MCMC for heavy-tailed distributions with known characteristic function and unknown density

Leonid Iosipoi

Higher School of Economics
Faculty of Computer Science
iosipoileonid@gmail.com
http://hdilab.ru/

This poster in three lines

If we are interested in computing the expectation of a function g with respect to a density π, i.e.

$$V = \mathsf{E}_\pi[g] = \int_{\mathbb{R}^d} g(x)\pi(x)\,dx,$$

we can use the Plancherel's theorem to represent V as

$$V = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \widehat{g}(-u)\widehat{\pi}(u)\,du$$

and then compute (using MCMC) the following expectation with respect to the density $p(u) \propto |\widehat{\pi}(u)|$

$$V = \frac{C_p}{(2\pi)^d}\mathsf{E}_p\left[\widehat{g}(-X)\frac{\widehat{\pi}(X)}{|\widehat{\pi}(X)|}\right],$$

where C_p is the normalizing constant of the density p.

Why this is interesting?

1. In many applications, especially if one works with heavy-tailed distributions,

   the density π is not known analytically
   but
   the characteristic function $\widehat{\pi}$ is known.

   Typical examples are stable distributions, infinite divisible distributions, and marginal distributions of Lévy and related processes.
2. For heavy-tailed distributions it is always so that

   the characteristic function $\widehat{\pi}$ decays faster than the density π
   (polynomially vs. exponentially).

   The faster is the decay, the "better" is the convergence of MCMC algorithms.

Generalization of this approach

The Plancherel's theorem is valid if $g, \pi \in L^1 \bigcap L^2$. If, for example, the tails of g do not decay and tails of π decay exponentially, one can damp the growth of g by multiplying it by $e^{-\langle x,a\rangle}$ for some vector $a \in \mathbb{R}^d$. Then the formula for V reads as

$$V = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \widehat{g}(ia-u)\widehat{\pi}(u-ia)\,du,$$

If $\widehat{\pi}(u - ia) \neq 0$, then it is, up to a constant, a probability density. Thus one can rewrite V as an expectation with respect to the density $p_a(u) \propto |\widehat{\pi}(ia-u)|$,

$$V = \frac{C_{p_a}}{(2\pi)^d}\mathsf{E}_{p_a}\left[\widehat{g}(ia-X)\frac{\widehat{\pi}(X-ia)}{|\widehat{\pi}(X-ia)|}\right],$$

where C_{p_a} is the normalizing constant of the density $p_a(u)$.

MCMC algorithms in Fourier domain

Algorithm 1: The Metropolis-Hastings algorithm in the Fourier domain

Initialize $X_0 = x_0$;
for $k = 1$ *to* $N + n$ **do**
  Sample $u \sim$ Uniform$[0,1]$ and $Y_k \sim Q(X_{k-1}, \cdot)$;
  if $u < \alpha(X_k, Y_{k+1}) = \min\{1, p_a(Y_{k+1})q(Y_{k+1}, X_k)/p_a(X_k)q(X_k, Y_{k+1})\}$ **then**
    $X_{k+1} = Y_{k+1}$;
  else
    $X_{k+1} = X_k$;
Set $V_{N,n} = \frac{C_{p_a}}{(2\pi)^d}\frac{1}{n}\sum_{k=N+1}^{N+n}\widehat{g}(ia - X_k)\frac{\widehat{\pi}(X_k - ia)}{|\widehat{\pi}(X_k - ia)|}$.

Algorithm 2: The Metropolis-Adjusted Langevin Algorithm in Fourier domain

Initialize $X_0 = x_0$;
for $k = 1$ *to* $N + n$ **do**
  Sample $u \sim$ Uniform$[0,1]$ and $Z_k \sim \mathcal{N}(0,1)$;
  Sample $Y_k = X_k + \gamma_{k+1}\nabla \log p_a(X_k) + \sqrt{2\gamma_{k+1}}Z_{k+1}$;
  if $u < \alpha(X_k, Y_{k+1}) = \min\{1, p_a(Y_{k+1})q(Y_{k+1}, X_k)/p_a(X_k)q(X_k, Y_{k+1})\}$ **then**
    $X_{k+1} = Y_{k+1}$;
  else
    $X_{k+1} = X_k$;
Compute $\Gamma_{N,n} = \sum_{k=N+1}^{n}\gamma_k$;
Put $V_{N,n} = \frac{C_{p_a}}{(2\pi)^d}\sum_{k=N+1}^{N+n}\frac{\gamma_k}{\Gamma_{N,n}}\widehat{g}(iR - X_k)\frac{\widehat{\pi}(X_k - ia)}{|\widehat{\pi}(X_k - ia)|}$.

