A Sufficient Statistics Construction of Exponential Family Lévy Measure Densities for Nonparametric Conjugate Models

Robert Finn
Ohio State University CSE Department

Brian Kulis
Ohio State University CSE Department

Abstract

Conjugate pairs of distributions over infinite dimensional spaces are prominent in machine learning, particularly due to the widespread adoption of Bayesian nonparametric methodologies for a host of models and applications. Much of the existing literature in the learning community focuses on processes possessing some form of computationally tractable conjugacy as is the case for the beta process and the gamma process (and, via normalization, the Dirichlet process). For these processes, conjugacy is proved via statistical machinery tailored to the particular model. We seek to address the problem of obtaining a general construction of prior distributions over infinite dimensional spaces possessing distributional properties amenable to conjugacy. Our result is achieved by generalizing Hjort's construction of the beta process via appropriate utilization of sufficient statistics for exponential families.

1 Introduction

Since Raiffa and Schlaifer [1] first formalized the notion of conjugate prior families in 1961, they have repeatedly earned their distinguished role as the key to the operability of Bayesian modeling. Indeed, in both the parametric and nonparametric cases, the ability to perform statistical inference in a computationally efficient manner hinges on the existence of such conjugate prior families.

Although conjugacy is key in both parametric and nonparametric modeling venues, the manners in which conjugacy assumes its role in these two arenas currently exist in unsettled contrast with another. To be sure, in the parametric setting, conjugate families such as the normal gamma, multinomial Dirichlet, and the Bernoulli beta, are familiar and often employed in the task of inference. Furthermore, such families afford the researcher convenient formulae for the updating of posterior parameters as a function of the prior parameters and observed data. Thus, the acceptance and application of conjugate families is received with a unified and uncontroversial disposition. Unfortunately, at the present, no such unification exists for the treatment of conjugate families in the case of nonparametric Bayesian models, a class of models which greatly enrich the researcher's toolkit and as such deserve the effort required to achieve a greater understanding of how to unify the notion of conjugacy in an infinite dimensional setting.

Bayesian nonparametric (BNP) modeling is a prominent and widely used technique in the machine learning community, providing a broad class of statistical models which are more flexible than classical nonparametric models and more robust than both classical and Bayesian parametric models. BNP models such as the Chinese restaurant process, the Indian buffet process, and Dirichlet process mixture models have obtained great success in problem domains such as clustering, dictionary learning, and density estimation, respectively [2–5]. This success is in large part due to the adaptive nature of BNP models. As such, this modeling framework permits the data to determine the level of model complexity rather than entail the specification of the complexity level by the researcher. In order for this adaptive framework to yield a computationally tractable model, one is commonly required to construct a pair of conjugate distributions defined over an infinite dimensional space.

Success in constructing such conjugate pairs in an infinite dimensional setting has been achieved in specific cases producing, for example, the Dirichlet, gamma, and beta processes. In each of these cases, the construction of a suitable likelihood/prior pair and sub-

Appearing in Proceedings of the 18^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2015, San Diego, CA, USA. JMLR: W&CP volume 38.

sequent illustration of the desired form for the posterior yielding conjugacy, is model specific. For example, in the case of the Dirichlet process, conjugacy of the multinomial and Dirichlet processes arises directly from the conjugacy of their marginals. In contrast, conjugacy in the case of the beta and Bernoulli processes is defined and obtained in a seemingly less direct manner, i.e. via the form of the densities for the processes' respective Lévy measures. This modification of the definition of conjugacy from the finite dimensional, or parametric, setting seems to yield a tractable definition relying on a specification of the distributional properties of the infinitesimal increments of the processes under consideration.

While this modified definition provides a framework for conjugacy in a number of cases, e.g. of the gamma process as well, the actual proof techniques employed in establishing this form of conjugacy are conceptually orthogonal to one another in each case and require lengthy, involved, and dense arguments. Obtaining a unified treatment of conjugacy in the infinite dimensional setting via updating formulae for the parameters governing the density of the processes' Lévy measure, analogous to the updating of the parameters of density of the random variable in the finite dimensional case, is a goal which we begin to work towards in this paper. We describe the steps achieved in this paper below.

The construction of families which are conjugate in the modified sense requisite for a nonparametric setting, entails both the construction of a prior with distributional qualities appropriate to the modeling task at hand *and* the ability to derive a conjugate posterior from this amenable form of the prior. With respect to the latter of the two constructions, i.e. the existence of a conjugate posterior, a number of distribution specific techniques have been successful, e.g. in the cases of the gamma and beta processes. A limited number of more general construction techniques either implicitly lurk in the theory of conditional measures over infinite dimensional spaces, or have explicitly been formulated to prove the existence of a conjugate prior, albeit with varying success. For example, recently Orbanz [6] has obtained a mathematical framework for proving the existence of a conjugate posterior over an infinite dimensional space *given* one has in hand an appropriate prior over that space. Prominent cases where such suitable priors exist are related to the class of *exponential families*. In other words, one may apply the Orbanz construction to produce a conjugate posterior *given* one has shown the existence of a prior whose distributional qualities are derived, i.e. appropriately related, to an exponential family. In this paper we address the issue of obtaining a *general* construction for priors over an infinite dimensional space whose distributional properties are determined, in a sense to be made clear in the sequel, by an exponential family. This provides the first of the two constructions required to produce a conjugate pair over an infinite dimensional space.

