Is the Volume of a Credal Set a Good Measure for Epistemic Uncertainty? (Supplementary Material)

Yusuf Sale[1,3] **Michele Caprio**[2] **Eyke Hüllermeier**[1,3]

[1]Institute of Informatics, University of Munich (LMU), Germany
[2]PRECISE Center, Department of Computer and Information Science, University of Pennsylvania, USA
[3]Munich Center for Machine Learning, Germany

A PROOFS

Proof of Proposition 1. Let $\mathcal{P}, \mathcal{Q} \subset \Delta(\mathcal{Y}, \sigma(\mathcal{Y}))$ be credal sets, and assume $|\mathcal{Y}| = 2$. Then we have the following.

- $\text{Vol}(\mathcal{P}) \geq 0$ and $\text{Vol}(\mathcal{P}) \leq \text{Vol}(\Delta^{2-1}) = \sqrt{2}$. Hence $\text{Vol}(\cdot)$ satisfies A1.
- The volume being a continuous functional is a well-known fact that comes from the continuity of the Lebesgue measure, so $\text{Vol}(\cdot)$ satisfies A2.
- $\mathcal{Q} \subset \mathcal{P} \implies \text{Vol}(\mathcal{Q}) \leq \text{Vol}(\mathcal{P})$. This comes from the fundamental property of the Lebesgue measure, so $\text{Vol}(\cdot)$ satisfies A3.
- Consider a sequence $(\mathcal{P}_n)$ of credal sets on $(\mathcal{Y}, \sigma(\mathcal{Y}))$ such that $\lim_{n\to\infty}[\overline{P}_n(A) - \underline{P}_n(A)] = 0$, for all $A \in \sigma(\mathcal{Y})$. Then, this means that there exists $N \in \mathbb{N}$ such that for all $n \geq N$, the geometric representation of $\mathcal{P}_n$ is a subset of the geometric representation of $\mathcal{P}_{n+1}$. In addition, the limiting element of $(\mathcal{P}_n)$ is a (multi)set $\mathcal{P}^\star$ whose elements are all equal to $P^\star$, so its geometric representation is a point and its volume is 0. Hence, probability consistency is implied by continuity A3, so $\text{Vol}(\cdot)$ satisfies A4'.
- The volume is invariant to rotation and translation. This is a well-known fact that comes from the fundamental property of the Lebesgue measure, so $\text{Vol}(\cdot)$ satisfies A7.

Let us now show that the volume operator satisfies sub-additivity A5. Let $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$. In addition, suppose we are in the general case in which $|\mathcal{Y}| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = 2$. In particular, let $\mathcal{Y} = \{(y_1, y_2), (y_3, y_4)\}$, so that $\mathcal{Y}_1 = \{y_1, y_3\}$ and $\mathcal{Y}_2 = \{y_2, y_4\}$. Suppose also $y_1 \neq y_3$ and $y_2 \neq y_4$. Now, pick any probability measure P on $\mathcal{Y}$. In general, we would have that its marginal $\text{marg}_{\mathcal{Y}_1}(P) = P'$ on $\mathcal{Y}_1$ is such that $P'(y_i) = \sum_j P((y_i, y_j))$. Similarly for marginal $\text{marg}_{\mathcal{Y}_2}(P) = P''$ on $\mathcal{Y}_2$. In our case, though, the computation is easier. To see this, fix y_1. Then, we should sum over j the probability of (y_1, y_j), $y_j \in \mathcal{Y}_2$. But the only pair (y_1, y_j) is (y_1, y_2). A similar argument holds if we fix y_3, or any of the elements of $\mathcal{Y}_2$. Hence, we have that

$$P'(y_1) = P((y_1, y_2)) = P''(y_2) \quad \text{and} \quad P'(y_3) = P((y_3, y_4)) = P''(y_4).$$

Let $\mathcal{P}'$ and $\mathcal{P}''$ denote the marginal convex sets of probability distributions on $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively, and let $\mathcal{P}$ denote the convex set of joint probability distributions on $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$ [Couso et al., 1999]. Then, given our argument above, we have that $\text{Vol}(\mathcal{P}) < \text{Vol}(\mathcal{P}') + \text{Vol}(\mathcal{P}'') = 2\text{Vol}(\mathcal{P})$. So in the general $|\mathcal{Y}| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = 2$ case where $y_1 \neq y_3$ and $y_2 \neq y_4$, the volume is subadditive. □

