On the Supremum of Certain Families of Stochastic Processes

Wenbo V. Li [*] Natesh S. Pillai[†] Robert L. Wolpert [‡]

May 4, 2009

Summary

We consider a family of stochastic processes $\{X_t^\epsilon, t \in T\}$ on a metric space T, with a parameter $\epsilon \downarrow 0$. We study the conditions under which

$$\lim_{\varepsilon \to 0} \mathbb{P}\Big(\sup_{t \in T} |X_t^\varepsilon| < \delta\Big) = 1$$

when one has the *a priori* estimate on the modulus of continuity and the value at one point. We compare our problem to the celebrated Kolmogorov continuity criteria for stochastic processes, and finally give an application of our main result for stochastic intergrals with respect to compound Poisson random measures with infinite intensity measures.

Key words: Compensated Poisson random measure, Generic chaining, Kolmogorov continuity criterion, Metric entropy, Suprema of stochastic processes

1 Introduction

Let (T, d) be a metric space with finite diameter,

$$D(T) = \sup\Big\{d(s,t) : s, t \in T\Big\} < \infty.$$

Let $N(T, d, \delta)$ denote the covering number, *i.e.*, for every $\delta > 0$, let $N(T, d, \delta)$ denote the minimal number of d-balls of radius δ required to cover T. The supremum of a stochastic process X_t defined on T, $\sup_{t \in T} X_t$ can be quantified in terms of $N(T, d, \delta)$ (see (Talagrand, 2005, Chapter 1) for instance) under various assumptions on the process X_t.

[*] Department of Mathematical Sciences, University of Delaware, wli@math.udel.edu
[†] CRiSM, University of Warwick, N.Pillai@warwick.ac.uk
[‡] Department of Statistical Science, Duke University, rlw@stat.duke.edu

In this paper we consider a family of stochastic processes X_t^ϵ on T, with a parameter $\epsilon > 0$. In certain applications in nonparametric statistics (see section 4) it is of interest to study the limiting behaviour of the supremum, $\lim_{\epsilon\to 0}\sup_{t\in T} X_t^\epsilon$ when one has the a priori estimate

$$\mathbb{E}\,|X_t^\varepsilon - X_s^\varepsilon|^\beta \le B_\varepsilon\, d(s,t)^\gamma \tag{1.1}$$

where $\beta, \gamma > 0$ and $B_\varepsilon \to 0$. In particular, we would like to identify conditions under which

$$\lim_{\varepsilon\to 0} \mathbb{P}\Big(\sup_{t\in T} |X_t^\varepsilon| < \delta\Big) = 1 \tag{1.2}$$

for every $\delta > 0$ as $\epsilon \to 0$. In our main result in section 2, we find conditions in terms of the covering number $N(T,d,\delta)$, so that (1.2) holds. Although our technique is based on the well known chaining methods, our principle result appears to be new. In section 3 we discuss briefly the optimality of our hypotheses and compare our theorem with the celebrated Kolmogorov criteria for continuity of stochastic processes. In section 4 we present an application of our main theorem on random fields constructed from Lévy random measures.

2 Main Result

Let (T,d) be a complete separable metric space and $(X_t^\epsilon)_{t\in T}$ a family of real-valued, centered, L_2 stochastic processes on T, indexed by $\epsilon > 0$.

Theorem 1. *Suppose that:*

1. *There exists a point $t_0 \in T$ such that*

$$\lim_{\varepsilon\to 0} \mathbb{E}\,(X_{t_0}^\varepsilon)^2 = 0. \tag{2.3}$$

2. *There exist $\alpha, \beta > 0$ and positive numbers $\{B_\epsilon\}$ with $\lim_{\epsilon\to 0} B_\epsilon = 0$ such that for any $s,t \in T$*

$$\mathbb{E}\,|X_t^\varepsilon - X_s^\varepsilon|^\beta \le B_\varepsilon\, d(s,t)^{1+\alpha}. \tag{2.4}$$

