Lennard-Jones potential estimation ¹

Jean-François Coeurjolly

Joint work with Frédéric Lavancier





^{1.} based on C-L, Parametric estimation of pairwise Gibbs point processes with infinite range interaction, Bernoulli, 2016



"The Lennard-Jones potential is a mathematically simple model that approximates the interaction between a pair of neutral atoms or molecules. A form of this interatomic potential was first proposed in 1924 by John Lennard-Jones."

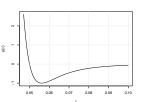




"The Lennard-Jones potential is a mathematically simple model that approximates the interaction between a pair of neutral atoms or molecules. A form of this interatomic potential was first proposed in 1924 by John Lennard-Jones."

1. Spatial point process context: Infinite range Gibbs model with potential

$$\Phi(r) = 4\varepsilon \left\{ (\sigma/r)^{12} - (\sigma/r)^6 \right\}$$

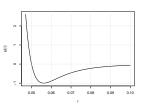




"The Lennard-Jones potential is a mathematically simple model that approximates the interaction between a pair of neutral atoms or molecules. A form of this interatomic potential was first proposed in 1924 by John Lennard-Jones."

1. Spatial point process context: Infinite range Gibbs model with potential

$$\Phi(r) = 4\varepsilon \left\{ (\sigma/r)^{12} - (\sigma/r)^6 \right\}$$



2. Pseudolikelihood (PL), logistic regression likelihood (LRL) are standard methods to estimate paramtetric Gibbs models.

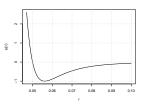




"The Lennard-Jones potential is a mathematically simple model that approximates the interaction between a pair of neutral atoms or molecules. A form of this interatomic potential was first proposed in 1924 by John Lennard-Jones."

1. Spatial point process context: Infinite range Gibbs model with potential

$$\Phi(r) = 4\varepsilon \left\{ (\sigma/r)^{12} - (\sigma/r)^6 \right\}$$



- 2. Pseudolikelihood (PL), logistic regression likelihood (LRL) are standard methods to estimate paramtetric Gibbs models.
- 3. PL and LRL estimates are consistent and asymptotically Gaussian for the Lennard-Jones potential.



"The Lennard-Jones potential is a mathematically simple model that approximates the interaction between a pair of neutral atoms or molecules. A form of this interatomic potential was first proposed in 1924 by John Lennard-Jones."

1. Spatial point process context : Infinite range Gibbs model with potential $\Phi(r) = 4\varepsilon \left\{ (\sigma/r)^{12} - (\sigma/r)^6 \right\}$

- 2. Pseudolikelihood (PL), logistic regression likelihood (LRL) are standard methods to estimate paramtetric Gibbs models.
- 3. PL and LRL estimates are consistent and asymptotically Gaussian for the Lennard-Jones potential.

Thank you for your attention



A few notation

- Let π_{Λ}^{β} , $\Lambda \in \mathbb{R}^d$ (bounded Borel set of \mathbb{R}^d), $\beta > 0$: Poisson point process in Λ with intensity β .
- Let $\Phi: \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be a potential function and $H_{\Lambda}: \Omega_T \to \mathbb{R} \cup \{\infty\}$ be the Hamiltonian given by

$$H_{\Lambda}(\mathbf{x}) = \frac{1}{2} \sum_{\substack{u,v \in \mathbf{x}, u \neq v, \\ \{u,v\} \cap \mathbf{x}_{\Lambda} \neq \emptyset}} \Phi(u-v)$$

 $\Omega = \{ \mathbf{x} \in \Omega_T, \forall \Lambda \in \mathbb{R}^d, H_{\Lambda}(\mathbf{X}) < \infty \}, \Omega_T \text{ set of tempered configurations.}$

Dobrushin-Landford-Ruelle formalism

P is a Gibbs measure if $P(\Omega) = 1$ and for P-a.e. \mathbf{x} and any Λ the cond. law of P given x_{Λ^c} is absolutely continuous w.r.t. π^{β}_{Λ} with density

$$\exp\left(-H_{\Lambda}(\mathbf{x})\right)/Z_{\Lambda}(\mathbf{x}_{\Lambda^c}).$$

Papangelou conditional intensity

Let $\lambda : \mathbb{R}^d \times \Omega \to \mathbb{R}_+$ defined for any $u \in \Lambda$ by

$$\lambda(u, \mathbf{x}) = \beta \frac{e^{-H_{\Lambda}(\mathbf{x} \cup u)}}{e^{-H_{\Lambda}(\mathbf{x})}} = \beta e^{-\sum_{v \in \mathbf{x}} \Phi(v - u)}.$$

= can be viewed as the conditional probability to have a point in a vicinity of u, given that the rest of the configuration is \mathbf{x} .

