Conjugate generalised Bayesian inference for discrete doubly intractable problems




Prof. François-Xavier Briol
Department of Statistical Science
University College London

Doubly-intractable problems

- We are interested in performing Bayesian inference with data defined on discrete space $\mathcal{X}$:

$$\pi(\theta \mid \{x_i\}_{i=1}^n) = \frac{1}{Z} \prod_{i=1}^{n} p_\theta(x_i) \pi(\theta)$$

Doubly-intractable problems

- We are interested in performing Bayesian inference with data defined on discrete space $\mathcal{X}$:

$$\pi(\theta \mid \{x_i\}_{i=1}^n) = \frac{1}{Z} \prod_{i=1}^{n} p_\theta(x_i)\pi(\theta)$$

Marginal likelihood; typically intractable!

Doubly-intractable problems

- We are interested in performing Bayesian inference with data defined on discrete space $\mathcal{X}$:

$$\pi(\theta \mid \{x_i\}_{i=1}^n) = \frac{1}{Z} \prod_{i=1}^n p_\theta(x_i)\pi(\theta)$$

Assume additionally:

$$p_\theta(x) = \frac{\tilde{p}_\theta(x)}{Z(\theta)}, \quad Z(\theta) := \sum_{x \in \mathcal{X}} \tilde{p}_\theta(x)$$

Doubly-intractable problems

- We are interested in performing Bayesian inference with data defined on discrete space $\mathcal{X}$:

$$\pi(\theta \mid \{x_i\}_{i=1}^n) = \frac{1}{Z} \prod_{i=1}^n p_\theta(x_i)\pi(\theta)$$

Assume additionally:

$$p_\theta(x) = \frac{\tilde{p}_\theta(x)}{Z(\theta)}, \quad Z(\theta) := \sum_{x \in \mathcal{X}} \tilde{p}_\theta(x)$$

Doubly-intractable!

Doubly-intractable problems

- We are interested in performing Bayesian inference with data defined on discrete space $\mathcal{X}$:

$$\pi(\theta \mid \{x_i\}_{i=1}^n) = \frac{1}{Z} \prod_{i=1}^n p_\theta(x_i)\pi(\theta)$$

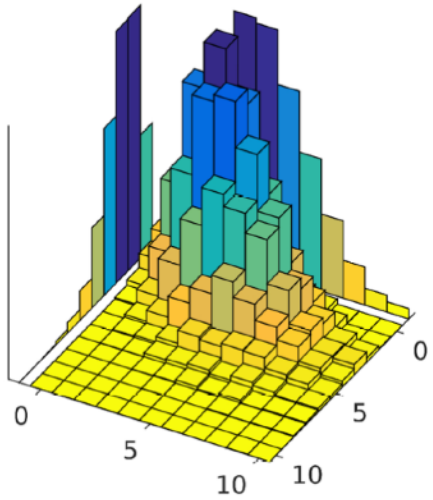
Assume additionally:

$$p_\theta(x) = \frac{\tilde{p}_\theta(x)}{Z(\theta)}, \quad Z(\theta) := \sum_{x \in \mathcal{X}} \tilde{p}_\theta(x)$$

- Sadly since the intractable $Z(\theta)$ depends on the parameter, we cannot use standard MCMC/VI.

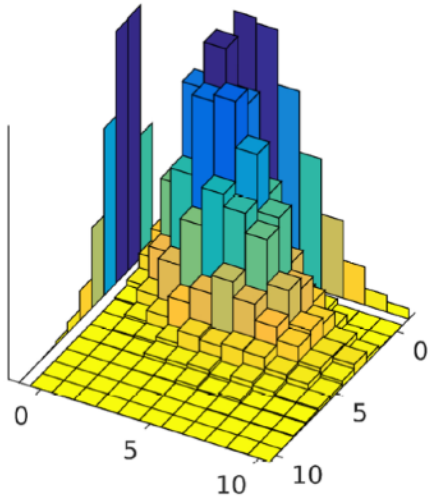
Discrete doubly-intractable problems

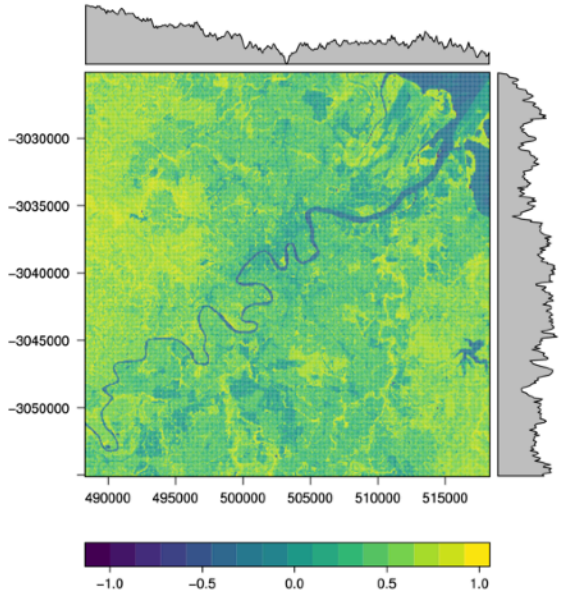
Discrete doubly-intractable problems



Inouye, D. I., Yang, E., Allen, G. I., & Ravikumar, P. (2017). A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdisciplinary Reviews: Computational Statistics*, *9*(3).

Discrete doubly-intractable problems

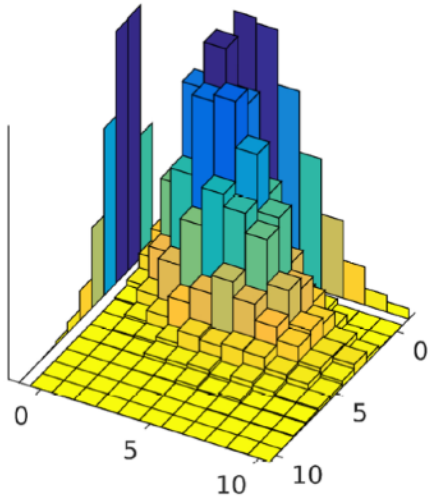


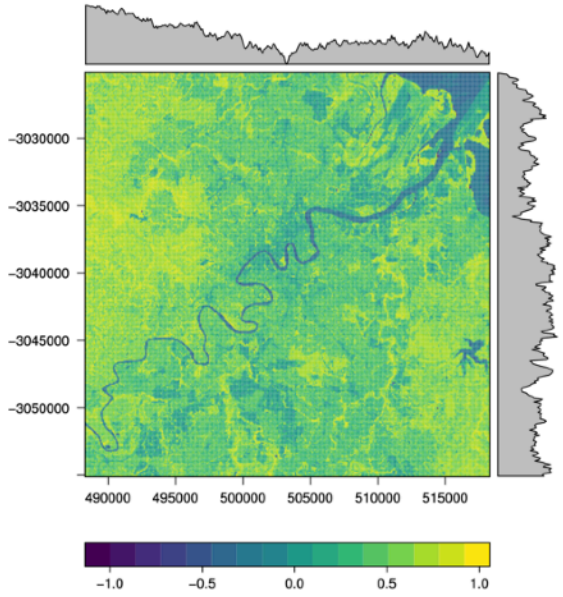


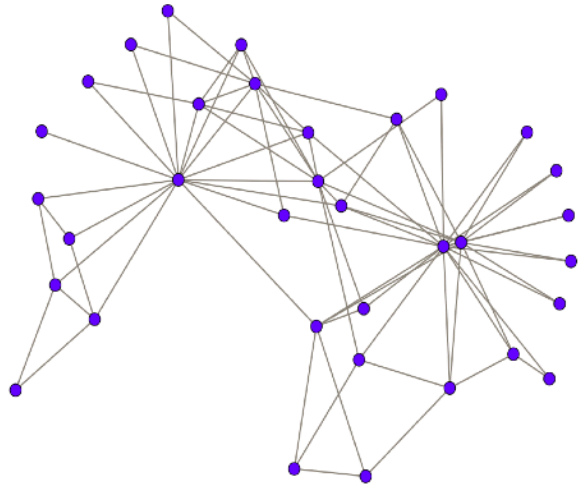
Inouye, D. I., Yang, E., Allen, G. I., & Ravikumar, P. (2017). A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdisciplinary Reviews: Computational Statistics*, *9*(3).

Moores, M. T., Nicholls, G. K., Pettitt, A. N., & Mengersen, K. (2020). Scalable Bayesian inference for the inverse temperature of a hidden Potts model. *Bayesian Analysis*, *15*(1), 1–27.

Discrete doubly-intractable problems







Inouye, D. I., Yang, E., Allen, G. I., & Ravikumar, P. (2017). A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdisciplinary Reviews: Computational Statistics*, *9*(3).

Moores, M. T., Nicholls, G. K., Pettitt, A. N., & Mengersen, K. (2020). Scalable Bayesian inference for the inverse temperature of a hidden Potts model. *Bayesian Analysis*, *15*(1), 1–27.

Bouranis, L., Friel, N., & Maire, F. (2018). Bayesian model selection for exponential random graph models via adjusted pseudolikelihoods. *Journal of Computational and Graphical Statistics*, *27*(3), 516–528.

A solution in continuous space

- One potential solution is to perform generalised Bayesian inference:

$$\pi^{\beta}_{\mathscr{L}}\left(\theta;\ \{x_i\}_{i=1}^{n}\right) \propto \exp\left(-\beta n\, \hat{\mathscr{L}}_n(\theta)\right) \pi(\theta)$$

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2022). Robust generalised Bayesian inference for intractable likelihoods. *Journal of the Royal Statistical Society: Series B: (Statistical Methodology)*, *84*(3), 997–1022.

