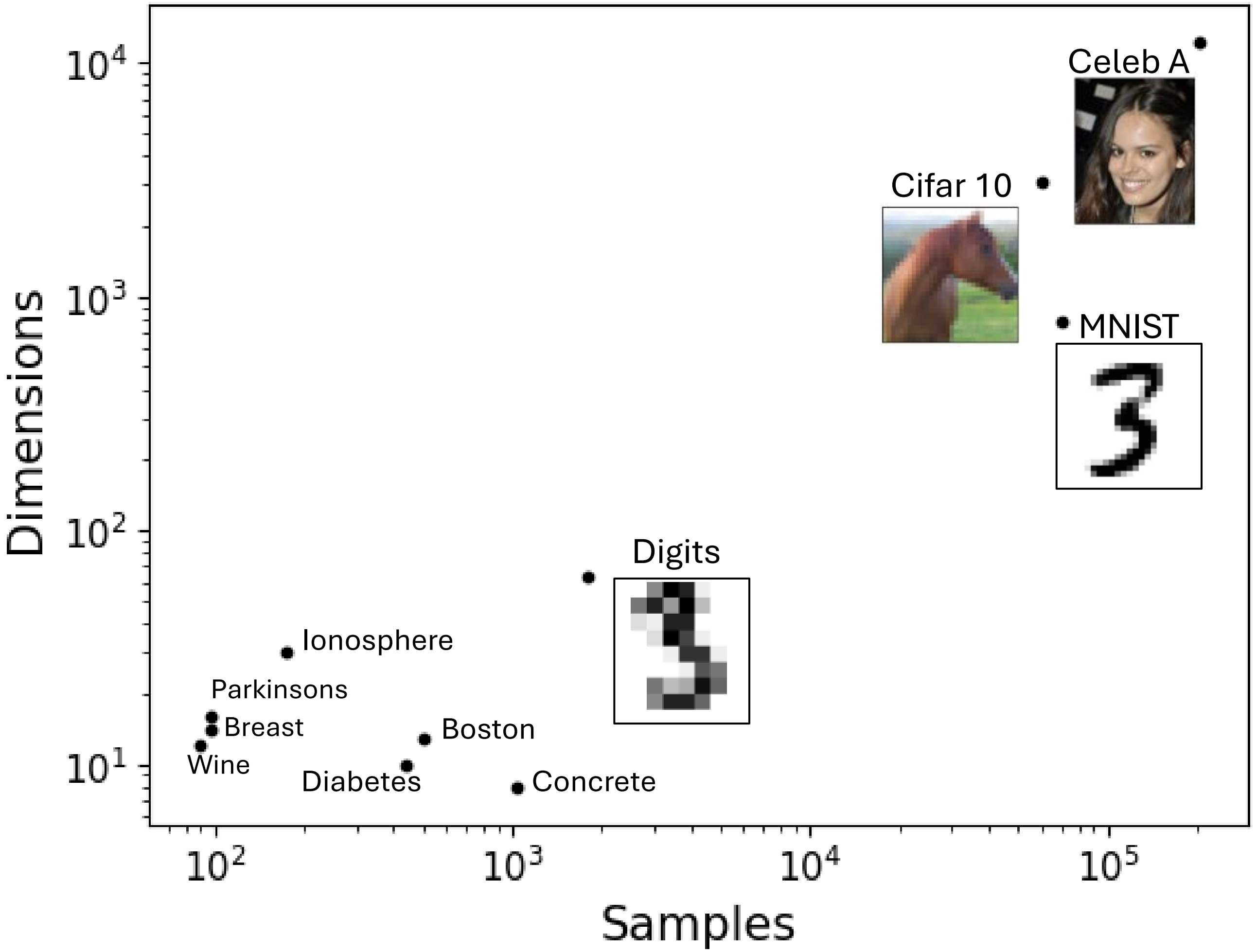
Quasi-Bayes meets Vines

David Huk, Yuanhe Zhang, Mark Steel, Ritabrata Dutta

Department of Statistics

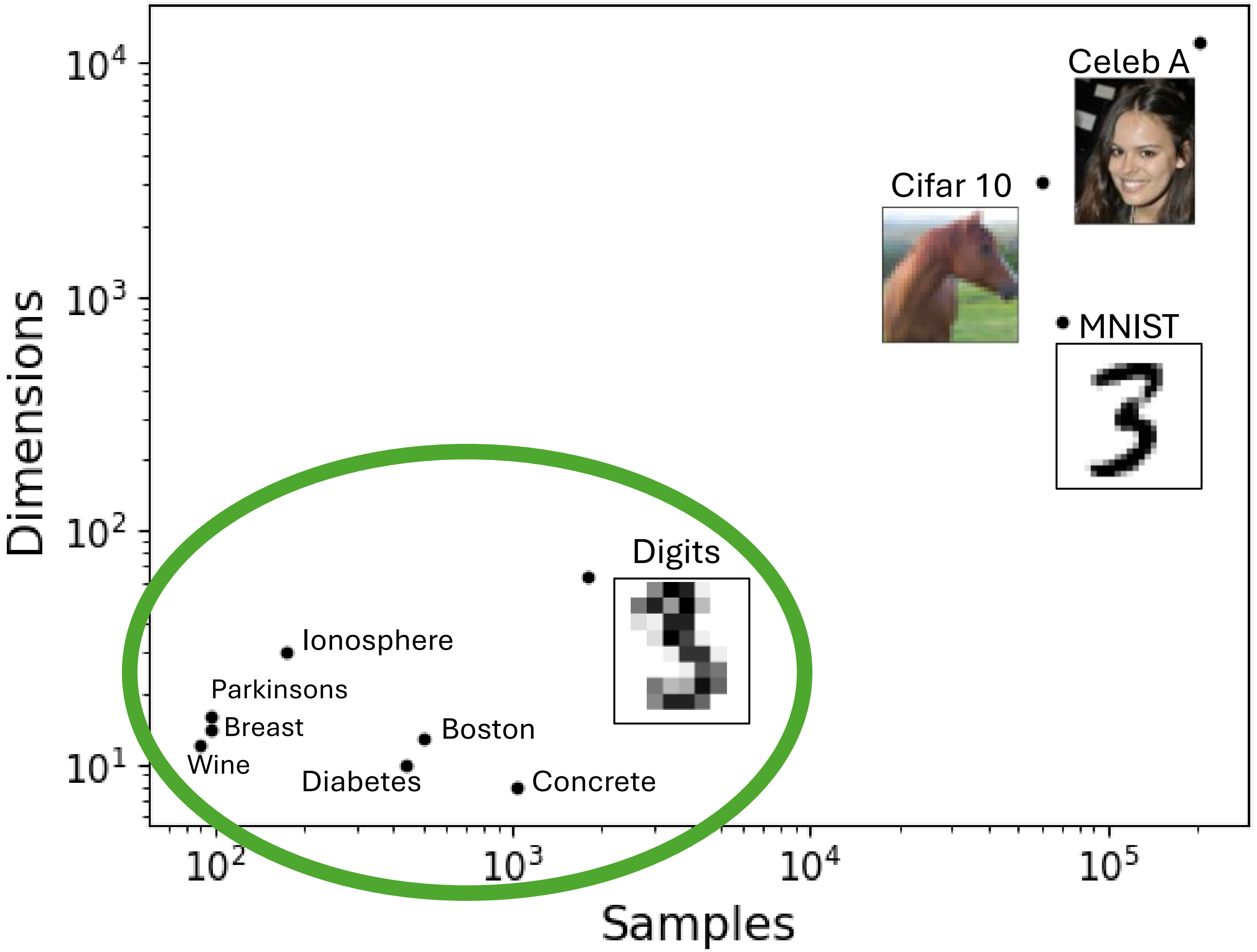
University of Warwick

Density estimation with small and high-dimensional data



