Noise contrastive estimation: asymptotic properties, formal comparison with MCMC-MLE

Lionel Riou-Durand
University of Warwick

Bayesian Computation Reading Group
14/01/2022

Joint work with

Nicolas Chopin

ENSAE - Paris

Outline

Intractable integrals in Statistics

- Many statistical models involve **intractable integrals** in both Bayesian and frequentist frameworks

Intractable integrals in Statistics

- Many statistical models involve **intractable integrals** in both Bayesian and frequentist frameworks
- Among others, these include:

Intractable integrals in Statistics

- Many statistical models involve **intractable integrals** in both Bayesian and frequentist frameworks
- Among others, these include:
  - non computable likelihoods
  - evidence in Bayesian model selection
  - partition functions in graphical models

Intractable integrals in Statistics

- Many statistical models involve **intractable integrals** in both Bayesian and frequentist frameworks
- Among others, these include:
  - non computable likelihoods
  - evidence in Bayesian model selection
  - partition functions in graphical models
- In this work, we focus on the inference of **un-normalized models**

Un-normalized models

What are they ?

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

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

Un-normalized models

What are they ?

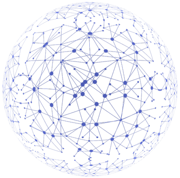
- Consider IID data $Y_1, \cdots, Y_n \sim \mathbb{P}_\theta$ on $\mathcal{X} \subset \mathbb{R}^p$ with density:

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

- The model $\{\mathbb{P}_\theta\}$ will be called **un-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.

Un-normalised models

Examples



- Example 1: Exponential Random Graph Model

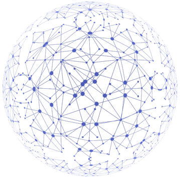
$$\mathbb{P}_\theta(Y = y) \propto \exp\left\{\theta^T S(y)\right\}, \qquad y \in \mathcal{X} = \{0,1\}^{C_n^2}$$

y : observation of a n-nodes network: a set of **random** connections between each of the C_n^2 pair of nodes.

$S(y)$: vector of **structural** statistics on the network (e.g. number of connexions, triangles, disconnected nodes...).

Un-normalised models

Examples



- Example 1: Exponential Random Graph Model

$$\mathbb{P}_\theta(Y = y) \propto \exp\left\{\theta^T S(y)\right\}, \qquad y \in \mathcal{X} = \{0, 1\}^{C_n^2}$$

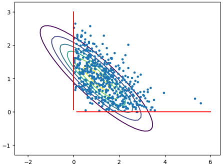
y : observation of a n-nodes network: a set of **random** connections between each of the C_n^2 pair of nodes.

$S(y)$: vector of **structural** statistics on the network (e.g. number of connexions, triangles, disconnected nodes...).

- Problem: $\mathcal{Z}(\theta) = \sum_{x \in \mathcal{X}} \exp\left\{\theta^T S(x)\right\}$ is not computable for large n, because $\mathcal{X}$ is a big set... ($|\mathcal{X}| = 2^{n(n-1)/2}$)

Un-normalised models

Examples



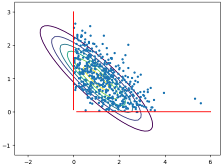
- Example 2: Truncated Gaussian Model

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

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

Un-normalised models

Examples



- Example 2: Truncated Gaussian Model

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

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

- 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)...

  (numerical approximations become inefficient for large p)

Un-normalised models

Examples

- Exponential families: a general framework.

  Assume that for some map $S(.)$ we have

$$h_\theta(x) = \exp\{\theta^\top S(x)\}, \qquad x \in \mathcal{X}.$$

Un-normalised models

Examples

- Exponential families: a general framework.

  Assume that for some map $S(.)$ we have

$$h_\theta(x) = \exp\{\theta^\top S(x)\}, \qquad x \in \mathcal{X}.$$

- Convenient properties, e.g. $S(x)$ sufficient statistic.

Un-normalised models

Examples

- Exponential families: a general framework.

  Assume that for some map $S(.)$ we have

$$h_\theta(x) = \exp\{\theta^\top S(x)\}, \qquad x \in \mathcal{X}.$$

- Convenient properties, e.g. $S(x)$ sufficient statistic.
- But no guarantee that $\mathcal{Z}(\theta) = \int_{\mathcal{X}} \exp\{\theta^\top S(x)\}\mu(\mathrm{d}x)$ is tractable.

Previous works

- Bayesian: Exchange algorithm (Murray, Ghahramani, and MacKay 2012), ABC, russian roulette (Lyne et al. 2015)...

