POSTERIOR CONSISTENCY OF BAYESIAN NONPARAMETRIC MODELS USING LÉVY RANDOM FIELD PRIORS

By Natesh S. Pillai, Robert L. Wolpert

Department of Statistical Science, Duke University

March 24, 2008

We show the posterior consistency of certain nonparametric regression models using Lévy Random field priors. An easily verifiable sufficient condition is derived for the posterior consistency to hold in popular models which use Lévy random fields for regression and function estimation. We apply our results to a Poisson regression model. Our calculations on the Poisson regression model are of independent interest.

1. Introduction. In this paper we study the posterior consistency of certain Bayesian nonparametric regression models using Lévy random measures. Nonparametric function estimation using Lévy random measures has received great attention in the literature during the last decade. They were introduced in [Wolpert and Ickstadt 1998, Wolpert. and Ickstadt 1998] in which the authors modeled spatial data using Lévy random fields. Other applications can be found in [Wolpert and Ickstadt 2004, Wolpert *et al.* 2003], also in which efficient numerical schemes for implementation are developed.

A similar approach is taken in Nieto-Barajas *et al.* [2004], in which normalized random measures derived from increasing additive processes are considered as priors for distribution functions. The main difference in our approach is that we consider random measures which are not necessarily positive. Moroever in our case, the random measures can be unbounded and need not even have finite total variation.

Recently [Tu *et al.* 2006] extended the methodology developed in the papers mentioned above to construct a richer class of models, called LARK (Lévy Adaptive Regression Kernels), for func-

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35

Keywords and phrases: Bayesian; posterior consistency; Lévy random Fields; nonparametric regression

tion estimation. Consider estimating an unknown function based on noisy data. The traditional approach is to represent the function in a series expansion using a linear combination of basis functions and then to estimate the coefficients from the data. In [Tu *et al.* 2006], the authors use overcomplete dictionaries instead of an orthonormal basis. Despite the non-uniqueness due to the over complete representation, this has the potential to lead to sparser representations by using fewer non-zero coefficients. Prior distributions based on infinitely divisible processes such as compound Poisson random fields and more generally Lévy random fields are ideally suited for construction of priors on functions using these overcomplete representations. While expressions for posterior distributions of quantities of interest are not available in closed form, one can take advantage of the independent increments property of Lévy random fields to simulate from the posterior distribution. Efficient computation is possible because updates based on adding/deleting or updating single dictionary elements bypass the need to invert large matrices. Furthermore, because dictionary elements are only computed as needed, memory requirements scale linearly with the sample size. See Clyde and Wolpert [2007] for a thorough exposition of the relative merits of these models and a comparison with other relevant models in the literature.

Even though these models perform very well in practice, the issue of their posterior consistency has not been studied in the literature. Posterior consistency of infinite dimensional non parametric models is not always guaranteed and the modeler needs to be careful to avoid unpleasant surprises (see Barron *et al.* [1999], Diaconis and Freedman [1986a,b]). One of the important criteria for the posterior consistency to hold is that the prior puts positive mass on the Kullback-Liebler (KL) neighborhoods of the true(?) parameter. This condition is easy to verify when the number of dictionary elements is almost surely finite. However, in the models we consider it is possible to have an infinite number of dictionary elements in the representation of the unknown function, and more importantly the coefficients corresponding to those elements need not be absolutely summable. Hence it requires more effort to verify the KL positivity condition mentioned above, and we accompolish this using a limiting arguement.

In addition to verifying the KL positivity condition, one has to show the existence of tests which separate points in the parameter space. This depends on the the specific model under consideration. Posterior consistency for independent but not identically distributed random variables are studied only recently in the literature. The general framework for proving such results was laid out by Amewou-Atisso *et al.* [2003], Ghosal and van der Vaart [2007a,b] and Choi and Schervish [2007]. Choi and Schervish [2007] prove the posterior consistency for a regression model with Gaussian error terms. Recently posterior consistency was proved for nonparametric binary regression by Ghosal and Roy [2006], who conjectured that similar results might hold for the Poisson regression. We verify that conjecture, proving posterior consistency for the Poisson regression model under moderate assumptions. Our results on the Poisson regression model are of independent interest.

The rest of the paper is organized as follows. Section 2 explains the basic frame work of Lévy random measures. Section 3 contains the main results of the paper. Section 4 has examples which use the results from the previous sections. Section 5 contains a brief discussion and conclusion.

2. Kernel Model.

Let $\mathcal{X}$ be a compact subset of $\mathbb{R}$ and Θ be a space of real valued functions on $\mathcal{X}$. Let Ω be a complete separable metric space and $K : \mathcal{X} \times \Omega \to \mathbb{R}_+$ a Borel measurable function. Set

$$\Theta_D \equiv \left\{ \theta \in \Theta : \theta(x) = \sum_{j=1}^{J} u_j K(x, \omega_j), \qquad J \in \mathbb{N},\ u_j \in \mathbb{R},\ \omega_j \in \Omega \right\}. \tag{2.1}$$

The function $K(\cdot,\cdot)$ will be chosen later to ensure that Θ_D is dense in Θ, in some specified topology. Let $M_D(\Omega)$ denote the space of finitely supported discrete signed measures on Ω. Let $\mathcal{K} : M_D(\Omega) \mapsto \Theta_D$ be the integral operator defined by

$$\mathcal{K}[\mathcal{L}](x) \equiv \int_\Omega K(x, \omega)\mathcal{L}(d\omega), \quad x \in \mathcal{X} \tag{2.2}$$

where $\mathcal{L} \in M_D(\Omega)$ is given by

$$\mathcal{L}(d\omega) \equiv \sum_{j=1}^{J} u_j \delta_{\omega_j}(d\omega), \quad J \in \mathbb{N},\ \{u_j, \omega_j\} \in \mathbb{R} \times \Omega. \tag{2.3}$$

The integral operator $\mathcal{K}$ will be extended later to countably supported measures $\mathcal{L}$ on Ω with possibly infinite total variation. We define a probability distribution Π on the measurable space $(\Theta_D^-, \mathcal{B}(\Theta_D^-))$ (here Θ_D^- denotes the closure of Θ_D in some topology to be defined later) by first specifying a joint distribution of $(\mathcal{L}(A_1), \mathcal{L}(A_2), \ldots, \mathcal{L}(A_n))$, for disjoint Borel sets $A_i \subset \Omega$, and define Π to be the corresponding probability measure induced on $(\Theta_D^-, \mathcal{B}(\Theta_D^-))$ by the integral operator $\mathcal{K}$.

The joint distribution of $(\mathcal{L}(A_1), \mathcal{L}(A_2), \ldots, \mathcal{L}(A_n))$ is determined by the joint distribution of the random variables J and $\{u_j, \omega_j\}$. Next, we introduce Lévy random fields for defining a joint distribution for these random variables [Tu *et al.* 2006].

2.1. *Lévy Random Measures.* Let $\pi(du\, d\omega)$ be a probability measure on $\mathbb{R} \times \Omega$ and $\nu_+ > 0$. Let $J \sim \mathrm{Po}(\nu_+)$ and, conditional on J, let $\{u_j, \omega_j\}_{j=1}^{J} \overset{iid}{\sim} \pi(du\, d\omega)$. The random measure $\mathcal{L}$ defined by

$$\mathcal{L}(A) = \sum_{j=1}^{J} 1_A(\omega_j)\ u_j \tag{2.4}$$

assigns independent infinitely-divisible (ID) random variables $\mathcal{L}(A_i)$ to disjoint Borel sets $A_i \subset \Omega$. For any Borel set $A \subset \Omega$, $\mathcal{L}(A)$ has characteristic function

$$\mathbb{E}\left[e^{it\mathcal{L}(A)}\right] = \exp\left\{\iint_{\mathbb{R}\times A} (e^{itu} - 1)\nu(du\, d\omega)\right\} \tag{2.5}$$

where $\nu(du\, d\omega) \equiv \nu_+ \pi(du\, d\omega)$. A sigma-finite random measure $\mathcal{L}$ with characteristic function given by (2.5) can be well defined for any sigma-finite positive measure $\nu(du\, d\omega)$ on $\mathbb{R} \times \Omega$ that satisfies

$$\iint_{\mathbb{R}\times K} (1 \wedge |u|)\ \nu(du\, dw) < \infty \tag{2.6}$$

for each compact $K \subset \Omega$. If a measure $\nu(\cdot)$ satisfying (2.6) is infinite, then the random measure $\mathcal{L}$ will have a countably-infinite support set. For relatively compact Borel sets A (*i.e.*, those with compact closure), the sum

$$\mathcal{L}(A) = \sum 1_A(\omega_j) u_j \tag{2.7}$$

converges absolutely.

More generally, for any positive sigma-finite measure $\nu(du\, d\omega)$ on $\mathbb{R} \times \Omega$ satisfying the weaker condition

$$\iint_{\mathbb{R}\times K} (1 \wedge u^2)\ \nu(du\, d\omega) < \infty \tag{2.8}$$

for each compact $K \subset \Omega$, and any sigma-finite signed measure $m(d\omega)$, there exists a random sigma-finite measure $\mathcal{L}(d\omega)$ with characteristic function

$$\mathbb{E}\left[e^{it\mathcal{L}(A)}\right] = \exp\left\{itm(A) + \iint_{\mathbb{R}\times A} (e^{itu} - 1 - ith_0(u))\nu(du\, d\omega)\right\} \tag{2.9}$$

with "compensator function" $h_0(u) \equiv u\mathbf{1}_{|u|\le 1}(u)$ (see Rajput and Rosiński [1989, Prop 2.1] for extending the classic Lévy Khintchine formula to random measures). Since $e^{itu} - 1 - ith_0(u) = O(u^2),\ u \to 0$, the integral in (2.9) converges for any relatively compact A. Equation (2.9) remains valid, if we replace $h_0(u)$ by any other bounded $h(u)$ satisfying $h(u) = u + O(u^2),\ u \to 0$, and $m(A)$ by $m(A) + \iint_{\mathbb{R}\times A}[h(u) - h_0(u)]\nu(du\,d\omega)$.

