Introduction to U-statistics

Fabian Baier

August 2024

1 Proofs and Derivations

Important Definitions, Proofs and Derivations for chapter 1.1

Definition (parametric functional): Let $\mathcal{P} = \{F\}$ be a class of probability distributions on $(X, \mathcal{X})$. By $\theta(F)$, we denote a functional defined on F and taking values in $\mathbb{R}$. The functional $\theta(F)$ is called a parametric (regular) functional if there exists an unbiased estimate of it, i.e., $\theta(F)$ can be represented as

$$\theta(F) = \int \cdots \int \Phi(x_1, \ldots, x_m) F(dx_1) \ldots F(dx_m) \tag{1.1.1}$$

(expected value of the kernel is equal to the parametric functional)

with some function $\Phi : (x_1, \ldots, x_m) \to \Phi(x_1, \ldots, x_m)$ and some integer $m \geq 1$ which are called the kernel of $\theta(F)$ and the degree of $\theta(F)$, respectively. The kernel is assumed to be symmetric in its arguments for this whole presentation, Otherwise, we can introduce a symmetric kernel Φ_0 by

$$\Phi_0(x_1, \ldots, x_m) = \frac{1}{m!} \sum_{\sigma \in S_m} \Phi(x_{\sigma(1)}, \ldots, x_{\sigma(m)}),$$

where S_m is the set of all permutations of m elements.

Quick example:

If $\Phi(x_1, x_2) = x_1^2 - x_2$, then

$$\Phi_0(x_1, x_2) = \frac{1}{2} \left(x_1^2 + x_2^2 - x_2 - x_1 \right).$$

Now we got:

$$\Phi_0(x_{\sigma(1)}, \ldots, x_{\sigma(m)}) = \Phi_0(x_1, \ldots, x_m)$$

for all permutations σ, and for all $F \in \mathcal{P}$:

$$E_F[\Phi_0] = E_F[\Phi] \Rightarrow \Phi_0 \text{ is an unbiased estimator of the original } \Phi.$$

Definition (U-statistics):

Let Φ be a symmetric kernel function of m variables of a parametric functional θ (F). The U-statistic based on a sample of size n, $(\mathrm{X}_1, \ldots, X_n)$, $is\ defined\ as:$

$$U_n := \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < i_2 < \cdots < i_m \le n} \Phi(X_{i_1}, X_{i_2}, \dots, X_{i_m}) \tag{1.1.3}$$

where the summation is over all subsets of size m chosen from the n sample points.

Quick example:

For $m = 2$, $n = 3$, let

$$U_3 = \frac{1}{\binom{3}{2}} \sum_{1 \le i < j \le 3} \Phi(X_i, X_j)$$

which simplifies to

$$= \frac{1}{3} \left[\Phi(X_1, X_2) + \Phi(X_2, X_3) + \Phi(X_1, X_3) \right].$$

Canonical functions: Assume that $E|\Phi| < \infty$ and

$$\Phi_c(x_1, \dots, x_c) := E\Phi(x_1, \dots, x_c, X_{c+1}, \dots, X_m)$$

$$= E(\Phi(X_1, \dots, X_m)|X_1 = x_1, \dots, X_c = x_c), \quad c = 0, 1, \dots, m. \tag{1.1.4}$$

Let $\Phi_0 = \theta(P)$, $\Phi_m = \Phi$; one can easily see that

$$\Phi_c(x_1, \dots, x_c) = E\Phi_{c+1}(x_1, \dots, x_c, X_{c+1}) \tag{1.1.5}$$

for $1 \le c \le m - 1$. We set

$$\tilde{\Phi} := \Phi - \theta(P), \quad \tilde{\Phi}_c := \Phi_c - \theta(P), \quad 1 \le c \le m.$$

