Hoeffding decompositions for exchangeable sequences
and chaotic representation of functionals of Dirichlet processes

Giovanni Peccati a,b

a Laboratoire de probabilités et modèles aléatoires, Universités de Paris VI & VII, Paris 75252, France
b Istituto di Metodi Quantitativi dell’Università ‘L. Bocconi’, 25, via Sarfatti, 20136 Milan, Italy

Received 3 September 2002; accepted 1 April 2003

Abstract

Consider an exchangeable sequence \( X = \{X_n; 1 \leq n < N\} \), where \( N \in \mathbb{N} \cup \{\infty\} \), and note \( X_n = (X_1, \ldots, X_n) \). We say that \( X \) is Hoeffding decomposable if, for each \( n \), every square integrable, centered and symmetric functional of \( X_n \) is the orthogonal sum of \( n \) U-statistics with degenerated and symmetric kernels. We state a necessary and sufficient condition for an exchangeable sequence to be Hoeffding decomposable, named weak independence. We show that a class of weakly independent sequences is given by generalized urn sequences and, specifically, by generalized Pólya urns. We point out that this yields an orthogonal decomposition of the space of square integrable functionals of Dirichlet–Ferguson processes into orthogonal subspaces of multiple integrals. Explicit formulae are provided.

Version française abrégée

Cette Note est un résumé des deux derniers chapitres dans [10]. Nous considérons une collection de variables aléatoires échangeables \( X = \{X_n; 1 \leq n < N\} \), avec \( N \in \mathbb{N} \cup \{+\infty\} \), à valeurs dans un espace de Borel \((A, \mathcal{A})\)
et définies sur $(\Omega, \mathcal{F}, \mathbb{P})$. On écrit $X_n = (X_1, \ldots, X_n)$, et on note $L^2_n(X_n)$ l’espace des fonctionnelles de carré intégrable et symétriques de $X_n$. On dit que $X$ est *dénomposable au sens d’Hoeffding* si, $\forall n$, chaque $T \in L^2_n(X)$ est telle que, p.s., $T = \sum_{i=0}^{n-1} T_i$, où $T_0 = \mathbb{E}(T)$, $T_i (i = 1, \ldots, n)$ est une $U$-statistique avec noyau dégénéré d’ordre $i$, et $\mathbb{E}(T_1 T_j) = 0$ pour $i \neq j$. On montre que $X$ est décomposable au sens d’Hoeffding ssi $X$ est *faiblement indépendante*. Par exemple, si $X = (X_1, X_2, X_3)$, ceci équivaut à la relation : $\forall \phi \in L^2_n(X_n)$ est t.q. $\mathbb{E}(\phi(X_2) | X_1) = 0$, alors $\mathbb{E}(\phi(X_2) | X_3) = 0$. Une classe remarquable de suites échangeables et faiblement indépendantes est celle des *suites d’urne généralisées*. Des exemples de telles suites sont : les variables aléatoires i.i.d., les extractions sans remise d’un ensemble fini et les processus de Pólya généralisés (cf. [2]). On prouve que dans le cas des suites d’urne, et pour chaque $n$, les $U$-statistiques qui décomposent une certaine $T \in L^2_n(X_n)$ sont des combinaisons linéaires d’espérances conditionnelles dont on peut déterminer explicitement les coefficients. Ceci généralise les formules classiques contenues dans [6] et [13]. Les résultats ainsi obtenus peuvent être appliqués à l’étude des fonctionnelles de carré intégrable d’un processus de Dirichlet–Ferguson $D$ sur un espace polonais (défini dans [5]), qu’on écrit $L^2(D)$. En effet, on utilise un résultat très connu de Blackwell et MacQueen (cf. [2]) qui relie suites de Pólya et processus de Dirichlet–Ferguson, pour prouver que chaque $F \in L^2(D)$ admet la représentation $F = \mathbb{E}(F) + \sum_{n \geq 1} \int_{A_n} h_n dD^{\otimes n}$, où les $h_n$ sont des noyaux dégénérés d’ordre $n$, qu’on peut encore une fois représenter explicitement.

