Some Bivariate Invariance Principles

Alexander Mariona, Homa Esfahanizadeh, Rafael G. L. D'Oliveira, and Muriel Médard
Research Laboratory of Electronics, Massachusetts Institute of Technology, USA
Email: {amariona, homaesf, rafaeld, medard}@mit.edu

***Abstract*—The Basic Invariance Principle is a nonlinear generalization of the Berry-Esseen Theorem which bounds the error in approximating the value of a Boolean function of an n-length sequence of random variables by the function of a differently distributed n-length sequence. We present two generalizations of the Basic Invariance Principle for a class of bivariate Boolean functions computed on two n-length sequences which we term separable functions. The first generalization yields a deterministic bound. The second generalization provides a bound that holds with a probability that can be made arbitrarily close to 1 at the cost of a constant factor. We introduce the notion of the participation of a coordinate on a given Boolean function, which is the number of terms in which that coordinate appears. When the maximum participation of a sequence of functions with increasing input length n remains constant, the probabilistic bound significantly improves upon the deterministic bound, regardless of the degree of the function. Indeed, there exist sequences of functions for which the deterministic bound does not converge to 0 while the probabilistic bound does.**

***Index Terms*—Basic Invariance Principle, Boolean functions, functional approximation**

I. Introduction

Boolean functions are ubiquitous in the fields of complexity theory [1], [2], cryptography [3], [4], social choice theory [5], [6], and digital electronics [7], [8]. In this work, we treat real-valued Boolean functions, which take the form $f : \{-1, 1\}^n \to \mathbb{R}$.[1]

One particularly significant result from the field of analysis of Boolean functions is the Basic Invariance Principle (BIP) [9]. The BIP is a nonlinear generalization of the Berry-Esseen Theorem [10], [11], which is in turn a quantitative version of the Central Limit Theorem. The Berry-Esseen Theorem provides an explicit bound on the difference between the distribution of a finite sum of independent, arbitrarily distributed random variables and a standard Gaussian distribution. Given a real-valued Boolean function $f : \{-1, 1\}^n \to \mathbb{R}$ and two sequences of independent, distinctly distributed random variables $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$, the BIP provides an explicit bound on the expected difference between $f(\mathbf{X})$ and $f(\mathbf{Y})$. This difference can be interpreted as the expected error incurred by approximating $f(\mathbf{X})$ as $f(\mathbf{Y})$. The BIP is quite general and holds for many different notions of what it means for the distributions of the random variables $F(\mathbf{X})$ and $F(\mathbf{Y})$ to be "close."

In order for the bound given by the BIP to be close to 0, the Boolean function under consideration must have relatively low influences. The influence of a coordinate on a Boolean function quantifies how sensitive the output is to a change in that particular input coordinate. The notion of influence originated in social choice and voting theory [12]. Qualitatively, the BIP states that low-influence functions are *invariant* to the distributions of the input sequences.

The BIP treats functions of a single sequence of random variables. A natural generalization of the BIP would be an invariance principle which treats functions of two sequences of random variables, i.e. functions of the form $F : \{-1, 1\}^n \times \{-1, 1\}^n \to \mathbb{R}$. We present two distinct bivariate invariance principles for a particular class of functions which we call *separable* functions. A separable function takes the form $F(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + g(\mathbf{y}) + h(\mathbf{xy})$. Separable functions are interesting because they generalize many different notions of noise that arise in communication channels.

One promiment example of a such channel is the binary symmetric channel (BSC) [13, ch. 7]. The BSC independently flips each bit in a binary sequence with a constant probability. This noisy communication process can be represented as the elementwise multiplication of two Boolean sequences.[2] Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the input sequence and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be the noise effect sequence. Then, the output sequence is $\mathbf{Z} = \mathbf{XY} = (X_1Y_1, \ldots, X_nY_n)$. A real-valued Boolean function $h : \{-1, 1\}^n \to \mathbb{R}$ can be used to describe any computation $h(\mathbf{Z})$ on the ouptut sequence. Such a function can be rewritten as a separable function of the form $F(\mathbf{X}, \mathbf{Y}) = h(\mathbf{XY})$ which depends on the input to and the noise effect of the BSC.

As another example, consider the computational wiretap channel [14], in which the transmitter has some sensitive data $\mathbf{X}$ and transmits a computation $f(\mathbf{X})$ on that data. An eavesdropper who is interested in some other computation $g(\mathbf{X})$ sees $f(\mathbf{X})$ and uses it to produce an estimate $\hat{g}(\mathbf{X})$. Such an estimate can be approximated with the separable function $F(\mathbf{X}, \mathbf{Y}) = g(\mathbf{X}) + f(\mathbf{Y}) - g(\mathbf{Y})$, where $\mathbf{Y}$ is a sequence used to approximate noise.

Directly analogous to the BIP, the bivariate invariance principles which we present provide an explicit bound on the expected difference between $F(\mathbf{X}^1, \mathbf{X}^2)$ and $F(\mathbf{Y}^1, \mathbf{Y}^2)$. The first invariance principle, which we denote BVIP1, is a direct extension of the BIP and treats the two sequences

This work was supported by MIT Welcome Trust Fellowship 2389724, Lincoln Lab Grant 6944494, MIT Portugal Grant 6942770, and NFS-CNS Grant 6944400.

[1] We follow the standard convention in analysis of Boolean functions of representing bits using -1 and 1. A representation using 0 and 1 is equivalent, but requires more complicated notation.

[2] Since we are using the -1 and 1 bit representation, multiplication by -1 results in a bit flip while multiplication by 1 leaves the input bit unchanged.

as a single sequence of random variables, while the second invariance principle, BVIP2, considers the two input sequences separately.

BVIP1 produces a deterministic bound, while BVIP2 produces a bound which holds in probability, but that probability can be made arbitrarily close to 1 at the cost of a constant factor. The bound of BVIP1 has a strong exponential dependence on the degree of the function of interest,[3] while the bound of BVIP2 has a weaker exponential dependence on the degree, along with an additional factor which depends on the number of terms[4] in the function of interest. These differences result in different asymptotic behaviors for the bounds produced by BVIP1 and BVIP2. We show that when the degree of the function is linear in the size of the input, the bound of BVIP2 is asymptotically stronger than that of BVIP1.

To characterize how the growth in the number of terms of the function affects the behavior of BVIP2, we introduce the concept of the *participation* of a coordinate on a Boolean function. The participation of a coordinate is equal to the number of terms in which it appears. Like the notion of influence, participation can be viewed as a way to quantify the extent to which a particular coordinate affects the output of a function. By limiting the maximum participation on the function of interest, the bound produced by BVIP2 can be tightened, reducing an exponential factor to a polynomial factor. With conditions on the maximum participation, the bound of BVIP2 can be made asymptotically stronger than that of BVIP1 even when the degree of the function is sublinear in the size of the input.

For both invariance principles, as with the BIP, the bounds can be tightened (i.e. brought closer to 0) by decreasing the influences of the function. We show that there exists sequence of functions for which sequences for which the bound of BVIP2 converges to 0 while that of BVIP1 diverges. In other words, there exists sequences for which BVIP1 cannot be used to bound the error in approximating $F(\mathbf{X}^1, \mathbf{X}^2)$ as $F(\mathbf{Y}^1, \mathbf{Y}^2)$, but for which BVIP2 states that the approximation error converges to 0. The convergence of BVIP2 does not depend on the constant factor which controls the probability with which the bound holds.

The remainder of this paper is organized as follows. In Section II we summarize relevant concepts and results from analysis of Boolean functions. In Section III and Section IV we present BVIP1 and BVIP2. In Section V we characterize the convergence behavior of each bound and illustrate that behavior with examples. We conclude in Section VI.

