Evaluation of Trajectory Distribution Predictions with Energy Score

Novin Shahroudi, Mihkel Lepson, Meelis Kull

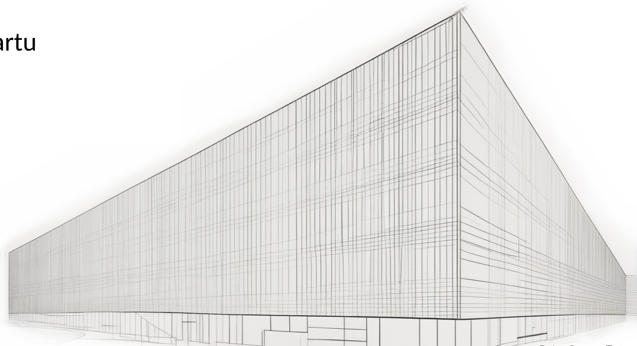
Institute of Computer Science, University of Tartu

ICML 2024

International Conference On Machine Learning

European Union European Social Fund

Investing in your future



Delta Center, Tartu

Table of Contents

- Introduction
- Methodology
- Experiments and Results
- Conclusion

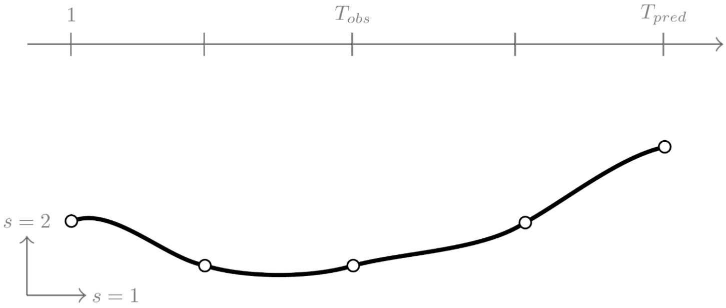
Motivation

- **Inherent uncertainty** in the movement of agents

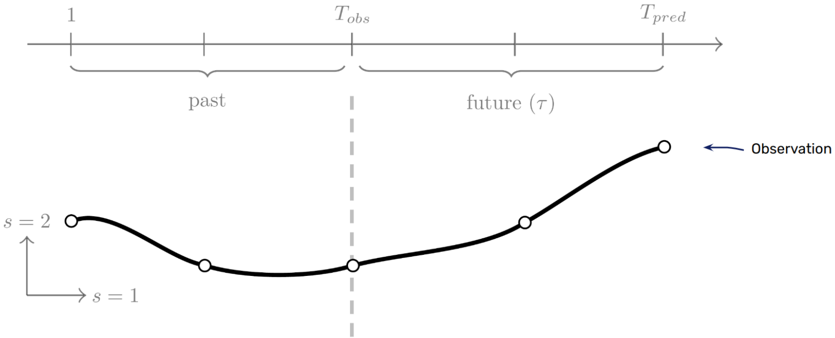
Motivation

- **Inherent uncertainty** in the movement of agents
- Estimating uncertainty is vital for **safe and reliable planning**

Trajectory

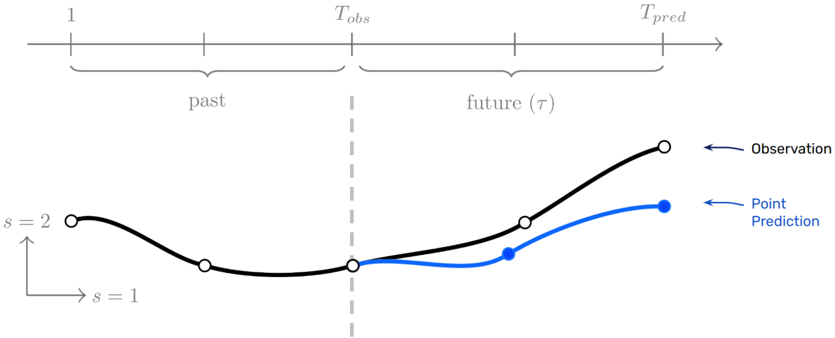


Trajectory



$$\mathbf{y}_i^{\text{past}} = (y_i^{-T_{obs}+1}, \ldots, y_i^0), \quad \mathbf{y}_i^{\tau} = (y_i^1, \ldots, y_i^T) \quad \text{where} \quad y_i^t \in \mathbb{R}^S \quad \text{and} \quad S = \{1, \mathbf{2}, \ldots\}.$$

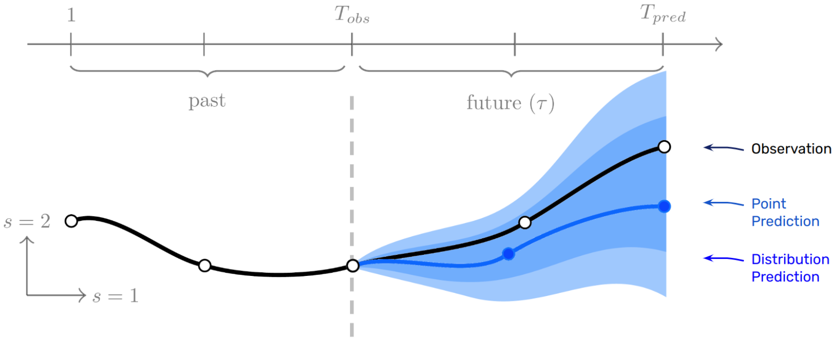
Trajectory Prediction



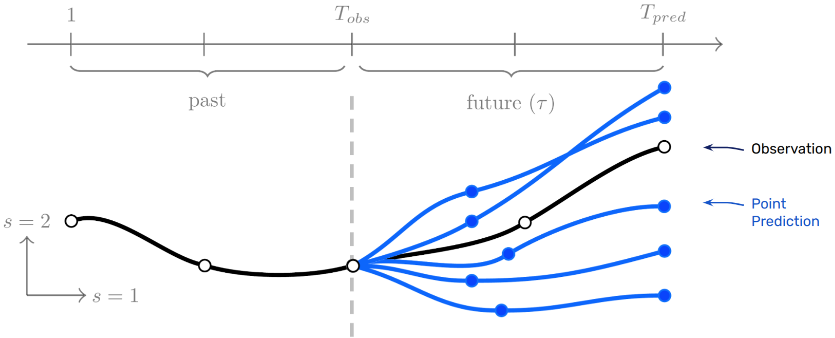
$$\mathbf{y}_i = (y_i^{t=1}, \ldots, y_i^{t=T}) \sim \mathbf{Y}_i \quad \text{where} \quad \mathbf{y}_i \in \mathbb{R}^{T \times S} \quad \text{and} \quad y_i^t \in \mathbb{R}^S$$

$$\mathbf{x}_i = (x_i^{t=1}, \ldots, x_i^{t=T}) \sim \mathbf{X}_i \quad \text{where} \quad \mathbf{x}_i \in \mathbb{R}^{T \times S} \quad \text{and} \quad x_i^t \in \mathbb{R}^S$$

Trajectory Distribution Prediction

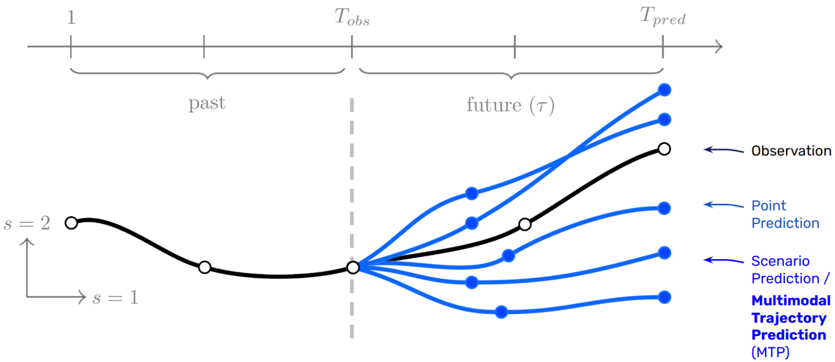


Trajectory Distribution Prediction



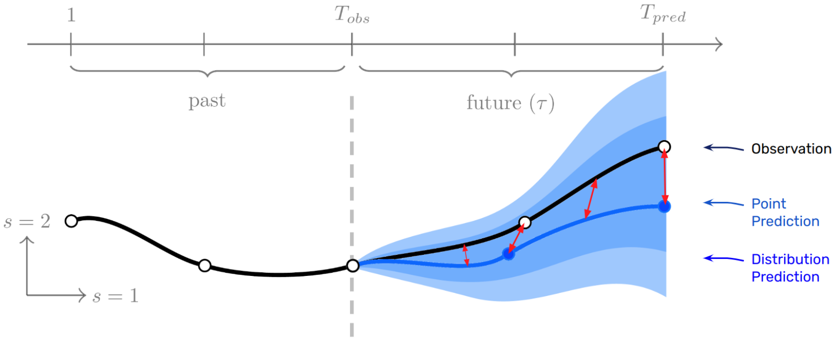
$$\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,K} \sim \mathbf{X}_i \quad \text{where} \quad \mathbf{x}_{i,k} \in \mathbb{R}^{T \times S}, \quad x^t_{i,k} \in \mathbb{R}^S$$

Trajectory Distribution Prediction



$$\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,K} \sim \mathbf{X}_i \quad \text{where} \quad \mathbf{x}_{i,k} \in \mathbb{R}^{T \times S}, \quad x_{i,k}^t \in \mathbb{R}^S$$

Trajectory Distribution Evaluation



Trajectory Distribution Evaluation

It measures the distance between the Predicted and ground truth distributions.

$$distance(\mathbf{F}_{\mathbf{X}_i}, \mathbf{F}_{\mathbf{Y}_i})$$

Where
$\mathbf{F}_{\mathbf{X}_i}$ is the CDF of the predicted.
$\mathbf{F}_{\mathbf{Y}_i}$ is the CDF of the ground truth.
i is the index of N instances in the dataset.