2.2. *Lévy Random Fields.* In this section we construct Lévy random fields. Let $\nu(du\,d\omega)$ be a positive sigma-finite measure on $\mathbb{R}\times\Omega$ satisfying the integrability condition (2.8). Let

$$N(du\,d\omega) \sim \mathrm{Po}(\nu) \tag{2.10}$$

be a Poisson random measure on $\mathbb{R}\times\Omega$ which assigns independent $\mathrm{Po}(\nu(B_i))$ distributions to disjoint Borel sets $B_i \subset \mathbb{R}\times\Omega$. Let

$$\tilde{N}(du\,d\omega) \equiv N(du\,d\omega) - \nu(du\,d\omega) \tag{2.11}$$

denote the *compensated* Poisson measure with mean 0, an isometry from $L_2(\mathbb{R}\times\Omega, \nu(du\,d\omega))$ to the square-integrable zero-mean random variables [iti Sato 1999, *pg.* 38]. For relatively compact Borel sets $A \subset \Omega$ and a sigma-finite signed measure m, define (following Wolpert and Taqqu [2005])

$$\mathcal{L}(A) \equiv m(A) + \iint_{\mathbb{R}\times A}(u - h(u))N(du\,d\omega) + \iint_{\mathbb{R}\times A} h(u)\tilde{N}(du\,d\omega) \tag{2.12}$$

where $h(u)$ is any bounded function satisfying $h(u) = u + O(u^2)$, $u \to 0$. The random variable $\mathcal{L}(A)$ has the characteristic function

$$\mathbb{E}\left[e^{it\mathcal{L}(A)}\right] = \exp\left\{itm(A) + \iint_{\mathbb{R}\times A}(e^{itu} - 1 - ith(u))\nu(du\,d\omega)\right\}. \tag{2.13}$$

Henceforth we denote the Lévy random measure $\mathcal{L}$ with intensity measure ν

$$\mathcal{L} \sim \text{Lévy}(\nu, m, h). \tag{2.14}$$

In what follows, unless otherwise specified, we set $m \equiv 0$ and $h(u) = h_0(u) \equiv u\mathbf{1}_{|u|<1}$. If in addition, the measure ν satisfies the stronger integrability condition (2.6), we also set $h \equiv 0$. In either of these instances, we further omit m, h from the notation and write

$$\mathcal{L} \sim \text{Lévy}(\nu). \tag{2.15}$$

Notice that if $\nu_+ \equiv \nu(\mathbb{R} \times \Omega) < \infty$, then the random measure $\mathcal{L}$ can be represented as in (2.4). A similar representation is available (with $J = \infty$) if ν is infinite but satisfies (2.6), but no such convergent sum is available if (2.6) fails.

Following [Taqqu and Wolpert 1983, section 3], we now extend the definition of the integral operator $\mathcal{K}$ as follows: For the random measure $\mathcal{L} \sim$ Lévy$(\nu(du\, d\omega))$, set

$$
\begin{aligned}
\mathcal{K}[\mathcal{L}](x) &\equiv \int_\Omega K(x,\omega)\mathcal{L}(d\omega),\ \ x \in \mathcal{X} &(2.16)\\
&\equiv \iint_{\mathbb{R}\times\Omega} K(x,\omega)(u - h(u))N(du\, d\omega) + \iint_{\mathbb{R}\times\Omega} K(x,\omega)h(u)\tilde{N}(du\, d\omega) &(2.17)
\end{aligned}
$$

for all Borel measurable $K(\cdot,\cdot)$ for which the integrals in (2.17) converge – *i.e.,* for those in "Musielak-Orlicz" (Rajput and Rosiński [1989]) space for (2.17) to be well defined. For a function $f : \mathbb{R} \times \Omega \mapsto \mathbb{R}$ define

$$
I(f) \equiv \iint_{\mathbb{R}\times\Omega} f(u,\omega)\tilde{N}(du\, d\omega). \tag{2.18}
$$

Define

$$
L_{\Psi_p}(\mathbb{R}\times\Omega,\nu) \equiv \left\{ f : f \text{ is Borel measurable and } \iint_{\mathbb{R}\times\Omega} \Psi_p(f(u,\omega))\nu(du\, d\omega) < \infty \right\} \tag{2.19}
$$

where

$$
\Psi_p(y) \equiv \begin{cases} |y| \wedge |y|^2, & 0 \le p \le 1, \\ |y|^p \wedge |y|^2, & 1 < p \le 2, \\ |y|^p \vee |y|^2, & p > 2. \end{cases} \tag{2.20}
$$

Theorem 2.1. *(Rajput and Rosiński [1989], Theorem 3.3, Gaigalas [2004a,b]) The integral $I(f)$ in equation (2.18) is well defined and $\mathbb{E}(|I(f)|^p) < \infty$ for some $p > 0$, if and only if $f \in L_{\Psi_p}(\mathbb{R}\times\Omega,\nu)$.*

Remark 2.2. *For any $A \subset \Omega$ with compact closure, $f(u,\omega) \equiv h_0(u)\, 1_A(\omega) \in L_{\Psi_0}(\mathbb{R}\times\Omega,\nu)$, since*

$$
\iint_{\mathbb{R}\times A} (|f(u,\omega)| \wedge |f(u,\omega)|^2)\nu(du\, d\omega) = \iint_{[-1,1]\times A} u^2 \nu(du\, d\omega) < \infty.
$$

Hence the compensated random measure $\mathcal{L}(A)$ given by (2.12) is well defined. Moreover for $p \ge 1$,

$$
\mathbb{E}(|\mathcal{L}(A)|^p) < \infty, \ \text{if and only if} \iint_{[-1,1]^c\times A} |u|^p \nu(du\, d\omega) < \infty. \tag{2.21}
$$

Remark 2.3. *Also*

$$L_p(\mathbb{R}\times\Omega,\nu) \subset L_{\Psi_p}(\mathbb{R}\times\Omega,\nu),\ 1\le p\le 2,$$

with strict inclusion unless $p = 2$.

2.3. *Truncation.* Let ν be an infinite Lévy measure and let $\mathcal{L}\sim$ Lévy(ν). Then $\mathcal{L}$ has an infinite number of elements in its support almost surely. Hence we will require (Robert: I don't have a good answer for why) a sequence of Lévy random measures $\{\mathcal{L}^\epsilon\}$ with almost surely finite number of points in their support and approximate (made more precise below) $\mathcal{L}$ as ϵ goes to 0. This finite approximation is also useful and routinely implemented in numerical procedures to sample Lévy random measures. Hence it is useful to study the weak convergence of the sequence of random measures $\{\mathcal{L}^\epsilon\}$ to a limiting $M(\Omega)$ valued random variable $\mathcal{L}$, where $M(\Omega)$ denotes the set of finite signed measures on Ω.

If Ω is compact, by the Reisz representation theorem [Albiac and Kalton 2006, Theorem 4.1.1, pg. 74], $M(\Omega)$ can be identified as the dual of the banach space $C(\Omega)$, the space of the continuous function on Ω. The duality is given by

$$\mu(f) = \int_\Omega f d\mu,\ f\in C(\Omega),\ \mu\in M(\Omega)$$

with the operator norm of the linear functional μ above given by $\|\mu\|\equiv|\mu|(\Omega)$, the total variation. A natural topology on $M(\Omega)$ is the usual weak star topology on $M(\Omega)$ [Reed and Simon 1980, pg. 113], the weakest topology for which all the functions $\mu\mapsto\mu(f),\ f\in C(\Omega)$ are continuous. Finite linear combinations of point measures are dense in $M(\Omega)$ [Reed and Simon 1980, pg. 114], so $M(\Omega)$ is the closure of $M_D(\Omega)$ (the set of finite discrete measures on $M(\Omega)$) in the weak star topology. Thus it is natural to approximate countably supported Lévy random measures with a sequence of measures belonging to $M_D(\Omega)$. Weak convergence of probability measures is well studied in the metric space setting [Billingsley 1999]. Alas, since the Banach space $C(\Omega)$ is infinite dimensional, the weak star topology on its dual $M(\Omega)$ is not metrizable (see [Dunford and Schwartz 1988, pg. 462] & [Wehausen 1938, Theorem 17]).

Hence to study the weak convergence of random variables in $M(\Omega)$, one option is to consider other natural topologies on $M(\Omega)$ which *are* metrizable. Another option is to consider $M(\Omega)$ as a subset of $D(\Omega)$, the set of tempered distributions ([Reed and Simon 1980, pg. 134]) on Ω. Instead we settle for a simpler approach: We find a sequence of Lévy random measures $\{\mathcal{L}^\epsilon\}$ such that for any precompact $A \subset \Omega$, the random variables $\{\mathcal{L}^\epsilon(A)\}$ converge in distribution to $\mathcal{L}(A)$.

Let Ω^ϵ be an increasing sequence of compact sets (as $\epsilon \to 0$) with $\cup_{\epsilon>0}\Omega^\epsilon = \Omega$. For any $A \times B \subset \mathbb{R} \times \Omega$, set

$$\nu^\epsilon(A \times B) \equiv \nu((A \cap (-\epsilon, \epsilon)^c) \times (B \cap \Omega^\epsilon)). \tag{2.22}$$

Consider

$$\mathcal{L}^\epsilon \sim \text{Lévy}(\nu^\epsilon, 0, h_0). \tag{2.23}$$

By equation (2.8), $\nu^\epsilon(\mathbb{R} \times \Omega) = \nu((-\epsilon, \epsilon)^c \times \Omega^\epsilon) < \infty$ and $\mathcal{L}^\epsilon$ will only have finitely many points (almost surely) in its support. We call $\mathcal{L}^\epsilon$ the compound Poisson approximation of $\mathcal{L}$. Also set

$$I^\epsilon(u, \omega) \equiv 1_{\{|u|>\epsilon, \omega \in \Omega^\epsilon\}}(u, \omega) \in L_1(\Omega, \nu(du\, d\omega)).$$

THEOREM 2.4. *For any precompact $A \subset \Omega$, the sequence of random variables $\mathcal{L}^\epsilon(A)$ converges in distribution to $\mathcal{L}(A)$, as ϵ goes to 0.*

Proof. The integrability condition $\iint_{\mathbb{R}\times A}(1 \wedge u^2)\nu(du\, d\omega) < \infty$ ensures that

$$\iint_{\mathbb{R}\times A} (e^{itu} - 1 - ith(u))\nu(du\, d\omega)$$

is well defined and hence $\mathbb{E}\left[e^{it\mathcal{L}(A)}\right]$ exists. For $\epsilon > 0, t \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}\left[e^{it\mathcal{L}^\epsilon(A)}\right] &= \exp\left\{\iint_{\mathbb{R}\times A} (e^{itu} - 1 - ith_0(u))\nu^\epsilon(du\, d\omega)\right\} \\ &= \exp\left\{\iint_{\mathbb{R}\times A} (e^{itu} - 1 - ith_0(u))I^\epsilon(u, \omega)\nu(du\, d\omega)\right\}. \end{aligned}$$

Since $I^\epsilon(u, \omega) \le 1$ and $\lim_{\epsilon\to 0} I^\epsilon(u, \omega) = 1$, the dominated convergence theorem ensures that,

$$\lim_{\epsilon\to 0} \mathbb{E}\left[e^{it\mathcal{L}^\epsilon(A)}\right] = \exp\left\{\iint_{\mathbb{R}\times A} (e^{itu} - 1 - ith_0(u))\left(\lim_{\epsilon\to 0} I^\epsilon(u, \omega)\right)\nu(du\, d\omega)\right\} = \mathbb{E}\left[e^{it\mathcal{L}(A)}\right] \tag{2.24}$$

and the theorem is proved. ∎

REMARK 2.5. *Notice that for $\epsilon > 0$, the random variable $\mathcal{L}(A) - \mathcal{L}^\epsilon(A)$ is independent of $\mathcal{L}^\epsilon(A)$, and has the characteristic function*

$$\mathbb{E}\left(e^{it\left(\mathcal{L}(A) - \mathcal{L}^\epsilon(A)\right)}\right) = \exp\left\{\iint_{(-\epsilon,\epsilon)\times A} (e^{itu} - 1 - ith_0(u))\nu(du\, d\omega)\right\}.$$

