LAMARR

INSTITUTE FOR
MACHINE LEARNING
AND ARTIFICIAL
INTELLIGENCE

In Machine Learning, *Unfolding* Is Known As *Quantification Learning*

Bridging the Gap: Unfolding & Quantification Learning for Physics Research

Mirko Bunse

EuCAIFCon 2025 – June 19th

Partner institutions:

technische universität dortmund

Fraunhofer IAIS

UNIVERSITÄT BONN

Fraunhofer IML

Institutionally funded by:

Bundesministerium für Bildung und Forschung

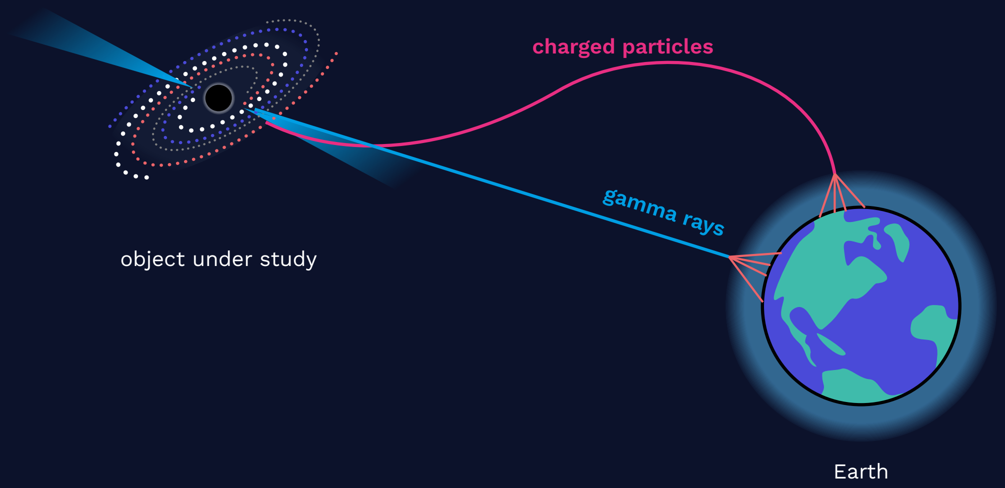
Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen

Why Bother?

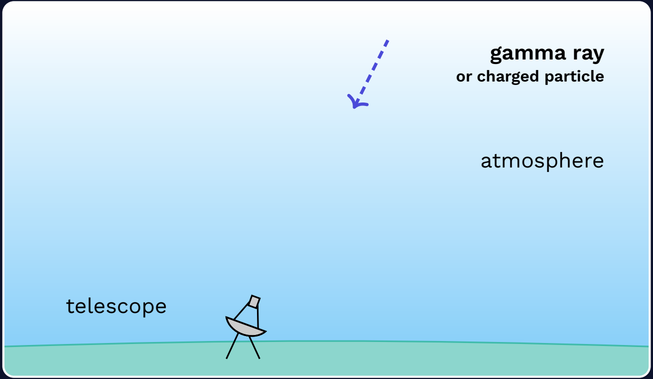
Research on quantification learning offers:

- Fast-paced improvements of methods
- Few limitations (e.g., no limitation on the number of observables)
- Comprehensive theoretical understanding
- Interdisciplinary community eager to explore new aspects
- Funding opportunities

Example Use Case: Ground-Based Gamma-Ray Observatories

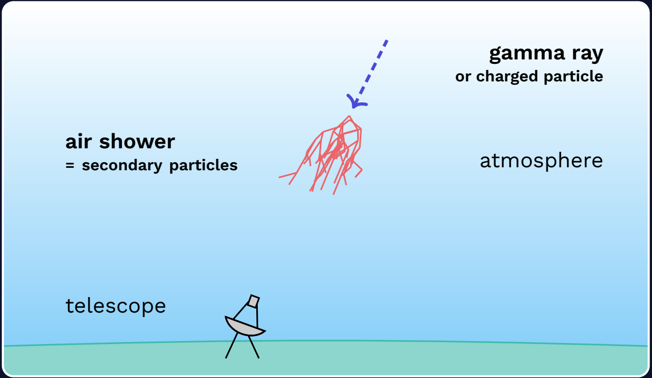


Example Use Case: Ground-Based Gamma-Ray Observatories



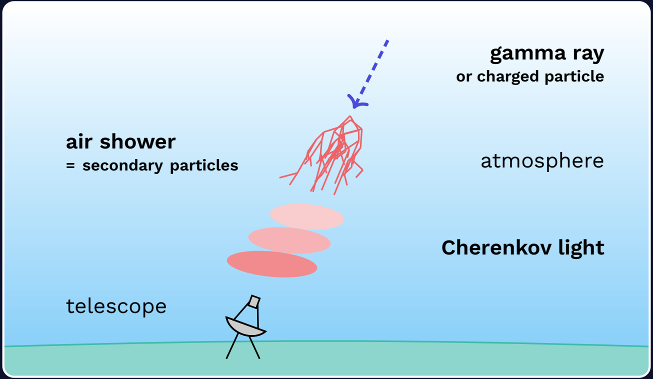
[1] **Fig:** Bockermann et al., "Online Analysis of High-Volume Data Streams in Astroparticle Physics", 2015

