Appendix A
The Proof of Some Results About Vapnik–Červonenkis Classes

Proof of Theorem 5.1. (Sauer's lemma). This result has several different proofs. Here I write down a relatively simple proof of P. Frankl and J. Pach which appeared in [17]. It is based on some linear algebraic arguments.

The following equivalent reformulation of Sauer's lemma will be proved. Let us take a set $S = S(n)$ consisting of n elements and a class $\mathcal{E}$ of subsets of S consisting of m subsets $E_1, \dots, E_m \subset S$. Assume that $m \ge m_0 + 1$ with $m_0 = m_0(n,k) = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{k-1}$. Then there exists a set $F \subset S$ of cardinality k which is shattered by the class of sets $\mathcal{E}$. Actually, it is enough to show that there exists a set F of cardinality greater than or equal to k which is shattered by the class of sets $\mathcal{E}$, because if a set has this property, then all of its subsets have it. This latter statement will be proved.

To prove this statement let us first list the subsets $X_0, \dots, X_{m_0}$ of the set S of cardinality less than or equal to $k-1$, and correspond to all sets $E_i \in \mathcal{E}$ the vector $e_i = (e_{i,1}, \dots, e_{i,m_0})$, $1 \le i \le m$, with elements

$$e_{i,j} = \begin{cases} 1 & \text{if } X_j \subseteq E_i \\ 0 & \text{if } X_j \not\subseteq E_i \end{cases} \qquad 1 \le i \le m, \text{ and } 1 \le j \le m_0.$$

Since $m > m_0$, the vectors $e_1, \dots, e_m$ are linearly dependent. Because of the definition of the vectors e_i, $1 \le i \le m$, this can be expressed in the following way: There is a non-zero vector $(f(E_1), \dots, f(E_m))$ such that

$$\sum_{E_i\colon E_i \supseteq X_j} f(E_i) = 0 \quad \text{for all } 1 \le j \le m_0. \tag{A.1}$$

Let F, $F \subset S$, be a *minimal* set with the property

$$\sum_{E_i\colon E_i \supseteq F} f(E_i) = \alpha \ne 0. \tag{A.2}$$

P. Major, *On the Estimation of Multiple Random Integrals and U-Statistics*, Lecture Notes in Mathematics 2079, DOI 10.1007/978-3-642-37617-7,

Such a set F really exists, since every maximal element of the family $\{E_i\colon 1 \le i \le m,\ f(E_i) \ne 0\}$ satisfies relation (A.2). The requirement that F should be a minimal set means that if F is replaced by some $H \subset F$, $H \ne F$, at the left-hand side of (A.2), then this expression equals zero. The inequality $|F| \ge k$ holds because of relation (A.1) and the definition of the sets X_j.

Introduce the quantities

$$Z_F(H) = \sum_{E_i\colon E_i \cap F = H} f(E_i)$$

for all $H \subseteq F$.

Then $Z_F(F) = \alpha$, and for any set of the form $H = F \setminus \{x\}$, $x \in F$,

$$Z_F(H) = \sum_{E_i\colon E_i \cap F = H} f(E_i) = \sum_{E_i\colon E_i \supseteq H} f(E_i) - \sum_{E_i\colon E_i \supseteq F} f(E_i) = 0 - \alpha = -\alpha$$

because of the minimality property of the set F.

Moreover, the identity

$$Z_F(H) = (-1)^p \alpha \quad \text{for all } H \subseteq F \text{ such that } |H| = |F| - p,\ 0 \le p \le |F|. \tag{A.3}$$

holds. To show relation (A.3) observe that

$$Z_F(H) = \sum_{E_i\colon E_i \cap F = H} f(E_i) = \sum_{j=0}^{p} (-1)^j \sum_{G\colon H \subset G \subset F,\ |G| = |H| + j} \ \sum_{E_i\colon E_i \supseteq G} f(E_i) \tag{A.4}$$

for all sets $H \subset F$ with cardinality $|H| = |F| - p$. Identity (A.4) holds, since the term $f(E_i)$ is counted at the right-hand side of (A.4) $\sum\limits_{j=0}^{l} (-1)^j \binom{l}{j} = (1-1)^l = 0$ times if $E_i \cap F = G$ with some $H \subset G \subseteq F$ with $|G| = |H| + l$ elements, $1 \le l \le p$, while in the case $E_i \cap F = H$ it is counted once. Relation (A.4) together with (A.2) and the minimality property of the set F imply relation (A.3).

It follows from relation (A.3) and the definition of the function $Z_F(H)$ that for all sets $H \subseteq F$ there exists some set E_i such that $H = E_i \cap F$, i.e. F is shattered by $\mathscr{E}$. Since $|F| \ge k$, this implies Theorem 5.1. □

Proof of Theorem 5.3. Let us fix an arbitrary set $F = \{x_1, \dots, x_{k+1}\}$ of the set X, and consider the set of vectors $\mathscr{G}_k(F) = \{(g(x_1), \dots, g(x_{k+1}))\colon g \in \mathscr{G}_k\}$ of the $k+1$-dimensional space R^{k+1}. By the conditions of Theorem 5.3 $\mathscr{G}_k(F)$ is an at most k-dimensional subspace of R^{k+1}. Hence there exists a non-zero vector $a = (a_1, \dots, a_{k+1})$ such that $\sum\limits_{j=1}^{k+1} a_j g(x_j) = 0$ for all $g \in \mathscr{G}_k$. We may assume that

the set $A = A(a) = \{j\colon a_j < 0,\ 1 \le j \le k+1\}$ is non-empty, by multiplying the vector a by -1 if it is necessary.

Thus the identity

$$\sum_{j \in A} a_j g(x_j) = \sum_{j \in \{1,\dots,k+1\} \setminus A} (-a_j) g(x_j), \qquad \text{for all } g \in \mathscr{G}_k \tag{A.5}$$

holds. Put $B = \{x_j\colon j \in A\}$. Then $B \subset F$, and $F \setminus B \ne \{x\colon g(x) \ge 0\} \cap F$ for all $g \in \mathscr{G}_k$. Indeed, if there were some $g \in \mathscr{G}_k$ such that $F \setminus B = \{x\colon g(x) \ge 0\} \cap F$, then the left-hand side of (A.5) would be strictly positive (as $a_j < 0$, $g(x_j) < 0$ if $j \in A$, and $A \ne \emptyset$) its right-hand side would be non-positive for this $g \in \mathscr{G}_k$, and this is a contradiction.

The above proved property means that $\mathscr{D}$ shatters no set $F \subset X$ of cardinality $k+1$. Hence Theorem 5.1 implies that $\mathscr{D}$ is a Vapnik–Červonenkis class. $\square$

Appendix B
The Proof of the Diagram Formula for Wiener–Itô Integrals

We start the proof of Theorem 10.2A (the diagram formula for the product of two Wiener–Itô integrals) with the proof of inequality (10.13). To show that this relation holds let us observe that the Cauchy inequality yields the following bound on the function $F_\gamma(f,g)$ defined in (10.11) (with the notation introduced there):

$$\begin{aligned} &F_\gamma^2(f,g)(x_{(1,j)},x_{(2,j')},\ \ (1,j)\in V_1(\gamma),\ (2,j')\in V_2(\gamma)) \\ &\quad\le \int f^2(x_{\alpha_\gamma(1,1)},\dots,x_{\alpha_\gamma(1,k)}) \prod_{(2,j)\in\{(2,1),\dots,(2,l)\}\setminus V_2(\gamma)} \mu(\,dx_{(2,j)}) \\ &\qquad \int g^2(x_{(2,1)},\dots,x_{(2,l)}) \prod_{(2,j)\in\{(2,1),\dots,(2,l)\}\setminus V_2(\gamma)} \mu(\,dx_{(2,j)}). \end{aligned} \tag{B.1}$$

The expression at the right-hand side of inequality (B.1) is the product of two functions with different arguments. The first function has arguments $x_{(1,j)}$ with $(1,j)\in V_1(\gamma)$ and the second one $x_{(2,j')}$ with $(2,j')\in V_2(\gamma)$. By integrating both sides of inequality (B.1) with respect to these arguments we get inequality (10.13).

Relation (10.14) will be proved first for the product of the Wiener–Itô integrals of two elementary functions. Let us consider two (elementary) functions $f(x_1,\dots,x_k)$ and $g(x_1,\dots,x_l)$ given in the following form: Let some disjoint sets $A_1,\dots,A_M$, $\mu(A_s)<\infty$, $1\le s\le M$, be given together with some real numbers $c(s_1,\dots,s_k)$ indexed with such k-tuples $(s_1,\dots,s_k)$, $1\le s_j\le M$, $1\le j\le k$, for which the numbers $s_1,\dots,s_k$ in a k-tuple are all different. Put $f(x_1,\dots,x_k)=c(s_1,\dots,s_k)$ if $(x_1,\dots,x_k)\in A_{s_1}\times\cdots\times A_{s_k}$ with some vector $(s_1,\dots,s_k)$ with different coordinates, and let $f(x_1,\dots,x_k)=0$ if $(x_1,\dots,x_k)$ is outside of these rectangles. Take similarly some disjoint sets $B_1,\dots,B_{M'}$, $\mu(B_t)<\infty$, $1\le t\le M'$, and some real numbers $d(t_1,\dots,t_l)$, indexed with such l-tuples $(t_1,\dots,t_l)$, $1\le t_{j'}\le M'$, $1\le j'\le l$, for which the numbers $t_1,\dots,t_l$ in an l-tuple are different. Put $g(x_1,\dots,x_l)=d(t_1,\dots,t_l)$ if $(x_1,\dots,x_l)\in B_{t_1}\times\cdots\times B_{t_l}$ with edges indexed with some of the above introduced l-tuples, and let $g(x_1,\dots,x_l)=0$ otherwise.

P. Major, *On the Estimation of Multiple Random Integrals and U-Statistics*, Lecture Notes in Mathematics 2079, DOI 10.1007/978-3-642-37617-7,

Let us take some small number $\varepsilon > 0$ and rewrite the above introduced functions $f(x_1,\dots,x_k)$ and $g(x_1,\dots,x_l)$ with the help of this number $\varepsilon > 0$ in the following way. Divide the sets $A_1,\dots,A_M$ to smaller sets $A_1^\varepsilon,\dots,A_{M(\varepsilon)}^\varepsilon$, $\bigcup\limits_{s=1}^{M(\varepsilon)} A_s^\varepsilon = \bigcup\limits_{s=1}^{M} A_s$, in such a way that all sets $A_1^\varepsilon,\dots,A_{M(\varepsilon)}^\varepsilon$ are disjoint, and $\mu(A_s^\varepsilon) \le \varepsilon$, $1 \le s \le M(\varepsilon)$. Similarly, take sets $B_1^\varepsilon,\dots,B_{M'(\varepsilon)}^\varepsilon$, $\bigcup\limits_{t=1}^{M'(\varepsilon)} B_t^\varepsilon = \bigcup\limits_{t=1}^{M'} B_t$, in such a way that all sets $B_1^\varepsilon,\dots,B_{M'(\varepsilon)}^\varepsilon$ are disjoint, and $\mu(B_t^\varepsilon) \le \varepsilon$, $1 \le t \le M'(\varepsilon)$. Beside this, let us also demand that two sets A_s^ε and B_t^ε, $1 \le s \le M(\varepsilon)$, $1 \le t \le M'(\varepsilon)$, are either disjoint or they agree. Such a partition exists because of the non-atomic property of measure μ. The above defined functions $f(x_1,\dots,x_k)$ and $g(x_1,\dots,x_l)$ can be rewritten by means of these new sets A_s^ε and B_t^ε. Namely, let $f(x_1,\dots,x_k) = c^\varepsilon(s_1,\dots,s_k)$ on the rectangles $A_{s_1}^\varepsilon \times\cdots\times A_{s_k}^\varepsilon$ with $1 \le s_j \le M(\varepsilon)$, $1 \le j \le k$, with different indices $s_1,\dots,s_k$, where $c^\varepsilon(s_1,\dots,s_k) = c(p_1,\dots,p_k)$ with those indices $(p_1,\dots,p_k)$ for which $A_{s_1}^\varepsilon \times\cdots\times A_{s_k}^\varepsilon \subset A_{p_1}\times\cdots\times A_{p_k}$. The function f disappears outside of these rectangles. The function $g(x_1,\dots,x_l)$ can be written similarly in the form $g(x_1,\dots,x_l) = d^\varepsilon(t_1,\dots,t_l)$ on the rectangles $B_{t_1}^\varepsilon\times\cdots\times B_{t_l}^\varepsilon$ with $1 \le t_{j'} \le M'(\varepsilon)$, $1 \le j' \le l$, and different indices, $t_1,\dots,t_l$. Beside this, the function g disappears outside of these rectangles.

The above representation of the functions f and g through a parameter ε is useful, since it enables us to give a good asymptotic formula for the product $k!Z_{\mu,k}(f)l!Z_{\mu,l}(g)$ which yields the diagram formula for the product of Wiener–Itô integrals of elementary functions with the help of a limiting procedure $\varepsilon \to 0$.

Fix a small number $\varepsilon > 0$, take the representation of the functions f and g with its help, and write

$$k!Z_{\mu,k}(f)l!Z_{\mu,l}(g) = \sum_{\gamma\in\Gamma(k,l)} Z_\gamma(f,g,\varepsilon) \tag{B.2}$$

with

$$\begin{aligned} Z_\gamma(f,g,\varepsilon) = \sum\nolimits^\gamma & c^\varepsilon(s_1,\dots,s_k)d^\varepsilon(t_1,\dots,t_l) \\ & \mu_W(A_{s_1}^\varepsilon)\dots\mu_W(A_{s_k}^\varepsilon)\mu_W(B_{t_1}^\varepsilon)\dots\mu_W(B_{t_l}^\varepsilon), \end{aligned} \tag{B.3}$$

where $\Gamma(k,l)$ denotes the class of diagrams introduced before the formulation of Theorem 10.2A, and $\sum^\gamma$ denotes summation for $k+l$-tuples $(s_1,\dots,s_k,t_1,\dots,t_l)$ such that $1 \le s_j \le M(\varepsilon)$, $1 \le j \le k$, $1 \le t_{j'} \le M'(\varepsilon)$, $1 \le j' \le l$, and $A_{s_j}^\varepsilon = B_{t_{j'}}^\varepsilon$ if $((1,j),(2,j')) \in E(\gamma)$, i.e. if it is an edge of γ, and otherwise all sets $A_{s_j}^\varepsilon$ and $B_{t_{j'}}^\varepsilon$ are disjoint. (This sum also depends on ε.) In the case of an empty sum $Z_\gamma(f,g,\varepsilon)$ equals zero.

We write the expression $Z_\gamma(f,g,\varepsilon)$ for all $\gamma \in \Gamma(k,l)$ in the form

$$Z_\gamma(f,g,\varepsilon) = Z_\gamma^{(1)}(f,g,\varepsilon) + Z_\gamma^{(2)}(f,g,\varepsilon), \quad \gamma \in \Gamma(k,l), \tag{B.4}$$

with

$$Z_\gamma^{(1)}(f,g,\varepsilon) = \sum\nolimits^\gamma c^\varepsilon(s_1,\dots,s_k)d^\varepsilon(t_1,\dots,t_l)$$
$$\prod_{j:(1,j)\in V_1(\gamma)} \mu_W(A^\varepsilon_{s_j}) \prod_{j':(2,j')\in V_2(\gamma)} \mu_W(B^\varepsilon_{t_{j'}})$$
$$\prod_{j':(2,j')\in\{(2,1),\dots,(2,l)\}\setminus V_2(\gamma)} \mu(B^\varepsilon_{t_{j'}}) \tag{B.5}$$

and

$$Z_\gamma^{(2)}(f,g,\varepsilon) = \sum\nolimits^\gamma c^\varepsilon(s_1,\dots,s_k)d^\varepsilon(t_1,\dots,t_l)$$
$$\prod_{j:(1,j)\in V_1(\gamma)} \mu_W(A^\varepsilon_{s_j}) \prod_{j':(2,j')\in V_2(\gamma)} \mu_W(B^\varepsilon_{t_{j'}})$$
$$\left[\prod_{j:(1,j)\in\{(1,1),\dots,(1,k)\}\setminus V_1(\gamma)} \mu_W(A^\varepsilon_{s_j})\right.$$
$$\prod_{j':(2,j')\in\{(2,1),\dots,(2,l)\}\setminus\in V_2(\gamma)} \mu_W(B^\varepsilon_{t_{j'}})$$
$$\left. - \prod_{j':(2,j')\in\{(2,1),\dots,(2,l)\}\setminus V_2(\gamma)} \mu(B^\varepsilon_{t_{j'}})\right], \tag{B.6}$$

where $V_1(\gamma)$ and $V_2(\gamma)$ (introduced before formula (10.9) during the preparation to the formulation of Theorem 10.2A) are the sets of vertices in the first and second row of the diagram γ from which no edge starts.

