IEEE International Conference on Acoustics, Speech, and Signal Processing
İstanbul, Turkey, June 5–9, 2000

GENERALIZE HIGHER-ORDER MOMENTS IN INDEPENDENT COMPONENT ANALYSIS

Jeffrey O. Coleman

jeffc@alum.mit.edu | U.S. Naval Research Laboratory, Washington, DC

ABSTRACT

In independent component analysis (ICA), random-variable independence is often equated with factorization of the joint moments, expectations of products of powers. This paper shows that many nonpower functions are equally useful: if $E[f(X)g(Y)]$ factors into $E[f(X)]E[g(Y)]$ for every f and g from an *independence class*, then random variables X and Y are independent. Examples of and sufficient conditions for independence classes are presented for bounded random variables.

I. INTRODUCTION

Formally, it is elementary that if $X_1, X_2, \ldots$ are independent random variables, then for any Borel functions $f_1, \ldots, f_n$,

$$E[f_1(X_1) \times \cdots \times f_n(X_n)] = E[f_1(X_1)] \times \cdots \times E[f_n(X_n)], \tag{1}$$

as long as either[1] $f_j(X_j) > 0$ a.s. $\forall j = 1, \ldots, n$ or $E[f_j(X_j)] < \infty, \forall j = 1, \ldots, n$. There is an obvious converse: If (1) holds for all n and for all Borel functions $f_1, \ldots, f_n$, then $X_1, X_2, \ldots$ are independent, because indicators $1_{B_1}, \ldots, 1_{B_n}$, for any Borel sets $B_1, \ldots, B_n$, are Borel functions, and (1) becomes either of these (the second usually defines random-variable independence):

$$\begin{aligned} &E[1_{B_1}(X_1) \times \cdots \times 1_{B_n}(X_n)] \\ &= E[1_{B_1}(X_1)] \times \cdots \times E[1_{B_n}(X_n)], \end{aligned}$$

$$\begin{aligned} &P[X_1 \in B_1, \cdots, X_n \in B_n] \\ &= P[X_1 \in B_1] \times \cdots \times P[X_n \in B_n]. \end{aligned}$$

Thus for independence it is enough that (1) hold for indicator functions, establishing one alternative to moment factorization. Would other families of functions serve as well? For bounded random variables relevant in ICA, this paper shows that such families exist, establishes sufficient identification criteria, and gives examples.

The footnotes are for the reader who might be missing some of the relevant mathematical background and can be safely ignored.

II. SUFFICIENT CONDITIONS FOR INDEPENDENCE

Definition 1 *For class $\mathcal{X}$ of random variables on a common probability space and family $\mathcal{F}$ of real Borel functions on $\mathbb{R}$, $\mathcal{F}$ is an* independence class *w.r.t. $\mathcal{X}$ if $E[f_1(X_1) \times \cdots \times f_n(X_n)] = E[f_1(X_1)] \times \cdots \times E[f_n(X_n)]$ $\forall f_1, \ldots, f_n \in \mathcal{F}$ implies the independence of arbitrary $X_1, \ldots, X_n \in \mathcal{X}$.*

So the Borel functions and the indicator functions on Borel sets are both independence classes w.r.t. any family of random variables. But the huge first class is trivial, and the second just reflects the definition of independence, so neither is of much interest. To find more interesting independence classes, we must find smaller ones.

First, the random variables of an infinite sequence are independent if those in every finite subfamily are,[2] which leads to this:

Proposition 2 *Let $\mathcal{F}$ be an independence class w.r.t. $\mathcal{X}$, and let $X_1, X_2, \ldots \in \mathcal{X}$. If $E[f_1(X_1) \times \cdots \times f_n(X_n)] = E[f_1(X_1)] \times \cdots \times E[f_n(X_n)]$, $\forall f_1, \ldots, f_n \in \mathcal{F}$ and for every $n = 2, 3, \ldots$, then $X_1, X_2, \ldots$ are independent.*

A π-system,[3] a family of subsets of $\mathbb{R}$ closed under finite intersection, then provides further independence classes. Here $\mathbb{B}$ refers to the Borel subsets of $\mathbb{R}$, and $\mathbb{B} \cap Y$ denotes $\{B \cap Y : B \in \mathbb{B}\}$. If π-system $\mathcal{D}$ generates $\mathbb{B}$, then $\mathcal{D} \cap Y$ is[4] a π-system generating $\mathbb{B} \cap Y$. The $Y = \mathbb{R}$ special case of this next theorem: the indicator functions of a π-system generating $\mathbb{B}$ form an independence class.