Variety Loss

Definition

$$L_{variety}(\mathbf{X}_i, \mathbf{y}_i) = \mathbb{E} \min_{k<K} \|\mathbf{x}_{i,k} - \mathbf{y}_i\|_2$$

Illustration adapted from Thiede, Luca Anthony, and Pratik Prabhanjan Brahma. "Analyzing the variety loss in the context of probabilistic trajectory prediction." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

Variety Loss

Definition

$$L_{variety}(\mathbf{X}_i, \mathbf{y}_i) = \mathbb{E} \min_{k<K} ||\mathbf{x}_{i,k} - \mathbf{y}_i||_2$$

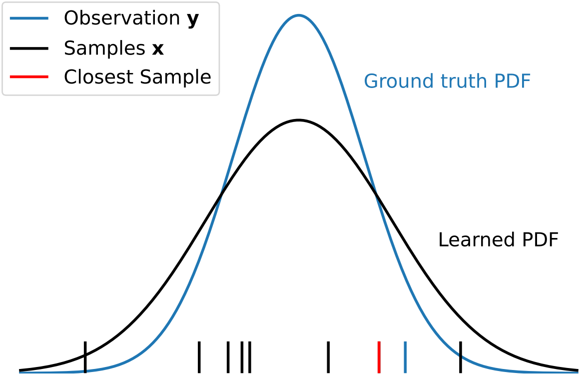


Illustration adapted from Thiede, Luca Anthony, and Pratik Prabhanjan Brahma. "Analyzing the variety loss in the context of probabilistic trajectory prediction." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

Variety Loss in the literature

- Probabilistic models are widely used, e.g., GANs, CVAEs, NFs

Variety Loss in the literature

- Probabilistic models are widely used, e.g., GANs, CVAEs, NFs
- Variety loss employed as a complementary loss

Variety Loss in the literature

- Probabilistic models are widely used, e.g., GANs, CVAEs, NFs
- Variety loss employed as a complementary loss
- Variety loss found its way as an evaluation metric (Minimum of N)

Variety Loss in the literature

- Probabilistic models are widely used, e.g., GANs, CVAEs, NFs
- Variety loss employed as a complementary loss
- Variety loss found its way as an evaluation metric (Minimum of N)
- minFDE/minADE are common instances of Minimum of N

Common Metrics for Single Trajectory Prediction

Average Displacement Error

$$ADE(\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{T} \sum_{t=1}^{T} \|x_i^t - y_i^t\|_2$$

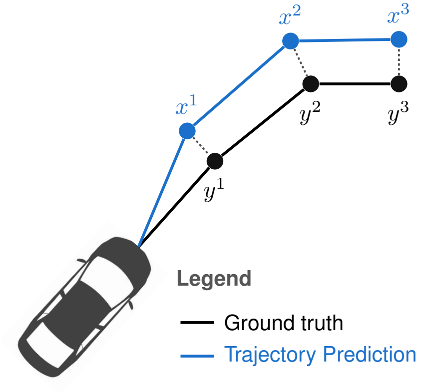


Illustration inspired by Boris, Ivanovic, and M. Pavone. "Rethinking trajectory forecasting evaluation." arXiv preprint arXiv:2107.10297 (2021).

Common Metrics for Single Trajectory Prediction

Final Displacement Error

$$FDE(\mathbf{x}_i, \mathbf{y}_i) = \|x_i^T - y_i^T\|_2$$

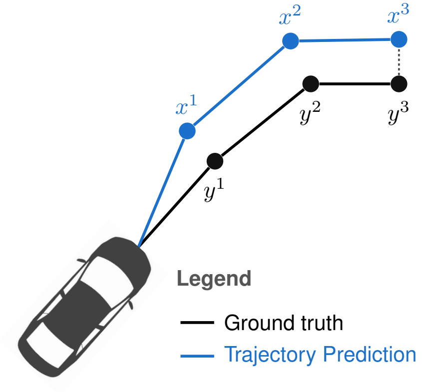


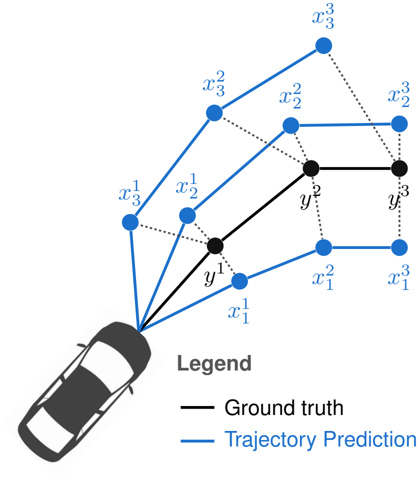
Illustration inspired by Boris, Ivanovic, and M. Pavone. "Rethinking trajectory forecasting evaluation." arXiv preprint arXiv:2107.10297 (2021).

ADE/FDE on Multimodal Trajectory Prediction (MTP)

Average Displacement Error on MTP

$$ADE(\mathbf{X}_i, \mathbf{Y}_i) = \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\|X_i^t - Y_i^t\|_2\right]$$

$$\widehat{ADE}(\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{KT}\sum_{k=1}^{K}\sum_{t=1}^{T}\|x_{i,k}^t - y_i^t\|_2$$

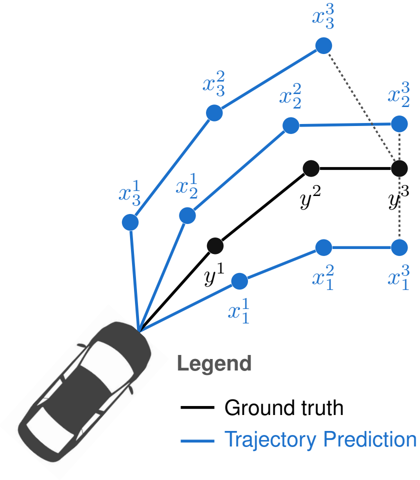


ADE/FDE on Multimodal Trajectory Prediction (MTP)

Final Displacement Error on MTP

$$FDE(\mathbf{X}_i, \mathbf{Y}_i) = \mathbb{E}\left[\|X_i^T - Y_i^T\|_2\right]$$

$$\widehat{FDE}(\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{K}\sum_{k=1}^{K} \|x_{i,k}^T - y_i^T\|_2$$

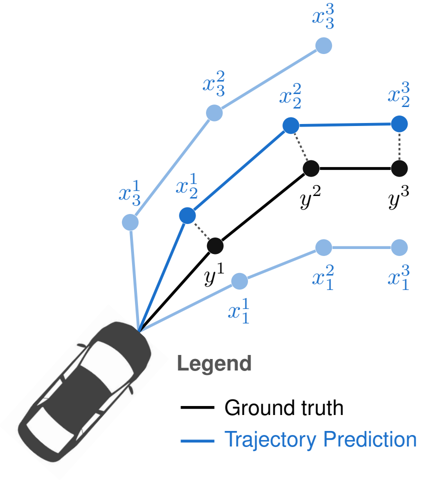


Common instances of Minimum of N (MoN)

Minimum Average Displacement Error

$$minADE(\mathbf{X}_i, \mathbf{Y}_i) = \mathbb{E}\left[\min_k \frac{1}{T}\sum_{t=1}^{T} \|X_{i,k}^t - Y_i^t\|_2\right]$$

$$\widehat{minADE}(\mathbf{x}_i, \mathbf{y}_i) = \min_k \frac{1}{T}\sum_{t=1}^{T} \|x_{i,k}^t - y_i^t\|_2$$

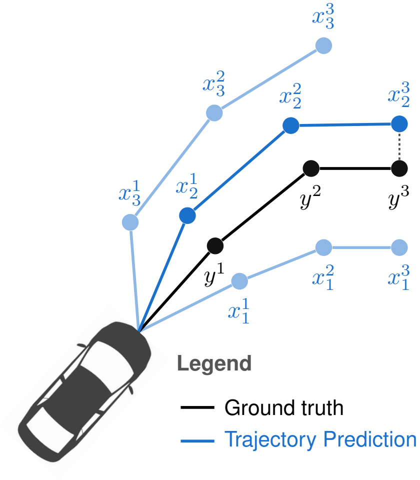


Common instances of Minimum of N (MoN)

Minimum Final Displacement Error

$$minFDE(\mathbf{X}_i, \mathbf{Y}_i) = \mathbb{E}\left[\min_k \|X_{i,k}^T - Y_i^T\|_2\right]$$

$$\widehat{minFDE}(\mathbf{x}_i, \mathbf{y}_i) = \min_k \|x_{i,k}^T - y_i^T\|_2$$



"L-lowest of N" (LoN)

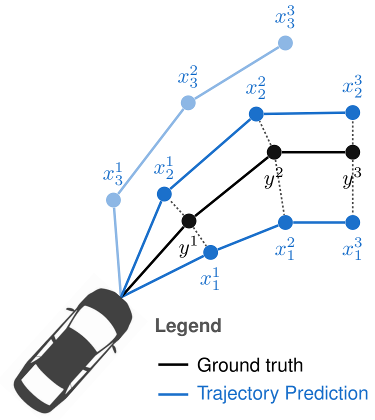
LoN as a more general form of "Minimum of N"

$$ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i) = \mathbb{E} \min_{\substack{\{k_1,\dots,k_L\} \\ k_i \neq k_j}} \frac{1}{LT} \sum_{l=1}^{L} \sum_{t=1}^{T} DE(X^t_{i,k_l}, Y^t_i)$$

$$FDE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i) = \mathbb{E} \min_{\substack{\{k_1,\dots,k_L\} \\ k_i \neq k_j}} \frac{1}{L} \sum_{l=1}^{L} DE(X^T_{i,k_l}, Y^T_i)$$

$$ADE_{(L=1)} \equiv minADE \quad , \quad FDE_{(L=1)} \equiv minFDE$$

$$ADE_{(L=K)} \equiv ADE \quad , \quad FDE_{(L=K)} \equiv FDE$$



Evaluation with MoN

- Variety loss aka ”Minimum of N” (MoN)

Evaluation with MoN

- Variety loss aka ”Minimum of N” (MoN)
- Does it identify the optimal solution, i.e., the true probability distribution?

