Boolean Functions with Biased Inputs: Approximation and Noise Sensitivity

Mohsen Heidari
University of Michigan, USA
mohsenhd@umich.edu

S. Sandeep Pradhan
University of Michigan, USA
pradhanv@umich.edu

Ramji Venkataramanan
University of Cambridge, UK
ramji.v@eng.cam.ac.uk

Abstract—This paper considers the problem of approximating a Boolean function f using another Boolean function from a specified class. Two classes of approximating functions are considered: k-juntas, and linear Boolean functions. The n input bits of the function are assumed to be independently drawn from a distribution that may be biased. The quality of approximation is measured by the mismatch probability between f and the approximating function g. For each class, the optimal approximation and the associated mismatch probability is characterized in terms of the biased Fourier expansion of f. The technique used to analyze the mismatch probability also yields an expression for the noise sensitivity of f in terms of the biased Fourier coefficients, under a general i.i.d. input perturbation model.

I. INTRODUCTION

Given a set of labeled data, we may wish to learn the optimal classifier within a specific class of functions. For example, given n-dimensional data with binary labels, one may wish to construct a classifier that depends on only k of the n input variables (where k may be much smaller than n). Such a parsimonious classifier would be less accurate on the training data than the optimal unconstrained classifier (which uses all n variables), but more robust to errors in the data. A useful measure to quantify this trade-off is the probability of mismatch between the optimal unconstrained and constrained classifiers, under some distribution on the input variables.

Motivated by such applications, we consider the problem of approximating a given Boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ using a simpler Boolean function from a specified class. The input set $\{-1, 1\}^n$ is equipped with a product distribution, where each of the n input bits $X_1, \ldots, X_n$ is drawn independently according to

$$\mathbb{P}(X_i = -1) = 1 - \mathbb{P}(X_i = 1) = p, \qquad i \in [n]. \quad (1)$$

The quality of approximation is measured by the *mismatch probability* $\mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{X}))$, where $\boldsymbol{X} \triangleq (X_1, \ldots, X_n)$.

We consider two classes of approximating functions: i) k-*juntas* where the Boolean function g depends on at most k of the n input variables (with $k < n$), and ii) linear Boolean functions which are parity functions or negations of a parity on a subset of the input variables. In each case, we characterize the optimal approximation and the associated mismatch probability in terms of the p-biased Fourier expansion of the original function f.

The standard Fourier expansion [1] of a Boolean function is a multilinear polynomial with real coefficients, where each term in the polynomial corresponds to a parity function on a subset of the input variables. The Fourier expansion has been used to analyze Boolean functions in wide range of applications, e.g., to characterize the learning complexity [2], [3], noise sensitivity [1], [4], [5], approximation [6], and other information-theoretic properties [7], [8], [9]. The parity functions form a set of orthonormal basis functions when the inputs to the Boolean function are uniformly random.

For $p \in (0, 1)$, the p-biased Fourier expansion [1, Chap. 8] generalizes the standard Fourier expansion by expressing the Boolean function as a linear combination of functions that form an orthonormal basis when the input variables are drawn i.i.d. according to the distribution in (1). p-biased Fourier analysis was used in [10] to show that a certain class of Boolean functions could be learnt efficiently using examples drawn from a biased input distribution. It has also been used to study threshold phenomena of random graphs [11]. In this paper, we use the p-biased expansion to study optimal approximation of Boolean functions with biased inputs.

The contributions of the paper are as follows.

1) In Section III, we obtain an expression (Lemma 1) for the mismatch probability $\mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{Y}))$, where f, g are Boolean functions with statistically dependent binary inputs $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Taking $f = g$ yields the noise sensitivity of a Boolean function under a general i.i.d. input perturbation model. Lemma 1 also generalizes a bound on the mismatch probability obtained in [12].

2) Next, by taking $\boldsymbol{X} = \boldsymbol{Y}$, Lemma 1 is used to establish the optimal approximation with k-juntas (Section IV), and with linear Boolean functions (Section V). We provide examples to illustrate how the optimal approximation within a class depends on the input bias.

We remark that some of the results (such as those in Section IV) hold for product distributions over any finite input alphabet. For concreteness, we focus on the binary input alphabet throughout the paper. We also mention that the worst-case circuit-size complexity of approximating Boolean functions with uniform inputs was analyzed in [13].

Notation: We use $[n]$ to denote the set $\{1, \ldots, n\}$. The cardinality of a set S is denoted by $|S|$. Given $S \subseteq [n]$ and a sequence of numbers $a_i, i \in [n]$, denote $a^S \triangleq (a_i)_{i \in S}$. We

use upper case to denote random variables, lower case for realizations, and boldface for vectors.

