Estimation et intervalles de crédibilité pour le taux de reproduction de la Covid19 paréchantillonnage Monte Carlo Langevin proximal

Patrice Abry, Gersende Fort, Barbara Pascal, Nelly Pustelnik

Estimation et intervalles de crédibilité pour le taux de reproduction de la Covid19 par échantillonnage Monte Carlo Langevin proximal

Patrice ABRY[1], Gersende FORT[2], Barbara PASCAL[3], Nelly PUSTELNIK[1,4]

[1]CNRS, ENS de Lyon, Laboratoire de Physique, Lyon, France (firstname.lastname@ens-lyon.fr)

[2]CNRS, Institut de Mathématiques de Toulouse, France (gersende.fort@math.univ-toulouse.fr)

[3]Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France (barbara.pascal@univ-lille.fr)

[4]ISPGroup/ICTEAM, UCLouvain, Belgium

Travail soutenu par la *Fondation Simone et Cino Del Duca, Institut de France.*

Résumé – Surveiller l'évolution temporelle de la Covid19 en temps réel et ce, malgré la qualité limitée des données disponibles, est un problème crucial et difficile. L'objectif de ce travail est de comparer six stratégies différentes d'échantillonneurs Monte Carlo de type Langevin proximal utilisées pour l'estimation par intervalles de crédibilité du taux de reproduction de la Covid19, et du compte débruité de nouvelles infections quotidiennes. La difficulté résulte de la formulation bayésienne utilisée qui, pour produire des estimées robustes, fait usage d'une loi a posteriori non régulière. L'application à des données réelles de la Covid19, de plusieurs pays, montre la pertinence des stratégies proposées dites *duales*.

Abstract – Monitoring the time evolution of the intensity of Covid19 pandemic within the pandemic and despite the limited quality of the available Covid19 data is both crucial and challenging. The present work compares six different Langevin proximal Monte Carlo samplers aiming to perform a credibility interval-based estimation of the pandemic Covid19 reproduction number and of the denoised daily new infection counts. The challenge stems from the Bayesian model which, to ensure robust estimates, makes uses of a non differentiable a posteriori distribution. Application to real Covid19 data from several countries shows the relevance of the so-named *dual* strategies proposed here.

1 Introduction.

Contexte. La surveillance continue de l'évolution temporelle de l'intensité de la pandémie de Covid19 constitue une tâche cruciale et difficile : cruciale car elle est un pré-requis pour la conception des mesures de politique sanitaire ; difficile, car elle doit être réalisée pendant que la pandémie se développe (et non rétrospectivement, après que la pandémie est terminée) et à partir de données rendues accessibles par les autorités nationales de santé publique qui restent de qualité très limitée, corrompues par des valeurs manquantes ou aberrantes ou par des effets pseudo-saisonniers. En phase de pandémie, les épidémiologistes ont souvent recours, pour mesurer l'intensité de la pandémie, à la notion de taux de reproduction dépendant du temps R_t, [1]. Des travaux conduits pendant la pandémie, [2, 3], ont permis de proposer une estimation robuste et efficace de ce R_t à partir d'une formulation de type *minimisation convexe non lisse*. En complément, et plus récemment, une relecture *bayésienne* de ce modèle a permis d'obtenir une estimation par intervalles de crédibilté de ce R_t via des stratégies d'échantillonnage Monte Carlo [4, 5]. Cependant, la conception de tels *échantillonneurs* dans le contexte de données de Covid19 de qualité très limitée s'avère délicate du fait des propriétés de la distribution *a posteriori* π qu'il faut reproduire. En effet, celle-ci prend la forme

$$\pi(\boldsymbol{\theta}) \propto \exp(-(f(\boldsymbol{\theta}) + g(\mathsf{A}\boldsymbol{\theta}))) \mathbb{1}_{\mathcal{D}}(\boldsymbol{\theta}); \qquad (1)$$

où $\boldsymbol{\theta}$ désigne les paramètres à estimer : la fonction f est différentiable, mais g est convexe non différentiable, $\mathbb{1}_{\mathcal{D}}$ est la fonction indicatrice de l'ensemble $\mathcal{D}$, à valeur dans $\{0, 1\}$. De plus, π n'est pas nécessairement log-concave ; la composition de g et de l'opérateur linéaire A rend non explicite le calcul de l'opérateur proximal de $g(\mathsf{A}\cdot)$ même si celui de g l'est ; les contraintes sur $\boldsymbol{\theta}$ exprimées par $\mathcal{D}$ doivent être satisfaites.

