MINIMAL ANGLE SPREAD IN THE PROBABILITY SIMPLEX WITH RESPECT TO THE UNIFORM DISTRIBUTION

Heinz H. Bauschke[*]   Peter A. V. DiBerardino[†]

Abstract   We compute the minimal angle spread with respect to the uniform distribution in the probability simplex. The resulting optimization problem is analytically solved. The formula provided shows that the minimal angle spread approaches zero as the dimension tends to infinity. We also discuss an application in cognitive science.

1 INTRODUCTION

Throughout this paper, we assume that

$$(1.1) \qquad X := \mathbb{R}^n \text{ with inner product } \langle \cdot, \cdot \rangle : X \times X \to \mathbb{R},$$

and induced Euclidean norm $\| \cdot \|$. We also define the probability simplex by

$$(1.2) \qquad \Delta := \Delta_n := \big\{ (x_1, \ldots, x_n) \in \mathbb{R}^n_+ \,\big|\, x_1 + x_2 + \cdots + x_n = 1 \big\},$$

where $\mathbb{R}_+ = \big\{ x \in \mathbb{R} \,\big|\, x \geq 0 \big\}$. The probability simplex is of central importance in Statistics, Optimization, and Information Theory; see, e.g., [4] and [11]. (We write Δ_n if we wish to emphasize the dimension n.) It will be convenient to set

$$(1.3) \qquad \mathbf{1} := (1, 1, \ldots, 1) \in \mathbb{R}^n \quad \text{and} \quad u := \big(\tfrac{1}{n}, \tfrac{1}{n}, \ldots, \tfrac{1}{n} \big) \in \Delta_n.$$

The problem we investigate is the following: Given $p \in \Delta \smallsetminus \{u\}$, there exist two unique points $a = a(p)$ and $b = b(p)$ in Δ such that $\{u, p\} \subseteq [a, b]$ and $\|a - b\|$ is maximal. (Let us note that when, e.g., $n = 3$, then the mapping $\Delta \smallsetminus \{u\} \colon \mathbb{R} : p \mapsto \|a(p) - b(p)\|$ is continuous; however, it is not possible to extend it continuously — let alone in a smooth manner — at the point u.) The quantity $\|a - b\|$ can be thought of as the "width" of Δ with respect to p and

$$(1.4) \qquad \cos\big(\angle(a, b) \big) = \frac{\langle a, b \rangle}{\|a\| \|b\|}$$

as the cosine of the "angle spread" with respect to p.

[*]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. heinz.bauschke@ubc.ca, ORCID: 0000-0002-4155-9930.

[†]Psychology, University of Waterloo, Waterloo, ON N2L 3G1, Canada. pavdiber@uwaterloo.ca, ORCID: 0000-0001-7607-9278.

The *aim of this paper* is to *minimize the angle spread*, which equivalently corresponds to *maximizing its cosine*

$$(\text{P}) \qquad \max_{p \in \Delta \smallsetminus \{u\}} \frac{\langle a(p), b(p) \rangle}{\|a(p)\| \|b(p)\|}.$$

When $n = 2$, it is clear that the maximum value of (P) is $\cos(\pi/2) = 0$ and that every $p \in \Delta \smallsetminus \{u\}$ solves (P). Thus, we assume henceforth that

$$(1.5) \qquad n \in \{3, 4, \ldots\}.$$

Our main result can now be stated:

Theorem 1.1. *The maximum value of* (P) *is*

$$(1.6) \qquad \frac{n-2}{n+2}$$

and a pair realizing this maximum is $a^* = \frac{1}{n}(2, 1, 1, \ldots, 1, 0)$ *and* $b^* = \frac{1}{n}(0, 1, 1, \ldots, 1, 2)$. *Consequently, the minimal angle spread is*

$$(1.7) \qquad \arccos\left(\frac{n-2}{n+2}\right).$$

See Figure 1 for an illustration of Theorem 1.1 when $n = 3$. We note that as the dimension of the space n increases to infinity, the minimal angle spread approaches $\arccos(1) = 0$. We also note that when $n = 1, 2, 4$, then $\pi - \arccos((n-2)/(n+2))$ coincides with the dihedral angle of the tetrahedron, cube, octahedron, respectively. However, dihedral angles of other classical polyhedra (see [16]) do not seem to be related to the angles provided by (1.7).

