Möbius inversion meets tensors: inference by Edgeworth series

Stephen Huan

Carnegie Mellon University
https://cgdct.moe

SIAM MDS24

Collaborators



Andreas Robertson,
Sandia



Youssef Marzouk,
MIT



Florian Schäfer,
Gatech



Overview

Introduction
- Knothe-Rosenblatt transport
- Cumulants

Edgeworth series

Cornish-Fisher
- Warmup: quantile matching

Multivariate procedure

Experimental results

Conclusion

A motivating question

How to sample from a multivariate Gaussian?

Multivariate Gaussian example

Find *transport map* pushing $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathrm{Id})$ to $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Theta)$

Multivariate Gaussian example

Find *transport map* pushing $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathrm{Id})$ to $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Theta)$

Match mean and covariance, consider ansatz $T(\boldsymbol{z}) = L\boldsymbol{z} + \boldsymbol{\mu}$

$$\mathbb{E}[T(\boldsymbol{z})] = \boldsymbol{\mu}$$
$$\mathbb{C}\mathrm{ov}[T(\boldsymbol{z})] = LL^\top$$

Multivariate Gaussian example

Find *transport map* pushing $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathrm{Id})$ to $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Theta)$

Match mean and covariance, consider ansatz $T(\boldsymbol{z}) = L\boldsymbol{z} + \boldsymbol{\mu}$

$$\mathbb{E}[T(\boldsymbol{z})] = \boldsymbol{\mu}$$
$$\mathbb{C}\mathrm{ov}[T(\boldsymbol{z})] = LL^\top$$

Pick L such that $\Theta = LL^\top$ (e.g. Cholesky factor)

This talk

How to sample from a *general* distribution?

Problem setup

Throughout: data X has density f_X (only through samples)
reference distribution Z has density f_Z (e.g. $\mathcal{N}(0, \mathrm{Id}_d)$)

Problem setup

Throughout: data X has density f_X (only through samples)
reference distribution Z has density f_Z (e.g. $\mathcal{N}(0, \mathrm{Id}_d)$)

Assuming f_Z and f_X smooth, infinitely many valid T

(VAE, GAN, normalizing flow, score-based/diffusion models)

Knothe-Rosenblatt transport

Restrict to *triangular* maps in the sense that

$$T(\boldsymbol{x}) = \begin{bmatrix} T_1(x_1) \\ T_2(x_1, x_2) \\ \vdots \\ T_d(x_1, \ldots, x_d) \end{bmatrix}$$

and control the sign of the diagonal by enforcing $\nabla T \succ 0$

Knothe-Rosenblatt transport

Restrict to *triangular* maps in the sense that

$$T(\boldsymbol{x}) = \begin{bmatrix} T_1(x_1) \\ T_2(x_1, x_2) \\ \vdots \\ T_d(x_1, \ldots, x_d) \end{bmatrix}$$

and control the sign of the diagonal by enforcing $\nabla T \succ 0$

Unique transport called the *Knothe-Rosenblatt rearrangement*

Knothe-Rosenblatt transport

Restrict to *triangular* maps in the sense that

$$T(\boldsymbol{x}) = \begin{bmatrix} T_1(x_1) \\ T_2(x_1, x_2) \\ \vdots \\ T_d(x_1, \ldots, x_d) \end{bmatrix}$$

and control the sign of the diagonal by enforcing $\nabla T \succ 0$

Unique transport called the *Knothe-Rosenblatt rearrangement*

Nonlinear generalization of the Cholesky factor

Computational benefits

Immediately gives an efficient sampler $z \sim \mathcal{N}(0, \mathsf{Id}),\ T(z) \sim X$

Computational benefits

Immediately gives an efficient sampler $z \sim \mathcal{N}(0, \mathsf{Id}),\ T(z) \sim X$

Density evaluation through the pushforward

$$f_X = T_\sharp f_Z := (f_Z \circ T^{-1}) |\det \nabla T^{-1}|$$

with efficient access to T^{-1} and $\det \nabla T$

Computational benefits

Immediately gives an efficient sampler $z \sim \mathcal{N}(0, \mathsf{Id}),\ T(z) \sim X$

Density evaluation through the pushforward

$$f_X = T_\sharp f_Z := (f_Z \circ T^{-1}) |\det \nabla T^{-1}|$$

with efficient access to T^{-1} and $\det \nabla T$

Compute T^{-1} by recursive substitution

Transport from samples

Previous work [Marzouk et al. 2016; Spantini, Bigoni, and Marzouk 2018; Katzfuss and Schäfer 2022]:
minimize empirical KL over parametric family

Transport from samples

Previous work [Marzouk et al. 2016; Spantini, Bigoni, and Marzouk 2018; Katzfuss and Schäfer 2022]:
minimize empirical KL over parametric family

This talk: use *cumulants*

Transport from samples

Previous work [Marzouk et al. 2016; Spantini, Bigoni, and Marzouk 2018; Katzfuss and Schäfer 2022]:
minimize empirical KL over parametric family

This talk: use *cumulants*

Motivations:

- Analogy to multivariate Gaussians (mean, covariance)
- Central limit theorem, decay of higher order cumulants
- Interpretability and material science (2nd order interactions)

Material microstructure

2-point statistics insufficient to capture structure (e.g. edges)

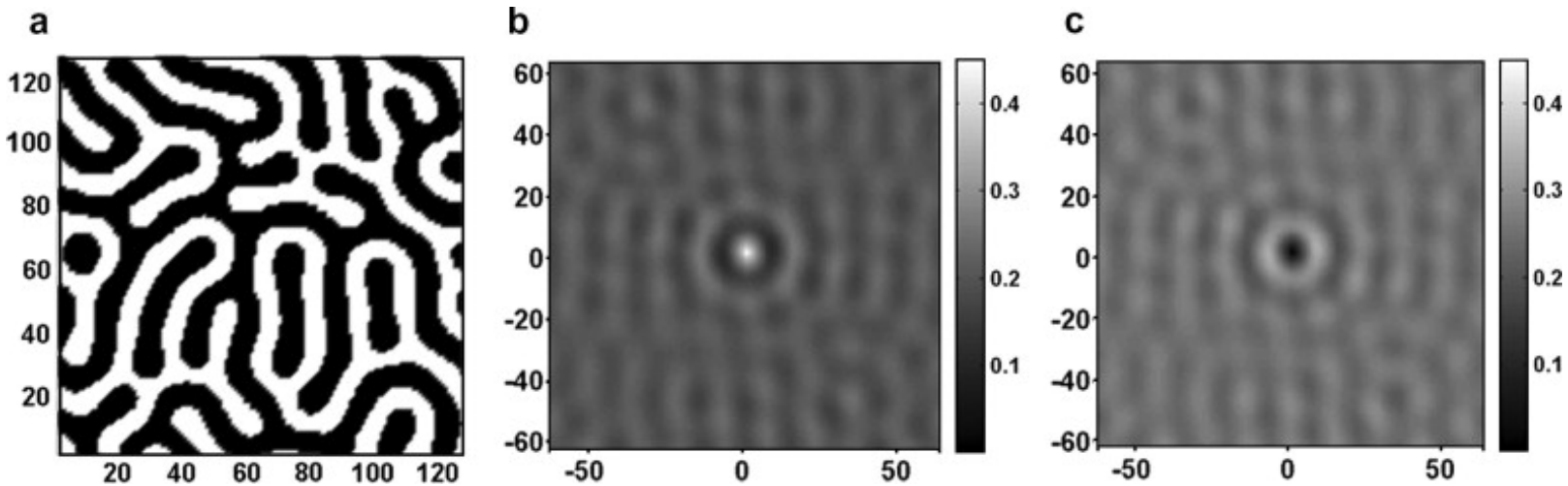


Figure: Fig. 1. of [Fullwood, Niezgoda, and Kalidindi 2008]

This talk

Framework of [Cornish and Fisher 1938; McCullagh 2018]

This talk

Framework of [Cornish and Fisher 1938; McCullagh 2018]

