Multidimensional conformal prediction with random ellipsoids

Iain Henderson[1,2], Adrien Mazoyer[2], Fabrice Gamboa[2]
ANR ROMEO

[1]ISAE-SUPAERO
[2]Institut de Mathématiques de Toulouse (IMT)

December 4, 2025

Context : ANR ROMEO

- Post-doc with Adrien Mazoyer and Fabrice Gamboa (IMT), ANR project with Thalès R&T (Palaiseau), finished August 2024.
- ANR ROMEO : anomaly detection in drone trajectories (e.g. rotor failure) using conformal inference $\rightarrow$ multivariate time series.
- Today's talk is taken from the paper

I.H, A. Mazoyer, F. Gamboa (2024).

Adaptive inference with random ellipsoids through Conformal Conditional Linear Expectation.

`https://arxiv.org/abs/2409.18508`

Outline of the talk

1. Background on conformal inference
   - Conformal inference
   - Split conformal inference in regression
2. A score based on the empirical covariance
   - Construction of adaptive scores
   - Non asymptotic ellipsoidal confidence region
3. Asymptotic study
   - Asymptotic ellipsoid $\mathcal{F}_\alpha$ (and $\mathcal{E}_\alpha$)
   - Comparison with the standard score for elliptical distributions
4. Numerical applications
   - Elliptical data : Gaussian data, Cauchy data
   - Non elliptical data : inverse Dirichlet data
5. Perspectives

Outline of the talk

1. Background on conformal inference
   - Conformal inference
   - Split conformal inference in regression
2. A score based on the empirical covariance
   - Construction of adaptive scores
   - Non asymptotic ellipsoidal confidence region
3. Asymptotic study
   - Asymptotic ellipsoid $\mathcal{F}_\alpha$ (and $\mathcal{E}_\alpha$)
   - Comparison with the standard score for elliptical distributions
4. Numerical applications
   - Elliptical data : Gaussian data, Cauchy data
   - Non elliptical data : inverse Dirichlet data
5. Perspectives

(Vanilla) conformal inference[1]

Framework : iid scalar data $\mathcal{Z} = (Z_1, \ldots, Z_n, Z_{n+1})$. Let $\alpha \in (0, 1)$.

- Aim : build $C_\alpha = C_\alpha(Z_1, \ldots, Z_n)$, a random set such that

$$\mathbb{P}(Z_{n+1} \in C_\alpha) \geq 1 - \alpha. \tag{1}$$

!!! Above we average over $(Z_1, \ldots, Z_{n+1})$!!!

- Construction of C_α : use known distribution. Conformal inference (CI) : discrete uniform distribution.
- Advantages of CI : valid whatever finite n, no assumption on the distribution of the Z_i.

[1]Da Veiga, S. (2024). *Tutorial on conformal prediction and related methods - ETICS 2024 Research School*. https://hal.science/hal-04690218 Fontana, M., Zeni, G., & Vantini, S. (2023).Conformal prediction: A unified review of theory and new challenges. *Bernoulli*, *29*(1), 1–23.

(Vanilla) conformal inference[1]

Framework : iid scalar data $\mathcal{Z} = (Z_1, \ldots, Z_n, Z_{n+1})$. Let $\alpha \in (0, 1)$.

- Aim : build $C_\alpha = C_\alpha(Z_1, \ldots, Z_n)$, a random set such that

$$\mathbb{P}(Z_{n+1} \in C_\alpha) \geq 1 - \alpha. \tag{1}$$

!!! Above we average over $(Z_1, \ldots, Z_{n+1})$!!!

- Construction of C_α : use known distribution. Conformal inference (CI) : discrete uniform distribution.
- Advantages of CI : valid whatever finite n, no assumption on the distribution of the Z_i.

[1]Da Veiga, S. (2024). *Tutorial on conformal prediction and related methods - ETICS 2024 Research School*. https://hal.science/hal-04690218 Fontana, M., Zeni, G., & Vantini, S. (2023).Conformal prediction: A unified review of theory and new challenges. *Bernoulli*, *29*(1), 1–23.

(Vanilla) conformal inference[1]

Framework : iid scalar data $\mathcal{Z} = (Z_1, \ldots, Z_n, Z_{n+1})$. Let $\alpha \in (0,1)$.

- Aim : build $C_\alpha = C_\alpha(Z_1, \ldots, Z_n)$, a random set such that

$$\mathbb{P}(Z_{n+1} \in C_\alpha) \geq 1 - \alpha. \tag{1}$$

!!! Above we average over $(Z_1, \ldots, Z_{n+1})$!!!

- Construction of C_α : use known distribution. Conformal inference (CI) : discrete uniform distribution.
- Advantages of CI : valid whatever finite n, no assumption on the distribution of the Z_i.

[1]Da Veiga, S. (2024). *Tutorial on conformal prediction and related methods - ETICS 2024 Research School*. https://hal.science/hal-04690218 Fontana, M., Zeni, G., & Vantini, S. (2023).Conformal prediction: A unified review of theory and new challenges. *Bernoulli*, *29*(1), 1–23.

Conformal inference : canonical example

- We assign a score to each Z_i. Here we set $S_i = S(Z_i) = |Z_i|$ and

$$\tau(Z_{n+1}) = \text{the rank of } S_{n+1} \text{ when sorting } (S_1, \ldots, S_{n+1})$$

(ascending order, assume no ties).

- From the exchangeability of the vector $(S_1, \ldots, S_{n+1})$, the distribution of $\tau(Z_{n+1})$ is $\mathcal{U}(\{1, \ldots, n+1\})$.

- We then set

$$C_\alpha := \{z \in \mathbb{R} : \tau(z) \leq n_\alpha\}, \tag{2}$$

where $n_\alpha = \lceil (1-\alpha)(n+1) \rceil$ and $\tau(z) = $ rank of $|z|$ when sorting $(S_1, \ldots, S_n, |z|)$. Indeed,

$$\mathbb{P}(Z_{n+1} \in C_\alpha) = \mathbb{P}(\tau(Z_{n+1}) \leq n_\alpha) \geq 1 - \alpha \; !$$

Conformal inference : canonical example

- We assign a score to each Z_i. Here we set $S_i = S(Z_i) = |Z_i|$ and

$$\tau(Z_{n+1}) = \text{the rank of } S_{n+1} \text{ when sorting } (S_1, \ldots, S_{n+1})$$

(ascending order, assume no ties).

- From the exchangeability of the vector $(S_1, \ldots, S_{n+1})$, the distribution of $\tau(Z_{n+1})$ is $\mathcal{U}(\{1, \ldots, n+1\})$.

- We then set

$$C_\alpha := \{z \in \mathbb{R} : \tau(z) \leq n_\alpha\}, \tag{2}$$

where $n_\alpha = \lceil (1-\alpha)(n+1) \rceil$ and $\tau(z) =$ rank of $|z|$ when sorting $(S_1, \ldots, S_n, |z|)$. Indeed,

$$\mathbb{P}(Z_{n+1} \in C_\alpha) = \mathbb{P}(\tau(Z_{n+1}) \leq n_\alpha) \geq 1 - \alpha \ !$$

Conformal inference : canonical example

- We assign a score to each Z_i. Here we set $S_i = S(Z_i) = |Z_i|$ and

$$\tau(Z_{n+1}) = \text{the rank of } S_{n+1} \text{ when sorting } (S_1, \ldots, S_{n+1})$$

  (ascending order, assume no ties).

- From the exchangeability of the vector $(S_1, \ldots, S_{n+1})$, the distribution of $\tau(Z_{n+1})$ is $\mathcal{U}(\{1, \ldots, n+1\})$.

- We then set

$$C_\alpha := \{z \in \mathbb{R} : \tau(z) \leq n_\alpha\}, \tag{2}$$

  where $n_\alpha = \lceil (1-\alpha)(n+1) \rceil$ and $\tau(z)$ = rank of $|z|$ when sorting $(S_1, \ldots, S_n, |z|)$. Indeed,

$$\mathbb{P}(Z_{n+1} \in C_\alpha) = \mathbb{P}(\tau(Z_{n+1}) \leq n_\alpha) \geq 1 - \alpha \ !$$

- C_α seems difficult to compute : to check whether $z \in C_\alpha$, we need to sort $(S_1, \ldots, S_n, |z|)$! $\exists$ trick : one shows that

$$C_\alpha = \left\{ z \in \mathbb{R} : |z| \leq S_{n_\alpha, n} \right\}, \tag{3}$$

where $S_{k,n}$ is the k^{th} order statistic of $(S_1, \ldots, S_n)$:

$$S_{1,n} < \ldots < S_{k,n} < \ldots < S_{n,n}. \tag{4}$$

- Because we've chosen $S(Z_i) = |Z_i|$, C_α is a ball.
- If we had set $S(Z_i) = Z_i$, C_α would be of the form $(-\infty, A]$.

Conformal inference : canonical example

- C_α seems difficult to compute : to check whether $z \in C_\alpha$, we need to sort $(S_1, \ldots, S_n, |z|)$! $\exists$ trick : one shows that

$$C_\alpha = \Big\{ z \in \mathbb{R} : |z| \leq S_{n_\alpha, n} \Big\}, \tag{3}$$

where $S_{k,n}$ is the k^{th} order statistic of $(S_1, \ldots, S_n)$:

$$S_{1,n} < \ldots < S_{k,n} < \ldots < S_{n,n}. \tag{4}$$

- Because we've chosen $S(Z_i) = |Z_i|$, C_α is a ball.
- If we had set $S(Z_i) = Z_i$, C_α would be of the form $(-\infty, A]$.

Conformal inference : canonical example

- C_α seems difficult to compute : to check whether $z \in C_\alpha$, we need to sort $(S_1, \ldots, S_n, |z|)$! $\exists$ trick : one shows that

$$C_\alpha = \Big\{ z \in \mathbb{R} : |z| \leq S_{n_\alpha, n} \Big\}, \tag{3}$$

where $S_{k,n}$ is the k^{th} order statistic of $(S_1, \ldots, S_n)$:

$$S_{1,n} < \ldots < S_{k,n} < \ldots < S_{n,n}. \tag{4}$$

- Because we've chosen $S(Z_i) = |Z_i|$, C_α is a ball.

- If we had set $S(Z_i) = Z_i$, C_α would be of the form $(-\infty, A]$.

Conformal inference : canonical example

- C_α seems difficult to compute : to check whether $z \in C_\alpha$, we need to sort $(S_1, \ldots, S_n, |z|)$! $\exists$ trick : one shows that

$$C_\alpha = \left\{ z \in \mathbb{R} : |z| \leq S_{n_\alpha, n} \right\}, \tag{3}$$

where $S_{k,n}$ is the k^{th} order statistic of $(S_1, \ldots, S_n)$:

$$S_{1,n} < \ldots < S_{k,n} < \ldots < S_{n,n}. \tag{4}$$

- Because we've chosen $S(Z_i) = |Z_i|$, C_α is a ball.

- If we had set $S(Z_i) = Z_i$, C_α would be of the form $(-\infty, A]$.

Conformal inference : multivariate framework

- If $Z_i \in \mathbb{R}^d$, no canonical order. We introduce a real-valued score function which generalises $S_i = |Z_i|$.
- Example : $S_i = \|Z_i\|_p \rightarrow p$-ball of $\mathbb{R}^d$.
- In general we can consider

$$S_i = S(Z_1, \dots, Z_{n+1}; Z_i), \quad 1 \leq i \leq n+1, \tag{5}$$

where S is a symmetrical function in its $n+1$ first arguments to preserve the exchangeability of $(S_1, \dots, S_{n+1}) \rightarrow$ different than a ball.

