Pseudo-Boolean Functions, Lovász Extensions, and Beta Distributions ⋆

Guoli Ding R. F. Lax *

Department of Mathematics, LSU, Baton Rouge, LA 70803

Abstract

Let $f : \{0,1\}^n \to \mathbb{R}$ be a pseudo-Boolean function and let $\hat{f} : [0,1]^n \to \mathbb{R}$ denote the Lovász extension of f. We show that the best linear approximation to f with respect to a binomial probability distribution on $\{0,1\}^n$ is the limit of the best linear approximations to $\hat{f}$ with respect to probability distributions on $[0,1]^n$ defined by certain independent beta random variables. When $n = 2$, we give an explicit formula for the best linear approximation to $\hat{f}$ that involves beta functions and hypergeometric series. In this case, we also show that when the two parameters of the beta random variables are equal, then the coefficients of x and y in the best linear approximation are each 1/2.

Key words: pseudo-Boolean function, Lovász extension, beta distribution, hypergeometric series, linear approximation

1 Introduction.

A pseudo-Boolean function of n variables is a function from $\{0,1\}^n$ to the real numbers. Such functions are used in 0-1 optimization problems, cooperative game theory, multicriteria decision making, and as fitness functions. Such a function $f(x_1, \ldots, x_n)$ has a unique expression as a multilinear polynomial

$$f(x_1, \ldots, x_n) = \sum_{T \subseteq N} \left[a_T \prod_{i \in T} x_i \right], \tag{1}$$

⋆ Research partially supported by NSF grant ITR-0326387 and AFOSR grants F49620-03-1-0238,F49620-03-1-0239, and F49620-03-1-0241

* Corresponding author

Email addresses: ding@math.lsu.edu (Guoli Ding), lax@math.lsu.edu (R. F. Lax).

where $N = \{1, \dots, n\}$ and the a_T are real numbers ([10, p. 22]). By the *degree* of a pseudo-Boolean function, we mean the degree of its multilinear polynomial representation.

Several authors have considered the problem of finding the best pseudo-Boolean function of degree $\leq k$ approximating a given pseudo-Boolean function f, where "best" means a least squares criterion. Hammer and Holzman [9] derived a system of equations for finding such a best degree $\leq k$ approximation, and gave explicit solutions when $k = 1$ and $k = 2$. Grabisch, Marichal, and Roubens [8] solved the system of equations derived by Hammer and Holzman, and gave the following explicit formula for the coefficients of the best degree $\leq k$ function.

Theorem 1 *[8] If $f_k^* = \sum_{T \subseteq N} [b_T^* \prod_{i \in T} x_i]$ is the best degree $\leq k$ approximation to f, then*

$$b_T^* = a_T + (-1)^{k-|T|} \sum_{R \supseteq T, |R| > k} (\frac{1}{2})^{|R|-|T|} \binom{|R| - |T| - 1}{k - |T|} a_R, \tag{2}$$

for all $T \subseteq N$ with $|T| \leq k$.

Given a pseudo-Boolean function f of n variables, L. Lovász [11] defined an extension $\hat{f}$ of f so that $\hat{f}$ is defined on all n-tuples of nonnegative real numbers. Let $\mathcal{S}_n$ denote the set of all permutations on N. For each $\sigma \in \mathcal{S}_n$, let A_σ denote the n-simplex given by

$$A_\sigma = \{(x_1, \dots, x_n) \in [0, 1]^n \mid x_{\sigma(1)} \leq \cdots \leq x_{\sigma(n)}\}.$$

The $n!$ simplices $A_\sigma, \sigma \in \mathcal{S}_n$, give a triangulation of $[0, 1]^n$, which is called the "standard" triangulation of $[0, 1]^n$ in [12]. I. Singer [15] showed that $\hat{f}$ is the unique affine function on A_σ that agrees with f at the $n + 1$ vertices of this simplex. By the "Lovász extension of f," we will mean here the extension $\hat{f}$ restricted to the n-cube $[0, 1]^n$. Lovász extensions of pseudo-Boolean functions also appear as discrete Choquet integrals in aggregation theory ([13]). Grabisch et al. [8] observed that, given f as in (1), the Lovász extension of f is given by

$$\hat{f}(x) = \sum_{T \subseteq N} \left[a_T \bigwedge_{i \in T} x_i \right], \tag{3}$$

where $\bigwedge$ denotes the *min* operation. As in [14], we will refer to the right side of the above equation as a "min-polynomial." The degree of such a min-polynomial is defined to be the degree of the multilinear polynomial given in (1).