A solution in continuous space

- One potential solution is to perform generalised Bayesian inference:

$$\pi^{\beta}_{\mathscr{L}}\left(\theta;\ \{x_i\}_{i=1}^{n}\right) \propto \exp\left(-\beta n\, \hat{\mathscr{L}}_n(\theta)\right) \pi(\theta)$$

Prior

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2022). Robust generalised Bayesian inference for intractable likelihoods. *Journal of the Royal Statistical Society: Series B: (Statistical Methodology)*, *84*(3), 997–1022.

A solution in continuous space

- One potential solution is to perform generalised Bayesian inference:

$$\pi^{\beta}_{\mathscr{L}}\left(\theta;\ \{x_i\}_{i=1}^{n}\right) \propto \exp\left(-\beta n\, \hat{\mathscr{L}}_n(\theta)\right) \pi(\theta)$$

Loss function Prior

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2022). Robust generalised Bayesian inference for intractable likelihoods. *Journal of the Royal Statistical Society: Series B: (Statistical Methodology)*, *84*(3), 997–1022.

A solution in continuous space

- One potential solution is to perform generalised Bayesian inference:

$$\pi^{\beta}_{\mathscr{L}}\left(\theta;\ \{x_i\}_{i=1}^{n}\right) \propto \exp\left(-\beta n\, \hat{\mathscr{L}}_n(\theta)\right) \pi(\theta)$$

'Generalised posterior' Loss function Prior

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2022). Robust generalised Bayesian inference for intractable likelihoods. *Journal of the Royal Statistical Society: Series B: (Statistical Methodology)*, *84*(3), 997–1022.

A solution in continuous space

- One potential solution is to perform generalised Bayesian inference:

$$\pi^{\beta}_{\mathscr{L}}\left(\theta;\ \{x_i\}_{i=1}^{n}\right) \propto \exp\left(-\beta n\, \hat{\mathscr{L}}_n(\theta)\right) \pi(\theta)$$

- Typically, $\mathscr{L}(\theta) = D(q_0 \| p_\theta)$ and $\hat{\mathscr{L}}_n(\theta)$ is a consistent estimator based on data $\{x_i\}_{i=1}^{n} \sim q_0$.

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2022). Robust generalised Bayesian inference for intractable likelihoods. *Journal of the Royal Statistical Society: Series B: (Statistical Methodology)*, *84*(3), 997–1022.

A solution in continuous space

- One potential solution is to perform generalised Bayesian inference:

$$\pi^{\beta}_{\mathscr{L}}\left(\theta;\ \{x_i\}_{i=1}^{n}\right) \propto \exp\left(-\beta n\, \hat{\mathscr{L}}_n(\theta)\right) \pi(\theta)$$

- Typically, $\mathscr{L}(\theta) = D(q_0 \| p_\theta)$ and $\hat{\mathscr{L}}_n(\theta)$ is a consistent estimator based on data $\{x_i\}_{i=1}^{n} \sim q_0$.
- One idea is to pick the divergence D so that it does note depend on $Z(\theta)$.

→ This is possible with the kernel Stein discrepancy (KSD)!

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2022). Robust generalised Bayesian inference for intractable likelihoods. *Journal of the Royal Statistical Society: Series B: (Statistical Methodology)*, *84*(3), 997–1022.

A solution in continuous space

- One potential solution is to perform generalised Bayesian inference:

$$\pi^{\beta}_{\mathscr{L}}\left(\theta;\ \{x_i\}_{i=1}^n\right) \propto \exp\left(-\beta n\, \hat{\mathscr{L}}_n(\theta)\right) \pi(\theta)$$

- Typically, $\mathscr{L}(\theta) = D(q_0 \| p_\theta)$ and $\hat{\mathscr{L}}_n(\theta)$ is a consistent estimator based on data $\{x_i\}_{i=1}^n \sim q_0$.
- One idea is to pick the divergence D so that it does note depend on $Z(\theta)$.

→ This is possible with the kernel Stein discrepancy (KSD)!

- When the likelihood is exponential family and the prior Gaussian, the approach is **fully conjugate**!

$$p_\theta(x) := \exp\left(\eta(\theta)^\top T(x) + B(x) - \log Z(\theta)\right)$$

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2022). Robust generalised Bayesian inference for intractable likelihoods. *Journal of the Royal Statistical Society: Series B: (Statistical Methodology)*, *84*(3), 997–1022.

Some more robust and conjugate Bayes

Altamirano, M., Briol, F.-X., & Knoblauch, J. (2023). Robust and scalable Bayesian online changepoint detection. *ICML*, 642–663.

Altamirano, M., Briol, F.-X., & Knoblauch, J. (2024). Robust and conjugate Gaussian process regression. *ICML,* 1155–1185.

Duran-Martin, G., Altamirano, M., Shestopaloff, A. Y., Knoblauch, J., Jones, M., Briol, F.-X., & Murphy, K. (2024). Outlier-robust Kalman filtering through generalised Bayes. *ICML*, 12138–12171.

Laplante, W., Altamirano, M., Duncan, A., Knoblauch, J., & Briol, F.-X. (2025). Robust and conjugate spatio-temporal Gaussian processes. *ICML*, 32562–32592.

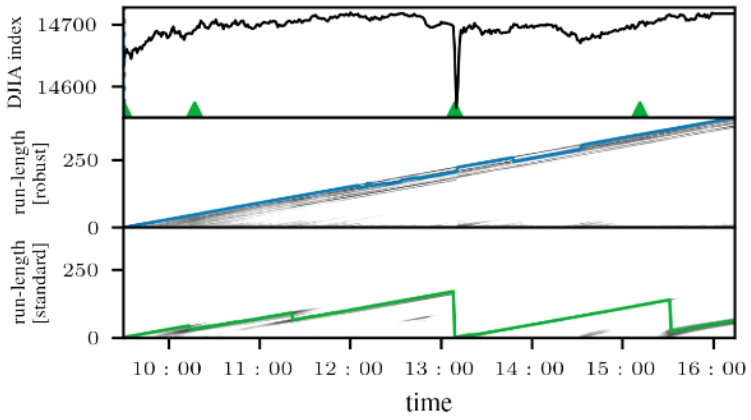
Rooijakkers, J., Ronneberg, L., Briol, F.-X., Knoblauch, J. & Altamirano, M. (2025+). Multi-output robust and conjugate Gaussian processes. *Under review*.

Sun, Z., Barp, A., & Briol, F-X. (2025+). Generalised Bayes for spherical data. In preparation.

Bharti, A., Dellaporta, C., Hikida, Y., & Briol, F-X (2025+). Amortised and provably-robust neural simulation-based inference. In preparation.

Some more robust and conjugate Bayes

Change-point detection



Altamirano, M., Briol, F.-X., & Knoblauch, J. (2023). Robust and scalable Bayesian online changepoint detection. *ICML*, 642–663.

Altamirano, M., Briol, F.-X., & Knoblauch, J. (2024). Robust and conjugate Gaussian process regression. *ICML,* 1155–1185.

Duran-Martin, G., Altamirano, M., Shestopaloff, A. Y., Knoblauch, J., Jones, M., Briol, F.-X., & Murphy, K. (2024). Outlier-robust Kalman filtering through generalised Bayes. *ICML*, 12138–12171.

Laplante, W., Altamirano, M., Duncan, A., Knoblauch, J., & Briol, F.-X. (2025). Robust and conjugate spatio-temporal Gaussian processes. *ICML*, 32562–32592.

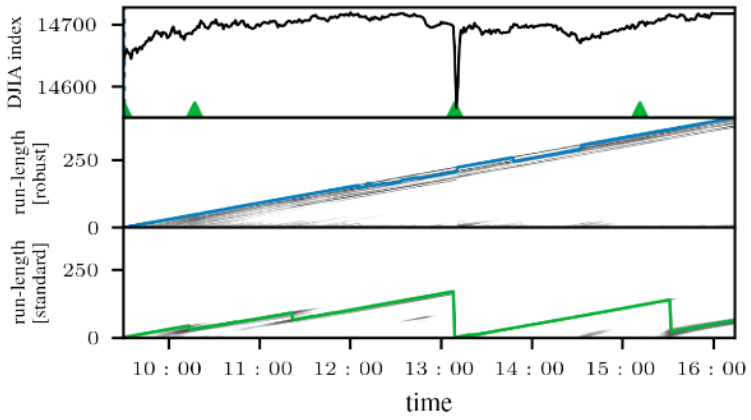
Rooijakkers, J., Ronneberg, L., Briol, F.-X., Knoblauch, J. & Altamirano, M. (2025+). Multi-output robust and conjugate Gaussian processes. *Under review*.

Sun, Z., Barp, A., & Briol, F-X. (2025+). Generalised Bayes for spherical data. In preparation.

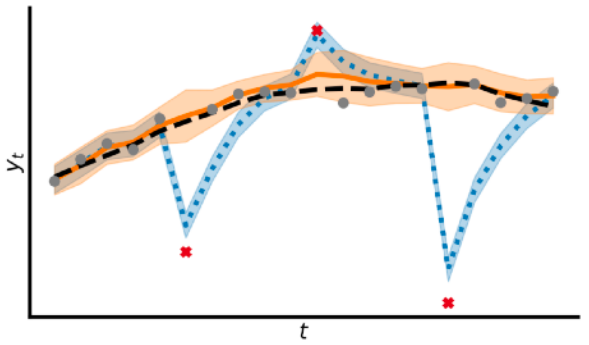
Bharti, A., Dellaporta, C., Hikida, Y., & Briol, F-X (2025+). Amortised and provably-robust neural simulation-based inference. In preparation.

Some more robust and conjugate Bayes

Change-point detection



Kalman filtering



Altamirano, M., Briol, F.-X., & Knoblauch, J. (2023). Robust and scalable Bayesian online changepoint detection. *ICML*, 642–663.

Altamirano, M., Briol, F.-X., & Knoblauch, J. (2024). Robust and conjugate Gaussian process regression. *ICML,* 1155–1185.

