A Contragredient View of Conformal Placement and Steinitz Balancing

Peter Cotton
peter.cotton@microprediction.com

July 2026

ROUGH DRAFT — NOT FOR DISSEMINATION

Abstract

We give the validity of split conformal prediction an exact contragredient form: the placement acceptance map is equivariant for the permutation action, and the Reynolds identity $\langle P^\vee f, v\rangle = \langle f, Pv\rangle$ reads its orbit average either as moving the evaluation functional or as averaging the induced acceptance vector over placements. A natural primal discrepancy construction under the same action is Steinitz balancing: ordering a zero-sum population so that every prefix sum stays bounded. The centered conformal placement residuals form such a population, and an explicit ordering keeps the running placement-acceptance average within $1/(2t)$ of the orbit level k/N, exact once all N candidate test placements have been traversed; the vector-valued lemma balances several coverage levels along one traversal. On the function–measure pairing, the measure-side counterpart of calibration is balanced sampling and exact cubature. A transitive distributional symmetry, exchangeability in the standard setting, turns the deterministic orbit average into the coverage probability of the designated test placement.

1 Introduction

Split conformal prediction validates its coverage claim by a symmetry argument over sample slots. This note gives that argument an exact dual form and identifies the deterministic construction that lives beside it under the same group action. The identity behind conformal validity is a contragredient statement for the permutation action on slots (Section 3). A natural primal discrepancy construction under that action is *Steinitz balancing*: ordering a finite zero-sum population so that every prefix sum stays uniformly bounded (Steinitz, 1913). The centered conformal placement residuals are themselves a scalar Steinitz population (Section 5), and admit a sharper balanced ordering than the general lemma provides. On the function–measure pairing, the measure-side counterpart of calibration is balanced sampling and exact cubature (Section 6). Section 7 records how the same two-ingredient structure appears in the dependent-data floor of Barber and Pananjady (2026), and Section 8 separates three rates that the shared algebra can blur. The group-symmetry view of predictive inference is itself established: Dobriban and Yu (2025) build predictive inference on general distributional group symmetries, of which exchangeability is the permutation case. What we believe is new here is narrower: the identification of the centered placement-residual vector as a Steinitz population, and the balanced traversals it admits, at one level and at several simultaneously. Herding (Welling, 2009), whose $O(1/n)$ bookkeeping resembles the constructions here, can be viewed as a with-replacement relaxation of the finite balanced-permutation problem (Harvey and Samadi, 2014); it is confined to Remark 2 and a kernel question in Section 6.

2 Conformal calibration and placement

Fix $\alpha \in (0,1)$, write $p = 1-\alpha$, let $S_1, \ldots, S_n$ be calibration scores and S_{n+1} a test score, almost surely distinct, and set $k = \lceil p\,(n+1) \rceil$ and $N = n+1$. Assume throughout that $\alpha \geq 1/N$, equivalently $k \leq n$, so that the k-th order statistic of the calibration scores exists; for $\alpha < 1/N$ the usual convention $q_n = +\infty$ applies and coverage is trivial. Split conformal declares coverage when $S_{n+1} \leq q_n$, with $q_n = S_{(k)}$ the k-th smallest calibration score. The scores are taken exchangeable, as they are in split conformal conditionally on an independent proper training sample once the fitted score rule is held fixed.

Lemma 1 (Steering identity)**.** *With $q_n = S_{(k)}$, $h_q(s) = \mathbf{1}\{s \leq q\}$, and the scores distinct,*

$$\sum_{i=1}^{n} \big(h_{q_n}(S_i) - p\big) \;=\; k - pn \;\in\; [\,p,\; 1+p\,).$$

Proof. Exactly k calibration scores satisfy $S_i \leq S_{(k)}$; and $p(n+1) \leq k < p(n+1)+1$. □

The material of this section is standard (Papadopoulos et al., 2002; Vovk et al., 2005; Lei et al., 2018).

