mathematics

Article

Exponential Inequality of Marked Point Processes

Chen Li [1] and Yuping Song [2,*]

[1] Institute for Financial Studies, Shandong University, Jinan 250100, China
[2] School of Finance and Business, Shanghai Normal University, Shanghai 200234, China
* Correspondence: songyuping@shnu.edu.cn; Tel.: +86-021-64324667

Abstract: This paper presents the uniform concentration inequality for the stochastic integral of marked point processes. We developed a new chaining method to obtain the results. Our main result is presented under an entropy condition for partitioning the index set of the integrands. Our result is an improvement of the work of van de Geer on exponential inequalities for martingales in 1995. As applications of the main result, we also obtained the uniform concentration inequality of functional indexed empirical processes and the Kakutani–Hellinger distance of the maximum likelihood estimator.

Keywords: marked point processes; entropy condition; uniform concentration inequalities

MSC: 60E15; 60G55

1. Introduction

Concentration inequalities play essential roles in probability and statistics. For a long time, many authors have studied concentration inequalities. The reader may refer to classic books by Boucheron et al. [1], Bercu et al. [2], and so on. In particular, there has been a renewed interest in uniform concentration inequalities in the last three decades. Uniform concentration inequalities are significant for statistical learning and related fields. Some classical ideas and results can be found in the work by Bartlett and Mendelson [3,4].

A collection of random variables $\{\xi_t\}_{t\in T}$, where t belongs to a certain index set T, can be regarded as a stochastic process. The study of uniform concentration inequalities is focused on the supremum of specific stochastic processes. Some significant theorems on the supremum of stochastic processes are presented in the work by Talagrand [5]. However, most of the results in this work are on the expectation of the supremum. Therefore, the tail probability is another crucial problem in studying the supremum of stochastic processes. In this paper, we studied the uniform concentration inequality of the stochastic integral of the marked point process. Specifically, we want to find the upper bound of the tail probability of the supremum of a class of martingales.

Before our work, most results of uniform concentration inequalities are on the tail inequalities for the suprema of empirical processes, or more precisely, for empirical process indices by functions. As far as we know, one of the most important and earliest results of uniform concentration inequality is Talagrand's seminal work [6]. Inspired by Talagrand's work, exponential inequalities for bounded empirical processes were obtained by Massart [7] and by Klein and Rio [8]. The derivation of these results relies on the entropy method, which provides an alternative approach. Furthermore, exponential tail bounds for unbounded functional indexed empirical processes are given by Adamczak in [9] and by Lederer and van de Geer in [10]. Recently, Chen and Wu obtained the uniform concentration inequality for empirical processes of linear time series in [11].

The deviation of uniform concentration inequalities usually relies on the chaining method and some metrics in the index set. The fundamental idea of chaining is to replace the index set with a sequence of finite approximations. Thus, the upper bound of the tail

check for updates

Citation: Li, C.; Song, Y. Exponential Inequality of Marked Point Processes. *Mathematics* **2023**, *11*, 881. https://doi.org/10.3390/math11040881

Academic Editor: Anatoliy Swishchuk

Received: 23 January 2023
Revised: 4 February 2023
Accepted: 7 February 2023
Published: 9 February 2023

probability of suprema is determined by the sizes of finite approximations and the distance between approximations. In addition, the choice of the metric in the index can impact the final result. We developed a new approach to drive uniform concentration inequality (independent of metrics). In this paper, we obtained the upper bound of the tail probability of the stochastic integral of the marked point process. On the one hand, the stochastic integral of marked point processes is an essential example of purely discontinuous local martingales, which play important roles in mathematical finance and stochastic analysis. On the other hand, the stochastic integral of marked point processes is a generalization of functional indexed empirical processes. We present a new inequality of the functional indexed empirical processes in Section 4 as a corollary of our main result.

The rest of the paper is organized as follows. Section 2 provides some mathematical materials and methods, which can be used in the presentation and the proof of our main result. Next, we make some necessary assumptions and then state our main result in Section 3. The proof of the main result is also given in Section 3, which consists of several lemmas and propositions. Next, some applications will be shown in Section 4. Finally, the conclusion and suggestions for future research are presented in Section 5. Throughout this paper, we will use $c_1, c_2, \cdots$ to denote some universal positive constants, which may differ from section to section.

2. Mathematical Materials and Methods

In this section, we provide some mathematical materials and methods, which can be used in the presentation and the proof of our main result.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P})$ be a stochastic basis. A stochastic process $M = (M_t)_{t\geq 0}$ is called a purely discontinuous local martingale if $M_0 = 0$ and M is orthogonal to all continuous local martingales. The reader may refer to the classic book by Jacod and Shiryayev [12] for more information. Stochastic integrals of marked point processes are essential examples of purely discontinuous local martingales.

