



HAL
open science

Linking the Hoeffding–Sobol and Möbius formulas through a decomposition of Kuo, Sloan, Wasilkowski, and Wozniakowski

Cécile Mercadier, Olivier Roustant, Christian Genest

► **To cite this version:**

Cécile Mercadier, Olivier Roustant, Christian Genest. Linking the Hoeffding–Sobol and Möbius formulas through a decomposition of Kuo, Sloan, Wasilkowski, and Wozniakowski. 2021. hal-03220809v2

HAL Id: hal-03220809

<https://hal.science/hal-03220809v2>

Preprint submitted on 31 Aug 2021 (v2), last revised 16 Feb 2022 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LINKING THE Hoeffding–Sobol AND MÖBIUS FORMULAS THROUGH A DECOMPOSITION OF KUO, SLOAN, WASILKOWSKI, AND WOŹNIAKOWSKI

CÉCILE MERCADIER*, OLIVIER ROUSTANT †, AND CHRISTIAN GENEST‡

Abstract. Extensions of a result of Kuo et al. (*Math. Comp.*, 79 (2010), 953–966) are presented which unify the derivation of the Hoeffding–Sobol and Möbius decompositions of a multivariate function as a sum of terms of increasing complexity.

Key words. Functional representation, Hoeffding–Sobol decomposition, Möbius decomposition

AMS subject classifications. Primary 41A63, Secondary 46N30

1. Introduction. The Hoeffding–Sobol and Möbius formulas are two common ways of expressing a real-valued function f of $d \geq 2$ variables acting on a domain $D \subseteq \mathbb{R}^d$ into a sum running over all subsets of $\{1, \dots, d\}$, viz.

$$(1.1) \quad f = \sum_{u \subseteq \{1, \dots, d\}} f_u .$$

Both formulas are well-known and useful in statistics because the terms in representation (1.1) become gradually simpler (in a specific sense) as the size $|u|$ of the set u decreases from d to zero. While many might have suspected that the two decompositions have a common origin, this suspicion does not appear to have been ever confirmed, possibly because the two formulas have been developed and used by nearly disjoint research communities. It is the purpose of this note to elucidate this connection.

The Hoeffding–Sobol decomposition goes back to the seminal work of Hoeffding on the theory of U -statistics [9], and by analogy with analysis of variance, Sobol [22] used it to derive sensitivity indices. This decomposition, which will be recalled in Section 2, has found many applications in statistics, notably in global sensitivity analysis; see, e.g., [2, 4, 10, 11, 17, 20, 21]. In particular, the Hoeffding–Sobol decomposition has been used extensively to compare the performance of Monte Carlo versus quasi-Monte Carlo integration in high-dimensional integration problems, where the notion of effective dimension reduces the dimension of the domain of integration; see, e.g., [14, 18, 19]. More recently, the Hoeffding–Sobol decomposition has also been used to define and study the tail dependograph, a graphical tool which captures the structure of asymptotic dependence within a multivariate random phenomenon; see [15, 16].

By comparison, the Möbius decomposition is relatively unknown and should not be confused with the celebrated Möbius inversion formula. The Möbius decomposition was originally suggested by Deheuvels [3] as a way to construct rank-based tests of independence among the components of a continuous random vector. More specifically, an application of Möbius’ formula leads to a representation of an empirical copula, and hence also of the empirical copula process, into a finite number of components which, under the null hypothesis of independence, are asymptotically independent. Cramér–von Mises statistics derived from the sub-processes prove to be very powerful, both asymptotically and in finite samples [6, 7, 8]. This approach has since been extended to the problem of testing for independence between random vectors [1, 12], and more recently to testing for dependence between arbitrary random variables [5].

In 2010, Kuo et al. [13] gave conditions which guarantee the existence and uniqueness of the general decomposition (1.1) and described how their result applies to the Hoeffding–Sobol formula. These findings

*Université de Lyon, Université Claude-Bernard Lyon 1, Institut Camille-Jordan, UMR CNRS 5208, 43, boulevard du 11 novembre 1918, F-69622 Villeurbanne, France. (mercadier@math.univ-lyon1.fr, <http://math.univ-lyon1.fr/~mercadier/>).

†Institut de Mathématiques de Toulouse, Université de Toulouse, INSA, Toulouse, 31077 cedex 4, France. (roustant@insa-toulouse.fr, <https://olivier-roustant.fr>).

