A Feynman–Wigner-Style Diagnostic for the Efficacy of Conformal Prediction via Signed de Finetti Representations

Peter Cotton

Draft — companion note

Abstract

Conformal prediction builds prediction sets that cover the truth at a rate you choose, finite-sample and distribution-free, assuming only exchangeable data. That guarantee is marginal. de Finetti's theorem describes the exchangeability it rests on, and in the finite form the mixing measure may be signed (Kerns and Székely, 2006). We present a short lemma that decomposes the slope of conformal's calibration-conditional coverage into two terms. One is a non-negative threshold term, the classical Beta-law fan. The other carries the sign of the de Finetti measure. Positive (extendable) mixtures make conformal conditionally adaptive. The signed corner—de-meaned, ranked, or compositional scores—makes it anti-adaptive. The marginal guarantee is the same either way. Only what it hides changes.

1 Setup

Split conformal returns a set $C(X)$ with $\mathbb{P}\big(Y \in C(X)\big) \geq 1 - \alpha$ for any distribution and any sample size, provided the data are exchangeable. That marginal guarantee is the whole appeal, and it is not in dispute here. The question is what it says about the case in front of you.

Scores $S_1, \ldots, S_n$ (calibration) and S_{n+1} (test) are jointly exchangeable with a continuous joint law (no ties). The conformal width is the order statistic $\hat{q} = S_{(k)}$ of the calibration scores, $k = \lceil (n+1)(1-\alpha) \rceil$, and a test point is covered when $S_{n+1} \leq \hat{q}$.

Two kinds of dependence live in any such problem, and they should not be confused.

- Within a row: inside a single (x, y), does the residual's shape depend on x? That is the subject of the companion paper (Cotton), where the log-score regret of a single-shape conformal system equals the mutual information $I(R; X)$. It is a property of one observation's law, present even when the rows are i.i.d.
- Across rows: how do distinct observations relate to one another, whether independent, sharing a latent cause, or mutually constrained? That is the subject of de Finetti's theorem, and of this note.

The two axes are orthogonal. This note is entirely about the second.

2 Finite de Finetti and the sign of the mixture

de Finetti's theorem pins down what infinite exchangeability is.

Theorem 1 (de Finetti; de Finetti, 1937; Hewitt and Savage, 1955)**.** *Let $X_1, X_2, \dots$ be an infinitely exchangeable sequence in a Polish space. Then there is a probability measure π on the space of distributions such that, for every n, the law of $(X_1, \dots, X_n)$ equals $\int P^{\otimes n}\,\mathrm{d}\pi(P)$. Equivalently, conditional on a random measure $\Theta \sim \pi$, the X_i are i.i.d. with law Θ.*

The mixing measure π is a genuine prior, and that forces a sign. A positive mixture has $\mathrm{Cov}(S_i, S_j) = \mathrm{Var}\big(\mathbb{E}[S_i \mid \Theta]\big) \geq 0$, so de Finetti mixtures are non-negatively correlated. Finite exchangeable sequences are not so constrained. Exchangeability alone forces only $\rho \geq -1/(n-1)$ (Aldous, 1985), and the negative band is real—sampling without replacement sits at its floor.

The finite theory completes the picture. Diaconis and Freedman (1980) show a finite exchangeable sequence is approximately a positive i.i.d. mixture, within total variation $O(k/n)$. Kerns and Székely (2006) make it exact, at the price of signing.

Theorem 2 (Kerns–Székely; Kerns and Székely, 2006)**.** *Let $(X_1, \dots, X_n)$ be exchangeable on a measurable space. Then there is a finite signed measure μ of total mass 1 on the space of distributions with $\mathrm{law}(X_1, \dots, X_n) = \int P^{\otimes n}\,\mathrm{d}\mu(P)$. One may take $\mu \geq 0$ (a genuine prior, recovering Theorem 1) if and only if the sequence extends to an infinitely exchangeable one (Leonetti, 2018).*

We picture this as the i.i.d. laws tracing a curve: positive mixtures are its convex hull, the extendable $\rho \geq 0$ laws, and signed mixtures its affine span, everything. A signed representation marks non-extendability—a finite exchangeable sequence that cannot be continued to an infinite one—which is the negative-correlation regime.