We shall restrict ourselves to the integer-valued random measure μ on $\mathbb{R}_+ \times \mathbb{R}$ induced by a $\mathbb{R}_+ \times \mathbb{R}$-valued marked point process. In this paper, let $(T_k, Z_k), k \geq 1$, be a marked point process, and define

$$\mu(dt, dx) = \sum_{k\geq 1} \mathbf{1}_{\{T_k<\infty\}} \varepsilon_{(T_k,Z_k)}(dt, dx), \tag{1}$$

where $\varepsilon_{(T_k,Z_k)}$ is the Dirac measure at point (T_k, Z_k), $\mathbf{1}_{\{T_k<\infty\}}$ is the indicator function of the set $\{T_k < \infty\}$. Then $\mu(\omega; [0, t] \times \mathbb{R}) < \infty$ for all $(\omega, t) \in \Omega \times \mathbb{R}$. Let $\tilde{\Omega} = \Omega \times \mathbb{R}_+ \times \mathbb{R}$, $\tilde{\mathcal{P}} = \mathcal{P} \otimes \mathcal{B}$, where $\mathcal{B}$ is a Borel σ-field on $\mathbb{R}$ and $\mathcal{P}$ a σ-field generated by all left continuous adapted processes on $\Omega \times \mathbb{R}_+$. The predictable function is a $\tilde{\mathcal{P}}$-measurable function on $\tilde{\Omega}$. Let ν be the unique predictable compensator of μ (up to a P-null set). Namely, ν is a predictable random measure, such that for any predictable function W, $W * \mu - W * \nu$ is a local martingale, where $W * \mu$ is defined by

$$W * \mu_t = \begin{cases} \int_0^t \int_{\mathbb{R}} W(s, x)\mu(ds, dx), & \text{if } \int_0^t \int_{\mathbb{R}} |W(s, x)|\mu(ds, dx) < \infty, \\ +\infty & \text{otherwise.} \end{cases}$$

Note the ν admits the disintegration

$$\nu(dt, dx) = dA_t K(\omega, t; dx), \tag{2}$$

where $K(\cdot, \cdot)$ is a transition kernel from $(\Omega \times \mathbb{R}_+, \mathcal{P})$ into $(\mathbb{R}, \mathcal{B})$, and $A = (A_t)_{t\geq 0}$ is an increasing càdlág predictable process. For μ in (1), which is defined through the marked point process, ν admits

$$\nu(dt, dx) = \sum_{n\geq 1} \frac{1}{G_n([t, \infty] \times \mathbb{R})} \mathbf{1}_{\{t\leq T_{n+1}\}} G_n(dt, dx),$$

where $G_n(\omega, ds, dx)$ is a regular version of the conditional distribution of (T_{n+1}, Z_{n+1}) with respect to $\sigma\{T_1, Z_1, \cdots, T_n, Z_n\}$. In particular, if $F_n(dt) = G_n(dt \times \mathbb{R})$, the point process $N = \sum_{n\geq 1} \mathbf{1}_{[T_n,\infty)}$ has the compensator $C_t = \nu([0,t] \times \mathbb{R})$, which satisfies

$$C_t = \sum_{n\geq 1} \int_0^{T_{n+1}\wedge t} \frac{1}{F_n([s,\infty])} F_n(ds).$$

Throughout the whole paper, we assume that $A = (A_t)_{t\geq 0}$ is a continuous process.

Next, let us consider the uniform concentration inequality for a family of stochastic integrals of predictable processes with respect to the marked point process. Let $\mathcal{G}$ be an index set and $\mathcal{W} = \{W^g : g \in \mathcal{G}\}$ a family of predictable functions on $\Omega \times \mathbb{R}_+ \times \mathbb{R}$. Fix a $T > 0$, denote $X^g = W^g * (\mu - \nu)_T$. In this paper, we will study the tail probability of $\sup_{g\in\mathcal{G}} |X^g|$. For simplicity, we assume $\sup_{g\in\mathcal{G}} |X^g|$ is measurable.

Some authors have already studied the uniform concentration inequality for $\sup_{g\in\mathcal{G}} |X^g|$. van de Geer first gave a inequality for $\sup_{g\in\mathcal{G}} |X^g|$ in [13]. In [13], μ is assumed as a counting process, which is a simple example of the marked point process. van de Geer introduced a metric d_0 in $\mathcal{W} = \{W^g : g \in \mathcal{G}\}$:

$$d_0(g_1, g_2) = \sqrt{\frac{1}{2}||(e^{W^{g_1}} - e^{W^{g_2}})^2 * \nu_T||_\infty}$$

where $||\cdot||_\infty$ stands for the norm of L^∞. Given $\delta > 0$, let $\{[W_j^{g^L}, W_j^{g^U}]\}_{j=1}^m \subseteq \mathcal{W} \times \mathcal{W}$ be a collection of pairs such that for each $g \in \mathcal{G}$ there exists a $j = j(g) \in \{1, 2, \cdots, m\}$ such that: $W_j^{g^L} \leq W^g \leq W_j^{g^U}$, $d\,(W_j^{g^L}, W_j^{g^U}) \leq \delta$. Let $\widehat{N}(\delta)$ be the smallest value of m for which such a set $\{[W_j^{g^L}, W_j^{g^U}]\}_{j=1}^m$ exists. van de Geer [13] obtained the following result.