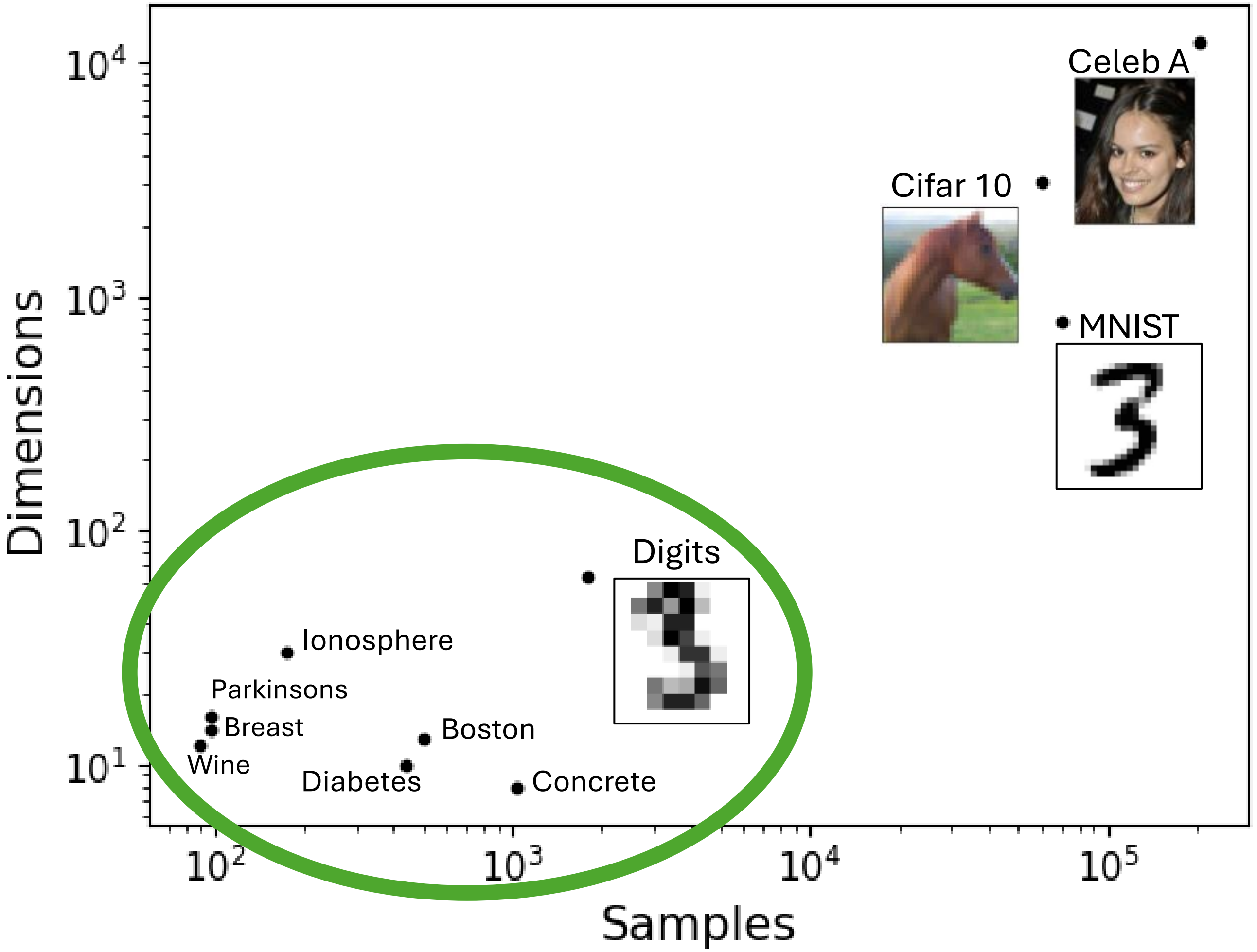
Density estimation with small and high-dimensional data

- **<1000 train samples**
- **10-64 Dimensions**



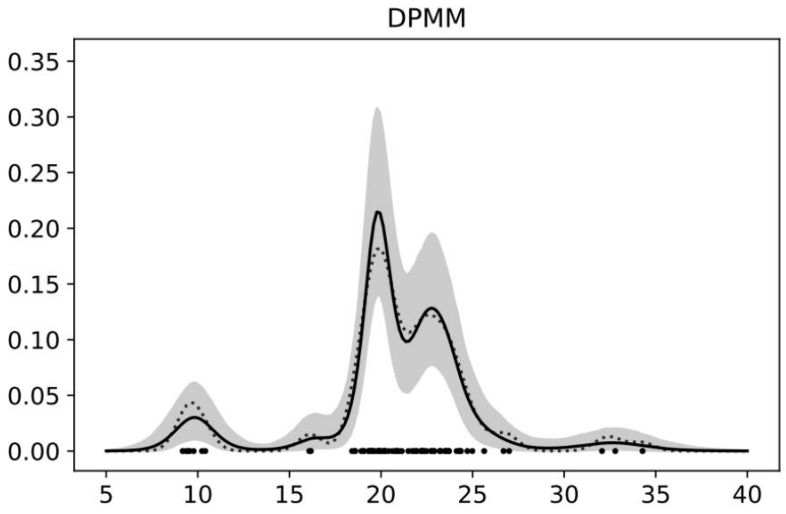
Density estimation with small and high-dimensional data