État-de-l'art. Ce travail est inscrit dans le contexte des échantillonneurs Monte Carlo de type Hastings-Metropolis (HM) - voir par exemple [6] - utilisant un mécanisme de proposition gaussien qui s'affranchit de la contrainte $\mathcal{D}$: étant donné le point courant $\boldsymbol{\theta}^n$ de la chaîne, un saut vers $\boldsymbol{\theta}^{n+1/2} := \boldsymbol{\mu}(\boldsymbol{\theta}^n) + \xi^{n+1}$, $\xi^{n+1} \sim \mathcal{N}(0, \mathsf{C})$ est proposé. La nouvelle valeur $\boldsymbol{\theta}^{n+1}$ est définie par une étape d'acceptation-rejet, laquelle, par définition de π n'accepte que des points dans $\mathcal{D}$. Lorsque $\ln \pi$ est régulière sur $\mathcal{D}$, une procédure standard pour estimer les statistiques de π consiste à mettre en œuvre une dynamique de Langevin, $\boldsymbol{\mu}(\boldsymbol{\theta}) := \boldsymbol{\theta} + \gamma \nabla \ln \pi(\boldsymbol{\theta})$, avec $\gamma > 0$ [7] : les déplacements proposés se dirigent vers des zones de forte probabilité pour π, en utilisant les informations du premier ordre sur π. Lorsque π n'est pas régulière, comme pour (1), les solutions proposées dans la littérature, que nous désignerons par *méthodes de Langevin proximal*, exploitent le gradient de f et l'opérateur proximal de $\rho\, g$, noté $\mathrm{prox}_{\rho g}$, pour $\rho > 0$ [8]. L'originalité des algorithmes HM considérés ici réside dans la dérive $\boldsymbol{\mu}$, choisie

pour exploiter les éléments f, g et A qui caractérisent la densité cible π. Les dérives $\boldsymbol{\mu}$ proposées dans [9] et [10] reposent sur une approximation de $g(\mathsf{A}\cdot)$ par son enveloppe de Moreau, qui conduit à une approximation régulière de $f + g(\mathsf{A}\cdot)$ dont le gradient fait intervenir l'opérateur proximal de g ; $\boldsymbol{\mu}$ est alors *la somme* d'un terme de gradient relatif à f et d'un terme proximal associé à g. Les dérives proposées par [4, 5] définissent $\boldsymbol{\mu}$ comme *la composition* d'un terme de gradient relatif à f et d'un terme proximal en lien avec g.

Objectifs et contributions. L'objectif de ce travail est de structurer les relations entre méthodes de Langevin proximal et comparer leur performances, à la fois sur un problème-jouet pédagogique et sur l'estimation de l'intensité de la pandémie de Covid19. La première contribution est, après avoir rappelé les différentes définitions de $\boldsymbol{\mu}$, de montrer que ces algorithmes peuvent être regroupés en deux familles, nommées *primales* et *duales* (cf. Section 2). La deuxième contribution réside dans la construction d'une densité-jouet bien choisie qui permet d'analyser les fonctionnements de ces différentes méthodes (cf. Section 3). La troisième contribution est de discuter leurs comportements sur des données réelles de Covid19 (cf. Section 4).

2 Monte Carlo Langevin proximal

Formulation. Dans une perspective plus générale que celle de la Covid19, nous supposons que les éléments de l'équation (1) satisfont les contraintes suivantes :

(i) l'opérateur proximal de g existe et a une expression explicite. En revanche, aucune hypothèse n'est faite sur l'opérateur proximal de $g(\mathsf{A}\cdot)$.

(ii) A est de taille $c \times d$ avec $c \leq d$ et de rang plein.

Par suite, A peut être augmentée en une matrice $\bar{\mathsf{A}}$ inversible par ajouts de $d-c$ lignes ; pour approcher π, il est alors possible de simuler une chaîne de Markov $\{\tilde{\boldsymbol{\theta}}^n, n \geq 0\}$ visant la densité *duale* π_{d} (où $\bar{g}(\tau) := g(\tau_{d-c+1:d})$) :

$$\pi_{\mathsf{d}}(\tilde{\boldsymbol{\theta}}) \propto \exp\left(-(f(\bar{\mathsf{A}}^{-1}\tilde{\boldsymbol{\theta}}) + \bar{g}(\tilde{\boldsymbol{\theta}}))\right) \mathbb{1}_{\mathcal{D}}(\bar{\mathsf{A}}^{-1}\tilde{\boldsymbol{\theta}}), \qquad (2)$$