Example Use Case: Ground-Based Gamma-Ray Observatories



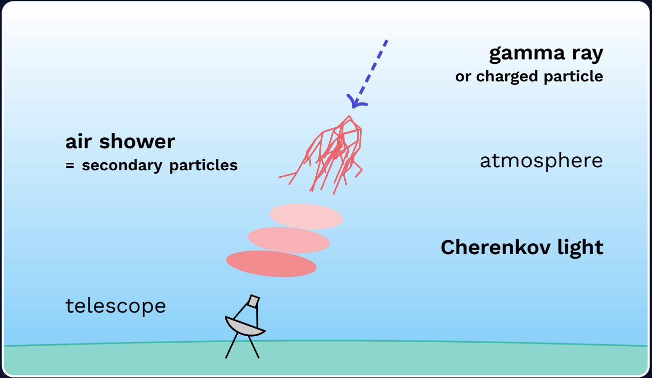
[1] **Fig:** Bockermann et al., “Online Analysis of High-Volume Data Streams in Astroparticle Physics”, 2015

Example Use Case: Ground-Based Gamma-Ray Observatories



[1] **Fig:** Bockermann et al., "Online Analysis of High-Volume Data Streams in Astroparticle Physics", 2015

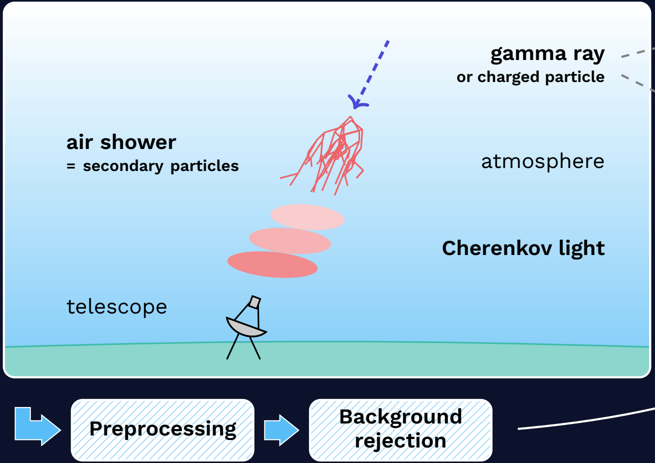
Example Use Case: Ground-Based Gamma-Ray Observatories

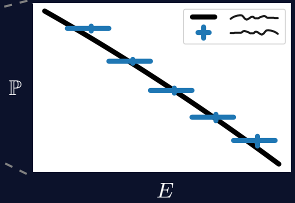


Preprocessing → Background rejection

[1] **Fig:** Bockermann et al., “Online Analysis of High-Volume Data Streams in Astroparticle Physics”, 2015

Example Use Case: Ground-Based Gamma-Ray Observatories



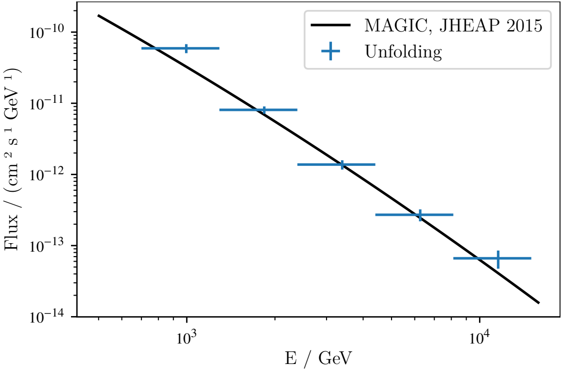


[1] **Fig:** Bockermann et al., “Online Analysis of High-Volume Data Streams in Astroparticle Physics”, 2015

Unfolding: the Reconstruction of Spectra

Goal: reconstruct the spectrum $p(y)$ of some quantity y from a measurement $q(x)$.