Evaluation with MoN

- Variety loss aka "Minimum of N" (MoN)
- Does it identify the optimal solution, i.e., the true probability distribution?
- Thiede and Brahma[1] show that

[1] Thiede, Luca Anthony, and Pratik Prabhanjan Brahma. "Analyzing the variety loss in the context of probabilistic trajectory prediction." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

Evaluation with MoN

- Variety loss aka "Minimum of N" (MoN)
- Does it identify the optimal solution, i.e., the true probability distribution?
- Thiede and Brahma[1] show that

$$arg \min_{f_X(x)} L_{variety}(f_X(x), f_Y(y)) \approx \frac{\sqrt{f_Y(y)}}{C} \quad \text{when} \quad K \to \infty$$

[1] Thiede, Luca Anthony, and Pratik Prabhanjan Brahma. "Analyzing the variety loss in the context of probabilistic trajectory prediction." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

Evaluation with MoN

- Variety loss aka "Minimum of N" (MoN)
- Does it identify the optimal solution, i.e., the true probability distribution?
- Thiede and Brahma[1] show that

$$arg \min_{f_X(x)} L_{variety}(f_X(x), f_Y(y)) \approx \frac{\sqrt{f_Y(y)}}{C} \quad \text{when} \quad K \to \infty$$

$$f_{m_i}(x) = \frac{(f_Y)^{m_i}}{C_{m_i}} \quad \text{where} \quad m_2(K_2) < m_1(K_1) \quad \text{when} \quad K_2 > K_1$$

[1] Thiede, Luca Anthony, and Pratik Prabhanjan Brahma. "Analyzing the variety loss in the context of probabilistic trajectory prediction." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

Implications of using Variety Loss or MoN

- **Loss function:** less sharp (high-variance density) or too sharp (low-variance density) depending on the value of K and dimensionality of the target distribution

Implications of using Variety Loss or MoN

- **Loss function:** less sharp (high-variance density) or too sharp (low-variance density) depending on the value of K and dimensionality of the target distribution
- **Evaluation:** probabilistic calibration not respected

Implications of using Variety Loss or MoN

- **Loss function:** less sharp (high-variance density) or too sharp (low-variance density) depending on the value of K and dimensionality of the target distribution
- **Evaluation:** probabilistic calibration not respected
- **Application:** Cost induced on the prevalent or extreme events

Table of Contents

2 Methodology

▶ Introduction

▶ Methodology

▶ Experiments and Results

▶ Conclusion

Our contribution

1. Study MoN as a scoring rule

Our contribution

1. Study MoN as a scoring rule
2. Propose energy score-based evaluation

Our contribution

1. Study MoN as a scoring rule
2. Propose energy score-based evaluation
3. Different ways energy score can be employed for the evaluation

Proper Scoring Rule

Definition

A (negatively-oriented) strictly proper scoring rule $\mathbf{S}$ maps a probability distribution $\mathbf{F_X}$ and an observation $\mathbf{y}$ to a real number, i.e., $\mathbf{S}(\mathbf{F_X}, \mathbf{y}) \in \mathbb{R}$. The expected value of $\mathbf{S}(\mathbf{F_X}, .)$ under $\mathbf{F_Y}$, is written as $\mathbf{S}(\mathbf{F_X}, \mathbf{F_Y}) = \mathbb{E}_{\mathbf{y} \sim \mathbf{F_Y}}[\mathbf{S}(\mathbf{F_X}, \mathbf{y})]$.

Proper Scoring Rule

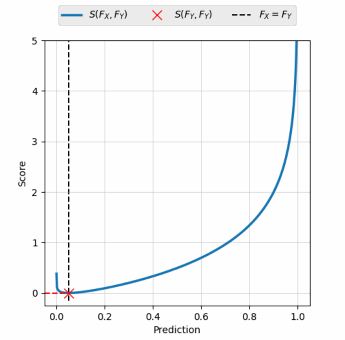
Definition

A (negatively-oriented) strictly proper scoring rule $\mathbf{S}$ maps a probability distribution $\mathbf{F_X}$ and an observation $\mathbf{y}$ to a real number, i.e., $\mathbf{S}(\mathbf{F_X}, \mathbf{y}) \in \mathbb{R}$. The expected value of $\mathbf{S}(\mathbf{F_X}, .)$ under $\mathbf{F_Y}$, is written as $\mathbf{S}(\mathbf{F_X}, \mathbf{F_Y}) = \mathbb{E}_{\mathbf{y} \sim \mathbf{F_Y}}[\mathbf{S}(\mathbf{F_X}, \mathbf{y})]$.
A scoring rule is proper if $\mathbf{S}(\mathbf{F_X}, \mathbf{F_Y}) \geq \mathbf{S}(\mathbf{F_Y}, \mathbf{F_Y})$ for all $\mathbf{F_X}$ and $\mathbf{F_Y}$, and strictly proper when the equality holds if and only if $\mathbf{F_X} = \mathbf{F_Y}$.

Proper Scoring Rule

Definition

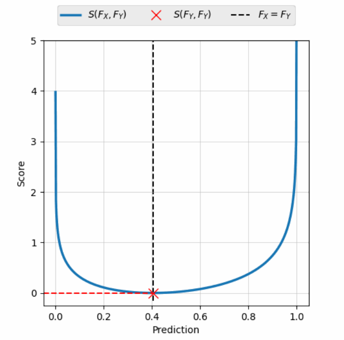
A scoring rule is proper if $\mathbf{S}(\mathbf{F_X}, \mathbf{F_Y}) \geq \mathbf{S}(\mathbf{F_Y}, \mathbf{F_Y})$ for all $\mathbf{F_X}$ and $\mathbf{F_Y}$, and strictly proper when the equality holds if and only if $\mathbf{F_X} = \mathbf{F_Y}$.



Proper Scoring Rule

Definition

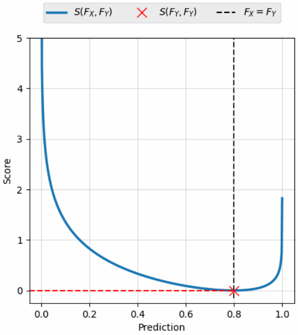
A scoring rule is proper if $\mathbf{S}(\mathbf{F_X}, \mathbf{F_Y}) \geq \mathbf{S}(\mathbf{F_Y}, \mathbf{F_Y})$ for all $\mathbf{F_X}$ and $\mathbf{F_Y}$, and strictly proper when the equality holds if and only if $\mathbf{F_X} = \mathbf{F_Y}$.



Proper Scoring Rule

Definition

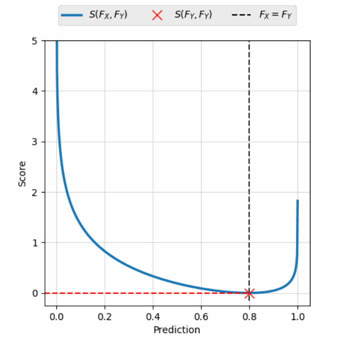
A scoring rule is proper if $\mathbf{S}(\mathbf{F_X}, \mathbf{F_Y}) \geq \mathbf{S}(\mathbf{F_Y}, \mathbf{F_Y})$ for all $\mathbf{F_X}$ and $\mathbf{F_Y}$, and strictly proper when the equality holds if and only if $\mathbf{F_X} = \mathbf{F_Y}$.

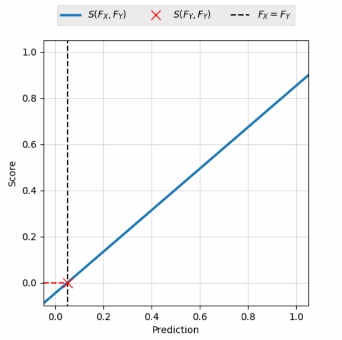


Proper Scoring Rule

Definition

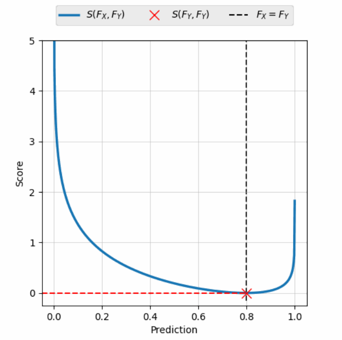
A scoring rule is proper if $\mathbf{S}(\mathbf{F_X}, \mathbf{F_Y}) \geq \mathbf{S}(\mathbf{F_Y}, \mathbf{F_Y})$ for all $\mathbf{F_X}$ and $\mathbf{F_Y}$, and strictly proper when the equality holds if and only if $\mathbf{F_X} = \mathbf{F_Y}$.

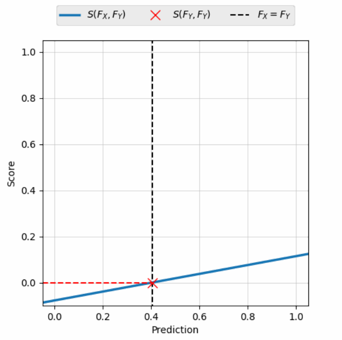




Definition

A scoring rule is proper if $\mathbf{S}(\mathbf{F_X}, \mathbf{F_Y}) \geq \mathbf{S}(\mathbf{F_Y}, \mathbf{F_Y})$ for all $\mathbf{F_X}$ and $\mathbf{F_Y}$, and strictly proper when the equality holds if and only if $\mathbf{F_X} = \mathbf{F_Y}$.

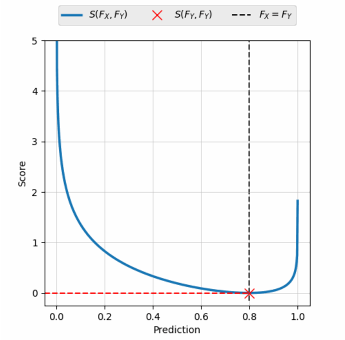


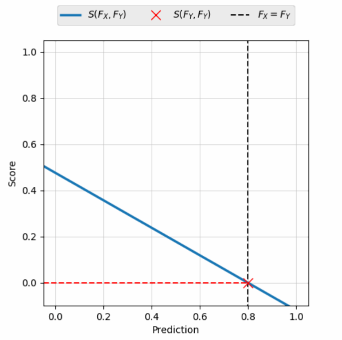


Proper Scoring Rule

Definition

A scoring rule is proper if $\mathbf{S}(\mathbf{F_X}, \mathbf{F_Y}) \geq \mathbf{S}(\mathbf{F_Y}, \mathbf{F_Y})$ for all $\mathbf{F_X}$ and $\mathbf{F_Y}$, and strictly proper when the equality holds if and only if $\mathbf{F_X} = \mathbf{F_Y}$.





MoN as a Scoring Rule

Proposition 4.1

Average Displacement Error $ADE(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper, meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$*, for which* $ADE(\mathbf{X}_i, \mathbf{Y}_i) < ADE(\mathbf{Y}_i, \mathbf{Y}_i)$.

MoN as a Scoring Rule

Proposition 4.1

Average Displacement Error $ADE(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper, meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$*, for which* $ADE(\mathbf{X}_i, \mathbf{Y}_i) < ADE(\mathbf{Y}_i, \mathbf{Y}_i)$.

Proof

For simplicity, we provide proof for $S = 1$.

MoN as a Scoring Rule

Proposition 4.1

Average Displacement Error $ADE(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper, meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$, *for which* $ADE(\mathbf{X}_i, \mathbf{Y}_i) < ADE(\mathbf{Y}_i, \mathbf{Y}_i)$.

