

Ismaël Castillo  
Professeur de statistique  
Responsable du projet ANR BACKUP  
Laboratoire de Probabilités, Statistique et Modélisation  
Sorbonne Université, Paris, France  
<https://perso.lpsm.paris/~castillo/>



## Sujet de thèse

### *Understanding Variational Bayes in high-dimensional models*

*Mots clés* : bayésien non-paramétrique, méthodes bayésiennes variationnelles, modèles de grande dimension et parcimonie, modèles semi-paramétriques.

*Profil recherché et pré-requis* : ayant des bases solides en probabilités et statistique, le/la candidat(e) aura obtenu d'excellents résultats en M1 et M2 et suivi des cours en statistique/apprentissage théorique tels que statistique non-paramétrique, bayésien non-paramétrique, statistique en grande dimension, apprentissage statistique, réseaux de neurones profonds.

## Détails

- *Encadrant principal* : Ismaël Castillo (PR, Sorbonne Université). Cette thèse s'inscrit dans le cadre du projet ANR BACKUP.
- *Discussions/collaborations prévues sur ce projet au sein de l'ANR BACKUP* : Kolyan Ray (Imperial College), Elisabeth Gassiat (Orsay), Sylvain Le Corff, Etienne Roquain, Fanny Villers (Sorbonne), Botond Szabo (Bocconi), Aad van der Vaart (TU Delft).
- *Début* : automne 2025; la thèse pourra être précédée d'un stage de Master 2 au printemps-été 2025.
- *Financement* : il est prévu que la thèse (3 ans) soit financée en totalité par le projet ANR BACKUP. L'allocation sera couplée à un monitorat effectué à Sorbonne Université.

## Description du projet

*Contexte et résultats antérieurs*. Les méthodes variationnelles [4] font partie des algorithmes les plus utilisées en statistique et apprentissage. Dans un cadre bayésien, l'idée consiste à approcher la loi a posteriori par un élément d'une famille de lois plus simples. Un des grands

exemples de telle famille variationnelle est la classe à *champ moyen* [4] constituée de lois dont les coordonnées dans une certaine paramétrisation sont indépendantes; un autre exemple est la méthode par *inducing points* [13, 11] pour les processus gaussiens.

Trouver l'élément de la famille variationnelle qui approche le mieux la loi a posteriori cible constitue un problème d'*optimisation*, pour lequel des algorithmes rapides, par exemple basés sur des descentes de gradient, sont souvent disponibles.

Parmi les exemples d'applications où cette méthode a été utilisée très récemment, on peut citer entre autres la régression linéaire en grande dimension [12, 7], les réseaux de neurones profonds [8, 3, 6], les processus gaussiens [11] et les modèles SBMs de graphes aléatoires [9].

D'un point de vue théorique, la plupart des résultats existants ont trait aux vitesses de convergence : sous certaines conditions, dans des modèles paramétriques ou non-paramétriques, on peut montrer que la loi variationnelle a posteriori converge à la même vitesse à constante près que la loi a posteriori cible [2, 16, 15, 1, 5]. Dans des modèles paramétriques (en dimension finie, donc), des résultats de type normalité asymptotique ont également été établis récemment [14, 10]. Cependant, la compréhension de cette méthode est très parcellaire lorsque l'on s'intéresse à des modèles de grande dimension tels que les modèles parcimonieux et les modèles semi-paramétriques (modèles statistique où l'on s'intéresse à un aspect fini-dimensionnel mais où les paramètres inconnus sont par exemple des fonctions), que ce soit du point de vue de la caractérisation de l'approximation variationnelle ou de la quantification de l'incertitude. Par exemple, d'un point de vue expérimental il est souvent observé que la loi a posteriori variationnelle 'sous-estime' l'incertitude, mais la compréhension de ce phénomène n'existe pour l'instant que dans des modèles de petite dimension.

*Objectifs de la thèse.* Le but du travail de thèse sera une meilleure compréhension des propriétés des lois a posteriori variationnelles dans des cadres pertinents pour les applications les plus récentes tels que les modèles de grande dimension, en obtenant des garanties mathématiques précises en terme de convergence, de lois limites et de quantification de l'incertitude.

