Sampling based inference for non normalized statistical models

Lionel Riou-Durand, joint work with Nicolas Chopin

CREST - ENSAE - Université Paris Saclay

Objectives

- Non normalized statistical models involve intractable partition functions.
- Sampling based approximations have been proposed for inference, e.g. MC-MLE (Geyer 1994), NCE (Gutmann and Hyvärinen 2012).
- We extend asymptotic guaranties of NCE and MC-MLE under two asymptotic regimes.
- We prove that NCE is more robust than MC-MLE to a bad choice of sampling distribution.
- We compare the sensibility of NCE and MC-MLE to the sampling distribution through numerical experiments.

Non normalized models

- Consider IID data $Y_1, \cdots, Y_n \sim \mathbb{P}_\theta$ on $\mathcal{X} \subset \mathbb{R}^p$ with density:

$$f_\theta(x) = \frac{h_\theta(x)}{\mathcal{Z}(\theta)}, \qquad \theta \in \Theta \subset \mathbb{R}^d$$

- The model $\{\mathbb{P}_\theta\}$ is called **non normalized** if
  - $h_\theta(x)$ can be computed pointwise, but...
  - $\mathcal{Z}(\theta) = \int_\mathcal{X} h_\theta(x)\mu(\mathrm{d}x)$ is intractable.

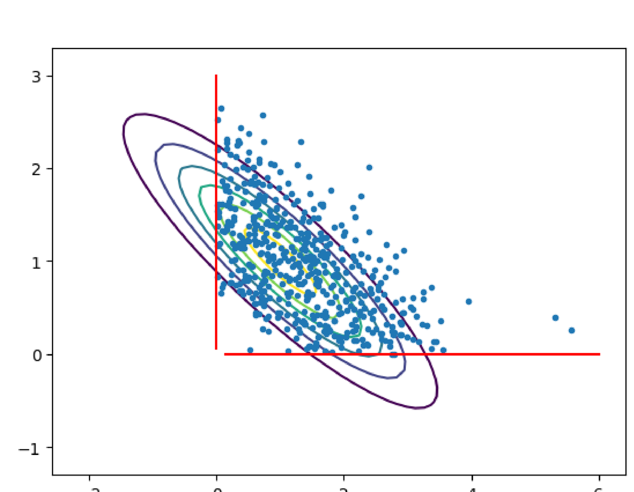
Example: Truncated Gaussian model

- Consider $Y_1, ..., Y_n \overset{i.i.d.}{\sim} \mathcal{N}_p(\mu, \Sigma)$ truncated to $]0, +\infty[^p$, with density

$$f_{\mu,\Sigma}(x) \propto \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} \mathbb{1}_{]0,+\infty[^p}(x)$$

- Problem: $\mathcal{Z}(\mu,\Sigma) = \mathbb{P}\left(\mathcal{N}_p(\mu,\Sigma) \in]0, +\infty[^p\right)$ is intractable for almost every (μ, Σ), and numerical approximations become inefficient for large p.

- $\mathcal{Z}(\mu, \Sigma)$ is intractable $\Rightarrow$ Maximum Likelihood Estimator is intractable.
- Question: How to infer μ and Σ?



Sampling based approximations

- Monte-Carlo Maximum Likelihood Estimation (Geyer 1994): approximating the likelihood by a **pointwise importance sampling estimate** of the partition function. $\rightarrow \hat{\theta}_{IS}$
- Noise Contrastive Estimation (Gutmann and Hyvärinen 2012): Learning parameters from a **logistic classification task** between data-points and 'noise'. $\rightarrow \hat{\theta}_{NCE}$
- Both methods require the sampling of r.v. $X_1, ..., X_m$ on $\mathcal{X}$ from a reference distribution $\mathbb{P}_\psi$ (e.g. Markov chain Monte Carlo sampling).

Previous results

- **Observed data:** $Y_1, ..., Y_n \overset{i.i.d.}{\sim} \mathbb{P}_\theta$
- **Monte Carlo sample:** $X_1, ..., X_m \sim \mathbb{P}_\psi$
- Geyer (1994): n is fixed and $m \rightarrow \infty$
  - $\hat{\theta}^{\mathrm{IS}}_{n,m}$ is an 'estimator' of the MLE, consistent and asymptotically normal.
- Gutmann, Hyvarïnen (2012): $m = \tau n \rightarrow +\infty$, $\tau > 0$
  - $\hat{\theta}^{\mathrm{NCE}}_{n,m}$ is an estimator of θ^*, consistent and asymptotically normal (when $X_1, ..., X_m$ IID).
  - Numerical evidence that NCE can outperform MC-MLE, especially for small $\tau = m/n$.