$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$



[2] Fig.: Morik and Rhode, *Machine Learning under Resource Constraints – Discovery in Physics*, 2023

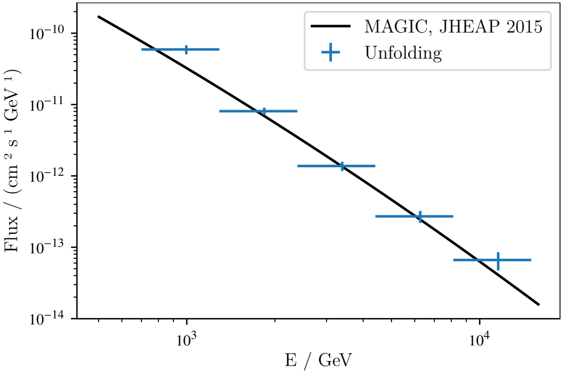
Unfolding: the Reconstruction of Spectra

Goal: reconstruct the spectrum $p(y)$ of some quantity y from a measurement $q(x)$.

$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$

Approach: set up a linear system of equations

$$\mathbf{q} = \mathbf{M}\mathbf{p} \qquad \text{where} \quad \begin{cases} \mathbf{q} = \frac{1}{|\mathrm{B}\,|} \sum_{x \in \mathrm{B}} \phi(x) \\ \mathbf{M}_i = \frac{1}{|\mathrm{D}_i|} \sum_{x \in \mathrm{D}_i} \phi(x) \end{cases}$$



[2] Fig.: Morik and Rhode, *Machine Learning under Resource Constraints – Discovery in Physics*, 2023

Unfolding: the Reconstruction of Spectra

Goal: reconstruct the spectrum $p(y)$ of some quantity y from a measurement $q(x)$.

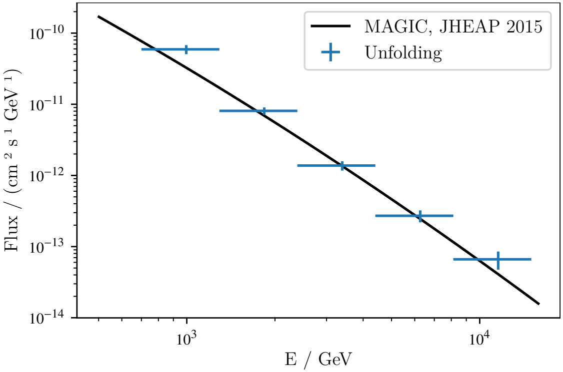
$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$

Approach: set up a linear system of equations

$$\mathbf{q} = \mathbf{M}\mathbf{p} \qquad \text{where} \quad \begin{cases} \mathbf{q} &= \frac{1}{|\mathrm{B}\,|} \sum_{x \in \mathrm{B}} \phi(x) \\ \mathbf{M}_i &= \frac{1}{|\mathrm{D}_i|} \sum_{x \in \mathrm{D}_i} \phi(x) \end{cases}$$

and solve it by minimizing some loss, i.e.,

$$\hat{\mathbf{p}} = \operatorname*{arg\,min}_{\mathbf{p} \in \Delta^{C-1}} \ell\left(\mathbf{p};\ \mathbf{q}, \mathbf{M}\right)$$



[2] Fig.: Morik and Rhode, *Machine Learning under Resource Constraints – Discovery in Physics*, 2023

Unfolding: the Reconstruction of Spectra

Goal: reconstruct the spectrum $p(y)$ of some quantity y from a measurement $q(x)$.

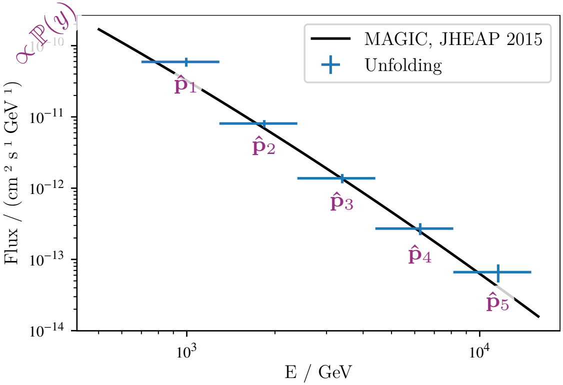
$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$

Approach: set up a linear system of equations

$$\mathbf{q} = \mathbf{M}\mathbf{p} \qquad \text{where} \begin{cases} \mathbf{q} & = \frac{1}{|\mathrm{B}\,|} \sum_{x \in \mathrm{B}} \; \phi(x) \\ \mathbf{M}_i & = \frac{1}{|\mathrm{D}_i|} \sum_{x \in \mathrm{D}_i} \phi(x) \end{cases}$$

and solve it by minimizing some loss, i.e.,

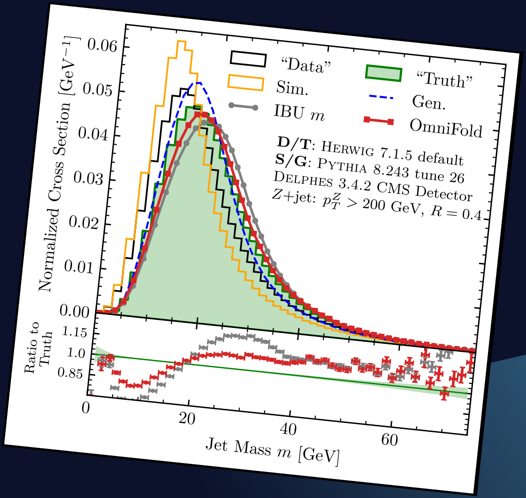
$$\hat{\mathbf{p}} \; = \; \underset{\mathbf{p} \in \Delta^{C-1}}{\arg\min} \; \ell\,(\mathbf{p};\; \mathbf{q}, \mathbf{M})$$