puis d'approcher π par la chaîne $\{\boldsymbol{\theta}^n := \bar{\mathsf{A}}^{-1}\tilde{\boldsymbol{\theta}}^n, n \geq 0\}$ (cf. [11, Section 3] et [12, Section 4.2.] pour des idées similaires). Cette approche conduit aux algorithmes Langevin proximaux que nous nommons *méthodes duales* par opposition aux stratégies usuelles qui simulent une chaîne dans l'espace d'origine, associée à l'équation (1), méthodes dites *primales*.

Méthodes primales adaptées à la densité (1).

• **Dérive de Moreau (`M`).** [10] propose la dérive $\boldsymbol{\mu}^{\mathrm{M}}(\boldsymbol{\theta}) := \boldsymbol{\theta} - \gamma\nabla f(\boldsymbol{\theta}) - \frac{\gamma}{\rho}\mathsf{A}^\top(\mathrm{I} - \mathrm{prox}_{\rho g})\mathsf{A}\boldsymbol{\theta}$ où ρ est le paramètre de l'enveloppe de Moreau ; on peut prendre $\rho := \gamma$ [10].

• **Dérive PGdec (`PGdec`).** Lorsque A vérifie $\mathsf{A}\mathsf{A}^\top = \nu\,\mathrm{I}$ pour $\nu > 0$, $\mathrm{prox}_{\gamma g(\mathsf{A}\cdot)}$ s'exprime sous forme explicite à l'aide de $\mathrm{prox}_{\nu\gamma g}$ [8, Propositions 23.25 et 23.345]. On peut définir une dérive correspondant à l'opérateur gradient-proximal de $f + g(\mathsf{A}\cdot)$: $\boldsymbol{\mu}^{\mathrm{PGdec}}(\boldsymbol{\theta}) := \mathrm{prox}_{\gamma\, g(\mathsf{A}\cdot)}(\boldsymbol{\theta} - \gamma\,\nabla f(\boldsymbol{\theta}))$. Cette approche est un cas particulier de ce que propose [5] pour un contexte plus général que (1). Lorsque $g(\mathsf{A}\cdot) = \sum_{i=1}^{I} g_i(\mathsf{A}_i\cdot)$ avec $\mathsf{A}_i\mathsf{A}_i^\top = \nu_i\,\mathrm{I}$, on peut choisir aléatoirement à chaque itération de l'algorithme, un des blocs $g_i(\mathsf{A}_i\cdot)$ [5].

• **Dérive marche aléatoire (`RW`).** La dérive simple de la marche aléatoire s'applique et correspond à $\boldsymbol{\mu}^{\mathrm{RW}}(\boldsymbol{\theta}) := \boldsymbol{\theta}$.

Méthodes duales adaptées à la densité duale (2).

• **Dérive Moreau (`Mdual`).** π_{d} est de la forme (1) où l'opérateur linéaire associé à g se réduit à I. La méthode de [9] s'applique, avec la dérive $\tilde{\boldsymbol{\mu}}^{\mathrm{M}}(\tilde{\boldsymbol{\theta}}) := \tilde{\boldsymbol{\theta}} - \gamma\bar{\mathsf{A}}^{-\top}\nabla f(\bar{\mathsf{A}}^{-1}\,\tilde{\boldsymbol{\theta}}) - \frac{\gamma}{\rho}(\mathrm{I} - \mathrm{prox}_{\rho\bar{g}})\tilde{\boldsymbol{\theta}}$ dans l'espace dual. On en déduit $\boldsymbol{\mu}^{\mathrm{Mdual}}(\boldsymbol{\theta}) := \bar{\mathsf{A}}^{-1}\tilde{\boldsymbol{\mu}}^{\mathrm{M}}(\bar{\mathsf{A}}\boldsymbol{\theta})$ dans l'espace d'origine. On peut prendre $\rho := \gamma$ [9].

