THE UNIQUE ALL-TIED DISTRIBUTION OF THE MULTI-LABEL JACCARD LOSS, AND THE EXHAUSTION OF THE FEASIBLE-SUBSPACE LOWER BOUND

KOREA SUPERINTELLIGENCE LABS

Abstract. For a multi-label problem with s labels the instance-wise Jaccard loss is the $2^s \times 2^s$ matrix $L^{\mathrm{Jac}} = U - S$ whose score entries are $S_{A,B} = \mathrm{Jac}(A,B) = |A \cap B|/|A \cup B|$, with $\mathrm{Jac}(\emptyset, \emptyset) = 1$. Zhang, and independently Dewasurendra, have recently proved $\operatorname{rank} S = \operatorname{rank} L^{\mathrm{Jac}} = 2^s$, $\operatorname{affdim} L^{\mathrm{Jac}} = 2^s - 1$ and $2^{s-1} \le \mathrm{CCdim}(L^{\mathrm{Jac}}) \le 2^s - 1$, both by the feasible-subspace lower bound of Ramaswamy and Agarwal, and both leave the remaining factor of two open. We show that the gap cannot be closed from below by that method. First, we determine every distribution at which all 2^s Jaccard reports are Bayes-optimal: there is exactly one, the uniform distribution $p^\star$ on the $s+1$ outcomes $\emptyset, \{1\}, \ldots, \{s\}$, for every s. Since its support holds only $s+1$ of the 2^s outcomes, the other Ramaswamy–Agarwal lower bound — the one that would give $2^s - 2$ — is unavailable at every $s \ge 2$, and the best value their feasible-subspace theorem can produce is at most $2^s - 2$. Second, an exhaustive exact search computes that best value outright at the three smallest label counts: it is $2, 4, 8$ at $s = 2, 3, 4$, that is, exactly the published lower bound 2^{s-1}, and exactly $2^{s-1} - 1$ at $s = 2, 3$ under the alternative convention $\mathrm{Jac}(\emptyset, \emptyset) = 0$. So at those label counts the surviving factor of two is a property of the method and not of the witness: the published witness is optimal for it. This is the opposite of what happens for the sibling F_1 loss. We also re-verify the rank formula at $s \le 4$, its one-entry perturbation at $s = 2, 3$, and the published witness at $s = 2, 3$, from explicit integer certificates. All the finite content is machine-checked in Lean 4; Section 7 says exactly what is and what is not formalised.

1. Introduction

The objects. Fix $s \ge 1$ labels. In multi-label classification an outcome and a report are both subsets of $[s] = \{1, \ldots, s\}$, so both range over the set $\mathcal{Y} = 2^{[s]}$ of size $N = 2^s$. The instance-wise Jaccard score — intersection over union — of the report B against the outcome A is

$$\mathrm{Jac}(A,B) \;=\; \begin{cases} \dfrac{|A \cap B|}{|A \cup B|}, & A \cup B \neq \emptyset, \\ 1, & A = B = \emptyset, \end{cases} \tag{1}$$

and the Jaccard loss is $\ell(A,B) = 1 - \mathrm{Jac}(A,B)$. Collecting these gives the $2^s \times 2^s$ score matrix S with $S_{A,B} = \mathrm{Jac}(A,B)$, the all-ones matrix $U = \mathbf{1}\mathbf{1}^{\mathsf{T}}$ and the Jaccard loss matrix

$$L \;=\; L^{\mathrm{Jac}} \;=\; U - S, \qquad \text{so} \qquad L - U = -S.$$

Following Ramaswamy and Agarwal [3], the *column-affine dimension* of a matrix M is

$$\operatorname{affdim} M \;=\; \operatorname{rank} \begin{bmatrix} M \\ \mathbf{1}^{\mathsf{T}} \end{bmatrix} - 1,$$

the dimension of the affine hull of its columns. Their *convex calibration dimension* $\mathrm{CCdim}(L)$ [3, Definition 10] is the least d for which there are a convex set $\mathcal{C} \subseteq \mathbb{R}^d$, a surrogate $\psi : \mathcal{C} \to \mathbb{R}_+^N$ with convex coordinates and a link from $\mathcal{C}$ to the reports that is L-calibrated. It is the number of real-valued functions a consistent convex surrogate method must learn, and it is the quantity this paper is about.

The sources and what they leave open. Zhang [1] determines the rank of the Jaccard matrices exactly. Its Theorem 4.2 states that for every $s \geq 1$,

$$\operatorname{rank} S \;=\; \operatorname{rank}(L-U) \;=\; \operatorname{rank} L \;=\; 2^s, \qquad \operatorname{affdim} L \;=\; 2^s - 1, \tag{2}$$

with a complete proof in its appendix: its Lemma 4.1 writes the nonempty block $K = S_{\mathcal{Y}_+,\mathcal{Y}_+}$ as a MinHash Gram matrix and uses Boolean Möbius inversion to show that it is positive definite. With [3, Theorem 12], which bounds $\mathrm{CCdim}(L)$ by $\operatorname{affdim} L$, this gives its Corollary 4.3,

$$\mathrm{CCdim}(L^{\mathrm{Jac}}) \;\leq\; 2^s - 1. \tag{3}$$

Its Theorem 5.2 is the matching exponential lower bound: for every $s \geq 1$,

$$2^{s-1} \;\leq\; \mathrm{CCdim}(L^{\mathrm{Jac}}) \;\leq\; 2^s - 1, \qquad \text{hence} \qquad \mathrm{CCdim}(L^{\mathrm{Jac}}) = \Theta(2^s), \tag{4}$$

proved by applying the feasible-subspace theorem [3, Theorem 16] at a factorially weighted witness recalled in Section 5. Its Remark 5.3 then opens, verbatim: *"The bounds in (7) differ by less than a factor of two. Determining the exact convex calibration dimension, or improving either constant, remains open."* Its equation (7) is (4) above. A concurrent and independent manuscript of Dewasurendra [2] proves the same two statements — its Proposition 2 gives $\operatorname{rank} S = \operatorname{rank} L = 2^s$ and $\operatorname{affdim} L = 2^s - 1$, and its Theorem 5 is (4) — by the same route through [3, Theorem 16] with the same factorial witness, and leaves the same gap. Numbered results of [1], [2] and [3] always carry their citation below; an unqualified number is this paper's own.