II. Preliminaries

We first summarize our conventions and notation. We denote random variables with uppercase letters, e.g. X. We denote vectors with bold-faced letters, e.g. $\mathbf{x}$. Accordingly, vectors (often referred to as sequences in our context) of random variables are denoted by uppercase bold-faced letters, e.g. $\mathbf{X}$. The coordinates, or elements, of vectors are denoted by indexed letters, e.g. x_i. We write out a vector in terms of its coordinates as $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Multiplication of two vectors is performed elementwise and results in a new vector:

$$\mathbf{xy} = (x_1y_1, x_2y_2, \ldots, x_ny_n).$$

The expectation of a random variable X is denoted by $\mathrm{E}[X]$. The notation $S \ni i$ describes a set S containing the element i. The notation $[n]$ describes the set $\{1, 2, \ldots, n\}$. The notation $2^{[n]}$ describes the power set of $[n]$.

We now summarize some fundamental results from analysis of Boolean functions [15]. The first theorem describes how Boolean functions can be written as multilinear polynomials.

Theorem 1. *Every Boolean function $f : \{-1,1\}^n \to \mathbb{R}$ can be uniquely expressed as a n-variate multilinear polynomial:*

$$f(\mathbf{x}) = \sum_{S \subseteq [n]} \hat{f}(S) \prod_{i \in S} x^i.$$

This expression is called the Fourier expansion *of f, and the real number $\hat{f}(S)$ is called the* Fourier coefficient *of f on S, with $\hat{f} : 2^{[n]} \to \mathbb{R}$. Collectively, the coefficients are called the* Fourier spectrum *of f.*

The degree of a Boolean function is the degree of its Fourier expansion. Since every Fourier expansion is multilinear, the degree k of a Boolean function is

$$k = \max_{\hat{f}(S) \neq 0} |S|.$$

An important property of a Boolean function is the *influence* of each coordinate on the output of the function. It is useful to first define the notion of a *pivotal* coordinate.

Definition 1. For a given $f : \{-1,1\}^n \to \{-1,1\}$, a particular coordinate $i \in [n]$ is *pivotal* on a particular input $\mathbf{x}$ if $f(\mathbf{x}) \neq f(\mathbf{x}^{\oplus i})$, where

$$\mathbf{x}^{\oplus i} = (x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n).$$

In words, a coordinate is pivotal for a particular input if changing that coordinate changes the output of the function. The influence of a coordinate quantifies how likely a particular coordinate is to be pivotal.

Definition 2. The *influence* of a coordinate i on a function $f : \{-1,1\}^n \to \{-1,1\}$ is defined to be the probability that i is pivotal for a random input:

$$\mathbf{Inf}_i[f] = \Pr_{\mathbf{x} \sim \{-1,1\}^n} \left[f(\mathbf{x}) \neq f(\mathbf{x}^{\oplus i}) \right].$$

Informally, if we consider f to be a voting rule in a two-party election, the influence of the ith coordinate can be thought of as the "influence" or "power" of the ith voter. The influences on a real-valued Boolean function can be defined in a more analytical fashion, but with a very similar meaning [15, ch. 2]. Such an approach leads to the next result relating the influences to the Fourier spectrum, which we treat as a definition.

[3] Every Boolean function can be written as a multilinear polynomial, and hence the degree of every Boolean function is well-defined.

[4] As with the degree, the number of terms in every Boolean function is well-defined for the same reason.

Definition 3. For a function $f : \{-1,1\}^n \to \mathbb{R}$ and $i \in [n]$, the influence of i on f is defined to be $\mathbf{Inf}_i[f] = \sum_{S \ni i} \hat{f}(S)^2$.

The following definitions describe two ways to capture the combined influence of all coordinates.

Definition 4. The *total influence* of $f : \{-1,1\}^n \to \mathbb{R}$ is defined to be $\mathbf{I}[f] = \sum_{i=1}^n \mathbf{Inf}_i[f]$.

Definition 5. The *total square influence* of $f : \{-1,1\}^n \to \mathbb{R}$ is defined to be $\mathbf{S}[f] = \sum_{i=1}^n \mathbf{Inf}_i[f]^2$.

We now present a few statements in anticipation of the BIP. First, the BIP only holds for sequences of random variables with certain distributions. In particular, each random variable in the sequence must satisfy the following assumption.

Assumption 1. The random variable X_i satisfies $\mathrm{E}[X_i] = 0$, $\mathrm{E}\left[X_i^2\right] = 1$, $\mathrm{E}\left[X_i^3\right] = 0$, and $\mathrm{E}\left[X_i^4\right] \leq 9$.

Two examples of random variables satisfying Assumption 1 are a uniform ± 1 random bit and a standard Gaussian.

The following two lemmas are useful for proving the BIP using the replacement method, which is also the method we adopt when proving BVIP2 in Section IV.

Lemma 1 (Bonami Lemma [16]). *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a n-variate multilinear polynomial with degree at most k and let $\mathbf{X}$ be a sequence of independent random variables satisfying Assumption 1. Then,* $\mathrm{E}\left[f(\mathbf{X})^4\right] \leq 9^k \cdot \mathrm{E}\left[f(\mathbf{X})^2\right]^2$.

Lemma 2 (Parseval [15, ch. 1]). *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a n-variate multilinear polynomial and let $\mathbf{X}$ be a sequence of independent random variables satisfying Assumption 1. Then,* $\mathrm{E}\left[f(\mathbf{X})^2\right] = \sum_{S \subseteq [n]} \hat{f}(S)^2$.

This leads us to the formal statement of the BIP.

Theorem 2 (Basic Invariance Principle). *Let the function $f : \{-1,1\}^n \to \mathbb{R}$ be a n-variate multilinear polynomial of degree at most $k \in \mathbb{N}$. Let $\mathbf{X}$ and $\mathbf{Y}$ be n-length sequences of independent random variables, each satisfying Assumption 1. Let $\psi : \mathbb{R} \to \mathbb{R}$ be $\mathcal{C}^4$ (i.e. the derivatives $\psi', \ldots, \psi''''$ exist and are continuous) with $\|\psi''''\|_\infty \leq C$. Then*

$$|\mathrm{E}[\psi(f(\mathbf{X}))] - \mathrm{E}[\psi(f(\mathbf{Y}))]| \leq \frac{C}{12} \cdot 9^k \cdot \mathbf{S}[f].$$

The function ψ used in the BIP is called a *test function* or a *distinguisher*, and is used to specify a particular notion of "closeness" between two random variables. A natural measure is *cdf-closeness*, which is used in the Berry-Esseen Theorem. Two random variables X and Y are cdf-close if $\Pr\{X \leq u\} \approx \Pr\{Y \leq u\}$ for all $u \in \mathbb{R}$. Equivalently, two random variables are cdf-close if $\mathrm{E}[\psi(X)] \approx \mathrm{E}[\psi(Y)]$ for $\psi(s) = 1_{s \leq u}$. The BIP is powerful enough to give bounds on cdf-closeness and many other notions of closeness.[5]

[5]Note that continuity of the fourth derivative of ψ is required. Smoothing techniques can be used to approximate functions like $\psi(s) = 1_{s \leq u}$, with the error incurred proportional to the fourth derivative of the smoothed function [15, ch. 11].

For convenience, we define a shorthand notation for the quantity with which the BIP is concerned. For $\psi : \mathbb{R} \to \mathbb{R}$, let

$$\mathrm{E}_\psi[X; Y] = |\mathrm{E}[\psi(X)] - \mathrm{E}[\psi(Y)]|,$$

Finally, we define the class of bivariate functions which will be treated in this work. We call these functions *separable*, owing to the fact that we can separate the dependences on the first variable, the second variable, and the product of the two.

Definition 6. A real-valued Boolean function of two variables $F : \{-1,1\}^n \times \{-1,1\}^n \to \mathbb{R}$ is *separable into f, g, and h* if it can be written in terms of real-valued Boolean functions of one variable $f, g, h : \{-1,1\}^n \to \mathbb{R}$ like

$$F(\mathbf{x}^1, \mathbf{x}^2) = f(\mathbf{x}^1) + g(\mathbf{x}^2) + h(\mathbf{x}^1\mathbf{x}^2).$$

Note that if F is separable into f, g, and h which are all of degree at most k, then F is of degree at most $2k$.