Let V_k denote the set of all min-polynomials $\hat{f}_k : [0, 1]^n \to \mathbb{R}$ of degree at most k. Given $\hat{f} \in V_n$, J.-L. Marichal and P. Mathonet [14] defined the best kth approximation of $\hat{f}$ to be the min-polynomial $\hat{f}_k \in V_k$ that minimizes

$$\int_{[0,1]^n} \left[\hat{f}(x_1, \ldots, x_n) - \hat{g}(x_1, \ldots, x_n)\right]^2 \, dx_1 \cdots dx_n \tag{4}$$

among all min-polynomials $\hat{g} \in V_k$, and they obtained the following result.

Theorem 2 *[14] If $\hat{f}_k^* = \sum_{T \subseteq N} [b_T^* \bigwedge_{i \in T} x_i]$ is the best kth approximation to $\hat{f}$, then*

$$b_T^* = a_T + (-1)^{k-|T|} \sum_{R \supseteq T, |R|>k} \frac{\binom{k+|T|+1}{k+1}\binom{|R|-|T|-1}{k-|T|}}{\binom{k+|R|+1}{k+1}} a_R, \tag{5}$$

for all $T \subseteq N$ with $|T| \leq k$.

The similarities between (2) and (5) raise the question of whether these two formulas might be special cases or limiting cases of a more general formula. While we will only be able to provide such a formula in the simplest possible case, we will show that these two approximation problems, one discrete and one continuous, are linked. The connection between the two involves jointly distributed beta random variables on the n-cube.

In the next section, we consider probability distributions on $\{0, 1\}^n$ and on $[0, 1]^n$, and we modify the notion of "best approximation" taking into account these distributions. In the final section, we consider the $n = 2$ case and we will see that this apparently simple case is quite nontrivial. We thank W. George Cochran for very helpful conversations.

2 Probability distributions

Put $\mathbb{B} = \{0, 1\}$. Let $\mathcal{F}$ denote the space of all pseudo-Boolean functions in n variables; i.e.,

$$\mathcal{F} = \{f : \mathbb{B}^n \to \mathbb{R}\}.$$

Then $\mathcal{F}$ has the structure of a 2^n-dimensional real vector space. A basis for this vector space is $\{\prod_{i \in T} x_i \, : \, T \subseteq N\}$.

As in [5] and [6], we wish to allow a weighting on the elements of $\mathbb{B}^n$. By scaling, we may assume this weighting defines a probability mass function $\mu(\mathbf{x})$ on $\mathbb{B}^n$. As in [5], define a pseudo-inner product $\langle \ , \ \rangle_\mu$ on $\mathcal{F}$ by

$$\langle f, g \rangle_\mu = \sum_{\mathbf{x} \in \mathbb{B}^n} f(\mathbf{x}) g(\mathbf{x}) \mu(\mathbf{x}).$$

This is a "pseudo" (or semidefinite) inner product because we may have $\langle f, g \rangle_\mu = 0$ for all g without f being identically zero. Indeed, if $\mu(\mathbf{x}) = 0$ and if f satisfies $f(\mathbf{x}) = 1$ and $f(\mathbf{y}) = 0$ for all $\mathbf{y} \neq \mathbf{x}$, then $\langle f, g \rangle_\mu = 0$ for all

g. On the other hand, it is easy to see that if $\mu(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbb{B}^n$, then this pseudo-inner product will be an inner product. For the remainder of this work, we assume that $\mu(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbb{B}^n$. This is not a serious practical restriction, since if one would like some n-tuples to have zero weight, then those n-tuples could be assigned an extremely small positive weight; moreover, we will be focusing on binomial distributions, which satisfy this positivity requirement. We note that $\langle f, g\rangle_\mu$ is the expected value $E_\mu(fg)$ of the random variable fg. Put $\| f \|_\mu = \sqrt{\langle f, f\rangle_\mu}$. Then $\| \quad \|_\mu$ is a norm, under our positivity assumption above.