II. THE p-BIASED FOURIER EXPANSION

We consider Boolean functions with the distribution on the entries of the input $\boldsymbol{X} = (X_1, \ldots, X_n)$ being i.i.d. according to (1). With this distribution, an inner product can be defined for the (larger) space of bounded functions with binary inputs and real-valued outputs. For any $f, g : \{-1, 1\}^n \to \mathbb{R}$, let

$$\langle f, g \rangle \triangleq \mathbb{E}[f(\boldsymbol{X})g(\boldsymbol{X})] = \sum_{\boldsymbol{x} \in \{-1,1\}^n} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x})f(\boldsymbol{x})g(\boldsymbol{x}). \tag{2}$$

The p-biased Fourier expansion [1, Chap. 8] of a function $f : \{-1, 1\}^n \to \mathbb{R}$ is

$$f(\boldsymbol{x}) = \sum_{S \subseteq [n]} \bar{f}(S)\phi_S(\boldsymbol{x}), \tag{3}$$

where

$$\phi_S(\boldsymbol{x}) = \prod_{i \in S} \frac{(x_i - \mu)}{\sigma}. \tag{4}$$

Here

$$\mu = (1 - 2p) \quad \text{and} \quad \sigma = 2\sqrt{p(1-p)} \tag{5}$$

are the mean and standard deviation, respectively, of each of the X_i's. For $S \subseteq [n]$, the p-biased Fourier coefficients can be computed as

$$\bar{f}(S) = \langle f, \phi_S \rangle = \mathbb{E}[f(\boldsymbol{X})\phi_S(\boldsymbol{X})], \tag{6}$$

where the entries of $\boldsymbol{X} = (X_1, \ldots, X_n)$ are i.i.d. according to (1). Under this inner product, the set of functions $\{\phi_S\}_{S \subseteq [n]}$ is an orthonormal basis. Indeed, using the independence of the X_i's, it can be shown that for any $S, T \subseteq [n]$, the inner product $\mathbb{E}[\phi_S(\boldsymbol{X})\phi_T(\boldsymbol{X})] = 1$ if $S = T$, and 0 otherwise.

Since (3) is an orthonormal expansion, the inner product between two functions can be expressed in terms of their p-biased Fourier coefficients. For any $f, g : \{-1, 1\}^n \to \mathbb{R}$

$$\langle f, g \rangle = \mathbb{E}[f(\boldsymbol{X})g(\boldsymbol{X})] = \sum_{S \subseteq [n]} \bar{f}(S)\bar{g}(S). \tag{7}$$

The standard Fourier expansion corresponds to the case where $p = \frac{1}{2}$. In this case, $\mu = 0, \sigma = 1$, and the basis functions are $\phi_S(\boldsymbol{x}) = \prod_{i \in S} x_i$, $S \subseteq [n]$.

For $f : \{-1, 1\}^n \to \mathbb{R}$ and any set $T \subseteq [n]$, let $\boldsymbol{X}^T$ denote the components of $\boldsymbol{X}$ indexed by T. We refer to $\mathbb{E}[f | \boldsymbol{X}^T]$ as the projection of f onto $\boldsymbol{X}^T$. This projection is denoted by $f^{\subseteq T}$, and is given by

$$f^{\subseteq T}(\boldsymbol{X}) \triangleq \mathbb{E}[f(\boldsymbol{X}) | \boldsymbol{X}^T] = \sum_{S \subseteq T} \bar{f}(S)\phi_S(\boldsymbol{X}^S). \tag{8}$$

The last equality is obtained from (3) by noting that for any set $S \not\subseteq T$, the conditional expectation $\mathbb{E}[\phi_S(\boldsymbol{X}) | \boldsymbol{X}^T] = 0$. We note that the projection $f^{\subseteq T}$ may have real-valued outputs, even when f is Boolean.

III. BOOLEAN FUNCTIONS OF JOINTLY DISTRIBUTED RANDOM VARIABLES

In this section we investigate Boolean functions, say f and g, whose inputs that are statistically correlated. We derive an expression for the mismatch probability in term of biased Fourier coefficients of the functions.

Let $X, Y \in \{-1, 1\}$ be jointly distributed Boolean random variables with joint pmf P_{XY} whose marginals satisfy

$$\mathbb{P}(X = -1) = p, \quad \mathbb{P}(Y = -1) = q. \tag{9}$$

Let $\rho \in [-1, 1]$ denote the correlation coefficient between X and Y. The joint pmf P_{XY} is uniquely determined by the triple (p, q, ρ). Let $(\boldsymbol{X}, \boldsymbol{Y})$ be a pair of sequences with entries $(X_i, Y_i)_{i \in [n]} \sim_{i.i.d.} P_{XY}$.