EXAMPLE 2.6. *Let* $\Omega = [0,1]$. *Let* $\mathbf{D}[0,1]$ *be the Skorohod space (the space of right continous functions with left limits, see [Billingsley 1999, Chapter 3]) and* $\mathcal{D}$ *be the associated Borel sigma field.*

Let $\nu(du\, d\omega)$ *satisfy equation (2.6). Set*

$$\mathcal{L}(d\omega) \equiv \sum_{j=1}^{\infty} u_j \delta_{\omega_j}(\omega),\ \ \{u_j, \omega_j\} \in \mathbb{R} \times [0,1].$$

Then the process X_t *on* $[0,1]$ *defined by*

$$X_t \equiv \sum_{j=1}^{\infty} u_j 1_{\{\omega_j \le t\}}$$

is infinitely divisible and the sample path $X \equiv \{X_t, t \in [0,1]\}$ *is a random element of* $\mathbf{D}[0,1]$. *The truncated random measure* $\mathcal{L}^\epsilon$ *is given by*

$$\mathcal{L}^\epsilon(d\omega) \equiv \sum_{j=1}^{\infty} u_j 1_{\{|u_j| > \epsilon\}} \delta_{\omega_j}(\omega),\ \ \{u_j, \omega_j\} \in \mathbb{R} \times [0,1]$$

and the process X_t^ϵ *defined analogously, is the usual compound Poisson process. The random measures* $\mathcal{L}$ *and* $\{\mathcal{L}^\epsilon\}$ *induce a distribution on the measure space* $(\mathbf{D}[0,1], \mathcal{D})$ *by the correspondence mentioned above. Theorem (2.4) implies that the random variable* X_t^ϵ *converges in distribution to* X_t *for any fixed* $t > 0$. *Hence by (give reference Bertoin ?), the sequence* $\{X_t^\epsilon,\ t \in [0,1]\}$ *of* $\mathbf{D}[0,1]$ *variables converges in distribution to* $\{X_t, t \in [0,1]\}$. *The ongoing construction needs a little more modification, when* ν *does not satisfy (2.6) as illustrated in the next example.*

This example also suggests another alternative to study weak convergence of Lévy random measures. The skorohod topology is defined for functions on the positive real line and we believe that generalizing it to functions on Polish spaces (Ω *in our case) will be a useful path to follow.*

EXAMPLE 2.7. *For the choice of the compensator* $h(u) \equiv \sin(u) 1_{|u|<1}$, *the one dimensional stable distribution* $\boldsymbol{St}(\alpha, \beta, \gamma, \delta)$ *on the real line has the characteristic function*

$$\Phi_{\boldsymbol{St}}(t) = \begin{cases} \exp\Big[-\gamma|t|^\alpha + it\Big\{\delta - \beta\gamma \tan\frac{\pi\alpha}{2}\ (1 - |t|^{\alpha-1})\Big\}\Big], & \alpha \ne 1, \\ \exp\Big[-\gamma|t| + it\Big\{\delta - \frac{2\beta\gamma}{\pi} \log|t|\Big\}\Big], & \alpha = 1, \end{cases} \tag{2.25}$$

with $0 < \alpha \le 2$, $-1 \le \beta \le 1$, $0 < \gamma < \infty$, $-\infty < \delta < \infty$, *with the Lévy measure*

$$\nu(du) = \begin{cases} \gamma\frac{\alpha}{\pi}\ \Gamma(\alpha)\ \ \sin(\pi\alpha/2)\ |u|^{-1-\alpha}\ (1 + \beta \operatorname{sgn} u)\ du, & \alpha \ne 1, \\ \frac{\gamma}{\pi}\ |u|^{-2}\ (1 + \beta \operatorname{sgn} u)\ du, & \alpha = 1. \end{cases} \tag{2.26}$$

Notice that

$$\int_{\mathbb{R}} (1 \wedge |u|)\ \nu(du)du < \infty, \text{ only for } \alpha \in (0,1). \tag{2.27}$$

The skewed Cauchy distribution, belongs to the stable family and is parametrized by ***St****(1,1,1,0). The corresponding Lévy measure* ν *is*

$$\nu(du) = \frac{1}{\pi}\ |u|^{-2} 1_{u>0}.$$

Let $(u_j, s_j) \in \mathbb{R}_+ \times [0,1]$ *be generated from a Poisson random field with intensity measure* $\nu(du)ds$. *Then for any fixed* $t \in (0,1]$, *the sequence of random variables defined by*

$$X_t^{\epsilon} \equiv \Big\{ \sum u_j : u_j > \epsilon,\ s_j \le t \Big\} \tag{2.28}$$

tends to ∞ *a.s. as* $\epsilon \to 0$. *However if we compensate* X_t^{ϵ} *to obtain,*

$$Y_t^{\epsilon} \equiv X_t^{\epsilon} - t\,\delta^{\epsilon},\ \ \delta^{\epsilon} \equiv \int_{\epsilon}^{\infty} \sin(u)\nu(du) < \infty, \tag{2.29}$$

then the sequence of random variables Y_1^{ϵ} *converges almost surely to a* ***St****(1,1,1,0) random variable. Moreover, the sequence of* $\mathbf{D}[0,1]$ *valued random variables*

$$Y^{\epsilon} \equiv \Big\{ Y_t^{\epsilon}, t \in [0,1] \Big\} \tag{2.30}$$

converges in distribution to a skewed cauchy process in $\mathbf{D}[0,1]$, *as* $\epsilon \downarrow 0$.

Let $Y_1^{\epsilon,1}, Y_1^{\epsilon,2}$ *be two independent copies of the random variable* Y_1^{ϵ} *defined in* (2.29). *Then the sequence of random variables*

$$\tilde{Y_1}^{\epsilon} \equiv Y_1^{\epsilon,1} - Y_1^{\epsilon,2} = X_1^{\epsilon,1} - X_1^{\epsilon,2}$$

converges almost surely to a Cauchy random variable $\Big($***St****(1,0,2,0)*$\Big)$. *(Robert: I need help in writing the bit clearly about why it does not seem like compensation is needed in defining* $\tilde{Y}_1^{\epsilon}$ *).*

Remark 2.8. *There exist methods in the literature, other than the truncation approach, for approximating Lévy random fields that have infinite Lévy measure with those that have finite Lévy measure.*

The Lévy measure for the Symmetric Alpha Stable (SαS) process on the unit interval is given by

$$\nu_{\alpha}(du\,d\omega) = c_{\alpha}\gamma\alpha\ |u|^{-\alpha-1} du\,d\omega,\ \ \{u,\omega\} \in \mathbb{R}\times[0,1],\ \ \alpha \in (0,2), \gamma > 0$$

with $c_\alpha \equiv \frac{\Gamma(\alpha)}{\pi}\sin(\frac{\pi\alpha}{2})$. *Notice that*

$$\nu_\alpha(\mathbb{R}\times[0,1]) = 2c_\alpha\gamma\alpha\int_0^\infty |u|^{-\alpha-1}du\,d\omega = \infty,\ \alpha\in(0,2).$$

The compound Poisson approximation has finite Lévy measure

$$\nu_\alpha^\epsilon(du\,d\omega) = \nu_\alpha(du\,d\omega)1_{|u|>\epsilon},\ \{u,\omega\}\in\mathbb{R}\times[0,1],\ \alpha\in(0,2),\gamma>0.$$

It can be shown that the Theorem(2.4) is also applicable to the sequence $\mathcal{L}^\epsilon \sim$ *Lévy*$(\nu_\alpha^{\prime\epsilon}(du\,d\omega))$ *with*

$$\nu_\alpha^{\prime\epsilon}(du\,d\omega) \equiv c_\alpha\gamma\alpha\,\Big(|u|\vee\epsilon\Big)^{-\alpha-1}du\,d\omega,\ \{u,\omega\}\in\mathbb{R}\times[0,1],\ \alpha\in(0,2),\gamma>0.$$

In (cite Zhi's paper), the authors use the following approximation for ν_α*:*

$$\tilde{\nu}_\alpha^\epsilon(du\,d\omega) \equiv c_\alpha\gamma\alpha\,\frac{1}{(\epsilon^2+|u|^2)^{(\alpha+1)/2}}\,du\,d\omega,\ \{u,\omega\}\in\mathbb{R}\times[0,1],\ \alpha\in(0,2),\gamma>0.$$

Notice that $\nu_\alpha^{\prime\epsilon}(\mathbb{R}\times[0,1])<\infty$ *and* $\tilde{\nu}_\alpha^\epsilon(\mathbb{R}\times[0,1])<\infty$.

3. Prior positivity. Let $\nu(du\,d\omega)$ be a positive Radon measure on Ω such that

$$\iint_{\mathbb{R}\times\Omega}(1\wedge u^2)\nu(du\,d\omega)<\infty \tag{3.1}$$

and set $\mathcal{L}\sim$ Lévy(ν). For $\epsilon>0$, let $\nu^\epsilon(du\,d\omega)$ denote the measure defined in (2.22) and let $\mathcal{L}^\epsilon\sim$ Lévy$(\nu^\epsilon,0,h_0)$. Let $K(\cdot,\omega)$ be a continuous function on $\mathcal{X}$ for each fixed $\omega\in\Omega$. Then the integral operator $\mathcal{K}$ given by (2.17) maps Lévy random measures into C$(\mathcal{X})$. Let Π, $\{\Pi^\epsilon\}$ be probability distributions on $(C(\mathcal{X}),\mathcal{B}(C))$ induced by the operator $\mathcal{K}$ via the random measures $\mathcal{L}$ and $\{\mathcal{L}^\epsilon\}$ respectively.

For a real valued function $f:\mathcal{X}\mapsto\mathbb{R}$ let $W_f(\delta)$ denote the modulus of continuity of $f(\cdot)$,

$$W_f(\delta)\equiv\sup_{|s-t|<\delta}\Big\{|f(s)-f(t)|,\ s,t\in\mathcal{X}\Big\}. \tag{3.2}$$

Define

$$W_K(\delta)\equiv\sup_{\omega\in\Omega}W_{K(\cdot,\omega)}(\delta). \tag{3.3}$$

Since the function W_K is non-decreasing in δ its right-continuous inverse $W_K^{\leftarrow}$ is well defined.

We now introduce two conditions, either of which (as we will see in theorem (3.2)) is sufficient to ensure the weak convergence of Π^ϵ.

Condition A: The family of functions $\Big\{K(\cdot,\omega),\ \omega\in\Omega\Big\}$ is equicontinuous with $\kappa\equiv\sup_{\mathcal{X}\times\Omega}K(x,\omega)<\infty$ and, in addition to (3.1), the Lévy measure satisfies

$$\iint_{\mathbb{R}\times\Omega}(1\wedge|u|)\nu(du\,d\omega)<\infty. \tag{3.4}$$

Condition B: There exists $r>0$ such that

$$\int_0^r \frac{1}{\sqrt{W_K^{\leftarrow}(a)}}\,da<\infty. \tag{3.5}$$

REMARK 3.1. *If the function $K(\cdot,\cdot)$ is uniformly continuous on $\mathcal{X}\times\Omega$ with $\kappa\equiv\sup_{\mathcal{X}\times\Omega}K(x,\omega)<\infty$, then the family $\Big\{K(\cdot,\omega),\ \omega\in\Omega\Big\}$ is equicontinuous. In addition if $K(\cdot,\omega)\in C^{\alpha}(\mathcal{X})$, for some $\alpha\in(1/2,1]$, i.e.*

$$W_K(\delta)\le C\delta^{\alpha},\ C>0,\ \alpha\in(1/2,1], \tag{3.6}$$

then condition B is satisfied as well.

