Bayesian ML: project topics

Julyan Arbel, Rémi Bardenet, Gabriel Victorino Cardoso

1 Nature of the project

Students should form groups of two, each group undertaking one project. We suggest in Section 4 a few scientific papers that can each lead to a project, but you can choose another paper, subject to our approval.

For the paper your group will have chosen, you should: (1) explain the theoretical, computational and/or empirical methods developed in the paper, (2) emphasize the strengths and weaknesses of the paper, and (3) apply it to real data of your choice when applicable. We expect you to be creative and add something insightful that is not in the original paper: this can be a theoretical point, an alternative illustrative experiment, etc. The whole point is to read the paper with a critical mind. Finally, do not hesitate to pick longer or more difficult papers: we are aware of the diversity in difficulty, and we will take it into account.

2 Assignment of papers

As a first step, we ask each group to fill the spreadsheet at

TBC: https://lite.framacalc.org/cdlr2k2rpq-9sey

with the title of the paper, a link to it (if available), and the composition of the group. We ask that you fill in the form **before Wednesday 25 February**.

3 Format of the deliverable

You can use either Python or R for the programming part. Please have each group send

- one report as a pdf ($\leq$ 5 pages, all included) in the NeurIPS template,
- the link to a GitHub or GitLab repository containing your code, a Jupyter notebook to demo the code, and a readme file with instructions to (compile/install and) run the code.

to all teachers[1] **no later than Monday 9 March, 23:59**. There will be no deadline

[1] if the above link is broken, this means: julyan.arbel@inria.fr, remi.bardenet@gmail.com, and gabriel.victorino_cardoso@minesparis.psl.eu

extension.

The last step of your project will be a presentation in front of the class on Thursday 12 March.

4 Proposed papers

Lecture on Bayesics

[A1] Peter Grünwald. The safe Bayesian. In *Proceedings of the International Conference on Algorithmic Learning Theory*, pages 169–183. Springer, 2012.

[A2] Simon Lacoste-Julien, Ferenc Huszár, and Zoubin Ghahramani. Approximate inference for the loss-calibrated Bayesian. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 416–424. JMLR Workshop and Conference Proceedings, 2011.

[A3] Romain Lopez, Pierre Boyeau, Nir Yosef, Michael Jordan, and Jeffrey Regier. Decision-making with auto-encoding variational Bayes. *Advances in Neural Information Processing Systems*, 33:5081–5092, 2020.

Lecture on MCMC

[B1] Sourav Chatterjee and Persi Diaconis. The sample size required in importance sampling. *The Annals of Applied Probability*, 28(2):1099–1135, 2018.

[B2] Matthew D Hoffman, Andrew Gelman, et al. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

[B3] Jun Liu. The Collapsed Gibbs Sampler in Bayesian Computations With Applications to a Gene Regulation Problem . *Journal of the American Statistical Association*, 1994.

[B4] Stephan Mandt, Matthew Hoffman, and David Blei. A variational analysis of stochastic gradient algorithms. In *International conference on machine learning*, pages 354–363. PMLR, 2016.

Lecture on variational inference

[C1] Pierre Alquier, James Ridgway, and Nicolas Chopin. On the properties of variational approximations of Gibbs posteriors. *Journal of Machine Learning Research*, 17(236):1–41, 2016.

[C2] George Casella and Christian P Robert. Rao-Blackwellisation of sampling schemes. *Biometrika*, 83(1):81–94, 1996.

[C3] Gintare Karolina Dziugaite and Daniel Roy. Entropy-SGD optimizes the prior of a PAC-Bayes bound: Generalization properties of entropy-sgd and data-dependent priors. In *International Conference on Machine Learning*, pages 1377–1386. PMLR, 2018.

[C4] Thomas P Minka. Expectation propagation for approximate Bayesian inference. In *Uncertainty in Artificial Intelligence*, 2001.

[C5] Kolyan Ray and Botond Szabó. Variational Bayes for high-dimensional linear regression with sparse priors. *Journal of the American Statistical Association*, pages 1–12, 2021.

[C6] Y. W. Teh, D. Newman, and M. Welling. A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In *Advances in neural information processing systems*, pages 1353–1360, 2007.

Lecture on foundations

[D1] Rianne Heide, Alisa Kirichenko, Peter Grunwald, and Nishant Mehta. Safe-Bayesian generalized linear regression. In *International Conference on Artificial Intelligence and Statistics*, pages 2623–2633. PMLR, 2020.

[D2] Kenichiro McAlinn and Kōsaku Takanashi. When is generalized Bayes Bayesian? a decision-theoretic characterization of loss-based updating. *arXiv preprint arXiv:2602.01573*, 2026.

