Easy conditioning far beyond Gaussian

Faul Antoine, Ginsbourger David, Spycher Ben

IMSV
University of Bern

October, 2025

1 Motivation

2 Stability by conditioning

3 Methodology for Estimating Conditional Distributions

4 Application

5 Conclusion

A motivating example in medicine

- Cohort of 443 female patients from the University hospital of Bern, provided by Pr. Stute.

n	Age	BMI	TotalC	HDLC	BP
1	58	22.15	5.93	1.85	100
2	52	30.25	3.65	1.06	120
3	45	29.75	5.07	2.03	130
4	46	21.23	NA	NA	NA
5	55	20.63	NA	NA	NA

A motivating example in medicine

- Cohort of 443 female patients from the University hospital of Bern, provided by Pr. Stute.

n	Age	BMI	TotalC	HDLC	BP
1	58	22.15	5.93	1.85	100
2	52	30.25	3.65	1.06	120
3	45	29.75	5.07	2.03	130
4	46	21.23	NA	NA	NA
5	55	20.63	NA	NA	NA

- Predict the risk (with uncertainty) of developing a cardio-vascular disease (CVD) in the next 10 years when the clinical measurements are systematically missing [Mühlemann et al., 2025]

A motivating example in medicine

- Cohort of 443 female patients from the University hospital of Bern, provided by Pr. Stute.

n	Age	BMI	TotalC	HDLC	BP
1	58	22.15	5.93	1.85	100
2	52	30.25	3.65	1.06	120
3	45	29.75	5.07	2.03	130
4	46	21.23	NA	NA	NA
5	55	20.63	NA	NA	NA

- Predict the risk (with uncertainty) of developing a cardio-vascular disease (CVD) in the next 10 years when the clinical measurements are systematically missing [Mühlemann et al., 2025]
- Apply risk calculators such as SCORE2 with imputed inputs.

A motivating example in medicine

- Cohort of 443 female patients from the University hospital of Bern, provided by Pr. Stute.

n	Age	BMI	TotalC	HDLC	BP
1	58	22.15	5.93	1.85	100
2	52	30.25	3.65	1.06	120
3	45	29.75	5.07	2.03	130
4	46	21.23	NA	NA	NA
5	55	20.63	NA	NA	NA

- Predict the risk (with uncertainty) of developing a cardio-vascular disease (CVD) in the next 10 years when the clinical measurements are systematically missing [Mühlemann et al., 2025]
- Apply risk calculators such as SCORE2 with imputed inputs.
- Data publicly available on Zenodo

Analytical conditioning approach

Goal: Sample from the conditional distribution $f_{\mathbf{X_1}|\mathbf{X_2}}$ of a random vector $\mathrm{X}_1 \in \mathbb{R}^l$ given an observation of x_2 of $\mathrm{X}_2 \in \mathbb{R}^m$, based on a sample of $\mathrm{X} = (x_{1i}, x_{2i})_{i=1}^n$ previously acquired.

- A possible approach is:

Analytical conditioning approach

Goal: Sample from the conditional distribution $f_{\mathbf{X_1}|\mathbf{X_2}}$ of a random vector $\mathrm{X}_1 \in \mathbb{R}^l$ given an observation of x_2 of $\mathrm{X}_2 \in \mathbb{R}^m$, based on a sample of $\mathrm{X} = (x_{1i}, x_{2i})_{i=1}^n$ previously acquired.

- A possible approach is:

  1. to estimate the joint density of $(\mathrm{X}_1, \mathrm{X}_2)$ by $\hat{f}_{\mathbf{X_1},\mathbf{X_2}}$.

Analytical conditioning approach

Goal: Sample from the conditional distribution $f_{\mathbf{X_1}|\mathbf{X_2}}$ of a random vector $\mathrm{X}_1 \in \mathbb{R}^l$ given an observation of x_2 of $\mathrm{X}_2 \in \mathbb{R}^m$, based on a sample of $\mathrm{X} = (x_{1i}, x_{2i})_{i=1}^n$ previously acquired.