Theorem 1. *Suppose $\mathcal{W} = \{W^g : g \in \mathcal{G}\}$ is a family of bounded predictable functions on $\Omega \times \mathbb{R}_+ \times \mathbb{R}$. For all $g \in \mathcal{G}$, there exist constants c_1, c_2, c_3, c_4, such that $W^g \geq c_1$, and $A_T \leq c_2$,*

$$\frac{\varepsilon v^2}{c_3} \geq \int_{\varepsilon v^2/c_4 \wedge \frac{v}{8}}^{v} \sqrt{\log(1 + \widehat{N}(\delta))} d\delta$$

where $d_0(W^g, 0) \leq v$ and $0 \leq \varepsilon \leq 1$. Then

$$\mathsf{P}\big(\sup_{g\in\mathcal{G}} |X^g| \geq \varepsilon v^2)\big) \leq c_5 \exp(-\frac{\varepsilon^2 v^2}{c_6}). \tag{3}$$

c_5 and c_6 are constants.

An extension of [13] can be found in work conducted by Le Guével [14]. Wang, Lin, and Su [15] extended van de Geer's result to a more general case. The generic chaining method from Talagrand [5] is employed in [15]. Suppose $\mathcal{W} = \{W^g : g \in \mathcal{G}\}$ is a family of bounded predictable functions on $\Omega \times \mathbb{R}_+ \times \mathbb{R}$. Two metrics are defined in [15] as follows:

$$d_1(g_1, g_2) = ||\Xi(W^{g_1} - W^{g_2})_T||_\infty, \tag{4}$$

$$d_2(g_1, g_2) = \sqrt{||C(W^{g_1} - W^{g_2})_T||_\infty} \tag{5}$$

where $\Xi(W)_T = \max\{0, W\} * \nu_T$.

For a given metric d on $\mathcal{G}$, an increasing sequence $(\mathcal{A}_n)_{n\geq 1}$ of partitions of $\mathcal{G}$ is called an admissible sequence if $\sharp\mathcal{A}_n \leq 2^{2^n}$. Denote by $A_n(g)$ the unique element of $(\mathcal{A}_n)$ containing g, and denote by $\Upsilon_d(A_n(g))$ the diameter of $A_n(g)$ under d. In addition, let

$$\gamma_\alpha(\mathcal{G}, d) = \inf \sup_{g \in \mathcal{G}} \sum_{n \geq 0} 2^{n/\alpha} \Upsilon_d(A_n(g)), \tag{6}$$

where the infimum takes over all admissible sequences. Wang, Lin, and Su [15] obtained the following result.

Theorem 2. *Suppose $\mathcal{W} = \{W^g : g \in \mathcal{G}\}$ is a family of bounded predictable functions on $\Omega \times \mathbb{R}_+ \times \mathbb{R}$ and a $g_0 \in \mathcal{G}$. Then*

$$P\big(\sup_{g \in \mathcal{G}} |X^g - X^{g_0}| \geq c_7 u(\gamma_0(\mathcal{G}, d_2) + \gamma_1(\mathcal{G}, d_1))\big) \leq c_8 \exp(-\frac{u}{2}). \tag{7}$$

where $u > 0$, c_7, and c_8 are constants.

The main contribution of this paper involves extending Theorem 1 and 2 to a sharper case. In (7), $\gamma_2(\mathcal{G}, d_2)$ and $\gamma_1(\mathcal{G}, d_1)$ are difficult to compute. In this paper, we will present our result in terms of an entropy condition for partitioning the index set of the integrands, and we find a new approach to drive the uniform concentration inequality independent of the metrics in the proof.

3. Main Result and Its Proof

In this section, the main result and its proof are presented. To state our main result, we need some more notations and make some technical assumptions.

Recall that $\mathcal{G}$ is an index set, $\mathcal{W} = \{W^g : g \in \mathcal{G}\}$ a family of predictable functions on $\Omega \times \mathbb{R}_+ \times \mathbb{R}$. Δ_Π a positive rational number. $\Pi = \{\Pi(\varepsilon)\}_{\varepsilon \in (0, \Delta_\Pi]}$, is called a decreasing series of finite partitions of $\mathcal{G}$ if

(i) each $\Pi(\varepsilon) = \{\mathcal{G}(\varepsilon; k) : 1 \leq k \leq N_\Pi(\varepsilon)\}$ is a finite partition of $\mathcal{G}$, namely

$$\mathcal{G} = \bigcup_{k=1}^{N_\Pi(\varepsilon)} \mathcal{G}(\varepsilon; k);$$

(ii) $N_\Pi(\Delta_\Pi) = 1$ and $\lim_{\varepsilon \downarrow 0} N_\Pi(\varepsilon) = \infty$;
(iii) $N_\Pi(\varepsilon) \geq N_\Pi(\varepsilon')$ as $\varepsilon \leq \varepsilon'$.

Given a $0 < \varepsilon \leq \Delta_\Pi$, define

$$H_\Pi(\varepsilon) = \log(1 + N_\Pi(\varepsilon)).$$

If $(\mathcal{X}, \mathcal{A}, \lambda)$ is a σ—finite measure space. For a given $\mathcal{A}$—measurable function F, we denote by $[F]_{\mathcal{A},\lambda}$ any $\mathcal{A}$—measurable function Y such that:

(i) $Y \geq F$ holds identically;
(ii) for every $\mathcal{A}$—measurable function $\check{Y}$, if $\check{Y} \geq F$ holds λ—almost everywhere, then $\check{Y} \geq Y$ holds λ—almost everywhere.