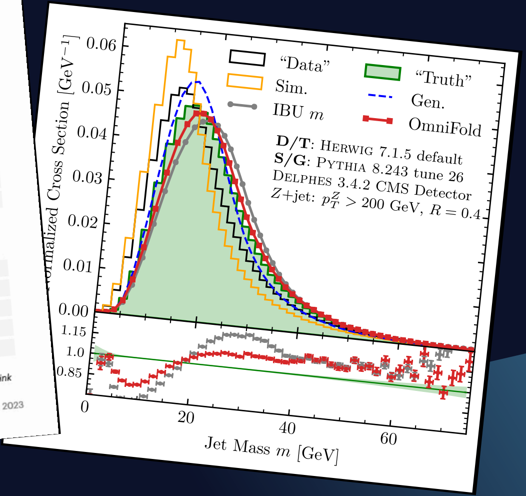
unknown $y \in [1, 5]$
$\equiv$ class label!

[2] Fig.: Morik and Rhode, *Machine Learning under Resource Constraints – Discovery in Physics*, 2023

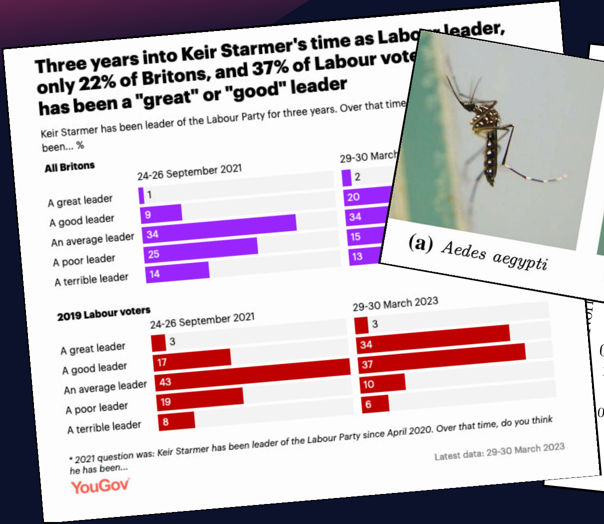
Use Cases Everywhere: Count + Predict

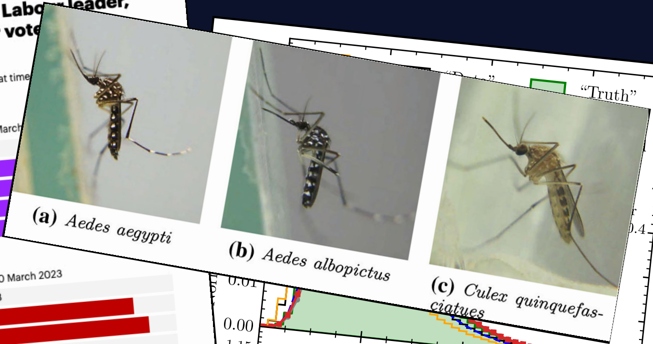


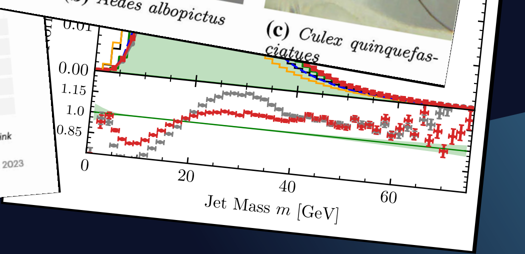
Use Cases Everywhere: Count + Predict



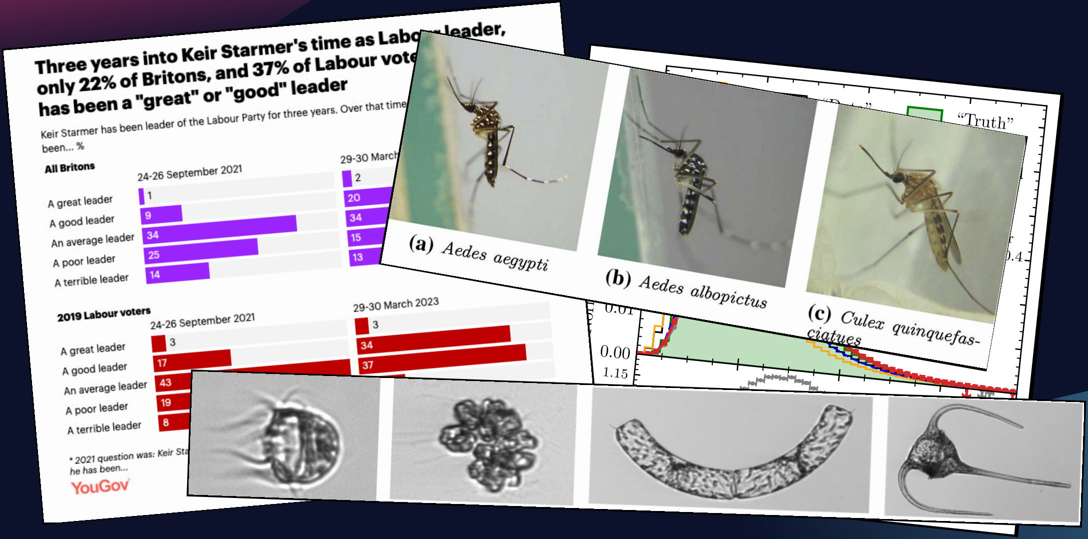
Use Cases Everywhere: Count + Predict




(a) *Aedes aegypti* (b) *Aedes albopictus* (c) *Culex quinquefasciatus*



Use Cases Everywhere: Count + Predict



Quantification

In Computer Science, unfolding-like problems are covered by **Quantification Learning**[3,4].

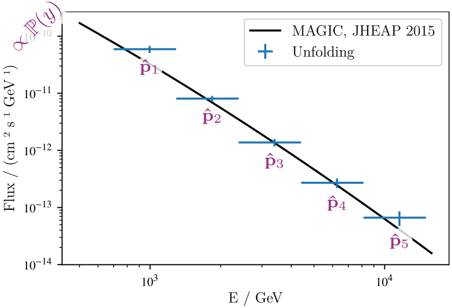
[3] Esuli et al., *Learning to Quantify*, 2023.
[4] Forman, “Quantifying counts and costs via classification”, 2008, .

Quantification

In Computer Science, unfolding-like problems are covered by **Quantification Learning**[3,4].

Goal: count the frequency $p(y)$, which quantifies how often class y occurs.

$$\underbrace{q(x)}_{\text{measurement}} = \sum_{y \in \mathcal{Y}} \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$



unknown $y \in [1, 5]$
$\equiv$ class label!

[3] Esuli et al., *Learning to Quantify*, 2023.
[4] Forman, “Quantifying counts and costs via classification”, 2008, .

Quantification

