PLN: beyond prediction and theory

LSD

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Bastien Batardière

10 février 2025

Aperçu

- Normalité asymptotique et inférence variationnelle
- Optimisation et inférence variationnelle

Section 1

Contexte et introduction

Contexte

Données

$\mathbf{Y} \in \mathbb{R}^{n \times p}$: n individus $\mathbf{Y}_i$ indépendants, de dimension p.

Modèle à variables latentes

$$\mathbf{Z}_i \overset{\text{indep}}{\sim} p_\theta(\cdot)$$
$$Y_{ij} \,|\, Z_{ij} \overset{\text{indep}}{\sim} p_\theta\left(\cdot | Z_{ij}\right)$$

Vraisemblance

$$p_\theta(\mathbf{Y}_i) = \int_{\mathbb{R}^p} p_\theta(\mathbf{Y}_i, \mathbf{Z}_i) \mathrm{d}\mathbf{Z}_i$$

Modèle Poisson Log Normal

Modèle

- Etant donné des covariables :
  - $\mathbf{x}_i$ covariables de taille m pour $i = \{1, \ldots, n\}$ (longueur de la séquence d'ADN, température, etc)
- un paramètre $\theta = (\mathbf{B}, \Sigma)$:
  - $\mathbf{B} \in \mathbb{R}^{m \times p}$ un paramètre de régression
  - $\Sigma \in \mathcal{S}_p^{++}$ une matrice de covariance

$$\mathbf{Z}_i \overset{\text{indep}}{\sim} \mathcal{N}(\mathbf{x}_i^\top \mathbf{B}, \Sigma)$$

$$Y_{ij} \,|\, Z_{ij} \overset{\text{indep}}{\sim} \mathcal{P}\left(\exp\{Z_{ij}\}\right)$$

Enjeux

Maximisation de la log vraisemblance

$$\hat{\theta}^{\text{MLE}} = \operatorname{argmax}_\theta \log p_\theta(\mathbf{Y}) = \operatorname{argmax}_\theta \sum_{i=1}^{n} \log p_\theta(\mathbf{Y}_i)$$

- Comment approcher $\hat{\theta}^{\text{MLE}}$ efficacement ?
- Quelles garanties ?

Garanties de $\hat{\theta}^{\text{MLE}}$

$$\sqrt{n}\left(\hat{\theta}^{\text{MLE}} - \theta^\star\right) \xrightarrow{d} \mathcal{N}(0, \mathcal{J}(\theta^\star)^{-1})$$

Rappel sur l'inférence variationnelle

- Log vraisemblance intractable
  - Estimation problématique
  - Approche alternative: approximation
- Borne inférieure de la vraisemblance (ELBO) :

$$J_{\mathbf{Y}}(\theta, \phi) \triangleq \log p_\theta(\mathbf{Y}) - \mathrm{KL}\left[\tilde{p}_\phi(\cdot) \| p_\theta(\cdot \mid \mathbf{Y})\right]$$

avec $\tilde{p}_\phi$ une distribution variationnelle connue approchant $p_\theta(\cdot \mid \mathbf{Y})$.

- Estimateur variationnel

$$\hat{\theta}^{\mathsf{VEM}} = \mathrm{argmax}_{\theta,\phi}\, J_{\mathbf{Y}}(\theta, \phi)$$

Algorithme VEM

- EM Variationnel (VEM) :
  - Maximisation alternée de $J_{\mathbf{Y}}(\theta, \phi)$.

    Etape VE :

$$\phi^{(t)} = \operatorname{argmax}_{\phi} J_{\mathbf{Y}}(\theta^{(t)}, \phi)$$

    Etape de maximisation (M) :

$$\theta^{(t+1)} = \operatorname{argmax}_{\theta} \; J_{\mathbf{Y}}(\theta, \phi^{(t)})$$

Garanties de $\widehat{\theta}^{\text{VEM}}$

- ELBO:

$$J_{\mathbf{Y}}(\theta, \phi) = \log p_\theta(\mathbf{Y}) - \mathrm{KL}\left[\tilde{p}_\phi(\cdot) \| p_\theta(\cdot \mid \mathbf{Y})\right] \tag{1}$$

$$= \mathbb{E}_{\tilde{p}_\phi}\left[\log p_\theta(\mathbf{Y}, \mathbf{Z}) - \log \tilde{p}_\phi(\mathbf{Z})\right] \tag{2}$$