Both sources also record what happens under the competing convention $\mathrm{Jac}(\emptyset, \emptyset) = 0$; write $S^{(0)}$ and $L^{(0)} = U - S^{(0)}$ for the matrices it produces. Then [1, Remark 4.4] states

$$\operatorname{rank} S^{(0)} \;=\; \operatorname{rank}(L^{(0)} - U) \;=\; 2^s - 1, \qquad \operatorname{rank} L^{(0)} \;=\; 2^s, \qquad \operatorname{affdim} L^{(0)} \;=\; 2^s - 1, \tag{5}$$

and [1, Remark 5.4], like [2, Appendix G], gives $2^{s-1} - 1 \leq \mathrm{CCdim}(L^{(0)}) \leq 2^s - 1$.

What this paper settles. The lower halves of (4) and of its convention variant are both instances of a single tool, the feasible-subspace bound [3, Theorem 16]: for a probability vector p and a Bayes-optimal report B_0 at p,

$$\mathrm{CCdim}(L) \;\geq\; \|p\|_0 - \mu_{Q^L_{B_0}}(p) - 1, \tag{6}$$

where μ is the dimension of the feasible subspace of the trigger set $Q^L_{B_0}$ at p. Section 2 recalls that the right-hand side of (6) depends on p only through the pair $(\operatorname{supp} p, \operatorname{opt}_L(p))$, so the best bound the theorem can give for a fixed L is a finite quantity, which we call $\mathrm{fsd}(L)$. Everything below is about that quantity for the Jaccard loss. We do not decide $\mathrm{CCdim}(L^{\mathrm{Jac}})$; we decide the method.

- **There is exactly one all-tied distribution, for every** s (Theorem 4.2). The uniform distribution $p^\star$ on the $s+1$ outcomes $\emptyset, \{1\}, \ldots, \{s\}$ makes all 2^s reports Bayes-optimal at the common risk $s/(s+1)$, and it is the only probability vector on $\mathcal{Y}$ that does. Its support has $s+1$ elements, so for $s \geq 2$ it is not of full support.
- **Hence the other Ramaswamy–Agarwal lower bound is unavailable at every** $s \geq 2$ (Corollary 4.3). Their Theorem 18 concludes $\mathrm{CCdim}(L) \geq \operatorname{affdim} L - 1$ — here $2^s - 2$, which would nearly close (4) — from a distribution of *full* support at which every report ties. No such distribution exists for the Jaccard loss when $s \geq 2$. In particular the uniform distribution on $\mathcal{Y}$ does not tie the reports, which is where Jaccard parts company with the 0-1 loss of [3, Example 11].
- **Hence also** $\mathrm{fsd}(L^{\mathrm{Jac}}_s) \leq 2^s - 2$ **for every** $s \geq 2$ (Corollary 4.4): the route through (6) can never certify the upper bound (3), whatever witness is used.
- **At the three smallest label counts the exact value of** fsd **is the published lower bound** (Section 6). An exhaustive exact-rational search gives $\mathrm{fsd}(L^{\mathrm{Jac}}_2) = 2$, $\mathrm{fsd}(L^{\mathrm{Jac}}_3) = 4$ and $\mathrm{fsd}(L^{\mathrm{Jac}}_4) = 8$, that is exactly 2^{s-1}; and $\mathrm{fsd}(L^{(0)}_2) = 1$, $\mathrm{fsd}(L^{(0)}_3) = 3$, exactly $2^{s-1} - 1$. Each value is attained by the source's own witness — that of [1, Theorem 5.2], and its [1, Remark 5.4] variant under the alternative convention. So at $s = 2, 3, 4$ the surviving factor

of two in (4) is a property of the method, not a weakness of the witness, and improving the lower constant needs an idea outside [3, Theorem 16].

- **The published facts are re-verified from explicit certificates** (Sections 3 and 5): (2) at $s = 1, 2, 3, 4$, (5) at $s = 2, 3$, the witness of [1, Theorem 5.2] at $s = 2, 3$ with its exact Bayes set and $\mu = 0$, and [3, Example 11] at $n = 4$ as a literature control. These are not new mathematics; they are the machine-checked base the new statements sit on, and they are stated here because Section 7 has to say precisely what was checked.

The contrast with F_1. The same programme was carried out for the instance-wise F_1 loss, whose calibration dimension is quadratic rather than exponential [4, 5], in the companion manuscript [6]. There the exhaustive search *beat* the source's own witness: at three labels $\mathrm{fsd}(L_3^{F_1}) = 3$ while [4, Theorem 5.1] evaluates to 2, so the published lower bound was not the ceiling of the method and could be improved inside it. Here the outcome is the opposite one, and it is worth stating plainly: for Jaccard at $s = 2, 3, 4$ nothing is left to gain. The published constant 2^{s-1} is already the largest number [3, Theorem 16] can return, so the contribution of this paper on the lower-bound side is a negative result about the method together with the exact values that prove it, and not an improvement of (4).

2. Preliminaries

Write $\Delta_N = \{p \in \mathbb{R}_+^N : \mathbf{1}^\mathsf{T} p = 1\}$ for the probability simplex on $\mathcal{Y}$, $\operatorname{supp} p = \{A : p_A > 0\}$ and $\|p\|_0 = |\operatorname{supp} p|$. The conditional risk of the report B under p is $p^\mathsf{T} L_{\cdot,B}$, the Bayes-optimal set is $\mathrm{opt}_L(p) = \arg\min_{B \in \mathcal{Y}} p^\mathsf{T} L_{\cdot,B}$, and the trigger set of a report B is $Q_B^L = \{p \in \Delta_N : p^\mathsf{T} L_{\cdot,B} \le p^\mathsf{T} L_{\cdot,B'} \ \forall B' \in \mathcal{Y}\}$. For $p \in Q$, $\mu_Q(p)$ is the dimension of the largest linear subspace of feasible directions of Q at p [3, Definition 14], and [3, Lemma 15] computes it as the nullity of the matrix of the constraints active at p.