For any Boolean functions $f, g : \{-1, 1\}^n \to \{-1, 1\}$, the p-biased Fourier expansion of f is given by (3)–(4), and the q-biased Fourier expansion of g is

$$g(\boldsymbol{y}) = \sum_{S \subseteq [n]} \tilde{g}(S)\psi_S(\boldsymbol{y}), \tag{10}$$

where $\psi_S(\boldsymbol{y}) = \prod_{i \in S} \frac{(y_i - \mu')}{\sigma'}$, with

$$\mu' = (1 - 2q), \quad \sigma' = 2\sqrt{q(1-q)}. \tag{11}$$

The q-biased Fourier coefficients of g are

$$\tilde{g}(S) = \langle g, \psi_S \rangle = \mathbb{E}[g(\boldsymbol{Y})\psi_S(\boldsymbol{Y})], \qquad \forall S \subseteq [n]. \tag{12}$$

The following result expresses the probability of mismatch between $f(\boldsymbol{X})$ and $g(\boldsymbol{Y})$ in terms of their biased Fourier coefficients.

Lemma 1. For $(\boldsymbol{X}, \boldsymbol{Y})$ with $(X_i, Y_i)_{i \in [n]} \sim_{i.i.d.} P_{XY}$,

$$\mathbb{E}[f(\boldsymbol{X})g(\boldsymbol{Y})] = \sum_{S \in [n]} \bar{f}(S)\tilde{g}(S)\rho^{|S|}, \tag{13}$$

$$\mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{Y})) = \frac{1}{2} - \frac{1}{2}\sum_{S \in [n]} \bar{f}(S)\tilde{g}(S)\rho^{|S|}. \tag{14}$$

Proof: Using the p-biased Fourier expansion for $f(\boldsymbol{X})$ and the q-biased one for $g(\boldsymbol{Y})$, we have

$$\mathbb{E}[f(\boldsymbol{X})g(\boldsymbol{Y})] = \sum_{S \subseteq [n], S' \subseteq [n]} \bar{f}(S)\tilde{g}(S')\mathbb{E}[\phi_S(\boldsymbol{X})\psi_{S'}(\boldsymbol{Y})]$$

$$= \sum_{S \subseteq [n], S' \subseteq [n]} \bar{f}(S)\tilde{g}(S')\mathbb{E}\Big[\prod_{i \in S, j \in S'}\Big(\frac{X_i - \mu}{\sigma}\Big)\Big(\frac{Y_j - \mu'}{\sigma'}\Big)\Big]$$

$$\stackrel{(a)}{=} \sum_{S \subseteq [n]} \bar{f}(S)\tilde{g}(S)\prod_{i \in S}\mathbb{E}\Big[\Big(\frac{X_i - \mu}{\sigma}\Big)\Big(\frac{Y_i - \mu'}{\sigma'}\Big)\Big]$$

$$= \sum_{S \subseteq [n]} \bar{f}(S)\tilde{g}(S)\rho^{|S|}. \tag{15}$$

Here (a) is obtained as follows, using the independence of the (X_i, Y_i) pairs across $i \in [n]$: when $S \neq S'$, there is at least one index that belongs to only one of these two sets. If $i \in S$ and $i \notin S'$, the term $\mathbb{E}[(X_i - \mu)/\sigma] = 0$; similarly if $j \in S'$ and $j \notin S$, then $\mathbb{E}[(Y_j - \mu')/\sigma'] = 0$.

Eq. (14) follows by observing that

$$\mathbb{E}[f(\boldsymbol{X})g(\boldsymbol{Y})] = 1 \cdot \mathbb{P}(\boldsymbol{X} = \boldsymbol{Y}) - 1 \cdot \mathbb{P}(\boldsymbol{X} \neq \boldsymbol{Y}).$$

∎

For $|\rho| < 1$, Lemma 1 shows that the biased Fourier coefficients corresponding to sets of small cardinality play a key role in determining probability of mismatch. Since f and g are Boolean, by Parseval's formula we have

$$\sum_{S \in [n]} |\bar{f}(S)|^2 = \sum_{S \in [n]} |\tilde{g}(S)|^2 = 1. \tag{16}$$

Suppose that the biased Fourier coefficients of f and g are both largely concentrated on sets S of small cardinality. Then if the coefficients have the same sign on these sets, then (14) shows that the probability of mismatch between $f(\boldsymbol{X})$ and $g(\boldsymbol{Y})$ will be small; if the coefficients have opposite signs on these sets, the probability of mismatch will be close to 1. On the other hand, if the biased Fourier coefficients of f, g are concentrated on sets S of large cardinality, then for $\rho < 1$, the probability of mismatch will be close to $1/2$.

