Scoring Point-Cloud Distributional Submissions

The nearest-the-pin parimutuel, jittered outcomes, and the heat ladder

Peter Cotton*

July 2026

Abstract

Point-cloud forecasts are often evaluated by smoothing the submitted samples into a kernel density estimate and scoring that density at the realised outcome. This apparently natural procedure is not proper: under logarithmic scoring at the raw outcome, a forecaster is generally rewarded for submitting samples from a deconvolution of their belief by the smoothing kernel, rather than from the belief itself. For Gaussian beliefs and Gaussian kernels this incentive has the simple form "shave h^2 from the covariance." We show that the defect is repaired by adding outcome noise from the same kernel used to smooth the submission. More generally, composing a proper score with a fixed Markov kernel preserves propriety, and strict propriety for the original report is recovered exactly when the induced channel is injective; for convolution kernels this is equivalent to the set where the characteristic function is nonzero being dense. For Gaussian kernels, repeating the repaired score over a ladder of smoothing scales decomposes log-score regret, via the relative de Bruijn identity, into Fisher-divergence bands plus a coarse-scale KL term. We also describe a high-dimensional alternative based on random one-dimensional projections, whose average CRPS is, up to an explicit dimension constant, the multivariate energy score. The results are population-level: finite clouds, endogenous bandwidths, and finite-player equilibria are left open.

Keywords: proper scoring rules; continuous ranked probability score; energy score; kernel density estimation; distributional forecasting contests; score matching; sliced scores

JEL classification: C53, C52, C14, D81

1 The problem

A classical parimutuel operates over a finite partition of outcomes. Bettors stake on outcomes; the pot is divided among backers of the realised outcome in proportion to stake; the implied probabilities are the pool fractions, and the operator bears no risk. Modern forecasting is rarely categorical: the object of interest is a full predictive distribution over a continuous, often multivariate quantity, and the natural generalisation replaces "a ticket on outcome j" with "probability mass placed near the point z".

Operationally, a contest accepts from each participant i a cloud $x_1^{(i)}, \ldots, x_m^{(i)} \in \mathbb{R}^d$, smooths it into a density $\widehat{q}_i = \frac{1}{m} \sum_j \varphi_h(\cdot - x_j^{(i)})$, and, when the outcome z is revealed, rewards $\log \widehat{q}_i(z)$ or splits a pot in proportion to $\widehat{q}_i(z)$. This is the reward rule of monteprediction.com; its launch note (Cotton 2024a) describes the pot as

*Microprediction. Email: peter.cotton@microprediction.com.

> "a splitting of the pot in proportion to the density that you ascribe to the truth z, [which] also depends on the density that others ascribe to z,"

and it is also the standard evaluation protocol for sample-based generative models. Not every contest smooths: in the MidOne contests at CrunchDAO (CrunchDAO 2024), participants supplied densities directly, through a shared convention for density specifications (Cotton 2024b), and the incentive analysed in §3 does not arise. This paper concerns cloud submissions.

Two questions organise what follows. What does the pool built on this rule look like, and in what sense is it truthful (§2)? And, more delicately: *what cloud should a rational participant submit?* The answer to the second is *not* "samples from your belief," and the failure has a closed form (§3), a one-line repair (§4), and, once repaired, a multi-scale structure (§5).

Two payout layers appear, and they are different mechanisms. The proportional pot split (NTP) of §2 is the parimutuel object: self-funding, limited by the pot, and truthful only in the Kelly, price-taking, symmetric-equilibrium sense of §2. The strict-propriety theorems of §4-§6 apply directly to additive stake-weighted score transfers, which preserve properness cleanly but are not automatically limited-liability for unbounded scores; applying them to a proportional pot split requires the separate §2 analysis.

Throughout, we idealise the cloud by its sampling law ρ ($m \to \infty$; finite m is Open Problem 1), so the mechanism observes $T_h\rho := \rho * \varphi_h$, where φ_h is a symmetric density on $\mathbb{R}^d$, taken strictly positive (Gaussian, canonically) so that log scores are finite; the strict-propriety results themselves need only the channel to be fixed and injective, so compactly supported kernels qualify under Lemma 1's characteristic-function condition with extended-score conventions. The truth p^* is absolutely continuous with finite differential entropy, and all expectations below are assumed finite (for Gaussian kernels and beliefs with finite second moments, $\log T_h\rho$ has at-worst-quadratic tails, so this is mild). Point-mass examples are understood in the obvious measure-theoretic extension.

2 The pool: a nearest-the-pin parimutuel

Why a *density* split, and not some other functional? Because the parimutuel already has a sharp incentive theory in the discrete case, and it is exactly the one we want. Consider n outcomes with true probabilities p_k, and a parimutuel in which a player allocates a unit stake as a distribution $b = (b_1, \ldots, b_n)$ over outcomes. If the rest of the pool's stake fractions are r_k and the player is small, a unit bet on k returns $1/r_k$ when k occurs. A player maximising the expected log growth of wealth solves

$$\max_{b \in \Delta} \sum_k p_k \log \frac{b_k}{r_k},$$

whose unique maximiser is $b_k = p_k$, *bet your beliefs* (Kelly 1956; Breiman 1961). The growth rate at the optimum is the Kullback–Leibler divergence $D(p \,\|\, r)$: the player profits exactly to the extent their belief beats the crowd's implied distribution. Log-wealth maximisation and truthful reporting coincide, and the quantity being maximised is the logarithmic score of the report against the realised outcome, net of the crowd.

Mechanism. Each of n participants holds wealth W_i and submits a predictive density q_i over $\mathbb{R}^d$. Each risks a stake $s_i = b\, W_i$ for a fixed fraction $b \in (0, 1)$. The outcome z is revealed.

The collected pot $S = \sum_i s_i$ is redistributed in proportion to $s_i\, q_i(z)$, stake times the density placed at z, so participant i's wealth change is

$$\boxed{\Delta W_i \;=\; S\,\frac{s_i\, q_i(z)}{\sum_j s_j\, q_j(z)} \;-\; s_i} \tag{NTP}$$

Self-funding. $\sum_i \Delta W_i = S - S = 0$: the pool is a pure transfer, the operator bears no risk, exactly as in the racetrack tote. (Implemented and tested in `pot_split`.)

