PLN: beyond prediction and theory

LSD

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Bastien Batardière

11 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 $i = \{1, \dots, n\}$ (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

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

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

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

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

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

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

Choix de la famille variationnelle

Famille gaussienne diagonale

$$\tilde{p}_{\phi_i} = \mathcal{N}\left(\mathbf{m}_i, \mathrm{Diag}(\mathbf{s}_i^2)\right), \quad \phi_i = (\mathbf{m}_i, \mathbf{s}_i), \quad \mathbf{m}_i \in \mathbb{R}^p, s_i \in \mathbb{R}^q, \tag{1}$$

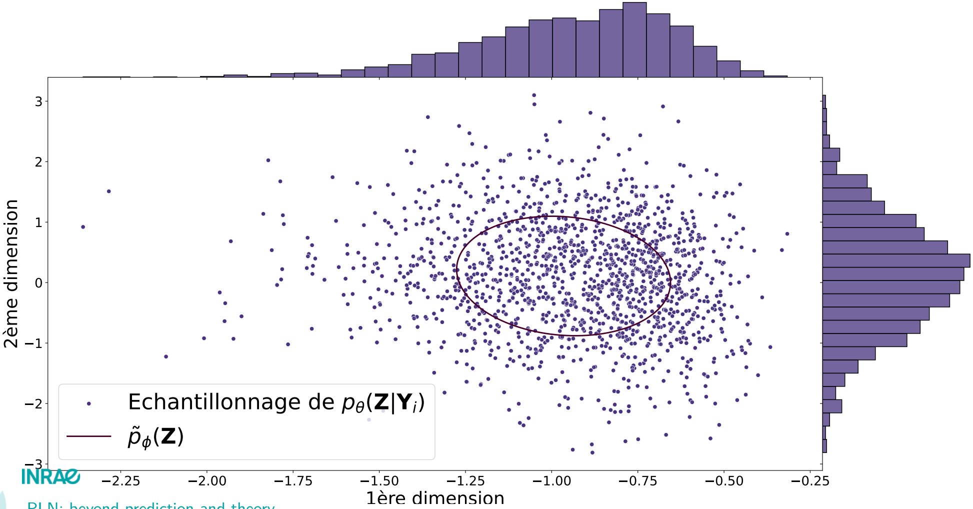
$$\phi = (\phi_1, \dots, \phi_n) = (\mathbf{m}_1, \dots, \mathbf{m}_n, \mathbf{s}_i, \dots, \mathbf{s}_n)$$

Un peu d'intuition

- ELBO:

$$
\begin{aligned}
J_{\mathbf{Y}}(\theta, \phi) &= \log p_\theta(\mathbf{Y}) - \mathrm{KL}\left[\tilde{p}_\phi(\cdot) \| p_\theta(\cdot \mid \mathbf{Y})\right] && (2) \\
&= \mathbb{E}_{\tilde{p}_\phi}\left[\log p_\theta(\mathbf{Y}, \mathbf{Z}) - \log \tilde{p}_\phi(\mathbf{Z})\right] && (3)
\end{aligned}
$$

▶ Un peu d'intuition



Section 2

Optimisation et inférence variationnelle

Optimisation et inférence variationnelle

Objectif

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

Leviers

- Changement de paramétrisation
- Méthodes d'optimisations

Algorithme Variational EM (VEM)

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

    Etape VE :

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

    Etape de maximisation (M) :

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

  - $J_{\mathbf{Y}}(\theta^{(t+1)}, \phi^{(t+1)}) > J_{\mathbf{Y}}(\theta^{(t)}, \phi^{(t)})$
- Maximisation alternée de $J_{\mathbf{Y}}(\theta, \phi)$.
- Un problème d'optimisation à chaque itération
- Très couteux

