Posterior convergence analysis of $\alpha$-stable processes

Neil Chada

National University of Singapore

IMS programme: Bayesian Computation for High-Dimensional Statistical Models

Sari Lasanen and Lassi Roininen
(Lappeenranta University of Technology)

September 18, 2018
Motivation

$\alpha$-stable processes

Convergence analysis

Current work
Inverse problems

- Abstract setting: \((\mathcal{X}, \langle \cdot \rangle, \| \cdot \|), (\mathcal{Y}, \langle \cdot \rangle, \| \cdot \|)\).
- **Aim**: The recovery of an unknown \(u \in \mathcal{X}\) from perturbed noisy measurements of data \(y \in \mathcal{Y}\) where

\[
y = G(u) + \eta. \tag{1}
\]

- Solution-to-parameter operator: \(G : \mathcal{O} \circ G : \mathcal{X} \to \mathbb{R}^k\).
- Forward operator: \(G : \mathcal{X} \to V\) (Solution space)
- Observational operator: \(\mathcal{O} : V \to \mathbb{R}^k\)
- Additive Gaussian noise: \(\eta \sim \mathcal{N}(0, \Gamma)\).

**Question**: How to solve for \(u\) from (1)???
Deterministic approach

- Construct functional with added regularization and minimize

\[ u^* := \text{argmin}_{u \in X} J(u), \]  
\[ J(u) := \frac{1}{2} |y - G(u)|_Y^2 + \frac{\lambda}{2} |u|_E^2, \quad \lambda > 0, \quad E \subset X. \]  

Numerically solved through various optimization methods:

(i) Least squares.
(ii) Conjugate gradient.
(iii) L-BFGS.

Issues that can arise:

- No guarantee of well-posedness.
- Regularization can be dependent on the problem.
- Account for uncertainty within system?
Bayesian approach

▶ Finite dimension

Unknown is now a probabilistic distribution of the random variable $u | y$ using Bayes’ formula

$$P(u | y) \propto P(y | u) P(u).$$

▶ $\infty$-dimension

We consider a posterior measure $\mu^y$ described through Radon-Nikodym derivative

$$\frac{d \mu^y}{d \mu_0}(u) = \frac{1}{Z} \exp(-\Phi(u; y)),$$

where

$$Z := \int_{\mathcal{X}} \exp(-\Phi(u; y)) \mu_0(du),$$

with misfit functional

$$\Phi(u; y) = \frac{1}{2} |y - G(u)|^2_\Gamma.$$
Bayesian approach

- Well-posedness theorem ✓
- Tackles uncertainty ✓

\[-\nabla \cdot (\kappa \nabla p) = f \quad \in D \quad \}
\[p = 0 \quad \in \partial D \quad\]

- \(\kappa \sim \mathcal{N}(0, \mathcal{C})\), \(\kappa \in L^\infty(D)\).
- \(\kappa = \sum_j \sqrt{\lambda_j} \xi_j \phi_j\), \(C_j \phi_j = \lambda_j \phi_j\).
- \(\sigma^2(I - \tau^2 \Delta)^{\alpha/2} \kappa = \sqrt{\beta} \tau^{d/2} \mathcal{W}\), \(\mathcal{W} \sim \mathcal{N}(0, I)\).

**Uncertainty** can arise such as (i) heterogenous field, (ii) level set/phase field construction, (iii) geometric.
Theorem

Assume that $\mu_0$ is defined as $\mathcal{N}(0, \mathcal{C})$, $y$ by (1) and $\Phi$ by $\frac{1}{2}|y - G(u)|_\Gamma^2$. If $\mu^y$ is the regular conditional probability measure on $u|y$, then $\mu^y \ll \mu_0$ with Radon-Nikodym derivative

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z} \exp(-\Phi(u; y)),$$

where

$$Z := \int_X \exp(-\Phi(u; y))\mu_0(du).$$

Furthermore $\mu^y$ is locally Lipschitz with respect to $y$ in the Hellinger distance: for all $y, y'$ with $\max\{|y|_\Gamma, |y'|_\Gamma\} \leq r$, there exists a $C = c(r) > 0$ such that

$$d_{\text{Hell}}(\mu^y, \mu^{y'}) \leq C|y - y'|_\Gamma.$$
Assumptions

The least squares functional $\Phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and probability measure $\mu_0$ on the space $(\mathcal{X}, \Sigma)$ satisfy the properties

1. Every $r > 0$ there is a $K = K(r)$, such that for all $u \in \mathcal{X}$, and $y \in \mathcal{Y}$, with $0 \leq \Phi(u; y) \leq K$.