Proof

For simplicity, we provide proof for $S = 1$. Let $\mathbf{Y} = (Y^1, \ldots, Y^T)$ be a random vector, where $Y^t \sim N(\mu_{Y^t}, \sigma^2_{Y^t})$ is distributed normally for $t = 1, \ldots, T$.

MoN as a Scoring Rule

Proposition 4.1

Average Displacement Error $ADE(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper, meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$*, for which* $ADE(\mathbf{X}_i, \mathbf{Y}_i) < ADE(\mathbf{Y}_i, \mathbf{Y}_i)$.

Proof

For simplicity, we provide proof for $S = 1$. Let $\mathbf{Y} = (Y^1, \ldots, Y^T)$ be a random vector, where $Y^t \sim N(\mu_{Y^t}, \sigma^2_{Y^t})$ is distributed normally for $t = 1, \ldots, T$. We show that there exists a random vector $\mathbf{X} = (X^1, \ldots, X^T)$ such that $ADE(\mathbf{X}, \mathbf{Y}) < ADE(\mathbf{Y}, \mathbf{Y})$.

MoN as a Scoring Rule

Proposition 4.1

Average Displacement Error $ADE(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper, meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$, *for which* $ADE(\mathbf{X}_i, \mathbf{Y}_i) < ADE(\mathbf{Y}_i, \mathbf{Y}_i)$.

Proof

For simplicity, we provide proof for $S = 1$. Let $\mathbf{Y} = (Y^1, \ldots, Y^T)$ be a random vector, where $Y^t \sim N(\mu_{Y^t}, \sigma^2_{Y^t})$ is distributed normally for $t = 1, \ldots, T$. We show that there exists a random vector $\mathbf{X} = (X^1, \ldots, X^T)$ such that $ADE(\mathbf{X}, \mathbf{Y}) < ADE(\mathbf{Y}, \mathbf{Y})$.

Choosing $X^t \sim N(\mu_{X^t}, \sigma^2_{X^t})$ such that $\mu_{X^t} = \mu_{Y^t}$ and $\sigma^2_{X^t} \to 0$, we get that

$$ADE(\mathbf{X}, \mathbf{Y}) \to \frac{1}{T} \sum_{t=1}^{T} \sqrt{\sigma^2_{Y^t}} \frac{\sqrt{2}}{\sqrt{\pi}} < \frac{1}{T} \sum_{t=1}^{T} \sqrt{\sigma^2_{Y^t} + \sigma^2_{Y^t}} \frac{\sqrt{2}}{\sqrt{\pi}} = ADE(\mathbf{Y}, \mathbf{Y}). \quad \square$$

MoN as a Scoring Rule

Proposition 4.1

Average Displacement Error $ADE(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper, meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$, *for which* $ADE(\mathbf{X}_i, \mathbf{Y}_i) < ADE(\mathbf{Y}_i, \mathbf{Y}_i)$.

Proof

For simplicity, we provide proof for $S = 1$. Let $\mathbf{Y} = (Y^1, \ldots, Y^T)$ be a random vector, where $Y^t \sim N(\mu_{Y^t}, \sigma^2_{Y^t})$ is distributed normally for $t = 1, \ldots, T$. We show that there exists a random vector $\mathbf{X} = (X^1, \ldots, X^T)$ such that $ADE(\mathbf{X}, \mathbf{Y}) < ADE(\mathbf{Y}, \mathbf{Y})$.

Choosing $X^t \sim N(\mu_{X^t}, \sigma^2_{X^t})$ such that $\mu_{X^t} = \mu_{Y^t}$ and $\sigma^2_{X^t} \to 0$, we get that

$$ADE(\mathbf{X}, \mathbf{Y}) \to \frac{1}{T}\sum_{t=1}^{T} \sqrt{\sigma^2_{Y^t}}\frac{\sqrt{2}}{\sqrt{\pi}} < \frac{1}{T}\sum_{t=1}^{T} \sqrt{\sigma^2_{Y^t} + \sigma^2_{Y^t}}\frac{\sqrt{2}}{\sqrt{\pi}} = ADE(\mathbf{Y}, \mathbf{Y}). \qquad \square$$

Because ADE is improper, FDE is improper too.

Proposition 4.2

L-lowest Average Displacement Error $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper for any values of* $L \leq 2$, *meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$, *for which* $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i) < ADE_{(L)}(\mathbf{Y}_i, \mathbf{Y}_i)$.

Proposition 4.2

L-lowest Average Displacement Error $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper for any values of* $L \leq 2$*, meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$*, for which* $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i) < ADE_{(L)}(\mathbf{Y}_i, \mathbf{Y}_i)$.

For the Proof, refer to the paper. We provide proof for $\mathbf{X} = X_1, \ldots, X_K \overset{i.i.d}{\sim} Ber(p_X)$ and a random variable $Y \sim Ber(p_Y)$ when $S = 1$ and $T = 1$.

Proposition 4.2

L-lowest Average Displacement Error $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper for any values of* $L \leq 2$, *meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$, *for which* $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i) < ADE_{(L)}(\mathbf{Y}_i, \mathbf{Y}_i)$.

For the Proof, refer to the paper. We provide proof for $\mathbf{X} = X_1, \ldots, X_K \overset{i.i.d}{\sim} Ber(p_X)$ and a random variable $Y \sim Ber(p_Y)$ when $S = 1$ and $T = 1$. Important results:

Proposition 4.2

L-lowest Average Displacement Error $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper for any values of* $L \leq 2$, *meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$, *for which* $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i) < ADE_{(L)}(\mathbf{Y}_i, \mathbf{Y}_i)$.

For the Proof, refer to the paper. We provide proof for $\mathbf{X} = X_1, \ldots, X_K \overset{i.i.d}{\sim} Ber(p_X)$ and a random variable $Y \sim Ber(p_Y)$ when $S = 1$ and $T = 1$. Important results:

- It is strictly proper only when $K = 2$.

MoN as a Scoring Rule

Proposition 4.2

L-lowest Average Displacement Error $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper for any values of* $L \leq 2$*, meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$*, for which* $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i) < ADE_{(L)}(\mathbf{Y}_i, \mathbf{Y}_i)$.

For the Proof, refer to the paper. We provide proof for $\mathbf{X} = X_1, \ldots, X_K \overset{i.i.d}{\sim} Ber(p_X)$ and a random variable $Y \sim Ber(p_Y)$ when $S = 1$ and $T = 1$. Important results:

- It is strictly proper only when $K = 2$.
- When $K = 1$ and $p_Y > 0.5$, then $p_X = 1$ gives the lowest ADE

MoN as a Scoring Rule

Proposition 4.2

L-lowest Average Displacement Error $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i)$ is improper for any values of $L \leq 2$, meaning there exist distributions $\mathbf{F}_{\mathbf{X}_i}$ and $\mathbf{F}_{\mathbf{Y}_i}$, for which $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i) < ADE_{(L)}(\mathbf{Y}_i, \mathbf{Y}_i)$.

For the Proof, refer to the paper. We provide proof for $\mathbf{X} = X_1, \ldots, X_K \overset{i.i.d}{\sim} Ber(p_X)$ and a random variable $Y \sim Ber(p_Y)$ when $S = 1$ and $T = 1$. Important results:

- It is strictly proper only when $K = 2$.
- When $K = 1$ and $p_Y > 0.5$, then $p_X = 1$ gives the lowest ADE
- When $K = 1$ and $p_Y < 0.5$, then $p_X = 0$ gives the lowest ADE

Proposition 4.2

L-lowest Average Displacement Error $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i)$ *is improper for any values of* $L \leq 2$, *meaning there exist distributions* $\mathbf{F}_{\mathbf{X}_i}$ *and* $\mathbf{F}_{\mathbf{Y}_i}$, *for which* $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i) < ADE_{(L)}(\mathbf{Y}_i, \mathbf{Y}_i)$.

For the Proof, refer to the paper. We provide proof for $\mathbf{X} = X_1, \ldots, X_K \overset{i.i.d}{\sim} Ber(p_X)$ and a random variable $Y \sim Ber(p_Y)$ when $S = 1$ and $T = 1$. Important results:

- It is strictly proper only when $K = 2$.
- When $K = 1$ and $p_Y > 0.5$, then $p_X = 1$ gives the lowest ADE
- When $K = 1$ and $p_Y < 0.5$, then $p_X = 0$ gives the lowest ADE
- When $K >= 3$ and the $p_Y \neq 0.5$, the lowest ADE is obtained by p_X value that is between p_Y and 0.5.

MoN as a Scoring Rule

Proposition 4.2

L-lowest Average Displacement Error $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i)$ is improper for any values of $L \leq 2$, meaning there exist distributions $\mathbf{F}_{\mathbf{X}_i}$ and $\mathbf{F}_{\mathbf{Y}_i}$, for which $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i) < ADE_{(L)}(\mathbf{Y}_i, \mathbf{Y}_i)$.

For the Proof, refer to the paper. We provide proof for $\mathbf{X} = X_1, \ldots, X_K \overset{i.i.d}{\sim} Ber(p_X)$ and a random variable $Y \sim Ber(p_Y)$ when $S = 1$ and $T = 1$. Important results:

- It is strictly proper only when $K = 2$.
- When $K = 1$ and $p_Y > 0.5$, then $p_X = 1$ gives the lowest ADE
- When $K = 1$ and $p_Y < 0.5$, then $p_X = 0$ gives the lowest ADE
- When $K >= 3$ and the $p_Y \neq 0.5$, the lowest ADE is obtained by p_X value that is between p_Y and 0.5.
- When $K \to \infty$, the optimal $p_X \to 0.5$.

Proposition 4.4

Let $\mathbf{X}_i \sim \mathbf{F}_{\mathbf{X}_i}$ *of length* K *and* $\mathbf{Y}_i \sim \mathbf{F}_{\mathbf{Y}_i}$. *If* $K \to \infty$, L *is fixed and supp*$(\mathbf{F}_{\mathbf{Y}_i}) \subset$ *supp*$(\mathbf{F}_{\mathbf{X}_i})$ *then* $ADE_{(L)}(\mathbf{X}_i, \mathbf{Y}_i) \to 0$.

For the Proof, refer to the paper.

Energy Score

- Energy Distance [1]

[1] Székely, G.J., Rizzo, M.L.: Energy statistics: A class of statistics based on distances. Journal of Statistical Planning and Inference. 143, 1249–1272 (2013).