Papangelou conditional intensity

Let $\lambda: \mathbb{R}^d \times \Omega \to \mathbb{R}_+$ defined for any $u \in \Lambda$ by

$$\lambda(u,\mathbf{x}) = \beta \, \frac{e^{-H_{\Lambda}(\mathbf{x} \cup u)}}{e^{-H_{\Lambda}(\mathbf{x})}} = \beta \, e^{-\sum_{v \in \mathbf{x}} \Phi(v-u)}.$$

= can be viewed as the conditional probability to have a point in a vicinity of u, given that the rest of the configuration is \mathbf{x} .

Statistical model: $\Phi \to \Phi_{\theta}$, $\lambda \to \lambda_{\theta}$, where $\theta \in \mathbb{R}^p$, $p \ge 1$

Exponential family model:

$$\lambda_{\theta}(u, \mathbf{x}) = \beta e^{-\sum_{v \in \mathbf{x}} \Phi_{\theta}(v - u)} = e^{-\theta^{\top} t(u, \mathbf{x})}$$

with $\theta_1 = -\log \beta$ and $t = (t_1, \dots, t_p)^{\mathsf{T}}$ where $t_1(u, \mathbf{x}) = 1$ and

$$t_m(u, \mathbf{x}) = \sum_{v \in \mathbf{x}} g_m(v - u), \quad m = 2, \dots, p.$$

$$\Rightarrow \Phi_{\theta} = \sum_{m=2}^{p} \theta_m g_m.$$



Existence of a Gibbs measure, Ruelle (1969)

If the potential Φ_{θ} is bounded from below and there exist $0 < r_1 < r_2 < \infty, \ c > 0$ and $\gamma_1, \gamma_2 > d$ such that $\Phi_{\theta}(u) \ge c||u||^{-\gamma_1}$ for $||u|| \le r_1$ and $|\Phi_{\theta}(u)| \le c||u||^{-\gamma_2}$ for $||u|| \ge r_2$, then \exists at least one stationary Gibbs measure.

Examples

- Hard-core with finite range potential!!
- $\Phi(u) = ||u||^{-\gamma}$ with $\gamma > d$; $\Phi(u) = e^{-||u||} ||u||^{-\gamma}$ with $\gamma > d$.
- Lennard-Jones type pair potential defined for some $d < \gamma_2 < \gamma_1$ and some A, B > 0 by $\Phi(u) = A||u||^{-\gamma_1} B||u||^{-\gamma_2}$.
- Standard LJ-model= d = 2, $\gamma_1 = 12$ and $\gamma_2 = 6$.

We assume [g] in the following

a collection of technical conditions on the g_m 's which (among other) imply that Φ_{θ} satisfies Ruelle conditions. (stisfied by previous examples)

Domain of observation

X is observed in W_n , where (W_n) is an increasing sequence of convex domains, such that $W_n \to \mathbb{R}^d$ as $n \to \infty$.

PL and LR

Popular alternatives to the ML (no normalizing constant). In particular, the PL estimator is defined as the maximum of

$$\mathsf{LPL}_{W_n}(\mathbf{X}; \theta) = \sum_{u \in \mathbf{X} \cap W_n} \log \lambda_{\theta}(u, \mathbf{X} \setminus u) - \int_{W_n} \lambda_{\theta}(u, \mathbf{X}) \mathrm{d}u.$$

<u>Problem</u>: $\lambda_{\theta}(u, \mathbf{X})$ depends on $\mathbf{X} \cap W_n^c$ which is not observed.

Domain of observation

X is observed in W_n , where (W_n) is an increasing sequence of convex domains, such that $W_n \to \mathbb{R}^d$ as $n \to \infty$.