- **<1000 train samples**
- **10-64 Dimensions**
- **Analytical density function**
- **Able to sample**



Non-parametric Bayesian prediction

Non-parametric Bayesian prediction



$$p_{(n)}(x|x_{1:n})$$

Predictive density

Non-parametric Bayesian prediction

$$\pi_{(n-1)}(\theta|x_{1:n-1})$$

Posterior density

$$p_{(n)}(x|x_{1:n})$$

Predictive density

Non-parametric Bayesian prediction

$$\pi_{(n-1)}(\theta | x_{1:n-1})$$

Posterior density

1-step updates
via MCMC

$$p_{(n)}(x | x_{1:n})$$

Predictive density

Non-parametric Bayesian prediction

Expensive and slow with the dimension!

$$\pi_{(n-1)}(\theta|x_{1:n-1})$$

Posterior density

1-step updates via MCMC

$$p_{(n)}(x|x_{1:n})$$

Predictive density

Recursive Quasi-Bayesian prediction

Hahn *et. al* (2018) + Fong *et. al* (2023): Recursive Bayesian Predictive (**R-BP**):

$$p_{(n)}(x|x_{1:n}) = p_{(n-1)}(x|x_{1:n-1}) \cdot$$ Update term

Recursive Quasi-Bayesian prediction

Hahn *et. al* (2018) + Fong *et. al* (2023): Recursive Bayesian Predictive (**R-BP**):

$$p_{(n)}(x|x_{1:n}) = p_{(n-1)}(x|x_{1:n-1}) \cdot$$

New predictive density

Old predictive density

Update term

Recursive Quasi-Bayesian prediction

Hahn *et. al* (2018) + Fong *et. al* (2023): Recursive Bayesian Predictive (**R-BP**):

$$p_{(n)}(x|x_{1:n}) = p_{(n-1)}(x|x_{1:n-1}) \cdot c_{(n)}\left(P_{(n-1)}(x), P_{(n-1)}(x_n)\right)$$

New predictive density | Old predictive density | Bivariate copula update

✓ **No need for MCMC**

Recursive Quasi-Bayesian prediction

Hahn *et. al* (2018) + Fong *et. al* (2023): Recursive Bayesian Predictive (**R-BP**):

$$p_{(n)}(x|x_{1:n}) = p_{(n-1)}(x|x_{1:n-1}) \cdot c_{(n)}\left(P_{(n-1)}(x), P_{(n-1)}(x_n)\right)$$

New predictive density — Old predictive density — Bivariate copula update

- **No need for MCMC**
- **Nonparametric**
- **Quasi-Bayesian** (nice Bayesian properties)
- **Very fast density evaluation and sampling**

Recursive Quasi-Bayesian prediction

Hahn *et. al* (2018) + Fong *et. al* (2023): Recursive Bayesian Predictive (**R-BP**):

$$p_{(n)}(x|x_{1:n}) = p_{(n-1)}(x|x_{1:n-1}) \cdot c_{(n)}\left(P_{(n-1)}(x), P_{(n-1)}(x_n)\right)$$

New predictive density | Old predictive density | Bivariate copula update

- **No need for MCMC**
- **Nonparametric**
- **Quasi-Bayesian** (nice Bayesian properties)
- **Very fast density evaluation and sampling**

A Predictive Approach to Bayesian Nonparametric Survival Analysis
Edwin Fong, University of Oxford; Brieuc Lehmann, University College London
AISTATS 22

MARTINGALE POSTERIOR NEURAL PROCESSES
Hyungi Lee[1], Eunggu Yun[1], Giung Nam[1], Edwin Fong[2], Juho Lee[1,3]
[1]KAIST, [2]Novo Nordisk, [3]AITRICS
[1]{lhk2708, eunggu.yun, giung, juholee}@kaist.ac.kr, [2]chef@novonordisk.com
ICLR 23

Quasi-Bayesian Nonparametric Density Estimation via Autoregressive Predictive Updates
Sahra Ghalebikesabi[1], Chris Holmes[2], Edwin Fong[*2], Brieuc Lehmann[*3]
[1]University of Oxford, [2]Novo Nordisk, [3]University College London
UAI 23

Is In-Context Learning in Large Language Models Bayesian? A Martingale Perspective
Fabian Falck[*1], Ziyu Wang[*1], Chris Holmes[1]
ICML 24

On Uncertainty Quantification for Near-Bayes Optimal Algorithms
Ziyu Wang, Chris Holmes, Department of Statistics, University of Oxford
WZY196@GMAIL.COM, CHOLMES@STATS.OX.AC.UK
AABI 24

Posterior Uncertainty Quantification in Neural Networks using Data Augmentation
Luhuan Wu[1], Columbia University; Sinead A. Williamson, Apple Machine Learning Research
AISTATS 24

Recursive Quasi-Bayesian prediction

Hahn *et. al* (2018) + Fong *et. al* (2023): Recursive Bayesian Predictive (**R-BP**):

$$p_{(n)}(x|x_{1:n}) = p_{(n-1)}(x|x_{1:n-1}) \cdot c_{(n)}\left(P_{(n-1)}(x), P_{(n-1)}(x_n)\right)$$

New predictive density | **Old predictive density** | **Bivariate copula update**

- ✓ **No need for MCMC**
- ✓ **Nonparametric**
- ✓ **Quasi-Bayesian** (nice Bayesian properties)
- ✓ **Very fast density evaluation and sampling**

- ✖ **Extensions to multivariate settings are non-trivial**
- ✖ **Restrictive assumptions on dependence structure**
- ✖ **Computed sequentially with the dimension**

Our solution: even more copulas!

Use **Sklar's Theorem** to split the joint predictive density:

$$\mathbf{p}_{(n)}(x^1, \ldots, x^d)$$

Our solution: even more copulas!

Use **Sklar's Theorem** to split the joint predictive density:

$$\underbrace{\mathbf{p}_{(n)}(x^1, \ldots, x^d)}_{\textbf{Joint}} = p^1_{(n)}(x^1) \cdot \ldots \cdot p^d_{(n)}(x^d) \cdot \mathbf{c}_{(n)}(P^1_{(n)}(x^1), \ldots, P^d_{(n)}(x^d))$$

Our solution: even more copulas!

