Möbius inversion meets tensors: inference by Edgeworth series

Stephen Huan[1] Andreas Robertson[2] Youssef Marzouk[3] Florian Schäfer[4]

[1]Carnegie Mellon University [2]Sandia National Laboratories [3]Massachusetts Institute of Technology [4]Georgia Institute of Technology

Generative modeling by measure transport

Data X (known only from samples) has density f_X, reference distribution Z with chosen density f_Z (e.g. $\mathcal{N}(0, \mathsf{Id}_d)$); Z easy to sample.
Infinitely many *transport maps* T pushing $z \sim Z$ to $T(z) \sim X$.
(VAEs, GANs, normalizing flows, score-based/diffusion models.)

Knothe-Rosenblatt transport

Restrict to *triangular* maps in the sense that

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

and control the sign of the diagonal by enforcing $\nabla T \succ 0$.
Resulting *unique* transport called the *Knothe-Rosenblatt rearrangement*.

Computational benefits

Immediate sampler $z \sim \mathcal{N}(0, \mathsf{Id})$, $T(z) \sim X$.
Density through change-of-variables formula

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

Compute T^{-1} by recursive substitution and 1-d root finding. Eigenvalues of ∇T and $\det \nabla T$ easily accessible from $\operatorname{diag}(\nabla T)$.

Example: Gaussian transport by Cholesky

Specialize to $x \sim \mathcal{N}(\boldsymbol{\mu}, \Theta)$. Cumulant generating function is

$$K_X(\xi) \coloneqq \log \mathbb{E}[\exp(\xi_i x^i)] = \xi_i \mu^i + \frac{1}{2} \xi_i \xi_j \Theta^{i,j}.$$

Suffices to match mean and covariance, affine ansatz $T(z) = Lz + \mu$,

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

Pick L such that $\Theta = LL^\top$ (e.g. Cholesky factor). Jacobian of T is L, by construction lower triangular with positive diagonal ($\nabla T \succ 0$).

How to estimate the map from samples?

Previous work [1]–[3]: minimize empirical KL over parametric family.
This work: 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 microstructure (e.g. edges).

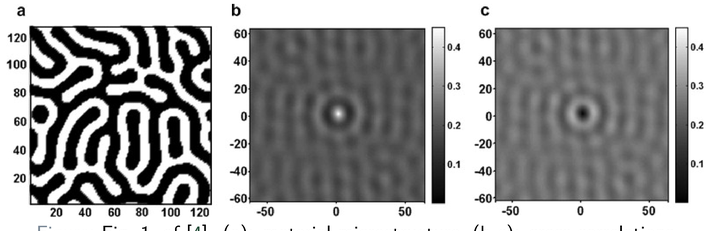


Figure: Fig. 1. of [4]. (a): material microstructure, (b-c): cross-correlations

Moments

The i-th *moment* captures interactions (product) of i variables

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

Can be estimated directly from samples with empirical expectation. *Moment generating function* collects moments into multivariate Taylor series

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

Cumulants

Cumulant generating function implicitly defines *cumulants* by

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

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

$$\kappa^i = \kappa^i,\ \kappa^{i,j} = \kappa^{ij} - \kappa^i \kappa^j,$$
$$\kappa^{i,j,k} = \kappa^{ijk} - \kappa^i \kappa^{jk} - \kappa^j \kappa^{ik} - \kappa^k \kappa^{ij} + 2\kappa^i \kappa^j \kappa^k, \ldots$$

Cumulant of sum of independent variables sum of cumulants.
Independent random variables have zero joint cumulants.

Edgeworth series

Derivation follows [5]. 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" $\eta^i, \eta^{ij}, \ldots$

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

By term-by-term inversion of MGF, density approximation

$$h_{i_1 \cdots i_n}(x) \coloneqq \frac{(-1)^n}{f_Z(x)} \frac{\partial^n f_Z(x)}{\partial x^{i_1} \cdots \partial x^{i_n}},$$
$$f_X(x) = f_Z(x)(1 + \eta^i h_i(x) + \eta^{ij} h_{ij}(x)/2! + \eta^{ijk} h_{ijk}(x)/3! + \cdots).$$

For $Z \sim \mathcal{N}(0, 1)$, h's are the Hermite polynomials.
Integrate term-by-term to get approximation of (1-d) CDF

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

Warmup: quantile matching

If X is 1-d, F_X^{-1} transports $\mathrm{Unif}(0, 1)$ 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 $z \sim Z$.