- A possible approach is:
  1. to estimate the joint density of $(\mathrm{X}_1, \mathrm{X}_2)$ by $\hat{f}_{\mathbf{X_1},\mathbf{X_2}}$.
  2. to perform analytical conditioning to obtain $\hat{f}_{\mathbf{X_1}|\mathbf{X_2}}$ for a new observation x_2

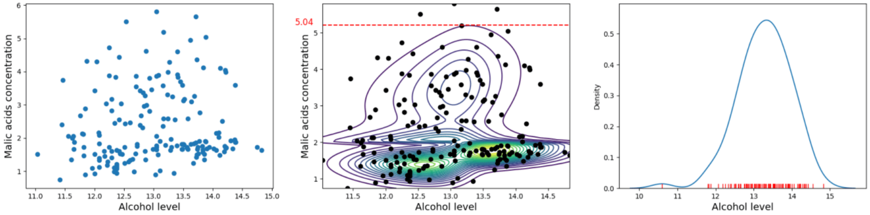
Analytical conditioning approach

Goal: Sample from the conditional distribution $f_{\mathbf{X_1}|\mathbf{X_2}}$ of a random vector $\mathrm{X}_1 \in \mathbb{R}^l$ given an observation of x_2 of $\mathrm{X}_2 \in \mathbb{R}^m$, based on a sample of $\mathrm{X} = (x_{1i}, x_{2i})_{i=1}^n$ previously acquired.

- A possible approach is:
  1. to estimate the joint density of $(\mathrm{X}_1, \mathrm{X}_2)$ by $\hat{f}_{\mathbf{X_1},\mathbf{X_2}}$.
  2. to perform analytical conditioning to obtain $\hat{f}_{\mathbf{X_1}|\mathbf{X_2}}$ for a new observation x_2
- We will assume a **parametric form** for the joint distribution and use **analytical formulas for conditioning** instead of approximations such as MCMC methods.

Example

Chemical analysis of wines grown in the same region in Italy, 178 observations.



1. Motivation
2. Stability by conditioning
3. Methodology for Estimating Conditional Distributions
4. Application
5. Conclusion

The Gaussian case

- Let $\mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma)$, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^t$ with $\mathbf{X}_1 \in \mathbb{R}^l$, $\mathbf{X}_2 \in \mathbb{R}^m$ and $l + m = d$ and
$$\Sigma = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{pmatrix}$$
with Σ_{22} non-singular.

  Then
$$\mathbf{X}_2 \sim \mathcal{N}_m(\mu_2, \Sigma_{22})$$
and
$$\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2 \sim \mathcal{N}_l(\mu_{1.2}, \Sigma_{11.2})$$

  with $\mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2)$ and $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

- Limitations: asymmetry, heavy-tail, tail dependence ...

  $\rightarrow$ Model the joint distribution with a broader class of multivariate distributions, while still enjoying stability by conditioning.

Multivariate Student t-distribution

Let $\mathbf{X} \sim t_d(\mu, \Sigma, \nu)$ and $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^t$ with $\mathbf{X}_1 \in \mathbb{R}^l$, $\mathbf{X}_2 \in \mathbb{R}^m$ and $l + m = d$.

Then by, Ding [2016] :

$$X_2 \sim t_q(\mu_2, \Sigma_{22}, \nu)$$

and

$$\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2 \sim t_p(\mu_{1.2}, \frac{\nu + d_2}{\nu + q} \Sigma_{11.2}, \nu + q)$$

with $\mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2)$, $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ and $d_2 = (\mathbf{X}_2 - \mu)^t \Sigma_{22}^{-1}(\mathbf{X}_2 - \mu)$, and Σ_{22} non-singular.

Beyond

- Other distributions with analytical formulas for their conditional distributions:

Beyond

- Other distributions with analytical formulas for their conditional distributions:
  - Multivariate unified Skew-Normal distribution, [Azzalini and Valle, 1996]

Beyond

- Other distributions with analytical formulas for their conditional distributions:
  - Multivariate unified Skew-Normal distribution, [Azzalini and Valle, 1996]
  - Multivariate unified Skew Student t-distribution [Wang et al., 2024]

Beyond

- Other distributions with analytical formulas for their conditional distributions:
  - Multivariate unified Skew-Normal distribution, [Azzalini and Valle, 1996]
  - Multivariate unified Skew Student t-distribution [Wang et al., 2024]
- Remaining limitations: multi-modal distributions, different type of marginals.

Transdimensional families of distributions

Let $d \geq 1$ and a product of spaces $\prod_{k=1}^{d} \mathcal{X}_k$, where $\mathcal{X}_k$ are assumed to be measure spaces, typically $\mathbb{R}$ or its subspaces.

