Dice Is Quadratic, Jaccard Is Exponential: Convex Calibration of Set Similarities

Pahan Dewasurendra
Johns Hopkins University

Abstract

Dice and Jaccard are monotone transforms on each prediction–outcome pair, yet their statistically calibrated convex surrogates have radically different dimension. For s-label instance-wise prediction, recent work places the convex calibration dimension of Dice/F1 at order s^2. We prove that Jaccard loss instead satisfies

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

Thus every distribution-free convex calibrated surrogate for example-based intersection over union needs exponentially many prediction coordinates. The score and loss matrices both have exact rank 2^s, and the loss has affine dimension $2^s - 1$. The lower bound is not a rank argument. We construct a full-support outcome law for which exactly $2^{s-1} + 1$ reports are Bayes optimal: the empty report and all reports containing one distinguished label. Its boundary form combines an empty-outcome atom with factorial mass proportional to $1/(|A| - 1)!$ on sets containing that label. A factorial cancellation makes all active reports tie, while strict positive definiteness of the Jaccard kernel makes their loss columns affinely independent. The feasible-subspace theorem then gives the lower bound. We also give a self-contained strict-definiteness proof by min-wise hashing and Möbius inversion, an explicit calibrated upper surrogate, and a regret transfer. The result explains why low-dimensional convex IoU extensions cannot be exactly calibrated without additional assumptions.

1 Introduction

The Jaccard index, also called intersection over union (IoU), evaluates a predicted set B against an outcome set A by $|A \cap B|/|A \cup B|$. It is a standard example-based metric in multi-label prediction and a central evaluation measure in segmentation (Waegeman et al., 2014; Berman et al., 2018; Wang et al., 2023). Its close relative Dice, or F1, replaces the denominator by $(|A| + |B|)/2$. Pointwise the two scores contain the same information:

$$J(A, B) = \frac{F_1(A, B)}{2 - F_1(A, B)}. \tag{1}$$

The transform is strictly increasing. It is therefore tempting to expect comparable surrogate complexity.

Expectation breaks this equivalence. A Bayes action maximizes an expected score, and a nonlinear transform cannot generally be moved through that expectation. This distinction was visible in decision-theoretic work showing that Bayes-optimal F1 prediction uses only $s^2 + 1$ statistics of the joint label law, while unrestricted Jaccard optimization remained tied to the full law (Waegeman et al., 2014). It becomes sharp at the level of convex calibration.

Convex calibration dimension is the smallest number of real prediction coordinates in any convex surrogate that is calibrated for a finite target loss (Ramaswamy and Agarwal, 2016). It does not fix a particular loss family, link, or output code. It therefore asks whether a low-dimensional convex objective can be statistically faithful at all. For multi-label F1, quadratic-dimensional calibrated surrogates are known (Zhang et al., 2020). Concurrent work proves a matching quadratic lower bound up to constants and explicitly asks for the corresponding analysis of Jaccard loss (Zhang, 2026).

We answer that question with an exponential separation. If L^{Jac} is the $2^s \times 2^s$ instance-wise Jaccard loss matrix, then

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

The upper bound estimates the full conditional law. The lower bound applies to every convex calibrated surrogate, including nonpolyhedral and nonsmooth ones. In particular, no surrogate with poly(s) prediction coordinates can be distribution-free calibrated for instance-wise IoU.

The proof combines kernel geometry and Bayes geometry. The Jaccard score matrix is strictly positive definite. We give a short proof from its minwise-hashing representation (Broder, 1997), followed by Möbius inversion. This recovers the full-rank fact previously proved by an infinite positive-semidefinite decomposition (Bouchard et al., 2013). Full rank alone does not lower-bound arbitrary convex calibration dimension.

The new ingredient is a trigger distribution with exponentially many independent ties. Fix label 1. On outcomes A containing 1, assign mass proportional to $1/(|A|-1)!$. Every report containing 1 has the same expected Jaccard score, while every nonempty report omitting 1 is worse. Adding a calibrated atom at the empty outcome makes the empty report tie as well. An exact linear correction then gives every remaining outcome positive mass while preserving all $2^{s-1}+1$ ties. The active loss columns have affine dimension 2^{s-1} by strict positive definiteness. The feasible-subspace lower bound of Ramaswamy and Agarwal (2016) yields (2).

This also clarifies the status of practical IoU losses. Lovász extensions and soft Jaccard objectives are useful low-dimensional optimization tools (Berman et al., 2018; Wang et al., 2023). Prior consistency analysis shows that, for targets induced by a fixed nonmodular submodular function of the error set, the Lovász hinge instead embeds a structured abstain loss (Finocchiaro et al., 2022). That theorem does not directly cover the ground-truth-dependent Jaccard matrix studied here. Our result does: it rules out every low-dimensional convex calibrated surrogate for this target, rather than one construction. It does not rule out nonconvex surrogates, restricted conditional laws, or approximate notions of calibration.