Use **Sklar's Theorem** to split the joint predictive density:

$$\underbrace{\mathbf{p}_{(n)}(x^1, \ldots, x^d)}_{\text{Joint}} = \underbrace{p^1_{(n)}(x^1)}_{\text{Marginal}} \cdot \ldots \cdot \underbrace{p^d_{(n)}(x^d)}_{\text{Marginal}} \cdot \mathbf{c}_{(n)}(P^1_{(n)}(x^1), \ldots, P^d_{(n)}(x^d))$$

Our solution: even more copulas!

Use **Sklar's Theorem** to split the joint predictive density:

$$\underbrace{\mathbf{p}_{(n)}(x^1, \ldots, x^d)}_{\text{Joint}} = \underbrace{p^1_{(n)}(x^1)}_{\text{Marginal}} \cdot \ldots \cdot \underbrace{p^d_{(n)}(x^d)}_{\text{Marginal}} \cdot \underbrace{\mathbf{c}_{(n)}(P^1_{(n)}(x^1), \ldots, P^d_{(n)}(x^d))}_{\text{High-dimensional copula}}$$

Our solution: even more copulas!

Use **Sklar's Theorem** to split the joint predictive density:

$$\underbrace{\mathbf{p}_{(n)}(x^1,\dots,x^d)}_{\text{Joint}} = \underbrace{p^1_{(n)}(x^1)}_{\text{Marginal}} \cdot \ldots \cdot \underbrace{p^d_{(n)}(x^d)}_{\text{Marginal}} \cdot \underbrace{\mathbf{c}_{(n)}(P^1_{(n)}(x^1),\dots,P^d_{(n)}(x^d))}_{\text{High-dimensional copula}}$$

Obtain **simple** recursive update:

$$\frac{\mathbf{p}_{(m)}(\mathbf{x})}{\mathbf{p}_{(m-1)}(\mathbf{x})} =$$

Update term

Our solution: even more copulas!

Use **Sklar's Theorem** to split the joint predictive density:

$$\underbrace{\mathbf{p}_{(n)}(x^1,\ldots,x^d)}_{\text{Joint}} = \underbrace{p^1_{(n)}(x^1)}_{\text{Marginal}} \cdot \ldots \cdot \underbrace{p^d_{(n)}(x^d)}_{\text{Marginal}} \cdot \underbrace{\mathbf{c}_{(n)}(P^1_{(n)}(x^1),\ldots,P^d_{(n)}(x^d))}_{\text{High-dimensional copula}}$$

Obtain **simple** recursive update:

$$\frac{\mathbf{p}_{(m)}(\mathbf{x})}{\mathbf{p}_{(m-1)}(\mathbf{x})} = \underbrace{\prod_{i=1}^{d}\left\{\frac{p^i_{(m)}(x^i)}{p^i_{(m-1)}(x^i)}\right\}}_{\text{Independent recursions}} \cdot \underbrace{\frac{\mathbf{c}_{(m)}\left(P^1_{(m)}(x^1),\ldots,P^d_{(m)}(x^d)\right)}{\mathbf{c}_{(m-1)}\left(P^1_{(m-1)}(x^1),\ldots,P^d_{(m-1)}(x^d)\right)}}_{\text{recursion on copulas}}$$

Benefits of even more copulas

$$\frac{\mathbf{p}_{(m)}(\mathbf{x})}{\mathbf{p}_{(m-1)}(\mathbf{x})} = \underbrace{\prod_{i=1}^{d}\left\{\frac{p^i_{(m)}(x^i)}{p^i_{(m-1)}(x^i)}\right\}}_{\text{Independent recursions}} \cdot \underbrace{\frac{\mathbf{c}_{(m)}\left(P^1_{(m)}(x^1),\ldots,P^d_{(m)}(x^d)\right)}{\mathbf{c}_{(m-1)}\left(P^1_{(m-1)}(x^1),\ldots,P^d_{(m-1)}(x^d)\right)}}_{\text{recursion on copulas}}$$

Benefits of even more copulas

$$\frac{\mathbf{p}_{(m)}(\mathbf{x})}{\mathbf{p}_{(m-1)}(\mathbf{x})} = \underbrace{\prod_{i=1}^{d}\left\{\frac{p^i_{(m)}(x^i)}{p^i_{(m-1)}(x^i)}\right\}}_{\text{Independent recursions}} \cdot \underbrace{\frac{\mathbf{c}_{(m)}\left(P^1_{(m)}(x^1),\ldots,P^d_{(m)}(x^d)\right)}{\mathbf{c}_{(m-1)}\left(P^1_{(m-1)}(x^1),\ldots,P^d_{(m-1)}(x^d)\right)}}_{\text{recursion on copulas}}$$

Known univariate R-BP recursion

$$p_{(n)}(x|x_{1:n}) = p_{(n-1)}(x|x_{1:n-1}) \cdot c_{(n)}\left(P_{(n-1)}(x), P_{(n-1)}(x_n)\right)$$

New predictive density — Old predictive density — Bivariate copula update

✓ Very **simple** and **fast**

Benefits of even more copulas

$$\frac{\mathbf{p}_{(m)}(\mathbf{x})}{\mathbf{p}_{(m-1)}(\mathbf{x})} = \underbrace{\prod_{i=1}^{d} \left\{ \frac{p^i_{(m)}(x^i)}{p^i_{(m-1)}(x^i)} \right\}}_{\text{Independent recursions}} \cdot \underbrace{\frac{\mathbf{c}_{(m)}\left(P^1_{(m)}(x^1), \ldots, P^d_{(m)}(x^d)\right)}{\mathbf{c}_{(m-1)}\left(P^1_{(m-1)}(x^1), \ldots, P^d_{(m-1)}(x^d)\right)}}_{\text{recursion on copulas}}$$

Known univariate R-BP recursion

$$\underbrace{p_{(n)}(x|x_{1:n})}_{\text{New predictive density}} = \underbrace{p_{(n-1)}(x|x_{1:n-1})}_{\text{Old predictive density}} \cdot \underbrace{c_{(n)}\left(P_{(n-1)}(x), P_{(n-1)}(x_n)\right)}_{\text{Bivariate copula update}}$$

- ✓ Very **simple** and **fast**
- ✓ Each **marginal** recursion is done in **parallel**