III. A Deterministic Bivariate Invariance Principle

In this section, we present our first bivariate invariance principle, which we refer to as BVIP1. This invariance principle is a direct extension of the BIP. Given a real-valued bivariate Boolean function $F : \{-1,1\}^n \to \mathbb{R}$, two input sequences $\mathbf{X}^1$ and $\mathbf{X}^2$, and two replacement sequences $\mathbf{Y}^1$ and $\mathbf{Y}^2$, we derive BVIP1 by considering the concatenations $\mathbf{X}$ of $\mathbf{X}^1$ and $\mathbf{X}^2$ and $\mathbf{Y}$ of $\mathbf{Y}^1$ and $\mathbf{Y}^2$. The $2n$-length sequences $\mathbf{X}$ and $\mathbf{Y}$ are used as inputs to a function $G : \{-1,1\}^{2n} \to \mathbb{R}$ which is equivalent to F. We then apply the BIP to $G(\mathbf{X})$ and $G(\mathbf{Y})$.

Theorem 3 (BVIP1). *Let $F : \{-1,1\}^n \times \{-1,1\}^n \to \mathbb{R}$ be a multilinear polynomial separable into f, g, and h such that f, g, and h are all of degree at most k. Let $\mathbf{X}^1$, $\mathbf{X}^2$, $\mathbf{Y}^1$, and $\mathbf{Y}^2$ be n-length sequences of independent random variables satisfying Assumption 1. Let $F_{\mathbf{XX}} = F(\mathbf{X}^1, \mathbf{X}^2)$ and $F_{\mathbf{YY}} = F(\mathbf{Y}^1, \mathbf{Y}^2)$. Assume $\psi : \mathbb{R} \to \mathbb{R}$ is $\mathcal{C}^4$ with $\|\psi''''\|_\infty \leq C$. Then*

$$\mathrm{E}_\psi[F_{\mathbf{XX}}; F_{\mathbf{YY}}] \leq \frac{C}{6} \cdot 9^{2k} \cdot (\mathbf{S}[f] + \mathbf{S}[g] + 2\,\mathbf{S}[h]).$$

Proof. Define the random variables $\mathbf{X}$ and $\mathbf{Y}$ to be

$$\mathbf{X} = (X_1^1, \ldots, X_n^1, X_1^2, \ldots, X_n^2)$$
$$\mathbf{Y} = (Y_1^1, \ldots, Y_n^1, Y_1^2, \ldots, Y_n^2).$$

Define the function $G : \{-1,1\}^{2n} \to \mathbb{R}$ to be

$$\begin{aligned} G(\mathbf{x}) &= F((x_1, \ldots, x_n), (x_{n+1}, \ldots, x_{2n})) \\ &= f((x_1, \ldots, x_n)) + g((x_{n+1}, \ldots, x_{2n})) \\ &\quad + h((x_1 x_{n+1}, \ldots, x_n x_{2n})). \end{aligned}$$

By definition, the function G is of degree at most $2k$. Note that $G(\mathbf{X}) = F_{\mathbf{XX}}$ and $G(\mathbf{Y}) = F_{\mathbf{YY}}$.

In order to apply the BIP to $G(\mathbf{X})$ and $G(\mathbf{Y})$, we must first bound the influences of G, which involves determining the Fourier coefficients of G. In the following argument, use S to refer to subsets of $[n]$ and T to refer to subsets of $[2n]$.

For a given set S, let $S^+ = \{i+n : i \in S\}$. For any given T, there are four possible values for $\hat{G}(T)$:

1) If $\exists S$ such that $T = S$, then $\hat{G}(T) = \hat{f}(S)$.
2) If $\exists S$ such that $T = S^+$, then $\hat{G}(T) = \hat{g}(S)$.
3) If $\exists S$ such that $T = S \cup S^+$, then $\hat{G}(T) = \hat{h}(S)$.
4) Otherwise, $\hat{G}(T) = 0$.

Hence, if $i \in [1, n]$, the value of $\mathbf{Inf}_i[G]$ only depends on $\mathbf{Inf}_i[f]$ and $\mathbf{Inf}_i[h]$. Similarly, if $i \in [n+1, 2n]$, the value of $\mathbf{Inf}_i[G]$ only depends on $\mathbf{Inf}_i[g]$ and $\mathbf{Inf}_i[h]$.

More explicitly, we can bound the influence of the ith coordinate on G as follows. First, assume $i \in [n]$. Then,

$$\begin{aligned}\mathbf{Inf}_i[G] &= \sum_{T \ni i} \hat{G}(T)^2 \\ &= \sum_{T=S\ni i} \hat{f}(S)^2 + \sum_{T=S\cup S^+ \ni i} \hat{h}(S)^2 \\ &= \sum_{S \ni i} \hat{f}(S)^2 + \sum_{S\ni i} \hat{h}(S)^2 \\ &= \mathbf{Inf}_i[f] + \mathbf{Inf}_i[h].\end{aligned}$$

Similarly, if $i \in [n+1, 2n]$,

$$\mathbf{Inf}_i[G] = \mathbf{Inf}_i[g] + \mathbf{Inf}_i[h].$$

We can then bound $\mathbf{S}[G]$ as

$$\begin{aligned}\mathbf{S}[G] &= \sum_{i=1}^{2n} \mathbf{Inf}_i[G]^2 \\ &= \sum_{i=1}^{n} (\mathbf{Inf}_i[f] + \mathbf{Inf}_i[h])^2 \\ &\quad + \sum_{i=1}^{n} (\mathbf{Inf}_i[g] + \mathbf{Inf}_i[h])^2 \\ &\le 2\,\mathbf{S}[f] + 2\,\mathbf{S}[g] + 4\,\mathbf{S}[h]. \qquad (1)\end{aligned}$$

Applying the Basic Invariance Principle to $G(\mathbf{X})$ and $G(\mathbf{Y})$,

$$\mathrm{E}_\psi[G(\mathbf{X}); G(\mathbf{Y})] \le \frac{C}{12} \cdot 9^{2k} \cdot \mathbf{S}[G].$$

Substitutions using (1), along with the fact that $G(\mathbf{X}) = F_{\mathbf{XX}}$ and $G(\mathbf{Y}) = F_{\mathbf{YY}}$, yield the desired inequality. $\square$

The key feature of BVIP1 is the fact that F is a polynomial of degree at most $2k$. As a result, the bound of BVIP1 incurs an exponential factor of 9^{2k} which comes from the factor of 9^k in the BIP. This can be a significant added cost for high-degree polynomials. BVIP2 takes a different approach and avoids a factor of 9^{2k}.

IV. A Probabilistic Bivariate Invariance Principle

In this section, we present our second bivariate invariance principle, which we refer to as BVIP2. This invariance principle, like BVIP1, is fundamentally an application of the BIP, but instead of applying it once, we apply it iteratively. Given a real-valued bivariate Boolean function F, two input sequences $\mathbf{X}^1$ and $\mathbf{X}^2$, and two replacement sequences $\mathbf{Y}^1$ and $\mathbf{Y}^2$, we derive BVIP2 using a method which is often referred to as the "replacement method" [17]. The replacement method is one way to prove the Berry-Esseen Theorem, as well as the BIP. A key characteristic of BVIP2 is that it only holds in probability, whereas BVIP1 and BIP provide deterministic bounds. Nonetheless, this probability can be made arbitrarily close to 1 at the cost of a constant factor in the bound.

Another key difference between BVIP2 and BVIP1 is their dependence on the maximum degree k of the function. As described in Section III, the bound of BVIP1 has a factor of 9^{2k}. Owing to the iterative application of the BIP, a factor of $D_{n,m}^2 9^k$ appears in BVIP2. The factor $D_{n,m}$ is related to the number of terms in $h : \{-1,1\}^n \to \mathbb{R}$, one of the functions into which F separates and which captures the dependence on the product of the input sequences. In general, $D_{n,m}$ can be proportional to 2^n. This results in an exponential dependency which can be comparable to the additional factor of 9^k in the bound of BVIP1. However, if the maximum degree of the function is linear in the size of the input (i.e. $k = an + b$ for $a, b \in \mathbb{R}$) then the bound of BVIP2 is asymptotically tighter than BVIP1. A complete comparison between the asymptotics of the bounds can be found in Section V.