PL and LR

Popular alternatives to the ML (no normalizing constant). In particular, the PL estimator is defined as the maximum of

$$\mathsf{LPL}_{W_n}(\mathbf{X}; \theta) = \sum_{u \in \mathbf{X} \cap W_n} \log \lambda_{\theta}(u, \mathbf{X} \setminus u) - \int_{W_n} \lambda_{\theta}(u, \mathbf{X}) \mathrm{d}u.$$

<u>Problem</u>: $\lambda_{\theta}(u, \mathbf{X})$ depends on $\mathbf{X} \cap W_n^c$ which is not observed.

Border-correction (Møller and Jensen (91), Jensen and Künsch (94), Billiot et al. (08), C and Drouilhet (10), Baddeley et al. (14))

X has a finite range $\Leftrightarrow \Phi_{\theta}$ is compactly supported in $B(0,R), R < \infty$ $\Leftrightarrow \lambda_{\theta}(u, \mathbf{x}) = \lambda_{\theta}(u, \mathbf{x}_{u,R}), \text{ where } \mathbf{x}_{u,R} = \mathbf{x} \cap B(u,R)$

$$\Rightarrow \mathsf{LPL}_{\textcolor{red}{W_n \ominus R}}(\mathbf{X};\theta) = \sum_{u \in \mathbf{X} \cap \textcolor{red}{W_n \ominus R}} \log \lambda_{\theta}(u,\mathbf{X}_{u,R} \setminus u) - \int_{\textcolor{red}{W_n \ominus R}} \lambda_{\theta}(u,\mathbf{X}_{u,R}) \mathrm{d}u$$

Pseudolikelihood for infinite range models

Let (α_n) and (R_n) be two sequences of real numbers. We define

$$\widetilde{\mathsf{LPL}}_{W_n \ominus \alpha_n, R_n}(\mathbf{X}; \theta) = \sum_{u \in \mathbf{X} \cap W_n \ominus \alpha_n} \log \lambda_{\theta}(u, \mathbf{X}_{u, R_n} \backslash u) - \int_{W_n \ominus \alpha_n} \lambda_{\theta}(u, \mathbf{X}_{u, R_n}) \mathrm{d}u$$

where now
$$\mathbf{X}_{u,R_n} = \mathbf{X} \cap (W_n \ominus \alpha_n) \cap B(u,R_n)$$



Pseudolikelihood for infinite range models

Let (α_n) and (R_n) be two sequences of real numbers. We define

$$\widetilde{\mathsf{LPL}}_{W_n \ominus \alpha_n, R_n}(\mathbf{X}; \theta) = \sum_{u \in \mathbf{X} \cap W_n \ominus \alpha_n} \log \lambda_{\theta}(u, \mathbf{X}_{u, R_n} \setminus u) - \int_{W_n \ominus \alpha_n} \lambda_{\theta}(u, \mathbf{X}_{u, R_n}) \mathrm{d}u$$

where now
$$\mathbf{X}_{u,R_n} = \mathbf{X} \cap (W_n \ominus \alpha_n) \cap B(u,R_n)$$

Interesting choices (further investigated)

- $\alpha_n = R_n \Leftrightarrow$ border correction for finite range interaction models with range R taking $R_n = R$;
- $R_n = \infty \Leftrightarrow$ accounting for the maximal possible range of interaction;
- $R_n = \infty$ and $\alpha_n = 0 \Leftrightarrow$ maximal possible range of interaction and no erosion is considered.

Consistency result

If $\alpha_n |W_n|^{-1/d} \to 0$ and $R_n \to \infty$ as $n \to \infty$ and if for any $y \in \mathbb{R} \setminus \{0\}$, $P\{y^\top t(0, \mathbf{X}) \neq 0\} > 0$, then as $n \to \infty$, $\widehat{\theta} \to \theta^*$ a.s.

Consistency result

If $\alpha_n |W_n|^{-1/d} \to 0$ and $R_n \to \infty$ as $n \to \infty$ and if for any $y \in \mathbb{R} \setminus \{0\}$, $P\{y^\top t(0, \mathbf{X}) \neq 0\} > 0$, then as $n \to \infty$, $\widehat{\theta} \to \theta^*$ a.s.