THEOREM 3.2. *The sequence of probability measures $\{\Pi^{\epsilon}\}$ converges weakly to Π in $C(\mathcal{X})$ if either (A) or (B) holds true.*

Proof. Set

$$f^{\epsilon}(x)\equiv\mathcal{K}[\mathcal{L}^{\epsilon}](x),\quad f(x)\equiv\mathcal{K}[\mathcal{L}](x)\quad x\in\mathcal{X}. \tag{3.7}$$

First we claim that each finite dimensional random vector of the form

$$\Big(f^{\epsilon}(x_1),f^{\epsilon}(x_2),\cdots,f^{\epsilon}(x_n)\Big),\ n\in\mathbb{N},\ \{x_k\}_{k=1}^n\in\mathcal{X},$$

converges in distribution to the random vector $(f(x_1),f(x_2),\cdots,f(x_n))$ when ϵ goes to 0. To see this, as in the proof of Theorem 2.4, use dominated convergence theorem to verify that

$$\lim_{\epsilon\to0}\mathbb{E}\left[e^{\sum_{k=1}^n ia_k f^{\epsilon}(x_k)}\right]=\mathbb{E}\left[e^{\sum_{k=1}^n ia_k f(x_k)}\right],\ n\in\mathbb{N},\ \{x_k\}_{k=1}^n\in\mathcal{X},\{a_k\}_{k=1}^n\in\mathbb{R}.$$

Hence the finite dimensional distributions of $\{\Pi^{\epsilon}\}$ converge to those of Π. To show that $\{\Pi^{\epsilon}\}$ converges weakly, by Prohorov's theorem ([Billingsley 1999], Thm 5.1, pg 59) it is enough to show the tightness of the sequence $\{\Pi^\epsilon\}$. This is equivalent to showing that ([Billingsley 1999],Thm 7.5), for every $\eta > 0$,

$$\lim_{\delta\to 0}\limsup_{\epsilon\to 0}\mathbb{P}(W_{f^\epsilon}(\delta) > \eta) = 0. \tag{3.8}$$

If condition A holds true, then

$$W_{f^\epsilon}(\delta) \le \int_\Omega \sup_{|x-y|<\delta} |K(x,\omega) - K(y,\omega)|\ |\mathcal{L}|^\epsilon(d\omega) \le W_K(\delta)\ |\mathcal{L}|^\epsilon(\Omega), \tag{3.9}$$

where $|\mathcal{L}|$ denotes the total variation norm of the measure $\mathcal{L}$. Hence

$$\lim_{\delta\to 0}\limsup_{\epsilon\to 0}\mathbb{P}(W_{f^\epsilon}(\delta) > \eta) \le \lim_{\delta\to 0}\limsup_{\epsilon\to 0}\mathbb{P}\Big(W_K(\delta)|\mathcal{L}|^\epsilon(\Omega) > \eta\Big).$$

Theorem (2.4) and (3.4) guarantee that the sequence $\{|\mathcal{L}|^\epsilon(\Omega)\}$ converges in distribution to $|\mathcal{L}|(\Omega)$, and that $|\mathcal{L}|(\Omega) < \infty$ a.s. Therefore

$$\lim_{\delta\to 0} W_K(\delta)|\mathcal{L}|(\Omega) = 0 \text{ a.s.}$$

Hence (3.8) holds, and the theorem under condition A.

Now suppose condition B holds. From equation (2.17), for every $x \in \mathcal{X}$, the random variable $f^\epsilon(x)$ can be written as the sum of two independent random variables

$$f^\epsilon(x) = \sum_j K(x,\omega_j)u_j 1_{|u_j|>1} + \iint_{[-1,1]\times\Omega} K(x,\omega)u\tilde{N}^\epsilon(du\,d\omega) \tag{3.10}$$

Therefore

(3.11)

$$\begin{aligned}\sup_{|x-y|<\delta}|f^\epsilon(x) - f^\epsilon(y)| &\le \sup_{|x-y|<\delta}\sum_j \Big|K(x,\omega_j) - K(y,\omega_j)\Big||u_j|1_{|u_j|>1} \\ &\quad + \sup_{|x-y|<\delta}\Big|\iint_{[-1,1]\times\Omega}\Big(K(x,\omega) - K(y,\omega)\Big)u\,\tilde{N}^\epsilon(du\,d\omega)\Big| \\ &\le W_K(\delta)\Big(\sum_j |u_j|1_{|u_j|>1}\Big) + \sup_{|x-y|<\delta}\Big|\iint_{[-1,1]\times\Omega}\Big(K(x,\omega) - K(y,\omega)\Big)u\,\tilde{N}^\epsilon(du\,d\omega)\Big|.\end{aligned}$$

The finite sum $\sum_j |u_j|1_{|u_j|>1}$ is almost surely finite, so as before it follows that

$$\lim_{\delta\to 0}\limsup_{\epsilon\to 0}\mathbb{P}^\epsilon[W_K(\delta)(\sum_j |u_j|1_{|u_j|>1}) > \eta] = 0. \tag{3.12}$$

By the L_2 isometry of the compensated Poisson integrals,

(3.13)
$$\mathbb{E}\left[\iint_{[-1,1]\times\Omega}\Big(K(x,\omega)-K(y,\omega)\Big)u\tilde{N}^{\epsilon}(du\,d\omega)\right]^2 = \iint_{[-1,1]\times\Omega}\Big(K(x,\omega)-K(y,\omega)\Big)^2 u^2\ \nu^{\epsilon}(du\,d\omega)$$
$$\leq CW_K^2(|x-y|),\ C\equiv\iint_{[-1,1]\times\Omega}u^2\ \nu(du\,d\omega)<\infty.$$

Let $N(a,\mathcal{X},d)$ denote the covering number of $\mathcal{X}$ with d-balls of radius a, for the semi-metric d on $\mathcal{X}$ defined by

$$d(x,y)\equiv W_K(|x-y|),\ x,y\in\mathcal{X}.$$

By theorem 2.2.4 of [van der Vaart and Wellner 1996, pg. 98], for a positive constant K independent of ϵ and any $r>0$

(3.14)
$$\mathbb{E}\sup_{|x-y|<\delta}\Big|\iint_{[-1,1]\times\Omega}\Big(K(x,\omega)-K(y,\omega)\Big)u\tilde{N}^{\epsilon}(du\,d\omega)\Big|^2\leq K\Big[\int_0^r\sqrt{N(a,\mathcal{X},d)}\ da+\delta N(r,\mathcal{X},d)\Big]^2$$

Notice that by the definition of the metric d,

$$N(a,\mathcal{X},d)\leq K_1\frac{1}{W_K^{\leftarrow}(a)},\ K_1>0.$$

Hence if condition B holds true, for small enough r

$$\text{(3.15)}\quad \mathbb{E}\sup_{|x-y|<\delta}\Big|\iint_{[-1,1]\times\Omega}\Big(K(x,\omega)-K(y,\omega)\Big)u\tilde{N}^{\epsilon}(du\,d\omega)\Big|^2\leq K\Big[\int_0^r\frac{1}{\sqrt{W_K^{\leftarrow}(a)}}\ da+\delta\frac{1}{W_K^{\leftarrow}(r)}\Big]^2.$$

The right hand side of the above equation can be made arbitrarily small by first choosing r and then δ. Therefore by Markov's inequality, for any $\eta>0$,

(3.16)
$$\lim_{\delta\to0}\limsup_{\epsilon\to0}\mathbb{P}^{\epsilon}\Big(\sup_{|x-y|<\delta}\Big|\iint_{[-1,1]\times\Omega}\Big(K(x,\omega)-K(y,\omega)\Big)u\tilde{N}^{\epsilon}(du\,d\omega)\Big|>\eta\Big)$$
$$\leq\lim_{\delta,r\to0}\frac{K}{\eta^2}\Big[\int_0^r\frac{1}{\sqrt{W_K^{\leftarrow}(a)}}\ da+\delta\frac{1}{W_K^{\leftarrow}(r)}\Big]^2=0.$$

Hence the tightness of the sequence $\{\Pi^{\epsilon}\}$ follows and the theorem is proved. ■

The next theorem is our main result in this section. It gives a simply verifiable sufficient condition for the Lévy random field priors so that posterior consistency holds in most nonparametric regression models.

Let ν be a positive Radon measure on $\mathbb{R}\times\Omega$ satisfying (3.1). Let Supp $\nu(\cdot)$ denote the topological support of the measure ν and $K : \mathcal{X}\times\Omega \mapsto \mathbb{R}$ be a Borel measurable function.

Recall that and $\mathcal{X}$ is a compact subset of $\mathbb{R}$ and Θ_D is defined as

$$\Theta_D \equiv \left\{\theta \in C(\mathcal{X}) : \theta(x) \equiv \sum_{j=1}^{J} u_j K(x,\omega_j), \qquad J\in\mathbb{N},\ u_j\in\mathbb{R},\ \omega_j\in\Omega.\right\}$$

Let Θ_D^- be its closure in the uniform topology and $\mathcal{B}(\Theta_D^-)$ be the associated Borel σ-field. For $\theta_0\in\Theta$, let $B_\delta(\theta_0)$ denote the uniform ball of radius δ centered at θ_0, *i.e.*,

$$B_\delta(\theta_0) \equiv \{\theta' : \|\theta-\theta'\|_* < \delta\}. \tag{3.17}$$

THEOREM 3.3. *Let $\mathcal{L}\sim$ Lévy(ν) and Π be the probability measure induced on $(\Theta_D^-, (\mathcal{B}(\Theta_D^-)))$ by the integral operator $\mathcal{K}$ defined as in (2.17). Suppose that*

1. *$K(\cdot,\cdot)$ is uniformly continuous in $\mathcal{X}\times\Omega$ and atleast one of the conditions A,B holds true.*
2. *Supp $\nu(\cdot)=\mathbb{R}\times\Omega$.*

Then for any $\theta_0\in\Theta_D^-$, $\Pi(B_\delta(\theta_0))>0$ for all $\delta>0$.

Proof. Fix $\delta>0$. If the Lévy measure ν is infinite ($\nu_+\equiv\nu(\mathbb{R}\times\Omega)=\infty$), the random measure $\mathcal{L}$ will almost surely have infinite support. Fix $\epsilon>0$, and set $\mathcal{L}^\epsilon\sim$ Lévy$(\nu^\epsilon(du\,d\omega))$ as described in (2.22) and (2.23). Let Π^ϵ be the distribution of the random function $\mathcal{K}[\mathcal{L}^\epsilon](x)$. Define the function

$$g^\epsilon(x) \equiv \iint_{\mathbb{R}\times\Omega} K(x,\omega)h(u)\nu^\epsilon(du\,d\omega). \tag{3.18}$$

Notice that

$$\|g^\epsilon(x)\|_* \le \kappa \iint_{\{\epsilon\le|u|<1\}\times\Omega^\epsilon} u\,\nu(du\,d\omega) < \infty.$$

Since ν^ϵ is a compactly supported Radon measure, it can be well approximated by linear combinations of delta measures and hence $g^\epsilon(x)\in\Theta_D^-$.

