The Width of the Conformal Fan: Dependence and the Variance of Realized Coverage

Peter Cotton

Draft — companion note

Abstract

Split conformal prediction's realized, calibration-conditional coverage fluctuates around the nominal $1-\alpha$. For exchangeable-but-independent calibration scores it follows the $\mathrm{Beta}(k, n-k+1)$ law, variance about $\alpha(1-\alpha)/n$, the classical fan. We show the width of that fan is governed by the sign of the cross-sample dependence. The mean stays pinned at the marginal level whatever the dependence; the variance, to leading order, is the average pairwise covariance of the exceedance indicators at the operating quantile times the independent fan. Positive (extendable) dependence adds a between-dataset term and widens the fan; negative dependence narrows it, to exactly zero at the maximally negatively associated contest floor $\rho = -1/(n-1)$, where the realized coverage equals the nominal level identically. We prove the positive side exactly through de Finetti and the law of total variance, the negative side exactly at the floor and to leading order in general, and reduce the remaining finite-sample inequality to a convex-order contraction of the exceedance count, supported numerically across dependence structures. The governing sign is the one the companion note attaches to the finite de Finetti measure: a genuine prior widens the fan, the signed corner collapses it.

1 Setup

Split conformal returns a set with marginal coverage $1-\alpha$. Fix the calibration set; the coverage you then realize on fresh test points is a random variable, and it is not $1-\alpha$ on the nose. Write the scores on the probability-integral scale, $U_i = F(S_i)$, so each U_i is marginally uniform whatever the model. With $k = \lceil (n+1)(1-\alpha) \rceil$ the conformal threshold is the order statistic $U_{(k)}$, and the calibration-conditional coverage against an independent test draw from the marginal is

$$c = U_{(k)}. \tag{1}$$

For independent (or merely exchangeable, de facto independent) calibration scores this is the classical result (Vovk, 2012; Marques F., 2024): $c \sim \mathrm{Beta}(k, n-k+1)$, with

$$\mathbb{E}[c] = \frac{k}{n+1}, \qquad \mathrm{Var}(c) = \frac{k(n-k+1)}{(n+1)^2(n+2)} \;\approx\; \frac{\alpha(1-\alpha)}{n}. \tag{2}$$

This is the fan: the per-dataset coverage scatters around $1-\alpha$ with a spread that shrinks like $1/n$. The question of this note is what the dependence among the calibration scores does to that spread. The companion notes treat the orthogonal, within-dataset axis (coverage uneven across x for a fixed dataset, the information gap $I(R; X)$ and its sign Cotton; Cotton); here every statement is about the across-dataset fluctuation of the single number c.

The whole problem passes through the count process $N(u) = \sum_{i=1}^{n} \mathbf{1}\{U_i \le u\}$, because $U_{(k)} = \int_0^1 \mathbf{1}\{N(u) < k\}\,\mathrm{d}u$, so

$$\mathrm{Var}(c) = \int_0^1 \int_0^1 \mathrm{Cov}\left(\mathbf{1}\{N(u) < k\},\, \mathbf{1}\{N(v) < k\}\right) \mathrm{d}u\, \mathrm{d}v, \tag{3}$$

a functional of the law of the count process alone. Two facts about N drive everything.

Lemma 1 (Count variance)**.** *If the scores are negatively associated (Joag-Dev and Proschan, 1983), then for every u, $\mathrm{Var}(N(u)) \le n\,u(1-u)$, with equality iff the exceedance indicators are pairwise uncorrelated.*

Proof. The indicators $\mathbf{1}\{U_i \le u\}$ are coordinatewise-monotone functions of single distinct coordinates, hence negatively associated, hence pairwise non-positively correlated. So $\mathrm{Var}(N(u)) = \sum_i \mathrm{Var} + \sum_{i \ne j} \mathrm{Cov} \le \sum_i u(1-u) = n\,u(1-u)$. □

2 The fan coefficient

Let $p = 1 - \alpha$ be the operating level and $\xi_i = \mathbf{1}\{U_i \le p\}$, so $N(p) = \sum_i \xi_i$. The Bahadur representation for the uniform quantile (Bahadur, 1966), which holds for negatively associated sequences under mild conditions, gives $U_{(k)} - p = p - N(p)/n + o_P(n^{-1/2})$, and hence