The growth of $D_{n,m}$ is controlled by the number of terms in h, but for many functions, the number of terms is significantly less than 2^n. In order to meaningfully and simply bound the number of terms in a function, we introduce the concept of the *participation* of a coordinate on a Boolean function.

Definition 7. The *participation* of a coordinate i on a function $f : \{-1,1\}^n \to \mathbb{R}$ is defined to be the number of non-zero Fourier coefficients of f for which i is involved:

$$\mathbf{Par}_i[f] = \left|\left\{S \subseteq [n] : i \in S,\ \hat{f}(S) \neq 0\right\}\right|.$$

The relevance of participation for BVIP2 is captured in the following lemma.

Lemma 3. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a function on which the maximum participation over all coordinates is m: $\max_{i\in[n]} \mathbf{Par}_i[f] = m$. Then, the number of terms in f is at most nm.*

Proof. Each of n coordinates can appear in at most m terms of f. The trivial upper bound nm is then achieved by assuming all terms are disjoint, although this bound is increasing loose as m grows. $\square$

We show in Section V that, under assumptions on the maximum participation, BVIP2 asymptotically outperfoms BVIP1 even if the degree of the function is not linear in the size of the input.

An alternative interpretation of both influence and participation on a given function $f : \{-1,1\}^n \to \mathbb{R}$ views both quantities through the norm of the function's Fourier spectrum. Let the set $S_i = \{S \subset [n] : i \in S\}$ be the set of subsets which include the coordinate i. For a particular coordinate i, consider a vector $\mathbf{r}$ indexed by $I \in S_i$ such that $\mathbf{r}_I = \hat{f}(I)$. Then, $\mathbf{Inf}_i[f] = \|\mathbf{r}\|_2^2$ and $\mathbf{Par}_i[f] = \|\mathbf{r}\|_0$, where $\|\cdot\|_p$ is the p-norm of a vector.

We now give the formal statement of BVIP2.

Theorem 4 (BVIP2). *Let $F : \{-1,1\}^n \times \{-1,1\}^n \to \mathbb{R}$ be a multilinear polynomial separable into f, g, and h such that f, g, and h are all of degree at most k, the maximum participation on h is at most m, and $\hat{f}(\emptyset) = \hat{g}(\emptyset) = \hat{h}(\emptyset) = 0$. Let $\mathbf{X}^1$, $\mathbf{X}^2$, $\mathbf{Y}^1$, and $\mathbf{Y}^2$ be n-length sequences of independent random variables satisfying Assumption 1. Let $F_{\mathbf{XX}} = F(\mathbf{X}^1, \mathbf{X}^2)$ and $F_{\mathbf{YY}} = F(\mathbf{Y}^1, \mathbf{Y}^2)$. Assume $\psi : \mathbb{R} \to \mathbb{R}$ is $\mathcal{C}^4$ with $\|\psi''''\|_\infty \leq C$. Then, the following inequality holds with probability at least $p \in (0,1)$:*

$$\mathrm{E}_\psi[F_{\mathbf{XX}}; F_{\mathbf{YY}}] \leq \frac{2C}{3} \cdot 9^k \cdot \left(\mathbf{I}[f]^2 + \mathbf{I}[g]^2 + 2D_{n,m}^2\, \mathbf{I}[h]^2\right),$$

where $D_{n,m} = (nm) / \left(1 - \sqrt{p}\right)$.

Before proving Theorem 4, we first present a few necessary lemmas. Lemma 4 is a similar statement to the BIP, but in a form that is applicable to the replacement process we will use to prove BVIP2.

Lemma 4. *Let the functions $f : \{-1,1\}^n \to \mathbb{R}$ and $g : \{-1,1\}^n \to \mathbb{R}$ be n-variate multilinear polynomials such that $\hat{f}(\emptyset) = \hat{g}(\emptyset) = 0$. Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be sequences of independent random variables satisfying Assumption 1. Assume $\psi : \mathbb{R} \to \mathbb{R}$ is $\mathcal{C}^4$ with $\|\psi''''\|_\infty \leq C$. Then,*

$$\mathrm{E}_\psi[f(\mathbf{X}) + g(\mathbf{Y}); f(\mathbf{X}) + g(\mathbf{Z})] \leq \frac{C}{12} \cdot 9^k \cdot \mathbf{I}[g]^2.$$

Proof. Take two 4th-order Taylor expansion for ψ. First,

$$\begin{aligned}\psi(f(\mathbf{X}) + g(\mathbf{Y})) = \psi(f(\mathbf{X})) &+ \psi'(f(\mathbf{X}))g(\mathbf{Y}) \\ &+ \frac{1}{2}\psi''(f(\mathbf{X}))g(\mathbf{Y})^2 \\ &+ \frac{1}{6}\psi'''(f(\mathbf{X}))g(\mathbf{Y})^3 \\ &+ \frac{1}{24}\psi''''(X^*)g(\mathbf{Y})^4,\end{aligned} \tag{2}$$

for some point X^* between $f(\mathbf{X})$ and $f(\mathbf{X}) + g(\mathbf{Y})$. Similarly,

$$\begin{aligned}\psi(f(\mathbf{X}) + g(\mathbf{Z})) = \psi(f(\mathbf{X})) &+ \psi'(f(\mathbf{X}))g(\mathbf{Z}) \\ &+ \frac{1}{2}\psi''(f(\mathbf{X}))g(\mathbf{Z})^2 \\ &+ \frac{1}{6}\psi'''(f(\mathbf{X}))g(\mathbf{Z})^3 \\ &+ \frac{1}{24}\psi''''(X^{**})g(\mathbf{Z})^4,\end{aligned} \tag{3}$$

for some point X^{**} between $f(\mathbf{X})$ and $f(\mathbf{X}) + g(\mathbf{Z})$. Let $A = \psi''''(X^*)$ and $B = \psi''''(X^{**})$. Let the difference between the two Taylor expansions be

$$D_T = \psi(f(\mathbf{X}) + g(\mathbf{Y})) - \psi(f(\mathbf{X}) + g(\mathbf{Z})). \tag{4}$$

Substituting (2) and (3) into (4),

$$\begin{aligned}D_T = \psi'(f(\mathbf{X}))&[g(\mathbf{Y}) - g(\mathbf{Z})] \\ &+ \frac{1}{2}\psi''(f(\mathbf{X}))[g(\mathbf{Y})^2 - g(\mathbf{Z})^2] \\ &+ \frac{1}{6}\psi'''(f(\mathbf{X}))[g(\mathbf{Y})^3 - g(\mathbf{Z})^3] \\ &+ \frac{1}{24}[A \cdot g(\mathbf{Y})^4 - B \cdot g(\mathbf{Z})^4].\end{aligned} \tag{5}$$

Define

$$Q = |\mathrm{E}[D_T]| = \mathrm{E}_\psi[f(\mathbf{X}) + g(\mathbf{Y}); f(\mathbf{X}) + g(\mathbf{Z})].$$

Using the linearity of expectation, the independence of $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, and the triangle inequality, we bound Q as

$$\begin{aligned}Q \leq |\mathrm{E}[\psi'(f(\mathbf{X}))]| &\cdot |\mathrm{E}[g(\mathbf{Y})] - \mathrm{E}[g(\mathbf{Z})]| \\ &+ \frac{1}{2}|\mathrm{E}[\psi''(f(\mathbf{X}))]| \cdot \left|\mathrm{E}\left[g(\mathbf{Y})^2\right] - \mathrm{E}\left[g(\mathbf{Z})^2\right]\right| \\ &+ \frac{1}{6}|\mathrm{E}[\psi'''(f(\mathbf{X}))]| \cdot \left|\mathrm{E}\left[g(\mathbf{Y})^3\right] - \mathrm{E}\left[g(\mathbf{Z})^3\right]\right| \\ &+ \frac{1}{24}\left|\mathrm{E}\left[A \cdot g(\mathbf{Y})^4 - B \cdot g(\mathbf{Z})^4\right]\right|.\end{aligned} \tag{6}$$