$\mathcal{F}^{(d)} = \{f^{(d)}(\cdot;\theta); \theta \in \Theta^{(d)}\}$ a parametric family of pdfs on $\prod_{k=1}^{d} \mathcal{X}_k$ with parameters belonging to a set $\Theta^{(d)}$.

We define trans-dimensional family of pdfs by $\mathcal{F} = \bigcup_{d=1}^{D} \mathcal{F}^{(d)}$ with $1 \leq D \leq +\infty$.

We will question the stability of such families by marginalization and conditioning with respect to some components, [Faul et al., 2024].

Stability by conditioning

For f a positive pdf on $\prod_{k=1}^{d} \mathcal{X}_k$, $(\mathrm{i},\mathrm{j}) \in I_{\ell,m}^{d}$ and $(x_{j_1},\dots,x_{j_m})$ observed:

$$f_{\mathrm{i}|\mathrm{j}}(\cdot|(x_{j_1},\dots,x_{j_m})) : (x_{i_1},\dots,x_{i_\ell}) \in \prod_{k=1}^{\ell} \mathcal{X}_{i_k} \to \frac{f_{[\mathrm{i},\mathrm{j}]}(x_{i_1},\dots,x_{i_\ell},x_{j_1},\dots,x_{j_m})}{f_{\mathrm{j}}(x_{j_1},\dots,x_{j_m})}$$

The conditioning operator is defined for $f^{(d)}(\cdot;\theta) \in \mathcal{F}^{(d)}$ as:

$$\mathrm{Cond}(f^{(d)}(\cdot,\theta)) = \left\{ f_{\mathrm{i}|\mathrm{j}}^{(d)}(\cdot|(x_{j_1},\dots,x_{j_m});\theta),\ (\mathrm{i},\mathrm{j}) \in I_{\ell,m}^{d},\ (x_{j_1},\dots,x_{j_m}) \in \prod_{k=1}^{m} \mathcal{X}_{i_k} \right\}$$

By applying it to the trans-dimensional family of pdf $\mathcal{F}$ we get:

$$\mathrm{Cond}(\mathcal{F}) = \bigcup_{d=1}^{D} \left\{ \mathrm{Cond}(f^{(d)}(\cdot;\theta)),\ \theta \in \Theta^{(d)} \right\}$$

A trans-dimensional family of probability distributions $\mathcal{F}$ is said to be stable by conditioning if $\mathrm{Cond}(\mathcal{F}) \subset \mathcal{F}$.

Extension to finite mixtures

For $\mathcal{F}^{(d)} = \{f^{(d)}(\cdot;\theta); \Theta \in \Theta^{(d)}\}$ define the mixing operator as:

$$\text{Mix}(\mathcal{F}^{(d)}) = \left\{ \sum_{k=1}^{K} \alpha_k f^{(d)}(\cdot;\theta_k), \quad K \geq 1, \quad \theta_1, \ldots, \theta_K \in \Theta^{(d)}, \right.$$
$$\left. , \ \alpha_1, \ldots, \alpha_K \geq 0 \quad \text{s.t.} \quad \sum_{k=1}^{K} \alpha_k = 1 \right\}$$

and for the transdimensional case:

$$\text{Mix}(\mathcal{F}) = \bigcup_{d=1}^{D} \text{Mix}(\mathcal{F}^{(d)})$$

Extension to finite mixtures

Theorem [Faul et al., 2024]

If a parametrized family of trans-dimensional probability distributions $\mathcal{F}$ is stable by conditioning (resp. marginalization) then $Mix(\mathcal{F})$ is also stable by conditioning (resp. marginalization).

Example of mixture of Gaussians:

$$f(\mathbf{x}_1|\mathbf{x}_2) = \sum_{i=1}^{K} \frac{\alpha_i \mathcal{N}_m(\mathbf{x}_2|\mu_{i,2}, \Sigma_{i,22})}{\sum_{j=1}^{K} \alpha_j \mathcal{N}_m(\mathbf{x}_2|\mu_{l,2}, \Sigma_{l,22})} \mathcal{N}_l(\mathbf{x}_1|\mu_{i,1.2}, \Sigma_{i,11.2})$$

The conditional distribution is a Gaussian mixture distribution with weights:

$$\tilde{\alpha}_i(\mathbf{x}_2) = \frac{\alpha_i \mathcal{N}_m(\mathbf{x}_2|\mu_{i,2}, \Sigma_{i,22})}{\sum_{j=1}^{K} \alpha_j \mathcal{N}_m(\mathbf{x}_2|\mu_{j,2}, \Sigma_{j,22})}$$

Furthermore, stability by conditioning is preserved by univariate transformations of the marginals.

