

Betting Against a Conformal Predictor: a Parimutuel Account of the Information Gap

Peter Cotton

Draft — companion note

Abstract

The companion paper shows that the log-score regret of a single-shape conformal predictor to the conditional oracle is the mutual information $I(R; X)$ between the residual and the input. Here we rederive that quantity from a betting mechanism rather than from the score. Treat the predictor as the crowd in a parimutuel pool on the residual: bettors put money in, and the pot is split among the winners in proportion to their stake on the realised outcome. In the continuous limit the payoff is the ratio of the bettor’s density to the crowd’s. An entrant who knows only the marginal breaks even, which is marginal coverage stated as wealth. An entrant who conditions on X grows his bankroll at rate exactly $I(R; X)$. The gap is the rent. This is not a metaphor: the nearest-the-pin pool of the microprediction platform, and the continuous density version run in the MidOne contest, are this mechanism, and a conformal predictor is the entrant that prices the pool flat in X . We then give two ways to measure the rent on a fitted predictor: a static lower bound from distance covariance, and a sequential e-process whose growth rate estimates $I(R; X)$ and which is an anytime-valid test for conditional miscoverage.

1 Setup

A point predictor leaves a residual R , and conformalization re-levels its marginal law. Write $U = G(R)$ for the probability integral transform, G the marginal CDF of R ; U is marginally uniform on $[0, 1]$ whatever the model. Let $g(u | x)$ be the conditional density of U given $X = x$. The companion paper (Cotton) identifies the wasted log-score of a single-shape conformal system as

$$\Delta = \mathbb{E}_X \text{KL}(P_{R|X} \| G) = I(R; X) = I(U; X) = \mathbb{E}_X \int_0^1 g(u | X) \log g(u | X) du. \quad (1)$$

The middle equalities are the invariance of mutual information under the deterministic, a.s. invertible map $R \mapsto G(R)$, and the fact that the marginal of U is uniform, so the information is the negative conditional differential entropy of U . We take (1) as given and ask what it is, operationally. The answer is a betting rate.

2 The continuous parimutuel

A parimutuel sets no odds. Everyone’s money goes into one pool; once the outcome is known the whole pool is paid to the holders of winning tickets, divided in proportion to their winning stake. With a continuous outcome the winning “ticket” is a density, and the payoff has a clean form.

Lemma 1 (Pool payoff). *Normalise the pool to 1 and let the crowd’s aggregate stake have density q on the outcome space, $\int q = 1$. An infinitesimal, price-taking entrant who stakes unit wealth with density b , $\int b = 1$, and takes no part in moving q , holds after outcome u the wealth*

$$W(u) = \frac{b(u)}{q(u)}. \quad (2)$$

Proof. Bin the outcome axis at width δ . The crowd stakes $q(u)\delta$ of the unit pool on the bin at u ; the entrant stakes $b(u)\delta$. If the outcome lands in that bin, the entire pool is split among its stakeholders in proportion to stake, so the per-unit payoff there is $1/(q(u)\delta)$ (the entrant is infinitesimal, so the crowd’s stake is the whole bin), and the entrant collects $b(u)\delta \cdot 1/(q(u)\delta) = b(u)/q(u)$. The bin width cancels, so the limit is well posed and equals the Radon–Nikodym derivative db/dq . \square

The cancellation is the point. A continuous lottery on a point outcome looks ill defined — every exact value has probability zero — but pool-splitting divides two vanishing quantities, and the ratio survives. No reference measure and no assumed “fair odds” enter.

Lemma 2 (Log-optimal growth). *Let the outcome be drawn from the true law p . The expected log-growth of an entrant staking b against a crowd staking q is*

$$G(b) = \mathbb{E}_{U \sim p} \left[\log \frac{b(U)}{q(U)} \right] = \text{KL}(p||q) - \text{KL}(p||b) \leq \text{KL}(p||q), \quad (3)$$

with equality iff $b = p$ a.e. The log-optimal stake is the truth, and the optimal growth rate is $\text{KL}(p||q)$.

Proof. Add and subtract $\log p$: $\int p \log(b/q) = \int p \log(p/q) - \int p \log(p/b)$. The first term is $\text{KL}(p||q)$, the second is $\text{KL}(p||b) \geq 0$ by Gibbs’ inequality, zero iff $b = p$. \square

Lemmas 1 and 2 are the parimutuel statement of the Kelly–Breiman log-optimality principle (Kelly, 1956; Cover and Thomas, 2006); the growth rate $\text{KL}(p||q)$ is the information the crowd’s price q is missing about the truth p . When the crowd prices the marginal, that is, offers fair odds, the averaged growth of an entrant with belief b is $I(R; X) - \mathbb{E}_X \text{KL}(g(\cdot | X) || b(\cdot | X))$. Kemp and Bettencourt (2022) write exactly this decomposition, $\gamma = I(E; S) - \mathbb{E}_s \text{KL}(P(E | s) || X(E | s))$, in a population-growth setting unconnected to conformal prediction; the next section is the observation that a conformal predictor is precisely such a fair-odds crowd.