1. Introduction, Hoeffding decompositions and weak independence

This Note is an abridged version of the last two chapters in [10] (see also [8] and [9]). For $N \in \mathbb{N} \cup \{+\infty\}$, consider a collection $X = \{X_n : 1 \leq n < N\}$ of exchangeable random observations, whose components take values in a Borel space $(A, \mathcal{A})$ and are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. When $N < +\infty$, we will suppose that $X$ is $2(N-1)$-extendible (see [1] for any notion concerning exchangeability). We also use the notation $X_n = (X_1, \ldots, X_n)$.

Fix $n \geq 1$. For any $m \in \{0, 1, \ldots, n\}$ we define

$$
V_m(m) := \{k(m) = (k_1, \ldots, k_m) : 1 \leq k_1 < \cdots < k_m \leq n\}
$$

with $k_0 = 0$ and $V_0(0) = \emptyset$. We also set $V_m(n) = \bigcup_{k=1}^{n} V_m(k)$. For $m \geq 1$, $l_1 = (l_1, \ldots, l_m) \in V_m(m)$ and $k_1 = (k_1, \ldots, k_0) \in V_m(n)$, $l_1 \wedge k_1$ stands for the class $\{l_i : l_i = k_j \text{ for some } j = 1, \ldots, n\}$, written as an element of $V_m(r)$, where $r := \text{Card}(l_1 \wedge k_1)$. Given $k_1 \in V_m(n)$ and a vector $h(m) = (h_1, \ldots, h_m)$, $h(m) \subset k_1$ means that $h(j) \in V_m(k_1)$, and that for every $1 \leq i \leq m$ there exists $1 \leq j \leq n$ such that $k_j = h_i$. For any $j(m) \in V_m(n)$, we write $X_{j(m)} = (X_{j_1}, \ldots, X_{j_m})$.

Consider eventually a symmetric and measurable function $T$ on $A^n$ s.t. $T(X_n) \in L^1(X_n)$. We recall that exchangeability implies that for every $0 \leq r \leq m \leq n$ there exists a measurable function $T^{(r)}_{n,m} : A^m \rightarrow \mathbb{R}$, symmetric in the first $r$ variables and in the last $m - r$, s.t. for every $j(m) \in V_m(n)$ and $i(m) \in V_m(m)$ satisfying $\text{Card}(i(m) \wedge j(m)) = r$, one has

$$
\mathbb{E}[T(X_{j(m)}) | X_{i(m)}] = [T]^{(r)}_{n,m}(X_{j(m)}|X_{i(m)}, X_{i(m)}|j(m)) \quad \text{a.s.-}\mathbb{P}.
$$

We will denote by $[T]^{(r)}_{n,m}$ the usual symmetrization of $[T]^{(r)}_{n,m}$.

For $1 \leq n < N$, we set $L^2_0(X_n)$ to be the subspace of square integrable and symmetric functionals of the vector $X_n$ and, for $i = 1, \ldots, n$, we set $SU_i = SH_i = \emptyset$, and

$$
SU_i(X_n) = \left\{ T : T = \sum_{j(m) \in V_m(i)} \phi(X_{j(m)}), \ \phi(X_{j(m)}) \in L^2(X) \right\},
$$

$SH_i(X_n) = SU_i(X_n) \cup SU_{i-1}(X_n)$.
The set \( \{\text{SH}_i(X_n) : i = 1, \ldots, n\} \) is the collection of symmetric Hoeffding spaces associated to \( X_n \). Given \( T \in L^2_d(X_n) \), for every \( i = 0, \ldots, n \) we use the symbol \( \pi[T, \text{SH}_i(X_n)] \) for the projection of \( T \) on \( \text{SH}_i(X_n) \). Plainly, \( L^2_d(X_n) = \bigoplus_{i=0}^{n} \text{SH}_i(X_n) \). We assume that, for every \( n \) and every \( i \leq n \), \( \text{SH}_i(X_n) \neq \{0\} \). We remark that, thanks to results contained in [10, Chapter 9], \( \text{SU}_i(X_n) \) is, for every \( i \), a \( L^2 \)-closed vector space, and moreover every element of \( \text{SU}_i(X_n) \) admits an essentially unique representation of the type \( T = \sum_{j(i)} \phi(X_{j(i)}) \), where \( \phi(X_i) \in L^2_d(X_i) \).

