Conditionally I.I.D. Models*

Drew Fudenberg† Giacomo Lanzani‡

January 13, 2026

Abstract

We characterize when an agent's initial forecast and one-step-ahead forecast revisions are consistent with a conditionally i.i.d. (CIID) model, i.e., Bayesian learning about a stable but unknown i.i.d. data-generating process. For two periods and binary outcomes, two simple conditions are necessary and sufficient: Symmetry (pairwise exchangeability) and Reinforcement (realized outcomes become weakly more likely). For two periods and arbitrary finite outcome sets, we show that a forecast system admits a CIID representation if and only if a forecast-derived matrix of joint probabilities is completely positive; with at most four outcomes, this reduces to positive semidefiniteness. We prove that one-step-ahead forecasts can never identify beliefs in positively autocorrelated outcomes, but some beliefs in negatively autocorrelated outcomes can be detected. For multi-period forecasts with binary outcomes, we derive an easily checked characterization of CIID representations by linking to the truncated moment problem, and show how the identified set of minimal-support rationalizations depends on the number of periods. Finally, we show that complete positivity of the associated moment tensor provides a general necessary and sufficient condition for multiple periods and multiple outcomes.

*We thank Ian Ball, Simone Cerreia-Vioglio, Anujit Chakraborty, Ben Golub, and Leeat Yariv for helpful conversations, and NSF grant SES-2417162 for financial support.

†Department of Economics, MIT

‡Department of Economics, UC Berkeley

1 Introduction

Economic models of learning typically assume that an agent learns about a fixed but unknown state of the world, so that observations are conditionally independent given that state. This conditional independence implies that the induced probability measure over sequences of observations is exchangeable, in the sense that the probability of a finite sequence is invariant under permutations. De Finetti [1937] and subsequent work show that exchangeability is both necessary and sufficient for a probability measure on infinite sequences of random variables to correspond to learning about a fixed state: any exchangeable distribution admits a Bayesian representation as a belief over a (possibly infinite) collection of i.i.d. data-generating processes. Thus, when the agent believes their data is exchangeable, it is as if they are learning about a fixed state of the world, whether or not they consciously think of the problem that way.

Exchangeability is defined purely in terms of the ex-ante distribution over sequences. In contrast, work by Shmaya and Yariv [2016], Bohren and Hauser [2024], and Molavi [2025] examine the consistency of beliefs across different periods, and provide necessary and sufficient conditions under which beliefs before and after receiving information in two-period models can be rationalized as the result of Bayesian updating. This paper lies at the intersection of these two approaches: As in the work following de Finetti [1937], we characterize the existence of a conditionally i.i.d. model in terms of beliefs over observable events. However, we follow the economic literature in supposing that, in addition to eliciting the agent's prior beliefs, the analyst can elicit the agent's beliefs after observing one or more realized outcomes.

In contrast to the papers cited in the previous paragraph, we do not allow the analyst to elicit the agent's beliefs about the data-generating process, but only their predictions about the outcome in the next period, given the outcomes so far. In contrast to the line of work started by de Finetti, which takes as given the decision maker's ex-ante specification of the complete probability distribution, we analyze how one-period-ahead forecasts are updated in response to realized observations. This is important because, although the agent may have initial beliefs that satisfy exchangeability, biases in updating could lead them to beliefs that are not consistent with a conditionally i.i.d. model. Our results show how to detect when this occurs.

They can be interpreted as a sort of converse of de Finetti's theorem: what properties of the forecasts *do* characterize finite sequences of random variables that are conditionally i.i.d.?

When there are only two periods and only two possible outcomes, the existence of a conditionally i.i.d. model consistent with the forecasts has a simple and intuitive characterization: Such a model exists if and only if the forecasting system satisfies the properties of *Symmetry* and *Reinforcement.* Symmetry requires that for every two outcomes i, j the probability of j given i, multiplied by the ex-ante probability of i, is equal to the probability of i given j, multiplied by the ex-ante probability of j, which is exactly the content of exchangeability in the two-period setting. Reinforcement requires that the conditional probability of outcome i weakly increases when outcome i is observed.

When there are two periods and more than two outcomes per period, Symmetry and Reinforcement are still necessary but are no longer sufficient. Instead, we show that a key role is played by a square matrix M derived from the agent's forecast, whose (i, j)-th entry is the product of the ex-ante probability of the j-th outcome and the probability of the i-th outcome given the j-th outcome. When there are four or fewer outcomes, (p, q) has a conditionally i.i.d. representation if and only if M is positive semidefinite. When there are more than four outcomes, the same characterization holds, but with positive semidefiniteness replaced by the (typically stronger) condition of *complete positivity.*[1]

We then develop an operational diagnostic that verifies consistency with a conditionally i.i.d. model by searching for a positive diagonal scaling that renders the rescaled forecast matrix diagonally dominant; success guarantees complete positivity.

We apply the general characterization to determine which departures from a belief in an i.i.d. process can be detected with these data. We show that it is impossible to detect a belief that the data generating process is persistent, i.e., the hot-hand fallacy, because a conditionally i.i.d. model can rationalize any next-period belief generated by such updating. In contrast, some cases of the opposite bias, in which the agent believes that the first-period outcome is less likely to be realized in the

[1] Roughly speaking, a matrix is completely positive if it can be built from a finite collection of nonnegative component vectors whose outer products add up to the matrix.

next period (as in the gambler's fallacy), can be detected, as can the beliefs of a decision maker who (correctly or not) believes they have observed a garbled version of the outcome.

For multi-period forecasts with binary outcomes, we derive an exact and easily checked characterization of the CIID representation by combining generalizations of the symmetry and reinforcement conditions with results for the truncated Hausdorff moment problem, and show how the identified set of minimal-support rationalizations depends on the number of periods. Finally, we show that complete positivity of the associated moment tensor provides a general necessary and sufficient condition for multiple periods and multiple outcomes.

Related work The seminal contributions on the characterizations of conditionally i.i.d. models are de Finetti [1937] and Hewitt and Savage [1955] respectively for the binary and general outcome case. Diaconis [1977] characterizes the implications of exchangeability on finite sequences, and Aldous, Ibragimov, and Jacod [2006] surveys subsequent results. In the case of binary outcomes, we also make use of Schoenberg [1932]'s theorem for the truncated moment problem, and Szegő [1975]'s results about orthogonal polynomials.

Molavi [2025] shows that beliefs about an unknown state are consistent with Bayesian updating if and only if the mean posterior belief about the state is absolutely continuous with respect to the prior. This finding generalizes the earlier work of Shmaya and Yariv [2016] by allowing the state space to be infinite and the decision maker's subjective beliefs to have support that does not match that of the true data-generating process. Bohren and Hauser [2024] characterizes the conditions under which a departure from Bayesian updating (e.g., underinference from signals) can be rationalized as a consequence of Bayesian updating within a misspecified model. Sarnoff [2025] highlights that it is more common for forecasts to violate "posterior statistical sufficiency"[2] than exchangeability. Catonini and Lanzani [2025] characterizes the only form of Dutch-book to which misspecified but Bayesian agents can be exposed.

The form of elicited beliefs we consider - predictions of the next outcome - is

[2] I.e., belief at period $t+1$ depends only on belief at period t and the period-t outcome.

elicited in the field in many settings, see, e.g., Weber, d'Acunto, Gorodnichenko, and Coibion [2022] and Greenwood and Shleifer [2014] for surveys on beliefs about inflation and stock returns, respectively.

Finally, this paper is related to decision-theoretic work on the dynamic consistency of optimal plans and how they force Bayesianism (e.g., Epstein and Le Breton, 1993, Green and Park, 1996, and Ghirardato, 2002).

2 The Two-Period Model

Objects There is a finite set $Y = \{1, ..., n\}$ of possible outcomes. In each period $t \in \{1, 2\}$, an outcome is realized and observed.

The agent's forecast of the period-1 outcome is $p \in \Delta(Y)$, and their forecast of the period-2 outcome conditional on observing outcome i is $q^{(i)} \in \Delta(Y)$. Together, we call this pair a *forecast.* We aim to characterize when these probabilities are consistent with a conditionally i.i.d. model. In our setting, an i.i.d. model is one where outcomes are drawn independently from a fixed distribution $\theta \in \Delta(Y)$. A conditionally i.i.d. model is then summarized by a probability measure $\mu \in \Delta(\Delta(Y))$.[3]

Definition 1. A forecast (p, q) has a *Bayes-rationalizing conditionally i.i.d. model (has a CIID representation)* if there exists a probability measure $\mu \in \Delta(\Delta(Y))$ such that:

(i) For all $i \in Y$:

$$p_i = \int_{\Delta(Y)} \theta_i \, d\mu(\theta). \tag{1}$$

(ii) For all $i, j \in Y$ with $p_j > 0$, the posterior measure $\mu(\cdot|j)$ is defined by Bayes' rule: for any Borel set $A \subseteq \Delta(Y)$,

$$\mu(A|j) = \frac{\int_{\theta \in A} \theta_j \, d\mu(\theta)}{p_j} \tag{2}$$

[3]For an arbitrary Borel-measurable set X in a Euclidean space, we let $\Delta(X)$ denote the probability distributions on X.

and the conditional forecast satisfies

$$q_i^{(j)} = \int_{\Delta(Y)} \theta_i \, d\mu(\theta|j) \tag{3}$$

(iii) For $j \in Y$ with $p_j = 0$, the value of $q^{(j)}$ is unrestricted.

That is, in a CIID model, the initial probability p_i is the expected value of the latent parameter θ_i, and the period-2 probability of outcome i conditional on the first outcome being j is the expected value of θ_i conditional on seeing outcome j.

Conditionally i.i.d. models generate exchangeable sequences of observations; this implies that forecasts must be symmetric in the following sense.

Definition 2. Forecast (p, q) satisfies *Symmetry* if for all $(i, j) \in Y^2$, $p_i q_j^{(i)} = p_j q_i^{(j)}$.

This is a direct consequence of exchangeability; the unconditional probability of seeing the sequence (i, j) must equal that of seeing (j, i). The following condition also holds in any conditionally i.i.d. model.

Definition 3. (p, q) satisfies *Reinforcement* if $p_i \leqslant q_i^{(i)}$ for all $i \in Y$.

Reinforcement requires that the probability of observing an outcome in the next period is not decreased by observing that outcome in the current period.[4] This distinguishes CIID models from other exchangeable models, such as the sampling without replacement examples discussed in Diaconis [1977]. CIID models imply reinforcement because the θ that assigns today's observed outcome relatively more probability will assign it relatively more probability tomorrow as well. Our formal proof uses Jensen's inequality.

Lemma 1. *If (p, q) has a CIID representation, then (p, q) satisfies Symmetry and Reinforcement.*

[4]Note that CIID representations do not imply that seeing a particular outcome makes all other outcomes less likely. For example, suppose there are three outcomes and that $\mu = (.1\theta', .9\theta'')$, with $\theta' = (.4, .4, .2)$ and $\theta'' = (.1, .1, .8)$. Then $p_1 = .13$ and $q_2^{(1)} = 5/26 > .13 = p_2$.

Proof. As noted above, Symmetry follows from the fact that conditionally i.i.d. models are exchangeable. To see why Reinforcement holds, observe that

$$\begin{aligned} q_i^{(i)} &= \frac{\int_{\Delta(Y)} \theta_i^2 \mathrm{d}\mu(\theta)}{p_i} \\ &\geqslant \frac{\left(\int_{\Delta(Y)} \theta_i \mathrm{d}\mu(\theta)\right)^2}{p_i} \\ &= p_i, \end{aligned}$$

where the first equality follows from the fact that $p_i q_i^{(i)}$ is the probability that outcome i occurs twice in a row, the inequality follows from Jensen's inequality, and the last equality follows from the definition of p_i. □

As we note in Section 7 below, Symmetry and Reinforcement remain necessary conditions for conditionally i.i.d. models when forecasts are elicited over more than two periods.

3 Binary outcomes

We begin by providing a simple characterization in the case where the outcome is binary (i.e., for Bernoulli random variables), as in de Finetti [1937]. The following result shows that in this case, Symmetry and Reinforcement are sufficient as well as necessary for the existence of a CIID representation.