Asymptotic normality result

Let α_n and R_n be such that $\alpha_n |W_n|^{-1/d} \to 0$, $\alpha_n^{-\gamma'} |W_n|^{1/2} \to 0$ and $R_n^{-\gamma'} |W_n|^{1/2} \to 0$ for some $0 < \gamma' < \gamma_2 - d$. Assume $U_\infty, \Sigma_\infty > 0$ and $\gamma_2 > 2d$, then

$$|W_n|^{1/2} (\widehat{\theta} - \theta^*) \stackrel{d}{\to} \mathcal{N} (0, U_\infty^{-1} \Sigma_\infty U_\infty^{-1})$$

- U_{∞} : asymptotic normalized sensitivity matrix.
- Σ_{∞} : asymptotic covariance matrix of the score function.



Consistency result

If $\alpha_n |W_n|^{-1/d} \to 0$ and $R_n \to \infty$ as $n \to \infty$ and if for any $y \in \mathbb{R} \setminus \{0\}$, $P\{y^\top t(0, \mathbf{X}) \neq 0\} > 0$, then as $n \to \infty$, $\widehat{\theta} \to \theta^*$ a.s.

Asymptotic normality result

Let α_n and R_n be such that $\alpha_n |W_n|^{-1/d} \to 0$, $\alpha_n^{-\gamma'} |W_n|^{1/2} \to 0$ and $R_n^{-\gamma'} |W_n|^{1/2} \to 0$ for some $0 < \gamma' < \gamma_2 - d$. Assume $U_\infty, \Sigma_\infty > 0$ and $\gamma_2 > 2d$, then

$$|W_n|^{1/2} (\widehat{\theta} - \theta^*) \xrightarrow{d} \mathcal{N} (0, U_\infty^{-1} \Sigma_\infty U_\infty^{-1})$$

- U_{∞} : asymptotic normalized sensitivity matrix.
- Σ_{∞} : asymptotic covariance matrix of the score function.

Remarks

- $\gamma_2 > 2d$ is the more important condition (satisfied for LJ).
- for the asymptotic normality result, α_n must tend to ∞ .



A few explanations of the difficulties: consistency

(for ease, assume $\alpha_n = R_n$)

- for finite range models, $LPL_{W_n \ominus R}(\mathbf{X}, \cdot)$ is a concave function and the score function is unbiased.
- for inifinite range models, $\overline{\mathsf{LPL}}_{W_n \cap R_n, R_n}(\mathbf{X}, \cdot)$ can still be shown to be a concave function.
- But its score (at θ^*) writes

• But its score (at
$$\theta^*$$
) writes
$$s_{W_n \ominus R_n}(\mathbf{X}, \theta^*) = \int_{W_n \ominus R_n} t(u, \mathbf{X}_{u, R_n}) \lambda_{\theta^*}(u, \mathbf{X}_{u, R_n}) du - \sum_{u \in \mathbf{X} \cap W_n \ominus R_n} t(u, \mathbf{X}_{u, R_n})$$

$$\Rightarrow \mathrm{E} s_{W_n \ominus R_n}(\mathbf{X}, \theta^\star) \stackrel{GNZ}{=} \int_{W_n \ominus R_n} t(u, \mathbf{X}_{u, R_n}) \big(\lambda_{\theta^\star}(u, \mathbf{X}_{u, R_n}) - \lambda_{\theta^\star}(u, \mathbf{X}) \big) \, \mathrm{d}u \neq 0$$

A few explanations of the difficulties : CLT

- CLT for $\widehat{\theta}$ (mainly) ensues from a CLT for the score function.
- Assume $W_n \ominus R_n = \bigcup_{j \in \mathcal{J}_n} \Delta_j$ where the Δ_j 's are unit cubes centered at $j \in \mathbb{Z}^d$.
- Let $Z_{n,j} = s_{\Delta_j}(\mathbf{X}, \theta^*) \mathrm{E}(s_{\Delta_j}(\mathbf{X}, \theta^*))$

$$s_{W_n \ominus R_n}(\mathbf{X}, \theta^{\star}) = \sum_{j \in \mathcal{J}_n} s_{\Delta_j}(\mathbf{X}, \theta^{\star}) = \sum_{j \in \mathcal{J}_n} Z_{n,j} + \underbrace{\sum_{j \in \mathcal{J}_n} \mathrm{E} s_{\Delta_j}(\mathbf{X}, \theta^{\star})}_{=o_P(|W_n|^{-1/2})}$$

• The $Z_{n,j}$'s are not α -mixing!! For finite-range models, typically, they are conditionally centered, i.e.

$$E(Z_{n,j} | \mathbf{X} \cap \Delta_i^c) = 0$$
 [conditional GNZ]

⇒ CLT for sums of conditionally centered for functionals of Markov random fields.