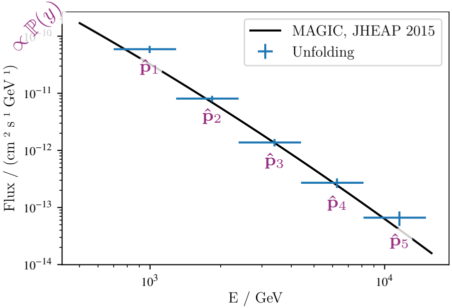
In Computer Science, unfolding-like problems are covered by **Quantification Learning**[3,4].

Goal: count the frequency $p(y)$, which quantifies how often class y occurs.

$$\underbrace{q(x)}_{\text{measurement}} = \sum_{y \in \mathcal{Y}} \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$

Approach: solve $\mathbf{q} = \mathbf{M}\mathbf{p}$ for $\mathbf{p}$ (just like before!)

Hence, unfolding $\equiv$ quantification.



unknown $y \in [1, 5]$
$\equiv$ class label!

[3] Esuli et al., *Learning to Quantify*, 2023.
[4] Forman, “Quantifying counts and costs via classification”, 2008, .

5 Learnings From Quantification Research

1st Learning: Statistical Consistency

Definition (Fisher Consistency for Prior Probability Shift):

If a consistent quantifier had access to the entire population $\mathbb{Q}(X)$ (i.e., to "unlimited data"), it would return the true class prevalences:

$$\underbrace{h'\big(\mathbb{Q}(X)\big)}_{\substack{\text{population}\\\text{analogue}\\\text{of } h(\mathrm{B})}} = \mathbb{Q}(Y) \quad \underbrace{\forall\ \mathbb{Q} : \mathbb{Q}(X \mid Y) = \mathbb{P}(X \mid Y)}_{\text{for any } \mathbb{Q} \text{ with PPS}}$$

[5] Blobel, "An unfolding method for high energy physics experiments", 2002, .

[6] Bunse, "Unification of Algorithms for Quantification and Unfolding", 2022.

