Diffusion and Flow-based Copulas: Forgetting and Remembering Dependencies

David Huk and Theodoros Damoulas

UNIVERSITY OF WARWICK

Warwick Statistics

Copula models of dependence

In the case of independent data x_1, x_2 we have that their joint distribution factorises as:

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$$

Copula models of dependence

In the case of independent data x_1, x_2 we have that their joint distribution factorises as:

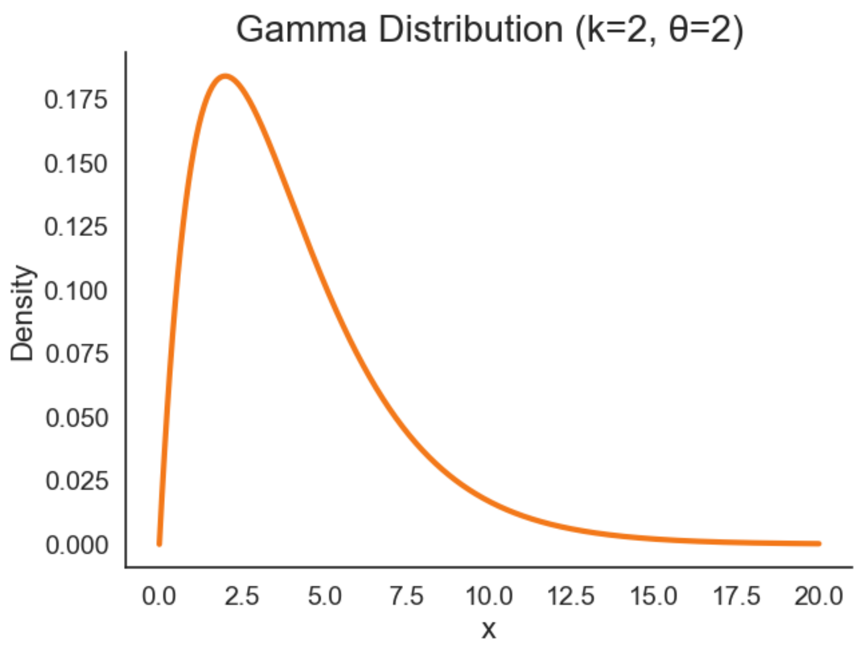
$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$$

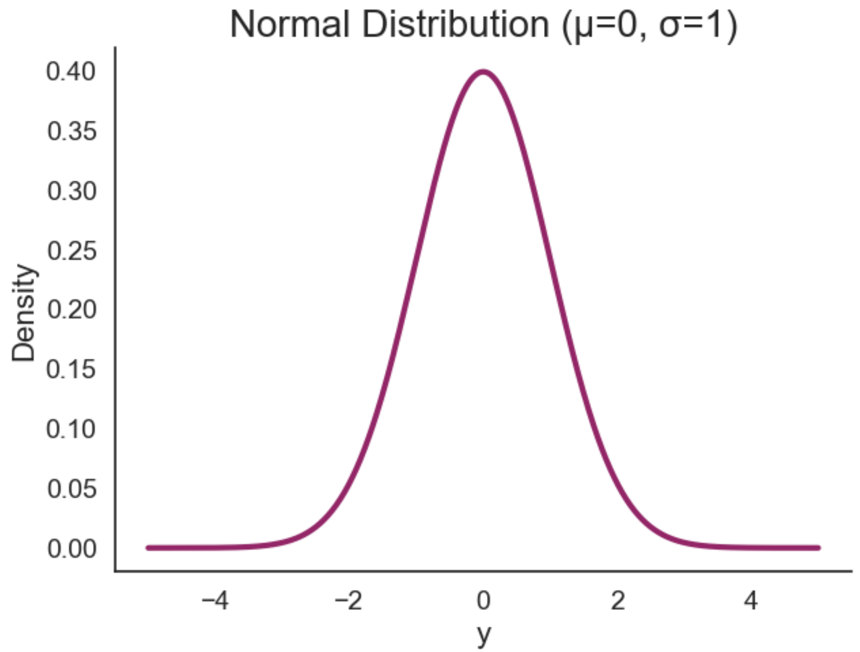
Product of marginals

Copula models of dependence

In the case of independent data x_1, x_2 we have that their joint distribution factorises as:

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$$

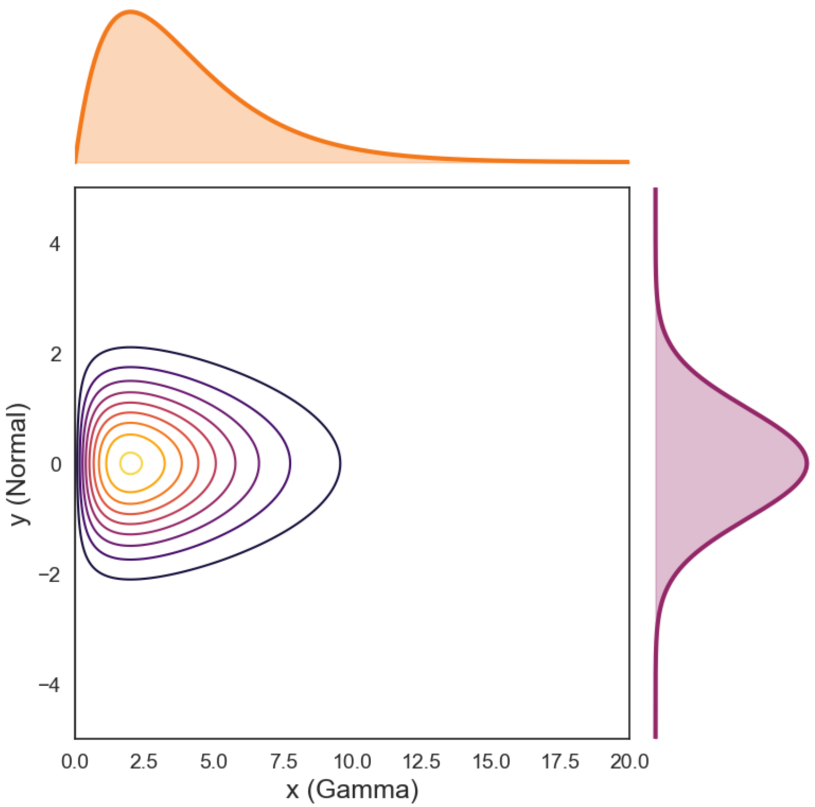




Copula models of dependence

In the case of independent data x_1, x_2 we have that their joint distribution factorises as:

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$$



Copula models of dependence

Now assume that x_1, x_2 are not independent. We then have:

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \cdot$$

Copula models of dependence

Now assume that x_1, x_2 are not independent. We then have:

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \cdot \ c(x_1, x_2)$$

Copula models of dependence

Now assume that x_1, x_2 are not independent. We then have:

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \cdot c(x_1, x_2)$$

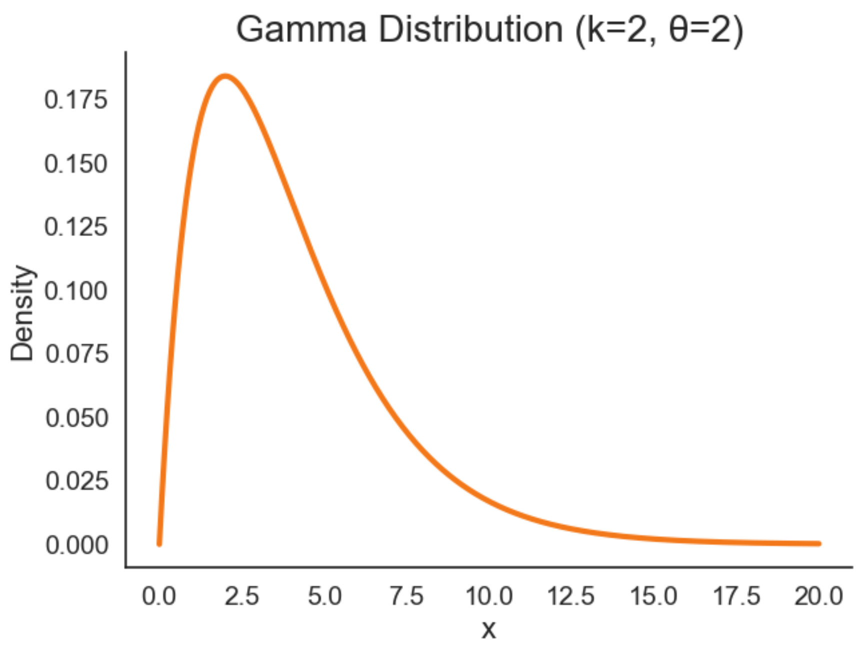
↑

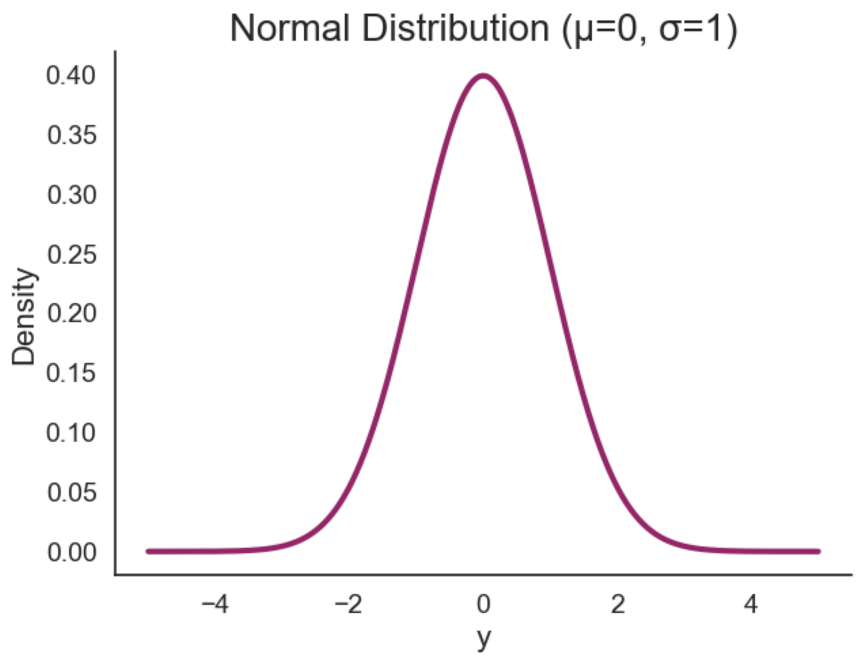
Copula = dependence

Copula models of dependence

Now assume that x_1, x_2 are not independent. We then have:

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \cdot c(x_1, x_2)$$

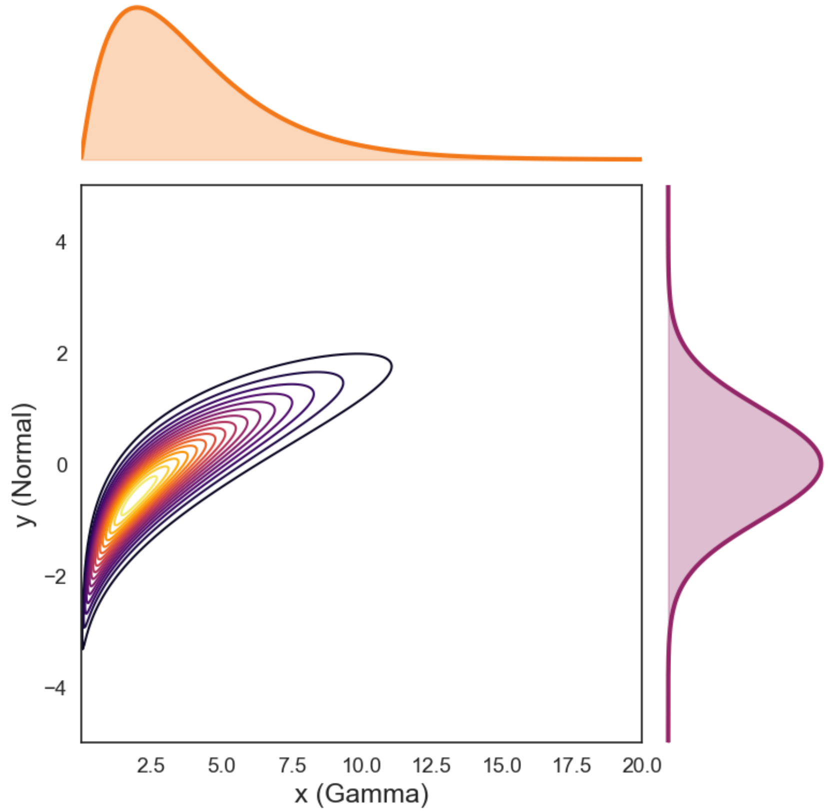




Copula models of dependence

Now assume that x_1, x_2 are not independent. We then have:

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \cdot c(x_1, x_2)$$



Copula models of dependence

Sklar's Theorem:

Any continuous joint density can always be **uniquely** decomposed into:

$$p(\mathbf{x}) = \prod_{i=1}^{d} \underbrace{\{p^i(x^i)\}}_{\textbf{Product of marginals}} \cdot \underbrace{\mathbf{c}(P^1(x^1), \ldots, P^d(x^d))}_{\textbf{Copula = dependence}}$$

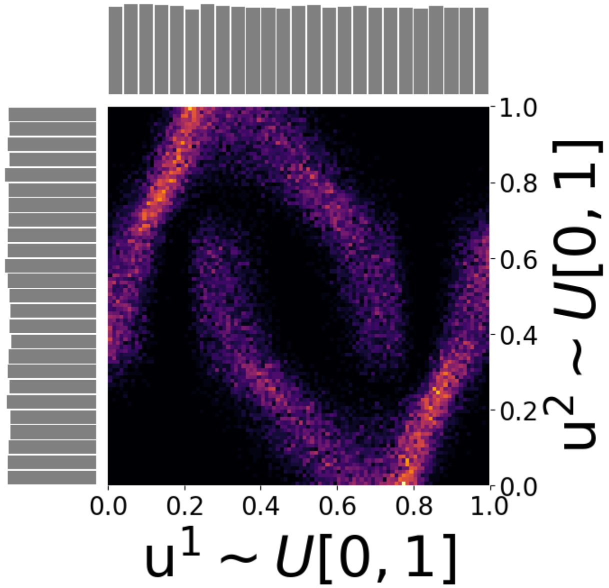
Copula models of dependence

Sklar's Theorem:

Any continuous joint density can always be **uniquely** decomposed into:

$$p(\mathbf{x}) = \underbrace{\prod_{i=1}^{d} \{p^i(x^i)\}}_{\text{Product of marginals}} \cdot \underbrace{\mathbf{c}(P^1(x^1), \ldots, P^d(x^d))}_{\text{Copula = dependence}}$$

Product of marginals **Copula = dependence**



Copula models of dependence

Sklar's Theorem:

Any continuous joint density can always be **uniquely** decomposed into:

$$p(\mathbf{x}) = \prod_{i=1}^{d} \underbrace{\{p^i(x^i)\}}_{\text{Product of marginals}} \cdot \underbrace{\mathbf{c}(P^1(x^1), \ldots, P^d(x^d))}_{\text{Copula = dependence}}$$

Part of SOTA in forecasting, causal inference, Bayesian methods, etc...

Copula models of dependence

Sklar's Theorem:

Any continuous joint density can always be **uniquely** decomposed into:

$$p(\mathbf{x}) = \prod_{i=1}^{d} \underbrace{\{p^i(x^i)\}}_{\text{Product of marginals}} \cdot \underbrace{\mathbf{c}(P^1(x^1), \ldots, P^d(x^d))}_{\text{Copula = dependence}}$$

Part of SOTA in forecasting, causal inference, Bayesian methods, etc...

Existing copulas have big **drawbacks**.

Copula models of dependence

Sklar's Theorem:

Any continuous joint density can always be **uniquely** decomposed into:

$$p(\mathbf{x}) = \underbrace{\prod_{i=1}^{d} \{p^i(x^i)\}}_{\text{Product of marginals}} \cdot \underbrace{\mathbf{c}(P^1(x^1), \ldots, P^d(x^d))}_{\text{Copula = dependence}}$$

Product of marginals **Copula = dependence**

Part of SOTA in forecasting, causal inference, Bayesian methods, etc...

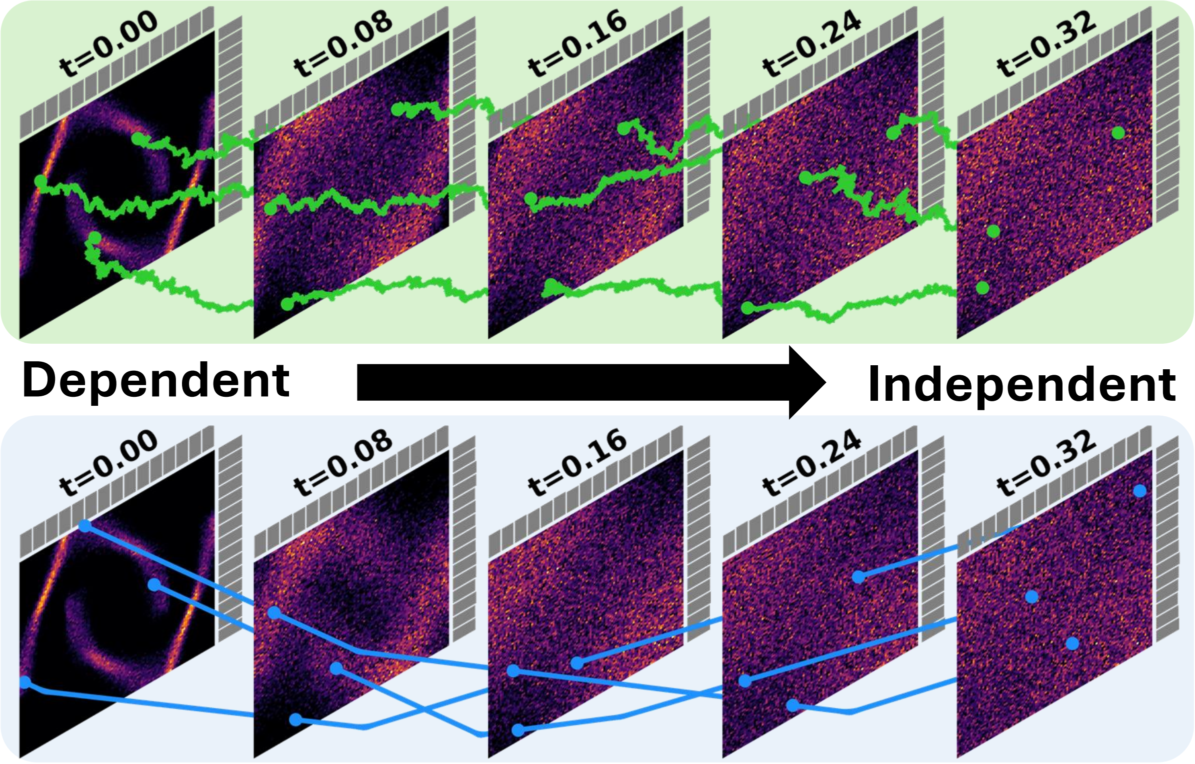
Existing copulas have big **drawbacks**.

→ **Develop better copulas for high-d and complex data.** ←

Forget and **Remember** Dependence

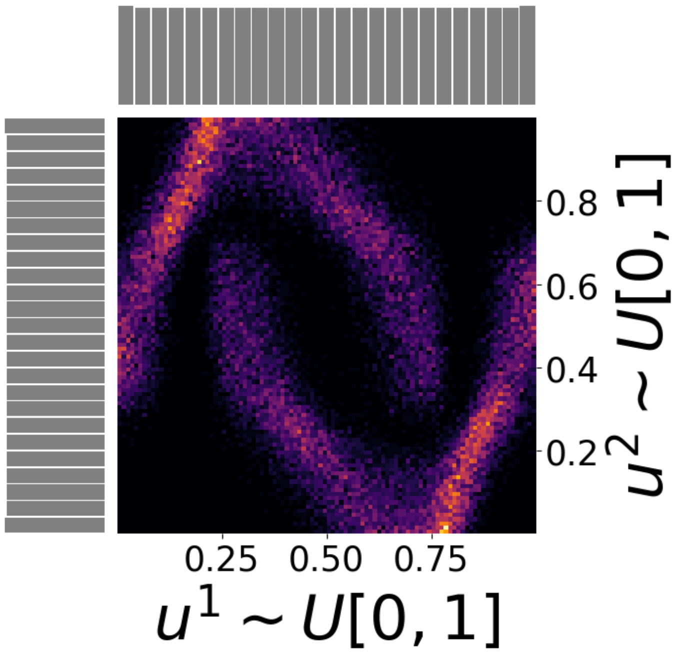
Goal 1. Only **forget dependence** but is always a **valid copula**

Goal 2. **Remember** forgotten dependencies to learn the copula.