Un volet principal d'investigation concernera les modèles de grande dimension sous contrainte de parcimonie, et ceci dans plusieurs directions. 1) De premiers résultats ont été obtenus en régression linéaire en grande dimension avec loi a priori spike-and-slab, du point de vue de la convergence globale du paramètre [12], puis pour l'inférence d'un nombre fini de paramètres [7]. Dans ce dernier cas, les résultats sont encore partiels, puisque du point de vue théorique, ils n'ont été obtenus que pour des matrices de design *peu corrélées*. Le cas plus *corrélé* est particulièrement intéressant dans les applications pratiques, et il devrait être possible d'obtenir des résultats rigoureux à partir de l'heuristique proposée dans [7]. 2) Par ailleurs, nous considérerons des aspects plus fins de la loi a posteriori variationnelle, comme par exemple la probabilité a posteriori d'inclusion d'un variable, avec en vue d'applications à la sélection de variables et aux tests multiples. 3) Le travail de thèse s'intéressera également à d'autres modèles de grande dimension, comme par exemple les *modèles graphiques* gaussiens, pour lesquels nous considérerons des résultats de convergence de la loi a posteriori variationnelle pour des classes de lois a priori parcimonieuses.

Le travail de thèse pourra également être complété par l'exploration de questions liées à l'estimation de fonctionnelles semi-paramétriques (par exemple dans des problème de causalité comme l'estimation de l'Average Treatment Effect - ATE). Des questions naturelles concerneront la quantification de l'incertitude par l'approximation variationnelle de la loi a posteriori, notamment la réduction de l'effet de 'diminution de la variance' évoquée plus haut.

Ces recherches pourront être combinées à l’investigation de nouvelles lois a priori adaptées à l’étude fonctionnelles spécifiques, dans la lignée du récent travail de thèse d’Alice l’Huillier sur l’estimation de fonctionnelles quadratiques.

## Bibliographie

- [1] P. Alquier. User-friendly introduction to PAC-Bayes bounds. *Foundations and Trends® in Machine Learning*, 17(2):174–303, 2024.
- [2] P. Alquier and J. Ridgway. Concentration of tempered posteriors and of their variational approximations. *Ann. Statist.*, 48(3):1475–1497, 2020.
- [3] J. Bai, Q. Song, and G. Cheng. Efficient variational inference for sparse deep learning with theoretical guarantee. In *NeurIPS conference*, 2020.
- [4] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, Apr. 2017.
- [5] I. Castillo. Bayesian nonparametrics: Saint–Flour lecture notes. *Lecture Notes in Mathematics, Springer*, 2024. To appear, arXiv eprint 2403.01737.
- [6] I. Castillo and P. Egels. Posterior and variational inference for deep neural networks with heavy-tailed weights. 2024. arXiv eprint 2406.03369.
- [7] I. Castillo, A. L’Huillier, K. Ray, and L. Travis. A variational Bayes approach to debiased inference for low-dimensional parameters in high-dimensional linear regression, 2024. arXiv eprint 2406.12659.
- [8] B.-E. Chérif-Abdellatif. Convergence rates of variational inference in sparse deep learning. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 1831–1842. PMLR, 13–18 Jul 2020.
- [9] S. Gaucher and O. Klopp. Optimality of variational inference for stochastic block model with missing links. In *NeurIPS*, pages 19947–19959, 2021.
- [10] M. A. Medina, J. L. M. Olea, C. Rush, and A. Velez. On the robustness to misspecification of  $\alpha$ -posteriors and their variational approximations. *Journal of Machine Learning Research*, 23(147):1–51, 2022.
- [11] D. Nieman, B. Szabo, and H. van Zanten. Uncertainty quantification for sparse spectral variational approximations in Gaussian process regression. *Electron. J. Stat.*, 17(2):2250–2288, 2023.
- [12] K. Ray and B. Szabó. Variational Bayes for high-dimensional linear regression with sparse priors. *J. Amer. Statist. Assoc.*, 117(539):1270–1281, 2022.
- [13] M. Titsias. Variational learning of inducing variables in sparse Gaussian processes. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 567–574, 2009.

- [14] Y. Wang and D. M. Blei. Frequentist consistency of variational Bayes. *J. Amer. Statist. Assoc.*, 114(527):1147–1161, 2019.
- [15] Y. Yang, D. Pati, and A. Bhattacharya.  $\alpha$ -variational inference with statistical guarantees. *Ann. Statist.*, 48(2):886–905, 2020.
- [16] F. Zhang and C. Gao. Convergence rates of variational posterior distributions. *Ann. Statist.*, 48(4):2180–2207, 2020.