Duran-Martin, G., Altamirano, M., Shestopaloff, A. Y., Knoblauch, J., Jones, M., Briol, F.-X., & Murphy, K. (2024). Outlier-robust Kalman filtering through generalised Bayes. *ICML*, 12138–12171.

Laplante, W., Altamirano, M., Duncan, A., Knoblauch, J., & Briol, F.-X. (2025). Robust and conjugate spatio-temporal Gaussian processes. *ICML*, 32562–32592.

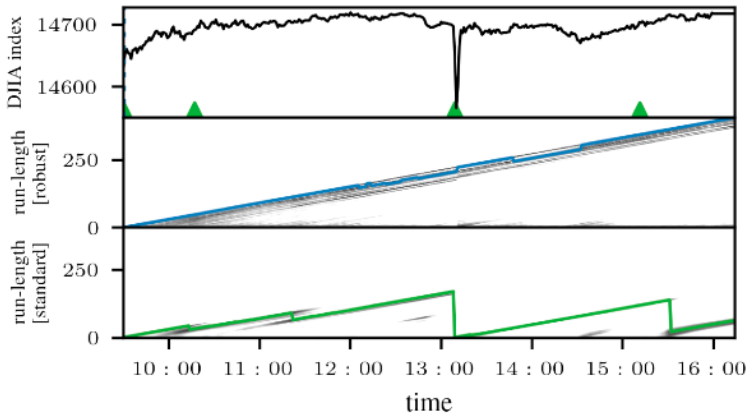
Rooijakkers, J., Ronneberg, L., Briol, F.-X., Knoblauch, J. & Altamirano, M. (2025+). Multi-output robust and conjugate Gaussian processes. *Under review*.

Sun, Z., Barp, A., & Briol, F-X. (2025+). Generalised Bayes for spherical data. In preparation.

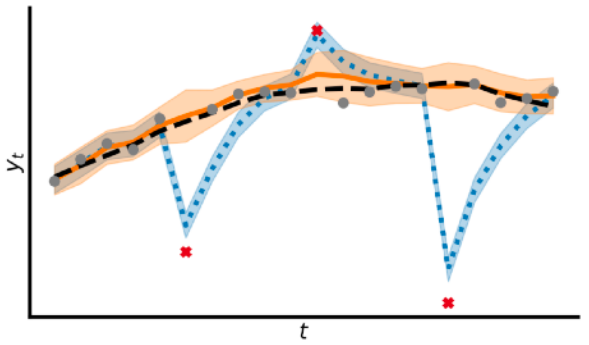
Bharti, A., Dellaporta, C., Hikida, Y., & Briol, F-X (2025+). Amortised and provably-robust neural simulation-based inference. In preparation.

Some more robust and conjugate Bayes

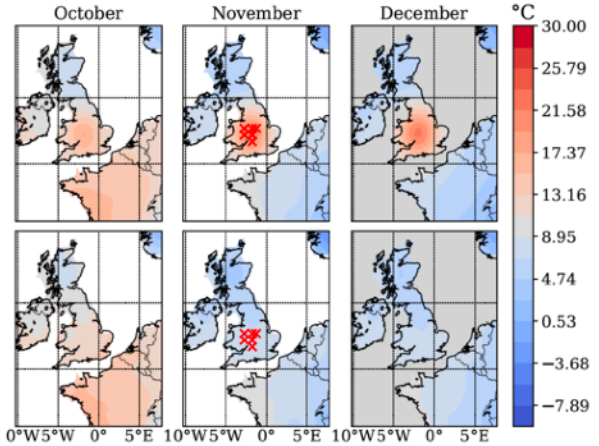
Change-point detection



Kalman filtering



Gaussian process regression



Altamirano, M., Briol, F.-X., & Knoblauch, J. (2023). Robust and scalable Bayesian online changepoint detection. *ICML*, 642–663.

Altamirano, M., Briol, F.-X., & Knoblauch, J. (2024). Robust and conjugate Gaussian process regression. *ICML,* 1155–1185.

Duran-Martin, G., Altamirano, M., Shestopaloff, A. Y., Knoblauch, J., Jones, M., Briol, F.-X., & Murphy, K. (2024). Outlier-robust Kalman filtering through generalised Bayes. *ICML*, 12138–12171.

Laplante, W., Altamirano, M., Duncan, A., Knoblauch, J., & Briol, F.-X. (2025). Robust and conjugate spatio-temporal Gaussian processes. *ICML*, 32562–32592.

Rooijakkers, J., Ronneberg, L., Briol, F.-X., Knoblauch, J. & Altamirano, M. (2025+). Multi-output robust and conjugate Gaussian processes. *Under review*.

Sun, Z., Barp, A., & Briol, F-X. (2025+). Generalised Bayes for spherical data. In preparation.

Bharti, A., Dellaporta, C., Hikida, Y., & Briol, F-X (2025+). Amortised and provably-robust neural simulation-based inference. In preparation.

DFD-Bayes

- What about for discrete spaces??

DFD-Bayes

- What about for discrete spaces??

$$D^{\mathsf{DFD}}(q\|p) := \mathbb{E}_{x\sim q}\left[\left\| \frac{\nabla^- p(x)}{p(x)} - \frac{\nabla^- q(x)}{q(x)} \right\|_2^2\right]$$

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2023). Generalised Bayesian inference for discrete intractable likelihood. *Journal of the American Statistical Association*, *119*(547), 2345–2355.

DFD-Bayes

- What about for discrete spaces??

$$\nabla^- q(x) := \left(q(x) - q(x^{1-}), \ldots, q(x) - q(x^{d-})\right)^\top$$

$$x^{j-} := (x_1, \ldots, x_j - 1, \ldots, x_d)$$

$$D^{\mathsf{DFD}}(q\|p) := \mathbb{E}_{x\sim q}\left[\left\| \frac{\nabla^- p(x)}{p(x)} - \frac{\nabla^- q(x)}{q(x)} \right\|_2^2\right]$$

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2023). Generalised Bayesian inference for discrete intractable likelihood. *Journal of the American Statistical Association*, *119*(547), 2345–2355.

DFD-Bayes

- What about for discrete spaces??

$$\nabla^- q(x) := \left(q(x) - q(x^{1-}), \ldots, q(x) - q(x^{d-})\right)^\top$$

$$x^{j-} := (x_1, \ldots, x_j - 1, \ldots, x_d)$$

$$D^{\mathsf{DFD}}(q\|p) := \mathbb{E}_{x\sim q}\left[\left\| \frac{\nabla^- p(x)}{p(x)} - \frac{\nabla^- q(x)}{q(x)} \right\|_2^2\right]$$

$$= \mathbb{E}_{x\sim q}\left[\sum_{j=1}^{d}\left(\frac{p(x^{j-})}{p(x)}\right)^2 - 2\left(\frac{p(x)}{p(x^{j+})}\right)\right] + C^{\mathsf{DFD}}(q),$$

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2023). Generalised Bayesian inference for discrete intractable likelihood. *Journal of the American Statistical Association*, *119*(547), 2345–2355.

DFD-Bayes

- What about for discrete spaces??

$$\nabla^- q(x) := \left(q(x) - q(x^{1-}), \ldots, q(x) - q(x^{d-})\right)^\top$$

$$x^{j-} := (x_1, \ldots, x_j - 1, \ldots, x_d)$$

$$D^{\mathrm{DFD}}(q\|p) := \mathbb{E}_{x\sim q}\left[\left\| \frac{\nabla^- p(x)}{p(x)} - \frac{\nabla^- q(x)}{q(x)} \right\|_2^2\right]$$

Crucially independent of p!

$$= \mathbb{E}_{x\sim q}\left[\sum_{j=1}^{d}\left(\frac{p(x^{j-})}{p(x)}\right)^2 - 2\left(\frac{p(x)}{p(x^{j+})}\right)\right] + C^{\mathrm{DFD}}(q),$$

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2023). Generalised Bayesian inference for discrete intractable likelihood. *Journal of the American Statistical Association*, *119*(547), 2345–2355.

DFD-Bayes

- What about for discrete spaces??

$$\nabla^- q(x) := \left(q(x) - q(x^{1-}), \ldots, q(x) - q(x^{d-})\right)^\top$$

$$x^{j-} := (x_1, \ldots, x_j - 1, \ldots, x_d)$$

$$D^{\mathrm{DFD}}(q\|p) := \mathbb{E}_{x\sim q}\left[\left\| \frac{\nabla^- p(x)}{p(x)} - \frac{\nabla^- q(x)}{q(x)} \right\|_2^2\right]$$

Crucially independent of p!

Can be approximated with Monte Carlo

$$= \mathbb{E}_{x\sim q}\left[\sum_{j=1}^{d}\left(\frac{p(x^{j-})}{p(x)}\right)^2 - 2\left(\frac{p(x)}{p(x^{j+})}\right)\right] + C^{\mathrm{DFD}}(q),$$

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2023). Generalised Bayesian inference for discrete intractable likelihood. *Journal of the American Statistical Association*, *119*(547), 2345–2355.

DFD-Bayes

- What about for discrete spaces??

$$\nabla^- q(x) := \left(q(x) - q(x^{1-}), \ldots, q(x) - q(x^{d-})\right)^\top$$

$$x^{j-} := (x_1, \ldots, x_j - 1, \ldots, x_d)$$

$$D^{\mathsf{DFD}}(q\|p) := \mathbb{E}_{x\sim q}\left[\left\| \frac{\nabla^- p(x)}{p(x)} - \frac{\nabla^- q(x)}{q(x)} \right\|_2^2\right]$$

$$= \mathbb{E}_{x\sim q}\left[\sum_{j=1}^{d}\left(\frac{p(x^{j-})}{p(x)}\right)^2 - 2\left(\frac{p(x)}{p(x^{j+})}\right)\right] + C^{\mathsf{DFD}}(q),$$

→ Does not depend on $Z(\theta)$ but….. is not conjugate!!