Lecture on Bayesian nonparametrics

[E1] Fadhel Ayed, Juho Lee, and Francois Caron. Beyond the chinese restaurant and Pitman-Yor processes: Statistical models with double power-law behavior. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 395–404. PMLR, 09–15 Jun 2019.

[E2] Michael Lavine. Some aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, pages 1222–1235, 1992.

[E3] Jeffrey W Miller and Matthew T Harrison. Inconsistency of Pitman-Yor process mixtures for the number of components. *The Journal of Machine Learning Research*, 15(1):3333–3370, 2014.

[E4] Jim Pitman and Marc Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 25(2):855–900, 1997.

Lecture on Bayesian deep learning

[F1] Paul Egels and Ismaël Castillo. Posterior and variational inference for deep neural networks with heavy-tailed weights. *Journal of Machine Learning Research*, 26(122):1–58, 2025.

[F2] Wenbo Guo, Sui Huang, Yunzhe Tao, Xinyu Xing, and Lin Lin. Explaining deep learning models-a Bayesian non-parametric approach. *NeurIPS*, 2018.

[F3] Mohammad Emtiyaz E Khan, Alexander Immer, Ehsan Abedi, and Maciej Korzepa. Approximate Inference Turns Deep Networks into Gaussian Processes. In *Advances in Neural Information Processing Systems*, pages 3088–3098, 2019.

[F4] A. Matthews, M. Rowland, J. Hron, R. Turner, and Z. Ghahramani. Gaussian process behaviour in wide deep neural networks. In *International Conference on Learning Representations*, volume 1804.11271, 2018.

[F5] Mikhail Yurochkin, Mayank Agarwal, Soumya Ghosh, Kristjan Greenewald, Nghia Hoang, and Yasaman Khazaeni. Bayesian nonparametric federated learning of neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 7252–7261. PMLR, 09–15 Jun 2019.

Lecture on Generative priors

[G1] Lisa Bedin, Gabriel Cardoso, Josselin Duchateau, Remi Dubois, and Eric Moulines. Leveraging an ECG beat diffusion model for morphological reconstruction from indirect signals. *Advances in Neural Information Processing Systems*, 37:84409–84446, December 2024.

[G2] Florentin Guth, Zahra Kadkhodaie, and Eero P. Simoncelli. Learning normalized image densities via dual score matching. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, October 2025.

[G3] Yazid Janati, Badr Moufad, Mehdi Abou El Qassime, Alain Oliviero Durmus, Eric Moulines, and Jimmy Olsson. A Mixture-Based Framework for Guiding Diffusion Models. In *Proceedings of the 42nd International Conference on Machine Learning*, pages 26830–26876. PMLR, October 2025.

[G4] Maxence Noble, Louis Grenioux, Marylou Gabrié, and Alain Oliviero Durmus. Learned Reference-based Diffusion Sampler for multi-modal distributions. In *The Thirteenth International Conference on Learning Representations*, October 2024.

[G5] Luhuan Wu, Brian Trippe, Christian Naesseth, David Blei, and John P. Cunningham. Practical and Asymptotically Exact Conditional Sampling in Diffusion Models. *Advances in Neural Information Processing Systems*, 36:31372–31403, December 2023.

Lecture on SBI

[H1] Maximilian Dax, Stephen R. Green, Jonathan Gair, Michael Pürrer, Jonas Wildberger, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Neural Importance Sampling for Rapid and Reliable Gravitational-Wave Inference. *Physical Review Letters*, 130(17):171403, April 2023.

[H2] Manuel Gloeckler, Michael Deistler, Christian Dietrich Weilbach, Frank Wood, and Jakob H. Macke. All-in-one simulation-based inference. In *Proceedings of the 41st International Conference on Machine Learning*, pages 15735–15766. PMLR, July 2024.

[H3] Pablo Lemos, Adam Coogan, Yashar Hezaveh, and Laurence Perreault-Levasseur. Sampling-Based Accuracy Testing of Posterior Estimators for General Inference. In *Proceedings of the 40th International Conference on Machine Learning*, pages 19256–19273. PMLR, July 2023.

[H4] Jonas Wildberger, Maximilian Dax, Simon Buchholz, Stephen Green, Jakob H. Macke, and Bernhard Schölkopf. Flow Matching for Scalable Simulation-Based Inference. *Advances in Neural Information Processing Systems*, 36:16837–16864, December 2023.

[H5] Yang Yang, Severi Rissanen, Paul E. Chang, Nasrulloh Loka, Daolang Huang, Arno Solin, Markus Heinonen, and Luigi Acerbi. PriorGuide: Test-Time Prior Adaptation for Simulation-Based Inference, October 2025.