Our strategy to prove Theorem 1.1 is to reduce the complexity of the problem in stages by exploiting its structure. Eventually, we are led to a one-dimensional problem which we then solve by Calculus.

The remainder of this paper is organized as follows. In Section 2, we collect a technical optimization result that will be used later. The computation of $[a, b]$ is carried out in Section 3. In Section 4, we set up the cosine quotient (1.4) in a more tractable form. The optimization is then tackled in Section 5 where we keep p_1 and p_n fixed. At last, the proof is completed in Section 6. In Section 7, we sketch an application of our results — which in fact motivated this note — in cognitive science. The presented results are a first contribution to using the minimal angle spread as a nonsmooth discrepancy term in this area.

Finally, we note that a reviewer pointed out that another possible departure point is to study sufficient conditions for strong convexity of integral functionals in two-stage stochastic linear programming. In particular, recent work by Claus and Spürkel [3] features minimal angles of similar though different kind. This is a promising direction to explore in future research.

The notation in this paper is fairly standard and follows largely [1].

2 AN AUXILIARY RESULT

It will be convenient to have the following result ready for future use.

Lemma 2.1. *Let* $m \in \{1, 2, \ldots\}$, *set* $\mathbf{1} := (1, 1, \ldots, 1) \in \mathbb{R}^m$, *and let* $\gamma \in \mathbb{R}$. *Set* $h \colon \mathbb{R}^m \to \mathbb{R}_+ \colon x \mapsto \|x - \gamma\mathbf{1}\|^2$, *let* $x \in \mathbb{R}^m$, *and set*

$$y := (\eta, \eta, \ldots, \eta) \in \mathbb{R}^m, \quad where \quad \eta := \tfrac{1}{m}(x_1 + x_2 + \cdots + x_m).$$
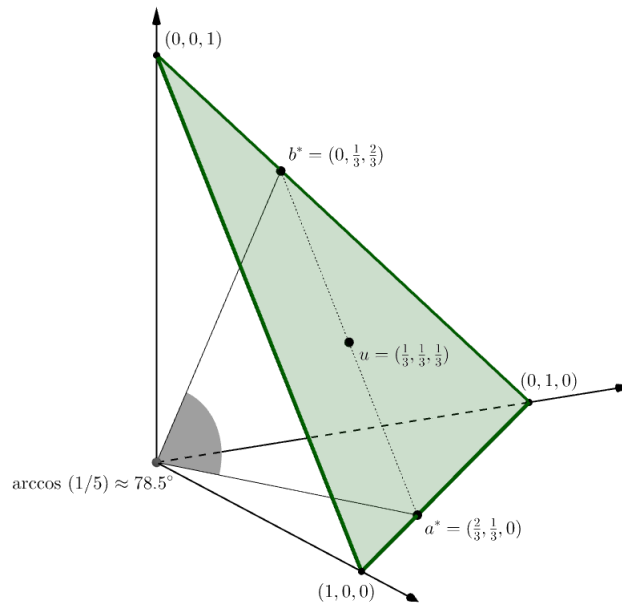
Then

$$h(y) \le h(x).$$

Figure 1: An illustration of Theorem 1.1 when $n = 3$.

Proof. Indeed, we have

$$
\begin{aligned}
h(y) - \gamma^2 \|\mathbf{1}\|^2 = \|y\|^2 - 2\gamma \langle y, \mathbf{1} \rangle &= m\eta^2 - 2\gamma m\eta \\
&= m\big(\tfrac{1}{m}(x_1 + x_2 + \cdots + x_m) \big)^2 - 2\gamma(x_1 + x_2 + \cdots + x_m) \\
&\leq m \tfrac{1}{m}\big(x_1^2 + x_2^2 + \cdots + x_m^2 \big) - 2\gamma \langle x, \mathbf{1} \rangle \\
&= \|x\|^2 - 2\gamma \langle x, \mathbf{1} \rangle \\
&= h(x) - \gamma^2 \|\mathbf{1}\|^2,
\end{aligned}
$$