3 The conformal lottery

Put the pool on the rank scale. Conformalization makes U marginally uniform, so the conformal predictor is exactly the crowd that prices the rank pool flat and uses nothing about X :

$$q(u) \equiv 1 \quad \text{on } [0, 1]. \quad (4)$$

Flat pricing on the rank scale is fair odds: the implied price equals the true marginal of U , which is what marginal calibration means. That is the hypothesis under which side information is worth its mutual information, so the rent is settled before we compute it. Two entrants step up.

Theorem 1 (Parimutuel rent of conformal). *Against the conformal crowd $q \equiv 1$:*

- (i) *an entrant who knows only the marginal stakes $b \equiv 1$ and grows at rate 0;*

(ii) an entrant who observes X and stakes the conditional $b = g(\cdot | X)$ grows, per round, at rate

$$G^* = \mathbb{E}_X \text{KL}(g(\cdot | X) \| 1) = I(R; X). \quad (5)$$

With a track take $\tau \in [0, 1)$ (payoff $(1 - \tau)/q$) the informed entrant grows at $I(R; X) + \log(1 - \tau)$, positive iff $I(R; X) > -\log(1 - \tau) \approx \tau$.

Proof. Apply Lemma 1 with $q \equiv 1$, then Lemma 2. For (i) the entrant's true law is the marginal, uniform, so $b = q = 1$ and $G = \text{KL}(1 \| 1) = 0$. For (ii) condition on $X = x$: the entrant's correct law is $g(\cdot | x)$, so his conditional growth is $\text{KL}(g(\cdot | x) \| 1)$; averaging over the X he sees each round gives $\mathbb{E}_X \text{KL}(g(\cdot | X) \| 1) = I(U; X) = I(R; X)$ by (1). The take multiplies every payoff by $(1 - \tau)$, adding $\log(1 - \tau)$ to the rate. \square

So the certificate and the leak are two readings of the same pool. Marginal coverage says the lottery is fair to anyone who knows only the average; part (i) is that fairness, as a break-even. Part (ii) is what the certificate cannot see: a side-informed entrant compounds at $I(R; X)$, bankroll growing like $e^{tI(R; X)}$, and no amount of re-leveling closes the gap, because re-leveling acts on the marginal where the leak is invisible. Conformal re-leveling takes no rake ($\tau = 0$), so any conditional information at all is pure profit. This is the lottery paradox in one line: a provably fair lottery that a player with side information beats with certainty.

4 The mechanism is real

The continuous parimutuel is not a thought experiment. It is the reward rule of a deployed prediction market.

The microprediction platform (Cotton, 2022) asks each entrant for 225 Monte Carlo samples of the next value of a live scalar stream. When the value z arrives, a stateless parimutuel splits the pot toward the entries nearest z — nearest-the-pin. Smoothed, an entry's reward is its sample density at z relative to the field's aggregate sample density at z , which is the payoff $b(z)/q(z)$ of Lemma 1 with the densities estimated from samples. The incentive is log-optimal by Lemma 2: an entrant maximises expected log-reward by sampling from his true conditional, $b = p$, so the aggregate crowd price q is driven toward the truth.

The continuous version drops the samples and prices a density directly. In the MidOne contest run by Crunch Labs (CrunchDAO), entrants submit an explicit predictive density — a mixture, in the conventions of the `density` package (microprediction/density), evaluated pointwise — and the wealth change splits the pot in proportion to the density the entrant ascribes to the realised outcome against the density the field ascribes to it, again $b(z)/q(z)$. MidOne scores attacks on small departures from martingality, so the priced object is a residual density: the R of this note.

A conformal predictor, dropped into either contest, is the entrant who prices the residual pool flat in X — correct on the margin, indifferent to the case in front of it. Theorem 1 then says any entrant who conditions on X holds a guaranteed log-growth rate $I(R; X)$ against it. In these markets the information gap is not a figure of speech. It is the realised bankroll of the better-informed competitor.

5 Measuring the rent

The rent is defined through a log-score, so it is not read off data directly. Two routes estimate it on a fitted predictor: a cheap lower bound, and an exact sequential estimate.

5.1 A static lower bound from distance covariance

The companion diagnostic (Cotton) tests $U \perp X$ with distance covariance, which is zero iff the gap is zero. The same statistic bounds the gap from below.

