On the Tail Behavior of the Distribution Function of Multiple Stochastic Integrals in Separable Metric Spaces

Péter Major[1], *Lídia Rejtő*,[1,2]

[1] Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary
[2] Department of Mathematical Sciences, University of Delaware,
Newark, Delaware, USA

Abstract. Let $\bar\mu_n(.)$ denote the empirical measure of a sample of i.i.d. random variables with values on a separable metric space $(X,\mathcal{X})$ with distribution μ. Define $\mu_n(\cdot)=\sqrt{n}\,(\bar\mu_n(\cdot)-\mu(\cdot))$ and consider the integrals

$$\mathcal{I}(t)=\int_{X_t}\int_X\cdots\int_X f(u_1,\dots,u_s)\mu_n(\,du_1)\dots\mu_n(\,du_s),$$

where $X_s\subseteq X_t$ for all $s\le t$, $X_0=\emptyset$, $X_1=X$, $\mu(X_t)=t$ and f is a bounded measurable function which disappears on the diagonal set of the product space X^s. We give a sharp upper bound on the probability $P\left(\sup\limits_{0\le t\le 1}|\mathcal{I}(t)|\ge x\right)$. This is a generalization of the result in paper [mp] where this estimate was proved in the special case when the space X is the interval $[0,1]$, and the empirical distribution of a sequence of independent and on $[0,1]$ uniformly distributed random variables is considered.

Key words and phrases. Stochastic integrals, empirical measure, separable metric space.

1. Introduction

In this note we consider the empirical distribution of a sequence of independent random variables on a separable metric space $(X,\mathcal{A})$, take its natural standardization and bound the tail distribution of the (random) integral of a bounded function which maps the s-fold product $\underbrace{X\times\cdots\times X}_{s\text{ times}}$ of the space X to the real line with respect to the s-fold product of the standardized empirical distribution. We give a sharp bound, only the universal constants appearing in this estimate will not be given. In paper [mp] such an estimate was proved in the special case when the space X is the interval, $[0,1]$, and the empirical distribution of a sequence of independent and on $[0,1]$ uniformly distributed random variables is considered. We shall recall this result in Theorem A. Actually we shall show that the case of random variables on a general metric space can be deduced from this special case by means of an isomorphism.

Our motivation for proving such a result comes from the estimation of a distribution function with the help of certain information. The result which we formulate in Theorem A turned out to be useful in the study of the so-called product limit estimation (see[mp]) when a complicated error term, which is a non-linear functional of the standardized empirical process, had to be bounded. In other problems, (see [mrt]), a multiple (random) integral of a function with respect to a standardized empirical measure on a higher dimensional Euclidean space has to be estimated. Since the proof of the result needed in this case is not harder for a general separable metric space, we shall work on a metric space. The above mentioned motivation can also explain why the estimates formulated below contain a supremum in the first coordinate, a generalization which may seem artificial at first sight. We want to bound the supremum of the difference between the estimate and the estimated function $F(t)$ for all t. The supremum appearing in our result will enable us to give a good bound on the error of the estimate not only for a fixed t, but for all t simultaneously. First we formulate Theorem A which will be the basis of our investigation.

Theorem A. *Let $\xi_1(\omega),\dots,\xi_n(\omega)$ be independent uniformly distributed random variables on $[0,1]$, $F_n(u)=F_n(u,\omega)=\frac{1}{n}\#\{k\colon 1\le k\le n,\ \xi_k\le u\}$, $0\le u\le 1$, their empirical distribution function and $\mu_n(u)=\sqrt{n}(F_n(u)-u)$ the standardization of this empirical distribution function. Let $f(u_1,\dots,u_s)$ be a function on $[0,1]^s$ such that $\sup\limits_{u_1,\dots,u_s}|f(u_1,\dots,u_s)|\le 1$, and $f(u_1,\dots,u_s)=0$ if $u_j=u_k$ with some $1\le j<k\le s$. There exist some universal constants C_s and α_s depending only on the dimension s in such a way that*

$$P\left(\sup_{0\le t\le 1}\left|\int_0^t\int_0^1\cdots\int_0^1 f(u_1,\dots,u_s)\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge x\right)\le C_s\exp\left\{-\alpha_s x^{2/s}\right\} \tag{1.1}$$

for all $x>0$, and function f with the above properties.

As it is remarked in paper [mp] this result is sharp. To formulate its generalization proved in this paper some notations have to be introduced.

Let a probability space $(\Omega,\mathcal{A},P)$ and a separable metric space $(X,\mathcal{X})$ be given. Let $\xi\colon\Omega\to X$ be an X valued random variable on $(\Omega,\mathcal{A},P)$. Let μ denote the distribution of the random variable ξ, i.e. let

$$\mu(B)=P(\xi\in B)=P(\xi^{-1}(B))\qquad\forall\, B\in\mathcal{X}\ .$$

Suppose that $\xi_1,\xi_2,\dots,\xi_n$ are independent, identically distributed random variables on $(\Omega,\mathcal{A},P)$ with values on the space $(X,\mathcal{X})$ and distribution μ. Let us introduce the empirical measure

$$\bar\mu_n(B)=\frac{1}{n}\sum_{i=1}^n I(\xi_i\in B)\qquad\forall\, B\in\mathcal{X}\ ,$$

and its standardization

$$\mu_n(B)=\sqrt{n}\,(\bar\mu_n(B)-\mu(B))\qquad\forall\, B\in\mathcal{X}\ .$$

Let X_t, $0\le t\le 1$ be a system of sets in $\mathcal{X}$ with the following property:

Property (i) $X_s\subseteq X_t$ for all $s\le t$, $X_0=\emptyset$, $X_1=X$, $\mu(X_t)=t$.