Theorem 1. *When $n = 2$, (p, q) has a CIID representation if and only if (p, q) satisfies Symmetry and Reinforcement.*

We establish the sufficiency of Symmetry and Reinforcement constructively. The special cases $p_1 = q_1^{(2)}$, $q_1^{(2)} = 0$, and $p_1 = 1$ correspond to degenerate priors. When $1 > p_1 > q_1^{(2)} > 0$, we show there is a CIID representation with a Beta prior with parameters

$$\alpha = \frac{p_1 q_1^{(2)}}{p_1 - q_1^{(2)}} \qquad \text{and} \qquad \beta = \frac{q_1^{(2)}(1-p_1)}{p_1 - q_1^{(2)}}. \tag{4}$$

Remark. The proof uses a Beta prior for the non-degenerate cases, but this is not the only possible rationalization. Indeed, as we show in Section 5, whenever $q_1^{(2)} \neq p_1$, there are infinitely many CIID representations, including many with finite support. The Beta prior is convenient here because it is the conjugate prior for Bernoulli outcomes, has two parameters, and the moment formulas are simple functions of (α, β). But, the key point here is that *some* CIID representation exists, not that it must be Beta.

The theorem establishes that, in the Bernoulli case, Symmetry and Reinforcement completely characterize the existence of a Bayes rationalizing conditionally i.i.d. model. The proof also shows that whenever a non-trivial CIID rationalization exists, a rationalization with a Beta distribution also exists.

4 General Characterization: Complete Positivity

4.1 Necessary conditions

Paralleling the development of the characterization of exchangeability provided by Hewitt and Savage [1955], we now move beyond the case of binary outcomes. We begin with the following necessary condition, which lets us demonstrate that Reinforcement and Symmetry are no longer sufficient when there are more than two outcomes. Define the $n \times n$ matrix $M(p, q)$ by

$$m_{ij}(p, q) = p_j \, q_i^{(j)} \qquad \forall i, j \in Y.$$

By construction, $M_{ij}(p, q)$ is the joint probability of observing outcome j in period 1 and outcome i in period 2. The next lemma says that in a CIID model, this matrix must coincide with the second-moment matrix $\big(\mathbb{E}_\mu[\theta_i \theta_j]\big)_{i,j}$ of the latent parameter θ.

Lemma 2. *$\mu \in \Delta(\Delta(Y))$ is a CIID representation for (p, q) if and only if*

$$m_{ij}(p, q) = \int_{\Delta(Y)} \theta_i \theta_j d\mu(\theta) \qquad \forall i, j \in Y. \tag{5}$$

In this case, $M(p, q)$ is positive semidefinite because it is a mixture of rank-one positive semidefinite matrices.

The proof is based on the observation that if (p, q) has a CIID measure μ, $M(p, q)$ is the matrix of that measure's second moments. Now consider the following example.

Example 1. *Suppose $n = 3$, the initial forecast is $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and the second-period forecasts are*

$$q^{(1)} = (0.4, 0.5, 0.1), \quad q^{(2)} = (0.5, 0.4, 0.1), \quad q^{(3)} = (0.1, 0.1, 0.8).$$

Clearly (p, q) satisfies Symmetry and Reinforcement with

$$M(p, q) = \begin{bmatrix} 4/30 & 5/30 & 1/30 \\ 5/30 & 4/30 & 1/30 \\ 1/30 & 1/30 & 8/30 \end{bmatrix}.$$

Let $e = (1, -1, 0)^\top$. Because $M(p, q) \cdot e = -\frac{1}{30}e$, $M(p, q)$ is not positive semidefinite, so (p, q) does not have a Bayes-rationalizing conditionally i.i.d. model.

This example reveals a failure of conditional independence that Symmetry and Reinforcement alone cannot detect. Imagine a voter assessing a politician who can produce one of three policy outcomes: Left, Center, and Right. The voter's forecasts in the example imply that observing a "Left" policy makes "Center" the most likely policy next period, while observing a "Center" outcome makes "Left" the most likely outcome, with "Right" having a symmetric effect on Left and Center. Although this is consistent with Reinforcement for each outcome and the Symmetry condition, the fact that "Center" and "Left" each boost the other the most makes the matrix $M(p, q)$ not positive semidefinite, so the forecasts cannot be reconciled with a model of learning about a politician with a stable ideological "type."

4.2 General Characterization: Complete Positivity

The complete positivity of $M(p, q)$ will play a central role in our characterization.

Definition 4. The matrix M is *completely positive* if $M = BB^\top$ for some $n \times r$ nonnegative matrix B, or equivalently $M = \sum_{\ell=1}^{r} b_\ell b_\ell^\top, b_\ell \in \mathbb{R}^n_{\geqslant 0}$. When M is completely positive, its *cp rank* $\mathrm{cpr}(M)$ is the smallest r for which such a B exists.

The decomposition of a completely positive matrix bears a resemblance to diagonalization, with the key difference that the b vectors are not orthogonal. It is a stronger condition than being positive semi-definite, and thus a completely positive $M(p, q)$ immediately implies Symmetry. We will momentarily see that it also implies Reinforcement. There is an extensive literature characterizing the properties of completely positive matrices see, e.g., Berman and Shaked-Monderer [2003]. In particular we will make use of the easily-shown facts that diagonal matrices with nonnegative diagonal entries are completely positive and that convex combinations of completely positive matrices are completely positive (Theorem 2.2 in Berman and Shaked-Monderer [2003]).

The next result shows that $M(p, q)$ encodes all the restrictions implied by CIID representations in two-period models.

Theorem 2. *The following are equivalent:*

1. *$M(p, q)$ is completely positive.*
2. *(p, q) has a Bayes-rationalizing conditionally i.i.d. model.*
3. *(p, q) has a Bayes-rationalizing conditionally i.i.d. model that has finite support.*

To prove the result, we introduce a strengthening of complete positivity called simplex complete positivity, and show that it is equivalent to positivity in our setting because the entries of any $M(p, q)$ sum up to one. An $n \times n$ matrix M is *simplex completely positive* if it admits the decomposition

$$M = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)} \pi^{(s)^\top}, \quad \gamma_s > 0, \; \sum_{s=1}^{r} \gamma_s = 1, \; \pi^{(s)} \in \Delta(Y) \tag{6}$$

for some integer r. It is immediate that a simplex completely positive matrix is completely positive. When $\sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} = 1$, the converse is also true.

Claim 1. *If M is completely positive and $\sum_{i=1}^n \sum_{j=1}^n M_{ij} = 1$, then M is simplex completely positive.*

The proof of the theorem uses this claim to establish the cycle of implications (2) $\Rightarrow$ (1) $\Rightarrow$ (3) $\Rightarrow$ (2). To show that (2) implies (1), we note that if μ is a CIID representation for (p, q), then by Lemma 2, $M(p, q)$ is a mixture of rank-one matrices $\theta\theta^\top$ with $\theta \in \Delta(Y)$. From Carathéodory's theorem, it can be written as a finite convex combination of them, so it is completely positive. The proof that (1) implies (3) uses Claim 1 to infer that M is simplex completely positive so that $M(p, q) = \sum_s \gamma_s \pi^{(s)} \pi^{(s)\top}$ for some $\pi^{(1)}, \ldots, \pi^{(r)} \in \Delta(Y)$ and weights $\gamma_s > 0$ with $\sum_s \gamma_s = 1$. Define the discrete measure $\mu := \sum_s \gamma_s \delta_{\pi^{(s)}}$. By construction, this measure has second moments $\int \theta_i \theta_j \, d\mu = M_{ij}(p, q)$, so by Lemma 2 it is a CIID representation with finite support. Finally, that (3) implies (2) is trivial. Note that combining this theorem with Lemma 1 shows that if $M(p, q)$ is completely positive, then (p, q) satisfies Reinforcement.

Corollary 1. *Let p be strictly positive. If there is a Bayes-rationalizing conditionally i.i.d. model for (p, q) with infinite support, then the Bayes-rationalizing model is not unique. In particular, with binary outcomes if $q_1^{(2)} \neq p_1$, any forecast that has a Bayes-rationalizing conditionally i.i.d. model has at least two of them.*

4.2.1 Small Number of Outcomes

A second corollary of the linear algebra characterization of CIID models is that for a small number of outcomes ($n \leqslant 4$), positive semidefiniteness of the matrix $M(p, q)$ captures all of the empirical implications of conditionally i.i.d. models. This result provides a computationally simple and definitive test.

Corollary 2. *Let $n \leqslant 4$. The following are equivalent:*

1. *(p, q) has a Bayes-rationalizing conditionally i.i.d. model.*
2. *$M(p, q)$ is positive semidefinite.*

The next example shows that the equivalence of (1) and (2) does not extend to $n > 4$ even when (p, q) satisfies Reinforcement.

Example 2. *Let* $n = 5$, $p = \left(\frac{3}{23}, \frac{4}{23}, \frac{4}{23}, \frac{4}{23}, \frac{8}{23}\right)$, $q^{(1)} = \left(\frac{1}{3}, \frac{1}{3}, 0, 0, \frac{1}{3}\right)$, $q^{(2)} = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0, 0\right)$, $q^{(3)} = \left(0, \frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0\right)$, $q^{(4)} = \left(0, 0, \frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$, *and* $q^{(5)} = \left(\frac{1}{8}, 0, 0, \frac{1}{8}, \frac{3}{4}\right)$. *It is immediate that* (p, q) *satisfies Reinforcement and Symmetry and that*

$$M(p,q) = \begin{bmatrix} 1/23 & 1/23 & 0 & 0 & 1/23 \\ 1/23 & 2/23 & 1/23 & 0 & 0 \\ 0 & 1/23 & 2/23 & 1/23 & 0 \\ 0 & 0 & 1/23 & 2/23 & 1/23 \\ 1/23 & 0 & 0 & 1/23 & 6/23 \end{bmatrix}.$$

Crucially, $M(p,q) = \frac{A}{23}$, *where* A *is the matrix given in Example 2.4 of Berman and Shaked-Monderer [2003]. Therefore, since the sets of positive semidefinite matrices and completely positive matrices are both cones (see, e.g., Theorem 2.2 in Berman and Shaked-Monderer, 2003)* $M(p,q)$ *is positive semidefinite, but it is not completely positive, so by Theorem 2* (p,q) *does not admit a CIID representation.*

4.3 Cycles

We can gain further insight into the CIID model by analyzing its implications for cyclical patterns in belief updating. The direction of belief updates is determined by the covariance structure of the latent variable θ, because

$$q_j^{(i)} = \frac{\mathbb{E}_\mu[\theta_j \theta_i]}{\mathbb{E}_\mu[\theta_i]} = p_j + \frac{\mathrm{Cov}_\mu(\theta_j, \theta_i)}{p_i}. \tag{7}$$

Thus, observing outcome i makes outcome j strictly more likely if and only if $\mathrm{Cov}_\mu(\theta_j, \theta_i) > 0$.

Definition 5. A forecast (p, q) exhibits a *full cycle* if, for a given ordering of outcomes, $q_{i+1}^{(i)} > p_{i+1}$ for all $i \in Y$, where indices are taken modulo n. This is equivalent to $\mathrm{Cov}_\mu(\theta_{i+1}, \theta_i) > 0$ for all i.

The components of θ sum to 1, so

$$0 = \mathrm{Var}\Big(\sum_{k=1}^{n} \theta_k\Big) = \sum_{i=1}^{n} \mathrm{Var}(\theta_i) + 2 \sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}(\theta_i, \theta_j). \tag{8}$$

This simple identity is sufficient to rule out full cycles for small n.

Proposition 1. *Let $n \leqslant 4$. If forecast (p, q) admits a CIID representation, then it does not exhibit a full cycle.*

The simplex constraint is less binding for larger n, as the necessary negative covariances from equation (8) can be assigned to non-adjacent pairs, leaving the cycle covariances free to be positive.[5]

Proposition 2. *If $n \geqslant 5$, there exists a forecast (p, q) that exhibits a full cycle and has a CIID representation.*

4.4 Scaled Diagonal Dominance

Complete positivity of M is both necessary and sufficient for a Bayes-rationalizing conditionally i.i.d. representation, but may be difficult to verify. This motivates the search for easier-to-verify sufficient conditions for complete positivity. Here is one.

Definition 6. Forecast (p, q) satisfies *scaled diagonal dominance* if there are weights $s \in \mathbb{R}^n_{++}$ with $\sum_i s_i = 1$ satisfying $p_i\, q_i^{(i)}\, s_i \;\geqslant\; \sum_{j \neq i} p_j\, q_i^{(j)}\, s_j$ for each $i \in Y$.

Scaled diagonal requires the existence of positive weights such that, after rescaling each outcome i by s_i, the diagonal term dominates the total influence of all other outcomes on i. If such a reweighting exists, then M can be written as a convex combination of nonnegative rank-one outer products, and hence is completely positive. This yields the following corollary of Theorem 2:

Corollary 3. *If (p, q) satisfies scaled diagonal dominance, then it admits a CIID representation.*

From a practical perspective, checking scaled diagonal dominance reduces to computing the spectral radius of the row-ratio matrix and verifying that it is less than

[5] It is not a coincidence that the same critical value of n appears here as in the relation between complete positivity and PSD: In both cases the issue is that the geometry of R^n for $n \geqslant 5$ is qualitatively different (cf. Berman and Shaked-Monderer, 2003).