Fix $\theta_0\in\Theta_D^-$. Since $(\theta_0+g^\epsilon)\in\Theta_D^-$, there exist a $J_\delta\in\mathbb{N}$ and $\{u_j^*,\omega_j^*\}_{j=1}^{J_\delta}$ such that

$$\|(\theta_0(x)+g^\epsilon(x)) - \sum_{j=1}^{J_\delta} K(x,\omega_j^*)u_j^*\|_* < \delta/4. \tag{3.19}$$

It is not guarenteed that $|u_j^*|>\epsilon$. Set $\tilde{u}_j^* = \mathrm{sgn}(u_j^*)(|u_j^*|\vee\epsilon)$. Notice that

$$\|\sum_{j=1}^{J_\delta} K(x,\omega_j^*)u_j^* - \sum_{j=1}^{J_\delta} K(x,\omega_j^*)\tilde{u}_j^*\|_* < J_\delta\kappa\epsilon \le \delta/4 \tag{3.20}$$

for $\epsilon \leq \frac{\delta}{4J_\delta \kappa}$. Hence

$$\|(\theta_0(x) + g^\epsilon(x)) - \sum_{j=1}^{J_\delta} K(x, \omega_j^*)\tilde{u}_j^*\|_* < \delta/2, \quad \epsilon \leq \frac{\delta}{4J_\delta \kappa}. \tag{3.21}$$

Set $\Psi \equiv \sum_{j=1}^{J_\delta} |\tilde{u}_j^*|$ and $\eta \equiv \frac{\delta}{2(\kappa+\Psi)}$. Since K in uniformly continuous in $\mathcal{X} \times \Omega$, there exists an $\eta' > 0$ such that $|\omega - \omega'| < \eta' \Rightarrow |K(x,\omega) - K(x,\omega')| < \eta, \ \forall x \in \mathcal{X}$. Define

$$B'_\delta(\theta_0 + g^\epsilon) \equiv \left\{\theta : \theta(x) = \sum_{j=1}^{J_\delta} u_j K(x, \omega_j), \ |u_j - \tilde{u}_j^*| < \eta, \ |\omega_j - \omega_j^*| < \eta' \right\}. \tag{3.22}$$

Lemma 3.4. $B'_\delta(\theta_0 + g^\epsilon) \subset B_\delta(\theta_0 + g^\epsilon)$

Proof. For any $\theta \in B'_\delta(\theta_0 + g^\epsilon)$ such that $\theta = \sum_{i=1}^{J_\delta} u_j K(x, \omega_j)$,

$$\begin{aligned}
|\theta(x) - \sum_{j=1}^{J_\delta} K(x,\omega_j^*)\tilde{u}_j^*| &\leq \sum_{j=1}^{J_\delta} |K(x,\omega_j)u_j - K(x,\omega_j^*)\tilde{u}_j^*| \\
&\leq \sum_{j=1}^{J_\delta} |K(x,\omega_j)u_j - K(x,\omega_j)\tilde{u}_j^*| + \sum_{j=1}^{J_\delta} |K(x,\omega_j)\tilde{u}_j^* - K(x,\omega_j^*)\tilde{u}_j^*| \\
&\leq \kappa \sum_{j=1}^{J_\delta} |u_j - \tilde{u}_j^*| + \sum_{j=1}^{J_\delta} |\tilde{u}_j^*|\eta \\
&\leq \kappa\eta + \Psi\eta = \delta/2.
\end{aligned}$$

Hence $\|\theta - \sum_{j=1}^{J_\delta} K(x,\omega_j^*)\tilde{u}_j^*\|_* < \delta/2$. Hence equation (3.21) and the triangle inequality imply that $\|\theta - (\theta_0 + g^\epsilon)\|_* < \delta$. Hence $B'_\delta(\theta_0 + g^\epsilon) \subset B_\delta(\theta_0 + g^\epsilon)$ and the lemma is proved. ∎

Notice that θ_0 is only assumed to be in Θ_D^- and hence need not belong to $C(\mathcal{X})$. Let $B_\delta^C(\theta_0) \subset C(\mathcal{X})$ be the closed ball around θ_0,

$$B_\delta^C(\theta_0) \equiv \left\{\theta \in C(\mathcal{X}) : \|\theta_0 - \theta\|_* \leq \delta\right\}.$$

$$\begin{aligned}
\Pi^\epsilon(B_\delta^C(\theta_0)) &= \mathbb{P}^\epsilon\left(\int_\Omega K(x,\omega)\mathcal{L}^\epsilon(d\omega) \in B_\delta^C(\theta_0)\right) \\
&\geq \mathbb{P}^\epsilon\left(\sum_{j=1}^{J_\delta} K(x,\omega_j^*)\tilde{u}_j^* - g^\epsilon(x) \in B_\delta(\theta_0)\right) \geq \mathbb{P}^\epsilon\left(\sum_{j=1}^{J_\delta} K(x,\omega_j^*)\tilde{u}_j^* \in B'_\delta(\theta_0 + g^\epsilon)\right) \\
&\geq \mathbb{P}^\epsilon\left((u_j,\omega_j)_{j=1}^{J_\delta} : |u_j - \tilde{u}_j^*| < \eta, \ |\omega_j - \omega_j^*| < \eta'\right) \\
&= \frac{(\nu^\epsilon(\mathbb{R}\times\Omega))^{J_\delta} e^{-\nu^\epsilon(\mathbb{R}\times\Omega)}}{J_\delta!}\left[\frac{\nu^\epsilon\left\{(u_j,\omega_j)_{j=1}^{J_\delta} : |u_j - \tilde{u}_j^*| < \eta, \ |\omega_j - \omega_j^*| < \eta'\right\}}{\nu^\epsilon(\mathbb{R}\times\Omega)}\right]. \qquad (3.23)
\end{aligned}$$

Since $|\tilde{u}_j| > \epsilon$ and Supp $(\nu^\epsilon) = (-\epsilon, \epsilon)^C \times \Omega^\epsilon$, for small enough ϵ (when $\epsilon < \frac{\delta}{4J_\delta \kappa}$ and Ω^ϵ contains all the finite number of points $\{\omega_j\}$)

$$\nu^\epsilon \Big\{ (u_j, \omega_j) : |u_j - \tilde{u}_j^*| < \eta,\ |\omega_j - \omega_j^*| < \eta' \Big\} > 0. \tag{3.24}$$

Therefore for small enough ϵ, $\Pi^\epsilon(B_\delta^C(\theta_0)) > 0$.

If conditions A or B are met, Theorem (2.4) implies that Π^ϵ converges weakly to Π in $C(\mathcal{X})$. For any δ, since $B_\delta^C(\theta_0)$ is a closed subset of $C(\mathcal{X})$ by Portmanteau theorem ([Billingsley 1999], Theorem 2.1),

$$0 < \limsup \Pi^\epsilon(B_\delta^C(\theta_0)) \leq \Pi(B_\delta^C(\theta_0)).$$

The theorem follows by observing that for any $\delta' < \delta$, $B_{\delta'}^C(\theta_0) \subset \Big\{ \theta : \|\theta_0 - \theta\|_\infty \leq \delta' \Big\} \subset B_\delta(\theta_0)$. ■

REMARK 3.5. *In the above theorem hypothesis (2) can be weakened by the following condition:* $\mathbb{R} \times \mathcal{X} = G[Supp\ \nu(du\, d\omega), +]$, *where* $G[Supp\ \nu(du\, d\omega), +]$ *is the additive group generated by the set Supp* $\nu(du\, d\omega)$.

REMARK 3.6. *If* Ω *is bounded (not necessarily compact) then the uniform continuity of* K *implies that* $\kappa = \sup K(x, \omega) < \infty$.

REMARK 3.7. *We considered the closure of* Θ_D *in the uniform topology. Our technique is applicable to other topologies as well. However, one would require a result analogous to Theorem (2.4). This would boil down to showing the tightness of the sequence* $\{\Pi^\epsilon\}$ *in that topology. Depending upon the topology, to show tightness, different assumptions about the Kernel* K *will be needed. See examples section.*

4. Examples.

In this section we consider some concrete examples in which the results from the previous section can be applied.

4.1. Poisson Regression.

We prove the posterior consistency of the Poisson regression model extending the results of Ghosal and Roy [2006]. Then we use the results of the previous section (Theorem 3.3) to show that the Lévy random field prior satisfies the required hypothesis for the posterior consistency to hold. The standard proof technique requires one to verify two conditions: positivity of prior mass for all information-metric (or equivalently Kullback-Leibler) neighborhoods of the true parameter, and existence of point-separating hypothesis tests with exponentially decaying error rates of types I and II. The first condition is prior-specific, and is relatively easy to verify. The existence of tests is independent of the prior and usually requires the calculation of the metric entropy of the parameter space. Since the Poisson distribution has support over the entire non-compact set of positive integers, we introduce a truncation approach, then utilize the Monotone Likelihood Ratio (MLR) property of the Poisson family to obtain the required exponential error rates.

Let $\mathcal{X}$ be a compact subset of $\mathbb{R}^d$. For any vector $k = (k_1, k_2, \cdots, k_d) \in \mathbb{N}^d$, let D^k denote the differential operator,

$$D^k \equiv \frac{\partial^{k_+}}{\partial x_1^{k_1} \cdots \partial x_d^{k_d}},$$

where $k_+ \equiv \sum_{i=1}^d k_i$. Denote by $\lfloor \alpha \rfloor$ the greatest integer less than or equal to any $\alpha \in \mathbb{R}_+$. For a function $f : \mathcal{X} \mapsto \mathbb{R}$, set

$$\|f\|_\alpha \equiv \max_{\tilde{k} \le \lfloor \alpha \rfloor} \sup_{x \in \mathcal{X}} |D^k f(x)| + \max_{\tilde{k} \le \lfloor \alpha \rfloor} \sup_{x,y \in \mathcal{X}} \frac{|D^k f(x) - D^k f(y)|}{||x-y||^{\alpha - \lfloor \alpha \rfloor}}, \tag{4.1}$$

where the supremum is taken over all x, y in the interior of $\mathcal{X}$ with $x \neq y$. Let $C^\alpha(\mathcal{X})$ be the set of all continuous functions $f : \mathcal{X} \mapsto \mathbb{R}$ with $\|f\|_\alpha < \infty$ and, for $\gamma > 0$, set

$$\Theta_\gamma \equiv \{\theta \in C^\alpha(\mathcal{X}) : \quad (\forall x)\ \theta(x) \ge \gamma\},$$

the set of C^α functions bounded below uniformly by $\gamma > 0$, with Borel σ-field $\mathcal{B}(\Theta_\gamma)$. We consider the following Poisson regression model:

$$\mathbb{P}(Y_i = y | X_i, \theta) = \exp\big(-\theta(X_i)\big) \frac{\theta(X_i)^y}{y!}, \qquad \theta \in \Theta_\gamma,\ y \in \mathbb{Z}_+,\ i \in \mathbb{N}. \tag{4.2}$$