where the inequality follows from the convexity of the square function.   □

Corollary 2.2. *Let* $m \in \{1, 2, \ldots\}$, *and let* $\gamma \in \mathbb{R}$. *Set* $A \colon \mathbb{R}^m \to \mathbb{R}^m \colon x \mapsto (\eta, \eta, \ldots, \eta)$, *where* $\eta = \frac{1}{m}(x_1 + x_2 + \cdots + x_m)$, *and let* C *be a nonempty subset of* $\mathbb{R}^m$ *such that* $A(C) \subseteq C$. *Set* $\mathbf{1} := (1, 1, \ldots, 1) \in \mathbb{R}^m$. *Then to*

$$
\text{(2.1)} \qquad\qquad \textit{minimize } \|x - \gamma\mathbf{1}\|^2 \;\; \textit{over } x \in C
$$

is the same as to

$$
\text{(2.2)} \qquad\qquad \textit{minimize } \|x - \gamma\mathbf{1}\|^2 \;\; \textit{over } x \in C \cap \mathbb{R}\mathbf{1}
$$

in the sense that the optimal values for both problems are identical and if the first problem has a solution, then it also has a solution that solves also the second problem.

Proof. Clear from Lemma 2.1.   □

Remark 2.3. We point out in passing that the operator A from Corollary 2.2 is the projection operator of the set $\mathbb{R}\mathbf{1}$.

3 DETERMINING THE RELATIVE BOUNDARY POINTS a AND b

In this section, we fix

$$(3.1) \qquad p \in \Delta \setminus \{u\}.$$

Then there exist indices i and j in $I := \{1, 2, \ldots, n\}$ such that $p_i < \frac{1}{n} < p_j$. Without loss of generality, we assume that

$$(3.2) \qquad 0 \le p_1 = \min_{i \in I} p_i < \frac{1}{n} < p_n = \max_{i \in I} p_i \le 1.$$

Set

$$(3.3) \qquad (\forall \lambda \in \mathbb{R}) \quad q(\lambda) := (1 - \lambda)u + \lambda p = u + \lambda(p - u).$$

Note that $q(0) = u$, $q(1) = p$,

$$(3.4) \qquad (\forall \lambda \in \mathbb{R})(\forall i \in I) \quad q_i(\lambda) = (1 - \lambda)u_i + \lambda p_i = u_i + \lambda(p_i - u_i),$$

and hence

$$(3.5) \qquad q_1(\lambda) + q_2(\lambda) + \cdots + q_n(\lambda) = 1.$$

We wish to find the smallest and largest $\lambda \in \mathbb{R}$ such that $q(\lambda) \in \Delta$. Let $i \in I$. Suppose that $p_i > u_i$. Then $(\forall \lambda \in \mathbb{R}_+)\ q_i(\lambda) > u_i$. So if $\lambda < 0$, then $q_i(\lambda) = 0 \Leftrightarrow -u_i = \lambda(p_i - u_i) \Leftrightarrow \lambda = -u_i/(p_i - u_i) = -(1/n)/(p_i - (1/n)) = -1/(np_i - 1) \le -1/(np_n - 1)$. Hence the smallest value λ still guaranteeing $q(\lambda) \in \Delta$ is

$$(3.6) \qquad \lambda_- := \frac{-1}{np_n - 1} = \frac{1}{1 - np_n} < 0.$$

Analogously, the largest value λ still guaranteeing $q(\lambda) \in \Delta$ is

$$(3.7) \qquad \lambda_+ := \frac{1}{1 - np_1} > 0.$$

The corresponding vectors

$$(3.8) \qquad a := q(\lambda_-) = u + \lambda_-(p - u)$$

and

$$(3.9) \qquad b := q(\lambda_+) = u + \lambda_+(p - u)$$

thus form the largest segment $[a, b] \subseteq \Delta$ such that $\{u, p\} \subseteq [a, b]$. Note that a and b are depending on p — when we want to stress this, then we'll write $a(p)$ and $b(p)$.