Proposition 1 (Counting identity)**.** *Let $s_1, \ldots, s_N$ be distinct reals and, for $i \in \{1, \ldots, N\}$, call placement i* covered *if s_i is at most the k-th smallest of $\{s_j\}_{j \neq i}$, writing $\chi_i(s) = \mathbf{1}\{s_i \leq Q_k(s_{-i})\}$ for the indicator. Then $\chi_i(s) = 1$ iff the rank of s_i among all N values is at most k, and consequently $\sum_{i=1}^{N} \chi_i(s) = k$.*

Proof. Placement i is covered iff at most $k-1$ of the other values are smaller than s_i, iff its overall rank is at most k; ranks are a permutation of $\{1, \ldots, N\}$. □

Corollary 1 (Marginal validity)**.** *If $(S_1, \ldots, S_N)$ is exchangeable with almost surely distinct components, then $\mathbb{P}\{S_N \leq q_n\} = k/N \geq 1-\alpha$.*

Proof. The coverage event is the event that placement N is covered. By exchangeability every placement indicator has the same expectation; average Proposition 1. □

The control-theoretic reading of the calibration step is established in the online literature: the level update of adaptive conformal inference (Gibbs and Candès, 2021) is integral feedback on realized miscoverage, and Angelopoulos et al. (2023) name the identification. There the controlled quantity is a time average along a sequence; Lemma 1 is the batch counterpart, a one-shot quantized correction: the rank count is exactly k, while the level error $k/n - p$ is quantized at $O(1/n)$.

3 The contragredient identity

Let $V = \mathbb{R}^N$, with $\mathfrak{S}_N$ acting by permuting slots, $(\rho(\pi)v)_j = v_{\pi^{-1}(j)}$, and let $\rho^\vee(\pi)f = f \circ \rho(\pi^{-1})$ be the contragredient action on the dual space V^* (for the dual representation and the averaging operator see, e.g., Serre, 1977). The coordinate functionals e_i^* and the prefix-sum functionals $L_t = \sum_{j \leq t} e_j^*$ are elements of V^*, and one has, exactly,

$$L_t\big(\rho(\pi)v\big) \;=\; \big(\rho^\vee(\pi^{-1})L_t\big)(v)\,: \tag{1}$$

permute the points holding the prefix functional fixed, or hold the points fixed and permute the functional. Averaging over the group, let $P = \frac{1}{N!}\sum_\pi \rho(\pi)$ be the Reynolds projector and $P^\vee = \frac{1}{N!}\sum_\pi \rho^\vee(\pi)$ its dual, so that

$$\langle P^\vee f,\, v\rangle \;=\; \langle f,\, Pv\rangle \qquad \text{for all } f \in V^*,\ v \in V. \tag{2}$$

For the permutation representation, $Pv = \bar v\,\mathbf{1}$ and $P^\vee e_N^* = \frac{1}{N}\mathbf{1}^*$, where $\mathbf{1}^* = \sum_{i=1}^N e_i^*$.

The object the projector acts on is not the score vector but the acceptance vector. Write $\mathbb{R}^N_{\neq}$ for the score vectors with distinct entries and $\chi : \mathbb{R}^N_{\neq} \to \{0,1\}^N \subset V$ for the placement-acceptance map of Proposition 1. Since $\chi_i(s)$ depends only on the rank of s_i among the entries of s, the map is equivariant:

$$\chi\big(\rho(\pi)s\big) \;=\; \rho(\pi)\,\chi(s) \qquad \text{for all } \pi \in \mathfrak{S}_N,\ s \in \mathbb{R}^N_{\neq}. \tag{3}$$

The distinction matters because χ is nonlinear and is defined here only on score vectors with distinct entries. The averaged score vector $Ps = \bar s\,\mathbf{1}$ leaves that domain and erases all rank information. One must therefore average the induced acceptance vectors $\chi(\rho(\pi)s)$, rather than apply χ after averaging the scores.