1. To see this, let $M(p,q)$ satisfy $m_{ii}(p,q) > 0$ for all $i \in Y$. Define the *row–ratio* matrix $R(p,q)$ by

$$R_{ij}(p,q) = \begin{cases} \dfrac{m_{ij}}{m_{ii}}, & i \neq j, \\ 0, & i = j. \end{cases}$$

Corollary 4. *Let (p,q) be such that $p_i q_i^{(i)} > 0$ for all $i \in Y$. If the spectral radius of $R(p,q)$ is less than 1, then (p,q) admits a CIID representation.*

Proof. The inequalities $m_{ii}(p,q)s_i \geqslant \sum_{j\neq i} m_{ij}(p,q)s_j$ are equivalent to $s \geqslant R(p,q)s$. If $\rho(R) < 1$ then

$$s := (I - R)^{-1}\mathbf{1} \ = \ \sum_{k=0}^{\infty} R^k \mathbf{1}$$

is well defined and strictly positive. Moreover, $(I - R)s = \mathbf{1}$ implies $s - Rs = \mathbf{1} \geqslant 0$, i.e. $s \geqslant Rs$, which implies that (p,q) satisfies scaled diagonal dominance. □

Remark. Row diagonal dominance of M requires $\sum_{j\neq i} R_{ij} \leqslant 1$ for every row, i.e. $\|R\|_\infty \leqslant 1$. Since $\rho(R) \leqslant \|R\|_\infty$, the spectral condition $\rho(R) < 1$ is strictly weaker and can hold even when some rows violate unscaled diagonal dominance.

5 Moments and Non-Uniqueness

When a two-period forecast has a CIID rationalization, it need not be unique. Moreover, uniqueness can fail in two ways: there may be CIID rationalizations with different support sizes, and there may be multiple CIID rationalizations with the same support size. This follows from the fact that CIID forecasts are determined by the first 2 moments of the latent variable θ, and many distributions can match the first two moments while differing at higher levels. This section illustrates this point for the case of binary outcomes and then discusses what can be said more generally.

5.1 Non-uniqueness with Binary Outcomes

Suppose there are two outcomes $Y = \{1, 2\}$, and suppose forecast (p,q) has a CIID representation μ where θ is unknown probability that $y = 1$, and $v = \mathbb{E}[\theta^2] -$

$(\mathbb{E}[\theta])^2 > 0$. Now consider a two-point prior:[6]

$$\Pr(\theta = \theta_L) = \lambda, \qquad \Pr(\theta = \theta_H) = 1 - \lambda, \qquad 0 < \lambda < 1,\ 0 \leqslant \theta_L \leqslant \theta_H \leqslant 1.$$

Let $d := \theta_H - \theta_L > 0$, then $\theta_L = p_1 - (1-\lambda)d, \quad \theta_H = p_1 + \lambda d$. Substituting into the second moment implies $d^2 = \dfrac{v}{\lambda(1-\lambda)}$, so the two-point priors that match (m_1, m_2) are

$$\theta_L = m_1 - (1-\lambda)\sqrt{\frac{v}{\lambda(1-\lambda)}}, \qquad \theta_H = m_1 + \lambda\sqrt{\frac{v}{\lambda(1-\lambda)}}$$

for any $\lambda \in (0,1)$ such that $0 \leqslant \theta_L \leqslant \theta_H \leqslant 1$.

Although there is an infinite continuum of two-point mixtures indexed by λ that reproduce the same two-period implications of any given Beta prior, these mixtures yield different third moments $\mathbb{E}[\theta^3] = \lambda\theta_L^3 + (1-\lambda)\theta_H^3$, and thus different predictions once a third outcome is observed. We discuss this further in the section on more than two periods.

5.2 Rank of $M(p,q)$ and support of the CIID representation

This section relates the forecasts (p,q) that can be rationalized by a prior with support of size r to the rank of the matrix $M(p,q)$. A consequence of this relation is that binary CIID models are characterized by the condition that $\operatorname{rank} M(p,q) = 2$.

Proposition 3. *If forecast (p,q) admits a CIID representation where μ has $r \geqslant 2$ point support then (p,q) satisfies Reinforcement and $M(p,q)$ is completely positive and* $\operatorname{rank} M \leqslant r$.

Proposition 3 is proved by establishing that if (p,q) has a CIID representation with support r, then the complete positive rank of $\operatorname{cpr}(M(p,q)) \leqslant r$. The result then follows from the general relation cpr $\geqslant$ rank. The next proposition uses the same connection between the size of the support and $\operatorname{cpr}(M(p,q))$, but now paired with the general inequality $\operatorname{cpr} \leqslant \operatorname{rank}(\operatorname{rank}+1)/2 - 1$.

[6]Then $p_1 = \mathbb{E}[\theta]$, $q_1^{(1)} = \mathbb{E}[\theta\theta]/\mathbb{E}[\theta]$, $q_1^{(2)} = \mathbb{E}[\theta(1-\theta)]/\mathbb{E}[1-\theta]$ and $\mathbb{E}[\theta] = \lambda\theta_L + (1-\lambda)\theta_H = p_1, \mathbb{E}[\theta^2] = \lambda\theta_L^2 + (1-\lambda)\theta_H^2$.

Proposition 4. *If forecast (p,q) satisfies Reinforcement and $M(p,q)$ is completely positive and has rank l, then (p,q) admits a CIID representation where μ has at most $l(l+1)/2-1$ point support.*

Say that a forecast is dogmatic if $p = q^{(i)}$ for every $i \in Y$.

Corollary 5. *For every nondogmatic forecast (p,q) the following statements are equivalent:*

1. *(p,q) admits a CIID representation where μ has a binary support;*

2. *(p,q) satisfies Reinforcement and $M(p,q)$ is completely positive and has rank 2.*

Proof. It immediately follows Propositions 3 and 4 and the observation that $\operatorname{rank} M(p,q) = 1$ for a dogmatic (p,q). □

Example 3 (A CIID model with support > rank M)**.** *Let $\theta \in [0,1]$ denote the Bernoulli success probability for outcome 1, and suppose the prior over θ has support $(\theta_L, \theta_M, \theta_H)$: $Pr(\theta = \theta_L) = \frac{1}{4}, \Pr(\theta = \theta_M) = \frac{1}{2}, \Pr(\theta = \theta_H) = \frac{1}{4}$, with $\theta_L = \frac{1}{3}, \theta_M = \frac{1}{2}, \theta_H = \frac{2}{3}$, and $p_1 = \mathbb{E}[\theta] = \frac{1}{4}\cdot\frac{1}{3} + \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{2}{3} = \frac{1}{2}$. Then $\mathbb{E}[\theta^2] = \frac{1}{4}\left(\frac{1}{9}\right) + \frac{1}{2}\left(\frac{1}{4}\right) + \frac{1}{4}\left(\frac{4}{9}\right) = \frac{19}{72}$, $\mathbb{E}[\theta(1-\theta)] = \frac{17}{72}$,*

$$q_1^{(0)} = \frac{\mathbb{E}[\theta(1-\theta)]}{\mathbb{E}[1-\theta]} = \frac{17/72}{1/2} = \frac{17}{36},$$

and $q_1^{(1)} = \frac{19}{36}$. Thus the implied second-period conditional forecasts satisfy Reinforcement, and

$$M(p,q) = \begin{pmatrix} p_1 q_{1|1} & (1-p_1) q_{1|0} \\ p_1(1-q_{1|1}) & (1-p_1)(1-q_{1|0}) \end{pmatrix} = \begin{pmatrix} \frac{19}{72} & \frac{17}{72} \\ \frac{17}{72} & \frac{19}{72} \end{pmatrix}$$

is symmetric, entrywise nonnegative, and has $\operatorname{rank} 2 < \operatorname{supp} \mu$.

6 Belief in Persistence or Reversal

We next apply our linear algebra characterization of CIID models to show that in the two-period setting, it is impossible to distinguish between an agent who has a

CIID model of the world and one who instead perceives persistence of the outcome process, a form of overreaction to the realized outcome.

Definition 7. (p, q) has a *persistent Bayesian representation* is there exists $\hat{q} = \{\hat{q}^{(1)}, ..., \hat{q}^{(N)}\}$ such that $(p, \hat{q})$ has a CIID representation and there is an $\alpha \in (0, 1)$ such that

$$q^{(i)} = \alpha e_i + (1 - \alpha)\hat{q}^{(i)} \quad \forall i \in \{1, ..., N\}, \tag{9}$$

where e_i is the unit vector corresponding to a point mass on the i-th outcome.

An agent whose forecasts have a persistent Bayesian representation, while in reality facing an i.i.d. environment, displays what has been called the *hot-hand fallacy* in the line of work pioneered by Gilovich, Vallone, and Tversky [1985].

Definition 8. (p, q) has a *reversing Bayesian representation* is there exists $\hat{q} = \{\hat{q}^{(1)}, ..., \hat{q}^{(N)}\}$ such that $(p, \hat{q})$ has a CIID representation and there is an $\alpha \in (0, 1)$ such that

$$\hat{q}^{(i)} = \alpha e_i + (1 - \alpha)q^{(i)} \qquad \forall i \in \{1, ..., N\}. \tag{10}$$

Having a reversing Bayesian representation when facing an i.i.d. environment corresponds to the *gambler's fallacy*, Tversky and Kahneman [1971].

Note that the only difference between these two definitions is that the roles of q and $\hat{q}$ are flipped, and that they both require that α is strictly between 0 and 1.[7] Finally, note that equation (10) can be rewritten as

$$\frac{\hat{q}^{(i)}}{(1 - \alpha)} - \frac{\alpha}{(1 - \alpha)} e_i = q^{(i)},$$

so that (p, q) has a reversing Bayesian representation if it can be derived from a CIID model that is modified so that after each outcome i, all outcomes $j \neq i$ receive a multiplicative boost of $1/(1-\alpha)$ to their probability, with the probability of outcome i decreasing accordingly.

[7]If we allowed $\alpha = 0$ in the definitions, every CIID representation would have both persistent and reversing Bayesian representations. Allowing $\alpha = 1$ in the persistent Bayesian representation would not change it, while allowing it in the reversing representation would make it trivially satisfied by every forecast.

Proposition 5. *If (p, q) has a persistent Bayesian representation then (p, q) has a CIID representation.*

Proof. Denote as m_{ij} the arbitrary entry of $M(p, q)$ and $\hat{m}_{ij}$ as the arbitrary entry of $M(p, \hat{q})$ where q and $\hat{q}$ are related as in equation (9). Then

$$m_{ij} = p_j q_i^{(j)} = p_j \left(\alpha \mathbb{I}_{i=j} + (1-\alpha)\, \hat{q}_i^{(j)}\right) = \alpha p_j \mathbb{I}_{i=j} + (1-\alpha)\, \hat{m}_{ij}$$

Therefore $M(p, q) = \alpha D + (1-\alpha) M(p, \hat{q})$ where D is the diagonal matrix with $d_{ii} = p_i$. By Example 2.1 in Berman and Shaked-Monderer [2003] D is completely positive. By our Theorem 2, $M(p, \hat{q})$ is also completely positive. By Theorem 2.2 of Berman and Shaked-Monderer [2003], $M(p, q)$ is also completely positive. Therefore (p, q) has a CIID representation by our Theorem 2. □

Thus (in the two-period case of this section), a belief that outcomes are somewhat persistent (i.e., positively correlated) is not distinguishable from a CIID model.