2 Setting and Calibration Dimension

Fix $s \geq 1$ and write $[s] = \{1, \dots, s\}$. Both outcomes and reports belong to $\mathcal{Y} = 2^{[s]}$, so $N = |\mathcal{Y}| = 2^s$. Define

$$J(A,B) = \begin{cases} 1, & A = B = \varnothing, \\ |A \cap B|/|A \cup B|, & \text{otherwise}, \end{cases} \tag{3}$$

$$L(A,B) = 1 - J(A,B). \tag{4}$$

The convention $J(\varnothing,\varnothing) = 1$ treats two empty sets as a perfect match and is standard for per-instance IoU (Berman et al., 2018). Section S6 treats the alternative value zero. Its lower bound is smaller by one, so the exponential order and all asymptotic conclusions are unchanged.

For $p \in \Delta_N$, let

$$\mathrm{opt}_L(p) = \arg\min_{B \in \mathcal{Y}} p^\top L_{\cdot,B} \tag{5}$$

be its Bayes reports. A d-dimensional convex surrogate consists of a convex set $C \subseteq \mathbb{R}^d$, a map $\psi : C \to \mathbb{R}_+^N$ whose coordinates are convex, and a link $\mathrm{pred} : C \to \mathcal{Y}$. It is L-calibrated when, for every $p \in \Delta_N$,

$$\inf_{u : \mathrm{pred}(u) \notin \mathrm{opt}_L(p)} p^\top \psi(u) > \inf_{u \in C} p^\top \psi(u). \tag{6}$$

The convex calibration dimension $\mathrm{CCdim}(L)$ is the smallest such d. We use two general results of Ramaswamy and Agarwal (2016). First,

$$\mathrm{CCdim}(L) \leq \mathrm{affdim}\{L_{\cdot,B} : B \in \mathcal{Y}\}. \tag{7}$$

Second, for the trigger set $Q_B = \{p : B \in \mathrm{opt}_L(p)\}$ and any $p \in Q_B$,

$$\mathrm{CCdim}(L) \geq \|p\|_0 - \mu_{Q_B}(p) - 1, \tag{8}$$

where $\mu_{Q_B}(p)$ is the dimension of the two-sided feasible-direction subspace of Q_B at p.

3 Exact Linear Algebra

We first establish the rank facts and the independent active directions needed later.

Lemma 1 (Strict Jaccard kernel)**.** *The matrix $J \in \mathbb{R}^{N \times N}$ in (4) is strictly positive definite.*

Proof. Restrict first to nonempty sets. If π is a uniformly random permutation of $[s]$, then

$$J(A,B) = \Pr\{\min_\pi A = \min_\pi B\}. \tag{9}$$

For $f_{\pi,i}(A) = \mathbf{1}\{\min_\pi A = i\}$, this expresses J as the average Gram matrix of the features $f_{\pi,i}$, so it is positive semidefinite.

Suppose a coefficient vector $(c_A)_{A \neq \varnothing}$ has zero quadratic form. Every squared feature sum must vanish:

$$\sum_{A : \min_\pi A = i} c_A = 0 \quad \text{for every } \pi, i. \tag{10}$$

Fix i and any $P \subseteq [s] \setminus \{i\}$. Choose π with exactly P before i. Equation (10) becomes

$$\sum_{S \subseteq ([s] \setminus \{i\}) \setminus P} c_{S \cup \{i\}} = 0. \tag{11}$$

As P varies, every lower-set sum of $S \mapsto c_{S\cup\{i\}}$ is zero. Möbius inversion gives $c_{S\cup\{i\}} = 0$ for every S. Varying i proves strictness on all nonempty sets. The empty set contributes a separate unit coordinate and zero cross terms, which proves the claim. □

Proposition 2 (Rank and affine dimension)**.** *For every $s \geq 1$,*

$$\operatorname{rank}(J) = \operatorname{rank}(L) = 2^s, \tag{12}$$

$$\operatorname{affdim}\{L_{\cdot,B} : B \in \mathcal{Y}\} = 2^s - 1. \tag{13}$$

Proof. Lemma 1 gives $\operatorname{rank}(J) = N$. Order the empty set first, so $J = 1 \oplus J_+$ with $J_+ \succ 0$. The matrix determinant lemma gives

$$\det(L) = \det(-J)\left(1 - \mathbf{1}^\top J^{-1}\mathbf{1}\right). \tag{14}$$

Here $\mathbf{1}^\top J^{-1}\mathbf{1} = 1 + \mathbf{1}^\top J_+^{-1}\mathbf{1} > 1$, so L is nonsingular. Its N columns are linearly independent and hence affinely independent, proving (13). □

Proposition 2 and (7) already give the upper bound $\operatorname{CCdim}(L) \leq 2^s - 1$. Rank itself does not give the lower bound in (2). The required Bayes face comes next.

4 A Factorial Bayes Face

Let $U = [s] \setminus \{1\}$ and let

$$\mathcal{A} = \{B \subseteq [s] : 1 \in B\}. \tag{15}$$

This family has size $r = 2^{s-1}$. Define a distribution p^0 supported on $\mathcal{A}$ by

$$p^0(\{1\} \cup S) = \frac{1}{Z_s |S|!}, \qquad Z_s = \sum_{S\subseteq U} \frac{1}{|S|!}. \tag{16}$$

Lemma 3 (Factorial ties)**.** *Under p^0, every report in $\mathcal{A}$ has the same expected Jaccard score. Every report outside $\mathcal{A}$ has smaller expected score, with gap at least $1/s$. The minimum gap is exactly $1/s$.*