For $\theta \in \Theta_\gamma$ let $\mathbb{P}_\theta$ be the distribution of conditionally-independent $\{Y_i\}$ generated according to the model (4.2) with mean function θ. Let $\|\cdot\|_*$ denote the uniform norm in Θ_γ, and for any $\theta \in \Theta_\gamma$ and $\delta > 0$ define

$$B_\delta(\theta) \equiv \{\tilde{\theta} \in \Theta_\gamma : \|\tilde{\theta} - \theta\|_* < \delta\}.$$

For any increasing sequence of positive numbers $\{M_n\}$ (we will choose a specific sequence in Section 4.3 below), set

$$\Theta_\gamma^n \equiv \{\theta \in \Theta_\gamma : \|\theta\|_\alpha \le M_n\}. \tag{4.3}$$

THEOREM 4.1. *Let Π be a (prior) probability measure on $\mathcal{B}(\Theta_\gamma)$ such that $\Pi(B_\delta(\theta)) > 0$ for each $\theta \in \Theta_\gamma$ and $\delta > 0$. Then for any $\epsilon > 0$, and any $\theta_0 \in \Theta_\gamma$,*

$$\lim_{n\to\infty} \Pi\left(\left\{ \theta \in \Theta_\gamma^n : \int_{\mathcal{X}} |\theta(x) - \theta_0(x)| \, Q_n(dx) > \epsilon \right\} \middle| Y_1, Y_2, \cdots, Y_n \right) = 0$$

in $\mathbb{P}_{\theta_0}$ probability, where $Q_n(dx) \equiv \frac{1}{n}\sum_{i=1}^n \delta_{X_i}(dx)$ denotes the empirical distribution.

Proof. Fix $\epsilon > 0$ and $\theta_0 \in \Theta_\gamma$, and let

$$\begin{aligned}\Lambda_i(\theta_0, \theta) &\equiv \log \frac{\exp\big(-\theta_0(X_i)\big)\big(\theta_0(X_i)\big)^{Y_i}}{\exp\big(-\theta(X_i)\big)\big(\theta(X_i)\big)^{Y_i}} \\ &= \big[\theta(X_i) - \theta_0(X_i)\big] + Y_i \log \frac{\theta_0(X_i)}{\theta(X_i)}\end{aligned}$$

denote the log ratio of two Poisson densities with means $\theta_0(X_i)$ and $\theta(X_i)$, evaluated at Y_i. Set

$$K_i(\theta_0, \theta) \equiv \mathbb{E}_{\theta_0}\big(\Lambda_i(\theta_0, \theta)\big)$$

(the Kullback-Leibler divergence for observation i), and

$$V_i(\theta_0, \theta) \equiv \mathsf{Var}_{\theta_0}\big(\Lambda_i(\theta_0, \theta)\big).$$

Also set

$$U_n^\epsilon \equiv \left\{ \theta \in \Theta_\gamma : \int |\theta(x) - \theta_0(x)| \; Q_n(dx) > \epsilon \right\}.$$

Our goal is to verify the following two conditions:

I) Positive prior probability: There exists a set $B \subset \Theta_\gamma$ with $\Pi(B) > 0$ such that

1. $(\forall \theta \in B)\ \sum_{i=1}^\infty V_i(\theta_0, \theta)/i^2 < \infty$.
2. $(\forall i \in \mathbb{N})\ \Pi\big(B \cap \{\theta : K_i(\theta_0, \theta) < \epsilon\}\big) > 0$.

II) Existence of tests: There exist non-negative bounded measurable functions $\{\Phi_n\}_{n=1}^\infty$ and positive constants $C_1, c_1 > 0$ such that

1. $\sum_{n=1}^\infty \mathbb{E}_{\theta_0}\Phi_n < \infty$.
2. $(\forall \theta \in U_n^\epsilon \cap \Theta_\gamma^n)\ \mathbb{E}_\theta(1 - \Phi_n) \le C_1 e^{-c_1 n}$.

By Theorem 1 of [Choi and Schervish 2007], these conditions imply that

$$\lim_{n\to\infty} \Pi\big(U_n^\epsilon \cap \Theta_\gamma^n | Y_1, Y_2, \cdots, Y_n\big) = 0, \quad \text{in } \mathbb{P}_{\theta_0} \text{ probability.}$$

4.2. *Positive prior probability.* First we consider Condition I. Fix any $\delta_0 > 0$ and set $B \equiv B_{\delta_0}(\theta_0)$. Then for any $\theta \in B$,

$$V_i(\theta_0, \theta) = \theta_0(X_i)\left[\log \frac{\theta_0(X_i)}{\theta(X_i)}\right]^2 \le \|\theta_0\|_* \left[\log \frac{\|\theta_0\|_* + \delta_0}{\gamma}\right]^2,$$

uniformly in $i \in \mathbb{N}$, so $\sum_{i=1}^{\infty} V_i(\theta_0, \theta)/i^2 < \infty$, for all $\theta \in B$ and I.1 holds.

Notice that

$$\begin{aligned} K_i(\theta_0, \theta) &= \theta(X_i) - \theta_0(X_i) + \theta_0(X_i) \log \frac{\theta_0(X_i)}{\theta(X_i)} \\ &= \theta(X_i)\left(1 - r + r \log r\right), \end{aligned}$$

where $r \equiv \frac{\theta_0(X_i)}{\theta(X_i)}$. The inequality $\log r \le r - 1$ implies that

$$\begin{aligned} K_i(\theta_0, \theta) &\le \theta(X_i)\big(1 - r + r(r-1)\big) = \theta(X_i)(r-1)^2 \\ &= \frac{\big(\theta(X_i) - \theta_0(X_i)\big)^2}{\theta(X_i)} \le \frac{1}{\gamma}\|\theta - \theta_0\|_*^2. \end{aligned}$$

Since $\Pi\big(B_\delta(\theta_0)\big) > 0$ for each $\delta > 0$ by hypothesis, and in particular for $\delta = \big(\delta_0 \wedge \sqrt{\gamma\epsilon}\big)$, condition I.2 is verified since $B_\delta(\theta_0) \subset B \cap \{\theta : K_i(\theta_0, \theta) < \epsilon\}$.

4.3. *Existence of tests.* Now we turn to the second condition. First we construct tests for the simple *vs.* simple hypotheses:

$$H_0 : Y_i \sim \mathsf{Po}\big(\theta_0(X_i)\big) \;\; vs. \;\; H_1 : Y_i \sim \mathsf{Po}\big(\theta_1(X_i)\big).$$

Below we construct test functions for testing a more general class of hypotheses, including those involving stochastically ordered positive random variables. For an integrable positive random variable Y with mean μ and an integer $n \in \mathbb{N}$, denote by $\mu^{(n)}$ the mean of the truncated random variable

$$\mu^{(n)} = \mathbb{E}\big[Y \wedge n\big]$$

where $\mathbf{1}_{\{A\}}$ represents the indicator function for the event A. For a Poisson-distributed random variable this is $\mu^{(n)} = n - [\Gamma(n+1, \mu) - \mu\Gamma(n, \mu)]/\Gamma(n) \le (\mu \wedge n)$, where $\Gamma(a, x)$ denotes the incomplete Gamma function [Abramowitz and Stegun 1974, § 6.5.3]. In particular, we define the "truncated mean" of any $\theta \in \Theta_\gamma$ to be

$$\theta^{(n)}(X_i) \equiv \mathbb{E}_\theta\big[Y_i \wedge n\big].$$

LEMMA 4.2. *Let $\{Y_i\}$ be independent positive random variables with means $\{\mu_i\}$, and let $\{X_i\} \subset \mathcal{X}$. Consider the following hypothesis testing problem:*

$$H_0 : \mu_i = \theta_0(X_i) \text{ vs. } H_1 : \mu_i = \theta_1(X_i)$$

for fixed elements $\theta_0, \theta_1 \in \Theta_\gamma$ that satisfy the inequality

$$\theta_1(x) > \theta_0(x) + \epsilon$$

for some number $\epsilon > 0$ and all $x \in \mathcal{X}$. Consider the sequence of indicator random variables,

$$\Phi_n \equiv \mathbf{1}_{\{A_n\}}, \qquad A_n \equiv \left\{ \frac{1}{n^2} \sum_{i=1}^{n} [(Y_i \wedge n) - \theta_0^{(n)}(X_i)] > \epsilon \right\}.$$

If the distributions of $\{Y_i\}$ are from a location family with location parameters μ_i, or they are absolutely continuous with respect to some reference measure and have density functions $\{p_{\mu_i}(y)\}$ with a monotone likelihood ratio (MLR), then for all sufficiently large n,

$$\mathbb{E}_{\theta_0}(\Phi_n) \leq \exp(-n\epsilon^2/2) \tag{4.4a}$$

$$\mathbb{E}_{\theta_1}(1 - \Phi_n) \leq \exp(-n\epsilon^2/2). \tag{4.4b}$$

Proof. The proof is deferred to the Appendix. ∎

REMARK 4.3. *In the above lemma if we had, $\theta_0(x) > \theta_1(x) + \epsilon$, with all the other conditions being the same, then the test*

$$\Phi_n = \mathbf{1}_{\{B_n\}}, \; B_n \equiv \left\{ \frac{1}{n^2} \sum_{i=1}^{n} [(Y_i \wedge n) - \theta_1^{(n)}(X_i)] < \epsilon \right\}$$

will satisfy (4.4), for all sufficiently large n (see Appendix for the details of the proof).