Let us consider the product space $\underbrace{X \times \cdots \times X}_{s \text{ times}} = X^s$ with product measure $\mu^{(s)}(\cdot)$ and the diagonal set $A \in X^s$ is defined as

$$A = \{(x_1, \ldots x_s) \colon x_i = x_j \quad \text{for some } i \neq j\}$$

Since $(X, \mathcal{X})$ is a separable metric space, A is a measurable set in the product space, and if μ has no atoms then $\mu^{(s)}(A) = 0$.

Let $\mathcal{F}$ denote the set of the real valued measurable functions $f(u_1, \ldots, u_s)$ defined on the space X^s whose absolute value is less than 1, and which disappear on the diagonal set A, i.e. let

$$\mathcal{F} = \{f(u_1, \ldots, u_s) \colon |f| \leq 1, \quad f(u_1, \ldots, u_s) = 0 \; \forall \, (u_1, \ldots, u_s) \in A\} \, . \tag{1.2}$$

We shall prove the following result:

Theorem 1. *There exist some universal constants C_s and α_s depending only on the dimension s in such a way that*

$$P\left(\sup_{0 \leq t \leq 1} \left| \int_{X_t} \int_X \cdots \int_X f(u_1, \ldots, u_s) \mu_n(\,du_1) \ldots \mu_n(\,du_s) \right| \geq x \right) \leq C_s \exp\left\{ -\alpha_s x^{2/s} \right\} \tag{1.3}$$

for all $f \in \mathcal{F}$ and $x > 0$, where the sets X_t, $0 \leq t \leq 1$ satisfy Property (i).

Remark: The condition that the function f in Theorem 1 disappears on the diagonal does not mean an important restriction in the applications we have in mind. On the other hand one can get rid of this condition by means of a more careful analysis if the contribution of the integral on the diagonal is separately estimated.

A special case of the theorem, when $s = 1$. In this case the result of Theorem 1 is the classical result of paper [dkw] or in multidimensional Euclidean space of paper [(kif]. However we have not determined the constant terms, whose calculation may demand more careful analysis.

The paper consists of three sections. In Section 2 we prove two lemmas, stating the theorem for finite system of sets. In Lemma 1 a weakened version of Theorem 1 is proved by means of Theorem A and in the case of simple step functions, a notion which will be introduced later. In Lemma 2 Theorem 1 is proved for general measurable functions under the restriction that the supremum is taken only on a finite set instead of all $0 \leq t \leq 1$. This is proved by means of an appropriate approximation of general measurable functions by simply step functions. Finally, Section 3 contains the proof of Theorem 1 in its original form.

2. Proof of some useful Lemmas

First we show that we may assume without violating the generality that the measure μ has no atoms. Indeed, by taking direct product with the space $([0,1],\mathcal{B}_1,\lambda)$, where $\mathcal{B}_1$ is the Borel σ-algebra on $[0,1]$, and λ is the Lebesgue measure we can reformulate the result of Theorem 1 in a space where the measure μ has no atoms. More explicitly, put

$$\begin{aligned}(\tilde{\Omega},\tilde{\mathcal{A}},\tilde{P})&=(\Omega\times[0,1]^n,\ \mathcal{A}\times\mathcal{B}_1^n\ P\times\lambda^n),\\(\tilde{X},\tilde{\mathcal{X}})&=(X\times[0,1],\ \mathcal{X}\times\mathcal{B}_1),\\\tilde{X}_t&=X_t\times[0,1]\ .\\\tilde{\xi}_k(\tilde{\omega})&=(\xi_k(\omega),x_k)\text{ if }\tilde{\omega}=(\omega,x_1,\dots,x_n),\ x_i\in[0,1],\end{aligned}$$

Given a function $f\in\mathcal{F}$, put

$$\tilde{f}(\tilde{u}_1,\dots,\tilde{u}_s)=f(u_1,\dots,u_s)\text{ if }\tilde{u}_j=(u_j,x_j),\ x_j\in[0,1],\quad j=1,\dots,n.$$

Then the distribution $\tilde{\mu}$ of $\tilde{\xi}_k$, $k=1,\dots,n$ has no atoms, and

$$\begin{aligned}&\int_{\tilde{X}_t}\int_{\tilde{X}}\cdots\int_{\tilde{X}}\tilde{f}(\tilde{u}_1\dots,\tilde{u}_s)\tilde{\mu}_n(d\tilde{u}_1)\tilde{\mu}_n(d\tilde{u}_2)\dots\tilde{\mu}_n(d\tilde{u}_s)\\&\qquad=\int_{X_t}\int_X\cdots\int_X f(u_1,\dots,u_s)\mu_n(du_1)\dots\mu_n(du_s).\end{aligned}$$

Hence it is enough to prove Theorem 1 to the $\tilde{}$ system whose measure $\tilde{\mu}$ has no atoms, since it also implies inequality (1.3) for the original system.

The strategy of our proof is the following. First we prove the estimate of Theorem 1 (when the supremum with respect of the first coordinate is dropped) by means of Theorem A for some special step functions which are constant on some special rectangles. The class of these step functions is sufficiently rich, so we can prove the estimate of Theorem 1 by taking their closure in an appropriate way. To prove the result with the supremum we shall consider such partitions of the space X which are adapted to the class of functions X_t appearing in Property (i.). Hence we introduce the following definitions.