Let $\mathcal{L} \subseteq \mathcal{F}$ be an affine space (a translation of a subspace; also known as a linear variety). For example, $\mathcal{L}$ might be the subspace of all pseudo-Boolean functions of degree at most k, for some fixed k. Given $f \in \mathcal{F}$, the best approximation to f with respect to μ by functions in $\mathcal{L}$ is the function $f^*_{\mathcal{L},\mu} \in \mathcal{L}$ that minimizes

$$\| f - g \|_\mu = \sqrt{\sum_{\mathbf{x}\in\mathbb{B}^n} (f(\mathbf{x}) - g(\mathbf{x}))^2 \mu(\mathbf{x})}$$

over all $g \in \mathcal{L}$. Notice that if we take the uniform distribution on $\mathbb{B}^n$, so that $\mu(\mathbf{x}) = (1/2)^n$ for all $\mathbf{x} \in \mathbb{B}^n$, then the best approximation to f in $\mathcal{L}$ is the function $f^* \in \mathcal{L}$ that also minimizes $\sum_{\mathbf{x}\in\mathbb{B}^n}(f(\mathbf{x}) - g(\mathbf{x}))^2$, over all $g \in \mathcal{L}$. This is the usual least squares condition used in [9] and [8], and in this case one may simply use the usual Euclidean inner product in $\mathbb{R}^{2^n}$.

If $x_1, \ldots, x_n$ are independent identically distributed Bernoulli random variables with p being the probability of a 1, then we call the resulting joint distribution on $\mathbb{B}^n$ the (n, p) binomial distribution. In [5], we proved the following generalization of Theorem 1.

Theorem 3 *[5] Let μ be the (n, p) binomial distribution on $\mathbb{B}^n$. If $f^*_{k,\mu} = \sum_{T\subseteq N} [b^*_T \prod_{i\in T} x_i]$ is the best degree $\leq k$ approximation to f with respect to μ, then*

$$b^*_T = a_T + (-1)^{k-|T|} \sum_{R\supseteq T, |R|>k} p^{|R|-|T|} \binom{|R| - |T| - 1}{k - |T|} a_R, \tag{6}$$

for all $T \subseteq N$ with $|T| \leq k$.

Now suppose we have an absolutely continuous probability measure on $[0, 1]^n$ with density function ν. Then we can define an inner product on the space V_n by

$$\langle \hat{f}, \hat{g}\rangle_\nu = \int_{[0,1]^n} \hat{f}\hat{g}\, \nu \, dx_1 \cdots dx_n;$$

that is, $\langle \hat{f}, \hat{g}\rangle_\nu = E_\nu(\hat{f}\hat{g})$. We also have the corresponding norm

$$\| \hat{f} \|_\nu = \sqrt{\langle \hat{f}, \hat{f}\rangle_\nu}.$$

Given $\hat{f} \in V_n$, we define the best kth approximation to $\hat{f}$ with respect to ν to be the min-polynomial $\hat{f}^*_{k,\nu} \in V_k$ that minimizes $\| \hat{f} - \hat{g} \|_\nu$ among all $\hat{g} \in V_k$.

If a and b are positive real numbers, then recall that the distribution determined by the beta-(a, b) random variable has density function

$$\begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\, dx$. Notice that the beta-$(1,1)$ distribution is simply the uniform distribution on $[0,1]$.

We wish to relate the best linear approximation of the Lovász extension of a pseudo-Boolean function f with the best linear approximation of f itself with respect to a binomial distribution, which we may consider as a probability distribution on the n-cube $[0,1]^n$ whose support is the set $\{0,1\}^n$ of vertices of the cube. We will be interested in probability distributions on $[0,1]^n$ determined by beta random variables because of the following result (cf. [7, pp. 288–289]).