Différentes paramétrisations

Paramétrisations équivalentes du modèle PLN:

$$
\begin{aligned}
Y_{ij}|Z_{ij} &\sim \mathcal{P}(\exp(Z_{ij})), & \mathbf{Z}_i &\sim \mathcal{N}(\mathbf{B}^\top \mathbf{x}_i, \Sigma), & \text{(PLN-closed)}\\
Y_{ij}|Z_{ij} &\sim \mathcal{P}(\exp(\mathbf{B}_j^\top \mathbf{X}_i + Z_{ij})), & \mathbf{Z}_i &\sim \mathcal{N}(\mathbf{0}_p, \Sigma), & \text{(PLN-0)}\\
Y_{ij}|\mathbf{Z}_i &\sim \mathcal{P}(\exp(\mathbf{B}_j^\top \mathbf{x}_i + \mathbf{C}_j^\top \mathbf{Z}_i)), & \mathbf{Z}_i &\sim \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p). & \text{(PLN-PF)}
\end{aligned}
$$

Formes close

En fonction de la paramétrisation, l'étape de maximisation peut être plus ou moins complexe.

- PLN-closed: Forme close sur $\mathbf{B}$ et Σ
- PLN-0: Forme close pour $\mathbf{B}$
- PLN-PF: Aucune forme close

Approche brutale

- Montée de gradient directe:

$$\theta^{(t+1)}, \phi^{(t+1)} = (\theta^{(t)}, \phi^{(t)}) + \eta_t \nabla_{\theta,\phi}(J_{\mathbf{Y}}(\theta^{(t)}, \phi^{(t)}))$$

  $\eta_t > 0$: pas d'apprentissage.
- Pas de garantie sur le signe de $J_{\mathbf{Y}}(\theta^{(t+1)}, \phi^{(t+1)}) - J_{\mathbf{Y}}(\theta^{(t)}, \phi^{(t)})$
- Convergence plus rapide en pratique, avec le bon optimiseur (choix de η_t).
- Très bon résultat avec l'optimiseur Rprop.

Approche mixte profilé

- Si l'étape E est une forme close : $(\mathbf{B}^{(t+1)}, \Sigma^{(t+1)}) = H(\phi)$ avec

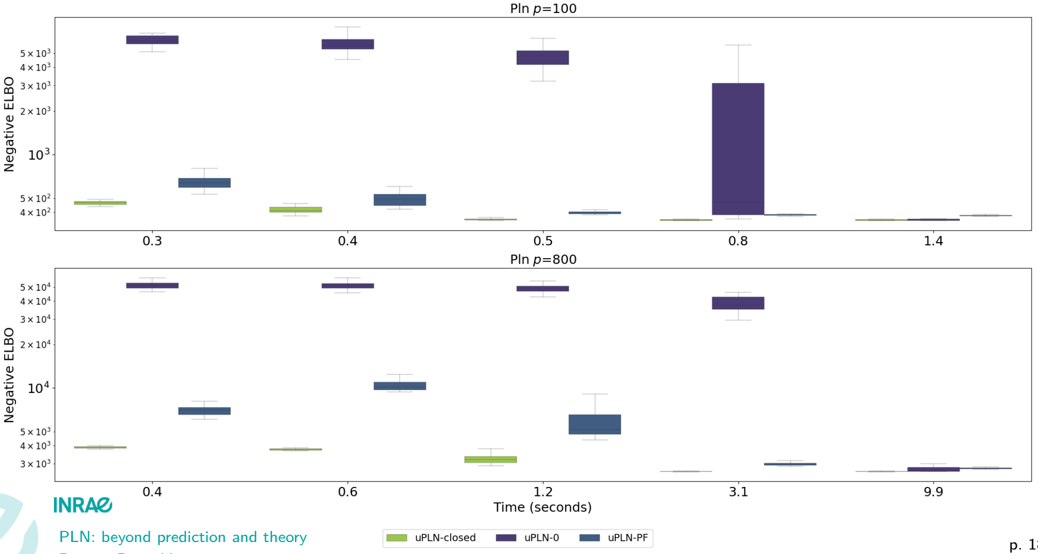
$$\nabla_\theta J(\theta, \phi) \mid_{\theta = H(\phi)} = 0.$$

- On peut définir l'ELBO profilé

$$\tilde{J}_{\mathbf{Y}}(\phi) = J_{\mathbf{Y}}(H(\phi), \phi)$$

- Réduction drastique du nombre de paramètres
- Gagne une inversion de matrice

Simulations



Section 3

Normalité asymptotique (Soumis à JCGS)