By Assumption 1, $\mathbf{Y}$ and $\mathbf{Z}$ have matching 1st, 2nd, and 3rd moments. Hence, for $n \in \{1,2,3\}$,

$$|\mathrm{E}[g(\mathbf{Y}^n)] - \mathrm{E}[g(\mathbf{Z}^n)]| = 0.$$

Using this result, another application of the triangle inequality, and the fact that $A \leq C$ and $B \leq C$, we can simplify (6) as

$$\begin{aligned}Q &\leq \frac{1}{24}\left|\mathrm{E}\left[A \cdot g(\mathbf{Y})^4 - B \cdot g(\mathbf{Z})^4\right]\right| \\ &\leq \frac{A}{24} \cdot \mathrm{E}\left[g(\mathbf{Y})^4\right] + \frac{B}{24} \cdot \mathrm{E}\left[g(\mathbf{Z})^4\right] \\ &\leq \frac{C}{24}\left(\mathrm{E}\left[g(\mathbf{Y})^4\right] + \mathrm{E}\left[g(\mathbf{Z})^4\right]\right).\end{aligned} \tag{7}$$

We continue by upper bounding the quantities on the left-hand side of (7). We have

$$\mathrm{E}\left[g(\mathbf{Y})^4\right] \leq 9^k \cdot \mathrm{E}\left[g(\mathbf{Y})^2\right]^2 \tag{8}$$

$$= 9^k \cdot \left(\sum_{S \subseteq [n]} \hat{g}(S)^2\right)^2 \tag{9}$$

$$\leq 9^k \cdot \left(\sum_{S \subseteq [n]} |S|\hat{g}(S)^2\right)^2 \tag{10}$$

$$= 9^k \cdot \mathbf{I}[g]^2, \tag{11}$$

where the inequality in (8) follows from Lemma 1 and the equality in (9) follows from Lemma 2. The inequality in (10) follows from the fact that $|S| \geq 1$, since $\hat{f}(\emptyset) = \hat{g}(\emptyset) = 0$ by assumption. The equality in (11) follows from the definitions of influence and total influence. When summing the influences over all $i \in [n]$ to compute the total influence, the Fourier weight $\hat{g}(S)^2$ is counted exactly $|S|$ times.

By the same argument,

$$\mathrm{E}\left[g(\mathbf{Z})^4\right] \leq 9^k \cdot \mathbf{I}[g]^2. \tag{12}$$

Substituting (11) and (12) into (7) yields the desired inequality. □

The next two lemmas will be used to bound the probability with which BVIP2 holds.

Lemma 5. *For n events $A_1, \ldots, A_n$,*

$$\Pr\left\{\bigcap_{i=1}^n A_i\right\} \geq \sum_{i=1}^n \Pr\{A_i\} - n + 1.$$

Proof. The proof is by induction. As a base case, consider two events A_1 and A_2. By definition,

$$\begin{aligned}\Pr\{A \cap B\} &= P(A) + P(B) - P(A \cup B) \\ &\geq P(A) + P(B) - 1.\end{aligned}$$

Now assume that the statement holds for $n-1$ events. Then,

$$\begin{aligned}\Pr\left\{\bigcap_{i=1}^{n} A_i\right\} &= \Pr\left\{\bigcap_{i=1}^{n-1} A_i \cap A_n\right\} \\ &= \Pr\left\{\bigcap_{i=1}^{n-1} A_i\right\} + \Pr\{A_n\} \\ &\quad - \Pr\left\{\bigcap_{i=1}^{n-1} A_i \cup A_n\right\} \\ &\geq \Pr\left\{\bigcap_{i=1}^{n-1} A_i\right\} + \Pr\{A_n\} - 1 \\ &\geq \sum_{i=1}^{n-1} \Pr\{A_i\} - (n-1) + 1 + \Pr\{A_n\} - 1 \\ &= \sum_{i=1}^{n} \Pr\{A_i\} - n + 1.\end{aligned}$$

Having proven the base case and the inductive step, the desired result holds by induction. □

In its usual form, Markov's inequality [18] provides an upper bound on the probability that a random variable exceeds a given value. We use a direct corollary of that result.

Lemma 6 (Variation on Markov's Inequality). *Let X be a nonnegative random variable and $a \in \mathbb{R}$ be a nonnegative constant. Then,*

$$\Pr\{X \leq a\} \geq 1 - \frac{\mathrm{E}[X]}{a}.$$

We now proceed to a proof of Theorem 4. As stated earlier, the proof of Theorem 4 uses the replacement method. The general idea is to first replace $\mathbf{X}^1$ by $\mathbf{Y}^1$ and use the BIP (more precisely, Lemma 4) to bound the approximation error between $F(\mathbf{X}^1, \mathbf{X}^2)$ and $F(\mathbf{Y}^1, \mathbf{X}^2)$. We then replace $\mathbf{X}^2$ by $\mathbf{Y}^2$ using an identical process and similarly bound the error between $F(\mathbf{Y}^1, \mathbf{X}^2)$ and $F(\mathbf{Y}^1, \mathbf{Y}^2)$. Finally, an application of the triangle inequality yields a bound on the overall approximation error between $F(\mathbf{X}^1, \mathbf{X}^2)$ and $F(\mathbf{Y}^1, \mathbf{Y}^2)$.

Proof of Theorem 4. We first replace $\mathbf{X}^1$ by $\mathbf{Y}^1$. To do so, we isolate the terms of $F_{\mathbf{XX}}$ involving $\mathbf{X}^1$.

$$\begin{aligned}F_{\mathbf{XX}} = &\sum_{S\subseteq[n],|S|\leq k} \hat{f}(S) \prod_{i\in S} X_i^1 \\ &+ \sum_{S\subseteq[n],|S|\leq k} \hat{h}(S) \prod_{i\in S} X_i^1 X_i^2 + g(\mathbf{X}^2) \\ = &\sum_{S\subseteq[n],|S|\leq k} \left(\hat{f}(S) + \hat{h}(S) \prod_{i\in S} X_i^2\right) \prod_{i\in S} X_i^1 \\ &+ g(\mathbf{X}^2).\end{aligned}$$

We then define a new, random function $r : \{-1, 1\}^n \to \mathbb{R}$ which captures the dependency of $F_{\mathbf{XX}}$ on $\mathbf{X}^1$. With respect to r, we treat the random variable $\mathbf{X}^2$ as a random parameter. Explicitly,

$$\begin{aligned}r(\mathbf{t}) &= f(\mathbf{t}) + h(\mathbf{t}\mathbf{X}^2) \\ &= \sum_{S\subseteq[n],|S|\leq k} \left(\hat{f}(S) + \hat{h}(S) \prod_{i\in S} X_i^2\right) \prod_{i\in S} t_i.\end{aligned}$$

Thus, $F_{\mathbf{XX}} = r(\mathbf{X}^1) + g(\mathbf{X}^2)$. Define $F_{\mathbf{XY}} = F(\mathbf{Y}^1, \mathbf{X}^2)$. It immediately follows that $F_{\mathbf{XY}} = r(\mathbf{Y}^1) + g(\mathbf{X}^2)$. In order to apply the BIP to $F_{\mathbf{XX}}$ and $F_{\mathbf{XY}}$, we must first bound the influences of r.

$$\begin{aligned}\mathbf{Inf}_i[r] &= \sum_{S\ni i} \left(\hat{f}(S) + \hat{h}(S) \prod_{j\in S} X_j^2\right)^2 \\ &\leq \sum_{S\ni i} \left(2\hat{f}(S)^2 + 2\hat{h}(S)^2 \prod_{j\in S} \left(X_j^2\right)^2\right) \\ &= 2\,\mathbf{Inf}_i[f] + 2\sum_{S\ni i} \left(\hat{h}(S)^2 \prod_{j\in S} \left(X_j^2\right)^2\right).\end{aligned} \tag{13}$$