Theorem 3 *For class $\mathcal{X}$ of random variables on a common probability space, let $Y \in \mathbb{B}$ be such that $\forall X \in \mathcal{X}$, $X \in Y$ a.s. If π-system $\mathcal{C}$ generates σ-field $\mathbb{B} \cap Y$, then the family of indicator functions $\{1_C : C \in \mathcal{C}\}$ is an independence class w.r.t. $\mathcal{X}$.*

PROOF: Let $X_1, \ldots, X_n$ be random variables on a common probability space, and initially let π-system $\mathcal{C}$ generate σ-field $\mathbb{B}$. By definition,[5] the independence of $X_1, \ldots, X_n$ is just the independence of $[X_1 \in B_1], \ldots, [X_n \in B_n]$, $\forall B_1, \ldots, B_n \in \mathbb{B}$. Suppose for the moment that $X_i \in Y$ always. Then these events can be written $[X_1 \in B_1 \cap Y], \ldots, [X_n \in B_n \cap Y]$, so independence requires just the independence of σ-fields[6] $X_1^{-1}(\mathbb{B} \cap Y), \ldots, X_n^{-1}(\mathbb{B} \cap Y)$. If $\mathcal{C}$ is a π-system that generates $\mathbb{B} \cap Y$, then $X_i^{-1}(\mathcal{C})$ is a π-system[7] that generates[8] $X_i^{-1}(\mathbb{B} \cap Y)$. Independent π-systems generate independent σ-fields,[9] so the independence of classes $X_1^{-1}(\mathcal{C}), \ldots, X_n^{-1}(\mathcal{C})$ is sufficient. This simply requires these equivalent conditions to hold:

$$\begin{aligned} &P[X_1 \in C_1, \ldots, X_n \in C_n] \\ &= P[X_1 \in C_1] \times \cdots \times P[X_n \in C_n] \end{aligned}$$

$$\begin{aligned} &E[1_{C_1}(X_1) \times \cdots \times 1_{C_n}(X_n)] \\ &= E[1_{C_1}(X_1)] \times \cdots \times E[1_{C_n}(X_n)], \end{aligned}$$

This work began when the author was with The Boeing Company.

[1] Loève [1, Section 16.2, p. 238], Billingsley [2, p. 284], or Chung [3, Corollary to Theorem 3.3.3, p. 53].

[2] Chung [3, Section 3.3, p. 50].

[3] Billingsley [2, p. 36]

[4] Billingsley [2, Thm 10.1(ii), p. 156].

[5] Billingsley [2, p. 267].

[6] Billingsley [2, Thm 10.1(i), p. 156].

[7] Billingsley [2, note A7, p. 565].

[8] Billingsley [2, problem 13.5(c), p. 187].

[9] Billingsley [2, Thm 4.2, p. 50] or Loève [1, Thm 16.1B, p. 237].

for all $C_1, \ldots, C_n \in \mathcal{C}$. This implies independence of $X_1, \ldots, X_n$, so $\{1_C : C \in \mathcal{C}\}$ is an independence class.

Now weaken the requirement $X_i \in Y$ to $X_i \in Y$ with probability one, or $X_i \in Y \cup N_i$ with $P[X_i \in N_i] = 0$. Again the independence of $X_1, \ldots, X_n$ first reduces to the independence of events $[X_1 \in B_1], \ldots, [X_n \in B_n]$, or

$$\begin{aligned} &P\left(X_1^{-1}(B_1) \cap \ldots \cap X_n^{-1}(B_n)\right) \\ &= P\left(X_1^{-1}(B_1)\right) \times \ldots \times P\left(X_n^{-1}(B_n)\right), \end{aligned} \tag{2}$$