Example: Elliptically contoured stable distributions

The elliptically contoured multivariate stable distribution is a special symmetric case of the multivariate stable distribution. We say that $X \in \mathbb{R}^d$ is elliptically contoured α-stable random vector if it has joint characteristic function

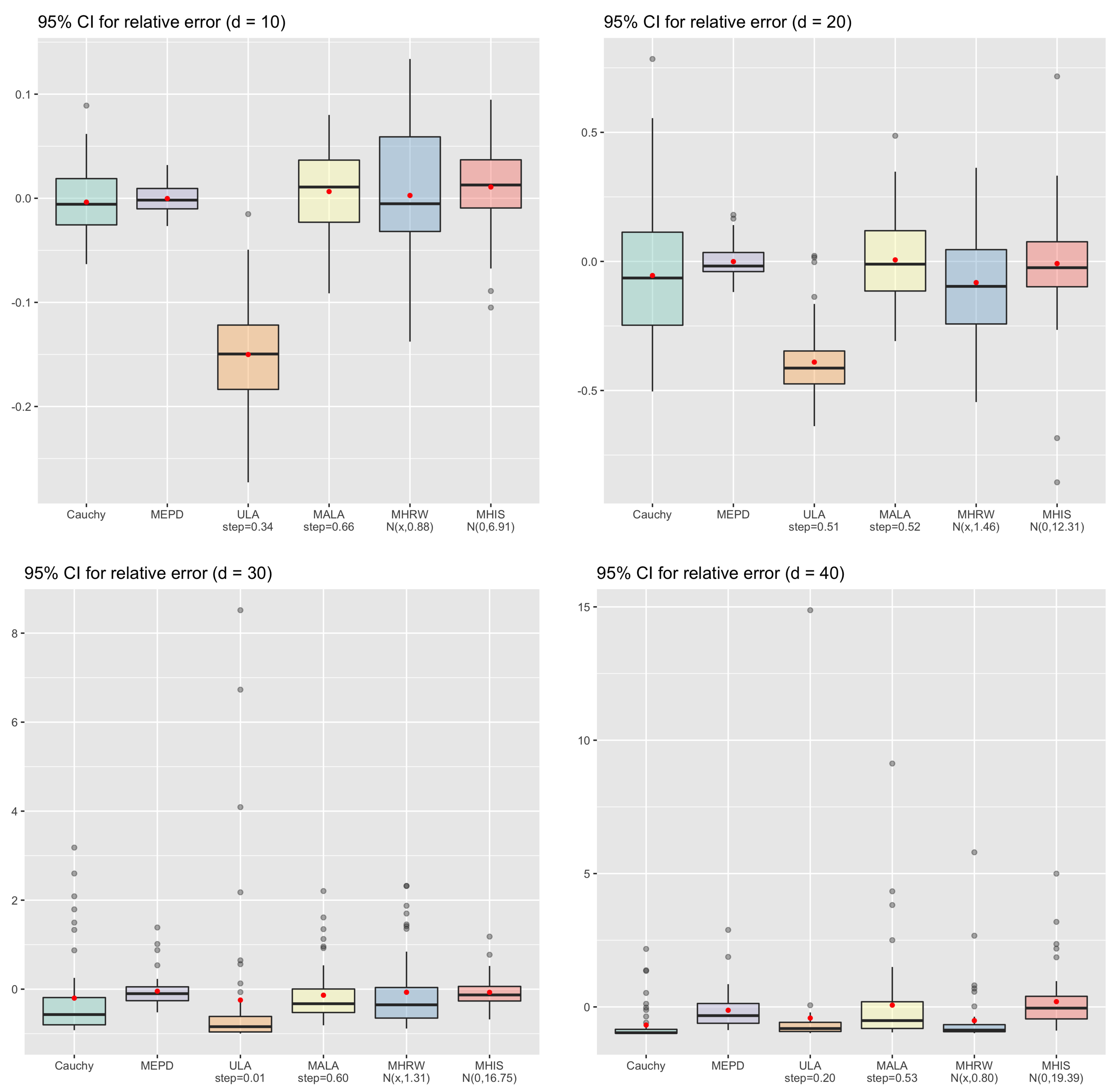
$$\widehat{\pi}(u) = \mathsf{E}_\pi\left[e^{iu^\top X}\right] = \exp\left(-(u^\top\Sigma u)^{\alpha/2} + iu^\top\mu\right),$$

for some $d \times d$ positive semidefinite symmetric matrix Σ and shift vector $\mu \in \mathbb{R}^d$.

Tails of $\widehat{\pi}$ decay exponentially, but tails of the density π decay polynomially

$$\pi(x) \sim |x|^{-(1+\alpha)}.$$

As a function $g(x)$ we take $\operatorname{sech}(t) = 2/(e^t + e^{-t})$ applied to each coordinate. This choice stems from the fact that sech is an eigenfunction for the Fourier Transform operator. Hence we will compute expectation of similar functions in the both Original and Fourier domains. We take $\alpha = 1$ (Cauchy distribution).



- Cauchy — sampling directly from the Cauchy distribution (Original domain)
- MEPD — sampling directly from the characteristic function (Fourier domain)
- ULA — Unadjusted Langevin Algorithm (Fourier domain)
- MALA — Metropolis Adjusted Langevin Algorithm (Fourier domain)
- MHRW — Metropolis-Hastings Random Walk (Fourier domain)
- MHIS — Metropolis-Hastings Independence Sampler (Fourier domain)

Conclusions

- The proposed approach is rather general and can be also used in combination with importance sampling and variance reduction methods.
- Even if there exists a direct sampling method from π, moving to Fourier domain may have its advantages.
- In many applications, there are no direct sampling methods from both π and $\widehat{\pi}$, so MCMC in the Fourier domain is the only way to proceed.

Acknowledgements

The author is greatly indebted to Denis Belomestny for suggesting the problem and for many stimulating conversations.