Theorem 1 (Fan coefficient)**.** *As $n \to \infty$,*

$$\mathrm{Var}(c) = \frac{1}{n^2}\,\mathrm{Var}(N(p))\,(1 + o(1)) = \frac{\alpha(1-\alpha)}{n}\,R\,(1 + o(1)), \qquad R = 1 + \frac{n-1}{\alpha(1-\alpha)}\,\bar{c}, \tag{4}$$

where $\bar{c} = \frac{1}{n(n-1)} \sum_{i \ne j} \mathrm{Cov}(\xi_i, \xi_j)$ is the average pairwise covariance of the exceedance indicators at the operating quantile.

The fan's leading coefficient is, exactly, the variance of the exceedance count at $1 - \alpha$. Under independence $\bar{c} = 0$ and $R = 1$, recovering (2). Negative association makes $\bar{c} < 0$ and $R < 1$: the fan narrows. At the contest floor $\bar{c} = -\alpha(1-\alpha)/(n-1)$ one gets $R = 0$. The reduction factor R is the single number to carry away.

3 Positive dependence widens the fan, exactly

When the scores are extendable, that is, drawn from an infinitely exchangeable sequence, de Finetti (de Finetti, 1937; Kerns and Székely, 2006) makes them conditionally independent given a latent directing measure Q. Conditional on Q the coverage c is an ordinary independent-sample order statistic, so the law of total variance gives an exact decomposition.

Theorem 2 (Positive dependence)**.** *For an extendable exchangeable calibration sequence with directing measure Q,*

$$\mathrm{Var}(c) = \underbrace{\mathbb{E}_Q\big[\,\mathrm{Var}(c \mid Q)\big]}_{\textit{average within-dataset fan}} + \underbrace{\mathrm{Var}_Q\big(\mathbb{E}[c \mid Q]\big)}_{\textit{between-dataset term}}. \tag{5}$$

The between-dataset term is non-negative and vanishes under independence. Positive cross-sample dependence is exactly this term: a per-dataset latent that shifts the conditional quantile, on top of the ordinary fan.

The decomposition is exact and needs no asymptotics. It says the danger of positive dependence is not bias but variance: you draw a dataset, the latent Q places the whole calibration sample high or low, and the realized coverage rides with it. Numerically, a one-factor Gaussian model at pairwise correlation 0.2 has $\mathrm{Var}(c)$ about 3.3 times the independent fan, of which the between-dataset term is more than half.

4 Negative dependence narrows the fan

The negative regime is the one a genuine prior cannot reach. By Kerns and Székely (2006) a finite exchangeable sequence with $\rho < 0$ has a *signed* directing measure, so the clean conditioning of Theorem 2 is unavailable: there is no Q to condition on, and the between-dataset term would be a negative quasi-variance. That is precisely the regime where the fan contracts.

Theorem 3 (Endpoints). *Independent scores give $c \sim \mathrm{Beta}(k, n-k+1)$, the widest fan at fixed marginals. At the maximally negatively associated floor, where the calibration scores are a uniformly random permutation of the grid $\{1/(n+1), \ldots, n/(n+1)\}$ (sampling without replacement, pairwise correlation $-1/(n-1)$), the order statistic is deterministic, $U_{(k)} = k/(n+1)$ almost surely, so $\mathrm{Var}(c) = 0$: the realized coverage equals the nominal level on every draw.*

Both endpoints agree with Theorem 1 at $R = 1$ and $R = 0$. Between them the fan contracts monotonically as the scores become more negatively dependent, and the contraction is exactly the $R < 1$ of the fan coefficient. We can prove this contraction exactly at the endpoints and to leading order everywhere; the exact finite-sample inequality is a statement about the spread of an order statistic under negative dependence.