It is easy to see that the converse need not be true: If $n = 2$ and $p = q^{(1)} = q^{(2)} = (1/2, 1/2)$, then (p, q) admits a CIID representation with a dogmatic rationalizing belief $\mu = \delta_{(1/2,1/2)}$. However, any persistent Bayesian representation (9) would require the associated CIID representation to have $q_1^{(1)} < p_1$, a violation of Reinforcement, which is not possible by Lemma 1. For the same reason, unlike the persistent Bayesian representation, some form of reversing Bayesian representation can be spotted from the agent's forecast.[8]

Proposition 6. *Let (p, q) be a forecast such that $q_i^{(i)} < 1$ for all i. If (p, q) has a CIID representation, then (p, q) has a reversing Bayesian representation.*

Proof. Suppose first that $q_i^{(i)} < 1$ for all $i \in \{1, ..., N\}$. Define $\alpha = 1 - \max_{i \in \{1,...,N\}} q_i^{(i)}$. Let $\hat{q}^{(i)} = \alpha e_i + (1-\alpha) q^{(i)}$. Then $(p, \hat{q})$ satisfies Reinforcement. By Theorem 2, $M(p, q)$ is completely positive. Observe that $M(p, \hat{q}) = \alpha D + (1-\alpha) M(p, q)$, where D is the diagonal matrix with $d_{ii} = p_i$. Since D is completely positive by Example 2.1 of Berman and Shaked-Monderer [2003], $M(p, \hat{q})$ is a convex combination of

[8] This can also be shown directly: Let $(p, \hat{q})$ be the CIID representation derived from the dogmatic belief $\mu = \delta_{(1/2,1/2)}$ and set $q^{(i)} = 1/2 e_{-i} + 1/2 \hat{q}$. Then (p, q) has a reversing Bayesian representation, but it violates Reinforcement, so it does not admit a CIID representation by Lemma 1.

completely positive matrices, so by Theorem 2.2 of Berman and Shaked-Monderer [2003] it is also completely positive. Therefore, $(p, \hat{q})$ has a CIID representation by our Theorem 2, so (p, q) has a reversing Bayesian representation. □

We conclude this section by discussing a different perturbation of the CIID model that can be detected using outcome forecasts, namely, when the agent believes in a CIID model, but also that the first period observation is a noisy (i.e., garbled) version of the realized outcome. This departure from a belief in a CIID model can be detected, even in the particular case of binary outcomes and belief in arbitrarily small garblings. Indeed, forecasts obtained from these models will typically fail Symmetry unless overall the environment is symmetric, as otherwise in general $p_i q_j^{(\tilde{i})} \neq p_i q_j^{(\tilde{j})}$ where the tilde denotes the garbled realization of the outcome.[9]

7 CIID Models for More than Two periods

When the analyst has elicited one-period-ahead forecasts over a longer time horizon T, the CIID representation imposes additional constraints. This means that fewer forecasts will have CIID representations, and those that do will have fewer of them. This section characterizes the additional implications that can be extracted when we observe forecasts over more than two dates.

First, in the binary case, observing one-step-ahead forecasts up to horizon T identifies the first T moments of the latent Bernoulli parameter θ. The forecasts after histories with many ones correspond to higher-order moments of θ. We use this to connect to the truncated Hausdorff moment problem, and thus show that the existence of a CIID representation is equivalent to a finite collection of simple sign conditions on forward differences of these moments. This yields an exact and easily checked characterization of CIID models in terms of observable forecasts. At the same time, the truncation point T determines how tightly the prior is pinned down: with T odd, the minimal-support rationalization is unique, whereas with T even there is a one-dimensional family of distinct minimal-support priors that generate

[9]For a concrete example, suppose that the agent believes that it is equally likely that the outcome is 1 with probability 1/5 or 3/4. Moreover, they believe that the garbling changes each outcome into the other with probability ε. Easy computations show that Symmetry fails.

the same finite sequence of forecasts.

Similarly, in the multinomial case, forecasts over more than two periods deliver information about higher-order joint probabilities of outcomes and thus about higher-order moments of the latent θ. The relevant objects are the sequence of moment tensors $\{M^{(k)}\}_{k\leqslant T}$ associated with the joint distributions of $(Y_1, \ldots, Y_k)$ and with the multinomial count moments. Theorem 4 shows that a forecast system up to horizon T has a CIID representation if and only if these tensors are simplex completely positive and satisfy the natural marginal consistency identities; the matrix test based on complete positivity of $M(p, q)$ is the special case $T = 2$.

Multiple periods allow us to discriminate between misspecified models that are observationally indistinguishable in two-period data. For example, in the two-period setting, we showed that an agent who believes in spurious persistence (a hot-hand bias implemented by shifting probability weight toward the most recent outcome) cannot be distinguished from a CIID learner: such a perturbation preserves complete positivity of $M(p, q)$. Once we observe forecasts over longer histories, this invariance breaks down. The additional restrictions imposed by higher-order moments can rule out persistent or reversing updating rules that would otherwise pass all two-period tests.

Finally, the multi-period perspective clarifies the relationship between finite-horizon CIID tests and the infinite-horizon exchangeability results of de Finetti [1937] and Hewitt and Savage [1955]. For each fixed horizon T, our characterization identifies the precise moment conditions on the latent θ that are implied by CIID models and shows that they can be expressed as complete positivity of the associated moment tensors. As T grows, these conditions become tighter and, in the limit they converge to the full set of exchangeability restrictions. Thus, additional periods buy both sharper falsifiability of the CIID benchmark and sharper identification of the underlying prior when the benchmark is not rejected.

7.1 Necessary Conditions for $T > 2$

Fix an horizon $T \geqslant 2$. For $t \geqslant 0$, a history of length t is $h_t = (y_1, \ldots, y_t) \in \mathcal{Y}^t$, with $\varnothing$ denoting the empty history ($t = 0$). Define the count map ν by $\nu_i(h) = \sum_{s=1}^{|h|} \mathbf{1}\{y_s = i\}, i \in Y$, write e_i for the ith unit vector, and note that when outcome i is observed

$\nu((h,i)) = \nu(h) + e_i$.

For any history $h = (y_1, \ldots, y_t) \in Y^t$ with $t < T$, let $q^{(h)} \in \Delta(Y)$ denote the elicited one–step–ahead forecast of the period-$(t+1)$ outcome. A CIID representation has three immediate, easily-tested implications,

First, it must depend only on the count; this was vacuously true in the two-period setting.

Definition 9 (Count sufficiency)**.** Forecast $\{q^{(\nu)}\}_{|\nu|<T}$ satisfies *count sufficiency* if for any t and h_t, h_t' such that $\nu(h_t) = \nu(h_t')$ we have

$$q^{(h_t)} = q^{(h_t')} = q^{(\nu)}. \tag{11}$$

Next, there are two easily testable implications that generalize the necessary Symmetry and Reinforcement conditions for $T = 2$:

Definition 10 (Pairwise exchangeability)**.** Forecast $\{q^{(\nu)}\}_{|\nu|<T}$ satisfies *pairwise exchangeability* if for every node ν with $|\nu| \leqslant T - 2$ and all $i \neq j \in Y$,

$$q_i^{(\nu)}\, q_j^{(\nu+e_i)} \;=\; q_j^{(\nu)}\, q_i^{(\nu+e_j)}. \tag{12}$$

Note that repeated applications of these pairwise conditions show that pairwise exchangeability is equivalent to exchangeability.

Definition 11 (Reinforcement)**.** Forecast $\{q^{(\nu)}\}_{|\nu|<T}$ satisfies *Reinforcement* if for every node ν with $|\nu| \leqslant T - 2$ and all $i \in Y$,

$$q_i^{(\nu+e_i)} \;\geqslant\; q_i^{(\nu)}. \tag{13}$$

The final condition is the martingale property of beliefs. Like count sufficiency, this condition only has bite when there are more than two periods.

Definition 12 (Martingale property)**.** Forecast $\{q^{(\nu)}\}_{|\nu|<T}$ satisfies the *Martingale property* if for every node ν with $|\nu| \leqslant T - 2$,

$$q^{(\nu)} \;=\; \sum_{j \in Y} q_j^{(\nu)}\, q^{(\nu+e_j)}. \tag{14}$$

Lemma 3. *If $\{q^{(\nu)}\}_{|\nu|<T}$ is induced by a CIID model, then (11), (12), (13), and (14) all hold.*

Example 4 (Martingale restriction in three periods)**.** *Consider the binary case $Y = \{1,2\}$ and forecasts up to horizon $T = 3$. Suppose the forecaster reports the following one–step–ahead beliefs:*

$$q_1^{(0,0)} = \tfrac{1}{2}, \qquad q_1^{(1,1)} = \tfrac{4}{5}, \qquad q_1^{(0,1)} = \tfrac{1}{5},$$

and for histories of length two

$$q_1^{(2,2)} = \tfrac{4}{5}, \qquad q_1^{(1,2)} = \tfrac{1}{2}, \qquad q_1^{(0,2)} = \tfrac{1}{5}.$$

Up to horizon $T = 2$, these forecasts are compatible with a CIID model. Indeed, the period–1 forecasts coincide with those generated by a Beta prior with parameters $(\alpha, \beta) = (\frac{1}{3}, \frac{1}{3})$:

$$q_1^{(0,0)} = \frac{\alpha}{\alpha+\beta} = \tfrac{1}{2}, \qquad q_1^{(1,1)} = \frac{\alpha+1}{\alpha+\beta+1} = \tfrac{4}{5}, \qquad q_1^{(0,1)} = \frac{\alpha}{\alpha+\beta+1} = \tfrac{1}{5},$$

so all two–period tests based on $M(p,q)$ are passed.

Once we elicit forecasts after two outcomes, the martingale condition (14) imposes additional restrictions. For instance, at the node with one success and no failures, $\nu = (1,0)$, the martingale condition requires

$$q_1^{(1,1)} \;=\; q_1^{(1,1)}\, q_1^{(2,2)} \;+\; q_2^{(1,1)}\, q_1^{(1,2)},$$

that is,

$$\tfrac{4}{5} \;=\; \tfrac{4}{5} \cdot q_1^{(2,2)} \;+\; \tfrac{1}{5} \cdot q_1^{(1,2)}.$$

Under the reported forecasts,

$$\tfrac{4}{5} \quad vs \quad \tfrac{4}{5} \cdot \tfrac{4}{5} + \tfrac{1}{5} \cdot \tfrac{1}{2} \;=\; \tfrac{16}{25} + \tfrac{1}{10} \;=\; \tfrac{37}{50},$$

so the equality fails.

Intuitively, after seeing outcome 1 once, the forecaster assigns probability $q_1^{(1,1)} =$

$\frac{4}{5}$ to a further success. If the process were CIID, their current belief $\frac{4}{5}$ would equal the expected value of their belief after the second period, averaging over the two possible second outcomes using her own current beliefs as weights. In this example, on average the forecaster expects to end up with a different belief than they currently hold. This violates the martingale property and therefore rules out any CIID representation on horizon $T = 3$, even though the forecasts up to $T = 2$ admit a CIID rationalization. This pattern can be interpreted as an overreaction to the first success or failure: the forecaster uses a simple rule that maps the empirical frequency of outcome 1 into a forecast, but this rule cannot arise from conditioning on a fixed latent θ, because it breaks the martingale restriction.

7.2 Binary Outcomes and More than Two Periods

The problem is particularly tractable in the case of binary outcomes, so we will start with that. Let $Y = \{1, 2\}$. For $t < T$ and $j \in \{0, \ldots, t\}$, write $q^{(j,t)} := q_1^{(\nu)}$ where ν has j ones and $t - j$ zeros. The multinomial conditions reduce to the following simple, testable restrictions.

Lemma 4 (Binary necessary conditions for $T > 2$)**.** *For every node (j, t) with $0 \leqslant j \leqslant t \leqslant T - 2$:*

$$q^{(j,t)}\big(1 - q^{(j+1,t+1)}\big) \;=\; \big(1 - q^{(j,t)}\big)\, q^{(j,t+1)}, \tag{15}$$

$$q^{(j,t)} \;\leqslant\; q^{(j+1,t+1)}, \tag{16}$$

Remark. There are only two equations here instead of four because notation here imposes count sufficiency, and with binary outcomes the symmetry condition is equivalent to the martingale condition.

7.2.1 Necessary and Sufficient Condition when $|Y| = 2$

In the binary-outcome case, we can work with a one-dimensional latent state θ where $\mathbb{E}_\mu(\theta) = p_1$ The following lemma shows that the moments $m_r = \mathbb{E}[\theta^r]$ are identified by the forecasts after observing a history of only successes (outcome "1").

Lemma 5 (Identification of Moments from Forecasts)**.** *Suppose $|Y| = 2$ and that a CIID representation exists for a sequence of one-step-ahead forecasts up to horizon*

T. For each $k \in \{1, \ldots, T-1\}$, let q^*_{k+1} be the one-step-ahead forecast for outcome 1 after observing k consecutive 1s:

$$p_1 = \mathbb{P}(X_1 = 1), \qquad q^*_{k+1} = \mathbb{P}(X_{k+1} = 1 \mid X_1 = \cdots = X_k = 1).$$

These forecasts uniquely identify the moments $\{m_1, \ldots, m_T\}$ of the latent variable $\theta \in [0,1]$ via the recursive formula $m_1 = p_1, m_{k+1} = m_k \cdot q^*_{k+1}$ for $k \in \{1, \ldots, T-1\}$.

This result shows that, given the forecasts, we can construct a unique candidate sequence of moments $\{m_r\}_{r=1}^T$. The remaining task, addressed by Lemma 6, is to determine whether this sequence of numbers could have been generated by a valid probability measure on $[0,1]$. This is precisely the truncated Hausdorff moment problem.