for arbitrary $B_1, \ldots, B_n \in \mathbb{B}$. In the development above, each B_i could at this point be replaced with the equivalent event $B_i \cap Y$, but here these events may not be equivalent. Consider the left side first. For brevity, let $Z \triangleq X_2^{-1}(B_2) \cap \ldots \cap X_n^{-1}(B_n)$. Then, since $X_1 \in Y \cup N_1$,

$$\begin{aligned} &P\left(X_1^{-1}(B_1) \cap Z\right) \\ &= P\left(X_1^{-1}(B_1 \cap (Y \cup N_1)) \cap Z\right) \\ &= P\left((X_1^{-1}(B_1 \cap Y) \cap Z) \cup (X_1^{-1}(B_1 \cap N_1) \cap Z)\right) \\ &\leq P\left(X_1^{-1}(B_1 \cap Y) \cap Z\right) + P\left(X_1^{-1}(B_1 \cap N_1) \cap Z\right) \\ &\leq P\left(X_1^{-1}(B_1 \cap Y) \cap Z\right) + P\left(X_1^{-1}(N_1)\right) \\ &= P\left(X_1^{-1}(B_1 \cap Y) \cap Z\right). \end{aligned}$$

In this chain[10] the last certainly cannot exceed the first,[11] so we have squeezed together the $m = 0$ and $m = 1$ forms of

$$P\left(\left\{\bigcap_{i=1}^{m} X_i^{-1}(B_i \cap Y)\right\} \cap \left\{\bigcap_{i=m+1}^{n} X_i^{-1}(B_i)\right\}\right).$$

A similar argument[12] begins with the $m = 1$ form and establishes equality with the $m = 2$ form. Continuing,

$$\begin{aligned} &P\left(X_1^{-1}(B_1) \cap \ldots \cap X_n^{-1}(B_n)\right) \\ &= P\left(X_1^{-1}(B_1 \cap Y) \cap \ldots \cap X_n^{-1}(B_n \cap Y)\right). \end{aligned}$$

Specializing with $B_j = \mathbb{R}$ for $j \neq i$ yields $P\left(X_i^{-1}(B_i)\right) = P\left(X_i^{-1}(B_i \cap Y)\right)$, so (2) becomes

$$\begin{aligned} &P\left(X_1^{-1}(B_1 \cap Y) \cap \ldots \cap X_n^{-1}(B_n \cap Y)\right) \\ &= P\left(X_1^{-1}(B_1 \cap Y)\right) \times \ldots \times P\left(X_n^{-1}(B_n \cap Y)\right). \end{aligned}$$

That this hold for arbitrary $B_1, \ldots, B_n \in \mathbb{B}$ is just the statement that σ-fields $X_1^{-1}(\mathbb{B} \cap Y), \ldots, X_n^{-1}(\mathbb{B} \cap Y)$ are independent. The rest of the proof proceeds as before. □

Many π-systems generate the Borels, including[13] the open sets, the closed sets, and the intervals (including the empty interval) of any one of eight forms, (a, b), $[a, b)$, $(a, b]$, $[a, b]$, $(-\infty, b)$, $(-\infty, b]$, $(a, \infty]$, and $[a, \infty)$, whose endpoints can be restricted to a set dense in $\mathbb{R}$, like the rationals. The b-adic intervals[14] $\{((k-1)b^{-n}, kb^{-n}] : n, k \in \mathbb{Z}\}$ with the null set adjoined yields a π-system generating the Borels[15] for any integer $b \geq 2$. Theorem 3 on $\{(-\infty, x] : x \in \mathbb{R}\}$ is this: a finite family of random variables are independent if their joint distribution function factors.

[10] The measure's subadditivity and monotonicity yield the inequalities.

[11] Because of the monotonicity of the measure.

[12] Distinguishing "2" now and redefining Z as the undistinguished part.

[13] Chung [3, p. 28], Loève [1, p. 104], Billingsley [2, p. 156], and Folland [4, Prop. 1.2, p. 21].

[14] Billingsley [2, p. 4].

[15] $\forall x \in \mathbb{R}, \exists \{x_n\} \subset \mathbb{Z} \ni: \frac{x_n}{b^n} \downarrow x$. Then $\bigcap_{n=0}^{\infty} \bigcup_{k=-\infty}^{x_n} ((k-1)b^{-n}, kb^{-n}] = (-\infty, x]$.

