Isotonic Distributional Regression

Johanna Ziegel

joint work with
Sebastian Arnold, Tilmann Gneiting, Alexander Henzi,
Gian-Reto Kleger, and Eva-Maria Walz

University of Bern

RiskDay, ETH Zurich

15 September, 2023

Motivation

Situation

Model provides prediction $x \in \mathbb{R}$ for outcome $Y \in \mathbb{R}$

- Physics-based numerical (weather prediction) model
- Statistical or ML model for mean of Y given covariates
- Expert opinion

"If the prediction increases we expect an increase of the outcome."

Goal

Quantify uncertainty by predictive distribution for Y

Provide probabilistic prediction for Y

Approach

Use distributional regression model for Y given x
that honors the isotonicity assumption:

Isotonic distributional regression (IDR)

Motivation

Situation

Model provides prediction $x \in \mathbb{R}$ for outcome $Y \in \mathbb{R}$

- Physics-based numerical (weather prediction) model
- Statistical or ML model for mean of Y given covariates
- Expert opinion

"If the prediction increases we expect an increase of the outcome."

Goal

Quantify uncertainty by predictive distribution for Y

Provide probabilistic prediction for Y

Approach

Use distributional regression model for Y given x

that honors the isotonicity assumption:

Isotonic distributional regression (IDR)

Isotonic distributional regression (IDR)

Goal'
Flexible distributional regression model for $Y \in \mathbb{R}$ given $x \in \mathcal{X}$ under the assumption:

"If the covariate increases we expect an increase of the outcome."

- $(\mathcal{X}, \leq)$ is a partially ordered set.
- Usual stochastic order on conditional distributions of Y.

Mathematical setup

"If the covariate increases we expect an increase of the outcome."

$$x \leq x' \implies \mathcal{L}(Y \mid X = x) \preceq_{\mathrm{st}} \mathcal{L}(Y \mid X = x')$$

$$\iff F_{Y|X=x}(y) \geq F_{Y|X=x'}(y), \quad y \in \mathbb{R}$$
$$\iff q_\alpha(Y|X = x) \leq q_\alpha(Y|X = x'), \quad \alpha \in (0,1)$$

IDR estimator (for $x \in \mathbb{R}$): Data $(x_i, y_i)_{i=1}^n$, $x_1 \leq \cdots \leq x_n$
Define $\hat{\mathbf{F}} = (\hat{F}_i)_{i=1}^n = (\hat{F}_{Y|X=x_i})_{i=1}^n$ as

$$\hat{\mathbf{F}} = \underset{F_1 \preceq_{\mathrm{st}} \cdots \preceq_{\mathrm{st}} F_n}{\operatorname{argmin}} \sum_{\ell=1}^{n} \mathrm{CRPS}(F_\ell, y_\ell).$$

Continuous ranked probability score (CRPS)

$$\mathrm{CRPS}(F, Y) = \int_{\mathbb{R}} (F(z) - \mathbb{1}\{Y \leq z\})^2 \, \mathrm{d}z$$

Mathematical setup

"If the covariate increases we expect an increase of the outcome."

$$x \leq x' \implies \mathcal{L}(Y \mid X = x) \preceq_{\text{st}} \mathcal{L}(Y \mid X = x')$$

$$\iff F_{Y|X=x}(y) \geq F_{Y|X=x'}(y), \quad y \in \mathbb{R}$$

$$\iff q_\alpha(Y|X = x) \leq q_\alpha(Y|X = x'), \quad \alpha \in (0,1)$$

IDR estimator (for $x \in \mathbb{R}$): Data $(x_i, y_i)_{i=1}^n$, $x_1 \leq \cdots \leq x_n$

Define $\hat{\mathbf{F}} = (\hat{F}_i)_{i=1}^n = (\hat{F}_{Y|X=x_i})_{i=1}^n$ as

$$\hat{\mathbf{F}} = \underset{F_1 \preceq_{\text{st}} \cdots \preceq_{\text{st}} F_n}{\operatorname{argmin}} \sum_{\ell=1}^{n} \text{CRPS}(F_\ell, y_\ell).$$

Continuous ranked probability score (CRPS)

$$\text{CRPS}(F, Y) = \int_{\mathbb{R}} \left(F(z) - \mathbb{1}\{Y \leq z\}\right)^2 \, dz$$