Conformal inference : multivariate framework

- If $Z_i \in \mathbb{R}^d$, no canonical order. We introduce a real-valued score function which generalises $S_i = |Z_i|$.
- Example : $S_i = \|Z_i\|_p \to p$-ball of $\mathbb{R}^d$.
- In general we can consider

$$S_i = S(Z_1, \dots, Z_{n+1}; Z_i), \quad 1 \leq i \leq n+1, \tag{5}$$

where S is a symmetrical function in its $n+1$ first arguments to preserve the exchangeability of $(S_1, \dots, S_{n+1}) \to$ different than a ball.

Outline of the talk

1. Background on conformal inference
   - Conformal inference
   - Split conformal inference in regression
2. A score based on the empirical covariance
   - Construction of adaptive scores
   - Non asymptotic ellipsoidal confidence region
3. Asymptotic study
   - Asymptotic ellipsoid $\mathcal{F}_\alpha$ (and $\mathcal{E}_\alpha$)
   - Comparison with the standard score for elliptical distributions
4. Numerical applications
   - Elliptical data : Gaussian data, Cauchy data
   - Non elliptical data : inverse Dirichlet data
5. Perspectives

Split conformal inference in regression[2]

In regression, $Z_i = (X_i, Y_i)$ and we wish to estimate $Y_i \in \mathbb{R}^\ell$ given $X_i \in \mathbb{R}^k$.

- We have a predictor $\hat{Y}_i = f(X_i)$ (considered deterministic in this talk).
- We consider $(Z_1, \ldots, Z_n, Z_{n+1})$ exchangeable, with $(Z_1, \ldots, Z_n)$ already observed.
- We consider X_{n+1}, we compute $\hat{Y}_{n+1} = f(X_{n+1})$. This time we wish to construct $C_\alpha = C_\alpha((X_1, Y_1), \ldots, (X_n, Y_n); X_{n+1})$ such that

$$\mathbb{P}(Y_{n+1} \in C_\alpha) \geq 1 - \alpha. \tag{6}$$

$\rightarrow$ Uncertainty quantification on the prediction $\widehat{Y}_{n+1}$.

- We still need to choose a score function. Typically,

$$S_i = \| Y_i - \hat{Y}_i \| = \| R_i \|_{\mathbb{R}^\ell}, \quad R_i = Y_i - \widehat{Y}_i. \tag{7}$$

[2]Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018).Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.*, *113*(523), 1094–1111.

Split conformal inference in regression[2]

In regression, $Z_i = (X_i, Y_i)$ and we wish to estimate $Y_i \in \mathbb{R}^\ell$ given $X_i \in \mathbb{R}^k$.

- We have a predictor $\hat{Y}_i = f(X_i)$ (considered deterministic in this talk).
- We consider $(Z_1, \ldots, Z_n, Z_{n+1})$ exchangeable, with $(Z_1, \ldots, Z_n)$ already observed.
- We consider X_{n+1}, we compute $\hat{Y}_{n+1} = f(X_{n+1})$. This time we wish to construct $C_\alpha = C_\alpha((X_1, Y_1), \ldots, (X_n, Y_n); X_{n+1})$ such that

$$\mathbb{P}(Y_{n+1} \in C_\alpha) \geq 1 - \alpha. \tag{6}$$

$\rightarrow$ Uncertainty quantification on the prediction $\widehat{Y}_{n+1}$.

- We still need to choose a score function. Typically,

$$S_i = \|Y_i - \hat{Y}_i\| = \|R_i\|_{\mathbb{R}^\ell}, \quad R_i = Y_i - \widehat{Y}_i. \tag{7}$$

[2]Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018).Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.*, *113*(523), 1094–1111.

Split conformal inference in regression[2]

In regression, $Z_i = (X_i, Y_i)$ and we wish to estimate $Y_i \in \mathbb{R}^\ell$ given $X_i \in \mathbb{R}^k$.

- We have a predictor $\hat{Y}_i = f(X_i)$ (considered deterministic in this talk).
- We consider $(Z_1, \ldots, Z_n, Z_{n+1})$ exchangeable, with $(Z_1, \ldots, Z_n)$ already observed.
- We consider X_{n+1}, we compute $\hat{Y}_{n+1} = f(X_{n+1})$. This time we wish to construct $C_\alpha = C_\alpha((X_1, Y_1), \ldots, (X_n, Y_n); X_{n+1})$ such that

$$\mathbb{P}(Y_{n+1} \in C_\alpha) \geq 1 - \alpha. \quad (6)$$

$\rightarrow$ Uncertainty quantification on the prediction $\widehat{Y}_{n+1}$.

- We still need to choose a score function. Typically,

$$S_i = \|Y_i - \hat{Y}_i\| = \|R_i\|_{\mathbb{R}^\ell}, \quad R_i = Y_i - \widehat{Y}_i. \quad (7)$$

[2]Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018).Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.*, *113*(523), 1094–1111.

Split conformal inference in regression[2]

In regression, $Z_i = (X_i, Y_i)$ and we wish to estimate $Y_i \in \mathbb{R}^\ell$ given $X_i \in \mathbb{R}^k$.

- We have a predictor $\hat{Y}_i = f(X_i)$ (considered deterministic in this talk).
- We consider $(Z_1, \ldots, Z_n, Z_{n+1})$ exchangeable, with $(Z_1, \ldots, Z_n)$ already observed.
- We consider X_{n+1}, we compute $\hat{Y}_{n+1} = f(X_{n+1})$. This time we wish to construct $C_\alpha = C_\alpha((X_1, Y_1), \ldots, (X_n, Y_n); X_{n+1})$ such that

$$\mathbb{P}(Y_{n+1} \in C_\alpha) \geq 1 - \alpha. \tag{6}$$

$\rightarrow$ Uncertainty quantification on the prediction $\widehat{Y}_{n+1}$.

- We still need to choose a score function. Typically,

$$S_i = \|Y_i - \hat{Y}_i\| = \|R_i\|_{\mathbb{R}^\ell}, \quad R_i = Y_i - \widehat{Y}_i. \tag{7}$$

[2]Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018).Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.*, *113*(523), 1094–1111.

Split conformal inference in regression[2]

In regression, $Z_i = (X_i, Y_i)$ and we wish to estimate $Y_i \in \mathbb{R}^\ell$ given $X_i \in \mathbb{R}^k$.

- We have a predictor $\hat{Y}_i = f(X_i)$ (considered deterministic in this talk).
- We consider $(Z_1, \ldots, Z_n, Z_{n+1})$ exchangeable, with $(Z_1, \ldots, Z_n)$ already observed.
- We consider X_{n+1}, we compute $\hat{Y}_{n+1} = f(X_{n+1})$. This time we wish to construct $C_\alpha = C_\alpha((X_1, Y_1), \ldots, (X_n, Y_n); X_{n+1})$ such that

$$\mathbb{P}(Y_{n+1} \in C_\alpha) \geq 1 - \alpha. \tag{6}$$

$\rightarrow$ Uncertainty quantification on the prediction $\widehat{Y}_{n+1}$.

- We still need to choose a score function. Typically,

$$S_i = \|Y_i - \hat{Y}_i\| = \|R_i\|_{\mathbb{R}^\ell}, \quad R_i = Y_i - \widehat{Y}_i. \tag{7}$$

[2]Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018).Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.*, *113*(523), 1094–1111.

Split conformal inference in regression

- For this score, C_α is a sphere centered at $\hat{Y}_{n+1}$

  $\rightarrow$ Huge set if the predictor $\hat{Y}$ is bad.

- This score only exploits the distribution of R_i, when in fact we observe (X_i, Y_i), or equivalently, (X_i, R_i). In the score, we can try to exploit the distribution of the pair (X, R), or ideally the distribution of R given X.

$\rightarrow$ "Adaptive score".

Typical questions in conformal inference

- Conditional coverage properties[3] : can we ensure that

$$\mathbb{P}(Y_{n+1} \in C_\alpha | Z_1, \dots, Z_n) \geq 1 - \alpha? \quad (8)$$

  Yes in a Probably Approximately Correct (PAC) sense. What about

$$\mathbb{P}(Y_{n+1} \in C_\alpha | X_{n+1}) \geq 1 - \alpha? \quad (9)$$

  Impssible without additional assumptions!
- CI for time series[4]
- For today : adaptivity of $C_\alpha \to$ can one choose the score function S so that it adapts to the distribution of the Z_i ?

[3]Vovk, V. (2012).Conditional validity of inductive conformal predictors. *Proceedings of the Asian Conference on Machine Learning*, *25*, 475–490.

[4]Lin, Z., Trivedi, S., & Sun, J. (2022).Conformal prediction intervals with temporal dependence. *Transactions on Machine Learning Research*.

Adaptive regions C_α in CI : some references

- Balls with adaptive radius :[5] (time series).
- Copulas :[6] (regression),[7] (time series).
- Ellipsoidal sets based on $\text{Cov}(R_1)$:[8] (regression),[9] (one time series $\sim$ stationary)
- Ellipsoidal sets based on $\text{Cov}(R_1|X_1 = x)$:[10], k-NN.

---

[5]Stankevičiūtė, K., Alaa, A. M., & van der Schaar, M. (2021).Conformal time-series forecasting. *NeurIPS*, *34*, 6216–6228.

[6]Messoudi, S., Destercke, S., & Rousseau, S. (2021).Copula-based conformal prediction for multi-target regression. *Pattern Recognition*, *120*, 108101.

[7]Sun, S., & Yu, R. (2022). *Copula conformal prediction for multi-step time series forecasting* [arXiv preprint arXiv:2212.03281].

[8]Johnstone, C., & Cox, B. (2021).Conformal uncertainty sets for robust optimization. *COPPA*, *152*, 72–90.

[9]Xu, C., Jiang, H., & Xie, Y. (2024).Conformal prediction for multi-dimensional time series by ellipsoidal sets. *ICML*.

[10]Messoudi, S., Destercke, S., & Rousseau, S. (2022).Ellipsoidal conformal inference for multi-target regression. *COPPA*, *179*, 294–306.

Content of today's talk

We introduce a score function for CI with the following properties.

- Provides "adaptive" ellipsoidal confidence regions C_α, centered at a linear correction $\widetilde{Y}_{n+1}$ of the predictor $\widehat{Y}_{n+1}$.
- We recover certain aspects of linear regression, but without any linear model assumption.
- Asymptotic study of the regions when $n \to \infty$.
- Theoretical study of the regions C_α, and volumetric comparison with the balls obtained from the score $\|Y_i - \widehat{Y}_i\|$.

Content of today's talk

We introduce a score function for CI with the following properties.

- Provides "adaptive" ellipsoidal confidence regions C_α, centered at a linear correction $\widetilde{Y}_{n+1}$ of the predictor $\widehat{Y}_{n+1}$.
- We recover certain aspects of linear regression, but without any linear model assumption.
- Asymptotic study of the regions when $n \to \infty$.
- Theoretical study of the regions C_α, and volumetric comparison with the balls obtained from the score $\|Y_i - \widehat{Y}_i\|$.

Content of today's talk

We introduce a score function for CI with the following properties.

- Provides "adaptive" ellipsoidal confidence regions C_α, centered at a linear correction $\widetilde{Y}_{n+1}$ of the predictor $\widehat{Y}_{n+1}$.
- We recover certain aspects of linear regression, but without any linear model assumption.
- Asymptotic study of the regions when $n \to \infty$.
- Theoretical study of the regions C_α, and volumetric comparison with the balls obtained from the score $\|Y_i - \widehat{Y}_i\|$.

Content of today's talk

We introduce a score function for CI with the following properties.