I claim that there is some constant $C > 0$ not depending on ε such that

$$E\left(|\gamma|!Z_{\mu,|\gamma|}(F_\gamma(f,g)) - Z_\gamma^{(1)}(f,g,\varepsilon)\right)^2 \le C\varepsilon \quad \text{for all } \gamma\in\Gamma(k,l) \tag{B.7}$$

with the Wiener–Itô integral with the kernel function $F_\gamma(f,g)$ defined in (10.9), (10.10) and (10.11), and

$$E\left(Z_\gamma^{(2)}(f,g,\varepsilon)\right)^2 \le C\varepsilon \quad \text{for all } \gamma\in\Gamma(k,l). \tag{B.8}$$

Relations (B.2), (B.4), (B.7) and (B.8) imply relation (10.14) if f and g are elementary functions. Indeed, (B.4), (B.7) and (B.8) imply that

$$\lim_{\varepsilon\to0} \left\| |\gamma|!Z_{\mu,|\gamma|}(F_\gamma(f,g)) - Z_\gamma(f,g,\varepsilon)\right\|_2 \to 0 \quad \text{for all } \gamma\in\Gamma(k,l),$$

and this relation together with (B.2) yield relation (10.14) with the help of a limiting procedure $\varepsilon\to0$.

To prove relation (B.7) let us introduce the function

$$\begin{aligned}
&F^{\varepsilon}_{\gamma}(f,g)(x_{(1,j)},x_{(2,j')},\ (1,j)\in V_1(\gamma),\ (2,j')\in V_2(\gamma))\\
&\quad = F_{\gamma}(f,g)(x_{(1,j)},x_{(2,j')},\ (1,j)\in V_1(\gamma),\ (2,j')\in V_2(\gamma))\\
&\qquad \text{if } x_{(1,j)}\in A^{\varepsilon}_{s_j},\ \text{for all } (1,j)\in V_1(\gamma),\\
&\qquad\quad x_{(2,j')}\in B^{\varepsilon}_{t_{j'}},\ \text{for all } (2,j')\in V_2(\gamma)),\quad \text{and}\\
&\qquad\qquad \text{all sets } A^{\varepsilon}_{s_j},\ (1,j)\in V_1(\gamma),\ \text{and } B^{\varepsilon}_{t_{j'}},\ (2,j')\in V_2(\gamma) \text{ are different.}
\end{aligned}$$

with the function $F_{\gamma}(f,g)$ defined in (10.10) and (10.11), and put

$$F^{\varepsilon}_{\gamma}(f,g)(x_{(1,j)},x_{(2,j')},\ (1,j)\in V_1(\gamma),\ (2,j')\in V_2(\gamma))=0\quad\text{otherwise.}$$

The function $F^{\varepsilon}_{\gamma}(f,g)$ is elementary, and a comparison of its definition with relation (B.5) and the definition of the function $F_{\gamma}(f,g)$ yields that

$$Z^{(1)}_{\gamma}(f,g,\varepsilon)=|\gamma|!Z_{\mu,|\gamma|}(F^{\varepsilon}_{\gamma}(f,g)). \tag{B.9}$$

The function $F^{\varepsilon}_{\gamma}(f,g)$ slightly differs from $F_{\gamma}(f,g)$, since the function $F_{\gamma}(f,g)$ may not disappear in such points $(x_{(1,j)},x_{(2,j')},\ (1,j)\in V_1(\gamma),\ (2,j')\in V_2(\gamma))$ for which there is some pair (j,j') with the property $x_{(1,j)}\in A^{\varepsilon}_{s_j}$ and $x_{(2,j')}\in B^{\varepsilon}_{t_{j'}}$ with some sets $A^{\varepsilon}_{s_j}$ and $B^{\varepsilon}_{t_{j'}}$ such that $A^{\varepsilon}_{s_j}=B^{\varepsilon}_{t_{j'}}$, while $F^{\varepsilon}_{\gamma}(f,g)$ must be zero in such points. On the other hand, in the case $|\gamma|=\max(k,l)-\min(k,l)$, i.e. if one of the sets $V_1(\gamma)$ or $V_2(\gamma)$ is empty, $F_{\gamma}(f,g)=F^{\varepsilon}_{\gamma}(f,g)$, $Z^{(1)}_{\gamma}(f,g,\varepsilon)=|\gamma|!Z_{\mu,|\gamma|}(F_{\gamma}(f,g))$, and relation (B.7) clearly holds for such diagrams γ.

In the case $|\gamma|=\max(k,l)-\min(k,l)>0$ we prove a good estimate on the measure of the set where $F_{\gamma}\neq F^{\varepsilon}_{\gamma}$ with respect to an appropriate power of the measure μ. Relation (B.7) will be proved with the help of this estimate and formula (B.9).

Let us define the sets $A=\bigcup\limits_{s=1}^{M(\varepsilon)}A^{\varepsilon}_s$ and $B=\bigcup\limits_{t=1}^{M'(\varepsilon)}B^{\varepsilon}_t$. These sets A and B do not depend on the parameter ε. Beside this $\mu(A)<\infty$, and $\mu(B)<\infty$. Define for all pairs (j_0,j_0') such that $(1,j_0)\in V_1(\gamma)$, $(2,j_0')\in V_2(\gamma)$ the set

$$\begin{aligned}
D(j_0,j_0')=\{&(x_{(1,j)},x_{(2,j')},\ (1,j)\in V_1(\gamma),\ (2,j')\in V_2(\gamma)):\\
&x_{(1,j_0)}\in A^{\varepsilon}_s, x_{(2,j_0')}\in B^{\varepsilon}_t \text{ with some } 1\le s\le M(\varepsilon) \text{ and } 1\le t\le M'(\varepsilon)\\
&\text{such that } A^{\varepsilon}_s=B^{\varepsilon}_t,\quad\text{and}\quad x_{(1,j)}\in A \text{ for all } (1,j)\in V_1(\gamma),\\
&\text{and } x_{(2,j')}\in B \text{ for all } (2,j')\in V_2(\gamma)\}.
\end{aligned}$$

Introduce the notation $x^{\gamma}=(x_{(1,j)},x_{(2,j')}),\ (1,j)\in V_1(\gamma),\ (2,j')\in V_2(\gamma))$, and consider only such vectors x^{γ} whose coordinates satisfy the conditions $x_{(1,j)}\in A$

for all $(1, j) \in V_1(\gamma)$ and $x_{(2,j')} \in B$ for all $(2, j') \in V_2(\gamma)$. Put

$$D_\gamma = \{x^\gamma\colon F^\varepsilon_\gamma(f,g)(x^\gamma) \neq F_\gamma(f,g)(x^\gamma)\}.$$

The relation $D_\gamma \subset \bigcup\limits_{j=1}^{k}\bigcup\limits_{j'=1}^{l} D(j_0, j_0')$ holds, since if $F^\varepsilon_\gamma(f,g)(x^\gamma) \neq F_\gamma(f,g)(x^\gamma)$ for some vector x^γ, then it has some coordinates $(1, j_0) \in V_1(\gamma)$ and $(2, j_0') \in V_2(\gamma)$ such that $x_{(1,j_0)} \in A^\varepsilon_s$ and $x_{(1,j_0')} \in B^\varepsilon_t$ with some sets $A^\varepsilon_s = B^\varepsilon_t$, and the relation in the last line of the definition of $D(j_0, j_0')$ must also hold for such a vector x^γ, since otherwise $F_\gamma(f,g)(x_\gamma) = 0 = F^\varepsilon_\gamma(f,g)(x_\gamma)$.

I claim that there is some constant C_1 such that

$$\mu^{|V_1(\gamma)|+|V_2(\gamma)|}(D(j_0, j_0')) \le C_1\varepsilon \quad \text{for all sets } D(j_0, j_0'),$$

where $\mu^{|V_1(\gamma)|+|V_2(\gamma)|}$ denotes the direct product of the measure μ on some copies of the original space $(X, \mathscr{X})$ indexed by $(1, j) \in V_1(\gamma)$ and $(2, j') \in V_2(\gamma)$. To see this relation one has to observe that $\sum\limits_{A^\varepsilon_s = B^\varepsilon_t} \mu(A^\varepsilon_s)\mu(B^\varepsilon_t) \le \sum \varepsilon\mu(A^\varepsilon_s) = \varepsilon\mu(A)$. Thus the set $D(j_0, j_0')$ can be covered by the direct product of a set whose μ measure is not greater than $\varepsilon\mu(A)$ and of a rectangle whose edges are either the set A or the set B.

The above relations imply that

$$\mu^{|V_1(\gamma)|+|V_2(\gamma)|}(D_\gamma) \le C_2\varepsilon \tag{B.10}$$

with some constant $C_2 > 0$.

Relation (B.9), estimate (B.10), the property (c) formulated in Theorem 10.1 for Wiener–Itô integrals and the observation that the function $F_\gamma(f,g)$ is bounded in supremum norm if f and g are elementary functions imply the inequality

$$\begin{aligned}
&E\left(|\gamma|!Z_{\mu,|\gamma|}(F_\gamma(f,g)) - Z^{(1)}_\gamma(f,g,\varepsilon)\right)^2 \\
&\quad = |\gamma!|^2 E\left(Z_{\mu,|\gamma|}(F_\gamma(f,g) - F^\varepsilon_\gamma(f,g))\right)^2 \le |\gamma|!\|F_\gamma(f,g) - F^\varepsilon_\gamma(f,g)\|_2^2 \\
&\quad \le K\mu^{|V_1(\gamma)|+|V_2(\gamma)|}(D_\gamma) \le C\varepsilon.
\end{aligned}$$

Hence relation (B.7) holds.

To prove relation (B.8) we rewrite $E\left(Z^{(2)}_\gamma(f,g,\varepsilon)\right)^2$ in the following form:

$$\begin{aligned}
E\left(Z^{(2)}_\gamma(f,g,\varepsilon)\right)^2 = {\sum}^{\gamma}{\sum}^{\gamma} & c^\varepsilon(s_1,\dots,s_k)d^\varepsilon(t_1,\dots,t_l)c^\varepsilon(\bar s_1,\dots,\bar s_k) \\
& d^\varepsilon(\bar t_1,\dots,\bar t_l)EU(s_1,\dots,s_k,t_1,\dots,t_l,\bar s_1,\dots,\bar s_k,\bar t_1,\dots,\bar t_l)
\end{aligned} \tag{B.11}$$

with

$$
\begin{aligned}
&U(s_1,\dots,s_k,t_1,\dots,t_l,\bar s_1,\dots,\bar s_k,\bar t_1,\dots,\bar t_l)\\
&\quad=\prod_{j\colon(1,j)\in V_1(\gamma)}\mu_W(A^{\varepsilon}_{s_j})\prod_{j'\colon(2,j')\in V_2(\gamma)}\mu_W(B^{\varepsilon}_{t_{j'}})\\
&\qquad\prod_{\bar j\colon(1,\bar j)\in V_1(\gamma)}\mu_W(A^{\varepsilon}_{\bar s_{\bar j}})\prod_{\bar j'\colon(2,\bar j')\in V_2(\gamma)}\mu_W(B^{\varepsilon}_{\bar t_{\bar j'}})\\
&\qquad\Bigg[\prod_{j\colon(1,j)\in\{(1,1),\dots,(1,k)\}\setminus V_1(\gamma)}\mu_W(A^{\varepsilon}_{s_j})\prod_{j'\colon(2,j')\in\{(2,1),\dots,(2,l)\}\setminus\in V_2(\gamma)}\mu_W(B^{\varepsilon}_{t_{j'}})\\
&\qquad-\prod_{j'\colon(2,j')\in\{(2,1),\dots,(2,l)\}\setminus V_2(\gamma)}\mu(B^{\varepsilon}_{t_{j'}})\Bigg]\\
&\qquad\Bigg[\prod_{\bar j\colon(1,\bar j)\in\{(1,1),\dots,(1,k)\}\setminus V_1(\gamma)}\mu_W(A^{\varepsilon}_{\bar s_{\bar j}})\prod_{\bar j'\colon(2,\bar j')\in\{(2,1),\dots,(2,l)\}\setminus\in V_2(\gamma)}\mu_W(B^{\varepsilon}_{\bar t_{\bar j'}})\\
&\qquad-\prod_{\bar j'\colon(2,\bar j')\in\{(2,1),\dots,(2,l)\}\setminus V_2(\gamma)}\mu(B^{\varepsilon}_{\bar t_{\bar j'}})\Bigg].
\end{aligned}\tag{B.12}
$$

The double sum $\sum^{\gamma}\sum^{\gamma}$ in (B.11) has to be understood in the following way. The first summation is taken for vectors $(s_1,\dots,s_k,t_1,\dots,t_l)$, and $\sum^{\gamma}$ is defined in the same way as in formula (B.3). The second summation is taken for vectors $(\bar s_1,\dots,\bar s_k,\bar t_1,\dots,\bar t_l)$, and again the summation $\sum^{\gamma}$ is taken as in (B.3), only here $\bar s_j$ plays the role of s_j and $\bar t_{j'}$ plays the role of $t_{j'}$.

Relation (B.8) will be proved by means of some estimates about the expectation of the above defined random variable $U(\cdot)$ which will be presented in the following Lemma B. To formulate this result I introduce the following Properties A and B.

Property A. *A sequence $s_1,\dots,s_k,t_1,\dots,t_l,\bar s_1,\dots,\bar s_k,\bar t_1,\dots,\bar t_l$, with elements $1\le s_j,\bar s_{\bar j}\le M(\varepsilon)$, for $1\le j,\bar j\le k$, and $1\le t_j,\bar t_{\bar j'}\le M'(\varepsilon)$ for $1\le j',\bar j'\le l$, satisfies Property A (depending on a fixed diagram γ and number $\varepsilon>0$) if the sequence of sets $A^{\varepsilon}_{s_j}$, $(1,j)\in V_1(\gamma)$, $B^{\varepsilon}_{t_{j'}}$, $(2,j')\in V_2(\gamma)$, and the sequence of sets $A^{\varepsilon}_{\bar s_{\bar j}}$, $(1,\bar j)\in V_1(\gamma)$, $B^{\varepsilon}_{\bar t_{\bar j'}}$, $(2,\bar j')\in V_2(\gamma)$, agree. (Here we say that two sequences agree if they contain the same elements in a possibly different order.)*

Property B. *A sequence $s_1,\dots,s_k,t_1,\dots,t_l,\bar s_1,\dots,\bar s_k,\bar t_1,\dots,\bar t_l$, with elements $1\le s_j,\bar s_{\bar j}\le M(\varepsilon)$, for $1\le j,\bar j\le k$, and $1\le t_j,\bar t_{\bar j'}\le M'(\varepsilon)$ for $1\le j',\bar j'\le l$, satisfies Property B (depending on a fixed diagram γ and number $\varepsilon>0$) if the sequences of sets*

$$A^{\varepsilon}_{s_j},\ (1,j)\in\{(1,1),\dots,(1,k)\}\setminus V_1(\gamma),\quad B^{\varepsilon}_{t_{j'}},(2,j')\in\{(2,1),\dots,(2,l)\}\setminus V_2(\gamma),$$

and

$$A^{\varepsilon}_{s_{\bar{j}}}, (1, \bar{j}) \in \{(1,1),\dots,(1,k)\} \setminus V_1(\gamma), \quad B^{\varepsilon}_{t_{\bar{j}'}}, (2, \bar{j}') \in \{(2,1),\dots,(2,l)\} \setminus V_2(\gamma),$$

have at least one common element.