Computational implementation of statistical formulas and "mathematical archaeology" of ancient methods

$$
\begin{aligned}
\xi - x = {} & a + \frac{1}{6} c\,(\xi^2 - 1) \\
& + \frac{1}{2} b\,\xi - \frac{1}{3} ac\,\xi + \frac{1}{24} d\,(\xi^3 - 3\,\xi) - \frac{1}{36} c^2\,(4\,\xi^3 - 7\,\xi) \\
& - \frac{1}{2} ab + \frac{1}{6} a^2 c - \frac{1}{12} bc\,(5\,\xi^2 - 3) - \frac{1}{8} ab\,(\xi^2 - 1) \\
& + \frac{1}{120} e\,(\xi^4 - 6\,\xi^2 + 3) + \frac{1}{36} ac^2\,(12\,\xi^2 - 7) \\
& - \frac{1}{144} cd\,(11\,\xi^4 - 42\,\xi^2 + 15) + \frac{1}{648} c^3\,(69\,\xi^4 - 187\,\xi^2 + 52) \\
& - \frac{3}{8} b^2 \xi + \frac{5}{6} abc\,\xi + \frac{1}{8} a^2 d\,\xi - \frac{1}{48} bd\,(7\,\xi^3 - 15\,\xi) - \frac{1}{30} ae\,(\xi^3 - \xi) \\
& + \frac{1}{720} f\,(\xi^5 - 10\,\xi^3 + 15\,\xi) - \frac{1}{3} a^2 c^2 \xi + \frac{1}{72} bc^2\,(36\,\xi^3 - 49\,\xi) \\
& - \frac{1}{384} d^2\,(5\,\xi^5 - 32\,\xi^3 + 35\,\xi) + \frac{1}{36} acd\,(11\,\xi^3 - 21\,\xi) \\
& - \frac{1}{360} ce\,(7\,\xi^5 - 48\,\xi^3 + 51\,\xi) - \frac{1}{324} ac^3\,(138\,\xi^3 - 187\,\xi) \\
& + \frac{1}{864} c^2 d\,(111\,\xi^5 - 547\,\xi^3 + 456\,\xi) - \frac{1}{7776} c^4\,(948\,\xi^5 - 3628\,\xi^3 + 2473\,\xi).
\end{aligned}
$$

This talk

Framework of [Cornish and Fisher 1938; McCullagh 2018]

Computational implementation of statistical formulas and "mathematical archaeology" of ancient methods

$$
\begin{aligned}
\xi - x = {} & a + \frac{1}{6} c\,(\xi^2 - 1) \\
& + \frac{1}{2} b\,\xi - \frac{1}{3} ac\,\xi + \frac{1}{24} d\,(\xi^3 - 3\,\xi) - \frac{1}{36} c^2\,(4\,\xi^3 - 7\,\xi) \\
& - \frac{1}{2} ab + \frac{1}{6} a^2 c - \frac{1}{12} bc\,(5\,\xi^2 - 3) - \frac{1}{8} ab\,(\xi^2 - 1) \\
& + \frac{1}{120} e\,(\xi^4 - 6\,\xi^2 + 3) + \frac{1}{36} ac^2\,(12\,\xi^2 - 7) \\
& - \frac{1}{144} cd\,(11\,\xi^4 - 42\,\xi^2 + 15) + \frac{1}{648} c^3\,(69\,\xi^4 - 187\,\xi^2 + 52) \\
& - \frac{3}{8} b^2 \xi + \frac{5}{6} abc\,\xi + \frac{1}{8} a^2 d\,\xi - \frac{1}{48} bd\,(7\,\xi^3 - 15\,\xi) - \frac{1}{30} ae\,(\xi^3 - \xi) \\
& + \frac{1}{720} f\,(\xi^5 - 10\,\xi^3 + 15\,\xi) - \frac{1}{3} a^2 c^2 \xi + \frac{1}{72} bc^2\,(36\,\xi^3 - 49\,\xi) \\
& - \frac{1}{384} d^2\,(5\,\xi^5 - 32\,\xi^3 + 35\,\xi) + \frac{1}{36} acd\,(11\,\xi^3 - 21\,\xi) \\
& - \frac{1}{360} ce\,(7\,\xi^5 - 48\,\xi^3 + 51\,\xi) - \frac{1}{324} ac^3\,(138\,\xi^3 - 187\,\xi) \\
& + \frac{1}{864} c^2 d\,(111\,\xi^5 - 547\,\xi^3 + 456\,\xi) - \frac{1}{7776} c^4\,(948\,\xi^5 - 3628\,\xi^3 + 2473\,\xi).
\end{aligned}
$$

Very much work in progress!

Moments

i-th *moment* captures product of i variables

$$\kappa^i := \mathbb{E}[X^i],\ \kappa^{ij} := \mathbb{E}[X^i X^j],\ \kappa^{ijk} := \mathbb{E}[X^i X^j X^k], \ldots$$

Moments

i-th *moment* captures product of i variables

$$\kappa^i := \mathbb{E}[X^i],\ \kappa^{ij} := \mathbb{E}[X^i X^j],\ \kappa^{ijk} := \mathbb{E}[X^i X^j X^k], \ldots$$

Estimated from samples with empirical expectation

Moments

i-th *moment* captures product of i variables

$$\kappa^i := \mathbb{E}[X^i],\ \kappa^{ij} := \mathbb{E}[X^i X^j],\ \kappa^{ijk} := \mathbb{E}[X^i X^j X^k], \ldots$$

Estimated from samples with empirical expectation

Moment generating function multivariate Taylor series

$$M_X(\xi) := \mathbb{E}[\exp(\xi_i X^i)] = 1 + \xi_i \kappa^i + \frac{\xi_i \xi_j \kappa^{ij}}{2!} + \frac{\xi_i \xi_j \xi_k \kappa^{ijk}}{3!} + \cdots$$

Cumulants

Cumulant generating function

$$K_X(\xi) := \log M_X(\xi) = \xi_i \kappa^i + \frac{\xi_i \xi_j \kappa^{i,j}}{2!} + \frac{\xi_i \xi_j \xi_k \kappa^{i,j,k}}{3!} + \cdots$$

Cumulants

Cumulant generating function

$$K_X(\xi) := \log M_X(\xi) = \xi_i \kappa^i + \frac{\xi_i \xi_j \kappa^{i,j}}{2!} + \frac{\xi_i \xi_j \xi_k \kappa^{i,j,k}}{3!} + \cdots$$

Match $M_X(\xi) = \exp(K_X(\xi)) = 1 + K_X(\xi) + K_X(\xi)^2/2! + \cdots$

$$\begin{aligned}
\kappa^i &= \kappa^i, \\
\kappa^{i,j} &= \kappa^{ij} - \kappa^i \kappa^j, \\
\kappa^{i,j,k} &= \kappa^{ijk} - \kappa^i \kappa^{jk} - \kappa^j \kappa^{ik} - \kappa^k \kappa^{ij} + 2\kappa^i \kappa^j \kappa^k \\
&\vdots
\end{aligned}$$

Properties of cumulants

Cumulant of sum of independent variables sum of cumulants

Properties of cumulants

Cumulant of sum of independent variables sum of cumulants

Gaussian only distribution with all $k > 2$ cumulants zero
Higher-order cumulants close to zero measure of Gaussianity?

Properties of cumulants

Cumulant of sum of independent variables sum of cumulants

Gaussian only distribution with all $k > 2$ cumulants zero
Higher-order cumulants close to zero measure of Gaussianity?

Independent random variables have zero joint cumulants
Approximate sparsity $\sim$ *conditional independence*

Properties of cumulants

Cumulant of sum of independent variables sum of cumulants

Gaussian only distribution with all $k > 2$ cumulants zero
Higher-order cumulants close to zero measure of Gaussianity?

