THE CONVEX CALIBRATION DIMENSION OF THE MULTI-LABEL F_1 LOSS AT TWO AND THREE LABELS, AND THE F_β SCALE

KOREA SUPERINTELLIGENCE LABS

ABSTRACT. For a multi-label problem with s labels, the instance-wise F_1 loss is the $2^s \times 2^s$ matrix $L^{F_1} = J - F$ with $F(A,B) = 2|A \cap B|/(|A| + |B|)$ and $F(\emptyset,\emptyset) = 1$. Zhang has recently determined $\operatorname{rank} L^{F_1} = s^2 - s + 2$ exactly, hence $\mathrm{CCdim}(L^{F_1}) \le s^2 - s + 1$, and has proved $\mathrm{CCdim}(L^{F_1}) \ge (2/3\sqrt{3})\, s^2 - O(s)$ for $s \ge 3$, leaving the exact value open. At the two smallest label counts the finite form of that lower bound is weak or absent: it assumes $s \ge 3$, and at $s = 3$ it evaluates to 2. We give explicit rational witnesses that improve both ends of the small-s window: $\mathrm{CCdim}(L_2^{F_1}) \ge 2$ and $\mathrm{CCdim}(L_3^{F_1}) \ge 3$. The first is a lower bound where none was available — Nowak-Vila, Bach and Rudi record that the standard sufficient condition fails at two labels — and we prove that failure, in the form that no distribution on the four outcomes ties all four reports, which caps the Ramaswamy–Agarwal feasible-subspace method at 2 there; so $\mathrm{CCdim}(L_2^{F_1}) \in \{2, 3\}$ and that method cannot decide which. The second is one more than the source's own theorem gives at $s = 3$, and an exhaustive exact search shows the gain comes entirely from the convention $F(\emptyset,\emptyset) = 1$: among witnesses giving the empty outcome no mass, the best bound at $s = 3$ is exactly the source's value, so its theorem is tight inside its own witness class. We also re-verify the rank formula at $s \le 4$ from explicit integer certificates.

This version adds the asymmetric F_β scale, $F_\beta(A,B) = (1+\beta^2)|A \cap B|/(\beta^2|A| + |B|)$ with the same convention, which the source names as an open direction. The rank $s^2 - s + 2$ holds for every $\beta \in (0,\infty)$, not only $\beta = 1$ — a sharpening by $s - 1$ of the published bound $s^2 + 1$ across the whole scale — and is machine-checked uniformly in β at $s \le 2$ by polynomial left-inverse certificates, one per matrix and each serving every β; for every s it is proved in prose, for the score matrix and hence for $L^{F_\beta} - J$, and recorded separately. At both endpoints of the scale, precision ($\beta \to 0$) and recall ($\beta \to \infty$), the rank collapses to $s + 1$ for every s, with $\operatorname{affdim}(J - P) = s$ and hence $\mathrm{CCdim}(L_s^{\mathrm{prec}}) \le s$. At two labels, $\mathrm{CCdim}(L_2^{F_\beta}) \ge 2$ for every β by one rational-function witness, no distribution ties all four reports for any β, and so $\mathrm{CCdim}(L_2^{F_\beta}) \in \{2, 3\}$ along the whole scale, undecidable by the feasible-subspace method. At three labels the feasible-subspace bound depends on β: explicit witnesses over $\mathbb{Q}(\beta^2)$ give at least 4 for $\beta^2 < 1/2$ and at least 3 for $\beta^2 \ge 1/2$, an exhaustive search at nine values of β^2 returns exactly those values, and the two witnesses meet at $\beta^2 = 1/2$. All the finite content is machine-checked in Lean 4; Section 9 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 $[s] = \{1, \dots, s\}$, so both range over a set of size $N = 2^s$. The instance-wise F_1 score of the report B against the outcome A is

$$F(A,B) \;=\; \begin{cases} 1, & |A| + |B| = 0, \\ \dfrac{2|A \cap B|}{|A| + |B|}, & \text{otherwise}, \end{cases} \tag{1}$$

and the F_1 loss is $\ell(A,B) = 1 - F(A,B)$. Collecting these gives the $2^s \times 2^s$ matrices F and $L = L^{F_1} = J - F$, where J is all-ones. Following Ramaswamy and Agarwal [2], 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.

The *convex calibration dimension* $\mathrm{CCdim}(L)$ of [2, Definition 10] is the least d for which there is a convex set $\mathcal{C} \subseteq \mathbb{R}^d$, a convex surrogate $\psi : \mathcal{C} \to \mathbb{R}^N_+$ and a link pred from $\mathcal{C}$ to the reports that is L-calibrated. It is the number of real-valued functions that a consistent convex surrogate method must learn, and it is the quantity this paper is about.

The source and what it leaves open. Zhang [1] determines the rank of the F_1 matrices exactly. Its Theorem 4.3 states that for every $s \geq 1$,

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

with a complete proof in its appendix (a Kronecker factorisation on the nonempty block, together with positive-definiteness of the Cauchy matrix $(2/(k+k'))_{k,k'}$). With [2, Theorem 12], which bounds $\mathrm{CCdim}(L)$ by $\operatorname{affdim} L$, this gives its Corollary 4.4, $\mathrm{CCdim}(L^{F_1}) \leq s^2 - s + 1$. Its Theorem 5.1 is a lower bound: for $s \geq 3$, with $t = \lfloor s/3 \rfloor$, $n = s - t$ and $h = \lceil \sqrt{st}\, \rceil - 1$,

$$\mathrm{CCdim}(L^{F_1}) \;\geq\; h\,n \;=\; \tfrac{2}{3\sqrt{3}}\, s^2 - O(s), \tag{3}$$

and its conclusion says, verbatim: *"The bounds do not determine the exact convex calibration dimension."* Numbered results of [1] and of [2] always carry their citation below; an unqualified number is this paper's own.

Two features of (3) matter here and are easy to miss behind the asymptotics. First, the finite value of hn is far below the leading term at small s: it is 2 at $s = 3$, 3 at $s = 4$, 8 at $s = 5$ and 12 at $s = 6$, against upper bounds $7, 13, 21, 31$. Second, Theorem 5.1 assumes $s \geq 3$, so the source proves nothing at all at $s = 2$, where the upper bound is 3. The honest windows at the two smallest label counts are therefore

$$\mathrm{CCdim}(L_2^{F_1}) \in \{?, \ldots, 3\}, \qquad \mathrm{CCdim}(L_3^{F_1}) \in \{2, \ldots, 7\}.$$

Nor does the earlier literature fill the first of these. Nowak-Vila, Bach and Rudi [3], studying the same loss with the same empty-set convention, say of the F_1 score that they "can't say anything about the potential existence of a convex calibrated surrogate with smaller dimension than $\operatorname{affdim}(L) - 1$", because the sufficient condition of [2, Theorem 18] "does not hold for any Π even for $m = 2$". Here Π is their name for the conditional label distribution and m for the number of labels, so "$m = 2$" is $s = 2$ and their $\operatorname{affdim}(L) - 1$ is 2; the two quoted sentences stand word for word in the arXiv source and in the published AISTATS supplementary alike, in the F-score appendix under the heading "Optimality of r".

The source treats $\beta = 1$ only. Its conclusion also says, verbatim: *"Another natural direction is to extend the exact rank and calibration-dimension analysis to the asymmetric F_β family and to related set-similarity losses such as Jaccard loss."* For general β the published rank statement is the upper bound $\operatorname{rank}(L^{F_\beta} - J) \leq s^2 + 1$ of Zhang, Ramaswamy and Agarwal [4, Proposition 3], of which the source says (related-work section, verbatim): *"Our rank theorem sharpens their bound for* $\beta = 1$ *and identifies the exact affine dimension."*

What this paper settles. We fix both windows from below, show that at $s = 2$ the standard method can do no better, and carry the rank and the two-label results to the whole F_β scale.

- *Two labels* (Section 4). An explicit rational witness gives $\mathrm{CCdim}(L_2^{F_1}) \geq 2$ (Theorem 4.1). We are aware of no previous lower bound at two labels: [1] assumes $s \geq 3$, [3] records the obstruction quoted above, and none of the three works indexed as citing [1] on 3 September 2026 computes a value of CCdim for the F_1 loss at any specific s — [5] and [6] are about the Jaccard measure and quote only $\mathrm{CCdim}(L^{F_1}) = \Theta(s^2)$, and the third has no calibration-dimension content. Complementing it, no probability vector on the four outcomes makes all four F_1 reports tie — the unique solution of the tie system is $(1/4, 3/4, 3/4, -3/4)$, which is not nonnegative (Proposition 4.2) — which is the failure [3] assert, here proved and in the stronger closed-simplex form, and which caps the Ramaswamy–Agarwal feasible-subspace bound at 2. Hence $\mathrm{CCdim}(L_2^{F_1}) \in \{2,3\}$, and no application of [2, Theorem 16] decides which (Corollary 4.3).

- *Three labels* (Section 5). A second explicit witness gives $\mathrm{CCdim}(L_3^{F_1}) \geq 3$ (Theorem 5.1), one more than (3) yields at $s = 3$. Section 8 records the exhaustive computation showing that the extra 1 comes entirely from the convention $F(\emptyset, \emptyset) = 1$ of (1): among witnesses that give the empty outcome no mass — which is what the construction of [1] does — the best available bound at $s = 3$ is exactly its $hn = 2$. So its Theorem 5.1 is tight at $s = 3$ inside its own witness class, and our improvement is not available to any convention-independent argument.
- *The published rank formula* (Section 3). Equation (2), and the perturbation recorded as Remark 4.5 of [1], re-verified at $s \leq 4$ from explicit integer certificates (Theorems 3.1 and 3.2). This is published mathematics, proved there for all s; what is new here is only the certification, and it is stated because the invertibility of $L_2^{F_1}$ is used in Proposition 4.2 and because the certificate format is what keeps the arithmetic small (Remark 3.3).
- *A control* (Section 6). The same machinery applied to the 4-class 0-1 loss returns $3 = n - 1$, reproducing [2, Example 11] — the one loss whose convex calibration dimension [2] computes by exactly the route used here (Theorem 6.1).
- *The F_β scale* (Section 7). Writing $b = \beta^2$, every entry of F_β is a rational function of b, and the statements below are uniform in $b \in (0, \infty)$. The rank is $s^2 - s + 2$ for every β: at $s \leq 2$ this is machine-checked from polynomial left inverses, one per matrix, whose scalars have no positive root (Theorem 7.2), and for every s it is proved in prose, for $\operatorname{rank} F_\beta = \operatorname{rank}(L^{F_\beta} - J)$, by a b-free left factor and two Cauchy systems (Section 8, not formalised); either way it sharpens [4, Proposition 3] by $s - 1$ along the whole scale. At the two endpoints the rank collapses: the precision and recall score and loss matrices have rank exactly $s + 1$ for every s, and $\mathrm{affdim}(J - P) = s$, so $\mathrm{CCdim}(L_s^{\mathrm{prec}}) \leq s$ (Theorem 7.3). At two labels a single rational-function witness gives $\mathrm{CCdim}(L_2^{F_\beta}) \geq 2$ for every β (Theorem 7.4), no distribution ties all four reports for any β (Theorem 7.5), and so $\mathrm{CCdim}(L_2^{F_\beta}) \in \{2,3\}$ for every β, undecidable by [2, Theorem 16] (Corollary 7.6). At three labels the computations of Section 8 show the feasible-subspace bound is β-dependent: explicit witnesses give at least 4 below $\beta^2 = 1/2$ and at least 3 at and above it, an exhaustive search at nine sampled values finds nothing better, and the two witnesses meet at $\beta^2 = 1/2$.