Energy Score

- Energy Distance [1]
- Other related measures

[1] Székely, G.J., Rizzo, M.L.: Energy statistics: A class of statistics based on distances. Journal of Statistical Planning and Inference. 143, 1249–1272 (2013).

Energy Score

- Energy Distance [1]
- Other related measures
  - Generalization of CRPS [2]

[1] Székely, G.J., Rizzo, M.L.: Energy statistics: A class of statistics based on distances. Journal of Statistical Planning and Inference. 143, 1249–1272 (2013).

[2] Gneiting, T., Raftery, A.E.: Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. 102, 359–378 (2007).

Energy Score

- Energy Distance [1]
- Other related measures
  - Generalization of CRPS [2]
  - Permutational Analysis of Variance (PERMANOVA) [3]

[1] Székely, G.J., Rizzo, M.L.: Energy statistics: A class of statistics based on distances. Journal of Statistical Planning and Inference. 143, 1249–1272 (2013).

[2] Gneiting, T., Raftery, A.E.: Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. 102, 359–378 (2007).

[3] Anderson, M. (2001). A new method for non-parametric multivariate analysis of variance. Australecology, 26(1), 32–46.

Energy Score

- Energy Distance [1]
- Other related measures
  - Generalization of CRPS [2]
  - Permutational Analysis of Variance (PERMANOVA) [3]
  - Sinkhorn distance and Maximum Mean Discrepancy [4]

[1] Székely, G.J., Rizzo, M.L.: Energy statistics: A class of statistics based on distances. Journal of Statistical Planning and Inference. 143, 1249–1272 (2013).

[2] Gneiting, T., Raftery, A.E.: Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. 102, 359–378 (2007).

[3] Anderson, M. (2001). A new method for non-parametric multivariate analysis of variance. Australecology, 26(1), 32–46.

[4] Ramdas A, Trillos NG, Cuturi M. On Wasserstein Two-Sample Testing and Related Families of Nonparametric Tests. Entropy. 2017; 19(2):47.

Definition

$$ES(\mathbf{F}_{\mathbf{X}_i}, \mathbf{y}_i) = \overbrace{\mathbb{E}\|\mathbf{X}_i - \mathbf{y}_i\|_p^\beta}^{ED} - \frac{1}{2}\overbrace{\mathbb{E}\|\mathbf{X}_i - \tilde{\mathbf{X}}_i\|_p^\beta}^{EI} \quad (1)$$

$$\overline{ES} = \frac{1}{N}\sum_{i=1}^{N} ES(\mathbf{F}_{\mathbf{X}_i}, \mathbf{y}_i) \quad (2)$$

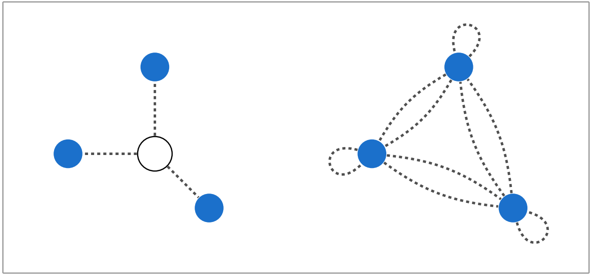
$$\widehat{ES} = \frac{1}{K}\sum_{k=1}^{K}\|x_{i,k} - y_i\|_2 - \frac{1}{2 \cdot K^2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\|x_{i,k} - x_{i,k'}\|_2 \quad \text{where} \quad \beta = 1, \quad p = 2$$

Energy Score Intuition

$$\widehat{ES} = \frac{1}{K}\sum_{k=1}^{K} \|x_{i,k} - y_i\|_2 - \frac{1}{2 \cdot K^2}\sum_{k=1}^{K}\sum_{k'=1}^{K} \|x_{i,k} - x_{i,k'}\|_2$$

Legend

- Observation y
- Prediction x

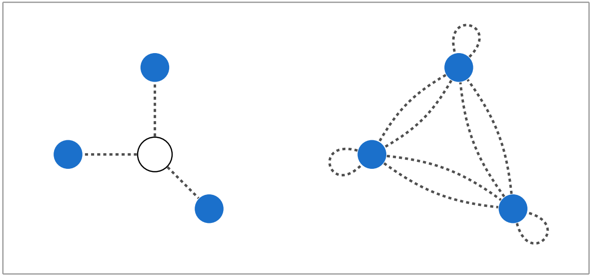


Energy Score Intuition

$$\widehat{ES} = \frac{1}{3}\sum_{k=1}^{3} \|x_{i,k} - y_i\|_2 - \frac{1}{2 \cdot 3^2}\sum_{k=1}^{3}\sum_{k'=1}^{3} \|x_{i,k} - x_{i,k'}\|_2$$

Legend

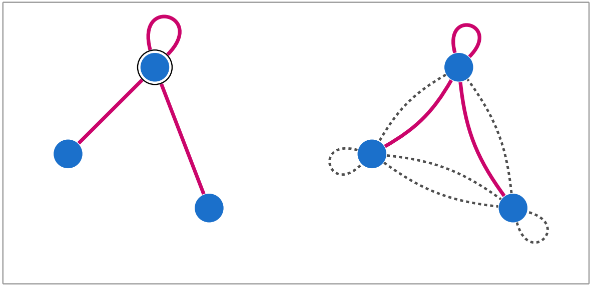
- Observation y
- Prediction x



Energy Score Intuition

$$\widehat{ES} = \frac{1}{3}\sum_{k=1}^{3} \|x_{i,k} - y_i\|_2 - \frac{1}{2 \cdot 3^2}\sum_{k=1}^{3}\sum_{k'=1}^{3} \|x_{i,k} - x_{i,k'}\|_2$$

$i = 1$



Legend

- Observation y
- Prediction x

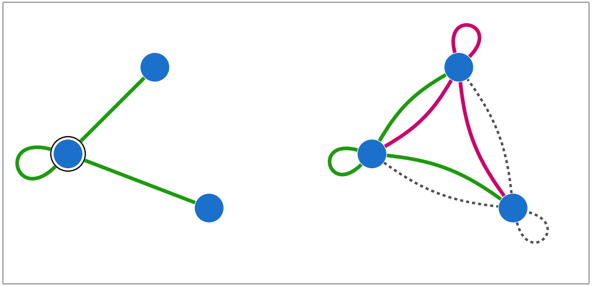
Energy Score Intuition

$$\widehat{ES} = \frac{1}{3}\sum_{k=1}^{3} \|x_{i,k} - y_i\|_2 - \frac{1}{2 \cdot 3^2}\sum_{k=1}^{3}\sum_{k'=1}^{3} \|x_{i,k} - x_{i,k'}\|_2$$

$i = 2$

Legend

- Observation y
- Prediction x



Energy Score Intuition

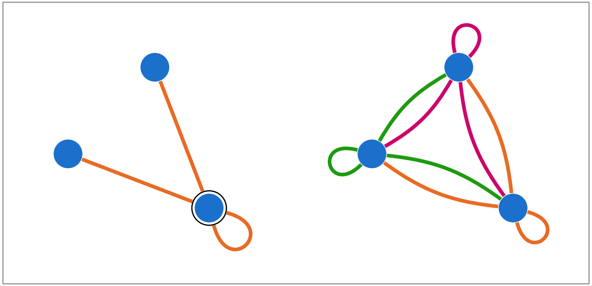
$$\widehat{ES} = \frac{1}{3}\sum_{k=1}^{3} \|x_{i,k} - y_i\|_2 - \frac{1}{2\cdot 3^2}\sum_{k=1}^{3}\sum_{k'=1}^{3} \|x_{i,k} - x_{i,k'}\|_2$$

$i = 3$

Legend

Observation y

Prediction x



Energy Score Intuition

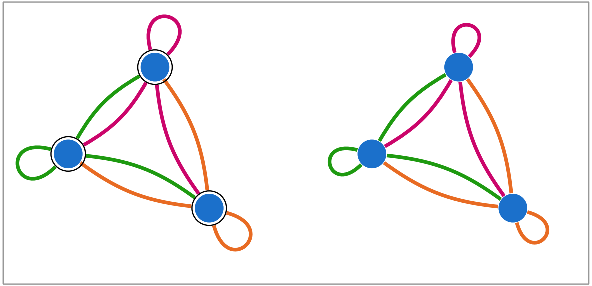
$$\widehat{ES} = \frac{1}{3}\sum_{k=1}^{3} \|x_{i,k} - y_i\|_2 - \frac{1}{2\cdot 3^2}\sum_{k=1}^{3}\sum_{k'=1}^{3} \|x_{i,k} - x_{i,k'}\|_2$$

$$i = \{1, 2, 3\}$$

Legend

Observation y

Prediction x



Energy Score Intuition

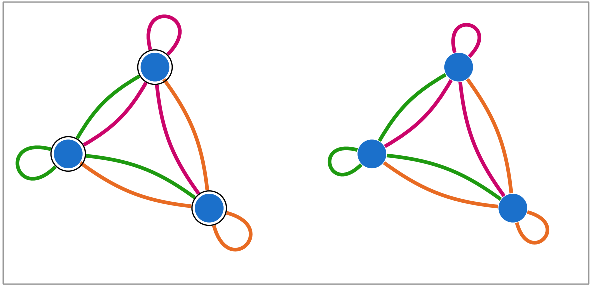
$$\widehat{ES} = \frac{1}{3}\sum_{i=1}^{3}\frac{1}{3}\sum_{k=1}^{3}\|x_{i,k} - y_i\|_2 - \frac{1}{3}\sum_{i=1}^{3}\frac{1}{2\cdot 3^2}\sum_{k=1}^{3}\sum_{k'=1}^{3}\|x_{i,k} - x_{i,k'}\|_2$$

$$i = \{1, 2, 3\}$$

Legend

Observation y

Prediction x



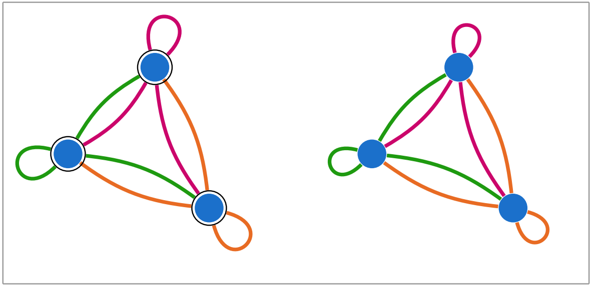
Energy Score Intuition

$$\widehat{ES} = \frac{1}{3^2}\sum_{i=1}^{3}\sum_{k=1}^{3}\|x_{i,k} - y_i\|_2 - \frac{1}{2}\frac{1}{3^2}\sum_{k=1}^{3}\sum_{k'=1}^{3}\|x_{i,k} - x_{i,k'}\|_2$$

$$i = \{1, 2, 3\}$$

Legend

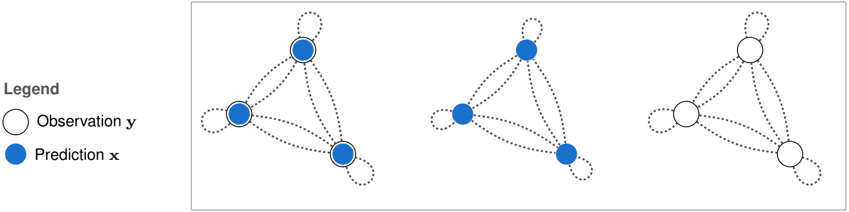
- Observation y
- Prediction x



Energy Score Intuition

$$ED(F_X, F_Y) = \mathbb{E}(\|X - Y\|_p^{\beta}) - \frac{1}{2}\mathbb{E}(\|X - \tilde{X}\|_p^{\beta}) - \frac{1}{2}\mathbb{E}(\|Y - \tilde{Y}\|_p^{\beta})$$