‡Department of Mathematics and Statistics, McGill University, 805, rue Sherbrooke ouest, Montréal (Québec) Canada H3A 0B9. (christian.genest@mcgill.ca, <https://www.math.mcgill.ca/cgenest/>).

will be summarized in Section 2. In Section 3, it will then be shown that the Möbius formula is also of the form (1.1) but that its existence and uniqueness do not follow from the result of Kuo et al. [13], whose conditions are too restrictive. This will lead us, in Sections 4 and 5, to state and prove two extensions of the result by Kuo et al. [13] which encompasses both the Hoeffding–Sobol and Möbius decompositions as special cases. Finally, Section 6 will highlight the intrinsic differences between the Hoeffding–Sobol and Möbius decompositions when applied to a multivariate dependence function, namely a copula.

2. A general decomposition formula. Kuo et al. [13] give conditions to ensure both the existence and uniqueness of the decomposition (1.1). They rely on projections P_1, \dots, P_d which induce the decomposition upon setting, for each $u \subseteq \{1, \dots, d\}$,

$$(2.1) \quad f_u = \left(\prod_{i \in u} (I_D - P_i) \prod_{i \notin u} P_i \right) (f).$$

Their result holds provided that each projection eliminates the dependence on a specific input variable to which the function is applied.

To describe their finding, let us introduce some notation. Let \mathcal{F} be the linear space of real-valued functions acting on a domain $D \subseteq \mathbb{R}^d$ and for each $i \in \{1, \dots, d\}$, let $P_i : \mathcal{F} \rightarrow \mathcal{F}$ be a projection, i.e., a linear and idempotent operator, the latter term meaning that the composition of P_i with itself, denoted $P_i P_i$, returns P_i . Further assume that these projections are commuting, i.e., for any distinct $i, j \in \{1, \dots, d\}$, $P_i P_j = P_j P_i$, and let I_D denote the identity operator on \mathcal{F} . For each set $u \subseteq \{1, \dots, d\}$, let \mathcal{F}_u denote the class of functions $f \in \mathcal{F}$ such that $f(x_1, \dots, x_d) = f(y_1, \dots, y_d)$ when $x_j = y_j$ for all $j \in u$. Elements of \mathcal{F}_u are then functions $f \in \mathcal{F}$ whose value at $\mathbf{x} = (x_1, \dots, x_d) \in D$ depends only on the coordinates of \mathbf{x} whose index is in u , and hence \mathcal{F}_\emptyset is the set of constant functions. To simplify notation, write $\mathcal{F}_{-i} = \mathcal{F}_{\{1, \dots, d\} \setminus \{i\}}$ for all $i \in \{1, \dots, d\}$ and, for every set $u \subseteq \{1, \dots, d\}$, write

$$P_u = \prod_{i \in u} P_i \quad \text{and} \quad P_{-u}(f) = P_{\{1, \dots, d\} \setminus u}(f) = \prod_{i \notin u} P_i.$$

The key assumption in Theorem 2.1 of Kuo et al. [13], formally stated below, means that for any $i \in \{1, \dots, d\}$ and $f \in \mathcal{F}$, the value of $P_i(f)$ at $\mathbf{x} \in D$ does not depend on the i th coordinate of \mathbf{x} and that if f does not depend at all on this coordinate, the projection P_i leaves the function unchanged.

Assumption 2.1. For all $f \in \mathcal{F}$ and $i \in \{1, \dots, d\}$, $P_i(f) \in \mathcal{F}_{-i}$ and $f \in \mathcal{F}_{-i} \Rightarrow P_i(f) = f$.

In what follows, for any set $u \subseteq \{1, \dots, d\}$ and vector $\mathbf{x} \in D$, \mathbf{x}_u stands for a vector whose length is the cardinality $|u|$ of u and whose components are those of \mathbf{x} whose indices are in u .