Proof. For $T \subseteq U$, write $B_T = \{1\} \cup T$ and omit the common factor Z_s^{-1}. Its expected score is

$$G(T) = \sum_{S\subseteq U} \frac{1 + |S \cap T|}{1 + |S \cup T|} \frac{1}{|S|!}. \tag{17}$$

We show that adding one element $x \notin T$ does not change G. Pair $S = A$ with $S = A \cup \{x\}$. After grouping by $a = |A \cap T|$ and $b = |A \setminus T|$, the difference vanishes by

$$\sum_{a=0}^{|T|} \binom{|T|}{a} \left\{ \frac{1}{(a+b+1)!} - \frac{a+1}{(|T|+b+1)(a+b)!} \right\} = 0. \tag{18}$$

Indeed, after multiplying by $|T| + b + 1$, the difference of the two sides reduces to

$$\sum_{a} \binom{|T|}{a} \left\{ \frac{a}{(a+b)!} - \frac{|T| - a}{(a+b+1)!} \right\} = 0, \tag{19}$$

where $a\binom{|T|}{a} = |T|\binom{|T|-1}{a-1}$ matches $(|T| - a)\binom{|T|}{a} = |T|\binom{|T|-1}{a}$ after shifting the index. Thus $G(T)$ is constant.

Compare B_T with the report T, which omits label 1. For every outcome $\{1\} \cup S$ their score difference is

$$\frac{1}{1 + |S \cup T|} \geq \frac{1}{s}. \tag{20}$$

Every inactive report has this form. Equality in (20) holds for $T = U$, proving both strictness and the exact minimum gap. □

Let $c_s > 0$ be the common expected score in Lemma 3, and enlarge the active family to

$$\mathcal{C} = \{\varnothing\} \cup \mathcal{A}, \qquad |\mathcal{C}| = r + 1. \tag{21}$$

Define a boundary law $\bar{p}^0$ by

$$\bar{p}^0(\varnothing) = \frac{c_s}{1 + c_s}, \qquad \bar{p}^0(A) = \frac{p^0(A)}{1 + c_s} \quad (A \in \mathcal{A}). \tag{22}$$

The empty report scores $c_s/(1 + c_s)$ under this law. Every report in $\mathcal{A}$ has the same score because it scores zero on the empty outcome. Every nonempty report outside $\mathcal{A}$ remains worse by at least $1/[s(1 + c_s)]$. Hence the exact Bayes family is $\mathcal{C}$.

The support of $\bar{p}^0$ has only $r+1$ outcomes. To use (8), we move this witness into the relative interior of the full N-outcome simplex.

Lemma 4 (Interiorization)**.** *There is a full-support distribution p for which exactly the reports in $\mathcal{C}$ are Bayes optimal.*

Proof. If $s = 1$, the law $\bar{p}^0$ already has full support and the claim follows. Suppose $s \geq 2$. Partition the score matrix with rows in $\mathcal{C}$ and $\mathcal{C}^c$, and columns in $\mathcal{C}$. Let $K = J_{\mathcal{C},\mathcal{C}}$ and $H = J_{\mathcal{C}^c,\mathcal{C}}$. Lemma 1 makes K invertible. Choose any strictly positive distribution q on $\mathcal{C}^c$. For $\epsilon > 0$, set $p_{\mathcal{C}^c} = \epsilon q$ and seek $p_{\mathcal{C}}$ such that

$$K^\top p_{\mathcal{C}} + \epsilon H^\top q = c_\epsilon \mathbf{1}. \tag{23}$$

The unique solution with total active mass $1 - \epsilon$ is

$$p_{\mathcal{C}} = c_\epsilon K^{-\top}\mathbf{1} - \epsilon K^{-\top} H^\top q, \tag{24}$$

$$c_\epsilon = \frac{1 - \epsilon + \epsilon \mathbf{1}^\top K^{-\top} H^\top q}{\mathbf{1}^\top K^{-\top}\mathbf{1}}. \tag{25}$$

At $\epsilon = 0$, this is exactly $\bar{p}^0_{\mathcal{C}}$ by (22) and uniqueness. Every coordinate is then positive. Continuity keeps all coordinates of (24) positive for sufficiently small ϵ. The active reports remain tied by (23), and their strict gap over inactive reports persists by continuity. □

5 Exponential Calibration Dimension

Theorem 5 (Convex calibration dimension of Jaccard)**.** *For the s-label instance-wise Jaccard loss,*

$$2^{s-1} \le \mathrm{CCdim}(L) \le 2^s - 1. \tag{26}$$

Consequently, $\mathrm{CCdim}(L) = \Theta(2^s)$.

Proof. The upper bound follows from Proposition 2 and (7). For the lower bound, take the full-support p from Lemma 4, fix $B_0 \in \mathcal{C}$, and put $d_B = L_{\cdot,B} - L_{\cdot,B_0}$. Exactly the comparisons indexed by $B \in \mathcal{C}$ are active at p. Since the corresponding columns of J are linearly independent, the differences $\{d_B : B \in \mathcal{C} \setminus \{B_0\}\}$ are linearly independent and have rank r.