(In the above definitions two sets A^{ε}_s and B^{ε}_t are identified if $A^{\varepsilon}_s = B^{\varepsilon}_t$.)

Now I formulate the following

Lemma B. *Let us consider the function $U(\cdot)$ introduced in formula (B.12). Assume that its arguments $s_1,\dots,s_k,t_1,\dots,t_l,\bar{s}_1,\dots,\bar{s}_k,\bar{t}_1,\dots,\bar{t}_l$ are chosen in such a way that the function $U(\cdot)$ with these arguments appears in the double sum $\sum^{\gamma}\sum^{\gamma}$ in formula (B.11), i.e. $A^{\varepsilon}_{s_j} = B^{\varepsilon}_{t_{j'}}$ if $((1,j),(2,j')) \in E(\gamma)$, otherwise all sets $A^{\varepsilon}_{s_j}$ and $B^{\varepsilon}_{t_{j'}}$ are disjoint, and an analogous statement holds if the coordinates $s_1,\dots,s_k,t_1,\dots,t_l$ are replaced by $\bar{s}_1,\dots,\bar{s}_k$ and $\bar{t}_1,\dots,\bar{t}_l$.*

If the sequence of the arguments in $U(\cdot)$ does not satisfies either Property A or Property B, then

$$EU(s_1,\dots,s_k,t_1,\dots,t_l,\bar{s}_1,\dots,\bar{s}_k,\bar{t}_1,\dots,\bar{t}_l) = 0. \tag{B.13}$$

If the sequence of the arguments in $U(\cdot)$ satisfies both Property A and Property B, then

$$|EU(s_1,\dots,s_k,t_1,\dots,t_l,\bar{s}_1,\dots,\bar{s}_k,\bar{t}_1,\dots,\bar{t}_l)| \le C\varepsilon {\prod}' \mu(A^{\varepsilon}_{\bar{s}_{\bar{j}}})\mu(B^{\varepsilon}_{\bar{t}_{\bar{j}'}}) \tag{B.14}$$

with some appropriate constant $C = C(k,l) > 0$ depending only on the number of variables k and l of the functions f and g. The prime in the product $\prod'$ at the right-hand side of (B.14) means that in this product the measure μ of those sets $A^{\varepsilon}_{\bar{s}_{\bar{j}}}$ and $B^{\varepsilon}_{\bar{t}_{\bar{j}'}}$ are considered, whose indices are listed among the arguments $\bar{s}_{\bar{j}}$ or $\bar{t}_{\bar{j}'}$ of $U(\cdot)$, and the measure μ of each such set appears exactly once. (This means that if $A^{\varepsilon}_{\bar{s}_{\bar{j}}} = B^{\varepsilon}_{\bar{t}_{\bar{j}'}}$ then one of the terms between $\mu(A^{\varepsilon}_{\bar{s}_{\bar{j}}})$ and $\mu(B^{\varepsilon}_{\bar{t}_{\bar{j}'}})$ is omitted from the product. For the sake of definitiveness let us preserve the set $\mu(A^{\varepsilon}_{\bar{s}_{\bar{j}}})$ in such a case.)

Remark. The content of Lemma B is that most terms in the double sum in formula (B.11) equal zero, and even the non-zero terms are small.

The proof of Lemma B. Let us prove first relation (B.13) in the case when Property A does not hold. It will be exploited that for disjoint sets the random variables $\mu_W(A_s)$ and $\mu_W(B_t)$ are independent, and this provides a good factorization of the expectation of certain products.

Let us carry out the multiplications in the expression $U(\cdot)$ defined (B.12). We get a sum consisting of four terms. We show that each of them has zero expectation. Indeed, if a sequence $s_1,\dots,s_k,t_1,\dots,t_l,\bar{s}_1,\dots,\bar{s}_k,\bar{t}_1,\dots,\bar{t}_l$ does not satisfy Property A, but it satisfies the remaining conditions of Lemma B, then each term in the sum expressing $U(\cdots)$ with these arguments is a product which contains a factor $\mu_W(A^{\varepsilon}_{s_{j_0}})$, $(1, j_0) \in V_1(\gamma)$ with the following property. It is independent of all those terms in this product which are in the following list: $\mu_W(A^{\varepsilon}_{s_j})$ with some

$j \neq j_0$, $1 \le j \le k$, or $\mu_W(B^\varepsilon_{t_{j'}})$, $1 \le j \le l$, or $\mu_W(A^\varepsilon_{s_{\bar j}})$ with $(1,\bar j) \in V_1(\gamma)$, or $\mu_W(B^\varepsilon_{t_{\bar j'}})$ with $(2,\bar j') \in V_2(\gamma)$. We will show with the help of this property that the expectation of the terms we consider can be written in the form of a product either with a factor of the form $E\mu_W(A^\varepsilon_{s_{j_0}}) = 0$ or with a factor of the form $E\mu_W(A^\varepsilon_{s_{j_0}})^3 = 0$. Hence this expectation equals zero.

Indeed, although the above properties do not exclude the existence of a set $A^\varepsilon_{t_{\bar j'}}$, $(1,\bar j') \in \{(1,1),\dots,(1,k) \setminus V_1(\gamma)$ or $B^\varepsilon_{t_{\bar j'}}$, $(2,\bar j') \in \{(2,1),\dots,(2,l)\} \setminus V_2(\gamma)$ such that $\mu_W(A^\varepsilon_{t_{\bar j'}})$ or $\mu_W(B^\varepsilon_{t_{\bar j'}})$, is not independent of $\mu_W(A^\varepsilon_{s_{j_0}})$, but this can only happen if $A^\varepsilon_{t_{\bar j}} = B^\varepsilon_{t_{\bar j'}} = A^\varepsilon_{s_{j_0}}$. This implies that in such a case when our term does not contain a factor of the form $E\mu_W(A^\varepsilon_{s_{j_0}})$, then it contains a factor of the form $E\mu_W(A^\varepsilon_{s_{j_0}})^3 = 0$. Hence $EU(\cdot) = 0$ if the arguments of $U(\cdot)$ do not satisfy Property A.

To finish the proof of relation (B.13) it is enough consider the case when the arguments of $U(\cdot)$ satisfy Property A, but they do not satisfy Property B. The validity of Property A implies that the sets $\{A^\varepsilon_{s_j},\ j \in V_1(\gamma)\} \cup \{B^\varepsilon_{t_{j'}},\ j' \in V_2(\gamma)\}$ and $\{A^\varepsilon_{\bar s_j},\ j \in V_1(\gamma)\} \cup \{B^\varepsilon_{\bar t_{j'}},\ j' \in V_2(\gamma)\}$ agree. The conditions of Lemma B also imply that the elements of these sets are disjoint of the sets $A^\varepsilon_{s_j}$, $B^\varepsilon_{t_{j'}}$, $A^\varepsilon_{\bar s_{\bar j}}$ and $B^\varepsilon_{\bar t_{\bar j'}}$ with indices $(1,j),(1,\bar j) \in \{(1,1),\dots,(1,k)\} \setminus V_1(\gamma)$ and $(2,j'),(2,\bar j') \in \{(2,1),\dots,(2,l)\} \setminus V_2(\gamma)$. If Property B does not hold, then we can divide the latter class of sets into two disjoint subclasses in an appropriate way. The first subclass consists of the sets $A^\varepsilon_{s_j}$ and $B^\varepsilon_{t_{j'}}$, and the second one of the sets $A^\varepsilon_{\bar s_{\bar j}}$ and $B^\varepsilon_{\bar t_{\bar j'}}$ with indices such that $(1,j),(1,\bar j) \in \{(1,1),\dots,(1,k)\} \setminus V_1(\gamma)$ and $(2,j'),(2,\bar j') \in \{(2,1),\dots,(2,l)\} \setminus V_2(\gamma)$. These facts imply that $EU(\cdot)$ has a factorization, which contains the term

$$E\Bigg[\prod_{j\colon (1,j)\in\{(1,1),\dots,(1,k)\}\setminus V_1(\gamma)} \mu_W(A^\varepsilon_{s_j}) \prod_{j'\colon (2,j')\in\{(2,1),\dots,(2,l)\}\setminus\in V_2(\gamma)} \mu_W(B^\varepsilon_{t_{j'}}) - \prod_{j'\colon (2,j')\in\{(2,1),\dots,(2,l)\}\setminus V_2(\gamma)} \mu(B^\varepsilon_{s_{j'}})\Bigg] = 0,$$

hence relation (B.13) holds also in this case. The last expression has zero expectation, since if we take such pairs $A^\varepsilon_{s_j}, B^\varepsilon_{t'_j}$ for the sets appearing in it for which that $((1,j),(2,j')) \in E(\gamma)$, i.e. these vertices are connected with an edge of γ, then $A^\varepsilon_{s_j} = B^\varepsilon_{t'_j}$ in a pair, and elements in different pairs are disjoint. This observation allows a factorization in the product whose expectation is taken, and then the identity $E\mu_W(A^\varepsilon_{s_j})\mu_W(B^\varepsilon_{t_{j'}}) = \mu(A^\varepsilon_{s_j})$ implies the desired identity.

To prove relation (B.14) if the arguments of the function $U(\cdot)$ satisfy both Properties A and B consider the expression (B.12) which defines $U(\cdot)$, carry out the term by term multiplication between the two differences at the end of this formula, take expectation for each term of the sum obtained in such a way and factorize them.

Since $E\mu_W(A)^2 = \mu(A)$, $E\mu_W(A)^4 = 3\mu(A)^2$ for all sets $A \in \mathscr{X}$, $\mu(A) < \infty$, some calculation shows that each term can be expressed as constant times a product whose elements are those probabilities $\mu(A^\varepsilon_{\bar{s}_j})$ and $\mu(B^\varepsilon_{\bar{t}_{\bar{j}'}})$ or their square which appear at the right-hand side of (B.14). Moreover, since the arguments of $U(\cdot)$ satisfy Property B, there will be at least one term of the form $\mu(A^\varepsilon_s)^2$ in this product. Since $\mu(A^\varepsilon_s)^2 \le \varepsilon\mu(A^\varepsilon_s)$, these calculations provide formula (B.14). Lemma B is proved. □

Relation (B.11) implies that

$$E\left(Z^{(2)}_\gamma(f, g, \varepsilon)\right)^2 \le K \sum{}^{\gamma} \sum{}^{\gamma} |EU(s_1, \dots, s_k, t_1, \dots, t_l, \bar{s}_1, \dots, \bar{s}_k, \bar{t}_1, \dots, \bar{t}_l)| \tag{B.15}$$

with some appropriate $K > 0$. By Lemma B it is enough to sum up only for such terms $U(\cdot)$ in (B.15) whose arguments satisfy both Properties A and B. Moreover, each such term can be bounded by means of inequality (B.14). Let us write up the upper bound we get on $E\left(Z^{(2)}_\gamma(f, g, \varepsilon)\right)^2$ in such a way. We get a sum consisting of terms of the form $\mu(A^\varepsilon_{s_1}) \cdots \mu(A^\varepsilon_{s_p})\mu(B^\varepsilon_{t_1}) \cdots \mu(B^\varepsilon_{\bar{t}_q})$ multiplied by constant times ε. The sets A^ε_s and B^ε_t whose measure μ appears in such a term are disjoint. Beside this $1 \le p \le k$, and $1 \le q \le l$.

In the above indicated estimation of $E\left(Z^{(2)}_\gamma(f, g, \varepsilon)\right)^2$ with the help of formula (B.15) and Lemma B we have exploited the following fact. A term

$$\mu(A^\varepsilon_{s_1}) \cdots \mu(A^\varepsilon_{s_p})\mu(B^\varepsilon_{t_1}) \cdots \mu(B^\varepsilon_{\bar{t}_q})$$

with prescribed indices $s_1, \dots, s_p$ and $t_1, \dots, t_q$ came up in the sum at the right-hand of our bound as a contribution of only finitely many expressions $|EU(\cdots)|$. Hence we get this term in the upper bound with a multiplying coefficient bounded by constant times ε.

We also have $\sum\limits_{s=1}^{M(\varepsilon)} \mu(A^\varepsilon_s) + \sum\limits_{t=1}^{M'(\varepsilon)} \mu(B^\varepsilon_t) = \mu(A) + \mu(B) < \infty$. The above relations imply that

$$\begin{aligned} E\left(Z^{(2)}_\gamma(f, g, \varepsilon)\right)^2 &\le C_1\varepsilon \sum_{\substack{1\le p\le k\\ 1\le q\le l}} \sum_{\substack{1\le s_l\le M\\ 1\le l\le p}} \sum_{\substack{1\le t_l\le M'\\ 1\le l\le q}} \mu(A^\varepsilon_{s_1}) \cdots \mu(A^\varepsilon_{s_p})\mu(B^\varepsilon_{t_1}) \cdots \mu(B^\varepsilon_{\bar{t}_q}) \\ &\le C_2\varepsilon \sum_{j=1}^{(k+l)} (\mu(A) + \mu(B))^j \le C\varepsilon. \end{aligned}$$

Hence relation (B.8) holds.

To prove Theorem 10.2A in the general case take for all pairs of functions $f \in \mathscr{H}_{\mu,k}$ and $g \in \mathscr{H}_{\mu,l}$ two sequences of elementary functions $f_n \in \bar{\mathscr{H}}_{\mu,k}$ and $g_n \in$

$\bar{\mathscr{H}}_{\mu,l}$, $n = 1, 2, \ldots$, such that $\|f_n - f\|_2 \to 0$ and $\|g_n - g\|_2 \to 0$ as $n \to \infty$. It is enough to show that

$$E|k!Z_{\mu,k}(f)l!Z_{\mu,l}(g) - k!Z_{\mu,k}(f_n)l!Z_{\mu,l}(g_n)| \to 0 \quad \text{as } n \to \infty, \tag{B.16}$$

and

$$|\gamma|!E\left|Z_{\mu,|\gamma|}(F_\gamma(f,g)) - Z_{\mu,|\gamma|}(F_\gamma(f_n,g_n))\right| \to 0 \text{ as } n \to \infty \quad \text{for all } \gamma \in \Gamma(k,l), \tag{B.17}$$

since then a simple limiting procedure $n \to \infty$, and the already proved part of the theorem for Wiener–Itô integrals of elementary functions imply Theorem 10.2A.

To prove relation (B.16) write with the help of Property c) in Theorem (10.1)

$$\begin{aligned}
&E|k!Z,\mu,k(f)l!Z_{\mu,l}(g) - k!Z_{\mu,k}(f_n)l!Z_{\mu,l}(g_n)| \\
&\quad \le k!l!\left(E|Z_{\mu,k}(f)Z_{\mu,l}(g-g_n)| + E|Z_{\mu,k}(f-f_n)Z_{\mu,l}(g_n)\right)| \\
&\quad \le k!l!\left(\left(EZ^2_{\mu,k}(f)\right)^{1/2}\left(EZ^2_{\mu,l}(g-g_n)\right)^{1/2}\right. \\
&\qquad \left. + \left(EZ^2_{\mu,k}(f-f_n)\right)^{1/2}\left(EZ^2_{\mu,l}(g_n)\right)^{1/2}\right) \\
&\quad \le (k!l!)^{1/2}\left(\|f\|_2\|g-g_n\|_2 + \|f-f_n\|_2\|g_n\|_2\right).
\end{aligned}$$

Relation (B.16) follows from this inequality with a limiting procedure $n \to \infty$.