- Provides "adaptive" ellipsoidal confidence regions C_α, centered at a linear correction $\widetilde{Y}_{n+1}$ of the predictor $\widehat{Y}_{n+1}$.
- We recover certain aspects of linear regression, but without any linear model assumption.
- Asymptotic study of the regions when $n \to \infty$.
- Theoretical study of the regions C_α, and volumetric comparison with the balls obtained from the score $\|Y_i - \widehat{Y}_i\|$.

On the volumetric study of CI

Up until $\sim$ 2022, was not of prime interest in the CI community. Since then new works have been done, e.g.

- [11] Conformal regions with minimum volume based on the likelihood;
- [12] use (unions of) p-norm balls where $p \in (0, +\infty)$ is learnt.

In our setting, we aim to perform CI for multivariate inputs ($X \in \mathbb{R}^k$) and outputs ($Y \in \mathbb{R}^\ell$). In this potentially high dimensional setting it is relevant to study parametric approaches. This is what we end up doing, by constructing CI prediction sets as tailored ellipsoids.

---

[11]Izbicki, R., Shimizu, G., & Stern, R. B. (2022).`CD-split` and `HPD-split`: Efficient conformal regions in high dimensions. *J. Mach. Learn. Res.*, *23*, Paper No. [87], 32.

[12]Braun, S., Aolaritei, L., Jordan, M. I., & Bach, F. (2025).Minimum volume conformal sets for multivariate regression. *arXiv preprint arXiv:2503.19068*.

Outline of the talk

1. Background on conformal inference
   - Conformal inference
   - Split conformal inference in regression
2. A score based on the empirical covariance
   - Construction of adaptive scores
   - Non asymptotic ellipsoidal confidence region
3. Asymptotic study
   - Asymptotic ellipsoid $\mathcal{F}_\alpha$ (and $\mathcal{E}_\alpha$)
   - Comparison with the standard score for elliptical distributions
4. Numerical applications
   - Elliptical data : Gaussian data, Cauchy data
   - Non elliptical data : inverse Dirichlet data
5. Perspectives

Construction of adaptive scores

- Goal : exploit the correlation between $R_i = Y_i - \widehat{Y}_i$ and X_i to obtain a narrower region $C_\alpha \to$ use $\mathrm{Cov}(V_1)$ in the score, $V_1 = (X_1, R_1)^\top$.

- Mahalanobis-type score : if $\mu = \mathbb{E}[V_1]$, $\Sigma = \mathrm{Cov}(V_1)$, we set

$$S_i = (V_i - \mu)^\top \Sigma^{-1} (V_i - \mu) = \|V_i\|^2_{\mathrm{Maha}}, \tag{10}$$

which generalises $S_i^{vanilla} = (R_i^\top R_i)^{1/2}$.

- But (μ, Σ) are unknown $\to$ use empirical counterparts :

$$\widehat{\mu}_n = \frac{1}{n+1} \sum_{i=1}^{n+1} V_i, \quad \widetilde{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n+1} (V_i - \widehat{\mu}_n)(V_i - \widehat{\mu}_n)^\top.$$

We then introduce our *joint* Mahalanobis score,

$$\begin{aligned} S_i &= (V_i - \widehat{\mu}_n)^\top \widetilde{\Sigma}_n^{-1} (V_i - \widehat{\mu}_n) = \|V_i\|^2_{\widehat{\mathrm{Maha}}} \\ &= S(V_1, \ldots, V_{n+1}; V_i). \end{aligned} \tag{11}$$

Construction of adaptive scores

- Goal : exploit the correlation between $R_i = Y_i - \widehat{Y}_i$ and X_i to obtain a narrower region $C_\alpha \to$ use $\mathrm{Cov}(V_1)$ in the score, $V_1 = (X_1, R_1)^\top$.

- Mahalanobis-type score : if $\mu = \mathbb{E}[V_1]$, $\Sigma = \mathrm{Cov}(V_1)$, we set

$$S_i = (V_i - \mu)^\top \Sigma^{-1} (V_i - \mu) = \|V_i\|^2_{\mathrm{Maha}}, \tag{10}$$

which generalises $S_i^{vanilla} = (R_i^\top R_i)^{1/2}$.

- But (μ, Σ) are unknown $\to$ use empirical counterparts :

$$\widehat{\mu}_n = \frac{1}{n+1} \sum_{i=1}^{n+1} V_i, \quad \widetilde{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n+1} (V_i - \widehat{\mu}_n)(V_i - \widehat{\mu}_n)^\top.$$

We then introduce our *joint* Mahalanobis score,

$$\begin{aligned} S_i = (V_i - \widehat{\mu}_n)^\top \widetilde{\Sigma}_n^{-1} (V_i - \widehat{\mu}_n) &= \|V_i\|^2_{\widehat{\mathrm{Maha}}} \\ &= S(V_1, \ldots, V_{n+1}; V_i). \end{aligned} \tag{11}$$

Construction of adaptive scores

- Goal : exploit the correlation between $R_i = Y_i - \widehat{Y}_i$ and X_i to obtain a narrower region $C_\alpha \rightarrow$ use $\mathrm{Cov}(V_1)$ in the score, $V_1 = (X_1, R_1)^\top$.

- Mahalanobis-type score : if $\mu = \mathbb{E}[V_1]$, $\Sigma = \mathrm{Cov}(V_1)$, we set

$$S_i = (V_i - \mu)^\top \Sigma^{-1} (V_i - \mu) = \| V_i \|^2_{\mathrm{Maha}}, \tag{10}$$

which generalises $S_i^{vanilla} = (R_i^\top R_i)^{1/2}$.

- But (μ, Σ) are unknown $\rightarrow$ use empirical counterparts :

$$\widehat{\mu}_n = \frac{1}{n+1} \sum_{i=1}^{n+1} V_i, \quad \widetilde{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n+1} (V_i - \widehat{\mu}_n)(V_i - \widehat{\mu}_n)^\top.$$

We then introduce our *joint* Mahalanobis score,

$$\begin{aligned} S_i &= (V_i - \widehat{\mu}_n)^\top \widetilde{\Sigma}_n^{-1} (V_i - \widehat{\mu}_n) = \| V_i \|^2_{\widehat{\mathrm{Maha}}} \\ &= S(V_1, \ldots, V_{n+1}; V_i). \end{aligned} \tag{11}$$

Construction of adaptive scores

- Goal : exploit the correlation between $R_i = Y_i - \widehat{Y}_i$ and X_i to obtain a narrower region $C_\alpha \to$ use $\mathrm{Cov}(V_1)$ in the score, $V_1 = (X_1, R_1)^\top$.
- Mahalanobis-type score : if $\mu = \mathbb{E}[V_1]$, $\Sigma = \mathrm{Cov}(V_1)$, we set

$$S_i = (V_i - \mu)^\top \Sigma^{-1} (V_i - \mu) = \| V_i \|^2_{\mathrm{Maha}}, \tag{10}$$

which generalises $S_i^{vanilla} = (R_i^\top R_i)^{1/2}$.

- But (μ, Σ) are unknown $\to$ use empirical counterparts :

$$\widehat{\mu}_n = \frac{1}{n+1} \sum_{i=1}^{n+1} V_i, \quad \widetilde{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n+1} (V_i - \widehat{\mu}_n)(V_i - \widehat{\mu}_n)^\top.$$

We then introduce our *joint* Mahalanobis score,

$$S_i = (V_i - \widehat{\mu}_n)^\top \widetilde{\Sigma}_n^{-1} (V_i - \widehat{\mu}_n) = \| V_i \|^2_{\widehat{\mathrm{Maha}}} \tag{11}$$

$$= S(V_1, \ldots, V_{n+1}; V_i).$$

Further interpretation of the score

If $W_{ij} = (V_i)_j - (n+1)^{-1} \sum_{k=1}^{n+1} (V_k)_j$ (centered data $W \in \mathcal{M}_{n+1,p}(\mathbb{R})$), then

$$\frac{S_i}{n} = [W(W^\top W)^{-1} W^\top]_{ii}, \quad \forall i \in \{1, \dots, n+1\}.$$

Thus, other than being a Mahalanobis metric, we see that

- S_i is similar to a *leverage score* $\rightarrow$ self-influence score in linear regression.
- S_i is related to an *orientation statistic of W*. Indeed, write the polar decomposition of W, $W = PT$ ($P =$ orientation, $T =$ modulus), then $PP^\top = W(W^\top W)^{-1} W^\top \in \mathcal{M}_{n+1}(\mathbb{R})$ is a random orthogonal projector with rank p ($V_i \in \mathbb{R}^p$).

Further interpretation of the score

If $W_{ij} = (V_i)_j - (n+1)^{-1} \sum_{k=1}^{n+1} (V_k)_j$ (centered data $W \in \mathcal{M}_{n+1,p}(\mathbb{R})$), then

$$\frac{S_i}{n} = [W(W^\top W)^{-1} W^\top]_{ii}, \quad \forall i \in \{1, \dots, n+1\}.$$

Thus, other than being a Mahalanobis metric, we see that

- S_i is similar to a *leverage score* $\rightarrow$ self-influence score in linear regression.
- S_i is related to an *orientation statistic of W*. Indeed, write the polar decomposition of W, $W = PT$ ($P =$ orientation, $T =$ modulus), then $PP^\top = W(W^\top W)^{-1} W^\top \in \mathcal{M}_{n+1}(\mathbb{R})$ is a random orthogonal projector with rank p ($V_i \in \mathbb{R}^p$).

Further interpretation of the score

If $W_{ij} = (V_i)_j - (n+1)^{-1} \sum_{k=1}^{n+1} (V_k)_j$ (centered data $W \in \mathcal{M}_{n+1,p}(\mathbb{R})$), then

$$\frac{S_i}{n} = [W(W^\top W)^{-1} W^\top]_{ii}, \quad \forall i \in \{1, \ldots, n+1\}.$$

Thus, other than being a Mahalanobis metric, we see that

- S_i is similar to a *leverage score* $\rightarrow$ self-influence score in linear regression.
- S_i is related to an *orientation statistic of* W. Indeed, write the polar decomposition of W, $W = PT$ ($P =$ orientation, $T =$ modulus), then $PP^\top = W(W^\top W)^{-1} W^\top \in \mathcal{M}_{n+1}(\mathbb{R})$ is a random orthogonal projector with rank p ($V_i \in \mathbb{R}^p$).

Outline of the talk

1. Background on conformal inference
   - Conformal inference
   - Split conformal inference in regression
2. A score based on the empirical covariance
   - Construction of adaptive scores
   - Non asymptotic ellipsoidal confidence region
3. Asymptotic study
   - Asymptotic ellipsoid $\mathcal{F}_\alpha$ (and $\mathcal{E}_\alpha$)
   - Comparison with the standard score for elliptical distributions
4. Numerical applications
   - Elliptical data : Gaussian data, Cauchy data
   - Non elliptical data : inverse Dirichlet data
5. Perspectives

Non asymptotic ellipsoidal confidence region

Using our score, we obtain

Theorem 1

Let $Z_1 = (X_1, Y_1), \ldots, Z_{n+1} = (X_{n+1}, Y_{n+1})$ be exchangeable, with $X_i \in \mathbb{R}^k$, $Y_i \in \mathbb{R}^\ell$. There exists explicit $\mathcal{A}_n \succ 0$, $\widetilde{Y}_{n+1}$ et $\rho_{n,\alpha} \in \mathbb{R}$ such that the ellipsoid

$$\mathcal{E}_\alpha^n := \{y \in \mathbb{R}^\ell : (y - \widetilde{Y}_{n+1})^\top \mathcal{A}_n^{-1} (y - \widetilde{Y}_{n+1}) \leq \rho_{n,\alpha}\} \tag{12}$$

verifies $\mathbb{P}(Y_{n+1} \in \mathcal{E}_\alpha^n) \geq 1 - \alpha$.