For integers $a, b \geqslant 0$, define the mixed moments $m_{a,b} := \mathbb{E}\big[\theta^a(1-\theta)^b\big]$. For any history with j realizations of 1 and $t-j$ realizations of 2, the condition for the one–step forecasts to be consistent with the given mixed moments is

$$q_1^{(j,t)} = \frac{m_{j+1,\,t-j}}{m_{j,\,t-j}}, \qquad \forall t \in \{1, ..., T-1\}, \forall j \in \{1, ..., t\}, \tag{17}$$

where the condition is deemed satisfied whenever the denominator is equal to 0. Also define forward differences $\Delta m_r := m_{r+1} - m_r$ and $\Delta^{s+1} m_r := \Delta(\Delta^s m_r)$.

Lemma 6 (Moment characterization of Binary CIID)**.** *Suppose $|Y| = 2$, and let $m_0, \ldots, m_T \in \mathbb{R}$ be the moments implied by a forecast system (p,q). The following are equivalent.*

(i) There exists a probability measure μ on $[0,1]$ and $\theta \sim \mu$ such that the forecasts admit a CIID representation up to horizon T.

(ii) Equation (17) is satisfied and the Hausdorff truncated moment conditions

$$(-1)^s \, \Delta^s m_r \;\geqslant\; 0 \qquad \text{for all integers } r, s \geqslant 0 \text{ with } r+s \leqslant T \tag{18}$$

hold.

The alternating sign pattern of the forward differences $\Delta^s m_r$ in equation (18) captures the requirement that m_r can be written as $\int_0^1 x^r d\mu(x)$ for some probability measure μ on $[0, 1]$. Thus, in the binary setting, when forecasts satisfy equation (17), checking for the existence of a CIID representation reduces to verifying these finite collections of inequalities on forward differences.

Combining Lemmas 5 and 6 yields the following theorem.

Theorem 3. *When $|Y| = 2$, a sequence of one-period forecasts up to period T has a CIID representation if and only if they satisfy equations (17) and (18).*

7.2.2 Parity

A striking feature is a parity effect: when we observe an odd number of moments, the minimal-support prior is unique; with an even number, there is a continuum of minimal-support priors.

Proposition 7 (Parity effect for binary CIID representations)**.** *Suppose $|Y| = 2$ and that a sequence of one-step-ahead forecasts up to horizon T admits a CIID representation, with associated moment sequence $\{m_r\}_{r=0}^{T}$ that strictly satisfy the Hausdorff moment inequalities.*

1. *If $T = 2k - 1$ is odd, then there is a unique CIID representation whose prior μ has support of size k, and no CIID representation exists with support strictly smaller than k.*
2. *If $T = 2k$ is even, then there is no CIID representation whose prior has support of size at most k. Moreover, the set of CIID representations with minimal support $k + 1$ is a one-dimensional family of distinct priors.*

The proof of this result is in Appendix A.13. If $T = 2k-1$ is odd and the Hausdorff moment inequalities hold strictly, then there is a unique CIID representation whose prior has support of size k, and no representation with smaller support exists. By contrast, if $T = 2k$ is even, then (under the same nondegeneracy condition) there is no CIID representation with support of size k, and the set of CIID representations with minimal support $k + 1$ is a one-dimensional family of distinct priors. For example, when we observe three moments ($T = 3$), the minimal–support CIID prior is unique and has support on two points. By contrast, with four observed moments ($T = 4$), under the same nondegeneracy conditions there is a one–dimensional continuum of distinct three–point priors that all generate the same finite sequence of forecasts. Intuitively, this "parity effect" comes from counting dimensions in the truncated moment problem: a k–point prior on $[0,1]$ has $2k-1$ free parameters (support locations and probabilities), so $T = 2k-1$ moments can pin it down uniquely, while $T = 2k$ moments does not pin down the $2k+1$ parameters of a $(k+1)$–point prior.

7.3 Many Outcomes and Many Periods

7.3.1 Necessary Conditions

When there are $n > 2$ outcomes and horizon $T \geqslant 1$, there are additional necessary conditions beyond those in Section 7.1. For a count vector $\nu = (\nu_1, \ldots, \nu_n) \in \mathbb{N}^n$ with $|\nu| := \sum_i \nu_i \leqslant T$, and a CIID model μ, write

$$\theta^\nu \;:=\; \prod_{i\in Y} \theta_i^{\nu_i}, \qquad m_\nu \;:=\; \mathbb{E}_\mu\big[\theta^\nu\big], \qquad \forall \theta = (\theta_1, \ldots, \theta_n) \in \Delta(Y).$$

In a CIID model μ, the forecasts depend only on ν, and

$$q_i^{(\nu)} \;=\; \mathbb{E}[\theta_i \mid \nu] \;=\; \frac{m_{\nu+e_i}}{m_\nu}, \qquad i \in Y, \;\; |\nu| < T.$$

Moreover, the moment sequence $\{m_\nu\}$ must satisfy the linear consistency identities that

$$\sum_{j\in Y} m_{\nu+e_j} = m_\nu \qquad \text{for all } |\nu| \leqslant T-1.$$

This ensures that $\sum_{j\in Y} q_j^{(\nu)} = 1$, for every count vector ν.[10] Define coordinate forward differences on the moment array by

$$(\Delta_i m)_\nu \;:=\; m_{\nu+e_i} - m_\nu, \qquad \Delta^\beta \;:=\; \prod_{i\in Y} \Delta_i^{\beta_i} \quad (\beta \in \mathbb{N}^n).$$

[10] Fix $k \geqslant 2$ and any multi-index $\nu = (\nu_i)_{i\in Y}$ with $|\nu| = k-1$. By definition, $m_\nu = \mathbb{E}\big[\prod_{i\in Y} \theta_i^{\nu_i}\big]$ and $m_{\nu+e_j} = \mathbb{E}\big[\theta_j \prod_{i\in Y} \theta_i^{\nu_i}\big]$ for each $j \in Y$. Hence

$$\sum_{j\in Y} m_{\nu+e_j} = \mathbb{E}\Big[\Big(\sum_{j\in Y} \theta_j\Big) \prod_{i\in Y} \theta_i^{\nu_i}\Big] = \mathbb{E}\big[\prod_{i\in Y} \theta_i^{\nu_i}\big] = m_\nu.$$

The forward-difference operators $\Delta^\beta m$ and the identities in the next lemma capture the two requirements that appeared in the binary case: (i) the existence of a non-negative representing measure on the simplex, and (ii) the constraint that $\sum_i \theta_i = 1$ almost surely. The inequalities $(-1)^{|\beta|}(\Delta^\beta m)_\nu \geqslant 0$ ensure that all mixed monomials $\theta^\nu \prod_i (1-\theta_i)^{\beta_i}$ have nonnegative expectation, while the consistency requirements ensure that the moments respect the simplex normalization. Together, these conditions are necessary for a CIID representation.

Lemma 7 (Necessary conditions for multinomial CIID representation up to horizon T)**.** *If there exists a probability measure on $\Delta(Y)$ with moments $\{m_\nu\}_{|\nu|\leqslant T}$ (hence forecasts $q_i^{(\nu)} = m_{\nu+e_i}/m_\nu$), then for all $\nu, \beta \in \mathbb{N}^n$ with $|\nu| + |\beta| \leqslant T$ the following hold.*

(a)

$$(-1)^{|\beta|}\,(\Delta^\beta m)_\nu \;=\; \mathbb{E}\Bigg[\Big(\prod_{i\in Y}\theta_i^{\nu_i}\Big)\Big(\prod_{i\in Y}(1-\theta_i)^{\beta_i}\Big)\Bigg] \;\geqslant\; 0.$$

(b) Simplex normalization (linear consistency across degrees): for every $b \in \{1,\dots,T-|\nu|\}$,

$$\sum_{|\beta|=b}\binom{b}{\beta} m_{\nu+\beta} \;=\; \mathbb{E}\Big[\theta^\nu\Big(\sum_{i\in Y}\theta_i\Big)^b\Big] \;=\; m_\nu,$$

in particular $\sum_{i\in Y} m_{\nu+e_i} = m_\nu$.

(c) Zeroth-degree normalization: $m_{\mathbf{0}} = \mathbb{E}[1] = 1$ (equivalently, $\sum_{i\in Y} m_{e_i} = 1$).

Remarks. 1) For $n = 2$ this reduces to the binary conditions: $m_{j,b} = \mathbb{E}[\theta^j(1-\theta)^b]$, $q_1^{(j,b)} = m_{j+1,b}/m_{j,b}$, and $(-1)^b\Delta_1^b m_{(j,0)} \geqslant 0$. 2) Unlike with binary outcomes, the Hausdorff inequalities are not sufficient here. For example, the necessary conditions above are satisfied by the forecast system in Example 1, which does not have a CIID representation. As the number of periods T grows, the necessary conditions in Proposition 7 become tighter, but (we conjecture that) they are only sufficient in the limit $T \to \infty$.

7.3.2 Tensor characterization

For binary outcomes, the key objects were the power moments $m_r = \mathbb{E}[\theta^r]$. For general outcomes, the observable counterparts are the order-k sequence tensors defined from joint probabilities of the data. Under a CIID representation, these coincide with the k-th moment tensors of the latent θ. Our main multi-period result shows that CIID rationalizability is equivalent to each of these tensors being simplex completely positive, together with a natural consistency condition implied by the law of total probability.

Definition 13 (Sequence and Count Tensors)**.** For a given stochastic process $(X_t)_{t=1}^T$ and $k \in \{1, \ldots, T\}$, the order-k *sequence tensor* $T^{(k)}$ is defined by its components:

$$T^{(k)}_{i_1 \cdots i_k} \;:=\; \mathbb{P}(X_1 = i_1, \ldots, X_k = i_k) \quad \text{for } (i_1, \ldots, i_k) \in Y^k.$$

If a CIID representation with latent variable $\theta \in \Delta(Y)$ drawn from a measure μ exists, these components are the moments of the outer product of θ:

$$T^{(k)}_{i_1 \cdots i_k} \;=\; \mathbb{E}_\mu\Big[\prod_{l=1}^{k} \theta_{i_l}\Big].$$

Under the CIID hypothesis, we will refer to $T^{(k)} = \mathbb{E}_\mu[\theta^{\otimes k}]$ as the *moment tensor* $M^{(k)}$, with entries $M^{(k)}_{i_1 \ldots i_k} = \mathbb{E}[\prod_{l=1}^{k} \theta_{i_l}]$.

The forecasts provided by an agent are the one-step-ahead conditional probabilities. If these forecasts depend only on the counts of past outcomes, as they must in a CIID model, we can express them directly in terms of the count moments.

Definition 14 (Simplex Completely Positive Tensors)**.** A symmetric order-k tensor S with components $S_{i_1 \cdots i_k}$ is *simplex completely positive (SCP)* if it can be written as a convex combination of rank-one tensors generated by vectors in the simplex. That is, there exist $r \in \mathbb{N}$, weights $\gamma_\ell \geqslant 0$ with $\sum_{\ell=1}^{r} \gamma_\ell = 1$, and points $\pi^{(\ell)} \in \Delta(Y)$ such that

$$S_{i_1 \cdots i_k} \;=\; \sum_{\ell=1}^{r} \gamma_\ell \prod_{j=1}^{k} \pi^{(\ell)}_{i_j}.$$

Also, for each $k \in \{1, \ldots, T\}$, define the *order-k count moment tensor* $M^{(k)}$ by

$$M^{(k)}_\nu := m_\nu \quad \text{for all } \nu \text{ with } |\nu| = k.$$

This can be viewed as symmetric tensor on Y^k with entries $M^{(k)}_{i_1 \cdots i_k} := m_{\nu(i_1, \ldots, i_k)}$, where $\nu(i_1, \ldots, i_k)$ is the count vector of the multi–index $(i_1, \ldots, i_k)$. When expressed using count moments $\{m_\nu\}_{|\nu|=k}$ corresponding to the symmetric tensor S, this is equivalent to:

$$m_\nu \;=\; \sum_{\ell=1}^{r} \gamma_\ell \left(\pi^{(\ell)}\right)^\nu \qquad \text{(for all } |\nu| = k\text{)}.$$

The main result of this section generalizes the CIID characterization of Theorem 2 to more than two periods.