CDC: Classification-Diffusion Copula

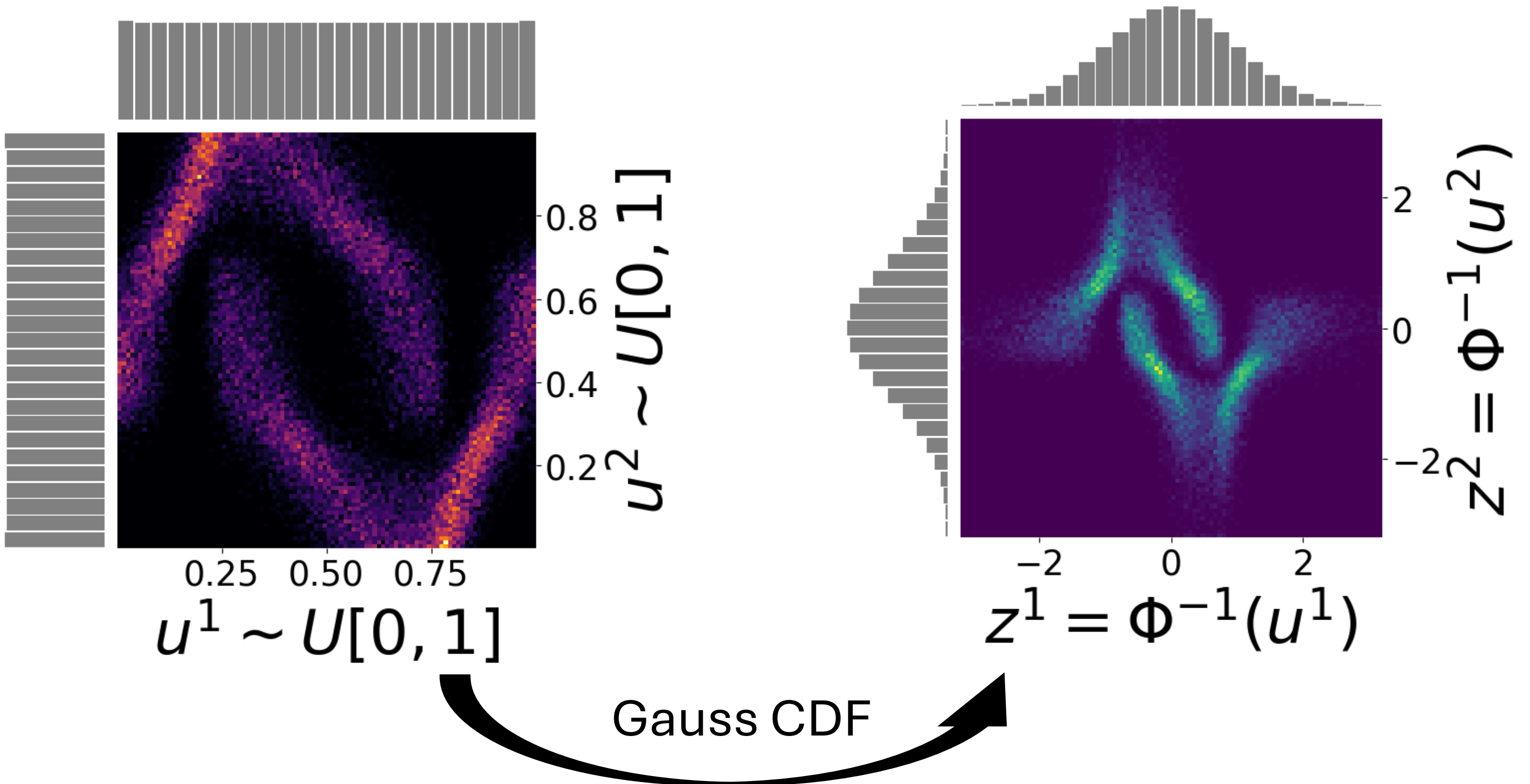
Map copula data to **Gaussian scale**



CDC: Classification-Diffusion Copula

Map copula data to **Gaussian scale**

$$z^i = \Phi^{-1}(u^i)$$



CDC: Classification-Diffusion Copula

Map copula data to **Gaussian scale**

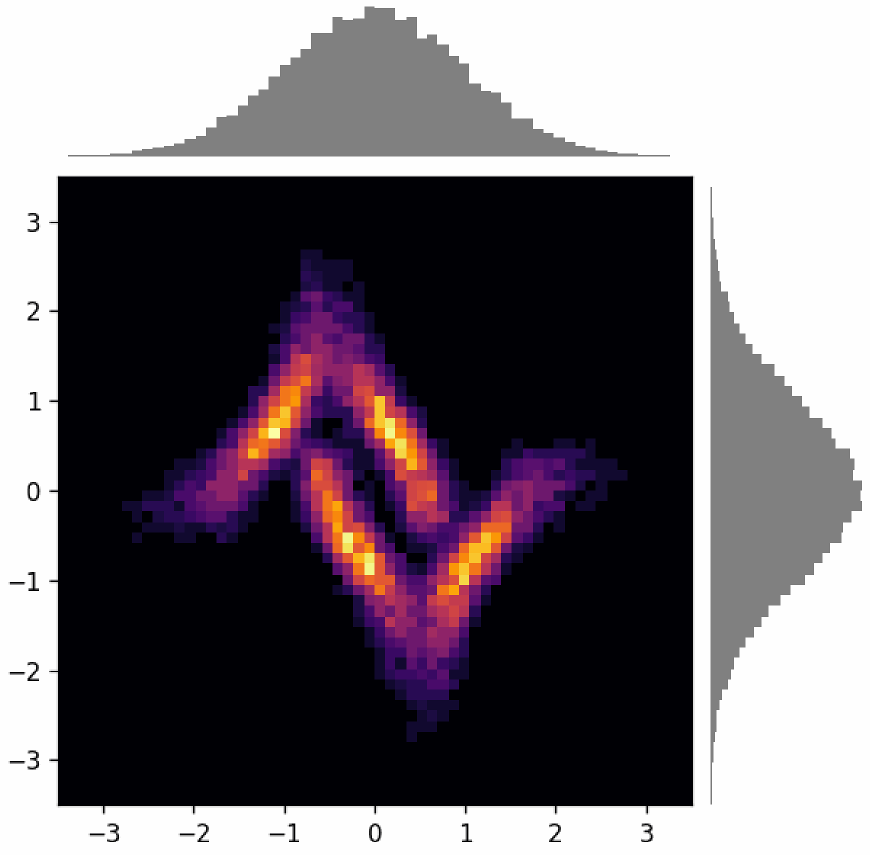
Apply Ornstein-Uhlenbeck

$$d\mathbf{z}_t = -\mathbf{z}_t \, dt + \sqrt{2} \, d\mathcal{B}_t$$

CDC: Classification-Diffusion Copula

Map copula data to **Gaussian scale**

Apply Ornstein-Uhlenbeck



CDC: Classification-Diffusion Copula

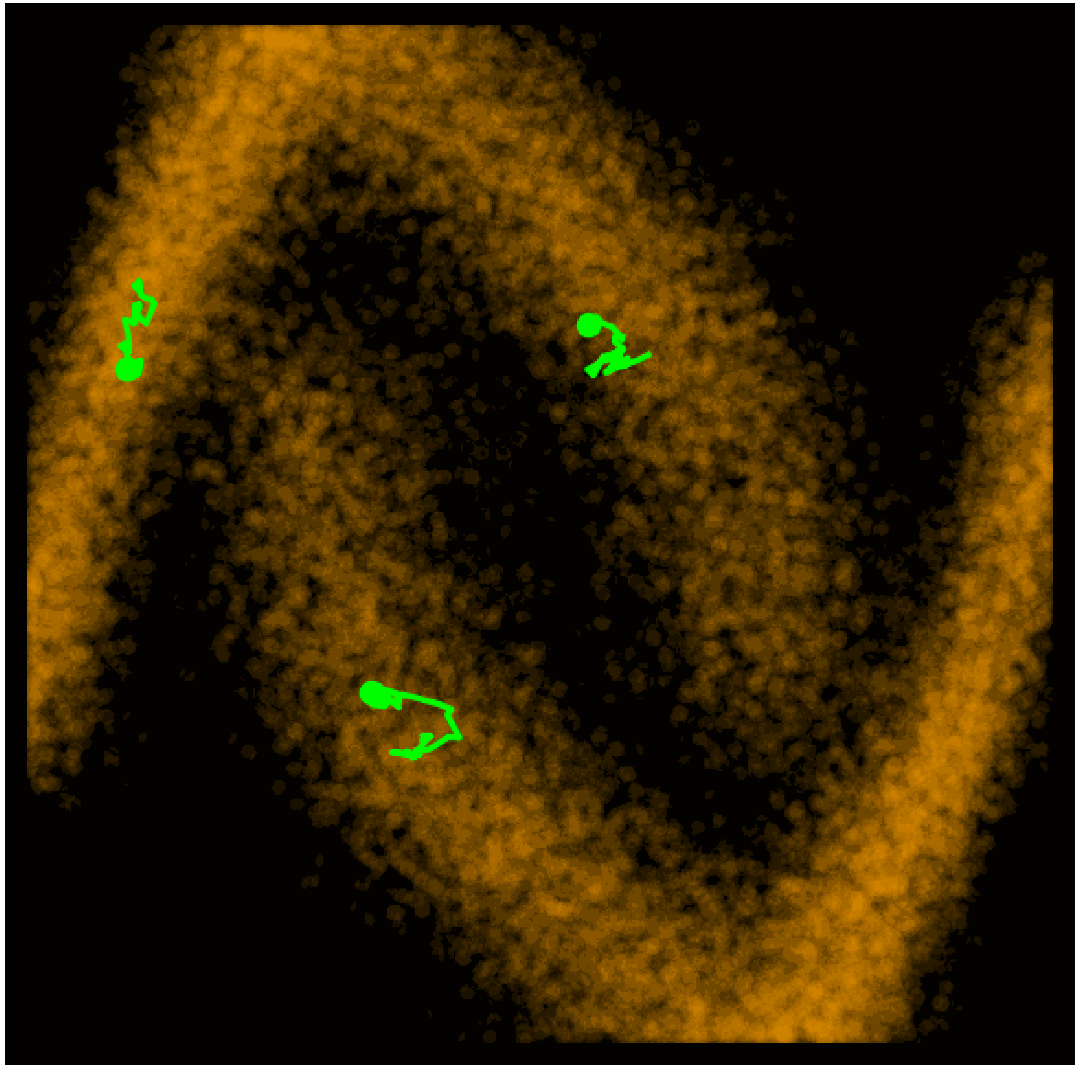
Map copula data to **Gaussian scale**

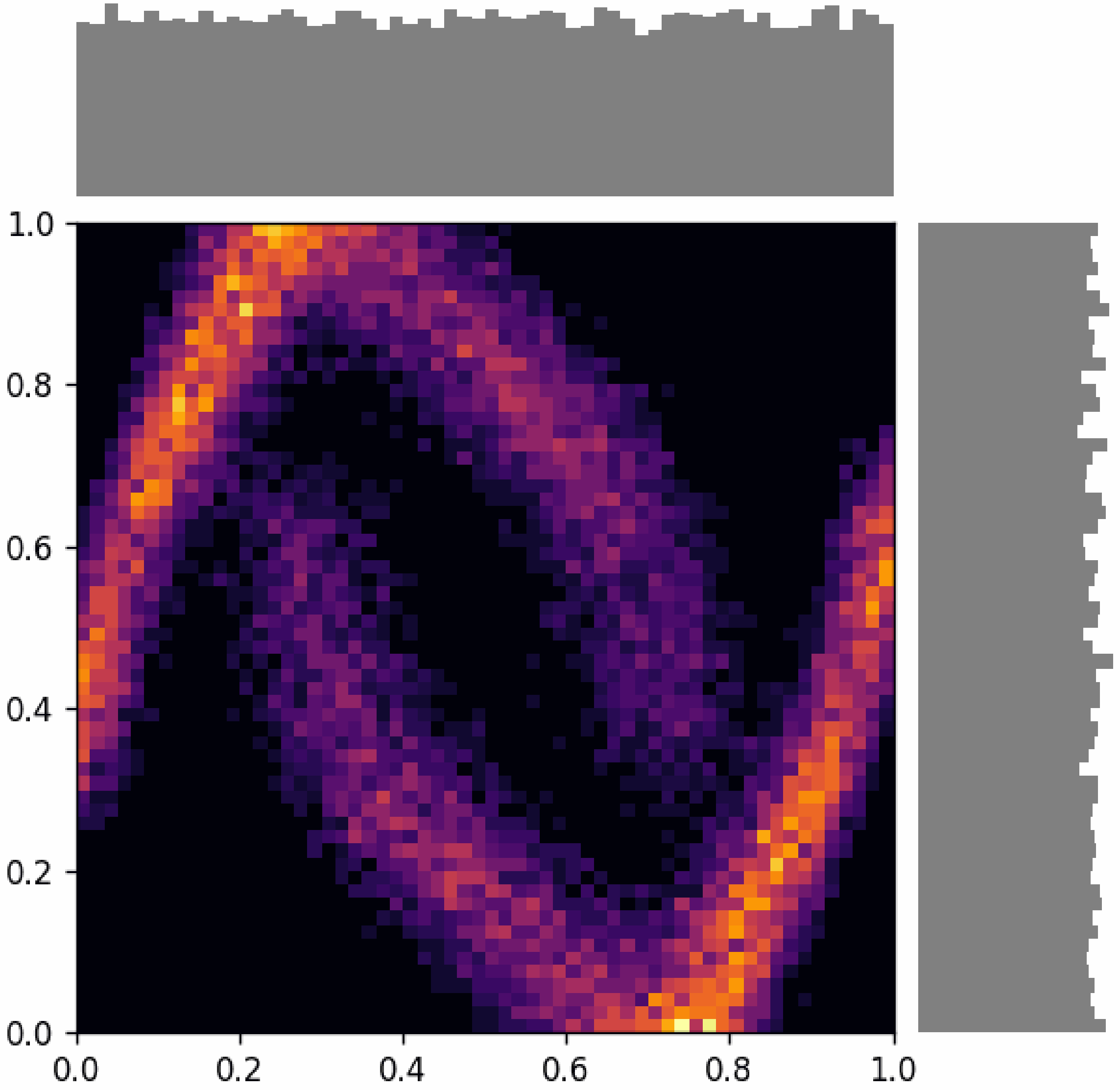
Apply Ornstein-Uhlenbeck

✓ **Goal 1**

The **copula scale** $\mathbf{u}_t^i = \Phi(\mathbf{z}_t^i)$ **is always uniform.**
The copula converges to **independence.**

