Credibility interval Design for Covid19 Reproduction Number from Nonsmooth Langevin-type Monte Carlo sampling

H. Artigas[1,†], B. Pascal[2], G. Fort[4,†], P. Abry[3,‡], N. Pustelnik[3,5]

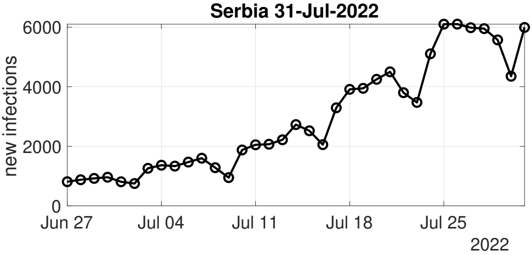
EUSIPCO 2022, Belgrade, Serbia

1. Ecole Polytechnique, Paris, France
2. CNRS, Université de Lille, CRIStAL, France
3. CNRS, ENS de Lyon, Laboratoire de Physique, France,
4. CNRS, Institut de Mathématique de Toulouse, France,
5. UC Louvain, Belgium

† Partly funded by the *Fondation Simone et Cino Del Duca, Institut de France*

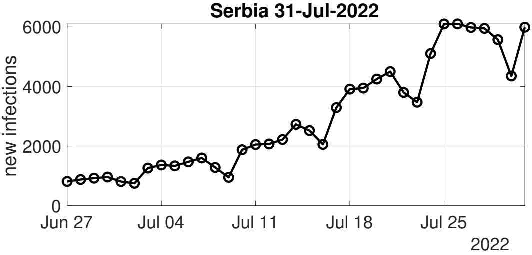
‡ Partially supported by Grant 80PRIME-2021 CNRS

Counts of daily new infections



data from National Health Agencies collected by Johns Hopkins University

Counts of daily new infections



data from National Health Agencies collected by Johns Hopkins University

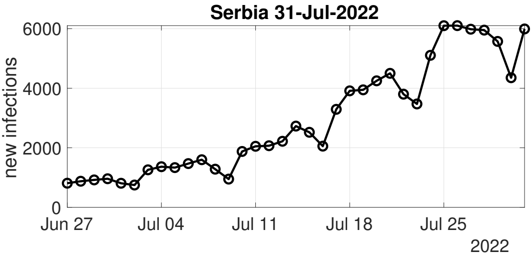
Design adapted counter measures and evaluate their effectiveness

$\longrightarrow$ efficient monitoring tools *epidemiological model,*

$\longrightarrow$ robust to low quality of the data *managing outliers,*

$\longrightarrow$ accompanied by reliable confidence level *credibility intervals.*

Counts of daily new infections



data from National Health Agencies collected by Johns Hopkins University

Design adapted counter measures and evaluate their effectiveness

$\longrightarrow$ efficient monitoring tools *epidemiological model,*

$\longrightarrow$ robust to low quality of the data *managing outliers,*

$\longrightarrow$ accompanied by reliable confidence level *credibility intervals.*

$\Longrightarrow$ number of cases not informative enough: need to capture the **dynamics**

Key concept: the reproduction number R_0 (Liu et al., 2018, *PNAS*)

"averaged number of secondary cases generated by a typical infectious individual"

$\Longrightarrow$ relaxed into **effective reproduction number** R_t at time t

Key concept: the reproduction number R_0 (Liu et al., 2018, *PNAS*)

"averaged number of secondary cases generated by a typical infectious individual"

$\Longrightarrow$ relaxed into **effective reproduction number** R_t at time t

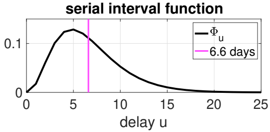
(Cori et al., 2013, *Am. Journal of Epidemiology* ; Abry et al., 2020, *PlosOne* ; Pascal et al., 2022, *Trans. Sig. Process.* ; Fort et al., 2022, *preprint*)

T observations $\mathbf{Z} = (Z_1, \ldots, Z_T)^\top \in \mathbb{N}^T$, Z_t: new infections on day t, $\theta = (\mathbf{R}, \mathbf{O})$

$$\mathbb{P}(Z_t | Z_1, \ldots, Z_{t-1}) = \text{Poiss}\left(p_t(\theta)\right), \quad p_t(\theta) = R_t \left(\Phi Z\right)_t + O_t$$