Independent random variables have zero joint cumulants
Approximate sparsity $\sim$ *conditional independence*

Central limit theorem implies higher-order cumulants decay

Table of Contents

Introduction
- Knothe-Rosenblatt transport
- Cumulants

Edgeworth series

Cornish-Fisher
- Warmup: quantile matching

Multivariate procedure

Experimental results

Conclusion

Algorithmic recipe

Recipe for transport in 1-d with cumulants

1. Representing densities (Edgeworth series)

Algorithmic recipe

Recipe for transport in 1-d with cumulants

1. Representing densities (Edgeworth series)
2. Transport (Cornish-Fisher)

Pseudo-variable [representation]

Let X have cumulants κ and Z have cumulants λ

Pseudo-variable [representation]

Let X have cumulants κ and Z have cumulants λ

Ansatz $X = Z + Y$ with *pseudo*-variable Y independent of Z

Pseudo-variable [representation]

Let X have cumulants κ and Z have cumulants λ

Ansatz $X = Z + Y$ with *pseudo*-variable Y independent of Z

Y has "cumulants" $\eta^i, \eta^{i,j}, \ldots$ with "moments" $\eta^i, \eta^{ij}, \ldots$

$$\eta^i = \kappa^i - \lambda^i,\ \eta^{i,j} = \kappa^{i,j} - \lambda^{i,j},\ \eta^{i,j,k} = \kappa^{i,j,k} - \lambda^{i,j,k}, \ldots$$

Pseudo-variable [representation]

Let X have cumulants κ and Z have cumulants λ

Ansatz $X = Z + Y$ with *pseudo*-variable Y independent of Z

Y has "cumulants" $\eta^i, \eta^{i,j}, \ldots$ with "moments" $\eta^i, \eta^{ij}, \ldots$

$$\eta^i = \kappa^i - \lambda^i,\ \eta^{i,j} = \kappa^{i,j} - \lambda^{i,j},\ \eta^{i,j,k} = \kappa^{i,j,k} - \lambda^{i,j,k}, \ldots$$

Density of X is

$$f_X(x) = f_Z * f_Y = \int f_Z(x - y)\, f_Y(y)\, \mathrm{d}y = \mathbb{E}_Y[f_Z(x - Y)]$$

Edgeworth series [representation]

Taking a Taylor expansion at $Y = 0$ and averaging (or using term-by-term inversion of the moment generating function),

Edgeworth series [representation]

Taking a Taylor expansion at $Y = 0$ and averaging (or using term-by-term inversion of the moment generating function),

$$h_{i_1 \cdots i_n}(x) := \frac{(-1)^n}{f_Z(x)} \frac{\partial^n f_Z(x)}{\partial x^{i_1} \cdots \partial x^{i_n}}$$

$$f_X(x) = f_Z(x)(1 + \eta^i h_i(x) + \eta^{ij} h_{ij}(x)/2! + \eta^{ijk} h_{ijk}(x)/3! + \cdots)$$

Edgeworth series [representation]

Taking a Taylor expansion at $Y = 0$ and averaging (or using term-by-term inversion of the moment generating function),

$$h_{i_1 \cdots i_n}(x) := \frac{(-1)^n}{f_Z(x)} \frac{\partial^n f_Z(x)}{\partial x^{i_1} \cdots \partial x^{i_n}}$$

$$f_X(x) = f_Z(x)(1 + \eta^i h_i(x) + \eta^{ij} h_{ij}(x)/2! + \eta^{ijk} h_{ijk}(x)/3! + \cdots)$$

For $Z \sim \mathcal{N}(0, 1)$, h's are the Hermite polynomials

Edgeworth series [representation]

Taking a Taylor expansion at $Y = 0$ and averaging (or using term-by-term inversion of the moment generating function),

$$h_{i_1 \cdots i_n}(x) := \frac{(-1)^n}{f_Z(x)} \frac{\partial^n f_Z(x)}{\partial x^{i_1} \cdots \partial x^{i_n}}$$

$$f_X(x) = f_Z(x)(1 + \eta^i h_i(x) + \eta^{ij} h_{ij}(x)/2! + \eta^{ijk} h_{ijk}(x)/3! + \cdots)$$

For $Z \sim \mathcal{N}(0, 1)$, h's are the Hermite polynomials

Integrate term-by-term to get approximation of (1-d) CDF

$$F_X(x) = F_Z(x) + f_Z(x)(\eta^i + \eta^{ij} h_i(x)/2! + \eta^{ijk} h_{ij}(x)/3! + \cdots)$$

Quantile matching [transport]

If X is 1-d, F_X^{-1} is a transport from $U \sim \mathrm{Unif}(0,1)$ to X

Quantile matching [transport]

If X is 1-d, F_X^{-1} is a transport from $U \sim \mathrm{Unif}(0,1)$ to X

Thus $z = [F_Z^{-1} \circ F_X](x)$ or $F_Z(z) = F_X(x)$
is a transport from $x \sim X$ to $z \sim Z$

An algorithm [transport]

Take Taylor series expansion of Φ

$$\begin{aligned}\Phi(z) &= \Phi(x) + \Phi'(x)(z-x) + \cdots + \frac{1}{k!}\Phi^{(k)}(x)(z-x)^k \\ &= \Phi(x) - \phi(x)[(x-z) + \cdots + \frac{1}{k!}h_{k-1}(x)(x-z)^k]\end{aligned}$$

An algorithm [transport]

Take Taylor series expansion of Φ

$$\begin{aligned}\Phi(z) &= \Phi(x) + \Phi'(x)(z-x) + \cdots + \frac{1}{k!}\Phi^{(k)}(x)(z-x)^k \\ &= \Phi(x) - \phi(x)[(x-z) + \cdots + \frac{1}{k!}h_{k-1}(x)(x-z)^k]\end{aligned}$$

and set equal to Edgeworth series for quantile matching

$$F_X(x) = \Phi(x) - \phi(x)[\eta_1 + \frac{1}{2!}\eta_2 h_1(x) + \cdots + \frac{1}{k!}\eta_k h_{k-1}(x)]$$

Formal inversion [transport]

$F_Z(z) = F_X(x)$ implies

$$(x - z) + \cdots + \frac{1}{k!} h_{k-1}(x)(x - z)^k = \eta_1 + \cdots + \frac{1}{k!} \eta_k h_{k-1}(x)$$

Formal inversion [transport]

$F_Z(z) = F_X(x)$ implies

$$(x - z) + \cdots + \frac{1}{k!} h_{k-1}(x)(x - z)^k = \eta_1 + \cdots + \frac{1}{k!} \eta_k h_{k-1}(x)$$

Let $y := x - z$, matching procedure $y^k = \eta_k$

Formal inversion [transport]

$F_Z(z) = F_X(x)$ implies

$$(x - z) + \cdots + \frac{1}{k!} h_{k-1}(x)(x - z)^k = \eta_1 + \cdots + \frac{1}{k!} \eta_k h_{k-1}(x)$$

Let $y := x - z$, matching procedure $y^k = \eta_k$

Treat $f(y) := y + \cdots + h_{k-1}(x) y^k / k!$ as polynomial in y

Define $p(x) := \eta_1 + \cdots + \eta_k h_{k-1}(x) / k!$ as polynomial in x

Formal inversion [transport]

$F_Z(z) = F_X(x)$ implies

$$(x - z) + \cdots + \frac{1}{k!} h_{k-1}(x)(x - z)^k = \eta_1 + \cdots + \frac{1}{k!} \eta_k h_{k-1}(x)$$

Let $y := x - z$, matching procedure $y^k = \eta_k$

Treat $f(y) := y + \cdots + h_{k-1}(x) y^k / k!$ as polynomial in y

Define $p(x) := \eta_1 + \cdots + \eta_k h_{k-1}(x)/k!$ as polynomial in x

Every polynomial of the form $f(x) = a_1 x + \cdots + a_k x^k$ with $a_1 \neq 0$ has inverse polynomial $g(x) = b_0 + \cdots + b_k x^k$ with

$$f \circ g = g \circ f = x + \mathcal{O}(x^{k+1})$$

Cornish-Fisher [transport]

From Möbius inversion/moment to cumulant procedure,

$$(g \circ f)_{[n]} = \sum_{\sigma \in \Upsilon_n} g_{\#\sigma} \prod_{b \in \sigma} f_{\#b}$$

Cornish-Fisher [transport]

From Möbius inversion/moment to cumulant procedure,

$$(g \circ f)_{[n]} = \sum_{\sigma \in \Upsilon_n} g_{\#\sigma} \prod_{b \in \sigma} f_{\#b}$$

Linear in g, back-substitute for coefficients of g

Cornish-Fisher [transport]

From Möbius inversion/moment to cumulant procedure,

$$(g \circ f)_{[n]} = \sum_{\sigma \in \Upsilon_n} g_{\#\sigma} \prod_{b \in \sigma} f_{\#b}$$

Linear in g, back-substitute for coefficients of g

Applying g to the Edgeworth CDF completes the procedure

$$f(x - z) = p(x)$$
$$z = x - g(p(x))$$

This is known as Cornish-Fisher [Cornish and Fisher 1938]

Generalizing Cornish-Fisher

How to generalize Cornish-Fisher to multivariate distributions?

