{(j - i + 1)} \sum_{t=i}^{j} \mathbb{1}\{Y_t > z\}.$$

Optimality properties of the IDR

- Let W-CRPS be a quantile- or threshold-weighted CRPS. The IDR $\hat{F}_X$ satisfies

$$\frac{1}{n}\sum_{\ell=1}^{n} \text{W-CRPS}(\hat{F}_{X_\ell}, Y_\ell) = \min_{G_X} \frac{1}{n}\sum_{\ell=1}^{n} \text{W-CRPS}(G_{X_\ell}, Y_\ell)$$

  where G_X runs over all isotone cdf-valued functions.
- The IDR is calibrated "if the partial order is strong enough/the training sample is large enough".

Using IDR for prediction

- Compute IDR for training dataset.
- For a new covariate value X, find nearest neighbors, choose suitable ones.
- Interpolate solution amongst nearest neighbors.

Application: Precipitation forecasts

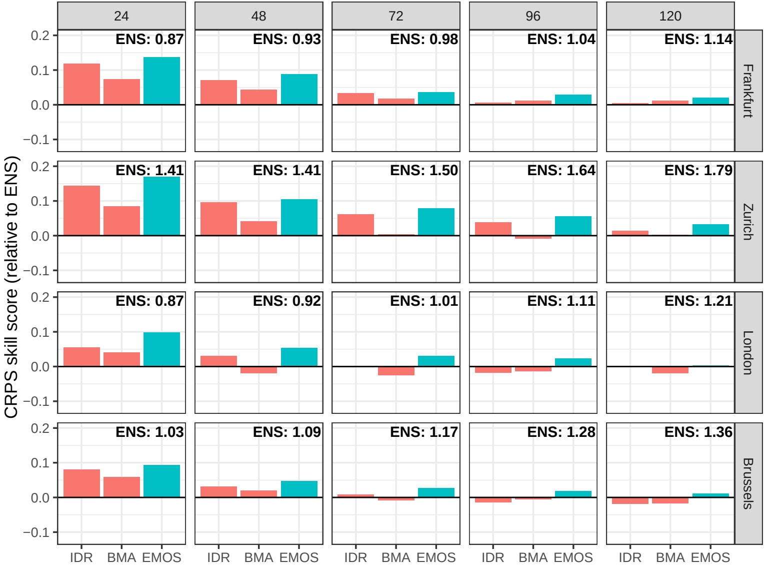
Dataset

- Precipitation forecasts and observations from 2007 to 2017

Airport	Available days (years)
London Heathrow	2256 (6.2)
Brussels	3406 (9.4)
Zurich Kloten	3241 (8.9)
Frankfurt	3617 (9.9)

- Observations: 24-hour accumulated precipitation amounts
- Forecasts: ECMWF ensemble
  52 members: high-resolution forecast (HRES), control forecast (CTRL), 50 perturbed members (PM)
- IDR using (HRES, CTRL, mean of PM)

Results: CRPSS



Discussion and outlook

- IDR is a new generic technique to generate calibrated probabilistic predictions.
- IDR can accomodate predictions from multiple models.
- IDR is in-sample optimal with respect to all weighted CRPS.
- IDR provides guarantees for calibration in-sample.
- IDR yields competitive predictions for precipitation using less information.
- R Package for IDR in preparation
- Paper in preparation, available upon request: Master Thesis of A. Henzi (2018).

Extensions/related methods:

- Semi-parametric IDR for outcomes with heavy tails.
- Isotonic regression for point predictions/specific parameters of the predictive distribution.
- Work in progress: Variable selection method for partial orders.