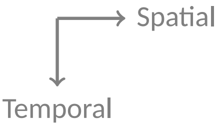
$$\widehat{ED} = \frac{1}{3^2}\sum_{i=1}^{3}\sum_{k=1}^{3}\|x_{i,k} - y_{i,k}\|_2 - \frac{1}{2}\frac{1}{3^2}\sum_{k=1}^{3}\sum_{k'=1}^{3} 3^2 \|x_{i,k} - x_{i,k'}\|_2 - \frac{1}{2}\frac{1}{3^2}\sum_{k=1}^{3}\sum_{k'=1}^{3} 3^2 \|y_{i,k} - y_{i,k'}\|_2$$



Székely, G.J., Rizzo, M.L.: Energy statistics: A class of statistics based on distances. Journal of Statistical Planning and Inference. 143, 1249–1272 (2013).

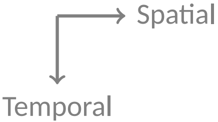
Energy Score for MTP Evaluation

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$



Energy Score for MTP Evaluation

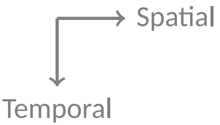
$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$



$$dim(\{\mathbf{x}_{i,k}\}_{k=1}^{K}) = K \times T \times S \qquad dim(\mathbf{y}_i) = 1 \times T \times S$$

Energy Score for MTP Evaluation

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$

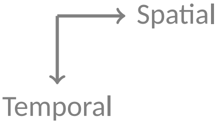


$$dim(\{\mathbf{x}_{i,k}\}_{k=1}^{K}) = K \times T \times S \qquad dim(\mathbf{y}_i) = 1 \times T \times S$$

- Entry-wise (jointly on spatial and temporal)

Energy Score for MTP Evaluation

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$

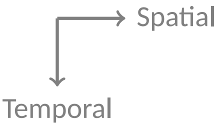


$$dim(\{\mathbf{x}_{i,k}\}_{k=1}^{K}) = K \times T \times S \qquad dim(\mathbf{y}_i) = 1 \times T \times S$$

- Entry-wise (jointly on spatial and temporal)
- Column-wise (marginalized on Spatial)

Energy Score for MTP Evaluation

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$



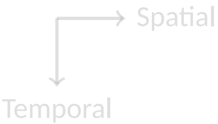
$$dim(\{\mathbf{x}_{i,k}\}_{k=1}^{K}) = K \times T \times S \qquad dim(\mathbf{y}_i) = 1 \times T \times S$$

- Entry-wise (jointly on spatial and temporal)
- Column-wise (marginalized on Spatial)
- Row-wise (marginalized on Temporal)

  remark: common metrics such as minADE are typically temporally marginalized

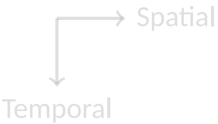
Entry-wise

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$



Entry-wise

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$

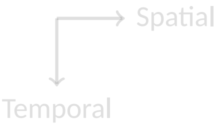


$$ES(\mathbf{F}_{\mathbf{X}_i}, \mathbf{y}_i) = \mathbb{E}_k\left(\|\mathbf{X} - \mathbf{y}\|_p^\beta\right) - \frac{1}{2}\mathbb{E}_k\left(\|\mathbf{X} - \tilde{\mathbf{X}}\|_p^\beta\right)$$

akin to Frobenius distance for $p = 2$.

Entry-wise

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$



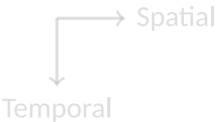
$$ES(\mathbf{F}_{\mathbf{x}_i}, \mathbf{y}_i) = \mathbb{E}_k\left(\|\mathbf{X} - \mathbf{y}\|_p^\beta\right) - \frac{1}{2}\mathbb{E}_k\left(\|\mathbf{X} - \tilde{\mathbf{X}}\|_p^\beta\right)$$

$$\widehat{ES}(\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{K}\sum_{k=1}^{K}\left(\sum_{t=1}^{T}\sum_{s=1}^{S}|x_k^{t,s} - y^{t,s}|^p\right)^{\beta/p} - \frac{1}{2}\frac{1}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\sum_{t=1}^{T}\sum_{s=1}^{S}|x_k^{t,s} - \tilde{x}_l^{t,s}|^p\right)^{\beta/p}$$

akin to Frobenius distance for $p = 2$.

Entry-wise

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \cdots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \cdots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \cdots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \cdots & y^{TS} \end{bmatrix}$$



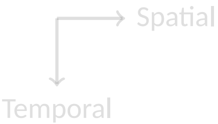
$$ES(\mathbf{F}_{\mathbf{X}_i}, \mathbf{y}_i) = \mathbb{E}_k\left(\|\mathbf{X} - \mathbf{y}\|_p^\beta\right) - \frac{1}{2}\mathbb{E}_k\left(\|\mathbf{X} - \tilde{\mathbf{X}}\|_p^\beta\right)$$

$$\widehat{ES}(\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{K}\sum_{k=1}^{K}\left(\sum_{t=1}^{T}\sum_{s=1}^{S}|x_k^{t,s} - y^{t,s}|^p\right)^{\beta/p} - \frac{1}{2}\frac{1}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\sum_{t=1}^{T}\sum_{s=1}^{S}|x_k^{t,s} - \tilde{x}_l^{t,s}|^p\right)^{\beta/p}$$

akin to Frobenius distance for $p = 2$.

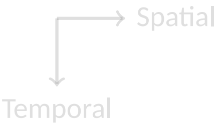
Column-wise

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$



Column-wise

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$

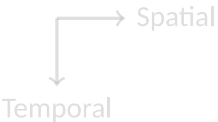


$$EST(\mathbf{F}_{\mathbf{X}_i}, \mathbf{y}_i) = \mathbb{E}_{k,s}\left(\|\mathbf{X} - \mathbf{y}\|_p^\beta\right) - \frac{1}{2}\mathbb{E}_{k,s}\left(\|\mathbf{X} - \tilde{\mathbf{X}}\|_p^\beta\right)$$

akin to Minkowski column distance.

Column-wise

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$



$$EST(\mathbf{F}_{\mathbf{X}_i}, \mathbf{y}_i) = \mathbb{E}_{k,s}\left(\|\mathbf{X} - \mathbf{y}\|_p^\beta\right) - \frac{1}{2}\mathbb{E}_{k,s}\left(\|\mathbf{X} - \tilde{\mathbf{X}}\|_p^\beta\right)$$

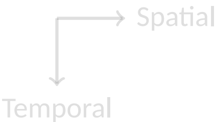
$$\widehat{EST}(\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{S}\sum_{s=1}^{S}\left(\sum_{t=1}^{T}|x_k^{t,s} - y^{t,s}|^p\right)^{\beta/p} - \frac{1}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\frac{1}{S}\sum_{s=1}^{S}\left(\sum_{t=1}^{T}|x_k^{t,s} - \tilde{x}_l^{t,s}|^p\right)^{\beta/p}$$

akin to Minkowski column distance.

Column-wise

2 Methodology

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$

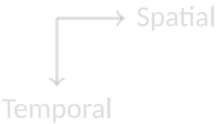


$$EST(\mathbf{F}_{\mathbf{x}_i}, \mathbf{y}_i) = \mathbb{E}_{k,s}\left(\|\mathbf{X} - \mathbf{y}\|_p^{\beta}\right) - \frac{1}{2}\mathbb{E}_{k,s}\left(\|\mathbf{X} - \tilde{\mathbf{X}}\|_p^{\beta}\right)$$

$$\widehat{EST}(\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{S}\sum_{s=1}^{S}\left(\sum_{t=1}^{T}|x_k^{t,s} - y^{t,s}|^p\right)^{\beta/p} - \frac{1}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\frac{1}{S}\sum_{s=1}^{S}\left(\sum_{t=1}^{T}|x_k^{t,s} - \tilde{x}_l^{t,s}|^p\right)^{\beta/p}$$

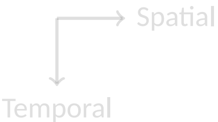
akin to Minkowski column distance.

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$



Row-wise

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$



$$ESS(\mathbf{F}_{\mathbf{X}_i}, \mathbf{y}_i) = \mathbb{E}_{k,t}\left(\|\mathbf{X} - \mathbf{y}\|_p^\beta\right) - \frac{1}{2}\mathbb{E}_{k,t}\left(\|\mathbf{X} - \tilde{\mathbf{X}}\|_p^\beta\right)$$

akin to Minkowski row distance.

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$

Spatial

Temporal

$$ESS(\mathbf{F}_{\mathbf{X}_i}, \mathbf{y}_i) = \mathbb{E}_{k,t}\left(\|\mathbf{X} - \mathbf{y}\|_p^\beta\right) - \frac{1}{2}\mathbb{E}_{k,t}\left(\|\mathbf{X} - \tilde{\mathbf{X}}\|_p^\beta\right)$$

$$\widehat{ESS}(\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{S}|x_k^{t,s} - y^{t,s}|^p\right)^{\beta/p} - \frac{1}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{S}|x_k^{t,s} - \tilde{x}_l^{t,s}|^p\right)^{\beta/p}$$

akin to Minkowski row distance.