$(\Phi Z)_t = \sum_{u=1}^{\tau_\phi} \Phi_u Z_{t-u}$, Φ: serial interval function

$\Rightarrow$ random delay between onset of symptoms in primary and secondary cases



Gamma distribution with

- mean 6.6 days
- standard deviation 3.5 days

Log-likelihood from Poisson model (by convention $0 \ln 0 = 0$)

$$f(\theta) := \begin{cases} -\sum_{t=1}^{T} \left(\mathsf{Z}_t \ln p_t(\theta) - p_t(\theta)\right) & \text{if } \theta \in \mathcal{D} = \{\theta \,|\, \forall t,\ p_t(\theta) \geq 0\}, \\ +\infty & \text{otherwise,} \end{cases}$$

Log-likelihood from Poisson model (by convention $0 \ln 0 = 0$)

$$f(\theta) := \begin{cases} -\sum_{t=1}^{T} (\mathsf{Z}_t \ln p_t(\theta) - p_t(\theta)) & \text{if } \theta \in \mathcal{D} = \{\theta \,|\, \forall t,\ p_t(\theta) \geq 0\}, \\ +\infty & \text{otherwise,} \end{cases}$$

A priori distribution of $\theta = (\mathbf{R}, \mathbf{O}) = (\mathsf{R}_1, \ldots, \mathsf{R}_T, \mathsf{O}_1, \ldots, \mathsf{O}_T) \in (\mathbb{R}_+)^T \times \mathbb{R}^T$

- reproduction number: $\mathsf{R}_t - 2\mathsf{R}_{t-1} + \mathsf{R}_{t-2} \sim \text{Laplace}(\lambda_{\mathrm{R}})$
- outliers $\mathsf{O}_t \sim \text{Laplace}(\lambda_{\mathrm{O}})$

$$\Rightarrow g(\theta) = \lambda_{\mathrm{R}} \|\mathsf{D}_2 \mathbf{R}\|_1 + \lambda_{\mathrm{O}} \|\mathbf{O}\|_1, \quad \mathsf{D}_2 = \underbrace{\frac{1}{\sqrt{6}} \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & -2 & 1 & 0 & \ldots & 0 \\ \ldots & & & & & & \ldots \\ 0 & \ldots & & & 1 & -2 & 1 \end{bmatrix}}_{\text{Laplacian}}$$

Bayesian model for the Covid19 reproduction number

Log-likelihood from Poisson model (by convention $0 \ln 0 = 0$)

$$f(\theta) := \begin{cases} -\sum_{t=1}^{T} \left(\mathsf{Z}_t \ln p_t(\theta) - p_t(\theta) \right) & \text{if } \theta \in \mathcal{D} = \{\theta \,|\, \forall t,\ p_t(\theta) \geq 0\}, \\ +\infty & \text{otherwise,} \end{cases}$$

A priori distribution of $\theta = (\mathbf{R}, \mathbf{O}) = (\mathsf{R}_1, \ldots, \mathsf{R}_T, \mathsf{O}_1, \ldots, \mathsf{O}_T) \in (\mathbb{R}_+)^T \times \mathbb{R}^T$

- reproduction number: $\mathsf{R}_t - 2\mathsf{R}_{t-1} + \mathsf{R}_{t-2} \sim \text{Laplace}(\lambda_{\mathrm{R}})$
- outliers $\mathsf{O}_t \sim \text{Laplace}(\lambda_{\mathrm{O}})$

$$\Rightarrow g(\theta) = \lambda_{\mathrm{R}} \|\mathsf{D}_2 \mathbf{R}\|_1 + \lambda_{\mathrm{O}} \|\mathbf{O}\|_1, \quad \mathsf{D}_2 = \underbrace{\frac{1}{\sqrt{6}} \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & -2 & 1 & 0 & \ldots & 0 \\ \ldots & & & & & & \ldots \\ 0 & \ldots & & & 1 & -2 & 1 \end{bmatrix}}_{\text{Laplacian}}$$

A posteriori distribution of unknown parameters $\theta = (\mathbf{R}, \mathbf{O})$

$$\pi(\theta) \propto \exp\left(-f(\theta) - g(\theta)\right) \mathbb{1}_{\mathcal{D}}(\theta)$$