2 Statistical Preliminaries

We first recall the definition of a completely random measure, and then provide a brief account of their connection to stochastic processes, while also supplying a representation theorem in terms of Poisson processes. In the second part of this section we will gather requisite facts regarding exponential families and their sufficient statistics.

2.1 Completely random measures

A random measure, Φ, is a function whose domain is a measure space $(\Omega, \mathcal{F}, \mu)$ and whose range is a space of measures over a state space (S, Σ), where we take Σ to be a σ-algebra of subsets of S. In other words, for each $\omega \in \Omega$ we have $\Phi(\omega, \cdot)$ is a measure on (S, Σ), and for each fixed $A \in \Sigma$, the function $\Phi(\cdot, A) : \Omega \longrightarrow \mathbf{R}^+$ is an $\mathcal{F}$-measurable function. In this way we may view a random measure Φ as a collection of random variables over Ω indexed by the elements of Σ.

A random measure Φ is said to be a *completely* random measure if for any finite collection $A_1, A_2, \ldots, A_n$ of elements of Σ which are pairwise disjoint, the corresponding random variables $\Phi(\cdot, A_1), \Phi(\cdot, A_2), \ldots, \Phi(\cdot, A_n)$ are independent. To motivate a bit of intuition in this situation, take for example $S = [0, +\infty)$ and Σ to be the collection of Borel measurable subsets of S. Then the condition of complete randomness implies for any $t_1 < t_2 < \ldots < t_n$ we have $\Phi(\cdot, (t_1, t_2]), \Phi(\cdot, (t_2, t_3]), \ldots, \Phi(\cdot, (t_{n-1}, t_n])$ are independent. This is reminiscent of the situation where $X(t)$ is a nondecreasing stochastic process with independent increments in the sense that for any $t_1 < t_2 < \ldots < t_n$ we have $X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$ are independent, and Φ is the unique Borel measure for which $\Phi((a, b]) = X(b+) - X(a+)$ for all $a < b$. As particular examples of this definition, both the beta and the gamma processes satisfy the conditions of a completely random measure.

Kingman showed [7, 8] that any completely random measure Φ has a decomposition into independent completely random measures of the form $\Phi = \Phi_f + \Phi_d + \Phi_o$, where Φ_f corresponds to the fixed atoms of Φ, Φ_d is the deterministic component of Φ, and Φ_o is a purely atomic measure. In general, it is the measure Φ_o that is of interest. One can show that for any $A \in \Sigma$ the random variable $\Phi_o(\cdot, A)$ is infinitely divisible in the sense that for any n there exists a decomposition of A into pairwise disjoint sets $A_1, A_2, \ldots, A_n \in \Sigma$ such that

$$\mathbb{E}[\exp(-\Phi_o(A_i))] = \{\mathbb{E}[\exp(-\Phi_o(A))]\}^{\frac{1}{n}} \quad i = 1, \ldots, n.$$

This property implies that the transform $\mathbb{E}[e^{-t\Phi_o(A)}]$ has the form

$$\mathbb{E}[e^{-t\Phi_o(A)}] = \exp\left(- \int_{A\times(0,\infty]} (1 - e^{-ts})\nu(dx, ds)\right).$$

The measure ν is referred to as the Lévy measure [9] of the random variable $\Phi_o(A)$ and is of great importance as it determines the random variable $\Phi_o(\cdot, A)$.

Kingman proved the Lévy measure ν and the completely random measure Φ_o have parallel decompositions of the form

$$\nu = \sum_n \nu_n, \qquad \Phi_0 = \sum_n \Phi_n,$$

$$\text{where for each } n, \quad \Phi_n = \sum_{(s,\phi(s))\in\Pi_n} \phi(s)\delta_s.$$

Here, for each index n, Π_n is a Poisson process on $S \times (0, \infty]$, and $s \in S$ is an atom of weight $\phi(s) \in (0, \infty]$. It is precisely this decomposition derived from the Lévy measure that permits simulation of the completely random measure via simulation of the Poisson processes which constitute the decomposition, where each ν_n is used as the mean measure of the Poisson process Π_n. The decomposition which allows one to view the completely random measure as a composition of atomic measures defined by Poisson processes is illustrated in Figure 1. Wang and Carin gave a detailed analysis of this decomposition and subsequent simulation techniques for the Lévy measures arising from the beta and gamma processes [10].