For $i \in I$, we simplify

$$(3.10) \qquad a_i = u_i + \lambda_-(p_i - u_i) = \frac{1}{n} + \frac{1}{1 - np_n}\left(p_i - \frac{1}{n}\right) = \frac{1 - np_n + n\left(p_i - \frac{1}{n}\right)}{n(1 - np_n)} = \frac{p_n - p_i}{np_n - 1}$$

and similarly

$$(3.11) \qquad b_i = \frac{p_i - p_1}{1 - np_1}.$$

Thus

$$(3.12) \qquad a = \frac{1}{np_n - 1}(p_n\mathbf{1} - p) \quad \text{and} \quad b = \frac{1}{1 - np_1}(p - p_1\mathbf{1}),$$

where $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$. Next,

$$(3.13\text{a}) \qquad \|a\|^2 = \frac{1}{(np_n - 1)^2}\|p_n\mathbf{1} - p\|^2 = \frac{1}{(np_n - 1)^2}\left(p_n^2\|\mathbf{1}\|^2 - 2p_n\langle\mathbf{1}, p\rangle + \|p\|^2\right)$$

$$(3.13\text{b}) \qquad = \frac{1}{(np_n - 1)^2}\left(np_n^2 - 2p_n + \|p\|^2\right)$$

$$(3.13\text{c}) \qquad = \frac{1}{(np_n - 1)^2}\left(p_1^2 + (p_2^2 + \cdots + p_{n-1}^2) + (n+1)p_n^2 - 2p_n\right)$$

and similarly

$$(3.14) \qquad \|b\|^2 = \frac{1}{(1 - np_1)^2}\left((n+1)p_1^2 - 2p_1 + (p_2^2 + \cdots + p_{n-1}^2) + p_n^2\right).$$

4 SETTING UP THE COSINE QUOTIENT

We uphold the assumptions and notation from the previous section. It is convenient to abbreviate

$$(4.1) \qquad z := (p_2, \ldots, p_{n-1}) \in \mathbb{R}^{n-2}.$$

Note that

$$(4.2) \qquad (\forall i \in \{1, 2, \ldots, n-2\}) \quad p_1 \le z_i \le p_n \quad \text{and} \quad \sum_{i=1}^{n-2} z_i = 1 - p_1 - p_n \ge 0$$

because of (3.1) and (3.2). This allows us to rewrite (3.13) and (3.14) more succinctly as

$$(4.3) \qquad \|a\|^2 = \frac{1}{(np_n - 1)^2}\left(p_1^2 + \|z\|^2 + (n+1)p_n^2 - 2p_n\right)$$

and

$$(4.4) \qquad \|b\|^2 = \frac{1}{(1 - np_1)^2}\left((n+1)p_1^2 - 2p_1 + \|z\|^2 + p_n^2\right).$$

Next, using (3.10) and (3.11), we have

$$(4.5\text{a}) \qquad \langle a, b\rangle = \sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n} \frac{p_n - p_i}{np_n - 1} \cdot \frac{p_i - p_1}{1 - np_1}$$

$$(4.5\text{b}) \qquad = \frac{1}{(np_n - 1)(1 - np_1)}\sum_{i=2}^{n-1}(p_n - p_i)(p_i - p_1)$$

$$(4.5\text{c}) \qquad = \frac{1}{(np_n - 1)(1 - np_1)}\sum_{i=1}^{n-2}(p_n - z_i)(z_i - p_1).$$

Using (4.5), (4.3), and (4.4), we now set up the quotient of interest from (1.4):

$$(4.6) \qquad \frac{\langle a, b\rangle}{\|a\|\|b\|} = \frac{\displaystyle\sum_{i=1}^{n-2}(p_n - z_i)(z_i - p_1)}{\sqrt{p_1^2 + \|z\|^2 + (n+1)p_n^2 - 2p_n}\sqrt{(n+1)p_1^2 - 2p_1 + \|z\|^2 + p_n^2}}.$$

5 MAXIMIZING THE COSINE QUOTIENT (WITH p_1 AND p_n FIXED)

We uphold the notation of the previous section. Now we turn toward maximizing the cosine quotient (4.6), with p_1 and p_n fixed. In view of (4.2), $z = (p_2, \ldots, p_{n-1})$ must belong to the compact convex set

$$(5.1) \qquad C := \left\{ w \in [p_1, p_n]^{n-2} \ \middle| \ \sum_{i=1}^{n-2} w_i = 1 - p_1 - p_n \right\}.$$

The continuity of (4.6) as a function of z coupled with the compactness of C guarantees the existence of a maximizer of the cosine quotient. Note that

$$(5.2) \qquad A(C) \subseteq C,$$

where $A \colon \mathbb{R}^{n-2} \to \mathbb{R}^{n-2} \colon w \mapsto (\eta, \eta, \ldots, \eta) \in \mathbb{R}^{n-2}$ and $\eta = \frac{1}{n-2}(w_1 + w_2 + \cdots + w_{n-2})$.