- f, g convex
- f smooth, g nonsmooth

Purpose: sampling the random variable $\theta = (\mathbf{R}, \mathbf{O}) \in \mathbb{R}^{2T}$ according to the posterior[†]

$$\pi(\theta) \propto \exp\left(-f(\theta) - g(\theta)\right) \mathbb{1}_{\mathcal{D}}(\theta)$$

[†] π is defined up to a normalizing constant

Markov Chain Monte Carlo sampling

Purpose: sampling the random variable $\theta = (\mathbf{R}, \mathbf{O}) \in \mathbb{R}^{2T}$ according to the posterior[†]

$$\pi(\theta) \propto \exp\left(-f(\theta) - g(\theta)\right) \mathbb{1}_{\mathcal{D}}(\theta)$$

Principle: 1) generate a random sequence $\{\theta^n,\, n \in \mathbb{N}\}$ such that

- θ^{n+1} only depends on θ^n,
- at convergence, i.e., as $n \to \infty$, $\theta^n \sim \pi$,

2) compute Bayesian estimators, e.g., credibility intervals, on samples $\{\theta^n,\, n \geq N\}$

[†] π is defined up to a normalizing constant

Markov Chain Monte Carlo sampling

Purpose: sampling the random variable $\theta = (\mathbf{R}, \mathbf{O}) \in \mathbb{R}^{2T}$ according to the posterior†

$$\pi(\theta) \propto \exp\left(-f(\theta) - g(\theta)\right) \mathbb{1}_{\mathcal{D}}(\theta)$$

Principle: 1) generate a random sequence $\{\theta^n,\ n \in \mathbb{N}\}$ such that

- θ^{n+1} only depends on θ^n,
- at convergence, i.e., as $n \to \infty$, $\theta^n \sim \pi$,

2) compute Bayesian estimators, e.g., credibility intervals, on samples $\{\theta^n,\ n \geq N\}$

State-of-the-art: *Hastings-Metropolis random walk*

(i) propose a random move according to

$$\theta^{n+\frac{1}{2}} = \theta^n + \sqrt{2\gamma}\Gamma\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_{2T}(0, \mathrm{I})$$

with γ positive step size, $\Gamma \in \mathbb{R}^{2T \times 2T}$

† π is defined up to a normalizing constant

Markov Chain Monte Carlo sampling

Purpose: sampling the random variable $\theta = (\mathbf{R}, \mathbf{O}) \in \mathbb{R}^{2T}$ according to the posterior[†]

$$\pi(\theta) \propto \exp\left(-f(\theta) - g(\theta)\right) \mathbb{1}_{\mathcal{D}}(\theta)$$

Principle: 1) generate a random sequence $\{\theta^n, \, n \in \mathbb{N}\}$ such that

- θ^{n+1} only depends on θ^n,
- at convergence, i.e., as $n \to \infty$, $\theta^n \sim \pi$,

2) compute Bayesian estimators, e.g., credibility intervals, on samples $\{\theta^n, \, n \geq N\}$

State-of-the-art: *Hastings-Metropolis random walk*

(i) propose a random move according to

$$\theta^{n+\frac{1}{2}} = \theta^n + \sqrt{2\gamma}\Gamma\xi^{n+1}, \quad \xi^{n+1} \sim \mathcal{N}_{2T}(0, \mathrm{I})$$

with γ positive step size, $\Gamma \in \mathbb{R}^{2T \times 2T}$

(ii) accept: $\theta^{n+1} = \theta^{n+\frac{1}{2}}$, with probability $1 \wedge \dfrac{\pi(\theta^{n+\frac{1}{2}})}{\pi(\theta^n)}$, or reject: $\theta^{n+1} = \theta^n$

† π is defined up to a normalizing constant

Metropolis Adjusted Langevin Algorithm (MALA)

Langevin dynamics: $\theta^{n+\frac{1}{2}} = \mu(\theta^n) + \sqrt{2\gamma}\xi^{n+1}$, (Kent, 1978, *Adv Appl Probab*)

$$\mu(\theta) \text{ adapted to } \pi(\theta) = \exp(-f(\theta) - g(\theta))\mathbb{1}_{\mathcal{D}}(\theta)$$

Metropolis Adjusted Langevin Algorithm (MALA)