Previous works

- Bayesian: Exchange algorithm (Murray, Ghahramani, and MacKay 2012), ABC, russian roulette (Lyne et al. 2015)...
- Frequentist: MCMC-MLE (Geyer 1994), noise contrastive estimation (Gutmann and Hyvärinen 2012)...

Previous works

- Bayesian: Exchange algorithm (Murray, Ghahramani, and MacKay 2012), ABC, russian roulette (Lyne et al. 2015)...
- Frequentist: MCMC-MLE (Geyer 1994), noise contrastive estimation (Gutmann and Hyvärinen 2012)...
- Here we focus on Monte Carlo approximations of the likelihood.

Summary

Inference for un-normalised statistical models

Intractable integrals in statistics

Un-normalised statistical models

Overview of the literature

Approximate likelihood via Monte Carlo

Monte Carlo Maximum Likelihood Estimation

Noise Contrastive estimation

Connections between NCE and MC-MLE

NCE: a formal comparison with MC-MLE

Two asymptotic regimes

Extension of asymptotic guarantees for NCE

Robustness to a bad choice of sampling distribution

Perspectives

Monte Carlo Maximum Likelihood Estimation

- Monte Carlo Maximum Likelihood Estimation (Geyer 1992)

Monte Carlo Maximum Likelihood Estimation

- Monte Carlo Maximum Likelihood Estimation (Geyer 1992)

Let $(X_j)_{j\geq 1}$ be a Markov Chain with stationary distribution

$$f_\psi(x) = \frac{h_\psi(x)}{\mathcal{Z}(\psi)}, \qquad \mathcal{Z}(\psi) = \int_{\mathcal{X}} h_\psi(x)\mu(\mathrm{d}x)$$

where $h_\psi(x)$ can be computed pointwise.

Monte Carlo Maximum Likelihood Estimation

- Monte Carlo Maximum Likelihood Estimation (Geyer 1992)

  Let $(X_j)_{j\geq 1}$ be a Markov Chain with stationary distribution

  $$f_\psi(x) = \frac{h_\psi(x)}{\mathcal{Z}(\psi)}, \qquad \mathcal{Z}(\psi) = \int_{\mathcal{X}} h_\psi(x)\mu(\mathrm{d}x)$$

  where $h_\psi(x)$ can be computed pointwise.

  If h_θ/h_ψ integrable, the law of large numbers applies: for any such $\theta \in \Theta$, almost surely

  $$\frac{1}{m}\sum_{j=1}^{m}\frac{h_\theta(X_j)}{h_\psi(X_j)} \underset{m\to\infty}{\longrightarrow} \frac{\mathcal{Z}(\theta)}{\mathcal{Z}(\psi)}.$$

Monte Carlo Maximum Likelihood Estimation

- Let $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \mathbb{P}_\theta$, and $\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}}\, \ell_n(\theta)$ where

$$\ell_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{h_\theta(y_i)}{h_\psi(y_i)} - \log \frac{\mathcal{Z}(\theta)}{\mathcal{Z}(\psi)}.$$

Monte Carlo Maximum Likelihood Estimation

- Let $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \mathbb{P}_\theta$, and $\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}}\, \ell_n(\theta)$ where

$$\ell_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{h_\theta(y_i)}{h_\psi(y_i)} - \log \frac{\mathcal{Z}(\theta)}{\mathcal{Z}(\psi)}.$$

Then for any fixed θ, almost surely

$$\ell^{\mathrm{IS}}_{n,m}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{h_\theta(y_i)}{h_\psi(y_i)} - \log \left\{ \frac{1}{m} \sum_{j=1}^{m} \frac{h_\theta(x_j)}{h_\psi(x_j)} \right\} \underset{m \to \infty}{\to} \ell_n(\theta).$$

Monte Carlo Maximum Likelihood Estimation

- Let $Y_1, \dots, Y_n \overset{i.i.d.}{\sim} \mathbb{P}_\theta$, and $\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}}\ \ell_n(\theta)$ where

$$\ell_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{h_\theta(y_i)}{h_\psi(y_i)} - \log \frac{\mathcal{Z}(\theta)}{\mathcal{Z}(\psi)}.$$

  Then for any fixed θ, almost surely

$$\ell_{n,m}^{\mathrm{IS}}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{h_\theta(y_i)}{h_\psi(y_i)} - \log \left\{ \frac{1}{m} \sum_{j=1}^{m} \frac{h_\theta(x_j)}{h_\psi(x_j)} \right\} \underset{m \to \infty}{\longrightarrow} \ell_n(\theta).$$

- Geyer (1994) proved that $\hat{\theta}_{n,m}^{\mathrm{IS}} = \underset{\theta \in \Theta}{\operatorname{argmax}}\ \ell_{n,m}^{IS}(\theta)$ is a consistent and asymptotically normal "estimator" of $\hat{\theta}_n$.