Now, we turn to the context of marked point processes, which plays a key role in this paper. The Doléans measure M_ν^{P} on $(\tilde{\Omega}, \tilde{\mathcal{P}})$ is given by

$$M_\nu^{\mathsf{P}} = \mathsf{P}(d\omega)\nu(\omega; dt, dx).$$

The predictable envelope $\overline{W}$of $\mathcal{W} = \{W^g : g \in \mathcal{G}\}$ is defined by $\overline{W} = [\sup_{g \in \mathcal{G}} |W^g|]_{\tilde{\mathcal{P}}, M_\nu^{\mathsf{P}}}$. For any subset $\mathcal{G}' \subset \mathcal{G}$, define

$$\mathcal{W}(\mathcal{G}') = [\sup_{g_1, g_2 \in \mathcal{G}'} |W^{g_1} - W^{g_2}|]_{\tilde{\mathcal{P}}, M_\nu^{\mathsf{P}}}.$$

Now, we present the main result of this paper.

Theorem 3. *Suppose $\mathcal{W} = \{W^g : g \in \mathcal{G}\}$ is a family of bounded predictable functions on $\Omega \times \mathbb{R}_+ \times \mathbb{R}$, and there exists a constant κ such that $0 < \overline{W} \leq \kappa$. Furthermore,*

$$\nu([0,T] \times \mathbb{R}) = O_{a.s.}(b_T).$$

b_T is an increasing function on T. If there exists a decreasing series of finite partitions of $\mathcal{G}$, such that

$$N_\Pi(\varepsilon) \leq \alpha \exp\{\frac{\beta}{\varepsilon^2}\},$$

for some constants $\alpha > 1$, $\beta > 0$, and for some constants c_9 and c_{10},

$$\{\sqrt{\beta}\log(\frac{b_T}{\frac{b_T^2 u}{c_9} \wedge \frac{b_T}{4}}) + \sqrt{\log\alpha}(b_T - \frac{b_T^2 u}{c_9} \wedge \frac{b_T}{4})\} \leq \frac{b_T^{3/2} u^2}{c_{10}},$$

Then for $u \in (0,1)$,

$$\mathsf{P}\big(\sup_{g\in\mathcal{G}} |X_T^g| \geq b_T^2 u)\big) \leq c_{11}\exp(-\frac{b_T^2 u}{c_{12}}). \tag{8}$$

where c_{11} and c_{12} are constants.

Remark 1. *Nishiyama [16,17] obtained the weak convergence of the stochastic integral of marked point processes under entropy conditions. His results can be regarded as the extension of some classical results of weak convergence for the empirical process, which can be found in van der Vaart and Wellner [18]. In this paper, the uniform concentration inequality of the stochastic integral of the marked point process is derived under the entropy condition of partitioning.*

Proof of the Theorem 3. For every integer $i \geq 0$, choose an element $g_{i,k}$ from each partition set $\mathcal{G}(2^{-i}b_T;k)$ such that

$$\{g_{i,k} : 1 \leq k \leq N_\Pi(2^{-i}b_T)\} \subset \{g_{i+1,k} : 1 \leq k \leq N_\Pi(2^{-i-1}b_T)\}$$

Define for every $g \in \mathcal{G}$,

$$\begin{cases} \pi_i g = g_{i,k}, \\ \Pi_i g = \mathcal{G}(\frac{b_T}{2^i};k), \end{cases}$$

for $g \in \mathcal{G}(\frac{b_T}{2^i};k)$. Furthermore, define

$$W(\Pi_i g) = [\sup_{g,g' \in \Pi_i g} |W^g - W^{g'}|]_{\widetilde{\mathcal{P}},M_\nu^{\mathsf{P}}}.$$

Let $l = \min\{i \geq 1 : 2^{-i} \leq \frac{b_T u}{2^3}\}$, for $i = 0,1,\cdots,l$, write $H_{T,i} = H_\Pi(\frac{b_T}{2^i})$. Set

$$\eta_i = \max\{\frac{(\sum_{k=0}^i H_{T,k})^{1/2}}{2^7 b_T^{1/2} u}, \sqrt{\frac{i}{2^i}}\}$$

and $a_i = \frac{2^{3-2i}}{\eta_{i+1}u}$. Furthermore, for $i = 0,1,\cdots,l$, we define

$$A_i(g) = 1_{\{W(\Pi_0 g) \leq a_0, \cdots, W(\Pi_{i-1}g) \leq a_{i-1}, W(\Pi_i g) \leq a_i\}}.$$

For $i = 1,\cdots,l$, we define

$$B_i(g) = 1_{\{W(\Pi_0 g) \leq a_0, \cdots, W(\Pi_{i-1}g) \leq a_{i-1}, W(\Pi_i g) > a_i\}},$$

and

$$B_0(g) = 1_{\{W(\Pi_0 g) > a_0\}}.$$

We observe the identity

$$\begin{aligned} W^g - W^{\pi_0 g} &= \sum_{i=0}^{l}(W^g - W^{\pi_i g})B_i(g) + \sum_{i=1}^{l}(W^{\pi_i g} - W^{\pi_{i-1} g})A_i(g) \\ &\quad +(W^g - W^{\pi_l g})A_l(g). \end{aligned}$$