Noise sensitivity: The noise sensitivity of a Boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ is defined as $\mathbb{P}(f(\boldsymbol{X}) \neq f(\boldsymbol{Y}))$, where $(X_i, Y_i)_{i \in [n]} \sim_{i.i.d.} P_{XY}$. It represents the mismatch probability under a perturbation model where the noisy input $\boldsymbol{Y}$ is assumed to be generated from the original input $\boldsymbol{X} \sim_{i.i.d.} P_X$ via a memoryless channel $P_{Y|X}$.

By taking $f = g$, Lemma 1 yields the noise sensitivity for a general bivariate distribution P_{XY} on a pair of Boolean random variables, parametrized by (p, q, ρ). From (14), the noise sensitivity of f can be expressed as

$$\mathsf{NS}_{(p,q,\rho)} = \frac{1}{2} - \frac{1}{2} \sum_{S \in [n]} \bar{f}(S)\tilde{f}(S)\rho^{|S|}, \tag{17}$$

where $\bar{f}(S)$ and $\tilde{f}(S)$ are the p-biased and q-biased Fourier coefficients, respectively. This generalizes previous characterizations of noise sensitivity [1], [6], which assumed a symmetric perturbation model with $p = q$.

In the following sections, we will use Lemma 1 to obtain the mismatch probability for approximations of Boolean functions. We will apply Lemma 1 taking g to be the approximating function, and with $\boldsymbol{X} = \boldsymbol{Y}$ (i.e., $\rho = 1$).

IV. APPROXIMATION WITH k-JUNTAS

In the set of Boolean functions with n input variables, k-juntas are Boolean functions whose output depends only on a subset of at most k input variables.

Definition 1. *A Boolean function $g : \{-1, 1\}^n \to \{-1, 1\}$ is a k-junta (with $k < n$), if there exist $i_1, i_2, ..., i_k \in [n]$ and a Boolean function $h : \{-1, 1\}^k \mapsto \{-1, 1\}$ such that*

$$g(\boldsymbol{x}) = h(x_{i_1}, x_{i_2}, ..., x_{i_k}), \quad \forall \boldsymbol{x} \in \{-1, 1\}^n.$$

In this section, we investigate approximation of Boolean functions by k-juntas. Given a Boolean function $f : \{-1, 1\}^n \mapsto \{-1, 1\}$, we wish to find a k-junta g that minimizes the mismatch probability $\mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{X}))$, where

the entries of $\boldsymbol{X} = (X_1, ..., X_n)$ are i.i.d. according to (1). Letting $\mathcal{B}_k$ denote the set of all k-juntas, the minimum mismatch probability is denoted by

$$\mathsf{P}^k[f] \triangleq \min_{g \in \mathcal{B}_k} \mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{X})). \tag{18}$$

The following theorem gives an expression for $\mathsf{P}^k[f]$ and an optimal k-junta function for approximation of f. For $x \in \mathbb{R}$, we define $\mathsf{sign}(x) = 1$ if $x \geq 0$, and -1 if $x < 0$.

Theorem 1. *Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be a Boolean function with input $\boldsymbol{X} = (X_i)_{i \in [n]}$ i.i.d. according to the distribution in (3). Then the minimum mismatch probability of a k-junta approximation of f (for $k < n$) is*

$$\mathsf{P}^k[f] = \frac{1}{2} \left[1 - \max_{J \subseteq [n], |J| \leq k} \| f^{\subseteq J} \|_1 \right], \tag{19}$$

where $f^{\subseteq J}$ is the projection defined in (8), and

$$\| f^{\subseteq J} \|_1 = \mathbb{E}[|f^{\subseteq J}(\boldsymbol{X})|]. \tag{20}$$

Furthermore, the minimum mismatch probability is achieved by the k-junta approximation $g = \mathsf{sign}(f^{\subseteq J^*})$, where J^* achieves the optimum in (19).