We say that $p \in \Delta_N$ is *all-tied* for L when $\mathrm{opt}_L(p) = \mathcal{Y}$, that is when every report has the same conditional risk. Two lower bounds of [3] are in play. The feasible-subspace bound (6) is [3, Theorem 16]. The other is [3, Theorem 18], quoted verbatim: *"Let $L \in \mathbb{R}_+^{n \times k}$. If $\exists p \in \mathrm{relint}(\Delta_n)$, $c \in \mathbb{R}_+$ such that $p^\mathsf{T} \ell_t = c \ \forall t \in [k]$, then* $\mathrm{CCdim}(L) \ge \mathrm{affdim}(L) - 1$.*"* Its hypothesis is precisely an all-tied distribution of full support; [3, Corollary 19] supplies one — the uniform distribution — whenever the columns of L are permutations of one another. The 0-1 loss is the model case: its columns are permutations of one another, the uniform distribution ties every report, and [3, Example 11] evaluates (6) at it directly ($\mu = 0$ by their Lemma 15 and $\|p\|_0 = n$) to get $\mathrm{CCdim}(L_n^{0\text{-}1}) \ge n - 1$, hence with their Lemma 11 the exact value $\mathrm{CCdim}(L_n^{0\text{-}1}) = n - 1$ — the one exact value that circle of ideas computes by this route, and the configuration [3, Theorem 18] later packages in general.

Writing out [3, Lemma 15] at a point p of a trigger set makes the right-hand side of (6) explicit. At $p \in Q_{B_0}^L$ the active constraints are the nonnegativities off $\operatorname{supp} p$, the ties $B \in \mathrm{opt}_L(p)$ and the normalisation, so with $T = \operatorname{supp} p$,

$$\mu_{Q_{B_0}^L}(p) \ = \ |T| - \mathrm{rank}\big(\{\mathbf{1}_T\} \cup \{(L_{\cdot,B} - L_{\cdot,B_0})|_T : B \in \mathrm{opt}_L(p)\}\big).$$

Since p_T is orthogonal to every active difference while $p_T^\mathsf{T} \mathbf{1}_T = 1$, that rank is one more than the dimension of the span of the differences, and therefore

Proposition 2.1. *For $p \in \Delta_N$ and $B_0 \in \mathrm{opt}_L(p)$, the right-hand side of* (6) *equals*

$$\mathrm{val}(p) \ = \ \mathrm{affdim}\{\, L_{\cdot,B}|_{\operatorname{supp} p} \ : \ B \in \mathrm{opt}_L(p) \,\},$$

which depends on p only through the pair $(\operatorname{supp} p, \mathrm{opt}_L(p))$*. Moreover* $\mathrm{val}(p) \le \|p\|_0 - 1$ *and* $\mathrm{val}(p) \le |\,\mathrm{opt}_L(p)| - 1$.

Proof. The displayed identity is the computation above. A family of m points has affine dimension at most $m - 1$, which gives $\mathrm{val}(p) \le |\,\mathrm{opt}_L(p)| - 1$; and $\mathrm{val}(p) = \|p\|_0 - \mu - 1 \le \|p\|_0 - 1$ because $\mu \ge 0$. □

Proposition 2.1 is a reformulation of [3, Lemma 15] and is not claimed as new; it is stated because it turns "the best bound (6) can give" into a finite maximum.

Definition 2.2. The *feasible-subspace value* of a loss matrix L is $\operatorname{fsd}(L) = \max_{p\in\Delta_N} \operatorname{val}(p)$, the largest lower bound on $\operatorname{CCdim}(L)$ obtainable from [3, Theorem 16].

By Proposition 2.1 the maximum ranges over the finitely many pairs $(\operatorname{supp} p, \operatorname{opt}_L(p))$ that are realised by some p, so $\operatorname{fsd}(L)$ is computable; and $\operatorname{CCdim}(L) \geq \operatorname{fsd}(L)$ always, since every $\operatorname{val}(p)$ is a lower bound on $\operatorname{CCdim}(L)$. For the 0-1 loss, [3, Example 11] both proves $\operatorname{CCdim}(L_n^{\text{0-1}}) = n-1$ and attains $n-1$ by this route, so $\operatorname{fsd}(L_n^{\text{0-1}}) = n-1$ as well: there the method is exact.

3. The rank of the Jaccard matrices, from integer certificates

The rank statement (2) is proved for every s in [1, Theorem 4.2] and again in [2, Proposition 2]. What is added here is a finite certificate at each small s, checked by a proof kernel rather than by floating-point or symbolic software; it is the base on which the witness computations of Section 5 rest, and it fixes the matrices themselves — entry by entry, over $\mathbb{Q}$, from (1) — as the objects everything else refers to.

The device is elementary. Every denominator $|A \cup B|$ in (1) is at most s, so $\Lambda = \operatorname{lcm}(1, \dots, s)$ clears all of them at once and ΛS is an integer matrix ($\Lambda = 1, 2, 6, 12$ for $s = 1, 2, 3, 4$, and the entries of ΛS never exceed Λ). Rank is invariant under the nonzero scaling, so it suffices to bound the rank of an integer matrix M from below, and for that we use a reduction read modulo a prime: if $XMY = I + pW$ over $\mathbb{Z}$ for integer matrices X, Y, W of size r, then $\det(XMY) \equiv 1 \pmod p$, so $\det(XMY) \neq 0$ and $\operatorname{rank} M \geq r$. Here $p = 1000003$, and the certificates are small: across all the certificates of this section every entry of Y is a residue below p and every entry of W lies between 0 and 96, where the adjugate route would need determinant-sized integers. When r is the width of M, the width already caps the rank from above, so a single certificate settles it; that is the case for all the matrices of Theorem 3.1, which are nonsingular. The one rank-deficient matrix below, in Theorem 3.2, needs an explicit factorisation for the upper bound as well.