Thus

$$\begin{aligned} & \mathsf{P}\big(\sup_{g\in\mathcal{G}} |X_T^g| \geq b_T^2 u)\big) \\ \leq\ & \mathsf{P}\big(\sup_{g\in\mathcal{G}} |W^{\pi_0 g} * (\mu-\nu)_T| \geq \frac{b_T^2 u}{4}) + \mathsf{P}\big(\sup_{g\in\mathcal{G}} |\sum_{i=0}^{l}(W^g - W^{\pi_i g})B_i(g) * (\mu-\nu)_T| \geq \frac{b_T^2 u}{4}) \\ & +\mathsf{P}\big(\sup_{g\in\mathcal{G}} |\sum_{i=1}^{l}(W^{\pi_i g} - W^{\pi_{i-1} g})A_i(g) * (\mu-\nu)_T| \geq \frac{b_T^2 u}{4}) \\ & +\mathsf{P}\big(\sup_{g\in\mathcal{G}} |(W^g - W^{\pi_l g})A_l(g) * (\mu-\nu)_T| \geq \frac{b_T^2 u}{4}) \\ =:\ & \mathsf{P}_I + \mathsf{P}_{II} + \mathsf{P}_{III} + \mathsf{P}_{IV}. \end{aligned}$$

First, if there exists a constant κ, such that $0 < \overline{W} \leq \kappa$, $\nu([0,T]\times\mathbb{R}) = O_{a.s.}(b_T)$, for $x > 0$, we have

$$\mathsf{P}(|W * (\mu-\nu)_T|) \geq x) \leq \exp\{-\frac{x^2}{2(\kappa x + \kappa^2 b_T)}\}$$

by Theorem 1.1 in Wang, Lin, and Su [15]. Thus, we have

$$\begin{aligned} \mathsf{P}_I &\leq N_\Pi(b_T)\exp\{-\frac{b_T^3 u^2}{8\kappa u + 32\kappa^2}\} \\ &\leq \exp\{H_{T,0} - \frac{b_T^3 u^2}{8\kappa b_T u + 32\kappa^2}\} = \exp\{H_{T,0} - \frac{b_T^2 u}{c_{10}}\}. \end{aligned}$$

With a proper choice of c_9,

$$\begin{aligned} \frac{3b_T}{4}H_{T,0}^{1/2} &\leq \int_{\frac{b_T^2 u}{c_9}\wedge\frac{b_T}{4}}^{b_T} \sqrt{H_\Pi(x)}dx \\ &\leq \int_{\frac{b_T^2 u}{c_9}\wedge\frac{b_T}{4}}^{b_T} \sqrt{\log\alpha + \beta x^{-2}}dx \\ &\leq \{\sqrt{\beta}\log(\frac{b_T}{\frac{b_T^2 u}{c_9}\wedge\frac{b_T}{4}}) + \sqrt{\log\alpha}(b_T - \frac{b_T^2 u}{c_9}\wedge\frac{b_T}{4})\} \\ &\leq \frac{b_T^{3/2}u^2}{c_{10}}, \end{aligned}$$

thus,

$$\mathsf{P}_I \leq \exp\{-\frac{b_T^2 u}{c_{13}}\}.$$

Now, we consider P_{II}.

$$[(W^{g}-W^{\pi_i g})B_i(g)]_{\tilde{\mathcal{P}},M_v^{\mathsf{P}}}\leq a_{i-1},$$

implies

$$\mathsf{P}_{II} \leq \sum_{i=0}^{l} N_{\Pi}(\frac{b_T}{2^i})\exp\{-\frac{b_T^3u^2}{8a_{i-1}b_Tu+32a_{i-1}^2}\}.$$

Thanks to

$$\begin{aligned}
\sum_{i=0}^{l}2^{-i}H_{T,i}^{1/2} &\leq \int_{\frac{b_T}{2^l}}^{b_T}\sqrt{H_{\Pi}(x)}dx \leq \int_{\frac{b_T^2u}{c_9}\wedge\frac{b_T}{4}}^{b_T}\sqrt{H_{\Pi}(x)}dx \\
&\leq \{\sqrt{\beta}\log(\frac{b_T}{\frac{b_T^2u}{c_9}\wedge\frac{b_T}{4}})+\sqrt{\log\alpha}(b_T-\frac{b_T^2u}{c_9}\wedge\frac{b_T}{4})\} \\
&\leq \frac{b_T^{3/2}u^2}{c_{10}},
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i=1}^{l}(\sum_{k=0}^{i}H_{T,k})^{1/2} &\leq 2^l\sum_{i=1}^{l}2^{-i}(\sum_{k=0}^{i}H_{T,k})^{1/2} \\
&\leq 2^{l+1}\sum_{i=1}^{l}2^{-i}H_{T,i}^{1/2}\leq\frac{\sqrt{b_T}u}{c_{14}},
\end{aligned}$$

we have

$$\begin{aligned}
\mathsf{P}_{II} &\leq \sum_{i=0}^{l}\exp\{H_{T,i}-\frac{b_T^2u\eta_i^2}{c_{15}}\} \\
&\leq \sum_{i=1}^{l}\exp\{\sum_{k=0}^{i}H_{T,k}-\frac{b_T^2u\eta_i^2}{c_{15}}\}\leq\sum_{i=1}^{l}\exp\{-\frac{b_T^2u\eta_i^2}{c_{16}}\} \\
&\leq \sum_{i=1}^{l}\exp\{-\frac{b_T^2ui}{2^ic_{16}}\}\leq\sum_{i=1}^{l}\exp\{-\frac{b_T^2ui}{c_{17}}\} \\
&\leq c_{18}\exp\{-\frac{b_T^2u}{c_{19}}\}
\end{aligned}$$