Proof: We apply Lemma 1 taking g to be a k-junta, and $\rho = 1$, i.e., $\boldsymbol{X} = \boldsymbol{Y}$. From (14), for any g the mismatch probability satisfies

$$\mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{X})) = \frac{1}{2} - \frac{1}{2} \sum_{S \in [n]} \bar{f}(S)\bar{g}(S), \tag{21}$$

where $\bar{f}(S), \bar{g}(S)$ are the p-biased Fourier coefficients of f and g, respectively. Suppose that $g(\boldsymbol{x})$ depends on the inputs $(x_i)_{i \in J}$, where J is a subset of $[n]$ with at most k elements. Then, $\bar{g}(S) = 0$ for any $S \not\subseteq J$. Hence, the mismatch probability in (21) equals

$$\mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{X})) = \frac{1}{2} - \frac{1}{2} \sum_{S \subseteq J} \bar{f}(S)\bar{g}(S) = \frac{1}{2} - \frac{1}{2}\langle f^{\subseteq J}, g \rangle$$

$$\geq \frac{1}{2} - \frac{1}{2}\langle |f^{\subseteq J}|, |g| \rangle = \frac{1}{2} - \frac{1}{2}\| f^{\subseteq J} \|_1. \tag{22}$$

The last equality in (22) holds because g is a Boolean function, hence $\|g\| = 1$. Since J is an arbitrary subset of $[n]$ with at most k elements, (22) implies

$$\mathsf{P}^k[f] \geq \frac{1}{2} - \frac{1}{2} \max_{J \subseteq [n], |J| \leq k} \| f^{\subseteq J} \|_1. \tag{23}$$

Next we obtain an upper bound on $\mathsf{P}^k[f]$ by specifying a k-junta approximation of f. Fix a subset $J \subseteq [n]$ with $|J| \leq k$, and let $g = \mathsf{sign}[f^{\subseteq J}]$. Note that for any f we have

$$\langle g, f^{\subseteq J} \rangle = \mathbb{E}[f^{\subseteq J}(\boldsymbol{x}) \cdot \mathsf{sign}(f^{\subseteq J}(\boldsymbol{x}))]$$
$$= \mathbb{E}[|f^{\subseteq J}(\boldsymbol{x})|] = \| f^{\subseteq J} \|_1. \tag{24}$$

Therefore, using (22), the mismatch probability of this approximation is

$$\mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{X})) = \frac{1}{2} - \frac{1}{2}\| f^{\subseteq J} \|_1. \tag{25}$$

Eq. (25) provides an upper-bound on $\mathsf{P}^k[f]$ for any J such that $|J| \leq k$. Taking $J = J^*$, where J^* achieves $\max_{J \subseteq [n], |J| \leq k} \|f^{\subseteq J}\|_1$, we obtain

$$\mathsf{P}^k[f] \leq \frac{1}{2} - \frac{1}{2} \max_{J \subseteq [n], |J| \leq k} \|f^{\subseteq J}\|_1. \tag{26}$$

Combining (26) and (23) completes the proof. ∎

Remark 1. The proof shows that for any $J \subseteq [n]$, the mismatch probability between f and $\mathrm{sign}[f^{\subseteq J}]$ is given by (25). The function $\mathrm{sign}[f^{\subseteq J}]$ is the maximum a posteriori probability (MAP) estimator of f given J. To see this, note that the MAP estimator of f given $\boldsymbol{X}^J$ is a Boolean function g such that $g(\boldsymbol{x}^J) = 1$ if

$$\mathbb{P}(f(\boldsymbol{X}) = 1 | \boldsymbol{X}^J = x^J) \geq \mathbb{P}(f(\boldsymbol{X}) = -1 | \boldsymbol{X}^J = \boldsymbol{x}^J),$$

and $g(\boldsymbol{x}^J) = -1$ otherwise. Since f is a Boolean function, by the definition of $f^{\subseteq J}$, we have

$$f^{\subseteq J}(\boldsymbol{X}^J) = \mathbb{E}[f | \boldsymbol{X}^J] = \mathbb{P}\{f = 1 | \boldsymbol{X}^J\} - \mathbb{P}\{f = -1 | \boldsymbol{X}^J\}.$$

Hence, $\mathrm{sign}[f^{\subseteq J}]$ equals the MAP estimator of f.

Remark 2. Eq. (25) shows that the mismatch probability for approximating f with $\mathrm{sign}[f^{\subseteq J}]$ is determined by $\|f^{\subseteq J}\|_1$. We can bound the mismatch probability from above and below in terms of $\|f^{\subseteq J}\|_2$, which depends only the weight of the p-biased Fourier coefficients of $S \subseteq J$.