Section 8 collects the computations and the one prose proof that are *not* part of the formal development and labels each as such: the exhaustive values of the feasible-subspace bound, the certificate-size measurement, two observations about the two-label case, the general-s rank theorem on the F_β scale, and the β-threshold at three labels. Section 9 describes the machine verification, Section 10 the frontier, and Appendix A reproduces the formal statements of Section 7 verbatim.

Changes from version 1. Version 1 of this paper contained Sections 1–6, the F_1 part of Section 8, and Sections 9–10. This version adds Section 7 (Proposition 7.1, Theorems 7.2–7.5, Corollary 7.6), the two F_β subsections of Section 8, the corresponding paragraphs of Sections 9 and 10, Appendix A, and the extended title and abstract. Every statement of version 1 is retained unchanged, including the corrections made at its refereeing (the numbering of the results of [1] as 4.3/4.4/4.5/5.1, and the wording of the control, of the [3] attribution and of Proposition 4.2); the only change to earlier text is that [4] is now cited from its own arXiv and PMLR versions rather than through the restatement in [1].

2. Preliminaries

Indexing. We index outcomes and reports by subsets of $[s]$ in binary order, $A \mapsto \sum_{i \in A} 2^{i-1}$, so that at $s = 2$ the order is $\emptyset, \{1\}, \{2\}, \{1,2\}$ and at $s = 3$ it is $\emptyset, \{1\}, \{2\}, \{1,2\}, \{3\}, \{1,3\}, \{2,3\}, \{1,2,3\}$.

Written out, the two-label loss matrix is

$$L_2^{F_1} \;=\; \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1/3 \\ 1 & 1 & 0 & 1/3 \\ 1 & 1/3 & 1/3 & 0 \end{pmatrix}, \tag{4}$$

rows indexed by the outcome A and columns by the report B. We write $\ell_B \in \mathbb{R}^N$ for the column of L at the report B, so that the conditional risk of B under $p \in \Delta_N$ is $R_p(B) = p^\mathsf{T}\ell_B$, and $\mathrm{opt}_L(p) = \arg\min_B R_p(B)$ for the set of Bayes-optimal reports at p. We write $\|p\|_0 = |\operatorname{supp} p|$.

The Ramaswamy–Agarwal feasible-subspace bound. Both [1] and this paper obtain lower bounds on CCdim from one theorem of [2]. For a report B_0 its *trigger set* is

$$Q_{B_0} \;=\; \big\{\, p \in \Delta_N \;:\; p^\mathsf{T}(\ell_B - \ell_{B_0}) \geq 0 \text{ for every report } B \,\big\},$$

the set of distributions at which B_0 is Bayes-optimal. For a convex set Q and $p \in Q$, the *feasible subspace dimension* $\mu_Q(p)$ of [2, Definition 14] is the dimension of the largest linear subspace contained in the cone of feasible directions of Q at p; by [2, Lemma 15] it equals the nullity of the matrix whose rows are the constraints of Q active at p. Their Theorem 16 then reads

$$\mathrm{CCdim}(L) \;\geq\; \|p\|_0 - \mu_{Q_{B_0}}(p) - 1 \qquad (p \in \Delta_N,\; B_0 \in \mathrm{opt}_L(p)). \tag{5}$$

Their Lemma 11 and Theorem 12 give the matching upper bounds $\mathrm{CCdim}(L) \leq \min\{N-1, \operatorname{affdim} L\}$, and their Example 11 computes $\mathrm{CCdim}(L_n^{0\text{-}1}) = n-1$ for the 0-1 loss.

The active constraints of Q_{B_0} at p are the nonnegativities $p_A \geq 0$ for $A \notin \operatorname{supp} p$, the normalisation $\mathbf{1}^\mathsf{T}p = 1$, and the ties $p^\mathsf{T}(\ell_B - \ell_{B_0}) = 0$ for $B \in \mathrm{opt}_L(p)$. Writing out the nullity of that matrix turns (5) into a statement about a finite quantity.

Proposition 2.1. *Let $p \in \Delta_N$, $S = \operatorname{supp} p$, and $B_0 \in \mathrm{opt}_L(p)$. Then the right-hand side of* (5) *equals*

$$\mathrm{val}(p) \;=\; \operatorname{affdim}\big\{\, \ell_B|_S \;:\; B \in \mathrm{opt}_L(p) \,\big\},$$

the dimension of the affine hull of the Bayes-optimal columns restricted to the support. In particular it does not depend on the choice of B_0, and it depends on p only through the pair $(\operatorname{supp} p, \mathrm{opt}_L(p))$. *Consequently the best bound obtainable from* (5) *for a fixed L is the finite quantity*

$$\mathrm{fsd}(L) \;=\; \max_{p \in \Delta_N} \mathrm{val}(p).$$

Proof. The rows e_A $(A \notin S)$ of the active-constraint matrix annihilate the coordinates outside S, so its nullity is $\mu = |S| - \operatorname{rank}\big(\{\mathbf{1}_S\} \cup \{(\ell_B - \ell_{B_0})|_S : B \in \mathrm{opt}_L(p)\}\big)$. Now $p|_S$ is orthogonal to every active difference while $p|_S^\mathsf{T}\mathbf{1}_S = 1$, so $\mathbf{1}_S$ is not in the span of the differences and that rank is $1 + \dim\operatorname{span}\{(\ell_B - \ell_{B_0})|_S\}$. Hence $\|p\|_0 - \mu - 1 = \dim\operatorname{span}\{(\ell_B - \ell_{B_0})|_S : B \in \mathrm{opt}_L(p)\}$, which is the dimension of the affine hull of $\{\ell_B|_S : B \in \mathrm{opt}_L(p)\}$. □

Proposition 2.1 is a rewriting of definitions of [2] and is not part of the formal development; what is machine-checked below is each individual application of (5), in the form "here is p, here are its Bayes ties, here is the nullity μ". The proposition is stated because it is what makes the exhaustive search of Section 8 finite, and because it explains why the empty-set convention can move the answer: it changes which reports tie.

3. The rank of the F_1 matrices at small s

The two statements of this section are published mathematics — [1, Theorem 4.3 and Remark 4.5] prove them for every s — re-derived here at $s \leq 4$ from explicit integer certificates. We state them because Theorem 3.1 at $s = 2$ is used in the proof of Proposition 4.2, and because the certificate format is what makes the verification cheap (Remark 3.3).

Theorem 3.1. *For $s = 1, 2, 3, 4$,*

$$\operatorname{rank} F \;=\; \operatorname{rank} L^{F_1} \;=\; s^2 - s + 2 \qquad \textit{and} \qquad \operatorname{affdim} L^{F_1} \;=\; s^2 - s + 1,$$

that is, the ranks are $2, 4, 8, 14$ and the affine dimensions $1, 3, 7, 13$. Since $L - J = -F$, the rank of $L - J$ is that of F.

Theorem 3.2. *Let F^0 agree with F except that $F^0(\emptyset, \emptyset) = 0$ (*[1, Remark 4.5]*). For $s = 2, 3$,*

$$\operatorname{rank} F^0 \;=\; s^2 - s + 1 \;\neq\; s^2 - s + 2 \;=\; \operatorname{rank} F,$$

that is, $3 \neq 4$ at $s = 2$ and $7 \neq 8$ at $s = 3$: changing the single entry $F(\emptyset, \emptyset)$ from 1 to 0 drops the rank by one.

Proof of Theorems 3.1 and 3.2. Each value is a pair of bounds on the rank of an integer matrix. Multiplying by $\Lambda = \operatorname{lcm}(1, \ldots, 2s)$ clears every denominator $|A| + |B|$ of (1) and does not change the rank, so it suffices to bound the rank of the integer matrix $M = \Lambda \cdot F$ (respectively ΛL, ΛF^0, or ΛL with an all-ones row appended, for affdim). The upper bound is an explicit factorisation $M = AB$ through $\mathbb{R}^r$: a matrix with such a factorisation has rank at most r. The lower bound is an integer reduction certificate

$$XMY \;=\; I_r + p\,W, \qquad p = 1000003,$$

with X, Y, W integer matrices: reading the identity modulo p gives $\det(XMY) \equiv 1$, so $\det(XMY) \neq 0$ over $\mathbb{Z}$, so XMY is invertible over $\mathbb{Q}$ and $r \leq \operatorname{rank} M$. Both lemmas are proved in general and the certificate data are then evaluated on the stated matrices; the only inputs specific to a value of s are the four integer matrices A, B and X, Y, W, which were produced by exact rational elimination outside the verification and are re-checked inside it. No search is involved and no correctness of the generator is used. $\square$

Remark 3.3 (The size of the certificates; not part of the formal development). The reason for the modular form of the lower-bound certificate is arithmetic size, and it was measured rather than estimated. The obvious alternative — exhibit the adjugate of a nonsingular $r \times r$ minor — needs integers of the size of that minor's determinant, and for the minors chosen here the determinants of ΛL^{F_1} have 26, 38 and 69 decimal digits at $s = 4, 5, 6$ (computed exactly by fraction-free elimination over $\mathbb{Z}$). The certificate $XMY = I + pW$ at $p = 1000003$ instead keeps every entry of Y below 10^6, with the largest entry of W equal to 5880, 35280 and 498960 at those same s, and at most 1.01×10^7 at $s = 7$; the entries of B lie in $\{-2, \ldots, 2\}$. Seven digits against sixty-nine is what keeps the whole verification inside ordinary kernel evaluation, with no compiled arithmetic anywhere (Section 9). The bound behind the observation is that the entries of such a certificate are bounded by p and by $r \cdot N \cdot \max_{i,j} |M_{ij}|$, independently of the loss; we have not looked for it in the literature on exact linear algebra, where it may well be folklore.

4. Two labels

At $s = 2$ there are four outcomes and four reports, the loss matrix is (4), and $\operatorname{affdim} L_2^{F_1} = 3 = N - 1$, so $\operatorname{CCdim}(L_2^{F_1}) \leq 3$; from below we found nothing in the literature.