2.2 Exponential families

We now turn to a brief accounting of exponential families and properties of their sufficient statistics. Standard references for material on exponential families are [11] and [12]. To begin, a family $\{P_\theta\}_{\theta\in\Theta}$ of distributions over a probability space $(\Omega, \mathcal{F})$ is said to constitute an n-dimensional exponential family if the distributions have densities of the form $p_\theta(x) = h(x)e^{(\langle\eta(\theta),T(x)\rangle - B(\theta))}\mu(dx)$ with respect to some common measure μ. In the above, $\eta(\theta) = (\eta_1(\theta), \ldots, \eta_n(\theta))$, and η_i and B are real-valued functions of the parameter θ. In addition, $T(x) = (T_1(x), \ldots, T_n(x))$, the T_i are real-valued statistics where x is in the support of the density, and $\langle\eta, T(x)\rangle$ denotes the usual inner product of η and $T(x)$. While

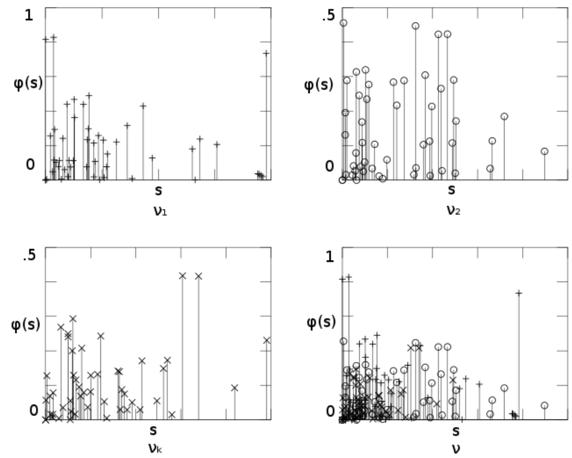

Figure 1: A Lévy measure ν decomposed into $\{\nu_n\}_{n=1}^{\infty}$. Each measure ν_n is constructed from a Poisson process, Π_n, on $S \times (0, \infty]$ yielding an atomic measure with with weights $\phi(s)$ assigned to the point mass at s where $(s, \phi(s)) \in \Pi_n$. The Lévy measure ν is the superposition of the measures $\nu_1, \nu_2, \ldots, \nu_k, \ldots$.

this is the formal definition of an exponential family, the form commonly used is obtained by employing the η_i, $i = 1 \ldots n$, as the parameters and writing the density in what is known as the *canonical form* $p_\theta(x|\eta) = h(x)e^{(\langle\eta,T(x)\rangle - A(\eta))}\mu(dx)$. The integrand of the density in its canonical form is a positive function and will yield a bona fide probability distribution if and only if

$$\int h(x)e^{(\langle\eta,T(x)\rangle - A(\eta))}\mu(dx) = 1,$$

which is equivalent to

$$\int h(x)e^{\langle\eta,T(x)\rangle}\mu(dx) = e^{A(\eta)} < +\infty.$$

The collection of all such η for which this holds is a convex set called the *natural parameter space* and is denoted by Ξ. Many common distributions belong to exponential families; for example, the normal, beta, and gamma distributions are all members of the exponential family of distributions.

The real valued functions, i.e. the statistics, $T_i(x)$, $i = 1 \ldots n$ appearing in the expression for the densities of an exponential family posses a pleasant property known as *sufficiency.* A statistic T for a random observation X is said to be *sufficient* for the family $\{P_\theta\}$ of possible distributions for X if the conditional expectation of X given $T = t$ is independent of θ for all t. The property of sufficiency has numerous and varied consequences which, although of great importance in many branches of statistics, only one of which will concern us in this paper. This property allows the computation of moments of T, a variant of which will supply a crucial step in the proof of our extension of Hjort's result. This well-known result relates the moments of T to the partial derivatives of the normalizing factor $\exp(A(\eta))$ with respect to the natural parameters. In fact, it can be shown that for any k where $1 \leq k \leq n$ and any $m \geq 1$ we have [11]

$$\mathbb{E}[T_k^m] = \int T_k^m(x)(h(x)\exp\big(\langle\eta, T(x)\rangle - A(\eta)\big)\mu(dx) \\ = e^{-A(\eta)}\frac{\partial^m[e^{A(\eta)}]}{\partial\eta_k^m}. \tag{1}$$

We require a variant of this result which replaces the natural parameter η by a function $\eta(z)$ defined on $(0, \infty)$. Thus, rather than taking the natural parameters to freely vary we require them to be in the range of the function $\eta(z)$. Given this modification, the variant of (1) involves functional derivatives, a common device from variational calculus. The statement of the result, whose proof appears in the supplemental material, is as follows:

Lemma 1 *Let $\eta(z) = (\eta_1(z), \ldots, \eta_n(z))$ be a piecewise continuous function on $(0, \infty)$ such that $\eta((0, \infty)) \subseteq \Xi$, where Ξ is the natural parameter space of the exponential family given by*

$$p(x|\eta)_\theta = h(x)\exp\big(\langle\eta, T(x)\rangle - A(\eta)\big)\mu(dx).$$

Then

1. *the functional m^{th}-derivative with respect to $\eta_k(z)$, $1 \leq k \leq n$*

   $$\frac{\partial^m}{\partial(\eta_k(z))^m}\left[\int h(x)\exp\big(\langle\eta(z), T(x)\rangle - A(\eta(z))\big)\mu(dx)\right]$$

   exists if and only if

   $$\int\left\{h(x)\frac{\partial^m}{\partial(\eta_k(z))^m}\left[\exp\big(\langle\eta(z), T(x)\rangle - A(\eta(z))\big)\right]\right\}\mu(dx)$$

   exists, and the two quantities are equal;