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, C. J. (2023). Generalised Bayesian inference for discrete intractable likelihood. *Journal of the American Statistical Association*, *119*(547), 2345–2355.

Log-ratio matching

$$D^{\mathsf{LRM}}(q\|p) := \mathbb{E}_{x\sim q}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p(x')}{p(x)} - \log\frac{q(x')}{q(x)}\right)^2\right]$$

Log-ratio matching

$$D^{\mathrm{LRM}}(q\|p) := \mathbb{E}_{x\sim q}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p(x')}{p(x)} - \log\frac{q(x')}{q(x)}\right)^2\right]$$

Intuition: $p_\theta \propto \exp(a\theta + b) \Rightarrow \log p_\theta \propto a'\theta + b' \Rightarrow (\log p_\theta)^2 \propto (a''\theta + b'')^2$

Log-ratio matching

$$D^{\mathrm{LRM}}(q\|p) := \mathbb{E}_{x\sim q}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p(x')}{p(x)} - \log\frac{q(x')}{q(x)}\right)^2\right]$$

Intuition: $p_\theta \propto \exp(a\theta + b) \Rightarrow \log p_\theta \propto a'\theta + b' \Rightarrow (\log p_\theta)^2 \propto (a''\theta + b'')^2$

Quadratic losses can give us conjugacy with Gaussian priors!

Log-ratio matching

$$D^{\mathrm{LRM}}(q\|p) := \mathbb{E}_{x\sim q}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p(x')}{p(x)} - \log\frac{q(x')}{q(x)}\right)^2\right]$$

Intuition: $p_\theta \propto \exp(a\theta + b) \Rightarrow \log p_\theta \propto a'\theta + b' \Rightarrow (\log p_\theta)^2 \propto (a''\theta + b'')^2$

Quadratic losses can give us conjugacy with Gaussian priors!

- Still a divergence under mild conditions.
- $M(x)$ allows us to compare against more than direct neighbours.

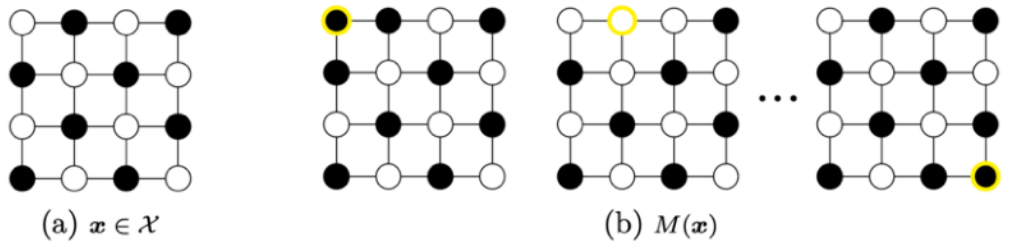
Log-ratio matching

$$D^{\mathrm{LRM}}(q\|p) := \mathbb{E}_{x\sim q}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p(x')}{p(x)} - \log\frac{q(x')}{q(x)}\right)^2\right]$$

Intuition: $p_\theta \propto \exp(a\theta + b) \Rightarrow \log p_\theta \propto a'\theta + b' \Rightarrow (\log p_\theta)^2 \propto (a''\theta + b'')^2$

Quadratic losses can give us conjugacy with Gaussian priors!

- Still a divergence under mild conditions.
- $M(x)$ allows us to compare against more than direct neighbours.



Log-ratio matching estimator

Loss:

$$\mathscr{L}^{\mathrm{LRM}}(\theta) := \mathbb{E}_{x\sim q_0}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p_\theta(x')}{p_\theta(x)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x)}\log\frac{q_0(x')}{q_0(x)}\right].$$

Log-ratio matching estimator

Loss:

$$\mathscr{L}^{\mathrm{LRM}}(\theta) := \mathbb{E}_{x\sim q_0}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p_\theta(x')}{p_\theta(x)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x)}\log\frac{q_0(x')}{q_0(x)}\right].$$

Empirical loss:

$$\hat{\mathscr{L}}^{\mathrm{LRM}}(\theta) := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|M(x_i)|}\sum_{x'\in M(x_i)}\left(\log\frac{p_\theta(x')}{p_\theta(x_i)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x_i)}\log\frac{q_0(x')}{q_0(x_i)}$$

Log-ratio matching estimator

Loss:

$$\mathscr{L}^{\mathrm{LRM}}(\theta) := \mathbb{E}_{x\sim q_0}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p_\theta(x')}{p_\theta(x)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x)}\log\frac{q_0(x')}{q_0(x)}\right].$$

Empirical loss:

$$\hat{\mathscr{L}}^{\mathrm{LRM}}(\theta) := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|M(x_i)|}\sum_{x'\in M(x_i)}\left(\log\frac{p_\theta(x')}{p_\theta(x_i)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x_i)}\log\frac{q_0(x')}{q_0(x_i)}$$

Hang on…. How on earth do we get $q_0(x)$??

Log-ratio matching estimator

Loss:

$$\mathscr{L}^{\mathrm{LRM}}(\theta) := \mathbb{E}_{x\sim q_0}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p_\theta(x')}{p_\theta(x)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x)}\log\frac{q_0(x')}{q_0(x)}\right].$$

Empirical loss:

$$\hat{\mathscr{L}}^{\mathrm{LRM}}(\theta) := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|M(x_i)|}\sum_{x'\in M(x_i)}\left(\log\frac{p_\theta(x')}{p_\theta(x_i)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x_i)}\log\frac{q_0(x')}{q_0(x_i)}\frac{\hat{q}_\alpha(x')}{\hat{q}_\alpha(x_i)}$$

Laplace smoothing:

$$\hat{q}_\alpha(x) = \frac{C_n(x) + \alpha\tilde{q}^\dagger(x)}{n + \alpha Z^\dagger}$$

Log-ratio matching estimator

Loss:

$$\mathscr{L}^{\mathrm{LRM}}(\theta) := \mathbb{E}_{x\sim q_0}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p_\theta(x')}{p_\theta(x)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x)}\log\frac{q_0(x')}{q_0(x)}\right].$$

Empirical loss:

$$\hat{\mathscr{L}}^{\mathrm{LRM}}(\theta) := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|M(x_i)|}\sum_{x'\in M(x_i)}\left(\log\frac{p_\theta(x')}{p_\theta(x_i)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x_i)}\log\frac{q_0(x')}{q_0(x_i)}\frac{\hat{q}_\alpha(x')}{\hat{q}_\alpha(x_i)}$$

Laplace smoothing:

$$\hat{q}_\alpha(x) = \frac{C_n(x) + \alpha\tilde{q}^\dagger(x)}{n + \alpha Z^\dagger}$$

Reference PMF: $q^\dagger(x) = \dfrac{\tilde{q}^\dagger}{Z^\dagger}$

Log-ratio matching estimator

Loss:

$$\mathscr{L}^{\mathrm{LRM}}(\theta) := \mathbb{E}_{x\sim q_0}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p_\theta(x')}{p_\theta(x)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x)}\log\frac{q_0(x')}{q_0(x)}\right].$$

Empirical loss:

$$\hat{\mathscr{L}}^{\mathrm{LRM}}(\theta) := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|M(x_i)|}\sum_{x'\in M(x_i)}\left(\log\frac{p_\theta(x')}{p_\theta(x_i)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x_i)}\log\frac{q_0(x')}{q_0(x_i)}\frac{\hat{q}_\alpha(x')}{\hat{q}_\alpha(x_i)}$$

Laplace smoothing:

$$\hat{q}_\alpha(x) = \frac{C_n(x) + \alpha\tilde{q}^\dagger(x)}{n + \alpha Z^\dagger}$$

Hyperparameter!

Reference PMF: $q^\dagger(x) = \dfrac{\tilde{q}^\dagger}{Z^\dagger}$

Log-ratio matching estimator

Loss:

$$\mathscr{L}^{\mathrm{LRM}}(\theta) := \mathbb{E}_{x\sim q_0}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)}\left(\log\frac{p_\theta(x')}{p_\theta(x)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x)}\log\frac{q_0(x')}{q_0(x)}\right].$$

Empirical loss:

$$\hat{\mathscr{L}}^{\mathrm{LRM}}(\theta) := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|M(x_i)|}\sum_{x'\in M(x_i)}\left(\log\frac{p_\theta(x')}{p_\theta(x_i)}\right)^2 - 2\log\frac{p_\theta(x')}{q_0(x_i)}\log\frac{q_0(x')}{q_0(x_i)}\frac{\hat{q}_\alpha(x')}{\hat{q}_\alpha(x_i)}$$

Laplace smoothing:

Count: $C_n(x) := \sum_{i=1}^{n}\delta(x_i = x)$

$$\hat{q}_\alpha(x) = \frac{C_n(x) + \alpha\tilde{q}^\dagger(x)}{n + \alpha Z^\dagger}$$

Hyperparameter!

Reference PMF: $q^\dagger(x) = \frac{\tilde{q}^\dagger}{Z^\dagger}$

Conjugacy with LRM!