Incentives of the pot split. Work in the price-taking model: fix the field, write r for the crowd's aggregate density, and treat the odds $1/r$ as unmoved by the participant's own stake (in (NTP) the player's $s_i q_i(z)$ also sits in the denominator, so a non-infinitesimal player moves their own odds; that equilibrium is Open Problem 3). A unit staked as q returns $q(z)/r(z)$ at the pin. One accounting identity organises everything: *in a parimutuel, cash is a bet on the crowd.* Staking the crowd density r returns exactly one whatever z realises, so a participant holding fraction $1-b$ in cash and b on submission q holds total exposure $\tilde{q} = (1-b)\,r + b\,q$, and the classical argument applies to $\tilde{q}$: the log-optimal investor's total exposure is p, whatever the odds. Equivalently, optimise over the exposure $\tilde{q} \geq (1-b)r$ directly and recover the submission as $q = (\tilde{q} - (1-b)r)/b$. Inverting the accounting, the optimal submission is

$$q_b^*(z) \;=\; \left[\frac{p(z)}{\lambda} - \frac{1-b}{b}\, r(z)\right]_+,$$

the KKT solution of maximising $\int p \log((1-b) + b\, q/r)$ over densities, with $\lambda > 0$ chosen so that $\int q_b^* = 1$ (assume throughout that $r > 0$ wherever $p > 0$ and that the log objective is integrable). When the bracket is nonnegative everywhere the multiplier is $\lambda = b$ and the solution is the affine formula

$$q^* \;=\; r + \frac{p-r}{b} \;:$$

the crowd plus the participant's disagreement, levered by $1/b$. Three regimes follow.

- At $b = 1$ the submission is p itself: all-in, bet your beliefs, whatever the crowd does (still within the price-taking model).
- For $0 < b < 1$, truth-telling $q = p$ is optimal iff $r = p$: truthful reporting is the symmetric-equilibrium report, not a dominant strategy. The numerical check in `test_nearest_the_pin.py`, a truthful reporter out-growing a biased one against a truthful field, is an equilibrium statement.
- As $b \to 0$ the lever blows up, the truncation binds, and the best response degenerates toward a spike at the argmax of p/r: a small-stake participant is paid to hunt the crowd's most underpriced point, not to report a density. (The objective becomes linear in q; there is no Gibbs argument in this limit.)

Proper-score payments do transfer cleanly to the *additive* stake-weighted transfer $\Delta W_i = s_i\,(S_i - \bar{S})$ of §5, which is the incentive-compatible form for fractional stakes; a pot split *proportional* to a score value is not an affine function of the score, and properness alone does not make it truthful.

Relationship to other mechanisms.

- It is the continuous, density-weighted limit of the discrete parimutuel above, and a density-pot-split generalisation of Pennock's dynamic parimutuel market (DPM) (Pennock 2004): the DPM prices shares on discrete outcomes via a cost function; NTP prices density mass on a continuum.
- Its truthfulness rests on the logarithmic score / log-wealth growth, the same object that, applied *sequentially*, gives Hanson's LMSR (Hanson 2007). NTP is the *pooled* reading; LMSR is the *sequential* reading.
- For directly reported densities the implicit score is the log score, strictly proper for the report. For clouds the KDE map changes the report space: §3–§4 analyse when raw settlement is improper for the cloud and how matched jitter repairs it, and §6 connects the projection version to the energy score.

These incentive statements concern the reported *density* q. In practice q is constructed from a sample cloud by KDE, so the report space is the cloud and the mechanism applies $q = \rho * \varphi_h$. The truthful object and the submitted object come apart, and the next two sections are about the gap.

3 The deconvolution incentive

Theorem 1 (improperness of raw-outcome KDE scoring). *Let $S_{\text{raw}}(\rho, z) = \log(T_h\rho)(z)$. Then*

$$\mathbb{E}_{z\sim p^*} S_{\text{raw}}(\rho, z) \;=\; -H(p^*) \;-\; \mathrm{KL}\big(p^* \,\|\, T_h\rho\big),$$

so the report solves $\min_{q\in T_h\mathcal{P}} \mathrm{KL}(p^*\|q)$, where $\mathcal{P}$ is the set of probability laws. Consequently:

(i) If $p^* \in T_h\mathcal{P}$, every ρ with $T_h\rho = p^*$ is optimal. If in addition T_h is injective — for the Gaussian kernel always, and in general whenever the characteristic function of φ_h is nonvanishing on a dense set (Lemma 1) — the optimal pre-smoothing law $\rho^\dagger := T_h^{-1}p^*$ is unique. For any non-degenerate kernel it differs from the truth, and truthful reporting is strictly suboptimal with truthfulness gap

$$\Delta \;=\; \mathrm{KL}\big(p^* \| p^* * \varphi_h\big) \;>\; 0.$$

(ii) If $p^* \notin T_h\mathcal{P}$, any attained optimum is a KL projection of p^* onto the convex class $T_h\mathcal{P}$ (the infimum need not be attained), and Δ above is only the loss of the truthful report relative to the unattainable ideal $q = p^*$; the gap to the attainable optimum is $\Delta - \inf_{\rho\in\mathcal{P}} \mathrm{KL}(p^*\|T_h\rho)$, which can vanish in degenerate cases (for a point-mass belief and a centered kernel whose density is maximized at the origin, the truthful cloud is optimal).

(iii) Gaussian closed form: $p^* = N(\mu, \Sigma)$, $\varphi_h = N(0, h^2 I)$. The deconvolution exists iff $\Sigma - h^2 I \succeq 0$, and then $\rho^\dagger = N(\mu, \Sigma - h^2 I)$. In one dimension with $\tau^2 > h^2$: shave h^2 off the variance, with gap $\Delta = \frac{1}{2}[\log(1 + h^2/\tau^2) - h^2/(\tau^2 + h^2)] \approx h^4/(4\tau^4)$.

Proof. The identity is the standard cross-entropy decomposition, $\int p^* \log q = -H(p^*) - \mathrm{KL}(p^*\|q)$, applied to $q = T_h\rho$; maximizing over ρ is minimizing KL over the convex image class $T_h\mathcal{P}$. (i) If $p^* \in T_h\mathcal{P}$ the KL term can be driven to zero, its unique minimum, and only by $q = p^*$; under the injectivity hypothesis, Lemma 1 lifts uniqueness of q to uniqueness of ρ. The truthful report attains $\mathrm{KL}(p^*\|p_h^*)$ where $p_h^* := p^* * \varphi_h$, and $p_h^* \neq p^*$ for any non-degenerate kernel: if $p^* = p^* * \varphi_h$ then $\widehat{p}^*(\omega)\,(1 - \widehat{\varphi}_h(\omega)) = 0$ for all ω; since $\widehat{p}^*$ is continuous with $\widehat{p}^*(0) = 1$ it is nonzero on a neighbourhood of the origin, so $\widehat{\varphi}_h = 1$ there, and a characteristic function equal to one on a neighbourhood of the origin is identically one (the standard inequality $1 - \mathrm{Re}\,\widehat{\varphi}(2\omega) \leq$

$4\,(1-\operatorname{Re}\widehat{\varphi}(\omega))$ propagates the equality outward), so φ_h is the point mass at zero. (ii) KL is jointly lower semicontinuous and strictly convex in its second argument where finite, and $T_h\mathcal{P}$ is convex, so the attainable optimum is the KL projection when attained; the point-mass example ($p^* = \delta_\mu$ with a kernel maximized at the origin: the truthful cloud maximises $(T_h\rho)(\mu)$) shows the gap to the attainable optimum can be zero. (iii) Gaussians: convolution adds covariances; the KL between $N(0,\tau^2)$ and $N(0,\tau^2+h^2)$ is the stated expression; Taylor expansion gives $h^4/(4\tau^4)$. ■

Remarks. (a) When the deconvolution fails to exist ($\Sigma - h^2 I \not\succeq 0$, or a rough p^*), the optimum is the KL projection of p^* onto $T_h\mathcal{P}$ and can degenerate toward atomic clouds — exactly the point-mass exploit of Theis et al. (2016), who observed the improperness (without the deconvolution characterization). (b) The gap is fourth order in h, which is why the flaw survives casual inspection — but it is a *slope*, and any optimising submitter, human or fitted, walks down it. (c) The phenomenon is the log-score/bandwidth sibling of the "fair scores" findings for the ensemble CRPS (Fricker et al. 2013; Ferro 2014), where the analogous cheat is derived along the ensemble-size axis.

