LAMARR
INSTITUTE FOR MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Quantification Learning for Inverse Problems

Mirko Bunse

Physics Monthly – Jan. 20th, 2025

Partner institutions:

technische universität dortmund
Fraunhofer IAIS
UNIVERSITÄT BONN
Fraunhofer IML

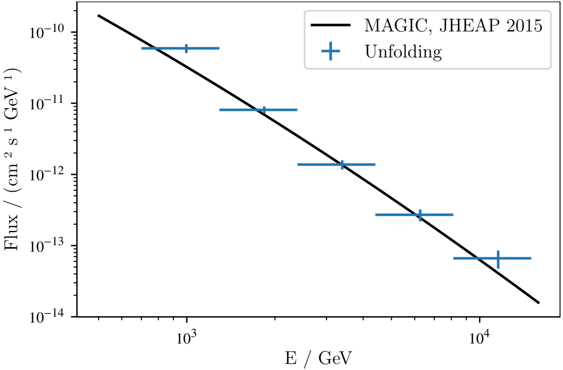
Institutionally funded by:

Bundesministerium für Bildung und Forschung
Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen

Inverse Problem: Reconstruction of Spectra

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

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



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

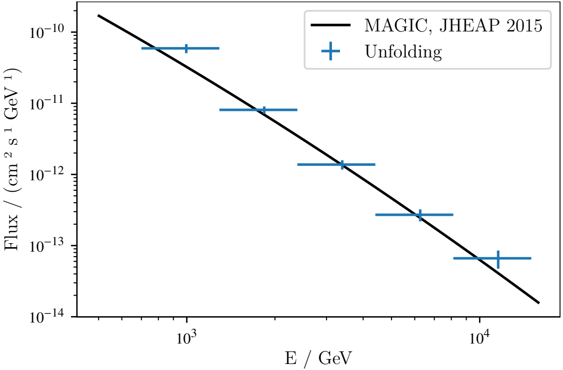
Inverse Problem: Reconstruction of Spectra

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

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



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

Inverse Problem: Reconstruction of Spectra

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

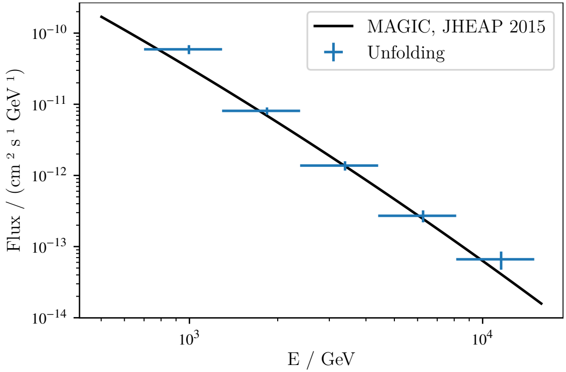
$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$

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.,

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



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

Inverse Problem: Reconstruction of Spectra

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

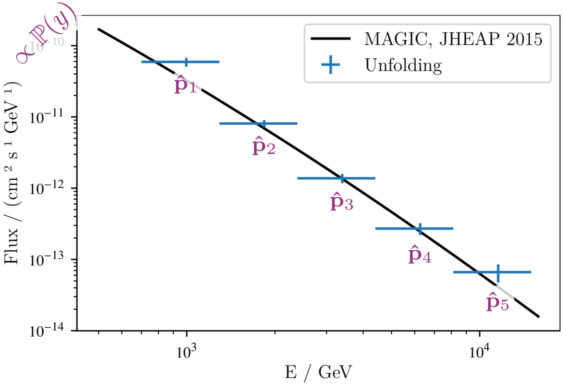
$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$

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!

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

Quantification

In Computer Science, this inverse problem is covered by **Quantification Learning**[2,3].

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

Quantification

In Computer Science, this inverse problem is covered by **Quantification Learning**[2,3].