Noise Contrastive Estimation

- Noise Contrastive Estimation (Gutmann et Hyvarïnen, 2012).

Noise Contrastive Estimation

- Noise Contrastive Estimation (Gutmann et Hyvarïnen, 2012).

  Define the following likelihood (logistic classifier):

$$\ell_{n,m}^{\mathrm{NCE}}(\theta,\nu) = \sum_{i=1}^{n} \log q_{\theta,\nu}(y_i) + \sum_{i=1}^{m} \log\left\{1 - q_{\theta,\nu}(x_i)\right\}$$

where $\log\left\{\frac{q_{\theta,\nu}(x)}{1-q_{\theta,\nu}(x)}\right\} = \log\left\{\frac{h_\theta(x)}{h_\psi(x)}\right\} + \nu + \log\left(\frac{n}{m}\right)$.

Noise Contrastive Estimation

- Noise Contrastive Estimation (Gutmann et Hyvarïnen, 2012).

  Define the following likelihood (logistic classifier):

$$\ell^{\mathrm{NCE}}_{n,m}(\theta,\nu)=\sum_{i=1}^{n}\log q_{\theta,\nu}(y_i)+\sum_{i=1}^{m}\log\left\{1-q_{\theta,\nu}(x_i)\right\}$$

  where $\log\left\{\frac{q_{\theta,\nu}(x)}{1-q_{\theta,\nu}(x)}\right\}=\log\left\{\frac{h_\theta(x)}{h_\psi(x)}\right\}+\nu+\log\left(\frac{n}{m}\right).$

- NCE estimator: $(\hat{\theta}^{\mathrm{NCE}}_{n,m},\hat{\nu}^{\mathrm{NCE}}_{n,m})=\underset{(\theta,\nu)\in\Theta\times\mathbb{R}}{\mathrm{argmax}}\ \ell^{NCE}_{n,m}(\theta,\nu).$

Noise Contrastive Estimation

- Gutmann and Hyvarïnen (2012) proved NCE consistency and asymptotic normality when $m = \tau n \to +\infty$, $\tau > 0$.
  (they assume $X_1, ..., X_m$ IID).

Noise Contrastive Estimation

- Gutmann and Hyvarïnen (2012) proved NCE consistency and asymptotic normality when $m = \tau n \to +\infty$, $\tau > 0$. (they assume $X_1, ..., X_m$ IID).
- They show through simulations that NCE can outperform MC-MLE, especially for small $\tau = m/n$.

Noise Contrastive Estimation

- Two frequentist methods of inference

Noise Contrastive Estimation

- Two frequentist methods of inference
  - MCMC-MLE (Geyer 1994): approximating the likelihood by a **pointwise importance sampling estimate** of the partition function. $\rightarrow \hat{\theta}_{IS}$

Noise Contrastive Estimation

- Two frequentist methods of inference
  - MCMC-MLE (Geyer 1994): approximating the likelihood by a **pointwise importance sampling estimate** of the partition function. $\rightarrow \hat{\theta}_{IS}$
  - NCE (Gutmann and Hyvärinen 2012): Learning parameters from a **logistic classification task** between data-points and "noise". $\rightarrow \hat{\theta}_{NCE}$

Noise Contrastive Estimation

- Two frequentist methods of inference
  - MCMC-MLE (Geyer 1994): approximating the likelihood by a **pointwise importance sampling estimate** of the partition function. $\rightarrow \hat{\theta}_{IS}$
  - NCE (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 $\mathbb{P}_{\psi}$.

Connections between NCE and MC-MLE

- **Observed data:** $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \mathbb{P}_{\theta^*}$

Connections between NCE and MC-MLE

- **Observed data:** $Y_1, ..., Y_n \overset{i.i.d.}{\sim} \mathbb{P}_{\theta^*}$
- **Monte Carlo sample:** $X_1, ..., X_m \sim \mathbb{P}_{\psi},$ (i.i.d. or MCMC)

Connections between NCE and MC-MLE

- **Observed data:** $Y_1, ..., Y_n \overset{i.i.d.}{\sim} \mathbb{P}_{\theta^*}$
- **Monte Carlo sample:** $X_1, ..., X_m \sim \mathbb{P}_{\psi}$, (i.i.d. or MCMC)

- Geyer (1994): n is fixed and $m \to \infty$
  - $\hat{\theta}^{\mathrm{IS}}_{n,m}$ is an "estimator" of the MLE, consistent and asymptotically normal.