4 The repair: jitter the pin

Definition (mollified log score). With $\varepsilon \sim N(0,I)$,

$$S_h(\rho,z) \;=\; \mathbb{E}_\varepsilon\big[\log(T_h\rho)(z+h\varepsilon)\big] \;=\; (\varphi_h * \log T_h\rho)(z).$$

Operationally: *settle at a jittered outcome* $z' = z + h\varepsilon$ — or use the right-hand form, which integrates the jitter analytically and makes settlement deterministic. The two have the same expectation, hence the same risk-neutral proper-score incentives; for risk-sensitive wealth dynamics (Kelly staking, the pot split of §2) the realized-jitter and integrated-jitter implementations differ in payoff variance.

Lemma 1 (injectivity criterion). *Convolution T_φ is injective on probability measures iff the zero set of the characteristic function has empty interior,*

$$\operatorname{int}\{\omega : \widehat{\varphi}(\omega) = 0\} \;=\; \emptyset$$

(equivalently, $\{\widehat{\varphi} \neq 0\}$ is dense in $\mathbb{R}^d$).

Proof. ($\Leftarrow$) $T_\varphi\rho = T_\varphi\rho'$ gives $(\widehat{\rho}-\widehat{\rho}')\widehat{\varphi} \equiv 0$, so $\widehat{\rho} = \widehat{\rho}'$ on a dense set, hence everywhere by continuity of characteristic functions, hence $\rho = \rho'$ by the uniqueness theorem. ($\Rightarrow$) If $\widehat{\varphi}$ vanishes on an open ball B (with $0 \notin B$, since $\widehat{\varphi}(0) = 1$), take $\psi \in C_c^\infty(B)$, $\psi \neq 0$, and set $\widehat{g}(\omega) = \psi(\omega) + \overline{\psi(-\omega)}$, so that g is a real Schwartz function (a signed integrable function with zero integral: B avoids the origin, so $\int g = \widehat{g}(0) = 0$) with $\widehat{g}$ supported where $\widehat{\varphi} = 0$. Choose a base density p with tails heavier than Schwartz decay (Cauchy), so $|\epsilon g| \le p$ pointwise for small $\epsilon > 0$. Then $q = p + \epsilon g$ is a probability density distinct from p, and the characteristic function of $T_\varphi q$ is $\widehat{\varphi}\,(\widehat{p}+\epsilon\widehat{g}) = \widehat{\varphi}\,\widehat{p}$: two distinct laws with identical smoothings. ■

This is Wiener-flavoured (Wiener's Tauberian theorem is the L^1 statement); for probability measures the continuity of characteristic functions buys the dense-support version. Practical reading: Gaussian and Laplace jitter are injective (nonvanishing characteristic functions); the uniform kernel is also fine (sinc zeros are isolated, hence the nonvanishing set is dense); band-limited kernels fail (e.g. Fejér-type kernels, whose characteristic function vanishes outside a bounded interval — two beliefs agreeing on low frequencies become indistinguishable). For compactly supported kernels Lemma 1 supplies the injectivity, but the log-score statement then requires the extended-real convention and a report class on which the expected score is well defined; the Gaussian case avoids the nuisance because every smoothed report is strictly positive.

Theorem 2 (kernel-channel properness). *Let S be a scoring rule, K a Markov kernel with push-forward T_K on laws, and define*

$$S_K(\rho, z) \;=\; \mathbb{E}_{z' \sim K(z,\cdot)}\big[S(T_K\rho,\, z')\big].$$

*Assume (a) $\mathbb{E}_{w \sim T_K p}\,|S(T_K\rho, w)| < \infty$ for the laws p, ρ under comparison (or adopt an extended-expectation convention), and (b) $T_K\mathcal{P}$ lies within the report class on which S is defined. If S is proper, S_K is proper. If S is strictly proper on $T_K\mathcal{P}$, then S_K is strictly proper on $\mathcal{P}$ **iff** T_K is injective on $\mathcal{P}$. In particular the mollified log score S_h is strictly proper for the pre-smoothing law whenever the zero set of $\widehat{\varphi}$ has empty interior — for Gaussian jitter, always.*

Proof. By Fubini, $\mathbb{E}_{z\sim p^*} S_K(\rho, z) = \mathbb{E}_{w\sim T_K p^*} S(T_K\rho, w)$: the expected score of the report $T_K\rho$ for the outcome law $T_K p^*$, under S. Properness of S maximizes this at $T_K\rho = T_K p^*$, which the truthful $\rho = p^*$ achieves — properness. Strict properness of S on the image class forces $T_K\rho = T_K p^*$ at any maximizer; injectivity of T_K lifts this to $\rho = p^*$, and conversely if T_K is not injective two distinct reports tie. For $S = \log$ and K = convolution by φ_h: the log score is strictly proper on densities, $T_K\rho = \rho * \varphi_h$ is a density, and Lemma 1 gives injectivity. ■

How much to jitter? Exactly as much as you smooth. The jitter is not a free parameter; it is pinned, kernel for kernel, to the mechanism's own smoothing. Jittering with s.d. j while smoothing with bandwidth h pairs the jittered truth $p^* * \varphi_j$ with the smoothed report $\rho * \varphi_h$, and in the all-Gaussian setting the optimal report variance is

$$v^* \;=\; \tau^2 + j^2 - h^2.$$

At $j = 0$ this is Theorem 1's shave. At $j = h$, and only at $j = h$, the optimum is the truth. At $j > h$ the mechanism pays padding by $j^2 - h^2$. The formula holds where admissible ($v^* \geq 0$); otherwise the Gaussian-family optimum sits on the boundary (the KL-projection case of Theorem 1(ii)). Under-jitter rewards sharpening, over-jitter rewards blurring, and matching the bandwidth is the unique fixed point (the general statement is Theorem 2 with the *same* kernel on both sides; a mismatched pair elicits $\arg\min_\rho \mathrm{KL}(p^* * \varphi_j \,\|\, \rho * \varphi_h)$, the j-blurred belief deconvolved by h).