For P_{III} and P_{IV}, denote

$$[(W^{\pi_i g}-W^{\pi_{i-1}g})A_i(g)]_{\tilde{\mathcal{P}},M_v^{\mathsf{P}}}\leq 2a_{i-1},$$

$$[(W^{g}-W^{\pi_l g})A_l(g)]_{\tilde{\mathcal{P}},M_v^{\mathsf{P}}}\leq a_l,$$

we can obtain

$$\mathsf{P}_{III}+\mathsf{P}_{IV}\leq c_{20}\exp\{-\frac{b_T^2u}{c_{21}}\}$$

by similar arguments. □

4. Applications

This section will first apply the previous result to functional index empirical processes. Consider a discrete-time process $(Y_n)_{n\geq 0}$ on$(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathsf{P})$. Let Ψ be the space of measurable functions in $\mathbb{R}$. For a $\psi \in \Psi$, define

$$X_n^{\psi} = \sum_{k=1}^{n} \big(\psi(Y_k) - \mathsf{E}(\psi(Y_k)|\mathcal{F}_{k-1})\big). \tag{9}$$

Obviously, for each ψ, $\{X_n^{\psi}\}_{n\geq 0}$ is a discrete-time martingale. We can study the concentration inequality by considering X_n^{ψ} as a stochastic integral of ψ with respect to a marked point process. In fact, let $\psi(k,x) = \psi(x_k)$, then X_n^{ψ} can be written as

$$X_n^{\psi} = \psi * (\mu - \nu)_n, \tag{10}$$

where

$$\mu(dt, dx) = \sum_{k\geq 1} \varepsilon_{(k,Y_k)}(dt, dx),$$

ν is the compensator of μ. A simple computation shows

$$\psi * \nu(dt, dx)_n = \sum_{k=1}^{n} \mathsf{E}[\psi(Y_k)|\mathcal{F}_{k-1}]. \tag{11}$$

By Theorem 3.1 in Wang Lin and Su [15]

$$\mathsf{P}(|X_n^{\psi}| > x) \leq \exp\{-\frac{x^2}{x + b_n}\}.$$

as a consequence of Theorem 3, we have

Theorem 4. *Suppose $\Psi = \{\psi\}$ is a family of bounded measurable functions on $\mathbb{R}$, and $|\psi| \leq 1$. Furthermore,*

$$\sum_{k=1}^{n} E[\psi(Y_k)|\mathcal{F}_{k-1}] = O_{a.s.}(b_n).$$

b_n is an increasing function on n. If there exists a decreasing series of finite partitions of Ψ, such that

$$N_{\Pi}(\varepsilon) \leq \alpha \exp\{\frac{\beta}{\varepsilon^2}\},$$

for some constants $\alpha > 1$, $\beta > 0$, and for some constants c_{22} and c_{23},

$$\{\sqrt{\beta}\log(\frac{b_n}{\frac{b_n^2 u}{c_{23}} \wedge \frac{b_n}{4}}) + \sqrt{\log\alpha}(b_n - \frac{b_n^2 u}{c_{23}} \wedge \frac{b_n}{4})\} \leq \frac{b_n^{3/2} u^2}{c_{22}},$$

Then for $u \in (0,1)$,

$$\mathsf{P}\big(\sup_{\psi\in\Psi} |X_n^{\psi}| \geq b_n^2 u)\big) \leq c_{24}\exp(-\frac{b_n^2 u}{c_{25}}). \tag{12}$$

where c_{24} and c_{25} are constants.

Furthermore, we consider the nonparametric maximum likelihood estimator below. Let $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ be a family of probability measures on $(\mathbb{R}, \mathcal{B})$. We assume that each P_θ is absolutely continuous with respect to the Lebesgue measure, and we denote the density function of P_θ by f_θ, namely $f_\theta = \frac{dP_\theta}{dx}$, $\theta \in \Theta$.

Now we are given a $\theta_0 \in \Theta$ with $f_{\theta_0} > 0$, and we want to make an estimate for θ_0. Let $X_1, X_2, \cdots$ be a sequence of i.i.d. observations from P_{θ_0}. Define the empirical distribution

$$P_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k} \tag{13}$$

on the basis of the first n observations, where δ_x stands for delta measure.