Define the functions

$$g_1(x_1) := \tilde{\Phi}_1(x_1),$$

$$g_2(x_1, x_2) := \tilde{\Phi}_2(x_1, x_2) - g_1(x_1) - g_1(x_2),$$

$$g_3(x_1, x_2, x_3) := \tilde{\Phi}_3(x_1, x_2, x_3) - \sum_{i=1}^{3} g_1(x_i) - \sum_{1 \le i < j \le 3} g_2(x_i, x_j),$$

$$\vdots$$

$$g_m(x_1, \dots, x_m) := \tilde{\Phi}_m(x_1, \dots, x_m) - \sum_{i=1}^{m} g_1(x_i) - \sum_{1 \le i < j \le m} g_2(x_i, x_j) - \cdots - \sum_{1 \le i_1 < \cdots < i_{m-1} \le m} g_{m-1}(x_{i_1}, \dots, x_{i_{m-1}}). \tag{1.1.6}$$

Functions defined as in (1.1.6) with the property of complete degeneracy are called canonical functions. The **property of complete degeneracy**:

$$Eg_1\left(X_1\right) = 0$$

$$Eg_2\left(x_1, X_2\right) = 0$$

$$Eg_m\left(x_1, \dots, x_{m-1}, X_m\right) = 0$$

Let's show this we use: We know that

$$E\left[E\left(\Phi|\mathcal{F}\right)\right] = E\left[\Phi\right] \quad \text{by the tower property}$$

and also that

$$E\left[\Phi\right] = \theta(F) = E\left[\Phi_0\right].$$

Derivation of canonical functions g_c in terms of the kernel Φ:
Let us express g_c explicitly in terms of Φ_c from (1.1.4). For this purpose, we define Dirac's δ-measure on a measurable space $(X, \mathcal{X})$

$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A \end{cases} \tag{1.1.9}$$

for any $x \in X$ and $A \subseteq X$.

With the help of (1.1.9), the kernel $\Phi(x_1, \dots, x_n)$ in (1.1.3) can be represented in the form

$$\Phi(x_1, \dots, x_m) = \int \cdots \int \Phi(y_1, \dots, y_m) \prod_{j=1}^{m} \delta_{x_j}(y_j). \tag{1.1.10}$$

Relation (1.1.10) enables us to use the multiplicative properties of the integrand in the following way. It is well-known that for arbitrary numbers $(a_1, \dots, a_m)$ and $(b_1, \dots, b_m)$,

$$\prod_{s=1}^{m}(a_s + b_s) = \prod_{s=1}^{m} a_s + \sum_{c=1}^{m} \sum_{1 \le j_1 < \cdots < j_c \le m} \prod_{s=1}^{m-c} b_{j_s} \prod_{s=1}^{c} a_{i_s}, \tag{1.1.11}$$

where $\{i_1, \dots, i_{m-c}\} = \{1, \dots, m\} \setminus \{j_1, \dots, j_c\}$.

Note that

$$d\delta_{x_j, y} = dP(y) + d(\delta_{x_j}(y) - P(y)), \quad j = 1, \dots, m.$$

In (1.1.11), we set

$$a_s = dP(y_s), \quad b_s = d(\delta_{x_s}(y_s) - P(y_s)), \quad s = 1, \dots, m,$$

and then insert this relation into (1.1.10). Then, taking into account notation (1.1.10) and the symmetry of the functions Φ_c, we obtain the formula

$$\Phi(x_1, \dots, x_m) = 0 + \sum_{c=1}^{m} \sum_{1 \le j_1 < \cdots < j_c \le m} \int \cdots \int \Phi(y_1, \dots, y_m) \prod_{s=1}^{c} d(\delta_{x_{j_s}}(y_{j_s}) - P(y_{j_s})). \tag{1.1.12}$$

Comparing (1.1.7) with (1.1.12), we find that for all $c = 1, \dots, m$,

$$g_c(x_1, \dots, x_c) = \int \cdots \int \Phi(y_1, \dots, y_c) \prod_{s=1}^{c} d(\delta_{x_{j_s}}(y_{j_s}) - P(y_{j_s})). \tag{1.1.13}$$