To prove relation (B.17) write

$$\begin{aligned}
&|\gamma|!E\left|Z_{\mu,|\gamma|}(F_\gamma(f,g)) - Z_{\mu,|\gamma|}(F_\gamma(f_n,g_n))\right| \\
&\quad \le |\gamma|!E\left|Z_{\mu,|\gamma|}(F_\gamma(f,g-g_n))\right| + |\gamma|!E\left|Z_{\mu,|\gamma|}(F_\gamma(f-f_n,g_n))\right| \\
&\quad \le |\gamma|!\left(EZ^2_{\mu,|\gamma|}(F_\gamma(f,g-g_n))\right)^{1/2} + |\gamma|!\left(EZ^2_{\mu,|\gamma|}(F_\gamma(f-f_n,g_n))\right)^{1/2} \\
&\quad \le (|\gamma|!)^{1/2}\left(\|F_\gamma(f,g-g_n)\|_2 + \|F_\gamma(f-f_n,g_n)\|_2\right),
\end{aligned}$$

and observe that by relation (10.13) $\|F_\gamma(f,g-g_n)\|_2 \le \|f\|_2\|g-g_n\|_2$, and $\|F_\gamma(f-f_n,g_n)\|_2 \le \|f-f_n\|_2\|g_n\|_2$. Hence

$$\begin{aligned}
&|\gamma|!E\left|Z_{\mu,|\gamma|}(F_\gamma(f,g)) - Z_{\mu,|\gamma|}(F_\gamma(f_n,g_n))\right| \\
&\quad \le (|\gamma|!)^{1/2}\left(\|f\|_2\|g-g_n\|_2 + \|f-f_n\|_2\|g_n\|_2\right).
\end{aligned}$$

The last inequality implies relation (B.17) with a limiting procedure $n \to \infty$. Theorem 10.2A is proved. □

Appendix C
The Proof of Some Results About Wiener–Itô Integrals

First I prove Itô's formula about multiple Wiener–Itô integrals (Theorem 10.3). The proof is based on the diagram formula for Wiener–Itô integrals and a recursive formula about Hermite polynomials proved in Proposition C. In Proposition C2 I present the proof of another important property of Hermite polynomials. This result states that the class of all Hermite polynomials is a *complete* orthogonal system in an appropriate Hilbert space. It is needed in the proof of Theorem 10.5 which provides an isomorphism between a Fock space and the Hilbert space generated by Wiener–Itô integrals with respect to a white noise with an appropriate reference measure. At the end of Appendix C the proof of Theorem 10.4, a limit theorem about degenerate U-statistics is given together with a version of this result about the limit behaviour of multiple integrals with respect to a normalized empirical distribution.

Proposition C About Some Properties of Hermite Polynomials. *The functions*

$$H_k(x) = (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}, \quad k = 0, 1, 2, \dots \tag{C.1}$$

are the Hermite polynomials with leading coefficient 1, i.e. $H_k(x)$ is a polynomial of order k with leading coefficient 1 such that

$$\int_{-\infty}^{\infty} H_k(x) H_l(x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 0 \quad \text{if } k \neq l. \tag{C.2}$$

Beside this,

$$\int_{-\infty}^{\infty} H_k^2(x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = k! \quad \text{for all } k = 0, 1, 2 \dots. \tag{C.3}$$

The recursive relation

$$H_k(x) = x H_{k-1}(x) - (k-1) H_{k-2}(x) \tag{C.4}$$

holds for all $k = 1, 2, \dots$.

Remark. It is more convenient to consider relation (C.4) valid also in the case $k=1$. In this case $H_1(x) = x$, $H_0(x) = 1$, and relation holds with an arbitrary function $H_{-1}(x)$.

Proof of Proposition C. It is clear from formula (C.1) that $H_k(x)$ is a polynomial of order k with leading coefficient 1. Take $l \ge k$, and write by means of integration by parts

$$\int_{-\infty}^{\infty} H_k(x)H_l(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} H_k(x)(-1)^l \frac{d^l}{dx^l}e^{-x^2/2}\,dx$$
$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{d}{dx} H_k(x)(-1)^{l-1} \frac{d^{l-1}}{dx^{l-1}}e^{-x^2/2}\,dx.$$

Successive partial integration together with the identity $\frac{d^k}{dx^k}H_k(x) = k!$ yield that

$$\int_{-\infty}^{\infty} H_k(x)H_l(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx = k!\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}(-1)^{l-k}\frac{d^{l-k}}{dx^{l-k}}e^{-x^2/2}\,dx.$$

The last relation supplies formulas (C.2) and (C.3).

To prove relation (C.4) observe that $H_k(x) - xH_{k-1}(x)$ is a polynomial of order $k-2$. (The term x^{k-1} is missing from this expression. Indeed, if k is an even number, then the polynomial $H_k(x) - xH_{k-1}(x)$ is an even function, and it does not contain the term x^{k-1} with an odd exponent $k-1$. Similar argument holds if the number k is odd.) Beside this, it is orthogonal (with respect to the standard normal distribution) to all Hermite polynomials $H_l(x)$ with $0 \le l \le k-3$. Hence $H_k(x) - xH_{k-1}(x) = CH_{k-2}(x)$ with some constant C to be determined.

Multiply both sides of the last identity with $H_{k-2}(x)$ and integrate them with respect to the standard normal distribution. Apply the orthogonality of the polynomials $H_k(x)$ and $H_{k-2}(x)$, and observe that the identity

$$\int H_{k-1}(x)xH_{k-2}(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx = \int H_{k-1}^2(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx = (k-1)!$$

holds. (In this calculation we have exploited that $H_{k-1}(x)$ is orthogonal to $H_{k-1}(x) - xH_{k-2}(x)$, because the order of the latter polynomial is less than $k-1$.) In such a way we get the identity $-(k-1)! = C(k-2)!$ for the constant C in the last identity, i.e. $C = -(k-1)$, and this implies relation (C.4). □

Proof of Itô's formula for multiple Wiener–Itô integrals. Let $K = \sum\limits_{p=1}^{m} k_p$, the sum of the order of the Hermite polynomials, denote the order of the expression in relation (10.24). Formula (10.24) clearly holds for expressions of order $K = 1$. It will be proved in the general case by means of induction with respect to the order K.

In the proof the functions $f(x_1)=\varphi_1(x_1)$ and

$$g(x_1,\dots,x_{K_m-1})=\prod_{j=1}^{K_1-1}\varphi_1(x_j)\cdot\prod_{p=2}^{m}\prod_{j=K_{p-1}}^{K_p-1}\varphi_p(x_j),$$

will be introduced and the product $Z_{\mu,1}(f)(K_m-1)!Z_{\mu,K_m-1}(g)$ will be calculated by means of the diagram formula. (The same notation is applied as in Theorem 10.3. In particular, $K=K_m$, and in the case $K_1=1$ the convention $\prod\limits_{j=1}^{K_1-1}\varphi_1(x_j)=1$ is applied.) In the application of the diagram formula diagrams with two rows appear. The first row of these diagrams contains the vertex $(1,1)$ and the second row contains the vertices $(2,1),\dots,(2,K_m-1)$. It is useful to divide the diagrams to three disjoint classes. The first class, Γ_0 contains only the diagram γ_0 without any edges. The second class Γ_1 consists of those diagrams which have an edge of the form $((1,1),(2,j))$ with some $1\le j\le k_1-1$, and the third class Γ_2 is the set of those diagrams which have an edge of the form $((1,1),(2,j))$ with some $k_1\le j\le K_m-1$. Because of the orthogonality of the functions φ_s for different indices s $F_\gamma\equiv 0$ and $Z_{\mu,K_m-2}(F_\gamma)=0$ for $\gamma\in\Gamma_2$. The class Γ_1 contains k_1-1 diagrams. Let us consider a diagram γ from this class with an edge $((1,1),(2,j_0))$, $1\le j\le k_1-1$. We have for such a diagram $F_\gamma=\prod\limits_{j\in\{1,\dots,K_1-1\}\setminus\{j_0\}}\varphi_1(x_{(2,j)})\prod\limits_{p=2}^{m}\prod\limits_{j=K_{p-1}}^{K_p-1}\varphi_p(x_{(2,j)})$, and by our inductive hypothesis $(K_m-2)!Z_{\mu,K_m-2}(F_\gamma)=H_{k_1-2}(\eta_1)\prod\limits_{p=2}^{m}H_{k_p}(\eta_p)$. Finally

$$K_m!Z_{\mu,K_m}(F_{\gamma_0})=K_m!Z_{\mu,K_m}\left(\prod_{p=1}^{m}\left(\prod_{j=K_{p-1}+1}^{K_p}\varphi_p(x_j)\right)\right)$$

for the diagram $\gamma_0\in\Gamma_0$ without any edge.

Our inductive hypothesis also implies the following identity for the expression we wanted to calculate with the help of the diagram formula.

$$Z_{\mu,1}(f)(K_m-1)!Z_{\mu,K_m-1}(g)=\eta_1H_{k_1-1}(\eta_1)\prod_{p=2}^{m}H_{k_p}(\eta_p).$$

The above calculations together with the observation $|\Gamma_1|=k_1-1$ yield the identity

$$
\begin{aligned}
&K_m!Z_{\mu,K_m}\left(\prod_{p=1}^{m}\left(\prod_{j=K_{p-1}+1}^{K_p}\varphi_p(x_j)\right)\right)\\
&\quad = K_m!Z_{\mu,K_m}(F_{\gamma_0})\\
&\quad = Z_{\mu,1}(f)(K_m-1)!Z_{\mu,K_m-1}(g)-\sum_{\gamma\in\Gamma_1}(K_m-2)!Z_{\mu,K_m-2}(F_\gamma)\\
&\quad = \eta_1 H_{k_1-1}(\eta_1)\prod_{p=2}^{m}H_{k_p}(\eta_p)-(k_1-1)H_{k_1-2}(\eta_1)\prod_{p=2}^{m}H_{k_p}(\eta_p)\\
&\quad = [\eta_1 H_{k_1-1}(\eta_1)-(k_1-1)H_{k_1-2}(\eta_1)]\prod_{p=2}^{m}H_{k_p}(\eta_p). \qquad \text{(C.5)}
\end{aligned}
$$

On the other hand, $\eta_1 H_{k_1-1}(\eta_1)-(k_1-1)H_{k_1-2}(\eta_1) = H_{k_1}(\eta_1)$ by formula (C.4). These relations imply formula (10.24), i.e. Itô's formula. □

I present the proof of another important property of the Hermite polynomials in the following Proposition C2.

Proposition C2 on the completeness of the orthogonal system of Hermite polynomials The Hermite polynomials $H_k(x)$, $k=0,1,2,\dots$, defined in formula (C.5) constitute a complete orthonormal system in the L_2-space of the functions square integrable with respect to the Gaussian measure $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx$ on the real line.

Proof of Proposition C2. Let us consider the orthogonal complement of the subspace generated by the Hermite polynomials in the space of the square integrable functions with respect to the measure $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx$. It is enough to prove that this orthogonal completion contains only the identically zero function. Since the orthogonality of a function to all polynomials of the form x^k, $k=0,1,2,\dots$ is equivalent to the orthogonality of this function to all Hermite polynomials $H_k(x)$, $k=0,1,2,\dots$, Proposition C2 can be reformulated in the following form:

If a function $g(x)$ on the real line is such that

$$\int_{-\infty}^{\infty} x^k g(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx = 0 \quad \text{for all } k=0,1,2,\dots \tag{C.6}$$

and

$$\int_{-\infty}^{\infty} g^2(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx < \infty, \tag{C.7}$$

then $g(x)=0$ for almost all x.

Given a function $g(x)$ on the real line whose absolute value is integrable with respect to the Gaussian measure $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx$ define the (finite) measure ν_g,

$$\nu_g(A) = \int_A g(x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

on the measurable sets of the real line together with its Fourier transform $\tilde{\nu}_g(t) = \int_{-\infty}^{\infty} e^{itx} \nu_g(\,dx)$. (This measure ν_g and its Fourier transform can be defined for all functions g satisfying relation (C.7), because their absolute value is integrable with respect to the Gaussian measure.) First I show that Proposition C2 can be reduced to the following statement: If a function g satisfies both (C.6) and (C.7) then $\tilde{\nu}_g(t) = 0$ for all $-\infty < t < \infty$.

Indeed, if there were a function g satisfying (C.6) and (C.7) which is not identically zero, then the non-negative functions $g^+(x) = \max(0, g(x))$ and $g^-(x) = -\min(0, g(x))$ would be different. Then also their Fourier transform $\tilde{\nu}_{g^+}(t)$ and $\tilde{\nu}_{g^-}(t)$ would be different, since a finite measure is uniquely determined by its Fourier transform. (This statement is equivalent to an important result in probability theory, by which a probability measure on the real line is determined by its characteristic function.) But this would mean that $\tilde{\nu}_g(t) = \tilde{\nu}_{g^+}(t) - \tilde{\nu}_{g^-}(t) \neq 0$ for some t. Hence Proposition C2 can be reduced to the above statement.

Since $\left| e^{itx} - 1 - (itx) - \cdots - \frac{(itx)^k}{k!} \right| \le \frac{|tx|^{(k+1)}}{(k+1)!}$ for all real numbers t, x and integer $k = 1, 2, \ldots$ we may write because of relation (C.6)

$$\begin{aligned} |\tilde{\nu}_g(t)| &= \left| \int_{-\infty}^{\infty} \left(e^{itx} - 1 - (itx) - \cdots - \frac{(itx)^k}{k!} \right) g(x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \right| \\ &\le \int_{-\infty}^{\infty} \frac{|t|^{(k+1)}}{(k+1)!} |x|^{k+1} |g(x)| \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \end{aligned}$$

for all $k = 1, 2, \ldots$ and real number t if the function g satisfies relation (C.6). If it satisfies both relation (C.6) and (C.7), then from the last relation and the Schwarz inequality

$$\begin{aligned} |\tilde{\nu}_g(t)|^2 &\le \text{const.} \frac{|t|^{2(k+1)}}{(k+1)!)^2} \int_{-\infty}^{\infty} |x|^{2(k+1)} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\ &= \text{const.} \frac{|t|^{2(k+1)}}{(k+1)!)^2} 1 \cdot 3 \cdot 5 \cdots (2k+1) \end{aligned}$$

for all real number t and integer $k = 1, 2, \ldots$. Simple calculation shows that the right-hand side of the last estimate tends to zero as $k \to \infty$. This implies that $\tilde{\nu}_g(t) = 0$ for all t, and Proposition C2 holds. □

I finish Appendix C with the proof of Theorem 10.4, a limit theorem about a sequence of normalized degenerate U-statistics. It is based on an appropriate representation of the U-statistics by means of multiple random integrals which makes possible to carry out an appropriate limiting procedure.