Then

$$\hat{F}_i(y) = \max_{j=i,\ldots,n} \min_{k=1,\ldots,j} \frac{1}{j-k+1} \sum_{\ell=k}^{j} \mathbb{1}\{y_\ell \le y\}.$$

Theorem (Universality of IDR)

$\hat{\mathbf{F}}$ *minimizes all criteria of the form*

$$\ell_\lambda^c(\mathbf{F}) = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}} \left(F_i(y) - \mathbb{1}\{y_i \le y\}\right)^2 \, d\lambda(y),$$

$$\ell_\nu^q(\mathbf{F}) = \frac{1}{n} \sum_{i=1}^{n} \int_{(0,1)} \left(\mathbb{1}\{y_i \le F_i^{-1}(\alpha)\} - \alpha\right) \left(F_i^{-1}(\alpha) - y_i\right) d\nu(\alpha),$$

over all stochastically ordered tuples of CDFs $\mathbf{F} = (F_1, \ldots, F_n)$,
where λ and ν are Borel measures.

Barlow, Bartholomew, Bremner, and Brunk (1972); Henzi, Ziegel, and Gneiting (2021)

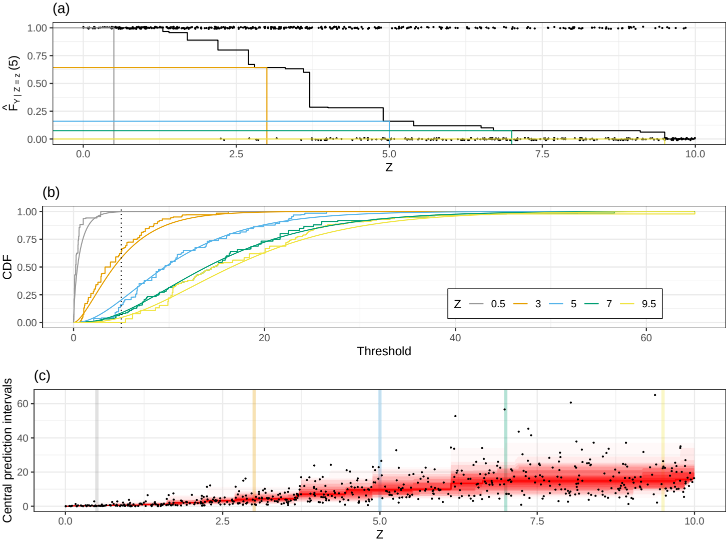


Illustration of IDR: $n = 600$ draws of $Z \sim \text{Unif}(0, 10)$ and $Y \sim \text{Gamma}(\text{sh} = \sqrt{Z}, \text{sc} = \min(\max(Z, 1), 6))$.

Computation

- Total order on the covariates: Pool adjacent violators (PAV) algorithm for each threshold $y \in \{y_1, \ldots, y_n\}$
- Partial order on the covariates: Quadratic programming problem for each $y \in \{y_1, \ldots, y_n\}$
- IDR can be combined with subsample aggregation: Computational gains and smoother estimated CDFs
- R package by Alexander Henzi available. Python implementation by Eva-Maria Walz.

Henzi, Mösching, and Dümbgen (2022)

Prediction

IDR solution $\hat{\mathbf{F}} = (\hat{F}_1, \ldots, \hat{F}_n)$ is defined at covariate values $x_1, \ldots, x_n$ only.
Procedure for prediction at new covariate value $x \notin \{x_1, \ldots, x_n\}$:

- Define predecessors and successors of x:

$$p(x) = \{i \mid x_i \preceq x_j \preceq x \implies x_j = x_i, j = 1, \ldots, n\}$$
$$s(x) = \{i \mid x \preceq x_j \preceq x_i \implies x_j = x_i, j = 1, \ldots, n\},$$