Theorem 3.1 ([1, Theorem 4.2], certified at $s \leq 4$)**.** *For $s = 1, 2, 3, 4$,*

$$\operatorname{rank} S \;=\; \operatorname{rank} L \;=\; \operatorname{rank}(L-U) \;=\; 2^s \qquad \textit{and} \qquad \operatorname{affdim} L \;=\; 2^s - 1.$$

Moreover $\operatorname{rank}(L-U) = \operatorname{rank} S$ *for* every s*, because* $L - U = -S$*.*

Proof sketch. The last sentence is the identity $L - U = -S$, recorded in [1] just after its equation (2), and holds for all s with no computation. For the rest: for each $s \leq 4$ one exhibits explicit integer matrices Y, W with $(\Lambda S)\, Y = I + 1000003\, W$, which forces $\operatorname{rank} S \geq 2^s$, and the width gives the reverse inequality; the same for ΛL. For the affine dimension, deleting the appended all-ones row is a left multiplication, so $\operatorname{rank} L \leq \operatorname{rank}\big[\begin{smallmatrix} L \\ \mathbf{1}^\top \end{smallmatrix}\big] \leq \operatorname{rank} L + 1$, and $\operatorname{rank} L = 2^s = N$ already fills the width, whence $\operatorname{affdim} L = N - 1$ with no second certificate. Every step is a finite identity between explicit rational matrices and is machine-checked; no numerical evaluation enters. $\square$

The one-entry perturbation of (5) is a useful control on that: it shows that the rank statement really is about the matrix defined by (1) and is not convention-free.

Theorem 3.2 ([1, Remark 4.4], certified at $s = 2, 3$)**.** *For $s = 2$ and $s = 3$, changing the single entry* $\operatorname{Jac}(\emptyset, \emptyset)$ *from* 1 *to* 0 *gives*

$$\operatorname{rank} S^{(0)} \;=\; 2^s - 1 \;\neq\; \operatorname{rank} S, \qquad \operatorname{rank} L^{(0)} \;=\; 2^s, \qquad \operatorname{affdim} L^{(0)} \;=\; 2^s - 1.$$

Proof sketch. $\operatorname{rank} L^{(0)}$ and $\operatorname{affdim} L^{(0)}$ are certified exactly as in Theorem 3.1. For $\operatorname{rank} S^{(0)} = 2^s - 1$ both directions are needed: an explicit factorisation $\Lambda S^{(0)} = AB$ through $\mathbb{Z}^{2^s-1}$ caps the rank, and a mod-p reduction of size $2^s - 1$ floors it. The inequality $\operatorname{rank} S^{(0)} \neq \operatorname{rank} S$ is then read off against Theorem 3.1. $\square$

4. The unique all-tied distribution

Let $p^\star \in \Delta_N$ be the uniform distribution on the $s+1$ outcomes of size at most one:

$$p^\star_A \;=\; \begin{cases} \dfrac{1}{s+1}, & |A| \le 1,\\ 0, & |A| \ge 2. \end{cases}$$

The proof of [1, Theorem 4.2] uses the identity

$$\sum_{j=1}^{s} \mathrm{Jac}(A,\{j\}) \;=\; 1 \qquad \text{for every nonempty } A \subseteq [s], \tag{7}$$

which holds because the j-th term is $1/|A|$ when $j \in A$ and 0 otherwise (the rank proof of [2, Proposition 2] goes through the matrix determinant lemma instead and does not state it). One step from (7) gives the following; neither source states it, and it is recorded here as the configuration that [3, Theorem 18] asks for, not as a new result.

Lemma 4.1. *For every $s \ge 1$ and every report $B \in \mathcal{Y}$, the Jaccard score of B at $p^\star$ is $1/(s+1)$, so every one of the 2^s reports has conditional risk $s/(s+1)$ and $\mathrm{opt}_L(p^\star) = \mathcal{Y}$. In particular $p^\star$ is all-tied.*

Proof. The score of B at $p^\star$ is $\big(\mathrm{Jac}(\emptyset,B) + \sum_{j=1}^{s} \mathrm{Jac}(\{j\},B)\big)/(s+1)$. If $B=\emptyset$ the bracket is $1+0$, since (1) gives $\mathrm{Jac}(\{j\},\emptyset)=0$. If $B \ne \emptyset$ then $\mathrm{Jac}(\emptyset,B)=0$, and Jac is symmetric, so the sum is $\sum_j \mathrm{Jac}(B,\{j\}) = 1$ by (7) applied to the nonempty set B; the bracket is $0+1$. Either way the score is $1/(s+1)$ and the risk is $1-1/(s+1) = s/(s+1)$. ☐

The converse is the first new statement of the paper, and it is what makes the negative results possible. It holds for every s, with no computation at any particular s.

Theorem 4.2. *Let $s \ge 1$ and let $p \in \Delta_N$ be all-tied for the Jaccard loss, that is $p \ge 0$, $\sum_A p_A = 1$, and all 2^s reports have the same conditional risk. Then $p = p^\star$.*

Proof. Write $f(B) = \sum_A p_A \,\mathrm{Jac}(A,B)$ for the conditional score of B. Since $\sum_A p_A(1-\mathrm{Jac}(A,B)) = 1 - f(B)$, the reports all tie in risk exactly when they all tie in score; let c be that common score. Three probe reports suffice.

(i) $B=\emptyset$. By (1) the empty report scores 0 against every nonempty outcome and 1 against $\emptyset$, so $c = f(\emptyset) = p_\emptyset$.

(ii) the s singleton reports. Summing and exchanging the order of summation, $\sum_j f(\{j\}) = \sum_A p_A \sum_j \mathrm{Jac}(A,\{j\}) = 1 - p_\emptyset$ by (7), the empty outcome contributing 0. The left side is $s\,c$, so $s\,c = 1-c$ and

$$c \;=\; p_\emptyset \;=\; \frac{1}{s+1}.$$

(iii) $B=[s]$. Here $\mathrm{Jac}(A,[s]) = |A|/s$, so $f([s]) = \frac{1}{s}\sum_A p_A|A| = c$, that is $\sum_A p_A|A| = s\,c = 1-p_\emptyset = \sum_{A\ne\emptyset} p_A$. Every nonempty outcome has $|A| \ge 1$, so this is the equality case of a term-by-term inequality: it forces $p_A = 0$ whenever $|A| \ge 2$.