THEOREM 2.2 (Theorem 2.1 of Kuo, Sloan, Wasilkowski, and Woźniakowski [13]). *Let \mathcal{F} be a linear space of real-valued functions acting on a domain $D \subseteq \mathbb{R}^d$ and let P_1, \dots, P_d be commuting projections on \mathcal{F} that satisfy Assumption 2.1. Then the following statements hold true.*

Part A: *Any function $f \in \mathcal{F}$ can be expressed in the form (1.1), where the function f_u defined in (2.1) is such that for any vector $\mathbf{x} \in D$, $f_u(\mathbf{x})$ only depends on \mathbf{x}_u .*

Part B: *Suppose that every $f \in \mathcal{F}$ can be written in the form (1.1) and that $f_u \in \mathcal{F}_u$ for every set $u \subseteq \{1, \dots, d\}$. Further assume that, for all $i \in \{1, \dots, d\}$,*

$$(2.2) \quad P_i(f_u) = \begin{cases} 0 & \text{if } i \in u, \\ f_u & \text{if } i \notin u. \end{cases}$$

When this annihilating property holds, the term f_u given in (2.1) satisfies the recursive relation

$$(2.3) \quad f_u = P_{-u}(f) - \sum_{v \subsetneq u} f_v.$$

Furthermore, one has

$$(2.4) \quad f_u = \sum_{v \subseteq u} (-1)^{|u \setminus v|} P_{-v}(f).$$

As mentioned by Kuo et al. [13], the Hoeffding–Sobol decomposition is a special case of Theorem 2.2. To be specific, let $\mathcal{F} = L^2(D)$ be the collection of square integrable functions on $D = [0, 1]^d$. For each $i \in \{1, \dots, d\}$, let the projection P_i be defined through integration with respect to the i th variable. That is, for every function $f \in \mathcal{F}$ and vector $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$, set

$$(2.5) \quad P_i(f)(\mathbf{x}) = \int_0^1 f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_d) dz.$$

Equivalently, the projection P_i could be defined as an expectation with respect to the i th component of a vector $\mathbf{Z} = (Z_1, \dots, Z_d)$ of mutually independent random variables uniformly distributed on the interval $[0, 1]$. Indeed, for every function $f \in \mathcal{F}$ and vector $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$,

$$P_i(f)(\mathbf{x}) = \mathbb{E}[f(x_1, \dots, x_{i-1}, Z_i, x_{i+1}, \dots, x_d)].$$

It is clear that the operators P_1, \dots, P_d so defined are commuting projections that satisfy Assumption 2.1. The well-known Hoeffding–Sobol decomposition then results upon injecting these projections into formula (2.1), as ensured by Part A of Theorem 2.2. Indeed, given that the terms in expression (2.1) are mutually orthogonal in the L^2 -sense, it is easily checked that

$$\text{var}[f(\mathbf{Z})] = \sum_{u \subseteq \{1, \dots, d\}} \text{var}[f_u(\mathbf{Z})],$$

so that the total variance of $f(\mathbf{Z})$ decomposes as a sum of contributive parts. Moreover, given that these operators also satisfy the annihilating property (2.2), formulas (2.3) and (2.4) yield the same decomposition by virtue of Part B of Theorem 2.2.

3. A look at the Möbius decomposition. Turning to the Möbius decomposition introduced by Deheuvels [3] for testing mutual independence in a vector of continuous random variables, first introduce the class \mathcal{C} of d -variate copulas, i.e., d -variate cumulative distribution functions whose margins are uniform on the interval $[0, 1]$. It is obviously included in the linear space \mathcal{F} of real-valued functions acting on $[0, 1]^d$.

When the focus is on dependence structures and the variables of interest are continuous, copulas are as informative as cumulative distribution functions. This is because if X_1, \dots, X_d are random variables with continuous cumulative distribution functions F_1, \dots, F_d , respectively, mutual independence between them occurs if and only if the uniform random variables $Z_1 = F_1(X_1), \dots, Z_d = F_d(X_d)$ are mutually independent, i.e., their joint cumulative distribution function, which is a copula, coincides with the independence copula, defined for all vectors $(z_1, \dots, z_d) \in D = [0, 1]^d$, by $\Pi(z_1, \dots, z_d) = z_1 \times \dots \times z_d$.

For arbitrary set $v \subseteq \{1, \dots, d\}$ and vector $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$, let $\mathbf{x}_{v,1}$ denote the vector (t_1, \dots, t_d) such that $t_i = x_i$ if $i \in v$ and $t_i = 1$ otherwise. The Möbius decomposition of a copula C is then given, for all $\mathbf{x} \in [0, 1]^d$, by

$$(3.1) \quad C(\mathbf{x}) = \Pi(\mathbf{x}) + \sum_{u \subseteq \{1, \dots, d\}, |u| \geq 2} \mathcal{M}_u(C)(\mathbf{x}) \times \prod_{k \notin u} x_k,$$

where

$$(3.2) \quad \mathcal{M}_u(C)(\mathbf{x}) = \sum_{v \subseteq u} (-1)^{|u \setminus v|} C(\mathbf{x}_{v,1}) \times \prod_{k \in u \setminus v} x_k.$$