Two-sided feasible directions v in the trigger set Q_{B_0} must satisfy normalization $\mathbf{1}^\top v = 0$ and preserve every active tie, so $v^\top d_B = 0$ for $B \in \mathcal{C}$. These conditions are also sufficient because p has full support and every inactive comparison is strict. The normalization row is independent of the r difference rows: every difference is orthogonal to p, while $\mathbf{1}^\top p = 1$. Hence

$$\mu_{Q_{B_0}}(p) = N - r - 1. \tag{27}$$

Substitution into (8) gives

$$\mathrm{CCdim}(L) \ge N - (N - r - 1) - 1 = r = 2^{s-1}. \tag{28}$$

□

Corollary 6 (Dice–Jaccard separation)**.** *Although Dice/F1 and Jaccard are related pointwise by* (1)*, their convex calibration dimensions are respectively* $\Theta(s^2)$ *and* $\Theta(2^s)$.

The Dice statement combines the quadratic calibrated construction of Zhang et al. (2020) with the concurrent lower bound of Zhang (2026). The contrast is not a rank-only artifact. Theorem 5 identifies an exponentially dimensional face in Jaccard Bayes geometry.

Corollary 7 (No low-dimensional convex IoU surrogate)**.** *For $s \ge 3$, no convex surrogate with s prediction coordinates is distribution-free calibrated for instance-wise Jaccard loss. More generally, no family of* $\mathrm{poly}(s)$*-dimensional convex surrogates is calibrated for all s.*

This conclusion covers every choice of convex objective and link. In particular, it rules out calibration of any s-coordinate convex Lovász-style IoU construction for $s \ge 3$. This implication comes directly from the dimension bound and does not require Jaccard loss to be a fixed function of the misprediction set.

6 An Explicit Upper Surrogate

The upper endpoint in (26) is constructive. Represent $q \in \Delta_N$ by $N - 1$ free coordinates and use the multiclass Brier surrogate

$$\psi_A(q) = \|q - e_A\|_2^2, \qquad \mathrm{pred}(q) \in \arg\min_{B \in \mathcal{Y}} q^\top L_{\cdot,B}. \tag{29}$$

Proposition 8 (Calibration and regret transfer)**.** *The surrogate in* (29) *is convex and L-calibrated in dimension $2^s - 1$. If the conditional outcome law is p, then*

$$\mathrm{reg}_L(\mathrm{pred}(q), p) \le 2^{s/2} \sqrt{\mathrm{reg}_\psi(q, p)}. \tag{30}$$

Proof. Conditional Brier regret is $\mathrm{reg}_\psi(q, p) = \|q - p\|_2^2$. Let B^* be Bayes optimal under p and $\widehat{B} = \mathrm{pred}(q)$. Optimality under q gives

$$p^\top (L_{\cdot,\widehat{B}} - L_{\cdot,B^*}) \le (p - q)^\top (L_{\cdot,\widehat{B}} - L_{\cdot,B^*}). \tag{31}$$

The loss difference has Euclidean norm at most $\sqrt{N}$. Cauchy–Schwarz yields (30). For fixed p, if any report is suboptimal, its smallest positive target regret γ_p is positive because $\mathcal{Y}$ is finite. Every q linked to a suboptimal report therefore has surrogate regret at least γ_p^2 / N. This proves the strict inequality in (6). If every report is optimal, that inequality has an empty left constraint and is vacuous. □

The upper surrogate estimates the complete conditional distribution. Theorem 5 shows that its exponential coordinate count is optimal within a factor of two among all convex calibrated constructions, even those that do not estimate probabilities.

7 Related Work and Scope

Convex calibration dimension. Ramaswamy and Agarwal (2016) introduced the general loss-matrix framework, the affine-dimension upper bound, and the feasible-subspace lower bound used here. Their applications include subset ranking losses. Exact rank alone does not imply their lower bound, which is why the factorial trigger and its full-support correction are necessary. Closely related property-elicitation work studies dimensions under additional regularity or loss classes. We use the unrestricted convex calibration definition.

F-measures and Bayes prediction. Waegeman et al. (2014) showed that Bayes-optimal F1 prediction can be computed from a quadratic collection of conditional statistics and analyzed F1 as an approximation to Jaccard. Zhang et al. (2020) converted the low-rank F-measure representation into quadratic-dimensional

calibrated surrogates. Zhang (2026) recently proved the exact F1 rank and a matching-order calibration lower bound. That paper names Jaccard as an open extension. The present result gives a different order rather than a parallel quadratic formula.

IoU objectives. The Lovász hinge and Lovász–Softmax use outcome-dependent convex extensions of Jaccard loss in misprediction coordinates (Berman et al., 2018). For the related setting where the target is induced by one fixed submodular function of the error set, Finocchiaro et al. (2022) proved that the Lovász hinge embeds a structured abstain target and is inconsistent unless that set function is modular. Jaccard is ground-truth dependent in error coordinates, so that theorem is not a direct Jaccard result. Soft Jaccard and Jaccard metric losses provide differentiable low-dimensional objectives and strong empirical performance (Wang et al., 2023). These constructions address optimization and empirical accuracy rather than unrestricted loss-matrix calibration. Theorem 5 shows that no convex objective with similarly few coordinates can satisfy the latter property.

Different meanings of Jaccard calibration. Dataset-level binary Jaccard is a ratio of population confusion-matrix entries. Bao and Sugiyama (2020) give calibrated direct optimization for that linear-fractional utility. Their prediction problem has one binary decision per instance and is distinct from the example-based multi-label loss matrix studied here, where one outcome is an entire subset and expectation is taken after computing its set similarity.