Learn: a quantifier $\lambda : \bigcup_{m=1}^{\infty} \mathcal{X}^m \rightarrow \Delta^{C-1}$ where

- $\mathcal{X}$ is the feature space (e.g., $\mathcal{X} = \mathbb{R}^d$)
- $\Delta^{C-1} = \left\{ \mathbf{p} \in \mathbb{R}^C \; : \; \mathbf{p}_i \geq 0 \, \forall i, \; \sum_{i=1}^{C} \mathbf{p}_i = 1 \right\}$ is the space of class prevalences
- for any bag $\mathrm{B} \sim \mathbb{Q}^m$, we want to achieve that $\lambda(\mathrm{B}) = \mathbb{Q}(Y)$

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

Quantification

In Computer Science, this inverse problem is covered by **Quantification Learning**[2,3].

Learn: a quantifier $\lambda : \bigcup_{m=1}^{\infty} \mathcal{X}^m \rightarrow \Delta^{C-1}$ where

- $\mathcal{X}$ is the feature space (e.g., $\mathcal{X} = \mathbb{R}^d$)
- $\Delta^{C-1} = \left\{ \mathbf{p} \in \mathbb{R}^C \,:\, \mathbf{p}_i \geq 0 \, \forall i, \, \sum_{i=1}^{C} \mathbf{p}_i = 1 \right\}$ is the space of class prevalences
- for any bag $\mathrm{B} \sim \mathbb{Q}^m$, we want to achieve that $\lambda(\mathrm{B}) = \mathbb{Q}(Y)$

Given: a labeled training set $\mathrm{D} = \{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}\}_{i=1}^{n} \sim \mathbb{P}^n$ where $\mathbb{P} \neq \mathbb{Q}$ (e.g., MCs)

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

Quantification

In Computer Science, this inverse problem is covered by **Quantification Learning**[2,3].

Learn: a quantifier $\lambda : \bigcup_{m=1}^{\infty} \mathcal{X}^m \to \Delta^{C-1}$ where

- $\mathcal{X}$ is the feature space (e.g., $\mathcal{X} = \mathbb{R}^d$)
- $\Delta^{C-1} = \left\{ \mathbf{p} \in \mathbb{R}^C \, : \, \mathbf{p}_i \geq 0 \, \forall i, \, \sum_{i=1}^{C} \mathbf{p}_i = 1 \right\}$ is the space of class prevalences
- for any bag $\mathrm{B} \sim \mathbb{Q}^m$, we want to achieve that $\lambda(\mathrm{B}) = \mathbb{Q}(Y)$

Given: a labeled training set $\mathrm{D} = \{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}\}_{i=1}^{n} \sim \mathbb{P}^n$ where $\mathbb{P} \neq \mathbb{Q}$ (e.g., MCs)

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

} Quantification $\equiv$ Unfolding

[2] Esuli et al., *Learning to Quantify*, 2023.

[3] Forman, “Quantifying counts and costs via classification”, 2008, .

5 Learnings From Quantification Research

1st Learning: 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:

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

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

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

1st Learning: 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:

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

- can also be defined for other types of data set shift
- does not indicate good performance on finite samples
- hence, not a sufficient but certainly **a necessary criterion** for quantifier selection

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

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

1st Learning: 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:

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

- can also be defined for other types of data set shift
- does not indicate good performance on finite samples
- hence, not a sufficient but certainly **a necessary criterion** for quantifier selection

1) RUN / TRUEE (and others) *are* Fisher consistent[4] ✓

2) DSEA & DSEA+ are *not* Fisher consistent[5] ✗

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

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

2nd Learning: The Anatomy of Prediction Errors

Prediction Error Bound:[6] describes the *impact of* and the *interplay between* causes of errors.

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

where

- $\lambda(\mathrm{B})$ 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

6 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[7]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^C}{\arg\min} \; \ell(\mathbf{p};\, \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗

[7] Blobel, *Unfolding methods in high-energy physics experiments*, 1985, .

[8] Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.

[9] Bunse, “On Multi-Class Extensions of Adjusted Classify and Count”, 2022, .