In general, maximizing a quotient is more involved; however, we will get a lucky break for our problem: It turns out we can maximize the numerator and minimize the denominator of (4.6) with respect to z and we luckily obtain the same optimal vector!

For convenience, set $\alpha := p_1, \beta := p_n, \gamma := (\alpha+\beta)/2$, and $\mathbf{1} := (1, 1, \ldots, 1) \in \mathbb{R}^{n-2}$. Then the numerator of (4.6) — as a function of z — is

$$(5.3a) \qquad \sum_{i=1}^{n-2}(\beta - z_i)(z_i - \alpha) = -\sum_{i=1}^{n-2} z_i^2 + (\alpha + \beta)\sum_{i=1}^{n-2} z_i + \text{constant}$$

$$(5.3b) \qquad = -\|z\|^2 + (\alpha + \beta)\langle \mathbf{1}, z \rangle + \text{constant}$$

$$(5.3c) \qquad = -\|z - \gamma\mathbf{1}\|^2 + \text{constant}.$$

We want to maximize the numerator over $z \in C$; equivalently, we want to minimize $\|z - \gamma\mathbf{1}\|^2$ over $z \in C$. By Corollary 2.2, we may restrict our attention to

$$(5.4) \qquad z \in C \cap \mathbb{R}\mathbf{1}.$$

On to the denominator of (4.6)! It is clear that the denominator becomes small when $\|z\|$ becomes small. So minimizing the denominator corresponds to minimizing $\|z\|^2 = \|z - 0\mathbf{1}\|^2$. Again by Corollary 2.2, we may restrict our attention to (5.4)! But if $z \in C \cap \mathbb{R}\mathbf{1}$, say $z = \zeta\mathbf{1}$ for $\zeta \in \mathbb{R}$, then the requirement that $\zeta\mathbf{1} \in C$ forces $\sum_{i=1}^{n-2} z_i = (n-2)\zeta = 1 - p_1 - p_n$, i.e., $\zeta = (1 - p_1 - p_n)/(n-2)$. Because $(n-1)p_1 + p_n \leq 1 \leq p_1 + (n-1)p_n$, it follows that $p_1 \leq \zeta \leq p_n$. Altogether,

$$(5.5) \qquad z = \frac{1 - p_1 - p_n}{n - 2}(1, 1, \ldots, 1) \in C \cap \mathbb{R}\mathbf{1} \subseteq \mathbb{R}^{n-2}$$

maximizes (4.6) and

$$(5.6) \qquad \|z\|^2 = \frac{(1 - p_1 - p_n)^2}{n - 2}.$$

To sum up, when restricted to the one-dimensional slice $C \cap \mathbb{R}\mathbf{1}$ (see (5.4)), we obtain the unique solution z given by (5.5). Our next step is to plug (5.5) and (5.6) back into (4.6). We start with the numerator of (4.6):

$$(5.7a) \qquad \langle a, b \rangle = \sum_{i=1}^{n-2}\left(p_n - \frac{1 - p_1 - p_n}{n - 2}\right)\left(\frac{1 - p_1 - p_n}{n - 2} - p_1\right)$$

$$(5.7b) \qquad = (n - 2)\left(\frac{(n-2)p_n - (1 - p_1 - p_n)}{n - 2}\right)\left(\frac{1 - p_1 - p_n - (n-2)p_1}{n - 2}\right)$$

$$(5.7c) \qquad = \frac{1}{n - 2}\left(p_1 + (n-1)p_n - 1\right)\left(1 - p_n - (n-1)p_1\right).$$

Using (5.6), we see that the square of the denominator of (4.6) is

$$(5.8a) \qquad \|a\|^2\|b\|^2 = \left(p_1^2 + \frac{(1-p_1-p_n)^2}{n-2} + (n+1)p_n^2 - 2p_n \right)$$

$$(5.8b) \qquad \times \left((n+1)p_1^2 - 2p_1 + \frac{(1-p_1-p_n)^2}{n-2} + p_n^2 \right)$$

$$(5.8c) \qquad = \frac{1}{(n-2)^2}$$

$$(5.8d) \qquad \times \left((n-2)p_1^2 + (1-p_1-p_n)^2 + (n-2)(n+1)p_n^2 - 2(n-2)p_n \right)$$

$$(5.8e) \qquad \times \left((n-2)(n+1)p_1^2 - 2(n-2)p_1 + (1-p_1-p_n)^2 + (n-2)p_n^2 \right).$$