• **Dérive PGdual (`PGdual`).** Elle est adaptée de [5] au cas simplifié (1) ; c'est une étape de gradient-proximal dans l'espace dual $\tilde{\boldsymbol{\mu}}^{\mathrm{PG}}(\tilde{\boldsymbol{\theta}}) := \mathrm{prox}_{\gamma\bar{g}}\left(\tilde{\boldsymbol{\theta}} - \gamma\bar{\mathsf{A}}^{-\top}\nabla f(\bar{\mathsf{A}}^{-1}\tilde{\boldsymbol{\theta}})\right)$; on en déduit $\boldsymbol{\mu}^{\mathrm{PGdual}}(\boldsymbol{\theta}) := \bar{\mathsf{A}}^{-1}\tilde{\boldsymbol{\mu}}^{\mathrm{PG}}(\bar{\mathsf{A}}\boldsymbol{\theta})$ dans l'espace d'origine.

• **Dérive marche aléatoire (`RWdual`).** Le choix $\tilde{\boldsymbol{\mu}}^{\mathrm{RW}}(\tilde{\boldsymbol{\theta}}) := \tilde{\boldsymbol{\theta}}$ induit $\boldsymbol{\mu}^{\mathrm{RWdual}}(\boldsymbol{\theta}) := \boldsymbol{\theta}$ dans l'espace d'origine.

Matrices de covariance. Les dérives primales, vues comme des extensions de celle de Langevin au cas de lois cibles non régulières, sont naturellement associées à la matrice de covariance $\mathsf{C} := 2\gamma\,\mathrm{I}$. Dans le cas des dérives duales, il est naturel de proposer une matrice de covariance égale à $2\gamma\,\mathrm{I}$, ce qui revient à poser $\mathsf{C} := 2\gamma\bar{\mathsf{A}}^{-1}\bar{\mathsf{A}}^{-\top}$ dans l'espace primal.

3 Exemple-jouet

Définition. Soit π_{t} définie sur $\mathcal{D} := \mathbb{R}^d$, la densité issue d'un critère de vraisemblance dans un modèle de régression logistique, incluant une pénalité sur le vecteur de régression $\boldsymbol{\theta}$:

$$\ln \pi_{\mathsf{t}}(\boldsymbol{\theta}) := \mathsf{Y}^\top\mathsf{X}\boldsymbol{\theta} - \sum_{j=1}^{N} \ln\left(1 + \exp((\mathsf{X}\boldsymbol{\theta})_j)\right) - \lambda\|\mathsf{D}_1\boldsymbol{\theta}\|_1;$$

$\mathsf{Y} \in \{0,1\}^N$ collecte le vecteur des réponses binaires, X est la matrice de taille $N \times d$ des covariables ; D_1 est la matrice de différentiation discrète de taille $(d-1) \times d$ et $\lambda > 0$.

Échantillonneurs. π_{t} est de la forme (1) et les six méthodes présentées en Section 2 s'appliquent. Pour `PGdec`, nous écrivons $\|\mathsf{D}_1\boldsymbol{\theta}\|_1 = \|\mathsf{D}_{1,p}\boldsymbol{\theta}\|_1 + \|\mathsf{D}_{1,i}\boldsymbol{\theta}\|_1$ où $\mathsf{D}_{1,p}$ (resp. $\mathsf{D}_{1,i}$) collecte les lignes d'indice pair (resp. impair) de D_1 ; ces deux matrices vérifient $\mathsf{D}_{1,x}\mathsf{D}_{1,x}^\top = \nu\,\mathrm{I}$ pour $x \in \{i,p\}$ et un $\nu > 0$ (ici $\nu = 1$). Pour la mise en œuvre des méthodes *duales*, $\bar{\mathsf{A}}$ est définie comme la matrice $\mathsf{A} = \mathsf{D}_1$ à laquelle on rajoute une première ligne égale à la projection du vecteur $(-1, 0, \cdots, 0) \in \mathbb{R}^d$ sur l'espace orthogonal aux lignes de D_1. Pour `M` et `Mdual`, nous prenons $\rho := \gamma$. Pour toutes ces méthodes, la valeur de γ est adaptée durant les $5\,10^3$ premières itérations pour atteindre un taux d'acceptation-rejet moyen de 0.25. La comparaison de `RW` et `RWdual` aux quatre autres approches permet d'étudier la pertinence de méthodes exploitant des informations d'ordre 1 sur $\mathcal{L} := \ln \pi_{\mathsf{t}}$.