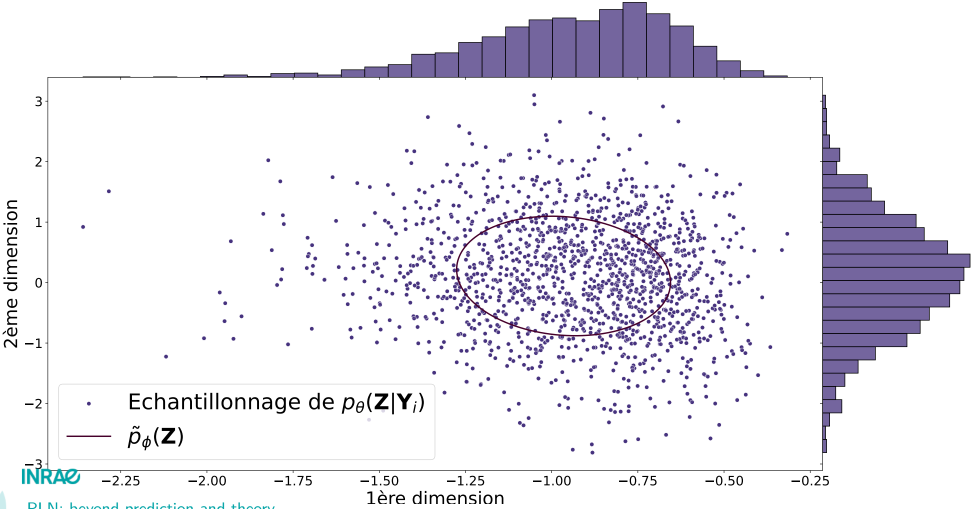
- Estimateur variationnel

$$\widehat{\theta}^{\text{VEM}}, \widehat{\phi} = \operatorname{argmax}_{\theta,\phi} J_{\mathbf{Y}}(\theta, \phi)$$

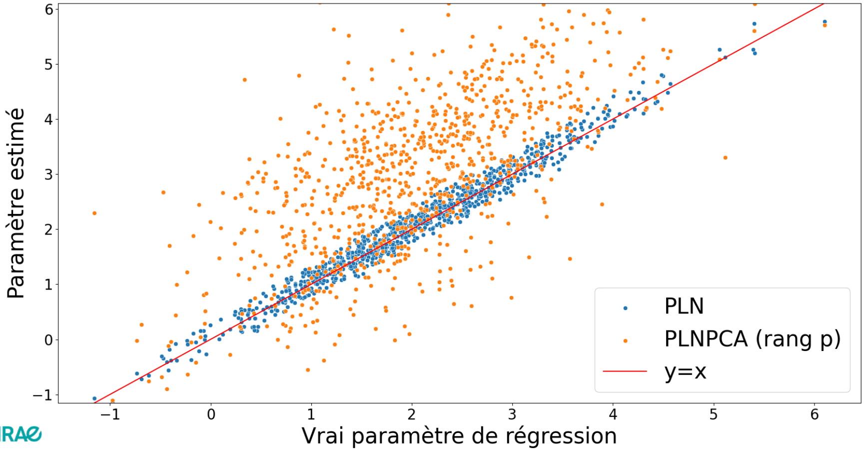
- Quelles garanties sur $\widehat{\theta}^{\text{VEM}}$?

$$\sqrt{n}\left(\widehat{\theta}^{\text{VEM}} - \theta^\star\right) \xrightarrow{d} ?$$

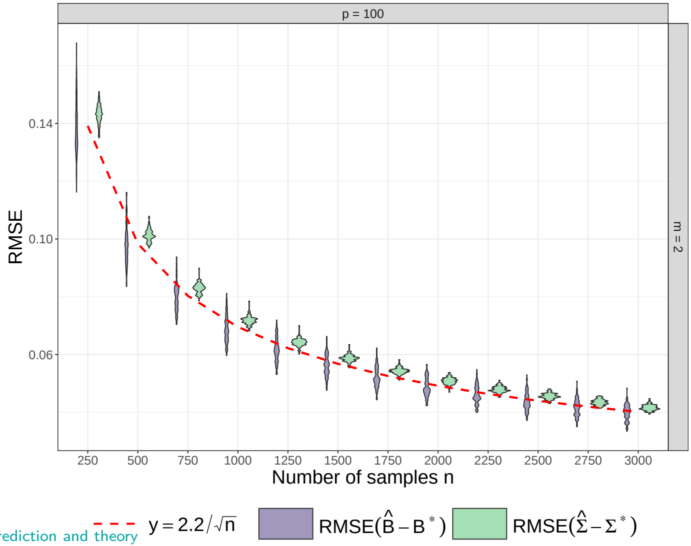
Un peu d'intuition



Estimateur biaisé ?



