The Analog Formulation of Sparsity Implies Infinite Divisibility and Rules Out Bernoulli-Gaussian Priors

Arash Amini, Ulugbek S. Kamilov, and Michael Unser
Biomedical Imaging Group (BIG)
École polytechnique fédérale de Lausanne (EPFL)
Email: {arash.amini, ulugbek.kamilov, and michael.unser}@epfl.ch

***Abstract*—Motivated by the analog nature of real-world signals, we investigate continuous-time random processes. For this purpose, we consider the stochastic processes that can be whitened by linear transformations and we show that the distribution of their samples is necessarily infinitely divisible. As a consequence, such a modeling rules out the Bernoulli-Gaussian distribution since we are able to show in this paper that it is not infinitely divisible. In other words, while the Bernoulli-Gaussian distribution is among the most studied priors for modeling sparse signals, it cannot be associated with any continuous-time stochastic process. Instead, we propose to adapt the priors that correspond to the increments of compound Poisson processes, which are both sparse and infinitely divisible.**

I. Introduction

For decades, the Gaussian distributions have been the dominant priors in many fields of research such as signal processing and machine learning. One of the strong arguments in favor of Gaussian models is the Central Limit Theorem (CLT) which states that the sum of a large number of independent and identically distributed (i.i.d.) random variables with finite variance converges to the Gaussian distribution. This is in fact compatible with many physical phenomena where the macroscopic behavior is governed by the interaction of a large number of small constituents. Moreover, the Gaussianity of random vectors is preserved when they are transformed linearly. Besides the physical motivation, Gaussian models have many practical advantages such as succinct parameterization, simplicity of statistical analysis, and links with linear regression.

One of the shortcomings of Gaussian priors is their poor fit for modeling compressible data such as multimedia signals. The existence of sparse (or almost sparse) representations of such signals allows for their efficient compression or acquisition. However, it is known that the compressibility of realizations of Gaussian models is very limited [1]. To accommodate sparsity while still benefiting from the advantages of Gaussian distributions, one might think of Bernoulli-Gaussian priors. The latter distribution is formed by adding a mass probability at point $x = 0$ to the conventional Gaussian density function. This means that random variables with such a distribution take the value 0 with some non-zero probability p, while the non-zero portion (with global probability $(1 - p)$) follows a Gaussian distribution. Thus, an i.i.d. sequence following a Bernoulli-Gaussian distribution will contain a fixed fraction of zero terms, which explains the sparse part of the prior.

This work is supported by the European Research Center (ERC) under ERC-2010-AdG 267439-FUN-SP grant.

In this paper, we investigate the fundamental question of knowing whether a non-Gaussian prior can naturally appear in the observations of a physical phenomenon. Since the observations are limited to a finite, or at most a countable, number of measurements, there are two approaches for the statistical modeling of the observations: 1) the direct adoption of discrete-time stochastic models and 2) the adoption of continuous-time stochastic models for the physical phenomena, along with their discretization to obtain the observation model. In this paper, we choose the second approach, with the extra assumption that the continuous-time process can be whitened with a linear operator. In other words, we assume an innovation-based model for continuous-time physical phenomena where the innovation model is not necessarily Gaussian.

Although the discrete and discretized models look the same at first glance, they are inherently different. In particular, we demonstrate that, unlike the discrete-time models where no restriction applies to the choice of priors, there is a fundamental constraint in the discretized models, namely, infinite-divisibility.

Definition 1. *The distribution of a random variable X is said to be infinitely divisible if, for all $n \in \mathbb{N}^*$, X can be written as the sum of n i.i.d. random variables (not necessarily distributed as X).*

It is known that not every distribution is infinitely divisible; for instance, distributions corresponding to density functions with finite support, such as uniform distributions, are not infinitely divisible [2], [3]. The major contribution of this paper is to show that the Bernoulli-Gaussian distribution is not infinitely divisible and therefore does not appear naturally in the observations of physical phenomena following any continuous-time stochastic model that would admit whitening. However, this issue could be resolved by slightly changing the distribution. For example, Bernoulli-Laplace can be used instead of Bernoulli-Gaussian.