Performances comparées. Pour comparer les performances des différents échantillonneurs, *Log* π mesure la distance relative au maximum $\mathcal{L}_\star$ ($-\mathcal{L}$ est fortement convexe) de $\mathcal{L}$: l'évolution de $\left(\mathcal{L}(\boldsymbol{\theta}^n) - \mathcal{L}_\star\right) / \left(\mathcal{L}(\boldsymbol{\theta}^1) - \mathcal{L}_\star\right)$ est tracée sur les $2\,500$ premières itérations de la chaîne. *ACF* représente la valeur absolue de la fonction d'autocorrélation en fonction du décalage (de 0 à 600), calculée à partir de $17\,500$ points obtenus après

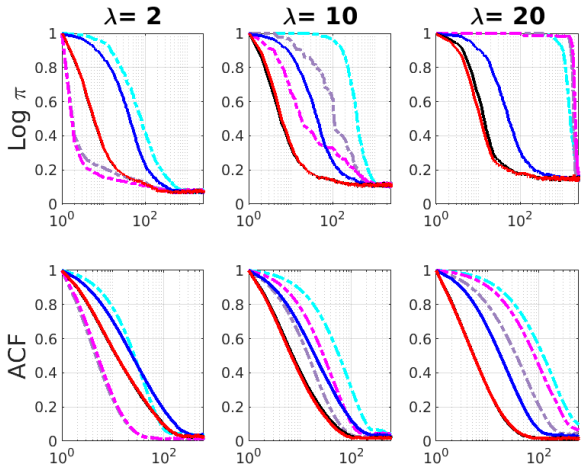

FIGURE 1 – Evolution de *Log* π en fonction du nombre d'itérations, et de *ACF* en fonction du décalage. Les méthodes primales sont en trait-point : `RW` en cyan, `M` en gris et `PGdec` en magenta. Les méthodes duales sont en trait plein : `RWdual` en bleu, `Mdual` en noir et `PGdual` en rouge. Les courbes `PGdual` et `Mdual` sont souvent superposées.

une période de chauffe (2 500 points). Ce critère est d'autant meilleur qu'il est petit, il reflète l'ergodicité de la chaîne [6].

Le jeu de données est simulé : $N = 2\,10^3$, $d = 20$; pour X, on tire des variables aléatoires (v.a.) indépendantes de Rademacher puis on normalise les lignes de X à 1 (donc $\nu = 1$) ; les composantes de Y sont des v.a. de Bernoulli indépendantes, de probabilité de succès $(1+\exp(-(\mathsf{X}\boldsymbol{\theta}^\star)_j))^{-1}$ où $\boldsymbol{\theta}^\star$ est constant par blocs (six composantes égales à 1, puis sept à -0.5, puis sept à 1). Enfin, trois valeurs différentes de λ sont considérées : $\lambda \in \{2, 10, 20\}$. Les performances rapportées en Figure 1 sont obtenues comme moyenne sur 50 réalisations indépendantes.

La figure 1 illustre le gain à exploiter des informations d'ordre 1 sur π pour accélérer le déplacement de l'échantillonneur vers les zones de plus forte densité. Pour les approches *primales*, ce gain faiblit lorsque λ est grand. L'exploitation d'informations partielles comme faite par `PGdec`, reste possible lorsque λ est petit ; pour des valeurs de λ plus grandes, `PGDec` est moins efficace que `M` dans le régime stationnaire de la chaîne (voir le critère *ACF*). Pour les approches *duales*, `PGdual` et `Mdual` présentent une excellente robustesse à la valeur de λ avec un léger avantage pour `PGdual` lorsque λ est moyen à grand. Enfin, l'ensemble de ces graphes incite à préférer l'approche *duale* à l'approche *primale*, et en particulier les échantillonneurs `PGdual` et `Mdual` dont les performances sont toujours bonnes, et les meilleures pour des valeurs λ moyennes à grandes.