Definition of simple step functions. *A function $f(u_1,\dots,u_s)\in\mathcal{F}$ is a* simple step function *if a partition $B_1,B_2,\dots,B_\ell$ of the space X and a set of numbers $c_{j_1,\dots,j_s}$ indexed by the s-tuples $(j_1,\dots,j_s)$, $1\le j_r\le\ell$, $j_r\ne j_p$ if $r\ne p$, can be given in such a way that*

$$f(u_1,\dots,u_s)=\begin{cases}c_{j_1,\dots,j_s} & \textit{if } u_1\in B_{j_1},\dots,u_s\in B_{j_s}\ B_{j_p}\cap B_{j_r}=\emptyset\\ & \qquad\textit{if } p\ne r,\ 1\le p,r\le s\\ 0 & \textit{otherwise.}\end{cases}$$

Definition of adapted partitions. *A finite partition $B_1,B_2,\dots,B_\ell$ of X is called* adapted *to a system of sets X_t, $0\le t\le1$, with Property (i) and fixed numbers $0=t_0<$*

$t_1 < \cdots < t_k = 1$ *if for all* B_i, $1 \le i \le \ell$, *there is some index* $j(i)$ *in such a way that* $B_i \in X_{t_{j(i)}} \setminus X_{t_{j(i)-1}}$.

Remark. Let us consider a system of sets X_t with Property (i), a finite set of numbers $0 = t_0 < t_1 < \cdots < t_k = 1$ and a finite partition $\mathcal{B}$ of X. Then it is always possible to define a finer partition $\mathcal{B}'$ which is adapted to X_t and $0 = t_0 < t_1 < \cdots < t_k = 1$. Indeed, the system of sets $B_i \cap (X_{t_j} \setminus X_{t_j - 1})$ for all i and j is a refinement of the partition $\mathcal{B}$ which is adapted to the system X_t and the numbers $0 = t_0 < t_1 < \cdots < t_k = 1$.

In the sequel we use the letters c_s, α_s, etc. for some appropriate constants. The same letter may denote different constant in different formulas.

Lemma 1. *Let* $f(u_1, \dots, u_s) \in \mathcal{F}$ *be a simple step function, where the class of functions* $\mathcal{F}$ *was defined in (1.2). Then*

(i)

$$P\left(\left|\int_X \cdots \int_X f(u_1, \dots, u_s)\mu_n(\,du_1) \dots \mu_n(\,du_s)\right| \ge x\right) \le C_s \exp\left\{-\alpha_s x^{2/s}\right\} \tag{2.1}$$

(ii) Let X_{t_j}, $j = 0, \dots, k$ *be finitely many elements of a system of sets with Property (i). Then*

$$\begin{aligned} P\left(\sup_{t_j\colon\, j=1,\dots,k} \left|\int_{X_{t_j}} \int_X \cdots \int_X f(u_1, \dots, u_s)\mu_n(\,du_1) \dots \mu_n(\,du_s)\right| \ge x\right) \\ \le C_s \exp\left\{-\alpha_s x^{2/s}\right\} \end{aligned} \tag{2.2}$$

Proof. (i) Let us consider $([0,1], \mathcal{B}_1, \lambda)$ where $\mathcal{B}_1$ is the Borel σ-algebra on $[0,1]$, λ is the Lebesgue measure and $\eta_1, \dots, \eta_n$ are independent uniformly distributed random variables on $[0,1]$ with empirical measure

$$\bar{\lambda}_n(D) = \frac{1}{n} \sum_{i=1}^{n} I(\eta_i \in D) \quad \forall\, D \in \mathcal{B}_1\ ,$$

and $\lambda_n(D) = \sqrt{n}(\bar{\lambda}_n(D) - \lambda(D))$ denotes its standardization. It is enough to construct a function $\tilde{f}\colon [0,1]^s \to [-1,1]$ in a such a way that

$$\begin{aligned} &\int_X \cdots \int_X f(u_1, \dots, u_s)\mu_n(\,du_1) \dots \mu_n(\,du_s) \\ &\quad \stackrel{\mathcal{D}}{=} \int_0^1 \cdots \int_0^1 \tilde{f}(x_1, \dots, x_s)\lambda_n(\,dx_1) \dots \lambda_n(\,dx_s) \end{aligned} \tag{2.3}$$

where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution, since the desired estimate follows from Theorem A. Let us define a partition $D_1, \dots, D_\ell$, of the interval $[0,1]$ in such a way that

$$\lambda(D_i) = \mu(B_i) = P(\xi \in B_i),\ i = 1, \dots, \ell,\ \bigcup_{i=1}^{\ell} D_i = [0,1] \tag{2.4}$$

and put

$$\tilde{f}(x_1,\dots,x_s)=\begin{cases} c_{j_1,\dots,j_s} & \text{if } x_1\in D_{j_1},\dots,x_s\in D_{j_s}.\ 1\le j_r\le \ell \\ & \qquad r=1,\dots \ell\ \ j_r\ne j_p\ \ \text{if } p\ne r \\ 0 & \text{otherwise.}\end{cases} \tag{2.5}$$

Then $\tilde{f}$ is a step function on $[0,1]^s$, and relation (2.3) holds.