Proposition 4 *Let X denote the Bernoulli random variable with $P(X = 1) = p, P(X = 0) = q$. For each positive real number a, let X_a denote the beta-$(a, (q/p)a)$ random variable. As $a \to 0^+$, the random variables X_a converge in distribution to X.*

PROOF. Put $c = q/p$. We have

$$\begin{aligned} E(X_a^n) &= \frac{1}{B(a, ca)} \int_0^1 x^{a+n-1}(1-x)^{ca-1}\, dx = \frac{B(a+n, ca)}{Ba, ca} \\ &= \frac{\Gamma(a+n)}{\Gamma(a)} \frac{\Gamma((c+1)a)}{\Gamma((c+1)a+n)} \\ &= \prod_{m=0}^{n-1} \frac{a+m}{(c+1)a+m}, \end{aligned}$$

where Γ denotes the gamma function. It follows that, for each $n \geq 1$, we have

$$\lim_{a \to 0^+} E(X_a^n) = \frac{1}{c+1} = p.$$

Since these moments are the same as the moments of our Bernoulli random variable, the proposition follows by the "method of moments" [4, Theorem 4.5.5]. $\Box$

Let μ denote the (n,p) binomial distribution on $\mathbb{B}^n$ and consider the probability distribution on $[0,1]^n$ whose density ν_a is the product of the densities of n independent beta-$(a,(q/p)a)$ random variables. Let f denote a pseudo-Boolean function of n variables and let $\hat{f}$ denote its Lovász extension. Then, as a consequence of the above result, we have the following corollary.

Corollary 5 $E_\mu(f) = \lim_{a\to 0^+} E_{\nu_a}(\hat{f})$.

Theorem 6 *If $f^*_{1,\mu}$ denotes the best linear approximation to f with respect to μ and $\hat{f}^*_{1,\nu_a}$ denotes the best linear approximation to $\hat{f}$ with respect to ν_a, then $f^*_{1,\mu} = \lim_{a\to 0^+} \hat{f}^*_{1,\nu_a}$.*

PROOF. Let V denote the real vector space generated by $1, x_1, \ldots, x_n$. We will consider orthonormal bases of V with respect to different inner products. First, we view each of these functions as being defined on $\{0,1\}$ and we consider the inner product $\langle\ ,\ \rangle_\mu$ on the pseudo-Boolean functions of n variables. Since the inner product of two functions is the expected value of their product, it is not hard to see, as in [6], that we obtain an orthonormal basis with respect to this inner product by standardizing $x_1, \ldots, x_n$ when we view each of these as a Bernoulli random variable on $\{0,1\}$. Specifically, since the expected value of each of these variables is p and the standard deviation is $\sqrt{pq}$, our orthonormal basis is $1, z_1, \ldots, z_n$, where

$$z_i = \frac{x_i - p}{\sqrt{pq}} \text{ for } i = 1, 2, \ldots, n.$$

The best linear approximation to f with respect to μ is then

$$f^*_{1,\mu} = \langle f, 1\rangle_\mu + \sum_{i=1}^{n} \langle f, z_i\rangle_\mu \, z_i. \tag{7}$$

Now we view each of the functions $1, x_1, \ldots, x_n$ as being continuous on $[0,1]$ and we consider the inner product $\langle\ ,\ \rangle_{\nu_a}$ on $V = V_1$, the vector space of min-polynomials of degree at most 1. We obtain an orthonormal basis with respect to this inner product by standardizing $x_1, \ldots, x_n$ when we view each of these as a beta-$(a,(q/p)a)$ random variable on $[0,1]$. The expected value of each of these variables is

$$\frac{a}{a + (q/p)a} = \frac{1}{1 + (q/p)} = p$$

and the standard deviation of each of these variables is

$$\sigma = \sqrt{\frac{a(a + (q/p)a)}{(a + (q/p)a)^2(a + (q/p)a + 1)}} = \sqrt{pq\frac{p}{a+p}}.$$

An orthonormal basis of V_1 with respect to this inner product is then $1, w_1, \dots, w_n$, where

$$w_i = \frac{x_i - p}{\sigma} \text{ for } i = 1, 2, \dots, n.$$

The best linear approximation to $\hat{f}$ with respect to ν_a is then

$$\hat{f}^*_{1,\nu_a} = \langle \hat{f}, 1 \rangle_{\nu_a} + \sum_{i=1}^{n} \langle \hat{f}, w_i \rangle_{\nu_a} w_i. \tag{8}$$

Now let $a \to 0^+$. Notice that at any vertex $(c_1, \dots, c_n)$ of the n-cube, we have $\lim_{a \to 0^+} w_i(c_1, \dots, c_n) = z_i(c_1, \dots, c_n)$. Then, since our inner products are expected values, we have from Corollary 5 that

$$\lim_{a \to 0^+} \langle \hat{f}, w_i \rangle_{\nu_a} = \langle f, z_i \rangle_{\mu},$$

and the theorem follows. $\square$

3 The $n = 2$ case

In order to get an explicit formula for the best linear approximation to a Lovász extension with respect to the distribution ν_a of Theorem 6, one would need to compute moments of order statistics of jointly distributed independent beta random variables on the n-cube. While some results in that direction have been obtained in [1] and [16], those authors assumed that the parameters of the beta random variables were integers (so that the binomial expansion of $(1-x)^{b-1}$ is a polynomial). We will be able to give such an explicit formula for the best linear approximation only in the $n = 2$ case.