2. *if the quantities in 1 exist, then the m^{th} moment of $T_k(x)$ exists and*

   $$\mathbb{E}[T_k^m] = e^{-A(\eta(z))}\frac{\partial^m}{\partial(\eta_k(z))^m}\big[e^{A(\eta(z))}\big]. \tag{2}$$

3 Sufficient statistics construction of a Lévy measure

We begin part three of this paper with a brief discussion of our motivation for presenting the central construction, and in particular the role we intend this construction to take as the first of two components necessary to extend the researcher's palette of nonparametric conjugate models. Following this, in the first of the two remaining sections we discuss Hjort's construction of the beta process. The conditions under which the construction is implemented and an essential step in his proof are explained. Our modifications of these which allow the construction to produce processes whose infinitesimal increments are distributed according to an exponential family are provided. In the final section we state and interpret our main result.

3.1 Conjugacy and densities

To begin, we remind the reader of the brief discussion contained in the introduction regarding the forms conjugacy assumes in the parametric and nonparametric cases. In the former, conjugacy is well known and accepted as the condition that the density of the posterior is *of the same form* as the density of the prior, *and* the parameters for the posterior are obtained as a function of the parameters for the prior and the sampled data. In contrast, for the nonparametric case, the form conjugacy now commonly assumes applies not to the densities of the random variables under consideration, but to the densities of the Lévy measures associated with the processes.

For example, the beta process is taken to be a stochastic process whose infinitesimal increments are, for a given base measure μ and concentration function $c(\omega)$, $\mathrm{BP}(c(d\omega)\mu(\omega), c(\omega)(1-\mu(\omega))$ distributed. This somewhat imprecise definition of the beta process is formalized via the fact that the associated Lévy measure of the process has the form $\nu(d\pi, d\omega) = c(\omega)\pi^{-1}(1 - \pi^{c(\omega)-1})d\pi\mu(d\omega)$, a form which is a degenerate beta density. To obtain the data generating process for a conjugate beta process model, one invokes the well known conjugacy of the beta and Bernoulli distributions from the parametric case, and then pushes the analogy to the infinite dimensional case. In fact, letting X be a Bernoulli process with base measure B, denoted as $X \sim \mathrm{Be}(B)$, one can show that if $B \sim \mathrm{BP}(c, B_0)$ and $X_1|B, X_2|B, \ldots, X_n|B \sim \mathrm{Be}(B)$ are independent observations, then

$$B|X_1, \ldots, X_n \sim \mathrm{BP}\left(c+n, \frac{c}{c+n}B_0 + \frac{1}{c+n}\sum X_i\right)$$

This result, while pointed out in [13], derives its proof from a result of [14] which in fact proves that the density of the Lévy measure associated with the posterior maintains the form of a beta density with parameters derived from the prior and the observed data.

Thus, in both the parametric and nonparametric cases, the active definition of conjugacy is one of a condition on the relationship between densities associated with the prior and posterior. In the parametric case the definition involves the densities of the random variables, and in the nonparametric case the definition involves the densities of the Lévy measures associated with the stochastic processes.

It is precisely this observation which drives us to consider extending Hjort's construction to encompass stochastic processes whose infinitesimal increments are distributed according to a positive exponential family. Or more precisely, to stochastic processes whose associated Lévy measures have densities from positive exponential families. This choice of density class takes aim directly at the problem which concerns the second part of our program, namely demonstrating that the processes yielded by our construction in this paper, do in fact, under appropriate regularity conditions, serve as *infinite dimensional conjugate priors*. As such, the second part of our program will establish that the associated infinite dimensional posteriors do in fact belong to the same class as the priors and provide analytical formulae for their computation. Establishing this fact will complete a program which defines a theory of conjugacy for positive exponential families in infinite dimensional spaces analogous to the current theory of conjugacy for exponential families in a finite dimensional setting. While space considerations for this paper will not allow a detailed accounting of this program, we provide some comments along this direction below.

Key to such a program is the ability to invoke well known conjugacy relationships from a parametric setting, and then, as in the case of the beta and Bernoulli processes mentioned above, analogously push the conjugacy to the infinite dimensional setting. In passing the conjugacy relationship from the finite dimensional setting to the infinite dimensional setting, one in fact transfers the conjugacy condition on densities of random variables to densities of Lévy measures of stochastic processes. A number of results concerning explicit relationships between these two densities have been demonstrated [15], and recent work has produced such results readily applicable to the case of densities from exponential families [16–18]. Theorem 1 of this paper, presented in section 3.3, explicitly constructs a stochastic process whose associated Lévy measure possesses a density derived from an exponential family. As such, the processes yielded by our construction are firmly positioned to take advantage of the above mentioned results in order to prove the existence of their conjugate posteriors.