We wish to find a way to upper bound the product $\prod_{j\in S}(X_j^2)^2$. In general, there is no such bound, since X_j^2 can be an unbounded random variable (e.g. a standard Gaussian). Nonetheless, we will show that we can bound the product in probability. Let D represent that probabilistic upper bound. Let

$$Z_S = \prod_{j\in S} (X_j^2)^2.$$

We will show that we can choose D such that

$$\Pr\left\{\bigcap_{S\in S^*} Z_S \leq D\right\} \geq \sqrt{p}, \tag{14}$$

where $S^* = \{S : \hat{h}(S)^2 \neq 0\}$. As a result, we will have shown that with probability at least $\sqrt{p}$,

$$2\sum_{S\ni i} \hat{h}(S)^2 Z_s \leq 2D \sum_{S\ni i} \hat{h}(S)^2 = 2D\,\mathbf{Inf}_i[h]. \tag{15}$$

Applying Lemma 5 to the left-hand side of (14).

$$\Pr\left\{\bigcap_{S\in S^*} Z_S \leq D\right\} \geq \sum_{S\in S^*} \Pr\{Z_S \leq D\} - |S^*| + 1. \tag{16}$$

We now bound $\Pr\{Z_S \leq D\}$ for a particular S. By Assumption 1, $\mathrm{E}\left[X_i^2\right] = 1$ for all $i \in [n]$. Since $\mathbf{X}^2$ is a sequence of independent random variables, $\mathrm{E}[Z_S] = 1$ for all S. Thus, by Lemma 6, for every S,

$$\Pr\{Z_S \leq D\} \geq 1 - \frac{1}{D}.$$

Substituting into (16),

$$\Pr\left\{\bigcap_{S\in S^*} Z_S \le D\right\} \ge \sum_{S\in S^*}\left(1-\frac{1}{D}\right) - |S^*| + 1$$
$$= |S^*|\left(1-\frac{1}{D}\right) - |S^*| + 1. \quad (17)$$

Since the maximum participation on h is m, by Lemma 3 we can bound $|S^*| \le nm$. Substituting into (17) and simplifying,

$$\Pr\left\{\bigcap_{S\in S^*} Z_S \le D\right\} \ge 1 - \frac{nm}{D}.$$

To achieve (14), we can choose

$$D = D_{n,m} = \frac{nm}{1-\sqrt{p}}. \quad (18)$$

Henceforth, we refer to this quantity as $D_{n,m}$ to emphasize that it depends on the size of the input n and the maximum participation m on h.

With this choice of $D_{n,m}$, the bound in (15) holds with probability at least $\sqrt{p}$, as desired. Thus, with probability at least $\sqrt{p}$,

$$\mathbf{Inf}_i[r] \le 2\,\mathbf{Inf}_i[f] + 2D_{n,m}\,\mathbf{Inf}_i[h].$$

By the definition of total influence, we have with the same probability that

$$\mathbf{I}[r] \le 2\,\mathbf{I}[f] + 2D_{n,m}\,\mathbf{I}[h]. \quad (19)$$

We can now apply the BIP to $F_{\mathbf{XX}}$ and $F_{\mathbf{XY}}$ using Lemma 4. With probability at least $\sqrt{p}$,

$$\begin{aligned}
\mathrm{E}_\psi[F_{\mathbf{XX}}; F_{\mathbf{XY}}] &= \mathrm{E}_\psi[r(\mathbf{X}^1) + g(\mathbf{X}^2); r(\mathbf{Y}^1) + g(\mathbf{X}^2)] \\
&\le \frac{C}{12}\cdot 9^k \cdot \mathbf{I}[r]^2 \\
&\le \frac{C}{12}\cdot 9^k \cdot (2\,\mathbf{I}[f] + 2D_{n,m}\,\mathbf{I}[h])^2 \\
&\le \frac{C}{12}\cdot 9^k \cdot \left(8\,\mathbf{I}[f]^2 + 8D_{n,m}^2\,\mathbf{I}[h]^2\right) \\
&= \frac{2C}{3}\cdot 9^k \cdot \left(\mathbf{I}[f]^2 + D_{n,m}^2\,\mathbf{I}[h]^2\right). \quad (20)
\end{aligned}$$

This completes the first step of the iterative replacement method approach.

Next, to replace $\mathbf{X}^2$ by $\mathbf{Y}^2$, we begin as before, separating the terms of $F_{\mathbf{XY}}$ involving $\mathbf{X}^2$.

$$\begin{aligned}
F_{\mathbf{XY}} = &\sum_{S\subseteq[n],|S|\le k}\left(\hat{f}(S) + \hat{h}(S)\prod_{i\in S} X_i^2\right)\prod_{i\in S} Y_i^1 \\
&+ \sum_{S\subseteq[n],|S|\le k} \hat{g}(S)\prod_{i\in S} X_i^2 \\
= &\sum_{S\subseteq[n],|S|\le k}\left(\hat{g}(S) + \hat{h}(S)\prod_{i\in S} Y_i^1\right)\prod_{i\in S} X_i^2 \\
&+ \sum_{S\subseteq[n],|S|\le k} \hat{f}(S)\prod_{i\in S} Y_i^1
\end{aligned}$$

We now define a new, random function $q : \{-1,1\}^n \to \mathbb{R}$ which captures the dependency of $F_{\mathbf{XY}}$ on $\mathbf{X}^2$. The function q serves the same role as the function r in the replacment process. As we treated $\mathbf{X}^2$ as a random parameter with respect to r, we now treat the random variable $\mathbf{Y}^2$ as a random parameter with respect to q. Explicitly,

$$\begin{aligned}
q(\mathbf{t}) &= g(\mathbf{t}) + h(\mathbf{Y}^1\mathbf{t}) \\
&= \sum_{S\subseteq[n],|S|\le k}\left(\hat{g}(S) + \hat{h}(S)\prod_{i\in S} Y_i^1\right)\prod_{i\in S} t_i.
\end{aligned}$$

We now seek to bound the influences of q in order to apply the BIP to $F_{\mathbf{XY}}$ and $F_{\mathbf{YY}}$.

$$\begin{aligned}
\mathbf{Inf}_i[q] &= \sum_{S\ni i}\left(\hat{g}(S) + \hat{h}(S)\prod_{j\in S} Y_j^1\right)^2 \\
&\le \sum_{S\ni i}\left(2\hat{g}(S)^2 + 2\hat{h}(S)^2\prod_{j\in S}\left(Y_j^1\right)^2\right) \\
&= 2\,\mathbf{Inf}_i[g] + 2\sum_{S\ni i}\left(\hat{h}(S)^2\prod_{j\in S}\left(Y_j^1\right)^2\right).
\end{aligned}$$

As before, we wish to probabilistically upper bound the product $\prod_{j\in S}\left(Y_j^1\right)^2$. Because $\mathbf{Y}^1$ satisfies the same assumptions as $\mathbf{X}^2$, the same approach we used to bound $\prod_{j\in S}\left(X_j^2\right)^2$ will suffice. By that same argument, with probability at least $\sqrt{p}$,

$$\mathbf{Inf}_i[q] \le 2\,\mathbf{Inf}_i[g] + 2D_{n,m}\,\mathbf{Inf}_i[h],$$

where $D_{n,m}$ is chosen as in (18). Hence, with probability at least $\sqrt{p}$,

$$\mathbf{I}[q] \le 2\,\mathbf{I}[g] + 2D_{n,m}\,\mathbf{I}[h]. \quad (21)$$

We then apply the BIP to $F_{\mathbf{XY}}$ and $F_{\mathbf{YY}}$ using Lemma 4. With probability at least $\sqrt{p}$,

$$\begin{aligned}
\mathrm{E}_\psi[F_{\mathbf{XY}}; F_{\mathbf{YY}}] &= \mathrm{E}_\psi[q(\mathbf{X}^2) + f(\mathbf{Y}^1); q(\mathbf{Y}^2) + f(\mathbf{Y}^1)] \\
&\le \frac{C}{12}\cdot 9^k \cdot \mathbf{I}[q]^2 \\
&\le \frac{C}{12}\cdot 9^k \cdot (2\,\mathbf{I}[g] + 2D_{n,m}\,\mathbf{I}[h])^2 \\
&\le \frac{C}{12}\cdot 9^k \cdot \left(8\,\mathbf{I}[g]^2 + 8D_{n,m}^2\,\mathbf{I}[h]^2\right) \\
&= \frac{2C}{3}\cdot 9^k \cdot \left(\mathbf{I}[g]^2 + D_{n,m}^2\,\mathbf{I}[h]^2\right). \quad (22)
\end{aligned}$$