Langevin dynamics: $\theta^{n+\frac{1}{2}} = \mu(\theta^n) + \sqrt{2\gamma}\xi^{n+1}$, (Kent, 1978, *Adv Appl Probab*)

$$\mu(\theta) \text{ adapted to } \pi(\theta) = \exp(-f(\theta) - g(\theta))\mathbb{1}_{\mathcal{D}}(\theta)$$

Case 1: $g = 0$ and $-\ln \pi = f$ is smooth (Roberts & Tweedie, 1996, *Bernoulli*)

$$\mu(\theta) = \theta - \gamma\Gamma\Gamma^\top \nabla f(\theta) = \theta + \gamma\Gamma\Gamma^\top \nabla \ln \pi(\theta)$$

$\Longrightarrow$ move towards areas of higher probability

Metropolis Adjusted Langevin Algorithm (MALA)

Langevin dynamics: $\theta^{n+\frac{1}{2}} = \mu(\theta^n) + \sqrt{2\gamma}\xi^{n+1}$, (Kent, 1978, *Adv Appl Probab*)

$\mu(\theta)$ adapted to $\pi(\theta) = \exp(-f(\theta) - g(\theta))\mathbb{1}_{\mathcal{D}}(\theta)$

Case 1: $g = 0$ and $-\ln \pi = f$ is smooth (Roberts & Tweedie, 1996, *Bernoulli*)

$$\mu(\theta) = \theta - \gamma\Gamma\Gamma^\top \nabla f(\theta) = \theta + \gamma\Gamma\Gamma^\top \nabla \ln \pi(\theta)$$

$\Longrightarrow$ move towards areas of higher probability

Case 2: $-\ln \pi = f + g$ is nonsmooth

$$\mu(\theta) = \text{prox}_{\gamma g}^{\Gamma\Gamma^\top}\left(\theta - \gamma\Gamma\Gamma^\top \nabla f(\theta)\right)$$

combining *Langevin* and *proximal*[†] approaches

[†] $\text{prox}_{\gamma g}^{\Gamma\Gamma^\top}(y) = \underset{x \in \mathbb{R}^d}{\text{argmin}} \left(\frac{1}{2}\|x - y\|^2_{\Gamma\Gamma^\top} + \gamma g(x)\right)$: preconditioned proximity operator of g

Proximal-Gradient dual sampler PGdual

Posterior density of $\theta = (\mathbf{R}, \mathbf{O})$: $\pi(\theta) \propto \exp\left(-f(\theta) - g(\theta)\right) \mathbb{1}_{\mathcal{D}}(\theta)$

- smooth negative log-likelihood

$$\text{if } \theta \in \mathcal{D}, \quad f(\theta) = -\sum_{t=1}^{T} \left(\mathrm{Z}_t \ln p_t(\theta) - p_t(\theta)\right), \quad p_t(\theta) = \mathrm{R}_t(\Phi \mathrm{Z})_t + \mathrm{O}_t$$

- nonsmooth convex lower-semicontinuous negative a priori log-distribution

$$g(\theta) = \lambda_{\mathrm{R}} \|\mathrm{D}_2 \mathbf{R}\|_1 + \lambda_{\mathrm{O}} \|\mathbf{O}\| = h(\mathrm{A}\theta)$$

$\mathrm{A} : \theta \mapsto (\mathrm{D}_2\mathbf{R}, \mathbf{O})$ linear operator, $h(\cdot_1, \cdot_2) = \lambda_{\mathrm{R}} \|\cdot_1\|_1 + \lambda_{\mathrm{O}} \|\cdot_2\|_1$

Proximal-Gradient dual sampler PGdual

Posterior density of $\theta = (\mathbf{R}, \mathbf{O})$: $\quad \pi(\theta) \propto \exp\left(-f(\theta) - g(\theta)\right) \mathbb{1}_{\mathcal{D}}(\theta)$

- smooth negative log-likelihood

$$\text{if } \theta \in \mathcal{D}, \quad f(\theta) = -\sum_{t=1}^{T} \left(\mathsf{Z}_t \ln p_t(\theta) - p_t(\theta)\right), \quad p_t(\theta) = \mathsf{R}_t(\Phi\mathsf{Z})_t + \mathsf{O}_t$$

- nonsmooth convex lower-semicontinuous negative a priori log-distribution

$$g(\theta) = \lambda_{\mathrm{R}} \|\mathsf{D}_2 \mathbf{R}\|_1 + \lambda_{\mathrm{O}} \|\mathbf{O}\| = h(\mathsf{A}\theta)$$