The result is distribution-free and finite-output. It does not assert computational hardness of Bayes decoding from a represented distribution. It also does not preclude low-dimensional calibrated surrogates on restricted conditional-law families, nonconvex calibrated objectives, or approximate calibration. The exponential obstruction concerns exact convex calibration over all conditional distributions. This is the regime in which convex surrogate minimization is a universal replacement for target-risk minimization.

References

Bao, H. and Sugiyama, M. (2020). Calibrated surrogate maximization of linear-fractional utility in binary classification. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 2337–2347. PMLR.

Berman, M., Rannen Triki, A., and Blaschko, M. B. (2018). The Lovász–Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4413–4421.

Bouchard, M., Jousselme, A.-L., and Doré, P.-E. (2013). A proof for the positive definiteness of the Jaccard index matrix. *International Journal of Approximate Reasoning*, 54(5):615–626.

Broder, A. Z. (1997). On the resemblance and containment of documents. In *Proceedings of Compression and Complexity of Sequences 1997*, pages 21–29. IEEE.

Finocchiaro, J. J., Frongillo, R., and Nueve, E. B. (2022). The structured abstain problem and the Lovász hinge. In *Proceedings of the Thirty-Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 3718–3740. PMLR.

Ramaswamy, H. G. and Agarwal, S. (2016). Convex calibration dimension for multiclass loss matrices. *Journal of Machine Learning Research*, 17(14):1–45.

Waegeman, W., Dembczyński, K., Jachnik, A., Cheng, W., and Hüllermeier, E. (2014). On the Bayes-optimality of F-measure maximizers. *Journal of Machine Learning Research*, 15(103):3513–3568.

Wang, Z., Ning, X., and Blaschko, M. B. (2023). Jaccard metric losses: Optimizing the Jaccard index with soft labels. In *Advances in Neural Information Processing Systems*, volume 36, pages 75259–75285.

Zhang, M. (2026). Exact rank and convex calibration dimension lower bounds for the multi-label F1 loss. *arXiv preprint arXiv:2608.08399*.

Zhang, M., Ramaswamy, H. G., and Agarwal, S. (2020). Convex calibrated surrogates for the multi-label F-measure. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 11246–11255. PMLR.

A Strict Positive Definiteness

We give a complete feature proof for the Jaccard matrix. Let $\mathcal{Y}_+ = 2^{[s]} \setminus \{\varnothing\}$. For a permutation π of $[s]$ and $i \in [s]$, define

$$f_{\pi,i}(A) = \mathbf{1}\{i \in A\}\mathbf{1}\{A \cap P_{\pi,i} = \varnothing\}, \tag{32}$$

where $P_{\pi,i}$ is the set of elements appearing before i in π. Thus $f_{\pi,i}(A) = 1$ exactly when i is the first element of A under π.

Lemma 9. *For nonempty $A, B \subseteq [s]$,*

$$\frac{|A \cap B|}{|A \cup B|} = \frac{1}{s!} \sum_{\pi} \sum_{i=1}^{s} f_{\pi,i}(A) f_{\pi,i}(B). \tag{33}$$

Proof. For a fixed permutation, the inner sum is one if the first element of $A \cup B$ lies in $A \cap B$ and zero otherwise. Under a uniform permutation, every element of $A \cup B$ is equally likely to appear first. The probability is therefore $|A \cap B|/|A \cup B|$. □

Equation (33) makes the nonempty Jaccard matrix K positive semidefinite. For strictness, suppose $c^\top K c = 0$. Expanding the average of squares gives

$$\sum_{A \in \mathcal{Y}_+} c_A f_{\pi,i}(A) = 0 \tag{34}$$

for every π, i. Fix i and a subset $P \subseteq [s] \setminus \{i\}$. There is a permutation whose elements before i are exactly P. For that permutation, (34) says

$$\sum_{S \subseteq ([s] \setminus \{i\}) \setminus P} c_{S \cup \{i\}} = 0. \tag{35}$$

Put $U_i = [s] \setminus \{i\}$ and $d_S = c_{S \cup \{i\}}$. As P ranges over subsets of U_i, its complement $Q = U_i \setminus P$ also ranges over all subsets. Hence (35) gives

$$\sum_{S \subseteq Q} d_S = 0 \qquad \text{for every } Q \subseteq U_i. \tag{36}$$

Boolean-lattice Möbius inversion yields

$$d_Q = \sum_{S \subseteq Q} (-1)^{|Q \setminus S|} \sum_{R \subseteq S} d_R = 0. \tag{37}$$

Thus $c_A = 0$ for every nonempty A containing i. Every nonempty set contains some i, so $c = 0$ and $K \succ 0$. Under $J(\varnothing, \varnothing) = 1$, the full score matrix is $1 \oplus K$ and is also strictly positive definite.

This proof uses the minwise representation introduced for document resemblance by Broder (1997). Bouchard et al. (2013) proved the same strict-definiteness conclusion for the complete nonempty Jaccard matrix by decomposing it into an infinite sum of positive-semidefinite matrices. The Möbius argument above is included to make the exact rank input self-contained.