Theorem 4.1. *Let $p = \frac{1}{21}(9, 7, 2, 3)$ on the outcomes $\emptyset, \{1\}, \{2\}, \{1, 2\}$. Then p is strictly positive with $\mathbf{1}^\mathsf{T} p = 1$; its Bayes-optimal reports are exactly $\emptyset, \{1\}, \{1, 2\}$, of common risk $4/7$, while the report $\{2\}$ has risk $17/21 > 4/7$; and the matrix of constraints of $Q_\emptyset$ active at p has nullity $\mu = 1$. Consequently* (5) *gives*

$$\operatorname{CCdim}(L_2^{F_1}) \;\geq\; \|p\|_0 - \mu - 1 \;=\; 4 - 1 - 1 \;=\; 2.$$

Proof. The four risks are read off (4):

$$R_p(\emptyset) = \tfrac{0 \cdot 9 + 7 + 2 + 3}{21} = \tfrac{4}{7}, \quad R_p(\{1\}) = \tfrac{9 + 0 + 2 + 1}{21} = \tfrac{4}{7}, \quad R_p(\{2\}) = \tfrac{9 + 7 + 0 + 1}{21} = \tfrac{17}{21},$$
$$R_p(\{1, 2\}) = \tfrac{1}{21}\big(9 + \tfrac{7}{3} + \tfrac{2}{3} + 0\big) = \tfrac{4}{7}.$$

So $\mathrm{opt}(p) = \{\emptyset, \{1\}, \{1,2\}\}$. Since p has full support there are no active nonnegativity constraints, and the active-constraint matrix of $Q_\emptyset$ at p is the 3×4 matrix with rows $\ell_{\{1\}} - \ell_\emptyset$, $\ell_{\{1,2\}} - \ell_\emptyset$ and $\mathbf{1}^\top$. That matrix has rank 3 — an integer reduction certificate of the shape used in Section 3 exhibits a nonsingular 3×3 reduction — so its nullity is $4 - 3 = 1$, and (5) applies with $B_0 = \emptyset$. □

The witness is asymmetric in the two labels and puts mass on the empty outcome; it was found by the exhaustive search of Section 8, not by an ansatz. The next proposition is the reason no better witness exists.

Proposition 4.2. *There is no $p \in \mathbb{R}^4$ with $p \geq 0$ and $\mathbf{1}^\top p = 1$ under which all four F_1 reports have the same conditional risk. The unique p with $\mathbf{1}^\top p = 1$ solving the tie system is $p = \left(\frac{1}{4}, \frac{3}{4}, \frac{3}{4}, -\frac{3}{4}\right)$, whose last coordinate is negative. Consequently the sufficient condition of* [2, Theorem 18] *— a full-support point at which every report ties, which would give $\mathrm{CCdim}(L) \geq \mathrm{affdim}\, L - 1$ — is unavailable at $s = 2$, and so is the hypothesis of* [2, Corollary 19]*, that all columns of L be permutations of one another, which* (4) *plainly fails. Moreover no application of* (5) *can exceed the value* 2 *there.*

Proof. Uniqueness: $L = L_2^{F_1}$ is invertible by Theorem 3.1 at $s = 2$, so $L^\top p = c\,\mathbf{1}$ forces $p = c\,(L^\top)^{-1}\mathbf{1}$, and $\mathbf{1}^\top p = 1$ fixes c; the resulting vector is $(\frac{1}{4}, \frac{3}{4}, \frac{3}{4}, -\frac{3}{4})$, at which all four risks equal $3/4$, as one checks against (4). Since it has a negative coordinate, no nonnegative p ties all four reports; the machine-checked statement is this last one, proved directly from the sixteen entries of (4) by linear arithmetic and so covering every $p \geq 0$, not only the relative interior of the simplex.

For the final sentence: by Proposition 2.1 a value of 3 from (5) would need $\|p\|_0 \geq 4$, hence full support, and four affinely independent Bayes-optimal columns, hence all four reports tied; and $N = 4$, so no value above 3 is possible at all. □

Corollary 4.3. $\mathrm{CCdim}(L_2^{F_1}) \in \{2, 3\}$*, and no application of* [2, Theorem 16] *decides which.*

Proof. The lower bound is Theorem 4.1; the upper bound is $\mathrm{CCdim}(L) \leq \min\{N-1, \mathrm{affdim}\, L\} = 3$ by [2, Lemma 11 and Theorem 12] with Theorem 3.1. Proposition 4.2 caps at 2 every bound obtainable from (5). □

Corollary 4.3 is where [3] stop: Proposition 4.2 proves the obstruction they name, and Theorem 4.1 supplies the lower bound they say they cannot give. What remains open at $s = 2$ is a genuinely new idea on one side or the other; Section 10 records what is ruled out. Section 7 shows that the whole of this section is a statement about the entire F_β scale, not about $\beta = 1$.

5. Three labels

At $s = 3$ the bound (3) of [1, Theorem 5.1] applies and evaluates to $hn = 1 \cdot 2 = 2$, against the upper bound $\mathrm{affdim}\, L_3^{F_1} = 7$.

Theorem 5.1. *Let*

$$p \;=\; \tfrac{1}{66}\,(25,\, 15,\, 0,\, 15,\, 11,\, 0,\, 0,\, 0),$$

that is, mass $\frac{25}{66}$ on $\emptyset$, $\frac{5}{22}$ on $\{1\}$, $\frac{5}{22}$ on $\{1,2\}$ and $\frac{1}{6}$ on $\{3\}$, and no mass on the other four outcomes; so $\|p\|_0 = 4$. Then the Bayes-optimal reports at p are exactly $\emptyset, \{1\}, \{1,2\}, \{1,2,3\}$, of common risk $41/66$, and the matrix of constraints of $Q_\emptyset$ active at p — four nonnegativity rows, three tie rows and the normalisation row — is nonsingular, so $\mu = 0$. Consequently (5) *gives*

$$\mathrm{CCdim}(L_3^{F_1}) \;\geq\; \|p\|_0 - \mu - 1 \;=\; 4 - 0 - 1 \;=\; 3,$$

one more than [1, Theorem 5.1] *gives at $s = 3$.*

Proof. Writing the risks over the denominator 66 and using (1), the eight reports carry the risks

$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$	$\{3\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$\frac{41}{66}$	$\frac{41}{66}$	$\frac{28}{33}$	$\frac{41}{66}$	$\frac{5}{6}$	$\frac{247}{396}$	$\frac{307}{396}$	$\frac{41}{66}$,

so exactly $\emptyset, \{1\}, \{1,2\}, \{1,2,3\}$ are optimal. (For instance, at the report $\{1,2,3\}$ the four supported outcomes contribute $25\cdot 1 + 15\cdot\frac{1}{2} + 15\cdot\frac{1}{5} + 11\cdot\frac{1}{2} = 41$.) The outcomes $\{2\}, \{1,3\}, \{2,3\}, \{1,2,3\}$ carry no mass, so their nonnegativity constraints are active; together with the three tie rows $\ell_{\{1\}} - \ell_\emptyset$, $\ell_{\{1,2\}} - \ell_\emptyset$, $\ell_{\{1,2,3\}} - \ell_\emptyset$ and the normalisation row $\mathbf{1}^\mathsf{T}$ this is an 8×8 matrix, and an integer reduction certificate of the shape of Section 3 shows it is nonsingular. Hence $\mu = 0$ and (5) applies with $B_0 = \emptyset$. Machine-checked are the four equalities and the four strict inequalities among the risks, the support pattern, and the nullity; the displayed rational values are arithmetic from (1). □

Note the shape of the witness: a report can be Bayes-optimal at an outcome of zero probability — $\{1,2,3\}$ here — and it is precisely the fourth tie, supplied by that report, that raises the affine dimension of Proposition 2.1 from 2 to 3. Note also that the witness gives the empty outcome mass 25/66, the largest of the four. That is not incidental: Section 8 records that no witness avoiding the empty outcome does better than the source's own value at $s = 3$.

6. A control: the 0-1 loss

The one loss whose convex calibration dimension [2] computes by exactly the route used here is the n-class 0-1 loss: their Example 11 evaluates (5) at the uniform distribution and combines it with Lemma 11 to get $\mathrm{CCdim} = n - 1$. Running the same machinery on it returns that published value.

Theorem 6.1. *For the* 4*-class* 0-1 *loss and the uniform distribution* $p = (1/4, 1/4, 1/4, 1/4)$*, every report has risk* $3/4$*, so all four tie and the support is full; the matrix of active constraints — the three tie rows* $\ell_j - \ell_0$ *and the normalisation row — is nonsingular, so* $\mu = 0$*, and* (5) *gives* $\mathrm{CCdim}(L_4^{0\text{-}1}) \geq 4 - 0 - 1 = 3 = n - 1$*, which with* [2, Lemma 11] *is the published value* $\mathrm{CCdim}(L_n^{0\text{-}1}) = n - 1$*.*

Proof. The risks are immediate, and the nonsingularity is again an integer reduction certificate. □

Remark 6.2 (An off-by-one in a published bound; observation, not machine-checked beyond the instance). Theorem 6.1 also makes visible a small slack in [2, Lemma 17 and Theorem 18]. Their Lemma 17 states, for $p \in \mathrm{relint}(\Delta_n)$ and $c \in \mathbb{R}_+$ with $p^\mathsf{T}\ell_t = c$ for every t, that $\mu_{Q_t}(p) \leq n - \mathrm{affdim}(L)$; their Theorem 18 concludes $\mathrm{CCdim}(L) \geq \mathrm{affdim}(L) - 1$ from it and Theorem 16, and nothing else in their statement or proof supplies a further 1. The rank appearing in their proof of Lemma 17 is in fact exactly $\mathrm{affdim}(L) + 1$ and not merely $\geq \mathrm{affdim}(L)$, by the same orthogonality used in Proposition 2.1: p annihilates every column difference while $p^\mathsf{T}\mathbf{1} = 1$, so the all-ones row is independent of the differences. Hence $\mu = n - \mathrm{affdim}(L) - 1$ and $\mathrm{CCdim}(L) \geq \mathrm{affdim}(L)$, which with their Theorem 12 gives $\mathrm{CCdim}(L) = \mathrm{affdim}(L)$ exactly whenever a full-support all-tied point exists — and that sharp form is what their own Example 11 computes by hand for the 0-1 loss, where $\mathrm{affdim} = n-1$ rather than $n-2$. Theorem 6.1 is the machine-checked instance: $\mu = 0 = 4-3-1$. The same 1 is available in [2, Corollary 19], whose hypothesis — that all columns of L are permutations of one another — gives such a p.