(ii) Using the previous *Remark* we can suppose that the simple step function is defined on a finite partition $\mathcal{B}$ which is adapted to the system X_t and the number $0\le t_1<\dots<t_k$. Let us observe that the adapted partition, $\{B_i\}_{i=1}^{\ell}$ is finer than the partition $\{X_{t_j}\setminus X_{t_{j-1}}\}_{j=1}^{k}$. This means that either $B_i\cap X_{t_j}=B_i$ or $B_i\cap X_{t_j}=\emptyset$ for all $i=1,\dots,\ell$, $j=1,\dots,k$. Because $\mu(X_t)=t$, for all $0\le t\le 1$ it follows that $\mu(B_i)\le t_{j(i)}-t_{j(i)-1}$. It is possible to construct a partition $D_1,\dots,D_\ell$ of the interval $[0,1]$ with property (2.4) in such a way that it is finer than the partition $[0,t_1],(t_1,t_2],\dots,(t_{k-1},1]$. Let us define a step function $\tilde{f}$ on $[0,1]^s$ with relation (2.5). Then

$$\begin{aligned}&\int_0^{t_j}\dots\int_0^1 \tilde{f}(x_1,\dots,x_s)\lambda_n(dx_1)\dots\lambda_n(dx_s)\\ &\quad\stackrel{\mathcal{D}}{=}\int_{X_{t_j}}\dots\int_X f(u_1,\dots,u_s)\mu_n(\,du_1)\dots\mu_n(\,du_s),\quad \forall\, j=1,\dots,k,\end{aligned}$$

where $\stackrel{\mathcal{D}}{=}$ means that the joint distribution of the vectors $j=1,\dots,k$ on the left and right side agree. Then statement (ii) of the lemma follows from Theorem A.

Lemma 2. *Statement (ii) of Lemma 1 holds for all measurable $f(u_1,\dots,u_s)\in\mathcal{F}$ functions, where the class of functions $\mathcal{F}$ was defined in (1.2).*

Proof. We prove this lemma in two steps. First we show that it is enough to prove the statement for a general step function, then we approximate a general step function by simple step functions for which the statement holds by Lemma 1.

For all bounded and measurable function f it is possible to construct a measurable step function f' which is close to f in the supremum norm. It can be supposed that the step function $f'\in\mathcal{F}$, i.e. it is equal to zero on the diagonal set A since $f\in\mathcal{F}$. More precisely for any $x>0$ and fixed dimension s and number of sample points n there exist a step function $f'\in\mathcal{F}$ on the product space $(X^s,\mathcal{X}^s)$ such that

$$\sup_{(u_1,\dots,u_s)\in X^s}|f'(u_1,\dots,u_s)-f(u_1,\dots,u_s)|\le\frac{x}{2(2\sqrt{n})^s}\,. \tag{2.6}$$

Then

$$\begin{aligned}&P\left(\sup_{t_j:\ j=1,\dots,k}\left|\int_{X_{t_j}}\cdots\int_X f(u_1,\dots,u_s)\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge x\right)\\ &\quad\le P\left(\sup_{t_j:\ j=1,\dots,k}\left|\int_{X_{t_j}}\cdots\int_X (f-f')\,\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge\frac{x}{2}\right)\\ &\qquad+P\left(\sup_{t_j:\ j=1,\dots,k}\left|\int_{X_{t_j}}\cdots\int_X f'\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge\frac{x}{2}\right).\end{aligned}\tag{2.7}$$

Since the total variation of μ_n is bounded by $2\sqrt{n}$ formula (2.6) implies that

$$\sup_{t_j:\, j=1,\dots,k} \left| \int_{X_{t_j}} \cdots \int_X (f-f')\, \mu_n(\,du_1)\dots\mu_n(\,du_s) \right|$$
$$\le \sup_{t_j:\, j=1,\dots,k} \int_{X_{t_j}} \cdots \int_X |f-f'|\, |\mu_n|(\,du_1)\dots|\mu_n|(\,du_s) \le \frac{x}{2}\,.$$

Thus the first term on the right-hand side of inequality (2.7) equals zero, and it is enough to prove Lemma 2 for step functions.

Let us consider a step function $f \in \mathcal{F}$

$$f(u_1,\dots,u_s) = \begin{cases} c_j & \text{if } (u_1,\dots,u_s) \in Y_j,\ Y_j \cap A = \emptyset,\ j=1,\dots,k \\ & \quad Y_j \cap Y_i = \emptyset \text{ if } j \ne i,\ 1 \le i,j \le k \\ 0 & \text{otherwise.} \end{cases}$$

Since $(X,\mathcal{X})$ is a separable metric space any measurable set in the product space can be approximated with finite union of rectangular type sets. More explicitly, for each Y_j and for any given $\delta > 0$ there exists finite union of disjoint rectangles R'_j such that

$$\mu^{(s)}(Y_j \triangle R'_j) \le \frac{\delta}{2}\,,$$

where $\triangle$ denotes symmetric difference and $\delta > 0$ is a sufficiently small number which will be chosen later.

If $Y_j \cap A = \emptyset$, $j=1,\dots,k$, where A denotes the diagonal on X^s, then there exists a good approximating set R_j of Y_j consisting of finitely many disjoint rectangles such that also the relation $R_j \cap A = \emptyset$ holds. Indeed, since $(X,\mathcal{X})$ is separable metric space and $\mu^{(s)}(A) = 0$, the diagonal set A can be covered by the union of finitely many rectangles R'' such that $\mu^{(s)}(R'') \le \frac{\delta}{2}$. Then the sets $R_j = R'_j \setminus R''$ also can be represented as the union of finitely many disjoint rectangles, they are disjoint from A and close to the original Y_j, i.e.

$$\mu^{(s)}(Y_j \triangle R_j) \le \delta, \quad j=1,\dots,k\,.$$

It is possible to define a simple step function $\tilde{f}$ with the help of the approximating set of rectangles for which the set where $\tilde{f} \ne f$ is small.