3. *There exists $0 < \gamma < \alpha$ such that*

$$\int_0^{D(T)} a^\gamma N(T,d,a)\; da < \infty. \tag{2.5}$$

Then for any $\delta > 0$,

$$\lim_{\varepsilon\to 0} \mathbb{P}\Big(\sup_{t\in T} |X_t^\varepsilon| < \delta\Big) = 1$$

For each fixed $\epsilon > 0$, Equation (2.4) and Komogorov's continuity criterion (see **?**, p. 375) implies that there exists a path-continuous version of (X_t^ϵ) (see Section 3 for more on this connection). Equation (2.5) is an entropy criterion that controls the size of the index set T.

Proof. Fix $\delta > 0$. First observe that

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\in T}|X_t^\varepsilon| < \delta\Big) &\geq \mathbb{P}\Big(\sup_{t\in T}|X_t^\varepsilon - X_{t_0}^\varepsilon| < \delta/2, |X_{t_0}^\varepsilon| < \delta/2\Big) \\ &\geq \mathbb{P}\Big(\sup_{t\in T}|X_t^\varepsilon - X_{t_0}^\varepsilon| < \delta/2\Big) - \mathbb{P}\Big(|X_{t_0}^\varepsilon| \geq \delta/2\Big)\end{aligned}$$

and

$$\mathbb{P}(|X_{t_0}^\varepsilon| \geq \delta/2) \leq 4\delta^{-2}\mathbb{E}\,|X_{t_0}^\varepsilon|^2 \to 0$$

as $\varepsilon \to 0$ by equation (2.3). Thus we only need to control $\sup_{t\in T}|X_t^\varepsilon - X_{t_0}^\varepsilon|$.

Our argument is based on the so-called generic chaining principle (see Ledoux (1996) and Talagrand (2005)). Let n_0 be the largest integer n such that $N(T, d, 2^{-n}) = 1$ (note $n_0 < 0$ is possible). For every $n \geq n_0$, we consider a family of cardinality $N_n = N(T, d, 2^{-n})$ of balls of radius 2^{-n} covering T. We can theqrefore construct a partition $\mathcal{A}_n$ of T of cardinality $N_n = |\mathcal{A}_n|$ on the basis of this covering with sets of diameter less than 2^{-n+1}. In each A of $\mathcal{A}_n$, one can fix a point of T and denote by T_n the collection of these points (without loss of generality, let $T_{n_0} = \{t_0\}$ be the selected point in the single element of partition A_{n_0}). For each $t \in T$, denote by $\mathcal{A}_n(t)$ the element of $\mathcal{A}_n$ that contains t. For every t and every n, let $s_n(t)$ be the element of T_n such that $t \in \mathcal{A}_n(s_n(t))$. It is clear that $d(t, s_n(t)) \leq 2^{-n+1}$ for every $t \in T$ and $n \geq n_0$. We also know

$$d(s_n(t), s_{n-1}(t)) \leq d(t, s_n(t)) + d(t, s_{n-1}(t)) \leq 2^{-n+1} + 2^{-n+2} = 6 \cdot 2^{-n} \tag{2.6}$$

The fundamental relation is the convergent telescoping sum

$$X_t - X_{t_0} = \sum_{n>n_0} \big(X_{s_n(t)} - X_{s_{n-1}(t)}\big)$$

for every $t \in T$, where we note that $s_{n_0}(t) = t_0$ for every $t \in T$. Since the partition elements are nested, $s_{n-1}(t) = s_{n-1}\big(s_n(t)\big)$ for all $t \in T$ and $n \geq n_0$, so

$$\begin{aligned}\sup_{t\in T}|X_t - X_{t_0}| &\leq \sup_{t\in T}\sum_{n>n_0}|X_{s_n(t)} - X_{s_{n-1}(t)}| \\ &= \sup_{t\in T}\sum_{n>n_0}|X_{s_n(t)} - X_{s_{n-1}(s_n(t))}| \\ &\leq \sum_{n>n_0}\max_{v\in T_n}|X_v - X_{s_{n-1}(v)}|\end{aligned}$$