Abbreviating $x = p_1$ and $y = p_n$, we use (5.7) and (5.8) to write the square of (4.6) as

$$(5.9) \qquad \frac{\langle a,b\rangle^2}{\|a\|^2\|b\|^2} = \frac{\dfrac{\left(x+(n-1)y-1\right)^2\left(1-y-(n-1)x\right)^2}{\left((n-2)x^2+(1-x-y)^2+(n-2)(n+1)y^2-2(n-2)y\right)}}{\left((n-2)(n+1)x^2 - 2(n-2)x + (1-x-y)^2 + (n-2)y^2\right)}.$$

In the next section, we will do the final maximization by setting p_1 and p_n, i.e., x and y, loose.

6 MAXIMIZING THE COSINE QUOTIENT (CONCLUDED)

While Section 5 looks like a huge mess, we can invoke one small but useful optimization: the objective function value of the cosine quotient is by construction constant on $[a, b]$. Now both a and b have at least one coordinate equal to 0, namely $a_n = 0$ and $b_1 = 0$ (see (3.12)). Because the objective function in (P) is constant on $\Delta \setminus \{u\}$ intersected with any line passing through u, we may and do finally assume that $a = p$ and thus $p_1 = x = 0$. Then Section 5 simplifies to

$$(6.1) \qquad \frac{\langle a,b\rangle^2}{\|a\|^2\|b\|^2} = \frac{\left((n-1)y-1\right)^2\left(1-y\right)^2}{\left((1-y)^2 + (n-2)(n+1)y^2 - 2(n-2)y\right)\left((1-y)^2 + (n-2)y^2\right)} =: Q(y).$$

The constraints on p are $0 \le p_1 \le p_2 = \cdots = p_{n-1} = (1 - p_1 - p_n)/(n-2) \le p_n \le 1$. With our assumption that $p_1 = 0 = x$ and $p_n = y$, these simplify to $0 \le (1-y)/(n-2) \le y \le 1$ and then to

$$(6.2) \qquad \frac{1}{n-1} \le y \le 1.$$

Because $n \ge 3 > 2$, we have

$$(6.3) \qquad \frac{1}{n-1} < \frac{2}{n} < 1.$$

So our remaining goal is to

$$(6.4) \qquad \text{maximize } Q(y) \text{ subject to } \frac{1}{n-1} \le y \le 1$$

where $Q(y)$ is defined in (6.1) Using the chain quotient rule to compute the derivative of $Q(y)$ followed by factoring yields

$$(6.5) \qquad Q'(y) = \frac{2(n-2)\left((n-1)^2y^2 - ny + 1\right)\left((n-1)y-1\right)(ny-2)(y-1)y}{\left((n^2-n-1)y^2 + 2(1-n)y + 1\right)^2\left((n-1)y^2 - 2y + 1\right)^2}.$$

Note that $Q'(y)$ has six roots, namely

$$(6.6) \qquad \left\{ \frac{n \pm \sqrt{-(3n-2)(n-2)}}{2(n-1)^2}, 0, \frac{1}{n-1}, \frac{2}{n}, 1 \right\}.$$

Because $n > 2$ implies $-(3n-2)(n-2) < 0$, we note that there are exactly four real roots. Our constraint (6.2) excludes 0. The remaining three roots include the endpoints of interval constraint. Thus, the maximum value of (6.4) is found by substituting for y the values $1/(n-1), 2/n, 1$ into $Q(y)$ which results in $0, (n-2)^2/(n+2)^2, 0$.

To sum up, $y = \frac{2}{n}$ is the solution of (6.4), with maximum value $(n-2)^2/(n+2)^2$. Using (5.5), we see this gives rise to the probability distribution

$$(6.7) \qquad p^* = \tfrac{1}{n}\big(0, 1, 1, \ldots, 1, 2\big).$$

We have shown that the angle spread is minimized for p^* given by (6.7). In view of (3.10) and (3.11), this gives rise to

$$(6.8) \qquad a^* := a(p^*) = \tfrac{1}{n}\big(2, 1, 1, \ldots, 1, 0\big) \ \text{ and } \ b^* := b(p^*) = \tfrac{1}{n}\big(0, 1, 1, \ldots, 1, 2\big) = p^*.$$

These vectors give the optimal value

$$(6.9) \qquad \cos\big(\measuredangle(a^*, b^*)\big) = \frac{\langle a^*, b^* \rangle}{\|a^*\|\|b^*\|} = \frac{n-2}{n+2}$$

which we claimed in the introduction. The proof of Theorem 1.1 is thus complete.