- Here, $\mathcal{E}_\alpha^n$ corresponds to the region C_α^n associated to the joint Mahalanobis score apart from a small conservative approximation.
- $\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widehat{A}_n X_{n+1} + \widehat{e}_n$, where $\widehat{A}_n X_{n+1} + \widehat{e}_n$ = linear regression of R_{n+1} on X_{n+1}, learnt on $(V_1, \ldots, V_n)$.

Non asymptotic ellipsoidal confidence region

Using our score, we obtain

Theorem 1

Let $Z_1 = (X_1, Y_1), \ldots, Z_{n+1} = (X_{n+1}, Y_{n+1})$ be exchangeable, with $X_i \in \mathbb{R}^k$, $Y_i \in \mathbb{R}^\ell$. There exists explicit $\mathcal{A}_n \succ 0$, $\widetilde{Y}_{n+1}$ et $\rho_{n,\alpha} \in \mathbb{R}$ such that the ellipsoid

$$\mathcal{E}_\alpha^n := \{y \in \mathbb{R}^\ell : (y - \widetilde{Y}_{n+1})^\top \mathcal{A}_n^{-1}(y - \widetilde{Y}_{n+1}) \leq \rho_{n,\alpha}\} \tag{12}$$

verifies $\mathbb{P}(Y_{n+1} \in \mathcal{E}_\alpha^n) \geq 1 - \alpha$.

- Here, $\mathcal{E}_\alpha^n$ corresponds to the region C_α^n associated to the joint Mahalanobis score apart from a small conservative approximation.
- $\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widehat{A}_n X_{n+1} + \widehat{e}_n$, where $\widehat{A}_n X_{n+1} + \widehat{e}_n$ = linear regression of R_{n+1} on X_{n+1}, learnt on $(V_1, \ldots, V_n)$.

Non asymptotic ellipsoidal confidence region

Using our score, we obtain

Theorem 1

Let $Z_1 = (X_1, Y_1), \ldots, Z_{n+1} = (X_{n+1}, Y_{n+1})$ be exchangeable, with $X_i \in \mathbb{R}^k$, $Y_i \in \mathbb{R}^\ell$. There exists explicit $\mathcal{A}_n \succ 0$, $\widetilde{Y}_{n+1}$ et $\rho_{n,\alpha} \in \mathbb{R}$ such that the ellipsoid

$$\mathcal{E}_\alpha^n := \{y \in \mathbb{R}^\ell : (y - \widetilde{Y}_{n+1})^\top \mathcal{A}_n^{-1} (y - \widetilde{Y}_{n+1}) \leq \rho_{n,\alpha}\} \tag{12}$$

verifies $\mathbb{P}(Y_{n+1} \in \mathcal{E}_\alpha^n) \geq 1 - \alpha$.

- Here, $\mathcal{E}_\alpha^n$ corresponds to the region C_α^n associated to the joint Mahalanobis score apart from a small conservative approximation.
- $\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widehat{A}_n X_{n+1} + \widehat{e}_n$, where $\widehat{A}_n X_{n+1} + \widehat{e}_n =$ linear regression of R_{n+1} on X_{n+1}, learnt on $(V_1, \ldots, V_n)$.

Non asymptotic ellipsoidal confidence region

Using our score, we obtain

Theorem 1

Let $Z_1 = (X_1, Y_1), \dots, Z_{n+1} = (X_{n+1}, Y_{n+1})$ *be exchangeable, with* $X_i \in \mathbb{R}^k,\ Y_i \in \mathbb{R}^\ell$*. There exists explicit* $\mathcal{A}_n \succ 0, \widetilde{Y}_{n+1}$ *et* $\rho_{n,\alpha} \in \mathbb{R}$ *such that the ellipsoid*

$$\mathcal{E}_\alpha^n := \{y \in \mathbb{R}^\ell : (y - \widetilde{Y}_{n+1})^\top \mathcal{A}_n^{-1}(y - \widetilde{Y}_{n+1}) \leq \rho_{n,\alpha}\} \tag{12}$$

verifies $\mathbb{P}(Y_{n+1} \in \mathcal{E}_\alpha^n) \geq 1 - \alpha$.

- Here, $\mathcal{E}_\alpha^n$ corresponds to the region C_α^n associated to the joint Mahalanobis score apart from a small conservative approximation.
- $\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widehat{A}_n X_{n+1} + \widehat{e}_n$, where $\widehat{A}_n X_{n+1} + \widehat{e}_n =$ linear regression of R_{n+1} on X_{n+1}, learnt on $(V_1, \dots, V_n)$.

Expression of the ellipsoid $\mathcal{E}_\alpha^n$

Consider the expression

$$\mathcal{E}_\alpha^n := \{y \in \mathbb{R}^\ell : (y - \widetilde{Y}_{n+1})^\top \mathcal{A}_n^{-1} (y - \widetilde{Y}_{n+1}) \leq \rho_{n,\alpha}\}. \tag{13}$$

We set

$$\overline{V}_n = \frac{1}{n}\sum_{i=1}^{n} V_i = \begin{pmatrix} \overline{X}_n \\ \overline{R}_n \end{pmatrix},$$

$$\widehat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n} (V_i - \overline{V}_n)(V_i - \overline{V}_n)^\top = \begin{pmatrix} \widehat{\Sigma}_n^{XX} & \widehat{\Sigma}_n^{XR} \\ \widehat{\Sigma}_n^{RX} & \widehat{\Sigma}_n^{RR} \end{pmatrix},$$

$$\mathcal{A}_n = \widehat{\Sigma}_n^{RR} - \widehat{\Sigma}_n^{RX}(\widehat{\Sigma}_n^{XX})^{-1}\widehat{\Sigma}_n^{XR} = \widehat{\Sigma}_n / \widehat{\Sigma}_n^{XX} \quad (Schur), \tag{14}$$

$$\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widetilde{R}_{n+1},$$

$$\widetilde{R}_{n+1} = \widehat{\Sigma}_n^{RX}(\widehat{\Sigma}_n^{XX})^{-1}(X_{n+1} - \overline{X}_n) + \overline{R}_n = \widehat{A}_n X_{n+1} + \widehat{e}_n \tag{15}$$

Expression of the ellipsoid $\mathcal{E}_\alpha^n$

Consider the expression

$$\mathcal{E}_\alpha^n := \{y \in \mathbb{R}^\ell : (y - \widetilde{Y}_{n+1})^\top \mathcal{A}_n^{-1}(y - \widetilde{Y}_{n+1}) \leq \rho_{n,\alpha}\}. \qquad (13)$$

We set

$$\overline{V}_n = \frac{1}{n}\sum_{i=1}^{n} V_i = \begin{pmatrix} \overline{X}_n \\ \overline{R}_n \end{pmatrix},$$

$$\widehat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n}(V_i - \overline{V}_n)(V_i - \overline{V}_n)^\top = \begin{pmatrix} \widehat{\Sigma}_n^{XX} & \widehat{\Sigma}_n^{XR} \\ \widehat{\Sigma}_n^{RX} & \widehat{\Sigma}_n^{RR} \end{pmatrix},$$

$$\mathcal{A}_n = \widehat{\Sigma}_n^{RR} - \widehat{\Sigma}_n^{RX}(\widehat{\Sigma}_n^{XX})^{-1}\widehat{\Sigma}_n^{XR} = \widehat{\Sigma}_n/\widehat{\Sigma}_n^{XX} \quad (Schur), \qquad (14)$$

$$\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widetilde{R}_{n+1},$$

$$\widetilde{R}_{n+1} = \widehat{\Sigma}_n^{RX}(\widehat{\Sigma}_n^{XX})^{-1}(X_{n+1} - \overline{X}_n) + \overline{R}_n = \widehat{A}_n X_{n+1} + \widehat{e}_n \qquad (15)$$

Expression of the ellipsoid $\mathcal{E}_\alpha^n$

Consider the expression

$$\mathcal{E}_\alpha^n := \{y \in \mathbb{R}^\ell : (y - \widetilde{Y}_{n+1})^\top \mathcal{A}_n^{-1}(y - \widetilde{Y}_{n+1}) \leq \rho_{n,\alpha}\}. \qquad (13)$$

We set

$$\overline{V}_n = \frac{1}{n}\sum_{i=1}^{n} V_i = \begin{pmatrix} \overline{X}_n \\ \overline{R}_n \end{pmatrix},$$

$$\widehat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n}(V_i - \overline{V}_n)(V_i - \overline{V}_n)^\top = \begin{pmatrix} \widehat{\Sigma}_n^{XX} & \widehat{\Sigma}_n^{XR} \\ \widehat{\Sigma}_n^{RX} & \widehat{\Sigma}_n^{RR} \end{pmatrix},$$

$$\mathcal{A}_n = \widehat{\Sigma}_n^{RR} - \widehat{\Sigma}_n^{RX}(\widehat{\Sigma}_n^{XX})^{-1}\widehat{\Sigma}_n^{XR} = \widehat{\Sigma}_n/\widehat{\Sigma}_n^{XX} \quad (Schur), \qquad (14)$$

$$\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widetilde{R}_{n+1},$$

$$\widetilde{R}_{n+1} = \widehat{\Sigma}_n^{RX}(\widehat{\Sigma}_n^{XX})^{-1}(X_{n+1} - \overline{X}_n) + \overline{R}_n = \widehat{A}_n X_{n+1} + \widehat{e}_n \qquad (15)$$

About the "radius" $\rho_{n,\alpha}$ ...

- With our score we obtain

$$\rho_{n,\alpha} = \widetilde{q}_{n,\alpha} - (X_{n+1} - \overline{X}_n)^\top (\widehat{\Sigma}_n^{XX})^{-1} (X_{n+1} - \overline{X}_n) \in \mathbb{R}.$$

In particular we can have $\rho_{n,\alpha} < 0$ (even for large n) i.e. $C_\alpha^n = \varnothing$...

- To solve this issue we introduce the *adjusted* Mahalanobis score : by writing $W = (W_X | W_R) \in \mathcal{M}_{n+1,k+\ell}$, $W_X \in \mathcal{M}_{n+1,k}$, set

$$S_i' = [W(W^\top W)^{-1} W^\top]_{ii} - [W_X (W_X^\top W_X)^{-1} W_X^\top]_{ii}.$$

This score yields a second ellipsoid $\mathcal{F}_\alpha^n$ with the same $\mathcal{A}_n$ and $\widetilde{Y}_{n+1}$ but a different radius $\widetilde{\rho}_{n,\alpha}$, for which we have $\widetilde{\rho}_{n,\alpha} \geq 0$ a.s., but

$$\mathbb{P}(\mathcal{F}_\alpha^n = \mathbb{R}^\ell) > 0, \quad \mathbb{P}(\mathcal{F}_\alpha^n = \mathbb{R}^\ell) \xrightarrow[n\to\infty]{} 0.$$

About the "radius" $\rho_{n,\alpha}$ ...

- With our score we obtain

$$\rho_{n,\alpha} = \widetilde{q}_{n,\alpha} - (X_{n+1} - \overline{X}_n)^\top (\widehat{\Sigma}_n^{XX})^{-1} (X_{n+1} - \overline{X}_n) \in \mathbb{R}.$$

In particular we can have $\rho_{n,\alpha} < 0$ (even for large n) i.e. $C_\alpha^n = \varnothing$...