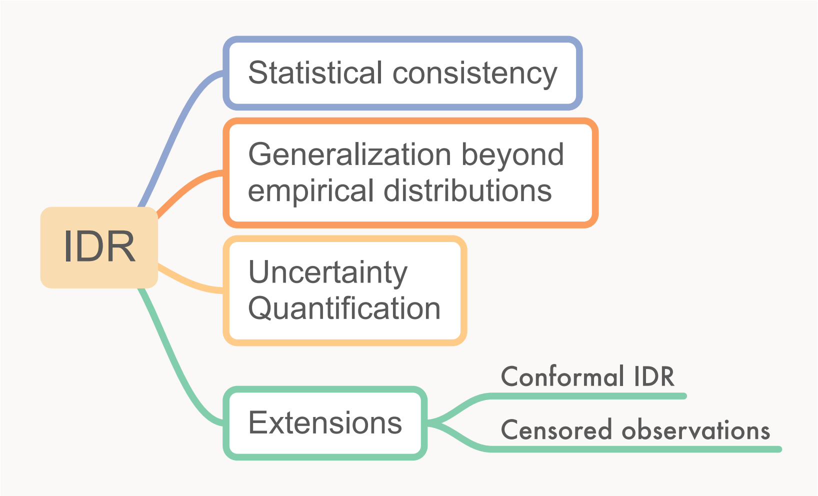
- Predictive CDF F that respects order constraints must satisfy

$$\max_{i \in s(x)} \hat{F}_i(y) \leq F(y) \leq \min_{j \in p(x)} \hat{F}_j(y), \quad y \in \mathbb{R}.$$

- Define

$$\hat{F}_x(y) = \frac{1}{2}\left(\max_{i \in s(x)} \hat{F}_i(y) + \min_{j \in p(x)} \hat{F}_j(y)\right).$$

What can we do with IDR?



Statistical consistency

Statistical consistency

Let $(X_{in}, Y_{in})_{in}$ be a triangular array of iid random variables, and suppose that the model assumption holds

$$x \leq x' \implies \mathcal{L}(Y \mid X = x) \preceq_{\text{st}} \mathcal{L}(Y \mid X = x').$$

Under reasonable assumptions, IDR is uniformly consistent

- for a categorical covariate
  El Barmi and Mukerjee (2005)
- for a real-valued covariate
  Mösching and Dümbgen (2020)
- for a vector-valued covariate
  Henzi, Ziegel, and Gneiting (2021)

What if the "covariate" is a prediction, that is, it depends on the data?

- Distributional (single) index models (DIM): Consistency still holds.
  Henzi, Kleger, and Ziegel (2023)

Statistical consistency

Let $(X_{in}, Y_{in})_{in}$ be a triangular array of iid random variables, and suppose that the model assumption holds

$$x \leq x' \implies \mathcal{L}(Y \mid X = x) \preceq_{\text{st}} \mathcal{L}(Y \mid X = x').$$

Under reasonable assumptions, IDR is uniformly consistent

- for a categorical covariate
  El Barmi and Mukerjee (2005)
- for a real-valued covariate
  Mösching and Dümbgen (2020)
- for a vector-valued covariate
  Henzi, Ziegel, and Gneiting (2021)

What if the "covariate" is a prediction, that is, it depends on the data?

- Distributional (single) index models (DIM): Consistency still holds.
  Henzi, Kleger, and Ziegel (2023)

Generalization beyond empirical distributions

Generalization of IDR: Isotonic conditional laws

$(\mathcal{X}, \leq)$ measurable space with partial order
Random vector (X, Y) with $X \in \mathcal{X}$, $Y \in \mathbb{R}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$

- Conditional law $\mathcal{L}(Y \mid X)$ of Y given X is distribution of Y "given all information about X".
- What is the conditional law of Y given X under the constraint that an increase in X results in an increase in Y?
- "Best" approximation $\mathcal{L}'$ to $\mathcal{L}(Y \mid X)$ that satisfies

$$x \leq x' \implies \mathcal{L}'(Y \mid X = x) \preceq_{\mathrm{st}} \mathcal{L}'(Y \mid X = x').$$

Solution: Isotonic conditional law of Y given X

$$\mathcal{L}(Y \mid \mathcal{A}(X)),$$

where $\mathcal{A}(X)$ is the σ-lattice generated by X.

σ-lattice $\mathcal{C} \subseteq \mathcal{F}$

$\mathcal{C}$ contains $\emptyset, \Omega$ and is closed under countable unions and intersections.