Approximating indicator functions with convergent function sequences leads to even smaller independence classes. Notation $X \sim F$ means random variable X has distribution function F:

Lemma 4 *Let $\mathcal{G}$ be an independence class w.r.t. $\mathcal{X}$, and let $\mathcal{F}$ be a family of real Borel functions on $\mathbb{R}$. Suppose that $\forall g \in \mathcal{G}, \forall X \in \mathcal{X}$, there is a sequence $\{f_k\} \subset \mathcal{F}$ such that $f_k \to g$ a.e. F as $k \to \infty$ and such that $|f_k| \leq M$ a.e. F for some $M < \infty$ (independent of k), where $X \sim F$. Then $\mathcal{F}$ is an independence class w.r.t. X.*

PROOF: With $\mathcal{X}$, $\mathcal{G}$, and $\mathcal{F}$ as above and $\forall h_1, \ldots, h_n \in \mathcal{F}$,

$$\begin{aligned} &E[h_1(X_1) \times \cdots \times h_n(X_n)] \\ &= E[h_1(X_1)] \times \cdots \times E[h_n(X_n)]. \end{aligned} \tag{3}$$

Fix $X_1, \ldots, X_n \in \mathcal{X}$ and $g_1, \ldots, g_n \in \mathcal{G}$. Then $\forall j = 1, \ldots, n$, there is a sequence $\{f_{j1}, f_{j2}, \ldots\} \subset \mathcal{F}$ such that $f_{jk} \to g_j$ a.e. F_j as $k \to \infty$ and $|f_{jk}| \leq M_j < \infty$ a.e. F_j, where $X_j \sim F_j$. Since $|f_{jk}(X_j)| \overset{\text{a.s.}}{\leq} M_j < \infty$ and $f_{jk}(X_j) \overset{\text{a.s.}}{\to} g_j(X_j)$, it follows that $E[f_{jk}(X_j)] \to E[g_j(X_j)]$ by the dominated convergence theorem[16] (DCT). Also $|E[g_j(X_j)]| \leq E|g_j(X_j)| < \infty$, where the latter inequality is guaranteed by the DCT. Therefore,

$$\begin{aligned} &E[f_{1k}(X_1)] \times \cdots \times E[f_{nk}(X_n)] \\ &\to E[g_1(X_1)] \times \cdots \times E[g_n(X_n)] \end{aligned} \tag{4}$$

as $k \to \infty$, because the product on the left must converge to the product of the finite limits of its factors.[17]

Similarly, the bounds on the functions give $|f_{1k}(X_1) \times \cdots \times f_{nk}(X_n)| \overset{\text{a.s.}}{\leq} M_1 \times \cdots \times M_n < \infty$. Since the f_{jk} are bounded a.e. F_j, so are the g_j, and therefore $f_{1k}(X_1) \times \cdots \times f_{nk}(X_n) \overset{\text{a.s.}}{\to} g_1(X_1) \times \cdots \times g_n(X_n)$. The DCT then implies that $E[f_{1k}(X_1) \times \cdots \times f_{nk}(X_n)]$ converges to $E[g_1(X_1) \times \cdots \times g_n(X_n)]$. The sequences here and in (4) are equal by (3), so their limits are also equal. Functions $g_1, \ldots, g_n \in \mathcal{G}$ were arbitrary, and $\mathcal{G}$ is an independence class w.r.t. $\mathcal{X}$, so $X_1, \ldots, X_n$ are independent. Then $\mathcal{F}$ is an independence class with respect to $\mathcal{X}$ because (3) implies independence for arbitrary $X_1, \ldots, X_n \in \mathcal{X}$. □

If bounded functions converge to the indicators a.e. with respect to the probability distributions, then Theorem 3 and Lemma 4 apply. The complex-valued continuous functions with compact support on X are denoted $C_c(X)$ in this corollary[18] to Lusin's theorem:[19]

Corollary 5 *(Lusin) Suppose that μ is a Radon measure[20] on a locally compact Hausdorff space X and that g is a complex measurable function on X with $|g| \leq 1$ and $\mu(\{x : g(x) \neq 0\}) < \infty$. Then there is a sequence $\{f_k\}$ with $f_k \in C_c(X)$ and $|f_k| \leq 1$ such that $f_k \to g$ a.e. μ.*

Does the corollary apply? Space X here is just $\mathbb{R}$ under the usual topology, which makes it a locally compact Hausdorff space.[21]

[16] Billingsley [2, Thm 5.3, page 72, or Theorem 16.4, page 213], Chung [3, (viii), page 42], or Rudin [5, Thm 1.34, p. 26].