4 Application à la pandémie de Covid19

Modèle bayésien. Nous avons proposé [3] une extension d'un modèle épidémiologique classique [1] destinée à estimer conjointement les taux de reproduction $\mathbf{R} := (\mathsf{R}_1, \cdots, \mathsf{R}_T)$ et les valeurs aberrantes $\mathbf{O} := (\mathsf{O}_1, \cdots, \mathsf{O}_T)$ dans les comptes de nouvelles infections, $\boldsymbol{\theta} := (\mathbf{R}, \mathbf{O})$, à partir des comptes de nouvelles infections quotidiennes observées $\mathbf{Z} := (\mathsf{Z}_1, \ldots, \mathsf{Z}_T)$. Nous avons établi [5] que la loi *a posteriori* prend la forme (1), où le terme d'attache aux données $-f$ s'écrit comme la log-vraisemblance d'un processus de Poisson (divergence de Kullback-Leibler), définie sur le domaine adapté $\mathcal{D}$: $f(\boldsymbol{\theta}) := \sum_{t=1}^{T} d_{\mathrm{KL}}\left(\mathsf{Z}_t \,\middle|\, \mathsf{R}_t \sum_{u=1}^{\tau_\phi} \Phi_u \mathsf{Z}_{t-u} + \mathsf{O}_t\right)$, [5]. La fonction d'intervalle de série Φ, approximée par une distribution Gamma tronquée à $\tau_\phi = 26$ jours, modélise le délai aléatoire entre infections primaire et secondaire. Le terme de régularisation $g(\mathsf{A}\boldsymbol{\theta}) := \lambda_{\mathsf{R}}\|\mathsf{D}_2\mathbf{R}\|_1 + \lambda_{\mathsf{O}}\|\mathbf{O}\|_1$ favorise simultanément le caractère parcimonieux de O et le comportement linéaire par morceaux de R, en pénalisant $(\mathsf{D}_2\mathbf{R})_t := (\mathsf{R}_{t+2} - 2\mathsf{R}_{t+1} + \mathsf{R}_t)/\sqrt{6}$. Celui-ci peut être réécrit à partir de $g = \lambda_{\mathsf{R}}\|\cdot\|$ et de la matrice par blocs $\mathsf{A} = [\mathsf{D}_2, 0; 0, \lambda_{\mathsf{O}}/\lambda_{\mathsf{R}}\mathsf{I}]$, où I désigne la matrice identité de taille T ; une augmentation inversible $\bar{\mathsf{A}}$ est obtenue en complétant D_2 en une matrice inversible $\overline{\mathsf{D}}_2$ [5].

Échantillonneurs duaux. Suivant les conclusions de la Section 3, nous concentrons la comparaison sur les méthodes duales, `MDual`, `PGdual` et `RWdual`, dont nous détaillons, pour les deux premières, les termes de dérive particularisés au modèle Covid19 bayésien. Avec le choix $\rho = \gamma$, `Mdual` devient :

$$\mathbf{R}^{n+\frac{1}{2}} := \mathbf{R}^n - \gamma\overline{\mathsf{D}}_2^{-1}\overline{\mathsf{D}}_2^{-\top}\nabla_{\mathsf{R}} f(\boldsymbol{\theta}^n) - \overline{\mathsf{D}}_2^{-1}\left[0\,; 0\,; \mathsf{D}_2\boldsymbol{\theta}^n - \mathrm{prox}_{\gamma\lambda_{\mathsf{R}}\|\cdot\|_1}(\mathsf{D}_2\mathbf{R}^n)\right] + \sqrt{2\gamma}\xi_{\mathsf{R}}^{n+1},$$

$$\mathbf{O}^{n+\frac{1}{2}} := \mathrm{prox}_{\gamma_{\mathsf{O}}\lambda_{\mathsf{O}}\|\cdot\|_1}(\mathbf{O}^{n+\frac{1}{2}}) - \gamma_{\mathsf{O}}\nabla_{\mathsf{O}} f(\boldsymbol{\theta}^n) + \sqrt{2\gamma_{\mathsf{O}}}\xi_{\mathsf{O}}^{n+1},$$

où ∇_{R} (resp. ∇_{O}) désigne le gradient partiel par rapport à $\mathbf{R}$ (resp. $\mathbf{O}$), $\gamma_{\mathsf{O}} := \gamma(\lambda_{\mathsf{R}}/\lambda_{\mathsf{O}})^2$, $\xi_{\mathsf{R}}^{n+1} \sim \mathcal{N}(0, \overline{\mathsf{D}}_2^{-1}\overline{\mathsf{D}}_2^{-\top})$ et $\xi_{\mathsf{O}}^{n+1} \sim \mathcal{N}(0, \mathsf{I})$. La dérive `PGdual` prend la forme :

$$\mathbf{R}^{n+\frac{1}{2}} := \overline{\mathsf{D}}_2^{-1}\,\mathrm{prox}_{\gamma\lambda_{\mathsf{R}}\|(\cdot)_{3:T}\|_1}\left(\overline{\mathsf{D}}_2\mathbf{R}^n - \gamma\overline{\mathsf{D}}_2^{-\top}\nabla_{\mathsf{R}} f(\boldsymbol{\theta}^n)\right) + \sqrt{2\gamma}\xi_{\mathsf{R}}^{n+1},$$

$$\mathbf{O}^{n+\frac{1}{2}} := \mathrm{prox}_{\gamma_{\mathsf{O}}\lambda_{\mathsf{O}}\|\cdot\|_1}\left(\mathbf{O}^n - \gamma_O\nabla_{\mathsf{O}} f(\boldsymbol{\theta}^n)\right) + \sqrt{2\gamma_{\mathsf{O}}}\xi_{\mathsf{O}}^{n+1}.$$