Recall: Knothe-Rosenblatt triangular map

$$T(\boldsymbol{x}) = \begin{bmatrix} T_1(x_1) \\ T_2(x_1, x_2) \\ \vdots \\ T_d(x_1, \ldots, x_d) \end{bmatrix}$$

Conditional properties

Recall: Knothe-Rosenblatt triangular map

$$T(\boldsymbol{x}) = \begin{bmatrix} T_1(x_1) \\ T_2(x_1, x_2) \\ \vdots \\ T_d(x_1, \ldots, x_d) \end{bmatrix}$$

i-th row encodes *conditional* distribution of i-th variable

Triangular transport

For i in $1, \ldots, d$

1. Partition into first $i-1$ entries and i-th entry

Triangular transport

For i in $1, \ldots, d$

1. Partition into first $i-1$ entries and i-th entry
2. i-th entry of transport: *conditional* distribution
   (how to condition: variety of strategies) [representation]

Triangular transport

For i in $1, \ldots, d$

1. Partition into first $i-1$ entries and i-th entry
2. i-th entry of transport: *conditional* distribution
   (how to condition: variety of strategies) [representation]
3. Reduce to 1-d Cornish-Fisher [transport]

cdf (cf. sampling from copula)

$$\frac{\partial}{\partial \boldsymbol{x}_1} \operatorname{cdf}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \operatorname{cdf}(\boldsymbol{x}_2 \mid \boldsymbol{x}_1) \operatorname{pdf}(\boldsymbol{x}_1)$$

Conditioning

cdf (cf. sampling from copula)

$$\frac{\partial}{\partial \boldsymbol{x}_1} \operatorname{cdf}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \operatorname{cdf}(\boldsymbol{x}_2 \mid \boldsymbol{x}_1) \operatorname{pdf}(\boldsymbol{x}_1)$$

logpdf and subtraction

$$\log \operatorname{pdf}(\boldsymbol{x_1}, \boldsymbol{x_2}) - \log \operatorname{pdf}(\boldsymbol{x_1}) = \log \operatorname{pdf}(\boldsymbol{x_2} \mid \boldsymbol{x_1})$$

Conditioning

cdf (cf. sampling from copula)

$$\frac{\partial}{\partial \boldsymbol{x}_1} \operatorname{cdf}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \operatorname{cdf}(\boldsymbol{x}_2 \mid \boldsymbol{x}_1) \operatorname{pdf}(\boldsymbol{x_1})$$

logpdf and subtraction

$$\log \operatorname{pdf}(\boldsymbol{x_1}, \boldsymbol{x_2}) - \log \operatorname{pdf}(\boldsymbol{x_1}) = \log \operatorname{pdf}(\boldsymbol{x_2} \mid \boldsymbol{x_1})$$

conditional cumulants

Implementation details

`jax` [Bradbury et al. 2018] and `sympy` [Meurer et al. 2017]

numerical computation $\iff$ symbolic computation

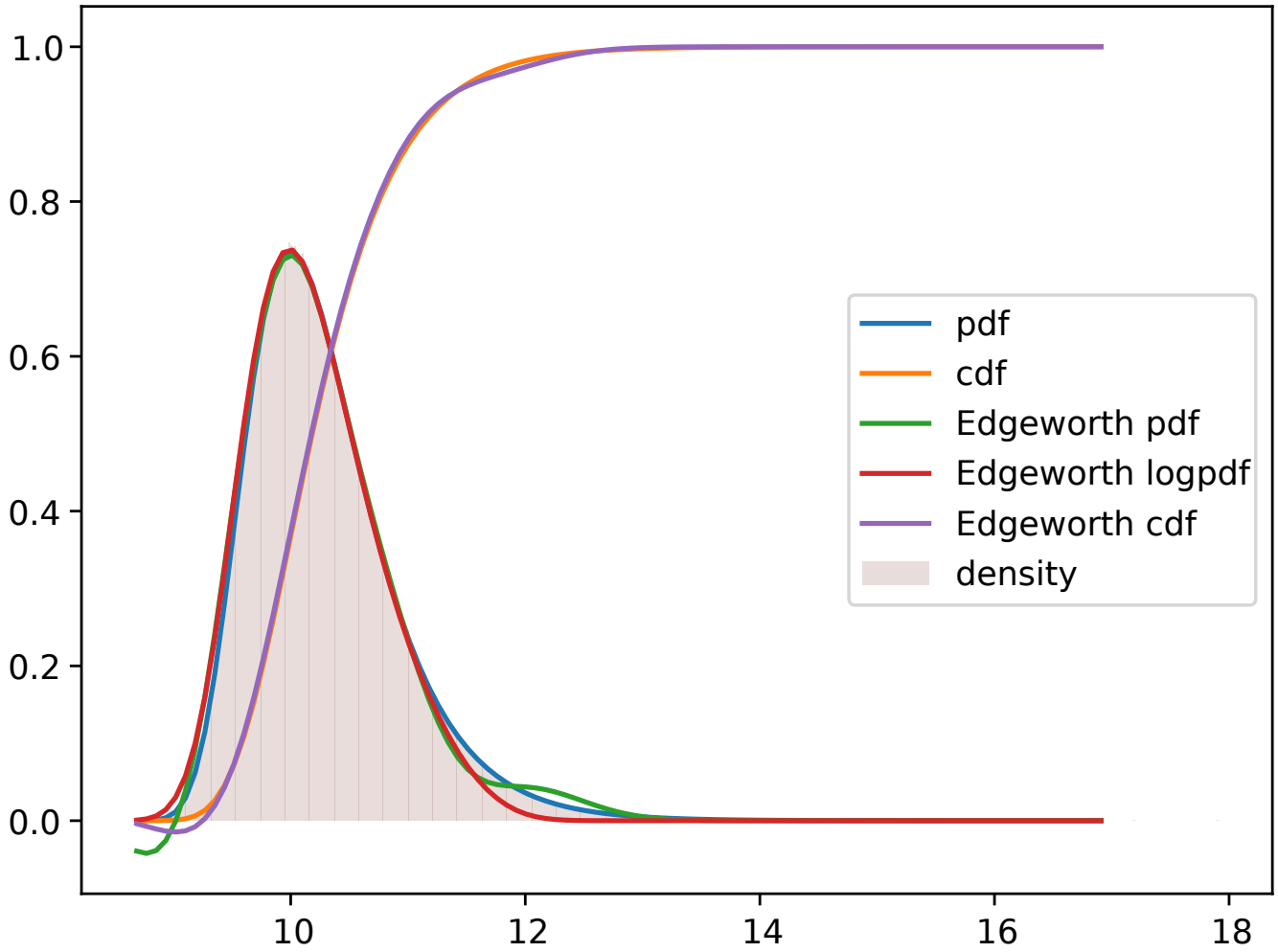
Implementation details

`jax` [Bradbury et al. 2018] and `sympy` [Meurer et al. 2017]