[17] Rudin [6, Thm 3.3(c), p. 49].

[18] Rudin [5, p. 56].

[19] Rudin [5, Thm 2.24, p. 55] or Folland [4, Thm 7.10, p. 211].

[20] Folland [4, p. 205].

[21] Every metric space is Hausdorff [5, p. 36], and the real line is locally compact by the Heine-Borel theorem [6, Thm 2.41, p. 40].

Measure μ is the Lebesgue-Stieltjes probability distribution (measure) corresponding[22] to F, the distribution function of random variable X. Since μ is regular[23] and finite, it is Radon.[24] The corollary's function g here is suitably bounded indicator 1_C. Since μ is finite, the support of g is trivially finite. So the corollary's conditions are satisfied and the Lemma 4 sequences exist. The real parts of the corollary's complex functions inherit the latter's properties, so the corollary ensures a real function sequence for a real limit function. From Theorem 3, Lemma 4, and Corollary 5 then:

Theorem 6 *The class of real functions in $C_c(\mathbb{R})$ forms an independence class w.r.t. any family of random variables.*

The celebrated Stone-Weierstrass Theorem will further narrow the function class. Some terminology is needed. The real linear space of real-valued continuous functions on compact Hausdorff space X is denoted by $C(X, \mathbb{R})$, and any subspace $\mathcal{A}$ closed under pointwise multiplication is a *subalgebra* of $C(X, \mathbb{R})$. A subset $\mathcal{S}$ of $C(X, \mathbb{R})$ *separates points* if for every $x, y \in X$ with $x \neq y$, some $f \in \mathcal{S}$ has $f(x) \neq f(y)$. This straightforward corollary of the Stone-Weierstrass Theorem[25] generalizes the famous Weierstrass Approximation Theorem[26] and is all that is needed here:

Corollary 7 *(Stone-Weierstrass) Let X be a compact Hausdorff space. If $\mathcal{A}$ is a subalgebra of $C(X, \mathbb{R})$ that contains the constant functions and separates points, then for every $g \in C(X, \mathbb{R})$, there is a sequence of functions $\{f_k\}$ in $\mathcal{A}$ such that $f_k \to g$ uniformly as $k \to \infty$.*

With Lemma 4 and Theorem 6 this corollary gives the following key theorem, in which $\mathcal{A}|_K$ indicates the family $\{f|_K : f \in \mathcal{A}\}$ of the restrictions to $K \subset \mathbb{R}$ of the functions in $\mathcal{A}$.

Theorem 8 *Let $\mathcal{X}$ be a class of random variables uniformly bounded a.s. on a common probability space. Let $\mathcal{A}$ be a family of real Borel functions on $\mathbb{R}$ with $\mathcal{A}|_K$ a subalgebra of $C(K, \mathbb{R})$ that contains the constant functions and separates points and with $K \subset \mathbb{R}$ a compact neighborhood[27] of zero. Then there is a positive α such that $\forall X \in \mathcal{X}$, $\alpha X \in K$ a.s. Further, for any such α, $\{f : f(x) = h(\alpha x), h \in \mathcal{A}\}$ is an independence class w.r.t. $\mathcal{X}$.*

PROOF: Let $\mathcal{A}$, K, and $\mathcal{X}$ be as above with $b > 0$ such that $\forall X \in \mathcal{X}$, $|X| \leq b$ a.s. Neighborhood K of zero contains[28] a closed interval $[-a, a]$ for some $a > 0$. Fix $\alpha = a/b$; then $|\alpha X| = \frac{a}{b}|X| \overset{\text{a.s.}}{<} a$, so $\alpha X \overset{\text{a.s.}}{\in} [-a, a] \subset K$.