In the following, we first explain the continuous-time stochastic model in Section II. The consequences of the process admitting whitening are described in Section III. We present the negative result regarding Bernoulli-Gaussian priors

978-1-4673-0223-4/12/$31.00 ©2012 IEEE

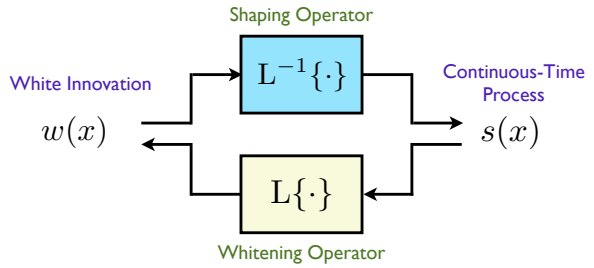

Fig. 1. Innovation-based definition of the stochastic model.

in Section IV, where we also propose alternatives.

II. Continuous-Time Stochastic Model

The stochastic models studied in this paper are continuous-time random processes that can be whitened through linear operators. We denote the stochastic process by s and the associated whitening operator by L. Therefore, $w = \mathrm{L}s$ defines the white innovation process. We further assume that L is either invertible or its kernel is of finite dimension. Thus, L has a linear right inverse L^{-1}, which we call the shaping operator. In cases where L has a nonempty but finite-dimensional null-space (e.g., differential operators), the definition of L^{-1} usually involves boundary conditions. We show in Figure 1 the schematic diagram of the stochastic model.

We adopt Gelfand's approach of generalized stochastic processes [4]. Hence, instead of the conventional point-wise definition of processes, we focus on the statistics of the inner products of the process with test functions. Let φ be a given test function. The inner product of the random process s with the test function φ, which we denote as $X_\varphi = \langle s, \varphi \rangle$, defines a random variable.

Definition 2. *The characteristic form of the process s is defined by*

$$\widehat{\mathscr{P}}_s(\varphi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle s,\varphi\rangle}\}, \tag{1}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator.

The role of the characteristic form $\widehat{\mathscr{P}}_s$ for the random process s is essentially the same as the role of the characteristic function for a random variable. The main difference is that the characteristic form is indexed by the test function rather than the frequency parameter. The domain from which φ is chosen should be a function space, so that linear combinations of test functions are valid test functions. It is common to initially use the Schwartz space of rapidly decreasing functions and extend it to larger sets whenever possible.

Let $\varphi_1, \ldots, \varphi_n$ be a finite number of valid test functions. As a result, any linear combination of them is again a valid test function. Particularly, for scalars $\omega_1, \cdots, \omega_n$, we know that

$$\begin{aligned} \widehat{\mathscr{P}}_s\Big(\sum_{i=1}^{n} \omega_i \varphi_i\Big) &= \mathbb{E}\big\{\mathrm{e}^{\mathrm{j}\langle s, \sum_{i=1}^{n} \omega_i \varphi_i\rangle}\big\} \\ &= \mathbb{E}\big\{\mathrm{e}^{\mathrm{j}\sum_{i=1}^{n} \omega_i X_{\varphi_i}}\big\}. \end{aligned} \tag{2}$$

The expectation in (2) can be written in integration form as

$$\begin{aligned} &\int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{j}\sum_{i=1}^{n} \omega_i x_i} p_{X_{\varphi_1},\ldots,X_{\varphi_n}}(x_1, \ldots, x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n \\ = \;& \mathcal{F}\big\{p_{X_{\varphi_1},\ldots,X_{\varphi_n}}(x_1, \ldots, x_n)\big\}(\omega_1, \ldots, \omega_n), \end{aligned} \tag{3}$$

where $p_{X_{\varphi_1},\ldots,X_{\varphi_n}}$ is the joint probability density function of the random variables $X_{\varphi_1}, \cdots, X_{\varphi_n}$ and $\mathcal{F}\{\cdot\}$ stands for the Fourier transform $\mathcal{F}\{f\}(\omega) = \int_{\mathbb{R}} f(x)\mathrm{e}^{\mathrm{j}x\omega}\mathrm{d}x$. Hence, we observe that the characteristic form implicitly contains all the finite-dimensional probability density functions of the process. In analogy with characteristic functions, characteristic forms are also positive-definite functionals [4].