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

- ELBO:

$$J_{\mathbf{Y}}(\theta, \phi) = \log p_\theta(\mathbf{Y}) - \mathrm{KL}\left[\tilde{p}_\phi(\cdot) \| p_\theta(\cdot \mid \mathbf{Y})\right] \tag{4}$$

$$= \mathbb{E}_{\tilde{p}_\phi}\left[\log p_\theta(\mathbf{Y}, \mathbf{Z}) - \log \tilde{p}_\phi(\mathbf{Z})\right] \tag{5}$$

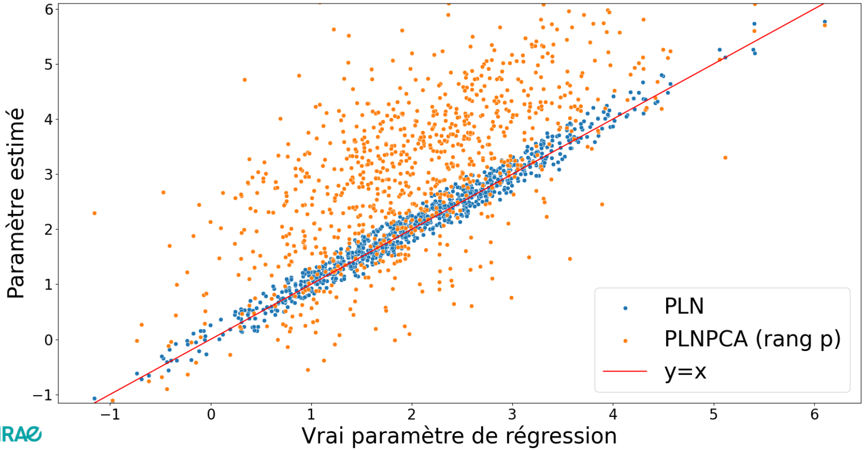
- Estimateur variationnel

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

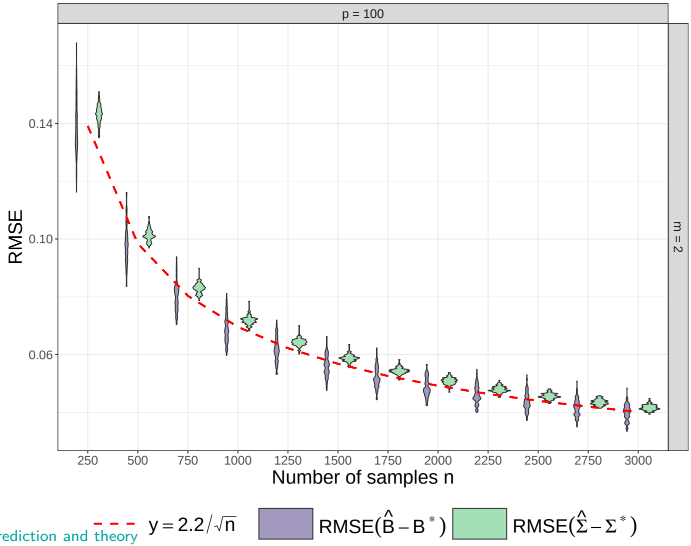
- Quelles garanties sur $\widehat{\theta}^{\mathsf{VEM}}$?

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

Estimateur biaisé ?



Estimation du bais



Section 4

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:

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

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

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

Estimateur sandwich

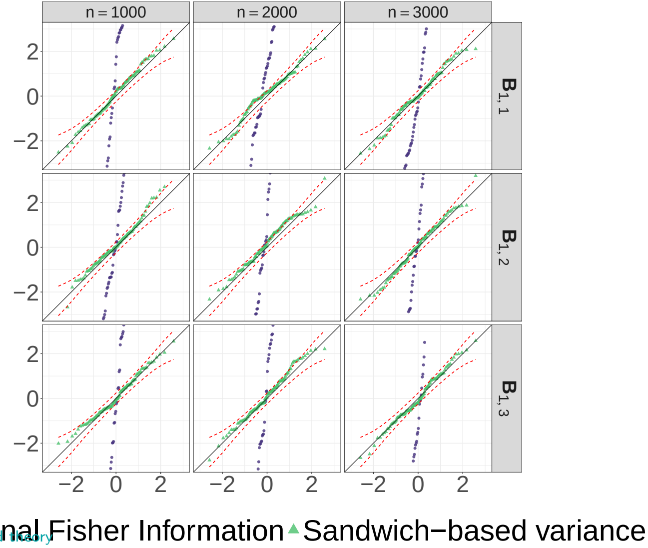
- Soit $\bar{\theta}^{\star\text{VEM}} = \lim_{n\to\infty} L(\theta)$, alors sous certaines conditions sur L,