CDC: Classification-Diffusion Copula





CDC: Classification-Diffusion Copula

Classify time based on dependence level:

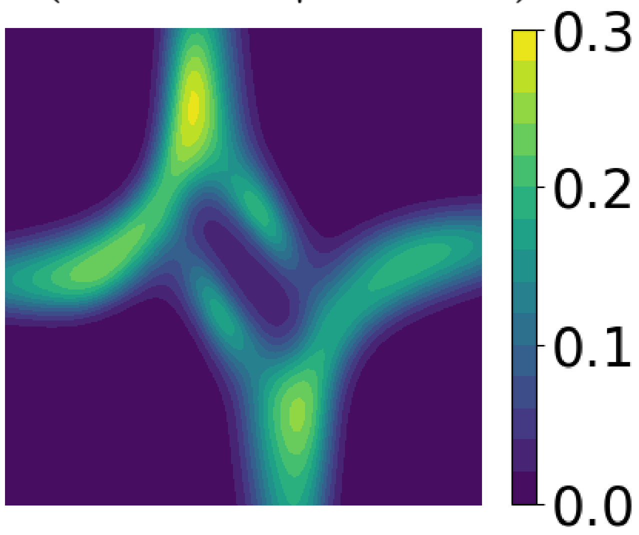
$$c_{dc}(\boldsymbol{z}) = \left(\mathbb{P}(t = T_1|\mathbf{z} = \boldsymbol{z}), \ldots, \mathbb{P}(t = T_k|\mathbf{z} = \boldsymbol{z})\right)$$

CDC: Classification-Diffusion Copula

Classify time based on dependence level:

$$c_{dc}(\boldsymbol{z}) = \left(\mathbb{P}(t = T_1 | \mathbf{z} = \boldsymbol{z}), \ldots, \mathbb{P}(t = T_k | \mathbf{z} = \boldsymbol{z})\right)$$

$\mathbb{P}(t = T_1 | \mathbf{z} = \boldsymbol{z})$

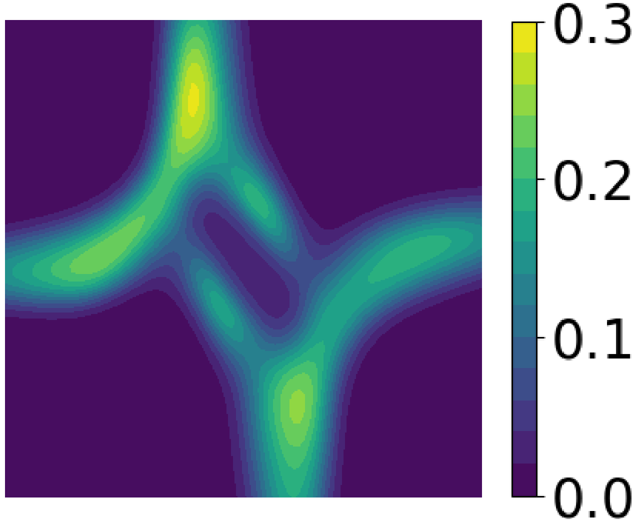


CDC: Classification-Diffusion Copula

Classify time based on dependence level:

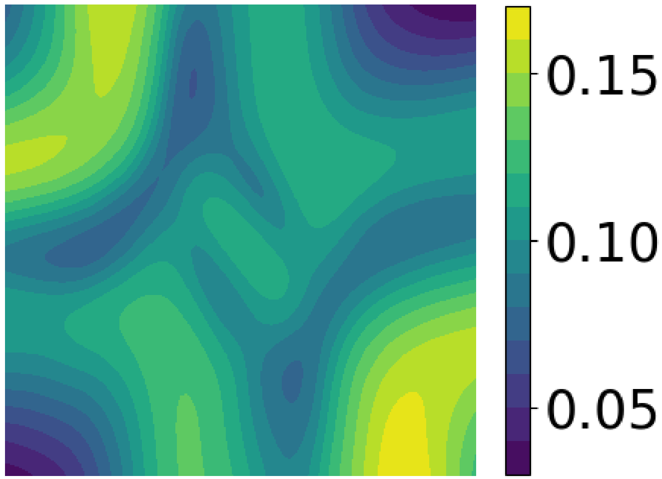
$$c_{dc}(z) = \left(\mathbb{P}(t = T_1 | \mathbf{z} = z), \ldots, \mathbb{P}(t = T_k | \mathbf{z} = z)\right)$$

$\mathbb{P}(t = T_1 | \mathbf{z} = z)$



...

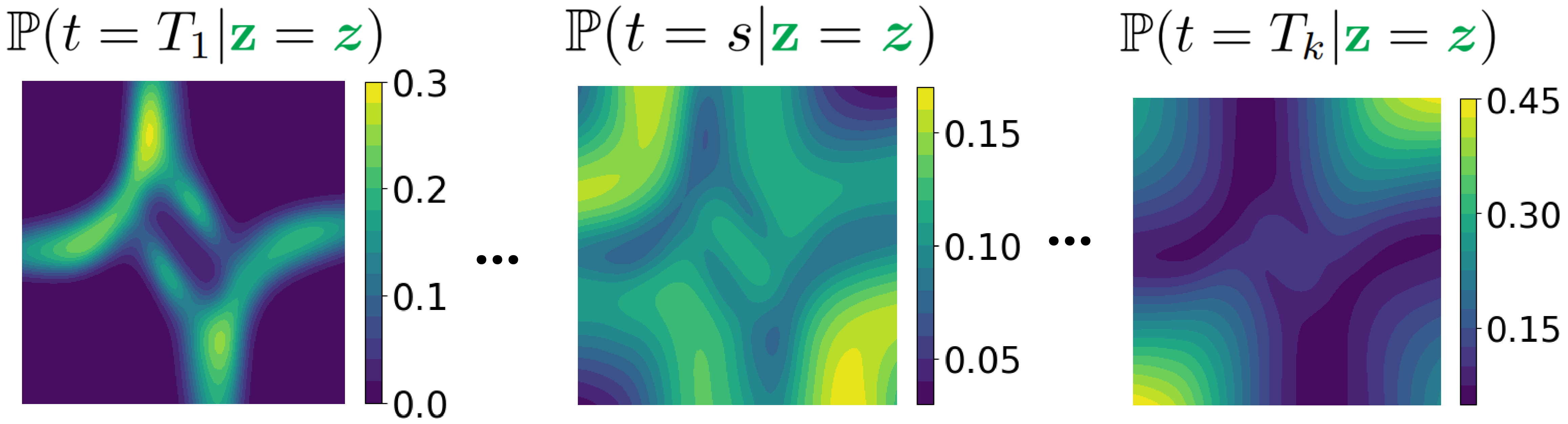
$\mathbb{P}(t = s | \mathbf{z} = z)$



CDC: Classification-Diffusion Copula

Classify time based on dependence level:

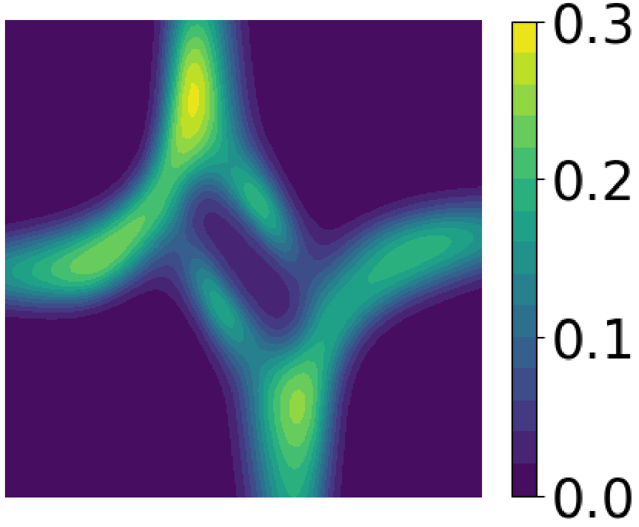
$$c_{dc}(\boldsymbol{z}) = \left(\mathbb{P}(t = T_1 | \mathbf{z} = \boldsymbol{z}), \ldots, \mathbb{P}(t = T_k | \mathbf{z} = \boldsymbol{z})\right)$$