Note that the definition (3.2) also makes sense for sets $u \subseteq \{1, \dots, d\}$ of cardinality 0 or 1, which yields $\mathcal{M}_u(C) = 1$ when $|u| = 0$ and $\mathcal{M}_u(C) = 0$ when $|u| = 1$. Therefore, formula (3.1) can be written alternatively,

for all $\mathbf{x} \in [0, 1]^d$, as

$$C(\mathbf{x}) = \sum_{u \subseteq \{1, \dots, d\}} C_u(\mathbf{x})$$

with

$$(3.3) \quad C_u(\mathbf{x}) = \mathcal{M}_u(C)(\mathbf{x}) \times \prod_{k \notin u} x_k = \sum_{v \subseteq u} (-1)^{|u \setminus v|} C(\mathbf{x}_{v, \mathbf{1}}) \times \prod_{k \notin v} x_k.$$

To see that this representation can be expressed in the form (1.1) with components defined through formula (2.1) in terms of commuting operators P_1, \dots, P_d , it suffices to let, for all $i \in \{1, \dots, d\}$, $C \in \mathcal{C}$, and $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$,

$$(3.4) \quad P_i C(\mathbf{x}) = x_i \times C(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_d).$$

The operators P_1, \dots, P_d so defined are clearly commutative, and it follows from their definition that for any set $v \subseteq \{1, \dots, d\}$ with $-v = \{1, \dots, d\} \setminus v$ and vector $\mathbf{x} \in [0, 1]^d$,

$$P_{-v} C(\mathbf{x}) = C(\mathbf{x}_{v, \mathbf{1}}) \Pi(\mathbf{x}_{-v}).$$

It then follows from definition (3.3) that for every set $u \subseteq \{1, \dots, d\}$ and vector $\mathbf{x} \in [0, 1]^d$,

$$(3.5) \quad C_u(\mathbf{x}) = \sum_{v \subseteq u} (-1)^{|u \setminus v|} P_{-v} C(\mathbf{x}).$$

Observe that with operators of the form (3.4) for all $i \in \{1, \dots, d\}$, the first term in the Möbius decomposition corresponds to the independence copula, Π , and that all first-order terms vanish. That is, for all $\mathbf{x} \in [0, 1]^d$, one has

$$C_\emptyset(\mathbf{x}) = P_{\{1, \dots, d\}} C(\mathbf{x}) = \Pi(\mathbf{x})$$

and for all $i \in \{1, \dots, d\}$,

$$C_{\{i\}}(\mathbf{x}) = P_{-i}(C)(\mathbf{x}) - P_{\{1, \dots, d\}} C(\mathbf{x}) = C(\mathbf{x}_{i, 1}) \Pi(\mathbf{x}_{-i}) - \Pi(\mathbf{x}) = 0.$$

Note, however, that Assumption 2.1 does not hold for the operators defined in (3.4). Yet the term (3.5) in the Möbius decomposition is consistent with representation (2.4). This suggests that Theorem 2.2 of Kuo et al. [13] is valid in a more general setting, as will be shown in Section 4.

Before proceeding, note that the restriction to a copula is actually unnecessary, as the above mentioned identity extends to any cumulative distribution function H whose support is $[0, 1]^d$. Indeed, Lemma 6 of [12] ensures that for any such cumulative distribution function and all $\mathbf{x} \in [0, 1]^d$, one has

$$H(\mathbf{x}) = \sum_{u \subseteq \{1, \dots, d\}} H_u(\mathbf{x})$$

with

$$H_u(\mathbf{x}) = \sum_{v \subseteq u} (-1)^{|u \setminus v|} P_{-v} H(\mathbf{x}).$$

In this more general setting, formula (3.4) is no longer appropriate and should be replaced by

$$P_i H(\mathbf{x}) = H(1, \dots, 1, x_i, 1, \dots, 1) \times H(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_d).$$

Thus, the operator P_i is no longer linear, even though the decomposition is still valid. Again, this suggests that Theorem 2.2 remains valid under a weaker set of conditions than that assumed by Kuo et al. [13].