3.2 Hjort's construction of the improper beta process

Hjort [14] provides an explicit construction proving the existence of a process $B(t)$ on $[0, \infty)$ such that $B(0) = 0$ and $B(t)$ possesses independent increments which are infinitesimally beta distributed. In addition, Hjort requires the condition that the sample paths of $(1 - e^{-B(t)})$ are all cumulative distribution functions. The construction is given relative to two fixed objects: 1. a nondecreasing, right continuous function, $A_0(t)$ on $[0, \infty)$ with $A_0(0) = 0$ and the quantity $(1 - e^{-A_0(t)})$ yielding a cumulative distribution function on $[0, \infty)$, and 2. a piecewise continuous function $c(z)$ on $(0, \infty)$. The properties required of the function $A_0(t)$ ensured the resulting process $(1 - e^{-B(t)})$ would have sample paths that were cumulative distribution functions. The function $c(z)$, which is termed the *concentration function*, in part determines the beta distribution of the increments of $B(t)$.

In the course of his construction, Hjort proposes a form for the Lévy measure for $B(t)$. Proving the correctness of the form for the Lévy measure relies heavily on the ability to find a convenient closed form for all moments of a beta distributed random variable. Fortunately it is known that for $X \sim \text{Beta}(\alpha, \beta)$ we have for all $m \geq 1$

$$\mathbb{E}[X^m] = \frac{\Gamma(\alpha + m)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + m)\Gamma(\alpha)}.$$

When attempting to generalize Hjort's construction to prove the existence of a process $X(t)$ on $[0, \infty)$ such that $X(0) = 0$ and X possesses independent increments which are distributed according to an exponential family, the lack of a closed form for the moments of a random variable distributed according to a general exponential family imposes a significant obstacle. However, equation (2) provides a formula for the moments of the sufficient statistics $T_k(x)$ relative to a density which is a modification of the density of X. It is for this reason that our extension of Hjort's construction uses a sufficient statistic $T_k(x)$ of X rather than X itself.

Additionally, we do not require the fixed function $A_0(t)$ to have the property that $(1 - e^{-A_0(t)})$ yields a cumulative distribution function on $[0, \infty)$. We only require that $A_0(t)$ corresponds to a unique Lebesgue-Stieltjes measure on $[0, \infty)$. Finally, we replace the piecewise continuous function $c(z)$ with the vector of piecewise continuous functions $(\eta_1(z), \ldots, \eta_n(z))$.

3.3 Sufficient statistics construction for positive valued exponential families

We now state our main result, the proof of which is provided in the supplementary material. In the light of the discussion of Hjort's construction above, the following paragraph provides few essential comments regarding the construction obtained in Theorem 1 and the objects it produces.

For a given m-dimensional exponential family, $p_\theta(x|\eta)$, defined on $[0,\infty)$, one has, in our case, the function $\eta(z) = (\eta_1(z),\ldots,\eta_m(z))$ supplying the natural parameters, and the sufficient statistics $T(s) = (T_1(s),\ldots,T_m(s))$. Our construction employs only *one* of the sufficient statistics, say $T_k(s)$ for a fixed k where $1 \le k \le m$. As long as the chosen $T_k(s)$ and the given function $\eta(z)$ satisfy the regularity conditions in the theorem, then for any positive, increasing $A_0(z)$ on $[0,\infty)$ which is right continuous with left hand limits, the construction of the theorem results in a Lévy process $T(t)$ and a corresponding Lévy measure given by the specified form. Indeed, the change of variables formula expressed in equation (3) in the theorem furnishes the machinery to accomplish this task. That this change of variables permits such a construction is shown in the example 4.3.

Theorem 1 *Let $T_k(x)$ be a sufficient statistic of an m-dimensional exponential family $p_\theta(x|\eta)$, $\eta(z) = (\eta_1(z),\ldots,\eta_m(z))$ a vector of piecewise continuous, nonnegative functions on $(0,\infty)$, and $A_0(z)$ a positive, increasing function on $[0,\infty)$ which is right continuous with left hand limits. Assume the following conditions hold:*

1. *$T_k^{-1}(u)$ exists and is differentiable;*
2. *for all $z \in (0,\infty)$ we have $\eta(z) \in \Xi$, the natural parameter space of $p_\theta(x|\eta)$;*
3. *if $(\eta_1(z),\ldots,\eta_m(z)) \in \Xi$ then for every $0 < \varepsilon < 1$ it follows that $(\eta_1(z),\ldots,\varepsilon\eta_k(z),\ldots,\eta_m(z)) \in \Xi$.*

Then there exists a Lévy process $T(t)$ with a Lévy representation given by

$$\mathbb{E}[\exp(-\theta T)] = \exp\left\{-\int (1-e^{-\theta u})dL_t(u)\right\} =$$

$$\exp\left\{-\int (1-e^{-\theta T_k(s)})dL_t(s)\right\},\ \text{where}$$

$$dL_t(u) = \left\{\int_0^t \exp\left(\langle \eta(z), U\rangle - A(\eta(z))\right)\frac{dT_k^{-1}}{du}dA_0(z)\right\}du \tag{3}$$

and

$$dL_t(s) = \left\{\int_0^t \exp\left(\langle \eta(z), T(s)\rangle - A(\eta(z))\right)dA_0(z)\right\}ds. \tag{4}$$

where $U = (T_1(T_k^{-1}(u)),...,T_m(T_k^{-1}(u)))$ and $u = T_k(s)$.