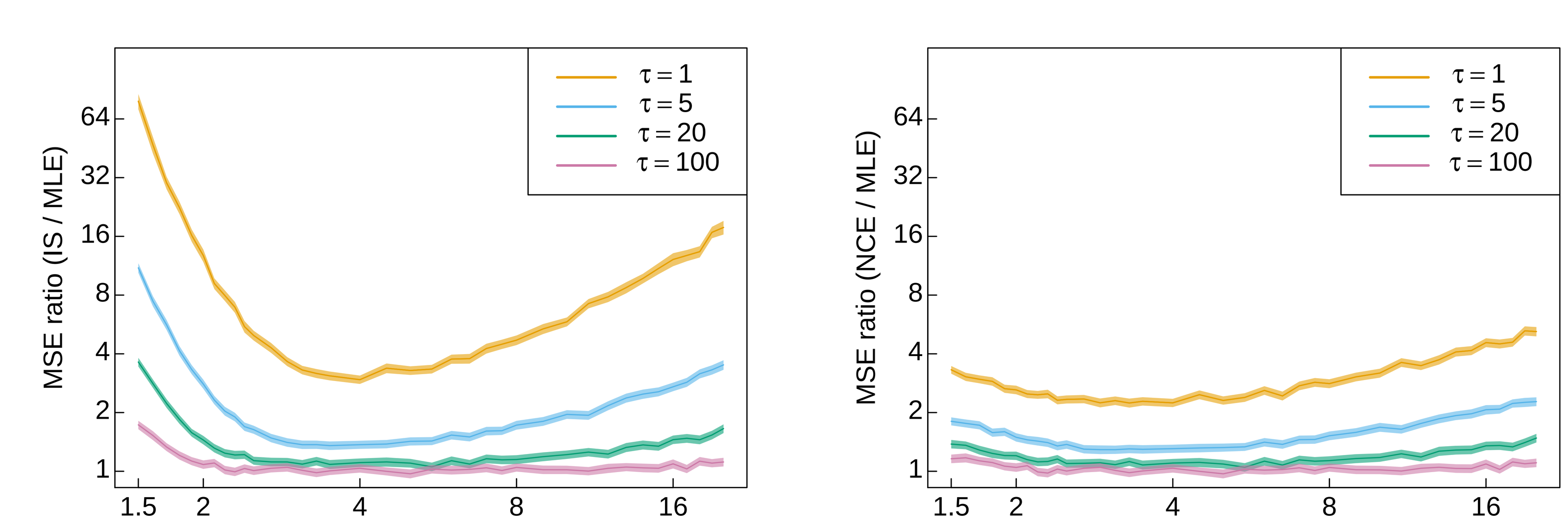
Asymptotic comparison

- Riou-Durand and Chopin 2018: Extension of asymptotic guaranties for MC-MLE and NCE (when $m \rightarrow \infty$)
  - consistency and $\sqrt{m}$-CLT's when (X_j) is sampled by MCMC
  - two asymptotic regimes

Monte Carlo error	Overall inferential error
n fixed	$m/n \rightarrow \tau \in (0, \infty)$
$m^{1-\varepsilon}(\hat{\theta}_{NCE} - \hat{\theta}_{IS}) \overset{\text{a.s.}}{\rightarrow} 0$	$\mathbb{V}_{as}(\hat{\theta}_{NCE}) \preccurlyeq \mathbb{V}_{as}(\hat{\theta}_{IS})$
asymp. equivalence	NCE is more 'robust'

Sensibility to the sampling distribution

- Numerical example (Gaussian Truncated Model):
  $Y_1, ..., Y_n \overset{i.i.d.}{\sim} \mathcal{N}_p(\mu, \Sigma)$ truncated to $]0, +\infty[^p$,
  $X_1, ...X_m \overset{i.i.d.}{\sim} \mathcal{N}_p(0, \lambda I_p)$ truncated to $]0, +\infty[^p$.
- Mean Square Error ratios of MC-MLE (left) and NCE (right):



References

[Gey94] Charles J Geyer. "On the convergence of Monte Carlo maximum likelihood calculations". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1994), pp. 261–274.

[GH12] M. U. Gutmann and A. Hyvärinen. "Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics". In: *J. Mach. Learn. Res.* 13.1 (2012), pp. 137–361.

[RC18] Lionel Riou-Durand and Nicolas Chopin. "Noise contrastive estimation: Asymptotic properties, formal comparison with MC-MLE". In: *Electronic Journal of Statistics* 12.2 (2018), pp. 3473–3518.

Note: The Truncated Gaussian model picture was taken from https://alan-turing-institute.github.io