Estimation du bais



Section 2

Estimateur de la variance

Estimateur de la variance

Fonction objective profilé

$$L(\theta; \mathbf{Y}) \triangleq \sup_{\phi} J_{\mathbf{Y}}(\theta, \phi) = J_{\mathbf{Y}}(\theta, \hat{\phi})$$

Estimateur naif de la variance

- Pour l'estimateur du maximum de vraisemblance:

$$\sqrt{n}\left(\hat{\theta}^{\text{MLE}} - \theta^\star\right) \xrightarrow{d} mathcalN(0, \mathcal{J}(\theta^\star)^{-1})$$

- Si $\hat{\theta}^{\text{VEM}}$ est assez proche de $\theta^\star$, on peut approximer $\mathcal{J}(\theta^\star)$ par $\mathcal{J}(\hat{\theta}^{\text{VEM}})$.

$$\begin{aligned}\mathcal{J}(\hat{\theta}^{\text{VEM}}) &\approx -\frac{1}{n}\nabla_\theta^2 \log p_\theta(\mathbf{Y}) \\ &\approx -\frac{1}{n}\nabla_\theta^2 L(\theta)\end{aligned}$$

Estimateur sandwich

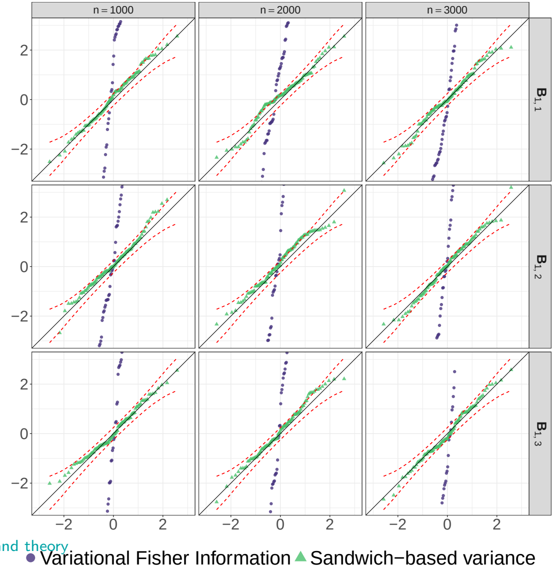
- Soit $\bar{\theta}^{\star\mathsf{VEM}} = \lim_{n\to\infty} L(\theta)$, alors sous certaines conditions sur L,

$$\sqrt{n}(\hat{\theta}^{\mathsf{VEM}} - \theta^{\star\mathsf{VEM}}) \xrightarrow{d} \mathcal{N}(0, V(\theta^{\star\mathsf{VEM}}))$$

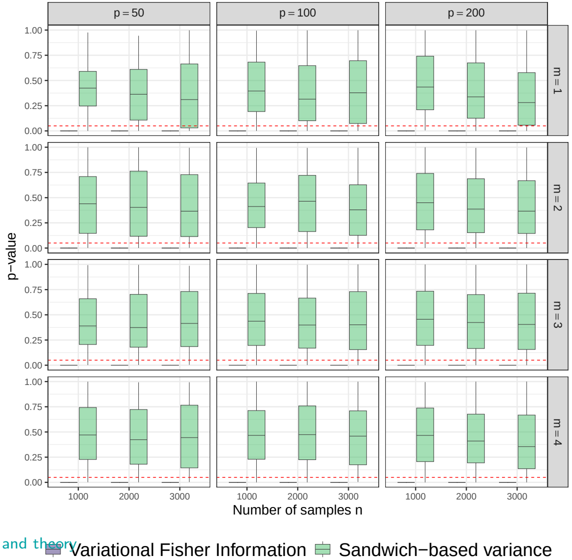
avec $V(\theta) = C(\theta)^{-1} D(\theta) C(\theta)^{-1}$ et

$$C(\theta) = \mathbb{E}[\nabla_{\theta\theta} L(\theta; Y)]$$
$$D(\theta) = \mathbb{E}\left[(\nabla_\theta L(\theta; Y))(\nabla_\theta L(\theta; Y))^\top\right]$$

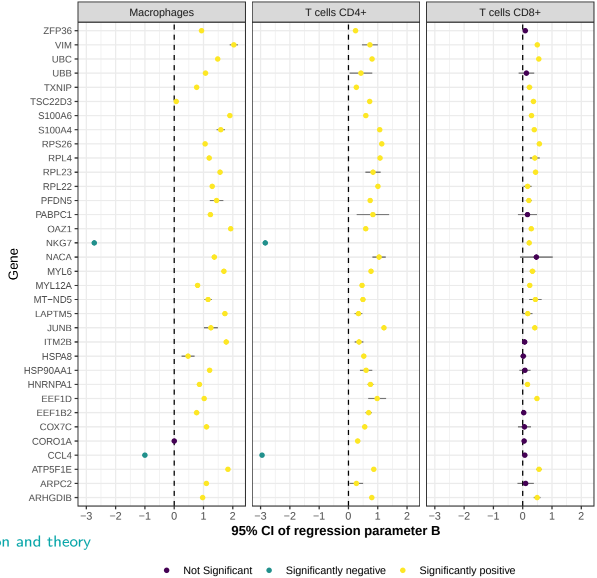
Simulations



Simulations



Application



Callouts

Note

A note

Astuce

A tip

Important

An important message

Slides

$$\operatorname{argmax} L(\theta) \mapsto \operatorname{argmax} \log p_\theta(\mathbf{Y})$$

$$\widehat{\theta}^{VEM} \mapsto \widehat{\theta}^{MLE} \mapsto \theta^\star$$