Lemma 3 (Certified lower bound). *Let $\text{HSIC}(U, X)$ use a bounded product kernel, $k \leq K$ on the product space. Then*

$$I(R; X) \geq \frac{\text{HSIC}(U, X)}{2K}. \quad (6)$$

Proof. Write $\text{MMD} = \text{HSIC}^{1/2} = \sup_{\|f\|_{\mathcal{H}} \leq 1} |\mathbb{E}_{P_{U,X}} f - \mathbb{E}_{P_U \otimes P_X} f|$, the integral probability metric over the unit ball of the tensor RKHS. A unit-ball witness satisfies $|f(z)| = |\langle f, k(z, \cdot) \rangle| \leq \sqrt{k(z, z)} \leq \sqrt{K}$. With total variation $\text{TV} = \frac{1}{2} \int |dP_{U,X} - dP_U dP_X|$,

$$|\mathbb{E}_P f - \mathbb{E}_Q f| \leq \|f\|_{\infty} \int |dP - dQ| \leq 2\sqrt{K} \text{TV}, \quad \text{so} \quad \text{MMD} \leq 2\sqrt{K} \text{TV}.$$

Pinsker's inequality (Cover and Thomas, 2006) gives $\text{KL}(P_{U,X} \| P_U \otimes P_X) \geq 2 \text{TV}^2$, and that KL is $I(U; X) = I(R; X)$. Combining, $I(R; X) \geq 2 (\text{MMD} / 2\sqrt{K})^2 = \text{HSIC} / (2K)$. \square

Distance covariance (Székely et al., 2007) is a fixed multiple of this HSIC (Sejdinovic et al., 2013), so the companion detector inherits the same bound and a measured value certifies a minimum wasted log-score, hence by Theorem 1 a minimum extractable rent. The matching upper bound does not exist with a fixed kernel, since MMD metrizes weak convergence (Simon-Gabriel et al., 2023) while KL does not: a sequence of laws can have $\text{HSIC} \rightarrow 0$ with $I(R; X) \rightarrow \infty$. Under a bounded conditional density ratio $g(u | x) \leq M$ one has the crude $I(R; X) \leq \log M$, refinable through χ^2 (Wang and Tay, 2022), but the exact value needs a different object.

5.2 A sequential estimate: the e-process

Run the informed entrant and watch the bankroll. At round t , having seen $(U_1, X_1), \dots, (U_{t-1}, X_{t-1})$, choose a betting density $b_t(\cdot | x) \geq 0$ with $\int_0^1 b_t(u | x) du = 1$ for every x , and set

$$W_t = \prod_{s \leq t} b_s(U_s | X_s), \quad W_0 = 1. \quad (7)$$

Proposition 1 (Anytime-valid test and consistent rent estimate).

(i) *Under $H_0 : U \perp X$ (the conformal price is correct, $I(R; X) = 0$), (W_t) is a non-negative martingale with $\mathbb{E} W_t = 1$, so by Ville's inequality $\mathbb{P}(\sup_t W_t \geq 1/\alpha) \leq \alpha$. Rejecting when $W_t \geq 1/\alpha$ is an anytime-valid, level- α test of conditional miscoverage.*

(ii) *If the entrant learns the conditional, $b_t \rightarrow g(\cdot | x)$, then $\frac{1}{t} \log W_t \rightarrow I(R; X)$ almost surely.*

Proof. (i) Under H_0 , $U_t | X_t$ is uniform and independent of the past, so $\mathbb{E}[b_t(U_t | X_t) | \mathcal{F}_{t-1}] = \int_0^1 b_t(u | X_t) du = 1$; the product of conditionally mean-one factors is a martingale with mean 1, and Ville's inequality applies. (ii) $\frac{1}{t} \log W_t = \frac{1}{t} \sum_s \log b_s(U_s | X_s) \rightarrow \mathbb{E}[\log g(U | X)]$ by the law of large numbers once $b_s \rightarrow g$, and $\mathbb{E}[\log g(U | X)] = \mathbb{E}_X \int g(u | X) \log g(u | X) du = I(R; X)$ by (1). \square

This is the betting story made operational. The wealth process is at once the test — it crosses $1/\alpha$ only if coverage is conditionally uneven — and the estimator, its growth rate converging to the rent the static bound could only lower. It is the log-optimal e-variable of safe testing (Grünwald et al., 2024; Vovk and Wang, 2021) for the null $U \perp X$, and the sequential, anytime sibling of the batch distance-covariance test. Choosing b_t is online conditional density estimation; any consistent learner gives the convergence, and a poor one still keeps the type-I guarantee in (i), since validity needs only $\int b_t = 1$.

A measurement caveat, shared with the companion diagnostic. The transform $U = G(R)$ uses the true marginal G ; with the empirical rank the null is exact only in the leave-one-out or full-conformal sense, which is what we use to keep $U_t | X_t$ uniform under H_0 .