The support therefore lies in $\{\emptyset\} \cup \{\{1\},\dots,\{s\}\}$, and on that set $f(\{j\}) = p_{\{j\}}$ because a singleton report scores 1 against itself and 0 against the other outcomes of the support. Hence $p_{\{j\}} = c = 1/(s+1)$ for every j, and with (i) this is $p^\star$. ☐

Corollary 4.3. *For every $s \ge 2$ there is no all-tied distribution of full support for the Jaccard loss: any all-tied p gives the outcome $[s]$ mass 0. In particular the uniform distribution on all 2^s outcomes is not all-tied, so the columns of L^{Jac} are not permutations of one another. Consequently, at every $s \ge 2$, the hypothesis of* [3, Theorem 18] *fails for L^{Jac}, and so does the (different) hypothesis of* [3, Corollary 19]*; the bound $\mathrm{CCdim}(L^{\mathrm{Jac}}) \ge \operatorname{affdim} L^{\mathrm{Jac}} - 1 = 2^s - 2$, with $\operatorname{affdim} L^{\mathrm{Jac}} = 2^s - 1$ from* [1, Theorem 4.2]*, is not available by that route.*

Proof. By Theorem 4.2 an all-tied p equals $p^\star$, and $p^\star_{[s]} = 0$ because $|[s]| = s \geq 2$. The uniform distribution on $\mathcal{Y}$ gives $[s]$ positive mass, so it is not all-tied; and if the columns of L^{Jac} were permutations of one another, the uniform distribution would tie all reports. Theorem 18 of [3] asks exactly for a full-support all-tied point, which Theorem 4.2 rules out; its Corollary 19 asks for permuted columns, which the previous sentence rules out. □

Corollary 4.3 is exactly the point at which the Jaccard loss separates from the 0-1 loss: the columns of $L^{0\text{-}1}_n$ are permutations of one another, the uniform distribution ties every report, and [3, Example 11] gets the exact value out of (6) evaluated at that full-support all-tied point — the configuration [3, Theorem 18] describes in general. Neither [1] nor [2] discusses this route, an all-tied distribution, or how large (6) can be for their loss. (The full-support witness of [2, Lemma 4] is not a counterexample to Theorem 4.2: it interiorises the factorial witness to full support, but only $2^{s-1} + 1$ of the 2^s reports are Bayes-optimal there, not all of them.)

The same two facts also cap the other route, and this is the sharper consequence.

Corollary 4.4. *For every $s \geq 2$, $\mathrm{fsd}(L^{\mathrm{Jac}}_s) \leq 2^s - 2$. The feasible-subspace bound* (6) *can therefore never certify the upper bound $\mathrm{CCdim}(L^{\mathrm{Jac}}) \leq 2^s - 1$ of* (3)*, at any $s \geq 2$ and from any witness.*

Proof. Let $p \in \Delta_N$. By Proposition 2.1, $\mathrm{val}(p) \leq |\operatorname{opt}_L(p)| - 1$, so $\mathrm{val}(p) \geq 2^s - 1$ would force $|\operatorname{opt}_L(p)| \geq 2^s$, that is $\operatorname{opt}_L(p) = \mathcal{Y}$ and p all-tied, hence $p = p^\star$ by Theorem 4.2. But $\|p^\star\|_0 = s + 1$, so $\mathrm{val}(p^\star) \leq s \leq 2^s - 2$ for $s \geq 2$, again by Proposition 2.1. Every other p has $|\operatorname{opt}_L(p)| \leq 2^s - 1$ and hence $\mathrm{val}(p) \leq 2^s - 2$. □

Remark 4.5. The bound $\mathrm{val}(p^\star) \leq s$ of the proof is attained: an exact computation of $\operatorname{affdim}\{L_{\cdot,B}|_{\operatorname{supp} p^\star} : B \in \mathcal{Y}\}$ gives $\mathrm{val}(p^\star) = s$ for $s = 1, \ldots, 7$. So the configuration [3, Theorem 18] asks for is, for the Jaccard loss, not merely unavailable but exponentially worse than the factorial witness of [1, Theorem 5.2], which reaches 2^{s-1}. This is a computation outside the formal development (Section 7).

5. The published witness, and the 0-1 control

The lower half of (4) is obtained in [1, Theorem 5.2] as follows. Fix the core label 1, put $O = [s] \setminus \{1\}$ and $\mathcal{U} = \{\{1\} \cup D : D \subseteq O\}$, a family of 2^{s-1} outcomes. Let q give $\{1\} \cup D$ mass proportional to $1/|D|!$; the factorial balancing identity [1, Lemma 5.1] makes all 2^{s-1} reports in $\mathcal{U}$ tie under q at a common score κ, and every nonempty report omitting the label 1 is improved by adding it, because every outcome in $\mathcal{U}$ contains 1. Mixing q with the point mass at $\emptyset$ with weights $1/(1 + \kappa)$ and $\kappa/(1 + \kappa)$ makes the empty report tie as well, so the support and the Bayes-optimal set are both $\mathcal{A} = \{\emptyset\} \cup \mathcal{U}$, of size $2^{s-1} + 1$. The active score submatrix is $\operatorname{diag}(1, S_{\mathcal{U},\mathcal{U}})$, nonsingular by [1, Lemma 4.1], so $\mu = 0$ and (6) returns $|\mathcal{A}| - 1 = 2^{s-1}$.

At the two smallest label counts that construction is the following pair of explicit rational distributions, and the finite content of the application — which reports tie, which are strictly worse, and the nullity of the active-constraint matrix — is certified.