Since the Poisson family has an MLR, the above Lemma certainly applies to the case of testing $H_0 : Y_i \sim \mathsf{Po}(\theta_0(X_i))$ *vs.* $H_1 : Y_i \sim \mathsf{Po}(\theta_j(X_i))$ for any fixed $\theta_j \in \Theta_\gamma$. The rest of our proof is based on a standard argument using covering numbers [Ghosal and Roy 2006]. First notice that for any $\theta_j \in \Theta_\gamma$ with $\int |\theta_j(x) - \theta_0(x)| Q_n(dx) > \epsilon$,

$$Q_n\{x : |\theta_j(x) - \theta_0(x)| > \epsilon/2\} \geq \frac{\epsilon}{2(\|\theta_0\|_* + \|\theta_j\|_*)}.$$

This can be seen from:

$$
\begin{aligned}
\epsilon &< \int |\theta_j(x) - \theta_0(x)|\, Q_n(dx) \\
&< (\|\theta_0\|_* + \|\theta_j\|_*)\, Q_n\{x : |\theta_j(x) - \theta_0(x)| > \epsilon/2\} + \epsilon/2.
\end{aligned}
$$

Hence the cardinality $\#\{i : |\theta_j(X_i) - \theta_0(X_i)| > \epsilon/2\} \geq \tilde{C}n$ for the constant $\tilde{C} \equiv \epsilon(2\|\theta_0\|_* + 2\|\theta_j\|_*)^{-1}$. This implies that one of the two sets $\{i : \theta_j(X_i) > \theta_0(X_i) + \epsilon/2\}$ or $\{i : \theta_0(X_i) > \theta_j(X_i) + \epsilon/2\}$ will have at least $\tilde{C}n/2$ points. For definiteness we will assume it is the former and set

$$
S^+_{\theta_j} \equiv \{x : \theta_j(x) > \theta_0(x) + \epsilon/2\} \tag{4.5}
$$

(otherwise see Remark 4.3, and make the corresponding change in equation (4.6) below). Now for testing θ_0 *vs.* θ_j such that $\int |\theta_j(x) - \theta_0(x)|\, Q_n(dx) > \epsilon$, we construct the test:

$$
\Psi_{nj} = \mathbf{1}_{\{E_{nj}\}},\ E_{nj} \equiv \left\{ \frac{1}{n^2} \sum_{i=1}^{n} \left[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\right] \mathbf{1}_{\{X_i \in S^+_{\theta_j}\}} > \epsilon/2 \right\}. \tag{4.6}
$$

Let $N \equiv N(\epsilon/2, \Theta^n_\gamma, \|\cdot\|_*)$ denote the covering number of the set Θ^n_γ under the uniform norm, *i.e.*, the smallest number N such that

$$
\Theta^n_\gamma \subset \bigcup_{j=1}^{N} B_{\epsilon/2}(\theta_j)
$$

for some $\{\theta_1, \theta_2, \cdots, \theta_N\} \subset \Theta^n_\gamma$. Then by Theorem 2.7.1 of van der Vaart and Wellner [1996],

$$
\log N \leq K_1 M_n^{d/\alpha} \epsilon^{-1/\alpha} \tag{4.7}
$$

for some constant $K_1 < \infty$. For each $\theta \in \Theta^n_\gamma$ there is some $j \leq N$ such that $\|\theta - \theta_j\|_* \leq \epsilon/2$. Recall the tests $\{\Psi_{nj}\}_{j=1}^N$, from (4.6). By Lemma 4.2 (with $\epsilon/2$ replacing ϵ),

$$
\begin{aligned}
\mathbb{E}_{\theta_0}(\Psi_{nj}) &\leq \exp(-n\epsilon^2/8), \\
\mathbb{E}_{\theta_j}(1 - \Psi_{nj}) &\leq \exp(-n\epsilon^2/8),
\end{aligned}
\qquad 1 \leq j \leq N.
$$

Now set $\Psi_n \equiv \max\{\Psi_{n,\theta_j} :\ 1 \leq j \leq N\}$. Then,

$$
\mathbb{E}_{\theta_0}(\Psi_n) \leq \sum_{j=1}^{N} \mathbb{E}_{\theta_j}(\Psi_{nj}) \leq \exp(\log N - n\epsilon^2/8). \tag{4.8}
$$

If we set $M_n \equiv K_2 n^{\alpha/d}$ (see equation (4.3)) for sufficiently small K_2, then equations (4.7, 4.8) imply

$$
\mathbb{E}_{\theta_0}(\Psi_n) \leq \exp(-K_3 n)
$$

for some constant $K_3 > 0$. In particular, $\sum_{n=1}^{\infty} \mathbb{E}_{\theta_0}(\Psi_n) < \infty$, verifying II.1. Now we turn to II.2.

For any $\theta \in U_n^{\epsilon} \cap \Theta_{\gamma}^n$, there is at least one $j \in \{1, 2, \cdots, N\}$ such that $\|\theta - \theta_j\|_* < \epsilon/2$. For $X_i \in S_{\theta_j}^+$ (see equation (4.5)), by the triangle inequality, $[\theta(X_i) - \theta_0(X_i)] \geq \epsilon/2$. Hence

$$\mathbb{E}_{\theta}(1 - \Psi_n) \leq \mathbb{E}_{\theta}(1 - \Psi_{nj}) \leq \exp(-n\epsilon^2/8).$$

Since this holds uniformly for all $\theta \in U_n^{\epsilon} \cap \Theta_{\gamma}^n$,

$$\sup_{\theta \in U_n^{\epsilon} \cap \Theta_{\gamma}^n} \mathbb{E}_{\theta}(1 - \Psi_n) \leq \exp(-n\epsilon^2/8)$$

verifying II.2. Hence we have proved the theorem. ■

REMARK 4.4. *If the prior* Π *also satisfies the condition* $\Pi(\Theta_{\gamma}^{nc}) \leq C_2 e^{-c_2 n}$ *for constants* $C_2, c_2 > 0$, *then the conclusion of Theorem 4.1 can be strengthened to:*

$$\lim_{n\to\infty} \Pi \left\{ \left[\theta \in \Theta_{\gamma} : \int_{\mathcal{X}} |\theta(x) - \theta_0(x)| \, Q_n(dx) > \epsilon \right] \Big| Y_1, Y_2, \cdots, Y_n \right\} = 0$$

in $\mathbb{P}_{\theta_0}$ *probability.*

REMARK 4.5. *The predictors* $\{X_i\}$, *instead of being fixed, can be sampled according to a distribution* D *on* $\mathcal{X}$. *With similar arguments as above, we can show that consistency holds in* L_1 *norm, instead of the empirical* L_1 *norm. See [Ghosal and Roy 2006] for a result of this type.*

REMARK 4.6. *It is interesting to note that we require the parameter space* Θ_{γ} *to be bounded away from zero. This is not a mere artifact of our proof: if almost all observations in a sample from a Poisson distribution are* $Y_i = 0$, *then it is hard to estimate the log mean if it is not bounded away from zero. We expect that our approach may be extended to obtain bounds on convergence rates for nonparametric Poisson regression models.*

4.3.1. *Lévy Random Field Priors.* Let $K(x, \omega) : \mathcal{X} \times \Omega \mapsto \mathbb{R}$ be a Borel measurable function on $\mathcal{X} \times \Omega$ such that the set

$$\left\{ \sum_{n=1}^{N} K(x, \omega_j) u_j, \ \{u_j, \omega_j\} \in \mathbb{R} \times \Omega, \ N \in \mathbb{N} \right\}$$

is dense in $C(\mathcal{X})$ in the uniform topology. Let ν be a positive sigma-finite measure on $\mathbb{R} \times \Omega$ satisfying the integrability condition (3.1). Consider the following prior $\Pi_L(d\theta)$ on Θ_γ:

$$\theta(x) = \int_\Omega K(x, \omega)\mathcal{L}(d\omega) \tag{4.9}$$

$$\mathcal{L}(d\omega) \sim \text{Lévy}(\nu). \tag{4.10}$$

Let $\Pi_L(\cdot|Y_1, Y_2, \cdots, Y_n)$ denote the corresponding posterior conditioned on the observations $\{Y_i\}$ sampled according to the Poisson regression model given in (4.2).

THEOREM 4.7. *If either condition A or B holds true, for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \Pi_L \left\{ \left[\theta \in \Theta_\gamma^n : \int_\mathcal{X} |\theta(x) - \theta_0(x)| \, Q_n(dx) > \epsilon \right] \Big| Y_1, Y_2, \cdots, Y_n \right\} = 0 \tag{4.11}$$

in $\mathbb{P}_{\theta_o}$ probability.

Proof. If condition A or B holds true, by theorem (3.3) it follows that $\Pi_L(B_\delta(\theta_0)) > 0$ for any $\theta_0 \in \Theta_\gamma$ and $\delta > 0$. Hence the result follows from theorem (4.1). ■

4.4. **Regression with Gaussian Errors.** Let $\Theta \equiv C^\alpha(\mathcal{X}), \alpha > 0$. We consider the following regression model on Θ. Let $X_i \in \mathcal{X} \subset \mathbb{R}$ and

$$Y_i \big| \theta, X_i = \theta(X_i) + \epsilon_i, \;\; \theta_0 \in \Theta, \;\; \epsilon_i \overset{iid}{\sim} N(0, \sigma^2). \tag{4.12}$$

Here the error variance $\sigma > 0$ is assumed to be unknown. Let $\Pi_\sigma(d\sigma)$ denote a (prior) probability distribution whose support is the postive real line. Let $\Pi_L(d\theta)$ be the Lévy random field prior on Θ defined in (4.9). Let $\Pi_L(d\theta) \otimes \Pi_\sigma(d\sigma)$ be the prior distribution on $\Theta \times \mathbb{R}_+$ and $\Pi_{L,\sigma}(\cdot|Y_1, Y_2, \cdots, Y_n)$ denote the corresponding posterior distribution. Define the sets

$$\Theta^n \equiv \left\{ \theta \in \Theta : \|\theta\|_\alpha \le n^\beta, \;\; \beta \in (1/2, 1] \right\}$$

and set $S_{n,\epsilon}$

$$S_{n,\epsilon} \equiv \left\{ \theta \in \Theta^n : \int_\mathcal{X} |\theta(x) - \theta_0(x)| \, Q_n(dx) < \epsilon, \; \left| \frac{\sigma}{\sigma_0} - 1 \right| < \epsilon \right\}, \;\; Q_n(dx) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}(dx). \tag{4.13}$$

THEOREM 4.8. *Let $\mathbb{P}_{\theta_0,\sigma_0}$ denote the distribution of $\{Y_i\}$ conditional on (θ_0, σ_0). If conditions A or B hold true, for any $\theta_0 \in \Theta$ and for any $\epsilon > 0$*

$$\lim_{n\to\infty} \Pi_{L,\sigma}\Big(S_{n,\epsilon}^C \;\Big|\; Y_1, Y_2, \cdots, Y_n \Big) = 0$$

in $\mathbb{P}_{\theta_0,\sigma_0}$ probability.

Proof. Direct calculations yield that (also see [Choi and Schervish 2007, pg. 1974])

$$K_i(\theta_0,\theta) = \frac{1}{2}\log\frac{\sigma^2}{\sigma_0^2} - \frac{1}{2}\left(1 - \frac{\sigma_0^2}{\sigma^2}\right) + \frac{1}{2}\frac{\left[\theta_0(X_i) - \theta(X_i)\right]^2}{\sigma^2}$$

and

$$V_i(\theta_0,\theta) = \frac{1}{2}\left(\frac{\sigma_0^2}{\sigma^2} - 1\right)^2 + \frac{\sigma_0^4}{\sigma^4}\Big(\theta_0(X_i) - \theta(X_i)\Big)^2.$$

Define

$$B_\delta^{\theta_0,\sigma_0} \equiv \left\{(\theta,\sigma) : \|\theta - \theta_0\|_* < \delta, \left|\frac{\sigma}{\sigma_0} - 1\right| < \delta\right\}.$$

Hence the above calculations show that for every $\epsilon > 0$, there exists a $\delta > 0$ such that for all $(\theta,\sigma) \in B_\delta^{\theta_0,\sigma_0}$,

$$K_i(\theta_0,\theta) < \epsilon \;\forall i, \;\; \sum_{i=1}^{\infty}\frac{V_i(\theta_0,\theta)}{i^2} < \infty.$$

Hence conditions I.1 and I.2 are satisfied if the prior probability of the set $B_\delta^{\theta_0,\sigma_0}$ is positive for every $\delta > 0$. If either condition A or B is satisfied, theorem (3.3) gaurantees that $\Pi_L(B_\delta(\theta_0)) > 0$. Since the prior $\Pi_\sigma(d\sigma)$ has support over the entire real line, $\Pi_\sigma\left(\left|\frac{\sigma}{\sigma_0} - 1\right| < \delta\right) > 0$. Hence for every $\delta > 0$, $\Pi_{L,\sigma}\left(B_\delta^{\theta_0,\sigma_0}\right) > 0$ and conditions I.1 and I.2 are satisfied.