Benefits of even more copulas

$$\frac{\mathbf{p}_{(m)}(\mathbf{x})}{\mathbf{p}_{(m-1)}(\mathbf{x})} = \underbrace{\prod_{i=1}^{d}\left\{\frac{p^i_{(m)}(x^i)}{p^i_{(m-1)}(x^i)}\right\}}_{\text{Independent recursions}} \cdot \underbrace{\frac{\mathbf{c}_{(m)}\left(P^1_{(m)}(x^1),\ldots,P^d_{(m)}(x^d)\right)}{\mathbf{c}_{(m-1)}\left(P^1_{(m-1)}(x^1),\ldots,P^d_{(m-1)}(x^d)\right)}}_{\text{Implicit recursion on copulas}}$$

Known univariate R-BP recursion

$$\underbrace{p_{(n)}(x|x_{1:n})}_{\text{New predictive density}} = \underbrace{p_{(n-1)}(x|x_{1:n-1})}_{\text{Old predictive density}} \cdot \underbrace{c_{(n)}\left(P_{(n-1)}(x), P_{(n-1)}(x_n)\right)}_{\text{Bivariate copula update}}$$

✓ Very **simple** and **fast**

✓ Each **marginal** recursion is done in **parallel**

- Only interested in the **final** predictive density
- Copula recursion is left implicit.

✓ Only fit a **single** copula at **final** step
Use a **vine copula model**

Benefits of even more copulas

$$\frac{\mathbf{p}_{(m)}(\mathbf{x})}{\mathbf{p}_{(m-1)}(\mathbf{x})} = \underbrace{\prod_{i=1}^{d}\left\{\frac{p^i_{(m)}(x^i)}{p^i_{(m-1)}(x^i)}\right\}}_{\text{Independent recursions}} \cdot \underbrace{\frac{\mathbf{c}_{(m)}\left(P^1_{(m)}(x^1), \ldots, P^d_{(m)}(x^d)\right)}{\mathbf{c}_{(m-1)}\left(P^1_{(m-1)}(x^1), \ldots, P^d_{(m-1)}(x^d)\right)}}_{\text{Implicit recursion on copulas}}$$

Known univariate R-BP recursion

$$\underbrace{p_{(n)}(x|x_{1:n})}_{\text{New predictive density}} = \underbrace{p_{(n-1)}(x|x_{1:n-1})}_{\text{Old predictive density}} \cdot \underbrace{c_{(n)}\left(P_{(n-1)}(x), P_{(n-1)}(x_n)\right)}_{\text{Bivariate copula update}}$$

✓ Very **simple** and **fast**

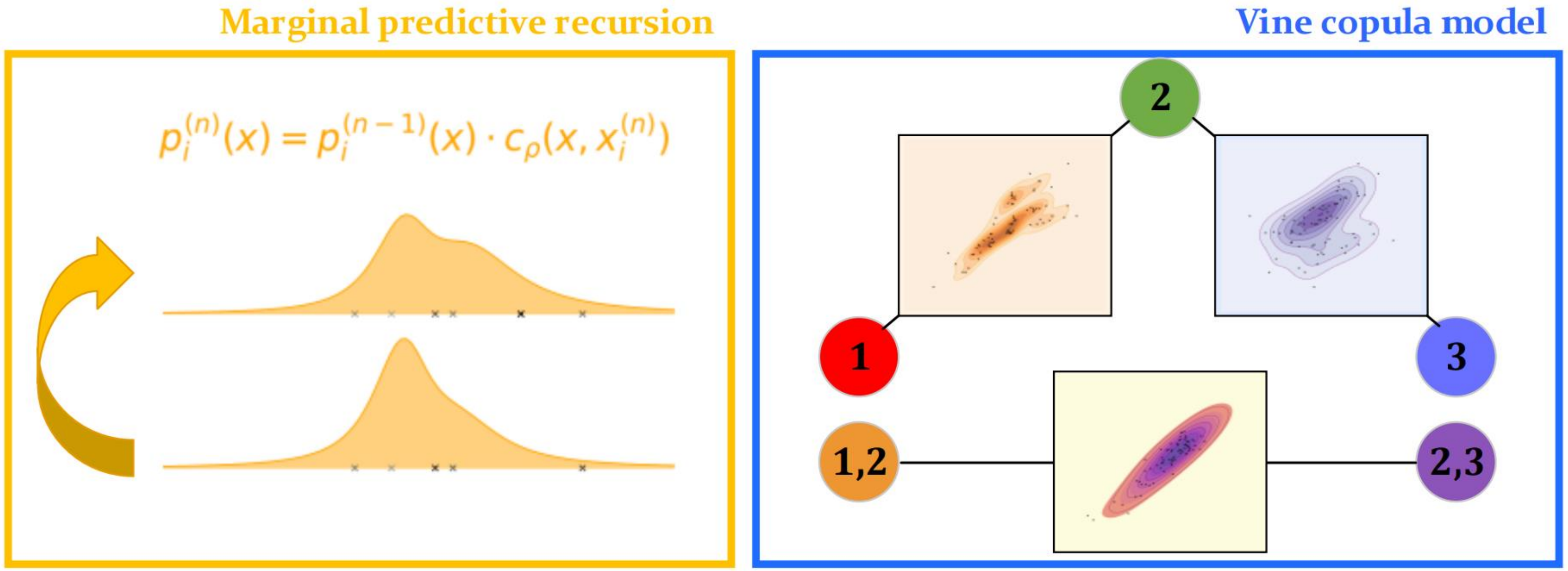
✓ Each **marginal** recursion is done in **parallel**

- Only interested in the **final** predictive density
- Copula recursion is left implicit.