Remark 6.1. We point out that the expression for the optimal value in (6.9) was discovered numerically using `Julia` [10]. The computations in this section were verified with `SageMath` [14]. Finally, Figure 1 was created using `Geogebra` [7].

7 AN APPLICATION IN COGNITIVE SCIENCE

Our main result (Theorem 1.1) provides an important theoretical and methodological advancement in the study of probabilistic belief models in humans. Human learning is often modeled with Bayesian statistics, where beliefs are represented as probability distributions over possible outcomes [8, 9, 13, 15].

A MOTIVATING EXAMPLE

Consider the expected returns on a $100 investment over 1 year. A "bull" investor believes market prices will rise, while a "bear" investor believes market prices will fall. However, a bull still surely accepts prices *could* fall, despite holding this belief to a lesser degree than the bear. In Bayesian terms, the beliefs of our two investors can be represented as probability distributions over the domain of potential investment returns. The bull's "prior" (initial belief) will have most of its probability mass over positive returns, while the opposite is true for the bear's prior. These priors change over time in light of new evidence. While it is theoretically optimal to update according to the laws of probability, the empirical question remains as to how people actually represent and update probabilistic beliefs.

PREVIOUS WORK

Much of the experimental work in this field infers prior and posterior beliefs from sequential participant actions, or elicits them with insufficient detail [2, 6, 12]. This requires experimenters to make unwarranted assumptions about the way that probabilistic beliefs are parameterized as probability

distributions, and the nature of how they are updated with new data. It may be that a strict Bayesian model of human cognition is inappropriate when these assumptions are relaxed. Recent work by DiBerardino, Filipowicz, Danckert, and Anderson used a computerized version of the game "Plinko", where at each trial a ball falls through an array of pegs into one of $n = 40$ slots below. Participants were tasked with estimating the distribution of future ball drops by explicitly drawing a probability distribution that represents their beliefs after each trial [5]. This work found that initial beliefs (priors) vary across individuals and that these differences indicated future ball drop learning accuracy, thus demonstrating the importance of directly measuring individual beliefs in a theoretically agnostic manner [5].

However, [5] cannot make any additional claims about *why* participants with some priors appear to learn better than others. Its main limitation is that some participants' priors are more similar to the forthcoming ball drop distribution than others. As a result, it cannot be determined whether or not the observed differences in learning ability across different priors are due to participant "state" or "trait" differences. It may be that some participants are better probabilistic learners, which is a trait that can be detected by the shape of their prior beliefs. But it may also be that the participants who performed the best only did so because they just happened to have an initial state of belief more amenable to learning the forthcoming distribution.

ON-GOING WORK

Our main result (Theorem 1.1) allows for a new Plinko experiment that presents each participant with a ball drop distribution of a set level of similarity to any arbitrary participant's prior. Specifically, when measuring similarity between discrete probability distributions as the angular spread between representative Euclidean vectors, as was done in [5], we can determine the maximum possible rotation (dissimilarity) for any given vector (participant's prior) with respect to the uniform distribution. We require the following result.

Corollary 7.1. *For every $p \in \Delta_n \setminus \{u\}$ there exist q and v in $\Delta_n \cap (u + \mathbb{R}(p - u))$ such that*

$$(7.1) \qquad \sphericalangle(p, q) = \frac{1}{2} \arccos\left(\frac{n-2}{n+2}\right) = \sphericalangle(u, v).$$

Proof. Let $p \in \Delta_n \setminus \{u\}$, and set $\alpha_n := \arccos((n - 2)/(n + 2))$. We have seen earlier (see Section 3) that there exist two unique points $a = a(p)$ and $b = b(p)$ in Δ_n such that $\{u, p\} \subseteq [a, b]$ and $\|a - b\|$ is maximal. Consequently,

$$(7.2) \qquad \sphericalangle(a, p) + \sphericalangle(p, b) = \sphericalangle(a, b) = \sphericalangle(a, u) + \sphericalangle(u, b).$$