Données de Covid19. Les données sont téléchargées de la base *Johns Hopkins University* [1], qui, remarquablement, depuis le début de la pandémie, collecte et organise les donnés Covid19 rendues disponibles par les autorités de santé publique de près de 200 pays. Nous n'utiliserons ici que les comptes quotidiens de nouvelles infections Z_t pour une période récente de 5 semaines ($T = 35$ jours) et deux pays, mais les outils décrits peuvent être appliqués à tous pays ou périodes d'intérêt.

Simulations MCMC. Les chaînes simulées comprennent 10^7 échantillons dont 30% pour la phase de chauffe et sont initialisées au point non informatif $\mathbf{R}^{\mathrm{init}} := (1, \ldots, 1)^\top$, $\mathbf{O}^{\mathrm{init}} := (0, \ldots, 0)^\top$. Le pas γ est ajusté pendant la phase de chauffe pour produire un taux d'acceptation/rejet de 0.25 [13]. Nous utilisons $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}}) = (3.5\,\sigma_{\mathbf{Z}}\sqrt{6}/4, 0.05)$ où $\sigma_{\mathbf{Z}}$ désigne l'écart-type de $\mathbf{Z}$, qui permet de s'affranchir des différences de tailles de population ou d'intensité de la pandémie entre pays [3].

Discussion. La figure 2 montre une décroissance plus lente pendant la phase de chauffe du critère *Log* π pour `RWdual` (comparée à `Mdual` et `PGdual`), ainsi qu'une décroissance plus lente du critère *ACF*, ce qui confirme ainsi l'intérêt du terme de dérive des méthodes `Mdual` et `PGdual`.

1. https://coronavirus.jhu.edu/

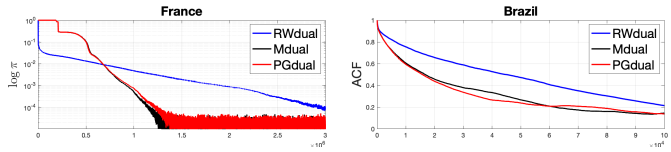

FIGURE 2 – **Covid19 : performances des échantillonneurs.** Critères *Log* π (gauche, phase de chauffe) et *ACF* (droite).

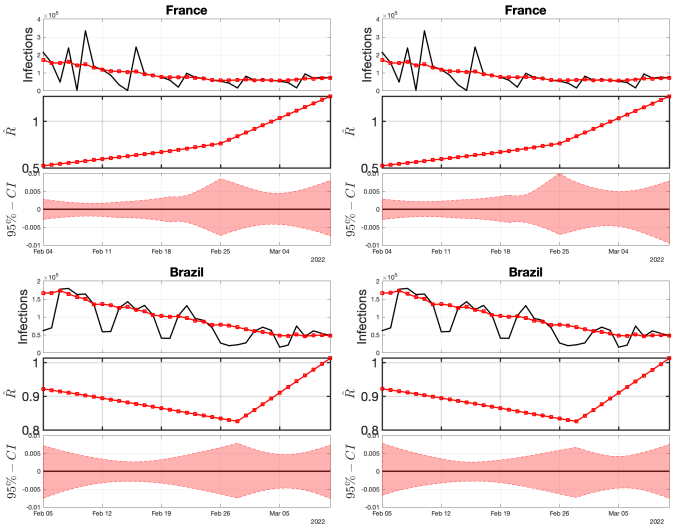

FIGURE 3 – **Covid19 : estimations et intervalles de crédibilité.** Pour `Mdual` (gauche) et `PGdual` (droite), superposition de comptes bruts $\mathbf{Z}$ (noir) et débruités $\mathbf{Z}^{(D)}$ (rouge) ; $\hat{\mathsf{R}}$; intervalles de crédibilité (95%) autour de $\hat{R}$.