In the setting shown in Figure 1, we are dealing with two characteristic forms, namely, $\widehat{\mathscr{P}}_w$ and $\widehat{\mathscr{P}}_s$. However, since each process can produce the other one by applying either the shaping or the whitening operator, their statistics are not independent. In [5], [6], and [7] the link between the two characteristic forms is established in terms of the adjoint operators

$$\begin{cases} \langle s, \varphi\rangle &= \langle \mathrm{L}^{-1}w, \varphi\rangle &= \langle w, \mathrm{L}^{-1^*}\varphi\rangle, \\ \langle w, \varphi\rangle &= \langle \mathrm{L}s, \varphi\rangle &= \langle s, \mathrm{L}^{*}\varphi\rangle, \end{cases} \tag{4}$$

where L^{-1^*} and L^* are the adjoint operators of the linear operators L^{-1} and L, respectively. Thus, we have that

$$\begin{cases} \widehat{\mathscr{P}}_s(\varphi) &= \widehat{\mathscr{P}}_w(\mathrm{L}^{-1^*}\varphi), \\ \widehat{\mathscr{P}}_w(\varphi) &= \widehat{\mathscr{P}}_s(\mathrm{L}^{*}\varphi). \end{cases} \tag{5}$$

Equation (5) reveals the fact that we require the characteristic form of only one of s and w to determine the other one. Due to the additional whiteness property of the innovation process, in practice, it is easier to characterize $\widehat{\mathscr{P}}_w$.

III. White Innovation Process

The distinguishing property of the white innovation process is that it is formed of independent and identical infinitesimal contributions. This can be mathematically described as

1) If φ_1 and φ_2 are two test functions with non-overlapping supports (i.e., $\varphi_1(x)\varphi_2(x) \equiv 0$), then, the random variables $\langle w, \varphi_1\rangle$ and $\langle w, \varphi_2\rangle$ are statistically independent.
2) If $\varphi_2(x) = \varphi_1(x - T)$ for some T, which is equivalent to say that φ_1 and φ_2 are equal except for a shift, then the random variables $\langle w, \varphi_1\rangle$ and $\langle w, \varphi_2\rangle$ have identical distributions but are not necessarily independent.

To characterize the white innovation process, we start by considering the test function $\varphi(x) = \mathrm{rect}(x)$, where $\mathrm{rect}(x)$ denotes the unit rectangular causal function taking the value 1 for $x \in [0, 1)$ and 0 otherwise.

Lemma 1. *For the white innovation process w, the distribution of the random variable $\langle w, \mathrm{rect}\rangle$ is infinitely divisible.*

Proof. The rectangular test function is treated separately in our analysis because its refinement equation is

$$\forall\, n \in \mathbb{N}^* : \quad \mathrm{rect}(x) = \sum_{i=0}^{n-1} \mathrm{rect}(nx - i). \tag{6}$$

Hence, we can write

$$\underbrace{\langle w, \mathrm{rect}\rangle}_{X} = \sum_{i=0}^{n-1} \underbrace{\langle w, \mathrm{rect}(n\cdot - i)\rangle}_{X_i}. \quad (7)$$

Note that, for $i \neq j$, the functions $\mathrm{rect}(nx-i)$ and $\mathrm{rect}(nx-j)$ have disjoint supports, while they are equal up to a shift operation. Thus, the random variables X_i for $i = 0, \ldots, n-1$ are independently and identically distributed. In other words, the random variable X can be represented as the sum of n i.i.d. random variables for arbitrary n. ■

Infinite divisibility (id) restricts the choice of the distribution in the sense that there exist non-id distributions. For instance, every stable law, such as the Gaussian distribution, is infinitely divisible, while the binomial and uniform distributions are excluded from the id family.