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \dots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \dots & x_k^{TS} \end{bmatrix} \qquad \mathbf{y}_i = \begin{bmatrix} y^{11} & \dots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \dots & y^{TS} \end{bmatrix}$$

Spatial

Temporal

$$ESS(\mathbf{F}_{\mathbf{X}_i}, \mathbf{y}_i) = \mathbb{E}_{k,t}\left(\|\mathbf{X} - \mathbf{y}\|_p^\beta\right) - \frac{1}{2}\mathbb{E}_{k,t}\left(\|\mathbf{X} - \tilde{\mathbf{X}}\|_p^\beta\right)$$

$$\widehat{ESS}(\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{S}|x_k^{t,s} - y^{t,s}|^p\right)^{\beta/p} - \frac{1}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{S}|x_k^{t,s} - \tilde{x}_l^{t,s}|^p\right)^{\beta/p}$$

akin to Minkowski row distance.

The three variations

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \cdots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \cdots & x_k^{TS} \end{bmatrix} \quad \mathbf{y}_i = \begin{bmatrix} y^{11} & \cdots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \cdots & y^{TS} \end{bmatrix}$$

ES Spatio-temporal (ES)

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \cdots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \cdots & x_k^{TS} \end{bmatrix} \quad \mathbf{y}_i = \begin{bmatrix} y^{11} & \cdots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \cdots & y^{TS} \end{bmatrix}$$

ES Temporal (EST)

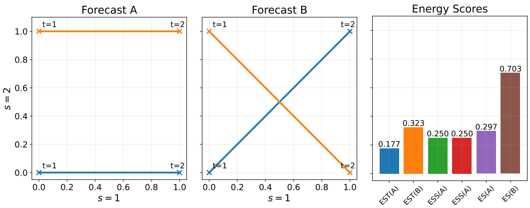
Averaged over the **spatial** dimension!

$$\mathbf{x}_{i,k} = \begin{bmatrix} x_k^{11} & \cdots & x_k^{1S} \\ \vdots & \ddots & \vdots \\ x_k^{T1} & \cdots & x_k^{TS} \end{bmatrix} \quad \mathbf{y}_i = \begin{bmatrix} y^{11} & \cdots & y^{1S} \\ \vdots & \ddots & \vdots \\ y^{T1} & \cdots & y^{TS} \end{bmatrix}$$

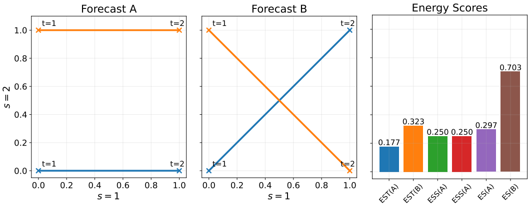
ES Spatial (ESS)

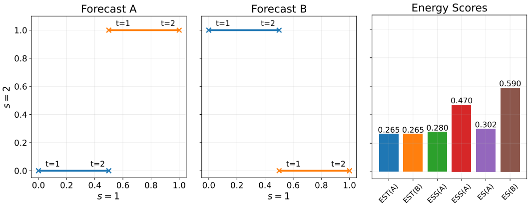
Averaged over the **temporal** dimension!

Toy Example Demonstration



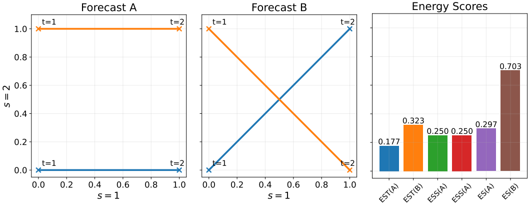
Toy Example Demonstration

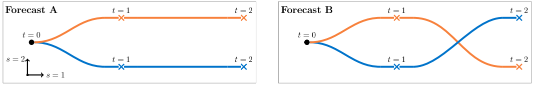


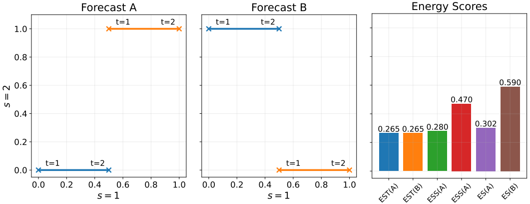


Toy Example Demonstration

2 Methodology







Toy Example Demonstration

2 Methodology

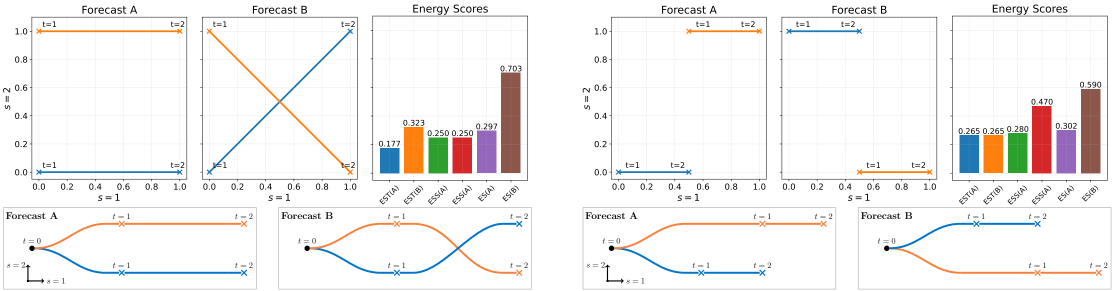


Table of Contents

▶ Introduction

▶ Methodology

▶ Experiments and Results

▶ Conclusion

Autoregressive Process

$$y_{i,k}^{t} = y_{i,k}^{t-1} + \mathcal{N}(\mu^{t} + a^{t}, (\sigma^{t} + b^{t})^{2})$$

where $t \in \{1, 2, 3\}, \quad i = [0, N), \quad k = [0, K), \quad y^{0} = 0$

Autoregressive Process

$$y_{i,k}^{t} = y_{i,k}^{t-1} + \mathcal{N}(\mu^{t} + a^{t}, (\sigma^{t} + b^{t})^{2})$$

where $t \in \{1, 2, 3\}, \quad i = [0, N), \quad k = [0, K), \quad y^{0} = 0$

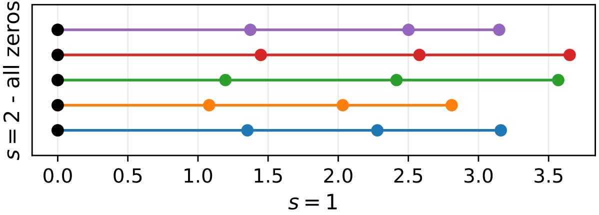
- $\mu^{t} = 1$, $\sigma^{t} = 0.2$, $a^{t} = 0$, and $b^{t} = 0$ for $t = \{1, 2, 3\}$.
- Generate $N = 5000$ observations and consider $K = \{10, 20, 50, 100, 300\}$.

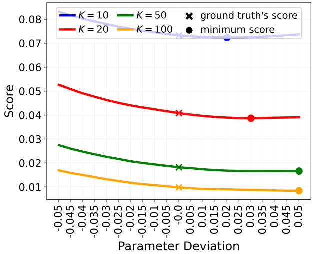
Autoregressive Process

$$y_{i,k}^{t} = y_{i,k}^{t-1} + \mathcal{N}(\mu^t + a^t, (\sigma^t + b^t)^2)$$

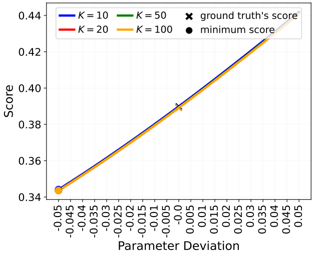
where $t \in \{1, 2, 3\}, \quad i = [0, N), \quad k = [0, K), \quad y^0 = 0$

- $\mu^t = 1$, $\sigma^t = 0.2$, $a^t = 0$, and $b^t = 0$ for $t = \{1, 2, 3\}$.
- Generate $N = 5000$ observations and consider $K = \{10, 20, 50, 100, 300\}$.

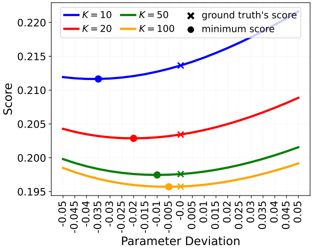




$FDE_{(L=1)}$



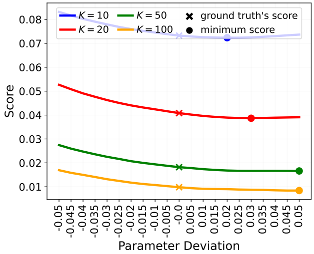
$FDE_{(L=K)}$



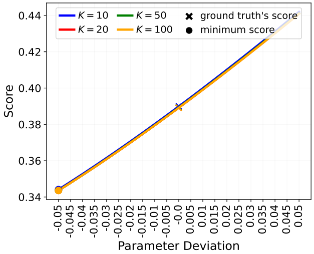
FES

- X-axis: predictions with different deviations.

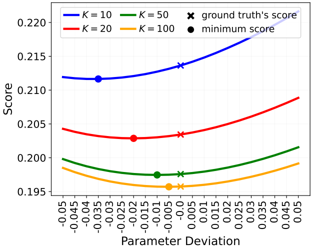
Propriety Showcase



$FDE_{(L=1)}$



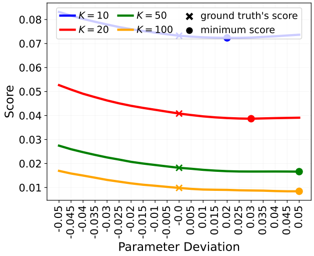
$FDE_{(L=K)}$



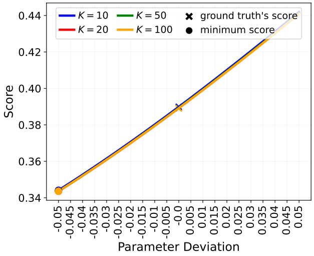
FES

- X-axis: predictions with different deviations.
- A strictly proper metric gets minimized at the optimal parameter ($deviation = 0$).