Proof of Theorem 10.4. For all $n = 1, 2, \dots$, the normalized degenerate U-statistics $n^{-k/2} k! I_{n,k}(f)$ can be written in the form

$$n^{-k/2} k! I_{n,k}(f) = n^{k/2} \int' f(x_1, \dots, x_k) \mu_n(\,dx_1) \dots \mu_n(\,dx_k)$$
$$= n^{k/2} \int' f(x_1, \dots, x_k)(\mu_n(\,dx_1) - \mu(\,dx_1))$$
$$\dots (\mu_n(\,dx_k) - \mu(\,dx_k)), \tag{C.8}$$

where μ_n is the empirical distribution of the sequence $\xi_1, \dots, \xi_n$ defined in (4.5), and the prime in $\int'$ denotes that the diagonals, i.e. the points $x = (x_1, \dots, x_k)$ such that $x_j = x_{j'}$ for some pairs of indices $1 \le j, j' \le k$, $j \ne j'$, are omitted from the domain of integration. The second identity in relation (C.8) can be justified by means of the identity

$$\int' f(x_1, \dots, x_k)(\mu_n(\,dx_1) - \mu(\,dx_1)) \dots (\mu_n(\,dx_k) - \mu(\,dx_k)) - I_{n,k}(f)$$
$$= \sum_{V\colon V \in \{1,\dots,k\},\, |V| \ge 1} (-1)^{|V|} \int' f(x_1, \dots, x_k)$$
$$\prod_{j \in V} \mu(\,dx_j) \prod_{j \in \{1,\dots,k\} \setminus V} \mu_n(\,dx_j)) = 0. \tag{C.9}$$

This identity holds for a function f canonical with respect to a non-atomic measure μ, because each term in the sum at the right-hand side of (C.9) equals zero. Indeed, the integral of a canonical function f with respect to $\mu(\,dx_j)$ with some index $j \in V$ equals zero for all fixed values $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k$. The non-atomic property of the measure μ was needed to guarantee that this integral equals zero also in the case when the diagonals are omitted from the domain of integration.

We would like to derive Theorem 10.4 from relation (C.8) by means of an appropriate limiting procedure which exploits the convergence of the random fields $n^{1/2}(\mu_n(A) - \mu(A))$, $A \in \mathscr{X}$, to a Gaussian field $\nu(A)$, $A \in \mathscr{X}$, as $n \to \infty$. But some problems arise if we want to carry out such a program, because the fields $n^{1/2}(\mu_n - \mu)$ converge to a non white noise type Gaussian field. The limit we get is similar to a Wiener bridge on the real line. Hence a relation between Wiener processes and Wiener bridges suggests to write the following version of formula (C.8).

Let us take a standard Gaussian random variable η, independent of the random sequence $\xi_1, \xi_2, \dots$. For a canonical function f the following version of (C.8) holds.

$$n^{-k/2} k! I_{n,k}(f) = J'_{n,k}(f) \tag{C.10}$$

with

$$J'_{n,k}(f)=\int' f(x_1,\dots,x_k)\left[\sqrt{n}(\mu_n(\,dx_1)-\mu(\,dx_1))+\eta\mu(\,dx_1)\right]$$
$$\dots\left[\sqrt{n}(\mu_n(\,dx_k)-\mu(\,dx_k))+\eta\mu(\,dx_k)\right]. \tag{C.11}$$

This relation can be seen similarly to (C.8).

The random measures $n^{1/2}(\mu_n-\mu)+\eta\mu$ converge to a white noise with reference measure μ. Hence Theorem 10.4 can be proved by means of formulas (C.10) and (C.11) with the help of an appropriate limiting procedure. More explicitly, I claim that the following slightly more general result holds. The expressions $J'_{n,k}(f)$ introduced in (C.11) converge in distribution to the Wiener–Itô integral $k!Z_{\mu,k}(f)$ as $n\to\infty$ for all functions f square integrable with respect to the product measure μ^k. This result also holds for non-canonical functions f. This limit theorem together with relation (C.10) imply Theorem 10.4.

The convergence of the random variables $J'_{n,k}(f)$ defined in (C.11) to the Wiener–Itô integral $k!Z_{\mu,k}(f)$ can be easily checked for elementary functions $f\in\bar{\mathscr{H}}_{\mu,k}$. Indeed, if $A_1,\dots,A_M$ are disjoint sets with $\mu(A_s)<\infty$, then the multi-dimensional central limit theorem implies that the random vectors $\{\sqrt{n}((\mu_n(A_s)-\mu(A_s))+\eta\mu(A_s),\ 1\le s\le M\}$ converge in distribution to the random vector $\{(\mu_W(A_s),\ 1\le s\le M\}$, i.e. to a set of independent normal random variables ζ_s, $E\zeta_s=0$, $1\le s\le M$, with variance $E\zeta_s^2=\mu(A_s)$ as $n\to\infty$. The definition of the elementary functions given in (10.2) shows that this central limit theorem implies the demanded convergence of the sequence $J'_{n,k}(f)$ to $k!Z_{\mu,k}(f)$ for elementary functions.

To show the convergence of the sequence $J'_{n,k}(f)$ to $k!Z_{\mu,k}(f)$ in the general case take for any function $f\in\mathscr{H}_{\mu,k}$ a sequence of elementary functions $f_N\in\bar{\mathscr{H}}_{\mu,k}$ such that $\|f-f_N\|_2\to 0$ as $N\to\infty$. Then $E(Z_{\mu,k}(f)-Z_{\mu,k}(f_N))^2=E(Z_{\mu,k}(f-f_N))^2\to 0$ as $N\to\infty$ by Property (c) in Theorem 10.1. Hence the already proved part of the theorem implies that there exists some sequence of positive integers, $N(n)$, $n=1,2,\dots$, in such a way that $N(n)\to\infty$, and the sequence $J'_{n,k}(f_{N(n)})$ converges to $k!Z_{\mu,k}(f)$ in distribution as $n\to\infty$. Thus to complete the proof of Theorem 10.4 it is enough to show that $E(J'_{n,k}(f_{N(n)})-J'_{n,k}(f))^2=E(J'_{n,k}(f_{N(n)}-f))^2\to 0$ as $n\to\infty$.

It is enough to show that

$$E(J'_{n,k}(f))^2\le C\|f\|_2^2 \quad \text{for all } f\in\mathscr{H}_{\mu,k} \tag{C.12}$$

with a constant $C=C_k$ depending only on the order k of the function f and to apply inequality (C.12) for the functions $f_{N(n)}-f$. Relation (C.12) is a relatively simple consequence of Corollary 1 of Theorem 9.4.

Indeed,

$$J'_{n,k}(f) = \sum_{V\subset\{1,\dots,k\}} \eta^{k-|V|}|V|!J_{n,|V|}(f_V)$$

with

$$f_V(x_j,\ j\in V) = \int f(x_1,\dots,x_k) \prod_{j'\in\{1,\dots,k\}\setminus V} \mu(dx_{j'})$$

and the random integral $J_{n,k}(\cdot)$ defined in (4.8), hence

$$E(J'_{n,k}(f))^2 \le 2^k \sum_{V\subset\{1,\dots,k\}} (|V|!)^2 E\eta^{2(k-|V|)}\cdot EJ^2_{n,|V|}(f_V). \tag{C.13}$$

Inequality $\|f_V\|_2 \le \|f\|_2$ holds for all sets $V\subset\{1,\dots,k\}$, hence an application of Corollary 1 of Theorem 9.4 to all random integrals $J_{n,|V|}(f)$ supplies (C.12). □

The above proof also yields the following slight generalization of Theorem 10.4. Let us consider a finite sequence of functions $f_j\in\mathscr{H}_{\mu,j}$, $1\le j\le k$, canonical with respect to a non-atomic probability measure μ. The vectors $\{n^{-j/2}I_{n,j}(f_j), 1\le j\le k\}$, consisting of normalized degenerate U-statistics defined with the help of a sequence of independent μ-distributed random variables converge to the random vector $\{Z_{\mu,j}(f_j), 1\le j\le k\}$ in distribution as $n\to\infty$. This result together with Theorem 9.4 imply the following limit theorem about multiple random integrals $J_{n,k}(f)$.

Theorem 10.4′ (Limit theorem about multiple random integrals with respect to a normalized empirical measure). *Let a sequence of independent and identically distributed random variables $\xi_1,\xi_2,\dots$ be given with some non-atomic distribution μ on a measurable space $(X,\mathscr{X})$ together with a function $f(x_1,\dots,x_k)$ on the k-fold product $(X^k,\mathscr{X}^k)$ of the space $(X,\mathscr{X})$ such that*

$$\int f^2(x_1,\dots,x_k)\mu(\,dx_1)\dots\mu(\,dx_k)<\infty.$$

Let us consider for all $n=1,2,\dots$ the random integrals $J_{n,k}(f)$ of order k defined in formulas (4.5) and (4.8) with the help of the empirical distribution μ_n of the sequence $\xi_1,\dots,\xi_n$ and the function f. These random integrals $J_{n,k}(f)$ converge in distribution, as $n\to\infty$, to the following sum $U(f)$ of multiple Wiener–Itô integrals:

$$\begin{aligned} U(f) &= \sum_{V\subset\{1,\dots,k\}} C(k,V)Z_{\mu,|V|}(f_V) \\ &= \sum_{V\subset\{1,\dots,k\}} \frac{C(k,V)}{|V|!} \int f_V(x_j,\ j\in V)\prod_{j\in V}\mu_W(dx_j), \end{aligned}$$

where the functions $f_V(x_j,\ j\in V)$*,* $V\subset\{1,\dots,k\}$*, are those functions defined in formula (9.3) which appear in the Hoeffding decomposition of the function* $f(x_1,\dots,x_k)$*, the constants* $C(k,V)$ *are the limits appearing in the limit relation* $\lim\limits_{n\to\infty} C(n,k,V)=C(k,V)$ *satisfied by the coefficients* $C(n,k,V)$ *in formula (9.16), and* μ_W *is a white noise with reference measure* μ*.*

An essential step of the proof of Theorem 10.4 was the reduction of the case of general kernel functions to the case of elementary kernel functions. Let me make some comments about it.

It would be simple to make such a reduction if we had a good approximation of a canonical function with such elementary functions which are also canonical. But it is very hard to find such an approximation. To overcome this difficulty we reduced the proof of Theorem 10.4 to a modified version of this result where instead of a limit theorem for degenerate U-statistics a limit theorem for the random variables $J'_{n,k}(f)$ introduced in formula (C.11) has to be proved. In the proof of such a version we could apply the approximation of a general kernel function with not necessarily canonical elementary functions. Theorem 9.4 helped us to work with such an approximation. Another natural way to overcome the above difficulty is to apply a Poissonian approximation of the normalized empirical measure. Such an approach was applied in [16,34], where some generalizations of Theorem 10.4 were proved.

Appendix D
The Proof of Theorem 14.3 About U-Statistics and Decoupled U-Statistics

The proof of Theorem 14.3. It will be simpler to formulate and prove a generalized version of Theorem 14.3 where such generalized U-statistics are considered in which different kernel functions may appear in each term of the sum. More explicitly, let $\ell = \ell(n,k)$ denote the set of all such sequences $l = (l_1, \dots, l_k)$ of integers of length k for which $1 \le l_j \le n$, $1 \le j \le k$. To define generalized U-statistics let us fix a set of functions $\{f_{l_1,\dots,l_k}(x_1,\dots,x_k),\ (l_1,\dots,l_k) \in \ell\}$ which map the space $(X^k, \mathcal{X}^k)$ to a separable Banach space B, and have the property $f_{l_1,\dots,l_k}(x_1,\dots,x_k) \equiv 0$ if $l_j = l_{j'}$ for some indices $j \ne j'$. (The last condition corresponds to that property of U-statistics that the diagonals are omitted from the summation in their definition.) Let us denote this set of functions by $f(\ell)$, and define, similarly to the U-statistics and decoupled U-statistics the generalized U-statistics and generalized decoupled U-statistics by the formulas

$$I_{n,k}(f(\ell)) = \frac{1}{k!} \sum_{(l_1,\dots,l_k)\colon 1 \le l_j \le n,\ j=1,\dots,k} f_{l_1,\dots,l_k}\left(\xi_{l_1},\dots,\xi_{l_k}\right) \tag{D.1}$$

and

$$\bar{I}_{n,k}(f(\ell)) = \frac{1}{k!} \sum_{(l_1,\dots,l_k)\colon 1 \le l_j \le n,\ j=1,\dots,k} f_{l_1,\dots,l_k}\left(\xi^{(1)}_{l_1},\dots,\xi^{(k)}_{l_k}\right) \tag{D.2}$$

(with the same independent and identically distributed random variables ξ_l and $\xi_l^{(j)}$, $1 \le l \le n$, $1 \le j \le k$, as in the definition of the original U-statistics and decoupled U-statistics.)

The following generalization of relation (14.13) will be proved.

$$P\left(\|I_{n,k}(f(\ell))\| > u\right) \le A(k) P\left(\|\bar{I}_{n,k}(f(\ell))\| > \gamma(k)u\right) \tag{D.3}$$

P. Major, *On the Estimation of Multiple Random Integrals and U-Statistics*, Lecture Notes in Mathematics 2079, DOI 10.1007/978-3-642-37617-7,

with some constants $A(k) > 0$ and $\gamma(k) > 0$ depending only on the order k of these generalized U-statistics. The sign $\|\cdot\|$ in (D.3) denotes the norm in the Banach space we are working in.

We concentrate mainly on the proof of the generalization (D.3) of relation (14.13). Formula (14.14) is a relatively simple consequence of it. Formula (D.3) will be proved by means of an inductive procedure which works only in this more general setting. It will be derived from the following statement.

Let us take two independent copies $\xi_1^{(1)}, \dots, \xi_n^{(1)}$ and $\xi_1^{(2)}, \dots, \xi_n^{(2)}$ of our original sequence of random variables $\xi_1, \dots, \xi_n$, and introduce for all sets $V \subset \{1, \dots, k\}$ the function $\alpha_V(j)$, $1 \le j \le k$, defined as $\alpha_V(j) = 1$ if $j \in V$ and $\alpha_V(j) = 2$ if $j \notin V$. Let us define with their help the following version of decoupled U-statistics:

$$I_{n,k,V}(f(\ell)) = \frac{1}{k!} \sum_{(l_1,\dots,l_k):\, 1\le l_j\le n,\ j=1,\dots,k} f_{l_1,\dots,l_k}\left(\xi_{l_1}^{(\alpha_V(1))}, \dots, \xi_{l_k}^{(\alpha_V(k))}\right)$$
$$\text{for all } V \subset \{1, \dots, k\}. \tag{D.4}$$

The following inequality will be proved: There are some constants $C_k > 0$ and $D_k > 0$ depending only on the order k of the generalized U-statistic $I_{n,k}(f(\ell))$ such that for all numbers $u > 0$

$$P\left(\|I_{n,k}(f(\ell))\| > u\right) \le \sum_{V\subset\{1,\dots,k\},\, 1\le|V|\le k-1} C_k P\left(D_k \|I_{n,k,V}(f(\ell))\| > u\right). \tag{D.5}$$

Here $|V|$ denotes the cardinality of the set V, and the condition $1 \le |V| \le k-1$ in the summation of formula (D.5) means that the sets $V = \emptyset$ and $V = \{1, \dots, k\}$ are omitted from the summation, i.e. the terms where either $\alpha_V(j) = 1$ or $\alpha_V(j) = 2$ for all $1 \le j \le k$ are not considered. Formula (D.3) can be derived from formula (D.5) by means of an inductive argument. The hard part of the problem is to prove formula (D.5). To do this first we prove the following simple lemma.