A. Characterization of Infinitely Divisible Distributions

The concept of infinite divisibility is a well-studied subject in probability theory since late 1920s. Interestingly, a complete characterization of the whole family is known as the Lévy-Khintchine representation theorem [2].

Theorem 1. *The distribution of a random variable X is infinitely divisible if and only if its characteristic function $\hat{p}_X(\omega)$ is of the form $\mathrm{e}^{f(\omega)}$ with*

$$\begin{aligned} f(\omega) &= \mathrm{j}b_1\omega - b_2^2\omega^2 \\ &\quad + \int_{\mathbb{R}\setminus\{0\}} \left(\mathrm{e}^{\mathrm{j}a\omega} - 1 - \mathrm{j}a\omega \mathbb{1}_{|a|<1}\right) V(\mathrm{d}a), \quad (8) \end{aligned}$$

where b_1, b_2 are real-valued constants, the notation $\mathbb{1}_{|a|<1}$ is the indicator function for $|a| < 1$, and V is a positive measure known as Lévy measure that satisfies

$$\int_{\mathbb{R}\setminus\{0\}} \min(1, a^2) V(\mathrm{d}a) < \infty. \quad (9)$$

The function $f(\omega)$ is usually referred to as the Lévy exponent. For $V \equiv 0$, the Lévy exponent would be a concave quadratic function which is the well-known form for the characteristic function of a normal distribution. When V is a non-zero finite measure, the resulting distribution corresponds to the increments of a compound Poisson process.

B. Characteristic Form

Up to this point we have characterized only the inner product of the white innovation process with rectangular functions. Gelfand and Vilenkin consider a natural way of extending this result to other test functions by sequentially approximating them with narrow and non-overlapping rectangular functions (zero-order hold approximation) [4]. This leads to

$$\begin{aligned} \varphi(x) &= \lim_{n\to\infty} \sum_{i\in\mathbb{Z}} \varphi\Big(\frac{i}{n}\Big)\, \mathrm{rect}(nx - i) \\ \Rightarrow \langle w, \varphi\rangle &= \lim_{n\to\infty} \sum_{i\in\mathbb{Z}} \varphi\Big(\frac{i}{n}\Big)\, \langle w, \mathrm{rect}(n\cdot - i)\rangle. \quad (10) \end{aligned}$$

It can be verified that this approach results in

$$\widehat{\mathscr{P}}_w(\varphi) = \exp\left(\int_{\mathbb{R}} f\big(\varphi(\tau)\big)\mathrm{d}\tau\right). \quad (11)$$

The following theorem by Gelfand and Vilenkin guarantees that this extension is acceptable [4]:

Theorem 2. *The functional $\widehat{\mathscr{P}}_w$ in (11) is a valid white characteristic form if and only if f is a valid Lévy exponent .*

C. Infinite Divisibility

We have a complete representation of the white innovation process by (11) and Theorem 2, where the Lévy exponent $f(\omega)$ is defined through (8). We next prove that, for all bounded test functions φ with finite support, the random variable $\langle w, \varphi\rangle$ is infinitely divisible. In other words, we show that the infinite divisibility of Lemma 1 is a general property and is not limited to the case where $\varphi = \mathrm{rect}$.