We frame the signed measure as a negative probability, the kind Feynman (1987) used partway through a calculation, or Wigner (1932)'s phase-space distribution, which dips below zero while every observable marginal stays a genuine probability (the idea reached statistics through Bartlett, 1945). You cannot sample a parameter from it. That is why the signed corner gives a diagnosis, not a model: recovering a latent means fitting a real one, not reading the quasi-probability off.

So the sign of the de Finetti mixing measure is a clean label: positive = extendable, non-negatively correlated, a genuine prior. Signed = finite, negatively correlated, no genuine prior.

3 Marginal coverage ignores the sign

For continuous exchangeable scores the rank of S_{n+1} among all $n+1$ is uniform, and $S_{n+1} \leq S_{(k)}$ iff that rank is at most k, so

$$\mathbb{P}\big(S_{n+1} \leq \hat{q}\big) = \frac{k}{n+1} \geq 1 - \alpha. \tag{1}$$

This uses nothing but exchangeability (Vovk et al., 2005). Positive or signed, the marginal number is the same. The sign enters only when we ask what the certificate hides.

4 The conditional-coverage decomposition

Condition on the realised width. For a value q of the width $\hat{q}$ and a threshold s, write $H(s, q) = \mathbb{P}(S_{n+1} \leq s \mid \hat{q} = q)$. Coverage holds when $S_{n+1} \leq \hat{q}$, so coverage conditional on the realised width is the diagonal $c(q) = H(q, q)$, where the threshold equals the width.

Lemma 1 (Conditional-coverage slope)**.** *If $(S_{n+1}, \hat{q})$ has a joint density, then*

$$c'(q) = \underbrace{f_{S_{n+1}\mid\hat{q}}(q \mid q)}_{\text{(i) threshold}} + \underbrace{\partial_q\,\mathbb{P}(S_{n+1} \leq q \mid \hat{q} = q)}_{\text{(ii) dependence}}. \tag{2}$$

Term (i) *is non-negative. Term* (ii) *is* ≤ 0 *when* S_{n+1} *is stochastically increasing in* $\hat{q}$*,* $= 0$ *under independence, and* ≥ 0 *when stochastically decreasing.*

Proof. Total derivative of $H(q, q)$ along the diagonal: $c'(q) = \partial_s H(s, q)\big|_{s=q} + \partial_q H(s, q)\big|_{s=q}$. The first equals $f_{S_{n+1}|\hat{q}}(q \mid q) \geq 0$. For the second, $\partial_q H(s, q) \leq 0$ for all s exactly when $\mathbb{P}(S_{n+1} \leq s \mid \hat{q} = q)$ is non-increasing in q, the definition of S_{n+1} being stochastically increasing in $\hat{q}$. The reverse inequality is stochastic decrease. □

Remark 1 (The independence case is classical)**.** *Under independence term (ii) vanishes and* $c(q) = F(q)$*, the test-score CDF, which already rises with the width purely from estimation. This is the known law of calibration-conditional coverage:* $F(\hat{q}) = F(S_{(k)}) \sim \mathrm{Beta}(k, n-k+1)$ *(Vovk, 2012; Marques F., 2024; Angelopoulos and Bates, 2023). Lemma 1 is its extension to dependent scores, with term (ii) the new piece.*

Remark 2 (Reading the sign)**.** *The sign of the cross-sample dependence sets the direction. Positive dependence makes conformal conditionally adaptive, negative dependence makes it anti-adaptive. The mechanism is short. The width* $\hat{q} = S_{(k)}$ *is non-decreasing in the calibration scores, and the test score is disjoint from them. A positive (proper) de Finetti mixture makes the score vector associated (Esary et al., 1967), so* $\hat{q}$ *and* S_{n+1} *move together: term (ii) is* ≤ 0*, it cancels the threshold term, and pooling the calibration scores conditions you into the realised world. A signed, negatively-correlated law makes the vector negatively associated (Joag-Dev and Proschan, 1983) (without-replacement samples, multinomials, ranks, compositional vectors), so term (ii) is* ≥ 0 *and adds to it: the calibration set anti-predicts the test, and the per-case coverage swings high and low while the marginal stays pinned. The order-statistic version uses a difference in place of the derivative and behaves the same.*

The reading is a go/no-go on conformal's coverage, and coverage only. The residual-information gap of the companion paper sits there whatever the sign, so even at its best this is an honest certificate, not a sharp forecast. With independent or positively-dependent data the certificate is at least conditionally honest. With negatively-dependent data it is not: the number holds on average while the per-case coverage swings high and low. The sharpness comes from modelling the latent, never from the wrapper.