Consider the probability distribution on the unit square $[0, 1]^2$ with density

$$\nu = \nu_{(a,b)} = \frac{1}{B(a, b)^2} x^{a-1} (1-x)^{b-1} y^{a-1} (1-y)^{b-1}.$$

To be able to find the best linear approximation (with respect to ν) of any min-polynomial in V_2, we need to find the best linear approximation to $x \wedge y$, the minimum of x and y. Our calculations of $\langle x \wedge y, 1 \rangle_\nu$ and $\langle x \wedge y, x \rangle_\nu$ will involve hypergeometric functions. Recall that the (generalized) hypergeometric function ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x)$ is given by the power series

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} \frac{x^k}{k!},$$

where $(a)_k = \Gamma(a + k)/\Gamma(k)$ is the "rising factorial" or Pochhammer symbol. When $p = q+1$, which will be of interest to us, this series converges absolutely for $|x| < 1$, and if $\mathrm{Re}(\sum b_i - \sum a_i) > 0$, then it also converges absolutely for $|x| = 1$ [2, Theorem 2.1.1].

Using symmetry, we have

$$\begin{aligned}\langle x \wedge y, 1\rangle_\nu &= \frac{1}{B(a,b)^2}\Big[\int_0^1 \int_0^x y\, x^{a-1}(1-x)^{b-1} y^{a-1}(1-y)^{b-1}\, dy\, dx + \\ &\quad \int_0^1 \int_0^y x\, x^{a-1}(1-x)^{b-1} y^{a-1}(1-y)^{b-1}\, dx\, dy\Big] \\ &= \frac{2}{B(a,b)^2}\int_0^1 x^{a-1}(1-x)^{b-1} B_x(a+1,b)\, dx,\end{aligned}$$

where $B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}\, dy$ is the incomplete beta function. An elementary calculation, using the binomial series, shows that

$$B_x(a,b) = \frac{x^a}{a}\, {}_2F_1(a, 1-b; a+1; x).$$

Hence, we have

$$\begin{aligned}\langle x \wedge y, 1\rangle_\nu &= \frac{2}{B(a,b)^2}\int_0^1 x^{a-1}(1-x)^{b-1}\frac{x^{a+1}}{a+1}\, {}_2F_1(a+1, 1-b; a+2; x)\, dx \\ &= \frac{2}{(a+1)B(a,b)^2}\int_0^1 x^{2a}(1-x)^{b-1}\, {}_2F_1(a+1, 1-b; a+2; x)\, dx.\end{aligned}$$

The last integral is a beta transform of a hypergeometric function and by [2, (2.2.2)], we get that

$$\langle x \wedge y, 1\rangle_\nu = \frac{2B(2a+1,b)}{(a+1)B(a,b)^2}\, {}_3F_2(2a+1, a+1, 1-b; 2a+b+1, a+2; 1). \quad (9)$$

We next compute $\langle x \wedge y, x\rangle_\nu$, which is the same as $\langle x \wedge y, y\rangle_\nu$ by symmetry.