Theorem (Jensen and Künsch (1994), Comets and Janzura (1998), Guyon and Gaetan (2004), Coeurjolly and Lavancier (2012))

 $X_{n,j}$ be a triangular array field in a measurable space S; For $n \in \mathbb{N}$, let $I_n \subset \mathbb{Z}^d$ and $\alpha_n \in \mathbb{R}_+$ such that $|I_n| \to \infty$ and $\alpha_n \to \infty$ as $n \to \infty$. Define $S_n = \sum_{j \in I_n} Z_{n,j}$ where $Z_{n,j} = f_{n,j}(X_{n,k}, k \in K_j)$ with $\mathcal{K}_j = \{k \in \mathbb{Z}^d, |k-j| \le A\}$ and where $f_{n,j} : S^{\mathcal{K}_{n,j}} \to \mathbb{R}^p$ is a measurable function. We define $\widehat{\Sigma}_n$ and Σ_n by $\widehat{\Sigma}_n = \sum_{j \in I_n} \sum_{\substack{k \in I_n \\ |k-j| \le A}} Z_{n,j} Z_{n,k}^{\top} \qquad \text{and} \qquad \Sigma_n = \mathrm{E}\widehat{\Sigma}_n.$

(a)
$$EZ_{n,j} = 0$$
 and $\sup_{n \ge 1} \sup_{j \in I_n} E||Z_{n,j}||^4 < \infty$.

Then, $|I_n|^{-1}(\widehat{\Sigma}_n - \Sigma_n) \to 0$ in L^2 . If in addition

- (c) $\exists Q > 0$ such that $|I_n|^{-1}\Sigma_n \ge Q$ for n sufficiently large,
- (d) $E(Z_{n,j} | X_{n,k}, k \neq j) = 0$

then

$$\Sigma_n^{-1/2} S_n \xrightarrow{d} \mathcal{N}(0, I_p). \tag{1}$$



Theorem (Coeurjolly and Lavancier (2016))

 $X_{n,j}$ be a triangular array field in a measurable space S; For $n \in \mathbb{N}$, let $I_n \subset \mathbb{Z}^d$ and $\alpha_n \in \mathbb{R}_+$ such that $|I_n| \to \infty$ and $\alpha_n \to \infty$ as $n \to \infty$. Define $S_n = \sum_{j \in I_n} Z_{n,j}$ where $Z_{n,j} = f_{n,j}(X_{n,k}, k \in \mathcal{K}_{n,j})$ with $\mathcal{K}_{n,j} = \{k \in \mathbb{Z}^d, |k-j| \le \alpha_n\}$ and where $f_{n,j} : S^{\mathcal{K}_{n,j}} \to \mathbb{R}^p$ is a measurable function. We define $\widehat{\Sigma}_n$ and Σ_n by $\widehat{\Sigma}_n = \sum_{j \in I_n} \sum_{\substack{k \in I_n \\ |k-j| \le \alpha_n}} Z_{n,j} Z_{n,k}^{\mathsf{T}} \qquad \text{and} \qquad \Sigma_n = \widehat{\mathrm{E}\Sigma}_n.$

- (a) $\mathrm{E} Z_{n,j} = 0$ and there exists $q \geq 1$ such that $\sup_{n \geq 1} \sup_{j \in \mathcal{I}_n} \mathrm{E} \|Z_{n,j}\|^{4q} < \infty$,
- (b) for any sequence $\mathcal{J}_n \subset I_n$ such that $|\mathcal{J}_n| \to \infty$ as $n \to \infty$, $|\mathcal{J}_n|^{-1} \sum_{j,k \in \mathcal{J}_n} \|\mathbb{E}(Z_{n,j} Z_{n,k}^\top)\| = O(1)$.