2. For any fixed $y \in \mathcal{Y}$, $\Phi(\cdot; y) : \mathcal{X} \to \mathbb{R}$ is continuous $\mu_0$-almost surely on the probability space $(\mathcal{X}, \Sigma, \mu_0)$.

3. For $y_1, y_2 \in \mathcal{X}$ with $\max\{|y_1|_\Gamma, |y_2|_\Gamma\} < r$, there exists a $C = c(r)$ such that, for all $u \in \mathcal{X}$

   $$|\Phi(u; y_1) - \Phi(u; y_2)| \leq C|y_1 - y_2|_\Gamma$$

4. Continuity of the map $G$. *(unrelated to misfit functional).*
Edge-preserving Bayesian inversion?
Brief history

- **Gaussian priors:**
  \[ u \sim \mathcal{N}(0, \mathcal{C}) \]
  (Lehtinen [1991], Fitzpatrick [1992], Knapik [2008], Agapiou [2011]).

- **Geometric priors:**
  \[ u = \sum_{i=1}^{n} u_i(x) \chi_{D_i}(x) \]
  (Somsersalo [2004], Iglesias [2013]).

- **Level set priors:**
  \[ w = w^+\mathbb{I}_{u>0}(x) + w^-\mathbb{I}_{u<0}(x) \]
  (Burger [1991], Iglesias [2011], Lu [2015]).

- **Total variation priors:**
  degenerate with mesh, (Lassas, Siltanen [2008]).

- **Besov priors:**
  \[ u = \sum_{j=1}^{n} \langle u_j, \phi \rangle \phi_j \]
  (Lassas [2009], Dashti [2011], Agapiou [2017]).

- **Laplace priors:**
  \[ u = \sum_{j=1}^{n} \sqrt{\lambda_j} \xi_j \phi_j \] Laplace noise
  (Hosseini [2016], [2017]).
Extension to $\alpha$-stables processes?
\( \alpha \)-stable distributions

- linear combination of two independent r.v's \( X_1, X_2 \) \( \Rightarrow \) stable distribution.

\[
aX_1 + aX_2 = cX + d.
\]

- a r.v. is stable if its distribution is stable.

\[
X \sim S_\alpha(\mu, \beta, \sigma).
\]

- \( \alpha \in (0, 2] \) - stability.
- \( \beta \in [-1, 1] \) - skewness.
- \( \mu \in (0, \infty) \) - location.
- \( \sigma \in (0, \infty) \) - scale.
- Gaussian case = \( S_2(\sigma, 0, \mu) \), Cauchy case = \( S_1(\sigma, 0, \mu) \).
What we consider

- Understanding theoretical properties of these processes, i.e. convergence.
- Finite convergence (expectation).
- For **simplicity**: finite dimensions, finite observations.
- \( \mathbb{R} \)-values stable processes.
- Domain will be fixed.
- Interested in the case of \( \alpha < 2 \).
Definition

An independently scattered \( \sigma \)-additive set function

\[ M : \epsilon_0 \to L^0(\Omega), \]

such that for any \( A \in \epsilon_0 \),

\[ M(A) \sim S_\alpha \left( (m(A))^{1/\alpha}, \frac{\int_A \beta(x)m(dx)}{m(A)}, 0 \right), \]

is called an \( \alpha \)-stable random measure on \((E, \epsilon)\) with control measure \( m \) and skewness parameter \( \beta \).
\(\alpha\)-stable random fields

- Special case of Brownian sheet.

**Definition**

A random field \(X\) is called a multivariable \(\alpha\)-stable sheet if

\[
X(t_1, \ldots, t_n) := \int_{[0,t_1] \times \ldots \times [0,t_n]} M(ds_1, \ldots, ds_n).
\]

A natural discretisation of (3) on \([0, 1]^n\) arises by considering a uniform grid \(\{t = kh : k \in \{0, \ldots, N\}^n\}\), \(h = 1/N\) and \(N \in \mathbb{N}\). Indeed,

\[
X(k_1h, \ldots, k_nh) = \sum_n \int 1_{I_n}(s_1, \ldots, s_n) M(ds_1, \ldots, ds_n),
\]

where \(I_n\) are disjoint hypercubes of Lebesgue measure \(|I_n| = h^n\) whose all vertices are on the grid and \(n\) represents some fixed vertex of the cube.
Convergence of sheets

- Integrand representation of stable processes.