- To solve this issue we introduce the *adjusted* Mahalanobis score : by writing $W = (W_X | W_R) \in \mathcal{M}_{n+1,k+\ell}$, $W_X \in \mathcal{M}_{n+1,k}$, set

$$S_i' = [W(W^\top W)^{-1} W^\top]_{ii} - [W_X(W_X^\top W_X)^{-1} W_X^\top]_{ii}.$$

This score yields a second ellipsoid $\mathcal{F}_\alpha^n$ with the same $\mathcal{A}_n$ and $\widetilde{Y}_{n+1}$ but a different radius $\widetilde{\rho}_{n,\alpha}$, for which we have $\widetilde{\rho}_{n,\alpha} \geq 0$ a.s., but

$$\mathbb{P}(\mathcal{F}_\alpha^n = \mathbb{R}^\ell) > 0, \quad \mathbb{P}(\mathcal{F}_\alpha^n = \mathbb{R}^\ell) \xrightarrow[n\to\infty]{} 0.$$

Where does this second score come from ?

With $W = (W_X | W_R) \in \mathcal{M}_{n+1,k+\ell}, W_X \in \mathcal{M}_{n+1,k}$, we defined

$$S_i' = [W(W^\top W)^{-1}W^\top]_{ii} - [W_X(W_X^\top W_X)^{-1}W_X^\top]_{ii}.$$

Why do this ? Using the **block inversion formula**, we have (in essence)

$$S_i' = [R_i - (\widehat{A}_n X_i + \widehat{e}_n)]^\top (\widehat{\Sigma}_n / \widehat{\Sigma}_n^{XX})^{-1} [R_i - (\widehat{A}_n X_i + \widehat{e}_n)] \geq 0.$$

That is, S_i' is the (squared) norm of the difference between R_i and its linear regressor $\widehat{A}_n X_i + \widehat{e}_n$, normalized by the corresponding residual covariance $\widehat{\Sigma}_n / \widehat{\Sigma}_n^{XX}$.

This score is thus much closer to a "conditional friendly score" than the first score S_i. In fact, in the Gaussian case,

$$\widehat{A}_n X_i + \widehat{e}_n = \text{ empirical estimate of } \mathbb{E}[R_i | X_i],$$

$$\widehat{\Sigma}_n / \widehat{\Sigma}_n^{XX} = \text{ empirical estimate of } \mathrm{Cov}(R_i | X_i).$$

Where does this second score come from ?

With $W = (W_X | W_R) \in \mathcal{M}_{n+1,k+\ell}, W_X \in \mathcal{M}_{n+1,k}$, we defined

$$S'_i = [W(W^\top W)^{-1} W^\top]_{ii} - [W_X(W_X^\top W_X)^{-1} W_X^\top]_{ii}.$$

Why do this ? Using the **block inversion formula**, we have (in essence)

$$S'_i = [R_i - (\widehat{A}_n X_i + \widehat{e}_n)]^\top (\widehat{\Sigma}_n / \widehat{\Sigma}_n^{XX})^{-1} [R_i - (\widehat{A}_n X_i + \widehat{e}_n)] \geq 0.$$

That is, S'_i is the (squared) norm of the difference between R_i and its linear regressor $\widehat{A}_n X_i + \widehat{e}_n$, normalized by the corresponding residual covariance $\widehat{\Sigma}_n / \widehat{\Sigma}_n^{XX}$.

This score is thus much closer to a "conditional friendly score" than the first score S_i. In fact, in the Gaussian case,

$$\widehat{A}_n X_i + \widehat{e}_n = \text{ empirical estimate of } \mathbb{E}[R_i | X_i],$$

$$\widehat{\Sigma}_n / \widehat{\Sigma}_n^{XX} = \text{ empirical estimate of } \operatorname{Cov}(R_i | X_i).$$

Where does this second score come from ?

With $W = (W_X|W_R) \in \mathcal{M}_{n+1,k+\ell}, W_X \in \mathcal{M}_{n+1,k}$, we defined

$$S_i' = [W(W^\top W)^{-1} W^\top]_{ii} - [W_X(W_X^\top W_X)^{-1} W_X^\top]_{ii}.$$

Why do this ? Using the **block inversion formula**, we have (in essence)

$$S_i' = [R_i - (\widehat{A}_n X_i + \widehat{e}_n)]^\top (\widehat{\Sigma}_n / \widehat{\Sigma}_n^{XX})^{-1} [R_i - (\widehat{A}_n X_i + \widehat{e}_n)] \geq 0.$$

That is, S_i' is the (squared) norm of the difference between R_i and its linear regressor $\widehat{A}_n X_i + \widehat{e}_n$, normalized by the corresponding residual covariance $\widehat{\Sigma}_n / \widehat{\Sigma}_n^{XX}$.

This score is thus much closer to a "conditional friendly score" than the first score S_i. In fact, in the Gaussian case,

$$\widehat{A}_n X_i + \widehat{e}_n = \text{ empirical estimate of } \mathbb{E}[R_i|X_i],$$
$$\widehat{\Sigma}_n / \widehat{\Sigma}_n^{XX} = \text{ empirical estimate of } \mathrm{Cov}(R_i|X_i).$$

Additional remarks

In the formula

$$\mathcal{E}_\alpha^n := \{y \in \mathbb{R}^\ell : (y - \widetilde{Y}_{n+1})^\top \mathcal{A}_n^{-1}(y - \widetilde{Y}_{n+1}) \leq \rho_{n,\alpha}\}$$

where $\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widetilde{R}_{n+1}$, we have without Gaussianity assumptions that

- $\widetilde{R}_{n+1}$ exactly corresponds to the estimator from linear multivariate regression (empirical conditional linear expectation).
- $\mathcal{A}_n$ exactly corresponds to the empirical linear conditional covariance (Schur complement), up to a mult. constant.

That is, the Mahalanobis score enables to recover formulas from linear regression, while also providing the non asymptotic probability coverage of CI, withour any assumptions on the distribution of the Z_i (no linear model nor normality assumption).

Additional remarks

In the formula

$$\mathcal{E}_\alpha^n := \{y \in \mathbb{R}^\ell : (y - \widetilde{Y}_{n+1})^\top \mathcal{A}_n^{-1}(y - \widetilde{Y}_{n+1}) \leq \rho_{n,\alpha}\}$$

where $\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widetilde{R}_{n+1}$, we have without Gaussianity assumptions that

- $\widetilde{R}_{n+1}$ exactly corresponds to the estimator from linear multivariate regression (empirical conditional linear expectation).
- $\mathcal{A}_n$ exactly corresponds to the empirical linear conditional covariance (Schur complement), up to a mult. constant.

That is, the Mahalanobis score enables to recover formulas from linear regression, while also providing the non asymptotic probability coverage of CI, withour any assumptions on the distribution of the Z_i (no linear model nor normality assumption).

Additional remarks

In the formula

$$\mathcal{E}_\alpha^n := \{y \in \mathbb{R}^\ell : (y - \widetilde{Y}_{n+1})^\top \mathcal{A}_n^{-1}(y - \widetilde{Y}_{n+1}) \leq \rho_{n,\alpha}\}$$

where $\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widetilde{R}_{n+1}$, we have without Gaussianity assumptions that

- $\widetilde{R}_{n+1}$ exactly corresponds to the estimator from linear multivariate regression (empirical conditional linear expectation).
- $\mathcal{A}_n$ exactly corresponds to the empirical linear conditional covariance (Schur complement), up to a mult. constant.

That is, the Mahalanobis score enables to recover formulas from linear regression, while also providing the non asymptotic probability coverage of CI, withour any assumptions on the distribution of the Z_i (no linear model nor normality assumption).

Sketch of proof for the first ellipsoid $\mathcal{E}_\alpha^n$ (1/2)

We show that our score does lead to the set $\mathcal{E}_\alpha^n$.

- Let $z \in \mathbb{R}^\ell$ be a candidate for R_{n+1} : $V_{n+1}(z) = \begin{pmatrix} X_{n+1} \\ z \end{pmatrix}$,

$$V(z) = \begin{pmatrix} V_1^\top \\ \vdots \\ V_n^\top \\ V_{n+1}(z)^\top \end{pmatrix}, \quad W(z) = \left(I - \frac{\mathbb{1}\mathbb{1}^\top}{n+1}\right) V(z) \quad \text{(centered data).}$$

- $S_i(z) = \left[W(z)\left(W(z)^\top W(z)\right)^{-1} W(z)^\top\right]_{ii} = [S(z)]_{ii}$.
- Sherman-Morrison : $W(z) = W(0) + v(Lz)^\top$, rewrite $S(z)$ as

$$S(z) = C - a(z)a(z)^\top, \quad S_i(z) = C_{ii} - a_i(z)^2 \quad (!).$$

Sketch of proof for the first ellipsoid $\mathcal{E}_\alpha^n$ (1/2)

We show that our score does lead to the set $\mathcal{E}_\alpha^n$.

- Let $z \in \mathbb{R}^\ell$ be a candidate for R_{n+1} : $V_{n+1}(z) = \begin{pmatrix} X_{n+1} \\ z \end{pmatrix}$,

$$V(z) = \begin{pmatrix} V_1^\top \\ \vdots \\ V_n^\top \\ V_{n+1}(z)^\top \end{pmatrix}, \quad W(z) = \left(I - \frac{\mathbb{1}\mathbb{1}^\top}{n+1}\right) V(z) \quad \text{(centered data).}$$

- $S_i(z) = \left[W(z)\left(W(z)^\top W(z)\right)^{-1} W(z)^\top\right]_{ii} = [S(z)]_{ii}$.

- Sherman-Morrison : $W(z) = W(0) + v(Lz)^\top$, rewrite $S(z)$ as

$$S(z) = C - a(z)a(z)^\top, \quad S_i(z) = C_{ii} - a_i(z)^2 \quad (!).$$

Sketch of proof for the first ellipsoid $\mathcal{E}_\alpha^n$ (1/2)

We show that our score does lead to the set $\mathcal{E}_\alpha^n$.

- Let $z \in \mathbb{R}^\ell$ be a candidate for R_{n+1} : $V_{n+1}(z) = \begin{pmatrix} X_{n+1} \\ z \end{pmatrix}$,

$$V(z) = \begin{pmatrix} V_1^\top \\ \vdots \\ V_n^\top \\ V_{n+1}(z)^\top \end{pmatrix}, \quad W(z) = \left(I - \frac{\mathbb{1}\mathbb{1}^\top}{n+1}\right) V(z) \quad \text{(centered data)}.$$

- $S_i(z) = \left[W(z)\big(W(z)^\top W(z)\big)^{-1} W(z)^\top\right]_{ii} = [S(z)]_{ii}$.
- Sherman-Morrison : $W(z) = W(0) + v(Lz)^\top$, rewrite $S(z)$ as

$$S(z) = C - a(z)a(z)^\top, \quad S_i(z) = C_{ii} - a_i(z)^2 \quad \text{(!)}.$$

Sketch of proof for the first ellipsoid $\mathcal{E}_\alpha^n$ (2/2)

- C_α^n is given by

$$C_\alpha^n = \widehat{Y}_{n+1} + \{z : S_{n+1}(z) \leq S_{n_\alpha,n}(z)\},$$

where $S_{k,n}(z)$ is the k^{th} order statistic of $(S_1(z), \ldots, S_n(z))$:

$$S_{1,n}(z) < \ldots < S_{k,n}(z) < \ldots < S_{n,n}(z).$$

- Trick : the conservative bound $S_i(z) \leq C_{ii}, 1 \leq i \leq n$, yields

$$C_\alpha^n \subset \mathcal{E}_\alpha^n := \widehat{Y}_{n+1} + \{z : S_{n+1}(z) \leq C_{n_\alpha,n}\}, \quad (16)$$

where $C_{k,n}$ is the k^{th} order statistic of $(C_{11}, \ldots, C_{nn})$.