For $v \in T$, let $\{w_n(v)\}_{n\geq n_0}$ be a sequence of non-negative real numbers such that $\sum_{n\geq n_0} w_n(v) = 1$. Notice that for any $\delta > 0$,

$$\bigcap_{n>n_0}\bigcap_{v\in T_n}\Big\{|X_v - X_{s_{n-1}(v)}| \leq w_n(v)\delta/2\Big\} \subset \Big\{\sup_{t\in T}|X_t - X_{t_0}| \leq \delta/2\Big\}.$$

Theqrefore,

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in T}|X^\varepsilon_t - X^\varepsilon_{t_0}| > \delta/2\right) &\le \mathbb{P}\Big(\bigcup_{n>n_0}\bigcup_{v\in T_n}\Big\{|X^\varepsilon_v - X^\varepsilon_{s_{n-1}(v)}| > w_n(v)\delta/2\Big\}\Big) \\ &\le \sum_{n>n_0}\sum_{v\in T_n}\mathbb{P}\left(|X^\varepsilon_v - X^\varepsilon_{s_{n-1}(v)}| > w_n(v)\delta/2\right)\end{aligned}$$

where the last inequality follows from the union bound. Next we use Equations (2.4) and (2.5) to find optimal choices for $w_n(v)$ (the so-called "majorizing measure", see (Talagrand, 2005, Chapter 1)). Set

$$w_n \equiv w_n(v) \equiv (1-2^{-h})2^{-h(n-n_0)}, \quad h = (\alpha-\gamma)/\beta,\ v \in T.$$

Notice that $\sum_{n\ge n_0} w_n = 1$. By Markov's inequality and (2.4), for $v \in T_n$,

$$\begin{aligned}\mathbb{P}\left(|X^\varepsilon_v - X^\varepsilon_{s_{n-1}(v)}| \ge w_n\delta/2\right) &\le (w_n\delta/2)^{-\beta}\mathbb{E}\,|X^\varepsilon_v - X^\varepsilon_{s_{n-1}(v)}|^\beta \\ &\le (\delta/2)^{-\beta}(1-2^{-h})^{-\beta}\,2^{\beta h(n-n_0)}\,B_\varepsilon\, d(v, s_{n-1}(v))^{1+\alpha} \\ &\le (\delta/2)^{-\beta}(1-2^{-h})^{-\beta}\,2^{-\beta h n_0}\,2^{\beta h n}\,B_\varepsilon\,(6\cdot 2^{-n})^{1+\alpha}\end{aligned}$$

where the last estimate follows from equation (2.6). Putting all the estimates together, using $|T_n| \le N(T, d, 2^{-n})$ and the monotonicity of $N(T, d,\delta)$,

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in T}|X^\varepsilon_t - X^\varepsilon_{t_0}| \ge \delta/2\right) &\le (\delta/2)^{-\beta}(1-2^{-h})^{-\beta}\,6^{1+\alpha}\,2^{-\beta h n_0}\,B_\varepsilon\Big(\sum_{n>n_0}\sum_{v\in T_n}2^{\beta h n}\,2^{-(1+\alpha)n}\Big) \\ &= CB_\varepsilon\sum_{n>n_0}N(T,d,2^{-n})2^{-n}2^{-\gamma n} \\ &\le C2^{1+\gamma}B_\varepsilon\int_0^{D(T)} a^\gamma N(T,d,a)\,da \qquad (2.7)\end{aligned}$$

where C is a positive constant. Since $B_\epsilon \to 0$ as $\epsilon \to 0$ and the integral in the above equation is finite by the assumption in (2.5),