To do this first we replace the sets R_j which approximate well the level sets Y_j and disjoint of the diagonal A by such sets $\bar{R}_j$ which also have this property and beside this, they are disjoint. We define $\bar{R}_j = R_j \setminus \bigcap\limits_{\ell=1}^{j-1} R_\ell$. These sets also can be represented as the union of disjoint rectangles, they are disjoint. Since the sets Y_j are disjoint, $\mu(Y_j \triangle R_j) \le \delta$ for all $1 \le j \le k$, $\mu(R_j \cap R_\ell) \le 2\delta$ if $j \ne \ell$, and

$$\mu^{(s)}(\bar{R}_j \triangle Y_j) \le 2j\delta \quad \text{for all } 1 \le j \le k \tag{2.8}$$

The simple step function $\tilde f$ will be defined as

$$\tilde f(u_1,\dots,u_s)=\begin{cases} c_j & \text{if } (u_1,\dots,u_s)\in \bar R_j, \quad j=1,\dots,k\\ 0 & \text{otherwise.}\end{cases}$$

Put $C=\{(u_1,\dots,u_s)\colon f\neq \tilde f\}$. Then because of relation (2.8)

$$\mu^{(s)}(C)\le 2(\delta+2\delta+\dots+k\delta)=k(k+1)\delta\ . \tag{2.9}$$

Obviously

$$\begin{aligned} P&\left(\sup_{t_j:\ j=1,\dots,k}\left|\int_{X_{t_j}}\cdots\int_X f(u_1,\dots,u_s)\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge x\right)\\ &\le P\left(\sup_{t_j:\ j=1,\dots,k}\left|\int_{X_{t_j}}\cdots\int_X \tilde f\mu_n(\,du_1)\dots\mu_n(du_s)\right|\ge \frac{x}{2}\right)\\ &\qquad + P\left(\sup_{t_j:\ j=1,\dots,k}\left|\int_{X_{t_j}}\cdots\int_X \left(f-\tilde f\right)\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge \frac{x}{2}\right). \end{aligned}\tag{2.10}$$

The first term at the right hand-side of (2.10) is already estimated in Lemma 1, and we will show that the second term is also small. Since $|\mu_n|=|\sqrt{n}(\bar\mu_n-\mu)|\le\sqrt{n}(\bar\mu_n+\mu)$ we get that

$$\begin{aligned} P&\left(\sup_{t_j:\ j=1,\dots,k}\left|\int_{X_{t_j}}\cdots\int_X \left(f-\tilde f\right)\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge \frac{x}{2}\right)\\ &\le P\left((2\sqrt{n})^s(\bar\mu_n+\mu)\times\dots\times(\bar\mu_n+\mu)(C)\ge\frac{x}{2}\right)\\ &= P\left((2\sqrt{n})^s\sum_{\ell=0}^s\binom{s}{\ell}\underbrace{\bar\mu_n\times\dots\times\bar\mu_n}_{\ell\text{ times}}\times\underbrace{\mu\times\dots\times\mu}_{s-\ell\text{ times}}(C)\ge\frac{x}{2}\right)\\ &\le\sum_{\ell=0}^s P\left(\underbrace{\bar\mu_n\times\dots\times\bar\mu_n}_{\ell\text{ times}}\times\underbrace{\mu\times\dots\times\mu}_{s-\ell\text{times}}(C)\ge\frac{x}{2(2\sqrt{n})^s(s+1)\binom{s}{\ell}}\right). \end{aligned}\tag{2.11}$$

Applying the Markov inequality to the last term of (2.11) results the following inequality

$$\begin{aligned} P&\left(\sup_{t_j:\ j=1,\dots,k}\left|\int_{X_{t_j}}\cdots\int_X \left(f-\tilde f\right)\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge \frac{x}{2}\right)\\ &\le\sum_{\ell=0}^s E(\underbrace{\bar\mu_n\times\dots\times\bar\mu_n}_{\ell\text{ times}}\times\underbrace{\mu\times\dots\times\mu}_{s-\ell\text{ times}}(C))\frac{2(2\sqrt{n})^s(s+1)\binom{s}{\ell}}{x}\ . \end{aligned}\tag{2.12}$$

Consider for fixed ℓ the expected value of the random variable

$$\underbrace{\bar{\mu}_n \times \cdots \times \bar{\mu}_n}_{\ell \text{ times}} \times \underbrace{\mu \times \cdots \times \mu}_{s-\ell \text{ times}}(\cdot)$$

on a rectangle $B = B_1 \times \cdots \times B_s$ for which $B \cap A = \emptyset$, i.e. none of the ξ_i-s is contained in more than one B_j. Notice that

$$\bar{\mu}_n \times \cdots \times \bar{\mu}_n(B_1 \times \cdots \times B_\ell) = \frac{1}{n^\ell} \sum_{(j_1,\dots,j_\ell):\ 1\le j_i \le n} I(\xi_{j_1} \in B_1, \dots, \xi_{j_\ell} \in B_\ell) \ .$$

The above sum contains exactly $n(n-1)\cdots(n-\ell+1)$ different terms and

$$\begin{aligned} &E\left(\underbrace{\bar{\mu}_n \times \cdots \times \bar{\mu}_n}_{\ell \text{ times}} \times \underbrace{\mu \times \cdots \times \mu}_{s-\ell \text{ times}}(B_1 \times \cdots \times B_s)\right) \\ &\qquad = \frac{n(n-1)\cdots(n-\ell+1)}{n^\ell}\mu(B_1)\cdots\mu(B_s) \\ &\qquad = \frac{n(n-1)\cdots(n-\ell+1)}{n^\ell}\mu^{(s)}(B) \ . \end{aligned} \tag{2.13}$$