- We show that $\mathcal{E}_\alpha^n$ is the announced ellipsoid, QED.
- If $\mathbb{E}[\|V_1\|^{4q}] < +\infty$ for $q > 1$, then $|\mathrm{Vol}(C_\alpha^n) - \mathrm{Vol}(\mathcal{E}_\alpha^n)| \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Sketch of proof for the first ellipsoid $\mathcal{E}_\alpha^n$ (2/2)

- C_α^n is given by

$$C_\alpha^n = \widehat{Y}_{n+1} + \{z : S_{n+1}(z) \leq S_{n_\alpha,n}(z)\},$$

where $S_{k,n}(z)$ is the k^{th} order statistic of $(S_1(z), \ldots, S_n(z))$:

$$S_{1,n}(z) < \ldots < S_{k,n}(z) < \ldots < S_{n,n}(z).$$

- Trick : the conservative bound $S_i(z) \leq C_{ii}, 1 \leq i \leq n$, yields

$$C_\alpha^n \subset \mathcal{E}_\alpha^n := \widehat{Y}_{n+1} + \{z : S_{n+1}(z) \leq C_{n_\alpha,n}\}, \quad (16)$$

where $C_{k,n}$ is the k^{th} order statistic of $(C_{11}, \ldots, C_{nn})$.

- We show that $\mathcal{E}_\alpha^n$ is the announced ellipsoid, QED.
- If $\mathbb{E}[\|V_1\|^{4q}] < +\infty$ for $q > 1$, then $|\mathrm{Vol}(C_\alpha^n) - \mathrm{Vol}(\mathcal{E}_\alpha^n)| \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Sketch of proof for the first ellipsoid $\mathcal{E}_\alpha^n$ (2/2)

- C_α^n is given by

$$C_\alpha^n = \widehat{Y}_{n+1} + \{z : S_{n+1}(z) \leq S_{n_\alpha,n}(z)\},$$

where $S_{k,n}(z)$ is the k^{th} order statistic of $(S_1(z), \dots, S_n(z))$:

$$S_{1,n}(z) < \dots < S_{k,n}(z) < \dots < S_{n,n}(z).$$

- Trick : the conservative bound $S_i(z) \leq C_{ii}, 1 \leq i \leq n$, yields

$$C_\alpha^n \subset \mathcal{E}_\alpha^n := \widehat{Y}_{n+1} + \{z : S_{n+1}(z) \leq C_{n_\alpha,n}\}, \tag{16}$$

where $C_{k,n}$ is the k^{th} order statistic of $(C_{11}, \dots, C_{nn})$.

- We show that $\mathcal{E}_\alpha^n$ is the announced ellipsoid, QED.
- If $\mathbb{E}[\|V_1\|^{4q}] < +\infty$ for $q > 1$, then $|\mathrm{Vol}(C_\alpha^n) - \mathrm{Vol}(\mathcal{E}_\alpha^n)| \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Outline of the talk

1. Background on conformal inference
   - Conformal inference
   - Split conformal inference in regression
2. A score based on the empirical covariance
   - Construction of adaptive scores
   - Non asymptotic ellipsoidal confidence region
3. Asymptotic study
   - Asymptotic ellipsoid $\mathcal{F}_\alpha$ (and $\mathcal{E}_\alpha$)
   - Comparison with the standard score for elliptical distributions
4. Numerical applications
   - Elliptical data : Gaussian data, Cauchy data
   - Non elliptical data : inverse Dirichlet data
5. Perspectives

Asymptotic confidence ellipsoid 1/3

We focus on $\mathcal{F}_\alpha^n$ (the one that is never empty). We study the regime $n \to +\infty$, assuming that $\mathbb{E}[\|V_1\|^2] < +\infty$.

- We show that $\mathcal{F}_\alpha^n$ "converges" towards a random ellipsoid $\mathcal{F}_\alpha$,

$$\mathcal{F}_\alpha = \widehat{Y}_1 + \{z : (z - Z_0)^\top \mathcal{A}^{-1}(z - Z_0) \leq \widetilde{\rho}_\alpha\}.$$

- Center $\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widehat{A}_n X_{n+1} + \widehat{e}_n \xrightarrow{\mathcal{L}} \widehat{Y}_1 + \underbrace{A^* X_1 + e^*}_{Z_0}$, where

$$\begin{aligned}(A^*, e^*) &= \arg\min_{(A,e)} \mathbb{E}\left[\|R_1 - AX_1 - e\|^2\right] \\ &= \arg\min_{\text{affine } f} \mathbb{E}\left[\|R_1 - f(X_1)\|^2\right],\end{aligned}$$

i.e. $A^* X_1 + e^*$ is the linear conditional expectation of R_1 given X_1.

Asymptotic confidence ellipsoid 1/3

We focus on $\mathcal{F}_\alpha^n$ (the one that is never empty). We study the regime $n \to +\infty$, assuming that $\mathbb{E}[\|V_1\|^2] < +\infty$.

- We show that $\mathcal{F}_\alpha^n$ "converges" towards a random ellipsoid $\mathcal{F}_\alpha$,

$$\mathcal{F}_\alpha = \widehat{Y}_1 + \{z : (z - Z_0)^\top \mathcal{A}^{-1}(z - Z_0) \leq \widetilde{\rho}_\alpha\}.$$

- Center $\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widehat{A}_n X_{n+1} + \widehat{e}_n \xrightarrow{\mathcal{L}} \widehat{Y}_1 + \underbrace{A^* X_1 + e^*}_{Z_0}$, where

$$(A^*, e^*) = \arg\min_{(A,e)} \mathbb{E}\left[\|R_1 - AX_1 - e\|^2\right]$$
$$= \arg\min_{\text{affine } f} \mathbb{E}\left[\|R_1 - f(X_1)\|^2\right],$$

i.e. $A^* X_1 + e^*$ is the linear conditional expectation of R_1 given X_1.

Asymptotic confidence ellipsoid 1/3

We focus on $\mathcal{F}_\alpha^n$ (the one that is never empty). We study the regime $n \to +\infty$, assuming that $\mathbb{E}[\|V_1\|^2] < +\infty$.

- We show that $\mathcal{F}_\alpha^n$ "converges" towards a random ellipsoid $\mathcal{F}_\alpha$,

$$\mathcal{F}_\alpha = \widehat{Y}_1 + \{z : (z - Z_0)^\top \mathcal{A}^{-1}(z - Z_0) \le \widetilde{\rho}_\alpha\}.$$

- Center $\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widehat{A}_n X_{n+1} + \widehat{e}_n \xrightarrow{\mathcal{L}} \widehat{Y}_1 + \underbrace{A^* X_1 + e^*}_{Z_0}$, where

$$(A^*, e^*) = \underset{(A,e)}{\arg\min}\, \mathbb{E}\left[\|R_1 - AX_1 - e\|^2\right]$$

$$= \underset{\text{affine } f}{\arg\min}\, \mathbb{E}\left[\|R_1 - f(X_1)\|^2\right],$$

i.e. $A^* X_1 + e^*$ is the linear conditional expectation of R_1 given X_1.

Asymptotic confidence ellipsoid 1/3

We focus on $\mathcal{F}_\alpha^n$ (the one that is never empty). We study the regime $n \to +\infty$, assuming that $\mathbb{E}[\|V_1\|^2] < +\infty$.

- We show that $\mathcal{F}_\alpha^n$ "converges" towards a random ellipsoid $\mathcal{F}_\alpha$,

$$\mathcal{F}_\alpha = \widehat{Y}_1 + \{z : (z - Z_0)^\top \mathcal{A}^{-1}(z - Z_0) \leq \widetilde{\rho}_\alpha\}.$$

- Center $\widetilde{Y}_{n+1} = \widehat{Y}_{n+1} + \widehat{A}_n X_{n+1} + \widehat{e}_n \xrightarrow{\mathcal{L}} \widehat{Y}_1 + \underbrace{A^* X_1 + e^*}_{Z_0}$, where

$$\begin{aligned}(A^*, e^*) &= \underset{(A,e)}{\arg\min}\, \mathbb{E}\left[\|R_1 - AX_1 - e\|^2\right] \\ &= \underset{\text{affine } f}{\arg\min}\, \mathbb{E}\left[\|R_1 - f(X_1)\|^2\right],\end{aligned}$$

i.e. $A^* X_1 + e^*$ is the linear conditional expectation of R_1 given X_1.

Asymptotic confidence ellipsoid 2/3

- Moreover, $\mathcal{A}_n \xrightarrow{a.s.} \mathcal{A}$, with

$$\mathrm{Cov}(R_1 - A^* X_1 - e^*) = \mathcal{A} = \Sigma^{RR} - \Sigma^{RX} \left(\Sigma^{XX}\right)^{-1} \Sigma^{RX},$$

where $\Sigma = \mathrm{Cov}(V_1)$.

- These results remain true for finite n when replacing $\mathbb{E}[\cdot]$ with an empirical average on the n available examples $(V_1, \ldots, V_n)$, e.g. :

$$(\widehat{A}_n, \widehat{e}_n) = \arg\min_{(A,e)} \frac{1}{n} \sum_{i=1}^{n} \|R_i - AX_i - e\|^2.$$

Asymptotic confidence ellipsoid 2/3

- Moreover, $\mathcal{A}_n \xrightarrow{a.s.} \mathcal{A}$, with

$$\mathrm{Cov}(R_1 - A^* X_1 - e^*) = \mathcal{A} = \Sigma^{RR} - \Sigma^{RX}\left(\Sigma^{XX}\right)^{-1}\Sigma^{RX},$$

where $\Sigma = \mathrm{Cov}(V_1)$.

- These results remain true for finite n when replacing $\mathbb{E}[\cdot]$ with an empirical average on the n available examples $(V_1, \ldots, V_n)$, e.g. :

$$(\widehat{A}_n, \widehat{e}_n) = \underset{(A,e)}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \|R_i - AX_i - e\|^2.$$

Asymptotic confidence ellipsoid 3/3

If additionally $\mathbb{E}[||X_1||^{4q}] < +\infty$ for some $q > 1$, then

$$\widetilde{\rho}_{n,\alpha} \xrightarrow[n\to\infty]{\mathbb{P}} q^{\mathcal{F}}_{1-\alpha},$$

where

$$\begin{aligned} q^{\mathcal{F}}_{1-\alpha} &= q_{1-\alpha}(T^\top \mathcal{A}^{-1} T), \\ T &= R_1 - \mathbb{E}[R_1] - \Sigma^{RX}(\Sigma^{XX})^{-1}(X_1 - \mathbb{E}[X_1]). \end{aligned}$$

Basically, $q^{\mathcal{F}}_{1-\alpha}$ is the scalar such that

$$\mathbb{P}(Y_1 \in \mathcal{F}_\alpha) = 1 - \alpha,$$

i.e. the limit set $\mathcal{F}_\alpha$ has exactly a coverage of $1 - \alpha$.

Asymptotic confidence ellipsoid 3/3

If additionally $\mathbb{E}[||X_1||^{4q}] < +\infty$ for some $q > 1$, then

$$\widetilde{\rho}_{n,\alpha} \xrightarrow[n\to\infty]{\mathbb{P}} q^{\mathcal{F}}_{1-\alpha},$$

where

$$\begin{aligned} q^{\mathcal{F}}_{1-\alpha} &= q_{1-\alpha}(T^\top \mathcal{A}^{-1} T), \\ T &= R_1 - \mathbb{E}[R_1] - \Sigma^{RX}(\Sigma^{XX})^{-1}(X_1 - \mathbb{E}[X_1]). \end{aligned}$$

Basically, $q^{\mathcal{F}}_{1-\alpha}$ is the scalar such that

$$\mathbb{P}(Y_1 \in \mathcal{F}_\alpha) = 1 - \alpha,$$

i.e. the limit set $\mathcal{F}_\alpha$ has exactly a coverage of $1 - \alpha$.