1 Motivation

2 Stability by conditioning

3 Methodology for Estimating Conditional Distributions

4 Application

5 Conclusion

Gaussian mixture copula model (GMCM)

- The Gaussian mixture copula model (GMCM) is the copula implicit in a Gaussian mixture model, it allows to capture multi-modality in the dependence structure

Gaussian mixture copula model (GMCM)

- The Gaussian mixture copula model (GMCM) is the copula implicit in a Gaussian mixture model, it allows to capture multi-modality in the dependence structure
- Let remind that the density of a Gaussian mixture model (GMM) with K components is given by:

$$\psi(x_1, \dots, x_d, \Theta) = \sum_{k=1}^{K} \alpha_j \phi(x_1, \dots, x_d; \mu_k, \Sigma_k), \ \forall \mathrm{x} = (x_1, \dots, x_n) \in \mathbb{R}^d$$

with $\Theta = \{\alpha_i, \mu_i, \Sigma_i\}_{i=1}^{K}$.

Gaussian mixture copula model (GMCM)

- The Gaussian mixture copula model (GMCM) is the copula implicit in a Gaussian mixture model, it allows to capture multi-modality in the dependence structure
- Let remind that the density of a Gaussian mixture model (GMM) with K components is given by:

$$\psi(x_1, \ldots, x_d, \Theta) = \sum_{k=1}^{K} \alpha_j \phi(x_1, \ldots, x_d; \mu_k, \Sigma_k), \;\; \forall \mathrm{x} = (x_1, \ldots, x_n) \in \mathbb{R}^d$$

  with $\Theta = \{\alpha_i, \mu_i, \Sigma_i\}_{i=1}^{K}$.
- The density of a GMCM with parameters Θ is given by:

$$c_{gmc}(u_1, \ldots, u_d; \Theta) = \frac{\psi(\Psi_1^{-1}(u_1), \ldots, \Psi_d^{-1}(u_d))}{\prod_{i=1}^{d} \psi_i(\Psi_i^{-1}(u_i))}, \;\; 0 \leq u_1, \ldots, u_d \leq 1$$

  where ψ_j and Ψ_j^{-1} denote the marginal df and inverse cdf of the GMM along the j-th dimension .

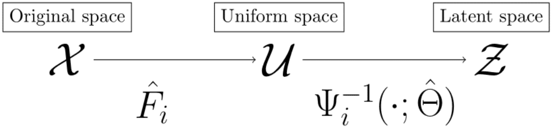
Estimate the joint multivariate distribution with copulas

From $\mathcal{D} = (x_{1i}, \ldots, x_{di})_{i=1}^{n}$ d-dimensional sample with i.i.d observations, estimate the joint distribution by:

- Estimate the marginal distribution by $\hat{F}_1, \ldots, \hat{F}_d$.
- By PIT, transform to approximately uniform pseudo-samples $\hat{u}_{ji} = \hat{F}_j(x_{ji}), \ \forall j \in \{1, \ldots, d\}$
- Fit a parametric copula by ML

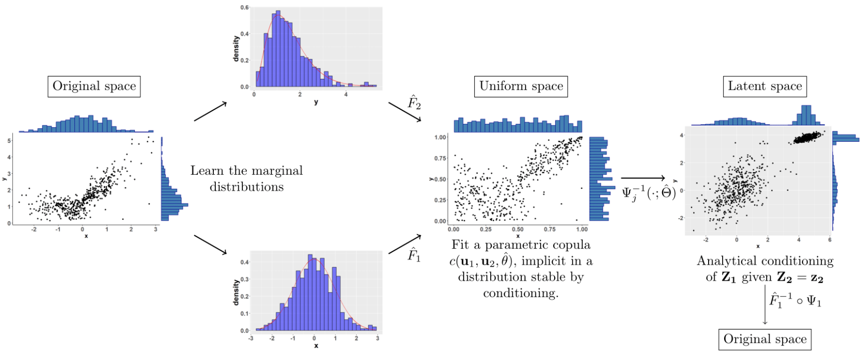
$$\hat{\theta} \in \underset{\theta \in \Theta}{\operatorname{argmin}} \prod_{i=1}^{n} c(\hat{u}_{1i}, \ldots, \hat{u}_{di}; \theta)$$

Conditioning in the latent space



- For a new point $x_2 \in \mathcal{X}$, transform to the a point in the latent space $z_2 \in \mathcal{Z}$.
- The joint distribution is approximately stable under conditioning, allowing us to perform conditioning effectively in the space $\mathcal{Z}$
- Transform back the samples to the original space by $x_1 = \hat{F}_1^{-1}(\Psi_1(z_1))$.