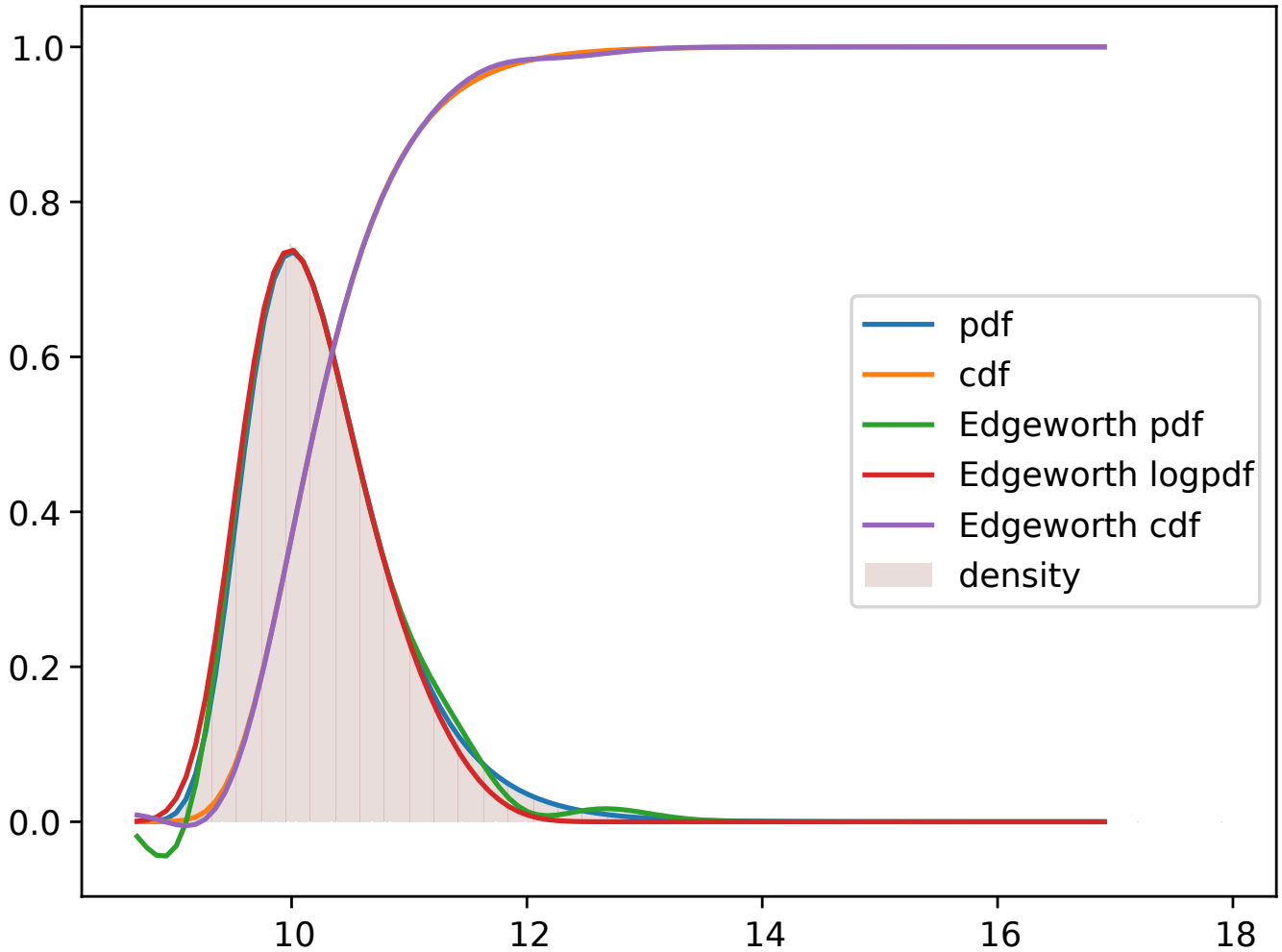
numerical computation $\iff$ symbolic computation

implicit throughout: truncation based on *cumulant order* from averaging in central limit theorem rather than *term order* (the i-th cumulant has order $\max(i-2,0)/2$)

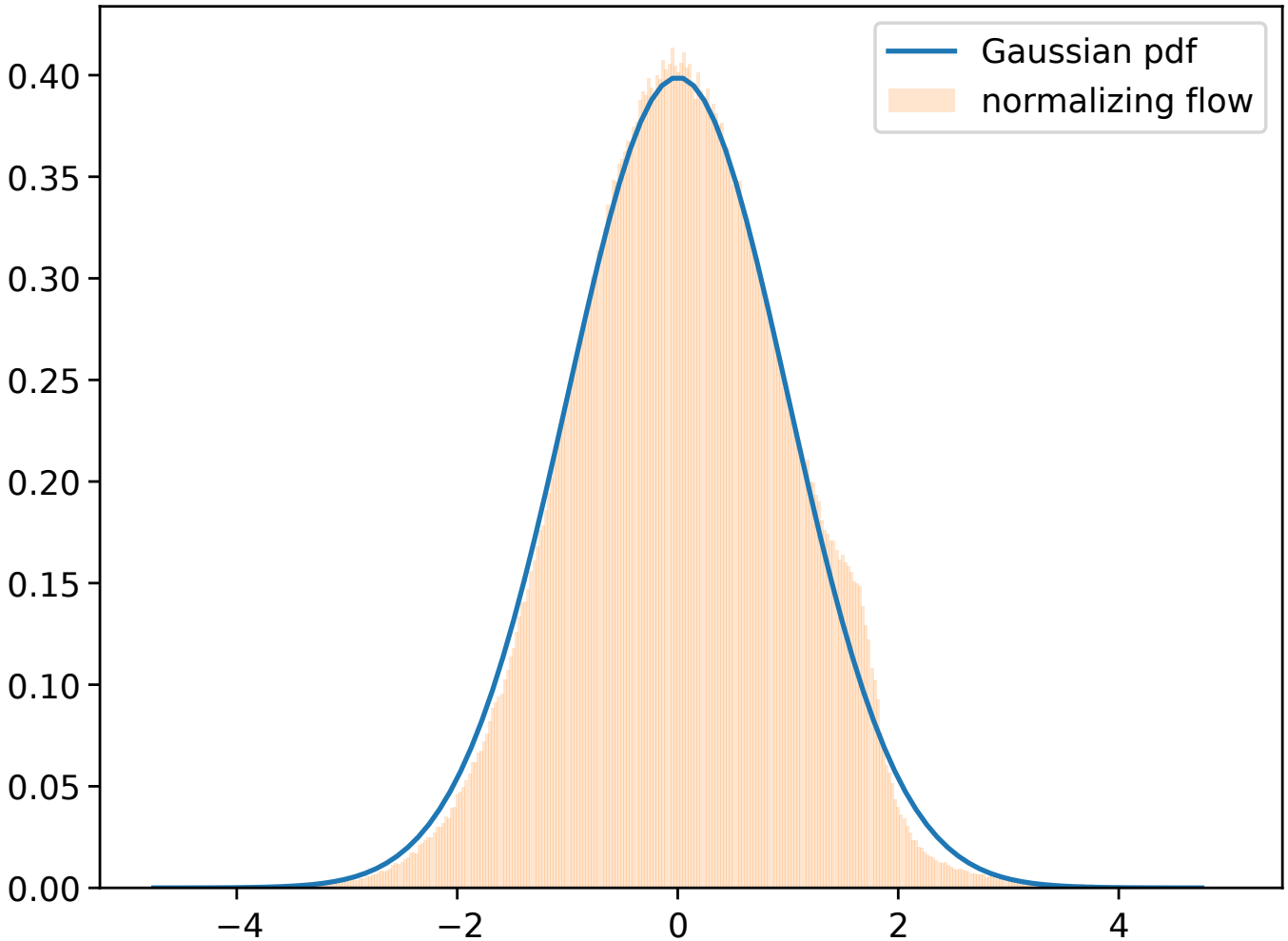
1-d result


1.0
0.8
0.6
0.4
0.2
0.0
10
12
14
16
18
pdf
cdf
Edgeworth pdf
Edgeworth logpdf
Edgeworth cdf
density