Theorem 4. *The forecast system (p, q) has a CIID representation if and only if the associated count moment tensors $\{M^{(k)}\}_{k=1}^T$ are simplex completely positive.*

A key step of the proof is Lemma B.1, which is another version of the truncated moment problem on the simplex. The Lemma gives two properties of the moment tensor that are necessary and sufficient for there to be a probability distribution over $\theta \in \Delta(Y)$ whose moments up to order T are (m_ν). The hard direction is showing these conditions are enough to reconstruct a measure μ. The proof builds a linear functional L that "pretends" to be integration against μ: $L(\theta^\nu) = m_\nu$. The consistency identities imply L treats the sum $S(\theta) = \sum_j \theta_j$ as if it were the constant 1, mimicking the fact that $S = 1$ on the simplex. The simplex-complete-positivity of the order-T tensor then forces $L(p) \geqslant 0$ for any polynomial p that is nonnegative on the simplex. With this positivity in hand, a separating-hyperplane/convex-hull argument shows the vector $m = (m_\nu)$ must lie in the convex hull of "moment vectors" (θ^ν) coming from actual points $\theta \in \Delta(Y)$. That means m can be written as a finite mixture of such pointwise moment vectors, which is exactly the same as saying there exists a probability measure μ on $\Delta(Y)$ with those moments.

8 Discussion

This paper characterizes when a sequence of one-step-ahead forecasts is consistent with a CIID model. For the two-period, binary-outcome case, the conditions are simple and intuitive: Symmetry and Reinforcement. For more outcomes, these conditions are necessary but not sufficient; the key object becomes the second-moment matrix $M(p, q)$, which must be completely positive. For more periods, the entire hierarchy of moment tensors must be simplex completely positive and satisfy linear consistency identities.

These results provide a clear, operational way to test whether observed forecasting behavior can be explained by a classic model of learning about a stable, unknown environment. For example, forecasts where outcome i is most reinforced after outcome j and vice versa, as in Example 1, can be immediately flagged as non-Bayesian in this sense. More generally, any failure of complete positivity (or, in the $n \leqslant 4$ case, positive semidefiniteness) is a definitive sign that the agent's updating rule is inconsistent with any CIID representation.

Conversely, our results on persistence show the limits of what one-step-ahead forecasts can reveal. An agent who believes in spurious positive autocorrelation (persistence) will generate forecasts that are indistinguishable from CIID models. This is because adding persistence to a CIID model preserves the complete positivity of the moment structures. Detecting such biases would require richer data, such as eliciting multi-step-ahead forecasts or beliefs about the underlying data-generating process itself.

The connection to the truncated moment problem in the multi-period binary case shows that observing an odd number of moments identifies a unique minimal-support prior can be identified, while an even number of moments leaves some indeterminacy. This has direct implications for applied work attempting to estimate belief structures from observed forecasts.

References

Aldous, D. J., I. A. Ibragimov, and J. Jacod (2006). *Ecole d'Ete de Probabilites de Saint-Flour XIII, 1983*. Vol. 1117. Springer.

Aliprantis, C. and K. Border (2013). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag.

Berman, A. and R. J. Plemmons (1994). *Nonnegative matrices in the mathematical sciences*. SIAM.

Berman, A. and N. Shaked-Monderer (2003). *Completely positive matrices*. World Scientific.

Bohren, J. A. and D. N. Hauser (2024). "The behavioral foundations of model misspecification: A decomposition." *Working paper*.

Catonini, E. and G. Lanzani (2025). "A Dutch-book trap for misspecification." *mimeo*.

Diaconis, P. (1977). "Finite forms of de Finetti's theorem on exchangeability." *Synthese*, 36, 271–281.

Diananda, P. H. (1962). "On non-negative forms in real variables some or all of which are non-negative." *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 58. Cambridge University Press, 17–25.

Epstein, L. G. and M. Le Breton (1993). "Dynamically consistent beliefs must be Bayesian." *Journal of Economic theory*, 61, 1–22.

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

Ghirardato, P. (2002). "Revisiting Savage in a conditional world." *Economic theory*, 20, 83–92.

Gilovich, T., R. Vallone, and A. Tversky (1985). "The hot hand in basketball: On the misperception of random sequences." *Cognitive psychology*, 17, 295–314.

Green, E. J. and I.-U. Park (1996). "Bayes contingent plans." *Journal of Economic Behavior & Organization*, 31, 225–236.

Greenwood, R. and A. Shleifer (2014). "Expectations of returns and expected returns." *The Review of Financial Studies*, 27, 714–746.

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

Kreuin, M., A. Nudel, et al. (1977). "The Markov moment problem and extremal problems." *Translations of mathematical monographs*.

Molavi, P. (2025). "Misspecified Bayesianism." *arXiv preprint arXiv:2109.07007*.

Sarnoff, K. (2025). "The Structure of Sequential Updating."

Schoenberg, I. (1932). "On finite and infinite completely monotonic sequences." *Bulletin of the American Mathematical Society*, 2, 72–76.

Shmaya, E. and L. Yariv (2016). "Experiments on decisions under uncertainty: A theoretical framework." *American Economic Review*, 106, 1775–1801.

Szegő, G. (1975). *Orthogonal Polynomials*. 4th. Vol. 23. Colloquium Publications. American Mathematical Society.

Tversky, A. and D. Kahneman (1971). "Belief in the law of small numbers." *Psychological bulletin*, 76, 105.

Weber, M., F. d'Acunto, Y. Gorodnichenko, and O. Coibion (2022). "The subjective inflation expectations of households and firms: Measurement, determinants, and implications." *Journal of Economic Perspectives*, 36, 157–184.

A Appendix

A.1 Proof of Theorem 1

Proof. Lemma 1 shows that Symmetry and Reinforcement are necessary. To show they are sufficient, suppose forecast (p, q) satisfies Symmetry and Reinforcement.

Case a If $1 > p_1 > q_1^{(2)} > 0$, consider a Beta distribution with parameters given by equation (4). The implied probability of outcome 1 in the first period is then is

$$\tilde{p}_1(\alpha, \beta) = \frac{\alpha}{\alpha + \beta} = \frac{\frac{p_1 q_1^{(2)}}{p_1 - q_1^{(2)}}}{\frac{p_1 q_1^{(2)}}{p_1 - q_1^{(2)}} + \frac{q_1^{(2)}(1-p_1)}{p_1 - q_1^{(2)}}} = \frac{p_1 q_1^{(2)}}{p_1 q_1^{(2)} + q_1^{(2)}(1 - p_1)} = p_1,$$

as desired.

The period-2 forecast probabilities are

$$\tilde{q}_1^{(1)}(\alpha, \beta) = \frac{\alpha + 1}{\alpha + \beta + 1}, \; \tilde{q}_1^{(2)} = \frac{\alpha}{\alpha + \beta + 1}, \text{ and}$$

$$\tilde{q}_1^{(2)}(\alpha, \beta) = \frac{\alpha}{\alpha + \beta + 1} = \frac{\frac{p_1 q_1^{(2)}}{p_1 - q_1^{(2)}}}{\frac{p_1}{p_1 - q_1^{(2)}}} = q_1^{(2)}, \text{ as desired.}$$

Note that $\tilde{q}_1^{(1)}(\alpha, \beta) = \frac{\alpha+1}{\alpha} \tilde{q}_1^{(2)}(\alpha, \beta)$, $\alpha + 1 = \frac{p_1 q_1^{(2)}}{p_1 - q_1^{(2)}} + 1 = \frac{p_1 q_1^{(2)} + p_1 - q_1^{(2)}}{p_1 - q_1^{(2)}}$, and $\frac{\alpha+1}{\alpha} = \frac{p_1 q_1^{(2)} + p_1 - q_1^{(2)}}{p_1 q_1^{(2)}}$. Thus

$$\tilde{q}_1^{(1)}(\alpha, \beta) = \frac{p_1 q_1^{(2)} + p_1 - q_1^{(2)}}{p_1} = q_1^{(2)} + 1 - q_1^{(2)}/p_1.$$

Symmetry implies that $p_1(1-q_1^{(1)}) = (1-p_1)q_1^{(2)}$, so

$$q_1^{(1)} = q_1^{(2)} + 1 - q_1^{(2)}/p_1 = \tilde{q}_1^{(1)}(\alpha, \beta),$$

which completes the proof for case a).

Case b If $p_1 = q_1^{(2)}$, Symmetry implies that $p_2 = q_2^{(1)}$. So specifying that μ is a point mass on $\theta = (p_1, p_2)$ recovers the specified (p, q).

Case c If $p_1 \neq q_1^{(2)}$ and $q_1^{(2)} = 0$, then by Symmetry $q_2^{(1)} = 0$. Then specifying $\mu = p_1\delta_1 + (1-p_1)\delta_2$ recovers the specified (p, q).

Case d If $p_1 = 1$, then by Reinforcement $q_1^{(1)} = 1$. Therefore, specifying that μ is a point mass on $\theta = (1, 0)$ recovers the specified (p, q). □

A.2 Proof of Lemma 2

Proof. *Only if.* Observe that by equations (1) and (3) if $\mu \in \Delta(\Delta(Y))$ is a CIID representation for (p, q) then for all $i, j \in Y$

$$\begin{aligned} m_{ij}(p,q) &= \int_{\Delta(Y)} \theta_j d\mu(\theta) \int_{\Delta(Y)} \theta_i d\mu(\theta|j) \\ &= \int_{\Delta(Y)} \theta_j d\mu(\theta) \int_{\Delta(Y)} \frac{\theta_i\theta_j}{\int_{\Delta(Y)} \theta_j d\mu(\theta)} d\mu(\theta) = \int_{\Delta(Y)} \theta_i\theta_j d\mu(\theta) \end{aligned} \tag{19}$$

if $p_j \neq 0$ and $m_{ij}(p,q) = 0 = \int_{\Delta(Y)} \theta_i\theta_j d\mu(\theta)$ if $p_j = 0$.

If. Suppose that there exists a $\mu \in \Delta(\Delta(Y))$ such that equation (5) is satisfied. Then

$$p_j = \sum_{i=1}^{n} p_j\, q_i^{(j)} = \sum_{i=1}^{n} m_{ij}(p,q) = \sum_{i=1}^{n} \int_{\Delta(Y)} \theta_i\theta_j d\mu(\theta) = \int_{\Delta(Y)} \sum_{i=1}^{n} \theta_i\theta_j d\mu(\theta) = \int_{\Delta(Y)} \theta_j d\mu(\theta).$$

Moreover, if $p_j \neq 0$

$$q_i^{(j)} = \frac{m_{ij}(p,q)}{p_j} = \frac{\int_{\Delta(Y)} \theta_i\theta_j d\mu(\theta)}{\int_{\Delta(Y)} \theta_j d\mu(\theta)} = \int_{\Delta(Y)} \theta_i d\mu(\theta|j),$$

so μ is a CIID model for (p, q) with $\mu(\cdot|j)$ defined as δ_q for those j such that $p_j = 0$.

□

A.3 Proof of Claim 1

Proof. Suppose that M is completely positive, so there exist $\left(\bar{\pi}^{(s)}\right)_{s=1}^{k} \in (\mathbb{R}_+^n)^k$ such that

$$M_{ij} = \sum_{s=1}^{k} \bar{\pi}_i^{(s)} \bar{\pi}_j^{(s)} \qquad \forall i, j \in \{1, \ldots, n\}.$$

For each s, let

$$Z_s := \sum_{l=1}^{n} \bar{\pi}_l^{(s)}, \qquad \hat{\pi}_i^{(s)} := \frac{\bar{\pi}_i^{(s)}}{Z_s}, \qquad \gamma_s := Z_s^2 \in \mathbb{R}_+.$$

Then $\hat{\pi}^{(s)} \in \Delta(Y)$ and

$$M_{ij} = \sum_{s=1}^{k} \bar{\pi}_i^{(s)} \bar{\pi}_j^{(s)} = \sum_{s=1}^{k} Z_s^2 \hat{\pi}_i^{(s)} \hat{\pi}_j^{(s)} = \sum_{s=1}^{k} \gamma_s \hat{\pi}_i^{(s)} \hat{\pi}_j^{(s)}.$$

Moreover,

$$\sum_{i=1}^{n} M_{ij} = \sum_{s=1}^{k} \gamma_s \sum_{i=1}^{n} \hat{\pi}_i^{(s)} \hat{\pi}_j^{(s)} = \sum_{s=1}^{k} \gamma_s \hat{\pi}_j^{(s)},$$

so

$$\sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} = \sum_{s=1}^{k} \gamma_s \sum_{j=1}^{n} \hat{\pi}_j^{(s)} = \sum_{s=1}^{k} \gamma_s = 1.$$

Thus $\sum_{s=1}^{k} \gamma_s = 1$ and $M_{ij} = \sum_{s=1}^{k} \gamma_s \hat{\pi}_i^{(s)} \hat{\pi}_j^{(s)}$ is a simplex completely positive representation, which proves the claim. □

A.4 Proof of Theorem 2

Proof. (2) ⇒ (1). By Lemma 2, there is $\mu \in \Delta(\Delta(Y))$ with $m_{ij}(p, q) = \int_{\Delta(Y)} \theta_i \theta_j d\mu(\theta)$ for every $i, j \in Y$. Since $\Delta(Y)$ is compact and $\pi \mapsto \pi\,\pi^\top$ is continuous,

$$M(p, q) \in \operatorname{conv}\{\pi\,\pi^\top : \pi \in \Delta(Y)\}.$$

By Carathéodory's theorem (see Aliprantis and Border, 2013, Theorem 5.32 with dimension $r = \frac{1}{2}n(n+1)$), we can write

$$M(p,q) = \sum_{s=1}^{r+1} \gamma_s \, \pi^{(s)} \pi^{(s)^\top}, \text{ where all } \gamma_s > 0, \sum_{s=1}^{r+1} \gamma_s = 1, \text{ and all } \pi^{(s)} \in \Delta(Y) \tag{20}$$

establishing complete positivity.