Since $E\delta_{x_s}(y_s) = P(y_s), s = 1, \dots, m$, the property of complete degeneracy (1.1.8) follows from (1.1.13) obviously. If we use formula (1.1.11) for

$$a_s = \delta_{x_s}(y_s), \quad b_s = -dP(y_s), \quad s = 1, \dots, m,$$

then expression (1.1.13) turns into

$$g_c(x_1, \dots, x_c) = \Phi_c(x_1, \dots, x_c) - \sum_{j=1}^{c} \Phi_{c-1}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_c)$$

$$+ \sum_{1 \le s_1 < s_2 \le c} \Phi_{c-2}(x_1, \dots, x_{s_1-1}, x_{s_1+1}, \dots, x_{s_2-1}, x_{s_2+1}, \dots, x_c) + (-1)^c \theta. \quad (1.1.14)$$

If we set

$$a_s = -dP(y_s), \quad b_s = \delta_{x_s}(y_s), \quad s = 1, \dots, m,$$

in (1.1.11), then (1.1.13) yields

$$g_c(x_1, \dots, x_c) = (-1)^c \theta + \sum_{d=1}^{c} (-1)^{c-d} \sum_{1 \le j_1 < \dots < j_d \le c} \Phi_d(x_{j_1}, \dots, x_{j_d}).$$

$$= \sum_{d=1}^{c} (-1)^{c-d} \sum_{1 \le j_1 < \dots < j_d \le c} \Phi_d(x_{j_1}, \dots, x_{j_d}). \quad (1.1.15)$$

Derivation Hoeffding representation:
The Hoeffding representation of U-statistics:
Let us write the U-statistic U_n with a kernel Φ in the form

$$U_n - \theta(F) = \binom{n}{m}^{-1} S_n, \quad (1.1.16)$$

where

$$S_n = \sum_{1 \le i_1 < \dots < i_m \le n} \Phi(X_{i_1}, \dots, X_{i_m}) - \theta, \quad (1.1.17)$$

For $c = 1, \dots, m$, we set

$$S_{nc} = \sum_{1 \le i_1 < \dots < i_c \le n} g_c(X_{i_1}, \dots, X_{i_c}). \quad (1.1.18)$$

Inserting (1.1.18) with $c = m$ into (1.1.17) and using:

$$\tilde{\Phi}_c(x_1, \dots, x_c) = \sum_{d=1}^{c} \sum_{1 \le i_1 < \dots < i_d \le c} g_d(x_{i_1}, \dots, x_{i_d}),$$

(1.1.7) We obtain

$$S_n = S_{nm} + \sum_{c=1}^{m-1} \binom{n}{m}^{-1} \sum_{1 \le i_1 < \dots < i_c \le n} g_c(X_{i_1}, \dots, X_{i_c}).$$

For $c = 1$, the term on the right-hand side can be written as follows

$$\sum_{1 \le i_1 < \cdots < i_n \le n} \sum_{s=1}^{m} g_1(X_s).$$

In this sum, each $g_1(X_s)$, $1 \le s \le n$, appears the same number of times. Since the sum contains $\binom{n}{m}$ terms, each term is encountered $n^{-1}\binom{n}{m}$ times.

Therefore, the sum $S_{n1} = \sum_{s=1}^{n} g_1(X_s)$ appears $\binom{n}{1}^{-1}\binom{n}{m}$ times. By repeating this argument for $c = 2, \dots, m-1$, we obtain

$$S_n = \sum_{c=1}^{m} \binom{n}{c} \binom{n}{m}^{-1} S_{nc}.$$

This and (1.1.16) imply the following formula

$$U_n - \theta = \sum_{c=1}^{m} \binom{n}{c} \binom{n}{m}^{-1} S_{nc}, \tag{1.1.19}$$

which is called the Hoeffding representation (or canonical decomposition) of a U-statistic.