The existence of tests as required in II.1 and II.2 are constructed in [Choi and Schervish 2007, Theorem 2, pg. 1973]. Hence the theorem is proved. ∎

Remark 4.9. *If the covariates X_i were sampled uniformly from $\mathcal{X}$ and σ_0 is known, it follows that*

$$K_i(\theta_0,\theta) = \frac{1}{2\sigma_0^2}\;\|\theta_0 - \theta\|_{L_2(\mathcal{X})} \le \frac{1}{2\sigma_0^2}\;\|\theta_0 - \theta\|_* \tag{4.14}$$

Hence the results of the previous case hold in this case as well. However, condition B for the function $K(\cdot,\cdot)$ is too strong. Notice that theorem (3.3) gives that $\Pi_L(\theta : \|\theta_0 - \theta\|_* < \delta) > 0$ whereas we only need $\Pi_L(\theta : \|\theta_0 - \theta\|_{L_2(\mathcal{X})} < \delta) > 0$. To show this and thereby relax the assumptions on condition B, we will have to show the weak convergence of probability measures on $L_2(\mathcal{X})$ (instead of $C(\mathcal{X})$ as in theorem (3.2)) induced by the truncated compound Poisson Lévy measures.

5. Appendix. Proof of Lemma 4.2. Notice that

$$\mathbb{E}_{\theta_0}(\Phi_n) = \mathbb{P}_{\theta_0}\left\{\sum_{i=1}^{n}\left[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\right] > n^2\epsilon\right\}.$$

Since $\mathbb{E}_{\theta_0}[(Y_i \wedge n) - \theta_0^{(n)}(X_i)] = 0$ and $-n \le [(Y_i \wedge n) - \theta_0^{(n)}(X_i)] \le n$, by Hoeffding's inequality [Hoeffding 1963, Theorem 1]

$$\mathbb{P}_{\theta_0}\left\{\sum_{i=1}^{n}\left[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\right] > n^2\epsilon\right\} \le \exp\big(-2n^4\epsilon^2/n(2n)^2\big) = \exp(-n\epsilon^2/2).$$

To obtain the type II error rate of equation (4.4b), first note

$$\begin{aligned}
\mathbb{E}_{\theta_1}(1-\Phi_n) &= \mathbb{P}_{\theta_1}\left\{\sum_{i=1}^{n}\left[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\right] \le n^2\epsilon\right\} \\
&= \mathbb{P}_{\theta_1}\left\{\sum_{i=1}^{n}\left[(Y_i \wedge n) - \theta_1^{(n)}(X_i)\right] + \sum_{i=1}^{n}\left[\theta_1^{(n)}(X_i) - \theta_0^{(n)}(X_i)\right] \le n^2\epsilon\right\} \\
&= \mathbb{P}_{\theta_1}\left\{-\sum_{i=1}^{n}\left[(Y_i \wedge n) - \theta_1^{(n)}(X_i)\right]\right. \\
&\qquad\qquad \left. \ge -n^2\epsilon + \sum_{i=1}^{n}\left[\theta_1^{(n)}(X_i) - \theta_0^{(n)}(X_i)\right]\right\}. \qquad (5.1)
\end{aligned}$$

Since $\theta_0(X_i) + \epsilon < \theta_1(X_i)$ for all $i \in \mathbb{N}$, under the (stated) hypothesis that θ_i is the location parameter for the distribution of Y_i, or that the family of p.d.f.'s have a Monotone Likelihood Ratio in θ_i, we have $\mathbb{P}_{\theta_0(X_i)}(Y_i \le c) \ge \mathbb{P}_{\theta_1(X_i)}(Y_i \le c)$ for each $c > 0$ and, since the function $f(y) \equiv (y \wedge n)$ is nondecreasing,

$$\theta_0^{(n)}(X_i) \equiv \mathbb{E}_{\theta_0(X_i)}[(Y_i \wedge n)] \le \mathbb{E}_{\theta_1(X_i)}[(Y_i \wedge n)] \equiv \theta_1^{(n)}(X_i) \qquad (5.2)$$

for all $n \in \mathbb{N}$ and all $i \in \mathbb{N}$. Thus each $[\theta_1^{(n)}(X_i) - \theta_0^{(n)}(X_i)] \ge 0$ is positive, so the inequality (5.1) can be bounded by

$$\begin{aligned}
\mathbb{E}_{\theta_1}(1-\Phi_n) &\le \mathbb{P}_{\theta_1}\left\{-\sum_{i=1}^{n}\left[(Y_i \wedge n) - \theta_1^{(n)}(X_i)\right] \ge -n^2\epsilon\right\} \\
&\le \exp(-n\epsilon^2/2) \qquad (5.3)
\end{aligned}$$

by another application of Hoeffding's inequality.

Proof of Remark 4.3. In this case the type II error rate follows from the following simple application of Hoeffding's inequality:

$$\mathbb{E}_{\theta_1}(1-\Phi_n) = \mathbb{P}_{\theta_1}\left\{\sum_{i=1}^n \left[(Y_i \wedge n) - \theta_1^{(n)}(X_i)\right] > n^2\epsilon\right\}$$
$$\leq \exp(-2n^4\epsilon^2/n(2n)^2) = \exp(-n\epsilon^2/2).$$

For the type I error,

$$\mathbb{E}_{\theta_0}(\Phi_n) = \mathbb{P}_{\theta_0}\left\{\sum_{i=1}^n \left[(Y_i \wedge n) - \theta_1^{(n)}(X_i)\right] \leq n^2\epsilon\right\}$$
$$= \mathbb{P}_{\theta_0}\left\{-\sum_{i=1}^n \left[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\right]\right.$$
$$\left. \geq -n^2\epsilon + \sum_{i=1}^n \left[\theta_0^{(n)}(X_i) - \theta_1^{(n)}(X_i)\right]\right\}. \tag{5.4}$$

Once again the differences $\left[\theta_0^{(n)}(X_i) - \theta_1^{(n)}(X_i)\right]$ will be positive, so equation (5.4) can be bounded above by

$$\mathbb{E}_{\theta_0}(\Phi_n) \leq \mathbb{P}_{\theta_0}\left\{-\sum_{i=1}^n \left[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\right] \geq -n^2\epsilon\right\} \leq \exp(-n\epsilon^2/2)$$

by another application of Hoeffding's inequality. ∎

References.

Abramowitz, M. and Stegun, I. A., eds. (1974). *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*. Dover, New York, NY.

Albiac, F. and Kalton, N. J. (2006). *Topics in Banach space theory*, *Graduate Texts in Mathematics*, vol. 233. Springer-Verlag, New York, NY.

Amewou-Atisso, M., Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (2003). Posterior consistency for semi-parametric regression problems. *Bernoulli* **9** 291–312.

Barron, A., Schervish, M. J. and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Stat.* **27** 536–561.

Billingsley, P. J. (1999). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, New York, NY, second edn.

Choi, T. and Schervish, M. J. (2007). On posterior consistency in nonparametric regression problems. *J. Multivariate Anal.* **98** 1969–1987.

Clyde, M. A. and Wolpert, R. L. (2007). Estimation using oveercomplete dictionaries. In *Bayesian Statistics 8* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.), 91–114. Oxford University Press, Oxford, UK.

Diaconis, P. and Freedman, D. A. (1986a). On inconsistent Bayes estimates of location. *Ann. Stat.* **14** 68–87.

Diaconis, P. and Freedman, D. A. (1986b). On the consistency of Bayes estimates. *Ann. Stat.* **14** 1–67.

Dunford, N. and Schwartz, J. T. (1988). *Linear operators. Part I.* Wiley Classics Library, John Wiley & Sons, New York, NY.

Gaigalas, R. (2004a). A Poisson bridge between fractional Brownian motion and stable Lévy motion. Tech. rep., Department of Mathematics, Uppsala University.

Gaigalas, R. (2004b). The "polish" integral with respect to compensated Poisson random measures. Personal communication.

Ghosal, S. and Roy, A. (2006). Posterior consistency of Gaussian process prior for nonparametric binary regression. *Ann. Stat.* **34** 2413–2429.

Ghosal, S. and van der Vaart, A. W. (2007a). Convergence rates of posterior distributions for noniid observations. *Ann. Stat.* **35** 192–223.

Ghosal, S. and van der Vaart, A. W. (2007b). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Stat.* **35** 697–723.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58** 13–30.

iti Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*, *Cambridge Studies in Advanced Mathematics*, vol. 68. Cambridge Univ. Press, Cambridge, UK.

Nieto-Barajas, L. E., Prünster, I. and Walker, S. G. (2004). Normalized random measures driven by increasing additive processes. *Ann. Stat.* **32** 2343–2360.

Rajput, B. S. and Rosiński, J. (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Rel.* **82** 451–487.

Reed, M. C. and Simon, B. (1980). *Methods of Modern Mathematical Physics. I. Functional Analysis*. Academic Press, New York, NY, second edn.

Taqqu, M. S. and Wolpert, R. L. (1983). Infinite variance self-similar processes subordinate to a Poisson measure. *Z. Wahrscheinlichkeit.* **62** 53–72.

Tu, C., Clyde, M. A. and Wolpert, R. L. (2006). Lévy adaptive regression kernels. Discussion Paper 2006-08, Duke University Department of Statistical Science.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics, Springer-Verlag, New York, NY.

Wehausen, J. V. (1938). Transformations in linear topological spaces. *Duke Math. J.* **4** 157–169.

Wolpert, R. L. and Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85** 251–267.

Wolpert., R. L. and Ickstadt, K. (1998). Simulation of Lévy random fields. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (D. K. Dey, P. Müller and D. Sinha, eds.), *Lecture Notes in Statistics*, vol. 133, 305–322. Springer-Verlag, New York, NY.

Wolpert, R. L. and Ickstadt, K. (2004). Reflecting uncertainty in inverse problems: A Bayesian solution using

Lévy processes. *Inverse Problems* **20** 1759–1771.

Wolpert, R. L., Ickstadt, K. and Hansen, M. B. (2003). A nonparametric Bayesian approach to inverse problems. In *Proceedings of the 7th Valencia International Meeting held in Tenerife, June 2–6, 2002* (J. M. Bernardo, A. P. Dawid, J. O. Berger, M. West, D. Heckerman, M. J. Bayarri and A. F. M. Smith, eds.), 403–417. Cambridge Univ. Press, New York, NY. With a discussion by Subhashis Ghosal and a rejoinder by the authors.

Wolpert, R. L. and Taqqu, M. S. (2005). Fractional Ornstein-Uhlenbeck Lévy processes and the Telecom process: Upstairs and downstairs. *Signal Processing* **85** 1523–1545.

Department of Statistical Sciences
Duke University, Durham
NC, 27708-0251
E-mail: nsp2@stat.duke.edu; rlw@stat.duke.edu