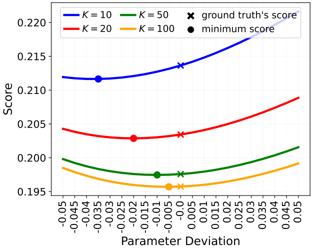
Propriety Showcase



$FDE_{(L=1)}$

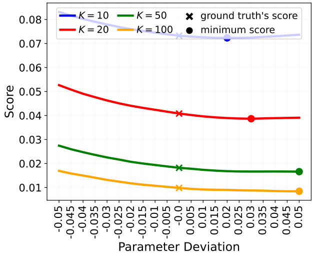


$FDE_{(L=K)}$

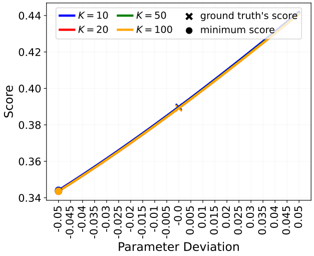


FES

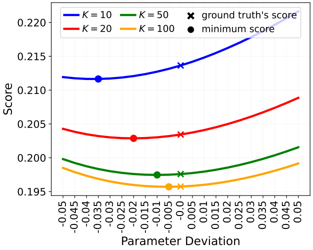
- X-axis: predictions with different deviations.
- A strictly proper metric gets minimized at the optimal parameter ($deviation = 0$).
- As a reminder: $FDE_{(L=1)} \equiv minFDE$ and $FDE_{(L=K)} \equiv FDE$.



$FDE_{(L=1)}$



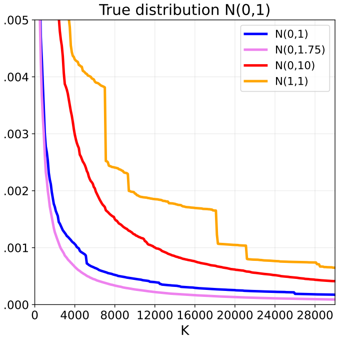
$FDE_{(L=K)}$



FES

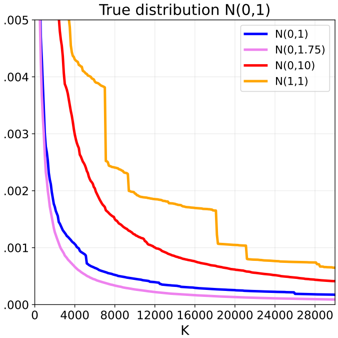
Results support Prepositions 4.1 and 4.2

Effect of Sample Size

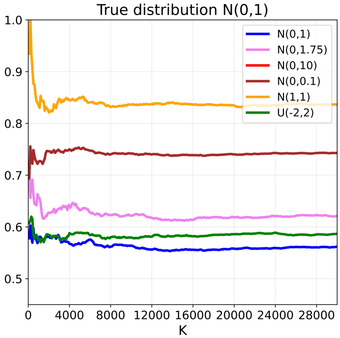


$$FDE_{(L=1)}$$

Effect of Sample Size

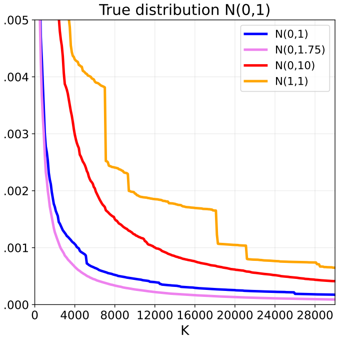


$FDE_{(L=1)}$

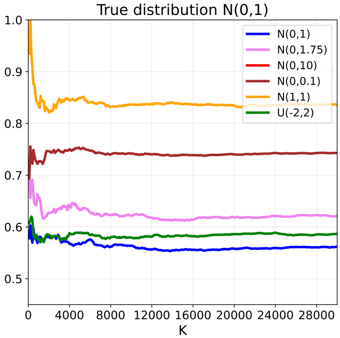


FES

Effect of Sample Size



$FDE_{(L=1)}$



FES

Results support Proposition 4.4

ETH/UCY Dataset

- Pretrained models from Bae et. al [5] on ETH/UCY human trajectory datasets.

[5] Bae, I., Park, J. H., & Jeon, H. G. (2022). Non-probability sampling network for stochastic human trajectory prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6477-6487).

ETH/UCY Dataset

- Pretrained models from Bae et. al [5] on ETH/UCY human trajectory datasets.
- Three generative models: STGCNN, PECNET, and SGCN.

[5] Bae, I., Park, J. H., & Jeon, H. G. (2022). Non-probability sampling network for stochastic human trajectory prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6477-6487).

ETH/UCY Dataset

- Pretrained models from Bae et. al [5] on ETH/UCY human trajectory datasets.
- Three generative models: STGCNN, PECNET, and SGCN.
- Three sampling methods Monte Carlo (MC), Quasi-Monte Carlo (QMC), Non-Probability Sampling Network (NPSN)

[5] Bae, I., Park, J. H., & Jeon, H. G. (2022). Non-probability sampling network for stochastic human trajectory prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6477-6487).

ETH/UCY Dataset

- Pretrained models from Bae et. al [5] on ETH/UCY human trajectory datasets.
- Three generative models: STGCNN, PECNET, and SGCN.
- Three sampling methods Monte Carlo (MC), Quasi-Monte Carlo (QMC), Non-Probability Sampling Network (NPSN)
- Combination of the three models and sampling methods were considered.

[5] Bae, I., Park, J. H., & Jeon, H. G. (2022). Non-probability sampling network for stochastic human trajectory prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6477-6487).

ETH/UCY Dataset

- Pretrained models from Bae et. al [5] on ETH/UCY human trajectory datasets.
- Three generative models: STGCNN, PECNET, and SGCN.
- Three sampling methods Monte Carlo (MC), Quasi-Monte Carlo (QMC), Non-Probability Sampling Network (NPSN)
- Combination of the three models and sampling methods were considered.
- **Goal:** how Energy Score ranks differently than its MoN counterpart?

[5] Bae, I., Park, J. H., & Jeon, H. G. (2022). Non-probability sampling network for stochastic human trajectory prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6477-6487).

ETH/UCY Dataset

Reported values: expected $minADE/ES$. AVG is the arithmetic average over all datasets. **Bold**: best model, underline: second best model. Baselines: *-MC.

	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
stgcnn-mc	0.65/1.44	0.50/1.05	0.44/0.96	0.34/0.77	0.30/0.67	0.45/0.98
pecnet-mc	0.61/1.64	0.22/0.70	0.33/0.89	0.25/0.74	0.19/0.65	0.32/0.92
sgcn-mc	0.57/1.34	0.31/0.73	0.37/0.85	0.29/0.68	0.22/0.53	0.35/<u>0.82</u>
stgcnn-qmc	0.61/<u>1.30</u>	0.34/0.98	0.36/0.89	0.32/0.74	0.29/0.65	0.38/0.91
pecnet-qmc	0.60/1.62	0.21/0.68	0.33/0.88	0.24/0.72	0.18/0.62	0.31/0.91
sgcn-qmc	0.49/**1.23**	0.21/0.66	0.31/**0.78**	0.25/**0.63**	0.19/**0.49**	0.29/**0.76**
stgcnn-npsn	<u>0.44</u>/1.48	0.21/0.88	<u>0.28</u>/0.88	0.25/0.83	0.22/0.73	<u>0.28</u>/0.96
pecnet-npsn	0.55/1.60	<u>0.19</u>/<u>0.63</u>	0.29/0.88	<u>0.21</u>/0.70	<u>0.16</u>/0.56	<u>0.28</u>/0.87
sgcn-npsn	**0.36/1.23**	**0.16/0.62**	**0.23**/<u>0.79</u>	**0.18**/<u>0.66</u>	**0.14**/<u>0.50</u>	**0.21/0.76**

ETH/UCY Dataset

Reported values: expected $minADE/ES$. AVG is the arithmetic average over all datasets. **Bold**: best model, underline: second best model. Baselines: *-MC.

	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
stgcnn-mc	0.65/1.44	0.50/1.05	0.44/0.96	0.34/0.77	0.30/0.67	0.45/0.98
pecnet-mc	0.61/1.64	0.22/0.70	0.33/0.89	0.25/0.74	0.19/0.65	0.32/0.92
sgcn-mc	0.57/1.34	0.31/0.73	0.37/0.85	0.29/0.68	0.22/0.53	0.35/0.82
stgcnn-qmc	0.61/1.30	0.34/0.98	0.36/0.89	0.32/0.74	0.29/0.65	0.38/0.91
pecnet-qmc	0.60/1.62	0.21/0.68	0.33/0.88	0.24/0.72	0.18/0.62	0.31/0.91
sgcn-qmc	0.49/**1.23**	0.21/0.66	0.31/**0.78**	0.25/**0.63**	0.19/**0.49**	0.29/**0.76**
stgcnn-npsn	0.44/1.48	0.21/0.88	0.28/0.88	0.25/0.83	0.22/0.73	0.28/0.96
pecnet-npsn	0.55/1.60	0.19/0.63	0.29/0.88	0.21/0.70	0.16/0.56	0.28/0.87
sgcn-npsn	**0.36/1.23**	**0.16/0.62**	**0.23**/0.79	**0.18**/0.66	**0.14**/0.50	**0.21/0.76**

ES favors SGCN-QMC as the best model over SGCN-NPSN on 3 out of 5 datasets, in contrast to $minADE$.

Table of Contents

- Introduction
- Methodology
- Experiments and Results
- Conclusion

Summary

- Adopted **Scoring Rules framework** for evaluation of trajectory distribution predictions.

Summary

4 Conclusion

- Adopted **Scoring Rules framework** for evaluation of trajectory distribution predictions.
- Examined $minADE/minFDE$ as common instances of the **Minimum of N** family of evaluation metrics and showed that they are **improper**.

Summary

- Adopted **Scoring Rules framework** for evaluation of trajectory distribution predictions.
- Examined $minADE/minFDE$ as common instances of the **Minimum of N** family of evaluation metrics and showed that they are **improper**.
- Proposed **Energy Score-based evaluation** as an alternative.

Summary

- Adopted **Scoring Rules framework** for evaluation of trajectory distribution predictions.
- Examined $minADE/minFDE$ as common instances of the **Minimum of N** family of evaluation metrics and showed that they are **improper**.
- Proposed **Energy Score-based evaluation** as an alternative.
- Demonstrated how **marginalized variations of Energy Score** can be useful for diagnosis of trajectory distribution predictions.

Limitations of Energy Score

- Energy Score calculation has $\mathcal{O}(K^2)$ computational complexity.

Limitations of Energy Score

- Energy Score calculation has $\mathcal{O}(K^2)$ computational complexity.
- Behavior of energy score for small values of K merits further investigation.

Evaluation of Trajectory Distribution Predictions with Energy Score

Thank you for listening!

Any questions?