Definition rank of a U-statistic:

Let $r \ge 1$ be the first integer for which the relations

$$g_1 = \cdots = g_{r-1} = 0, \quad g_r \ne 0 \tag{1.1.20}$$

hold. It is obvious that r takes values $1, \dots, m$. The integer r satisfying (1.1.20) is called the rank of a U-statistic (or of a kernel Φ). If $r = 1$, a U-statistic (or a kernel Φ) is called nondegenerate. If $r \ge 2$, then a U-statistic (or a kernel Φ) is called degenerate, and r is called the order of degeneracy. If $r = m$, we say that the kernel Φ possesses the property of complete degeneracy (e.g. the canonical functions possess this property).
We define

$$U_{nc} := \binom{n}{c}^{-1} \sum_{1 \le i_1 < \cdots < i_c \le n} g_c(X_{i_1}, \dots, X_{i_c})$$

(1.1.21) then the Hoeffding representation of a U-statistic can be written as:

$$U_n - \theta = \sum_{c=r}^{m} \binom{m}{c} U_{nc}$$

(1.1.22) , i.e. any U-statistic (1.1.3) is a linear combination of U -statistics (1.1.21), whose kernels have the property of complete degeneracy (1.1.8), and r denotes the rank from (1.1.20). **Lemma 1.1.1:**
Suppose that $E|\Phi| < \infty$, $\mathcal{F}_k = \sigma(\omega; X_1, \dots, X_k)$ is a σ-algebra generated by $X_1, \dots, X_k$, and that $\mathcal{F}_0 = \{\emptyset, X\}$. Then, for every $c = 1, \dots, m$,

$$E(S_{nc}|\mathcal{F}_k) = S_{kc}, \quad c \le k \le n,$$

(1.1.25) **Proof:**

Definition of a martingale:

(M1) $S_{n,c}$ is $\mathcal{F}_n$-measurable.

(M2) $E\left[|S_{n,c}|\right] < \infty$ for all n.

(M3) $E\left[S_{n,c}|\mathcal{F}_k\right] = S_{k,c}$.

The functions g_c are symmetric with respect to their arguments $x_1, \ldots, x_c$ and possess the **property of complete degeneracy**. According to this property , we have

$$E\left(g_c\left(X_{i_1}, \ldots, X_{i_c}\right) \mid X_1, \ldots, X_k\right) = 0$$

if at least one of the indices $i_1, \ldots, i_c$ does not belong to $\{1, \ldots, k\}$. For example, if $i_1 \notin \{1, \ldots, k\}$, then

$$\begin{aligned}
&E\left(g_c\left(X_{i_1}, \ldots, X_{i_c}\right) \mid X_1, \ldots, X_k\right) \\
&= E\left\{E\left[g_c\left(X_{i_1}, \ldots, X_{i_c}\right)\middle| X_1, \ldots, X_k, X_{i_2}, \ldots, X_{i_c}\right| \mid X_1, \ldots, X_k\right\} \\
&= E\left\{E\left[g_c\left(X_{i_1}, \ldots, X_{i_c}\right) \mid X_{i_2}, \ldots, X_{i_c}\right] \mid X_1, \ldots, X_k\right\} \\
&= E\left\{0 \mid X_1, \ldots, X_k\right\} = 0
\end{aligned}$$

We show it for the case c=2, m=2 the rest follows analogously: Let for example be $i_1 = 3$, i.e. we want to show

$$E\left[g_2(X_3, X_2)\right] = 0 \quad \text{(case } m = 2, c = 2)$$

$$E\left[g_2(X_3, X_2)\right] = E\left[E\left[\Phi(X_1, X_2)|X_1 = X_3, X_2 = X_2\right]\right]$$

$$-E\left[E\left[\Phi(X_1, X_2)|X_1 = X_3\right]\right] - E\left[E\left[\Phi(X_1, X_2)|X_2 = X_2\right]\right]$$

$-\theta + \theta + \theta = 0$ by property of complete degeneracy and by the fact that X_i are i.i.d.

hence

$$E\left[\Phi(X_3, X_2)\right] = E\left[\Phi(X_1, X_1)\right] = \theta.$$

Thus,

$$E\left(S_{nc} \mid X_1, \ldots, X_k\right) = \sum_{1 \le i_1 < \ldots < j_c \le k} g_c\left(X_{i_1}, \ldots, X_{i_c}\right) = S_{kc}$$

q.e.d.