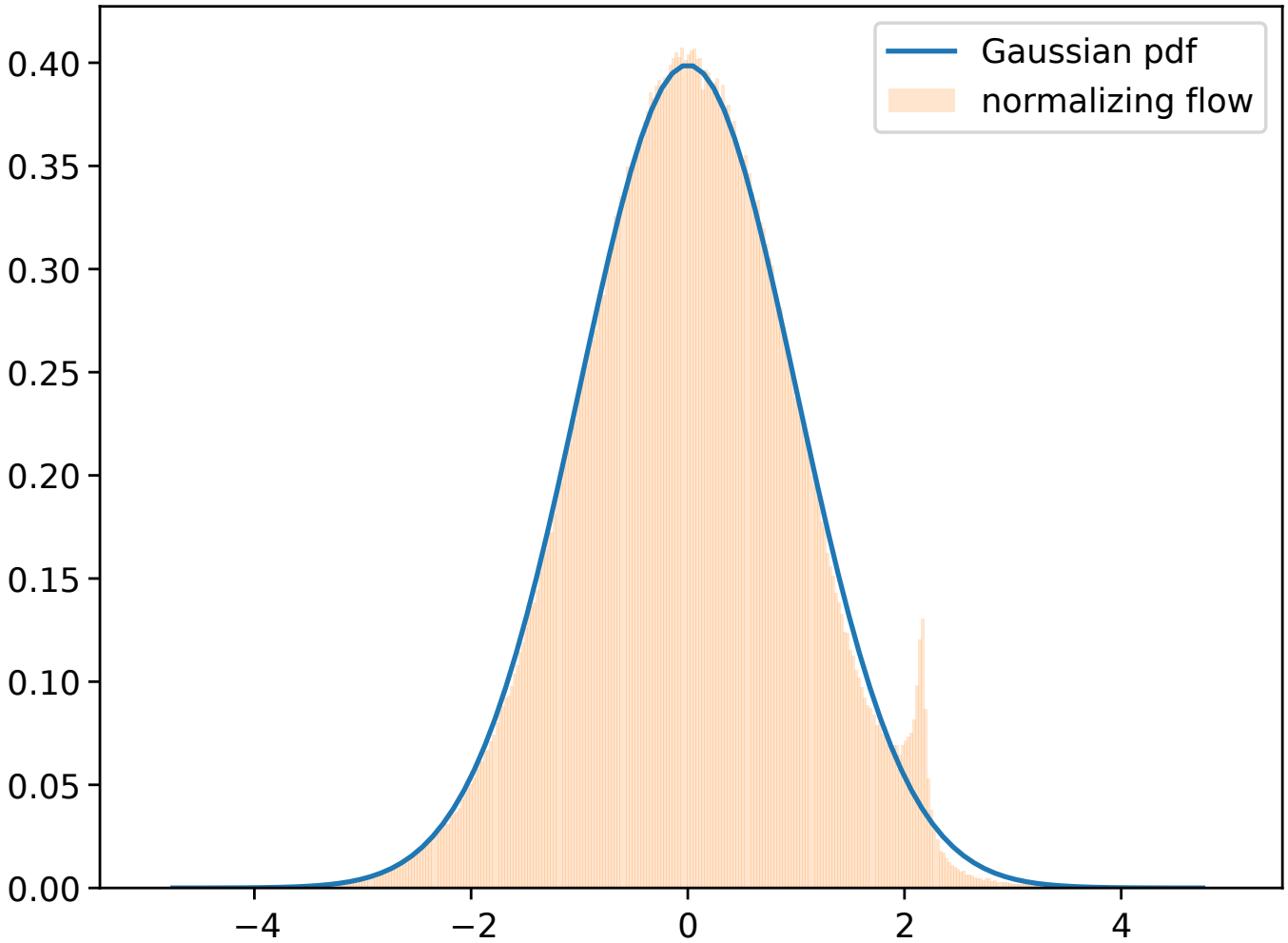
1-d result


1.0
0.8
0.6
0.4
0.2
0.0
10
12
14
16
18
pdf
cdf
Edgeworth pdf
Edgeworth logpdf
Edgeworth cdf
density

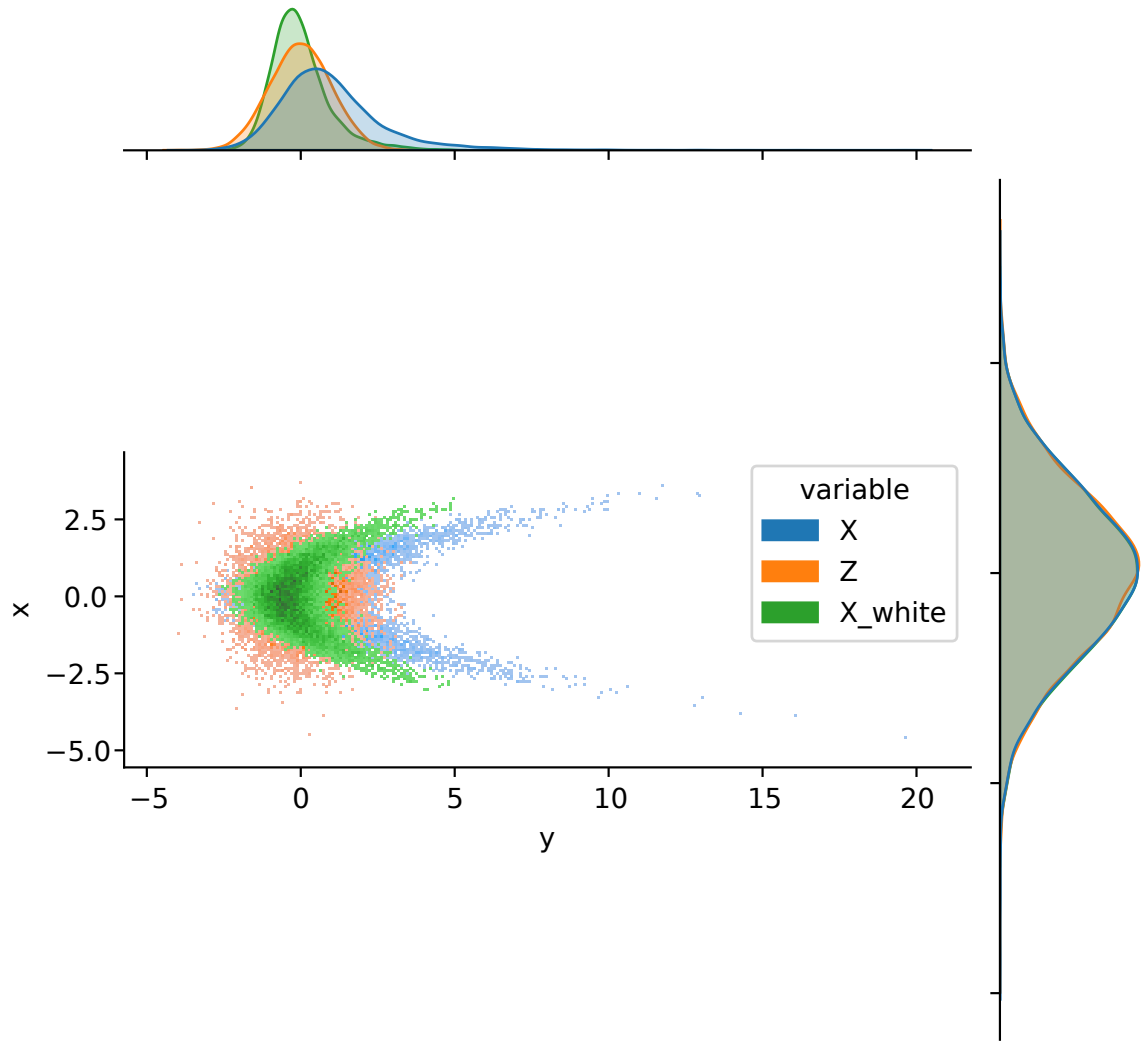
1-d result


Gaussian pdf
normalizing flow
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−4
−2
0
2
4