Theorem 5.1 ([1, Theorem 5.2], certified at $s = 2, 3$)**.** *Let $s = 2$ and $p = \frac{1}{7}(3, 2, 0, 2)$ in the order $\emptyset, \{1\}, \{2\}, \{1, 2\}$. Then $p \in \Delta_4$ with $\operatorname{supp} p = \operatorname{opt}_L(p) = \{\emptyset, \{1\}, \{1, 2\}\}$: those three reports have risk $4/7$ and the remaining report $\{2\}$ has risk $6/7$. The matrix of the constraints active at p is nonsingular, so $\mu = 0$ and* (6) *gives $\|p\|_0 - \mu - 1 = 3 - 0 - 1 = 2 = 2^{s-1}$.*

Let $s = 3$ and $p = \frac{1}{34}(13, 6, 0, 6, 0, 6, 0, 3)$ in the order $\emptyset, \{1\}, \{2\}, \{1, 2\}, \{3\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$. Then $p \in \Delta_8$ with $\operatorname{supp} p = \operatorname{opt}_L(p) = \{\emptyset, \{1\}, \{1, 2\}, \{1, 3\}, \{1, 2, 3\}\}$: those five reports have risk $21/34$, while $\{2\}$, $\{3\}$ and $\{2, 3\}$ have risks $15/17$, $15/17$ and $14/17$. Again the active-constraint matrix is nonsingular, so $\mu = 0$ and (6) *gives $5 - 0 - 1 = 4 = 2^{s-1}$.*

Proof sketch. The risks are exact rational identities: multiplying through by Λ and by the denominator of p turns each of them into an integer computation, and the equalities and strict inequalities above are then decided over $\mathbb{Z}$. The nullity is $\mu = N - \operatorname{rank}(\text{active matrix})$ by [3, Lemma 15], and the rank of the active matrix is N by one mod-p certificate as in Section 3 (4×4 at $s = 2$, 8×8 at $s = 3$). The passage from $\|p\|_0 - \mu - 1$ to a bound on CCdim is [3, Theorem 16], quoted and not reproved. $\square$

Theorem 5.2 ([3, Example 11], reproduced at $n = 4$)**.** *For the* 4*-class* 0-1 *loss the uniform distribution has full support, ties all four reports at risk* $3/4$*, and has a nonsingular active-constraint matrix. So* $\mu = 0$ *and* (6) *returns* $4-0-1 = 3 = n-1$*, which is the exact value* $\mathrm{CCdim}(L_n^{0\text{-}1}) = n-1$ *of* [3, Example 11].

Theorem 5.2 is a control in two directions at once. It shows that the machinery of Theorem 5.1 reproduces the one published exact value obtainable by this route; and it is the configuration that Corollary 4.3 shows to be unavailable for Jaccard, so it makes visible what the Jaccard loss lacks.

6. The exact value of the method, by exhaustive search

Definition 2.2 makes $\mathrm{fsd}(L)$ a finite maximum, so at small s it can be computed outright. Doing so answers a question neither [1] nor [2] raises: is their witness the best one for their method?

What was run. By Proposition 2.1 it is enough to enumerate pairs $(T, \mathcal{B})$ of a candidate support and a candidate Bayes set, evaluate $\mathrm{affdim}\{L_{\cdot,B}|_T : B \in \mathcal{B}\}$ exactly over $\mathbb{Q}$, sort by that value, and test each pair from the top for *strict* realisability: is there $p \in \Delta_N$ with $\mathrm{supp}\, p = T$ exactly and $\mathrm{opt}_L(p) = \mathcal{B}$ exactly? That test is a linear program — maximise a slack δ subject to $\sum_T p = 1$, ties on $\mathcal{B}$, $p_A \geq \delta$ on T and risk gap $\geq \delta$ off $\mathcal{B}$ — and it was solved in exact rational arithmetic by a two-phase simplex with Bland's rule written from scratch for this purpose, so no floating-point comparison enters any of the numbers below. The first realisable pair encountered gives the value; the pairs above it, all infeasible, certify that nothing larger occurs.

At $s = 4$ the direct enumeration is out of reach: $2^{16} \times 2^{16} \approx 4.3 \times 10^9$ pairs. It was replaced by a search over tie sets alone. For any $\mathcal{B} \subseteq \mathrm{opt}_L(p)$ the relaxed polytope $P(\mathcal{B}) = \{p \in \Delta_N :$ ties on $\mathcal{B}$, risk gaps ≥ 0 off $\mathcal{B}\}$ contains p, so $\mathrm{supp}\, p$ lies inside the largest support $T_{\max}(\mathcal{B})$ attainable in $P(\mathcal{B})$, which is computed by an incremental exact LP (repeatedly maximise the mass outside the support found so far; a zero optimum certifies that those coordinates vanish on all of $P(\mathcal{B})$), two to four LPs per set. A value above 8 forces $|\,\mathrm{opt}_L(p)| \geq 10$, so by monotonicity of $T_{\max}$ it suffices to run every one of the $\binom{16}{10} = 8\,008$ ten-element tie sets.

What came out.

loss	fsd	published bound	how the upper half was settled
L_2^{Jac}	2	$2^{s-1} = 2$	1 pair above the value, infeasible
L_3^{Jac}	4	4	all 1 288 pairs of value ≥ 5 infeasible
L_4^{Jac}	8	8	support collapse, 8 008 tie sets
$L_2^{(0)}$	1	$2^{s-1} - 1 = 1$	24 pairs, all infeasible
$L_3^{(0)}$	3	3	8 127 pairs, all infeasible
L_2^{Jac}, $\emptyset \notin \mathrm{supp}\, p$	1	—	5 pairs, all infeasible
L_3^{Jac}, $\emptyset \notin \mathrm{supp}\, p$	3	—	2 652 pairs, all infeasible
$L_n^{0\text{-}1}$, $n = 3, 4, 5$	$n-1$	[3, Ex. 11]	control
$L_2^{F_1}$, $L_3^{F_1}$	2, 3	[6]	cross-check

The last two rows are controls, run with the same code before the Jaccard rows were believed: the search reproduces the published exact value $\mathrm{fsd}(L_n^{0\text{-}1}) = n - 1$ of [3, Example 11] at $n = 3, 4, 5$, and it reproduces the independently obtained F_1 values of [6] at $s = 2, 3$. Every attained value in the table is realised by the witness of [1, Theorem 5.2] itself: the mixed factorial witness of Section 5 has $\mathrm{val} = 2^{s-1}$ (checked at $s = 1, \ldots, 5$), and the same witness without the empty-outcome mixture — q alone, supported on $\mathcal{U}$ — has $\|q\|_0 = |\,\mathrm{opt}_L(q)| = 2^{s-1}$ and $\mathrm{val} = 2^{s-1} - 1$ under both conventions (checked at $s = 2, \ldots, 5$), which is what the four middle rows report.