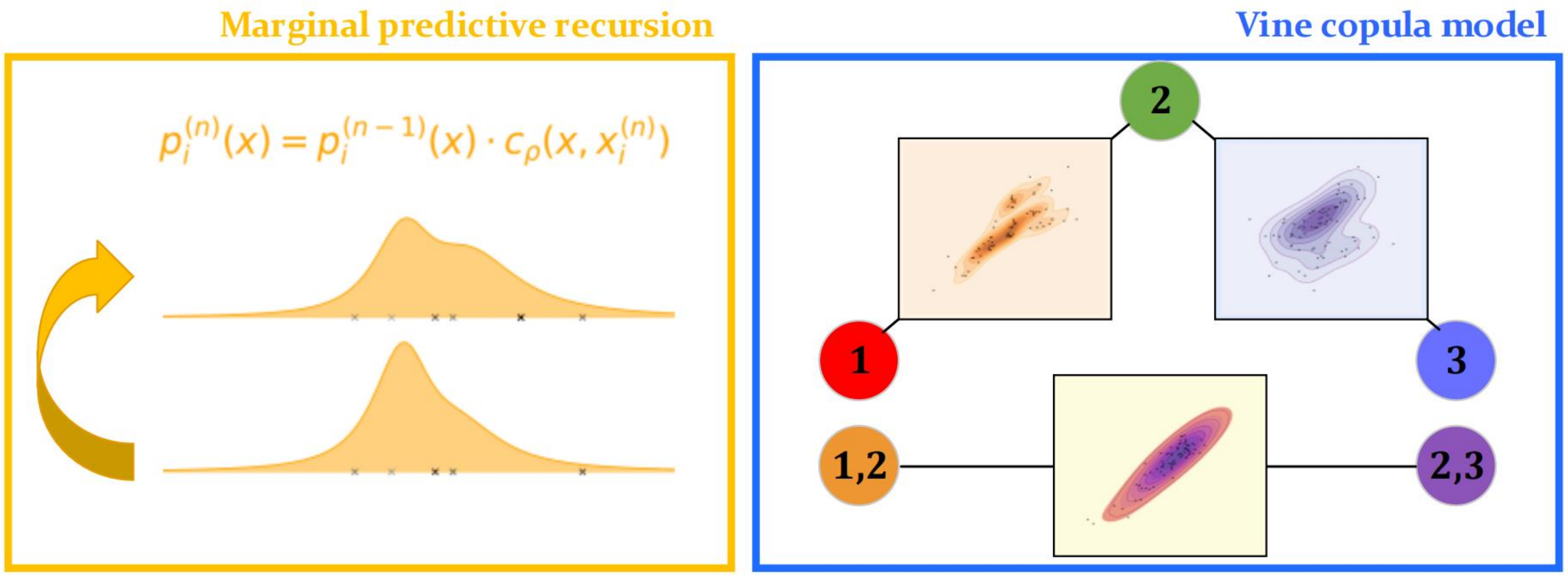
✓ Only fit a **single** copula at **final** step
Use a **vine copula model**