The widely used maximum likelihood estimator $\hat{\theta}_n$ of θ_0 is defined by

$$\int \log(f_{\hat{\theta}_n}) dP_n = \max_{\theta\in\Theta} \int \log(f_\theta) dP_n. \tag{14}$$

The Kakutani–Hellinger distance is used to compute the distance between two probability measures. In particular, for any two probability measures P_1 and P_2 with density functions f_1 and f_2 on $(\mathbb{R}, \mathcal{B})$, the Kakutani–Hellinger distance $h(f_1, f_2)$ is defined by

$$h^2(f_1, f_2) = \frac{1}{2}\int_{\mathbb{R}} \left(\sqrt{f_1(x)} - \sqrt{f_2(x)}\right)^2 dx. \tag{15}$$

We shall give the rate of convergence for $f_{\hat{\theta}_n}$ to f_{θ_0} by means of $h^2(f_{\hat{\theta}_n}, f_{\theta_0})$ below. Note $\int_{\mathbb{R}} f_\theta dx = \int_{\mathbb{R}} f_{\theta_0} dx$, and $\log(1+x) \le x$ for any $x > -1$,

$$\begin{aligned}
0 &\le \frac{1}{2}\int_{\mathbb{R}} \log\frac{f_{\hat{\theta}_n}}{f_{\theta_0}} dP_n \le \int_{\mathbb{R}} \Big(\sqrt{\frac{f_{\hat{\theta}_n}}{f_{\theta_0}}} - 1\Big) dP_n \\
&\le \int_{\mathbb{R}} \Big(\sqrt{\frac{f_{\hat{\theta}_n}}{f_{\theta_0}}} - 1\Big) d(P_n - P_{\theta_0}) + \int_{\mathbb{R}} \Big(\sqrt{\frac{f_{\hat{\theta}_n}}{f_{\theta_0}}} - 1\Big) dP_{\theta_0} - \frac{1}{2}\int (f_{\hat{\theta}_n} - f_{\theta_0}) dx \\
&= \int_{\mathbb{R}} \Big(\sqrt{\frac{f_{\hat{\theta}_n}}{f_{\theta_0}}} - 1\Big) d(P_n - P_{\theta_0}) + \int_{\mathbb{R}} \sqrt{f_{\hat{\theta}_n} f_{\theta_0}} dx - \frac{1}{2}\int_{\mathbb{R}} (f_{\hat{\theta}_n} + f_{\theta_0}) dx \\
&= \int_{\mathbb{R}} \Big(\sqrt{\frac{f_{\hat{\theta}_n}}{f_{\theta_0}}} - 1\Big) d(P_n - P_{\theta_0}) - h^2(f_{\hat{\theta}_n}, f_{\theta_0}).
\end{aligned} \tag{16}$$

Then for any $x > 0$, we have

$$\begin{aligned}
\mathsf{P}(h^2(f_{\hat{\theta}_n}, f_{\theta_0}) \ge x) &\le \mathsf{P}\Big(\int_{\mathbb{R}} \Big(\sqrt{\frac{f_{\hat{\theta}_n}}{f_{\theta_0}}} - 1\Big) d(P_n - P_{\theta_0}) \ge x\Big) \\
&\le \mathsf{P}\Big(\sup_{\theta\in\Theta} \int_{\mathbb{R}} \Big(\sqrt{\frac{f_\theta}{f_{\theta_0}}} - 1\Big) d(P_n - P_{\theta_0}) \ge x\Big).
\end{aligned}$$

Let $\mathcal{G} = \{g_\theta := \sqrt{\frac{f_\theta}{f_{\theta_0}}} - 1,\ \theta \in \Theta\}$, and note $g_{\theta_0} = 0$. Since $X_1, X_2, \cdots$, is a sequence of i.i.d. samples from P_{θ_0}, then

$$\begin{aligned}
\int_{\mathbb{R}} \Big(\sqrt{\frac{f_\theta}{f_{\theta_0}}} - 1\Big) d(P_n - P_{\theta_0}) &= \frac{1}{n}\sum_{k=1}^{n} \big(g_\theta(X_k) - Eg_\theta(X_k)\big) \\
&=: \frac{1}{n} X_n^{g_\theta}.
\end{aligned}$$

At last, we have the following result.

Theorem 5. *Suppose* $\mathcal{G} = \{g_\theta := \sqrt{\frac{f_\theta}{f_{\theta_0}}} - 1,\ \theta \in \Theta\}$ *is a family of bounded measurable functions on* $\mathbb{R}$*, and* $|g_\theta| \leq 1$*. Furthermore,*

$$\sum_{k=1}^{n} \mathsf{E}[g_\theta(X_k)] = O(b_n),$$

b_n *is an increasing function on* n*. If there exists a decreasing series of finite partitions of* Ψ*, such that*

$$N_\Pi(\varepsilon) \leq \alpha \exp\{\frac{\beta}{\varepsilon^2}\},$$

for some constants $\alpha > 1$*,* $\beta > 0$*, and for some constants* c_{26} *and* c_{27}*,*

$$\{\sqrt{\beta}\log(\frac{b_n}{\frac{b_n^2 u}{c_{27}} \wedge \frac{b_n}{4}}) + \sqrt{\log\alpha}(b_n - \frac{b_n^2 u}{c_{27}} \wedge \frac{b_n}{4})\} \leq \frac{b_n^{3/2} u^2}{c_{26}},$$