[7] Gövert, "Fisher-Konsistenz für Quantification-Algorithmen", 2023.

1^{st} Learning: Statistical Consistency

Definition (Fisher Consistency for Prior Probability Shift):

If a consistent quantifier had access to the entire population $\mathbb{Q}(X)$ (i.e., to "unlimited data"), it would return the true class prevalences:

$$\underbrace{h'\big(\mathbb{Q}(X)\big)}_{\substack{\text{population}\\ \text{analogue}\\ \text{of } h(\mathrm{B})}} = \mathbb{Q}(Y) \quad \underbrace{\forall\ \mathbb{Q} : \mathbb{Q}(X \mid Y) = \mathbb{P}(X \mid Y)}_{\text{for any } \mathbb{Q} \text{ with PPS}}$$

- can also be defined for other types of data set shift
- not a sufficient but certainly **a necessary criterion** for quantifier selection

[5] Blobel, "An unfolding method for high energy physics experiments", 2002, .

[6] Bunse, "Unification of Algorithms for Quantification and Unfolding", 2022.

[7] Gövert, "Fisher-Konsistenz für Quantification-Algorithmen", 2023.

1st Learning: Statistical Consistency

Definition (Fisher Consistency for Prior Probability Shift):

If a consistent quantifier had access to the entire population $\mathbb{Q}(X)$ (i.e., to “unlimited data”), it would return the true class prevalences:

$$\underbrace{h'\big(\mathbb{Q}(X)\big)}_{\substack{\text{population}\\\text{analogue}\\\text{of } h(\mathrm{B})}} = \mathbb{Q}(Y) \quad \underbrace{\forall\ \mathbb{Q} : \mathbb{Q}(X \mid Y) = \mathbb{P}(X \mid Y)}_{\text{for any } \mathbb{Q} \text{ with PPS}}$$

- can also be defined for other types of data set shift
- not a sufficient but certainly **a necessary criterion** for quantifier selection

1) RUN[5] **/ TRUEE** (and others) *are* Fisher consistent[6] ✓

2) DSEA & DSEA+ are *not* Fisher consistent[7] ✗

[5] Blobel, “An unfolding method for high energy physics experiments”, 2002, .

[6] Bunse, “Unification of Algorithms for Quantification and Unfolding”, 2022.

[7] Gövert, “Fisher-Konsistenz für Quantification-Algorithmen”, 2023.

2nd Learning: The Anatomy of Prediction Errors

Prediction Error Bound:[8] describes the *impact of* and the *interplay between* causes of errors.

$$\underbrace{\left\| h(\mathrm{B}) - \mathbf{p}^* \right\|_2}_{\text{prediction error}} \leq \underbrace{\frac{2k(2 + \sqrt{2 \log \frac{2C}{\delta}})}{\sqrt{\lambda_2}}}_{\text{representation } \phi} \cdot \left(\underbrace{\left\| \frac{\mathbf{p}^*}{\mathbf{p}_{\mathrm{trn}}} \right\|_2}_{\text{shift}} \cdot \underbrace{\frac{1}{\sqrt{|\mathrm{D}|}}}_{\text{volume } \mathrm{D}} + \underbrace{\frac{1}{\sqrt{|\mathrm{B}|}}}_{\text{volume } \mathrm{B}} \right)$$

where

- $h(\mathrm{B})$ is the solution of $\mathbf{q} = \mathbf{M}\mathbf{p}$
- k is a constant s.t. $\|\phi(x)\|_2 \leq k \;\forall\, x \in \mathcal{X}$
- λ_2 is the second-smallest eigenvalue of some particular $\mathbf{G}$
- δ is the desired probability

[8] Dussap, Blanchard, and Chérief-Abdellatif, "Label Shift Quantification with Robustness Guarantees via Distribution Feature Matching", 2023, .

3rd Learning: Improved Optimization Techniques

Algorithm	Estimate	Validity
RUN[5]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^C}{\arg\min} \; \ell(\mathbf{p};\ \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗

[9] Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.
[10] Bunse, “On Multi-Class Extensions of Adjusted Classify and Count”, 2022, .

3rd Learning: Improved Optimization Techniques

Algorithm	Estimate	Validity
RUN[5]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^C}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗
TRUEE[9]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \geq \mathbf{0}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗

[9] Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.
[10] Bunse, “On Multi-Class Extensions of Adjusted Classify and Count”, 2022, .

3rd Learning: Improved Optimization Techniques

Algorithm	Estimate	Validity
RUN[5]	$\hat{\mathbf{p}} = \underset{\boxed{\mathbf{p} \in \mathbb{R}^C}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗
TRUEE[9]	$\hat{\mathbf{p}} = \underset{\boxed{\mathbf{p} \geq \mathbf{0}}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗
Constrained[10]	$\hat{\mathbf{p}} = \underset{\boxed{\mathbf{p} \in \Delta^{C-1}}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	valid ✓

[9] Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.
[10] Bunse, “On Multi-Class Extensions of Adjusted Classify and Count”, 2022, .