We have

$$
\begin{aligned}
\langle x \wedge y, x\rangle_\nu =& \frac{1}{B(a,b)^2}\Big[\int_0^1\int_0^x xy\, x^{a-1}(1-x)^{b-1}y^{a-1}(1-y)^{b-1}\,dy\,dx+\\
&\int_0^1\int_0^y x^2\, x^{a-1}(1-x)^{b-1}y^{a-1}(1-y)^{b-1}\,dx\,dy\Big]\\
=& \frac{1}{B(a,b)^2}\Big[\int_0^1 x^a(1-x)^{b-1}\,B_x(a+1,b)\,dx+\\
&\int_0^1 y^{a-1}(1-y)^{b-1}B_y(a+2,b)\,dy\Big]\\
=& \frac{1}{B(a,b)^2}\Big[\frac{1}{a+1}\int_0^1 x^{2a+1}(1-x)^{b-1}\,{}_2F_1(a+1,1-b;a+2;x)\,dx+\\
&\frac{1}{a+2}\int_0^1 y^{a+1}(1-y)^{b-1}\,{}_2F_1(a+2,1-b;a+3;y)\,dy\Big]\\
=& \frac{B(2a+2,b)}{B(a,b)^2}\Big[\frac{1}{a+1}\,{}_3F_2(2a+2,a+1,1-b;2a+b+2;a+2;1)+\\
&\frac{1}{a+2}\,{}_3F_2(2a+2,a+2,1-b;2a+b+2,a+3;1)\Big]
\end{aligned}
\tag{10}
$$

The expected value and standard deviation of a beta-(a,b) random variable are $a/(a+b)$ and

$$
\sigma = \sigma_{(a,b)} = \sqrt{\frac{ab}{(a+b)^2(a+b+1)}},
$$

respectively. Hence, an orthonormal basis (with respect to $\langle\ ,\ \rangle_\nu$) for V_1 is

$$
1, \frac{x-\frac{a}{a+b}}{\sigma}, \frac{y-\frac{a}{a+b}}{\sigma}.
$$

Using (8), we then have the following result.

Theorem 7 *The best linear approximation to $x\wedge y$ with respect to $\nu=\nu_{(a,b)}$ is $c_0(a,b)+c_1(a,b)x+c_1(a,b)y$, where*

$$
\begin{aligned}
c_0(a,b) &= \langle x\wedge y,1\rangle_\nu - \frac{2a}{(a+b)\sigma^2}\left[\langle x\wedge y,x\rangle_\nu - \frac{a}{a+b}\langle x\wedge y,1\rangle_\nu\right]\\
c_1(a,b) &= \frac{1}{\sigma^2}\left[\langle x\wedge y,x\rangle_\nu - \frac{a}{a+b}\langle x\wedge y,1\rangle_\nu\right],
\end{aligned}
$$

and where these inner products are given in terms of beta functions and hypergeometric series in (9) and (10).

We now consider the special case of the uniform distribution with $n=2$. For the uniform distribution on $\{0,1\}^n$, the best linear approximation to the pseudo-Boolean function xy (by [9] or Theorem 1) is

$$-\frac{1}{4}+\frac{1}{2}x+\frac{1}{2}y. \tag{11}$$

For the uniform distribution on $[0,1]^n$, the best linear approximation to the Lovász extension $x \wedge y$ (by Theorem 2 or Theorem 7 with $a=b=1$) is

$$-\frac{1}{6}+\frac{1}{2}x+\frac{1}{2}y. \tag{12}$$

By our Theorem 6, these two approximations are related by considering the $\nu_{(a,a)}$ distribution on $[0,1]^n$, with the uniform distribution on $[0,1]^n$ being the case $a=1$ and the uniform distribution on $\{0,1\}^n$ being the limiting distribution as a approaches 0. The above two formulas suggest that perhaps the coefficient of x (and y) in the best linear approximation to $x \wedge y$ is *always* $1/2$ with respect to the $\nu_{(a,a)}$ distribution. Our final result will be a proof of this. From Theorem 7 and our inner product calculations, we have the following.

Corollary 8 *The coefficient of x in the best linear approximation to $x \wedge y$ with respect to the $\nu_{(a,a)}$ distribution is*

$$\begin{aligned}\frac{4(2a+1)B(2a+1,a)}{B(a,a)^2}\Big[\frac{2a+1}{3a+1}\Big(&\frac{1}{a+1}{}_3F_2(2a+2,a+1,1-a;3a+2,a+2;1)\\ &\frac{1}{a+2}{}_3F_2(2a+2,a+2,1-a;3a+2,a+3;1)\Big)-\\ &\frac{1}{a+1}{}_3F_2(2a+1,a+1,1-a;3a+1,a+2;1)\Big].\end{aligned} \tag{13}$$

Our claim is that (13) is $1/2$ for a any positive real number. The three hypergeometric series involved in (13) are contiguous ${}_3F_2(1)$ series. As noted by W. N. Bailey [3], any three contiguous series of the type ${}_3F_2(1)$ satisfy a linear dependence relation. We will find such a relation by using Wilson's method, as explained in [2, 3.7].