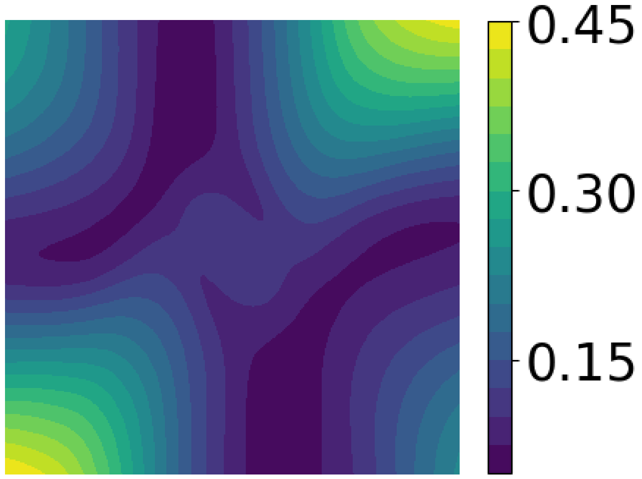
CDC: Classification-Diffusion Copula

Copula Likelihood: $c(\boldsymbol{u})$

$\mathbb{P}(t = T_1 | \mathbf{z} = \boldsymbol{z})$

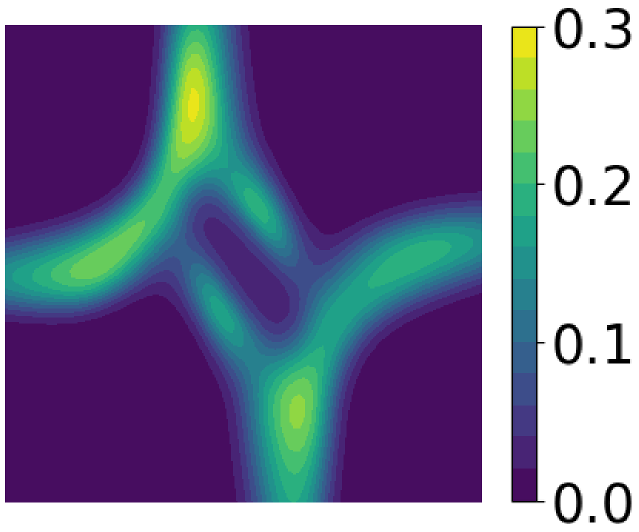


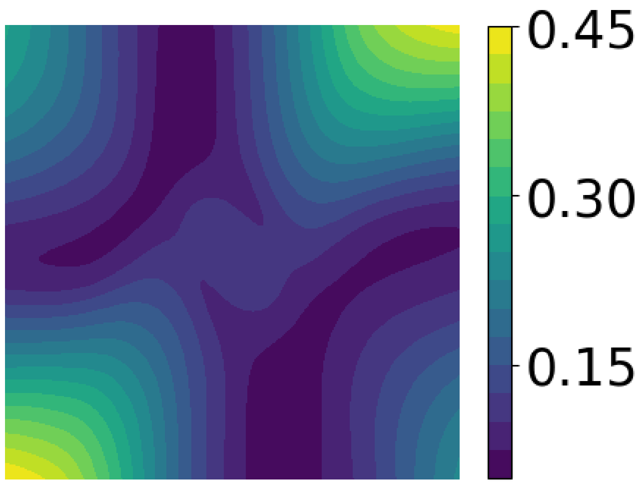
$\mathbb{P}(t = T_k | \mathbf{z} = \boldsymbol{z})$



CDC: Classification-Diffusion Copula

Copula Likelihood: $c(u) = \frac{\mathbb{P}(t = T_1 | \mathbf{z} = \bar{\Phi}^{-1}(u))}{\mathbb{P}(t = T_k | \mathbf{z} = \bar{\Phi}^{-1}(u))}$

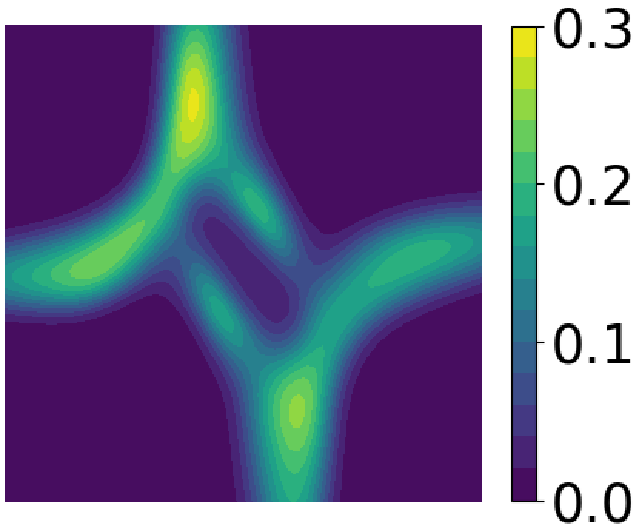


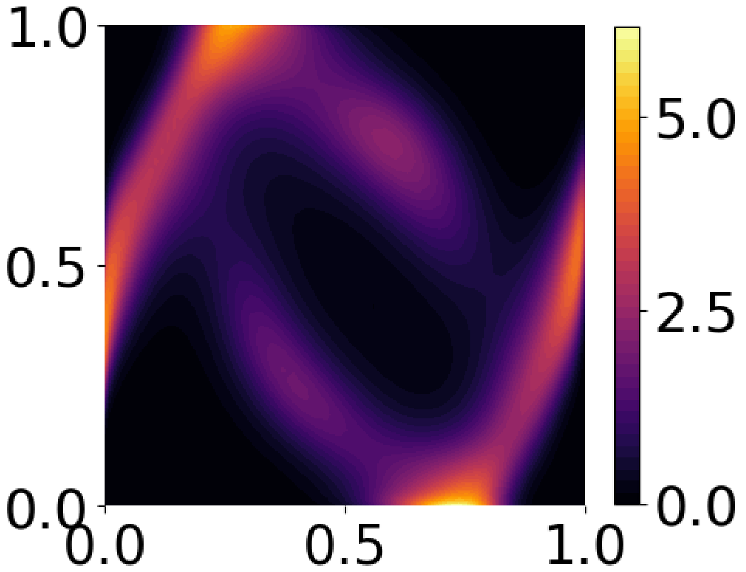


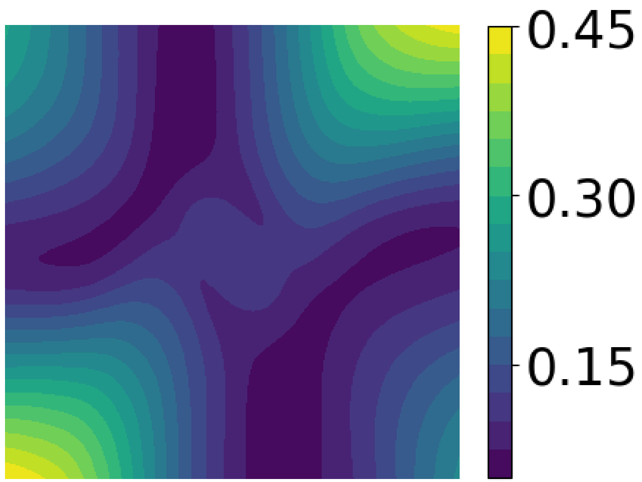
CDC: Classification-Diffusion Copula

Copula Likelihood: $$c(u) = \frac{\mathbb{P}(t = T_1 | \mathbf{z} = \bar{\Phi}^{-1}(u))}{\mathbb{P}(t = T_k | \mathbf{z} = \bar{\Phi}^{-1}(u))}$$

✓ **Goal 2**



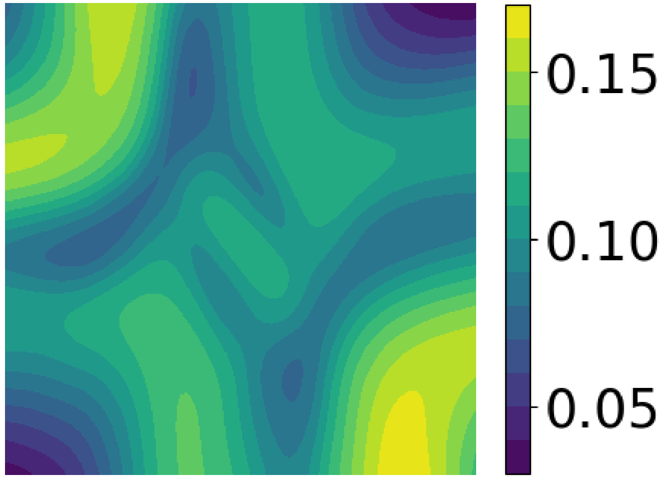




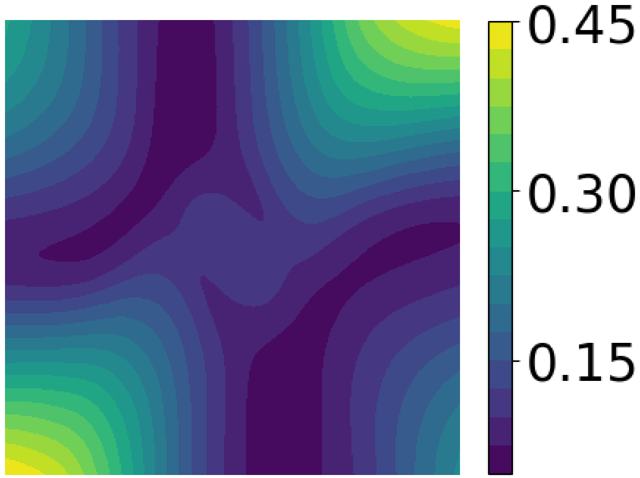
CDC: Classification-Diffusion Copula

Score-based copula sampling $\nabla_{\boldsymbol{u}} \log c_s(\boldsymbol{u}) =$

$\mathbb{P}(t = s | \mathbf{z} = \boldsymbol{z})$



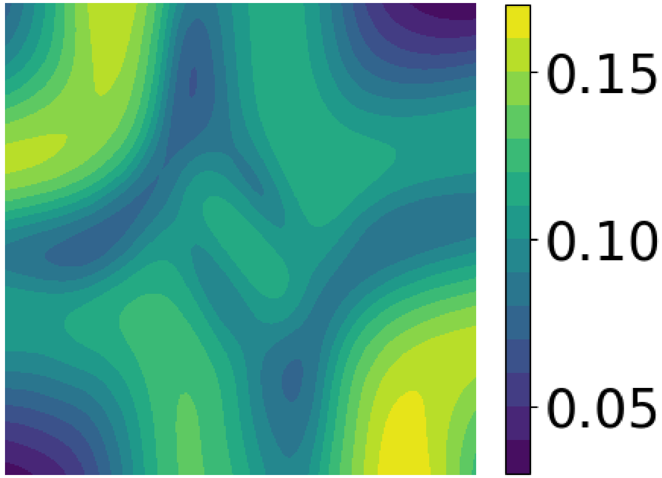
$\mathbb{P}(t = T_k | \mathbf{z} = \boldsymbol{z})$

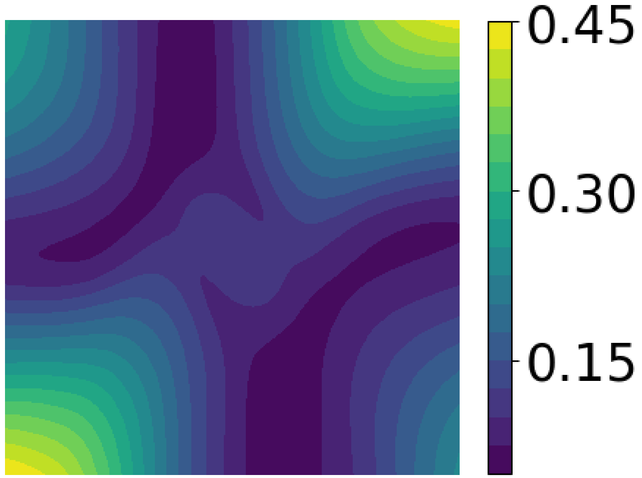


CDC: Classification-Diffusion Copula

Score-based copula sampling

$$\nabla_{\boldsymbol{u}} \log c_s(\boldsymbol{u}) = \nabla_{\boldsymbol{u}} \frac{\log \mathbb{P}\big(t = T_s | \mathbf{z} = \bar{\Phi}^{-1}(\boldsymbol{u})\big)}{\log \mathbb{P}\big(t = T_k | \mathbf{z} = \bar{\Phi}^{-1}(\boldsymbol{u})\big)}$$