Lemma 1.1.2:

Assume that $E|\Phi| < \infty$. Then the U-statistic U_n can be represented in the form of the sum of martingale-differences. In particular, for $m = 2$, we have in (1.1.27)

$$\xi_{nk} = 2n^{-1} g_1(X_k) + 2n^{-1}(n-1)^{-1} \sum_{j=1}^{k-1} g_2(X_j, X_k).$$

Proof:

We obtain the representation for the U-statistic U_n from (1.1.23):

$$U_n - \theta = \xi_{n1} + \xi_{n2} + \cdots + \xi_{nm}, \tag{1.1.26}$$

where ξ_{nk} is defined as:

$$\xi_{nk} = E(U_n|X_1, \ldots, X_k) - E(U_n|X_1, \ldots, X_{k-1}).$$

We need to express this in terms of the sum of martingale differences. From (1.1.19) and (1.1.25), we can represent the conditional expectation of U_n as:

$$E(U_n|X_1, \ldots, X_k) = \theta + \sum_{c=1}^{m} \binom{n}{c} \binom{n}{m}^{-1} S_{kc},$$

where S_{kc} represents the sum over the g_c-functions applied to subsets of size c from $X_1, \ldots, X_k$.

Thus, we can write the martingale difference ξ_{nk} as:

$$\xi_{nk} = \sum_{c=1}^{m} \binom{n}{c} \binom{n}{m}^{-1} (S_{kc} - S_{k-1,c}).$$

This gives the decomposition of ξ_{nk} in terms of the differences in sums over functions g_c. We denote k=1:n, c=1:m

$$\eta_{kc} = \sum_{1 \le i_1 < \cdots < i_c \le k-1} g_c(X_{i_1}, \ldots, X_{i_{c-1}}, X_k),$$

with $\eta_{k1} = g_1(X_k)$, which represents the contribution of the k-th sample to the sum S_{kc}.

Since

$$S_{kc} - S_{k-1,c} = \eta_{kc}.$$

Substituting this into the expression for ξ_{nk}, we obtain:

$$\xi_{nk} = \sum_{c=1}^{m} \binom{n}{c} \binom{n}{m}^{-1} \eta_{kc}. \tag{1.1.27}$$

This shows that the sequence $\{\xi_{nk}, \mathcal{F}_k\}_{k \ge 1}$, where $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$, forms a martingale-difference. q.e.d.

Lemma 1.1.3: The sequence U_n is a reversed martingale w.r.t. the sequence (B_n)= (σ (U_n,U_{n+1}, ...)) n=1, 2, ... of σ-algebras .

Proof:

We want to show that U_n is a reversed martingale.

Definition (reversed martingale): A sequence (ξ_n) is a reversed martingale if it satisfies the following:

(M1) ξ_n is B_n measurable

(M2) $E|\xi_n| < \infty$ for all n

(M3) $E(\xi_n|B_{n+1})= \xi_{n+1}$ for all n

We already know that (M1) and (M2) are satisfied, so it is only left to proof (M3) for reversed martingales. M1 is satisfied because the σ-algebras are chosen in a way that U_n is B_n measurable. Any permutation of the variables $X_1, X_2, \dots, X_n$ does not change the sequence $U_n, U_{n+1}, \dots$, since a U statistic is a symmetric function (since the kernel Φ is symmetric). This fact together with the independence of $X_i, i = 1, \dots, n$, for any $\left(i_1, \dots, i_m\right)$ such that $1 \leq i_1 < \dots < i_m \leq n$, leads to the relation

$$E\left(\Phi\left(X_{i_1}, \dots, X_{i_m} \mid B_n\right) = E\left(\Phi\left(X_1, \dots, X_m\right) \mid B_n\right) (1)$$

Note that $\Phi(X_{i_1}, .., X_{i_m})$ is not B_n-measurable. Carrying out the summation over all the indices $1 \leq i_1 < \dots < i_m \leq n$ on the both sides of (1) and seeing that the right hand side doesn't depend on the the summation indices, we obtain

$$E\left(U_n \mid B_n\right) = E\left(\Phi\left(X_1, \dots, X_m\right) \mid B_n\right)$$