4. An extension of Theorem 2.2 by Kuo et al. (2010). The following result relaxes Assumption 2.1 in Theorem 2.2, so that the Hoeffding–Sobol decomposition and the Möbius decompositions can both be derived from a common representation of functions of several variables. This generalization is stated below *in extenso* to facilitate understanding and referencing. For internal consistency, however, the equation numbers appearing therein are repeated from their first occurrence above. Recall that for any $u \subseteq \{1, \dots, d\}$, $P_u = \prod_{i \in u} P_i$ and $P_{-u} = P_{\{1, \dots, d\} \setminus u}$.

PROPOSITION 4.1. *Let \mathcal{F} be a linear space of real-valued functions acting on a domain D of \mathbb{R}^d and let P_1, \dots, P_d be commuting and idempotent operators on \mathcal{F} . Then, the following statements hold true.*

Part A: *Any function $f \in \mathcal{F}$ can be written as*

$$(1) \quad f = \sum_{u \subseteq \{1, \dots, d\}} f_u,$$

in which the term f_u , defined by

$$(2) \quad f_u = \left(\prod_{i \in u} (I_D - P_i) \prod_{i \notin u} P_i \right) (f),$$

is such that the annihilating property is satisfied, meaning that, for all $i \in \{1, \dots, d\}$,

$$(5) \quad P_i(f_u) = \begin{cases} 0 & \text{if } i \in u, \\ f_u & \text{if } i \notin u. \end{cases}$$

Part B: *Suppose that every function $f \in \mathcal{F}$ can be expressed as in (1.1) and that the annihilating property (2.2) holds for the operators P_1, \dots, P_d , which are further assumed to be linear. Then the term f_u is given by*

$$(6) \quad f_u = P_{-u}(f) - \sum_{v \subsetneq u} f_v,$$

which is equivalent to

$$(7) \quad f_u = \sum_{v \subseteq u} (-1)^{|u \setminus v|} P_{-v}(f),$$

and hence also to formula (2.1).

Proof. To recover the desired decomposition, first call on the multinomial formula to see that

$$\prod_{i=1}^d I_D = \prod_{i=1}^d \{(I_D - P_i) + P_i\} = \sum_{u \subseteq \{1, \dots, d\}} \left(\prod_{i \in u} (I_D - P_i) \right) \left(\prod_{i \notin u} P_i \right).$$

Accordingly, one finds

$$(4.1) \quad f = \left(\prod_{i=1}^d I_D \right) (f) = \sum_{u \subseteq \{1, \dots, d\}} \left(\prod_{i \in u} (I_D - P_i) \right) \left(\prod_{i \notin u} P_i \right) (f) = \sum_{u \subseteq \{1, \dots, d\}} f_u,$$

with f_u defined as in Eq. (2.1). Furthermore, note that for whatever $u \subseteq \{1, \dots, d\}$, the function f_u satisfies the annihilating property. Indeed, if $k \in u$, then

$$P_k(f_u) = \left(\prod_{i \in u \setminus \{k\}} (I_D - P_i) \right) \left(\prod_{i \notin u} P_i \right) P_k(I_D - P_k)(f) = 0$$

while if $k \notin u$, then

$$(I_D - P_k)(f_u) = \left(\prod_{i \in u} (I_D - P_i) \right) \left(\prod_{i \notin u, i \neq k} P_i \right) P_k (I_D - P_k)(f) = 0.$$

This proves Part A, namely the existence of the decomposition.

To establish uniqueness under the added assumption that P_1, \dots, P_d are linear stated in Part B, write

$$f = \sum_{v \subseteq \{1, \dots, d\}} f_v,$$

and apply the linearity of P_{-u} . It follows from the annihilating property (2.2) that

$$P_{-u}(f) = \sum_{v \subseteq u} f_v = f_u + \sum_{v \subsetneq u} f_v,$$

which gives the recursive formula. The well-known Möbius inversion then implies formula (2.4). Finally, the product version can be deduced from

$$f_u = \sum_{v \subseteq u} (-1)^{|u \setminus v|} P_{-v}(f) = \sum_{v \subseteq u} (-1)^{|u \setminus v|} \left(\prod_{i \in u \setminus v} P_i \right) \left(\prod_{i \notin u} P_i \right) (f) = \left(\prod_{i \in u} (I_D - P_i) \right) \left(\prod_{i \notin u} P_i \right) (f),$$

which concludes the argument. \square

In view of Part A of Proposition 4.1, the Möbius decomposition can also be derived for any cumulative distribution function whose support is $[0, 1]^d$ from a set of maps and its terms can be computed from (2.4) or equivalently from (2.1) or (2.3). This observation does not appear to have been documented previously.