5 Why the signed corner is not exotic: constrained scores

The negative corner arises whenever the nonconformity score is defined relative to the batch, because a linear constraint forces it. De-meaning ($\sum_i s_i = 0$), differencing, ranks or percentiles-within-batch, and compositional or budgeted targets all impose such a constraint, and the constraint is the negative association (Joag-Dev and Proschan, 1983). For these scores conformal remains marginally valid—if the relativisation is symmetric across calibration and test, exchangeability is preserved—but its conditional coverage is the fanning, anti-adaptive kind of Lemma 1.

A caution: if the relativisation treats the test point asymmetrically (de-meaning by a training-only mean, say), exchangeability itself fails and even the marginal guarantee is no longer assured. This is the familiar reason split conformal uses out-of-fold rather than in-sample residuals (Lei et al., 2018); in-sample regression residuals are both constrained ($\sum_i e_i = 0$) and heteroscedastic (variance $1 - h_{ii}$), and so are not exchangeable to begin with.

Remark 3 (Rank, choice, and market-implied scores)**.** *The signed corner is the native habitat of contest data. In a Thurstone or Luce choice model a field of competitors carries latent values and* *the observed outcome is a winner or a full ranking. The winner indicator is a single multinomial trial, and the rank vector a permutation law, both negatively associated (Joag-Dev and Proschan, 1983). Any nonconformity score read off such a field—a rank, a within-field relative score, or a market-implied win probability—therefore sits in the negatively-associated regime above, where the marginal certificate is least informative about conditional reliability. The converse direction is the modelling the certificate cannot supply: recovering latent means from observed prices or winning probabilities (Cotton, 2021) is exactly the de Finetti latent-recovery of Section 2. In betting terms the market price is the crowd's Q, the recovered latent mean is an estimate of the true P, and the divergence between them is the rent of the companion paper. So the same contest data that lands conformal in its anti-adaptive corner is also where the de Finetti latent, once estimated, pays.*

6 A runnable detector

The gap is defined through the log-score, so it is not something you read off data. Its energy-statistics counterpart is, and it gives a test you can run on any fitted conformal predictor.

Lemma 2 (A distance-covariance detector)**.** *Let $U = \widehat{G}(R)$ be the conformal rank of the residual among the calibration residuals. Then U is marginally uniform, and $\mathrm{dCov}(U, X) = 0$ iff $U \perp X$ iff $R \perp X$ iff $I(R; X) = 0$. The permutation test that reshuffles the U labels against X is a valid test of $R \perp X$, and rejecting it certifies $I(R; X) > 0$: the conformal rank is conditionally non-uniform, so the marginal coverage is uneven across x.*

Two reasons this is the right object. U is standardized to uniform whatever the model, so the test probes only the conditional structure conformal cannot see. And distance covariance (Székely et al., 2007) catches dependence in any moment: under symmetric heteroscedastic noise the conditional mean of U stays at $\frac{1}{2}$ while its spread moves with x, so a correlation test sees nothing and the distance covariance sees it.

A small experiment makes the point (one run, reproduced by `check_dcov.py`). The residual is $R = \sigma(X)\,\varepsilon$ with the location correct, so any dependence is pure conditional shape.

regime	marginal coverage	$\mathrm{dCov}(U, X)$	permutation p	Pearson
homoscedastic	0.90	0.001	0.33	+0.02
heteroscedastic, mild	0.89	0.011	0.002	+0.05
heteroscedastic, strong	0.90	0.023	0.002	−0.03

Marginal coverage is on target throughout, so it cannot separate the regimes. The test fires only where the gap is positive, and the Pearson column shows that a correlation test would miss it.

The idea of a dependence measure as a conditional-coverage diagnostic is due to Feldman et al. (2021), who use HSIC between the miscoverage indicator and the interval length, and Braun et al. (2025), who classify the indicator on x. Testing the rank against the features is the rank-based, level-free variant, and it is the energy counterpart of the gap: $I(R; X)$ is the dependence of R and X in KL (Cover and Thomas, 2006), and the distance covariance, equivalently HSIC (Sejdinovic et al., 2013), is the same dependence in energy. The paper's gap is the one you prove; this is the one you run.