Then for $u \in (0,1)$*,*

$$\mathsf{P}\big(h^2(f_{\hat{\theta}_n}, f_{\theta_0}) \geq n b_n^2 u\big) \leq c_{28} \exp(-\frac{n b_n^2 u}{c_{29}}). \tag{17}$$

where c_{28} *and* c_{29} *are constants.*

Remark 2. *The uniform concentration inequality of the functional index empirical processes and the Kakutani–Hellinger distance of the maximum likelihood estimator were studied by Wang, Lin, and Su [15]. We give our new inequalities for these two particular cases in Theorems 4 and 5.*

5. Conclusions

This paper explores the uniform concentration inequality for the stochastic integral of the marked point process. We can obtain an upper bound of tail probability for the supremum of a class of purely discontinuous local martingales. In other words, our result implies that the supremum of a class of stochastic integrals with respect to marked point processes has a sharp bound in high probability. Our inequality is an extension of classical results on uniform concentration inequalities for functional indexed empirical processes. Furthermore, the deviation of our inequality is a new approach to drive the uniform concentration inequality under an entropy condition for partitioning the index set. Thus, our result is independent of metrics in the index set.

Several issues can be further explored. First, the assumption imposed on the entropy can be relaxed to a more general case. Different entropy conditions can give other tail bounds of the suprema of stochastic processes. Furthermore, it would be an interesting future topic to extend the chaining method developed in this paper to obtain a concentration around the mean for the maxima of the stochastic integral for the marked point process. Such results in empirical processes have been studied by Klein and Rio [8] and by Lederer and van de Geer [10]. It is also worthwhile to further explore the application of uniform concentration inequalities in statistical learning.

Author Contributions: Conceptualization, Y.S.; methodology, C.L. and Y.S.; writing, C.L. and Y.S.; C.L. worked out the details of the theorem; Y.S. designed the framework of this study and substantively revised it. All authors have read and agreed to the published version of the manuscript.

Funding: This research work is supported by the Ministry of Education, Humanities and Social Sciences project (18YJCZH153), National Statistical Science Research Project (2018LZ05), Youth Academic Backbone Cultivation Project of Shanghai Normal University (310-AC7031-19-003021), General Research Fund of Shanghai Normal University (SK201720), and the Key Subject of Quantitative Economics (310-AC7031-19-004221) and Academic Innovation Team (310-AC7031-19-004228) of Shanghai Normal University.

Data Availability Statement: Data sharing not applicable to this article as no datasets were generated or analysed during the current (theoretical) study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Boucheron, S.; Lugosi, G.; Massart, P. *Concentration Inequalities: A Nonasymptotic Theory of Independence*; Oxford University Press: Oxford, UK, 2013.
2. Bercu, B.; Delyon, B.; Rio, E. *Concentration Inequalities for Sums and Martingales*; Springer: Cham, Switzerland, 2015.
3. Bartlett, P.; Mendelson, S. Rademacher and Gaussian complexities: Risk bounds and structural results. *J. Mach. Learn. Res.* **2002**, *2*, 463–482.
4. Bartlett, P.; Mendelson, S. Empirical minimization. *Probab. Theory Related Fields* **2006**, *135*, 311–334. [CrossRef]
5. Talagrand, M. *Upper and Lower Bounds for Stochastic Processes. Modern Methods and Classical Problems*; Springer: Berlin/Heidelberg, Germany, 2014.
6. Talagrand, M. New concentration inequalities in product spaces. *Invent. Math.* **1996**, *126*, 505–563. [CrossRef]
7. Massart, P. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **2000**, *28*, 863–884. [CrossRef]
8. Klein, T.; Rio, E. Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **2005**, *33*, 1060–1077. [CrossRef]
9. Adamczak, R. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **2008**, *34*, 1000–1034. [CrossRef]
10. Lederer, J.; van de Geer, S. New concentration inequalities for suprema of empirical processes. *Bernoulli* **2014**, *20*, 2020–2038. [CrossRef]
11. Chen, L.; Wu, W. Concentration inequalities for empirical processes of linear time series. *J. Mach. Learn. Res.* **2017**, *18*, 8639–8684.
12. Jacod, J.; Shiryaev, A. *Limit Theorems for Stochastic Processes*, 2nd ed.; Grundlehren der Mathematischen Wissenschaften; Springer: Berlin/Heidelberg, Germany, 2003; Volume 288.
13. van de Geer, S. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. *Ann. Statist.* **1995**, *23*, 1779–1801. [CrossRef]
14. Le Guével, R. Exponential inequalities for the supremum of some counting processes and their square martingales. *C. R. Math. Acad. Sci. Paris* **2021**, *359*, 1969–982. [CrossRef]
15. Wang, H.; Lin, Z.; Su, Z. On Bernstein type inequalities for stochastic integrals of multivariate point processes. *Stoch. Process. Appl.* **2019**, *129*, 1605–1621. [CrossRef]
16. Nishiyama, Y. Some central limit theorems for l^∞—valued semimartingales and their applications. *Probab. Theory Related Fields* **1997**, *108*, 459–494. [CrossRef]
17. Nishiyama, Y. Weak convergence of some classes of martingales with jumps. *Ann. Probab.* **2000**, *28*, 685–712. [CrossRef]
18. van der Vaart, A.; Wellner, J. *Weak Convergence and Empirical Processes*; Springer: New York, NY, USA, 1996.