**Definition 1.1.** The exchangeable sequence \( X \) is Hoeffding decomposable if, for every \( 1 \leq i \leq n < N \) and every \( 1 \leq i \leq n \), \( T \in \text{SH}_i(X_n) \) iff there exists \( \phi_T^{(i)} : A' \mapsto \mathbb{R} \) such that \( \phi_T^{(i)}(X_i) \in L^2_d(X_i) \).

\[
E[\phi_T^{(i)}(X_i) | X_{i-1}] = 0, \quad \mathbb{P}\text{-a.s.} \tag{2}
\]

and \( T = \sum_{j(i) \in V(i)} \phi_T^{(i)}(X_{j(i)}) \). When \( \phi_T^{(i)} \) is such that \( \phi_T^{(i)}(X_i) \in L^2_d(X_i) \) and satisfies (2), we write \( \phi_T^{(i)} \in \mathcal{E}_{i}(X) \).

Note that the only examples of Hoeffding decomposable sequences explicitly studied in the literature are i.i.d. random variables (see [6], for the original result, and [7]) and extractions without replacement from a finite population (see [13], [3] and [4]).

**Definition 1.2.** The exchangeable sequence \( X \) is weakly independent if, for every \( 1 \leq i \leq n < N \) and every \( T \in L^2_d(X_n) \), the following implication holds for every \( 0 \leq r \leq n-1 \) such that \( 2n-r \leq N \)

\[
\{T^{(r)(n-1)}(X_{n-1}) = 0, \quad \mathbb{P}\text{-a.s.} \} \quad \Rightarrow \quad \{T^{(r)}_{n,n-1}(X_{n-1}) = 0, \quad \mathbb{P}\text{-a.s.} \}.
\]

The proof of the next result is contained in [10, Chapter 9] and [8].

**Theorem 1.3.** Under the previous assumptions and notation, \( X \) is Hoeffding decomposable if, and only if, it is weakly independent.

There exist exchangeable sequences that are not weakly independent. To have an example of such a situation, just take a mixture of independent Bernoulli trials with common random parameter uniformly distributed on \([0, 1/2]\) (see [8] for further details).

### 2. The case of GUS

From now on, \((A, \mathcal{A})\) is a Polish space endowed with its Borel \( \sigma \)-field. For \( N \in \mathbb{N} \cup \{+\infty\} \), and writing \( \mathcal{M}(A) \) for the class of finite and positive measures on \( A \), we say that a sequence \( X^{(\alpha,c)}_n = \{X^{(\alpha,c)}_n : 1 \leq n < N\} \) is a GUS of parameters \( \alpha \in \mathcal{M}(A) \) and \( c \in \mathbb{R} \), if \( \alpha(A) + 2c(N-1) > 0 \) and if, for every \( k \) and every \( j(k) \in V_{N-1}(k) \),

\[
\mathbb{P}(X^{(\alpha,c)}_j \in dx_1, \ldots, X^{(\alpha,c)}_k \in dx_k) = \prod_{i=1}^{k} \frac{\alpha(dx_i) + c \sum_{h=1}^{i-1} \delta_{x_h}(dx_i)}{\alpha(A) + c(i-1)}, \tag{3}
\]