La figure 3 compare, de plus, pour deux pays, les deux échantillonneurs duaux `Mdual` et `PGdual`, la superposition de comptes bruts $\mathbf{Z}$ et débruités $\mathbf{Z}^{(D)}$ obtenus par soustraction de $\hat{\mathsf{O}}$ (défini comme la médiane *a posteriori*) à $\mathbf{Z}$, $\hat{\mathsf{R}}$ (estimateur de la médiane *a posteriori*) et intervalles de crédibilité (95%) autour de $\hat{\mathsf{R}}$, définis par les quantiles empiriques 0.025 et 0.975. La figure 3 montre que les estimées des méthodes `Mdual` et `PGdual` sont comparables, en accord avec les résultats de la section 3. Les différences entre les $\hat{\mathsf{R}}$ ou entre les tailles des intervalles de crédibilité obtenus par les méthodes `Mdual` et `PGdual` sont de l'ordre de 10^{-3}, pour des intervalles de crédibilité de taille moyenne de 10^{-2}. `Mdual` induit cependant un surcoût en temps de calcul d'environ 45% par rapport à `PGdual`. La figure 3 montre aussi que l'estimation des valeurs aberrantes et donc des comptes débruités de nouvelles infections est parfaitement opérationnelle. Elle montre enfin que les intervalles de crédibilité s'élargissent aux occurrences des changements de pente des estimées de R_t.

La figure 3 montre enfin que, à l'heure où nous écrivons (14 mars 2022), la pandémie est à nouveau en phase croissante en France et ce depuis plus de 5 semaines, avec une accélération depuis le 25 février. Cette recrudescence de la pandémie est également observable sur les données du Brésil, avec un retournement de tendance le 27 février, et, de manière concomitante pour de nombreux autres pays. L'estimation linéaire par morceaux promue ici permet donc *naturellemment* une prévision de tendance à très court terme, en plus de l'analyse rétrospective de l'impact des mesures sanitaires sur l'évolution de la pandémie.

Ces estimations et intervalles de crédibilité sont mis à jour quotidiennement pour la France [2].

Références

[1] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, "A new framework and software to estimate time-varying reproduction numbers during epidemics," *Am. J. Epidemiol.*, vol. 178, pp. 1505–1512, 2013.

[2] P. Abry et al., "Spatial and temporal regularization to estimate COVID-19 reproduction number $R(t)$: Promoting piecewise smoothness via convex optimization," *PLOS One*, vol. 15, 2020, e0237901.

[3] B. Pascal, P. Abry, N. Pustelnik, S. Roux, R. Gribonval, and P. Flandrin, "Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data," Tech. Rep., arXiv 2109.09595, 2022.

[4] H. Artigas, B. Pascal, G. Fort, P. Abry, and N. Pustelnik, "Credibility interval design for Covid19 reproduction number from nonsmooth Langevin-type Monte Carlo sampling," Tech. Rep., hal-03371837, 2021.

[5] G. Fort, B. Pascal, P. Abry, and N. Pustelnik, "Covid19 Reproduction Number : Credibility Intervals by Blockwise Proximal Monte Carlo Samplers," Tech. Rep. hal-03611079v1, 2022.

[6] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, Berlin, Heidelberg, 2005.

[7] G.O. Roberts and O. Stramer, "Langevin Diffusions and Metropolis-Hastings Algorithms," *Methodol. Comput. Appl. Probab.*, vol. 4, pp. 337–357, 2002.

[8] H. H Bauschke and P.-L. Combettes, "Convex Analysis and Monotone Operator Theory in Hilbert Spaces," Springer International Publishing, 2017.

[9] A. Durmus, É. Moulines, and M. Pereyra, "Efficient Bayesian Computation by Proximal Markov Chain Monte Carlo : When Langevin Meets Moreau," *SIAM J Imaging Sci*, vol. 11, pp. 473–506, 2018.

[10] T.D. Luu, J. Fadili, and C. Chesneau, "Sampling from Non-smooth Distributions Through Langevin Diffusion," *Methodol Comput Appl Probab*, 2020.

[11] R.J. Tibshirani and J. Taylor, "The solution path of the generalized lasso," *Ann. Stat.*, vol. 39, no. 3, pp. 1335 – 1371, 2011.

[12] A.S. Dalalyan, "Theoretical guarantees for approximate sampling from smooth and log-concave densities," *J. Roy. Statist. Soc. B*, vol. 79, no. 3, pp. 651–676, 2017.

[13] G.O. Roberts and J.S. Rosenthal, "Optimal scaling for various Metropolis-Hastings algorithms," *Statistical Science*, vol. 16, pp. 351 – 367, 2001.

2. perso.math.univ-toulouse.fr/gfort/project/opsimore/