$$\lim_{\epsilon\to 0}\mathbb{P}\left(\sup_{t\in T}|X^\varepsilon_t - X^\varepsilon_{t_0}| \ge \delta/2\right) = \lim_{\epsilon\to 0}B_\varepsilon\Big[C'\int_0^{D(T)} a^\gamma N(T,d,a)\,da\Big] = 0$$

and the theorem is proved. □

3 Optimality of Our Hypothesis

Recall the Kolmogorov's continuity criterion: If for a stochastic process $X_t, t \in [0,1]$, there exist $\alpha, \beta > 0$, such that (compare with (2.4))

$$\mathbb{E}\Big(|X_t - X_s|^\beta\Big) \le C|t-s|^{1+\alpha}. \qquad (3.8)$$

then there exists a continuous version of the process $\{X_t\}$. The following well known example shows that the restriction $\alpha > 0$ is critical. On the interval $[0, 1]$ define a continuous process

$$U \sim \text{Unif}[0,1],\ X_t = \mathbf{1}_{\{U \le t\}}.$$

Then it follows that for any $\beta > 0$ and $s, t \in [0, 1]$

$$\mathbb{E}\Big(|X_t - X_s|^\beta\Big) = C|t-s|, \tag{3.9}$$

for some $C > 0$. However, the process $\{X_t\}$ is almost surely discontinuous on $[0, 1]$.

In the spirit of the above example, here we construct a stochastic process which shows that our hypothesis (2) in Theorem 1 is "very close" to the optimum.

Let $U \sim U[0, 1]$, $\epsilon > 0$ and $X_t^\varepsilon = \mathbf{1}_{\{t<U\le t+\varepsilon\}}$, $0 \le t \le 1$. Then for any fixed $t \in [0, 1]$,

$$\mathbb{E}\,(X_t^\varepsilon)^2 = \mathbb{P}(t < U \le t+\varepsilon) = \min(\varepsilon, 1-t) \to 0 \quad \text{as} \quad \varepsilon \to 0.$$

Since

$$\mathbb{E}\,X_t^\varepsilon X_s^\varepsilon = \begin{cases} 0 & \text{if} \quad |t-s| > \varepsilon \\ \varepsilon - |t-s| & \text{if} \quad |t-s| \le \varepsilon, \min(s,t) \le 1-\varepsilon \\ 1-\max(s,t) & \text{if} \quad |t-s| \le \varepsilon, \min(s,t) \ge 1-\varepsilon \end{cases}$$

it follows that

$$\mathbb{E}\,(X_t^\varepsilon - X_s^\varepsilon)^2 \le 2\min(\varepsilon, |t-s|). \tag{3.10}$$

On the other hand, for any $\epsilon > 0$, $\sup_{0\le t\le 1} X_t^\varepsilon = 1$ almost surely! Although by (3.10) we have bounds for $\mathbb{E}[(X_t^\epsilon - X_s^\epsilon)^2]$ of the form $B|t-s|$ and B_ϵ, there is no bound of the form $B_\epsilon|t-s|$.

Conjecture 1. *Theorem 1 is not true if hypothesis (2) (equation* (2.4)*) is replaced by*

$$\mathbb{E}\,|X_t^\varepsilon - X_s^\varepsilon|^\beta \le B_\varepsilon\, d(s,t). \tag{3.11}$$

4 An Application: Compensated Poisson Random Measures

In this section we present an application of Theorem 1 to a stochastic process constructed from compensated Poisson random measures.