Theorem 3. *If $X_\varphi = \langle w, \varphi\rangle$, where φ is a bounded test function with finite support, then X_φ is infinitely divisible.*

Proof. By using Theorem 2 and an argument similar to (2), we obtain the characteristic function of X_φ as

$$\hat{p}_{X_\varphi}(\omega) = \widehat{\mathscr{P}}_w(\omega\varphi) = \exp\left(\int_{\mathbb{R}} f\big(\omega\,\varphi(\tau)\big)\mathrm{d}\tau\right). \quad (12)$$

In order to show the infinite divisibility of X_φ, we need to show that $\int_{\mathbb{R}} f\big(\omega\,\varphi(\tau)\big)\mathrm{d}\tau$ can be represented in the same way as (8). First, let us define the modified Lévy measure for X_φ as

$$\bar{V}(\mathrm{d}a) = \int_{\Pi_\varphi} V\left(\frac{\mathrm{d}a}{\varphi(\tau)}\right)\mathrm{d}\tau, \quad (13)$$

where Π_φ denotes the support set of φ. Similar to V, $\bar{V}$ is also a positive measure. Furthermore,

$$\begin{aligned} &\int_{\mathbb{R}\setminus\{0\}} \min(1, a^2)\bar{V}(\mathrm{d}a) \\ &= \int_{\Pi_\varphi}\int_{\mathbb{R}\setminus\{0\}} \min(1, a^2) V\left(\frac{\mathrm{d}a}{\varphi(\tau)}\right)\mathrm{d}\tau \\ &= \int_{\Pi_\varphi}\int_{\mathbb{R}\setminus\{0\}} \min\left(1, \tilde{a}^2\varphi^2(\tau)\right) V(\mathrm{d}\tilde{a})\mathrm{d}\tau \\ &\leq \underbrace{\int_{\mathbb{R}\setminus\{0\}} \min\left(1, \tilde{a}^2\right) V(\mathrm{d}\tilde{a})}_{<\infty} \underbrace{\int_{\Pi_\varphi} \big(1 + \varphi^2(\tau)\big)\mathrm{d}\tau}_{<\infty}. \quad (14) \end{aligned}$$

Hence, $\bar{V}$ is a valid Lévy measure. We still need another simplification before showing the claim. It is given by

$$\begin{aligned} &\int_{\mathbb{R}}\int_{\mathbb{R}\setminus\{0\}} \left(\mathrm{e}^{\mathrm{j}a\omega\varphi(\tau)} - 1 - \mathrm{j}a\omega\varphi(\tau)\mathbb{1}_{|a|<1}\right) V(\mathrm{d}a)\mathrm{d}\tau \\ &\quad = \int_{\mathbb{R}\setminus\{0\}} \left(\mathrm{e}^{\mathrm{j}\bar{a}\omega} - 1 - \mathrm{j}\bar{a}\omega\mathbb{1}_{|\bar{a}|<1}\right)\bar{V}\left(\mathrm{d}\bar{a}\right) \\ &+ \mathrm{j}\omega \underbrace{\left(\int_{\mathbb{R}\setminus\{0\}} \bar{a}\Big(\mathbb{1}_{|\bar{a}|<1} - \int_{\Pi_\varphi} \mathbb{1}_{|\bar{a}|<|\varphi(\tau)|}\mathrm{d}\tau\Big)\bar{V}(\mathrm{d}\bar{a})\right)}_{b^*} \quad (15) \end{aligned}$$

We can finally conclude that

$$\begin{aligned}&\int_{\mathbb{R}} f\big(\omega\varphi(\tau)\big)\mathrm{d}\tau\\ =\ & \mathrm{j}\omega\underbrace{\Big(b^* + b_1\int_{\mathbb{R}}\varphi(\tau)\mathrm{d}\tau\Big)}_{\bar{b}_1} - \mathrm{j}\omega^2\underbrace{b_2^2\int_{\mathbb{R}}\varphi^2(\tau)\mathrm{d}\tau}_{\bar{b}_2^2}\\ &+\int_{\mathbb{R}\setminus\{0\}}\big(\mathrm{e}^{\mathrm{j}\bar{a}\omega}-1-\mathrm{j}\bar{a}\omega\mathbb{1}_{|\bar{a}|<1}\big)\bar{V}(\mathrm{d}\bar{a})\mathrm{d}\tau. \quad (16)\end{aligned}$$