Condition 1 is required so that an explicit form of the density function of T_k can be found, which in turn permits the computation of the the transform $\mathbb{E}[\exp(-\theta T_k)]$ linking the random variable T_k to the Lévy measure

$$\nu(du, dA_0(z)) = \exp\left(\langle \eta(z), u\rangle - A(\eta(z))\right)\frac{dT_k^{-1}}{du}dA_0(z)du.$$

As noted in section 2.1 this linkage between T_k and $\nu(du, dA_0(z))$ completely establishes the distributional properties of the random variable via a Poisson process with intensity measure given by the Lévy measure.

Condition 2 of the theorem simply requires that $\eta(z)$ does in fact determine a well defined exponential family. Condition 3 is a technical requirement which is directly tied to the construction of the Lévy process $T(t)$. The condition, loosely interpreted, states that the natural parameter space is closed under *contraction towards* 0, i.e. if one takes any point in the natural parameter space, and shrinks it in absolute value by an amount ε, then the resulting value is still in the natural parameter space. Note that this is not quite equivalent to the well known property of convexity of the natural parameter space [12], as the element 0 need not be in the space. This is precisely the case for a beta distributed random variable. Finally, we note that if there are multiple sufficient statistics satisfying condition 1 of the theorem, then the Lévy measures resulting from different choices of $T_k(s)$ will all be absolutely continuous with respect to one another, i.e. all measures will be equivalent.

4 Examples

In this section we demonstrate that the Lévy measures for the beta and gamma processes are obtainable from the Lévy measure representation derived in Theorem 1. In addition, we compute the Lévy measure for a process with infinitesimally Pareto distributed increments.

The methodology employed in the examples which follow stands in contrast to the Lévy measure decomposition procedure demonstrated in [10]. This decomposition procedure requires one to have the Lévy measure for the completely random measure in hand, and then after examining the particular form of the measure, apply a number of series expansions and identities particular to the moments of the process to arrive at the decomposition. Thus, starting with a completely random measure, the researcher arrives at a decomposition which, while permitting simulation of the process, does not allow the researcher to specify the Lévy measures which are used to generate the simulation. Wang and Carin acknowledge this in [10] by noting that all that can be said about the components of the resulting decomposition are that they are Lévy processes.

In comparison, employing the construction in Theorem 1, our procedure allows the researcher to specify *a priori* the specific forms of the Lévy measures, prove the existence of the corresponding processes, and then arrive at the desired completely random measure via an infinite sum.

4.1 The beta process

In Wang and Carin [10] the beta process B with corresponding Lévy measure given by $\nu(ds,dz) = c(z)s^{-1}(1-s)^{c(z)-1}dsd\mu(z)$ is decomposed into infinite sums $B=\sum_n B_n$ and $\nu=\sum_n \nu_n$ where the B_n are Lévy processes with Lévy measures $\nu_n(ds,dz)$, given by

$$\nu_n(ds,dz) = \text{Beta}(1, c(z)+n)ds\frac{c(z)}{c(z)+n}d\mu(z). \quad (5)$$

The function $c(z)$ in the above is assumed to be a piecewise continuous positive function on $(0,\infty)$. The decomposition of the Lévy measure ν into an infinite series with components given by (5) is achieved by simply writing the s^{-1} term in ν as an infinite series and then distributing the remaining terms, as well as employing identities for the gamma function.

Rather than beginning with the measure ν and then subsequently arriving at the form of the decomposition, the Lévy measures in (5) can be constructed directly from Theorem 1, hence we will verify that the conditions of Theorem 1 are satisfied.

First, the sufficient statistics for the beta distribution are $T_1(x) = \ln(x)$ and $T_2(x) = \ln(1-x)$ and both of these functions has an infinitely differentiable inverse, hence condition 1 of the theorem is satisfied. Second, defining $\eta(z) = (1, c(z)+n)$ yields a function satisfying condition 2 of the theorem. Finally, since the natural parameter space Ξ of the beta distribution consists of all (α,β) where $\alpha,\beta > 0$ we see that for any $\varepsilon > 0$ if $(\alpha,\beta)\in\Xi$ then both $(\varepsilon\alpha,\beta)$ and $(\alpha,\varepsilon\beta)$ are in Ξ. Thus condition 3 is satisfied. As all the requisite conditions are satisfied, we can construct processes B_n for $n = 1,\ldots$ whose Lévy measures have the form in (5). This is achievable by Theorem 1 as each ν_n is of the form in (3) with $A_0(z) = (c(z)/(c(z)+n))F(z)$, where $F(z)$ is the cumulative distribution of μ, resulting in an $A_0(z)$ which is a positive, increasing function which is right continuous with left hand limits.