Recall exponential families: $p_\theta(x) := \exp\left(\eta(\theta)^\top T(x) + B(x) - \log Z(\theta)\right)$

Conjugacy with LRM!

$$\Lambda_n := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{|M(x_i)|} \sum_{x' \in M(x_i)} \big(T(x') - T(x_i)\big)\big(T(x') - T(x_i)\big)^\top$$

$$\nu_n := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{|M(x_i)|} \sum_{x' \in M(x_i)} \big(T(x') - T(x_i)\big)\left(\log \frac{\hat{q}_\alpha(x')}{\hat{q}_\alpha(x_i)} - \big(B(x') - B(x_i)\big)\right)$$

Recall exponential families: $p_\theta(x) := \exp\left(\eta(\theta)^\top T(x) + B(x) - \log Z(\theta)\right)$

For those, the LRM becomes quadratic: $\hat{\mathscr{L}}^{\mathrm{LRM}}(\theta) = \eta(\theta)^\top \Lambda_n \eta(\theta) - 2\eta(\theta)^\top \nu_n + c_n,$

Conjugacy with LRM!

$$\Lambda_n := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{|M(x_i)|} \sum_{x' \in M(x_i)} \big(T(x') - T(x_i)\big)\big(T(x') - T(x_i)\big)^\top$$

$$\nu_n := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{|M(x_i)|} \sum_{x' \in M(x_i)} \big(T(x') - T(x_i)\big)\left(\log \frac{\hat{q}_\alpha(x')}{\hat{q}_\alpha(x_i)} - \big(B(x') - B(x_i)\big)\right)$$

Recall exponential families: $p_\theta(x) := \exp\left(\eta(\theta)^\top T(x) + B(x) - \log Z(\theta)\right)$

For those, the LRM becomes quadratic: $\hat{\mathscr{L}}^{\mathsf{LRM}}(\theta) = \eta(\theta)^\top \Lambda_n \eta(\theta) - 2\eta(\theta)^\top \nu_n + c_n,$

Suppose we have a Gaussian prior: $\pi(\eta) \propto \exp\left(-\frac{1}{2}(\eta - \mu)^\top \Sigma^{-1}(\eta - \mu)\right)$

Conjugacy with LRM!

$$\Lambda_n := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{|M(x_i)|} \sum_{x' \in M(x_i)} \big(T(x') - T(x_i)\big)\big(T(x') - T(x_i)\big)^\top$$

$$\nu_n := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{|M(x_i)|} \sum_{x' \in M(x_i)} \big(T(x') - T(x_i)\big)\left(\log \frac{\hat{q}_\alpha(x')}{\hat{q}_\alpha(x_i)} - \big(B(x') - B(x_i)\big)\right)$$

Recall exponential families: $p_\theta(x) := \exp\left(\eta(\theta)^\top T(x) + B(x) - \log Z(\theta)\right)$

For those, the LRM becomes quadratic: $\hat{\mathscr{L}}^{\mathrm{LRM}}(\theta) = \eta(\theta)^\top \Lambda_n \eta(\theta) - 2\eta(\theta)^\top \nu_n + c_n,$

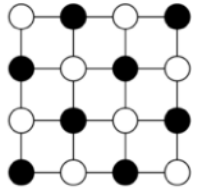
Suppose we have a Gaussian prior: $\pi(\eta) \propto \exp\left(-\frac{1}{2}(\eta - \mu)^\top \Sigma^{-1}(\eta - \mu)\right)$

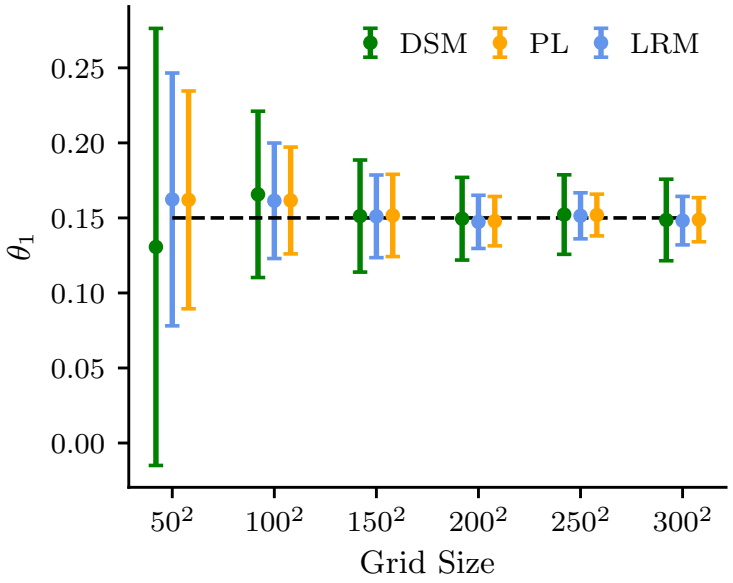
Conjugate Gaussian posterior: $\pi^{\beta}_{\mathrm{LRM}}\left(\eta;\ \{x_i\}_{i=1}^n\right) = \mathcal{N}(\eta; \mu_n, \Sigma_n)$

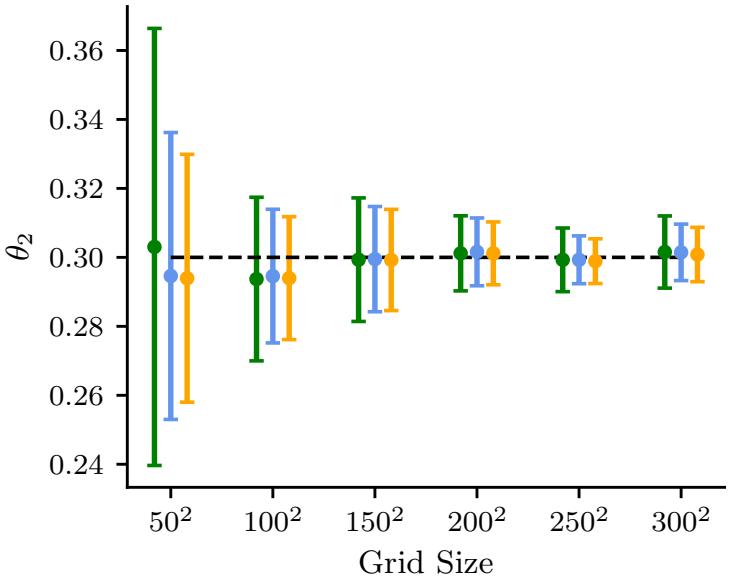
$$\Sigma_n := \left(\Sigma^{-1} + 2\beta\Lambda_n\right)^{-1}$$
$$\mu_n := \Sigma_n\left(\Sigma^{-1}\mu + 2\beta\nu_n\right)$$

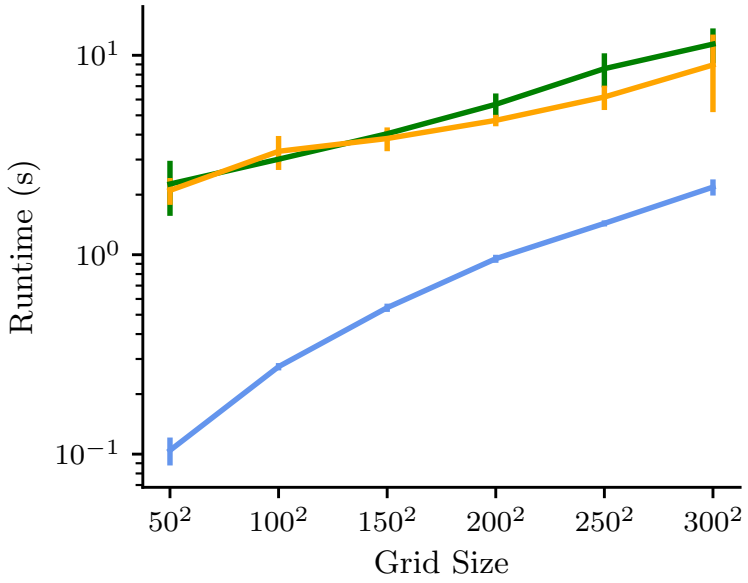
Ising model

$$p_\theta(x) = \frac{1}{Z(\theta)} \exp\left(\theta_1 \sum_{i=1}^{N} x_i + \theta_2 \sum_{i \sim j} x_i x_j\right)$$





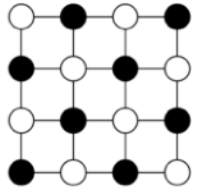


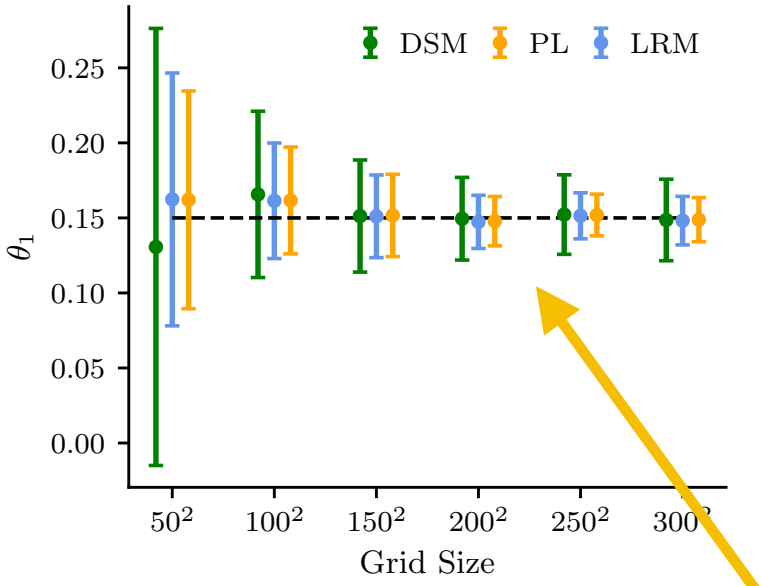


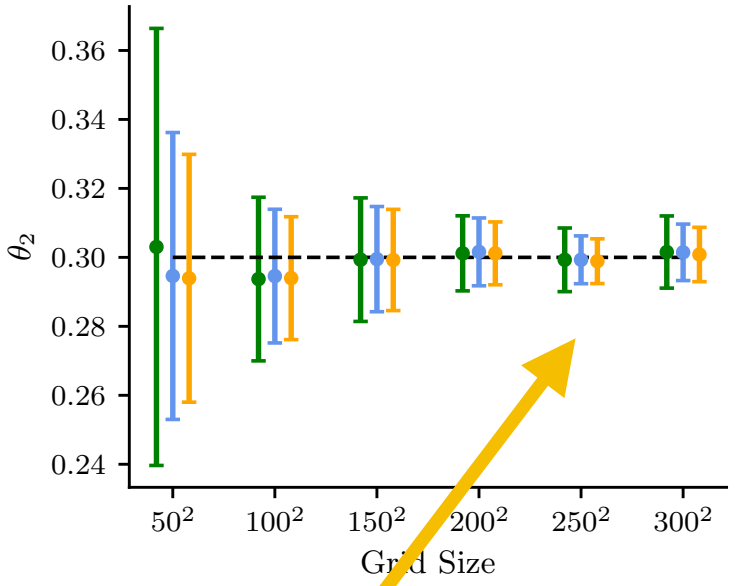
Dimensionality of data: up to 90000
Number of parameters: 2