Therefore, the distribution of X_φ belongs to the infinite-divisible family. ■

IV. Revisiting Bernoulli-Gaussian Priors

Let us assume that we have a continuous-time stochastic model from which we obtain linear measurements. More precisely, if s represents the random process of interest, the measurement model is given by $\langle s,\varphi\rangle$, where φ corresponds to the generalized sampling kernel. Assuming the innovation-based model of Figure 1, we can write that

$$\langle s,\varphi\rangle = \langle \mathrm{L}^{-1}w,\varphi\rangle = \langle w,\underbrace{\mathrm{L}^{-1^*}\varphi}_{\bar{\varphi}}\rangle. \quad (17)$$

As a result, the distribution of the linear measurement $\langle w,\bar{\varphi}\rangle$ is infinitely divisible. This is the main distinction between discretizing a continuous-time model and directly assuming a discrete-time model. When L^{-1} (or L^{-1^*}) is a smoothing operator (e.g., an integrator), even a delta function is an admissible test function φ. The resulting $\bar{\varphi}$ would still be bounded.

Next, we show that the popular Bernoulli-Gaussian sparse prior is not infinitely divisible. Thus, it cannot represent the distribution of a discretized continuous-time process.

We make use of Theorem 4 proved in [2] (Theorem 26.1).

Theorem 4. *Let X be an id random variable corresponding to the Lévy measure V. Define*

$$c = \inf\Big\{a>0:\ S_V\subset\{x:\ |x|\le a\}\Big\}, \quad (18)$$

where S_V denotes the support set of the Lévy measure V. The parameter c is permitted to take the values 0 and ∞. Then, for the super-exponential moments we have that

$$\begin{cases} 0<\alpha<\frac{1}{c}: & \mathbb{E}_X\big\{\mathrm{e}^{\alpha|x|\ln|x|}\big\} < \infty,\\ \frac{1}{c}<\alpha: & \mathbb{E}_X\big\{\mathrm{e}^{\alpha|x|\ln|x|}\big\} = \infty.\end{cases} \quad (19)$$

Now consider the Gaussian-Bernoulli prior. Since the tail of the distribution is essentially the same as that of the Gaussian distribution, it decays in proportion to $\mathrm{e}^{-\frac{\sigma^2x^2}{2}}$ for some σ. Thus, the super-exponential moments are all finite, with

$$\forall\ \alpha>0:\quad \mathbb{E}_X\big\{\mathrm{e}^{\alpha|x|\ln|x|}\big\}<\infty. \quad (20)$$

According to Theorem 4, this happens for an id distribution only if $\frac{1}{c}=\infty$, or $c=0$. This implies that the support set of the corresponding Lévy measure is at most limited to the origin. Nevertheless, the origin is always excluded in the integrals involving the Lévy measure. This implies that $V\equiv 0$. However, we have already mentioned that $V\equiv 0$ corresponds to the Gaussian distributions, which excludes the Bernoulli-Gaussian prior. A similar argument shows that all non-Gaussian priors that decay at least as fast as a Gaussian law cannot be infinitely divisible.

The exclusion of the Bernoulli-Gaussian prior from the id family is due to its tail, which decays too fast. It is now interesting to examine whether there exists an infinitely divisible prior with a slower decay rate but still a mass probability at $x=0$. The answer is positive. In fact, the distribution of the increments of any compound Poisson process has a mass probability and is infinitely divisible.

Let A be a random variable with a bounded and symmetric pdf $p_A(a)$, and let $\lambda\ge 0$ be an arbitrary nonnegative real number. Define the id random variable X by the characteristic function (8) with $b_1=b_2=0$ and $V(\mathrm{d}a)=\lambda p_A(a)\mathrm{d}a$. Next, we show that the pdf of X contains a mass probability at $X=0$. First, due to the symmetry of $p_A(a)$ (V, respectively), we have that

$$\begin{aligned}\hat{p}_X(\omega) &= \exp\Big(\lambda\int\big(\mathrm{e}^{\mathrm{j}a\omega}-1\big)p_A(a)\mathrm{d}a\Big)\\ &= \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda\hat{p}_A(\omega)} = \mathrm{e}^{-\lambda}\Big(1+\sum_{k=1}^{\infty}\frac{\lambda^k\hat{p}_A^k(\omega)}{k!}\Big). \quad (21)\end{aligned}$$