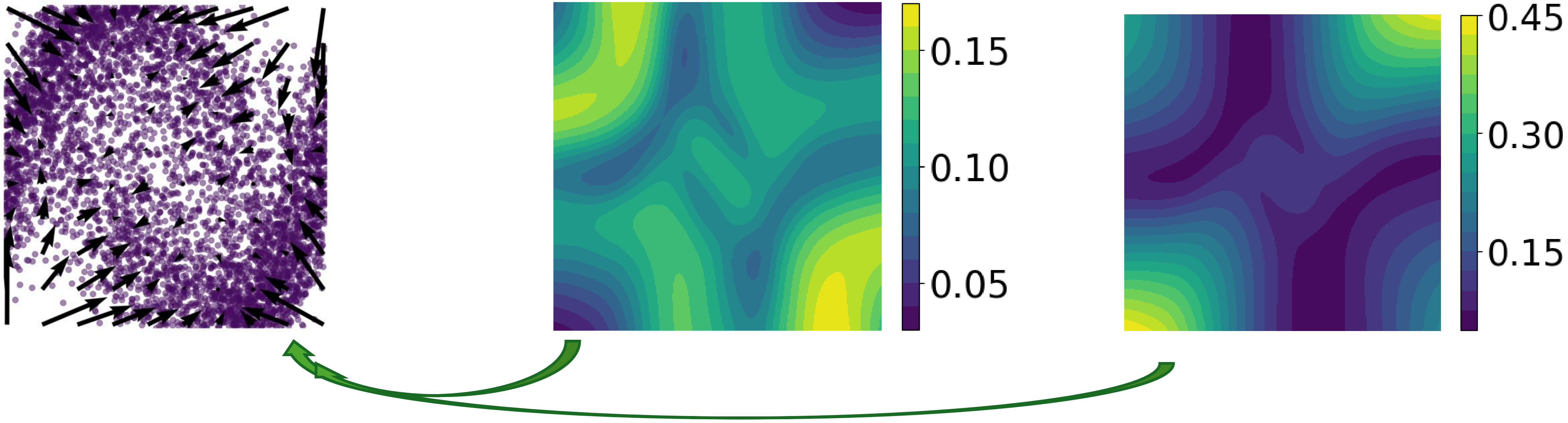


CDC: Classification-Diffusion Copula

Score-based copula sampling

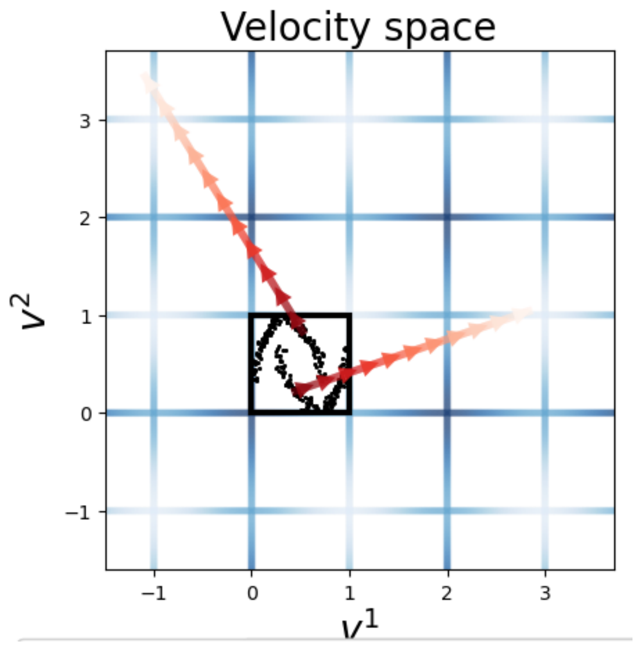
$$\nabla_{\boldsymbol{u}} \log c_s(\boldsymbol{u}) = \nabla_{\boldsymbol{u}} \frac{\log \mathbb{P}\left(t = T_s | \mathbf{z} = \bar{\Phi}^{-1}(\boldsymbol{u})\right)}{\log \mathbb{P}\left(t = T_k | \mathbf{z} = \bar{\Phi}^{-1}(\boldsymbol{u})\right)}$$

✓ **Goal 2**



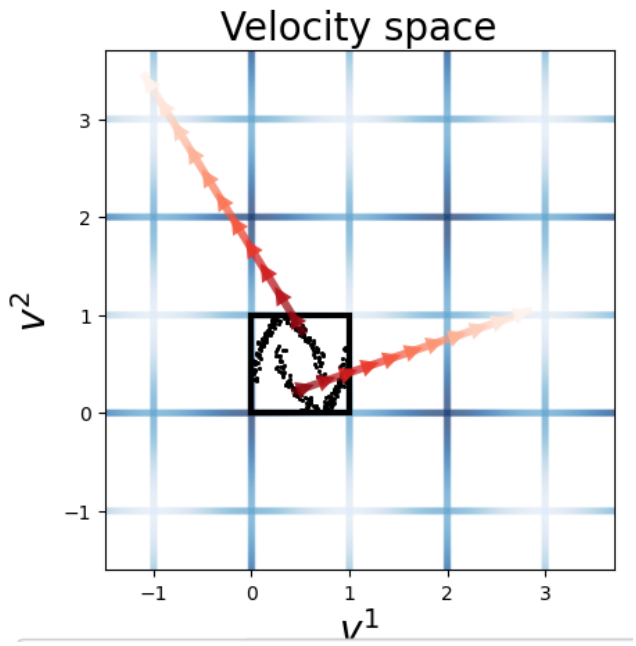
Reflection Copula

Augment data into **sample-velocity** pairs $(\mathbf{u}_t, \mathbf{v}_t)$, with $\mathbf{v}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$.



Reflection Copula

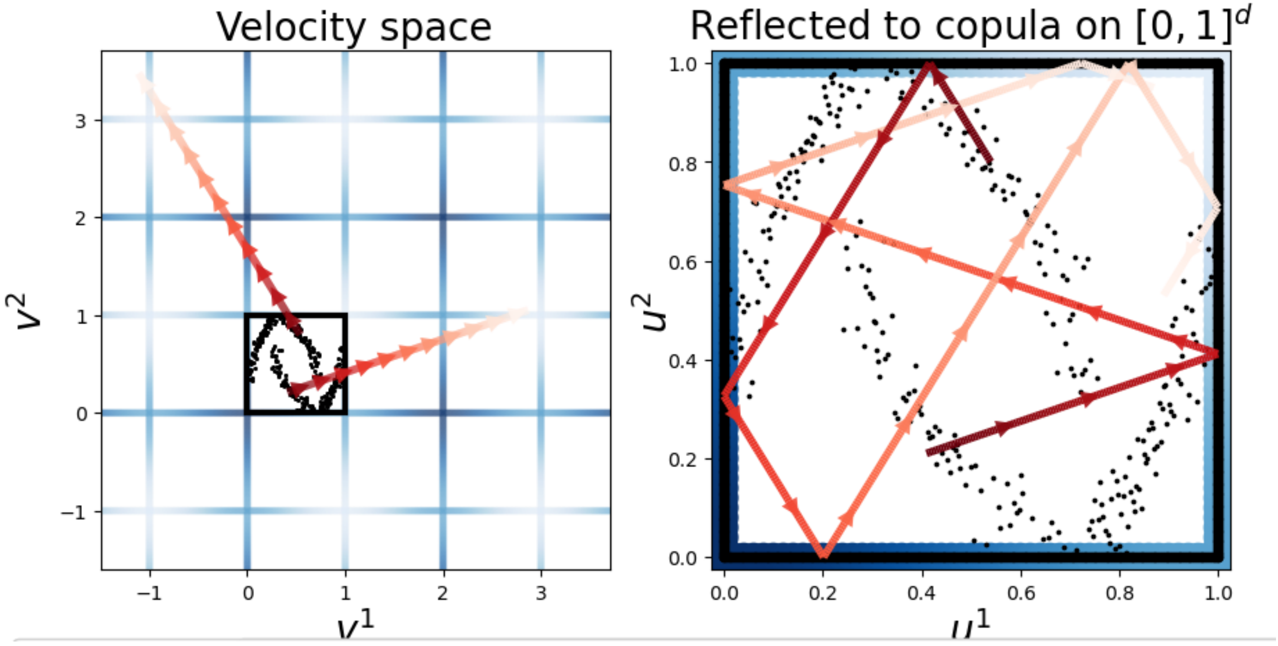
Augment data into **sample-velocity** pairs $(\mathbf{u}_t, \mathbf{v}_t)$, with $\mathbf{v}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$.



Reflection Copula

Augment data into **sample-velocity** pairs $(\mathbf{u}_t, \mathbf{v}_t)$, with $\mathbf{v}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$.

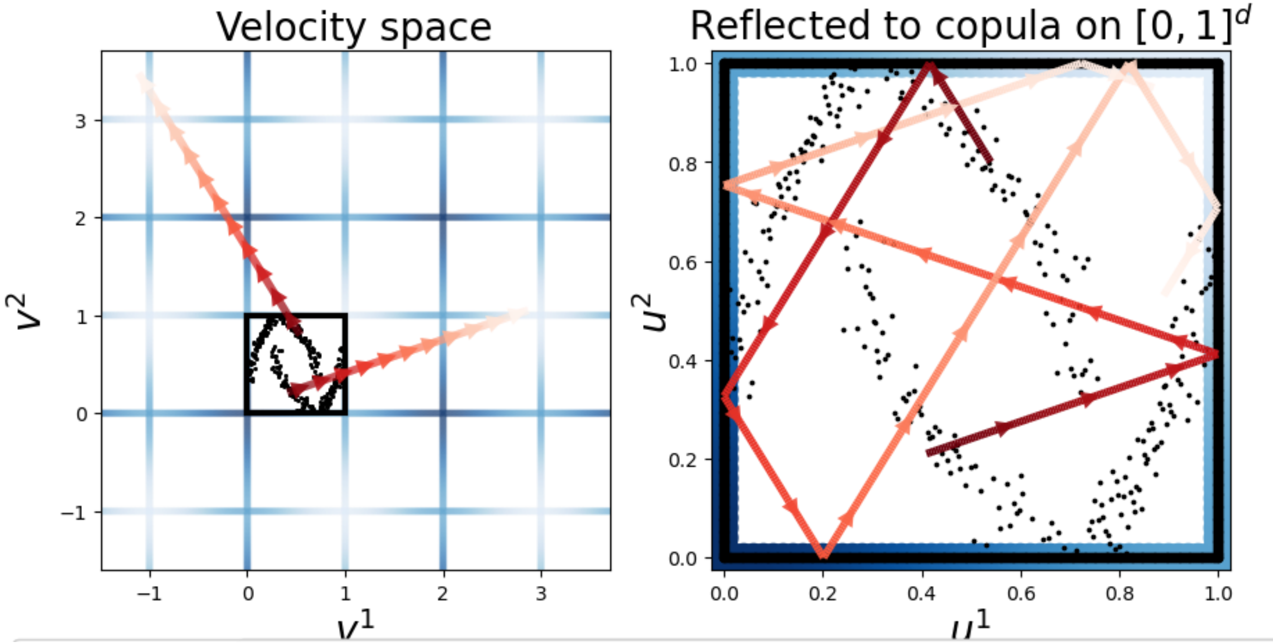
Let $\mathbf{u}_t$ **reflect** and bounce in $[0,1]^{\mathrm{d}}$.