Corollary 1. *With the assumptions of Theorem 1, the minimum mismatch probability satisfies*

$$\frac{1}{2} \left[1 - \max_{J \subseteq [1,n], |J| \leq k} \|f^{\subseteq J}\|_2 \right]$$
$$\leq \mathsf{P}^k[f] \leq \frac{1}{2} \left[1 - \max_{J \subseteq [1,n], |J| \leq k} \|f^{\subseteq J}\|_2^2 \right], \tag{27}$$

where

$$\|f^{\subseteq J}\|_2^2 = \mathbb{E}[f^{\subseteq J}(\boldsymbol{X})^2] = \sum_{S \subseteq J} |\bar{f}(S)|^2. \tag{28}$$

Proof: Since $-1 \leq f^{\subseteq J} \leq 1$, we have $|f^{\subseteq J}(\boldsymbol{x})| \geq |f^{\subseteq J}(\boldsymbol{x})|^2$. Thus $\|f^{\subseteq J}\|_1 \geq \|f^{\subseteq J}\|_2^2$, which yields the upper bound by substituting in (19). Next, from Jensen's inequality we have

$$\mathbb{E}\left[|f^{\subseteq J}(\boldsymbol{X})|^2 \right] \geq \left[\mathbb{E}|f^{\subseteq J}(\boldsymbol{X})| \right]^2.$$

This implies that $\|f^{\subseteq J}\|_2 \geq \|f^{\subseteq J}\|_1$, which establishes the lower-bound. ∎

Given $k < n$, Theorem 1 specifies the optimal k-junta approximation for f. The problem may be viewed from another perspective: given $\epsilon > 0$, find the smallest k such that there exists a k-junta function whose mismatch probability with f is at most ϵ. When f depends on all n input variables, there is a trade-off between k and ϵ: the lower the tolerance ϵ, the larger the required value of k. As discussed in Section VI, this formulation can be useful in the context of learning arbitrary Boolean functions to within a specified mismatch probability.
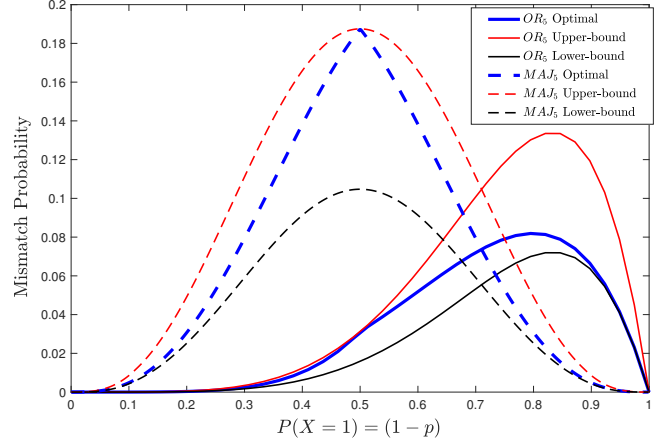

Fig. 1: Minimum mismatch probability for approximation of OR_5 and MAJ_5 using 4-juntas.

Examples: We examine k-junta approximations of the 'or' function OR_n, and the majority function MAJ_n. The function $\mathrm{OR}_n : \{-1, 1\}^n \mapsto \{-1, 1\}$ is defined as $\mathrm{OR}_n(\boldsymbol{x}) = 1$ if $\boldsymbol{x} = (1, 1, \ldots, 1)$, and $\mathrm{OR}_n(\boldsymbol{x}) = -1$ otherwise. The majority function is defined as $\mathrm{MAJ}_n(\boldsymbol{x}) = \mathrm{sign}(\sum_{i=1}^n x_i)$ for all $\boldsymbol{x} \in \{-1, 1\}^n$. Figure 1 shows the minimum mismatch probability as function of $P_X(1) = (1 - p)$ for the approximation of OR_5 and MAJ_5 using 4-juntas (i.e., $n = 5$ and $k = 4$). The bounds given in Corollary 1 are also plotted.

Using the symmetry between the inputs, we can show that

$$\mathrm{OR}_n^{\subseteq J}(\boldsymbol{x}^J) = -1 + (1 - p)^{n - |J|} \left[1 + \mathrm{OR}_{|J|}(\boldsymbol{x}^J) \right].$$

For $(1 - p) < \frac{1}{2}$, the optimal approximation $\mathrm{sign}(\mathrm{OR}_n^{\subseteq J})$ is therefore the constant function -1 (for $|J| < n$). For MAJ_n, the projection does not have a compact closed form expression and is computed as

$$\mathrm{MAJ}_n^{\subseteq J} = \mathbb{E}\left[\mathrm{sign}\left(\sum_{i=1}^n X_i \right) \mid \boldsymbol{X}^J \right].$$

V. APPROXIMATION WITH LINEAR BOOLEAN FUNCTIONS

A linear Boolean function is either a parity or a negation of a parity. More precisely, a Boolean function $f : \{-1, +1\}^n \mapsto \{-1, +1\}$ is linear if it is of the form $f(\boldsymbol{x}) = c \prod_{i \in S} x_i$ for some subset $S \subseteq [1, n]$ and constant $c \in \{-1, 1\}$.