3^{rd} Learning: Improved Optimization Techniques

Algorithm	Estimate	Validity
RUN[7]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^C}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗
TRUEE[8]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \geq \mathbf{0}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗

[7] Blobel, *Unfolding methods in high-energy physics experiments*, 1985, .

[8] Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.

[9] Bunse, “On Multi-Class Extensions of Adjusted Classify and Count”, 2022, .

3rd Learning: Improved Optimization Techniques

Algorithm	Estimate	Validity
RUN[7]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^C}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗
TRUEE[8]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \geq \mathbf{0}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗
Constrained[9]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \Delta^{C-1}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	valid ✓

7 Blobel, *Unfolding methods in high-energy physics experiments*, 1985, .

8 Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.

9 Bunse, “On Multi-Class Extensions of Adjusted Classify and Count”, 2022, .

3rd Learning: Improved Optimization Techniques

Algorithm	Estimate	Validity
RUN[7]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^C}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗
TRUEE[8]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \geq \mathbf{0}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta^{C-1}$ ✗
Constrained[9]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \Delta^{C-1}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	valid ✓
Soft-Max[9]	$\hat{\mathbf{p}} = \sigma(\mathbf{l}^*), \quad \mathbf{l}^* = \underset{\mathbf{l} \in \mathbb{R}^{C-1}}{\arg\min} \; \ell(\sigma(\mathbf{l});\; \mathbf{q}, \mathbf{M})$	valid ✓

[7] Blobel, *Unfolding methods in high-energy physics experiments*, 1985, .
[8] Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.
[9] 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.

[10] Bella et al., "Quantification via Probability Estimators", 2010, .
[11] González-Castro, Alaíz-Rodríguez, and Alegre, "Class distribution estimation based on the Hellinger distance", 2013, .
[12] Börner et al., "Measurement/Simulation Mismatches and Multivariate Data Discretization in the Machine Learning Era", 2020, .
[13] Kawakubo, Plessis, and Sugiyama, "Computationally Efficient Class-Prior Estimation under Class Balance Change Using Energy Distance", 2016, .

4th 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.

Representations: hard[3] & soft[10] classification, histograms[11], tree-based binnings[12], kernel means[13], ...

Loss Functions: least squares[3,10], Hellinger distance[11], energy distance[13], Poisson likelihood[7], ...

10 Bella et al., “Quantification via Probability Estimators”, 2010, .
11 González-Castro, Alaíz-Rodríguez, and Alegre, “Class distribution estimation based on the Hellinger distance”, 2013, .
12 Börner et al., “Measurement/Simulation Mismatches and Multivariate Data Discretization in the Machine Learning Era”, 2020, .
13 Kawakubo, Plessis, and Sugiyama, “Computationally Efficient Class-Prior Estimation under Class Balance Change Using Energy Distance”, 2016, .

4^{th} Learning: Methods Are Numerous

Likelihood Principle:

$$\begin{aligned}\mathcal{L}(\mathbf{p} \mid \mathrm{B}) &= \mathbb{Q}(\mathrm{B} \mid \mathbf{p}) \\ &= \prod_{\mathbf{x} \in \mathrm{B}} \mathbb{Q}(\mathbf{x} \mid \mathbf{p}) \\ &\overset{\text{PPS}}{=} \prod_{\mathbf{x} \in \mathrm{B}} \sum_{y \in \mathcal{Y}} \mathbb{P}(\mathbf{x} \mid y) \cdot \mathbf{p}_y\end{aligned}$$

[14] Alexandari, Kundaje, and Shrikumar, “Maximum Likelihood with Bias-Corrected Calibration is Hard-To-Beat at Label Shift Adaptation”, 2020.