The second conclusion of the theorem is based on Lemma 4. Consider $X \in \mathcal{X}$ with $X \sim F$ and a real-valued $g \in C_c(\mathbb{R})$. Define $g' \in C(K, \mathbb{R})$ by $g'(x) = g(\frac{x}{\alpha})$. By Corollary 7, there is a sequence $\{h_k\}$ in $\mathcal{A}$ with $h_k \to g'$ uniformly on K. Let Borel function f_k on $\mathbb{R}$ be such that $f_k(x) = h_k(\alpha x)$ for $x \in \mathbb{R}$. Then $f_k(x) = h_k(\alpha x) \to g'(\alpha x) = g(x)$ uniformly on $\{x : \alpha x \in K\}$. But $\alpha X \overset{\text{a.s.}}{\in} K$, so $f_k(X) \overset{\text{a.s.}}{\to} g(X)$, and $f_k \to g$ uniformly a.e. F.

Also required is a uniform bound on the $\{f_k\}$ a.e. F. Each h_k is continuous on compact set K and so[29] is bounded on K, and f_k is bounded on $\{x : \alpha x \in K\}$. But $\alpha X \in K$ a.s., so for some $M_k < \infty$, $f_k(X) \leq M_k$ a.s., so f_k is essentially bounded[30] w.r.t. F. The family of such functions is a Banach space[31] with the essential-supremum norm $\|\cdot\|_\infty$. But $g \in C(K, \mathbb{R})$ implies[32] $\|g\|_\infty < \infty$, so $f_k \to g$ uniformly a.e. F yields convergence in the norm trivially, and the boundedness of the norms follows.[33] So for some $M < \infty$, $\|f_k\|_\infty \leq M$ and hence $|f_k| \leq M$ a.e. F. The theorem follows from Lemma 4. ☐

In this theorem, the functions in $\mathcal{A}$ are horizontally scaled to form an independence class. If a horizontal shift is permitted as well, then K need not contain a neighborhood of zero.

The following lemma will permit "the linear span of $\mathcal{A}|_K$" to replace "$\mathcal{A}|_K$" in Theorem 8, greatly reducing the size of the independence classes. This lemma is also useful with Theorem 3.

Lemma 9 *If the linear span of $\mathcal{F}$ is an independence class with respect to $\mathcal{X}$, then so is $\mathcal{F}$.*

PROOF: Fix finite family $\mathcal{Y} \subset \mathcal{X}$ and write $\mathcal{Y} = \{X_1, \ldots, X_n\}$ for convenience. Let $\mathcal{G}_\mathcal{Y}$ be the class of Borel functions with $E[g_1(X_1) \times \cdots \times g_n(X_n)] = E[g_1(X_1)] \times \cdots \times E[g_n(X_n)]$ for all $g_1, \ldots, g_n \in \mathcal{G}_\mathcal{Y}$. Then $\mathcal{G}_\mathcal{Y}$ is closed under scalar multiplication, because

$$
\begin{aligned}
&E[cg_1(X_1) \times \cdots \times g_n(X_n)] \\
&= E[cg_1(X_1)] \times E[g_2(X_2)] \times \cdots \times E[g_n(X_n)].
\end{aligned}
$$

Also, $\mathcal{G}_\mathcal{Y}$ is closed under addition, because $h, g_1, \ldots, g_n \in \mathcal{G}_\mathcal{Y}$ implies

$$
\begin{aligned}
&E[(h(X_1) + g_1(X_1)) \times g_2(X_2) \times \cdots \times g_n(X_n)] \\
&= E[h(X_1) + g_1(X_1)] \\
&\quad \times E[g_2(X_2)] \times \cdots \times E[g_n(X_n)]\,.
\end{aligned}
$$

So $\mathcal{G}_\mathcal{Y}$ is closed under linear combination.