Finally, we combine the bounds in (20) and (22) using the triangle inequality. Since the random variables $\mathbf{X}^2$ and $\mathbf{Y}^1$ are independent and the bounds in (19) and (21) each hold with probability $\sqrt{p}$, both bounds hold with probability at least p. Thus, with probability at least p,

$$\begin{aligned}
\mathrm{E}_\psi[F_{\mathbf{XX}}; F_{\mathbf{YY}}] &\leq \mathrm{E}_\psi[F_{\mathbf{XX}}; F_{\mathbf{XY}}] + \mathrm{E}_\psi[F_{\mathbf{XY}}; F_{\mathbf{YY}}] \\
&\leq \frac{2C}{3} \cdot 9^k \cdot \left(\mathbf{I}[f]^2 + D_{n,m}^2\, \mathbf{I}[h]^2\right) \\
&\quad + \frac{2C}{3} \cdot 9^k \cdot \left(\mathbf{I}[g]^2 + D_{n,m}^2\, \mathbf{I}[h]^2\right) \\
&= \frac{2C}{3} \cdot 9^k \cdot \left(\mathbf{I}[f]^2 + \mathbf{I}[g]^2 + 2D_{n,m}^2\, \mathbf{I}[h]^2\right).
\end{aligned}$$

□

V. Comparison

In this section, we compare the two bounds of BVIP1 and BVIP2, characterize their asymptotic behavior, and offer some examples which illustrate that behavior. We begin by formally describing the conditions under which, given a sequence of bivariate functions, each bound converges to 0.

A. Conditions for Convergence

For any separable function $F : \{-1,1\}^n \times \{-1,1\}^n \to \mathbb{R}$, let $B_1(F)$ and $B_2(F)$ be the bounds of BVIP1 and BVIP2 respectively:

$$\begin{aligned}
B_1(F) &= \frac{C}{6} \cdot 9^{2k} \cdot (\mathbf{S}[f] + \mathbf{S}[g] + 2\,\mathbf{S}[h]) \\
B_2(F) &= \frac{2C}{3} \cdot 9^k \cdot \left(\mathbf{I}[f]^2 + \mathbf{I}[g]^2 + 2D_{n,m}^2\, \mathbf{I}[h]^2\right).
\end{aligned}$$

Given a sequence of functions (F_n), with F_n separable into f_n, g_n, and h_n, the asymptotic behavior of $B_1(F_n)$ is determined by the behavior of the maximum degree and the influences of F_n as n grows. The asymptotic behavior of $B_2(F_n)$ is determined by the maximum degree k, the influences, the size n of the inputs, and the maximum participation m on h_n as n grows.

For the purposes of bounding the influences of each F_n, it will be useful to define the concept of maximum influence.

Definition 8. The *maximum influence* on $f : \{-1,1\}^n \to \mathbb{R}$ is defined to be the maximum influence on f over all coordinates:

$$\mathbf{MaxInf}[f] = \max_{1 \leq i \leq n} \mathbf{Inf}_i[f].$$

We formally state the conditions for convergence for each bound in the following two lemmas.

Lemma 7 (Convergence Conditions for BVIP1). *Let (F_n), $F_n : \{-1,1\}^n \times \{-1,1\}^n \to \mathbb{R}$, be a sequence of multilinear polynomials separable into f_n, g_n, and h_n and let k_n be the maximum degree of F_n. Then, the sequence $B_1(F_n)$ converges to 0 if all of the following asymptotic relations hold:*

$$\begin{aligned}
\mathbf{MaxInf}[f_n] &= o\left(\left(9^{2k_n} n\right)^{-1/2}\right) \\
\mathbf{MaxInf}[g_n] &= o\left(\left(9^{2k_n} n\right)^{-1/2}\right) \\
\mathbf{MaxInf}[h_n] &= o\left(\left(9^{2k_n} n\right)^{-1/2}\right).
\end{aligned}$$

Proof. In order for $B_1(F)$ to converge to 0, we must have

$$\lim_{n\to\infty} 9^{2k_n}\, \mathbf{S}[f_n] = 0.$$

For any function $f : \{-1,1\}^n \to \mathbb{R}$,

$$\mathbf{S}[f] \leq n\, \mathbf{MaxInf}[f]^2.$$

Thus, it is sufficient to require that

$$\lim_{n\to\infty} 9^{2k_n} n\, \mathbf{MaxInf}[f_n]^2 = 0.$$

This limit is achieved when

$$\mathbf{MaxInf}[f_n] = o\left(\left(9^{2k_n} n\right)^{-1/2}\right).$$

The same argument applies to g_n and h_n, yielding the desired asymptotic results. □

Lemma 8 (Convergence Conditions for BVIP2). *Let (F_n), $F_n : \{-1,1\}^n \times \{-1,1\}^n \to \mathbb{R}$, be a sequence of multilinear polynomials separable into f_n, g_n, and h_n, let k_n be the maximum degree of F_n, and let m_n be the maximum participation on h_n. Then, the sequence $B_2(F_n)$ converges to 0 if all of the following asymptotic relations hold:*

$$\begin{aligned}
\mathbf{MaxInf}[f_n] &= o\left(\left(9^{k_n} n^2\right)^{-1/2}\right) \\
\mathbf{MaxInf}[g_n] &= o\left(\left(9^{k_n} n^2\right)^{-1/2}\right) \\
\mathbf{MaxInf}[h_n] &= o\left(\left(9^{k_n} n^4 m_n^2\right)^{-1/2}\right).
\end{aligned}$$

Proof. In order for $B_2(F)$ to converge to 0, we must have

$$\lim_{n\to\infty} 9^{k_n}\, \mathbf{I}[f_n]^2 = 0.$$

For any function $f : \{-1,1\}^n \to \mathbb{R}$,

$$\mathbf{I}[f] \leq n^2\, \mathbf{MaxInf}[f]^2.$$

Thus, it is sufficient to require that

$$\lim_{n\to\infty} 9^{k_n} n^2\, \mathbf{MaxInf}[f_n]^2 = 0.$$

This limit is achieved when

$$\mathbf{MaxInf}[f_n] = o\left(\left(9^{k_n} n^2\right)^{-1/2}\right).$$

The same argument applies to g_n. The behavior of h_n is slightly different. We must have

$$\lim_{n\to\infty} 9^{k_n} D^2\, \mathbf{I}[h_n]^2 = 0.$$

Thus, it is sufficient to require that

$$\lim_{n\to\infty} 9^{k_n} D^2 n^2\, \mathbf{MaxInf}[h_n]^2 = 0.$$

Substituting in the value of D from (18),

$$\lim_{n\to\infty} 9^{k_n} n^4 m^2\, \mathbf{MaxInf}[h_n]^2 = 0.$$

This limit is achieved when

$$\mathbf{MaxInf}[h_n] = o\left(\left(9^{k_n} n^4 m_n^2\right)^{-1/2}\right).$$

□

As a result of these two lemmas, if the sequence of maximum degrees (k_n) grows linearly in n, the sequence $B_2(F_n)$ will always converge if $B_1(F_n)$ does. This follows from the fact that the asymptotic conditions in Lemma 8 are stronger than those in Lemma 7 when k_n is linear in n.