Cornish-Fisher

[6] equates Taylor series of standard Gaussian CDF to Edgeworth CDF

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

Let $y \coloneqq x - z$, treat $f(y) \coloneqq y + \cdots + h_{k-1}(x) y^k/k!$ as a polynomial in y and define $p(x) \coloneqq \eta_1 + \cdots + \eta_k h_{k-1}(x)/k!$ as a polynomial in x. Apply formal inversion: 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

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

Applying g to the Edgeworth CDF completes the procedure,

$$f(x - z) = p(x) \qquad z = x - q(x) \qquad q = g \circ p.$$

Sampling by truncated polynomial inversion

Cornish-Fisher transports Gaussian to data. What about sampling?

$$q(x) = q(z + (x - z)) = q(z + q(x)); \quad x = z + q(x)$$
$$q(x) = q(z) + q'(z) q(x) + \frac{1}{2} q''(z) q(x)^2 + \frac{1}{6} q'''(z) q(x)^3 + \cdots$$

Recursively substituting $q(x)$ into the Taylor expansion and truncating based on cumulant order gives an "inverse" polynomial (not formal).

Generalization to multivariate distributions

Multivariate Cornish-Fisher by Knothe-Rosenblatt. For i in $1, \ldots, d$

1. Partition into first $i - 1$ variables and i-th variable
2. i-th output of transport: *conditional* distribution
3. Reduce to 1-d Cornish-Fisher

Conditioning by implicit knowledge of joint distribution from cumulants.

Normalizing map from data to Gaussian

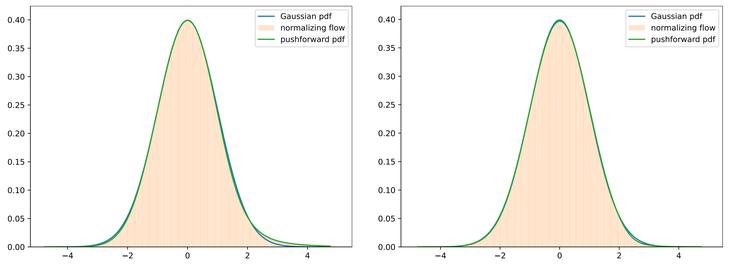


Figure: Cornish-Fisher series applied to a Gumbel distribution with $\mu = 10$, $\beta = 0.5$. Left panel shows the computed expansion with 4 moments and right with 6 moments.

Sampling map from Gaussian to data

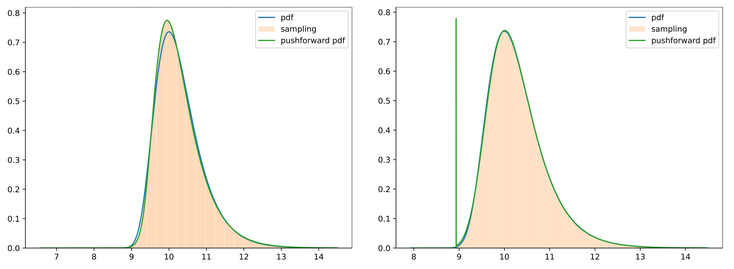


Figure: Inverse Cornish-Fisher used for sampling the same distribution.

References

[1] Y. Marzouk, T. Moselhy, M. Parno, and A. Spantini, "An introduction to sampling via measure transport," *arXiv:1602.05023 [math, stat]*, pp. 1–41, 2016. DOI: 10.1007/978-3-319-11259-6_23-1. arXiv: 1602.05023 [math, stat].

[2] A. Spantini, D. Bigoni, and Y. Marzouk, *Inference via low-dimensional couplings*, Jul. 2018. DOI: 10.48550/arXiv.1703.06131. arXiv: arXiv:1703.06131.

[3] M. Katzfuss and F. Schäfer, *Scalable Bayesian transport maps for high-dimensional non-Gaussian spatial fields*, Feb. 2022. arXiv: 2108.04211 [stat].

[4] D. T. Fullwood, S. R. Niezgoda, and S. R. Kalidindi, "Microstructure reconstructions from 2-point statistics using phase-recovery algorithms," *Acta Materialia*, vol. 56, no. 5, pp. 942–948, Mar. 2008, ISSN: 1359-6454. DOI: 10.1016/j.actamat.2007.10.044.

[5] P. McCullagh, *Tensor Methods in Statistics: Monographs on Statistics and Applied Probability*. Boca Raton, FL: CRC Press, 2018, ISBN: 978-1-351-08556-4 978-1-351-07711-8 978-1-351-09401-6 978-1-315-89801-8.

[6] E. A. Cornish and R. A. Fisher, "Moments and Cumulants in the Specification of Distributions," *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, vol. 5, no. 4, pp. 307–320, 1938, ISSN: 0373-1138. DOI: 10.2307/1400905. JSTOR: 1400905.