Let Ω be a Polish space and $\nu(du\,d\omega)$ be a positive sigma-finite measure on $(-1, 1) \times \Omega$ such that

$$\nu\Big((-a,a)\times\Omega\Big) = \infty, \quad \forall a \in (0,1] \tag{4.12}$$

$$\iint_{(-1,1)\times\Omega} u^2\ \nu(du\,d\omega) < \infty. \tag{4.13}$$

Let

$$N(du\, d\omega) \sim \mathrm{Po}(\nu) \tag{4.14}$$

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

$$\tilde{N}(du\, d\omega) \equiv N(du\, d\omega) - \nu(du\, d\omega) \tag{4.15}$$

denote the compensated Poisson measure with mean 0, an isometry from $L_2\big(\mathbb{R}\times\Omega, \nu(du\, d\omega)\big)$ to the square-integrable zero-mean random variables (**?**, *pg.* 38).

Let $K(t,\omega) : [0,1] \times \Omega \mapsto \mathbb{R}$ be a Borel measurable function such that

$$\iint_{(-1,1)\times\Omega} K^2(t,\omega)\; u^2\; \nu(du\, d\omega) < \infty. \tag{4.16}$$

For $\epsilon > 0$, define a stochastic process X_t^ϵ,

$$X_t^\epsilon \equiv \iint_{\{0<|u|<\epsilon\}\times\Omega} K(t,\omega)\, u\, \tilde{N}(du\, d\omega), \quad t \in [0,1]. \tag{4.17}$$

For every $t \in [0,1]$ the stochastic integral in equation (4.17) is well defined because of (4.16) (see Wolpert & Taqqu (2005); Rajput & Rosiński (1989)). For $t \in [0,1]$, we have the following identities:

$$\mathbb{E}\,[X_t^\epsilon] = 0 \tag{4.18}$$

$$\mathbb{E}\left[(X_t^\epsilon)^2\right] = \iint_{(-\epsilon,\epsilon)\times\Omega} K^2(t,\omega)\; u^2\; \nu(du\, d\omega) < \infty \tag{4.19}$$

$$\mathbb{E}\left[e^{i\zeta X_t^\epsilon}\right] = \exp\left\{ \iint_{(-\epsilon,\epsilon)\times\Omega} \left[e^{i\zeta K(t,\omega)u} - 1 - i\zeta K(t,\omega)u\right] \nu(du\, d\omega)\right\} \tag{4.20}$$

analogous to the classic Lévy-Khintchine formula for the characteristic function of an infinitely divisible random variable.

The stochastic process $\Big\{X^\epsilon \equiv X_t^\epsilon, t \in [0,1]\Big\}$ is the discretization error of approximation of certain stochastic processes with discrete sums (see Pillai & Wolpert (2008) and the references therein). It is of interest to know the limiting behaviour of the process X^ϵ, when ϵ goes to zero (see Pillai & Wolpert (2008), section 3). In particular we would like to identify the conditions on the function K under which

$$\lim_{\epsilon\to 0} \mathbb{P}\Big(\sup_{t\in[0,1]} |X_t^\epsilon| > \delta\Big) = 0 \tag{4.21}$$

for any $\delta > 0$. Concentration equalities similar to (4.21) were studied in Reynaud-Bouret (2006) for finite intensity measures (*i.e.,* $\nu((-1,1) \times \Omega) < \infty$) using methods that are not applicable to our infinite intensity case.

In the next proposition we apply theorem 1 to identify conditions for the kernel $K(\cdot,\cdot)$ under which (4.21) holds.

Proposition 1. *Let X^ϵ be a stochastic process on $[0,1]$ defined as in (4.17). Let $K(t,\omega): [0,1]\times\Omega \mapsto \mathbb{R}$, $\alpha>0$ and a measurable $C_\omega : \Omega \mapsto \mathbb{R}_+$ satisfy:*

$$|K(t,\omega)-K(s,\omega)|^2 \le C_\omega^2 \ |t-s|^{1+\alpha}, \quad s,t\in[0,1], \tag{4.22}$$

$$\iint_{(-1,1)\times\Omega} C_\omega^2 \ u^2 \ \nu(du\, d\omega) < \infty \tag{4.23}$$