Two readings. First, at $s = 2, 3, 4$ the published lower bound 2^{s-1} is exactly $\mathrm{fsd}(L_s^{\mathrm{Jac}})$: the factor-of-two gap in (4) is a property of [3, Theorem 16], not a weakness of the witness. Under the alternative convention the published $2^{s-1} - 1$ is likewise exactly the ceiling at $s = 2, 3$; so the convention shifts the bound and the ceiling by the same 1 and nothing is left over. Second, together with Corollary 4.4 — which is a statement for all s — this says that at $s = 2, 3, 4$ the lower half of (4) cannot be improved without leaving the method, and that at every $s \geq 2$ the upper half cannot be met by it at all.

Remark 6.1 (measured, not proved). The $s = 4$ sweep produced a stronger fact than it needed. At $s = 3$ and $s = 4$, *every* tie set of size $2^{s-1} + 2$ collapses the attainable support onto $\operatorname{supp} p^\star$: all 28 six-element tie sets at $s = 3$ give $|T_{\max}| = 4$, and all 8 008 ten-element tie sets at $s = 4$ give $|T_{\max}| = 5$, with the histogram of $|T_{\max}|$ concentrated on the single value in each case. This is what makes the $s = 4$ computation affordable at all. If it holds for every s then, with $\mathrm{val} \leq |\,\mathrm{opt}\,| - 1$, it gives $\mathrm{fsd}(L_s^{\mathrm{Jac}}) = 2^{s-1}$ for every s. That general statement is a conjecture here, not a result.

Remark 6.2 (the standing of Section 6). Everything in this section is exact: integer and rational arithmetic throughout, no floating point, and every "no larger value exists" claim is a finite list of infeasibility certificates. It is nevertheless *not* machine-checked in the sense of Section 7: the enumeration and the simplex are ordinary programs, and the results are reported as computations, not as kernel-checked theorems. What *is* kernel-checked is that each attained value is attained — that is Theorem 5.1 at $s = 2, 3$.

7. Verification

Theorems 3.1, 3.2, 4.2, 5.1 and 5.2, Lemma 4.1 and Corollary 4.3 are formalised and machine-checked in Lean 4 against *Mathlib*, in four files of 1 410 lines whose bulk is certificate data. Proposition 2.1, Definition 2.2, Corollary 4.4, Remark 4.5 and the whole of Section 6 with its two remarks are not, and are marked as such where they appear. What is formalised is stated precisely, because the split matters here. The convex calibration dimension itself is *not* formalised — there is no calibration theory in *Mathlib* — so the passage from a witness to a bound on CCdim is [3, Theorem 16] quoted, and likewise fsd is a definition of this paper and not of the development; consequently Corollary 4.4 is the one-line argument printed above, resting on the formalised Theorem 4.2, and is not itself a formal theorem. What is machine-checked is: the Jaccard score and loss matrices exactly as in (1), over $\mathbb{Q}$ and with no per-entry numerical assertion (the integer rescaling by $\Lambda = \mathrm{lcm}(1, \ldots, s)$ is proved to agree with the rational definition for all s, A, B); the certificate lemmas of Section 3; the identity $L - U = -S$ and the rank and affine-dimension values of Theorems 3.1 and 3.2; Lemma 4.1 and Theorem 4.2 in full generality, quantified over all s and all $p \geq 0$ with $\sum_A p_A = 1$; the two finite assertions of Corollary 4.3, namely that every all-tied p gives $[s]$ mass zero when $s \geq 2$ and that the uniform distribution on $\mathcal{Y}$ is not all-tied when $s \geq 2$ (the reading of these as the failure of the hypotheses of [3, Theorem 18] and [3, Corollary 19] is prose, since CCdim is not formalised); and, for each witness of Theorem 5.1 and Theorem 5.2, that it is a probability vector with exactly the stated support, exactly which reports tie and which are strictly worse, and the nullity μ of the active-constraint matrix. Negative controls are part of the development, not an afterthought: the rank perturbation of Theorem 3.2 (a one-entry change really does move the rank), the failure of the uniform distribution to tie the Jaccard reports for $s \geq 2$ (so that [3, Corollary 19] is not silently in play), the strict inequalities inside Theorem 5.1 (which pin $\mathrm{opt}_L(p)$ to exactly the stated family, a larger one changing μ), and a witness showing the hypothesis of Theorem 4.2 is satisfiable at every s. The reasoning layer uses no compiled evaluation: axiom sets were printed for the thirty-nine declarations carrying the statements above, and each is exactly propositional extensionality, quotient soundness and choice — no `native_decide`, and no incomplete proof, anywhere in the development. The certificate checks run inside ordinary kernel evaluation: the largest file takes between eighty and ninety seconds of processor time and about eight and a half gigabytes of memory. The repository is private: repository reference to be added at submission.

8. The frontier

Question 8.1. Is $\mathrm{fsd}(L_s^{\mathrm{Jac}}) = 2^{s-1}$ for every s?

Corollary 4.4 gives fsd $\le 2^s - 2$ for all s and the search gives equality with 2^{s-1} at $s \le 4$, but the two do not meet. Remark 6.1 points at the missing step: if every tie set of size $2^{s-1}+2$ forces the support into $\operatorname{supp} p^\star$ — measured at $s = 3, 4$ — then val $\le |\,\mathrm{opt}\,| - 1$ closes the question. That looks like an extension of the three-probe argument of Theorem 4.2 from complete ties to partial ones, and it is the natural next theorem. Extending the search itself is not the way in: at $s = 5$ the collapse route needs every tie set of size $2^{s-1} + 2 = 18$ out of 32, that is $\binom{32}{18} \approx 4.7 \times 10^8$ LP groups, some 6×10^4 times the 8 008 sets of the $s = 4$ sweep; no price for it has been measured.