B Score Rank, Loss Rank, and Affine Dimension

Let $N = 2^s$ and order the empty set first. Under the unit empty-set convention,

$$J = \begin{bmatrix} 1 & 0 \\ 0 & K \end{bmatrix}, \qquad K \succ 0. \tag{38}$$

This gives $\operatorname{rank}(J) = N$. Since $L = \mathbf{1}\mathbf{1}^\top - J$, the matrix determinant lemma gives

$$\det(L) = \det(-J) \det\left(1 + \mathbf{1}^\top (-J)^{-1} \mathbf{1}\right) \tag{39}$$

$$= \det(-J) \left(1 - \mathbf{1}^\top J^{-1} \mathbf{1}\right). \tag{40}$$

The block form gives

$$\mathbf{1}^\top J^{-1} \mathbf{1} = 1 + \mathbf{1}^\top K^{-1} \mathbf{1} > 1, \tag{41}$$

where strict positivity follows from $K^{-1} \succ 0$. Thus (40) is nonzero and $\operatorname{rank}(L) = N$.

If vectors $v_1, \ldots, v_N$ are linearly independent, then they are affinely independent. Indeed,

$$\sum_{j=2}^{N} a_j (v_j - v_1) = 0 \tag{42}$$

is a linear relation with coefficient a_j on v_j and coefficient $-\sum_{j=2}^{N} a_j$ on v_1. Linear independence forces every $a_j = 0$. Applying this to the columns of L gives affine dimension $N - 1$.

We will also need the following direct consequence. If $\mathcal{A}$ is any set of r reports, then its score columns are linearly independent. Fixing $B_0 \in \mathcal{A}$, the $r-1$ vectors

$$J_{\cdot,B} - J_{\cdot,B_0}, \qquad B \in \mathcal{A} \setminus \{B_0\}, \tag{43}$$

are linearly independent. Negating them gives the corresponding loss differences.

C The Factorial Identity

Fix a distinguished label 1 and put $U = [s] \setminus \{1\}$, so $|U| = n = s - 1$. Define the unnormalized mass

$$w(S) = \frac{1}{|S|!}, \qquad S \subseteq U, \tag{44}$$

at outcome $\{1\} \cup S$. Let $T \subseteq U$. The unnormalized expected Jaccard score of report $\{1\} \cup T$ is

$$G(T) = \sum_{S \subseteq U} \frac{1 + |S \cap T|}{1 + |S \cup T|} \frac{1}{|S|!}. \tag{45}$$

Lemma 10. *The quantity $G(T)$ is independent of T.*

Proof. It suffices to prove $G(T \cup \{x\}) = G(T)$ whenever $x \in U \setminus T$. Pair each $A \subseteq U \setminus \{x\}$ with the two outcomes $S = A$ and $S = A \cup \{x\}$. Write

$$a = |A \cap T|, \qquad b = |A \setminus T|, \qquad t = |T|. \tag{46}$$

Then $|A| = a + b$ and $|A \cup T| = t + b$. The paired contribution to $G(T \cup \{x\}) - G(T)$ is

$$D_{a,b} = -\frac{1+a}{(1+t+b)(2+t+b)(a+b)!} \tag{47}$$

$$+\frac{1}{(2+t+b)(a+b+1)!}. \tag{48}$$

There are $\binom{t}{a}\binom{n-t-1}{b}$ sets A with these values. For fixed b, the factor $\binom{n-t-1}{b}/(2+t+b)$ is common. It remains to show

$$\sum_{a=0}^{t} \binom{t}{a} \left\{ \frac{1}{(a+b+1)!} - \frac{a+1}{(t+b+1)(a+b)!} \right\} = 0. \tag{49}$$

Multiplying by $t+b+1$ and using

$$\frac{t+b+1}{(a+b+1)!} = \frac{1}{(a+b)!} + \frac{t-a}{(a+b+1)!} \tag{50}$$

reduces (49) to

$$\sum_{a=0}^{t} \binom{t}{a} \frac{a}{(a+b)!} = \sum_{a=0}^{t} \binom{t}{a} \frac{t-a}{(a+b+1)!}. \tag{51}$$

The identities

$$a\binom{t}{a} = t\binom{t-1}{a-1}, \qquad (t-a)\binom{t}{a} = t\binom{t-1}{a} \tag{52}$$

make the two sides of (51) identical after replacing a by $a-1$ on the left. Summing over b proves the claim. □

Normalizing (44) gives the law in the main paper. The common active score has the explicit form

$$c_s = \frac{\displaystyle\sum_{r=0}^{s-1} \binom{s-1}{r} (r+1)!^{-1}}{\displaystyle\sum_{r=0}^{s-1} \binom{s-1}{r} r!^{-1}}. \tag{53}$$

For a report $T \subseteq U$ that omits label 1, compare it with the active report $\{1\} \cup T$. At outcome $\{1\} \cup S$ their score difference is exactly

$$\frac{1+|S \cap T|}{1+|S \cup T|} - \frac{|S \cap T|}{1+|S \cup T|} = \frac{1}{1+|S \cup T|}. \tag{54}$$

This lies in $[1/s, 1]$. It equals $1/s$ for every S when $T = U$. Hence all active reports tie, all inactive reports are strictly worse, and the minimum expected gap is exactly $1/s$.