5. Another extension of Theorem 2.2 by Kuo et al. (2010). While Proposition 4.1 achieves the main objective of providing a common representation for the Hoeffding–Sobol and Möbius decompositions, the need to impose linearity in Part B implies that existence and uniqueness are no longer obtained under a common set of assumptions. The following result, which remains valid for non-linear operators, shows that one can find a unique assumption and can even bypass the idempotent assumption.

PROPOSITION 5.1. *Let P_1, \dots, P_d be commuting operators. Then any real-valued function f acting on a domain $D \subseteq \mathbb{R}^d$ can be written in a unique way in the form*

$$(1) \quad f = \sum_{u \subseteq \{1, \dots, d\}} f_u,$$

under the assumption that for all $u \in \{1, \dots, d\}$,

$$(5.1) \quad P_{-u}(f) = \sum_{v \subseteq u} f_v.$$

Furthermore, the three following identities hold for all subsets $u \subseteq \{1, \dots, d\}$:

$$(2) \quad f_u = \left(\prod_{i \in u} (I_D - P_i) \prod_{j \notin u} P_j \right) (f),$$

$$(6) \quad f_u = P_{-u}(f) - \sum_{v \subsetneq u} f_v,$$

and

$$(7) \quad f_u = \sum_{v \subseteq u} (-1)^{|u \setminus v|} P_{-v}(f).$$

Proof. First invoke the multinomial identity to deduce Eq. (4.1) as detailed at the beginning of the proof of Proposition 4.1. This equation shows that the decomposition (1.1) holds with f_u defined by Eq. (2.1) for all $u \subseteq \{1, \dots, d\}$. Next observe that for this choice of functions and any $u \subseteq \{1, \dots, d\}$, one has

$$f = \prod_{j \in u} \{(I_D - P_j) + P_j\}(f) = \left(\sum_{v \subseteq u} \prod_{j \in v} (I_D - P_j) \prod_{k \in u \setminus v} P_k \right) (f)$$

and hence

$$\begin{aligned} P_{-u}(f) &= \prod_{i \notin u} P_i(f) = \prod_{i \notin u} P_i \prod_{j \in u} \{(I_D - P_j) + P_j\}(f) \\ &= \left(\prod_{i \notin u} P_i \sum_{v \subseteq u} \prod_{j \in v} (I_D - P_j) \prod_{k \in u \setminus v} P_k \right) (f). \end{aligned}$$

Further note that if $v \subseteq u$, then $k \notin v$ if and only if $k \in u \setminus v$ or if $k \notin u$. Therefore,

$$P_{-u}(f) = \left(\sum_{v \subseteq u} \prod_{j \in v} (I_D - P_j) \prod_{k \notin v} P_k \right) (f) = \sum_{v \subseteq u} f_v,$$

which is Eq. (5.1). This identity immediately implies Eq. (2.3). Moreover, Eq. (2.4) results from the fact that

$$f_u = \left(\prod_{i \in u} (I_D - P_i) \right) \left(\prod_{j \notin u} P_j \right) (f) = \sum_{v \subseteq u} (-1)^{|u \setminus v|} \left(\prod_{i \in u \setminus v} P_i \right) \left(\prod_{j \notin u} P_j \right) (f) = \sum_{v \subseteq u} (-1)^{|u \setminus v|} P_{-v}(f).$$

This concludes the argument. \square

Assumption (5.1), which is new, implies that for any fixed set $u \subseteq \{1, \dots, d\}$ of indices, an application of the operator P_{-u} to f retains the terms in the decomposition (1.1) whose variables have indices contained in u and filters out the others. This condition is very intuitive when the terms do not overlap: in particular, it precludes a term f_u associated to a set u to be considered as (part of) a term of higher complexity that would involve variables whose indices belong to a larger set $v \supset u$. In effect, the *no-overlap condition* (5.1) is equivalent to Eq. (2.3), which makes it possible to define the terms f_u recursively in a unique way, starting from the terms of smallest complexity.

In their paper, Kuo et al. (2021) assumed that the operators P_1, \dots, P_d are both idempotent and linear, and that they satisfy the annihilating property (2.2). In that situation, the no-overlap condition is equivalent to the annihilating property. More precisely, the following statements hold true:

- (i) Condition (2.2) holds for each idempotent operator P_i when f_u is defined as per Eq. (2.1) for all $u \subseteq \{1, \dots, d\}$.
- (ii) If the operators P_1, \dots, P_d are linear, then Condition (2.2) implies Eq. (2.1).