$\mathsf{A} : \theta \mapsto (\mathsf{D}_2 \mathbf{R}, \mathbf{O})$ linear operator, $h(\cdot_1, \cdot_2) = \lambda_{\mathrm{R}} \|\cdot_1\|_1 + \lambda_{\mathrm{O}} \|\cdot_2\|_1$

Case 3: $-\ln \pi = f + h(\mathsf{A}\cdot)$ (Fort et al., 2022, *preprint*)

closed-form expression of $\mathrm{prox}_{\gamma h}$ but not of $\mathrm{prox}_{\gamma h(\mathsf{A}\cdot)}$

1) extend A into **invertible** $\overline{\mathsf{A}}$, and h in $\bar{h}$ such that $\bar{h}(\overline{\mathsf{A}}\theta) = h(\mathsf{A}\theta)$

2) reason on the **dual** variable $\tilde{\theta} = \overline{A}\theta$

Proximal-Gradient dual sampler PGdual

Langevin: drift toward higher probability regions

$$\underset{\theta \in \mathbb{R}^{2T}}{\operatorname{argmax}} \ln \pi(\theta) = \underset{\theta \in \mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\theta) + \bar{h}(\overline{\mathrm{A}}\theta) = \mathrm{A}^{-1} \underset{\tilde{\theta} \in \mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\overline{A}^{-1}\tilde{\theta}) + \bar{h}(\tilde{\theta})$$

Proximal-Gradient dual sampler PGdual

Langevin: drift toward higher probability regions

$$\underset{\theta\in\mathbb{R}^{2T}}{\operatorname{argmax}} \ln \pi(\theta) = \underset{\theta\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\theta) + \bar{h}(\overline{\mathsf{A}}\theta) = \mathsf{A}^{-1} \underset{\tilde{\theta}\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\overline{A}^{-1}\tilde{\theta}) + \bar{h}(\tilde{\theta})$$

$$\Longrightarrow \quad \mu(\theta) = \underbrace{\overline{\mathsf{A}}^{-1}}_{\text{back to } \theta}\ \underbrace{\operatorname{prox}_{\gamma\bar{h}}\left(\overline{\mathsf{A}}\theta - \gamma\overline{A}^{-\top}\nabla f(\theta)\right)}_{\text{proximal-gradient on } \tilde{\theta}}$$

Langevin: drift toward higher probability regions

$$\underset{\theta\in\mathbb{R}^{2T}}{\operatorname{argmax}} \ln \pi(\theta) = \underset{\theta\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\theta) + \bar{h}(\overline{\mathsf{A}}\theta) = \mathsf{A}^{-1} \underset{\tilde{\theta}\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\overline{A}^{-1}\tilde{\theta}) + \bar{h}(\tilde{\theta})$$

$$\implies \quad \mu(\theta) = \underbrace{\overline{\mathsf{A}}^{-1}}_{\text{back to } \theta}\ \underbrace{\operatorname{prox}_{\gamma\bar{h}}\left(\overline{\mathsf{A}}\theta - \gamma\overline{A}^{-\top}\nabla f(\theta)\right)}_{\text{proximal-gradient on } \tilde{\theta}}$$

Two strategies to extend $\mathsf{A} = \begin{pmatrix} \mathsf{D}_2 & 0 \\ 0 & \mathsf{I} \end{pmatrix} \in \mathbb{R}^{(2T-1)\times 2T}$ into $\overline{\mathsf{A}} = \begin{pmatrix} \overline{\mathsf{D}} & 0 \\ 0 & \mathsf{I} \end{pmatrix} \in \mathbb{R}^{2T\times 2T}$:

Proximal-Gradient dual sampler PGdual

Langevin: drift toward higher probability regions

$$\underset{\theta\in\mathbb{R}^{2T}}{\operatorname{argmax}} \ln \pi(\theta) = \underset{\theta\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\theta) + \bar{h}(\overline{\mathsf{A}}\theta) = \mathsf{A}^{-1} \underset{\tilde{\theta}\in\mathbb{R}^{2T}}{\operatorname{argmin}}\ f(\overline{A}^{-1}\tilde{\theta}) + \bar{h}(\tilde{\theta})$$

$$\Longrightarrow \quad \mu(\theta) = \underbrace{\overline{\mathsf{A}}^{-1}}_{\text{back to } \theta} \ \underbrace{\operatorname{prox}_{\gamma\bar{h}}\left(\overline{\mathsf{A}}\theta - \gamma\overline{A}^{-\top}\nabla f(\theta)\right)}_{\text{proximal-gradient on } \tilde{\theta}}$$