The total probability assigned to outcomes with $|S| = r$ is

$$\frac{\binom{s-1}{r}/r!}{\sum_{j=0}^{s-1}\binom{s-1}{j}/j!}. \tag{55}$$

This layer formula explains the integer patterns in the verification output. After multiplying by $(s-1)!$, the unnormalized layer weights are $\binom{s-1}{r}(s-1)!/r!$.

Now let $\mathcal{A} = \{B : 1 \in B\}$ and $\mathcal{C} = \{\varnothing\} \cup \mathcal{A}$. Denote the factorial law by p^0 and its common active score by c_s. Define

$$\bar{p}^0(\varnothing) = \frac{c_s}{1+c_s}, \qquad \bar{p}^0(A) = \frac{p^0(A)}{1+c_s} \quad (A \in \mathcal{A}). \tag{56}$$

The empty report has expected score $c_s/(1+c_s)$. Every report in $\mathcal{A}$ has the same expected score because its score against the empty outcome is zero. A nonempty report outside $\mathcal{A}$ has its original expected score divided by $1+c_s$, so it is worse by at least

$$\frac{1}{s(1+c_s)}. \tag{57}$$

Consequently, the exact Bayes family under $\bar{p}^0$ is $\mathcal{C}$, which has size $2^{s-1}+1$.

D Moving the Witness into the Interior

Let $\mathcal{D} = \mathcal{Y} \setminus \mathcal{C}$ and write $r = |\mathcal{A}| = N/2$, so $|\mathcal{C}| = r+1$. If $s = 1$, then $\mathcal{C} = \mathcal{Y}$ and the law (56) already has full support. Suppose henceforth that $s \geq 2$. Partition the score columns indexed by $\mathcal{C}$ into

$$K = J_{\mathcal{C},\mathcal{C}} \in \mathbb{R}^{(r+1)\times(r+1)}, \tag{58}$$

$$H = J_{\mathcal{D},\mathcal{C}} \in \mathbb{R}^{(N-r-1)\times(r+1)}. \tag{59}$$

The principal matrix K is strictly positive definite. The preceding section gives

$$K^\top \bar{p}^0_{\mathcal{C}} = \bar{c}_0 \mathbf{1}, \qquad \mathbf{1}^\top \bar{p}^0_{\mathcal{C}} = 1, \tag{60}$$

Because K is invertible, (60) has the unique solution

$$\bar{p}^0_{\mathcal{C}} = \bar{c}_0 K^{-\top}\mathbf{1}, \qquad \bar{c}_0 = (\mathbf{1}^\top K^{-\top}\mathbf{1})^{-1}. \tag{61}$$

Choose an arbitrary $q \in \mathrm{relint}(\Delta_{\mathcal{D}})$. For $\epsilon > 0$, set

$$p^\epsilon_{\mathcal{D}} = \epsilon q. \tag{62}$$

Define

$$u = K^{-\top}\mathbf{1}, \qquad v = K^{-\top}H^\top q, \tag{63}$$

$$c_\epsilon = \frac{1-\epsilon+\epsilon\mathbf{1}^\top v}{\mathbf{1}^\top u}, \qquad p^\epsilon_{\mathcal{C}} = c_\epsilon u - \epsilon v. \tag{64}$$

Then $\mathbf{1}^\top p^\epsilon_{\mathcal{C}} = 1-\epsilon$ and

$$K^\top p^\epsilon_{\mathcal{C}} + H^\top p^\epsilon_{\mathcal{D}} = c_\epsilon \mathbf{1}. \tag{65}$$

At $\epsilon = 0$, (64) equals the coordinatewise positive vector $\bar{p}^0_{\mathcal{C}}$. For all sufficiently small positive ϵ, every active coordinate remains positive, while every coordinate in $\mathcal{D}$ is positive by construction. Thus $p^\epsilon \in \operatorname{relint}(\Delta_N)$.

Equation (65) makes all reports in $\mathcal{C}$ tie exactly. At $\epsilon = 0$, every inactive report is below them by at least (57). Expected score is linear and there are finitely many reports, so for all sufficiently small ϵ every inactive gap remains positive. Therefore the Bayes set under p^ϵ is exactly $\mathcal{C}$.

E Feasible-Subspace Calculation

Fix $B_0 \in \mathcal{C}$ and define loss differences

$$d_B = L_{\cdot,B} - L_{\cdot,B_0}. \tag{66}$$

The trigger set for B_0 is

$$Q_{B_0} = \{z \in \Delta_N : z^\top d_B \geq 0 \text{ for every } B \in \mathcal{Y}\}. \tag{67}$$

At the full-support witness $p = p^\epsilon$, equality holds exactly for $B \in \mathcal{C}$. Strict positive definiteness gives

$$E = \operatorname{span}\{d_B : B \in \mathcal{C}\} \quad \text{with} \quad \dim E = r. \tag{68}$$

A vector $v \in \mathbb{R}^N$ is a two-sided feasible direction at p exactly when

$$\mathbf{1}^\top v = 0, \qquad v \perp E. \tag{69}$$

Necessity follows because $p \pm \eta v$ must stay normalized and must satisfy every active inequality in both signs. For sufficiency, full support keeps every coordinate nonnegative for sufficiently small η, the conditions in (69) preserve every active comparison, and finitely many strict inactive comparisons remain strict.