On the other hand, Theorem 1.1 yields $\sphericalangle(a, b) \geq \alpha_n$. Altogether,

$$(7.3) \qquad \sphericalangle(a, p) + \sphericalangle(p, b) = \sphericalangle(a, u) + \sphericalangle(u, b) \geq \alpha_n = \arccos\left(\frac{n-2}{n+2}\right).$$

Thus if $\sphericalangle(a, p) \geq \alpha_n/2$, then rotating p towards a yields q. If $\sphericalangle(a, p) < \alpha_n/2$, then $\sphericalangle(p, b) > \alpha_n/2$ and we rotate p towards b to obtain q. Similarly, if $\sphericalangle(a, u) \geq \alpha_n/2$, then rotating u towards a yields v. Finally, if $\sphericalangle(a, u) < \alpha_n/2$, then $\sphericalangle(u, b) > \alpha_n/2$ and we rotate u towards b to obtain v. $\qquad \square$

By Corollary 7.1, if a participant produces a prior $p \in \Delta_n \setminus \{u\}$, then the experimenter presents a ball drop distribution q where $\sphericalangle(p, q) = \alpha_n/2$. If a participant produces the uniform prior u, then the experimenter presents a ball drop distribution v where $\sphericalangle(u, v) = \alpha_n/2$ arising from an arbitrary $p \in \Delta_n \setminus \{u\}$ via Corollary 7.1.

More psychological research is required to determine the precise measure of similarity humans use to compare probability distributions. Angular similarity is only one such possibility. This result also poses

an interesting theoretical development: Does the shrinking minimal angle spread as the dimension tends to infinity create any difficulties for people to perceive or act upon any particular set of probability distributions? It may be that thinking about opposite/dissimilar probability distributions becomes more difficult as the number of discrete histogram bins approaches infinity if these mathematics correspond to mechanisms of human inference.

REFERENCES

[1] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, second edition, Springer, 2017.

[2] G. Charness, U. Gneezy, and V. Rasocha, Experimental methods: Eliciting beliefs, *Journal of Economic Behavior & Organization* 189 (2021), 234–256.

[3] M. Claus and K. Spürkel, Improving constants of strong convexity in linear stochastic programming, *Operations Research Letters* 50 (2022), 76–83.

[4] T. A. Cover and J. M. Thomas, *Elements of Information Geometry*, second edition, Wiley, 2006.

[5] P. A. V. DiBerardino, A. L. S. Filipowicz, J. Danckert, and B. Anderson, Plinko: Eliciting beliefs to build better models of statistical learning and mental model updating, https://arxiv.org/abs/2107.11477.

[6] P. H. Garthwaite, J. B. Kadane, and A. O'Hagan, Statistical methods for eliciting probability distributions, *Journal of the American Statistical Association* 100(470) (2005), 680–701.

[7] The GeoGebra Developers, *GeoGebra*, https://www.geogebra.org.

[8] C. M. Glaze, A. L. Filipowicz, J. W. Kable, V. Balasubramanian, and J. I. Gold, A bias–variance trade-off governs individual differences in on-line learning in an unpredictable environment, *Nature Human Behaviour* 2(3) (2018), 213–224.

[9] T. L. Griffiths and J. B. Tenenbaum, Optimal predictions in everyday cognition, *Psychological Science* 17(9) (2006), 767–773.

[10] The Julia Developers, *Julia*, https://www.julialang.org.

[11] K. Lange, *Optimization*, second edition, Springer, 2013.

[12] C. F. Manski, Measuring expectations, *Econometrica* 72(5) (2004), 1329–1376.

[13] M. R. Nassar, R. C. Wilson, B. Heasly, and J. I. Gold, An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment, *Journal of Neuroscience* 30(37) (2010), 12366–12378.

[14] The Sage Developers, *SageMath*, https://www.sagemath.org.

[15] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, How to grow a mind: Statistics, structure, and abstraction, *Science*, 331(6022) (2011), 1279–1285.

[16] Wikipedia, Table of polyhedron dihedral angles, https://en.wikipedia.org/wiki/Table_of_polyhedron_dihedral_angles, retrieved February 20, 2022.