4.2 The gamma process

Similar to the decomposition of the Lévy measure for the beta process, in [10] a decomposition for the gamma process Lévy measure

$$\nu(ds,dz) = s^{-1}\exp\left(\frac{-s}{c(z)}\right)dsd\mu(z) \quad (6)$$

is obtained. In this case the Lévy measure is decomposed into a doubly indexed infinite sum of Lévy measures $\nu_{k,h}(ds,dz)$, where for every k,h

$$\nu_{k,h}(ds,dz) = \text{Gamma}\left(h, \frac{c(z)}{k+1}\right)ds\frac{d\alpha(z)}{(k+1)^h h}. \quad (7)$$

The decomposition of the Lévy measure in (6) into a doubly indexed infinite series with components given by (7) is achieved not by a simple series expansion of a term in (6), but rather by an iterative procedure of rewriting the exponential term in the measure as a product of terms, one of which may be series expanded. This step decomposes the Lévy measure into a sum of two measures and repeated application of this process produces the doubly indexed infinite series of measures.

Again, we may avoid decomposition techniques particular to a given Lévy measure, and instead construct each $\nu_{k,h}$ directly from Theorem 1. We see that each $\nu_{k,h}$ is of the form in (3) with

$$\eta(z) = \left(h, \frac{c(z)}{k+1}\right) \quad \text{and} \quad A_0(z) = \frac{d\alpha(z)}{(k+1)^h h}F(z),$$

where $c(z)$ is a positive piecewise continuous function, and $F(z)$ is the cumulative distribution of α. Noting that the sufficient statistics for the gamma distribution are $T_1(x) = \ln(x)$ and $T_2(x) = x$,both of which have an infinitely differentiable inverse, condition 1 of Theorem 1 is satisfied. Again, similar to the case for the beta distribution, the natural parameter space Ξ of the gamma distribution consists of all (α,β) where $\alpha,\beta > 0$. From this it follows that conditions 2 and 3 of Theorem 1 are also satisfied.

4.3 A completely random measure based on the Pareto distribution

We illustrate the construction of a completely random measure whose infinitesimal increments are *Pareto distributed*. We will demonstrate the procedure using the change of variables formula from Theorem 1 explicitly.

To begin consider the density function for a Pareto distributed random variable $p(u|\alpha) = \frac{\alpha u_m^\alpha}{u^{\alpha+1}}$. In this form $\alpha > 0$ is the shape parameter, u_m is the scale parameter, and the support of $p(u|\alpha)$ is $[u_m, \infty)$. Writing $p(u|\alpha)$ in canonical form yields $p(u|\alpha) = \exp\big(-(\alpha+1)\ln(u) - (-\ln(\alpha) - \alpha\ln(u_m)\big)$. From this we see that there is one natural parameter, $-(\alpha+1)$, and the corresponding sufficient statistic is $T(u) = \ln(u)$. Our goal is to choose an initial exponential family, apply the construction in Theorem 1, and then arrive, after the change of variable, at a form for the density of the Lévy measure in (3) which conforms to the density of a Pareto distribution. Thus, we consider the exponential family with sufficient statistics $T_1(x) = \ln(x), T_2(x) = \ln(\ln(x))$, and natural parameters $\eta_1 = -1, \eta_2 = -(\alpha+1)$. We note that on $(1, +\infty)$ both sufficient statistics have a well defined inverse which is differentiable. Second, integration by parts on the integrand $x^{-1}(\ln(x))^{-(\alpha+1)}$ shows that (η_1, η_2) is in the natural parameter space of $p(x|\eta)$. Similarly, we also see that the natural parameter space is closed under contraction towards 0. Hence we may apply the construction of Theorem 1 to produce a Lévy process T so that the density of the corresponding Lévy measure, after performing the change of variables to express the Lévy measure as in equation (3) of the theorem, has the form

$$\exp\big\{-(\alpha+1)\ln(u) - u - (-\ln(\alpha) - \alpha\ln(u_m)\big\}e^u = \frac{\alpha u_m^\alpha}{u^{\alpha+1}}$$

which is a Pareto density.

More generally, taking $\alpha = \alpha(z)$, a positive piecewise continuous function on $(0,\infty)$, we see that conditions 2 and 3 of Theorem 1 are satisfied. Likewise, the sufficient statistic $T(x)$ chosen in the previous paragraph satisfies condition 1. Thus, there exists a Lévy process whose Lévy measure is given by $\alpha(z)u_m^{\alpha(z)}u^{-(\alpha(z)+1)}dA_0(z)du$ where $A_0(z)$ is any function satisfying the conditions of Theorem 1.