Proof of Proposition 2. Immediate from the assumption on the instance of SI. □

Proof of Theorem 1. Pick any compact set $\mathcal{P} \subset \mathcal{M}(\Omega, \mathcal{F})$ and any set $\mathcal{Q}$ satisfying (a)-(c). Let $B_r^d \subset \mathbb{R}^d$ denote a generic ball in $\mathbb{R}^d$ of radius $r > 0$. Notice that $N_{r-\epsilon}^{\text{pack}}(\mathcal{Q}') = N_{r-\epsilon}^{\text{pack}}(\mathcal{P}) - N_{r-\epsilon}^{\text{pack}}(\mathcal{Q}) \geq 0$ because $\mathcal{P} \supset \mathcal{Q}$. Then, the proof goes as

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

follows

$$
\begin{aligned}
\frac{\mathrm{Vol}(\mathcal{P})-\mathrm{Vol}(\mathcal{Q}')}{\mathrm{Vol}(\mathcal{P})} &= \frac{\frac{1}{c(r,d,\mathcal{P})}\mathrm{Vol}(\tilde{\mathcal{P}}_r)-\frac{1}{c(r-\epsilon,d,\mathcal{Q}')}\mathrm{Vol}(\tilde{\mathcal{Q}}'_{r-\epsilon})}{\frac{1}{c(r,d,\mathcal{P})}\mathrm{Vol}(\tilde{\mathcal{P}}_r)} && (1)\\
&\geq \frac{\mathrm{Vol}(\tilde{\mathcal{P}}_r)-\mathrm{Vol}(\tilde{\mathcal{Q}}'_{r-\epsilon})}{\mathrm{Vol}(\tilde{\mathcal{P}}_r)} && (2)\\
&= \frac{N_r^{\text{pack}}(\mathcal{P})\mathrm{Vol}(B_r^d)-N_{r-\epsilon}^{\text{pack}}(\mathcal{Q}')\mathrm{Vol}(B_{r-\epsilon}^d)}{N_r^{\text{pack}}(\mathcal{P})\mathrm{Vol}(B_r^d)} && (3)\\
&= \frac{N_r^{\text{pack}}(\mathcal{P})\mathrm{Vol}(B_1^d)r^d-N_{r-\epsilon}^{\text{pack}}(\mathcal{Q}')\mathrm{Vol}(B_1^d)(r-\epsilon)^d}{N_r^{\text{pack}}(\mathcal{P})\mathrm{Vol}(B_1^d)r^d} && (4)\\
&= \frac{N_r^{\text{pack}}(\mathcal{P})r^d-N_{r-\epsilon}^{\text{pack}}(\mathcal{Q}')(r-\epsilon)^d}{N_r^{\text{pack}}(\mathcal{P})r^d}\\
&= 1-\frac{N_{r-\epsilon}^{\text{pack}}(\mathcal{Q}')}{N_r^{\text{pack}}(\mathcal{P})}\left(1-\frac{\epsilon}{r}\right)^d\\
&\geq 1-\left(1-\frac{\epsilon}{r}\right)^d, && (5)
\end{aligned}
$$

where (1) comes from equation (3), (2) comes from the fact that $r-\epsilon \leq r \implies c(r-\epsilon,d,\mathcal{Q}') \geq c(r,d,\mathcal{P})$ by (4), (3) comes from $\tilde{\mathcal{P}}_r$ being the union of pairwise disjoint balls of radius r, (4) comes from properties of the volume of a ball of radius r in $\mathbb{R}^d$, and (5) comes from property (c) of $\mathcal{Q}$. □

B HIGH-DIMENSIONAL PROBABILITY

Since Theorem 1 in Section 4.2 is intimately related with Carl-Pajor's Theorem [Ball and Pajor, 1990], we state (a version) of the theorem here.

Theorem 1 (Carl-Pajor)**.** *Let $B_{1,d}$ denote the d-dimensional unit euclidean ball, and let $\mathcal{P} \subset B_{1,d}$ be a polytope with $m \in \mathbb{N}$ vertices. Then, we have*

$$
\frac{Vol(\mathcal{P})}{Vol(B_{1,d})} \leq \left(4\sqrt{\frac{\log m}{d}}\right)^d. \tag{6}
$$

For further results connecting high-dimensional probability and data science, see Vershynin [2018].