4^{th} Learning: Methods Are Numerous

Likelihood Principle:

$$
\begin{aligned}
\mathcal{L}(\mathbf{p} \mid \mathrm{B}) &= \mathbb{Q}(\mathrm{B} \mid \mathbf{p}) \\
&= \prod_{\mathbf{x} \in \mathrm{B}} \mathbb{Q}(\mathbf{x} \mid \mathbf{p}) \\
&\overset{\text{PPS}}{=} \prod_{\mathbf{x} \in \mathrm{B}} \sum_{y \in \mathcal{Y}} \mathbb{P}(\mathbf{x} \mid y) \cdot \mathbf{p}_y
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \quad -\log \mathcal{L}(\mathbf{p} \mid \mathrm{B}) &= -\sum_{\mathbf{x} \in \mathrm{B}} \log \sum_{y \in \mathcal{Y}} \mathbb{P}(\mathbf{x} \mid y) \cdot \mathbf{p}_y \\
&\propto -\sum_{\mathbf{x} \in \mathrm{B}} \log \sum_{y \in \mathcal{Y}} \frac{\mathbb{P}(y|\mathbf{x})}{\mathbb{P}(y)} \cdot \mathbf{p}_y
\end{aligned}
$$

[14] Alexandari, Kundaje, and Shrikumar, “Maximum Likelihood with Bias-Corrected Calibration is Hard-To-Beat at Label Shift Adaptation”, 2020.

4^{th} Learning: Methods Are Numerous

Likelihood Principle:

$$\begin{aligned}\mathcal{L}(\mathbf{p} \mid \mathrm{B}) &= \mathbb{Q}(\mathrm{B} \mid \mathbf{p}) \\ &= \prod_{\mathbf{x} \in \mathrm{B}} \mathbb{Q}(\mathbf{x} \mid \mathbf{p}) \\ &\overset{\text{PPS}}{=} \prod_{\mathbf{x} \in \mathrm{B}} \sum_{y \in \mathcal{Y}} \mathbb{P}(\mathbf{x} \mid y) \cdot \mathbf{p}_y\end{aligned}$$

$$\begin{aligned}\Rightarrow \quad -\log \mathcal{L}(\mathbf{p} \mid \mathrm{B}) &= -\sum_{\mathbf{x} \in \mathrm{B}} \log \sum_{y \in \mathcal{Y}} \mathbb{P}(\mathbf{x} \mid y) \cdot \mathbf{p}_y \\ &\propto -\sum_{\mathbf{x} \in \mathrm{B}} \log \sum_{y \in \mathcal{Y}} \frac{\mathbb{P}(y|\mathbf{x})}{\mathbb{P}(y)} \cdot \mathbf{p}_y\end{aligned}$$

Maximum Likelihood Method:[14] choose $\hat{\mathbf{p}} = \arg\min_{\mathbf{p} \in \Delta^{C-1}} -\sum_{\mathbf{x} \in \mathrm{B}} \log \sum_{y \in \mathcal{Y}} \underbrace{\frac{\hat{\mathbb{P}}(y \mid \mathbf{x})}{\hat{\mathbb{P}}(y)}}_{\text{event-wise contributions}} \cdot \mathbf{p}_y$

[14] Alexandari, Kundaje, and Shrikumar, “Maximum Likelihood with Bias-Corrected Calibration is Hard-To-Beat at Label Shift Adaptation”, 2020.

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[6])

[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[6])
- 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[6])
- 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.

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[6])
- 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[6])
- 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 $\lambda(\mathrm{B})$ (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 $q(x)$.

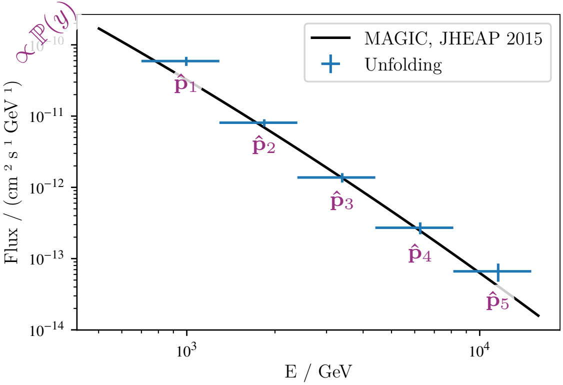
$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$

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!

[1] 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—are still to be evaluated within physics

5) Physics applications motivate further developments in quantification research