Lemma D1. *Let ξ and η be two independent and identically distributed random variables taking values in a separable Banach space B. Then*

$$3P\left(|\xi + \eta| > \frac{2}{3}u\right) \ge P(|\xi| > u) \quad \textit{for all } u > 0.$$

Proof of Lemma D1. Let ξ, η and ζ be three independent, identically distributed random variables taking values in B. Then

$$
\begin{aligned}
3P\left(|\xi+\eta|>\frac{2}{3}u\right) &= P\left(|\xi+\eta|>\frac{2}{3}u\right)+P\left(|\xi+\zeta|>\frac{2}{3}u\right)\\
&\quad +P\left(|-(\eta+\zeta)|>\frac{2}{3}u\right)\\
&\ge P(|\xi+\eta+\xi+\zeta-\eta-\zeta|>2u)=P(|\xi|>u).
\end{aligned}
$$

□

To prove formula (D.5) we introduce the random variable

$$
\begin{aligned}
T_{n,k}(f(\ell)) &= \frac{1}{k!}\sum_{\substack{(l_1,\dots,l_k),\ (s_1,\dots,s_k):\\ 1\le l_j\le n,\ s_j=1 \text{ or } s_j=2,\ j=1,\dots,k,}} f_{l_1,\dots,l_k}\left(\xi_{l_1}^{(s_1)},\dots,\xi_{l_k}^{(s_k)}\right)\\
&= \sum_{V\subset\{1,\dots,k\}} I_{n,k,V}(f(\ell)). \qquad\qquad (D.6)
\end{aligned}
$$

The random variables $I_{n,k}(f(\ell))$, $I_{n,k,\emptyset}(f(\ell))$ and $I_{n,k,\{1,\dots,k\}}(f(\ell))$ are identically distributed, and the last two random variables are independent of each other. Hence Lemma D1 yields that

$$
\begin{aligned}
P(\|I_{n,k}(f(\ell))\|>u) &\le 3P\left(\|I_{n,k,\emptyset}(f(\ell))+I_{n,k,\{1,\dots,k\}}(f(\ell))\|>\frac{2}{3}u\right)\\
&= 3P\left(\left\|T_{n,k}(f(\ell))-\sum_{V\colon V\subset\{1,\dots,k\},\ 1\le|V|\le k-1} I_{n,k,|V|}(f(\ell))\right\|>\frac{2}{3}u\right)\\
&\le 3P(3\cdot 2^{k-1}\|T_{n,k}(f(\ell))\|>u)\\
&\quad+\sum_{V\colon V\subset\{1,\dots,k\},\ 1\le|V|\le k-1} 3P(3\cdot 2^{k-1}\|I_{n,k,|V|}(f(\ell))\|>u). \qquad (D.7)
\end{aligned}
$$

To derive relation (D.5) from relation (D.7) we need a good upper bound on the probability $P(3\cdot 2^{k-1}\|T_{n,k}(f(\ell))\|>u)$. To get such an estimate we shall compare the tail distribution of $\|T_{n,k}(f(\ell))\|$ with that of $\|I_{n,k,V}(f(\ell))\|$ for an arbitrary set $V\subset\{1,\dots,k\}$. This will be done with the help of Lemmas D2 and D4 formulated below.

In Lemma D2 such a random variable $\|\hat I_{n,k,V}(f(\ell))\|$ will be constructed whose distribution agrees with that of $\|I_{n,k,V}(f(\ell))\|$. The expression $\hat I_{n,k,V}(f(\ell))$, whose norm will be investigated will be defined in formulas (D.8) and (D.9). It is a random polynomial of some Rademacher functions $\varepsilon_1,\dots,\varepsilon_n$. The coefficients of this polynomial are random variables, independent of the Rademacher functions $\varepsilon_1,\dots,\varepsilon_n$. Beside this, the constant term of this polynomial equals $T_{n,k}(f(\ell))$. These properties of the polynomial $\hat I_{n,k,V}(f(\ell))$ together with Lemma D4 formulated below enable

us prove such an estimate on the distribution of $\|T_{n,k}(f(\ell))\|$ that together with formula (D.7) imply relation (D.5). Let us formulate these lemmas.

Lemma D2. *Let us consider a sequence of independent random variables $\varepsilon_1,\dots,\varepsilon_n$, $P(\varepsilon_l=1)=P(\varepsilon_l=-1)=\frac{1}{2}$, $1\le l\le n$, which is also independent of the random variables $\xi_1^{(1)},\dots,\xi_n^{(1)}$ and $\xi_1^{(2)},\dots,\xi_n^{(2)}$ appearing in the definition of the modified decoupled U-statistics $I_{n,k,V}(f(\ell))$ given in formula (D.4). Let us define with their help the sequences of random variables $\eta_1^{(1)},\dots,\eta_n^{(1)}$ and $\eta_1^{(2)},\dots,\eta_n^{(2)}$ whose elements $(\eta_l^{(1)},\eta_l^{(2)})=(\eta_l^{(1)}(\varepsilon_l),\eta_l^{(2)}(\varepsilon_l))$, $1\le l\le n$, are defined by the formula*

$$(\eta_l^{(1)}(\varepsilon_l),\eta_l^{(2)}(\varepsilon_l))=\left(\frac{1+\varepsilon_l}{2}\xi_l^{(1)}+\frac{1-\varepsilon_l}{2}\xi_l^{(2)},\frac{1-\varepsilon_l}{2}\xi_l^{(1)}+\frac{1+\varepsilon_l}{2}\xi_l^{(2)}\right),$$

i.e. let

$$(\eta_l^{(1)}(\varepsilon_l),\eta_l^{(2)}(\varepsilon_l))=(\xi_l^{(1)},\xi_l^{(2)})\quad \text{if } \varepsilon_l=1,$$

and

$$(\eta_l^{(1)}(\varepsilon_l),\eta_l^{(2)}(\varepsilon_l))=(\xi_l^{(2)},\xi_l^{(1)})\quad \text{if } \varepsilon_l=-1,\quad 1\le l\le n.$$

Then the joint distribution of the pair of sequences of random variables $\xi_1^{(1)},\dots,\xi_n^{(1)}$ and $\xi_1^{(2)},\dots,\xi_n^{(2)}$ agrees with that of the pair of sequences $\eta_1^{(1)},\dots,\eta_n^{(1)}$ and $\eta_1^{(2)},\dots,\eta_n^{(2)}$, which is also independent of the sequence $\varepsilon_1,\dots,\varepsilon_n$.

Let us fix some $V\subset\{1,\dots,k\}$, and introduce the random variable

$$\hat I_{n,k,V}(f(\ell))=\frac{1}{k!}\sum_{(l_1,\dots,l_k)\colon 1\le l_j\le n,\ j=1,\dots,k} f_{l_1,\dots,l_k}\left(\eta_{l_1}^{(\alpha_V(1))},\dots,\eta_{l_k}^{(\alpha_V(k))}\right), \tag{D.8}$$

where similarly to formula (D.4) $\alpha_V(j)=1$ if $j\in V$, and $\alpha_V(j)=2$ if $j\notin V$. Then the identity

$$\begin{aligned}&2^k\hat I_{n,k,V}(f(\ell)) \\ &=\frac{1}{k!}\sum_{\substack{(l_1,\dots,l_k),\ (s_1,\dots,s_k)\colon\\ 1\le l_j\le n,\ s_j=1 \text{ or } s_j=2,\\ j=1,\dots,k,}}(1+\kappa_{s_1,V}^{(1)}\varepsilon_{l_1})\cdots(1+\kappa_{s_k,V}^{(k)}\varepsilon_{l_k})f_{l_1,\dots,l_k}\left(\xi_{l_1}^{(s_1)},\dots,\xi_{l_k}^{(s_k)}\right)\end{aligned} \tag{D.9}$$

holds, where $\kappa_{1,V}^{(j)}=1$ and $\kappa_{2,V}^{(j)}=-1$ if $j\in V$, and $\kappa_{1,V}^{(j)}=-1$ and $\kappa_{2,V}^{(j)}=1$ if $j\notin V$, i.e. $\kappa_{1,V}^{(j)}=3-2\alpha_V(j)$ and $\kappa_{2,V}^{(j)}=-\kappa_{1,V}^{(j)}$.

Before the formulation of Lemma D4 another Lemma D3 will be presented which will be applied in its proof.

Lemma D3. *Let Z be a random variable taking values in a separable Banach space B with expectation zero, i.e. let $E\kappa(Z)=0$ for all $\kappa\in B'$, where B' denotes*

the (Banach) space of all (bounded) linear transformations of B to the real line. Then $P(\|v+Z\|\ge\|v\|)\ge\inf\limits_{\kappa\in B'}\frac{(E|\kappa(Z)|)^2}{4E\kappa(Z)^2}$ for all $v\in B$.

Lemma D4. *Let us consider a positive integer n and a sequence of independent random variables $\varepsilon_1,\dots,\varepsilon_n$, $P(\varepsilon_l=1)=P(\varepsilon_l=-1)=\frac12$, $1\le l\le n$. Beside this, fix some positive integer k, take a separable Banach space B and choose some elements $a_s(l_1,\dots,l_s)$ of this Banach space B, $1\le s\le k$, $1\le l_j\le n$, $l_j\ne l_{j'}$ if $j\ne j'$, $1\le j,j'\le s$. With the above notations the inequality*

$$P\left(\left\|v+\sum_{s=1}^k\sum_{\substack{(l_1,\dots,l_s)\colon 1\le l_j\le n,\ j=1,\dots,s,\\ l_j\ne l_{j'} \text{ if } j\ne j'}}a_s(l_1,\dots,l_s)\varepsilon_{l_1}\cdots\varepsilon_{l_s}\right\|\ge\|v\|\right)\ge c_k \tag{D.10}$$

holds for all $v\in B$ with some constant $c_k>0$ which depends only on the parameter k. In particular, it does not depend on the norm in the separable Banach space B.

Proof of Lemma D2. Let us consider the conditional joint distribution of the sequences of random variables $\eta_1^{(1)},\dots,\eta_n^{(1)}$ and $\eta_1^{(2)},\dots,\eta_n^{(2)}$ under the condition that the random vector $\varepsilon_1,\dots,\varepsilon_n$ takes the value of some prescribed ±1 series of length n. Observe that this conditional distribution agrees with the joint distribution of the sequences $\xi_1^{(1)},\dots,\xi_n^{(1)}$ and $\xi_1^{(2)},\dots,\xi_n^{(2)}$ for all possible conditions. This fact implies the statement about the joint distribution of the sequences $(\eta_l^{(1)},\eta_l^{(2)})$, $1\le l\le n$ and their independence of the sequence $\varepsilon_1,\dots,\varepsilon_n$.

To prove identity (D.9) let us fix a set $M\subset\{1,\dots,n\}$, and consider the case when $\varepsilon_l=1$ if $l\in M$ and $\varepsilon_l=-1$ if $l\notin M$. Put $\beta_{V,M}(j,l)=1$ if $j\in V$ and $l\in M$ or $j\notin V$ and $l\notin M$, and let $\beta_{V,M}(j,l)=2$ otherwise. Then we have for all $(l_1,\dots,l_k)$, $1\le l_j\le n$, $1\le j\le k$, and our fixed set V

$$\begin{aligned}&\sum_{\substack{(s_1,\dots,s_k)\colon\\ s_j=1 \text{ or } s_j=2,\ j=1,\dots,k}}(1+\kappa^{(1)}_{s_1,V}\varepsilon_{l_1})\cdots(1+\kappa^{(k)}_{s_k,V}\varepsilon_{l_k})f_{l_1,\dots,l_k}\left(\xi_{l_1}^{(s_1)},\dots,\xi_{l_k}^{(s_k)}\right)\\ &\qquad=2^kf_{l_1,\dots,l_k}\left(\xi_{l_1}^{(\beta_{V,M}(1,l_1))},\dots,\xi_{l_k}^{(\beta_{V,M}(k,l_k))}\right),\end{aligned}\tag{D.11}$$

since the product $(1+\kappa^{(1)}_{s_1,V}\varepsilon_{l_1})\cdots(1+\kappa^{(k)}_{s_k,V}\varepsilon_{l_k})$ equals either zero or 2^k, and it equals 2^k for that sequence $(s_1,\dots,s_k)$ for which $\kappa^{(j)}_{s_j,V}\varepsilon_{l_j}=1$ for all $1\le j\le k$, and the relation $\kappa^{(j)}_{s_j,V}\varepsilon_{l_j}=1$ is equivalent to $\beta_{V,M}(j,l_j)=s_j$ for all $1\le j\le k$. (In relation (D.11) it is sufficient to consider only such products for which $l_j\ne l_{j'}$ if $j\ne j'$ because of the properties of the functions $f_{l_1,\dots,l_k}$.)

Beside this, $\xi_l^{\beta_{V,M}(l,j)}=\eta_l^{\alpha_V(j)}$ for all $1\le l\le n$ and $1\le j\le k$, and as a consequence

$$f_{l_1,\dots,l_k}\left(\xi_{l_1}^{(\beta_{V,M}(1,l_1))},\dots,\xi_{l_k}^{(\beta_{V,M}(k,l_k))}\right)=f_{l_1,\dots,l_k}\left(\eta_{l_1}^{(\alpha_V(1))},\dots,\eta_{l_k}^{(\alpha_V(k))}\right).$$

Summing up the identities (D.11) for all $1\le l_1,\dots,l_k\le n$ and applying the last identity we get relation (D.9), since the identity obtained in such a way holds for all $M\subset\{1,\dots,n\}$. □

Proof of Lemma D3. Let us first observe that if ξ is a real valued random variable with zero expectation, then $P(\xi\ge 0)\ge\frac{(E|\xi|)^2}{4E\xi^2}$ since $(E|\xi|)^2=4(E(\xi I(\{\xi\ge 0\}))^2\le 4P(\xi\ge 0)E\xi^2$ by the Schwarz inequality, where $I(A)$ denotes the indicator function of the set A. (In the above calculation and in the subsequent proofs I apply the convention $\frac{0}{0}=1$. We need this convention if $E\xi^2=0$. In this case we have the identities $P(\xi=0)=1$ and $E|\xi|=0$, hence the above proved inequality holds in this case, too.)

Given some $v\in B$, let us choose a linear operator κ such that $\|\kappa\|=1$, and $\kappa(v)=\|v\|$. Such an operator exists by the Banach–Hahn theorem. Observe that $\{\omega\colon \|v+Z(\omega)\|\ge\|v\|\}\supset\{\omega\colon \kappa(v+Z(\omega))\ge\kappa(v)\}=\{\omega\colon\kappa(Z(\omega))\ge 0\}$. Beside this, $E\kappa(Z)=0$. Hence we can apply the above proved inequality for $\xi=\kappa(Z)$, and it yields that $P(\|v+Z\|\ge\|v\|)\ge P(\kappa(Z)\ge 0)\ge\frac{(E|\kappa(Z)|)^2}{4E\kappa(Z)^2}$. Lemma D3 is proved. □

Proof of Lemma D4. Take the class of random polynomials

$$Y=\sum_{s=1}^{k}\sum_{\substack{(l_1,\dots,l_s)\colon 1\le l_j\le n,\ j=1,\dots,s,\\ l_j\ne l_{j'}\text{ if } j\ne j'}} b_s(l_1,\dots,l_s)\varepsilon_{l_1}\cdots\varepsilon_{l_s},$$

where ε_l, $1\le l\le n$, are independent random variables with $P(\varepsilon_l=1)=P(\varepsilon_l=-1)=\frac12$, and the coefficients $b_s(l_1,\dots,l_s)$, $1\le s\le k$, are arbitrary real numbers. The proof of Lemma D4 can be reduced to the statement that there exists a constant $c_k>0$ depending only on the order k of these polynomials such that the inequality

$$(E|Y|)^2\ge 4c_k EY^2. \tag{D.12}$$

holds for all such polynomials Y. Indeed, consider the polynomial

$$Z=\sum_{s=1}^{k}\sum_{\substack{(l_1,\dots,l_s)\colon 1\le l_j\le n,\ j=1,\dots,s,\\ l_j\ne l_{j'}\text{ if } j\ne j'}} a_s(l_1,\dots,l_s)\varepsilon_{l_1}\cdots\varepsilon_{l_s},$$

and observe that $E\kappa(Z)=0$ for all linear functionals κ on the space B. Hence Lemma D3 implies that the left-hand side expression in (D.10) is bounded

from below by $\inf\limits_{\kappa\in B'} \frac{(E|\kappa(Z)|)^2}{4E\kappa(Z)^2}$. On the other hand, relation (D.12) implies that $\inf\limits_{\kappa\in B'} \frac{(E|\kappa(Z)|)^2}{4E\kappa(Z)^2} \ge c_k$.