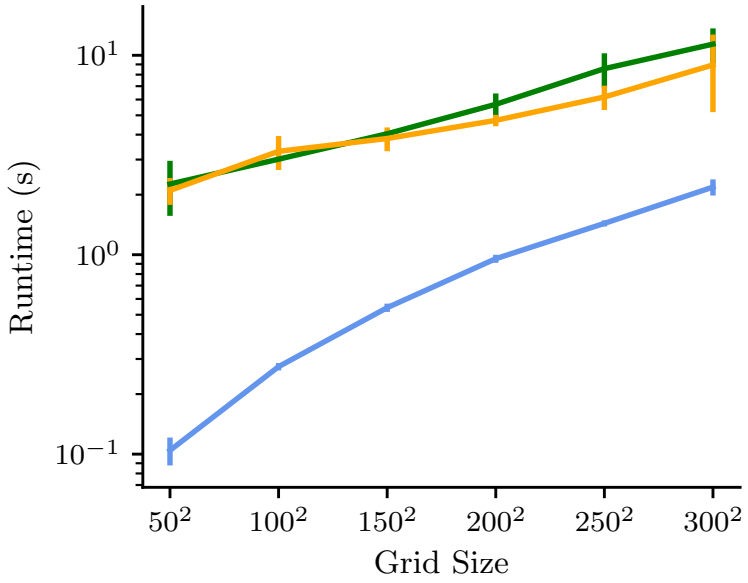
Ising model

$$p_\theta(x) = \frac{1}{Z(\theta)} \exp\left(\theta_1 \sum_{i=1}^{N} x_i + \theta_2 \sum_{i \sim j} x_i x_j\right)$$







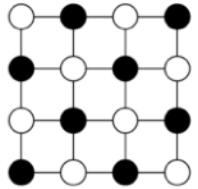


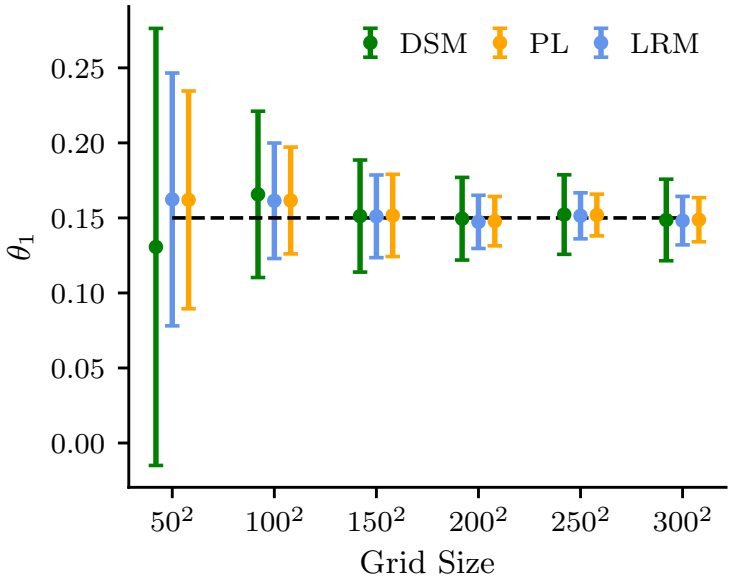
Almost identical fit

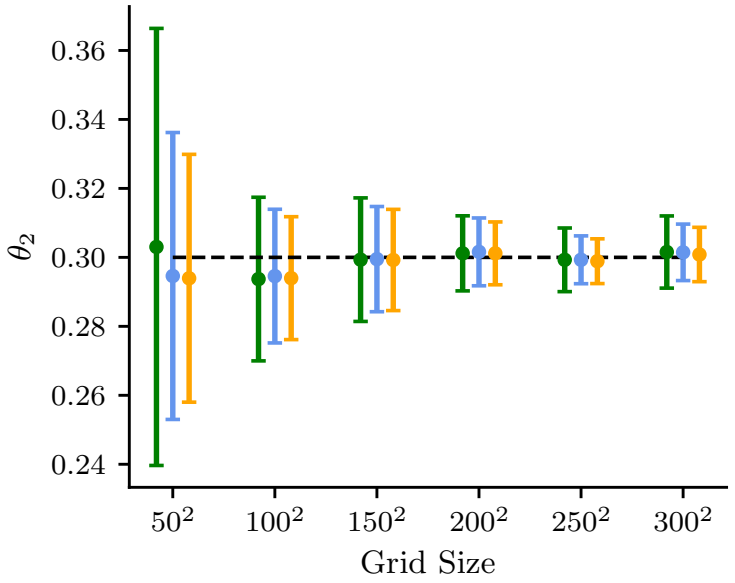
Dimensionality of data: up to 90000
Number of parameters: 2

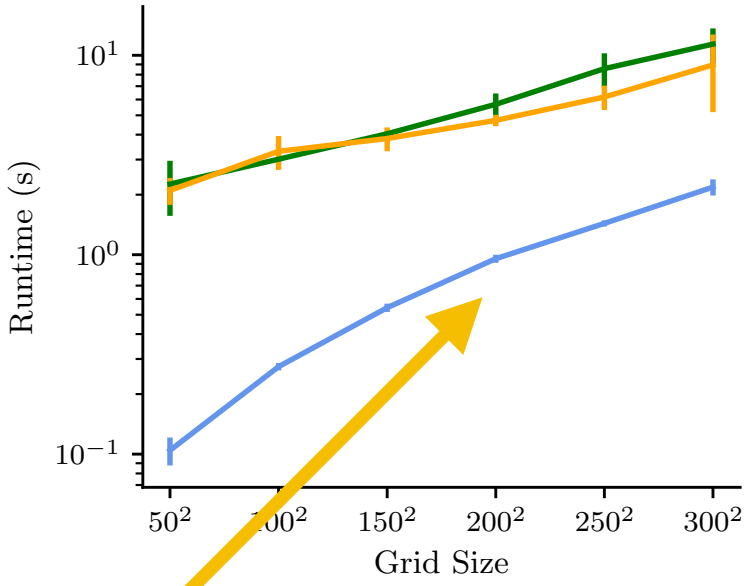
Ising model

$$p_\theta(x) = \frac{1}{Z(\theta)} \exp\left(\theta_1 \sum_{i=1}^{N} x_i + \theta_2 \sum_{i \sim j} x_i x_j\right)$$









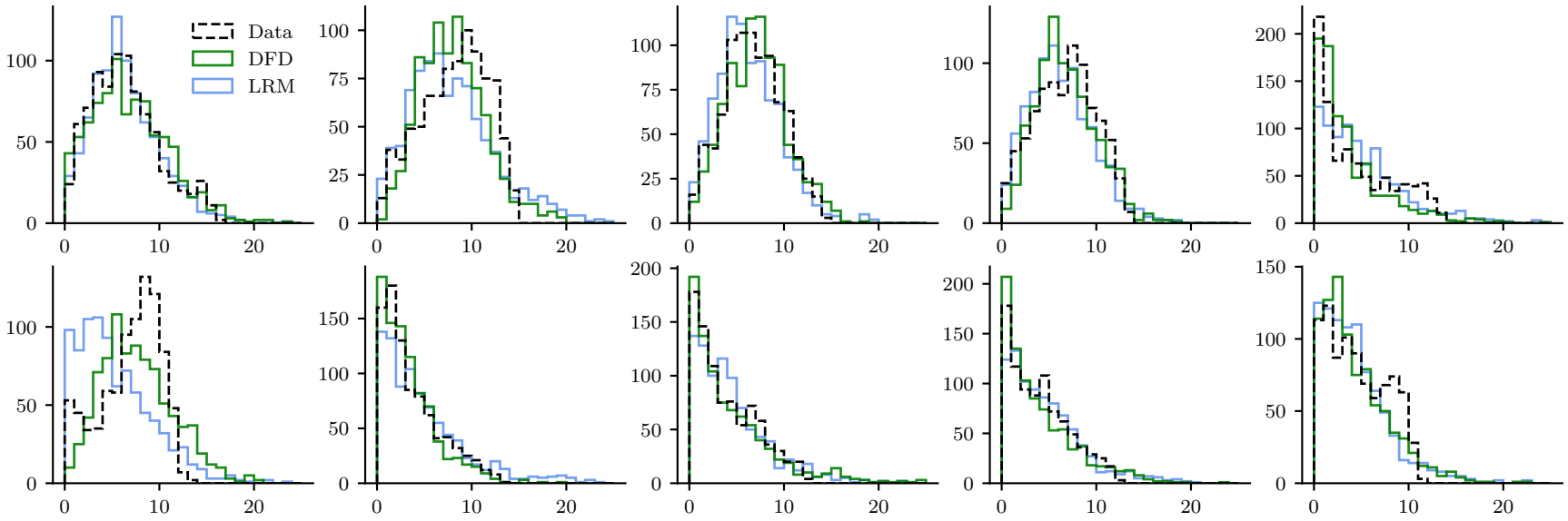
Almost identical fit … at a fraction of the cost!