The bandwidth must be exogenous. Theorem 2 assumes a *fixed* channel: the smoothing/jitter kernel must not depend on the submission being scored. A data-driven bandwidth computed *from the participant's own cloud* (Scott's rule on the submission, as reference KDE implementations default to) puts $h(\rho)$ under the participant's control, the score becomes $\mathbb{E}_\varepsilon \log(T_{h(\rho)}\rho)(z + h(\rho)\varepsilon)$, and strict properness no longer follows from Theorem 2. In a contest, freeze the bandwidth before submissions are observed (from the outcome history, a reference climatology, or a posted rule) and jitter with that frozen kernel. The incentives of participant-endogenous bandwidths are an open problem (§10).

Attribution. The properness of the convolved score against a noised outcome is not new: Bröcker and Smith (2007, sec. 5) prove it for a general observation-noise channel, and Ferro (2017, Prop. 3) works out exactly the white-noise/Gaussian case. What both left open is strictness. Bröcker & Smith, verbatim: "If S is strictly proper though, $\bar{S}$ is not necessarily strictly proper, because if $\bar{q}(z) = \bar{p}(z)$, this does not necessarily mean equality of $p(x)$ and $q(x)$."

Lemma 1 closes exactly that gap. The discrete-outcome analog is the label-noise literature's forward loss correction with an invertible transition matrix (Patrini et al. 2017, Thm. 2; Rooyen and Williamson 2018). Verification treats outcome noise as a nuisance to be endured, and Ferro explicitly doubts the "efficacy" of perturbing observations; read as mechanism design, jitter matched to the mechanism's own smoothing is what makes the point-cloud game strictly proper.

5 The heat ladder

Let $p_t := p * N(0, tI)$, the heat flow ($\partial_t p_t = \frac{1}{2}\Delta p_t$). To keep the bookkeeping explicit, define the totally smoothed laws

$$p_t^{(h)} := p^* * N(0, (h^2 + t)I), \qquad \rho_t^{(h)} := \rho * N(0, (h^2 + t)I),$$

so the rung at flow time t is the §4 score with kernel $\varphi_{\sqrt{h^2+t}}$ — for clouds this is free to compute, since it is *the same cloud* scored with bandwidth $\sqrt{h^2+t}$ against a pin jittered by the same kernel. Every rung the mechanism runs has total scale $h^2 + t \geq h^2 > 0$; below, p_t^* and ρ_t always abbreviate $p_t^{(h)}$ and $\rho_t^{(h)}$, the flow started from the already-smoothed laws.

Theorem 3 (scale decomposition of the log-score edge). *Let p^*, ρ have finite second moments, and suppose the standard relative de Bruijn regularity conditions hold on $[s, T]$: $\mathrm{KL}(p_t^* \| \rho_t)$ finite along the flow, the relative Fisher divergence integrable on the interval, and enough decay to justify the integrations by parts below. Then for $0 \leq s < T$,*

$$\frac{d}{dt}\mathrm{KL}(p_t^* \| \rho_t) \;=\; -\tfrac{1}{2}\, D_F(p_t^* \| \rho_t), \qquad \mathrm{KL}(p_s^* \| \rho_s) = \mathrm{KL}(p_T^* \| \rho_T) + \tfrac{1}{2}\int_s^T D_F(p_t^* \| \rho_t)\, dt,$$

where $D_F(p\|q) = \int p\, \|\nabla \log p - \nabla \log q\|^2$ is the Fisher divergence.

Proof. Write $p = p_t^*$, $q = \rho_t$, both solutions of the heat equation. Then

$$\frac{d}{dt}\int p \log\frac{p}{q} \;=\; \int (\partial_t p)\log\frac{p}{q} \;+\; \underbrace{\int \partial_t p}_{=0} \;-\; \int \frac{p}{q}\,\partial_t q.$$

First term: $\frac{1}{2}\int \Delta p\, \log(p/q) = -\frac{1}{2}\int \nabla p \cdot \nabla \log(p/q) = -\frac{1}{2}\int p\, \nabla \log p \cdot (\nabla \log p - \nabla \log q)$. Third term: $-\frac{1}{2}\int \frac{p}{q}\,\Delta q = \frac{1}{2}\int \nabla\!\left(\frac{p}{q}\right) \cdot \nabla q = \frac{1}{2}\int p\,(\nabla \log p - \nabla \log q) \cdot \nabla \log q$. Summing,

$$\frac{d}{dt}\mathrm{KL} = -\tfrac{1}{2}\int p\, \|\nabla \log p - \nabla \log q\|^2 = -\tfrac{1}{2} D_F(p\|q),$$

and integrating over $[s, T]$ gives the display. After smoothing by $u \geq h^2 > 0$ both densities are positive and smooth, which is necessary but not by itself sufficient: Gaussian convolution does *not* replace heavy tails by Gaussian tails, so the vanishing of boundary terms is an assumption (the stated regularity), automatic for the Gaussian and compactly supported cases used in the examples. ■

The differential identity is de Bruijn's, in relative form (Stam 1959; Barron 1986; Lyu 2009); its integral form prices the likelihood of diffusion models (Song et al. 2021). Note the bandwidth floor does real work: at $t = 0$ an empirical cloud has no density and the identity is vacuous; every rung the mechanism actually runs starts at $t \geq h^2$, where everything is smooth.

The heat-ladder pool. Fix scales $0 = t_0 < t_1 < \cdots < t_K = T$ and weights $w_k \geq 0$. Participant i stakes s_i and submits one cloud. Two payment schedules must be distinguished, because they buy different things.

Level schedule. Rung k settles by the stake-weighted additive transfer driven by the rung score $S_i^{(k)} := S_{\sqrt{h^2+t_k}}(\rho_i, z)$,

$$\Delta W_i^{(k)} = w_k\, s_i\,(S_i^{(k)} - \bar{S}^{(k)}),$$

with $\bar{S}^{(k)}$ the stake-weighted mean (Lambert et al. 2008, 2015). The transfer is budget-balanced ($\sum_i \Delta W_i^{(k)} = 0$) but not limited-liability: with unbounded scores such as the log score a transfer can exceed the stake, so a bounded pot split requires bounded scores, clipping, collateral, or a Lambert-style bounded transformation. Each rung is strictly proper (Theorem 2), so any nonnegative-weighted sum is. But the expected-regret decomposition carries *cumulative* weights on the Fisher bands, not the rung weights themselves: writing $\mathrm{KL}_k := \mathrm{KL}(p_{t_k}^{(h)} \| \rho_{t_k}^{(h)})$,

$$\sum_{k=0}^{K} w_k\,\mathrm{KL}_k \;=\; \Big(\sum_{k=0}^{K} w_k\Big)\mathrm{KL}_K \;+\; \tfrac{1}{2}\sum_{\ell=0}^{K-1}\Big(\sum_{k=0}^{\ell} w_k\Big)\int_{t_\ell}^{t_{\ell+1}} D_F(p_t^*\|\rho_t)\,dt.$$

Difference schedule. To pay a Fisher band directly, use the score *differences* as the payment driver.