Let $F = F(A,B,C;D,E;1)$ denote a general ${}_3F_2(1)$. Using the equations and method on page 157 of [2], it is not hard to derive the following three contiguous relations:

$$
\begin{aligned}
&F(A,B+1,C;D,E+1;1)-F(A,B,C;D,E;1)= \\
&\qquad\qquad \frac{AC(E-B)}{DE(E+1)}F(A+1,B+1,C+1;D+1,E+2;1) \\
&DF(A,B,C;D,E;1)-(D-A)F(A,B+1,C+1;D+1,E+1;1)= \\
&\qquad\qquad \frac{A(E-B)(E-C)}{E(E+1)}F(A+1,B+1,C+1;D+1,E+2;1) \\
&F(A-1,B,C;D-1,E;1)-F(A,B,C;D,E;1)= \\
&\qquad\qquad \frac{BC(A-D)}{(D-1)DE}F(A,B+1,C+1;D+1,E+1;1)
\end{aligned}
$$

Eliminating $F(A+1,B+1,C+1;D+1,E+2;1)$ and $F(A,B+1,C+1;D+1,E+1;1)$ from these equations gives us the contiguous relation we will need:

$$
\begin{aligned}
&(D-B-1)F(A,B,C;D,E;1)+(1-D)F(A-1,B,C;D-1,E;1)= \\
&\qquad\qquad \frac{B(C-E)}{E}F(A,B+1,C;D,E+1;1)
\end{aligned}
\tag{14}
$$

Now, put

$$
\begin{aligned}
F &= {}_3F_2(2a+2,a+1,1-a;3a+2,a+2;1) \\
G &= {}_3F_2(2a+2,a+2,1-a;3a+2,a+3;1) \\
H &= {}_3F_2(2a+1,a+1,1-a;3a+1,a+2;1)
\end{aligned}
$$

Applying (14) and doing some algebra, we obtain

$$
\frac{2a+1}{(3a+1)(a+2)}G-\frac{1}{a+1}H=\frac{2a}{(a+1)(3a+1)}F. \tag{15}
$$

Substituting (15) into (13), we see that the coefficient of x in the best linear approximation to $x \wedge y$ with respect to the $\nu_{(a,a)}$ distribution can be written as

$$
\frac{4(2a+1)(4a+1)B(2a+1,a)}{(a+1)(3a+1)B(a,a)^2}{}_3F_2(2a+2,a+1,1-a;3a+2,a+2;1). \tag{16}
$$

Now, some, but not all (cf. [17]), ${}_3F_2(1)$'s have a closed form as a rational expression involving evaluating the gamma function at the images of the parameters of the ${}_3F_2$ under certain affine mappings. The ${}_3F_2(1)$ in (16) is one that has such a closed form. Indeed, by Dixon's Formula [2, Theorem 3.4.1], we have

$$ {}_3F_2(2a+2, a+1, 1-a; 3a+2, a+2; 1) = \frac{\Gamma(a+2)^2\Gamma(3a+2)\Gamma(a)}{\Gamma(2a+3)\Gamma(2a+1)^2}. \quad (17) $$

Using the facts that $\Gamma(x+1) = x\Gamma(x)$ and $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, it is easy to see that

$$ \frac{B(2a+1,a)\Gamma(a+2)^2\Gamma(3a+2)\Gamma(a)}{B(a,a)^2\Gamma(2a+3)\Gamma(2a+1)^2} = \frac{(a+1)(3a+1)}{8(2a+1)}. \quad (18) $$

Finally, by (16), (17), and (18), we see that the coefficient of x in the best linear approximation to $x \wedge y$ with respect to the $\nu_{(a,a)}$ distribution is $1/2$. Combining this result with Theorem 7 and our calculation of $\langle x \wedge y, 1\rangle_\nu$ gives us the following corollary.