7. The F_β scale

With $P = |A\cap B|/|B|$ the precision of the report B against the outcome A and $R = |A \cap B|/|A|$ its recall, the weighted F-measure is the weighted harmonic mean $F_\beta = (1+\beta^2)PR/(\beta^2 P + R)$, that is,

$$F_\beta(A,B) \;=\; \begin{cases} 1, & |A| + |B| = 0, \\ \dfrac{(1+\beta^2)\,|A\cap B|}{\beta^2|A| + |B|}, & \text{otherwise}, \end{cases} \qquad \beta \in (0,\infty), \tag{6}$$

with the loss $\ell_\beta(A,B) = 1 - F_\beta(A,B)$ and the matrix $L^{F_\beta} = J - F_\beta$. This is equation (1) of [4], including the value at $(\emptyset,\emptyset)$ (their proof of Proposition 3 carries the isolated term $-\mathbf{1}[\|y\|_1 = 0]\,\mathbf{1}[\|\hat{y}\|_1 = 0]$), and at $\beta = 1$ it is (1). Throughout this section we write

$$b \;=\; \beta^2 \in (0,\infty),$$

the variable in which every entry of F_β is a rational function with denominator $b|A| + |B| > 0$; we write F_b, L_b for the matrices when the dependence matters. All statements below are uniform in b: each is a single theorem quantified over every rational $b > 0$, checked once, not a collection of instances.

The definition, and two structural facts.

Proposition 7.1. (a) *At $\beta = 1$ the score* (6) *coincides entrywise with* (1)*, so $F_1 = F$ and L^{F_1} is the matrix of the preceding sections.* (b) *For every s and every $b > 0$, $F_{1/b} = (F_b)^\mathsf{T}$: replacing β by $1/\beta$ transposes the score matrix. Consequently every rank statement about F_b or L_b is invariant under $\beta \mapsto 1/\beta$, and the two ends of the scale — precision as $\beta \to 0$, recall as $\beta \to \infty$ — are transposes of one another.*

Proof. (a) is the substitution $b = 1$ in (6). (b): for $|A| + |B| > 0$, $(1 + b^{-1})|A \cap B|/(b^{-1}|A| + |B|) = (b + 1)|A \cap B|/(|A| + b|B|) = F_b(B, A)$, and the entry at $(\emptyset, \emptyset)$ is 1 on both sides. Both parts are machine-checked; they are stated because (a) is the control that the definitions used in this section have not drifted from (1), and (b) is used to transfer every statement about precision to recall. □

At $s = 2$ the loss matrix, in the order $\emptyset, \{1\}, \{2\}, \{1, 2\}$, is

$$L_b = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & \dfrac{1}{2+b} \\ 1 & 1 & 0 & \dfrac{1}{2+b} \\ 1 & \dfrac{b}{1+2b} & \dfrac{b}{1+2b} & 0 \end{pmatrix}, \tag{7}$$

which at $b = 1$ is (4); the two off-diagonal blocks are the entries $1-(1+b)/(2+b)$ and $1-(1+b)/(1+2b)$.

The rank along the scale.

Theorem 7.2. *For every $b \in (0, \infty)$,*

$$\operatorname{rank} F_b = \operatorname{rank} L_b = 2 \quad (s = 1), \qquad \operatorname{rank} F_b = \operatorname{rank} L_b = 4 \quad \textit{and} \quad \operatorname{affdim} L_b = 3 \quad (s = 2);$$

that is, $\operatorname{rank} F_\beta = \operatorname{rank} L^{F_\beta} = s^2 - s + 2$ *and* $\operatorname{affdim} L^{F_\beta} = s^2 - s + 1$ *hold for every $\beta \in (0, \infty)$ at $s \le 2$, where $s^2 - s + 2 = 2^s$ and the statement says the score and loss matrices are nonsingular along the whole scale. Since $L_b - J = -F_b$, the rank of $L_b - J$ is that of F_b.*

Proof. A single polynomial left inverse serves every b. With the row scaling $\Lambda_b(A) = \prod_{l=1}^{s}(b|A| + l)$, which is positive for $b > 0$, the matrix $H(b) = \operatorname{diag}(\Lambda_b)\, M(b)$ ($M = F$ or L) has polynomial entries, and $X(b) = \operatorname{adj}(H(b)) \operatorname{diag}(\Lambda_b)$ satisfies $X(b)\, M(b) = c(b)\, I$ with $c(b) = \det H(b)$ up to sign. What is checked is the identity $X(b)M(b) = c(b)I$ as N^2 identities of rational functions of b, and that $c(b) > 0$ for every $b > 0$; the latter is immediate because every coefficient of each $c(b)$ produced is positive. At $s = 1$, $c(b) = 1 + b$ for both matrices. At $s = 2$,

$$c_F(b) = 8b + 28b^2 + 36b^3 + 20b^4 + 4b^5 = 4b\,(b+1)^3(b+2), \qquad c_L(b) = 3\,c_F(b) = 12b\,(b+1)^3(b+2),$$

consistent with the closed form $\det F_b = b/((2 + b)(1 + 2b))$ that one reads off (7). A left inverse up to a nonzero scalar makes $M(b)$ invertible, so its rank is N; and a nonsingular $N \times N$ matrix has affine dimension $N - 1$, which is 3 at $s = 2$. The certificates were produced by exact adjugate computation over $\mathbb{Q}(b)$ outside the verification and are re-checked inside it; no property of the generator is used. □

Two remarks on what the theorem does and does not say. The hypothesis $b > 0$ is not decorative: at $b = 0$ the formula (6) is the precision score of Theorem 7.3, of rank $s + 1 < 2^s$ at $s = 2$, so the statement is false at the endpoint, and $c_F(b)$ vanishes there. And the rank $s^2 - s + 2$ is not a generic-rank statement about the function field: it holds at every $b > 0$, which is what a bound on

$\mathrm{CCdim}(L^{F_\beta})$ needs. The published general-β statement, $\mathrm{rank}(L^{F_\beta}-J) \le s^2+1$ of [4, Proposition 3], is weaker by $s-1$ for every $s \ge 2$; at $s=2$ it says ≤ 5 for a 4×4 matrix. The first s at which $s^2-s+2 < 2^s$, i.e. at which F_b is genuinely singular, is $s=4$ ($14<16$); the uniform-in-β certificates at $s=3$ and $s=4$ are computed but not checked. For general s the prose proof of Section 8 gives $\mathrm{rank}\, F_b = \mathrm{rank}(L_b - J)$ only; the loss rank and the affine dimension are computed there at $s \le 5$.

The endpoints. For a fixed pair (A,B) the entry (6) has limits at both ends of the scale: $b \to 0^+$ gives $|A\cap B|/|B|$ and $b\to\infty$ gives $|A\cap B|/|A|$, with the constant value 1 at $(\emptyset,\emptyset)$ and the constant value 0 at $(A,\emptyset)$, $A\ne\emptyset$, and at $(\emptyset,B)$, $B\ne\emptyset$. (This is elementary and is not formalised; the two limit matrices are defined directly.) Let

$$P(A,B) = \begin{cases} 1, & A=B=\emptyset,\\ 0, & B=\emptyset\ne A,\\ |A\cap B|/|B|, & B\ne\emptyset, \end{cases} \qquad R(A,B)=P(B,A),$$

the classical precision and recall score matrices, and $J-P$, $J-R$ the corresponding losses; $R=P^{\mathsf T}$ is Proposition 7.1(b) at the endpoints.

Theorem 7.3. *For every $s \ge 0$, $\mathrm{rank}\, P = \mathrm{rank}\, R = s+1$. For every $s\ge 1$, $\mathrm{rank}(J-P) = \mathrm{rank}(J-R) = s+1$ and $\mathrm{affdim}(J-P)=s$; hence, by* [2, Theorem 12], $\mathrm{CCdim}(L_s^{\mathrm{prec}}) \le s$.

Proof. Proved for all s at once, with no case split on s. Index $\mathbb{R}^{s+1}$ by the labels $1,\dots,s$ and one extra symbol e. Put $A'[X,j] = \mathbf{1}[j\in X]$, $A'[X,e]=\mathbf{1}[X=\emptyset]$ and $B'[j,Y]=\mathbf{1}[j\in Y]/|Y|$, $B'[e,Y]=\mathbf{1}[Y=\emptyset]$ (with $B'[j,\emptyset]=0$). Then $P=A'B'$, so $\mathrm{rank}\,P\le s+1$; and the submatrix of P on the rows and columns $\emptyset,\{1\},\dots,\{s\}$ is the identity I_{s+1}, so $\mathrm{rank}\,P\ge s+1$. Every column of B' sums to 1, so $J-P=(\mathbf{1}-A')B'$ also factors through $\mathbb{R}^{s+1}$, and on the same $s+1$ indices it is $J_{s+1}-I_{s+1}$, whose inverse is $J_{s+1}/s - I_{s+1}$ for $s\ge1$; hence $\mathrm{rank}(J-P)=s+1$. For the affine dimension, the all-ones row is $(1,\dots,1)\,B'$ by the same column sums, so $\left[\begin{smallmatrix} J-P\\ \mathbf{1}^{\mathsf T}\end{smallmatrix}\right]$ factors through $\mathbb{R}^{s+1}$ with the left factor $\left[\begin{smallmatrix} \mathbf{1}-A'\\ \mathbf{1}^{\mathsf T}\end{smallmatrix}\right]$ and contains the same invertible $(s+1)\times(s+1)$ submatrix, so its rank is $s+1$ and $\mathrm{affdim}(J-P)=s$. The statements for R and $J-R$ follow by transposition, which preserves rank. The upper bound on CCdim is [2, Theorem 12] applied to $\mathrm{affdim}(J-P)=s$. □

The point of the theorem is the contrast with Theorem 7.2: the rank of L^{F_β} is s^2-s+2 at every interior β — proved at $s\le 2$, computed at $s\le5$ — and $s+1$ at both ends for every s, so it is discontinuous at both endpoints of the scale, and the convex-calibration upper bound of [2, Theorem 12] is quadratic along the scale and linear at its ends. The individual bounds are short; a referee may reasonably call each of them routine, and we found no statement of them in the literature searched (Section 9). The affine dimension is *not* symmetric between the ends: $\mathrm{affdim}(J-R)=s+1$ for $s=2,\dots,7$, computed exactly but not formalised, because $\mathbf{1}^{\mathsf T}$ lies in the row space of $J-P$ and not in that of $J-R$ (at $s=1$, $P=R$ and both affine dimensions are 1; Section 8).

	$\mathrm{rank}\, L_s^{F_\beta}$	$\mathrm{affdim}\, L_s^{F_\beta}$	status
$\beta\to 0$ (precision)	$s+1$	s	every s, formal (Theorem 7.3)
$\beta\in(0,\infty)$	s^2-s+2	s^2-s+1	$s\le2$, every β: formal (Theorem 7.2); $\beta=1$, every s: [1]; every s and β: prose (Section 8), for $\mathrm{rank}\, F_\beta = \mathrm{rank}(L^{F_\beta}-J)$ only
$\beta\to\infty$ (recall)	$s+1$	$s+1$ ($s\ge2$)	rank: every s, formal; affdim: computed at $s\le 7$ (it is 1 at $s=1$, where $P=R$)

Two labels, for every β. At $s=2$, Theorem 7.2 gives $\mathrm{affdim}\, L_b = 3 = N-1$ for every b, so $\mathrm{CCdim}(L_2^{F_\beta}) \le 3$ along the whole scale. The source's Theorem 5.1 assumes $s\ge3$ and treats $\beta=1$, so it says nothing here for any β; Theorem 4.1 is the case $b=1$ of the following.