Theorem 1 (Orbit–contragredient form of conformal validity)**.** *For every $s \in \mathbb{R}^N_{\neq}$,*

$$\frac{1}{N!}\sum_{\pi\in\mathfrak{S}_N} \chi_N\big(\rho(\pi)s\big) \;=\; \langle e_N^*,\, P\chi(s)\rangle \;=\; \langle P^\vee e_N^*,\, \chi(s)\rangle \;=\; \frac{1}{N}\sum_{i=1}^N \chi_i(s) \;=\; \frac{k}{N}.$$

Consequently, whenever the law of S is invariant under a subgroup of $\mathfrak{S}_N$ acting transitively on the slots, in particular under exchangeability, the designated test placement has coverage probability exactly k/N.

Proof. The first equality is (3) and linearity of e_N^*; the second is (2); the third is $P^\vee e_N^* = \frac{1}{N}\mathbf{1}^*$; the fourth is Proposition 1. For a transitive subgroup $H \le \mathfrak{S}_N$, set $P_H = \frac{1}{|H|}\sum_{h\in H}\rho(h)$; transitivity gives $P_H v = \bar v\,\mathbf{1}$, so the same Reynolds calculation runs verbatim with P_H in place of P and pins each $\mathbb{E}\chi_i(S)$ to k/N. □

Define also the centered placement residual vector $r(s) = (\chi_i(s) - k/N)_{i\le N}$; Proposition 1 says $\langle \mathbf{1}^*, r(s)\rangle = 0$, i.e.

$$\langle P^\vee e_N^*,\, r(s)\rangle \;=\; 0 \qquad \text{for every distinct } s. \tag{4}$$

The identity is exact and requires no analogy; it is the defining identity of the dual representation, applied to the acceptance vector. It does not remove the distributional assumption: Theorem 1 is a statement about the whole orbit, and a transitive symmetry of the law, exchangeability being the standard case, is what makes the designated placement inherit the orbit average. The systematic development of predictive inference from distributional group symmetries is due to Dobriban and Yu (2025).

4 Steinitz balancing

A natural primal discrepancy construction under the same $\mathfrak{S}_N$ action is the ordering of a finite zero-sum population. Let E be a real d-dimensional normed vector space; finite dimensionality is essential, as no constant independent of N is available in general otherwise. If $z_1, \dots, z_N \in E$ satisfy $\sum_i z_i = 0$ and $\|z_i\| \le 1$, the Steinitz lemma (Steinitz, 1913) supplies a permutation π with every prefix bounded, $\|\sum_{j\le t} z_{\pi(j)}\| \le C_E$ for all t, with C_E depending on the space and norm but not on N (a bound linear in d holds for arbitrary norms; Grinberg and Sevast'yanov, 1980); hence running averages within C_E/t of zero. This is the bounded-ledger $O(1/t)$ phenomenon, obtained by an element of the same permutation group that Section 3 acts with.

Two precisions. First, the vector-valued setting scalarizes cleanly: on $W = E^N$ define, for each $\lambda \in E^*$, the functional $L_{t,\lambda}(z) = \lambda(\sum_{j \le t} z_j) \in W^*$; then (1) holds verbatim for $L_{t,\lambda}$, and $\| \sum_{j \le t} z_{\pi(j)} \| = \sup_{\|\lambda\|_* \le 1} \big| L_{t,\lambda}(\rho(\pi^{-1})z) \big|$ recovers the norm bound. Second, Steinitz balancing is not itself the contragredient of conformal averaging. Every permutation, balanced or not, satisfies the transport identity (1); the Reynolds average gives the exact orbit identity of Theorem 1; what Steinitz contributes is the choice of one π, data-dependent, that makes the whole nested family of prefix functionals small at once.

For the relation to herding, see Remark 2.

5 Balanced conformal placements

The centered conformal placement residuals form a scalar Steinitz population: by Proposition 1, $r_i(s) = \chi_i(s) - k/N$ satisfies $\sum_i r_i(s) = 0$ and $|r_i(s)| \le 1$. Because they take only two values, the ordering can be made sharper than the general lemma provides. The balancing step is a statement about binary vectors, so we state it that way; the conformal content is that Proposition 1 produces such a vector.