Question 8.2. Is $\mathrm{CCdim}(L_2^{\mathrm{Jac}})$ equal to 2 or to 3, and where in $\{4, \ldots, 7\}$ does $\mathrm{CCdim}(L_3^{\mathrm{Jac}})$ lie?

This is no longer a computational question. From below, Section 6 says the feasible-subspace bound is exhausted at 2^{s-1}, so raising the lower end needs a technique that [3, Theorem 16] does not contain, and Corollary 4.3 rules out the obvious alternative inside [3]. From above, lowering the upper end needs an explicit convex calibrated surrogate in fewer than $2^s - 1$ coordinates, and neither source supplies one. [2, Section 6] constructs a calibrated surrogate in exactly $2^s - 1$ coordinates, and [2] records (its Sections 1 and 7) that its lower bound does not rule out approximate notions of calibration; [1, Section 6] does construct MinHash surrogates of polynomial dimension, but only for a nonzero approximation floor α, their dimension diverging as $\alpha \downarrow 0$, and it says itself that there is no conflict with the exact bound.

Two further directions are routine rather than open. The certificates of Section 3 stay small at $s = 5$ — the modulus is fixed and $\Lambda = \mathrm{lcm}(1, \ldots, 5) = 60$ — and generating them is cheap; only the kernel check is expensive, its ring operations growing like 2^{3s}, so one $s = 5$ certificate should cost about eight times an $s = 4$ one (an extrapolation, not a measurement) and splitting the file by matrix would keep it inside an ordinary budget. And [1, Theorem 4.2] is proved for all s by an argument — a MinHash Gram representation plus Boolean Möbius inversion — whose ingredients are within reach of a direct formalisation, which would replace the per-s certificates of Theorem 3.1 by a single theorem.

References

[1] M. Zhang, *Exponential Convex Calibration Dimension for the Multi-Label Jaccard Measure*, arXiv:2608.13549v1 (submitted 13 August 2026), primary classification cs.LG, cross-listed stat.ML. Read in full, appendices included, on 3 September 2026, when the arXiv record showed v1 as the only version, not withdrawn, with no comments line and no journal reference. All numbering here is that of v1 and was read from its compiled e-print (the numbers resolved in the `.aux` file, not counted by hand): its equation (1) is (1) above, its Lemma 4.1 the positive definiteness of the nonempty block, its Theorem 4.2 the ranks (2), its Corollary 4.3 the upper bound (3), its Remark 4.4 the perturbation (5), its Lemma 5.1 the factorial balancing identity, its Theorem 5.2 the window (4), its Remark 5.3 the sentence quoted in Section 1 and its Remark 5.4 the alternative-convention window.

[2] P. Dewasurendra (Johns Hopkins University), *Dice Is Quadratic, Jaccard Is Exponential: Convex Calibration of Set Similarities*, unpublished manuscript, undated. Indexed by Semantic Scholar as CorpusId 291141794, with no arXiv identifier, no venue and no year in that record; retrieved on 3 September 2026 from `https://andotheror.github.io/pdf/dice-is-quadratic-jaccard-is-exponential-convex-calibration-of-set-sim.pdf`, the open-access link that record then carried, and read in full; all numbering here is that of the version retrieved on that date. A concurrent, independent treatment of the same loss by a different author: its Proposition 2 is (2), its Lemma 1 the strict positive definiteness of the Jaccard score matrix, its Lemmas 3 and 4 the factorial ties and their interiorisation, its Theorem 5 the window (4), its Corollary 6 the $\Theta(s^2)$ versus $\Theta(2^s)$ separation from Dice, and its Appendix G the alternative-convention window. It computes no value of fsd and makes no statement about the ceiling of the method.

[3] H. G. Ramaswamy and S. Agarwal, *Convex Calibration Dimension for Multiclass Loss Matrices*, Journal of Machine Learning Research **17** (2016), no. 14, 1–45. Definition 10 (the convex calibration dimension), Lemma 11 (the general upper bound $n-1$), Theorem 12 (the upper bound affdim L), Definition 14 and Lemma 15 (the feasible subspace dimension and its computation as a nullity), Theorem 16 (the bound (6)), Lemma 17, Theorem 18 and Corollary 19 (the full-support all-tied sufficient condition), Example 11 ($\mathrm{CCdim}(L_n^{0\text{-}1}) = n - 1$, from Theorem 16

at the uniform distribution and Lemma 11). The volume, article number and page range are those of the JMLR record for the article; the numbering was read from the JMLR PDF.

[4] M. Zhang, *Exact Rank and Convex Calibration Dimension Lower Bounds for the Multi-Label F_1 Loss*, arXiv: 2608.08399v1 (submitted 9 August 2026). The companion source for the F_1 loss, cited by [1] for $\mathrm{CCdim}(L^{F_1}) = \Theta(s^2)$. Its numbering is that of v1 as read from its compiled e-print and from the arXiv v1 PDF: its Theorem 4.3 is the rank formula $s^2 - s + 2$, its Corollary 4.4 the upper bound $s^2 - s + 1$, and its Theorem 5.1 the lower bound used for the comparison in Section 1.

[5] M. Zhang, H. G. Ramaswamy and S. Agarwal, *Convex Calibrated Surrogates for the Multi-Label F-Measure*, in: Proceedings of the 37th International Conference on Machine Learning (ICML 2020), Proceedings of Machine Learning Research vol. 119, 11246–11255. Cited here only for the quadratic-dimensional calibrated F_1 surrogate that supplies the upper half of the $\Theta(s^2)$ statement; not read directly for this paper.

[6] Korea Superintelligence Labs, *The convex calibration dimension of the multi-label F_1 loss at two and three labels*, companion manuscript, 2026. Its exhaustive search, written independently of the one in Section 6, supplies the F_1 cross-check values $\mathrm{fsd}(L_2^{F_1}) = 2$ and $\mathrm{fsd}(L_3^{F_1}) = 3$ used there, and is the source of the contrast drawn in Section 1.

[7] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction — CADE 28, Lecture Notes in Computer Science, Springer, 2021, 625–635; DOI `10.1007/978-3-030-79876-5_37`.

[8] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, 367–381; DOI `10.1145/3372885.3373824`.