Proposition (band scores are strictly proper). *For a positive-width band $t_{k+1} > t_k$, Gaussian smoothing, and the regularity of Theorem 3, the difference score $S^{(k)} - S^{(k+1)}$ has expected regret $\mathrm{KL}_k - \mathrm{KL}_{k+1} = \frac{1}{2}\int_{t_k}^{t_{k+1}} D_F(p_t^*\|\rho_t)\,dt \geq 0$, with equality iff $\rho = p^*$. It is therefore itself a strictly proper score, even though a difference of proper scores is not proper in general.*

Proof. The expected value of $S^{(k)}$ under p^* is $-H(p_{t_k}^*) - \mathrm{KL}_k$; the entropy terms are report-independent, so the regret of the difference is $\mathrm{KL}_k - \mathrm{KL}_{k+1}$, which is the band integral by Theorem 3. It vanishes iff $D_F(p_t^*\|\rho_t) = 0$ on the band, i.e. $\nabla \log \rho_t = \nabla \log p_t^*$ p_t^*-a.e.; after Gaussian smoothing both densities are everywhere positive, so the log-ratio is constant, and both integrating to one forces $\rho_t = p_t^*$, hence $\rho = p^*$ by Lemma 1. ■

Corollary. *(i) Each rung (or band), hence the tower, is budget-balanced in the additive stake-weighted form. (ii) Each level score is strictly proper for the cloud law (Theorem 2) and each band score is strictly proper (the Proposition), so truthful submission is optimal for a risk-neutral participant at any stake short of the whole pool: the own score enters the stake-weighted transfer with coefficient $s_i(1-s_i/S) > 0$. The small-stake caveat belongs to the multiplicative pot split of §2, which requires the log-wealth (Kelly) model stated separately. (iii) Band payments purchase (integrated) Fisher divergence, the Hyvärinen-scored shape of the smoothed submission; the regret is insensitive to multiplicative constants in the smoothed density, though the implemented band score remains a difference of normalized log scores unless the report class is widened to unnormalized densities. The top level pays $\mathrm{KL}(p_T^*\|\rho_T)$, which at mode-connecting scales carries the between-mode mass that score matching is blind to (Wenliang and Kanagawa 2020; Zhang et al. 2022; Koehler et al. 2023). Under the level schedule the band weights are the cumulative sums above.*

Practicalities: rung scores are positively correlated (one cloud, re-smoothed), so discriminative value concentrates in a few well-separated scales; K of order 3–5, geometrically spaced, mirrors the noise ladders of annealed score matching (Song and Ermon 2019).

6 The projection route

The high-dimensional obstruction. In $d = 11$ dimensions, recovering a participant's density $q_i(z)$ from a finite sample cloud by KDE is fragile: the bandwidth, the curse of dimensionality, and the heavy tails of financial returns all bite. monteprediction's contest has participants submit *a million* scenarios precisely because dense coverage is needed to pin down $q_i(z)$. And the incentive analysis of §3 gets *worse* with dimension: Scott's-rule bandwidths grow toward the signal scale as d rises, so the optimal shave h^2 becomes first-order rather than a subtle correction.

Slicing. The projection (sliced) version sidesteps the d-dimensional density. Draw random unit directions $u \in S^{d-1}$, project every participant's cloud and the outcome onto each u, and score the resulting *one-dimensional* forecasts, where density estimation, the CRPS, and quantiles are all easy and robust; then average over directions. This is exactly aligned with how the energy score decomposes. For a uniformly random u on the sphere and any $x \in \mathbb{R}^d$,

$$\mathbb{E}_u\, |\langle u, x\rangle| \;=\; c_d\, \|x\|, \qquad c_d = \frac{\Gamma(d/2)}{\sqrt{\pi}\,\Gamma((d+1)/2)},$$

so $\|x\| = c_d^{-1}\, \mathbb{E}_u|\langle u, x\rangle|$. Substituting into the energy score $\mathrm{ES}(P, y) = \mathbb{E}\|X - y\| - \frac{1}{2}\mathbb{E}\|X - X'\|$ gives the projection identity

$$\boxed{\mathrm{ES}(P, y) \;=\; c_d^{-1}\, \mathbb{E}_u\big[\,\mathrm{CRPS}(P_u,\ \langle u, y\rangle)\,\big]} \tag{PROJ}$$

where P_u is the law of the projected sample $\langle u, X\rangle$. The multivariate energy score *is* the average over random directions of the one-dimensional CRPS, and the energy score is strictly proper for the full distribution (Gneiting and Raftery 2007). The sliced quantity is therefore a proper score that needs only 1-D evaluations. We verify (PROJ) numerically: in `energy_score_via_projection` the sliced estimate matches the exact multivariate energy score within a few percent at a few thousand directions, and equals the CRPS exactly in 1-D (`test_nearest_the_pin.py`).

The projection-scored pool. For each direction u, compute the 1-D CRPS of participant i's projected cloud at $\langle u, z\rangle$ and average over u: the sliced energy score ES_i. In the population idealization, replacing the projected cloud by its sampling law, this is the energy score and hence strictly proper for laws with a finite first moment; for a finite cloud the ordinary empirical CRPS is the plug-in score for the empirical distribution and is not, without a fair-score correction, the finite-m elicitation rule for iid sampling from the belief. This wagering claim, like the others here, is population-level. The score is negatively oriented (smaller is better), so the incentive-compatible wager is the additive stake-weighted transfer of §5 driven by its negation, $\Delta W_i = s_i\big((-\mathrm{ES}_i) - \overline{(-\mathrm{ES})}\big)$: self-funding, and properness transfers cleanly because the payoff is affine in a proper score. A pot split *proportional* to a transformed score value is a different mechanism and is not covered by the properness theorem (§2). Three further cautions. Uniform directions give the exact identity (PROJ); a non-uniform full-support direction law still yields a strictly proper sliced score in expectation, but it is an anisotropic projection score, no longer the Euclidean energy score. The directions must be drawn after submissions are in, from a full-support law: a fixed, pre-announced finite set elicits only those one-dimensional marginals, not the joint law, and with finitely many realized directions strict propriety is ex ante over the mechanism's randomization, while conditional on the realized set only the projected marginals are scored. And because no kernel smoothing is applied, the deconvolution incentive of §3 does not arise; the finite-m analog is the fairness correction of Ferro (2014), applied slice-wise in one dimension. This is, in spirit, the projection version at montepredicton: score the eleven-dimensional cloud through its one-dimensional shadows.