Outline of the talk

Comparison with the score $\|R_i\|^2$ for elliptical distributions

Assume that $\mathbb{E}[\|V_1\|^2] < +\infty$.

- We have shown that for the Mahalanobis score, $\mathcal{E}_\alpha^n$ "converges" towards a random ellipsoid $\mathcal{E}_\alpha$.
- Likewise for the score $S_i = \|R_i\|^2 = \|Y_i - \widehat{Y}_i\|^2$, we obtain a ball $\mathcal{B}_\alpha$ when $n \to \infty$.

We wish to compare the volumes of $\mathcal{F}_\alpha$ and $\mathcal{B}_\alpha$ → when V_1 follows an elliptical distribution.

- $V_1 \in \mathbb{R}^{k+\ell}$ has an elliptical distribution is V_1 has a density of the form

$$f(z) = g((z-\mu)^\top \Sigma^{-1}(z-\mu)) \quad (17)$$

where $g : \mathbb{R}^+ \to \mathbb{R}^+,\ \mu \in \mathbb{R}^{k+\ell},\ \Sigma \succ 0$ (elliptical level sets).

- Ex : multidim. Gaussian, Cauchy, Student...

Assume that $\mathbb{E}[\|V_1\|^2] < +\infty$.

- We have shown that for the Mahalanobis score, $\mathcal{E}_\alpha^n$ "converges" towards a random ellipsoid $\mathcal{E}_\alpha$.
- Likewise for the score $S_i = \|R_i\|^2 = \|Y_i - \widehat{Y}_i\|^2$, we obtain a ball $\mathcal{B}_\alpha$ when $n \to \infty$.

We wish to compare the volumes of $\mathcal{F}_\alpha$ and $\mathcal{B}_\alpha$ → when V_1 follows an elliptical distribution.

- $V_1 \in \mathbb{R}^{k+\ell}$ has an elliptical distribution is V_1 has a density of the form

$$f(z) = g((z-\mu)^\top \Sigma^{-1}(z-\mu)) \tag{17}$$

where $g : \mathbb{R}^+ \to \mathbb{R}^+,\ \mu \in \mathbb{R}^{k+\ell},\ \Sigma \succ 0$ (elliptical level sets).

- Ex : multidim. Gaussian, Cauchy, Student...

Comparison with the score $\|R_i\|^2$ for elliptical distributions

Assume that $\mathbb{E}[\|V_1\|^2] < +\infty$.

- We have shown that for the Mahalanobis score, $\mathcal{E}_\alpha^n$ "converges" towards a random ellipsoid $\mathcal{E}_\alpha$.
- Likewise for the score $S_i = \|R_i\|^2 = \|Y_i - \widehat{Y}_i\|^2$, we obtain a ball $\mathcal{B}_\alpha$ when $n \to \infty$.

We wish to compare the volumes of $\mathcal{F}_\alpha$ and $\mathcal{B}_\alpha$ $\to$ when V_1 follows an elliptical distribution.

- $V_1 \in \mathbb{R}^{k+\ell}$ has an elliptical distribution is V_1 has a density of the form

$$f(z) = g((z-\mu)^\top \Sigma^{-1}(z-\mu)) \tag{17}$$

  where $g : \mathbb{R}^+ \to \mathbb{R}^+,\ \mu \in \mathbb{R}^{k+\ell},\ \Sigma \succ 0$ (elliptical level sets).

- Ex : multidim. Gaussian, Cauchy, Student...

Comparison with the score $\|R_i\|^2$ for elliptical distributions

Assume that $k = 0$ (X_i empty and $\mathcal{E}_\alpha = \mathcal{F}_\alpha$) : $V_i = R_i = Y_i - \widehat{Y}_i$, and that V_1 follows an elliptical distrib. ($f(z) = g((z-\mu)^\top \Sigma^{-1}(z-\mu))$).

Theorem 2

If g is decreasing, then $\mathrm{Vol}(\mathcal{E}_\alpha) \leq \mathrm{Vol}(\mathcal{B}_\alpha)$.

This result is a consequence of a more general result on CDFs.

Theorem 3 (F. Barthe, IMT)

Assume that $U \in \mathbb{R}^\ell$ has a density of the form $g(\|z\|^2)$ (spherical distrib.). Let $t \geq 0$. If g is decreasing, then

$$(0, I_\ell) = \underset{\mu\in\mathbb{R}^\ell,\ \Sigma\succ 0,\ \det(\Sigma)=1}{\arg\max} \mathbb{P}(\|\Sigma^{1/2}U + \mu\|^2 \leq t). \quad (18)$$

This is a consequence of the fact that centered balls are level sets of centered spherical distributions.

Comparison with the score $\|R_i\|^2$ for elliptical distributions

Assume that $k = 0$ (X_i empty and $\mathcal{E}_\alpha = \mathcal{F}_\alpha$) : $V_i = R_i = Y_i - \widehat{Y}_i$, and that V_1 follows an elliptical distrib. ($f(z) = g((z-\mu)^\top \Sigma^{-1}(z-\mu))$).

Theorem 2

If g is decreasing, then $\mathrm{Vol}(\mathcal{E}_\alpha) \leq \mathrm{Vol}(\mathcal{B}_\alpha)$.

This result is a consequence of a more general result on CDFs.

Theorem 3 (F. Barthe, IMT)

Assume that $U \in \mathbb{R}^\ell$ has a density of the form $g(\|z\|^2)$ (spherical distrib.). Let $t \geq 0$. If g is decreasing, then

$$(0, I_\ell) = \underset{\mu \in \mathbb{R}^\ell,\ \Sigma \succ 0,\ \det(\Sigma)=1}{\arg\max} \mathbb{P}(\|\Sigma^{1/2}U + \mu\|^2 \leq t). \tag{18}$$

This is a consequence of the fact that centered balls are level sets of centered spherical distributions.

Comparison with the score $\|R_i\|^2$ for elliptical distributions

Assume that $k = 0$ (X_i empty and $\mathcal{E}_\alpha = \mathcal{F}_\alpha$) : $V_i = R_i = Y_i - \widehat{Y}_i$, and that V_1 follows an elliptical distrib. ($f(z) = g((z-\mu)^\top \Sigma^{-1}(z-\mu))$).

Theorem 2

If g is decreasing, then $\mathrm{Vol}(\mathcal{E}_\alpha) \leq \mathrm{Vol}(\mathcal{B}_\alpha)$.

This result is a consequence of a more general result on CDFs.

Theorem 3 (F. Barthe, IMT)

Assume that $U \in \mathbb{R}^\ell$ has a density of the form $g(\|z\|^2)$ (spherical distrib.). Let $t \geq 0$. If g is decreasing, then

$$(0, I_\ell) = \underset{\mu \in \mathbb{R}^\ell,\ \Sigma \succ 0,\ \det(\Sigma)=1}{\arg\max} \mathbb{P}(\|\Sigma^{1/2} U + \mu\|^2 \leq t). \tag{18}$$

This is a consequence of the fact that centered balls are level sets of centered spherical distributions.

Comparison with the score $\|R_i\|^2$ for elliptical distributions

Note : In the case $k > 0$ the same arguments also work for $\mathcal{F}_\alpha$.

Theorem 4

If $(X, R)^\top \in \mathbb{R}^{k+\ell}$ has an elliptical distribution and if g is decreasing, then $\mathrm{Vol}(\mathcal{E}_\alpha) \leq \mathrm{Vol}(\mathcal{B}_\alpha)$.

For $\mathcal{E}_\alpha$ the condition is more complex; see paper if interested...

Finally we have the additional conditional coverage result for $\mathcal{F}_\alpha$.

Proposition 1

If $V_1 \sim \mathcal{N}(\mu, \Sigma)$ and $\min Spec\, \Sigma > 0$, then

$$\forall x \in \mathbb{R}^k, \quad \mathbb{P}(Y_1 - \widehat{Y}_1 \in \mathcal{F}_\alpha^\infty | X_1 = x) = \mathbb{P}(Y_1 - \widehat{Y}_1 \in \mathcal{F}_\alpha^\infty) = 1 - \alpha.$$

This result is false for general elliptical distributions.

Some conclusions

- Under certain assumptions, we have shown that for elliptical distributions,

$$\mathrm{Vol}(\mathcal{F}_\alpha) \leq \mathrm{Vol}(\mathcal{B}_\alpha).$$

- We expect that some of our results remain true when $\mathbb{E}[\|V_1\|] = +\infty$ (see e.g. orthogonal projector interpretation, Cauchy data experiment).
- We expect that these results remain true when V_1 is close to an elliptical distrib. (...).
- Still, for non elliptical distrib., we may construct examples for which $\mathrm{Vol}(\mathcal{B}_\alpha) < \mathrm{Vol}(\mathcal{E}_\alpha)$, whatever k.

Some conclusions

- Under certain assumptions, we have shown that for elliptical distributions,

$$\mathrm{Vol}(\mathcal{F}_\alpha) \leq \mathrm{Vol}(\mathcal{B}_\alpha).$$

- We expect that some of our results remain true when $\mathbb{E}[\|V_1\|] = +\infty$ (see e.g. orthogonal projector interpretation, Cauchy data experiment).
- We expect that these results remain true when V_1 is close to an elliptical distrib. (...).
- Still, for non elliptical distrib., we may construct examples for which $\mathrm{Vol}(\mathcal{B}_\alpha) < \mathrm{Vol}(\mathcal{E}_\alpha)$, whatever k.

Some conclusions

- Under certain assumptions, we have shown that for elliptical distributions,

$$\mathrm{Vol}(\mathcal{F}_\alpha) \leq \mathrm{Vol}(\mathcal{B}_\alpha).$$

- We expect that some of our results remain true when $\mathbb{E}[\|V_1\|] = +\infty$ (see e.g. orthogonal projector interpretation, Cauchy data experiment).
- We expect that these results remain true when V_1 is close to an elliptical distrib. (...).
- Still, for non elliptical distrib., we may construct examples for which $\mathrm{Vol}(\mathcal{B}_\alpha) < \mathrm{Vol}(\mathcal{E}_\alpha)$, whatever k.