Theorem 7.4. *For $b \in (0,\infty)$ let*

$$p(b) \;=\; \frac{1}{D(b)}\big(5+20b+18b^2+4b^3,\;\; 2+15b+16b^2+4b^3,\;\; 2+7b+2b^2,\;\; 3+8b+4b^2\big),$$
$$D(b) \;=\; 12+50b+40b^2+8b^3,$$

on the outcomes $\emptyset, \{1\}, \{2\}, \{1,2\}$. Then for every $b>0$: $p(b)$ is strictly positive with $\mathbf{1}^\mathsf{T} p(b) = 1$; its Bayes-optimal reports under L_b are exactly $\emptyset, \{1\}, \{1,2\}$, of common risk $(7+30b+22b^2+4b^3)/D(b)$, while the report $\{2\}$ has risk $(7+38b+36b^2+8b^3)/D(b)$, larger by $(8b+14b^2+4b^3)/D(b) > 0$; and the matrix of constraints of $Q_\emptyset$ active at $p(b)$ has nullity $\mu = 1$. Consequently (5) *gives*

$$\mathrm{CCdim}(L_2^{F_\beta}) \;\ge\; \|p(b)\|_0 - \mu - 1 \;=\; 4-1-1 \;=\; 2 \qquad \textit{for every } \beta \in (0,\infty).$$

Proof. Positivity and the sum are read off the coefficients: every coefficient of every numerator and of D is positive, and the four numerators add up to $D(b)$. The risks are read off (7): $R(\emptyset) = p_{\{1\}} + p_{\{2\}} + p_{\{1,2\}}$; for $R(\{1\})$ and $R(\{2\})$ note $3+8b+4b^2 = (1+2b)(3+2b)$, so $p_{\{1,2\}} \cdot b/(1+2b) = (3b+2b^2)/D(b)$; for $R(\{1,2\})$ note $4+22b+18b^2+4b^3 = (2+b)(2+10b+4b^2)$, so $(p_{\{1\}}+p_{\{2\}})/(2+b) = (2+10b+4b^2)/D(b)$. Adding gives the three equal risks and the fourth, larger one. The witness has full support, so the active matrix of $Q_\emptyset$ is the 3×4 matrix with rows $\ell_{\{1\}} - \ell_\emptyset$, $\ell_{\{1,2\}} - \ell_\emptyset$ and $\mathbf{1}^\mathsf{T}$; its columns at $\emptyset, \{1\}, \{2\}$, scaled by the polynomials 2, $2+3b+b^2$, $2+3b+b^2$, form a 3×3 polynomial matrix with an explicit polynomial left inverse up to the scalar $12+38b+44b^2+22b^3+4b^4 > 0$, so the active matrix has rank 3 and nullity $4-3=1$ for every $b>0$; (5) applies with $B_0 = \emptyset$. All of this is machine-checked uniformly in b. ☐

At $b=1$ the witness is $\frac{1}{110}(47,37,11,15)$, a different point from the $\frac{1}{21}(9,7,2,3)$ of Theorem 4.1: the tie system for the reports $\emptyset, \{1\}, \{1,2\}$ has a one-dimensional solution line inside the simplex, and the point above was chosen on that line, by solving the tie system over $\mathbb{Q}(b)$ and shifting along the line, so that all four coordinates have positive coefficients and hence stay positive on the whole of $(0,\infty)$. Before the witness was written down, the exhaustive search of Section 8, run on L_b at $b = 1/10, 1/4, 1/2, 1, 2, 4, 10$, returned the value 2 each time.

Theorem 7.5. *For every $b \in (0,\infty)$ there is no $p \in \mathbb{R}^4$ with $p \ge 0$ and $\mathbf{1}^\mathsf{T} p = 1$ under which all four F_β reports have the same conditional risk: the three tie equations already force $p_\emptyset(1+2b) + p_{\{1,2\}} = 0$, so a nonnegative solution vanishes identically. The unique p with $\mathbf{1}^\mathsf{T} p = 1$ solving the tie system is*

$$p \;=\; \Big(\tfrac{1}{4},\;\; \tfrac{2+b}{4},\;\; \tfrac{2+b}{4},\;\; -\tfrac{1+2b}{4}\Big),$$

whose last coordinate is negative for every $b>0$. Consequently the sufficient condition of [2, Theorem 18] *is unavailable at $s=2$ for every β, so is the hypothesis of* [2, Corollary 19]*, and no application of* (5) *can exceed the value 2 at $s=2$ for any β.*

Proof. From (7), $R(\{1\}) = R(\{2\})$ forces $p_{\{1\}} = p_{\{2\}}$; $R(\emptyset) = R(\{1\})$ reads $p_{\{1\}} + p_{\{1,2\}}(1+b)/(1+2b) = p_\emptyset$; and $R(\emptyset) = R(\{1,2\})$ reads $2p_{\{1\}}(1+b)/(2+b) + p_{\{1,2\}} = p_\emptyset$. Eliminating $p_\emptyset$ gives $p_{\{1\}}\, b/(2+b) = -p_{\{1,2\}}\, b/(1+2b)$, i.e. $p_{\{1,2\}} = -p_{\{1\}}(1+2b)/(2+b)$, and then $p_\emptyset = p_{\{1\}}/(2+b)$, which is the relation $p_\emptyset(1+2b) + p_{\{1,2\}} = 0$; normalising, $p_{\{1\}} \cdot 4/(2+b) = 1$ gives the displayed vector, at which all four risks equal $3/4$. Uniqueness is the invertibility of L_b (Theorem 7.2), exactly as in Proposition 4.2, of which this is the case $b=1$. The machine-checked statement is the nonexistence of a nonnegative tie: from the sixteen entries of (7), clearing denominators, a linear combination of the three tie equations gives $b\,(p_\emptyset(1+2b) + p_{\{1,2\}}) = 0$, hence $p_\emptyset = p_{\{1,2\}} = 0$ for $p \ge 0$, hence $p_{\{1\}} = p_{\{2\}} = 0$, contradicting $\mathbf{1}^\mathsf{T} p = 1$; this covers every $p \ge 0$, not only the relative interior. The final sentence is the argument of Proposition 4.2: a value of 3 from (5) would need full support and four tied reports, and $N = 4$. ☐

Corollary 7.6. *For every $\beta \in (0,\infty)$, $\mathrm{CCdim}(L_2^{F_\beta}) \in \{2,3\}$, and no application of* [2, Theorem 16] *decides which.*

Proof. Theorem 7.4 from below; $\min\{N-1, \operatorname{affdim} L_b\} = 3$ from above by [2, Lemma 11 and Theorem 12] with Theorem 7.2; Theorem 7.5 caps every bound obtainable from (5) at 2. ☐

So the whole of Section 4 — the lower bound, the obstruction [3] name, and the exhaustion of the feasible-subspace method — is a statement about the entire F_β scale. At three labels the picture changes with β: the exhaustive computation recorded in Section 8 finds a threshold at $\beta^2 = 1/2$ in the value of the feasible-subspace bound, with $\mathrm{CCdim}(L_3^{F_\beta}) \geq 4$ below it; those statements are computed over $\mathbb{Q}(b)$ and are not part of the formal development.

8. Outside the formal development

Everything in this section was computed in exact rational arithmetic, or (in the second F_β subsection) proved in prose, and is *not* machine-checked; each statement says what was run or what was argued.

The exhaustive value of the feasible-subspace bound. By Proposition 2.1, $\mathrm{fsd}(L)$ depends on p only through the pair $(\operatorname{supp} p, \mathrm{opt}_L(p))$, of which there are at most $2^N \times 2^N$; so $\mathrm{fsd}(L)$ is computable by enumerating those pairs, computing the affine dimension exactly, and testing each pair for strict feasibility — that is, for the existence of a p with exactly that support and exactly that set of Bayes-optimal reports. We did this in exact rational arithmetic (an equality solve first, a two-phase simplex over $\mathbb{Q}$ only when the tie system is positive-dimensional), for the F_1 loss at $s = 2, 3$ under both empty-set conventions and with and without the restriction that the support avoid the empty outcome. The results:

loss	fsd	source's finite bound	upper bound
$L_2^{F_1}$, convention (1)	**2**	— (Thm. 5.1 needs $s \geq 3$)	3
$L_3^{F_1}$, convention (1)	**3**	$hn = 2$	7
$L_2^{F_1}$, convention $F(\emptyset,\emptyset) = 0$	≤ 1	—	3
$L_2^{F_1}$, support avoiding $\emptyset$	≤ 1	—	3
$L_3^{F_1}$, support avoiding $\emptyset$	2	$hn = 2$	7
$L_3^{F_1}$, convention $F(\emptyset,\emptyset) = 0$	≤ 2	$hn = 2$	7
$L_n^{0\text{-}1}$, $n = 3, 4, 5$	$n-1$	Ex. 11: $\mathrm{CCdim} = n-1$	$n-1$

The last line is the control: it reproduces the published $\mathrm{CCdim}(L_n^{0\text{-}1}) = n - 1$ of [2, Example 11], and it is the check that the search computes what we think it computes. (An earlier version of the exact simplex minimised where it should have maximised; the control returned 1 for every n and stopped the run before any claim was made.) The two rows in bold are the exhaustive maxima matching Theorems 4.1 and 5.1 above: at $s = 3$ all $9 + 282 + 2658 + 10690 = 13{,}639$ pairs of affine dimension at least 4 are certified infeasible, in 112 seconds. The entries recorded as inequalities are upper bounds only: at $s = 2$ they are what the scan certifies, and the $s = 3$ row under $F(\emptyset,\emptyset) = 0$ was stopped by a 400-second budget after it had cleared all values ≥ 3, so only ≤ 2 is claimed there.

Two readings of the table matter.

Remark 8.1 (The method is exhausted at two labels)*.* $\mathrm{fsd}(L_2^{F_1}) = 2$: the bound (5) cannot give 3 at $s = 2$ under any choice of p. The kernel-checked half of this is Proposition 4.2, which is all that is needed since $N = 4$; the scan agrees. So deciding $\mathrm{CCdim}(L_2^{F_1}) \in \{2, 3\}$ requires either an explicit 2-dimensional convex calibrated surrogate or a lower-bound technique beyond [2, Theorem 16].