Link to the random-projections literature. Slicing a high-dimensional problem into random 1-D projections is a recurring, theoretically-backed device:

- Johnson–Lindenstrauss (Johnson and Lindenstrauss 1984): for a *fixed finite* cloud, random projections approximately preserve the pairwise Euclidean distances entering the empirical energy score. Population-level and participant-ranking guarantees need separate concentration arguments (Open Problem 7).
- Sliced Wasserstein distances (Rabin et al. 2012; Bonneel et al. 2015): average 1-D optimal-transport costs over random projections, a now-standard, cheap surrogate for the multivariate Wasserstein distance — the optimal-transport cousin of (PROJ).
- Sliced score matching (Song et al. 2020): estimate high-dimensional score functions through random projections, for the same computational reasons.
- Energy distance (Székely and Rizzo 2013) is itself an integral of squared characteristic-function differences and, via the identity above, of absolute projected differences; the projection representation is intrinsic, not a heuristic.

The projection version is not an approximation bolted onto the mechanism; it is the *native* high-dimensional form of the additive proper-score wagering route, with the energy score represented, by construction, as an average of one-dimensional CRPS scores over random projections.

7 Two routes to high-dimensional scoring

Step back from the pool to the statistical problem underneath: how do you score a joint distributional forecast in $\mathbb{R}^d$ when d is large relative to the data you can condition on? This is exactly the regime ($d > n$) in which the naïve held-out Gaussian likelihood, the density-based score, becomes unreliable, because the estimated covariance is rank-deficient and its inverse is nonsense.

The portfolio/spatial-statistics literature answers this from the density side. *Two Sides of Schur Damping* (Cotton 2026) and the underlying Schur complementary allocation (Cotton 2024c) observe that a Gaussian density factorises through a Vecchia / conditional pseudo-likelihood (Vecchia 1988) $\prod_k \mathcal{N}(y_k; b_k^\top y_c, S_k)$ whose conditional covariances S_k are *Schur complements*, and that the reliable score in the undersampled regime is a damped version of this factorisation, the Schur pseudo-likelihood, with a closed-form James–Stein reliability damping $\gamma^\star$. When the raw joint density is untrustworthy, score it through a structured, positive-definiteness-preserving factorisation rather than the full inverse covariance. (This is the basis of the `precise` library's covariance assessors; see the schur project.)

The nearest-the-pin parimutuel needs precisely such a score: it must turn each participant's joint forecast into a number at z. There are then two routes, and they are the two sides of the same coin:

route	how the joint forecast is scored	regime it suits
density	structured / damped joint density — Schur pseudo-likelihood, Vecchia factorisation	a parametric or covariance-shaped forecast; $p > n$ handled by damping
projection	average 1-D CRPS over random directions — the sliced energy score (PROJ)	a free-form sample cloud; high d handled by slicing

The organising claim of the high-dimensional story: the density route (Schur / Vecchia) and the projection route (energy / sliced) are alternative ways to make a joint forecast scoreable in high dimensions, and the pool can be run with either. A speculative synthesis, flagged as conjecture, is a *Schur-damped* projection score, in which the directions u are not isotropic but shaped by a damped estimate of a reference covariance (project more often along the well-estimated directions), interpolating between the two columns with a single reliability dial γ exactly as in the Schur work. Its analysis is open, and a properness boundary is worked out in the anisotropic sliced scores note: strict properness survives only if the anisotropy comes from a reference covariance fixed in advance, not from the forecast under test. The same caution applies to the density column: a pseudo-likelihood score is proper only if the map from submitted forecast to scored factorisation — ordering, conditioning sets, damping, reference covariance — is fixed ex ante and sufficiently identifying. Letting any of them depend on the forecast under test reopens the endogenous-channel failure of §4, so the density route is a design direction here, not an established properness theorem.

8 Microprediction, and a historical note

The nearest-the-pin parimutuel is one of the mechanisms of the microprediction vision (Cotton 2022): a web-scale network of autonomous forecasters continuously submitting *distributional* predictions and being paid by self-funding, truth-eliciting pools. Two further pieces close the loop:

- Calibration via Z-streams. The crowd's aggregate density Q induces, for a scalar quantity, z-scores $\Phi^{-1}(F(x))$; if the market is calibrated these are standard normal over time, and any departure (fat tails, skew, autocorrelation) is an exploitable, self-correcting anomaly. The pool's payouts push the aggregate back toward calibration.
- Aggregation. The crowd density Q is itself a forecast, a wealth-weighted pool of the participants' densities: a (log-)opinion pool whose weights are endogenously set by past accuracy.

The microprediction platform, launched in 2019, collected a billion predictions over the course of its operation, 225 samples at a time, and added a small amount of noise to submissions and to the ground truth before settling its cloud-based lotteries. This was *merely intuitive*, a fairness-and-anti-gaming instinct about discreteness and ties, with no incentive theorem attached; the platform paper recorded the practice in one line and moved on. The verification literature, meanwhile, treated outcome noise as a defect: something to be modelled away (Saetra et al. 2004; Candille and Talagrand 2008), with Ferro (2017) doubting the value of perturbing observations.

Theorem 2 is the missing theorem: symmetrized jitter is what makes a smoothed-submission game strictly proper for the submitted law. The theory also sharpens the intuition into design guidance it could not supply on its own: the jitter distribution matters. Kernels whose characteristic function has dense support (Gaussian, Laplace, even uniform) preserve strict properness; band-limited kernels do not (Lemma 1). Intuition chose jitter; the theorem chooses *which* jitter.

9 Related work

Improperness of sample-based scoring: Theis et al. (2016), who call the KDE log-likelihood "an improper scoring function"; the fair-scores line for the ensemble CRPS (Bröcker 2012; Fricker et al. 2013; Ferro 2014), with the over-dispersion direction for fitted ensembles in Siegert et al. (2019); the estimator view of KDE log scores (Krüger et al. 2021); ensemble dressing (Bröcker and Smith 2008); discrete impossibility and randomized repair (Kimpara et al. 2023); sample elicitation via variational divergences (Wei et al. 2021). Convolved scores under observation error: (Bröcker and Smith 2007; Ferro 2017; Bessac and Naveau 2021). Noisy-channel learning: (Patrini et al. 2017; Rooyen and Williamson 2018); surrogate scoring rules (Liu et al. 2022). Transform-properness: Allen et al. (2023, Prop. 4 published, Prop. 3 in arXiv v1) and Pic et al. (2025, Prop. 1) cover deterministic injective transforms; Theorem 2 is the Markov-kernel extension. The identity: (Stam 1959; Barron 1986; Lyu 2009; Song et al. 2021), with the denoising connection in Vincent (2011). Local scores: (Hyvärinen 2005; Parry et al. 2012). Nearest mechanism neighbours, each missing a leg: Lang et al. (2025), a Gaussian scale-ladder of proper scores as a training loss; Dudík et al. (2021), a multi-resolution market by partition refinement, subsidized rather than budget-balanced; and self-funding cloud wagering at a single scale (Lambert et al. 2008, 2015). A full audit with verdicts is in the companion research note.