Then, for any $\delta>0$,

$$\lim_{\epsilon\to 0} \mathbb{P}\Big(\sup_{t\in[0,1]} |X_t^\epsilon| > \delta \Big) = 0. \tag{4.24}$$

Proof. Notice that for any $t\in[0,1]$, by equation (4.19),

$$\lim_{\epsilon\to 0} \mathbb{E}\Big[\big(X_t^\epsilon\big)^2 \Big] = \lim_{\epsilon\to 0} \int_{\{0<|u|<\epsilon\}\times\Omega} K^2(t,\omega) \ u^2 \ \nu(du\, d\omega) = 0, \tag{4.25}$$

verifying hypothesis 1 (equation (2.3)) of Theorem 1. For $t,s\in[0,1]$, by (4.22) and by the isometric property of $\tilde N(du\, d\omega)$,

$$\mathbb{E}\Big[\big(X_t^\epsilon - X_s^\epsilon\big)^2 \Big] = \iint_{\{0\le u\le\epsilon\}\times\Omega} |k(t,\omega)-K(s,\omega)|^2 \ u^2 \ \nu(du\, d\omega) \tag{4.26}$$

$$\le B_\epsilon |t-s|^{1+\alpha}, \tag{4.27}$$

$$B_\epsilon \equiv \iint_{\{0\le u\le\epsilon\}\times\Omega} C_\omega^2 \ u^2 \ \nu(du\, d\omega). \tag{4.28}$$

By the assumption in (4.23), $\lim_{\epsilon\to 0} B_\epsilon = 0$ and hypothesis 2 (equation 2.4) of theorem 1 is satisfied with the Euclidean metric $d(t,s)\equiv|t-s|$. Since $N([0,1],d,a) \le \frac{2}{a}$ for any $0<\gamma<\alpha$ (say $\gamma=\alpha/2$),

$$\int_0^1 a^\gamma N([0,1],d,a)\, da \le 2\int_0^1 a^{\gamma-1} da < \infty. \tag{4.29}$$

Theqrefore by (4.25),(4.28),(4.29) and Theorem 1, it follows that for any $\delta>0$,

$$\lim_{\epsilon\to 0} \mathbb{P}\Big(\sup_{t\in[0,1]} |X_t^\epsilon| > \delta \Big) = 0$$

and we are done. □

Remark: It is not known whether the conclusion of the above proposition still holds if (4.22) is weakened to

$$|K(t,\omega)-K(s,\omega)|^2 \le C_\omega^2 \ |t-s|^\alpha, \ x,y\in[0,1], \ 0<\alpha\le 1 \tag{4.30}$$

Acknowledgments

This work was supported in part by the National Science Foundation under Grant Numbers DMS-0805929, DMS-0757549, DMS–0635449 and the CRiSM research fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Ledoux, Michel. 1996. Isoperimetry and Gaussian analysis. *Pages 165–294 of: Lectures on probability theory and statistics (Saint-Flour, 1994)*. Lecture Notes in Math., vol. 1648. Berlin: Springer.

Pillai, Natesh S., & Wolpert, Robert L. 2008. *Posterior Consistency of Bayesian Nonparametric Models Using Lévy Random Field Priors*. Discussion Paper 2008-08. Duke University Department of Statistical Science.

Rajput, Balram S., & Rosiński, Jan. 1989. Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields*, **82**(3), 451–487.

Reynaud-Bouret, P. 2006. Compensator and exponential inequalities for some suprema of counting processes. *Statist. Probab. Lett.*, **76**(14), 1514–1521.

Talagrand, Michel. 2005. *The generic chaining*. Springer Monographs in Mathematics. Berlin: Springer-Verlag. Upper and lower bounds of stochastic processes.

Wolpert, Robert L., & Taqqu, Murad S. 2005. Fractional Ornstein-Uhlenbeck Lévy Processes and the Telecom Process: Upstairs and Downstairs. *Signal Processing*, **85**(8), 1523–1545.