- Random variable Z is $\mathcal{C}$-measurable if $\{Z > a\} \in \mathcal{C}$ for all $a \in \mathbb{R}$.
- $\mathbb{E}(Y \mid \mathcal{C})$ is L^2-projection of Z onto closed convex cone of $\mathcal{C}$-measurable random variables. Brunk (1965)

$\mathcal{L}(Y \mid \mathcal{C})$

- Markov kernel from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- $\omega \mapsto \mathcal{L}(Y \mid \mathcal{C})(\omega, (a, \infty))$ is a version of $\mathbb{E}(\mathbb{1}\{Y > a\} \mid \mathcal{C})$ for any $a \in \mathbb{R}$.

σ-lattices and increasing functions

- Collection $\mathcal{U}$ of all upper sets in $(\mathcal{X}, \leq)$ is a σ-lattice
- $f : \mathcal{X} \to \mathbb{R}$ is increasing if and only f is $\mathcal{U}$-measurable, that is, $\{f > a\} \in \mathcal{U}$ for all $a \in \mathbb{R}$.

σ-lattice generated by X

$(\mathcal{X}, d, \leq)$ ordered metric space

$$\mathcal{A}(X) = \{X^{-1}(B) \mid B \in \mathcal{B}(\mathcal{X}) \cap \mathcal{U}\}$$

- IDR is the isotonic conditional law of Y given X if the joint distribution of (X, Y) has finite support.
- Isotonic conditional laws are also CRPS minimizers in a suitable sense, and share the calibration properties of IDR.

Arnold and Ziegel (2023)

Uncertainty quantification

IDR for uncertainty quantification: EasyUQ approach

- Training data of predictions x and outcomes Y.

"If the prediction increases we expect an increase of the outcome."

Basic EasyUQ

- Apply IDR to training data with x as covariate.
- For new prediction x, obtain predictive cdf $\hat{F}_x$ by interpolation.

Smooth EasyUQ

- Apply IDR to training data with x as covariate.
- For new prediction x, obtain smooth predictive cdf as

$$\check{F}_x(y) = \int_{\mathbb{R}} \hat{F}_x(t) K_h(y - t)\, \mathrm{d}t, \quad y \in \mathbb{R},$$

where $K_h(u) = (1/h)\kappa(u/h)$ for some kernel κ.

Henzi, Kleger, and Ziegel (2023); Walz, Henzi, Ziegel, and Gneiting (2023+)

Case study: Intensive care unit length of stay

Predicting length of stay (LoS) of patients in intensive care units (ICUs) has always been useful for . . .

- planning
- quality improvement
- benchmarking (risk-adjusted LoS)

Ideally: Predict LoS using data available 24 hours after admission.

With corona pandemic: ICU LoS has suddently become a topic of general public interest.

Case study: Intensive care unit length of stay

Dataset

Single patient data from 86 ICUs in Switzerland (2007-2018) [1]

- Between 700 and 46'000 observations per ICU
- Mean (median) LoS ranging from 1 (0.7) to 8 (2.2) days
- Variables used for prediction available 24 hours after admission
- ICU patient groups are highly heterogeneous.

Age	Sex	From	SAPS	NEMS	Interv	Diag	Admission	Discharge	LoS
81	M	emergency	28	9	I00	T1	2017-01-21 15:40:00	2017-01-23 11:20:48	1.82
29	M	op	39	9	I80	T1	2017-08-28 16:20:00	2017-09-16 19:56:00	19.15
64	M	op	34	20	I30	A1	2018-04-15 15:00:00	2018-04-16 15:57:36	1.04

Model for mean log-LoS (Verburg et al., 2017)

$$\log(\text{LoS} + 1) \sim \mathcal{N}(\theta(X), \sigma^2),$$

$$\theta(X) = \text{spline}(\textit{age}) + \text{spline}(\textit{severity})$$

$$+ \text{Dummy variables for diagnosis, admission source, } \ldots$$

[1] Provided by *G.-R. Kleger and Schweizerische Gesellschaft für Intensivmedizin*. Data is internal hospital data and not publicly available. ICU identifiers are generated randomly.