(1) $\Rightarrow$ (3). By Claim 1, $M(p,q)$ is simplex completely positive. Given $M = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)} \pi^{(s)^\top}$, define the discrete measure $\mu(\theta) = \sum_{s=1}^{r} \gamma_s \, \delta_{\pi^{(s)}}(\theta)$. Then

$$\int_{\Delta(Y)} \pi_i \, \pi_j \, \mu(d\pi) = \sum_{s=1}^{r} \gamma_s \, \pi_i^{(s)} \, \pi_j^{(s)} = m_{ij},$$

establishing that $m_{ij}(p,q) = \int_{\Delta(Y)} \theta_i \theta_j d\mu(\theta)$ for every $i, j \in \{1, ..., n\}$.

(3) $\Rightarrow$ (2). Trivial. □

A.5 Proof of Corollary 1

Proof. The first part of the statement follows from Theorem 2, which shows (p,q) has a Bayes-rationalizing conditionally i.i.d. model with finite support. For the second part, suppose by contradiction that (p,q) has a unique CIID model. By the first part, it must have finite support. By Lemma 1 (p,q) satisfies Symmetry and Reinforcement, so the proof of Theorem 1 implies there is a CIID model with a Beta distribution. □

A.6 Proof of Corollary 2

Proof. Theorem 2 shows that (1) is equivalent to $M(p,q)$ being completely positive. That (1) $\Rightarrow$ (2) then follows from the immediate fact that a completely positive matrix is positive semidefinite. That (2) implies a completely positive $M(p,q)$ follows from the fact that when $n \leqslant 4$, every positive semidefinite matrix with non-negative entries is completely positive. (See Diananda [1962] or Theorem 2.4 in Berman and Shaked-Monderer [2003].) □

A.7 Proof of Proposition 1

Proof. A full cycle requires $\mathrm{Cov}_\mu(\theta_{i+1}, \theta_i) > 0$ for all i along the cycle.

Case n=2: From Lemma 1, (p, q) satisfies Reinforcement, which in the binary outcome case immediately rules out $q_i^{(j)} > i$ for $i \neq j$.

Case n=3: From equation (8):

$$\mathrm{Var}(\theta_1) + \mathrm{Var}(\theta_2) + \mathrm{Var}(\theta_3) + 2\,(\mathrm{Cov}(\theta_1, \theta_2) + \mathrm{Cov}(\theta_2, \theta_3) + \mathrm{Cov}(\theta_3, \theta_1)) = 0.$$

Since variances are non-negative, if all three covariances were strictly positive, the left-hand side would be strictly positive, a contradiction. So at least one of the cycle covariances must be non-positive, and a full cycle cannot occur.

Case n=4: The sum of the four edge covariances is

$$\begin{aligned}
&\mathrm{Cov}(\theta_1, \theta_2) + \mathrm{Cov}(\theta_2, \theta_3) + \mathrm{Cov}(\theta_3, \theta_4) + \mathrm{Cov}(\theta_4, \theta_1) \\
&= \mathrm{Cov}(\theta_1 + \theta_3, \theta_2 + \theta_4) = \mathrm{Cov}(\theta_1 + \theta_3, 1 - (\theta_1 + \theta_3)) = -\mathrm{Var}(\theta_1 + \theta_3) \leqslant 0.
\end{aligned}$$

A full 4-cycle is therefore impossible. □

A.8 Proof of Proposition 2

Proof. Fix $n \geqslant 5$. Choose numbers H, L with $0 < L < H < 1$ satisfying

$$2H + (n-2)L \;=\; 1. \tag{21}$$

For $r \in Y$ define $\theta^{(r)} \in \Delta(Y)$ by

$$\theta_k^{(r)} \;=\; \begin{cases} H, & k \in \{r,\ r+1 \bmod n\}, \\ L, & \text{otherwise.} \end{cases}$$

Let μ be the uniform distribution on $\{\theta^{(1)}, \ldots, \theta^{(n)}\}$. By equation (21),

$$p_j = \mathbb{E}_\mu[\theta_j] = \frac{2H + (n-2)L}{n} = \frac{1}{n} \qquad \text{for all } j \in Y,$$

so the period-1 forecast is uniform.

Fix $i \in Y$. Across the n support points, the pair (θ_i, θ_{i+1}) takes values (H, H)

once, (H, L) and (L, H) once each, and (L, L) the remaining $n-3$ times. Hence

$$\mathbb{E}_\mu[\theta_i\theta_{i+1}] = \frac{1}{n}\Big(H^2 + 2HL + (n-3)L^2\Big), \qquad \mathbb{E}_\mu[\theta_i] \;=\; \frac{1}{n}.$$

Therefore, by equation (7),

$$q^{(i)}_{i+1} = \frac{\mathbb{E}_\mu[\theta_i\theta_{i+1}]}{\mathbb{E}_\mu[\theta_i]} = H^2 + 2HL + (n-3)L^2.$$

We need $q^{(i)}_{i+1} > p_{i+1} = 1/n$. Define $f_n(L) := H(L)^2 + 2H(L)L + (n-3)L^2$, where $H(L) := \frac{1-(n-2)L}{2}$. A direct calculation gives

$$f_n(L) = \frac{1}{4} - \frac{n-4}{2}L + \frac{n(n-4)}{4}L^2, \qquad \text{so } f_n(0) = \frac{1}{4}.$$

For $n \geqslant 5$, we have $f_n(0) = 1/4 > 1/n$. By continuity of $f_n(L)$, there exists $\varepsilon > 0$ such that $f_n(L) > 1/n$ for all $L \in (0, \varepsilon)$. Picking such an L and setting H by equation (21) yields $q^{(i)}_{i+1} > 1/n = p_{i+1}$ for every i, which constitutes a full cycle. $\square$

A.9 Proof of Corollary 3

Proof. By assumption there is a positive diagonal matrix $D = \operatorname{diag}(s_1, \ldots, s_n)$ with $s \in \mathbb{R}^n_{++}$ such that

$$(DM(p,q)D)_{ii} \;\geqslant\; \sum_{j\neq i}(DM(p,q)D)_{ij} \qquad \forall i \in \{1, \ldots, n\}.$$

Therefore $A := DM(p,q)D$ is symmetric, nonnegative, and diagonally dominant, hence it is completely positive by Theorem 2.5 in Berman and Shaked-Monderer [2003]. As a consequence, it can be written as $A = \sum_{k=1}^{r} \alpha_k u^{(k)}(u^{(k)})^\top$ for some $\alpha \in \mathbb{R}^k$, $(u^{(k)})_{k=1}^r \in (\mathbb{R}^n)^k$. Conjugating by D^{-1} yields

$$M(p,q) \;=\; D^{-1}AD^{-1} \;=\; \sum_k \alpha_k \big(D^{-1}u^{(k)}\big)\big(D^{-1}u^{(k)}\big)^\top,$$

which is a sum of nonnegative rank-one outer products. Hence $M(p,q)$ is completely positive. With this, (p,q) admits a CIID representation by Theorem 2. $\square$

A.10 Proof of Proposition 3

Proof. By Theorem 2, (p, q) satisfies Reinforcement and $M(p, q)$ is completely positive. Denote as $(\theta(1), ..., \theta(r))$ the elements of the support of μ, and define $\phi(i) = \mu(\theta(i))$. Define the vectors $f_i \in \mathbb{R}^r$ by

$$f_i = \left(\sqrt{\phi(1)}\theta_i(1), ..., \sqrt{\phi(r)}\theta_i(r)\right)^\top.$$

Let $M_{ij} = \langle f_i, f_j \rangle$. By construction,

$$M_{ij} = \sum_{k=1}^{r} (\sqrt{\phi(k)}\theta_i(k))(\sqrt{\phi(k)}\theta_j(k)) = \sum_{k=1}^{r} \phi(k)\,\theta_i(k)\,\theta_j(k).$$

Also, Bayes rule in the CIID model with prior μ gives

$$q_j^{(i)} = \sum_{k=1}^{r} \frac{\phi(k)\,\theta_i(k)}{\sum_{\kappa=1}^{r} \phi(\kappa)\,\theta_i(\kappa)} \theta_j(k) \tag{22}$$

and

$$M(p, q)_{ij} = p_i q_j^{(i)} = \sum_{k=1}^{r} \phi(k)\,\theta_i(k)\,q_j^{(i)} = \sum_{k=1}^{r} \phi(k)\,\theta_i(k)\,\theta_j(k) = M_{ij},$$

where the third equality follows from equation (22).

Therefore, the CP-rank of M is no more than r and by Proposition 3.2 in Berman and Plemmons [1994], rank $M \leqslant$ cpr$M$. □

A.11 Proof of Proposition 4

Proof. Since $M(p, q)$ is completely positive, $M_{ij}(p, q) = \langle f_i, f_j \rangle$ and $f_i = (x(1), ..., x(r))$ with $x_i(j) \geqslant 0$, for all $i \in \{1, ..., n\}$. Moreover, by Theorem 3.5 in Berman and Plemmons [1994], we can pick these f such that $r \leqslant l(l+1)/2 - 1$, and it is without loss of generality to have $\sum_i x_i(k) > 0$ for all $k \in \{1, ..., r\}$.[11] Let

$$S_k = \sum_i x_i(k).$$

[11]To see this, let $\tilde{f}$ be the $r-1$ dimensional vector with entries equal to f except for not having entry k. We have $M_{ij}(p, q) = \langle f_i, f_j \rangle 0 \sum_{l=1}^{r} x_i(l)x_j(l) = \sum_{l \neq k}^{r} x_i(l)x_j(l) = \langle \tilde{f}_i, \tilde{f}_j \rangle$, showing that the zero entry k could be directly omitted to begin with.

We have $\sum_{i,j} M(p,q)_{ij} = \sum_j p_j = 1$. We also have $\sum_{i,j} M(p,q)_{ij} = \sum_{i,j}\sum_{k=1}^r x_i(k)\,x_j(k) = \sum_{k=1}^r (\sum_i x_i(k))(\sum_j x_j(k)) = \sum_{k=1}^r S_k^2$. Thus $\sum_{k=1}^r S_k^2 = 1$.

Define $\phi(k) = S_k^2$. Since $S_k > 0$ and $\sum_{k=1}^r S_k^2 = 1$, we have $\phi(k) \in (0,1)$. Define

$$\theta_i(k) = \frac{x_i(k)}{S_k} \qquad \forall k \in \{1, ..., r\}.$$

Since $\sum_i x_i(k) = S_k$, both $\theta(k)$ are probability vectors in Δ^{n-1}. Since the vectors $(x(1), ..., x(r))$ are linearly independent, $(\theta(1), ..., \theta(r))$ are distinct. We now check that the CIID model with prior μ supported on $(\theta(1), ..., \theta(r))$ and with $\mu(\theta(k)) = \phi(k)$ induces (p,q).

We also have:

$$\begin{aligned} p_j &= \sum_i M_{ij} = \sum_i \sum_{k=1}^r x_i(k)\,x_j(k) = \sum_{k=1}^r x_j(k) \sum_i x_i(k) = \sum_{k=1}^r x_j(k)\,S_k \\ &= \sum_{k=1}^r x_j(k)\sqrt{\phi(k)} = \sum_{k=1}^r \theta_j(k)\sqrt{\phi(k)}\sqrt{\phi(k)} = \sum_{k=1}^r \theta_j(k)\,\phi(k). \end{aligned}$$

Identity (ii) holds by construction:

$$M_{ij}(p,q) = \sum_{k=1}^r x_i(k)\,x_j(k) = \sum_{k=1}^r \sqrt{\phi(k)}\theta_i(k)\,(\sqrt{\phi(k)}\theta_j(k)) = \sum_{k=1}^r \phi(k)\,\theta_i(k)\,\theta_j(k).$$

Therefore,

$$q_j^{(i)} = \frac{M_{ij}(p,q)}{p_i} = \frac{\sum_{k=1}^r \phi(k)\,\theta_i(k)\,\theta_j(k)}{\sum_{k=1}^r \theta_i(k)\,\phi(k)} = \sum_{k=1}^r \mu(\theta(k)\,|i)\,\theta_j(k)$$

proving that (p,q) is represented by the CIID model with prior μ.