For a proof of claim (i), refer to the proof of Part A of Proposition 4.1. To establish claim (ii), fix any set $u \subseteq \{1, \dots, d\}$ and invoke the linearity of the operators to write

$$P_{-u}(f) = \sum_{v \subseteq \{1, \dots, d\}} P_{-u}(f_v) = \sum_{v \subseteq u} P_{-u}(f_v) + \sum_{v \not\subseteq u} P_{-u}(f_v).$$

Next observe that if $v \not\subseteq u$, there exists $i \in v$ such that $i \notin u$ and that for such an index i , one has $P_i(f_v) = 0$ by the annihilating property and therefore

$$P_{-u}(f_v) = \prod_{i \notin u} P_i(f_v) = 0.$$

This concludes the argument.

6. Discussion and perspectives. The Hoeffding–Sobol and Möbius formulas are two well-known ways of decomposing a function of several variables as a sum of terms of increasing complexity. As they were developed and used by distinct research communities, their suspicious resemblance had never been investigated. In this paper, a link between these two formulas was uncovered using an extension of a result due to Kuo et al. [13].

Beyond its intrinsic interest, the existence of a relation between these two famous formulas opens the door to cross-fertilization in their respective domains of application. Dependence modeling is a domain where this could happen, as both decompositions could be applied to copulas. In this context, the projection P_u (2.5) associated to the Hoeffding–Sobol decomposition cancels the influence of variables in the subset u through integration. The terms are of increasing complexity with $|u|$ in the sense that f_u depends on a vector \mathbf{x}_u whose length grows with $|u|$. In that decomposition, f_\emptyset is a constant.

In contrast, the projection P_u associated with the Möbius decomposition (3.5) erases the stochastic dependence between the variables with indices belonging to the set u . More precisely, if a random vector \mathbf{Z} has distribution $P_u(C)$, the subvectors \mathbf{Z}_u and \mathbf{Z}_{-u} are then independent. While there is no increasing complexity in the terms of the decomposition, in the sense that the value of C_u at any $\mathbf{x} \in D$ depends on both \mathbf{x}_u and \mathbf{x}_{-u} , it is noteworthy that only the stochastic dependence embodied within \mathbf{x}_u is retained. It can be said, therefore, that the terms in the Möbius decomposition are also increasing in complexity, but in the sense of probabilistic dependence. Further note that while the term f_\emptyset does not vanish in this decomposition, it does not contain any information about dependence given that $f_\emptyset = \Pi$.

Beyond the relative degree of complexity of their terms, the Hoeffding–Sobol and Möbius decompositions each have their comparative advantage. One key feature of the Hoeffding–Sobol decomposition is that it provides orthogonal terms, so that the structure of the function of interest can be analyzed through variances. In contrast, the terms in the Möbius decomposition are not generally orthogonal. However, a strong point of the Möbius decomposition is the ease with which any term can be computed as a simple alternate combination of evaluations of the function C .

Acknowledgments. Funding in support of this work was provided by the Canada Research Chairs Program and the Natural Sciences and Engineering Research Council of Canada.

References.

REFERENCES

- [1] Beran, R., Bilodeau, M., Lafaye de Micheaux, P., 2007. Nonparametric tests of independence between random vectors. *J. Multivariate Anal.* 98 (9), 1805–1824.
<https://doi.org/10.1016/j.jmva.2007.01.009>
- [2] Chastaing, G., Gamboa, F., Prieur, C., 2015. Generalized Sobol’ sensitivity indices for dependent variables: Numerical methods. *J. Stat. Comput. Simul.* 85 (7), 1306–1333.
<https://doi.org/10.1080/00949655.2014.960415>
- [3] Deheuvels, P., 1981. An asymptotic decomposition for multivariate distribution-free tests of independence. *J. Multivariate Anal.* 11 (1), 102–113.
[http://dx.doi.org/10.1016/0047-259X\(81\)90136-6](http://dx.doi.org/10.1016/0047-259X(81)90136-6)
- [4] Gamboa, F., Janon, A., Klein, T., Lagnoux, A., Prieur, C., 2016. Statistical inference for Sobol’ pick-freeze Monte Carlo method. *Statistics* 50 (4), 881–902.
<https://doi.org/10.1080/02331888.2015.1105803>
- [5] Genest, C., Nešlehová, J. G., Rémillard, B., Murphy, O. A., 2019. Testing for independence in arbitrary distributions. *Biometrika* 106 (1), 47–68.
<https://doi.org/10.1093/biomet/asy059>
- [6] Genest, C., Quessy, J.-F., Rémillard, B., 2007. Asymptotic local efficiency of Cramér–von Mises tests for multivariate independence. *Ann. Statist.* 35 (1), 166–191.
<https://doi.org/10.1214/009053606000000984>
- [7] Genest, C., Rémillard, B., 2004. Tests of independence and randomness based on the empirical copula process. *Test* 13, 335–369.