**Theorem [C., Lasanen, Roininen 18]**

Let $X^N(t_1, \ldots, t_n) = \sum_{k_1=1}^{\lceil t_1/h \rceil} \cdots \sum_{k_n=1}^{\lceil t_n/h \rceil} \int 1_{I_k}(s) dM_s$ for $n < \infty$, then $X^N(t) \to X(t)$ in probability when $N \to \infty$.

**Theorem [C., Lasanen, Roininen 18]**

Let $X^N(t_1, \ldots, t_n) = \sum_{k_1=1}^{\lceil t_1/h \rceil} \cdots \sum_{k_n=1}^{\lceil t_n/h \rceil} \int 1_{I_k}(s) dM_s$ for $n < \infty$, then $X^N(t) \to X(t)$ almost surely when $N \to \infty$.

- Theorems follow nicely from the properties of stable processes.
- Show convergence of other representations?
Representations

- Consider other forms of $\alpha$-stable processes.
- $\alpha$-stable random measures.
- Poisson process measures.

We can represent as: Let $\Gamma_i$ be arrivals times of a Poisson process with arrival rate 1. Let $(V_i, \gamma_i)$ form an i.i.d. sequence of random vectors independent of $\Gamma_i$ that consist of uniformly distributed $d$-dimensional random vectors $V_i$ on $[0, 1]^n$, and $\{-1, 1\}$-valued random variables $\gamma_i$

$$\tilde{X}(t) := C_\alpha^{1/\alpha} \sum_{i=1}^{\infty} \gamma_i \Gamma_i^{-1/\alpha} 1_{[0,t_1] \times \cdots \times [0,t_n]}(V_i).$$  \hspace{1cm} (4)

with

$$C_\alpha = \left( \int_0^\infty x^{-\alpha} \sin(x) \, dx \right)^{-1}.$$
Convergence of random series

Lemma [C., Lasanen, Roininen 18]

The random series

$$\tilde{X}(t) := C_1^{1/\alpha} \sum_{i=1}^{\infty} \gamma_i \Gamma_i^{-1/\alpha} 1_{[0,t_1] \times \cdots \times [0,t_n]}(V_i),$$

which converges a.s. for $t = (t_1, \ldots, t_n) \in [0,1]^n$ and a.s. in $L^p([0,1]^n)$ for $\max(1,\alpha) < p < \infty$. Moreover, the distribution of $\tilde{X}$ on $L^p([0,1]^n)$ is identical to the distribution of $X(t_1, \ldots, t_n)$. 

Proof (sketch)

[1.] $\sum_{i=1}^{\infty} \gamma_i \Gamma_i^{-1/\alpha}$ converges a.s. when $0 < \kappa < 1$.

[2.] Itô-Nisio Theorem, a.s. convergence $\to$ weak convergence.

[3.] Various inequalities: Jensen, Hölder.
Convergence of random series

Lemma [C., Lasanen, Roininen 18]

The random series

\[ \tilde{X}(t) := C_\alpha^{1/\alpha} \sum_{i=1}^{\infty} \gamma_i \Gamma_i^{-1/\alpha} 1_{[0,t_1] \times \cdots \times [0,t_n]}(V_i), \]

which converges a.s. for \( t = (t_1, \ldots, t_n) \in [0,1]^n \) and a.s. in \( L^p([0,1]^n) \) for \( \max(1, \alpha) < p < \infty \). Moreover, the distribution of \( \tilde{X} \) on \( L^p([0,1]^n) \) is identical to the distribution of \( X(t_1, \ldots, t_n) \).

Proof (sketch)

[1.] \( \sum_{i=1}^{\infty} \Gamma_i^{-1/\kappa} \) converges a.s. when \( 0 < \kappa < 1 \).
[2.] Itô-Nisio Theorem, a.s. convergence \( \rightarrow \) weak convergence.
[3.] Various inequalities: Jensen, Hölder.
**$L^p$-sample path continuity**

- **Question:** If $\tilde{X}$ and its sample paths are in $L^p([0, 1]^n)$ is it a random variable in $L^p([0, 1]^n)$?
- The case of $1 \leq \alpha < 2$ is cadlag.
- Convergence will differ for this form.