Thus the above expected value is equal to a measure on the set of rectangles disjoint from A. This measure can be extended for all measurable set of the product space disjoint from A. Therefore relation (2.13) holds for any measurable set disjoint from A. From relations (2.9) and (2.11)–(2.13) it follows that

$$\begin{aligned} &P\left(\sup_{t_j:\ j=1,\dots,k} \left| \int_{X_{t_j}} \cdots \int_X \left(f-\tilde{f}\right)\mu_n(\,du_1)\dots\mu_n(\,du_s)\right| \ge \frac{x}{2}\right) \\ &\qquad \le \frac{2^{2s}n^{s/2}(s+1)k(k+1)\delta}{x} \ . \end{aligned} \tag{2.14}$$

With the choice of a

$$\delta \le \frac{xC_s\exp\{-\alpha_s x^{2/s}\}}{k(k+1)(s+1)2^{2s}n^{s/2}} \tag{2.15}$$

the statement of the lemma follows from relations (2.10), (2.14) and (2.15).

3. Proof of Theorem 1.

To prove Theorem 1 consider a system of sets X_t with Property (i). Let $N = N(n,x,s)$ be an integer, and consider the sets $X_{j/N}$, $j = 0,1,\dots,N$. Then for any t: $\frac{j}{N} \le t < \frac{j+1}{N}$

$$\begin{aligned} &\sup_{0\le t\le 1}\left|\int_{X_t}\cdots\int_X f\,\mu_n(\,du_1)\dots\mu_n(\,du_s)\right| \\ &\qquad \le \sup_{0\le j\le N}\left|\int_{X_{j/N}}\cdots\int_X f\,\mu_n(\,du_1)\dots\mu_n(\,du_s)\right| \\ &\qquad\qquad + \sup_{0\le j<N}\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\cdots\int_X f\,\mu_n(\,du_1)\dots\mu_n(\,du_s)\right| \end{aligned} \tag{3.1}$$

Hence

$$\begin{aligned} &P\left(\sup_{0\le t\le 1}\left|\int_{X_t}\cdots\int_X f\,\mu_n(\,du_1)\dots\mu_n(\,du_s)\right| \ge x\right) \\ &\quad \le P\left(\max_{0\le j\le N}\left|\int_{X_{j/N}}\cdots\int_X f\,\mu_n(\,du_1)\dots\mu_n(\,du_s)\right| \ge \frac{x}{2}\right) \\ &\qquad + P\left(\max_{0\le j<N}\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\cdots\int_X f\,\mu_n(\,du_1)\dots\mu_n(\,du_s)\right| \ge \frac{x}{2}\right). \end{aligned} \tag{3.2}$$

It is enough to prove the theorem for the second term of the right-hand side of (3.2), since the desired estimation is proved for the first term in Lemma 2.

Let B denote the event that the set $X_{j/N}\setminus X_{(j-1)/N}$ does not contain more than one ξ_k for all $1\le j\le N$. Then

$$B^c = \{\exists\, k,\ell,j\colon \xi_k,\xi_\ell \in (X_{j/N}\setminus X_{(j-1)/N}),\ k\ne\ell,\ 1\le k,\ell\le n,\ 1\le j\le N\}\,. \tag{3.3}$$

For a fixed j, $(1\le j\le N)$, it is obvious that $P(\xi_k\in(X_{j/N}\setminus X_{(j-1)/N})) = \frac{1}{N}$ for any $1\le k\le n$. Therefore for any fixed j, $1\le j\le N$

$$P(\exists\, k,\ell\colon \xi_k,\, \xi_\ell \in (X_{j/N}\setminus X_{(j-1)/N}),\ k\ne\ell,\ 1\le k,\ell\le n) \le \left(\frac{n}{N}\right)^2\ ,$$

and

$$P(B^c) \le N\left(\frac{n}{N}\right)^2 = \frac{n^2}{N}\,. \tag{3.4}$$

Choosing

$$N \ge n^2\exp\left\{x^{2/s}\right\}, \tag{3.5}$$

it follows that

$$P(B^c) \le \exp\left\{-x^{2/s}\right\}. \tag{3.6}$$

Thus we need to prove, that

$$I = P\left(\left\{\max_{0\le j<N}\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\cdots\int_X f\,\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge\frac{x}{2}\right\}\cap B\right) \\ \le c_1\exp\left\{-c_2x^{2/s}\right\}\ . \tag{3.7}$$

Define the events

$$\begin{aligned} B_{i,j} &= \{\xi_i\in X_{(j+1)/N}\setminus X_{j/N},\ \xi_\ell\notin X_{(j+1)/N}\setminus X_{j/N},\ \ell\ne i,\ 1\le\ell\le n\}\ ,\\ B_{0,j} &= \{\xi_i\notin X_{(j+1)/N}\setminus X_{j/N},\ 1\le i\le n\}\end{aligned}$$

for $i=1,\dots,n$, $j=0,\dots,N-1$. The event B is the union of the above disjoint events

$$B=\bigcup_{j=0}^{N-1}\bigcup_{i=1}^{n}B_{i,j}.$$

Thus

$$I\le\sum_{j=0}^{N-1}P\left(\left\{\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\cdots\int_X f\,\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge\frac{x}{2}\right\}\cap B\right) \\ \sum_{j=0}^{N-1}\sum_{i=0}^{n}P\left(\left\{\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\cdots\int_X f\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge\frac{x}{2}\right\}\cap B_{i,j}\right). \tag{3.8}$$