with \( \delta_x(\cdot) \) the Dirac mass at \( x \). Note that for any choice of \( \alpha \) and \( c \) the sequence \( X \) is exchangeable, and that the assumption \( \alpha(A) + 2c(N-1) > 0 \) ensures that \( X \) is \( 2(N-1) \)-extendible. If \( c = 0 \) the \( X_i \)’s are i.i.d. variables with law \( \alpha(\cdot)/\alpha(A) \); if \( c = -1 \), \( A \) is finite and \( \alpha \) is integer valued the \( X_i \)’s have the same law as \( N-1 \) extractions without replacement from \( A \); if \( c = 1 \) the \( X_i \)’s constitute a generalized Pólya sequence with parameter \( \alpha(\cdot) \), as introduced in [2]. The following result stems from Proposition 8 in [10, Chapter 9]. It generalizes some classic computations contained in [6] and [13].
Proposition 2.1. For any choice of $\alpha$ and $c$ consistent with the above assumptions, the GUS $X^{(a,c)}$ is Hoeffding decomposable. Moreover, if $T, V \in \Sigma_n(X^{(a,c)})$, for every $J(n), I(n) \in V_{N-1}(n)$ such that $\text{Card}(I(n) \cap J(n)) = r$

\[
E[T(X^{(a,c)}_{\eta}))V(X^{(a,c)}_{j})] = e^{a-r} \prod_{l=1}^{n-r} \frac{n-r-l+1}{\alpha(A) + c(n+l-1)} E[T(X^{(a,c)}_{\eta}))V(X^{(a,c)}_{j})].
\]

(4)

Now fix $1 \leq M < N$ and set

\[
\Phi(n, m, r, p) := c^p(m-r)(m-r-p) \prod_{l=1}^{m-r+p-1} \frac{\alpha(A) + c(r+p+s-1)}{\alpha(A) + c(n+s-1)},
\]

where $1 \leq m \leq n \leq M$, $0 \leq r \leq m$, $0 \leq p \leq m-r$, $\alpha(A) + c(n+m-r) > 0$, $(a)_{(b)} := a!/b!$ for $a \geq b$ and $\prod_{s=1}^{0} = 1 = 0^1$, and, for $1 \leq q \leq m \leq n \leq M$,

\[
\Psi_M(q, n, m) := \sum_{r=0}^{q} \binom{q}{r} \binom{M-n}{m-r} \Phi(n, m, q-r)
\]

(6)

with $(\binom{q}{r}) := (\binom{q}{r})1_{(a \geq b)}$. The next result characterizes Hoeffding decompositions for GUS, thus unifying and generalizing the calculations contained in [6] for i.i.d. random variables, and in [13] and [3] for extractions without replacement from a finite set. Note that the assumptions on the $\Psi$’s are immaterial in the case $c \geq 0$.

Theorem 2.2. Under the previous notation and assumptions, fix $\alpha \in \mathcal{M}(A)$ such that $\Psi_M(q, n, q)$ differs from zero for every $n = 1, \ldots, M$ and every $1 \leq q \leq n$. For every $k = 1, \ldots, M - 1$ the following equality holds a.s.-$\mathbb{P}$ for any $T \in L_2^2(X^{(a,c)}_M)$ with $E(T) = 0$:

\[
\pi[T, SH_{\alpha}](X^{(a,c)}_M) = \sum_{J(k) \in V_M(k)} \left[ \sum_{a=1}^{k} \theta^{(k,a)}_{M,a} \sum_{J(a) \in J(k)} [T]_{M,a}(X^{(a,c)}_{j(a)}) \right],
\]

(7)

where $\theta^{(k,a)}_{M,a} := \theta^{(k,a)}_{M,M-a}^{-1}$ and the coefficients $\theta^{(k,a)}_{M}$ are recursively defined by the set of conditions

\[
[S_{M}(k), k = 1, \ldots, M - 1]
\]

given by

\[
\theta^{(k,k)}_{M} = (\Psi_M(k, k, k))^{-1},
\]

\[
\sum_{l=q}^{k} \sum_{j=q}^{l} \theta^{(l,j)}_{M} \Psi_M(q, k, j) = 0, \quad q = 1, \ldots, k - 1
\]

and consequently

\[
\pi[T, SH_{M}](X^{(a,c)}_M) = \sum_{a=1}^{M} \sum_{J(a) \in V_M(a)} \theta^{(M,a)}_{M} [T]_{M,a}(X^{(a,c)}_{j(a)}),
\]

where $\theta^{(M,a)}_{M} := - \sum_{l=1}^{M-1} \theta^{(l,a)}_{M}$ for $a = 1, \ldots, M - 1$ and $\theta^{(M,M)}_{M} = \Psi_M(M, M) = 1$.