3rd Learning: Improved Optimization Techniques

Algorithm	Estimate	Validity
RUN[5]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^C}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗
TRUEE[9]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \geq \mathbf{0}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗
Constrained[10]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \Delta^{C-1}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	valid ✓
Soft-Max[10]	$\hat{\mathbf{p}} = \sigma(\mathbf{l}^*) , \quad \mathbf{l}^* = \underset{\mathbf{l} \in \mathbb{R}^{C-1}}{\arg\min} \; \ell(\sigma(\mathbf{l})\; ; \; \mathbf{q}, \mathbf{M})$	valid ✓

[9] Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.
[10] Bunse, “On Multi-Class Extensions of Adjusted Classify and Count”, 2022, .

4^{th} Learning: Methods Are Numerous

Most methods are combinations of

- a data representation $\phi : \mathcal{X} \rightarrow \mathcal{Z}$
- a loss function $\ell : \mathcal{Z} \times \mathcal{Z} \rightarrow \mathbb{R}$
- an optimization algorithm

These components can be recombined to even more methods.

[11] Bella et al., “Quantification via Probability Estimators”, 2010, .

[12] González-Castro, Alaíz-Rodríguez, and Alegre, “Class distribution estimation based on the Hellinger distance”, 2013, .

[13] Börner et al., “Measurement/Simulation Mismatches and Multivariate Data Discretization in the Machine Learning Era”, 2020, .

[14] Kawakubo, Plessis, and Sugiyama, “Computationally Efficient Class-Prior Estimation under Class Balance Change Using Energy Distance”, 2016, .

4^{th} Learning: Methods Are Numerous

Most methods are combinations of

- a data representation $\phi : \mathcal{X} \rightarrow \mathcal{Z}$
- a loss function $\ell : \mathcal{Z} \times \mathcal{Z} \rightarrow \mathbb{R}$
- an optimization algorithm

github.com/mirkobunse/qunfold

These components can be recombined to even more methods.

Representations: hard[4] & soft[11] classification, histograms[12], tree-based binnings[13], kernel means[14], ...

Loss Functions: least squares[4,11], Hellinger distance[12], energy distance[14], Poisson likelihood[5], ...

[11] Bella et al., “Quantification via Probability Estimators”, 2010, .

[12] González-Castro, Alaíz-Rodríguez, and Alegre, “Class distribution estimation based on the Hellinger distance”, 2013, .

[13] Börner et al., “Measurement/Simulation Mismatches and Multivariate Data Discretization in the Machine Learning Era”, 2020, .

[14] Kawakubo, Plessis, and Sugiyama, “Computationally Efficient Class-Prior Estimation under Class Balance Change Using Energy Distance”, 2016, .

5th Learning: Open Issues in Quantification Research

Complications of experimental physics:

- ordinality: $y_i \prec y_{i+1} \ \forall \, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[15])

[15] Bunse et al., "Regularization-based Methods for Ordinal Quantification", 2024.

5th Learning: Open Issues in Quantification Research

Complications of experimental physics:

- ordinality: $y_i \prec y_{i+1} \;\forall\, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[15])
- background: $\mathbb{Q}(\mathbf{x}) = \mathbb{Q}(\mathbf{x}, \varnothing) + \sum_{y=1}^{C} \mathbb{Q}(\mathbf{x}, y)$ (PPS with a noise class[8])

[15] Bunse et al., “Regularization-based Methods for Ordinal Quantification”, 2024.

5th Learning: Open Issues in Quantification Research

Complications of experimental physics:

- ordinality: $y_i \prec y_{i+1} \; \forall \, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[15])
- background: $\mathbb{Q}(\mathbf{x}) = \mathbb{Q}(\mathbf{x}, \varnothing) + \sum_{y=1}^{C} \mathbb{Q}(\mathbf{x}, y)$ (PPS with a noise class[8])
- acceptance / class-conditional selection bias: $\mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_i) \neq \mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_j) \; \exists \, i \neq j$

[15] Bunse et al., “Regularization-based Methods for Ordinal Quantification”, 2024.