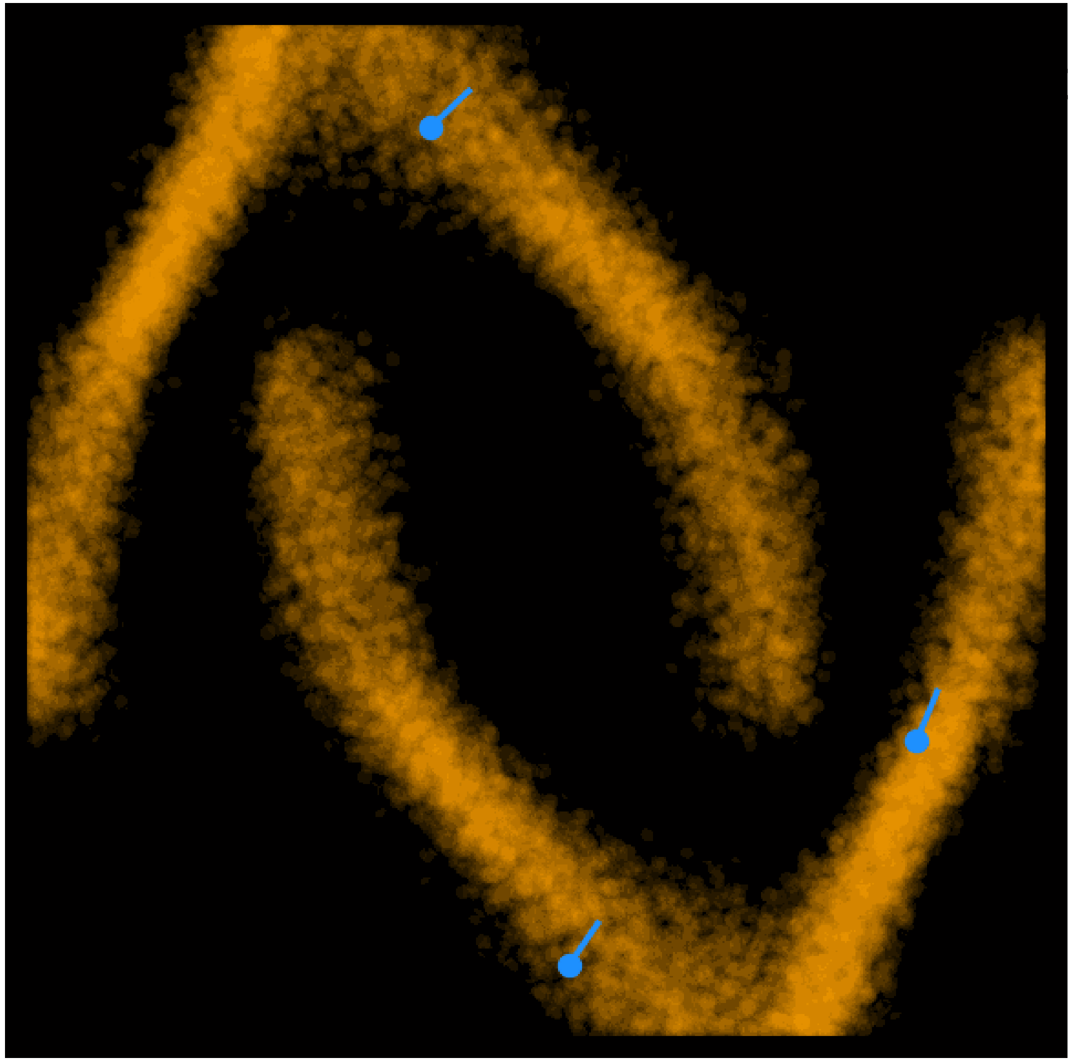
Reflection Copula

Reflected marginals stay **uniform** $\mathbf{u}_t^i \sim U(0,1)$.
Copula converges to **independence**.

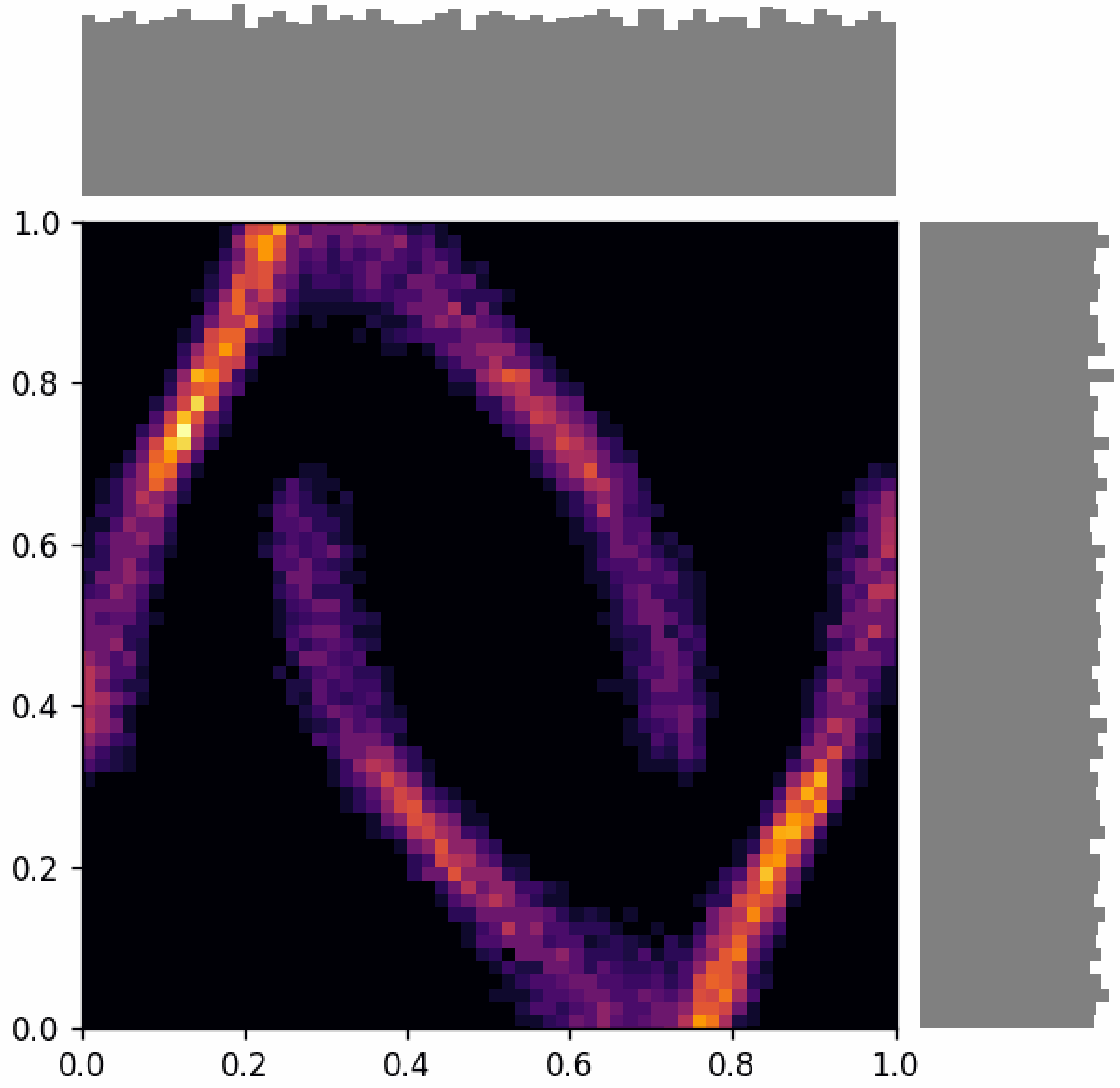
✓ **Goal 1**



Reflection Copula



is



Reflection Copula

Train a **velocity** prediction model: $v_{\theta^*}(\boldsymbol{u}, t) = \mathbb{E}[\mathbf{v}_t | \mathbf{u}_t = \boldsymbol{u}]$

Reflection Copula

Train a **velocity** prediction model: $v_{\theta^*}(\boldsymbol{u}, t) = \mathbb{E}[\mathbf{v}_t | \mathbf{u}_t = \boldsymbol{u}]$

Start from independence $\mathbf{u}_T \sim c_T(\mathbf{u}_T)$ and follow $\frac{d}{dt}\boldsymbol{u}_t = v^*(\boldsymbol{u}, t)$

in reverse to $T \to 0$. Then $\boldsymbol{u}_0 \sim c(\mathbf{u})$.

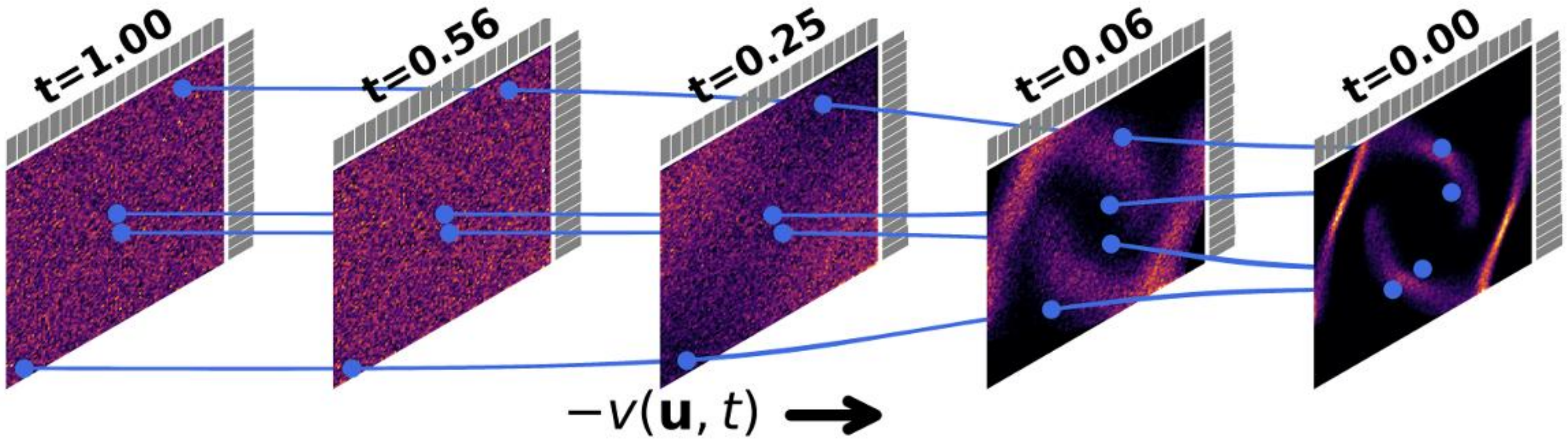
Reflection Copula

Train a **velocity** prediction model: $v_{\theta^*}(\boldsymbol{u}, t) = \mathbb{E}[\mathbf{v}_t | \mathbf{u}_t = \boldsymbol{u}]$

Start from independence $\mathbf{u}_T \sim c_T(\mathbf{u}_T)$ and follow $\frac{d}{dt}\boldsymbol{u}_t = v^*(\boldsymbol{u}, t)$

in reverse to $T \to 0$. Then $\boldsymbol{u}_0 \sim c(\mathbf{u})$.

✓ **Goal 2**



Results

- **CDC** produces **direct likelihoods** and does score-based sampling.
- **Reflection copula** produces **fast copula samples**.

Results

- **CDC** produces **direct likelihoods** and does score-based sampling.
- **Reflection copula** produces **fast copula samples**.