Conjecture 1 (Finite-sample contraction). *If the calibration scores are negatively associated then $\mathrm{Var}(c)$ is at most the independent value* (2)*.*

One ingredient is already exact. The exceedance indicators are negatively associated with $\mathrm{Bernoulli}(u)$ marginals, so by Shao (2000) the count is convex-order dominated by the independent binomial, $N(u) \leq_{\mathrm{cx}} \mathrm{Bin}(n, u)$, with the same mean nu: a mean-preserving contraction at every level. Carrying this to $\mathrm{Var}(c)$ needs the level comparison $\mathbb{P}(N(u) \geq k)$ to cross the binomial tail once in u, which the total positivity of the binomial tail in (k, u) (Karlin, 1968) makes plausible, followed by a single-crossing-implies-less-spread step. We claim only the variance statement, and deliberately not the stronger dispersive ordering: an order statistic of a dependent sample need not be dispersively smaller than the independent one even at fixed marginals (Panja et al., 2022), so the result is specific to the variance and is not a corollary of a dispersive theorem. The natural sufficient condition on the dependence is negative supermodular dependence (Christofides and Vaggelatou, 2004), which all the structures below satisfy. We verify the conjecture across them (one run each, reproduced by `check_fan.py`; $n = 20$, $\alpha = 0.1$, independent fan variance 3.92×10^{-3}).

calibration dependence	$\mathbb{E}[c]$	$\mathrm{Var}(c)$	ratio to fan	sign
independent ($\rho = 0$)	0.904	3.92×10^{-3}	1.00	—
exchangeable, $\rho = -1/(n-1)$	0.923	1.16×10^{-3}	0.29	neg. assoc.
MA(1), lag-1 corr $-\frac{1}{2}$	0.912	2.71×10^{-3}	0.69	neg. assoc.
random negatively-paired blocks	0.909	3.14×10^{-3}	0.80	neg. assoc.
contest / without replacement	0.905	0	0.00	floor
one-factor, $\rho = +0.2$	0.863	1.33×10^{-2}	3.41	positive

Every negatively associated structure, exchangeable or not, comes in under the independent fan; the positive control widens it. The mean drifts up slightly under negative dependence and down under positive dependence, but stays near $1-\alpha$; the action is in the variance.

Remark 1 (A closed form, and what an exact proof must dominate)**.** *The variance is an explicit functional of $g(u) = \mathbb{P}(N(u) \geq k)$,*

$$\mathrm{Var}(c) = 2\int_0^1 (1-u)\, g(u)\, \mathrm{d}u - \Big(\int_0^1 g(u)\, \mathrm{d}u\Big)^2, \tag{6}$$

so writing $\delta = g - g^{\mathrm{iid}}$, $A = \int_0^1 g^{\mathrm{iid}} = 1 - \mathbb{E}[c^{\mathrm{iid}}]$, and $\Delta = \int_0^1 \delta$,

$$\mathrm{Var}(c) - \mathrm{Var}(c^{\mathrm{iid}}) = 2\int_0^1 \big[(1-u) - A\big]\, \delta(u)\, \mathrm{d}u - \Delta^2. \tag{7}$$

The $-\Delta^2$ term is non-positive. The weight $(1-u)-A$ changes sign at $u = \mathbb{E}[c^{\mathrm{iid}}] = k/(n+1)$, and under negative association δ changes sign once, at the level $u \approx k/n$ where the operating quantile crosses the count mean and the contraction of Lemma 1 flips the tail comparison. The two sign changes agree up to the conformal offset between k/n and $k/(n+1)$, a band of width $O(1/n)$ on which the integrand is positive; everywhere else it is negative. So the identity proves the contraction up to that single $O(1/n)$ boundary term, a second route to the leading order of Theorem 1, and isolates exactly what a finite-sample proof must control: the boundary remainder against the slack Δ^2.

Remark 2 (The signed measure as a negative between-term)**.** *Theorem 2 reads the fan as within-dataset variance plus a non-negative between-dataset variance. In the signed-de-Finetti picture of the companion note, the negative corner is exactly where that between-dataset object turns negative, a Feynman–Wigner quasi-variance that subtracts from the within-dataset fan rather than adding to it, reaching the full cancellation $R = 0$ at the contest floor. This is heuristic, since a signed measure has no honest conditional variance, but it is the right cartoon: the same sign that decides conditional adaptivity in Cotton decides whether the fan is inflated or cancelled here.*

5 A constructive corollary

If negative dependence narrows the fan, one can engineer it. A balanced calibration set whose scores behave like the grid of Theorem 3 concentrates the realized coverage far inside the independent fan, so the same reliability is bought with fewer calibration points, or the same number of points gives less run-to-run wobble. The classical variance-reduction designs all apply in principle: stratification never increases variance (Cochran, 1977), antithetic pairing helps for the monotone empirical CDF, and determinantal or repulsive designs give negative association with faster-than-independent concentration (Hough et al., 2006; Bardenet and Hardy, 2020).