Corollary 9 *The best linear approximation to $x \wedge y$ with respect to the $\nu_{(a,a)}$ distribution is $c_0(a) + \frac{1}{2}x + \frac{1}{2}y$, where*

$$ c_0(a) = \frac{2B(2a+1,a)}{(a+1)B(a,a)^2}\, {}_3F_2(2a+1, a+1, 1-a; 3a+1, a+2; 1) - \frac{1}{2}. $$

Notice that when $a = 1$, which is the uniform distribution on $[0,1]^2$, we get

$$ c_0(1) = \frac{2B(3,1)}{2B(1,1)^2} - \frac{1}{2} = \frac{1}{3} - \frac{1}{2} = -\frac{1}{6}, $$

in agreement with (12). We also remark that as a consequence of Theorem 6, (11), and the above Corollary, we have that

$$ \lim_{a \to 0^+} \frac{2B(2a+1,a)}{(a+1)B(a,a)^2}\, {}_3F_2(2a+1, a+1, 1-a; 3a+1, a+2; 1) = \frac{1}{4}. $$

It is not clear to us how to prove this directly.

We note that the above Corollary about the coefficients of x and y in the best linear approximation to $x \wedge y$ being $1/2$ no longer holds with respect to the $\nu_{(a,b)}$ distribution if $a \neq b$. For example, a calculation with Mathematica shows that the coefficient of x in the best linear approximation to $x \wedge y$ with respect to the $\nu_{(\frac{1}{2},\frac{1}{3})}$ distribution is approximately 0.58477.

References

[1] Y. H. Abelkader, Computing the moments of order statistics from independent nonidentically distributed Beta random variables, Stat. Papers doi:10.1007/s00362-008-0161-0, 2008

[2] G. E. Andrews, R. Askey, and R. Roy, *Special Functions*. Encyclopedia of Mathematics and its Applications, 71. Cambridge University Press, Cambridge, 1999

[3] W. N. Bailey, Contiguous hypergeometric functions of the type ${}_3F_2(1)$, Proc. Glasgow Math. Assoc. 2, (1954), 62–65

[4] K. L. Chung, *A Course in Probability Theory, 2nd edition* Academic Press, New York, 1974

[5] G. Ding, R. F. Lax, J. Chen, and P. P. Chen, Formulas for approximating pseudo-Boolean random variables, Discrete Appl. Math. 156 (2008) 1581–1597

[6] G. Ding, R. F. Lax, J. Chen, P. P. Chen, and B. D. Marx, Transforms of pseudo-Boolean random variables, Discrete Appl. Math., to appear

[7] J. R. van Dorp and T. A. Mazzuchi, Parameter specification of the beta distribution and its Dirichlet extensions utilizing quantiles, in: *Handbook of Beta Distributions and Its Applications*, edited by A. K. Gupta and S. Nadarajah, Dekker, New York, 2004, 283–318

[8] M. Grabisch, J.-L. Marichal, and M. Roubens, Equivalent representations of set functions, Mathematics of Operations Research 25 (2000) 157–178

[9] P. L. Hammer and R. Holzman, Approximations of pseudo-Boolean functions; applications to game theory, ZOR–Methods and Models of Operations Research 36 (1992) 3–21

[10] P. L. Hammer and S. Rudeanu, Boolean Methods in Operations Research and Related Areas, Springer-Verlag, New York, 1968

[11] L. Lovász, Submodular functions and convexity, in: Mathematical programming: the state of the art (Bonn, 1982), Springer, Berlin, 1983, 235–257

[12] P. S. Mara, Triangulations for the cube, J. Combinatorial Theory Ser. A 20 (1976), no. 2, 170–177

[13] J.-L. Marichal, Aggregation of interacting criteria by means of the discrete Choquet integral, In: Aggregation Operators: New Trends and Applications, Physica, Heidelberg, 2002, pp. 224–244

[14] J.-L. Marichal and P. Mathonet, Approximations of Lovász extensions and their induced interaction index, Discrete Applied Math. 156 (2008) 11–24

[15] I. Singer, Extensions of functions of 0-1 variables and applications to combinatorial optimization, Numer. Funct. Anal. Optim. 7 (1984/1985) 23–62

[16] P. Y. Thomas and P. Samuel, Recurrence relations for the moments of order statistics from a beta distribution, Stat. Papers 49 (2008), 139–146

[17] J. Wimp, Irreducible recurrences and representation theorems for ${}_3F_2(1)$, Comput. Math. Appl. 9 (1983), 669–678