Corollary 1. *Let (F_n), $F_n : \{-1,1\}^n \times \{-1,1\}^n \to \mathbb{R}$, be a sequence of multilinear polynomials separable into f_n, g_n, and h_n and let k_n be the maximum degree of F_n. If $k_n = an + b$ for some $a, b \in \mathbb{R}$ and $B_1(F_n)$ converges to 0, then $B_2(F_n)$ also converges.*

Proof. By Lemma 7 and the assumption that $B_1(F_n)$ converges, we have

$$\mathbf{MaxInf}[f_n] = o\left(\left(9^{2(an+b)}n\right)^{-1/2}\right) \tag{23}$$

and likewise for g_n and h_n. We verify that the asymptotic requirements in Lemma 8 are met. We must have that

$$\mathbf{MaxInf}[f_n] = o\left(\left(9^{an+b}n^2\right)^{-1/2}\right), \tag{24}$$

and likewise for g_n. For h_n, we must have that

$$\mathbf{MaxInf}[h_n] = o\left(\left(9^{an+b}4^n n^4\right)^{-1/2}\right). \tag{25}$$

Here, we make the worst-case assumption that $m_n = 2^n$. Since $k = an + b$ is positive, $9^{an+b} > n$ and $9^{an+b} > 4^n n^3$ for sufficiently large n. Thus, given that (23) is satisfied for f_n, g_n, and h_n, (24) is satisfied for f_n and g_n, and (25) is satisfied by h_n. Hence, $B_2(F_n)$ converges to 0 by Lemma 8. □

Corollary 1 can also be interpreted as stating that, for k_n linear in n, the bound of BVIP2 is asymptotically tighter than that of BVIP1, since $B_2(F_n)$ grows more slowly than $B_1(F_n)$.

Having described the conditions under which the bounds of BVIP1 and BVIP2 converge to 0, we now present two examples which illustrate cases in which the bound of BVIP1 diverges while the bound of BVIP2 converges. The first example uses the fact that m_n is constant for the family of interest, while the second example shows that even without any assumptions on m_n, there exists cases for which BVIP1 diverges and BVIP2 converges.

Example 1. Let $f_n : \{-1,1\}^n \to \mathbb{R}$ and $g_n : \{-1,1\}^n \to \mathbb{R}$ be a weighted linear sum over n elements:

$$f_n(\mathbf{x}) = g_n(\mathbf{x}) = \left(9^n n^3\right)^{-1/4} \sum_{i=1}^{n} x_i.$$

Let $h_n : \{-1,1\}^n \to \mathbb{R}$ be a weighted product over n elements:

$$h(\mathbf{x}) = \left(9^n n^5\right)^{-1/4} \prod_{i=1}^{n} x_i.$$

Let $F_n : \{-1,1\}^n \times \{-1,1\}^n \to \mathbb{R}$ be the bivariate function which is separable into f_n, g_n, and h_n. Note that f_n, g_n, and h_n are all of degree at most $k_n = n$ (due to h_n) and that the maximum participation on h_n is $m_n = 1$.

The maximum influences on f_n, g_n, and h_n are

$$\mathbf{MaxInf}[f_n] = \mathbf{MaxInf}[g_n] = \left(9^n n^3\right)^{-1/2}$$
$$\mathbf{MaxInf}[h_n] = \left(9^n n^5\right)^{-1/2}.$$

Comparing these influences to the required asymptotics given by Lemma 7 and Lemma 8, we see that the bound of BVIP1 diverges for this family while the bound for BVIP2 converges.

Example 2. Let $f_n : \{-1,1\}^n \to \mathbb{R}$ and $g_n : \{-1,1\}^n \to \mathbb{R}$ be defined as in Example 1. Let $h : \{-1,1\}^n \to \mathbb{R}$ be a different weighted product over n elements:

$$h(\mathbf{x}) = \left((9 \cdot 4)^n n^5\right)^{-1/4} \prod_{i=1}^{n} x_i.$$

Let $F_n : \{-1,1\}^n \times \{-1,1\}^n \to \mathbb{R}$ be the bivariate function which is separable into f_n, g_n, and h_n. Note that f_n, g_n, and h_n are all of degree at most $k_n = n$ (due to h_n). While the maximum participation on h_n is 1, in this case we choose to not make use of that fact, instead treating the maximum participation as $m_n = 2^n$. Because the influences of this family of functions decay quickly enough, the maximum participation does not affect whether or not the bound of BVIP2 will be tighter than that of BVIP1.

The maximum influences on f_n, g_n, and h_n are

$$\mathbf{MaxInf}[f_n] = \mathbf{MaxInf}[g_n] = \left(9^n n^3\right)^{-1/2}$$
$$\mathbf{MaxInf}[h_n] = \left((9 \cdot 4)^n n^5\right)^{-1/2}.$$

Comparing these influences to the required asymptotics given by Lemma 7 and Lemma 8, we see that the bound of BVIP1 diverges for this family while the bound for BVIP2 converges.

It is important to note that h_n is of maximal degree in Example 2, i.e. $k_n = n$. It is not generally true that there exist functions f_n, g_n, and h_n, all of degree at most $k < n$, such that for the sequence of bivariate functions (F_n) where F_n is separable into f_n, g_n, and h_n, the bound of BVIP1 will diverge while the bound of BVIP2 will converge without any assumptions on the maximum participation m_n on h_n.

VI. Conclusion

We have presented two invariance principles for separable bivariate functions, one which gives a deterministic bound and one which gives a probabilistic bound. We have shown that there exist sequences of functions for which the probabilistic bound converges to 0 while the deterministic bound diverges.

We leave the investigation of bivariate and multivariate invariance principles for general functions to future work. In addition, we have made no claims on the optimality of these principles. In particular, probabilistic techniques may not be necessary for results comparable to those achieved here.

Finally, the notion of participation was shown to be connected to the notion of influence in the sense that they are the 0-norm and squared 2-norm respectively of the same vector. Future work might consider whether there are other meaningful relations between participation and influence and whether other p-norms yield interesting quantities.

REFERENCES

[1] S. Skyum and L. G. Valiant, "A complexity theory based on Boolean algebra," *Journal of the ACM*, vol. 32, no. 2, pp. 484–502, 1985.

[2] L. Babai, P. Frankl, and J. Simon, "Complexity classes in communication complexity theory," in *27th Annual Symposium on Foundations of Computer Science (SFCS 1986)*, 1986, pp. 337–347.

[3] C. Carlet, *Boolean Functions for Cryptography and Coding Theory*. Cambridge University Press, 2021.

[4] T. W. Cusick and P. Stanica, *Cryptographic Boolean Functions and Applications*. Academic Press, 2017.

[5] G. Kalai and E. Mossel, "Sharp thresholds for monotone non-Boolean functions and social choice theory," *Mathematics of Operations Research*, vol. 40, no. 4, pp. 915–925, 2015.

[6] A. M. Alturki and A. M. Ali Rushdi, "Weighted voting systems: A threshold-Boolean perspective," *Journal of Engineering Research*, vol. 4, no. 1, pp. 1–19, 2016.

[7] C. E. Shannon, "A symbolic analysis of relay and switching circuits," *Electrical Engineering*, vol. 57, no. 12, pp. 713–723, 1938.

[8] R. E. Bryant, "On the complexity of VLSI implementations and graph representations of Boolean functions with application to integer multiplication," *IEEE transactions on Computers*, vol. 40, no. 2, pp. 205–213, 1991.

[9] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, "Noise stability of functions with low influences: invariance and optimality," in *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, 2005, pp. 21–30.

[10] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Transactions of the american mathematical society*, vol. 49, no. 1, pp. 122–136, 1941.

[11] C.-G. Esseen, "On the Liapounoff limit of error in the theory of probability," *Arkiv för matematik, astronomi och fysik*, vol. 28, no. 9, pp. 1–19, 1942.

[12] L. S. Penrose, "The elementary statistics of majority voting," *Journal of the Royal Statistical Society*, vol. 109, no. 1, pp. 53–57, 1946.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2006.

[14] R. G. D'Oliveira, S. El Rouayheb, and M. Médard, "The computational wiretap channel," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2018, pp. 1136–1140.

[15] R. O'Donnell, *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[16] A. Bonami, "Étude des coefficients de Fourier des fonctions de $L^p(G)$," in *Annales de l'institut Fourier*, vol. 20, no. 2, 1970, pp. 335–402.

[17] J. W. Lindeberg, "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung," *Mathematische Zeitschrift*, vol. 15, no. 1, pp. 211–225, 1922.

[18] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013, ch. 2, p. 19.