Given a Boolean function f, we wish to find a linear Boolean function g that minimizes the mismatch probability $\mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{X}))$. Let $\mathcal{L}_n$ denote the set of linear Boolean functions with n input variables. The minimum mismatch probability is denoted by

$$\mathsf{P}^{\mathrm{lin}}[f] \triangleq \min_{g \in \mathcal{L}_n} \mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{X})), \tag{29}$$

where the entries of $\boldsymbol{X} = (X_i)_{i \in [n]}$ are i.i.d. according to (1).

For any Boolean function f and $S \subseteq [n]$, let

$$I_S[f] \triangleq \sum_{S' \subseteq S} \bar{f}(S') \sigma^{|S'|} \mu^{|S| - |S'|}, \tag{30}$$

where μ, σ are the mean and standard deviation of the $(X_i)_{i \in [n]}$, defined in (5).

Theorem 2. *Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be a Boolean function with input $\boldsymbol{X} = (X_i)_{i \in [n]}$ i.i.d. according to the distribution in (3). Then the linear Boolean function $g(\boldsymbol{x}) = c^* \boldsymbol{x}^{S^*}$ minimizes the mismatch probability where*

$$S^* = \arg\max_{S \subseteq [n]} |I_S[f]|, \ and \ c^* = \mathsf{sign}(I_{S^*}[f]). \quad (31)$$

The minimum mismatch probability is $\mathsf{P}^{\mathsf{lin}}[f] = \frac{1 - |I_{S^*}[f]|}{2}$.

Proof: We apply Lemma 1 with $\rho = 1$ (i.e., $\boldsymbol{X} = \boldsymbol{Y}$), and g a linear Boolean function. From (13)–(14), we have

$$\mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{X})) = \frac{1}{2} - \frac{1}{2}\mathbb{E}[f(\boldsymbol{X})g(\boldsymbol{X})]. \quad (32)$$

Since g is linear Boolean, $g(\boldsymbol{x}) = c\,\boldsymbol{x}^S$, for some $S \subseteq [n]$, and $c \in \{-1, 1\}$. Thus

$$\mathbb{E}[f(\boldsymbol{X})g(\boldsymbol{X})] = \mathbb{E}[f(\boldsymbol{X})\,c\boldsymbol{X}^S] = c \sum_{S' \subseteq S} \bar{f}(S)\mathbb{E}[\phi_{S'}(\boldsymbol{X})\boldsymbol{X}^S]$$

$$= c \sum_{S' \subseteq S} \bar{f}(S) \prod_{i \in S'} \mathbb{E}\left[\frac{(X_i - \mu)X_i}{\sigma}\right] \prod_{i \in S \setminus S'} \mathbb{E}X_i$$

$$= c \sum_{S' \subseteq S} \bar{f}(S)\sigma^{|S'|}\mu^{|S \setminus S'|} = c\,I_S[f]. \quad (33)$$

Substituting in (32), we deduce

$$\mathbb{P}(f(\boldsymbol{X}) \neq g(\boldsymbol{X})) = (1 - c\,I_S[f])/2. \quad (34)$$

The mismatch probability in (34) is minimized by taking $S = S^*$ and $c = c^*$, where $S^* = \arg\max_{S \subseteq [n]} |I_S[f]|$, and $c^* = \mathsf{sign}(I_{S^*}[f])$. $\blacksquare$

For uniformly random inputs ($p = \frac{1}{2}$), we have $\mu = 0, \sigma = 1$, which implies $I_S[f] = \bar{f}(S)$. The optimal linear approximation can be succinctly characterized in this case.

Corollary 2. *If the inputs $(X_i)_{i \in [n]}$ are uniformly random, then the mismatch probability with f is minimized by the linear Boolean function $g(\boldsymbol{x}) = c^* \boldsymbol{x}^{S^*}$ with*

$$S^* = \arg\max_{S \subseteq [n]} |\bar{f}(S)|, \ and \ c^* = \mathsf{sign}(\bar{f}(S)). \quad (35)$$

Here $\bar{f}(S)$ is the standard Fourier coefficient for the set S.