Lemma 2 (Balanced ordering of a binary population). *Let $b \in \{0,1\}^N$ with $\sum_i b_i = k$. There is a permutation $\pi \in \mathfrak{S}_N$ such that, simultaneously for every $t = 1, \dots, N$, $\big|\sum_{j \le t} b_{\pi(j)} - tk/N\big| \le \frac{1}{2}$.*

Proof. Define the balanced binary word $w_t = \lfloor tk/N + \frac{1}{2} \rfloor - \lfloor (t-1)k/N + \frac{1}{2} \rfloor$. Since $0 \le k/N \le 1$, each $w_t \in \{0,1\}$, and $\sum_{j \le t} w_j = \lfloor tk/N + \frac{1}{2} \rfloor$, so $| \sum_{j \le t} w_j - tk/N| \le \frac{1}{2}$ for every t, with exactly k ones among $w_1, \dots, w_N$. Send the indices with $b_i = 1$ to the positions with $w_t = 1$ and the rest to the rest. □

Corollary 2 (Balanced placement ordering). *For every distinct score vector s there is a permutation $\pi \in \mathfrak{S}_N$ such that, simultaneously for every $t = 1, \dots, N$,*

$$\Big| \frac{1}{t} \sum_{j=1}^{t} \chi_{\pi(j)}(s) - \frac{k}{N} \Big| \;\le\; \frac{1}{2t},$$

with equality to zero at $t = N$, and hence $\big|\frac{1}{t} \sum_{j \le t} \chi_{\pi(j)}(s) - p\big| < \frac{1}{2t} + \frac{1}{N}$ against the nominal level p.

Proof. Apply Lemma 2 to $b = \chi(s)$, using Proposition 1; the second bound adds $|k/N - p| < 1/N$. □

The lemma is one-dimensional balanced rounding; prefix bounds of this type are classical, going back to the chairman assignment problem (Tijdeman, 1980), and no specifically conformal structure enters once the count $\sum_i \chi_i(s) = k$ is known. The contribution is the identification: the conformal placement vector is an object to which balanced-word machinery applies, and the traversal it yields tracks the orbit level k/N, not the nominal p, exactly once all N candidate test placements have been traversed, and within $1/(2t)$ before that. The traversal orders the N coordinates $\chi_1(s), \dots, \chi_N(s)$, that is the N-element coordinate orbit $\mathfrak{S}_N/\mathrm{Stab}(N)$ (a coset space, $\mathrm{Stab}(N)$ not being normal), not the $N!$-element score orbit; the group average of Theorem 1 reduces to this N-placement average because each candidate test coordinate occurs equally often. In the notation of Section 3 the two readings of the traversal are literal: $\langle L_t, \rho(\pi^{-1}) r(s) \rangle = \langle \rho^\vee(\pi) L_t, r(s) \rangle$ — move the residual vector through the permutation holding the prefix test fixed, or hold the residuals fixed and move the nested family of prefix functionals.

The ordering is post hoc: constructing it requires knowing which placements are covered, hence the full score vector or its ranks. It is a deterministic traversal for orbit computations and repeated-placement audits, not an online algorithm, and it does not replace exchangeability in one-shot conformal prediction, where a single placement is designated as the test; only the distributional symmetry makes the designated coordinate inherit the orbit-average coverage probability.

Several coverage levels can be balanced at once, and here the vector-valued lemma of Section 4 is what does the work. The sharp one-dimensional $\frac{1}{2}$ bound does not in general survive at two levels simultaneously: with $N = 3$ and levels $k_1 = 1$, $k_2 = 2$, the $t = 1$ constraints force the first traversed rank to be 2, and then either completion violates $\frac{1}{2}$ at $t = 2$, so no ordering of the three placements meets the scalar bound at both levels, though each level alone admits two.