To prove relation (D.12) first we compare the moments EY^2 and EY^4. Let us introduce the random variables

$$Y_s = \sum_{\substack{(l_1,\dots,l_s)\colon 1\le l_j\le n,\ j=1,\dots,s,\\ l_j\ne l_{j'} \text{ if } j\ne j'}} b_s(l_1,\dots,l_s)\varepsilon_{l_1}\cdots\varepsilon_{l_s} \quad 1\le s\le k.$$

We shall show that the estimates of Chap. 13 imply that

$$EY_s^4 \le 2^{4s}\left(EY_s^2\right)^2 \tag{D.13}$$

for these random variables Y_s.

Relation (D.13) together with the uncorrelatedness of the random variables Y_s, $1\le s\le k$, imply that

$$\begin{aligned} EY^4 = E\left(\sum_{s=1}^k Y_s\right)^4 &\le k^3\sum_{s=1}^k EY_s^4 \le k^3 2^{4k}\sum_{s=1}^k (EY_s^2)^2 \\ &\le k^3 2^{4k}\left(\sum_{s=1}^k EY_s^2\right)^2 = k^3 2^{4k}(EY^2)^2. \end{aligned}$$

This estimate together with the Hölder inequality with $p=3$ and $q=\frac{3}{2}$ yield that

$$EY^2 = E|Y|^{4/3}\cdot|Y|^{2/3} \le (EY^4)^{1/3}(E|Y|)^{2/3} \le k2^{4k/3}(EY^2)^{2/3}(E|Y|)^{2/3},$$

i.e. $EY^2 \le k^3 2^{4k}(E|Y|)^2$, and relation (D.12) holds with $4c_k = k^{-3}2^{-4k}$. Hence to complete the proof of Lemma D4 it is enough to check relation (D.13).

In the proof of relation (D.13) we may assume that the coefficients $b_s(l_1,\dots,l_s)$ of the random variable Y_s are symmetric functions of the arguments $l_1,\dots,l_s$, since a symmetrization of these coefficients does not change the value of Y. Put

$$B_s^2 = \sum_{\substack{(l_1,\dots,l_s)\colon 1\le l_j\le n,\ j=1,\dots,s,\\ l_j\ne l_{j'} \text{ if } j\ne j'}} b_s^2(l_1,\dots,l_s), \quad 1\le s\le k.$$

Then

$$EY_s^2 = s!B_s^2,$$

and

$$EY_s^4 \le 1\cdot 3\cdot 5\cdots(4s-1)B_s^4 = \frac{(4s)!}{2^{2s}(2s)!}B_s^4$$

by Lemmas 13.4 and 13.5 with the choice $M = 2$ and $k = s$. Inequality (D.13) follows from the last two relations. Indeed, to prove formula (D.13) by means of these relations it is enough to check that $\frac{(4s)!}{2^{2s}(2s)!(s!)^2} \le 2^{4s}$. But it is easy to check this inequality with induction with respect to s. (Actually there is a well-known inequality in the literature, known under the name Borell's inequality, which implies inequality (D.13) with a better coefficient at the right hand side of this estimate.) We have proved Lemma D4. □

Let us turn back to the estimation of the probability $P(3 \cdot 2^{k-1}\|T_{n,k}(f)\| > u)$. Let us introduce the σ-algebra $\mathscr{F} = \mathscr{B}(\xi_l^{(1)}, \xi_l^{(2)}, 1 \le l \le n)$ generated by the random variables $\xi_l^{(1)}, \xi_l^{(2)}$, $1 \le l \le n$, and fix some set $V \subset \{1, \dots, k\}$. I show with the help of Lemma D4 and formula (D.9) that there exists some constant $c_k > 0$ such that the random variables $T_{n,k} f(\ell))$ defined in formula (D.6) and $\hat{I}_{n,k,V}(f(\ell))$ defined in formula (D.8) satisfy the inequality

$$P\left(\|2^k \hat{I}_{n,k,V}(f(\ell))\| > \|T_{n,k}(f(\ell))\| | \mathscr{F}\right) \ge c_k \qquad \text{with probability 1.} \tag{D.14}$$

In the proof of (D.14) we shall exploit that in formula (D.9) $2^k \hat{I}_{n,k,V}(f(\ell))$ is represented by a polynomial of the Rademacher functions $\varepsilon_1, \dots, \varepsilon_n$ whose constant term is $T_{n,k}(f(\ell))$. The coefficients of this polynomial are functions of the random variables $\xi_l^{(1)}$ and $\xi_l^{(2)}$, $1 \le l \le n$. The independence of these random variables from ε_l, $1 \le l \le n$, and the definition of the σ-algebra $\mathscr{F}$ yield that

$$\begin{aligned} &P\left(\|2^k \hat{I}_{n,k,V}(f(\ell))\| > \|T_{n,k}(f(\ell))\| | \mathscr{F}\right) \\ &\quad = P_{\varepsilon_V}\left(\left\| \frac{1}{k!} \sum_{\substack{(l_1,\dots,l_k),\, (s_1,\dots,s_k): \\ 1 \le l_j \le n, s_j = 1 \text{ or } s_j = 2, \\ j = 1,\dots,k,}} (1 + \kappa_{s_1,V}^{(1)} \varepsilon_{l_1}) \cdots (1 + \kappa_{s_k,V}^{(k)} \varepsilon_{l_k}) \right.\right. \\ &\qquad\qquad \left. f_{l_1,\dots,l_k}\left(\xi_{l_1}^{(s_1)}, \dots, \xi_{l_k}^{(s_k)}\right) \right\| \\ &\quad \left. > \|T_{n,k}(f(\ell))(\xi_l^{(j)},\ 1 \le l \le n,\ j = 1, 2)\| \right), \end{aligned} \tag{D.15}$$

where P_{ε_V} means that the values of the random variables $\xi_l^{(1)}$, $\xi_l^{(2)}$, $1 \le l \le n$, are fixed, (their value depend on the atom of the σ-algebra $\mathscr{F}$ we are considering) and the probability is taken with respect to the remaining random variables ε_l, $1 \le l \le n$. At the right-hand side of (D.15) the probability of such an event is considered that the norm of a polynomial of order k of the random variables $\varepsilon_1, \dots, \varepsilon_n$ is larger than $\|T_{n,k}(f(\ell))(\xi_l^{(j)},\ 1 \le l \le n,\ j = 1, 2)\|$. Beside this, the constant term of this polynomial equals $T_{n,k}(f(\ell))(\xi_l^{(j)},\ 1 \le l \le n,\ j = 1, 2)$. Hence this probability can be bounded by means of Lemma D4, and this result yields relation (D.14).

The distributions of $I_{n,k,V}(f(\ell))$ and $\hat I_{n,k,V}(f(\ell))$ agree by the first statement of Lemma D2 and a comparison of formulas (D.4) and (D.8). Hence relation (D.14) implies that

$$\begin{aligned}
P\left(\|2^k I_{n,k,V}(f(\ell))\| \ge \frac{1}{3}\cdot 2^{1-k}u\right) &= P\left(\|2^k \hat I_{n,k,V}(f(\ell))\| \ge \frac{1}{3}\cdot 2^{1-k}u\right)\\
&\ge P\left(\|2^k \hat I_{n,k,V}(f(\ell))\| \ge \|T_{n,k}(f(\ell))\|,\ \|T_{n,k}(f(\ell))\| \ge \frac{1}{3}\cdot 2^{1-k}u\right)\\
&= \int_{\{\omega\colon \|T_{n,k}(f(\ell))(\omega)\|\ge\frac13\cdot 2^{1-k}u\}} P\left(\|2^k \hat I_{n,k,V}(f(\ell))\| > \|T_{n,k}(f(\ell))\| | \mathscr{F}\right)\,dP\\
&\ge c_k P(3\cdot 2^{k-1}\|T_{n,k}(f(\ell))\| \ge u).
\end{aligned}$$

The last inequality with the choice of any set $V\subset\{1,\dots,k\}$, $1\le |V|\le k-1$, together with relation (D.7) imply formula (D.5).

We shall formulate an inductive hypothesis, and relation (D.3) will be proved together with it by means of an induction procedure with respect to the order k of the U-statistic. In the proof of this inductive procedure we shall apply the already proved relation (D.5). To formulate it some new quantities will be introduced.

Let $\mathscr{W}=\mathscr{W}(k)$ denote the set of all partitions of the set $\{1,\dots,k\}$. Let us fix k independent copies $\xi_1^{(j)},\dots,\xi_n^{(j)}$, $1\le j\le k$, of the sequence of random variables $\xi_1,\dots,\xi_n$. Given a partition $W=(U_1,\dots,U_s)\in\mathscr{W}(k)$ let us introduce the function $s_W(j)$, $1\le j\le k$, which tells for all arguments j the index of that element of the partition W which contains the point j, i.e. the value of the function $s_W(j)$, $1\le j\le k$, in a point j is defined by the relation $j\in V_{s_W(j)}$. Let us introduce the expression

$$I_{n,k,W}(f(\ell)) = \frac{1}{k!}\sum_{(l_1,\dots,l_k)\colon 1\le l_j\le n,\ j=1,\dots,k} f_{l_1,\dots,l_k}\left(\xi_{l_1}^{(s_W(1))},\dots,\xi_{l_k}^{(s_W(k))}\right)$$
$$\text{for all } W\in\mathscr{W}(k).$$

An expression of the form $I_{n,k,W}(f(\ell))$, $W\in\mathscr{W}_k$, will be called a decoupled U-statistic with generalized decoupling. Given a partition $W=(U_1,\dots,U_s)\in\mathscr{W}_k$ let us call the number s, i.e. the number of the elements of this partition the rank both of the partition W and of the decoupled U-statistic $I_{n,k,W}(f(\ell))$ with generalized decoupling.

Now I formulate the following hypothesis. For all $k\ge 2$ and $2\le j\le k$ there exist some constants $C(k,j)>0$ and $\delta(k,j)>0$ such that for all $W\in\mathscr{W}_k$ a decoupled U-statistic $I_{n,k,W}(f(\ell))$ with generalized decoupling satisfies the inequality

$$P(\|I_{n,k,W}(f(\ell))\|>u)\le C(k,j)P\left(\|\bar I_{n,k}(f(\ell))\|>\delta(k,j)u\right)$$
$$\text{for all } 2\le j\le k \text{ if the rank of } W \text{ equals } j. \tag{D.16}$$

It will be proved by induction with respect to k that both relations (D.3) and (D.16) hold for U-statistics of order k. Let us observe that for $k = 2$ relation (D.3) follows from (D.5). Relation (D.16) also holds for $k = 2$, since in this case we have to consider only the case $j = k = 2$. Relation (D.16) also holds in this case with $C(2,2) = 1$ and $\delta(2,2) = 1$. Hence we can start our inductive proof with $k = 3$. First I prove relation (D.16).

In relation (D.16) the tail-distribution of decoupled U-statistics with generalized decoupling is compared with that of the decoupled U-statistic $\bar I_{n,k}(f(\ell))$ introduced in (D.2). Given the order k of these U-statistics it will be proved by means of a backward induction with respect to the rank j of the decoupled U-statistics $I_{n,k,W}(f(\ell))$ with generalized decoupling.

Relation (D.16) clearly holds for $j = k$ with $C(k,k) = 1$ and $\delta(k,k) = 1$. If we already know that these relations hold up to $k-1$, then we prove first relation (D.16) for generalized decoupling U-statistics of order k with respect to backward induction for the rank $2 \le j < k$.

For this goal the following observation will be made. If the rank j of a partition $W = (U_1,\dots,U_j)$ satisfies the relation $2 \le j \le k-1$, then it contains an element with cardinality strictly less than k and strictly greater than 1. For the sake of simpler notation let us assume that the element U_j of this partition is such an element, and $U_j = \{t,\dots,k\}$ with some $2 \le t \le k-1$. The investigation of general U-statistics of rank j, $2 \le j \le k-1$, can be reduced to this case by a reindexation of the arguments in the U-statistics if it is necessary. Let us consider the partition $\bar W = (U_1,\dots,U_{j-1},\{t\},\dots,\{k\})$ and the decoupled U-statistic $I_{n,k,\bar W}(f(\ell))$ with generalized decoupling corresponding to this partition $\bar W$. It will be shown that our inductive hypothesis implies the inequality

$$P(\|I_{n,k,W}(f(\ell))\| > u) \le \bar A(k) P\left(\|I_{n,k,\bar W}(f(\ell))\| > \bar\gamma(k)u\right) \tag{D.17}$$

with $\bar A(k) = \sup\limits_{2\le p\le k-1} A(p)$, $\bar\gamma(k) = \inf\limits_{2\le p\le k-1} \gamma(p)$ if the rank j of W is such that $2 \le j \le k-1$, where the constants $A(p)$ and $\gamma(p)$ agree with the corresponding coefficients in formula (D.3).

To prove relation (D.17) (where $U_j = \{t,\dots,k\}$ is the last element of the partition W) let us define the σ-algebra $\mathscr{F}$ generated by the random variables appearing in the first $t-1$ coordinates of these U-statistics, i.e. by the random variables $\xi_{l_j}^{s_W(j)}$,$1 \le j \le t-1$, and $1 \le l_j \le n$ for all $1 \le j \le t-1$. We have $2 \le t \le k-1$. By our inductive hypothesis relation (D.3) holds for U-statistics of order $p = k-t+1$, since $2 \le p \le k-1$. I claim that this implies that

$$P(\|I_{n,k,W}(f(\ell))\| > u|\mathscr{F}) \le A(k-t+1) P\left(\|I_{n,k,\bar W}(f(\ell))\| > \gamma(k-t+1)u|\mathscr{F}\right) \tag{D.18}$$

with probability 1. Indeed, by the independence properties of the random variables $\xi_l^{s_W(j)}$ (and $\xi_l^{s_{\bar W}(j)}$), $1 \le j \le k$, $1 \le l \le n$,

$$P(\|I_{n,k,W}(f(\ell))\| > u|\mathscr{F}) = P_{\xi_l^{s_W(j)},1\le j\le t-1}(\|I_{n,k,W}(f(\ell)\| > u)$$

and

$$\begin{aligned} &P\left(\|I_{n,k,\bar W}(f(\ell))\| > \gamma(k-t+1)u|\mathscr{F}\right) \\ &\quad = P_{\xi_l^{s_W(j)},1\le j\le t-1}(\|I_{n,k,\bar W} f(\ell)\| > \gamma(k-t+1)u), \end{aligned}$$

where $P_{\xi_l^{s_W(j)},1\le j\le t-1}$ denotes that the values of the random variables $\xi_l^{s_W(j)}(\omega)$, $1 \le j \le t-1$, $1 \le l \le n$, are fixed, and we consider the probability that the appropriate functions of these fixed values and of the remaining random variables $\xi^{s_W(j)}$ and $\xi^{s_{\bar W}(j)}$, $t \le j \le k$, satisfy the desired relation. These identities and the relation between the sets W and $\bar W$ imply that relation (D.18) is equivalent to the identity (D.3) for the generalized U-statistics of order $2 \le k-t+1 \le k-1$ with kernel functions

$$\begin{aligned} &f_{l_t,\dots,l_k}(x_t,\dots,x_k) \\ &\quad = \sum_{(l_1,\dots,l_{t-1}):\,1\le l_j\le n,\ 1\le j\le t-1} f_{l_1,\dots,l_k}(\xi_{l_1}^{s_W(1)}(\omega),\dots,\xi_{l_{t-1}}^{s_W(t-1)}(\omega),x_t,\dots,x_k). \end{aligned}$$

Relation (D.17) follows from inequality (D.18) if expectation is taken at both sides. As the rank of $\bar W$ is strictly greater than the rank of W, relation (D.17) together with our backward inductive assumption imply relation (D.16) for all $2 \le j \le k$.