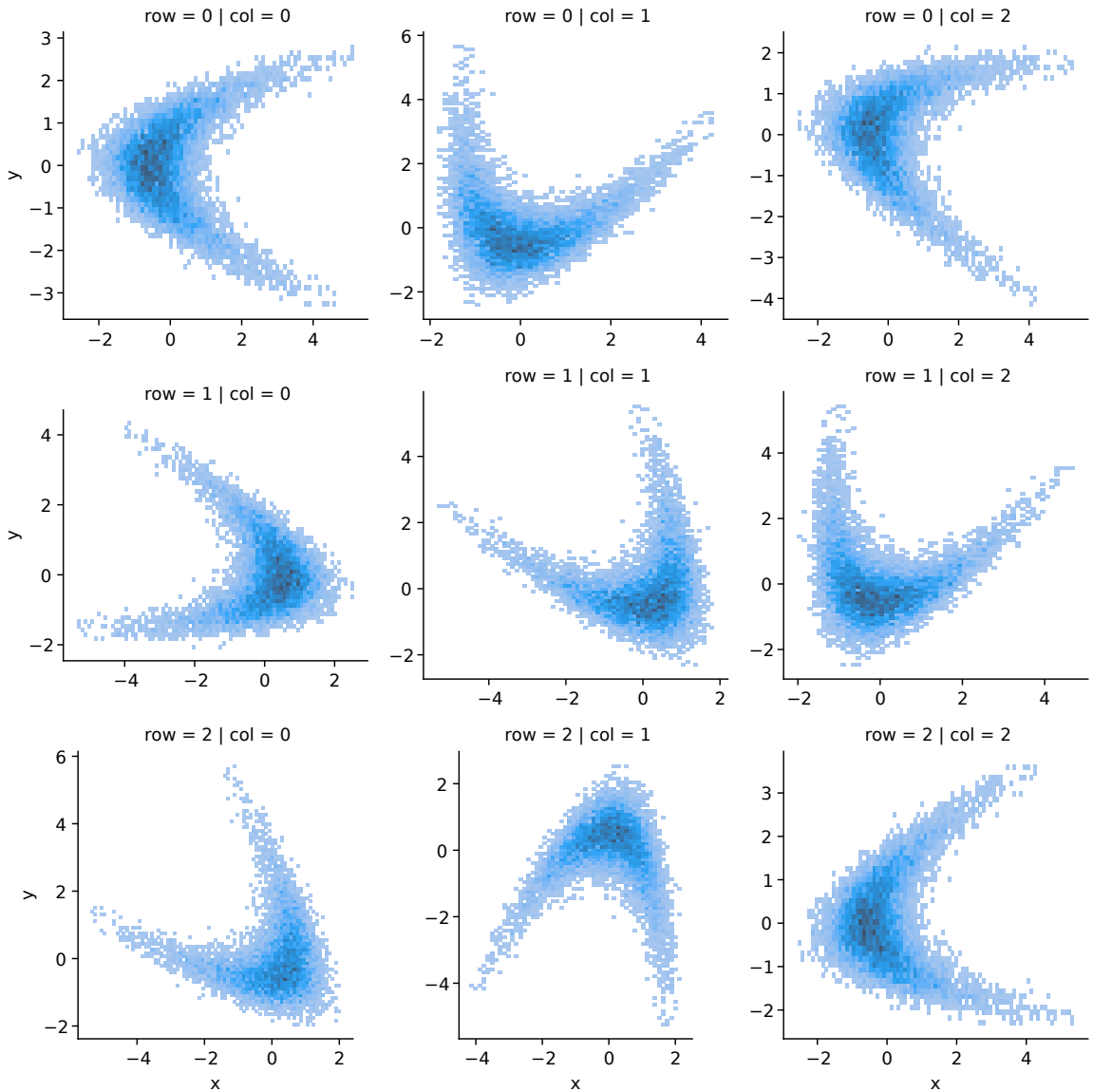
1-d result


Gaussian pdf
normalizing flow
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−4
−2
0
2
4