Now suppose that the linear span of $\mathcal{F}$ is an independence class with respect to $\mathcal{X}$. But $\mathcal{F} \subset \mathcal{G}_\mathcal{Y} \Rightarrow \text{span}\{\mathcal{F}\} \subset \mathcal{G}_\mathcal{Y} \Rightarrow$ the members of $\mathcal{Y}$ are independent. But $\mathcal{Y} \subset \mathcal{X}$ was arbitrary, so $\mathcal{F}$ is an independence class w.r.t. $\mathcal{X}$. ☐

Theorem 10 *Let class $\mathcal{X}$ of random variables be uniformly bounded a.s. on a common probability space. Let $\mathcal{A}$ be a family of real Borel functions on $\mathbb{R}$ for which the linear span of $\mathcal{A}|_K$ includes a subalgebra of $C(K, \mathbb{R})$ that contains the constant functions and separates points, where $K \subset \mathbb{R}$ is a compact neighborhood[34] of zero. Then there is a positive α such that $\forall X \in \mathcal{X}$, $\alpha X \in K$ a.s. Further, for any such α, $\{f : f(x) = h(\alpha x), h \in \mathcal{A}\}$ is an independence class w.r.t. $\mathcal{X}$.*

This is just Theorem 8 modified by Lemma 9, and a corollary yields the obvious independence classes of continuous functions:

[22] Folland [4, Thm 1.16, p. 34].
[23] Folland [4, Thm 1.18, p. 35].
[24] Folland [4, p. 205].
[25] Folland [4, Thm 4.45, p. 133].
[26] Folland [4, Thm 4.50, p. 135].
[27] Folland [4, p. 108].
[28] Since K is neighborhood of 0, $\exists$ open $V \subset K$ with $0 \in V$. On the real line, open sets are unions of disjoint open intervals, so $\exists c, d > 0 \ni$: $0 \in (c, d) \subset V \subset K$. Take $a = \frac{1}{2}\min\{c, d\}$.
[29] Rudin [6, Thm 4.15, p. 89].
[30] Rudin [5, p. 66].
[31] Rudin [5, Thm 3.11, p. 67].
[32] Rudin [6, Thm 4.15, p. 89].
[33] Rudin [6, Thm 3.2(c), p. 48].
[34] Folland [4, p. 108].

No.	independence class $\mathcal{A}$	span of $\mathcal{A}$
1	$\{1, x, x^2, \ldots\}$	polynomials in x
2	Chebychev polynomials in x	polynomials in x
3	$\{e^{-nx} : n = 0, 1, \ldots\}$	polynomials in e^{-x}
4	$\{1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots\}$	trig polynomials in x
5	$\{e^{inx} : n = 0, 1, \ldots\}$	polynomials in e^{ix}
6	$\{1, f(x), f^2(x), \ldots\}$ (f monotonic)	polynomials in $f(x)$

Table I: Examples of Theorem 10 independence classes.

Corollary 11 *On a common probability space let $\mathcal{X}$ be a class of random variables with range a.s. contained in closed interval I. Let $\mathcal{A}$ be a family of real, continuous functions on I that separates points. Then the (algebraic) closure of $\mathcal{A}$ under pointwise multiplication, when the unit constant function* 1 *is adjoined, is an independence class w.r.t. $\mathcal{X}$.*

The family $\mathcal{A}$ in Corollary 11 might contain, in the simplest case, a single strictly monotonic, continuous function!

Theorem 3 modified by Lemma 9 becomes Theorem 12, which with Theorem 10 and Corollary 11 is the paper's primary result:

Theorem 12 *Let $\mathcal{X}$ be a class of random variables on a common probability space, and let $Y \in \mathbb{B}$ be such that $\forall X \in \mathcal{X}, X \overset{\text{a.s.}}{\in} Y$. If π-system $\mathcal{C}$ generates $\mathbb{B} \cap Y$ and if indicator functions $\{1_C : C \in \mathcal{C}\}$ are included in the linear span of family $\mathcal{F}$ of real-valued Borel functions on $\mathbb{R}$, then $\mathcal{F}$ is an independence class w.r.t. $\mathcal{X}$.*

III. EXAMPLES

Classes of indicator functions meeting the Theorem 12 criteria are easily found for bounded random variables. The dyadic intervals $\mathcal{D} = \{(k2^{-n}, (k+1)2^{-n}] : n, k \in \mathbb{Z}\}$, for example, are a π-system that generates the Borels, so taking Y to be some interval $(-2^N, 2^N]$ so that $\mathcal{C} = \mathcal{D} \cap Y$ comprises those dyadic intervals contained in Y makes the indicator funtions of the sets in $\mathcal{C}$ an independence class. Further, these functions are in the span of an independence class of square-wave basis functions related[35] to Walsh functions.[36] They can take values in $\{\pm 1\}$ or $\{0, 1\}$ (as 1_Y is already in the basis). An independence class can be similarly constructed from any integer-coefficient basis $W_1, W_2, \ldots$ for $\mathbb{R}^n$.