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

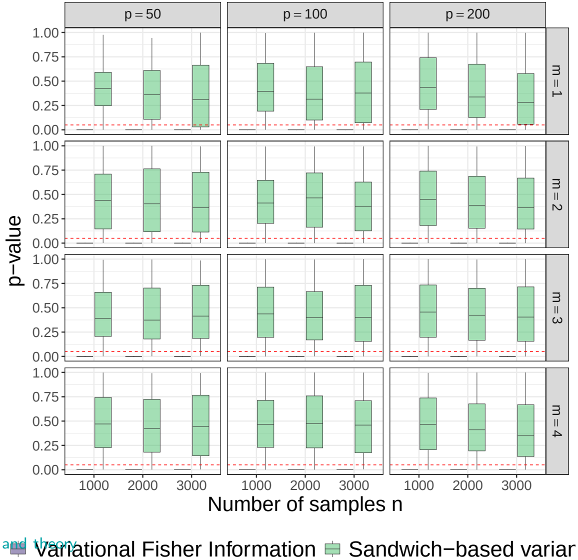
avec $V(\theta) = C(\theta)^{-1}D(\theta)C(\theta)^{-1}$ et

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

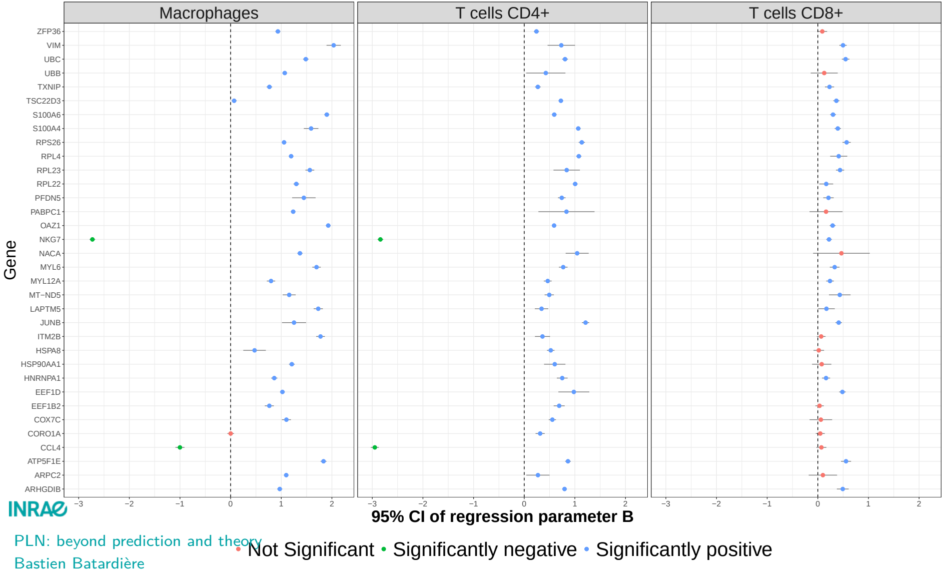
Simulations



Simulations



Application



Limites et perspectives

Limites

- Calcul de l'estimateur Sandwich laborieux
- Estimateur biaisé pour certains modèles

Perspectives

- Automatisation de l'estimation de la variance (autograd, `jax`, `functorch`) à d'autres modèles PLN.
- Calcul pour d'autres paramètres que celui de la régression.

Perspective

> **Note**
>
> A note

> **Astuce**
>
> A tip

> **Important**
>
> An important message

Slides

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

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