Figure 2 shows $\mathsf{P}^{\mathsf{lin}}$ for OR_5 and MAJ_5 as a function of $P_X(1)$. The optimal linear approximation for OR_5 is found to be a degree 5 linear function for $P_X(1) \in [0.815, 0.927]$, and the constant -1 function for other values of $P_X(1)$. For MAJ_5, the optimal linear approximation is a degree 1 function for $P_X(1) \in [0.389, 0.611]$, the constant -1 function for $P_X(1) < 0.389$, and the constant 1 for $P_X(1) > 0.611$. (The end points of these intervals are accurate up to 3 decimal places.)

VI. Discussion and Future Work

An interesting open question is whether we can efficiently *learn* the optimal approximation of an unknown function, using a small (polynomial in n) number of samples from the function. These samples may be generated from either
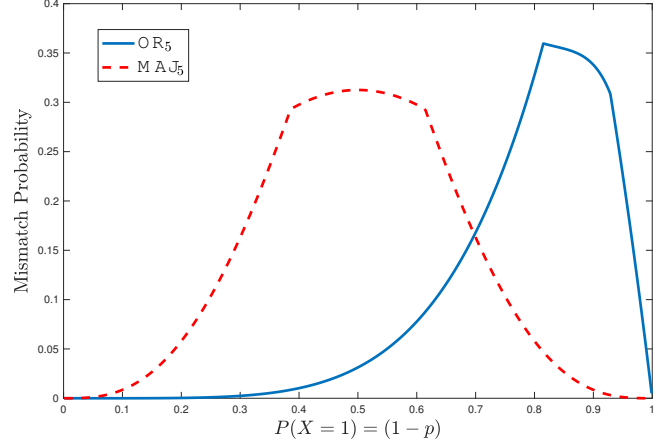

Fig. 2: Mismatch probability for approximation of MAJ_5 and OR_5 using linear Boolean functions.

uniformly distributed or biased inputs. For example, we may wish to learn the optimal k-junta approximation of a function, where k is large enough to achieve a desired mismatch probability. It is known that any k-junta can be learned with high probability with complexity of order $n^{\alpha k + \mathcal{O}(1)}$, where $\alpha < 1$ [3]. However this result is for the setting where the learning algorithm uses examples from the k-junta function. The question of how to efficiently learn the optimal k-junta approximation using examples from the original function is open. Similar questions may be posed for other useful classes of approximating functions such as linear threshold functions.

References

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

[2] N. Linial, Y. Mansour, and N. Nisan, "Constant depth circuits, Fourier transform, and learnability," *J. ACM*, vol. 40, no. 3, pp. 607–620, 1993.

[3] E. Mossel, R. O'Donnell, and R. A. Servedio, "Learning functions of k relevant variables," *J. Comput. Syst. Sci*, vol. 69, no. 3, pp. 421–434, 2004.

[4] G. Kalai, "Noise sensitivity and chaos in social choice theory," tech. rep., Hebrew University, 2005.

[5] J. Li and M. Médard, "Boolean functions: Noise stability, non-interactive correlation, and mutual information," in *Proc. IEEE ISIT*, 2018.

[6] E. Blais, R. O'Donnell, and K. Wimmer, "Polynomial regression under arbitrary product distributions," *Machine learning*, vol. 80, no. 2-3, pp. 273–294, 2010.

[7] T. A. Courtade and G. R. Kumar, "Which Boolean functions maximize mutual information on noisy inputs?," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4515–4525, 2014.

[8] N. Weinberger and O. Shayevitz, "On the optimal Boolean function for prediction under quadratic loss," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4202–4217, 2017.

[9] N. Weinberger and O. Shayevitz, "Self-predicting Boolean functions," in *Proc. IEEE ISIT*, 2018.

[10] M. L. Furst, J. C. Jackson, and S. W. Smith, "Improved learning of AC^0 functions," in *COLT*, vol. 91, pp. 317–325, 1991.

[11] M. Talagrand, "On Russo's approximate zero-one law," *The Annals of Probability*, vol. 22, no. 3, pp. 1576–1587, 1994.

[12] F. S. Chaharsooghi and S. S. Pradhan, "On the correlation between Boolean functions of sequences of random variables," in *Proc. IEEE Int. Symp. Inf. Theory*, 2017.

[13] A. E. Andreev, A. E. Clementi, and J. P. Rolim, "Optimal bounds for the approximation of Boolean functions and some applications," *Theoretical Computer Science*, vol. 180, no. 1-2, pp. 243–268, 1997.