- <https://doi.org/10.1007/BF02595777>
- [8] Ghoudi, K., Kulperger, R., Rémillard, B., 2001. A nonparametric test of serial independence for time series and residuals. *J. Multivariate Anal.* 79 (2), 191–218.
<https://doi.org/10.1006/jmva.2000.1967>
- [9] Hoeffding, W., 1948. A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* 19 (3), 293–325.
<https://doi.org/10.1214/aoms/1177730196>
- [10] Iooss, B., Prieur, C., 2019. Shapley effects for sensitivity analysis with correlated inputs: Comparisons with Sobol’ indices, numerical estimation and applications. *Int. J. Uncertain. Quantif.* 9 (5), 493–514.
<https://doi.org/10.1615/Int.J.UncertaintyQuantification.2019028372>
- [11] Janon, A., Klein, T., Agnoux, A., Nodet, M., Prieur, C., 2014. Asymptotic normality and efficiency of two Sobol’ index estimators. *ESAIM: PS* 18, 342–364.
<https://doi.org/10.1051/ps/2013040>
- [12] Kojadinovic, I., Holmes, M., 2009. Tests of independence among continuous random vectors based on Cramér–von Mises functionals of the empirical copula process. *J. Multivariate Anal.* 100, 1137–1154.
<https://doi.org/10.1016/j.jmva.2008.10.013>
- [13] Kuo, F. Y., Sloan, I. H., Wasilkowski, G. W., Woźniakowski, H., 2010. On decompositions of multivariate functions. *Math. Comp.* 79 (270), 953–966.
<https://doi.org/10.1090/S0025-5718-09-02319-9>
- [14] Liu, R., Owen, A., 2006. Estimating mean dimensionality of analysis of variance decompositions. *J. Amer. Statist. Assoc.* 101 (474), 712–721.
<https://doi.org/10.1198/016214505000001410>
- [15] Mercadier, C., Ressel, P., 2021. Hoeffding–Sobol decomposition of homogeneous co-survival functions: From Choquet representation to extreme value theory application. *Dependence Modeling*, in press.
<https://hal.archives-ouvertes.fr/hal-03200817/document>
- [16] Mercadier, C., Roustant, O., 2019. The tail dependograph. *Extremes* 22 (2), 343–372.
<https://doi.org/10.1007/s10687-019-00345-3>
- [17] Ökten, G., Liu, Y., 2021. Randomized quasi-Monte Carlo methods in global sensitivity analysis. *Reliab. Eng. Syst. Saf.* 210, 107520.
<https://doi.org/10.1016/j.res.2021.107520>
- [18] Owen, A., 1994. Lattice sampling revisited: Monte Carlo variance of means over randomized orthogonal arrays. *Ann. Statist.* 22 (2), 930–945.
<https://doi.org/10.1214/aos/1176325504>
- [19] Owen, A., 2019. Effective dimension of some weighted pre-Sobolev spaces with dominating mixed partial derivatives. *SIAM J. Numer. Anal.* 57 (2), 547–562.
<https://doi.org/10.1137/17M1158975>
- [20] Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J. H. A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., Smith, S., Sheikholeslami, R., Hosseini, N., Asadzadeh, M., Puy, A., Kucherenko, S., Maier, H. R., 2021. The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environ. Model. Softw.* 137, 104954.
doi.org/10.1016/j.envsoft.2020.104954
- [21] Roustant, O., Gamboa, F., Iooss, B., 2020. Parseval inequalities and lower bounds for variance-based sensitivity indices. *Electron. J. Statist.* 14 (1), 386–412.
<https://doi.org/10.1214/19-EJS1673>
- [22] Sobol, I. M., 1993. Sensitivity estimates for nonlinear mathematical models. *Math. Modeling Comput. Experiment.* 1 (4), 407–414 (1995).