**Lemma [C., Lasanen, Roininen 18]**

There exists $c(\omega), C(\omega) > 0$ and $K(\omega) \in \mathbb{N}$ so that $c(\omega)k \leq \Gamma_k(\omega) \leq C(\omega)k$ for all $k \geq K(\omega)$ and for $\mathbb{P}$-almost every $\omega$. Moreover, the series

$$\sum_{k=1}^{\infty} \Gamma_k^{-\kappa},$$

converges almost sure for all $\kappa < 1$. 
**$L^p$-sample path continuity**

- **Question:** If $\tilde{X}$ and its sample paths are in $L^p([0, 1]^n)$ is it a random variable in $L^p([0, 1]^n)$?
- The case of $1 \leq \alpha < 2$ is cadlag.
- Convergence will differ for this form.

<table>
<thead>
<tr>
<th>Lemma [C., Lasanen, Roininen 18]</th>
</tr>
</thead>
<tbody>
<tr>
<td>There exists $c(\omega), C(\omega) &gt; 0$ and $K(\omega) \in \mathbb{N}$ so that $c(\omega)k \leq \Gamma_k(\omega) \leq C(\omega)k$ for all $k \geq K(\omega)$ and for $\mathbb{P}$-almost every $\omega$. Moreover, the series $\sum_{k=1}^{\infty} \Gamma_k^{-\kappa}$ converges almost sure for all $\kappa &lt; 1$.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proof (sketch)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1.] Poisson process: $\Gamma_k = \sum_{j=1}^{k} \lambda_j$ with LLN.</td>
</tr>
<tr>
<td>[2.] $\Gamma_k \sim k$ and $c(\omega)k \leq \Gamma_k(\omega) \leq C(\omega)k$ for all $k &gt; K(\omega) \implies$ a.s. convergence.</td>
</tr>
</tbody>
</table>
Theorem [C., Lasanen, Roininen 18]

Let $A_k \subset [0,1]^n$, $k = 1, \ldots, N$, be such hypercubes with equal edge lengths $h$ that $\bigcup_{k=1}^N A_k = [0,1]^n$ and $|A_k \cap A_{k'}| = 0$ for all $k \neq k'$. Choose a point $t_k$ from each hypercube $A_k$.
If $0 < \alpha < 1$, the approximations

$$\tilde{X}^N(t) = \sum_{k=1}^N \tilde{X}(t_k) 1_{A_k}(t),$$

converge a.s. to $\tilde{X}$ in $L^p([0,1]^n)$ when $N \to \infty$. If $1 \leq \alpha < 2$, the approximations $\tilde{X}^N$ converge to $\tilde{X}$ in $L^1([0,1]^n)$ in distribution.
Proof

- For $0 < \alpha < 1$: by changing the order of the sums, we get

$$X^N(t) = C_{\alpha}^{1/\alpha} \sum_{i=1}^{\infty} \gamma_i \Gamma_i^{-1/\alpha} \sum_{k=1}^{N} 1_{[V_i \cdot e_1, 1]} \times \cdots \times [V_i \cdot e_n, 1](t_k) 1_{A_k^N}(t).$$

- Applying previous lemma and DCT we have

$$\lim_{N \to \infty} X^N(t) = C_{\alpha}^{1/\alpha} \sum_{i=1}^{\infty} \gamma_i \Gamma_i^{-1/\alpha} \lim_{N \to \infty} \sum_{k=1}^{N} 1_{[V_i \cdot e_1, 1]} \times \cdots \times [V_i \cdot e_n, 1](t_k) 1_{A_k}(t),$$

in $L^p([0, 1]^n)$.

- for $1 \leq \alpha < 2$: Aim to show

$$\lim_{N \to \infty} \mathbb{E}[f(X^N)] = \mathbb{E}[f(X)],$$

for all bounded Lipschitz functions on $L^1([0, 1]^d)$

- Split $X^N$ into $X^N = X_1^N(t) + X_2^N(t) \to$ conditional expectation + Khintchine inequality.
Back to well-posedness!

- We begin with assumptions on $\Phi(u; y)$ and the prior form.