To prove Theorem 1 it is enough to show that all terms of (3.8) can be bounded as

$$P\left(\left\{\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\cdots\int_X f\,\mu_n(\,du_1)\cdots\mu_n(\,du_s)\right|\ge\frac{x}{2}\right\}\cap B_{i,j}\right) \\ \le\frac{c_1}{n\,N}e^{-c_2x^{2/s}}. \tag{3.9}$$

Since $\mu_n(du_1)=\sqrt{n}(\bar\mu_n(du_1)-\mu(du_1))$ the integral (3.9) can be estimated in the following way:

$$\begin{aligned} &P\left(\left\{\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\cdots\int_X f\,\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge\frac{x}{2}\right\}\cap B_{i,j}\right)\le II+III\\ &\quad=P\left(\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\cdots\int_X f\,\sqrt{n}\mu(\,du_1)\dots\mu_n(\,du_s)\right|\ge\frac{x}{4}\right)\\ &\qquad+P\left(\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\cdots\int_X f\sqrt{n}\bar\mu_n(\,du_1)\dots\mu_n(\,du_s)\right|\ge\frac{x}{4}\,\middle|\,B_{i,j}\right)P(B_{i,j}).\end{aligned} \tag{3.10}$$

Because the integral in II can be bounded as

$$\sup_{\frac{j}{N}<t\le\frac{j+1}{N}} \left| \int_{X_t\setminus X_{j/N}} \cdots \int_X f\,\sqrt{n}\mu(\,du_1)\dots\mu_n(\,du_s)\right| \le \frac{\sqrt{n}}{N}(2\sqrt{n})^{s-1}\,, \tag{3.11}$$

and $II=0$ if N is chosen in such a way that

$$N\ge 2^{s+1}\frac{(\sqrt{n})^s}{x}\,.$$

In the estimation of III we may assume that $i>0$ because $III=0$ if there is no ξ in the set $X_{(j+1)/N}\setminus X_{j/N}$. Introduce the notation $Y_j=X_{(j+1)/N}\setminus X_{j/N}$, and $Y_j^c=X\setminus Y_j$. Split the domain of integration $(X_t\setminus X_{j/N})\times X^{s-1}$ in the way

$$(X_t\setminus X_{j/N})\times X^{s-1}=\bigcup_{j(\ell)=\pm1,\ \ell=2,\dots,s}(X_t\setminus X_{j/N})\times Z(j(2))\times\cdots\times Z(j(s))\,,$$

where $Z(1)=Y_j$, $Z(-1)=Y_j^c$. In this way we get a sum of 2^{s-1} stochastic integrals. Consider a term containing at least one coordinate Y_j. For the sake of simplicity consider such a term where the second coordinate is Y_j. In this case we get by splitting the measure $\mu_n(du_2)=\sqrt{n}(\bar\mu_n(du_2)-\mu(du_2))$ into two parts that

$$\begin{aligned}
&\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\int_{Y_j}\cdots\int_{Y_j^c} f\sqrt{n}\bar\mu_n(\,du_1)\mu_n(\,du_2)\dots\mu_n(\,du_s)\right|\\
&\quad\le \sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\int_{Y_j}\cdots\int_{Y_j^c} f\sqrt{n}\bar\mu_n(\,du_1)\sqrt{n}\bar\mu_n(\,du_2)\dots\mu_n(\,du_s)\right|\\
&\qquad+\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\int_{Y_j}\cdots\int_{Y_j^c} f\sqrt{n}\bar\mu_n(\,du_1)\sqrt{n}\mu(\,du_2)\dots\mu_n(\,du_s)\right|\,.
\end{aligned}\tag{3.12}$$

Notice that in the first term on the right-hand side of (3.12) the first two coordinates of the domain of integration are contained in the set Y_j, and in these coordinates integration is taken with respect to the empirical measure $\bar\mu_n$. Because there is only one sample element ξ_i in Y_j, and $f=0$ on the diagonal set, this implies that the first term at the right-hand side of (3.12) equals zero. The probability associated with the second term of the right-hand side of (3.12) can be bounded by using similar argument as in the estimation of the term II in formula (3.10). This time we have

$$P\left(\sup_{\frac{j}{N}<t\le\frac{j+1}{N}}\left|\int_{X_t\setminus X_{j/N}}\int_{Y_j}\cdots\int_{Y_j^c} f\sqrt{n}\bar\mu_n(\,du_1)\sqrt{n}\mu(\,du_2)\dots\mu_n(\,du_s)\right|\ge\frac{x}{2^{s+1}}\right)\,, \tag{3.13}$$

and choosing

$$N>2^{s+1}\frac{(\sqrt{n})^s}{x} \tag{3.14}$$

the probability in (3.13) is equal to 0. To complete the proof of Theorem 1 we have to give a good estimate of the following probability:

$$P\left(\sup_{\frac{j}{N}<t\le\frac{j+1}{N}} \left| \int_{X_t\setminus X_{j/N}} \int_{Y_j^c} \cdots \int_{Y_j^c} f\sqrt{n}\bar{\mu}_n(\,du_1)\mu_n(\,du_2)\dots\mu_n(\,du_s) \right| \ge \frac{x}{2^{s+1}} \,\middle|\, B_{i,j} \right). \tag{3.15}$$