Algorithm Overview



1 Motivation

2 Stability by conditioning

3 Methodology for Estimating Conditional Distributions

4 Application

5 Conclusion

Analytical example: Meta Gaussian mixture model

2-dimensional meta-Gaussian mixture model i.e Gaussian mixture copula with different marginals.

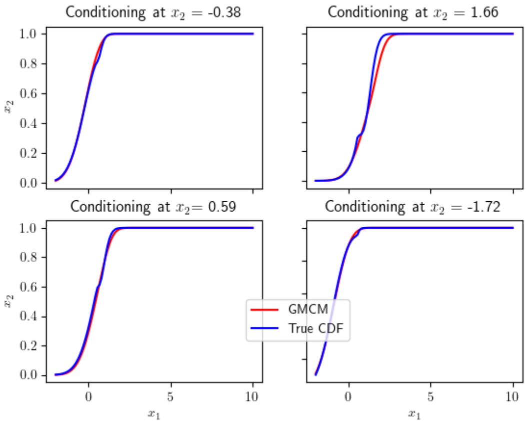


Figure 1: Comparison of true conditional cdf and estimated conditional cdf at various locations. The data follows a meta-Gaussian mixture model distribution.

Real Data: Wisconsin Cancer Data

569 patients on 30 features of the cell nuclei obtained from a digitized image of a fine needle aspirate (FNA) of a breast mass. For each patient the cancer was diagnosed as malignant or benign. We selected subset of 5 variables.

	ES	VS
GC	1.896	0.848
GMCM	**1.845**	**0.823**
TGMM	1.854	0.826
CKDE	2.007	1.979

Table 1: Comparison of various methods based on the mean of scoring rules over the test set for the Wisconsin dataset.

Application: Missing value imputation

Wisconsin Breast Cancer Dataset, adding 10% of missing values at random.

Metric	GMCM	MICE
Energy Score	**0.1908**	0.1989
Computational Time (s)	**112.3**	211.7

1. Motivation
2. Stability by conditioning
3. Methodology for Estimating Conditional Distributions
4. Application
5. Conclusion

Conclusion and perspectives

- Extend stability by conditioning properties of families of multivariate distributions far beyond the Gaussian case.

Conclusion and perspectives

- Extend stability by conditioning properties of families of multivariate distributions far beyond the Gaussian case.
- Design a versatile algorithm to estimate/sample from a conditional distribution.

Conclusion and perspectives

- Extend stability by conditioning properties of families of multivariate distributions far beyond the Gaussian case.
- Design a versatile algorithm to estimate/sample from a conditional distribution.
- Apply the methodology to other types of copulas such Student t-mixture copulas.

Conclusion and perspectives

- Extend stability by conditioning properties of families of multivariate distributions far beyond the Gaussian case.
- Design a versatile algorithm to estimate/sample from a conditional distribution.
- Apply the methodology to other types of copulas such Student t-mixture copulas.
- Extend the algorithm to handle discrete variables and missing values.

Thank You

THANK YOU !

ANY QUESTIONS ?

References

Adelchi Azzalini and A Dalla Valle. The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726, 1996.

Peng Ding. On the conditional distribution of the multivariate t distribution. *The American Statistician*, 70(3):293–295, 2016.

Antoine Faul, David Ginsbourger, and Ben Spycher. Easy conditioning far beyond gaussian. *arXiv preprint arXiv:2409.16003*, 2024.

Anja Mühlemann, Philip Stange, Antoine Faul, Serena Lozza-Fiacco, Rowan Iskandar, Manuela Moraru, Susanne Theis, Petra Stute, Ben D Spycher, and David Ginsbourger. Comparing imputation approaches to handle systematically missing inputs in risk calculators. *PLOS Digital Health*, 4(1):e0000712, 2025.

Kesen Wang, Maicon J Karling, Reinaldo B Arellano-Valle, and Marc G Genton. Multivariate unified skew-t distributions and their properties. *Journal of Multivariate Analysis*, 203:105322, 2024.