Proposition 2 (Simultaneously balanced levels)**.** *Fix levels $k_1, \dots, k_d \in \{1, \dots, N\}$ and set $z_i(s) = \big(\mathbf{1}\{\mathrm{rank}(s_i) \le k_\ell\} - k_\ell/N\big)_{\ell \le d} \in \mathbb{R}^d$. Then $\sum_i z_i(s) = 0$ and $\|z_i(s)\|_\infty \le 1$, so the Steinitz lemma supplies one permutation π with*

$$\max_{\ell \le d} \Big| \frac{1}{t} \sum_{j=1}^{t} \mathbf{1}\{\mathrm{rank}(s_{\pi(j)}) \le k_\ell\} - \frac{k_\ell}{N} \Big| \;\le\; \frac{C_d}{t} \qquad \text{for every } t = 1, \dots, N,$$

with C_d the Steinitz constant of $(\mathbb{R}^d, \|\cdot\|_\infty)$, independent of N.

Proof. Each coordinate of $z_i(s)$ lies in $[-1, 1]$. Ranks are a permutation of $\{1, \dots, N\}$, so $\sum_i \mathbf{1}\{\mathrm{rank}(s_i) \le k_\ell\} = k_\ell$ directly, for every $k_\ell \in \{1, \dots, N\}$, and summing coordinate ℓ over i gives $k_\ell - N \cdot k_\ell/N = 0$. Apply the Steinitz lemma in $(\mathbb{R}^d, \|\cdot\|_\infty)$ and divide the prefix bound by t. ☐

For $k_\ell \le N - 1$ the level is the placement-acceptance count of Proposition 1; $k_\ell = N$ is the trivial always-accepted rank level, whose reading as a conformal threshold needs the $+\infty$ endpoint convention.

A single traversal is therefore balanced against every selected orbit level simultaneously, at the price of the dimensional constant in place of the sharp $\frac{1}{2}$.

6 Balanced sampling: the measure-side counterpart

If the pairing of interest is $\langle h, \nu \rangle = \int h \, d\nu$ rather than the permutation representation, calibration has a measure-side counterpart obtained by reversing the roles of the two arguments. Calibration holds the empirical score measure ν_n fixed and chooses a test: find a threshold h_q with $\langle h_q, \nu_n \rangle = k/n$. The role-reversed problem holds tests $h_1, \dots, h_d$ and targets $m_1, \dots, m_d$ fixed and chooses a measure: find ν with $\langle h_j, \nu \rangle = m_j$. (This is an inverse moment problem under the pairing, not the transpose of an operator: the transpose under the pairing of the moment map $T(\nu) = (\langle h_j, \nu \rangle)_{j \le d}$ is $T^* a = \sum_j a_j h_j$, which lives on the function side.) Over a finite candidate population, fix weights $a_i > 0$ and set $A_{ji} = a_i h_j(x_i)$, $b = (m_1, \dots, m_d)$, and $\nu_z = \sum_{i=1}^{N} a_i z_i \delta_{x_i}$. Then $T(\nu_z) = Az$, so the measure-side problem is the selection problem

$$Az = b, \qquad z \in \{0, 1\}^N.$$

Horvitz–Thompson balancing is the case $a_i = 1/\pi_i$ with population totals in b. Equation $Az = b$ is the balancing constraint on a *realized* sample; a balanced sampling design is a distribution on $z \in \{0, 1\}^N$ with $\mathbb{E} z_i = \pi_i$ satisfying that constraint exactly when feasible and approximately otherwise, which is what the cube method delivers (Deville and Tillé, 2004). Normalized equal weights $1/\sum_i z_i$ do not fit the fixed-weight representation above, but for a nonempty selection the normalized constraints homogenize,

$$\frac{\sum_i z_i h_j(x_i)}{\sum_i z_i} = m_j \quad \Longleftrightarrow \quad \sum_i z_i\big(h_j(x_i) - m_j\big) = 0,$$