✓ **No restrictive assumptions needed!**

Final model: The Quasi-Bayesian Vine

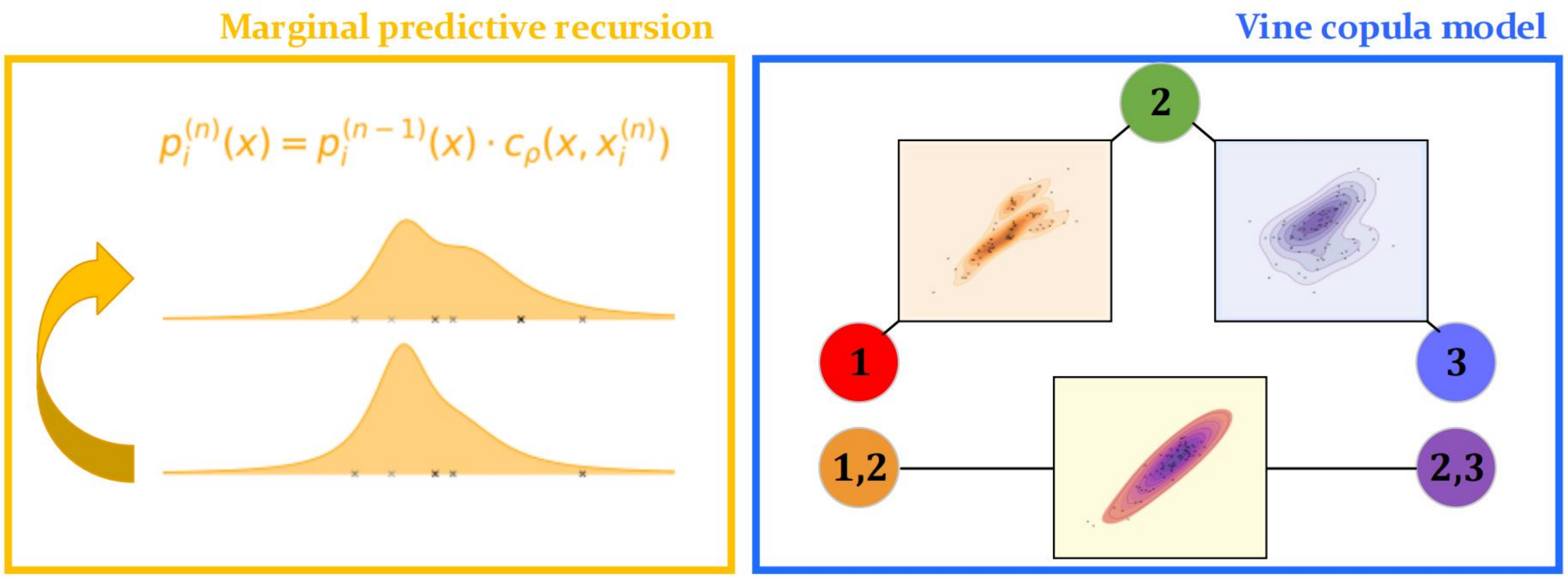


Final model: The Quasi-Bayesian Vine



Marginal convergence rate: $\mathcal{O}_p\left(n^{-1/2}\right)$

Final model: The Quasi-Bayesian Vine



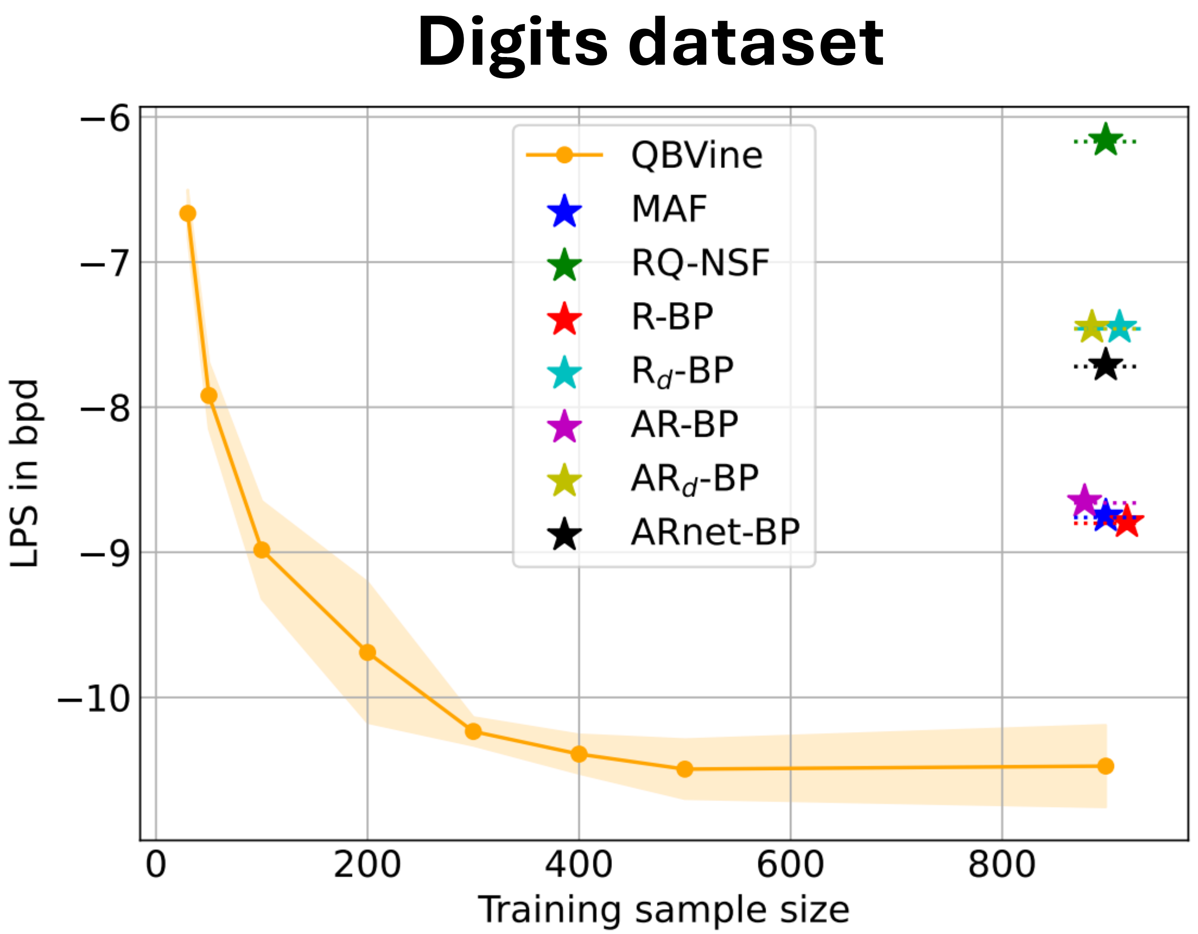
Marginal convergence rate: $\mathcal{O}_p\left(n^{-1/2}\right)$

Copula convergence rate:

$$\mathcal{O}_p(n^{-1/3})$$

for certain dependence structures only.