Dimensionality of data: up to 90000
Number of parameters: 2

Breast cancer data set

$$p_\theta(x) \propto \exp\left(\sum_{i=1}^{d}\theta_i x_i - \sum_{i=1}^{d}\sum_{i<j}\theta_{i,j}x_i x_j - \sum_{i=1}^{d}\theta_{0,i}\log(x_i!)\right)$$

Conway-Maxwell-Poisson Graphical Model

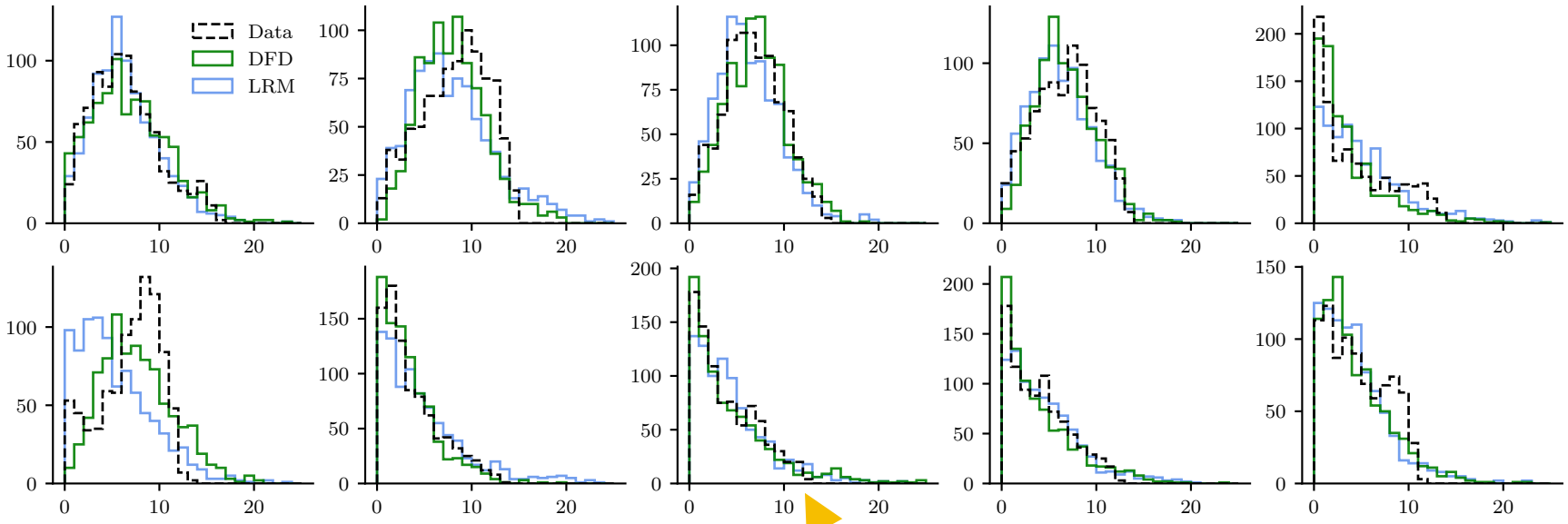


Dimension of data: 10,
Number of parameters: 64

Breast cancer data set

$$p_\theta(x) \propto \exp\left(\sum_{i=1}^{d}\theta_i x_i - \sum_{i=1}^{d}\sum_{i<j}\theta_{i,j}x_i x_j - \sum_{i=1}^{d}\theta_{0,i}\log(x_i!)\right)$$

Conway-Maxwell-Poisson Graphical Model



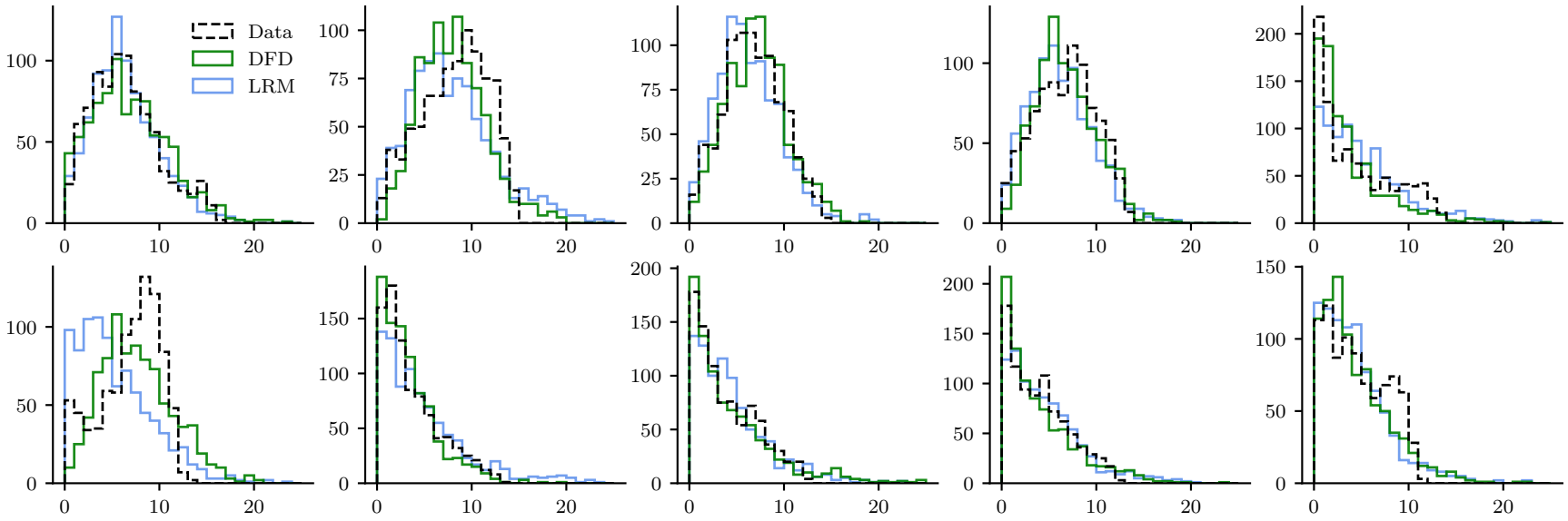
Almost identical fit

Dimension of data: 10,
Number of parameters: 64

Breast cancer data set

$$p_\theta(x) \propto \exp\left(\sum_{i=1}^{d} \theta_i x_i - \sum_{i=1}^{d}\sum_{i<j} \theta_{i,j} x_i x_j - \sum_{i=1}^{d} \theta_{0,i} \log(x_i!)\right)$$

Conway-Maxwell-Poisson Graphical Model



Almost identical fit … at a fraction of the cost!

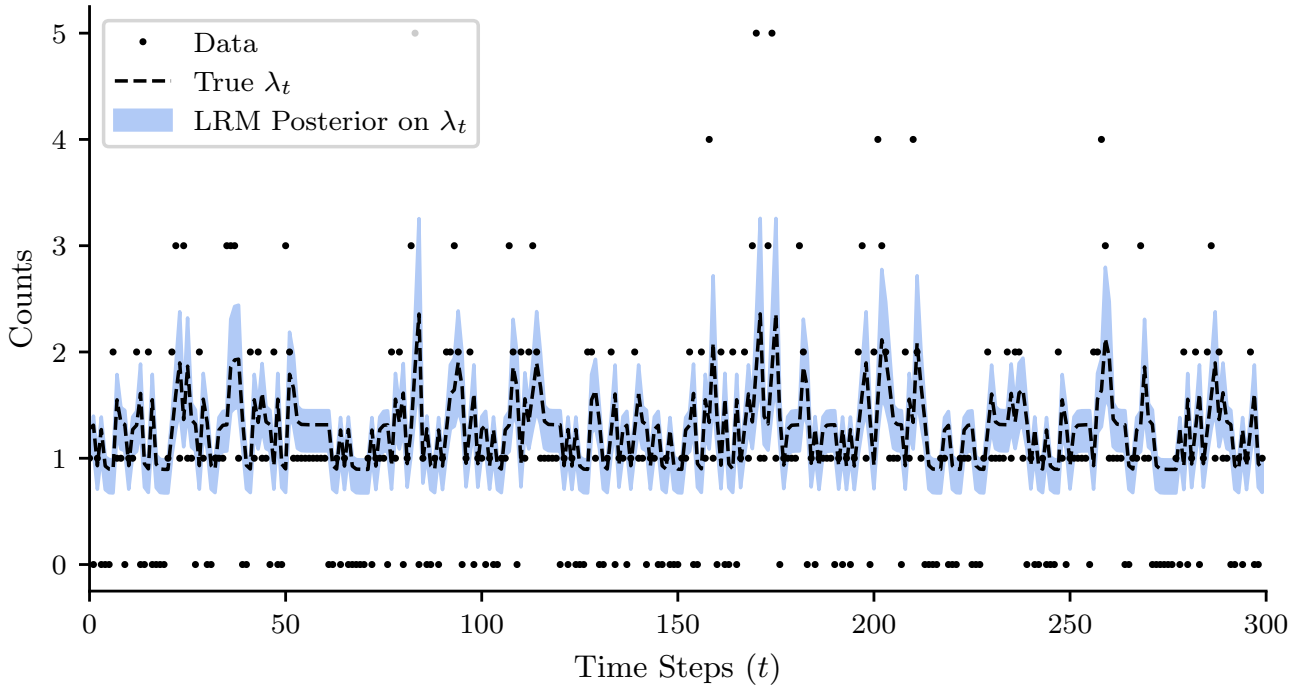
Dimension of data: 10,
Number of parameters: 64