Since the random variables ξ-s are conditionally independent under the condition $B_{i,j}$, and only the variable ξ_i takes its values in Y_j, we are able to get rid of the supremum taking the integral according to u_1:

$$\begin{aligned} &P\left(\sup_{\frac{j}{N}<t\le\frac{j+1}{N}} \left| \int_{X_t\setminus X_{j/N}} \int_{Y_j^c} \cdots \int_{Y_j^c} f\sqrt{n}\bar{\mu}_n(\,du_1)\mu_n(\,du_2)\dots\mu_n(\,du_s) \right| \ge \frac{x}{2^{s+1}} \,\middle|\, B_{i,j} \right) \\ &\qquad \le P\left(\left| \frac{1}{\sqrt{n}} \int_{Y_j^c} \cdots \int_{Y_j^c} f(\xi_i,u_2,\dots,u_s)\mu_n(\,du_2)\dots\mu_n(\,du_s) \right| \ge \frac{x}{2^{s+1}} \,\middle|\, B_{i,j} \right) \end{aligned} \tag{3.16}$$

Let us denote by $\mu\prime$ the restriction of measure μ and by μ_{n1} the standardization of the empirical measure on Y_j^c

$$\begin{aligned} \mu^* &= \frac{N}{N-1}\mu\prime, \\ \mu_{n1}(C) &= \sqrt{n-1}\left(\frac{1}{n-1} \sum_{k\ne i,\, k=1}^{n} I(\xi_k\in C) - \mu^*(C) \right). \end{aligned} \tag{3.17}$$

for any measurable set $C\in Y_j^c$. On the set Y_j^c

$$\mu_n = \frac{\sqrt{n-1}}{\sqrt{n}}\mu_{n1} - \frac{1}{\sqrt{n}}\frac{N-n}{N}\mu^*. \tag{3.18}$$

Using the above decomposition of μ_n the $s-1$ integrals in formula (3.16) can be written as the sum of 2^{s-1} integrals and the sum

$$\begin{aligned} &\sum_{\ell=0}^{s-1} \sum_{\substack{2\le i_1<\dots<i_\ell\le s \\ 2\le j_1<\dots<j_{s-1-\ell}\le s}} \\ &\quad P\left(\left| \int\limits_{Y_j^c} \cdots \int\limits_{Y_j^c} f\mu^*(du_{j_1})\dots\mu^*(du_{j_{s-1-l}})\mu_{n1}(\,du_{i_1})\dots\mu_{n1}(\,du_{i_\ell}) \right| \ge \frac{xn^{(s-\ell)/2}}{2^{2s+1}} \,\middle|\, B_{i,j} \right) \end{aligned} \tag{3.19}$$

is an upper estimation of (3.17). If $\ell=0$ only a deterministic integral appears, and the associated probability is equal to 0 with the choice of a $x\ge c$. This condition is not a restriction. The statement of the theorem is trivial for small x since the right-hand side of (1.3) is equal to 1 for $x \le c$. For $\ell \ge 1$ we shall bound each term of (3.19) by first integrating with respect to $u_{i_1}, \dots, u_{i_\ell}$, and then for $\ell \ge 1$ estimating the integrals arising this way with the help of Lemma 2. Introduce the functions

$$h(\xi_i, u_{i_1}, \dots, u_{i_\ell}) = \int_{Y_j^c} \cdots \int_{Y_j^c} f \mu^*(\,du_{j_1}) \dots \mu^*(\,du_{j_{s-1-l}})\,.$$

Since $|h| \le 1$, and the functions $h \in \mathcal{F}$, Lemma 2 is applicable for a general term of (3.19) with $s' = \ell$. Thus we get the following inequality:

$$\begin{aligned} &P\left(\left| \int_{Y_j^c} \cdots \int_{Y_j^c} f \mu^*(\,du_{j_1}) \dots \mu^*(\,du_{j_{s-1-\ell}}) \mu_{n1}(\,du_{i_1}) \dots \mu_{n1}(\,du_{i_\ell}) \right| \ge \frac{x(\sqrt{n})^{s-\ell}}{2^{2s+1}} |B_{i,j} \right) \\ &\quad \le P\left(\left| \int_{Y_j^c} \dots \int_{Y_j^c} h\, \mu_{n1}(du_{i_1}) \dots \mu_{n1}(du_{i_\ell}) \right| \ge \frac{x\,(\sqrt{n})^{s-\ell}}{2^{2s+1}} |B_{i,j} \right) \\ &\quad \le c_1 \exp\left(-c_2 x^{2/\ell} n^{(s-\ell)/\ell} \right). \end{aligned} \tag{3.20}$$

Again, if $x \le c$

$$c_1 \exp\left\{ -c_2 x^{2/\ell} n^{(s-\ell)/\ell} \right\} \le \frac{1}{n} c_1' \exp\left\{ -c_2' x^{2/s} \right\}. \tag{3.21}$$

Since $P(B_{ij}) \le \frac{1}{N}$, inequality (3.9) follows from (3.10), (3.12), (3.15), (3.16) and (3.19)–(3.21). Theorem 1 is proved.

References:

dkw Dvoretzky, A., Kiefer, J.C. and Wolfowitz, J. (1956) Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, *Annals of Mathematical Statistics*, **27**, 642–669.

kif Kiefer, J.C. (1961) On large deviation of the empiric D.F. of vector chance variables and a law of iterated logarithm, *Pacific J. Mathematics*, **2**, 649–660.

mp Major, P. (1988) On the tail behaviour of the distribution function of multiple stochastic integrals, *Probability Theory and Related Fields*, **78**, 419–435.

mr Major, P. and Rejtő, L. (1988) Strong embedding of the distribution function under random censorship, *Annals of Statistics*, **16**, 1113–1132.

mrt Major, P., Rejtő, L. and Tusnády, G. (1997) Strong embedding of a baseline function estimator in the Cox model, (manuscript).