Case study: Intensive care unit length of stay

Dataset

Single patient data from 86 ICUs in Switzerland (2007-2018) [1]

- Between 700 and 46'000 observations per ICU
- Mean (median) LoS ranging from 1 (0.7) to 8 (2.2) days
- Variables used for prediction available 24 hours after admission
- ICU patient groups are highly heterogeneous.

Age	Sex	From	SAPS	NEMS	Interv	Diag	Admission	Discharge	LoS
81	M	emergency	28	9	I00	T1	2017-01-21 15:40:00	2017-01-23 11:20:48	1.82
29	M	op	39	9	I80	T1	2017-08-28 16:20:00	2017-09-16 19:56:00	19.15
64	M	op	34	20	I30	A1	2018-04-15 15:00:00	2018-04-16 15:57:36	1.04

Model for mean log-LoS (Verburg et al., 2017)

$$\log(\mathrm{LoS} + 1) \sim \mathcal{N}(\theta(X), \sigma^2),$$

$$\theta(X) = \mathrm{spline}(\mathit{age}) + \mathrm{spline}(\mathit{severity})$$
$$+ \text{Dummy variables for diagnosis, admission source, } \ldots$$

[1] Provided by *G.-R. Kleger and Schweizerische Gesellschaft für Intensivmedizin*. Data is internal hospital data and not publicly available. ICU identifiers are generated randomly.

Model for mean log-LoS

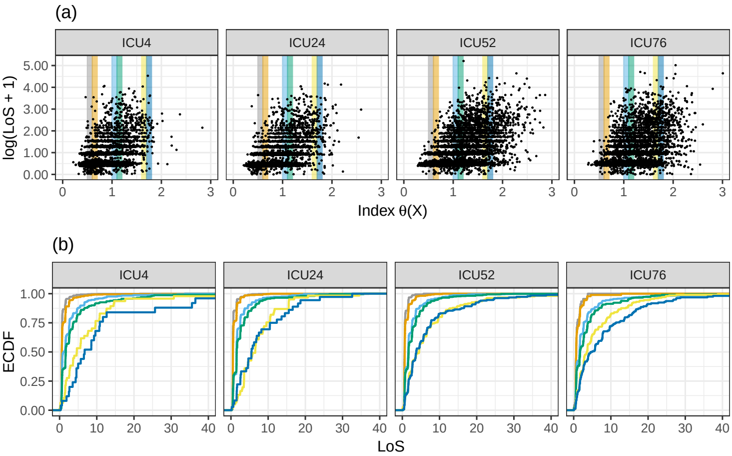


Figure: In-sample predicted and observed LoS (top) and empirical distribution function (ECDF) of LoS stratified by prediction (bottom).

Patient	Age	Sex	From	SAPS	NEMS	Interv	Diag	LoS
1	81	M	emergency	28	9	I00	T1	1.82
2	29	M	op	39	9	I80	T1	19.15
3	64	M	op	34	20	I30	A1	1.04

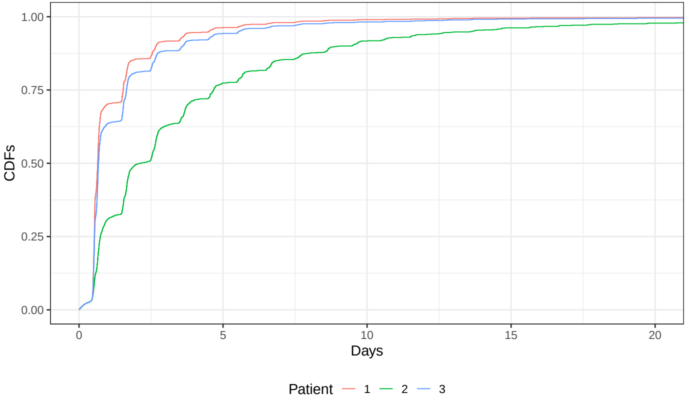


Figure: Predictive CDFs for the selected patients.

Results: Computed on validation dataset (last 20%)

Compare distributional index model (DIM, EasyUQ approach) predictions to

- empirical distribution of LoS on training dataset,
- Cox proportional hazards regression,
- quantile regression.