Note that, analogously to the previous two examples, we may consider the case where we wish to choose specific forms for $\alpha(z)$ for all $n = 1, 2, \ldots,$ yielding $\{\alpha_n(z)\}_{n=1}^\infty$. Applying the above construction to each $\alpha_n(z)$ produces $\nu_n(dz, du) = \alpha_n(z)u_m^{\alpha_n(z)}u^{-(\alpha_n(z)+1)}dA_0(z)du$ which are the Lévy measures acting as the intensity parameters for the countably many Poisson processes used to simulate the completely random measure whose corresponding Lévy measure is given by

$$\sum_{n=1}^{\infty} \alpha(z)_n u_m^{\alpha_n(z)} u^{-(\alpha(z)+1)} dA_0(z)du.$$

Fixing an $\alpha(z)$ so that defining $dA_{0,n} = \frac{1}{n\alpha(z)}dz$ for $n = 1, \ldots,$ all $A_{0,n}(z)$ satisfy the condition of Theorem 1, the above expression becomes

$$\sum_{n=1}^{\infty} n\alpha(z)u_m^{n\alpha(z)}u^{-(\alpha(z)+1)}\frac{1}{n\alpha(z)}dzdu = \left(1 + \frac{u_m^{\alpha(z)}}{u^{-(\alpha(z)+1)} - u_m^{\alpha(z)}}\right)dzdu.$$

Algorithm 1 Sampling algorithm for a CRM with Lévy measure Pareto density

Input: N, $A_{0,n}(z)$, $\alpha(z)_n x_m^{\alpha_n(z)} u^{(-\alpha_n(z)+1)}$ for $n = 1 \ldots N$
  for $n = 1 \to N$ **do**
    $m_n \leftarrow \text{Poisson}(\int A_{0,n}(dz))$
    for $j = 1 \to m_n$ **do**
      $z_{j,n} \xleftarrow{i.i.d.} \frac{A_{0,n}(dz)}{\int A_{0,n}(dz)}$
      $u_{j,n} \xleftarrow{i.i.d.} \alpha(z)_n u_m^{\alpha_n(z)} u^{(-\alpha_n(z)+1)}$
    end for
  end for
  return $\bigcup_{n=1}^N \{(z_{j,n}, u_{j,n})\}_{j=1}^{m_n}$

Thus, we have proved the existence of a completely random measure whose Lévy measure decomposition consists of the Pareto densities of our choice and whose composition of Lévy measures is obtained by the closed form expression

$$dL_t(u) = \left\{\int_0^t \left(1 + \frac{u_m^{\alpha(z)}}{u^{-(\alpha(z)+1)} - u_m^{\alpha(z)}}\right)dz\right\}du.$$

Employing the Lévy measures $\{\nu_n\}_{n=1}^N$, where N is a chosen level of truncation, simulation of this completely random measure may be achieved by Algorithm 1, analogous to the sampling algorithms based on decompositions of the beta and gamma processes.

5 Conclusion and future work

Through our generalization of Hjort's construction via sufficient statistics we have addressed the problem of obtaining a general construction of prior distributions over infinite dimensional spaces possessing distributional properties amenable to conjugacy. Thus, we have completed the first part in a two stage program which aims to define a theory of conjugacy for positive exponential families in infinite dimensional spaces analogous to the current theory of conjugacy for exponential families in a finite dimensional setting. A forthcoming article will target the second stage of this theory and as such will delineate appropriate conditions on the Lévy measures associated with the processes constructed in this paper that will allow the construction of their conjugate priors.

References

[1] Howard Raiffa and Robert Schlaifer. *Applied Statistical Decision Theory.* Division of Research, Harvard Business School, Boston, 1961.

[2] D. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII*, 1983.

[3] Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230, 1973.

[4] T.L. Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems 18*, 2005.

[5] Dahua Lin, Eric Grimson, and John Fisher. Construction of dependent Dirichlet processes based on Poisson processes. In *Advances in Neural Information Processing Systems 23*, 2010.

[6] P. Orbanz. Construction of nonparametric Bayesian models from parametric Bayes equations. In *Advances in Neural Information Processing Systems 22*, 2009.

[7] J.F.C. Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21:59 – 78, 1967.

[8] J.F.C. Kingman. *Poisson Processes.* Oxford University Press, Oxford, 1993.

[9] P. Lévy. *Théorie de L'Addition des Variables Aléatoires.* Gauthier-Villars, Paris, 1937.

[10] Yingjian Wang and Lawrence Carin. Lévy measure decompositions for the beta and gamma processes. In *Proceedings of the 29 th International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012.

[11] E.L. Lehmann and George Casella. *Theory of Point Estimation, Second Edition.* Springer-Verlag Inc., New York, 1998.

[12] E.L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses, Third Edition.* Springer Science+Business Media, Inc., New York, 2005.

[13] Romain Thibaux and Michael I. Jordan. Hierarchical beta processes and the Indian buffet process. *Technical Report 719, Department of Statistics, University of California, Berkeley*, 2006.

[14] Nils Lid Hjort. Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18:1259 – 1294, 1990.

[15] K. Sato. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge Univ. Press, Cambridge, 1999.

[16] O. E. Barndorff-Nielsen and F. Hubalek. Probability measures, Lévy measures and analycity in time. *Bernoulli*, 14:764–790, 2008.

[17] F. Hubalek. On a conjecture of Barndorff-Nielsen relating probability densities and Lévy densities. In *Proceedings of the 2nd MaPhySto Conference on Lévy Processes: Theory and Applications*, 2002.

[18] E.V. Burnaev. An inversion formula for infinitely divisible distributions. *Uspekhi Matematicheskikh Nauk*, 61:187–188, 2006.