It detects, it does not certify. A large p means nothing was found at this power, not conditional validity, by the no-go results discussed in the companion paper (Cotton). The ranks share the calibration ECDF, so leave-one-out or full-conformal ranks make the null exact. The test ships as `conformalguide.guardrails`.

7 Related work

The marginal-coverage fact is standard (Vovk et al., 2005). The Beta law for coverage given the calibration set is Vovk (2012), refined by Marques F. (2024); training-conditional coverage is Bian and Barber (2023). Magnitude-based bounds for dependent data are Barber et al. (2023) and Oliveira et al. (2024). Association and negative association are Esary et al. (1967) and Joag-Dev and Proschan (1983); finite and signed de Finetti are Diaconis and Freedman (1980) and Kerns and Székely (2006); Leonetti (2018).

Barber and Pananjady (2026) are closest: they show conformal can over- or under-cover under temporal dependence. That account is marginal, it grades a departure by the size of the dependence rather than its sign, and it does not condition on the calibration set. The sign, and its identification with the sign of the finite de Finetti measure, is what this note adds.

de Finetti describes a law. Conformal runs without one. Using a de Finetti measure means estimating it, and that estimation is ill-posed: the latent is recovered from finite data, so in general one regularizes, shrinking toward the field. Conformal skips this, and the residual-information gap is the price of skipping it. A signed measure gives nothing to estimate from at all, so the representation tells you how conformal behaves rather than standing in for it. Datta et al. (2025) argue conformal is not a de Finetti conditional at all.

References

Aldous, D. J. (1985). Exchangeability and related topics. *École d'Été de Probabilités de Saint-Flour XIII*, Lecture Notes in Mathematics 1117, Springer.

Angelopoulos, A. N. and Bates, S. (2023). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *Foundations and Trends in Machine Learning*. arXiv:2107.07511.

Barber, R. F., Candès, E. J., Ramdas, A. and Tibshirani, R. J. (2023). Conformal prediction beyond exchangeability. *The Annals of Statistics* 51(2):816–845.

Barber, R. F. and Pananjady, A. (2026). Predictive inference for time series: why is split conformal effective despite temporal dependence? Preprint, arXiv:2510.02471.

Bartlett, M. S. (1945). Negative probability. *Mathematical Proceedings of the Cambridge Philosophical Society* 41(1):71–73.

Bian, M. and Barber, R. F. (2023). Training-conditional coverage for distribution-free predictive inference. *Electronic Journal of Statistics* 17(2):2044–2066.

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

Cotton, P. (2021). Inferring relative ability from winning probability in multi-entrant contests. doi:10.1137/19M1276261.

Datta, J., Polson, N. G., Sokolov, V. and Zantedeschi, D. (2025). Conformal Prediction = Bayes? Preprint, arXiv:2512.23308.

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

Diaconis, P. and Freedman, D. (1980). Finite exchangeable sequences. *The Annals of Probability* 8(4):745–764.

Esary, J. D., Proschan, F. and Walkup, D. W. (1967). Association of random variables, with applications. *The Annals of Mathematical Statistics* 38(5):1466–1474.

Feynman, R. P. (1987). Negative probability. In Hiley, B. J. and Peat, F. D. (eds.), *Quantum Implications: Essays in Honour of David Bohm*, 235–248. Routledge & Kegan Paul.

Hewitt, E. and Savage, L. J. (1955). Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society* 80:470–501.

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.

Leonetti, P. (2018). Finite partially exchangeable laws are signed mixtures of product laws. *Sankhyā A* 80(2):195–214.

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

Marques F., P. C. (2024). Universal distribution of the empirical coverage in split conformal prediction. Preprint, arXiv:2303.02770.

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

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

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

Wigner, E. P. (1932). On the quantum correction for thermodynamic equilibrium. *Physical Review* 40:749–759.

Braun, Holzmüller, Jordan and Bach (2025). Conditional coverage diagnostics for conformal prediction. Preprint, arXiv:2512.11779.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*, 2nd ed. Wiley.

Feldman, S., Bates, S. and Romano, Y. (2021). Improving conditional coverage via orthogonal quantile regression. *Advances in Neural Information Processing Systems* 34. arXiv:2106.00394.

Sejdinovic, D., Sriperumbudur, B., Gretton, A. and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics* 41(5):2263–2291.

Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* 35(6):2769–2794.