Two strategies to extend $\mathsf{A} = \begin{pmatrix} \mathsf{D}_2 & 0 \\ 0 & \mathsf{I} \end{pmatrix} \in \mathbb{R}^{(2T-1)\times 2T}$ into $\overline{\mathsf{A}} = \begin{pmatrix} \overline{\mathsf{D}} & 0 \\ 0 & \mathsf{I} \end{pmatrix} \in \mathbb{R}^{2T\times 2T}$:

Invert

$$\overline{\mathsf{D}}_2 := \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -2/\sqrt{5} & 1/\sqrt{5} & 0 & \cdots & 0 \\ & \mathsf{D}_2 & & & \end{bmatrix}$$

Langevin: drift toward higher probability regions

$$\operatorname*{argmax}_{\theta\in\mathbb{R}^{2T}} \ln\pi(\theta) = \operatorname*{argmin}_{\theta\in\mathbb{R}^{2T}} f(\theta) + \bar{h}(\overline{\mathrm{A}}\theta) = \mathrm{A}^{-1}\operatorname*{argmin}_{\tilde{\theta}\in\mathbb{R}^{2T}} f(\overline{A}^{-1}\tilde{\theta}) + \bar{h}(\tilde{\theta})$$

$$\Longrightarrow \quad \mu(\theta) = \underbrace{\overline{\mathrm{A}}^{-1}}_{\text{back to } \theta} \underbrace{\operatorname{prox}_{\gamma\bar{h}}\left(\overline{\mathrm{A}}\theta - \gamma\overline{A}^{-\top}\nabla f(\theta)\right)}_{\text{proximal-gradient on } \tilde{\theta}}$$

Two strategies to extend $\mathrm{A} = \begin{pmatrix} \mathrm{D}_2 & 0 \\ 0 & \mathrm{I} \end{pmatrix} \in \mathbb{R}^{(2T-1)\times 2T}$ into $\overline{\mathrm{A}} = \begin{pmatrix} \overline{\mathrm{D}} & 0 \\ 0 & \mathrm{I} \end{pmatrix} \in \mathbb{R}^{2T\times 2T}$:

`Invert`

$$\overline{\mathrm{D}}_2 := \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -2/\sqrt{5} & 1/\sqrt{5} & 0 & \cdots & 0 \\ & \mathrm{D}_2 & & & \end{bmatrix}$$

`Ortho`

$$\overline{\mathrm{D}}_o := \begin{bmatrix} v_1^\top \\ v_2^\top \\ \mathrm{D}_2 \end{bmatrix} \qquad \begin{array}{c} v_1, v_2 \in \mathbb{R}^{2T} \\ v_1 \perp v_2,\ v_1, v_2 \in (\mathrm{D}_2^\top)^\perp \end{array}$$

Proximal-Gradient dual sampler `PGdual`

Langevin: drift toward higher probability regions

$$\underset{\theta\in\mathbb{R}^{2T}}{\text{argmax}} \ln \pi(\theta) = \underset{\theta\in\mathbb{R}^{2T}}{\text{argmin}}\ f(\theta) + \bar{h}(\overline{\mathsf{A}}\theta) = \mathsf{A}^{-1} \underset{\tilde{\theta}\in\mathbb{R}^{2T}}{\text{argmin}}\ f(\overline{A}^{-1}\tilde{\theta}) + \bar{h}(\tilde{\theta})$$

$$\Longrightarrow \quad \mu(\theta) = \underbrace{\overline{\mathsf{A}}^{-1}}_{\text{back to } \theta} \underbrace{\text{prox}_{\gamma\bar{h}}\left(\overline{\mathsf{A}}\theta - \gamma \overline{A}^{-\top} \nabla f(\theta)\right)}_{\text{proximal-gradient on } \tilde{\theta}}$$