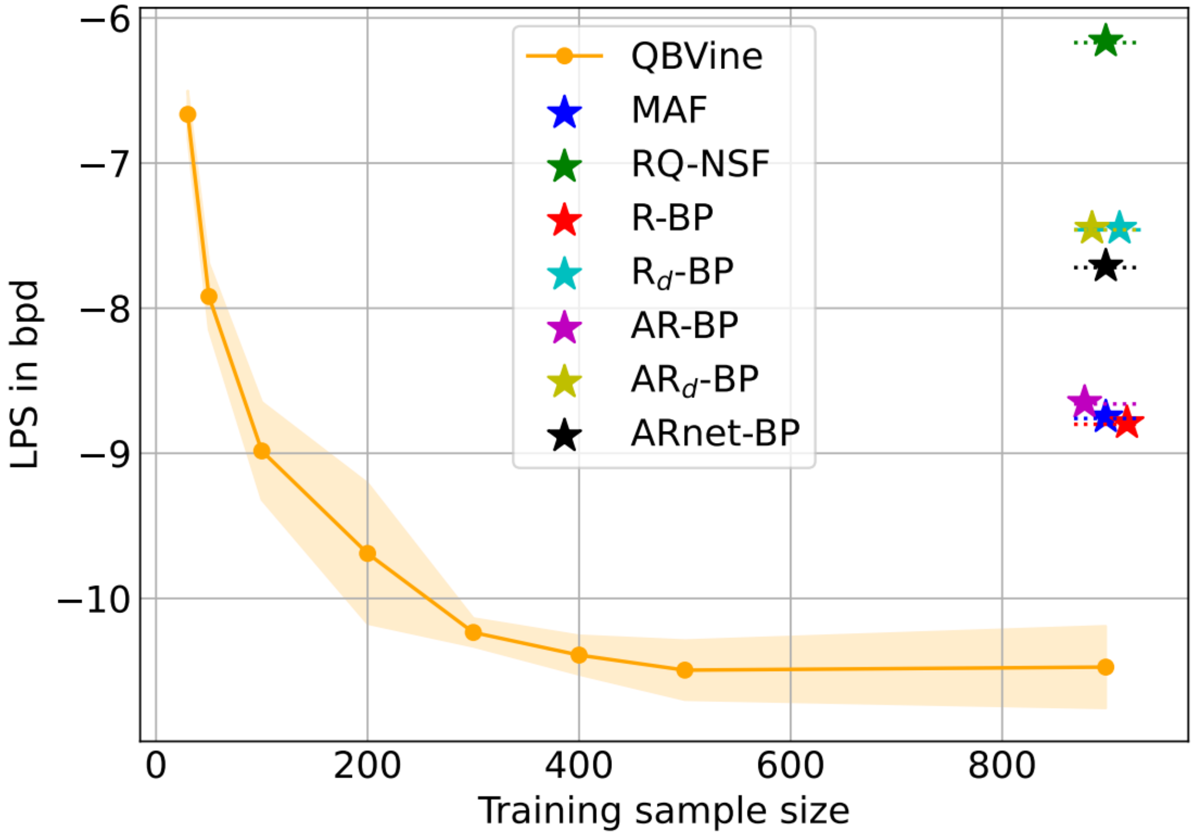
Experiments



Our paper

Experiments

Digits dataset



Regression

n/d	WINE 89/12	BREAST 97/14	PARKIN 97/16	IONO 175/30	BOSTON 506/13
KDE	$13.69_{\pm 0.00}$	$10.45_{\pm 0.24}$	$12.83_{\pm 0.27}$	$32.06_{\pm 0.00}$	$8.34_{\pm 0.00}$
DPMM (Diag)	$17.46_{\pm 0.60}$	$16.26_{\pm 0.71}$	$22.28_{\pm 0.66}$	$35.30_{\pm 1.28}$	$7.64_{\pm 0.09}$
DPMM (Full)	$32.88_{\pm 0.82}$	$26.67_{\pm 1.32}$	$39.95_{\pm 1.56}$	$86.18_{\pm 10.22}$	$9.45_{\pm 0.43}$
MAF	$39.60_{\pm 1.41}$	$10.13_{\pm 0.40}$	$11.76_{\pm 0.45}$	$140.09_{\pm 4.03}$	$56.01_{\pm 27.74}$
RQ-NSF	$38.34_{\pm 0.63}$	$26.41_{\pm 0.57}$	$31.26_{\pm 0.31}$	$54.49_{\pm 0.65}$	$-2.20_{\pm 0.11}$
PRticle Filter	$23.89,_{\pm 0.93}$	$25.98_{\pm 1.06}$	$34.79_{\pm 3.95}$	$79.22_{\pm 9.87}$	$27.18_{\pm 3.12}$
R-BP	$13.57_{\pm 0.04}$	$7.45_{\pm 0.02}$	$9.15_{\pm 0.04}$	$21.15_{\pm 0.04}$	$4.56_{\pm 0.04}$
R$_d$-BP	$13.32_{\pm 0.01}$	$6.12_{\pm 0.05}$	$7.52_{\pm 0.05}$	$19.82_{\pm 0.08}$	$-13.50_{\pm 0.59}$
AR-BP	$13.45_{\pm 0.05}$	$6.18_{\pm 0.05}$	$8.29_{\pm 0.11}$	$17.16_{\pm 0.25}$	$-0.45_{\pm 0.77}$
AR$_d$-BP	$\mathbf{13.22}_{\pm \mathbf{0.04}}$	$6.11_{\pm 0.04}$	$7.21_{\pm 0.12}$	$16.48_{\pm 0.26}$	$-14.75_{\pm 0.89}$
ARnet-BP	$14.41_{\pm 0.11}$	$6.87_{\pm 0.23}$	$8.29_{\pm 0.17}$	$15.32_{\pm 0.35}$	$-5.71_{\pm 0.62}$
QB-Vine	$13.76_{\pm 0.13}$	$\mathbf{4.67}_{\pm \mathbf{0.31}}$	$\mathbf{4.93}_{\pm \mathbf{0.20}}$	$\mathbf{-16.08}_{\pm \mathbf{2.12}}$	$\mathbf{-31.04}_{\pm \mathbf{1.02}}$

Classification

	Regression			Classification	
n/d	BOSTON 506/13	CONCR 1,030/8	DIAB 442/10	IONO 351/33	PARKIN 195/22
Linear	$0.87_{\pm 0.03}$	$0.99_{\pm 0.01}$	$1.07_{\pm 0.01}$	$0.33_{\pm 0.01}$	$0.38_{\pm 0.01}$
GP	$0.42_{\pm 0.08}$	$0.36_{\pm 0.02}$	$1.06_{\pm 0.02}$	$0.30_{\pm 0.02}$	$0.42_{\pm 0.02}$
MLP	$1.42_{\pm 1.01}$	$2.01_{\pm 0.98}$	$3.32_{\pm 4.05}$	$0.26_{\pm 0.05}$	$0.31_{\pm 0.02}$
R-BP	$0.76_{\pm 0.09}$	$0.87_{\pm 0.03}$	$1.05_{\pm 0.03}$	$0.26_{\pm 0.01}$	$0.37_{\pm 0.01}$
R$_d$-BP	$0.40_{\pm 0.03}$	$0.42_{\pm 0.00}$	$1.00_{\pm 0.02}$	$0.34_{\pm 0.02}$	$0.27_{\pm 0.03}$
AR-BP	$0.52_{\pm 0.13}$	$0.42_{\pm 0.01}$	$1.06_{\pm 0.02}$	$0.21_{\pm 0.02}$	$0.29_{\pm 0.02}$
AR$_d$-BP	$0.37_{\pm 0.10}$	$0.39_{\pm 0.01}$	$0.99_{\pm 0.02}$	$0.20_{\pm 0.02}$	$0.28_{\pm 0.03}$
ARnet-BP	$0.45_{\pm 0.11}$	$\mathbf{-0.03}_{\pm \mathbf{0.00}}$	$1.41_{\pm 0.07}$	$0.24_{\pm 0.04}$	$0.26_{\pm 0.04}$
QB-Vine	$\mathbf{-0.81}_{\pm \mathbf{1.26}}$	$0.54_{\pm 0.34}$	$\mathbf{0.87}_{\pm \mathbf{0.20}}$	$\mathbf{-1.85}_{\pm \mathbf{1.16}}$	$\mathbf{-0.76}_{\pm \mathbf{0.28}}$

Summary and future directions

Drop by **Poster Session 6,** Friday 13th
Chat & collaborate!

Summary and future directions

In a Quasi-Bayesian framework,
copulas are a useful tool for obtaining:

- More **general** models.
- Suitable for **parallelisation.**
- Effective on **high-dimensional data**.

Drop by **Poster Session 6,** Friday 13th
Chat & collaborate!

Summary and future directions

In a Quasi-Bayesian framework,
copulas are a useful tool for obtaining:

- More **general** models.
- Suitable for **parallelisation.**
- Effective on **high-dimensional data**.

Drop by **Poster Session 6,** Friday 13^{th}
Chat & collaborate!

Future directions include:

- Using more **effective copula** models.
- Applying the QB-Vine on dependent data such as in **time series, weather** and **RL**.
- Integrate **new** Quasi-Bayesian methods.

Summary and future directions

In a Quasi-Bayesian framework,
copulas are a useful tool for obtaining:

- More **general** models.
- Suitable for **parallelisation.**
- Effective on **high-dimensional data**.

Future directions include:

- Using more **effective copula** models.
- Applying the QB-Vine on dependent data such as in **time series, weather** and **RL**.
- Integrate **new** Quasi-Bayesian methods.

Drop by **Poster Session 6,** Friday 13th
Chat & collaborate!

Paper and more!

https://warwick.ac.uk/fac/sci/statistics/staff/research_students/huk