**Theorem**

Assume that $\mu_0$ is defined as random measure, $y$ by (1) and $\Phi$ by $\frac{1}{2}|y - G(u)|^2$. If $\mu^y$ is the regular conditional probability measure on $u|y$, then $\mu^y \ll \mu_0$ with Radon-Nikodym derivative

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z} \exp(-\Phi(u; y)),$$

where

$$Z := \int_{\mathcal{X}} \exp(-\Phi(u; y))\mu_0(du).$$

Furthermore $\mu^y$ is locally Lipschitz with respect to $y$ in the Hellinger distance: for all $y, y'$ with $\max\{|y|_\Gamma, |y'|_\Gamma\} \leq r$, there exists a $C = c(r) > 0$ such that

$$d_{Hell}(\mu^y, \mu^{y'}) \leq C|y - y'|_\Gamma.$$
Cauchy difference priors

- Continuous stochastic processes \(X(\cdot)\) is Lévy stable process, starting from 0, if \(X\) has independent increments such that
  \[
  X(t) - X(s) \sim S_\alpha((t-s)^\frac{1}{\alpha}, \beta, 0)
  \]

- Discrete random walk at \(t = jh\) by \(X_j\), where \(j \in \mathbb{Z}^+\) and \(h > 0\)
  \[
  X_j - X_{j-1} \sim S_\alpha(h^\frac{1}{\alpha}, \beta, 0).
  \]

- We have the following density
  \[
  D(x) = C \prod_{j=1}^{j} \left( \frac{\lambda_j h}{(\lambda_j h)^2 + (X_j - X_{j-1})^2} \right), \quad \lambda_j > 0.
  \]
  Can be extended to 2D case easily
we discuss various approaches for sampling the statistically dependent stable random vectors \((X(s_1), \ldots, X(s_k))\), where \(s_1, \ldots, s_k \in [0, 1]^d\). A well-known approach is to reduce the sampling to independent increments, where in the 2D case we have

\[
X(t_1, t_2) = \int_{[0,t_1] \times [0,t_2]} M(ds).
\]

When the measure \(M\) is discretised into

\[
M^N(ds) = \sum_{k=1}^{N} \frac{1}{|A_k|} \left( \int 1_{A_k}(r) M(dr) \right) 1_{A_k}(s) ds,
\]

we obtain for the 2D case

\[
X^N(t_1, t_2) = \sum_{k=1}^{N} \frac{1}{|A_k|} \left( \int 1_{A_k}(r) M(dr) \right) 1_{A_k \cap [0,t_1] \times [0,t_2]}(s) ds,
\]
\[ X^N(t_1, t_2) = \sum_{k=1}^{N} \frac{1}{|A_k|} \left( \int 1_{A_k}(r) M(dr) \right) 1_{A_k \cap [0, t_1] \times [0, t_2]}(s) ds, \]  

Cauchy difference priors

\[ X(hp, hr) - X(hp, h(r - 1)) - X(h(p - 1), hr) \]
\[ + X(h(p - 1), h(r - 1)) \sim S_\alpha(h^{d/\alpha}, 0, 0), \]

Key question: Can one show (6) is consistent with (5)?

Aim: Show this limit analysis in the context of numerics.
Concluding remarks

- Vast literature on various priors.
- Considerable work on both theory and application.
- Edge-preserving Bayesian inversion (lack of analysis).

- Aim was to analyze this with $\alpha$-stable processes for $\mathbb{R}^d$.
- Convergence results of different forms.
- Work: contraction, convergence, numerical study.
Consistency and contraction

- **Question:** How close is the posterior measure $\mu^y$ close to $u^\dagger$?

- **Posterior consistency,** which states that the posterior measure contracts around the true solution $u^\dagger$ as $n \to \infty$. Mathematically if posterior consistency is achieved then, for all $\epsilon > 0$

  $$
  \mathbb{E}^y \mu^y \{ u : \| u - u^\dagger \| \geq \epsilon \} \to 0.
  $$

- Alternatively viewed as

  $$
y_j = G_j(u^\dagger) + \eta_j, \quad j, \ldots, N.
  $$

- We aim to show that $G(u_n) \to G(u^\dagger)$.

- Can we determine the rate $M_n \epsilon_n$ such that

  $$
  \mathbb{E}^y \mu^y \{ u : \| u - u^\dagger \| \geq M_n \epsilon_n \} \to 0, \quad \forall M_n \to \infty.
  $$
Random fields

Gaussian random field (above), Cauchy random field (below).