still linear in z; at fixed sample size M this is $\sum_i z_i h_j(x_i) = M m_j$. When arbitrary positive weights are allowed the problem changes and convex geometry answers it exactly. Let Q be a probability measure on $\mathcal{X}$ (the letter P is reserved for the Reynolds projector), let $h_1, \ldots, h_d : \mathcal{X} \to \mathbb{R}$ be measurable and Q-integrable, and set $\tilde{\phi} = (1, h_1, \ldots, h_d)$. By the generalized Tchakaloff theorem for measurable integrable maps (Tchakaloff, 1957; Bayer and Teichmann, 2006) there are $m \le d+1$ points $x_1, \ldots, x_m$ in any specified measurable Q-full set and weights $w_r > 0$ with $\int \tilde{\phi}\, dQ = \sum_{r=1}^m w_r \tilde{\phi}(x_r)$: the first coordinate gives $\sum_r w_r = 1$ and the rest give $\int h_j\, dQ = \sum_r w_r h_j(x_r)$ for every j. In this formulation, balanced sampling and exact cubature are the measure-side counterpart of threshold calibration.

A designed calibration set sits inside this frame. If designed scores $S_1, \ldots, S_n$, taken distinct, have empirical distribution function within Kolmogorov distance D of a continuous reference law $\bar{F}$, and the test score is drawn from $\bar{F}$ independently of the design, then $|\bar{F}(S_{(k)}) - k/(n+1)| \le D + 1/(n+1)$ and $|\bar{F}(S_{(k)}) - p| < D + (1+p)/n$, deterministically in the design (the proof is $\hat{F}_n(S_{(k)}) = k/n$ plus Lemma 1). The midpoint quantile grid $S_i = \bar{F}^{-1}\big((i - \frac{1}{2})/n\big)$, $i = 1, \ldots, n$, with $\bar{F}^{-1}(u) = \inf\{x : \bar{F}(x) \ge u\}$, achieves $D \le 1/(2n)$ for continuous $\bar{F}$. This is design-based rather than distribution-free: it suppresses the $n^{-1/2}$ fluctuation of calibration-conditional coverage at the price of assuming the target law. Two limitations should be stated. The formulation treats the score values as design variables; realizing such a design through selectable observations, whose scores are outputs of a predictor and unknown responses, is a separate problem. And no implication from the reproducing-kernel discrepancy guarantees of kernel herding or kernel thinning (Chen et al., 2010; Dwivedi and Mackey, 2021) to the Kolmogorov discrepancy required here is established.

7 The Barber–Pananjady floor

The two-ingredient structure of the preceding sections reappears in the dependent-data coverage floor of Barber and Pananjady (2026). For $S = (S_1, \ldots, S_N)$ they define subvectors $\Delta^0_{k,\tau}(S)$ and $\Delta^1_{k,\tau}(S)$ of length $N - \tau$, deleting a block of τ entries while keeping S_N (respectively rotating S_k) in the final slot, and set

$$\Psi_{k,\tau}(S) = d_{\mathrm{TV}}\big(\mathcal{L}(\Delta^0_{k,\tau}(S)), \mathcal{L}(\Delta^1_{k,\tau}(S))\big), \qquad \bar{\Psi}_\tau(S) = \frac{1}{N}\sum_{k=1}^N \Psi_{k,\tau}(S),$$

with $d_{\mathrm{TV}}(P, Q) = \sup_A |P(A) - Q(A)|$. Their Theorem 5 gives, for a pretrained memoryless score,

$$\mathbb{P}\{S_N \le q_n\} \;\ge\; 1 - \alpha - \min_{0 \le \tau \le n}\Big\{\tfrac{\tau}{N} + \bar{\Psi}_\tau(S)\Big\},$$

their Proposition 3 bounds the individual switch coefficients by $2\beta(\tau)$ for stationary β-mixing data, and, after data processing and averaging, their Corollary 6 inserts this control into the coverage bound. Their Theorem 8 is a corresponding upper bound with the same switch-coefficient structure, under almost surely distinct scores; their Theorem 7 shows tightness up to the factor $(1-\alpha)/4$ when

$(1-\alpha)N$ is integral and the data space is infinite, a finite space carrying an additional cardinality remainder.[1]

The argument separates into an exact insertion–deletion stability statement for empirical quantile *levels* (deleting τ entries moves the rank of a fixed threshold by at most τ, so the level shifts by at most τ/N: order-statistic geometry, controlling levels rather than values, since empirical quantile values are not stable under small total-variation perturbations) and a total-variation comparison between switched laws, which is the definition of the switch coefficient. Both ingredients are linear in their perturbation sizes and no concentration step appears; by comparison the blocking and empirical-process bound of Oliveira et al. (2024) has square-root dependence on the corresponding block-and-mixing expression, up to logarithmic factors. This decomposition is thematically parallel to the discrepancy constructions above, but it is not derived from the contragredient identity.