10 Open problems

1. **Fair rungs.** The finite-m correction making each rung's expected score optimized by *sampling* from the belief, the log-score/KDE analog of the fair CRPS of Ferro (2014), and its interaction with the jitter. All results here are population-level; the finite-cloud game is not covered by the theorems.
2. **Endogenous bandwidth.** If the bandwidth is computed from the participant's own submission (Scott's rule on the cloud), the channel is no longer fixed and Theorem 2 does not apply; the participant controls both ρ and $h(\rho)$. Characterize the optimal joint deviation, and whether any self-referential bandwidth rule preserves truthfulness.
3. **Equilibrium of the proportional split.** For stake fraction $b < 1$ the best response to crowd r is $q = r + (p - r)/b$ truncated at zero (§2); truth is the symmetric equilibrium and the $b \to 0$ limit degenerates. Characterise the equilibrium with finitely many strategic participants, and, more generally, the incentives of pot splits proportional to (transformed) score values, which are not affine in the score and are not covered by Theorem 2.
4. **Choice of score.** Density pot-split (jittered, §4) vs. projection (sliced energy, §6) are both proper but reward different aspects of a forecast. Which yields better calibration and faster wealth concentration on skilled forecasters?
5. **Optimal scale weights.** For wealth-concentration objectives, is there a closed-form optimal $w(t)$ in the ladder of §5, and does it recover the likelihood weighting of Song et al. (2021)?

6. **Schur-damped projections.** Does anisotropic, covariance-shaped slicing with a reliability dial γ (§7) dominate isotropic slicing in the $p > n$ regime, and does it inherit the closed-form $\gamma^{\star}$?
7. **Variance of the sliced estimator.** How many directions are needed for the sliced score of §6 to rank participants correctly, as a function of d and the cloud size? (A Johnson–Lindenstrauss-style bound.)

References

Allen, Sam, David Ginsbourger, and Johanna Ziegel. 2023. "Evaluating Forecasts for High-Impact Events Using Transformed Kernel Scores." *SIAM/ASA Journal on Uncertainty Quantification* 11 (3): 906–40. https://doi.org/10.1137/22M1532184.

Barron, Andrew R. 1986. "Entropy and the Central Limit Theorem." *The Annals of Probability* 14 (1): 336–42. https://doi.org/10.1214/aop/1176992632.

Bessac, Julie, and Philippe Naveau. 2021. "Forecast Score Distributions with Imperfect Observations." *Advances in Statistical Climatology, Meteorology and Oceanography* 7: 53–71. https://doi.org/10.5194/ascmo-7-53-2021.

Bonneel, Nicolas, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. 2015. "Sliced and Radon Wasserstein Barycenters of Measures." *Journal of Mathematical Imaging and Vision* 51 (1): 22–45. https://doi.org/10.1007/s10851-014-0506-3.

Breiman, Leo. 1961. "Optimal Gambling Systems for Favorable Games." *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1: 65–78.

Bröcker, Jochen. 2012. "Evaluating Raw Ensembles with the Continuous Ranked Probability Score." *Quarterly Journal of the Royal Meteorological Society* 138 (667): 1611–17. https://doi.org/10.1002/qj.1891.

Bröcker, Jochen, and Leonard A. Smith. 2007. "Scoring Probabilistic Forecasts: The Importance of Being Proper." *Weather and Forecasting* 22 (2): 382–88. https://doi.org/10.1175/WAF966.1.

Bröcker, Jochen, and Leonard A. Smith. 2008. "From Ensemble Forecasts to Predictive Distribution Functions." *Tellus A* 60 (4): 663–78. https://doi.org/10.1111/j.1600-0870.2008.00333.x.

Candille, Guillem, and Olivier Talagrand. 2008. "Impact of Observational Error on the Validation of Ensemble Prediction Systems." *Quarterly Journal of the Royal Meteorological Society* 134 (633): 959–71. https://doi.org/10.1002/qj.268.

Cotton, Peter. 2022. *Microprediction: Building an Open AI Network*. MIT Press.

Cotton, Peter. 2024a. *A New Generative AI Competition in Eleven Dimensions*. https://microprediction.medium.com/a-new-generative-model-competition-to-test-the-old-against-the-new-189d3d460669.

Cotton, Peter. 2024b. *Density: Conventions for Densities Used in Various Games*. https://github.com/microprediction/density.

Cotton, Peter. 2024c. *Schur Complementary Allocation: A Unification of Hierarchical Risk Parity and Minimum Variance Portfolios*. arXiv:2411.05807. https://arxiv.org/abs/2411.05807.

Cotton, Peter. 2026. *Two Sides of Schur Damping: High-Dimensional Pseudo-Likelihoods and Portfolio Allocation*. arXiv:2606.14798. https://arxiv.org/abs/2606.14798.

CrunchDAO. 2024. *The MidOne Contest*. https://mid-one.crunchdao.com/.

Dudík, Miroslav, Xintong Wang, David M. Pennock, and David M. Rothschild. 2021. "Log-Time Prediction Markets for Interval Securities." *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 465–73. https://doi.org/10.65109/bbzi6501.

Ferro, Christopher A. T. 2014. "Fair Scores for Ensemble Forecasts." *Quarterly Journal of the Royal Meteorological Society* 140 (683): 1917–23. https://doi.org/10.1002/qj.2270.

Ferro, Christopher A. T. 2017. "Measuring Forecast Performance in the Presence of Observation Error." *Quarterly Journal of the Royal Meteorological Society* 143 (708): 2665–76. https://doi.org/10.1002/qj.3115.

Fricker, Thomas E., Christopher A. T. Ferro, and David B. Stephenson. 2013. "Three Recommendations for Evaluating Climate Predictions." *Meteorological Applications* 20 (2): 246–55. https://doi.org/10.1002/met.1409.

Gneiting, Tilmann, and Adrian E. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102 (477): 359–78. https://doi.org/10.1198/016214506000001437.

Hanson, Robin. 2007. "Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation." *The Journal of Prediction Markets* 1 (1): 3–15. https://doi.org/10.5750/jpm.v1i1.417.

Hyvärinen, Aapo. 2005. "Estimation of Non-Normalized Statistical Models by Score Matching." *Journal of Machine Learning Research* 6: 695–709.

Johnson, William B., and Joram Lindenstrauss. 1984. "Extensions of Lipschitz Mappings into a Hilbert Space." In *Conference on Modern Analysis and Probability*, vol. 26. Contemporary Mathematics. American Mathematical Society. https://doi.org/10.1090/conm/026/737400.

Kelly, Jr., J. L. 1956. "A New Interpretation of Information Rate." *The Bell System Technical Journal*

35 (4): 917–26. https://doi.org/10.1002/j.1538-7305.1956.tb03809.x.

Kimpara, Dhamma, Rafael Frongillo, and Bo Waggoner. 2023. "Proper Losses for Discrete Generative Models." *International Conference on Machine Learning (ICML)*, PMLR, vol. 202: 17015–40. https://arxiv.org/abs/2211.03761.