Connections between NCE and MC-MLE

- **Observed data:** $Y_1, ..., Y_n \overset{i.i.d.}{\sim} \mathbb{P}_{\theta^*}$
- **Monte Carlo sample:** $X_1, ..., X_m \sim \mathbb{P}_\psi,$ (i.i.d. or MCMC)

- Geyer (1994): n is fixed and $m \to \infty$
  - $\hat{\theta}^{\mathrm{IS}}_{n,m}$ is an "estimator" of the MLE, consistent and asymptotically normal.

- Gutmann, Hyvarïnen (2012): $m = \tau n \to +\infty$, $\tau > 0$
  - $\hat{\theta}^{\mathrm{NCE}}_{n,m}$ is an estimator of θ^*, consistent and asymptotically normal (when $X_1, ..., X_m$ IID).

Connections between NCE and MC-MLE

- **Observed data:** $Y_1, ..., Y_n \overset{i.i.d.}{\sim} \mathbb{P}_{\theta^*}$
- **Monte Carlo sample:** $X_1, ..., X_m \sim \mathbb{P}_{\psi}$, (i.i.d. or MCMC)

- Geyer (1994): n is fixed and $m \to \infty$
  - $\hat{\theta}^{\mathrm{IS}}_{n,m}$ is an "estimator" of the MLE, consistent and asymptotically normal.

- Gutmann, Hyvarïnen (2012): $m = \tau n \to +\infty$, $\tau > 0$
  - $\hat{\theta}^{\mathrm{NCE}}_{n,m}$ 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 $\tau = m/n$.

Summary

Formal comparison, two asymptotic regimes

- Can we do a formal comparison between NCE and MC-MLE ? Two asymptotic regimes...

Formal comparison, two asymptotic regimes

- Can we do a formal comparison between NCE and MC-MLE ? Two asymptotic regimes...
- n is fixed, $m \to +\infty$ (Geyer 1994).
  Observed data is fixed. We study asymptotics of the Monte Carlo error to the MLE.

Formal comparison, two asymptotic regimes

- Can we do a formal comparison between NCE and MC-MLE ? Two asymptotic regimes...
- n is fixed, $m \to +\infty$ (Geyer 1994).
  Observed data is fixed. We study asymptotics of the Monte Carlo error to the MLE.
- $m = \tau n \to +\infty$, $\tau > 0$ (Gutmann, Hyvarïnen, 2012).
  We study the overall inferential error to the true parameter, when both m and n go to infinity.

Monte Carlo error

- Theorem: for any $\varepsilon > 0$, almost surely

$$m^{1-\varepsilon}(\hat{\theta}^{\mathrm{NCE}}_{n,m} - \hat{\theta}^{\mathrm{IS}}_{n,m}) \underset{m\to\infty}{\to} 0.$$

Monte Carlo error

- Theorem: for any $\varepsilon > 0$, almost surely

$$m^{1-\varepsilon}(\hat{\theta}^{\mathrm{NCE}}_{n,m} - \hat{\theta}^{\mathrm{IS}}_{n,m}) \underset{m\to\infty}{\to} 0.$$

- The difference between the two estimators converges faster than the $m^{-1/2}$ rate of convergence to the MLE.
- When n is fixed, MC-MLE and NCE are asymptotically equivalent for approximating the MLE.

Overall error

- Theorem: Under milder assumptions than Gutmann & Hyvarïnen, 2012 (e.g. $(X_j)_{j\geq 1}$ can be a Markov Chain), as $m = \tau n \to +\infty$ we have

$$\sqrt{n}\left(\hat{\theta}^{\mathrm{IS}}_{n,m} - \theta^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0_d, V_\tau^{IS})$$

$$\sqrt{n}\left(\hat{\theta}^{\mathrm{NCE}}_{n,m} - \theta^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0_d, V_\tau^{NCE})$$

Overall error

- Theorem: For every sampling distribution f_ψ and every $\tau > 0$, if $X_1, \ldots, X_m$ are IID, then we have $V_\tau^{IS} \succcurlyeq V_\tau^{NCE}$.

Overall error

- Theorem: For every sampling distribution f_ψ and every $\tau > 0$, if $X_1, \ldots, X_m$ are IID, then we have $V_\tau^{IS} \succcurlyeq V_\tau^{NCE}$.
- Remark: if $f_\psi = f_{\theta^*}$ then $V_\tau^{IS} = V_\tau^{NCE} = (1 + \tau^{-1}) V^{MLE}$ where V^{MLE} is the asymptotic variance of the MLE.