Table I lists some continuous-function classes that meet the Theorem 10 requirements; for any given family of uniformly bounded random variables, each is within horizontal scaling of an independence class. The first example says, for example, that independence of bounded random variables X and Y is equivalent to decorrelation of all (nonnegative) integral powers X^n and Y^m. The second example reminds us that there are many generators of the polynomials. Replacing n by a general real λ in the third example gives a family of functions on any compact subset of $\mathbb{R}$ that can be used to prove the uniqueness of the Laplace transform or moment-generating function.[37] Here it says that a separable (factorable) vector moment-generating function (on negative integer arguments) implies component independence for a suitably bounded random vector. Changing "real Borel functions" to "complex Borel functions" in Definition 1 makes example 4 and example 5 equivalent, because expectation factors for both classes if it factors for either. Example 5 is related to a proof of characteristic-function uniqueness,[38] and it tells us that if the vector characteristic function (on positive integer arguments) is separable for a suitably bounded random vector, then its elements are independent. Examples 1, 3, and 5 suggest example 6, in which any f for which $\mathcal{A}$ separates points will do, as this implies that f must be strictly monotonic. Examples 1 and 3 are special cases of example 6, and the independence classes in examples 1, 3, 5, and 6 can each be described using a single generator, as in Corollary 11. Example 4 is covered by Corollary 11 also, but it requires both $\cos x$ and $\sin x$ as generators in order to separate points in $[-\pi, \pi - \epsilon]$. Example 2 is not covered by the corollary, because the Chebychev polynomials are not the closure under multiplication of a smaller set. They are not even closed under multiplication. But the linear span of the Chebychev polynomials equals that of the class $\mathcal{A}$ of example 1, which is a subalgebra of the requisite sort specified in Theorem 10.

[35] Horizontally shift and scale left-continuous Walsh functions.

[36] de Coulon [7, p. 68].

[37] Chung [3, Thm 6.6.2, p. 189].

[38] Billingsley [2, Problem 26.19, p. 365].

IV. SUMMARY

If expectation distributes over products of functions of random variables for all functions in an *independence class* (w.r.t. an appropriate class of random variables), then the random variables are independent. When the functions are powers, this says only that factorable higher-order joint moments imply independence. But higher-order moments themselves are not distinguished in any way. For bounded random variables, this paper offers small independence classes and thus a variety of alternatives to higher-order-moments for establishing independence via expectation factorization. Necessary conditions for an independence class remain undiscovered, but sufficient conditions here yield interesting classes of indicator functions and of continuous functions.

ACKNOWLEDGEMENTS

A discussion of [8] with David Feinstein motivated this work. The careful reading by Eric Mokole is much appreciated.

REFERENCES

[1] M. Loève, *Probability Theory I*, Springer-Verlag, New York, fourth edition, 1977.

[2] P. Billingsley, *Probability and Measure*, Wiley, New York, second edition, 1986.

[3] K. L. Chung, *A Course in Probability Theory*, Academic Press, Orlando, second edition, 1974.

[4] Gerald B. Folland, *Real Analysis*, Wiley, New York, 1984.

[5] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, third edition, 1987.

[6] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York, third edition, 1976.

[7] F. de Coulon, *Signal Theory and Processing*, Artech House, Dedham, MA, 1986.

[8] E. A. Vittoz and X. Arreguit, "CMOS integration of Herault-Jutten cells for separation of sources," in *Analog VLSI Implementation of Neural Systems*, C. Mead and M. Ismail, Eds., 1989, (This is the proceedings of the Pre-conference Workshop on Analog Integrated Neural Systems, 1989 IEEE International Symposium on Circuits and Systems, Portland, Oregon, May 8, 1989.).