Relation (D.16) implies in particular (with the applications of partitions of order k and rank 2) that the terms in the sum at the right-hand side of (D.5) satisfy the inequality

$$P\left(D_k\|I_{n,k,V}(f(\ell))\| > u\right) \le \bar C(k,j)P\left(\|\bar I_{n,k}(f(\ell))\| > \bar D_k u\right)$$

with some appropriate $\bar C_k > 0$ and $\bar D_k > 0$ for all $V \subset \{1,\dots,k\}$, $1 \le |V| \le k-1$. This inequality together with relation (D.5) imply that inequality (D.3) also holds for the parameter k.

In such a way we get the proof of relation (D.3) and its special case, relation (14.13). Let us prove formula (14.14) with its help first in the simpler case when the supremum of finitely many functions is taken. If $M < \infty$ functions $f_1,\dots,f_M$ are considered, then relation (14.14) for the supremum of the U-statistics and decoupled U-statistics with these kernel functions can be derived from formula (14.13) if it is applied for the function $f = (f_1,\dots,f_M)$ with values in the separable Banach space B_M which consists of the vectors $(v_1,\dots,v_M)$, $v_j \in B$, $1 \le j \le M$, and the norm $\|(v_1,\dots,v_M)\| = \sup\limits_{1\le j\le m} \|v_j\|$ is introduced in it. The application of formula (14.13) with this choice yields formula (14.14) for this supremum. Let us emphasize that the constants appearing in this estimate do not depend on the number M. (We took only $M < \infty$ kernel functions,

because with such a choice the Banach space B_M defined above is also separable.) Since the distribution of the random variables $\sup\limits_{1\le s\le M} \|I_{n,k}(f_s)\|$ converge to that of $\sup\limits_{1\le s<\infty} \|I_{n,k}(f_s)\|$, and the distribution of the random variables $\sup\limits_{1\le s\le M} \left\|\bar I_{n,k}(f_s)\right\|$ converge to that of $\sup\limits_{1\le s<\infty} \left\|\bar I_{n,k}(f_s)\right\|$ as $M\to\infty$, relation (14.14) in the general case follows from its already proved special case and a limiting procedure $M\to\infty$. $\square$

Remark. The above proved formula (D.3) can be slightly generalized. It also holds if the expressions $I_{n,k}(f(\ell))$ and $\bar I_{n,k}(f(\ell))$ appearing in this inequality are defined in a more general way. Namely, they are the random functions introduced in formulas (D.1) and (D.2), but the sequences $\xi_1,\dots,\xi_n$ and their independent copies $\xi_1^{(j)},\dots,\xi_n^{(j)}$ in these formulas are independent random variables which may also be non-identically distributed. Such a generalization can be proved without any essential change in the original proof.

References

1. R. Adamczak, Moment inequalities for U-statistics. Ann. Probab. **34**, 2288–2314 (2006)
2. M. Ajtai, J. Komlós, G. Tusnády, On optimal matchings. Combinatorica **4**(4), 259–264 (1984)
3. K. Alexander, The central limit theorem for empirical processes over Vapnik–Červonenkis classes. Ann. Probab. **15**, 178–203 (1987)
4. M.A. Arcones, E. Giné, Limit theorems for U-processes. Ann. Probab. **21**, 1494–1542 (1993)
5. M.A. Arcones, E. Giné, U-processes indexed by Vapnik–Červonenkis classes of functions with application to asymptotics and bootstrap of U-statistics with estimated parameters. Stoch. Process. Appl. **52**, 17–38 (1994)
6. G. Bennett, Probability inequality for the sum of independent random variables. J. Am. Stat. Assoc. **57**, 33–45 (1962)
7. A. Bonami, Étude des coefficients de Fourier des fonctions de $L^p(G)$. Ann. Inst. Fourier (Grenoble) **20**, 335–402 (1970)
8. C. Borell, On the integrability of Banach space valued Walsh polynomials, in *Séminaire de Probabilités XIII*. Lecture Notes in Mathematics, vol 721 (Springer, Berlin, 1979), pp. 1–3
9. V.H. de la Peña, E. Giné, in *Decoupling. From Dependence to Independence.* Springer Series in Statistics. Probability and Its Application (Springer, New York, 1999)
10. V.H. de la Peña, S. Montgomery–Smith, Decoupling inequalities for the tail-probabilities of multivariate U-statistics. Ann. Probab. **23**, 806–816 (1995)
11. R.L. Dobrushin, Gaussian and their subordinated fields. Ann. Probab. **7**, 1–28 (1979)
12. R.M. Dudley, Central limit theorems for empirical measures. Ann. Probab. **6**, 899–929 (1978)
13. R.M. Dudley, A course on empirical processes, in *Lecture Notes in Mathematics* vol 1097 (Springer, New York, 1984), pp. 1–142
14. R.M. Dudley, *Real Analysis and Probability.* (Wadsworth & Brooks, Pacific Grove, 1989)
15. R.M. Dudley, *Uniform Central Limit Theorems* (Cambridge University Press, Cambridge, 1998)
16. E.B. Dynkin, A. Mandelbaum, Symmetric statistics, Poisson processes and multiple Wiener integrals. Ann. Stat. **11**, 739–745 (1983)
17. P. Frankl, J. Pach, On the number of sets in null-t-design. Eur. J. Combin. **4**, 21–23 (1983)
18. E. Giné, A. Guillou, On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptive intervals. Ann. Inst. Henri Poincaré PR **37**, 503–522 (2001)
19. E. Giné, S. Kwapień, R. Latała, J. Zinn, The LIL for canonical U-statistics of order 2. Ann. Probab. **29**, 520–527 (2001)
20. E. Giné, R. Latała, J. Zinn, Exponential and moment inequalities for U-statistics, in *High Dimensional Probability II.* Progress in Probability, vol 47 (Birkhäuser, Boston, 2000), pp. 13–38

P. Major, *On the Estimation of Multiple Random Integrals and U-Statistics*, Lecture Notes in Mathematics 2079, DOI 10.1007/978-3-642-37617-7,

21. L. Gross, Logarithmic Sobolev inequalities. Am. J. Math. **97**, 1061–1083 (1975)
22. A. Guionnet, B. Zegarlinski, Lectures on Logarithmic Sobolev inequalities, in *Lecture Notes in Mathematics*, vol. 1801 (Springer, New York, 2003), pp. 1–134
23. D.L. Hanson, F.T. Wright, A bound on the tail probabilities for quadratic forms in independent random variables. Ann. Math. Stat. **42**, 52–61 (1971)
24. W. Hoeffding, A class of statistics with asymptotically normal distribution. Ann. Math. Stat. **19**, 293–325 (1948)
25. W. Hoeffding, Probability inequalities for sums of bounded random variables. J. Am. Math. Soc. **58**, 13–30 (1963)
26. K. Itô, Multiple Wiener integral. J. Math. Soc. Jpn. **3**, 157–164 (1951)
27. S. Janson, *Gaussian Hilbert Spaces* (Cambridge University Press, Cambridge, 1997)
28. E.L. Kaplan, P. Meier, Nonparametric estimation from incomplete data. J. Am. Stat. Assoc. **53**, 457–481 (1958)
29. R. Latała, Estimates of moments and tails of Gaussian chaoses. Ann. Probab. **34**, 2315–2331 (2006)
30. M. Ledoux, On Talagrand deviation inequalities for product measures. ESAIM Probab. Stat. **1**, 63–87 (1996). Available at http://www.emath./fr/ps/
31. M. Ledoux, The concentration of measure phenomenon, in *Mathematical Surveys and Monographs*, vol 89 (American Mathematical Society, Providence, 2001)
32. P. Major, Multiple Wiener–Itô integrals, in *Lecture Notes in Mathematics*, vol 849 (Springer, Berlin, 1981)
33. P. Major, On the tail behaviour of the distribution function of multiple stochastic integrals. Probab. Theory Related Fields **78**, 419–435 (1988)
34. P. Major, Asymptotic distributions for weighted U-statistics. Ann. Probab. **22**, 1514–1535 (1994)
35. P. Major, An estimate about multiple stochastic integrals with respect to a normalized empirical measure. Stud. Scientarum Mathematicarum Hung. **42**(3), 295–341 (2005)
36. P. Major, Tail behaviour of multiple random integrals and U-statistics. Probab. Rev. **2**, 448–505 (2005)
37. P. Major, An estimate on the maximum of a nice class of stochastic integrals. Probab. Theory Related Fields **134**, 489–537 (2006)
38. P. Major, A multivariate generalization of Hoeffding's inequality. Electron. Commun. Probab. **2**, (220–229) (2006)
39. P. Major, On a multivariate version of Bernstein's inequality. Electron. J. Probab. **12**, 966–988 (2007)
40. P. Major, L. Rejtő, Strong embedding of the distribution function under random censorship. Ann. Stat. **16**, 1113–1132 (1988)
41. P. Major, L. Rejtő, A note on nonparametric estimations. In the conference volume to the 65. birthday of Miklós Csörgő (1998), pp. 759–774
42. V.A. Malyshev, R.A. Minlos, *Gibbs Random Fields. Method of Cluster Expansion* (Kluwer, Academic, Dordrecht, 1991)
43. P. Massart, About the constants in Talagrand's concentration inequalities for empirical processes. Ann. Probab. **28**, 863–884 (2000)
44. H.P. Mc. Kean, Wiener's theory of non-linear noise, in *Stochastic Differential Equations*. SIAM-AMS Proceedings, vol. 6, (1973), pp. 197–209
45. E. Nelson, The free Markov field. J. Funct. Anal. **12**, 211–227 (1973)
46. I. Nourdin, G. Peccati, in *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality.* Cambridge Tracts in Mathematics, vol 192 (Cambridge University Press, Cambridge, 2012)
47. D. Nualart, *Malliavin Calculus and Related Topics of Probability and Its Applications*, 2nd edn. (Springer, Berlin, 2006)
48. G. Peccati, M.S. Taqqu, *Wiener Chaos: Moments, Cumulants and Diagrams* (Springer, New York, 2010)
49. D. Pollard, *Convergence of Stochastic Processes* (Springer, New York, 1984)

50. G.-C. Rota, C. Wallstrom, Stochastic integrals: a combinatorial approach. Ann. Probab. **25**(3), 1257–1283 (1997)
51. D. Surgailis, On multiple Poisson stochastic integrals and associated Markov semigroups. Probab. Math. Stat. **2**(3), 217–239 (1984)
52. D. Surgailis, Long-range dependence and Appell rank. Ann. Probab. **28**, 478–497 (2000)
53. G. Szegő, *Orthogonal Polynomials*. American Mathematical Society Colloquium Publications, vol. 23 (American Mathematical Society, Providence, RI, 1967)
54. A. Takemura, Tensor Analysis of ANOVA decomposition. J. Am. Stat. Assoc. **78**, 894–900 (1983)
55. M. Talagrand, Sharper bounds for Gaussian and empirical processes. Ann. Probab. **22**, 28–76 (1994)
56. M. Talagrand, New concentration inequalities in product spaces. Invent. Math. **126**, 505–563 (1996)
57. M. Talagrand, *Spin Glasses: A Challenge for Mathematicians* (Springer, Berlin, 2003)
58. M. Talagrand, in *The General Chaining*. Springer Monographs in Mathematics (Springer, Berlin, 2005)
59. V.N. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1995)
60. N. Wiener, The homogeneous chaos. Am. J. Math. **60**, 879–936 (1838)

Index

P. Major, *On the Estimation of Multiple Random Integrals and U-Statistics*,
Lecture Notes in Mathematics 2079, DOI 10.1007/978-3-642-37617-7,

LECTURE NOTES IN MATHEMATICS Springer

Edited by J.-M. Morel, B. Teissier; P.K. Maini

Editorial Policy (for the publication of monographs)

1. Lecture Notes aim to report new developments in all areas of mathematics and their applications - quickly, informally and at a high level. Mathematical texts analysing new developments in modelling and numerical simulation are welcome.
   Monograph manuscripts should be reasonably self-contained and rounded off. Thus they may, and often will, present not only results of the author but also related work by other people. They may be based on specialised lecture courses. Furthermore, the manuscripts should provide sufficient motivation, examples and applications. This clearly distinguishes Lecture Notes from journal articles or technical reports which normally are very concise. Articles intended for a journal but too long to be accepted by most journals, usually do not have this "lecture notes" character. For similar reasons it is unusual for doctoral theses to be accepted for the Lecture Notes series, though habilitation theses may be appropriate.

2. Manuscripts should be submitted either online at www.editorialmanager.com/lnm to Springer's mathematics editorial in Heidelberg, or to one of the series editors. In general, manuscripts will be sent out to 2 external referees for evaluation. If a decision cannot yet be reached on the basis of the first 2 reports, further referees may be contacted: The author will be informed of this. A final decision to publish can be made only on the basis of the complete manuscript, however a refereeing process leading to a preliminary decision can be based on a pre-final or incomplete manuscript. The strict minimum amount of material that will be considered should include a detailed outline describing the planned contents of each chapter, a bibliography and several sample chapters.
   Authors should be aware that incomplete or insufficiently close to final manuscripts almost always result in longer refereeing times and nevertheless unclear referees' recommendations, making further refereeing of a final draft necessary.
   Authors should also be aware that parallel submission of their manuscript to another publisher while under consideration for LNM will in general lead to immediate rejection.

3. Manuscripts should in general be submitted in English. Final manuscripts should contain at least 100 pages of mathematical text and should always include

   – a table of contents;
   – an informative introduction, with adequate motivation and perhaps some historical remarks: it should be accessible to a reader not intimately familiar with the topic treated;
   – a subject index: as a rule this is genuinely helpful for the reader.

   For evaluation purposes, manuscripts may be submitted in print or electronic form (print form is still preferred by most referees), in the latter case preferably as pdf- or zipped psfiles. Lecture Notes volumes are, as a rule, printed digitally from the authors' files. To ensure best results, authors are asked to use the LaTeX2e style files available from Springer's web-server at:

   ftp://ftp.springer.de/pub/tex/latex/svmonot1/ (for monographs) and
   ftp://ftp.springer.de/pub/tex/latex/svmultt1/ (for summer schools/tutorials).

Additional technical instructions, if necessary, are available on request from lnm@springer.com.

4. Careful preparation of the manuscripts will help keep production time short besides ensuring satisfactory appearance of the finished book in print and online. After acceptance of the manuscript authors will be asked to prepare the final LaTeX source files and also the corresponding dvi-, pdf- or zipped ps-file. The LaTeX source files are essential for producing the full-text online version of the book (see http://www.springerlink.com/openurl.asp?genre=journal&issn=0075-8434 for the existing online volumes of LNM). The actual production of a Lecture Notes volume takes approximately 12 weeks.

5. Authors receive a total of 50 free copies of their volume, but no royalties. They are entitled to a discount of 33.3 % on the price of Springer books purchased for their personal use, if ordering directly from Springer.

6. Commitment to publish is made by letter of intent rather than by signing a formal contract. Springer-Verlag secures the copyright for each volume. Authors are free to reuse material contained in their LNM volumes in later publications: a brief written (or e-mail) request for formal permission is sufficient.

Addresses:

Professor J.-M. Morel, CMLA,
École Normale Supérieure de Cachan,
61 Avenue du Président Wilson, 94235 Cachan Cedex, France
E-mail: morel@cmla.ens-cachan.fr

Professor B. Teissier, Institut Mathématique de Jussieu,
UMR 7586 du CNRS, Équipe “Géométrie et Dynamique”,
175 rue du Chevaleret
75013 Paris, France
E-mail: teissier@math.jussieu.fr

For the “Mathematical Biosciences Subseries” of LNM:

Professor P. K. Maini, Center for Mathematical Biology,
Mathematical Institute, 24-29 St Giles,
Oxford OX1 3LP, UK
E-mail : maini@maths.ox.ac.uk

Springer, Mathematics Editorial, Tiergartenstr. 17,
69121 Heidelberg, Germany,
Tel.: +49 (6221) 4876-8259

Fax: +49 (6221) 4876-8259
E-mail: lnm@springer.com