Outline of the talk

1. Background on conformal inference
   - Conformal inference
   - Split conformal inference in regression
2. A score based on the empirical covariance
   - Construction of adaptive scores
   - Non asymptotic ellipsoidal confidence region
3. Asymptotic study
   - Asymptotic ellipsoid $\mathcal{F}_\alpha$ (and $\mathcal{E}_\alpha$)
   - Comparison with the standard score for elliptical distributions
4. Numerical applications
   - Elliptical data : Gaussian data, Cauchy data
   - Non elliptical data : inverse Dirichlet data
5. Perspectives

Framework : multivariate regression

In this example, $X \in \mathbb{R}^6$, $Y \in \mathbb{R}^3$, $\rho \in (0,1)$, $\Sigma \in \mathcal{M}_{9,9}$, "Matérn 3/2" type :

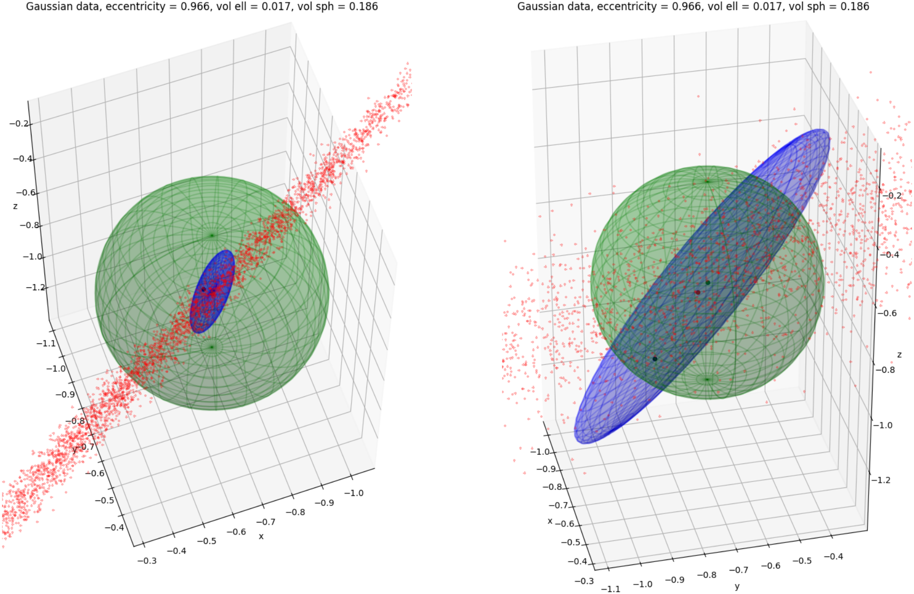
$$\Sigma_{ij} = (1 + |i-j| \ln \rho)\rho^{-|i-j|}. \tag{19}$$

Two cases :

- Gaussian data, $(X, Y) \sim \mathcal{N}(0, \Sigma)$
- Cauchy data, $(X, Y) \sim \mathcal{C}(0, \Sigma)$

With this choice of Σ, (X, Y) corresponds to the 9 first values of an AR(2) process ($\simeq$ solution of a noisy linear second order ODE).

Gaussian data 1/2



Gaussian data 2/2

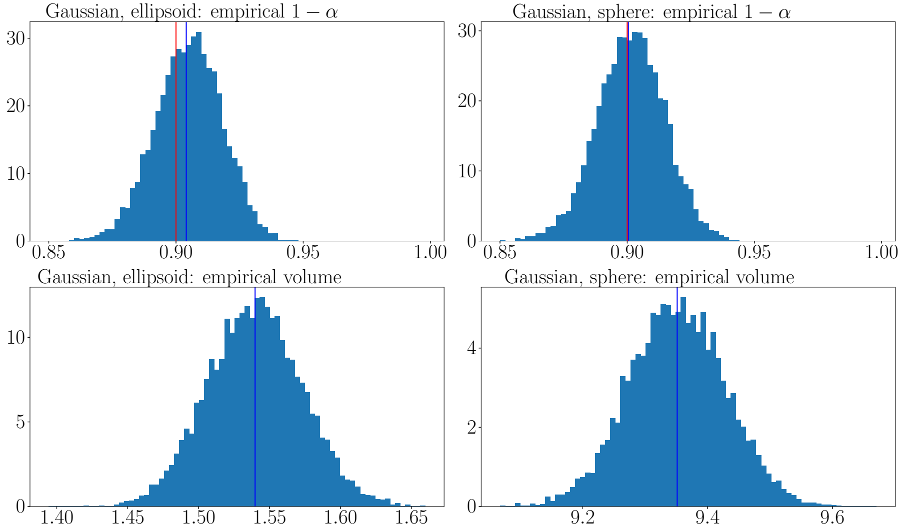
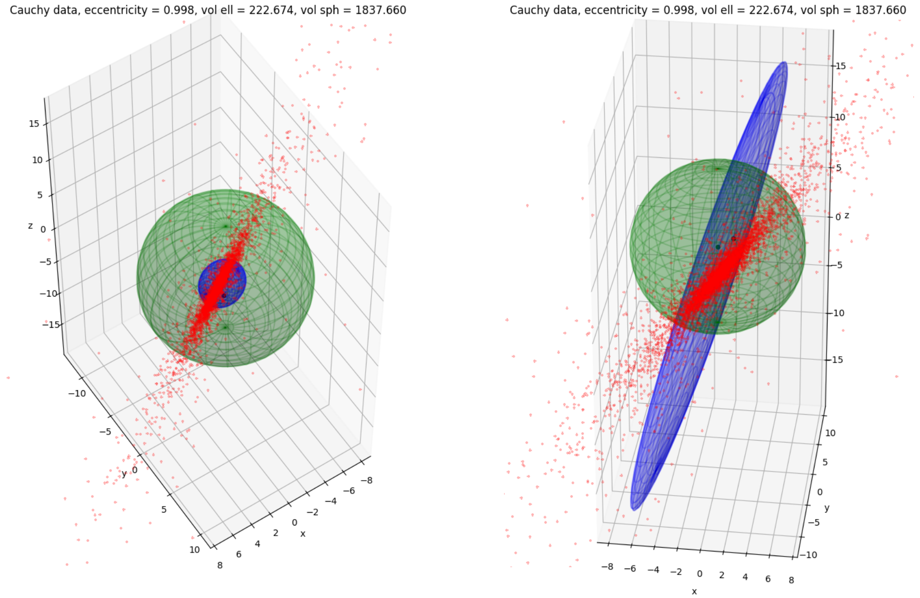


Figure 2: Empirical volumes and coverage. $n_{calib} = 200$, $n_{test} = 500$, $n_{histo} = 10000$. 75 bins were used.

Cauchy data (infinite var. and expectation) 1/2



Cauchy data (infinite var. and expectation) 2/2

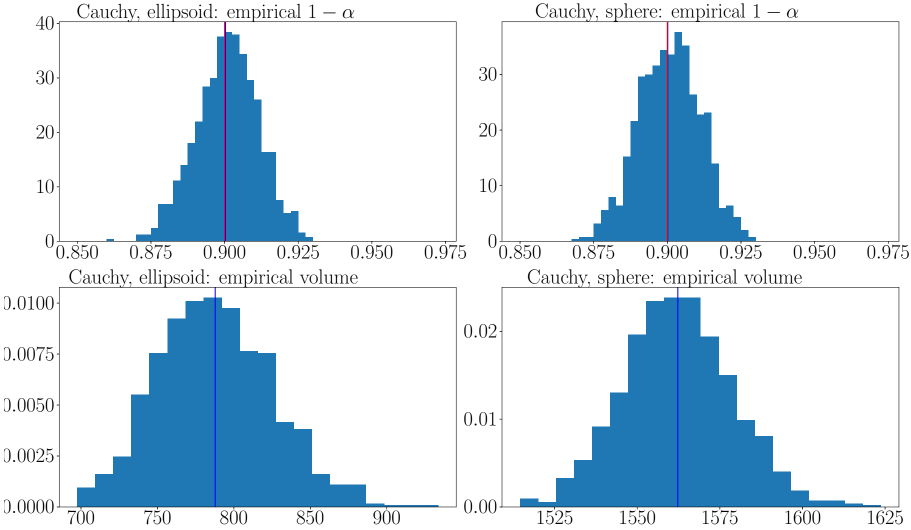


Figure 4: Empirical volumes and coverage $n = 10000$, $n_{test} = 800$ and $n_{histo} = 1000$. 50 bins were used.

Outline of the talk

1. Background on conformal inference
   - Conformal inference
   - Split conformal inference in regression
2. A score based on the empirical covariance
   - Construction of adaptive scores
   - Non asymptotic ellipsoidal confidence region
3. Asymptotic study
   - Asymptotic ellipsoid $\mathcal{F}_\alpha$ (and $\mathcal{E}_\alpha$)
   - Comparison with the standard score for elliptical distributions
4. Numerical applications
   - Elliptical data : Gaussian data, Cauchy data
   - Non elliptical data : inverse Dirichlet data
5. Perspectives

The inverse Dirichlet distribution

- Assume that the distribution of $(X_1^\top \ Y_1^\top)^\top$ is supported on the positive orthant $\mathbb{R}_+^p$, with a density of the form

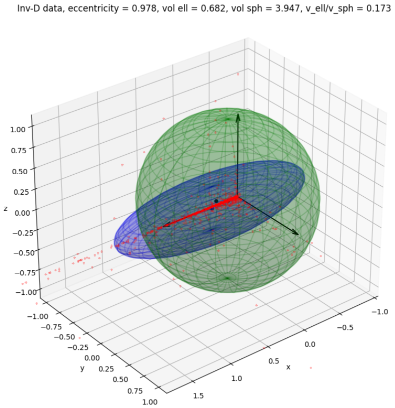
$$f(v) \propto \frac{1}{(1+|v|_1)^{b+|a|_1}} \prod_{i=1}^{p} v_i^{a_i-1}, \quad v \in \mathbb{R}_+^p.$$

  Above, $b > 0$, $(a_1, \ldots, a_p)$ is a vector of parameters with $a_i > 0$, and $|x|_1 = \sum_{i=1}^p |x_i|$.

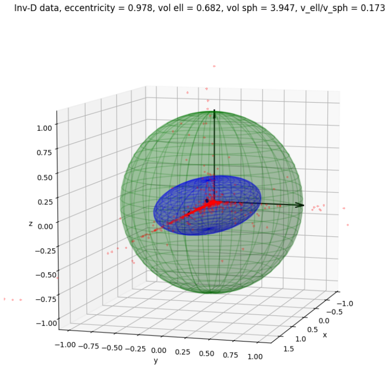
- V_1 is said to follow the inverted Dirichlet distribution which we denote by $V_1 \sim \text{IDirichlet}(a_1, \ldots, a_p; b)$.

- This distribution is mildly to highly non-elliptical.

A favourable case

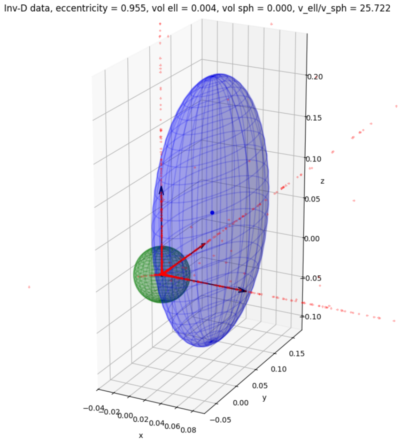


(a) Favourable inv. Dirichlet distribution (side)

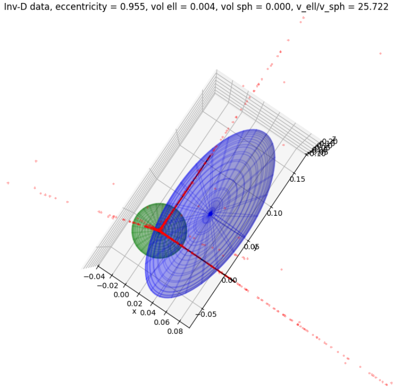


(b) Favourable inv. Dirichlet distribution (side, rotated)

A non favourable case



(a) Unfavourable inv. Dirichlet distribution (side)



(b) Unfavourable inv. Dirichlet distribution (top)

Perspectives

- Study our method when applied to time series + Python package (current work).
- Study for elliptical heavy tailed data : our asymptotic ellipsoid, is invariant w.r.t. dilatation of the pseudo-covariance matrix, cf Tyler dispersion matrix estimation.
- Functional framework $\rightarrow$ ellipsoids in function spaces (can be compact, contrarily to balls!).
- Conformal inference in the presence of covariate shift : training data according to $(X, Y) \sim P$, test data according to $(\widetilde{X}, \widetilde{Y}) \sim \widetilde{P}$ with the assumption that $P_{Y|X} = P_{\widetilde{Y}|\widetilde{X}}$ (e.g. $Y = f(X) + \varepsilon$, change P_X).

Thank you for your attention! :)