Summary

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

Summary

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

Monte Carlo error	Overall inferential error
n fixed	$m/n \to \tau \in (0, \infty)$

Summary

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

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

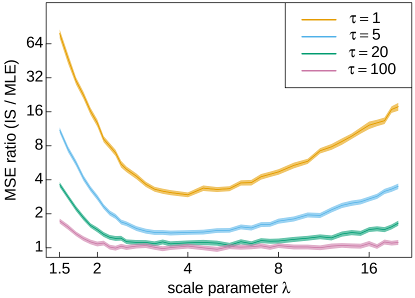
Robust to what?

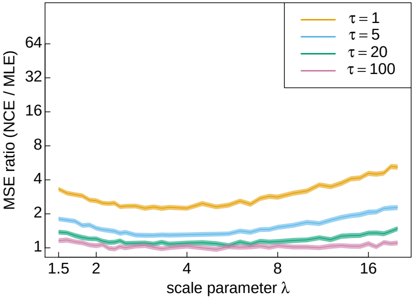
- Numerical example (Gaussian Truncated Model):
  $Y_1, ..., Y_n$ IID from $\mathcal{N}_p(\mu, \Sigma)$ truncated to $]0, +\infty[^p$,
  $X_1, ...X_m$ IID from $\mathcal{N}_p(0, \lambda I_p)$ truncated to $]0, +\infty[^p$.

Robust to what?

- Numerical example (Gaussian Truncated Model):
  $Y_1, \ldots, Y_n$ IID from $\mathcal{N}_p(\mu, \Sigma)$ truncated to $]0, +\infty[^p$,
  $X_1, \ldots X_m$ IID from $\mathcal{N}_p(0, \lambda I_p)$ truncated to $]0, +\infty[^p$.

- Mean Square Error ratios of MC-MLE (left) and NCE (right):





Summary

Summary of contributions

- NCE and MC-MLE are asymptotically equivalent when it comes to approximate the MLE.

Summary of contributions

- NCE and MC-MLE are asymptotically equivalent when it comes to approximate the MLE.
- If m and n are of the same order, NCE is more robust than MC-MLE to a "bad choice" of sampling distribution, especially when m/n is low.

Summary of contributions

- NCE and MC-MLE are asymptotically equivalent when it comes to approximate the MLE.
- If m and n are of the same order, NCE is more robust than MC-MLE to a "bad choice" of sampling distribution, especially when m/n is low.
- We prove that NCE always dominates MC-MLE in terms of asymptotic variance, when $X_1, ..., X_m$ IID. (we believe it is essentially true for MCMC sampling as well)

Perspectives

- extending nce to other inference frameworks.

Perspectives

- extending nce to other inference frameworks.
- non i.i.d. models: to exploit the structure of the dependence.

Perspectives

- extending nce to other inference frameworks.
- non i.i.d. models: to exploit the structure of the dependence.
- methodology on how to select $\mathbb{P}_\psi$.

Perspectives

- extending nce to other inference frameworks.
- non i.i.d. models: to exploit the structure of the dependence.
- methodology on how to select $\mathbb{P}_\psi$.
- developing tools for (un-normalized) model selection.

Summary

Inference for un-normalised statistical models

- Intractable integrals in statistics
- Un-normalised statistical models
- Overview of the literature

Approximate likelihood via Monte Carlo

- Monte Carlo Maximum Likelihood Estimation
- Noise Contrastive estimation
- Connections between NCE and MC-MLE

NCE: a formal comparison with MC-MLE

- Two asymptotic regimes
- Extension of asymptotic guarantees for NCE
- Robustness to a bad choice of sampling distribution

Perspectives

References I

[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.

[Lyn+15] Anne-Marie Lyne et al. "On Russian roulette estimates for Bayesian inference with doubly-intractable likelihoods". In: *Statistical science* 30.4 (2015), pp. 443–467.

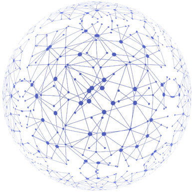
References II

[MGM12] Iain Murray, Zoubin Ghahramani, and David MacKay. "MCMC for doubly-intractable distributions". In: *arXiv preprint arXiv:1206.6848* (2012).

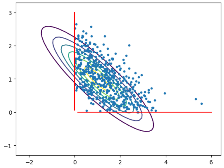
[RC18] Lionel Riou-Durand and Nicolas Chopin. "Noise contrastive estimation: Asymptotic properties, formal comparison with MC-MLE". In: *Electronic Journal of Statistics* 12.2 (2018), pp. 3473–3518.

Pictures

- ERGM: https://pixabay.com/illustrations/network-connections-communication-3537400/



- Truncated Gaussian Model: https://alan-turing-institute.github.io/



Thank you !