3. Application to Dirichlet–Ferguson processes

Let $(A, \mathcal{A})$ be a Polish space endowed with its Borel $\sigma$-field, and take $\alpha \in \mathcal{M}(A)$. Following [5], given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say that a random probability measure $\{D(C; \omega) : C \in \mathcal{A}\}$ is a Dirichlet–Ferguson process (in the sequel, DF process) with parameter $\alpha$ if for every finite measurable partition $(C_1, \ldots, C_n)$ of $A$ the vector $(D(C_1; \cdot), \ldots, D(C_n; \cdot))$ has a Dirichlet distribution with parameters $(\alpha(C_1), \ldots, \alpha(C_n))$. The link with the
previous section is given by the following classic result, contained in [2]: for every \( \alpha \in \mathcal{M}(A) \) and for every DF process \( D \) of parameter \( \alpha \), there exists a GUS \( X^{(\alpha,1)} \) s.t. \( D \) is the directing measure of \( X^{(\alpha,1)} \), and the sequence of empirical measures generated by \( X^{(\alpha,1)} \) converges a.s. to \( D \). The convergence can be interpreted in the sense of total variation (see [11]). Now write \( L^2(D) \) for the space of square integrable functionals of \( D \). The following result establishes a chaotic decomposition of \( L^2(D) \). It is proved in [9] and [10] by using Theorem 2.2, as well as the above quoted result of Blackwell and MacQueen.

**Theorem 3.1.** Every \( F \in L^2(D) \) admits a unique representation of the type

\[
F = \mathbb{E}(F) + \sum_{n \geq 1} \int h(F,n)(a_1, \ldots, a_n) D^\otimes n (da_1, \ldots, da_n),
\]

where \( D^\otimes n, n \geq 1, \) is the product measure associated to \( D \), the series converges in \( L^2 \), and \( h(F,n) \in \mathcal{L}_n(X^{(\alpha,1)}), n \geq 1 \). Moreover,

\[
\mathbb{E}\left(\int h(F,n) dD^\otimes n \int h(F,m) dD^\otimes m\right) = \delta_{m,n} c(n, \alpha(A)) \mathbb{E}\left(h(F,n)(X_n)^2\right),
\]

where \( c(n, \alpha(A)) = \prod_{l=1}^{n}(n-l+1)/(\alpha(A) + n + l - 1) \) and \( \delta \) is the Kronecker symbol. Eventually, there exist real coefficients \( \{\theta(n,k) : n \geq 1, 1 \leq k \leq n\} \) satisfying the relations

\[
\theta(n,k) = \sum_{1 \leq j_1 < \cdots < j_k \leq n} \mathbb{E}(F - \mathbb{E}(F) | X_1 = a_{j_1}, \ldots, X_k = a_{j_k}),
\]

where the notation is that of Theorem 2.2, for every \( F \in L^2(D) \).

It is worth noting that (11) is quite explicit since, according, e.g., to [5, Theorem 1], for every \( k \geq 1 \) and every \( (a_1, \ldots, a_k) \in A^k \) the law of \( D \) under the conditioned measure \( \mathbb{P}(\cdot | X_1 = a_1, \ldots, X_k = a_k) \) is that of a DF process with parameter \( \alpha + \sum_{i=1}^k \delta_{a_i} \). Observe also that the obtention of the explicit formula (11) does not require any regularity condition for the functional \( F \), as often happens in the case of other chaotic decompositions (see, e.g., [12] for the Brownian case, where the regularity conditions are related to Shigekawa–Malliavin differentiability).

**References**