The normalization vector $\mathbf{1}$ does not belong to E. Indeed, every $d_B \in E$ is orthogonal to p because its two reports tie under p. Hence every vector in E is orthogonal to p, while $p^\top \mathbf{1} = 1$. The constraints in (69) therefore have rank

$$1 + \dim E = r + 1. \tag{70}$$

It follows that

$$\mu_{Q_{B_0}}(p) = N - r - 1. \tag{71}$$

Since p has full support, Theorem 16 of Ramaswamy and Agarwal (2016) and (71) give

$$\operatorname{CCdim}(L) \geq N - (N - r - 1) - 1 = r = 2^{s-1}. \tag{72}$$

The affine-dimension upper bound from their Theorem 12 and Section S2 gives $\operatorname{CCdim}(L) \leq N - 1$.

F The Alternative Empty-Set Convention

Some software sets $J(\varnothing, \varnothing) = 0$. Let $J^{(0)}$ denote that score matrix and $L^{(0)} = \mathbf{1}\mathbf{1}^\top - J^{(0)}$. The nonempty principal block J_+ remains strictly positive definite, so

$$J^{(0)} = 0 \oplus J_+, \qquad \operatorname{rank}(J^{(0)}) = N - 1. \tag{73}$$

The loss matrix remains nonsingular. If $L^{(0)}x = 0$, its empty-outcome row gives $\mathbf{1}^\top x = 0$. Every nonempty-outcome row then gives $J_+ x_+ = 0$, where x_+ contains the nonempty-report coordinates. Thus $x_+ = 0$, and normalization gives the empty coordinate zero. Hence $\operatorname{rank}(L^{(0)}) = N$ and its affine dimension is $N - 1$.

The empty report cannot join the augmented Bayes family because it always has score zero under this convention. The original factorial witness still has exact Bayes family $\mathcal{A}$, and its gap over every report outside $\mathcal{A}$ is at least $1/s$. The interior perturbation through the unchanged invertible block $J^{(0)}_{\mathcal{A},\mathcal{A}}$ therefore gives a full-support law with exactly these r active reports. Repeating the feasible-subspace calculation with $r-1$ independent active differences gives

$$2^{s-1} - 1 \leq \operatorname{CCdim}(L^{(0)}) \leq 2^s - 1. \tag{74}$$

Thus the standard convention improves the proved lower bound by one. The exponential order and the asymptotic separation from Dice/F1 are unchanged.

G The Brier Upper Surrogate

Let $p \in \Delta_N$ be the conditional outcome law and let $q \in \Delta_N$ be the surrogate report. For outcome A, define

$$\psi_A(q) = \|q - e_A\|_2^2. \tag{75}$$

Its conditional risk is

$$\sum_A p_A \psi_A(q) = \|q\|_2^2 - 2p^\top q + 1 \tag{76}$$

$$= \|q - p\|_2^2 + 1 - \|p\|_2^2. \tag{77}$$

The unique minimizer is $q = p$, and surrogate regret is $\|q - p\|_2^2$.

Decode q to any

$$\widehat{B} \in \arg\min_B q^\top L_{\cdot,B}. \tag{78}$$

If B^* minimizes target risk under p, then

$$p^\top L_{\cdot,\widehat{B}} - p^\top L_{\cdot,B^*} = (p-q)^\top (L_{\cdot,\widehat{B}} - L_{\cdot,B^*}) \tag{79}$$

$$+ q^\top (L_{\cdot,\widehat{B}} - L_{\cdot,B^*}) \tag{80}$$

$$\leq (p-q)^\top (L_{\cdot,\widehat{B}} - L_{\cdot,B^*}) \tag{81}$$

$$\leq \sqrt{N}\|p-q\|_2. \tag{82}$$

The final inequality uses that every loss difference coordinate lies in $[-1, 1]$. Combining (77) and (82) gives the stated $2^{s/2}$ regret transfer. For fixed p, finiteness of the report set makes the smallest positive regret γ_p of any suboptimal report strictly positive. Thus every q linked to a suboptimal report has surrogate regret at least γ_p^2/N, which is exactly the required strict calibration gap. If every report is optimal, the claim is vacuous. The simplex has affine dimension $N - 1$, so this is an $(N - 1)$-dimensional convex calibrated surrogate.

H Deterministic Verification

The supplied file `verify_jaccard_theorems.py` uses exact rational arithmetic for the factorial witness, SymPy for small exact ranks, and NumPy for interior perturbations and eigenvalues. It requires Python 3, NumPy, SciPy, and SymPy. From the paper directory, run

```
python verify_jaccard_theorems.py
```

The script checks the following:

1. exact score rank, loss rank, affine-difference rank, augmented active-face rank, original factorial-face rank, and both ranks under the alternative empty-set convention for $1 \le s \le 5$.
2. exact equality and gaps for the 2^{s-1}-report factorial face and the $2^{s-1} + 1$-report augmented face for $1 \le s \le 10$.
3. positive full-support perturbations, augmented active tie error below 10^{-9}, and strict inactive gaps for $1 \le s \le 8$.
4. positive minimum eigenvalues of the Jaccard score matrix for $1 \le s \le 10$, including the 1024×1024 case.
5. the Brier regret transfer on 500 fixed-seed random pairs of distributions for each $1 \le s \le 6$.

The checks are not used as proofs. They independently exercise every finite-dimensional identity and matrix orientation used by the analytic argument.