Two strategies to extend $\mathsf{A} = \begin{pmatrix} \mathsf{D}_2 & 0 \\ 0 & \mathsf{I} \end{pmatrix} \in \mathbb{R}^{(2T-1)\times 2T}$ into $\overline{\mathsf{A}} = \begin{pmatrix} \overline{\mathsf{D}} & 0 \\ 0 & \mathsf{I} \end{pmatrix} \in \mathbb{R}^{2T\times 2T}$:

`Invert`

$$\overline{\mathsf{D}}_2 := \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -2/\sqrt{5} & 1/\sqrt{5} & 0 & \cdots & 0 \\ & \mathsf{D}_2 & & & \end{bmatrix}$$

`Ortho`

$$\overline{\mathsf{D}}_o := \begin{bmatrix} v_1^\top \\ v_2^\top \\ \mathsf{D}_2 \end{bmatrix} \qquad \begin{matrix} v_1, v_2 \in \mathbb{R}^{2T} \\ v_1 \perp v_2,\ v_1, v_2 \in (\mathsf{D}_2^\top)^\perp \end{matrix}$$

Proposed `PGdual` **drift terms** on $\theta = (\mathbf{R}, \mathbf{O})$:

reproduction numbers $\mu_{\mathrm{R}}(\theta) = \overline{\mathsf{D}}^{-1} \operatorname{prox}_{\gamma_{\mathrm{R}}\lambda_{\mathrm{R}}\|(\cdot)_{3:T}\|_1}\left(\overline{\mathsf{D}}\,\mathbf{R} - \gamma_{\mathrm{R}}\overline{\mathsf{D}}^{-\top}\nabla_{\mathrm{R}} f(\theta)\right)$

outliers $\mu_{\mathrm{O}}(\theta) = \operatorname{prox}_{\gamma_{\mathrm{O}}\lambda_{\mathrm{O}}\|\cdot\|_1}\left(\mathbf{O} - \gamma_{\mathrm{O}}\nabla_{\mathrm{O}} f(\theta)\right)$

Algorithm 1: Proximal-Gradient dual: `PGdual Invert` and `PGdual Ortho`

Data: $\overline{\mathsf{D}} = \overline{\mathsf{D}}_2$ (`Invert`) or $\overline{\mathsf{D}} = \overline{\mathsf{D}}_o$ (`Ortho`)

$\gamma_{\mathsf{R}}, \gamma_{\mathsf{O}} > 0$, $N_{\max} \in \mathbb{N}_\star$, $\theta^0 = (\mathbf{R}^0, \mathbf{O}^0) \in \mathcal{D}$

Result: A $\mathcal{D}$-valued sequence $\{\theta^n = (\mathbf{R}^n, \mathbf{O}^n), n \in 0, \ldots, N_{\max}\}$

for $n = 0, \ldots, N_{\max} - 1$ **do**

Sample $\xi_{\mathsf{R}}^{n+1} \sim \mathcal{N}_T(0, \mathsf{I})$ and $\xi_{\mathsf{O}}^{n+1} \sim \mathcal{N}_T(0, \mathsf{I})$;

Set $\mathbf{R}^{n+\frac{1}{2}} = \mu_{\mathsf{R}}(\theta^n) + \sqrt{2\gamma_{\mathsf{R}}\overline{\mathsf{D}}^{-1}\overline{\mathsf{D}}^{-\top}}\, \xi_{\mathsf{R}}^{n+1}$;

$\mathbf{O}^{n+\frac{1}{2}} = \mu_{\mathsf{O}}(\theta^n) + \sqrt{2\gamma_{\mathsf{O}}}\, \xi_{\mathsf{O}}^{n+1}$;

$\theta^{n+\frac{1}{2}} = (\mathbf{R}^{n+\frac{1}{2}}, \mathbf{O}^{n+\frac{1}{2}})$;

Set $\theta^{n+1} = \theta^{n+\frac{1}{2}}$ with probability

$$1 \wedge \frac{\pi(\theta^{n+\frac{1}{2}})}{\pi(\theta^n)} \frac{q_{\mathsf{R}}(\theta^{n+\frac{1}{2}}, \theta_{\mathsf{R}}^n)}{q_{\mathsf{R}}(\theta^n, \theta_{\mathsf{R}}^{n+\frac{1}{2}})} \frac{q_{\mathsf{O}}(\theta^{n+\frac{1}{2}}, \theta_{\mathsf{O}}^n)}{q_{\mathsf{O}}(\theta^n, \theta_{\mathsf{O}}^{n+\frac{1}{2}})},$$

$q_{\mathsf{R}/\mathsf{O}}$: Gaussian kernel stemming from nonsymmetric proposal

and $\theta^{n+1} = \theta^n$ otherwise.