8 Three rates

The shared algebra can blur three quantities. The marginal coverage k/N of Corollary 1 is an exact symmetry identity; its $O(1/n)$ deviation from p is rank-grid rounding, not estimation error. The calibration-sample-conditional coverage $\mathbb{P}\{S_{n+1} \le q_n \mid S_1, \ldots, S_n\} = \bar{F}(q_n)$, defined when the test score is drawn from a continuous $\bar{F}$ independently of the calibration scores, fluctuates around nominal at the sampling scale: for $S_1, \ldots, S_n$ i.i.d. from $\bar{F}$ one has exactly $\bar{F}(S_{(k)}) \sim \mathrm{Beta}(k,\, n+1-k)$, with standard deviation $O(n^{-1/2})$ (Vovk et al., 2005). The design of Section 6 suppresses this at the price of assuming $\bar{F}$. Steinitz discrepancies are genuine approximation errors, $O(1/t)$ with constants from the Steinitz lemma. The ledgers match; the objects differ.

9 Remarks

Remark 1 (The pinball dual). Threshold calibration also has an exact convex dual. The conformal threshold $q_n = S_{(k)}$ is the empirical quantile at level $\gamma = k/n$, not generally at p, and it minimises the pinball loss (Koenker and Bassett, 1978) at that level, $\sum_i \rho_\gamma(S_i - q)$ with $\rho_\gamma(r) = \sup_{u \in [\gamma-1,\gamma]} ur$. Fenchel duality gives

$$\max\Big\{\textstyle\sum_i u_i S_i :\ u_i \in [\gamma-1, \gamma],\ \sum_i u_i = 0\Big\},$$

and with distinct scores $u_i = \gamma - h_{q_n}(S_i)$ is feasible for every i, including at the threshold, since $\sum_i u_i = n\gamma - k = 0$. The dual optimum is a balanced signed measure on the scores: calibration's exact optimization dual, though not itself a point-generation algorithm.

Remark 2 (Herding). Herding (Welling, 2009) generates points by the recursion $w_t = w_{t-1} + \mu - \phi(x_t)$, $x_t = \arg\max_x \langle w_{t-1}, \phi(x)\rangle$, and attains $O(1/n)$ feature-mean discrepancy conditional on bounded weights; the identification with Frank–Wolfe (Bach et al., 2012) shows that the standard fast-rate guarantee follows under interiority-type conditions, with Chen et al. (2010) treating the kernel case. Harvey and Samadi (2014) introduce a Permutation Problem, the Euclidean finite-population instance of the balancing problem of Section 4, and show that periodically repeating a good permutation yields a solution to their with-replacement Sampling Problem; generic herding runs with repeated selection (Welling, 2009; Chen et al., 2010), and on the pairing of Section 6 it is an iterative with-replacement method for approximate moment matching, or equal-weight cubature; this is related to, but not identical with, balanced sampling in the Deville–Tillé sense, which fixes

[1] Result numbering follows the ALT 2026 (PMLR volume 313) version of Barber and Pananjady (2026); the arXiv versions number the same results differently.

inclusion probabilities and keeps the 0–1 without-replacement structure. The quadrature triangle is older on that side: optimally-weighted herding is Bayesian quadrature (Huszár and Duvenaud, 2012), and Snell and Griffiths (2025) recently recast conformal risk and threshold calculations through Bayesian quadrature of the loss quantile function.

