Posterior convergece analysis of α-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 Technnology)

September 18, 2018

Finnish Centre of Excellence in Inverse Problems Research

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

$$y = \mathcal{G}(u) + \eta. \tag{1}$$

- Solution-to-parameter operator: $\mathcal{G} : \mathcal{O} \circ G : \mathcal{X} \to \mathbb{R}^k$.
- 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

$$u^* := \operatorname{argmin}_{u \in \mathcal{X}} J(u), \tag{2}$$

such that

$$J(u) := \frac{1}{2}|y - \mathcal{G}(u)|^2_{\mathcal{Y}} + \frac{\lambda}{2}|u|^2_E, \quad \lambda > 0, \quad E \subset \mathcal{X}.$$

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

$$\underbrace{\mathbb{P}(u|y)}_{\text{posterior}} \propto \underbrace{\mathbb{P}(y|u)}_{\text{likelihood}} \underbrace{\mathbb{P}(u)}_{\text{prior}}.$$

- ∞-dimension

We consider a posterior measure μ^y described through Radon-Nikodym derivative

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

where

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

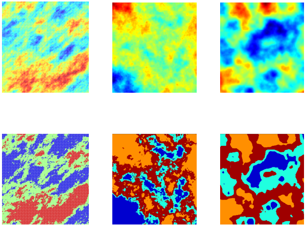
with misfit functional

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

Bayesian approach

- Well-posedness theorem ✓
- Tackles uncertainty ✓

$$\left.\begin{aligned} -\nabla\cdot(\kappa\nabla p) &= f & &\in D \\ p &= 0 & &\in \partial D \end{aligned}\right\}$$



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

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

Theorem

Assume that μ_0 is defined as $\mathcal{N}(0, \mathcal{C})$, y by (1) and Φ by $\frac{1}{2}|y - \mathcal{G}(u)|_\Gamma^2$. If μ^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 μ^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

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

Assumptions

The least squares functional $\Phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and probability measure μ_0 on the space $(\mathcal{X}, \Sigma)$ satisfy the properties

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

$$0 \leq \Phi(u; y) \leq K.$$

2. For ay fixed $y \in \mathcal{Y}$, $\Phi(\cdot; y) : \mathcal{X} \to \mathbb{R}$ is continuous μ_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 $u \in \mathcal{X}$

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

4. Continuity of the map $\mathcal{G}$. **(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 α-stables processes?

α-stable distributions

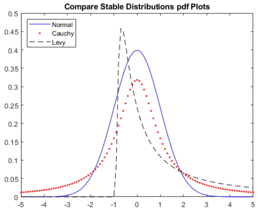
- linear combination of two independent r.v's $X_1, X_2 \implies$ stable distribution.

$$aX_1 + aX_2 = cX + d.$$

- 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 $\alpha < 2$.

α-stable measures

Definition

An independently scattered σ-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 α-stable random measure on (E, ϵ) with control measure m and skewness parameter β.

α-stable random fields

- Special case of Brownian sheet.

Definition

A random field X is called a multivariable α-stable sheet if

$$X(t_1, \dots, t_n) := \int_{[0,t_1]\times\dots\times[0,t_n]} M(ds_1, \dots, ds_n). \tag{3}$$

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

$$X(k_1 h, \dots, k_n h) = \sum_{n} \int 1_{I_n}(s_1, \dots, s_n) M(ds_1, \dots, ds_n),$$

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 vertice 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 α-stable processes.
- α-stable random measures.
- Poisson process measures.

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

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

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

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

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 $\widetilde{X}$ on $L^p([0,1]^n)$ is identical to the distribution of $X(t_1, \ldots, t_n)$.

Convergence of random series

Lemma [C., Lasanen, Roininen 18]

The random series

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

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 $\widetilde{X}$ 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 ω. 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.

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 ω. Moreover, the series

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

converges almost sure for all $\kappa < 1$.

Proof (sketch)

[1.] Poisson process: $\Gamma_k = \sum_{j=1}^{k} \lambda_j$ with LLN.
[2.] $\Gamma_k \sim k$ and $c(\omega)k \leq \Gamma_k(\omega) \leq C(\omega)k$ for all $k > K(\omega) \implies$ a.s. convergence.

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

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

converge a.s. to $\widetilde{X}$ 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

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

- 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 μ_0 is defined as random measure, y by (1) and Φ by $\frac{1}{2}|y - \mathcal{G}(u)|_\Gamma^2$. If μ^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 μ^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

$$X(t) - X(s) \sim S_\alpha\left((t-s)^{\frac{1}{\alpha}}, \beta, 0\right)$$

- Discrete random walk $t = jh$ by X_j, where $j \in \mathbb{Z}^+$ and $h > 0$

$$X_j - X_{j-1} \sim S_\alpha\left(h^{\frac{1}{\alpha}}, \beta, 0\right).$$

- 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

Random vectors

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

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

$$X^N(t_1,t_2) = \sum_{k=1}^{N} \frac{1}{|A_k|} \left(\int 1_{A_k}(r) M(dr) \right) 1_{A_k \cap [0,t_1]\times[0,t_2]}(s) ds, \tag{5}$$

- Cauchy difference priors

$$\begin{aligned} X(hp, hr) - X(hp, h(r-1)) - X(h(p-1), hr) \\ + X(h(p-1), h(r-1)) \sim S_\alpha(h^{d/\alpha}, 0, 0), \end{aligned} \tag{6}$$

- 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 α-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 μ^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 $\epsilon > 0$

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

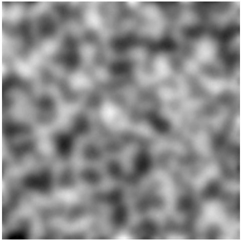
- Alternatively viewed as

$$y_j = \mathcal{G}_j(u^\dagger) + \eta_j, \quad j, \ldots, N.$$

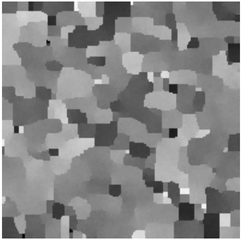
- we aim to show that $\mathcal{G}(u_n) \to \mathcal{G}(u^\dagger)$.
- Can we determine the rate $\mathcal{M}_n \epsilon_n$ such that

$$\mathbb{E}^y \mu^y \left\{ u : \|u - u^\dagger\| \geq \mathcal{M}_n \epsilon_n \right\} \to 0, \quad \forall \mathcal{M}_n \to \infty.$$

Random fields



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



References

W. Karcher, and H-P. Scheffler, E. Spodarev, *Simulation of infinitely divisible random fields*, Communications in Statistics - Simulation and Computation, Volume 42, p215- 246, 2013.

S. Lasanen, *Non-Gaussian statistical inverse problems. Part I: posterior distributions*, Inverse Problems & Imaging,Vol. 6 Issue 2, 2012.

S. Lasanen, *Non-Gaussian statistical inverse problems. Part II: posterior convergence for approximated unknowns*, Inverse Problems & Imaging,Vol. 6 Issue 2, 2012.

G. Samorodnitsky, M. S. Taqqu, *Stable non-Gaussian Random Processes*, Chapman & Hall, New York, 1994.

A. M. Stuart, *Bayesian approach for inverse problems*, Acta Numerica, 2010.