Then if $a_n^{\frac{4q-1}{2q-1}d} = o(|\mathcal{I}_n|)$ as $n \to \infty$, $|\mathcal{I}_n|^{-1}(\widehat{\Sigma}_n - \Sigma_n) \to 0$ in L^{2q} . If in addition

- (c) $\exists Q > 0$ such that $|I_n|^{-1}\Sigma_n \geq Q$ for n sufficiently large,
- (d) as $n \to \infty$

$$|I_n|^{-1/2} \sum_{j \in I_-} \mathbb{E} \left\| \mathbb{E} \left(Z_{n,j} | X_{n,k}, k \neq j \right) \right\| \to 0,$$

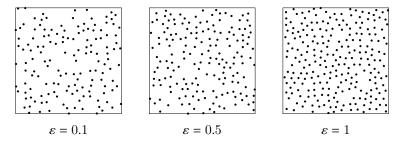
then

$$\Sigma_n^{-1/2} S_n \xrightarrow{d} \mathcal{N}(0, I_n).$$





- LJ model with parameters : $\beta = 100$, $\sigma = 0.1$ and $\varepsilon = 0.1, 0.5, 1$ (low, moderate and high rigidity models).
- $W_n = [-1/2, 1/2]^2, [-1, 1]^2$ and $[-2, 2]^2$.



• RWMSE = $\sqrt{\text{WMSE}}$ = root-weighted empirical mean squared error, where

WMSE =
$$\frac{\widehat{E}\left\{\left(\widehat{\log \beta} - \log \beta\right)^{2}\right\}}{(\log \beta)^{2}} + \frac{\widehat{E}\left\{\left(\widehat{\sigma} - \sigma\right)^{2}\right\}}{\sigma^{2}} + \frac{\widehat{E}\left\{\left(\widehat{\varepsilon} - \varepsilon\right)^{2}\right\}}{\varepsilon^{2}}$$

		RWMSE	
	$[-1/2, 1/2]^2$	$[-1,1]^2$	$[-2,2]^2$
Low $(\varepsilon = 0.1)$			
$\alpha_n = R_n \in [0.05, 0.3]$	3.26(0.13)	1.25 (0.13)	0.62(0.12)
$\alpha_n \in [0.05, 0.3], R_n = \infty$	3.72(0.05)	1.79(0.05)	$0.63 \ (0.06)$
$\alpha_n = 0, R_n = \infty$	3.5	1.66	0.69
Moderate ($\varepsilon = 0.5$)			
$\alpha_n = R_n \in [0.05, 0.3]$	0.65(0.12)	0.34(0.14)	0.2(0.15)
$\alpha_n \in [0.05, 0.3], R_n = \infty$	0.68 (0.05)	$0.38 \ (0.05)$	0.19(0.05)
$\alpha_n = 0, R_n = \infty$	0.59	0.33	0.18
High $(\varepsilon = 1)$			
$\alpha_n = R_n \in [0.05, 0.3]$	1.04(0.08)	0.42(0.16)	0.13(0.16)
$\alpha_n \in [0.05, 0.3], R_n = \infty$	1.34(0.05)	$0.36 \ (0.05)$	0.16 (0.05)
$\alpha_n = 0, R_n = \infty$	1.23	0.27	0.17

- 100 replications
- α_n and/or $R_n \in [0.05, 0.3] \Leftrightarrow 30$ values regularly sampled.
- When an interval is considered, we report the optimal α_n value between brackets.

No further question? Mine: what's the social event?



(Level imposed by Aalborg was amazingly high)



Assumption [g]

For convenience we let $g_1 = 0$ and we denote by g the p-dimensional vector $g = (0, g_2, \dots, g_p)^{\mathsf{T}}$. We make the following assumption on g. [g] For all $m \geq 2$, g_m is bounded from below and there exist $\gamma_1, \gamma_2 > d$ and $c_g, r_0 > 0$ such that

- (i) $\forall ||x|| < r_0 \text{ and } \forall \theta \in \Theta, \ \theta_2 \ g_2(x) \ge c_q ||x||^{-\gamma_1}$
- (ii) $\forall m \ge 3, \ g_m(x) = o(||x||^{-\gamma_1}) \text{ as } ||x|| \to 0$
- (iii) $\forall m \ge 2 \text{ and } \forall ||x|| \ge r_0, |g_m(x)| \le c||x||^{-\gamma_2}.$