6 Scope

This detects; it does not certify. By the distribution-free no-go results (Lei and Wasserman, 2014; Foygel Barber et al., 2021; Vovk, 2012) no procedure can prove conditional coverage from data alone, so a quiet e-process or a small distance covariance means “no rent found at this power,” never conditional validity. What the betting view adds is a positive reading of the other side: when the gap is there, it is money, and it is extractable by anyone who conditions on X . Closing it is a modelling job — a sharper conditional spread, not a finer threshold — and the rent is the exact bounty for doing it.

7 Related work

The gap $\Delta = I(R; X)$ is the companion paper (Cotton); the distance-covariance diagnostic and the signed de Finetti reading are (Cotton). That optimal log-growth with side information equals the mutual information is classical (Kelly, 1956; Cover and Thomas, 2006; Barron and Cover, 1988), and Kemp and Bettencourt (2022) write the fair-odds form $\gamma = I - \mathbb{E} \text{KL}$ that Lemma 2 specialises, for population growth rather than prediction. Log-loss as negative log-wealth is the prequential and game-theoretic tradition (Dawid, 1984; Shafer and Vovk, 2019), and the growth-optimal e-variable’s identity with a KL is from safe testing and the e-value calculus (Grünwald et al., 2024; Vovk and Wang, 2021); conformal test martingales are Vovk et al. (2003).

Closest are Arnold et al. (2023), who build e-processes for marginal PIT calibration, and Shaer et al. (2023), who test conditional independence by betting — the mechanics of Proposition 1(i) — but neither identifies the growth rate with $I(R; X)$ or frames it as the conformal rent. The parimutuel mechanism and its log-optimal incentive are Cotton (2022, 2021); microprediction/density. The contribution is the assembly: a self-contained parimutuel derivation of $I(R; X)$ in the conformal idiom, the recognition that conformal calibration is the fair-odds condition under which side information is worth its mutual information, the identification of the gap with the rent in a deployed market, and the pairing of a certified lower bound with an anytime-valid growth-rate estimate.

References

Arnold, S., Henzi, A. and Ziegel, J. F. (2023). Sequentially valid tests for forecast calibration. *The Annals of Applied Statistics* 17(3):1909–1935. arXiv:2109.11761.

- Barron, A. R. and Cover, T. M. (1988). A bound on the financial value of information. *IEEE Transactions on Information Theory* 34(5):1097–1100.
- 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.
- Cotton, P. (2021). Inferring relative ability from winning probability in multi-entrant contests. doi:10.1137/19M1276261.
- Cotton, P. (2022). *Microprediction: Building an Open AI Network*. MIT Press.
- Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*, 2nd ed. Wiley. (Chapter 6, gambling and side information; Pinsker’s inequality.)
- Dawid, A. P. (1984). Statistical theory: the prequential approach. *Journal of the Royal Statistical Society A* 147(2):278–292.
- Foygel Barber, R., Candès, E. J., Ramdas, A. and Tibshirani, R. J. (2021). The limits of distribution-free conditional predictive inference. *Information and Inference* 10(2):455–482. arXiv:1903.04684.
- Grünwald, P., de Heide, R. and Koolen, W. M. (2024). Safe testing. *Journal of the Royal Statistical Society B* 86(5):1091–1128. arXiv:1906.07801.
- Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal* 35(4):917–926.
- Kemp, J. T. and Bettencourt, L. M. A. (2022). The Bayesian origins of growth rates in stochastic environments. Preprint, arXiv:2209.09492.
- Lei, J. and Wasserman, L. (2014). Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society B* 76(1):71–96.
- Cotton, P. `density`: conventions for densities used in various games. <https://github.com/microprediction/density>.
- 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.
- Shaer, S., Maman, G. and Aharoni, E. (2023). Model-X sequential testing for conditional independence via testing by betting. *AISTATS*. arXiv:2210.00354.
- Shafer, G. and Vovk, V. (2019). *Game-Theoretic Foundations for Probability and Finance*. Wiley.
- Simon-Gabriel, C.-J., Barp, A., Schölkopf, B. and Mackey, L. (2023). Metrizing weak convergence with maximum mean discrepancies. *Journal of Machine Learning Research* 24. arXiv:2006.09268.
- 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.
- Vovk, V. (2012). Conditional validity of inductive conformal predictors. *Asian Conference on Machine Learning*, PMLR 25:475–490.

- Vovk, V., Nouretdinov, I. and Gammerman, A. (2003). Testing exchangeability on-line. *International Conference on Machine Learning*.
- Vovk, V. and Wang, R. (2021). E-values: calibration, combination, and applications. *The Annals of Statistics* 49(3):1736–1754.
- Wang, C. X. and Tay, W. P. (2022). Practical bounds of Kullback–Leibler divergence using maximum mean discrepancy. Preprint, arXiv:2204.02031.