Since $E\left(U_n \mid B_n\right) = U_n$, because we know U_n is B_n measurable and the property of conditional expectation, we have

$$U_n = E\left(\Phi\left(X_1, \dots, X_m\right) \mid \boldsymbol{B}_n\right) (2)$$

for all $n = m, m + 1, \dots$

Let us take the conditional expectation with respect to the σ-algebra $\mathfrak{B}_{n+1}$ in (2). Then

$$\begin{aligned} E\left(U_n \mid B_{n+1}\right) &= E\left(E\left(\Phi\left(X_1, \dots, X_m\right) \mid B_n\right) \mid B_{n+1}\right) \\ &= E\left(\Phi\left(X_1, \dots, X_m\right) \mid B_{n+1}\right) = U_{n+1} \end{aligned}$$

where the second equality follows from the property of conditional expectation and the inclusion $B_n \supseteq B_{n+1}$ since we know that the **smaller σ-algebra wins**, which in the case of a reverse martingale is the σ-algebra of knowledge from everything after time n, so n+1 is a smaller σ-algebra (by Probability Theory lecture). Hence, by virtue of (2) we obtain

$$E\left(U_n \mid B_{n+1}\right) = U_{n+1}$$

This shows U_n is a reverse martingale. q.e.d.

Definition Dispersion of a U-statistic:

$$\sigma^2(U_n) = \binom{n}{m}^{-1} \sum_{c=1}^{m} \binom{m}{c} \binom{n-m}{m-c} \eta_c. \tag{1.1.34}$$

with η_c being:

$$E(\tilde{\Phi}(X_{i_1}, \dots, X_{i_m})\tilde{\Phi}(X_{j_1}, \dots, X_{j_m})) =: \eta_c$$

and

$$\delta_c := Eg_c^2$$

Lemma 1.1.4: Let r be the rank of a U-statistic defined by (1.1.20). We set

$$\gamma_r(m,n) := \binom{n}{m}^{-1} \sum_{c=r}^{m} \binom{m-1}{c-1}\binom{n-m}{m-c}.$$

The following relations hold:

$$\eta_c \le \frac{c}{d}\eta_d \quad \text{for } 1 \le c < d \le m, \tag{1.1.39}$$

$$\frac{m^2}{n}\eta_1 \le \sigma^2(U_n) \le \frac{m}{n}\eta_m, \tag{1.1.40}$$

The variable $n\sigma^2(U_n)$ is decreasing with respect to n, i.e.,

$$(n+1)\sigma^2(U_{n+1}) \le n\sigma^2(U_n), \tag{1.1.41}$$

and furthermore,

$$\frac{m}{d}\gamma_r(m,n)\eta_d \le \sigma^2(U_n) \le \gamma_r(m,n)\eta_m. \tag{1.1.42}$$

Proof:

$$\eta_c = \delta_c + \binom{c}{1}\delta_{c-1} + \ldots + \binom{c}{c-1}\delta_1 (1.1.37)$$

for $c = 1, \ldots, m$. For $c \le d$, we can write, using (1.1.37),

$$\begin{aligned} c\eta_d - d\eta_c &= c\sum_{i=1}^{d}\binom{d}{i}\delta_i - d\sum_{i=1}^{c}\binom{c}{i}\delta_i \\ &= \sum_{i=1}^{c}\left[c\binom{d}{i} - d\binom{c}{i}\right]\delta_i + c\sum_{i=c+1}^{d}\binom{d}{i}\delta_i \end{aligned}$$

and this implies (1.1.39), since $c\binom{d}{i} - d\binom{c}{i} \ge 0$ for $1 \le i \le c \le d$.
From (1.1.39) we obtain

$$c\eta_1 \le \eta_c \le \frac{c}{m}\eta_m$$

for $c = 1, \ldots, m$.
Applying these inequalities to each term in (1.1.34) and using the following identity:

$$\left(\binom{n}{m}\right)^{-1} \sum_{c=1}^{m} c \cdot \binom{m}{c} \cdot \binom{n-m}{m-c} = \frac{m^2}{n}$$

we get (1.1.40)-(1.1.42). q.e.d.
Definition Projection of a U-statistic:
Assume E|Φ|< ∞. Then the projection of a U-statistic is defined as

$$\hat{U}_n := \sum_{i=1}^{n} E(U_n|X_i) - (n-1)\theta, \tag{1.1.45}$$

In terms of Φ_1 we have:

$$\hat{U}_n - \theta = \frac{m}{n} \sum_{j=1}^{n} \tilde{\Phi}_1(X_j)$$

(1.1.46) In the general case where η_0=...=η_{r-1}=0<η_r we get the following definition:

$$\hat{U}_n - \theta = \sum_{1 \leq i_1 < \cdots < i_r \leq n} E(U_n | X_{i_1}, \ldots, X_{i_r}) - \binom{n}{r}\theta$$

(1.1.47). Applying the Hoeffding representation (1.1.19) and the martingale relation (1.1.25)

$$E(S_{nc}|\mathcal{F}_k) = S_{kc}, \quad c \leq k \leq n,$$

we can rewrite the general definition as:

$$\hat{U}_n - \theta = \binom{m}{r}\binom{n}{r}^{-1} S_{nr}$$

(1.1.48) since S_{n0}=...=$S_{n,r-1}$=0, and for r=1 we get (1.1.46).

The Hoeffding Formula: Let $k = \left\lfloor \frac{n}{m} \right\rfloor$ be the integral part of the number n/m. We set

$$h(x_1, \ldots, x_n) := k^{-1}\big(\Phi(x_1, \ldots, x_m) + \Phi(x_{m+1}, \ldots, x_{2m}) + \cdots + \Phi(x_{km-m+1}, \ldots, x_{km})\big). \tag{1.1.49}$$

Denote by $\sum_{\sigma \in S_n}$ the summation over all $n!$ permutations $(i_1, \ldots, i_n)$ of the numbers $(1, \ldots, n)$ and $\sum_c$ summation over all $\binom{n}{m}$ combinations $(i_1, \ldots, i_m)$ of $(1, \ldots, n)$, respectively. Then by using the symmetry of the kernel Φ we get:

$$k \sum_{\sigma \in S_n} h(x_{\sigma(i_1)}, \ldots, x_{\sigma(i_n)}) = km!(n-m)! \sum_c \Phi(x_{i_1}, \ldots, x_{i_m}),$$

or using (1.1.3):

$$\sum_{\sigma \in S_n} h(X_{\sigma(i_1)}, \ldots, X_{\sigma(i_n)}) = m!(n-m)!\binom{n}{m} U_n.$$

Thus,

$$U_n = \frac{1}{n!} \sum_{\sigma \in S_n} h(X_{\sigma(i_1)}, \ldots, X_{\sigma(i_n)}), \tag{1.1.50}$$

i.e. U_n is represented as an average of $n!$ terms, each being, in turn, an average of k independent identically distributed random variables.

Important Definitions and Derivations for chapter 1.3 :

Definition Von Mises' functionals (or von Mises' statistics) :

$$V_n = n^{-m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \Phi(X_{i_1}, \dots, X_{i_m}). \tag{1.3.1}$$

With the help of the Dirac measure, we define the empirical measure P_n by the formula

$$P_n(A) = \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j}(A), \quad A \in \mathcal{X}. \tag{1.3.2}$$

Then we can write (1.3.1) in the integral form

$$V_n = \int \cdots \int \Phi(x_1, \dots, x_m) dP_n(x_1) \cdots dP_n(x_m). \tag{1.3.3}$$