Mean CRPS of different LoS forecasts

ICU	p	DIM	Quantile reg.	Cox reg.	ECDF	Point
ICU4	$1.18 \cdot 10^{-11}$	1.074	1.076	1.089	1.191	1.399
ICU24	0	1.099	1.111	1.141	1.265	1.416
ICU52	$7.40 \cdot 10^{-5}$	1.845	1.866	1.868	2.121	2.580
ICU76	0	2.420	2.448	2.458	2.783	3.468
$\vdots$						
Mean		1.359	1.372	1.392	1.607	1.853

Henzi, Kleger, and Ziegel (2023)

Case study: Precipitation data

- 24-hour accumulated precipitation at Frankfurt airport, Germany, lead times 1–5 days
- Single-valued forecasts:
  - HRES model output of ECMWF
  - weakly climatology
- Training data 2007–2014
- Evaluation data 2016–2017

Results: Computed on evaluation data

Type	Forecast Method	CRPS 1 Day	CRPS 3 Days	CRPS 5 Days
Single-v.	Climatology	2.187	2.187	2.187
	HRES	1.125	1.412	1.686
Distrib.	CP on Climatology	1.382	1.382	1.382
	CP on HRES	0.886	1.063	1.129
	Censored CP on Climatology	1.324	1.324	1.324
	Censored CP on HRES	0.850	1.031	1.100
Distrib.	EasyUQ on Climatology	1.242	1.242	1.242
	EasyUQ on HRES	0.732	0.876	1.001
Distrib.	ECMWF Ensemble	0.752	0.856	0.981

CP: Conformal predictive system (Least squares prediction machine)

Walz, Henzi, Ziegel, and Gneiting (2023+)

Discussion of EasyUQ approach

- EasyUQ approch converts single-valued predictions into distributional ones: output-based uncertainty quantification technique.
- Basic EasyUQ is "just" IDR and does not involve tuning parameters. It finds the support of the distributions automatically.
- Extensive comparison of uncertainty quantification methods in machine learning examples in Walz, Henzi, Ziegel, and Gneiting (2023+)
- EasyUQ/IDR shows better performance than CP with respect to CRPS but CP comes with calibration guarantees (under exchangeability.)

Summary

- IDR is a non-parametric distributional regression technique under order constraints.
- IDR is in-sample optimal with respect to all weighted CRPS.
- IDR provides guarantees for calibration in-sample.
- Population version of IDR are isotonic conditional laws.
- IDR can be used to "add" a distribution to a given model for a univariate functional such as the mean: EasyUQ approach.
- Extensions (work in progress):
  - IDR for censored observations
  - Conformal IDR

Thank you!

References

S. Arnold and J. Ziegel. Isotonic conditional laws. *Preprint, arXiv:2307.09032*, 2023.

R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk. *Statistical inference under order restrictions*. John Wiley & Sons Ltd., London, 1972.

H. Brunk. Conditional expectation given a σ-lattice and applications. *Annals of Mathematical Statistics*, 36: 1339–1350, 1965.

H. El Barmi and H. Mukerjee. Inferences under a stochastic ordering constraint. *Journal of the American Statistical Association*, 100:252–261, 2005.

A. Henzi, J. F. Ziegel, and T. Gneiting. Isotonic distributional regression. *Journal of the Royal Statistical Society: Series B*, 85:963–993, 2021.

A. Henzi, A. Mösching, and L. Dümbgen. Accelerating the pool-adjacent-violators algorithm for isotonic distributional regression. *Methodology and Computing in Applied Probability*, 24:2633–2645, 2022.

A. Henzi, G.-R. Kleger, and J. F. Ziegel. Distributional (single) index models. *Journal of the American Statistical Association*, 118:489–503, 2023.

A. Mösching and L. Dümbgen. Monotone least squares and isotonic quantiles. *Electronic Journal of Statistics*, 14: 24–49, 2020.

I. W. M. Verburg, A. Atashi, S. Eslami, R. Holman, A. Abu-Hanna, E. de Jonge, N. Peek, and N. F. de Keizer. Which models can I use to predict adult ICU length of stay? A systematic review. *Critical Care Medicine*, 45: e222–e231, 2017.

E.-M. Walz, A. Henzi, J. Ziegel, and T. Gneiting. Easy uncertainty quantification (EasyUQ): Generating predictive distributions from single-valued model output. *SIAM Review*, 2023+. To appear.