Comparison of MCMC sampling schemes

Gaussian proposal: $\theta^{n+\frac{1}{2}} = \mu(\theta^n) + \sqrt{2\gamma}\Gamma\xi^{n+1}$

- random walks: $\mu(\theta) = \theta$

  `RW`: $\Gamma = \mathrm{I}$; `RW Invert`: $\Gamma = \overline{\mathrm{D}}_2^{-1}\overline{\mathrm{D}}_2^{-\top}$; `RW Ortho`: $\Gamma = \overline{\mathrm{D}}_o^{-1}\overline{\mathrm{D}}_o^{-\top}$

- Proximal-Gradient dual: $\mu_{\mathrm{R}}(\theta)$, $\mu_{\mathrm{O}}(\theta)$, $\Gamma = \overline{\mathrm{D}}^{-1}\overline{\mathrm{D}}^{-\top}$

  `PGdual Invert`: $\overline{\mathrm{D}} = \overline{\mathrm{D}}_2$; `PGdual Ortho`: $\overline{\mathrm{D}} = \overline{\mathrm{D}}_o$

Practical settings: $N_{\mathrm{max}} = 10^7$ iterations, 15 independent runs

Gaussian proposal: $\theta^{n+\frac{1}{2}} = \mu(\theta^n) + \sqrt{2\gamma}\Gamma\xi^{n+1}$

- random walks: $\mu(\theta) = \theta$

  RW: $\Gamma = \mathrm{I}$; RW Invert: $\Gamma = \overline{\mathrm{D}}_2^{-1}\overline{\mathrm{D}}_2^{-\top}$; RW Ortho: $\Gamma = \overline{\mathrm{D}}_o^{-1}\overline{\mathrm{D}}_o^{-\top}$

- Proximal-Gradient dual: $\mu_{\mathrm{R}}(\theta)$, $\mu_{\mathrm{O}}(\theta)$, $\Gamma = \overline{\mathrm{D}}^{-1}\overline{\mathrm{D}}^{-\top}$

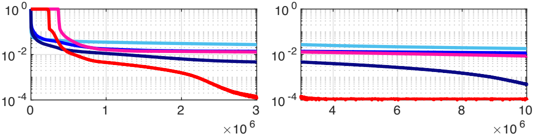
  PGdual Invert: $\overline{\mathrm{D}} = \overline{\mathrm{D}}_2$; PGdual Ortho: $\overline{\mathrm{D}} = \overline{\mathrm{D}}_o$

Practical settings: $N_{\max} = 10^7$ iterations, 15 independent runs

log-density

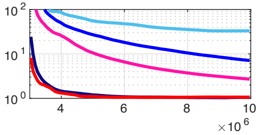
burn in

post burn in



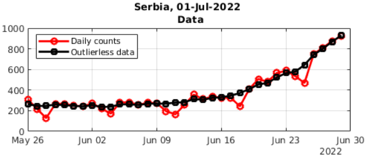
$\left(\ln \pi(\theta^n) - \max \ln \pi\right) / \left(\ln \pi(\theta^0) - \max \ln \pi\right)$

Gelman-Rubin

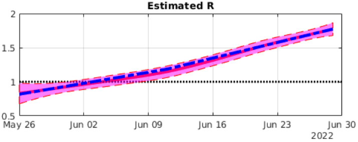


ANOVA-type criterion

Outlierless data: $\mathbf{Z}^{(D)} = \mathbf{Z} - \mathbf{O}$



95% credibility intervals: empirical quantiles



$\Longrightarrow$ significant step toward actual use of the estimate R_T

Denoised counts and reproduction number estimates for several countries available at

`https://perso.ens-lyon.fr/patrice.abry/`

`https://perso.math.univ-toulouse.fr/gfort/`

Further investigations and perspectives:

- comparisons with other MCMC strategies, such as Gibbs samplers (Fort et al., 2022, *preprint*)
- automated data-driven tuning of hyperparameters ($\gamma_{\mathrm{R/O}}, \lambda_{\mathrm{R}}, \lambda_{\mathrm{O}}, \dots$)