Consequently, $V_n = \theta(P_n)$, where the functional $\theta(F)$ is defined by (1.1.1). Though P_n is an unbiased estimate of F, the functional V_n is not an unbiased estimate of $\theta(F)$.
Canonical decompositon: We set

$$V_{nc} := \int \cdots \int \Phi_c(x_1, \dots, x_c) \prod_{j=1}^{c} d(P_n(x_j) - P(x_j)) \tag{1.3.6}$$

for $c = 1, \dots, m$; in notation (1.1.13):

$$g_c(x_1, \dots, x_c) = \int \cdots \int \Phi_c(y_1, \dots, y_c) \prod_{s=1}^{c} d(\delta_{x_s}(y_s) - P(y_s)).$$

this equality takes the form

$$V_{nc} = 1/n^c \sum_{i_1=1}^{n} \cdots \sum_{i_c=1}^{n} g_c(X_{i_1}, \dots, X_{i_c}). \tag{1.3.7}$$

For every $c = 1, \dots, m$, V_{nc} is a von Mises' functional with the kernel g_c possessing the property of complete degeneracy.

With the help of (1.1.11), we obtain

$$\theta(P_n) - \theta(P) = \sum_{c=1}^{m} \binom{m}{c} V_{nc}. \tag{1.3.8}$$

Here, we also take into account that $V_n = \theta(P_n)$.
The representation of von Mises' functional in the form of a linear combination of U-statistics:
For $k = 1, \dots, m$, let us consider the U-statistics

$$U_{nk} = \binom{n}{k}^{-1} \sum_{1 \le i_1 < \cdots < i_k \le n} \Phi_{mk}(X_{i_1}, \dots, X_{i_k}). \tag{1.3.9}$$

with symmetric kernels

$$\Phi_{mk}(x_1, \dots, x_k) = \sum_{\substack{v_1+\dots+v_k=m \\ v_j \ge 1}} \frac{m!}{v_1! \dots v_k!} \Phi(x_1^{v_1}, \dots, x_k^{v_k}), \tag{1.3.10}$$

where $x_j^{v_j} = (x_j, \dots, x_j)$ is a vector with v_j coordinates. Note that

$$\Phi_{mm}(x_1, \dots, x_m) = m!\Phi(x_1, \dots, x_m), \quad \Phi_{m1}(x_1) = \Phi(x_1, \dots, x_1).$$

Then the following representation is valid for $n \ge m$

$$V_n = \sum_{k=1}^{m} \binom{n}{k} n^{-m} U_{nk}. \tag{1.3.11}$$

Assume that $\theta_{mk} = \mathbb{E}\Phi_{mk}(X_1, \dots, X_k)$, $k = 1, \dots, m$, and

$$\tilde{\Phi}_{mk} = \Phi_{mk} - \theta_{mk}.$$

If $\tilde{U}_{nk}$ is a U-statistic with the kernel $\tilde{\Phi}_{mk}$, then (1.3.11) turns into

$$V_n - EV_n = \sum_{k=1}^{m} \binom{n}{k} n^{-m} \tilde{U}_{nk}. \tag{1.3.12}$$

The validity of (1.3.11) and (1.3.12) follows from the simple combinatorial formula

$$\sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \Phi(x_{i_1}, \dots, x_{i_m}) = \sum_{k=1}^{\min(n,m)} \sum_{\Sigma_1} \sum_{\Sigma_2} \frac{m!}{v_1! \dots v_k!} \Phi\left(\underbrace{x_{i_1}, \dots, x_{i_1}}_{v_1 \text{ times}}, \dots, \underbrace{x_{i_k}, \dots, x_{i_k}}_{v_k \text{ times}} \right). \tag{1.3.13}$$

where the summation in $\sum_{\Sigma_1}$ is carried out over all indices $1 \le i_1 < \dots < i_k \le n$, $1 \le k \le \min(n, m)$, and the summation in $\sum_{\Sigma_2}$ is carried out over all vectors $(v_1, \dots, v_k)$ such that $v_1 + \dots + v_k = m$, $v_j \ge 1$, $j = 1, \dots, m \ge 1$, and $m \ge 1$.

In (1.3.11) and (1.3.12), the relation (1.3.13) is used in that case $n \ge m$, i.e., $\min(n, m) = m$.