5th Learning: Open Issues in Quantification Research

Complications of experimental physics:

- ordinality: $y_i \prec y_{i+1} \ \forall\, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[15])
- background: $\mathbb{Q}(\mathbf{x}) = \mathbb{Q}(\mathbf{x}, \varnothing) + \sum_{y=1}^{C} \mathbb{Q}(\mathbf{x}, y)$ (PPS with a noise class[8])
- acceptance / class-conditional selection bias: $\mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_i) \neq \mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_j) \ \exists\, i \neq j$
- changing environment: $\mathbb{Q}(\mathbf{x}, y) = \sum_{e \in \mathcal{E}} \mathbb{Q}(\mathbf{x}, y, e)$

15 Bunse et al., “Regularization-based Methods for Ordinal Quantification”, 2024.

5^{th} Learning: Open Issues in Quantification Research

Complications of experimental physics:

- ordinality: $y_i \prec y_{i+1} \;\forall\, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[15])
- background: $\mathbb{Q}(\mathbf{x}) = \mathbb{Q}(\mathbf{x}, \varnothing) + \sum_{y=1}^{C} \mathbb{Q}(\mathbf{x}, y)$ (PPS with a noise class[8])
- acceptance / class-conditional selection bias: $\mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_i) \neq \mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_j) \;\exists\, i \neq j$
- changing environment: $\mathbb{Q}(\mathbf{x}, y) = \sum_{e \in \mathcal{E}} \mathbb{Q}(\mathbf{x}, y, e)$
- data-MC mismatches / concept shift: $\mathbb{Q}(\mathbf{x} \mid y) \neq \mathbb{P}(\mathbf{x} \mid y)$ (in addition to PPS)

[15] Bunse et al., “Regularization-based Methods for Ordinal Quantification”, 2024.

5th Learning: Open Issues in Quantification Research

Complications of experimental physics:

- ordinality: $y_i \prec y_{i+1} \;\forall\, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[15])
- background: $\mathbb{Q}(\mathbf{x}) = \mathbb{Q}(\mathbf{x}, \varnothing) + \sum_{y=1}^{C} \mathbb{Q}(\mathbf{x}, y)$ (PPS with a noise class[8])
- acceptance / class-conditional selection bias: $\mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_i) \neq \mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_j) \;\exists\, i \neq j$
- changing environment: $\mathbb{Q}(\mathbf{x}, y) = \sum_{e \in \mathcal{E}} \mathbb{Q}(\mathbf{x}, y, e)$
- data-MC mismatches / concept shift: $\mathbb{Q}(\mathbf{x} \mid y) \neq \mathbb{P}(\mathbf{x} \mid y)$ (in addition to PPS)
- inspect contributions of individual data items $\mathbf{x} \in \mathrm{B}$ to $h(\mathrm{B})$ (data selection, human in the loop)

Hence, there are substantial opportunities for quantification-related research in Computer Science.

15 Bunse et al., “Regularization-based Methods for Ordinal Quantification”, 2024.

Recap: Reconstruction of Spectra

Goal: reconstruct the spectrum $p(y)$ of some quantity y from a measurement $q(x)$.

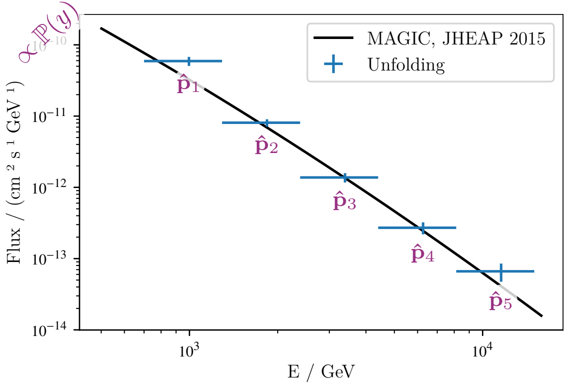
$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$

Approach: set up a linear system of equations

$$\mathbf{q} = \mathbf{M}\mathbf{p} \qquad \text{where} \begin{cases} \mathbf{q} = \frac{1}{|\mathrm{B}\,|} \sum_{x \in \mathrm{B}} \phi(x) \\ \mathbf{M}_i = \frac{1}{|\mathrm{D}_i|} \sum_{x \in \mathrm{D}_i} \phi(x) \end{cases}$$

and solve it by minimizing some loss, i.e.,

$$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \Delta^{C-1}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$$



unknown $y \in [1, 5]$

$\equiv$ class label!

[2] Fig.: Morik and Rhode, *Machine Learning under Resource Constraints – Discovery in Physics*, 2023

Conclusion: 5 Learnings From Quantification

Understanding of the Problem Statement:

1) Consistency is a necessary criterion for algorithm selection

2) The prediction error is governed by the representation, the amount of shift, and the data volumes

Conclusion: 5 Learnings From Quantification

Understanding of the Problem Statement:

1) Consistency is a necessary criterion for algorithm selection

2) The prediction error is governed by the representation, the amount of shift, and the data volumes

Improvements of the Methods:

3) Contraints must be implemented, either explicitly or via soft-max

4) Many methods—or aspects thereof—have a potential for improving physics analyses

5) Physics applications motivate further developments in quantification research