Remark 8.2 (The gain at three labels is exactly the empty-set convention)*.* Restricting the search at $s = 3$ to witnesses with $p_\emptyset = 0$ — which is what the construction of [1, Theorem 5.1] does, and is why that source can call its bound convention-independent — returns exactly $hn = 2$, attained at $p = \frac{1}{41}(15, 15, 11)$ on $\{1\}, \{1,2\}, \{3\}$; and the whole search under the convention $F(\emptyset,\emptyset) = 0$ returns at most 2. So [1, Theorem 5.1] is tight at $s = 3$ inside its own witness class, and the improvement to 3 of Section 5 is not available to any argument that is indifferent to the value of $F(\emptyset,\emptyset)$. Both directions are sharper statements about that theorem than the source makes.

The two-label trigger fan is not a pullback from the plane. The following obstruction is worth recording for whoever attacks $\mathrm{CCdim}(L_2^{F_1}) \in \{2,3\}$ from above. It was verified by exact linear programming over $\mathbb{Q}$ — each of the three inequalities $p^\mathsf{T}(\ell_B - \ell_{B_0}) \geq 0$ defining $Q_{B_0} \cap \Delta_4$ was dropped in turn and its violation maximised over the rest — and is not machine-checked. For $B_0 = \emptyset$ all three inequalities are irredundant (each violable by 1/5 when dropped), and for $B_0 = \{1,2\}$ all three are irredundant (by 1, 1/3, 1/3); for $B_0 = \{1\}$ and $B_0 = \{2\}$ one is redundant. A pullback $T^{-1}(C)$ of a 2-dimensional convex cone $C \subseteq \mathbb{R}^2$ along a linear map $T : \mathbb{R}^4 \to \mathbb{R}^2$ is an intersection of at most two halfspaces, so it can contribute at most two irredundant non-simplex facets. Hence no linear $T : \mathbb{R}^4 \to \mathbb{R}^2$ and fan in $\mathbb{R}^2$ reproduce the F_1 trigger partition at $s = 2$; in particular the calibrated-surrogate constructions whose positive-normal sets are pullbacks of a planar normal fan — which includes the affine construction behind [2, Theorem 12], and any affine $\psi : \mathcal{C} \subseteq \mathbb{R}^2 \to \mathbb{R}^4_+$ — cannot achieve $d = 2$. A genuinely nonlinear convex ψ is not excluded by this argument, and that is the open half of the question.

The F_β rank for every s: a prose proof. The following is the general form of Theorem 7.2. It is proved here on paper and is *not* formalised; the machine-checked instances are $s \leq 2$ uniformly in β (Theorem 7.2) and $\beta = 1$ for every s ([1, Theorem 4.3], re-certified at $s \leq 4$ in Theorem 3.1). Before it was written down, the statement was checked in exact arithmetic at every $s \leq 5$ for the ten values $b = 1/1000, 1/9, 1/4, 1/2, 7/11, 1, 2, 4, 9, 100$ (each giving $\operatorname{rank} F_b = \operatorname{rank} L_b = s^2 - s + 2$ and $\operatorname{affdim} L_b = s^2 - s + 1$), and symbolically over the function field $\mathbb{Q}(b)$ at $s \leq 4$ (rank $2, 4, 8, 14$).

Remark 8.3 ($\operatorname{rank} F_\beta = s^2 - s + 2$ for every $s \geq 1$ and every $\beta \in (0, \infty)$; prose proof, not machine-checked). Write $k_X = |X|$. Index a set ι by the pairs $(k,j) \in \{1, \dots, s-1\} \times [s]$ together with two extra symbols m and e, so $|\iota| = s(s-1) + 2 = s^2 - s + 2$, and define $\mathsf{A} \in \mathbb{R}^{2^s \times \iota}$ and $\mathsf{B}(b) \in \mathbb{R}^{\iota \times 2^s}$ by

$$\mathsf{A}[X,(k,j)] = \mathbf{1}[k_X = k]\,\mathbf{1}[j \in X], \qquad \mathsf{A}[X,m] = \mathbf{1}[k_X = s], \qquad \mathsf{A}[X,e] = \mathbf{1}[X = \emptyset],$$

$$\mathsf{B}(b)[(k,j),Y] = \frac{(1+b)\,\mathbf{1}[j \in Y]}{bk + k_Y}, \qquad \mathsf{B}(b)[m,Y] = \frac{(1+b)\,k_Y}{bs + k_Y}, \qquad \mathsf{B}(b)[e,Y] = \mathbf{1}[Y = \emptyset].$$

Then $F_b = \mathsf{A}\,\mathsf{B}(b)$: for $1 \leq k_X \leq s-1$ only the block $k = k_X$ contributes, giving $(1+b)|X \cap Y|/(bk_X + k_Y)$; for $X = [s]$ the column m gives $(1+b)k_Y/(bs + k_Y)$ and $X \cap Y = Y$; for $X = \emptyset$ only the column e survives and gives $\mathbf{1}[Y = \emptyset]$, which is the convention. Hence $\operatorname{rank} F_b \leq s^2 - s + 2$ for every b. The left factor A does not depend on b and has full column rank: columns with different k have disjoint supports; inside a layer $1 \leq k \leq s-1$ the columns are those of the subset-incidence matrix of k-subsets against labels, of rank s ([1, Lemma 4.1]); and the two extra columns are supported on the single rows $[s]$ and $\emptyset$. Therefore $\operatorname{rank} F_b = \operatorname{rank} \mathsf{B}(b)$, and it remains to show that $\mathsf{B}(b)$ has full row rank. Let c be a row combination vanishing on every column. The column $Y = \emptyset$ gives $c_e = 0$. For $k_Y = l \geq 1$, dividing by $1 + b$,

$$\sum_{j \in Y} d_j^{(l)} + c_m \frac{l}{bs + l} = 0, \qquad d_j^{(l)} := \sum_{k \leq s-1} \frac{c_{k,j}}{bk + l}.$$

For $l \leq s-1$ one may exchange a single element between two l-subsets, so $d_j^{(l)}$ does not depend on j; comparing two labels $j \neq j'$ gives $\sum_{k \leq s-1} (c_{k,j} - c_{k,j'})/(bk + l) = 0$ for $l = 1, \dots, s-1$, a Cauchy system with nodes bk (distinct since $b > 0$) and l, hence nonsingular, so $c_{k,j} = c_k$ does not depend on j. Substituting back, for $l = 1, \dots, s-1$, $\sum_{k \leq s-1} c_k/(bk + l) + c_m/(bs + l) = 0$, and the single column $Y = [s]$ gives the same equation at $l = s$; with $\gamma_k = c_k$ ($k \leq s-1$), $\gamma_s = c_m$ this is the $s \times s$ Cauchy system $\sum_k \gamma_k/(bk + l) = 0$, $l = 1, \dots, s$, again nonsingular. So $c = 0$, and $\operatorname{rank} F_b = s^2 - s + 2$; $\operatorname{rank}(L_b - J) = \operatorname{rank} F_b$ since $L_b - J = -F_b$.

At $b = 1$ this is [1, Theorem 4.3], proved differently: the source factors the nonempty block as $P(C \otimes I_s)P^\mathsf{T}$ with the symmetric Cauchy matrix $C_{kk'} = 2/(k + k')$ and uses its positive-definiteness through the identity $\operatorname{rank}(PMP^\mathsf{T}) = \operatorname{rank} P$ for $M \succ 0$. For $b \neq 1$ the middle matrix $(1+b)/(bk+l)$

is not symmetric, and no diagonal rescaling of its rows and columns makes it so (the consistency condition already fails at $b = 3$ on the sizes $1, 2, 5$), so that identity is unavailable and the two-sided Cauchy argument above replaces it. That is the mathematical content of the extension: the number is the same, the route is not. We do not claim $\operatorname{rank} L_b = s^2 - s + 2$ or $\operatorname{affdim} L_b = s^2 - s + 1$ for general s from this remark; those were checked at $s \le 5$ as stated above and are proved at $s \le 2$ by Theorem 7.2.

Two computations bear on the endpoints of Theorem 7.3 and are recorded here because they are not formalised. The drop of the rank happens exactly at the endpoint and nowhere inside: at $s = 2$, $\det F_b = b/((2 + b)(1 + 2b))$, whose only zero is $b = 0$, and at $s = 3$ the row-scaled determinant $\det H(b) = 23328\, b^5 + 209952\, b^6 + \cdots + 1152\, b^{16}$ has all twelve coefficients positive (its exact Sturm count of roots in $(0, \infty)$ is 0), so it never vanishes for $b > 0$. And the affine dimensions at the two ends differ: $\operatorname{affdim}(J - P) = s$ (formal) but $\operatorname{affdim}(J - R) = s + 1$ for $s = 2, \dots, 7$, computed exactly (at $s = 1$ the two matrices coincide); under the other convention $F(\emptyset, \emptyset) = 0$ the endpoint score ranks drop again, from $s + 1$ to s (computed at $s \le 6$), parallel to Theorem 3.2 in the interior.

At three labels the feasible-subspace bound depends on β. The exhaustive search of this section, run on L_b at $s = 3$, returns

$b = \beta^2$	1/10	1/4	49/100	1/2	51/100	1	2	4	10
$\operatorname{fsd}(L_2^{F_\beta})$	2	2	—	2	—	2	2	2	2
$\operatorname{fsd}(L_3^{F_\beta})$	**4**	**4**	**4**	3	3	3	3	3	3

(at $b = 1/2$ all 10,690 candidate pairs of value 4 are certified infeasible, at $b = 51/100$ all 10,672; the $s = 2$ row is the computation mentioned after Theorem 7.4). So the best bound that [2, Theorem 16] can give at $s = 3$ is not a β-independent quantity, in contrast to $s = 2$. Both sides of the jump are settled symbolically over $\mathbb{Q}(b)$, not only sampled, and the threshold is exactly $b = 1/2$.

Remark 8.4 (The threshold $\beta^2 = 1/2$; exact computation over $\mathbb{Q}(b)$, not machine-checked). *Below the threshold.* On the support consisting of $\emptyset$, $\{1\}$, $\{1, 2\}$, $\{3\}$, $\{2, 3\}$, with the five reports $\emptyset$, $\{1\}$, $\{1, 2\}$, $\{1, 3\}$, $\{1, 2, 3\}$ tied, the tie system together with $\mathbf{1}^\mathsf{T} p = 1$ has the unique solution

$$p(b) = \frac{1}{18 + 42b + 26b^2 + 4b^3} \big(6 + 15b + 11b^2 + 2b^3,\ \ 2 + 7b + 7b^2 + 2b^3,\ \ 0,\ \ 4 + 12b + 8b^2,\ \ 4 + 10b + 4b^2,\ \ 0,\ \ 2 - 2b - 4b^2,\ \ 0\big)$$

in the order of Section 2. Every numerator except the seventh has only positive coefficients; the seventh is $2(1 - 2b)(1 + b)$, positive exactly for $b < 1/2$. The three other reports are strictly worse for every $b > 0$ (their gaps have all-positive numerators), and the 8×8 active-constraint matrix — three nonnegativity rows, four tie rows, normalisation — is nonsingular over $\mathbb{Q}(b)$, with determinant numerator $-(b/4)(3 + 5b + b^2)$, which has no positive root. So $\|p\|_0 = 5$, $\mu = 0$, and (5) gives $\operatorname{CCdim}(L_3^{F_\beta}) \ge 5 - 0 - 1 = 4$ for every β with $\beta^2 < 1/2$ — twice the value $hn = 2$ of [1, Theorem 5.1] at $s = 3$, which is a $\beta = 1$ statement.