Therefore, by applying the inverse Fourier operator on both sides with respect to ω, we obtain

$$p_X(x) = \mathrm{e}^{-\lambda}\delta(x) + \mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}\big(\underbrace{p_A * \cdots * p_A}_{k-\text{times}}\big)(x). \quad (22)$$

It is not hard to verify that the infinite sum in (22) is convergent and finite when p_A is bounded. This confirms that X is an id random variable that has the mass probability $\mathrm{e}^{-\lambda}$ at $x=0$. Due to the finiteness of the Lévy measure V, the random variable X is associated with the increments of a compound Poisson process. The jump distribution of such a process is determined by A.

Remark 1. *The distribution of the increments of a compound Poisson process with Gaussian jumps is not Bernoulli-Gaussian. Although the infinite sum in (22) is composed of weighted Gaussian functions, the decay of the final result is much slower than that of a Gaussian.*

Remark 2. *By setting*

$$p_A(a) = \frac{\exp\big(-q|a|\big)-\exp\big(-q\mathrm{e}^{\frac{\lambda}{2}}|a|\big)}{\lambda|a|}$$

in (21), where q and λ are nonnegative real numbers, one can check that

$$p_X(x) = \mathrm{e}^{-\lambda}\delta(x) + (1-\mathrm{e}^{-\lambda})\frac{q\mathrm{e}^{-q|x|}}{2},$$

which is a Bernoulli-Laplace distribution. In other words, unlike the Bernoulli-Gaussian prior, the Bernoulli-Laplace is infinitely divisible, while it enjoys having a mass probability at the origin.

In summary, the Bernoulli-Laplace prior seems to be an appropriate infinitely divisible substitute for the inappropriate Bernoulli-Gaussian.

V. Conclusion

We adopted a statistical point of view to study the difference between the discrete-time models and the discretized continuous-time models. Although there is almost no restriction in discrete-time models, the probability distributions extracted from the continuous-time models are limited to the family of infinitely divisible distributions. Although this family is still large enough to include most of the conventional models, it excludes many distributions with fast decay. In particular, we showed that the common sparsity prior of Bernoulli-Gaussian is one of the excluded cases. To obtain a more realistic sparse prior, we propose to adapt the decay rate such that the prior becomes infinitely divisible. A simple example would be the Bernoulli-Laplace prior. Although this substitution does not greatly affect the conventional sparsity-based results such as those in compressed sensing, it might tighten the gap between the physical measurements and simulations on synthetic data.

References

[1] A. Amini, M. Unser, and F. Marvasti, "Compressibility of deterministic and random infinite sequences," *IEEE Trans. Sig. Proc.*, vol. 59, no. 11, pp. 5193–5201, Nov. 2011.

[2] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Chapman & Hall, 1994.

[3] F. W. Steutel and K. V. Harn, *Infinite Divisibility of Probability Distributions on the Real Line*. Marcel Dekker Inc., 2004.

[4] I. Gelfand and N. Y. Vilenkin, *Generalized Functions. Vol. 4: Applications of Harmonic Analysis*. New York, USA: Academic Press, 1964.

[5] M. Unser, P. Tafti, and Q. Sun, "A unified formulation of Gaussian *vs.* sparse stochastic processes: Part I—Continuous-domain theory," *arXiv:1108.6150v1*.

[6] M. Unser, P. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian *vs.* sparse stochastic processes: Part II—Discrete-domain theory," *arXiv:1108.6152v1*.

[7] A. Amini, U. Kamilov, and M. Unser, "Bayesian denoising of generalized Poisson processes with finite rate of innovation," in *ICASSP2012*, Kyoto, Japan, Mar. 25–30, 2012.