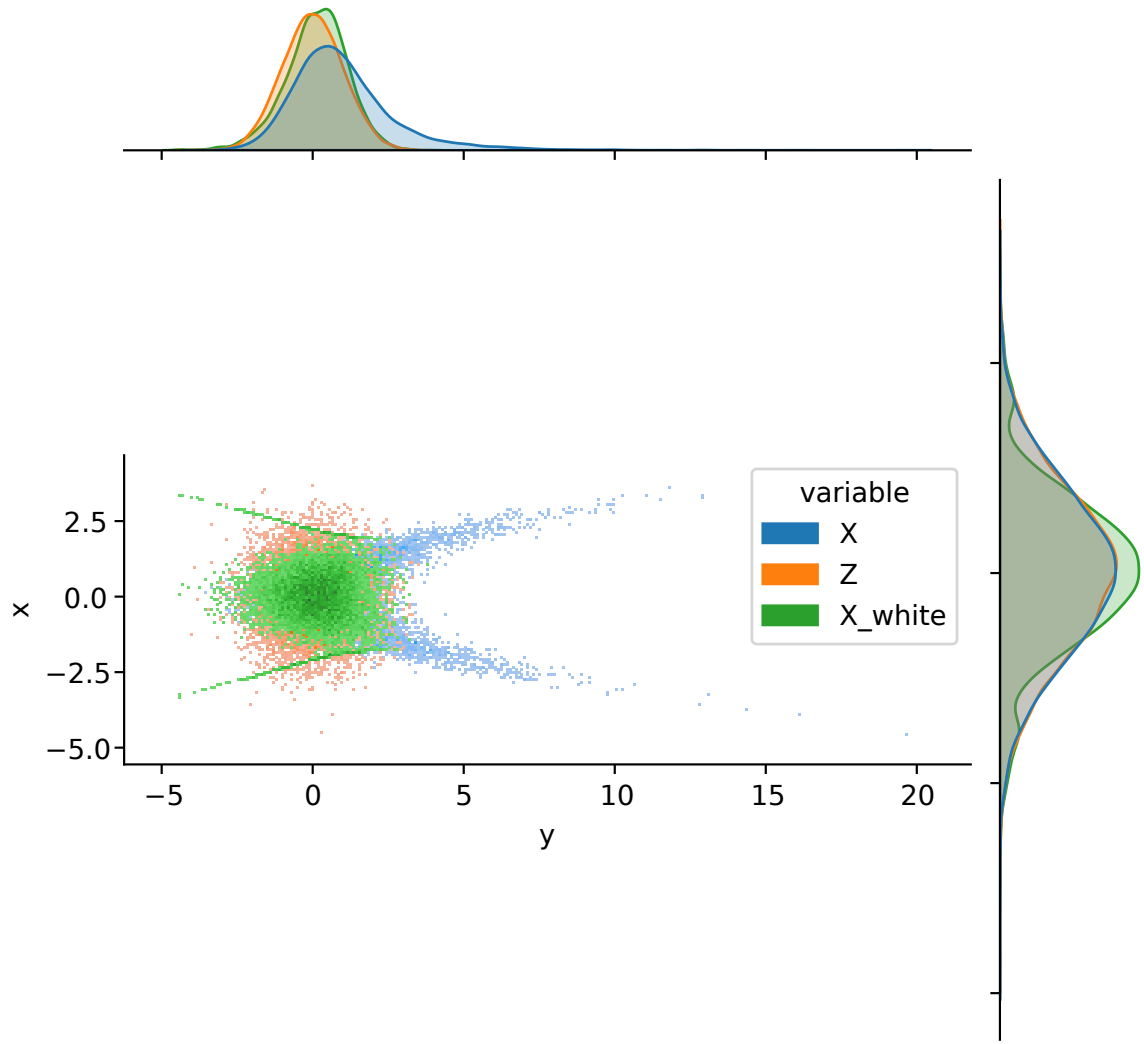
Multivariate result



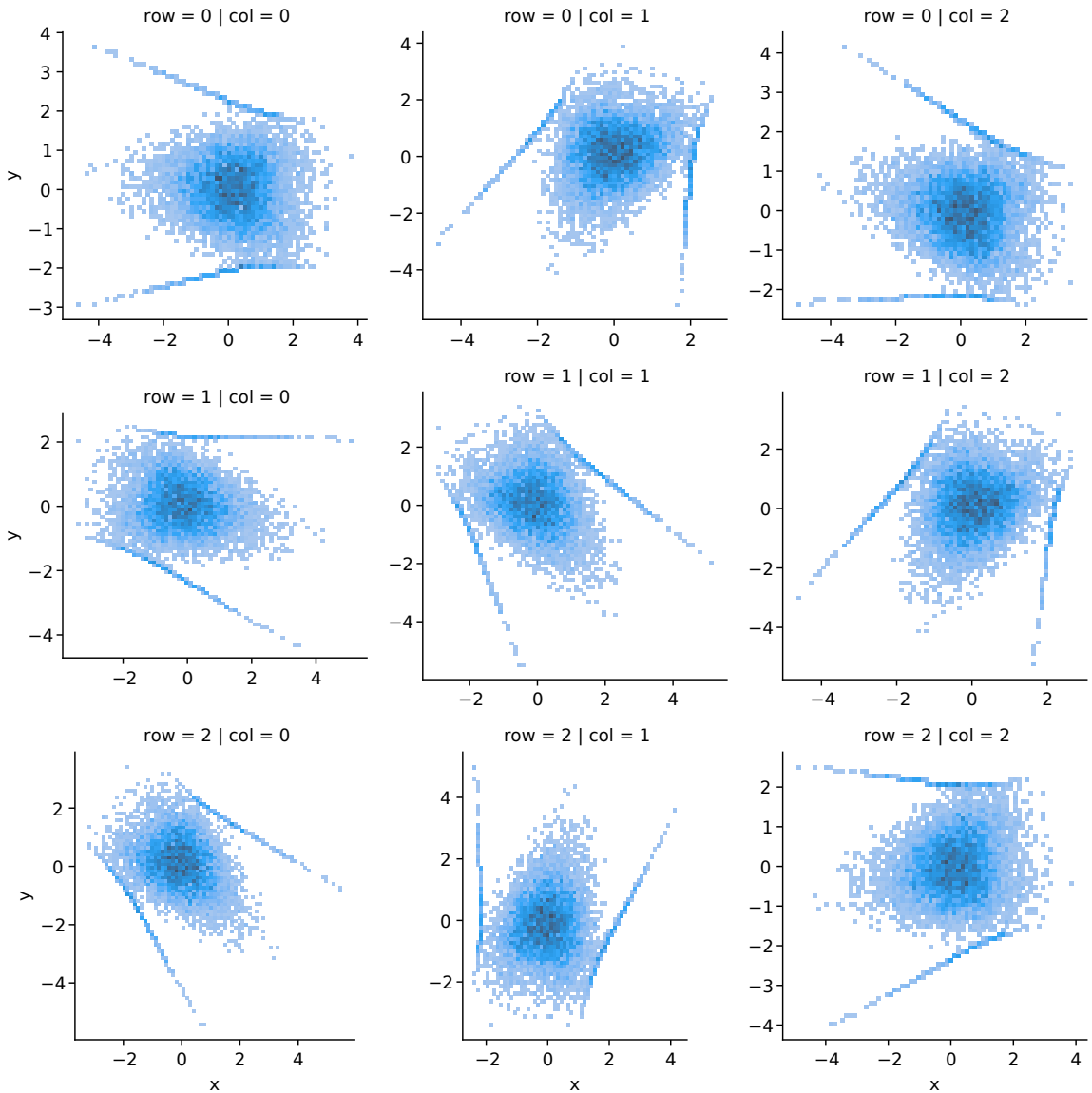
Multivariate result



Multivariate result



Multivariate result



Conclusion

Generative modeling through transport

Canonical Knothe-Rosenblatt map

Map estimated from samples through cumulants

Interpretable: polynomial and symbolic expressions

References I

- Bradbury, James et al. (2018). *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. URL: http://github.com/jax-ml/jax.
- Cornish, E. A. and R. A. Fisher (1938). "Moments and Cumulants in the Specification of Distributions". In: *Revue de l'Institut International de Statistique / Review of the International Statistical Institute* 5.4, pp. 307–320. ISSN: 0373-1138. DOI: 10.2307/1400905. JSTOR: 1400905.
- Fullwood, David T., Stephen R. Niezgoda, and Surya R. Kalidindi (Mar. 2008). "Microstructure Reconstructions from 2-Point Statistics Using Phase-Recovery Algorithms". In: *Acta Materialia* 56.5, pp. 942–948. ISSN: 1359-6454. DOI: 10.1016/j.actamat.2007.10.044.

References II

- Katzfuss, Matthias and Florian Schäfer (Feb. 2022). *Scalable Bayesian Transport Maps for High-Dimensional Non-Gaussian Spatial Fields*. arXiv: 2108.04211 [stat].
- Marzouk, Youssef et al. (2016). "An Introduction to Sampling via Measure Transport". In: *arXiv:1602.05023 [math, stat]*, pp. 1–41. DOI: 10.1007/978-3-319-11259-6_23-1. arXiv: 1602.05023 [math, stat].
- McCullagh, P (2018). *Tensor Methods in Statistics: Monographs on Statistics and Applied Probability*. Boca Raton, FL: CRC Press. ISBN: 978-1-351-08556-4 978-1-351-07711-8 978-1-351-09401-6 978-1-315-89801-8.
- Meurer, Aaron et al. (Jan. 2017). "SymPy: symbolic computing in Python". In: *PeerJ Computer Science* 3, e103. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.103. URL: https://doi.org/10.7717/peerj-cs.103.

References III



Spantini, Alessio, Daniele Bigoni, and Youssef Marzouk (July 2018). *Inference via Low-Dimensional Couplings*. DOI: 10.48550/arXiv.1703.06131. arXiv: arXiv:1703.06131.

Partition lattice on indices (subset)

Let $\Upsilon_n :=$ all partitions of n indices,

$$\mathsf{mom}([n]) = \sum_{\sigma \in \Upsilon_n} \prod_{b \in \sigma} \mathsf{cum}(b)$$

$$\mathsf{cum}([n]) = \sum_{\sigma \in \Upsilon_n} (-1)^{|\sigma|-1} (|\sigma| - 1)! \prod_{b \in \sigma} \mathsf{mom}(b)$$

Special case of composition of polynomials f, g

$$(f \circ g)_{[n]} = \sum_{\sigma \in \Upsilon_n} f_{\#\sigma} \prod_{b \in \sigma} g_{\#b}$$

Edgeworth series

Let X have cumulants κ and Z have cumulants λ

Define the *pseudo*-cumulants

$$\eta^i = \kappa^i - \lambda^i \quad \eta^{i,j} = \kappa^{i,j} - \lambda^{i,j} \quad \eta^{i,j,k} = \kappa^{i,j,k} - \lambda^{i,j,k} \quad \dots$$

as well as the resulting *pseudo*-moments $\eta^i, \eta^{ij}, \eta^{ijk}, \dots,$

Edgeworth series

X and Z have cumulant generating functions

$$K_X(\xi) = \xi_i \kappa^i + \xi_i \xi_j \kappa^{i,j}/2! + \xi_i \xi_j \xi_k \kappa^{i,j,k}/3! + \cdots$$
$$K_Z(\xi) = \xi_i \lambda^i + \xi_i \xi_j \lambda^{i,j}/2! + \xi_i \xi_j \xi_k \lambda^{i,j,k}/3! + \cdots$$

Subtracting gives

$$K_X(\xi) = K_Z(\xi) + \xi_i \eta^i + \xi_i \xi_j \eta^{i,j}/2! + \xi_i \xi_j \xi_k \eta^{i,j,k}/3! + \cdots$$

and exponentiation gives

$$M_X(\xi) = M_Z(\xi)(1 + \xi_i \eta^i + \xi_i \xi_j \eta^{ij}/2! + \xi_i \xi_j \xi_k \eta^{ijk}/3! + \cdots)$$

Invert $M_X(\xi) := \mathbb{E}[\exp(\xi_i x^i)]$ term-by-term to recover f_X

Edgeworth series

Taking a Taylor expansion at $Y = 0$ and averaging,

$$\mathbb{E}_Y \left[f_Z(x) - y_i \frac{\partial f_Z}{\partial y^i} + \frac{y_i y_j}{2!} \frac{\partial^2 f_Z}{\partial y^i \partial y^j} - \frac{y_i y_j y_k}{3!} \frac{\partial^3 f_Z}{\partial y^i \partial y^j \partial y^k} + \cdots \right]$$

Letting $f_{i_1 \cdots i_n}(x) := (-1)^n \partial^n f_Z(x) / \partial x^{i_1} \cdots \partial x^{i_n}$,

$$f_X(x) = f_Z(x) + \eta^i f_i(x) + \eta^{ij} f_{ij}(x)/2! + \eta^{ijk} f_{ijk}(x)/3! + \cdots$$

Letting $h_{i_1 \cdots i_n}(x) := f_{i_1 \cdots i_n}(x) / f_Z(x)$,

$$f_X(x) = f_Z(x)(1 + \eta^i h_i(x) + \eta^{ij} h_{ij}(x)/2! + \eta^{ijk} h_{ijk}(x)/3! + \cdots)$$

For $Z \sim \mathcal{N}(0, 1)$, h's are the Hermite polynomials

Cholesky factor

Consider multivariate Gaussian with covariance matrix $\begin{pmatrix} \Theta_{1,1} & \Theta_{1,2} \\ \Theta_{2,1} & \Theta_{2,2} \end{pmatrix}$ and Cholesky factor $\begin{pmatrix} L_{1,1} & 0 \\ L_{2,1} & L_{2,2} \end{pmatrix}$

We have for $z_1 := L_{1,1}^{-1} x_1$,

$$
\begin{aligned}
L_{2,1} z_1 + L_{2,2} z_2 &\sim \mathcal{N}(L_{2,1} z_1, L_{2,2} L_{2,2}^\top) \\
L_{2,1} z_1 + L_{2,2} z_2 &\sim \mathcal{N}(\Theta_{2,1} \Theta_{1,1}^{-1} x_1, \Theta_{2,2} - \Theta_{2,1} \Theta_{1,1}^{-1} \Theta_{1,2}) \\
L_{2,1} &= \Theta_{2,1} L_{1,1}^{-\top} \\
L_{2,2} &= \mathrm{chol}(\Theta_{2,2} - L_{2,1} L_{2,1}^\top)
\end{aligned}
$$

Allowing for iterative computation of L