Remark 3 (The exchangeability constraint)**.** *The marginal guarantee needs each calibration score exchangeable with the test score; it does not need the calibration scores independent of one another. So negative association is admissible, but only if it is induced by a map symmetric across all $n+1$ points, a pooled de-meaning or ranking, which preserves the exchangeability of the joint. Selecting or curating the calibration set by its scores, the obvious way to balance it, breaks exchangeability with the test point and voids the guarantee. The viable construction is transductive and symmetric; that is also why the contest grid, a symmetric function of the pooled sample, both keeps the marginal guarantee and zeroes the fan.*

6 Related work

The Beta$(k, n-k+1)$ law of calibration-conditional coverage and its variance are Vovk (2012) and Marques F. (2024); the entire prior literature treats the calibration scores as independent or exchangeable and reads off a single number, never varying the dependence to move it. Negative association and its consequences are Joag-Dev and Proschan (1983); that negatively associated sums are dominated in convex order by independent ones is Shao (2000), and sampling without replacement concentrates at least as tightly as with replacement (Serfling, 1974). The total positivity behind the single-crossing step is Karlin (1968); that an order statistic of a dependent sample need not be dispersively smaller than the independent one, which is why we claim only the variance, is Panja et al. (2022); the connection between supermodular ordering and negative dependence is Christofides and Vaggelatou (2004); the Bahadur representation is Bahadur (1966). Determinantal and repulsive designs are Hough et al. (2006); Bardenet and Hardy (2020) and stratification is Cochran (1977). The finite signed de Finetti representation is Kerns and Székely (2006), and the within-dataset, across-x reading of its sign is the companion note (Cotton); the information gap is (Cotton). The contribution here is to make the dependence the control variable for the *variance* of realized coverage: the exact fan coefficient R, the exact de Finetti decomposition for the positive side, the exact zero at the contest floor, and the reduction of the general negative case to a dispersive ordering.

References

Bahadur, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics* 37(3):577–580.

Bardenet, R. and Hardy, A. (2020). Monte Carlo with determinantal point processes. *The Annals of Applied Probability* 30(1):368–417. arXiv:1605.00361.

Christofides, T. C. and Vaggelatou, E. (2004). A connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis* 88(1):138–151.

Cochran, W. G. (1977). *Sampling Techniques*, 3rd ed. Wiley.

Cotton, P. Marginally Useful: Formalizing the Information Gap in Conformal Prediction. Companion paper.

Cotton, P. A Feynman–Wigner-Style Diagnostic for the Efficacy of Conformal Prediction via Signed de Finetti Representations. Companion note.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7(1):1–68.

Hough, J. B., Krishnapur, M., Peres, Y. and Virág, B. (2006). Determinantal processes and independence. *Probability Surveys* 3:206–229.

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables, with applications. *The Annals of Statistics* 11(1):286–295.

Kerns, G. J. and Székely, G. J. (2006). De Finetti's theorem for abstract finite exchangeable sequences. *Journal of Theoretical Probability* 19(3):589–608.

Marques F., P. C. (2024). Universal distribution of the empirical coverage in split conformal prediction. *Statistics & Probability Letters* 219:110350 (2025). arXiv:2303.02770.

Serfling, R. J. (1974). Probability inequalities for the sum in sampling without replacement. *The Annals of Statistics* 2(1):39–48.

Karlin, S. (1968). *Total Positivity, Vol. 1*. Stanford University Press.

Panja, A., Kundu, P. and Pradhan, B. (2022). Dispersive and star ordering of sample extremes from dependent random variables following the proportional odds model. Preprint, arXiv:2006.04454.

Shao, Q.-M. (2000). A comparison theorem on moment inequalities between negatively associated and independent random variables. *Journal of Theoretical Probability* 13(2):343–356.

Vovk, V. (2012). Conditional validity of inductive conformal predictors. *Asian Conference on Machine Learning*, PMLR 25:475–490.