At and above the threshold. On the support of Theorem 5.1, $\{\emptyset, \{1\}, \{1, 2\}, \{3\}\}$, with the reports $\{\emptyset, \{1\}, \{1, 2\}, \{1, 2, 3\}\}$, the unique solution is

$$p(b) = \frac{1}{9 + 41b + 52b^2 + 26b^3 + 4b^4} \big(3 + 14b + 20b^2 + 11b^3 + 2b^4,\ \ 6b + 13b^2 + 9b^3 + 2b^4,\ \ 0,\ \ 3 + 11b + 12b^2 + 4b^3,\ \ 3 + 10b + 7b^2 + 2b^3,\ \ 0,\ \ 0,\ \ 0\big),$$

all coefficients nonnegative, which at $b = 1$ is $\frac{1}{66}(25, 15, 0, 15, 11, 0, 0, 0)$. Its only obstruction is the report $\{1, 3\}$, whose gap is $(b/8)(2b - 1)(b + 1)$ over a positive denominator: nonnegative exactly for $b \ge 1/2$. Its active matrix is nonsingular for every $b > 0$, so $\mu = 0$, $\|p\|_0 = 4$ and $\operatorname{CCdim}(L_3^{F_\beta}) \ge 3$ there. At $b = 1/2$ the two witnesses meet: the fifth coordinate of the first vanishes exactly where the fifth tie of the second appears. Hence $\operatorname{fsd}(L_3^{F_\beta}) \ge 4$ for $b < 1/2$ and ≥ 3 for $b \ge 1/2$, for every b; the matching upper bounds are the exhaustive scan at the nine values tabulated above, not a statement for every b. (The two witnesses' $\mu = 0$ certificates over $\mathbb{Q}(b)$ — 8×8 adjugates of degree at most 28 with coefficients up to 2.7×10^{11} — are computed and are what a formalisation would check; see Section 10.)

9. Formal verification

Theorems 3.1, 3.2, 4.1, 5.1 and 6.1 and Proposition 4.2, and in this version Proposition 7.1 and Theorems 7.2, 7.3, 7.4 and 7.5, are formalised and machine-checked in Lean 4 [7] against *Mathlib* [8], in six modules totalling 2,149 lines and 418 declarations, most of the latter being certificate data (the three modules of version 1 are 1,161 lines and 282 declarations; the three added here 988 lines and 136 declarations). 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 [2, Theorem 16], quoted. What is machine-checked is the finite content of each application: the F_1 score and loss matrices exactly as in (1), over $\mathbb{Q}$ and with no per-entry numerical assertion (the integer rescaling by $\Lambda = \mathrm{lcm}(1, \dots, 2s)$ is proved to agree with the rational definition for all s, A, B); the two general certificate lemmas of Section 3; the rank and affine-dimension values of Theorems 3.1 and 3.2; and, for each witness, that it is nonnegative and sums to 1, exactly which reports tie and which are strictly worse, and the nullity μ of the active-constraint matrix, from which $\|p\|_0 - \mu - 1$ is the stated integer. Proposition 4.2 is checked in the strong form quantified over all $p \geq 0$ with $\mathbf{1}^\mathsf{T} p = 1$. Negative controls are part of the development: that the two-label nullity is neither 0 nor 2 (so the bound there is neither 3 nor 1), that the three-label witness really has proper support (so no full-support hypothesis is silently in play), and the rank perturbation of Theorem 3.2.

For the F_β scale, the score (6) is defined over $\mathbb{Q}$ in the variable b exactly as displayed, with the value 1 at $(\emptyset, \emptyset)$, and every theorem of Section 7 is a single statement quantified over all rational $b > 0$: Proposition 7.1(a) is the equality of the two definitions at $b = 1$, so the F_β definitions cannot have drifted from (1); Theorem 7.2 is the polynomial left-inverse identity $X(b)M(b) = c(b)I$, entry by entry, with the positivity of $c(b)$; Theorem 7.3 is the factorisation through $\mathbb{R}^{s+1}$, the identity submatrix, and the inverse $J/s - I$, for a variable s with no case split and no finite-case evaluation; Theorem 7.4 is the positivity and the sum of the witness, the two equalities and the strict inequality among its risks as rational functions of b, and the nullity 1 from a polynomial certificate on a column-scaled submatrix; Theorem 7.5 is the nonexistence of a nonnegative total tie, in the closed form. The precision and recall matrices are defined directly as such (the division by $|B|$ at $B = \emptyset$ returns 0, which is the entrywise limit); the limit statement is not formalised. Appendix A reproduces the fifteen formal statements verbatim. Not formalised, and marked as such above: Proposition 2.1; the numerical risk values displayed in the proofs, as opposed to the equalities and strict inequalities among them; the explicit tie-system solutions of Proposition 4.2 and Theorem 7.5 (the machine-checked statement in each is the nonexistence of a nonnegative tie); the upper bounds quoted from [2]; everything in Section 8; and Remarks 3.3 and 6.2.

The reasoning layer uses no compiled evaluation: axiom sets were printed for the twenty-three declarations that carry the statements of version 1 and for the two nullity computations they use, and again for the fifteen declarations that carry the statements of Section 7 and the nullity computation of Theorem 7.4, 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 and are cheap. Re-measured for this version by `lake env lean` on each module against its already-built dependencies, the largest file of version 1 checks in about eighty seconds of wall time and each of the three modules added here in about ten, forty-five and eighty-five seconds; the figures recorded when the files were first written, on a much more heavily loaded machine, were larger. The repository is private: repository reference to be added at submission.

The literature position of Section 7 was established by the same searches as for version 1, re-run on 3 September 2026 with the F_β phrasings added (a full-text corpus of arXiv through 2 September

2026, OpenAlex forward citations of [1], the three Semantic Scholar citing records, and the OEIS entries for $s^2 - s + 2$ and $s^2 - s + 1$): no paper states an exact rank for the F_β loss matrix at any $\beta \neq 1$, or a convex-calibration-dimension value or bound for it at a specific s; the only general-β rank statement found is [4, Proposition 3].

10. The frontier

Question 10.1. Is $\mathrm{CCdim}(L_2^{F_1})$ equal to 2 or to 3?

This is not a computational question any more. From below, Remark 8.1 says the feasible-subspace bound is exhausted at 2, so a proof of 3 needs a technique that [2, Theorem 16] does not contain. From above, a proof of 2 needs an explicit 2-dimensional convex calibrated surrogate, and Section 8 rules out the linear constructions: the trigger partition at $s = 2$ is not the pullback of any fan in the plane. What is left is a nonlinear convex $\psi : \mathcal{C} \subseteq \mathbb{R}^2 \to \mathbb{R}^4_+$, or a new lower-bound idea. By Corollary 7.6 the same question now stands, in the same terms, for $\mathrm{CCdim}(L_2^{F_\beta})$ at every $\beta \in (0, \infty)$.

Question 10.2. Is $\mathrm{fsd}(L_4^{F_1}) > 3$?

At $s = 4$ the source's Theorem 5.1 gives $hn = 3$ and the upper bound is 13. The enumeration used at $s = 3$ does not scale — $2^{16} \times 2^{16}$ pairs, about 58 CPU hours before a single feasibility test, by the measured $s = 3$ rate — so we did not run it. A structured search (tie sets of a fixed shape, supports confined to one or two cardinality layers) would already answer Question 10.2, and by Remark 8.2 the interesting comparison is with witnesses that do and do not use the empty outcome. More generally, both $\mathrm{fsd}(L_s^{F_1})$ and $\mathrm{CCdim}(L_s^{F_1})$ are unknown for every $s \geq 2$; the present paper closes the small-s window from $[?, 3]$ to $\{2, 3\}$ at $s = 2$ and from $[2, 7]$ to $[3, 7]$ at $s = 3$.

Question 10.3. Is $\mathrm{fsd}(L_s^{F_\beta})$ monotone in β for every s, and does the threshold $\beta^2 = 1/2$ of Remark 8.4 have an analogue at $s = 4$?

At $s = 2$ the bound is 2 for every β (Section 7); at $s = 3$ it is 4 below $\beta^2 = 1/2$ and 3 at the nine sampled values at and above it; and at $s = 4$ the enumeration is the one priced above. The duality of Proposition 7.1(b) transposes the loss, and the feasible-subspace bound is not obviously transposition-invariant, so nothing forces symmetry of $\mathrm{fsd}(L_s^{F_\beta})$ under $\beta \mapsto 1/\beta$; the $s = 3$ values above (4 at $b = 1/4$, 3 at $b = 4$) show it is not symmetric.

Three further directions are routine rather than open. The rank certificates of Section 3 exist as data at $s = 5, 6, 7$ and are small (Remark 3.3); only the verification is expensive in one file, and splitting it by matrix would bring $s = 5$ inside the budget. On the F_β scale, the uniform-in-β certificate at $s = 3$ (the scalar of degree 16 displayed in Section 8, adjugate entries of degree at most 16) is computed and priced at some tens of CPU-minutes, and the one at $s = 4$ — the first rank deficit, a 14×14 adjugate of degree 37 with coefficients to 1.5×10^{18} — at one to three CPU-hours; the two 8×8 nullity certificates of Remark 8.4 would make the β-threshold statement machine-checked at some tens of CPU-minutes each. And both [1, Theorem 4.3] and Remark 8.3 are proved for all s by arguments — a Kronecker factorisation plus positive-definiteness of a Cauchy matrix, or a b-free left factor plus nonsingularity of Cauchy matrices — whose ingredients are all in *Mathlib*, so the general statements are within reach of a direct formalisation rather than a certificate at each s; the upper bound $\mathrm{rank}\, F_b \leq s^2 - s + 2$ of Remark 8.3 is a three-case factorisation and is the cheap half.

Finally, the same machinery applies to the sibling losses. The Jaccard measure has recently been shown to have $2^{s-1} \leq \mathrm{CCdim}(L^{\mathrm{Jac}}) \leq 2^s - 1$ [5, 6], by the same route, with its own factor-of-two gap left open; $\mathrm{fsd}(L_s^{\mathrm{Jac}})$ is computable by the search of Section 8 at $s \leq 3$, and there the answer comes out the other way: that search returns $\mathrm{fsd}(L_2^{\mathrm{Jac}}) = 2$ and $\mathrm{fsd}(L_3^{\mathrm{Jac}}) = 4$, which is exactly the published lower bound 2^{s-1}, so for Jaccard at those s the feasible-subspace route is already exhausted at the source's own witness, whereas at $s = 3$ for F_1 it beats the source's witness by one. (A companion computation, on the same footing as Section 8: exact, exhaustive, and not machine-checked.)