Remark 3 (Limits). The passage from the exact orbit identity to exact one-shot coverage is supplied by transitive invariance of the law, exchangeability in the standard case. Under dependence, Section 7 instead uses approximate switched-law symmetry and pays an explicit coverage penalty; no deterministic bookkeeping supplies either conclusion. Nor does control of thresholds and labels create information: it re-levels coverage while conditional sharpness remains a property of the scores (Cotton). We do not know whether this particular placement-balancing construction can be localized conditionally, selecting or reweighting calibration slots to balance the rank count given the input; localized, weighted, Mondrian and multivalid conformal methods address the conditional target by other means.

References

A. N. Angelopoulos, E. Candès, and R. J. Tibshirani. Conformal PID control for time series prediction. In *Advances in Neural Information Processing Systems 36*, 2023.

F. Bach, S. Lacoste-Julien, and G. Obozinski. On the equivalence between herding and conditional gradient algorithms. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*, pages 1355–1362, 2012.

R. F. Barber and A. Pananjady. Predictive inference for time series: why is split conformal effective despite temporal dependence? In *Proceedings of the 37th International Conference on Algorithmic Learning Theory (ALT)*, volume 313 of *Proceedings of Machine Learning Research*, pages 1–24, 2026. arXiv:2510.02471.

Christian Bayer and Josef Teichmann. The proof of tchakaloff's theorem. *Proceedings of the American Mathematical Society*, 134(10):3035–3040, 2006.

Y. Chen, M. Welling, and A. Smola. Super-samples from kernel herding. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 109–116, 2010.

P. Cotton. Marginally useful: Formalizing the information gap in conformal prediction. Companion paper.

J.-C. Deville and Y. Tillé. Efficient balanced sampling: the cube method. *Biometrika*, 91(4):893–912, 2004.

Edgar Dobriban and Mengxin Yu. SymmPI: Predictive inference for data with group symmetries. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(5), 2025. arXiv:2312.16160.

R. Dwivedi and L. Mackey. Kernel thinning. In *Proceedings of the 34th Conference on Learning Theory (COLT)*, 2021. arXiv:2105.05842.

I. Gibbs and E. Candès. Adaptive conformal inference under distribution shift. In *Advances in Neural Information Processing Systems 34*, pages 1660–1672, 2021.

V. S. Grinberg and S. V. Sevast'yanov. Value of the Steinitz constant. *Functional Analysis and Its Applications*, 14:125–126, 1980.

N. Harvey and S. Samadi. Near-optimal herding. In *Proceedings of the 27th Conference on Learning Theory (COLT)*, volume 35 of *Proceedings of Machine Learning Research*, pages 1165–1182, 2014.

F. Huszár and D. Duvenaud. Optimally-weighted herding is Bayesian quadrature. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 377–386, 2012. arXiv:1204.1664.

R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

R. I. Oliveira, P. Orenstein, T. Ramos, and J. V. Romano. Split conformal prediction and non-exchangeable data. *Journal of Machine Learning Research*, 25, 2024.

H. Papadopoulos, K. Proedrou, V. Vovk, and A. Gammerman. Inductive confidence machines for regression. In *Machine Learning: ECML 2002*, pages 345–356. Springer, 2002.

J.-P. Serre. *Linear Representations of Finite Groups*, volume 42 of *Graduate Texts in Mathematics*. Springer, 1977.

J. C. Snell and T. L. Griffiths. Conformal prediction as bayesian quadrature. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, 2025. arXiv:2502.13228.

E. Steinitz. Bedingt konvergente Reihen und konvexe Systeme. *Journal für die reine und angewandte Mathematik*, 143:128–176, 1913.

V. Tchakaloff. Formules de cubatures mécaniques à coefficients non négatifs. *Bulletin des Sciences Mathématiques*, 81:123–134, 1957.

R. Tijdeman. The chairman assignment problem. *Discrete Mathematics*, 32:323–330, 1980.

V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

M. Welling. Herding dynamical weights to learn. In *Proceedings of the 26th International Conference on Machine Learning (ICML)*, pages 1121–1128, 2009.