Koehler, Frederic, Alexander Heckett, and Andrej Risteski. 2023. "Statistical Efficiency of Score Matching: The View from Isoperimetry." *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/2210.00726.

Krüger, Fabian, Sebastian Lerch, Thordis Thorarinsdottir, and Tilmann Gneiting. 2021. "Predictive Inference Based on Markov Chain Monte Carlo Output." *International Statistical Review* 89 (2): 274–301. https://doi.org/10.1111/insr.12405.

Lambert, Nicolas S., John Langford, Jennifer Wortman, et al. 2008. "Self-Financed Wagering Mechanisms for Forecasting." *Proceedings of the 9th ACM Conference on Electronic Commerce (EC)*, 170–79. https://doi.org/10.1145/1386790.1386820.

Lambert, Nicolas S., John Langford, Jennifer Wortman Vaughan, et al. 2015. "An Axiomatic Characterization of Wagering Mechanisms." *Journal of Economic Theory* 156: 389–416. https://doi.org/10.1016/j.jet.2014.03.012.

Lang, Simon, Martin Leutbecher, and Pedro Maciel. 2025. *A Multi-Scale Loss Formulation for Learning a Probabilistic Model with Proper Score Optimisation*. arXiv:2506.10868. https://arxiv.org/abs/2506.10868.

Liu, Yang, Juntao Wang, and Yiling Chen. 2022. "Surrogate Scoring Rules." *ACM Transactions on Economics and Computation* 10 (3): 1–36. https://doi.org/10.1145/3565559.

Lyu, Siwei. 2009. "Interpretation and Generalization of Score Matching." *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI)*, 359–66.

Parry, Matthew, A. Philip Dawid, and Steffen Lauritzen. 2012. "Proper Local Scoring Rules." *The Annals of Statistics* 40 (1): 561–92. https://doi.org/10.1214/12-AOS971.

Patrini, Giorgio, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. 2017. "Making Deep Neural Networks Robust to Label Noise: A Loss Correction Approach." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2233–41. https://doi.org/10.1109/CVPR.2017.240.

Pennock, David M. 2004. "A Dynamic Pari-Mutuel Market for Hedging, Wagering, and Information Aggregation." *Proceedings of the 5th ACM Conference on Electronic Commerce (EC)*, 170–79. https://doi.org/10.1145/988772.988799.

Pic, Romain, Clément Dombry, Philippe Naveau, and Maxime Taillardat. 2025. "Proper Scoring Rules for Multivariate Probabilistic Forecasts Based on Aggregation and Transformation."

Advances in Statistical Climatology, Meteorology and Oceanography 11 (1): 23–58. https://doi.org/10.5194/ascmo-11-23-2025.

Rabin, Julien, Gabriel Peyré, Julie Delon, and Marc Bernot. 2012. "Wasserstein Barycenter and Its Application to Texture Mixing." *Scale Space and Variational Methods (SSVM)*, LNCS, vol. 6667: 435–46. https://doi.org/10.1007/978-3-642-24785-9_37.

Rooyen, Brendan van, and Robert C. Williamson. 2018. "A Theory of Learning with Corrupted Labels." *Journal of Machine Learning Research* 18 (228): 1–50.

Saetra, Øyvind, Hans Hersbach, Jean-Raymond Bidlot, and David S. Richardson. 2004. "Effects of Observation Errors on the Statistics for Ensemble Spread and Reliability." *Monthly Weather Review* 132 (6): 1487–501. https://doi.org/10.1175/1520-0493(2004)132%3C1487:EOOEOT%3E2.0.CO;2.

Siegert, Stefan, Christopher A. T. Ferro, David B. Stephenson, and Martin Leutbecher. 2019. "The Ensemble-Adjusted Ignorance Score." *Quarterly Journal of the Royal Meteorological Society* 145 (S1): 129–39. https://doi.org/10.1002/qj.3447.

Song, Yang, Conor Durkan, Iain Murray, and Stefano Ermon. 2021. "Maximum Likelihood Training of Score-Based Diffusion Models." *Advances in Neural Information Processing Systems (NeurIPS)*. https://arxiv.org/abs/2101.09258.

Song, Yang, and Stefano Ermon. 2019. "Generative Modeling by Estimating Gradients of the Data Distribution." *Advances in Neural Information Processing Systems (NeurIPS)* 32. https://arxiv.org/abs/1907.05600.

Song, Yang, Sahaj Garg, Jiaxin Shi, and Stefano Ermon. 2020. "Sliced Score Matching: A Scalable Approach to Density and Score Estimation." *Proceedings of the 35th Uncertainty in Artificial Intelligence Conference (UAI)*, PMLR, vol. 115: 574–84. https://arxiv.org/abs/1905.07088.

Stam, A. J. 1959. "Some Inequalities Satisfied by the Quantities of Information of Fisher and Shannon." *Information and Control* 2 (2): 101–12. https://doi.org/10.1016/S0019-9958(59)90348-1.

Székely, Gábor J., and Maria L. Rizzo. 2013. "Energy Statistics: A Class of Statistics Based on Distances." *Journal of Statistical Planning and Inference* 143 (8): 1249–72. https://doi.org/10.1016/j.jspi.2013.03.018.

Theis, Lucas, Aäron van den Oord, and Matthias Bethge. 2016. "A Note on the Evaluation of Generative Models." *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1511.01844.

Vecchia, A. V. 1988. "Estimation and Model Identification for Continuous Spatial Processes." *Journal of the Royal Statistical Society: Series B* 50 (2): 297–312. https://doi.org/10.1111/j.2517-6161.1988.tb01729.x.

Vincent, Pascal. 2011. "A Connection Between Score Matching and Denoising Autoencoders." *Neural Computation* 23 (7): 1661–74. https://doi.org/10.1162/NECO_a_00142.

Wei, Jiaheng, Zuyue Fu, Yang Liu, Xingyu Li, Zhuoran Yang, and Zhaoran Wang. 2021. "Sample Elicitation." *International Conference on Artificial Intelligence and Statistics (AISTATS)*, PMLR, vol. 130: 2692–700. https://arxiv.org/abs/1910.03155.

Wenliang, Li Kevin, and Heishiro Kanagawa. 2020. *Blindness of Score-Based Methods to Isolated Components and Mixing Proportions*. arXiv:2008.10087. https://arxiv.org/abs/2008.10087.

Zhang, Mingtian, Oscar Key, Peter Hayes, David Barber, Brooks Paige, and François-Xavier Briol. 2022. *Towards Healing the Blindness of Score Matching*. arXiv:2209.07396. https://arxiv.org/abs/2209.07396.

Implementation: `mechanisms/nearest_the_pin.py`*. Tests (self-funding, truthfulness, the projection identity, and Theorem 1 in both directions):* `tests/test_nearest_the_pin.py`*.*