DFD Bayes: 1896 seconds (i.e. 31mins)
LRM Bayes: **56 seconds**

Time-series of count data

$$x_t \sim \text{CMP}(\lambda_t, \theta_2)$$

$$\ln \lambda_t = \theta_0 + \varphi \ln \lambda_{t-1} + \theta_1 \ln(1 + x_{t-1}),$$

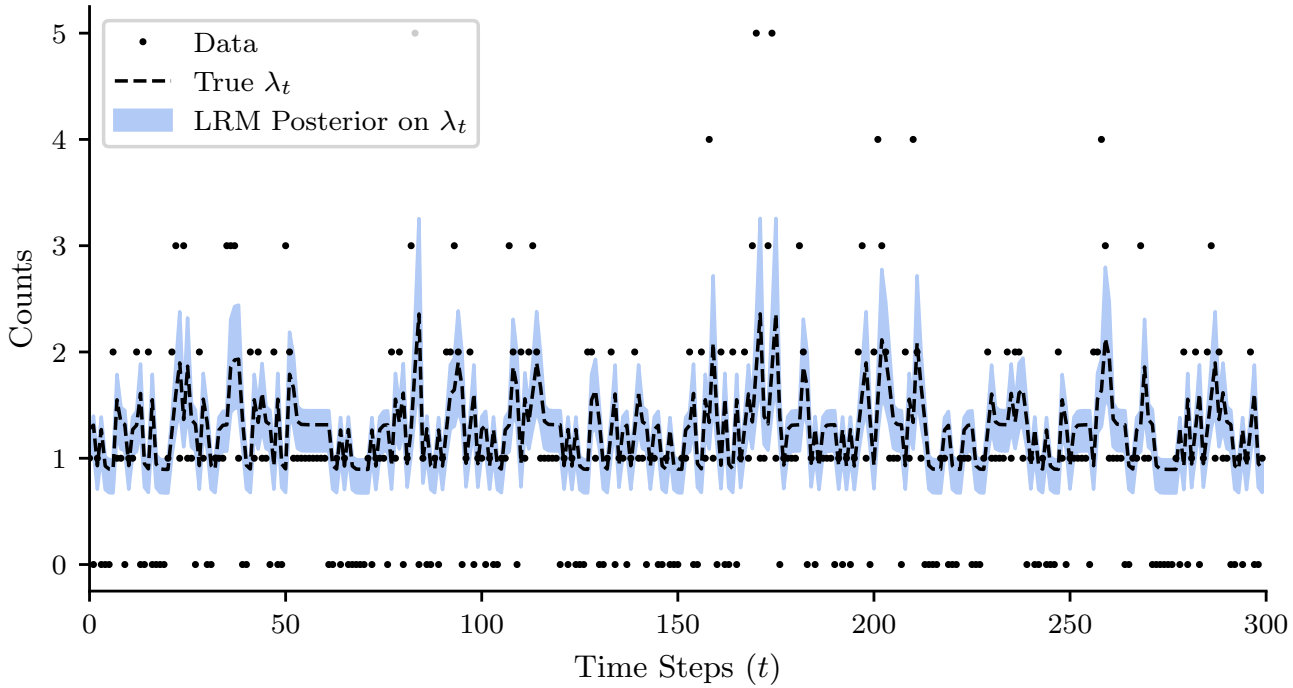


Dimension of data: T (length of time series),
Number of parameters: 3

Time-series of count data

$$x_t \sim \text{CMP}(\lambda_t, \theta_2)$$

$$\ln \lambda_t = \theta_0 + \varphi \ln \lambda_{t-1} + \theta_1 \ln(1 + x_{t-1}),$$



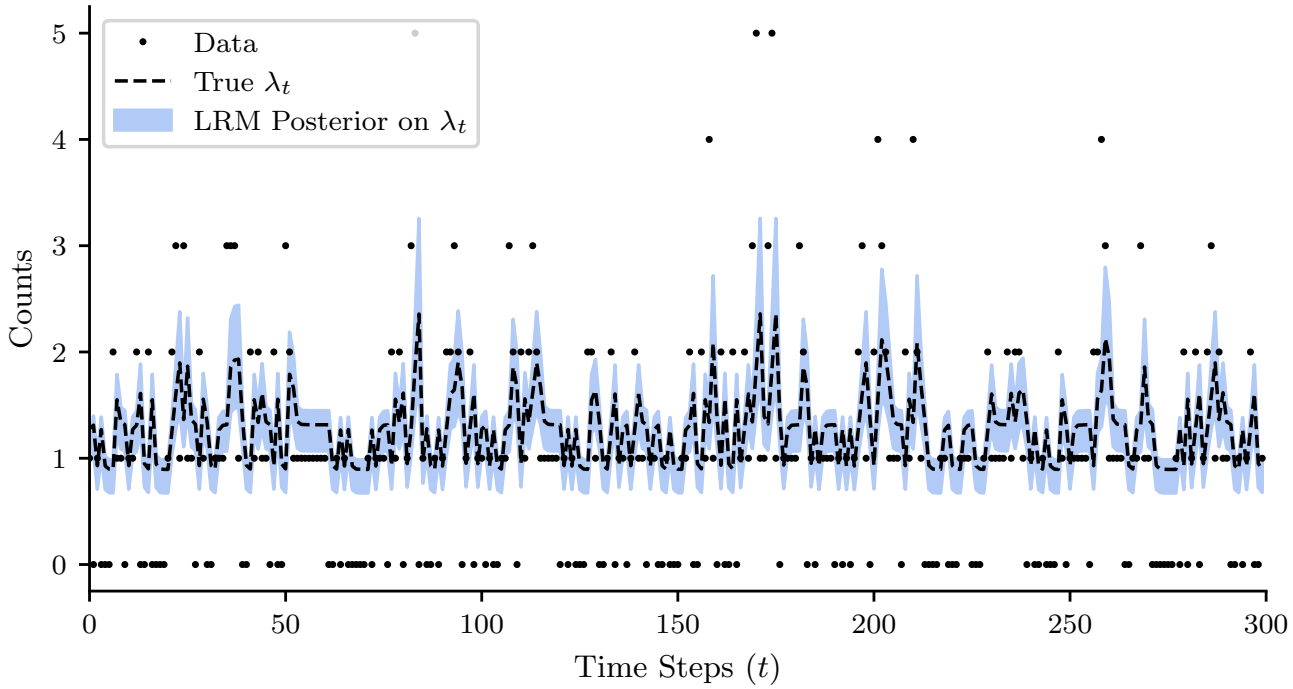
Need to approximate doubly-intractable posterior at each time step! **Completely intractable to do online for most problems**

Dimension of data: T (length of time series),
Number of parameters: 3

Time-series of count data

$$x_t \sim \text{CMP}(\lambda_t, \theta_2)$$

$$\ln \lambda_t = \theta_0 + \varphi \ln \lambda_{t-1} + \theta_1 \ln(1 + x_{t-1}),$$



Need to approximate doubly-intractable posterior at each time step! **Completely intractable to do online for most problems**

Dimension of data: T (length of time series),
Number of parameters: 3

Our method does this in 0.1 seconds..!

What I haven't had time to talk about….

- **Theory:** We can show fairly standard consistency and BvM under mild conditions!

$$\int_{B_\epsilon(\theta_\star)} \pi^{\beta}_{\text{LRM}}(\theta \,|\, \{x_i\}_{i=1}^n) d\theta \xrightarrow{\text{a.s}} 1.$$

What I haven't had time to talk about….

- **Theory:** We can show fairly standard consistency and BvM under mild conditions!

$$\int_{B_\epsilon(\theta_\star)} \pi^{\beta}_{\text{LRM}}(\theta \,|\, \{x_i\}_{i=1}^n) d\theta \xrightarrow{\text{a.s}} 1.$$

- **Robustness:** Straightforward to add some weights to enforce outlier-robustness.

$$D^{\text{w-LRM}}(q\|p) := \mathbb{E}_{x\sim q}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)} w(x)\left(\log\frac{p(x')}{p(x)} - \log\frac{q(x')}{q(x)}\right)^2\right]$$

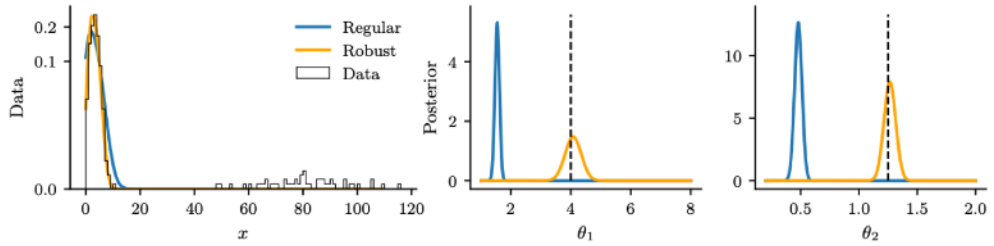
What I haven't had time to talk about….

- **Theory:** We can show fairly standard consistency and BvM under mild conditions!

$$\int_{B_\epsilon(\theta_\star)} \pi^{\beta}_{\text{LRM}}(\theta \,|\, \{x_i\}_{i=1}^n) d\theta \xrightarrow{\text{a.s}} 1.$$

- **Robustness:** Straightforward to add some weights to enforce outlier-robustness.

$$D^{\text{w-LRM}}(q\|p) := \mathbb{E}_{x\sim q}\left[\frac{1}{|M(x)|}\sum_{x'\in M(x)} w(x)\left(\log\frac{p(x')}{p(x)} - \log\frac{q(x')}{q(x)}\right)^2\right]$$



1D Conway-Maxwell-Poisson model

Any Questions?

Laplante, W., Altamirano, M., Duncan, A. D., Knoblauch, J. & Briol, F.-X. (2025+). *Conjugate generalised Bayesian Inference for discrete doubly intractable problems*. To appear.