New **SOTA copula models** in sampling and density

Model	Magic ($n=19020, d=10$)			Dry_Bean ($n=13611, d=16$)			Robocup ($n=135607, d=20$)		
	LL ↑	W2 ↓	Frob ↓	LL ↑	W2 ↓	Frob ↓	LL ↑	W2 ↓	Frob ↓
Gaussian	$3.92_{\pm 0.06}$	$1.76_{\pm 0.02}$	$0.27_{\pm 0.06}$	$40.09_{\pm 0.29}$	$1.57_{\pm 0.02}$	$0.40_{\pm 0.04}$	$0.22_{\pm 0.00}$	$3.96_{\pm 0.01}$	$\mathbf{0.45_{\pm 0.03}}$
Vine	$6.59_{\pm 0.07}$	$1.44_{\pm 0.01}$	$0.30_{\pm 0.05}$	$32.75_{\pm 0.14}$	$1.35_{\pm 0.03}$	$0.95_{\pm 0.07}$	$1.80_{\pm 0.00}$	$3.96_{\pm 0.01}$	$0.60_{\pm 0.04}$
Ratio	$6.76_{\pm 0.38}$	$2.26_{\pm 0.79}$	$1.24_{\pm 0.76}$	$48.21_{\pm 0.89}$	$2.54_{\pm 0.27}$	$2.25_{\pm 0.55}$	$2.30_{\pm 0.33}$	$3.93_{\pm 0.08}$	$0.59_{\pm 0.05}$
IGC	–	$1.69_{\pm 0.04}$	$1.24_{\pm 0.20}$	–	$1.66_{\pm 0.01}$	$2.31_{\pm 0.02}$	–	$4.13_{\pm 0.02}$	$2.85_{\pm 0.12}$
C_{dc} (ours)	$\mathbf{18.65_{\pm 4.85}}$	$\mathbf{1.33_{\pm 0.03}}$	$\mathbf{0.21_{\pm 0.05}}$	$\mathbf{50.21_{\pm 0.82}}$	$\mathbf{1.12_{\pm 0.03}}$	$\mathbf{0.35_{\pm 0.08}}$	$\mathbf{3.40_{\pm 0.37}}$	$\mathbf{3.87_{\pm 0.03}}$	$0.51_{\pm 0.02}$
Reflection (ours)	–	$\mathbf{1.34_{\pm 0.03}}$	$\underline{0.28_{\pm 0.07}}$	–	$\mathbf{1.35_{\pm 0.08}}$	$\underline{0.47_{\pm 0.16}}$	–	$\mathbf{3.84_{\pm 0.03}}$	$\underline{0.49_{\pm 0.02}}$

Results

- **CDC** produces **direct likelihoods** and does score-based sampling.
- **Reflection copula** produces **fast copula samples**.

New **SOTA copula models** in sampling and density

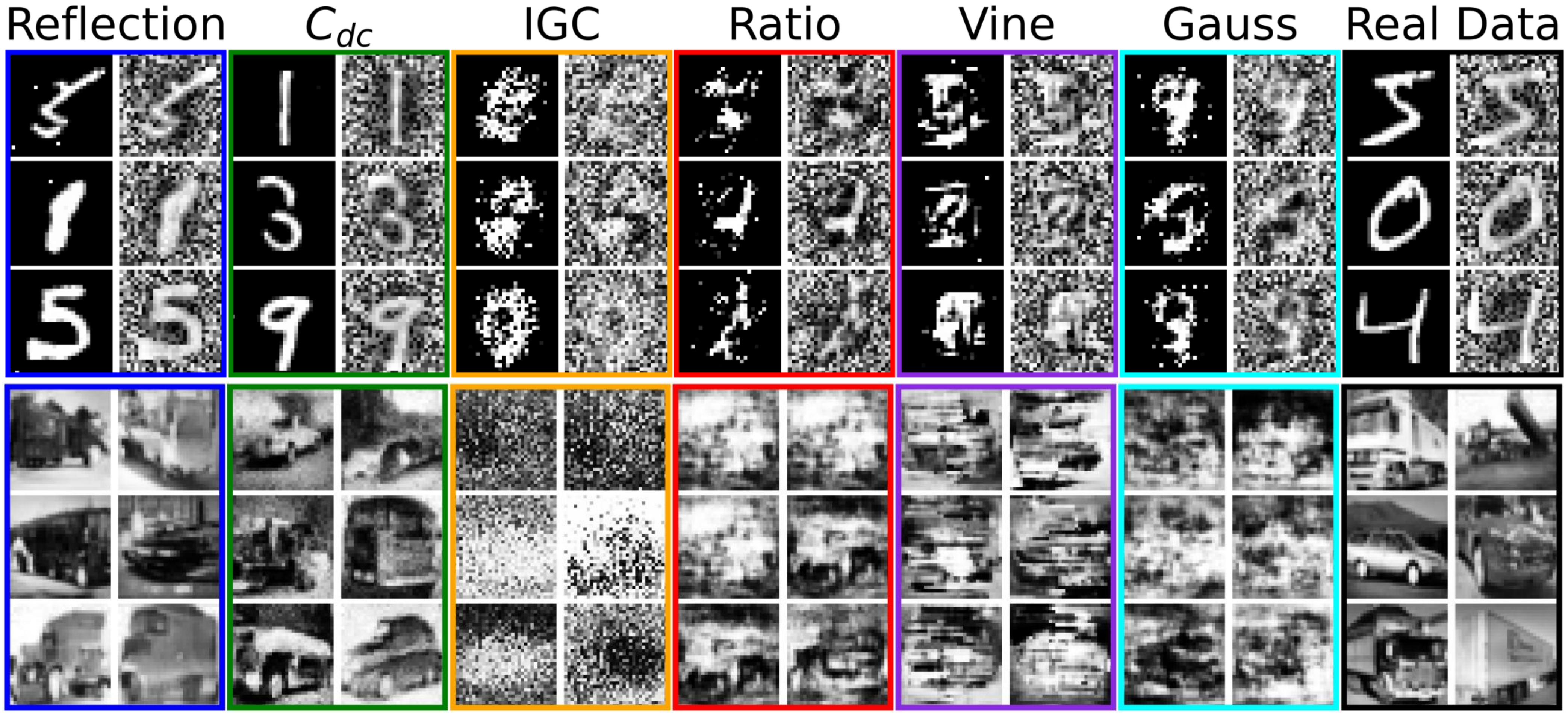
Model	Magic ($n = 19020, d = 10$)			Dry_Bean ($n = 13611, d = 16$)			Robocup ($n = 135607, d = 20$)		
	LL ↑	W2 ↓	Frob ↓	LL ↑	W2 ↓	Frob ↓	LL ↑	W2 ↓	Frob ↓
Gaussian	$3.92_{\pm0.06}$	$1.76_{\pm0.02}$	$0.27_{\pm0.06}$	$40.09_{\pm0.29}$	$1.57_{\pm0.02}$	$0.40_{\pm0.04}$	$0.22_{\pm0.00}$	$3.96_{\pm0.01}$	**$0.45_{\pm0.03}$**
Vine	$6.59_{\pm0.07}$	$1.44_{\pm0.01}$	$0.30_{\pm0.05}$	$32.75_{\pm0.14}$	$1.35_{\pm0.03}$	$0.95_{\pm0.07}$	$1.80_{\pm0.00}$	$3.96_{\pm0.01}$	$0.60_{\pm0.04}$
Ratio	$6.76_{\pm0.38}$	$2.26_{\pm0.79}$	$1.24_{\pm0.76}$	$48.21_{\pm0.89}$	$2.54_{\pm0.27}$	$2.25_{\pm0.55}$	$2.30_{\pm0.33}$	$3.93_{\pm0.08}$	$0.59_{\pm0.05}$
IGC	–	$1.69_{\pm0.04}$	$1.24_{\pm0.20}$	–	$1.66_{\pm0.01}$	$2.31_{\pm0.02}$	–	$4.13_{\pm0.02}$	$2.85_{\pm0.12}$
C_{dc} (ours)	**$18.65_{\pm4.85}$**	**$1.33_{\pm0.03}$**	**$0.21_{\pm0.05}$**	**$50.21_{\pm0.82}$**	**$1.12_{\pm0.03}$**	**$0.35_{\pm0.08}$**	**$3.40_{\pm0.37}$**	**$3.87_{\pm0.03}$**	$0.51_{\pm0.02}$
Reflection (ours)	–	**$1.34_{\pm0.03}$**	$0.28_{\pm0.07}$	–	**$1.35_{\pm0.08}$**	$0.47_{\pm0.16}$	–	**$3.84_{\pm0.03}$**	$0.49_{\pm0.02}$

First copulas to scale to **1000+ dimensions**.

Model	digits ($n = 1797, d = 64$)			MNIST ($n = 60000, d = 784$)			Cifar ($n = 10000, d = 1024$)		
	LL ↑	W2 ↓	FID ↓	LL ↑	W2 ↓	FID ↓	LL ↑	W2 ↓	FID ↓
Gaussian	$10.74_{\pm0.13}$	$8.13_{\pm0.02}$	$5.74_{\pm0.71}$	$115.84_{\pm0.14}$	$35.59_{\pm0.03}$	$102.56_{\pm2.61}$	$1258.40_{\pm5.69}$	$30.62_{\pm0.08}$	$140.12_{\pm2.41}$
Vine	$11.20_{\pm0.86}$	$8.20_{\pm0.01}$	$6.06_{\pm0.73}$	$198.10_{\pm0.40}$	$36.30_{\pm0.05}$	$86.48_{\pm2.82}$	NaN	$33.84_{\pm0.15}$	$100.04_{\pm2.52}$
Ratio	$13.29_{\pm2.75}$	$8.42_{\pm0.42}$	$6.04_{\pm0.97}$	$334.42_{\pm45.91}$	$35.98_{\pm0.38}$	$66.56_{\pm17.76}$	$1348.18_{\pm12.31}$	$49.91_{\pm19.22}$	$134.41_{\pm33.99}$
IGC	–	$9.52_{\pm0.15}$	$25.41_{\pm4.35}$	–	$36.37_{\pm0.14}$	$128.87_{\pm5.31}$	–	$33.09_{\pm0.30}$	$269.68_{\pm8.03}$
C_{dc} (ours)	**$13.80_{\pm1.30}$**	**$6.97_{\pm0.03}$**	$15.24_{\pm0.92}$	**$346.70_{\pm2.52}$**	**$33.64_{\pm0.03}$**	**$7.38_{\pm0.19}$**	**$1470.75_{\pm24.90}$**	**$28.67_{\pm0.50}$**	**$80.51_{\pm17.32}$**
Reflection (ours)	–	**$7.86_{\pm0.07}$**	**$5.50_{\pm1.36}$**	–	**$35.02_{\pm0.40}$**	**$9.13_{\pm0.90}$**	–	$32.40_{\pm2.08}$	**$42.14_{\pm3.23}$**

Results

First copulas to scale to **images**.



Conclusion

- Developed better copulas for high-d and complex data.

Paper and more!

Conclusion

- Developed better copulas for high-d and complex data.
- Formalised a forget-and-remember paradigm for copula models.

Paper and more!

Conclusion

- Developed better copulas for high-d and complex data.
- Formalised a forget-and-remember paradigm for copula models.
- **CDC** for likelihoods, **Reflection copula** for sampling.

Paper and more!

Conclusion

- Developed better copulas for high-d and complex data.
- Formalised a forget-and-remember paradigm for copula models.
- **CDC** for likelihoods, **Reflection copula** for sampling.
- Achieve a new SOTA in copula modelling.

Paper and more!

Conclusion

- Developed better copulas for high-d and complex data.
- Formalised a forget-and-remember paradigm for copula models.
- **CDC** for likelihoods, **Reflection copula** for sampling.
- Achieve a new SOTA in copula modelling.

Drop by Poster session 4 on Friday, April 24th!

Happy chat & collaborate.

Paper and more!