Appendix A. The formal statements of Section 7

The Lean statements corresponding to the results of Section 7, verbatim (statements only; proofs, attributes and documentation strings omitted), preceded by the definitions they use. Subsets of $[s]$ are the natural numbers below 2^s through their binary expansion, `pc s a` is $|A|$, `a &&& c` is $A \cap B$, `score s a c` is (1) and `scoreMat s` its matrix, `affdim M` is $\operatorname{rank}\left[\begin{smallmatrix} M \\ \mathbf{1}^\top \end{smallmatrix}\right] - 1$, `risk M p j` is $\sum_i p_i M_{ij}$, and `muOf A` is the dimension of the kernel of A (the nullity), all from the modules of version 1. In ℚ, `x / 0 = 0`.

Definitions

```
def fbScore (s : ℕ) (b : ℚ) (a c : ℕ) : ℚ :=
  if pc s a + pc s c = 0 then 1
  else (1 + b) * (pc s (a &&& c) : ℚ) / (b * (pc s a : ℚ) + (pc s c : ℚ))
def fbScoreMat (s : ℕ) (b : ℚ) : Matrix (Fin (2 ^ s)) (Fin (2 ^ s)) ℚ := fun i j => fbScore s b i j
def fbLossMat (s : ℕ) (b : ℚ) : Matrix (Fin (2 ^ s)) (Fin (2 ^ s)) ℚ :=
  fun i j => 1 - fbScore s b i j
def precScore (s a c : ℕ) : ℚ :=
  if pc s a + pc s c = 0 then 1 else (pc s (a &&& c) : ℚ) / (pc s c : ℚ)
def recScore (s a c : ℕ) : ℚ :=
  if pc s a + pc s c = 0 then 1 else (pc s (a &&& c) : ℚ) / (pc s a : ℚ)
def precMat (s : ℕ) : Matrix (Fin (2 ^ s)) (Fin (2 ^ s)) ℚ := fun i j => precScore s i j
def recMat (s : ℕ) : Matrix (Fin (2 ^ s)) (Fin (2 ^ s)) ℚ := fun i j => recScore s i j
def precLossMat (s : ℕ) : Matrix (Fin (2 ^ s)) (Fin (2 ^ s)) ℚ := fun i j => 1 - precScore s i j
def recLossMat (s : ℕ) : Matrix (Fin (2 ^ s)) (Fin (2 ^ s)) ℚ := fun i j => 1 - recScore s i j
def wb2 (b : ℚ) : Fin 4 → ℚ := fun i =>
  (if (i : ℕ) = 0 then 5 + 20 * b + 18 * b ^ 2 + 4 * b ^ 3
   else if (i : ℕ) = 1 then 2 + 15 * b + 16 * b ^ 2 + 4 * b ^ 3
   else if (i : ℕ) = 2 then 2 + 7 * b + 2 * b ^ 2
   else 3 + 8 * b + 4 * b ^ 2) / (12 + 50 * b + 40 * b ^ 2 + 8 * b ^ 3)
def actb2 (b : ℚ) : Matrix (Fin 3) (Fin 4) ℚ := fun i j =>
  if (i : ℕ) = 0 then fbLossMat 2 b j 1 - fbLossMat 2 b j 0
  else if (i : ℕ) = 1 then fbLossMat 2 b j 3 - fbLossMat 2 b j 0
  else 1
```

Proposition 7.1

```
theorem fbScore_one (s a c : ℕ) : fbScore s 1 a c = score s a c
theorem fbScoreMat_one (s : ℕ) : fbScoreMat s 1 = scoreMat s
theorem fbScoreMat_inv (s : ℕ) (b : ℚ) (hb : 0 < b) :
    fbScoreMat s b⁻¹ = (fbScoreMat s b)ᵀ
```

Theorem 7.2

```
theorem fbScoreMat_rank_1 (b : ℚ) (hb : 0 < b) : (fbScoreMat 1 b).rank = 2
theorem fbLossMat_rank_1 (b : ℚ) (hb : 0 < b) : (fbLossMat 1 b).rank = 2
theorem fbScoreMat_rank_2 (b : ℚ) (hb : 0 < b) : (fbScoreMat 2 b).rank = 4
theorem fbLossMat_rank_2 (b : ℚ) (hb : 0 < b) : (fbLossMat 2 b).rank = 4
theorem fbLossMat_affdim_2 (b : ℚ) (hb : 0 < b) : affdim (fbLossMat 2 b) = 3
```

Theorem 7.3

```
theorem precMat_rank (s : ℕ) : (precMat s).rank = s + 1
theorem recMat_rank (s : ℕ) : (recMat s).rank = s + 1
theorem precLossMat_rank (s : ℕ) (hs : 1 ≤ s) : (precLossMat s).rank = s + 1
theorem recLossMat_rank (s : ℕ) (hs : 1 ≤ s) : (recLossMat s).rank = s + 1
theorem precLossMat_affdim (s : ℕ) (hs : 1 ≤ s) : affdim (precLossMat s) = s
```

Theorem 7.4

```
theorem ccdim_witness_two_beta (b : ℚ) (hb : 0 < b) :
    (∀ i, 0 < wb2 b i) ∧ (∑ i, wb2 b i = 1) ∧
    (risk (fbLossMat 2 b) (wb2 b) 0 = risk (fbLossMat 2 b) (wb2 b) 1) ∧
    (risk (fbLossMat 2 b) (wb2 b) 0 = risk (fbLossMat 2 b) (wb2 b) 3) ∧
    (risk (fbLossMat 2 b) (wb2 b) 0 < risk (fbLossMat 2 b) (wb2 b) 2) ∧
    muOf (actb2 b) = 1 ∧ 4 - muOf (actb2 b) - 1 = 2
```

Theorem 7.5

```
theorem no_total_tie_two_beta (b : ℚ) (hb : 0 < b) (p : Fin 4 → ℚ) (hp : ∀ i, 0 ≤ p i)
    (hs : ∑ i, p i = 1) :
    ¬ (risk (fbLossMat 2 b) p 0 = risk (fbLossMat 2 b) p 1 ∧
       risk (fbLossMat 2 b) p 0 = risk (fbLossMat 2 b) p 2 ∧
       risk (fbLossMat 2 b) p 0 = risk (fbLossMat 2 b) p 3)
```

References

[1] M. Zhang, *Exact Rank and Convex Calibration Dimension Lower Bounds for the Multi-Label F_1 Loss*, arXiv: 2608.08399v1 (submitted 9 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, read from the arXiv v1 PDF and confirmed by compiling its own e-print: its equation (1) is (1) above; its Theorem 4.3 is (2), its Corollary 4.4 the upper bound $s^2 - s + 1$, its Remark 4.5 the perturbation of Theorem 3.2, and its Theorem 5.1 the lower bound (3); and its Lemma 4.1, "Rank of subset-incidence matrices", gives $\operatorname{rank}(W_k) = s$ for $1 \le k \le s-1$ and $\operatorname{rank}(W_s) = 1$, where $(W_k)_{A,j} = \mathbf{1}[j \in A]$ on the k-subsets A of $[s]$ — the fact used in Remark 8.3.

[2] H. G. Ramaswamy and S. Agarwal, *Convex Calibration Dimension for Multiclass Loss Matrices*, Journal of Machine Learning Research **17** (2016), no. 14, 1–45. Read in full from the JMLR site on 3 September 2026; the volume, article number and page range are those of the JMLR record for the article. Definition 10 (the convex calibration dimension), Lemma 11 and Theorem 12 (the upper bounds $N-1$ and $\operatorname{affdim} L$), Definition 14 and Lemma 15 (the feasible subspace dimension and its computation as a nullity), Theorem 16 (the bound (5)), Lemma 17, Theorem 18 and Corollary 19 (the full-support all-tied sufficient condition), Example 11 ($\operatorname{CCdim}(L_n^{0\text{-}1}) = n-1$).

[3] A. Nowak-Vila, F. Bach and A. Rudi, *Sharp Analysis of Learning with Discrete Losses*, in: Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics (AISTATS 2019), Proceedings of Machine Learning Research vol. 89, 1920–1929; e-print arXiv:1810.06839 (v1, the only version). Its appendix section on the F-score uses the same empty-set convention as (1) and is the prior-art anchor for Section 4: it states that the sufficient condition of [2, Theorem 18] fails for the F-score at two labels and that nothing can then be said about calibrated surrogates of dimension below $\operatorname{affdim}(L) - 1$. The sentences quoted in Section 1 were checked word for word against both the arXiv v1 LaTeX source and the published PMLR supplementary file, where they agree; the surname is spelt Nowak-Vila in the arXiv record and in the byline of the published paper, and Nowak in the PMLR index metadata. Read on 3 September 2026.

[4] 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 (page range confirmed against the PMLR volume index on 3 September 2026; PMLR record `v119/zhang20w`); e-print arXiv:2009.07801 (v1, submitted 16 September 2020, the only version on 4 September 2026). Read on 3–4 September 2026 from the arXiv v1 PDF and from the PMLR PDF, in both of which its equation (1) defines $F_\beta(y, \hat{y}) = (1+\beta^2) \sum_j y_j \hat{y}_j / (\beta^2 \|y\|_1 + \|\hat{y}\|_1)$, the convention (6), and its Proposition 3 states $\operatorname{rank}(L^{F_\beta} - \mathbf{1}) \le s^2 + 1$, the only published rank statement for general β; it also constructs a convex calibrated surrogate of dimension $s^2 + 1$ with a regret transfer. [1, Theorem 4.3] sharpens the rank half at $\beta = 1$, and Theorem 7.2 and Remark 8.3 above at every β. Version 1 of this paper cited it only through the restatement in [1].

[5] M. Zhang, *Exponential Convex Calibration Dimension for the Multi-Label Jaccard Measure*, arXiv:2608.13549v1 (submitted 13 August 2026). Read in full on 3 September 2026. It proves $2^{s-1} \le \operatorname{CCdim}(L^{\mathrm{Jac}}) \le 2^s - 1$ by the same route through [2, Theorem 16], cites [1] for $\operatorname{CCdim}(L^{F_1}) = \Theta(s^2)$, and computes no F_1 value at any specific s.

[6] P. Dewasurendra (Johns Hopkins University), *Dice Is Quadratic, Jaccard Is Exponential: Convex Calibration of Set Similarities*. Indexed by Semantic Scholar as CorpusId 291141794, with no arXiv identifier, no venue and no year in that record; retrieved through the open-access link in that record and read in full on 3 September 2026. It is a concurrent, independent proof of the same Jaccard bounds as [5] by a different author; its F_1 statements are the asymptotic ones quoted from [4] and [1], and it computes no F_1 value at any specific s.

[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`.