□

A.12 Proof of Lemma 3

Proof. We will show that the four stated properties hold. Throughout the proof, we consider an arbitrary count vector ν with $|\nu| \leqslant T-2$ and let $t = |\nu|$.

(1) **Count sufficiency**

Let h_t and h'_t be two histories of the same length t with the same count vector ν. The likelihood of observing h_t given θ is $L(h_t \mid \theta) = \prod_{i=1}^n \theta_i^{\nu_i(h_t)}$ and the likelihood of h_t' is $L(h_t' \mid \theta) = \prod_{i=1}^n \theta_i^{\nu_i(h_t')}$. Since h_t and h_t' have the same count vector, $\nu(h_t) = \nu(h_t')$, and so $L(h_t \mid \theta) = L(h_t' \mid \theta)$ for all θ Hence the posteriors coincide and this implies the forecasts do too. This proves count sufficiency.

(2) **Pairwise exchangeability.**

Using count sufficiency, $q_j^{(\nu+e_i)} = \Pr(Y_{t+2} = j \mid Y_{t+1} = i, \nu)$, where e_i is the unit vector with 1 in component i and 0 elsewhere. Therefore

$$q_i^{(\nu)}\, q_j^{(\nu+e_i)} = \Pr(Y_{t+1} = i \mid \nu)\ \Pr(Y_{t+2} = j \mid Y_{t+1} = i, \nu) = \Pr(Y_{t+1} = i, Y_{t+2} = j \mid \nu).$$

Similarly, $q_j^{(\nu)}\, q_i^{(\nu+e_j)} = \Pr(Y_{t+1} = j, Y_{t+2} = i \mid \nu)$.

Given θ, the sequence after time t is i.i.d. with

$$\Pr(Y_{t+1} = i, Y_{t+2} = j \mid \theta, \nu) = \theta_i\theta_j = \theta_j\theta_i = \Pr(Y_{t+1} = j, Y_{t+2} = i \mid \theta, \nu).$$

Integrating with respect to the posterior yields

$$\Pr(Y_{t+1} = i, Y_{t+2} = j \mid \nu) = \Pr(Y_{t+1} = j, Y_{t+2} = i \mid \nu).$$

Combining with the expressions above gives $q_i^{(\nu)}\, q_j^{(\nu+e_i)} = q_j^{(\nu)}\, q_i^{(\nu+e_j)}$, so pairwise exchangeability holds.

(3) **Reinforcement.**

From count sufficiency

$$q_i^{(\nu+e_i)} = \Pr(Y_{t+2} = i \mid Y_{t+1} = i, \nu) = \frac{\Pr(Y_{t+1} = i, Y_{t+2} = i \mid \nu)}{\Pr(Y_{t+1} = i \mid \nu)}.$$

Conditional on θ and ν, Y_{t+1} and Y_{t+2} are independent and both have distribution θ. Thus $\Pr(Y_{t+1} = i, Y_{t+2} = i \mid \nu) = E[\theta_i^2 \mid \nu]$ and . $\Pr(Y_{t+1} = i \mid \nu) = E[\theta_i \mid \nu]$, so $q_i^{(\nu+e_i)} = \frac{E[\theta_i^2|\nu]}{E[\theta_i|\nu]}$.

We want to show that $q_i^{(\nu+e_i)} \geqslant q_i^{(\nu)}$, that is,

$$\frac{E[\theta_i^2 \mid \nu]}{E[\theta_i \mid \nu]} \geqslant E[\theta_i \mid \nu].$$

Whenever $E[\theta_i \mid \nu] > 0$, this is equivalent to $E[\theta_i^2 \mid \nu] \geqslant (E[\theta_i \mid \nu])^2$. And because $\mathrm{Var}(\theta_i \mid \nu) = E[\theta_i^2 \mid \nu] - (E[\theta_i \mid \nu])^2 \geqslant 0$ this inequality holds. In the case $E[\theta_i \mid \nu] = 0$, we have $q_i^{(\nu)} = 0$, and $E[\theta_i^2 \mid \nu] = 0$ as well, so $q_i^{(\nu+e_i)} = 0$ and the inequality still holds. This proves reinforcement.

(4) **Martingale property.** Conditional on θ and ν,Y_{t+2} is independent of the past and has distribution θ, so $\Pr(Y_{t+2} = i \mid \nu) = E[\theta_i \mid \nu] = q_i^{(\nu)}$.

By the law of total probability,

$$\Pr(Y_{t+2} = i \mid \nu) = \sum_{j=1}^{n} \Pr(Y_{t+2} = i \mid Y_{t+1} = j, \nu) \Pr(Y_{t+1} = j \mid \nu).$$

Using count sufficiency and the definition of q, we get $\Pr(Y_{t+2} = i \mid \nu) = \sum_{j=1}^{n} q_j^{(\nu)} q_i^{(\nu+e_j)}$. Thus $q_i^{(\nu)} = \sum_{j=1}^{n} q_j^{(\nu)} q_i^{(\nu+e_j)}$ or equivalently

$$q^{(\nu)} = \sum_{j=1}^{n} q_j^{(\nu)} q^{(\nu+e_j)},$$

which is the martingale property.

We have shown that count sufficiency, pairwise exchangeability, reinforcement, and the martingale property all follow from the assumption of a CIID representation. This completes the proof of Lemma 3. □

A.13 Uniqueness and Multiplicity

Definition 15 (Hankel Moment Matrix)**.** Given moments $\{m_0, \ldots, m_{2j-2}\}$, the *Hankel matrix* H_j is the $j \times j$ matrix with entries $(H_j)_{rc} = m_{r+c}$ for $r, c \in \{0, \ldots, j-1\}$. A measure μ can generate these moments only if H_j is positive semidefinite (PSD). H_j is positive definite (PD) if and only if the minimal support of μ contains at least j points.

Theorem 5 (Uniqueness with Odd Moments)**.** *Let $T = 2j - 1$ for an integer $j \geqslant 1$. If the associated $j \times j$ Hankel matrix H_j is positive definite, then there exists a unique discrete probability distribution with exactly j support points that generates the moments $\{m_0, \ldots, m_{2j-1}\}$.*

Theorem 6 (Non-Uniqueness with Even Moments)**.** *Let $T = 2k$ for an integer $k \geqslant 1$. If the moments $\{m_0, \ldots, m_{2k}\}$ are such that the Hankel matrix H_{k+1} is positive definite, then there is no rationalizing prior with k or fewer support points.*

Furthermore, there exists a one-parameter family of distinct $(k+1)$-point distributions that all generate these moments.

The proofs rely on the following theorems.

Theorem A (Orthogonal Polynomial Roots)**.** (See Szegő, 1975, Theorem 3.3.1)

Let μ be a positive measure on $[a, b]$ with at least j points in its support. Let $\{P_k(x)\}$ be the sequence of monic orthogonal polynomials with respect to μ. Then the roots of $P_j(x)$ are all real, distinct, and lie in the interior (a, b).

Theorem B (Gaussian Quadrature)**.** (See Szegő, 1975 Theorem 3.4.1)

Let the nodes $\{p_1, \ldots, p_j\}$ be the roots of the j-th orthogonal polynomial $P_j(x)$. Then there exist unique positive weights $\{\lambda_1, \ldots, \lambda_j\}$ such that for any polynomial $f(x)$ of degree at most $2j - 1$:

$$\int f(x)d\mu(x) = \sum_{k=1}^{j} \lambda_k f(p_k)$$

Theorem C (Range of Next Moment)**.** Kreuin, Nudel, et al., 1977 Given a moment sequence $\{m_0, \ldots, m_{2k}\}$ for which H_{k+1} is positive definite, the set of all possible values for the next moment, m_{2k+1}, consistent with a positive measure on $[0, 1]$, forms a non-degenerate closed interval $[m^-_{2k+1}, m^+_{2k+1}]$.

A.13.1 Proof of Uniqueness with Odd Moments

Let $T = 2j - 1$, giving moments $\{m_0, \ldots, m_{2j-1}\}$. Assume H_j is PD.

The PD condition on H_j ensures a well-defined inner product $\langle f, g\rangle = \int f g \, d\mu$. This allows the construction of a unique sequence of monic orthogonal polynomials $\{P_k(x)\}$, where each $P_j(x)$ is uniquely determined by moments $m_0, \ldots, m_{2j-1}$.

By **Theorem A**, $P_j(x)$ has j distinct real roots $\{p_1, \ldots, p_j\}$ in $(0, 1)$. These are our candidate support points.

By **Theorem B**, there exist unique positive weights $\{\lambda_k\}$ corresponding to these nodes such that the integration rule is exact for all polynomials of degree up to $2j - 1$. By choosing the polynomial $f(x) = x^r$ for each $r \in \{0, \ldots, 2j - 1\}$, we get $m_r = \int x^r d\mu(x) = \sum_{k=1}^{j} \lambda_k p_k^r$. This confirms the existence of a j-point distribution matching all $2j$ moments (m_0 to m_{2j-1}).

Uniqueness of the j-point Representation The proof is by contradiction. Assume there exists a second, different j-point distribution with support $\{q_k\}$ and weights $\{w_k\}$ that also generates the moments $m_0, \ldots, m_{2j-1}$.

Construct a monic polynomial $Q(x) = \prod(x - q_k)$. Since the support set is different, $Q(x) \neq P_j(x)$. For any polynomial $R(x)$ of degree less than j, the inner product is $\langle Q, R\rangle = \int QR\, d\mu$. We can compute this using the alternative distribution:

$$\langle Q, R\rangle = \sum_{k=1}^{j} w_k Q(q_k) R(q_k) = \sum_{k=1}^{j} w_k \cdot 0 \cdot R(q_k) = 0$$

This shows that $Q(x)$ is also a monic orthogonal polynomial of degree j.

The sequence of monic orthogonal polynomials is unique. Therefore, we must have $Q(x) = P_j(x)$. This implies their roots are identical, so $\{q_k\} = \{p_k\}$, which contradicts the assumption that the distributions were different. The support points are thus unique. The uniqueness of the weights follows from the unique solution to the invertible Vandermonde system defined by these points.

A.13.2 Proof of Non-Uniqueness with Even Moments

Let $T = 2k$, giving moments $\{m_0, \ldots, m_{2k}\}$. Assume H_{k+1} is PD. This implies the minimal support size must be at least $k+1$. The proof shows that the rationalizing $(k+1)$-point prior is not unique by leveraging the uniqueness result of Theorem 5.

By **Theorem C**, there is a non-degenerate closed interval $[m^-_{2k+1}, m^+_{2k+1}]$ of possible values for the next moment. Choose any two distinct values from the interior of this interval for the next moment:

- Let $m'_{2k+1} \in (m^-_{2k+1}, m^+_{2k+1})$.
- Let $m''_{2k+1} \in (m^-_{2k+1}, m^+_{2k+1})$, with $m'_{2k+1} \neq m''_{2k+1}$.

This allows us to form two different, valid moment sequences of odd length $2k+1$:

- **Sequence A:** $\{m_0, \ldots, m_{2k}, m'_{2k+1}\}$
- **Sequence B:** $\{m_0, \ldots, m_{2k}, m''_{2k+1}\}$

Let $j = k+1$. We now have two distinct moment sequences of length $2j$. We can now apply Theorem 5 to each sequence:

- For Sequence A, there exists a *unique* $(k+1)$-point distribution, μ_A, that generates its moments.

- For Sequence B, there exists a *unique* $(k+1)$-point distribution, μ_B, that generates its moments.

Since Sequence A and Sequence B disagree on the final moment, their unique minimal representations, μ_A and μ_B, must also be different. However, by construction, both μ_A and μ_B generate the same first $2k+1$ moments $\{m_0, \ldots, m_{2k}\}$, which are the moments corresponding to the original observed data. Since there is a continuum of choices for the next moment, there is a continuum of corresponding unique $(k+1)$-point priors, so the rationalizing prior is not unique.

A.14 Proof of Lemma 6

Proof. (i) $\Rightarrow$ (ii) Assume there exists a probability measure μ on $[0,1]$ such that $m_r = \mathbb{E}_\mu[\theta^r]$. For any $r, s \geqslant 0$ with $r+s \leqslant T$, the function $f(\theta) = \theta^r(1-\theta)^s$ is non-negative on $[0,1]$. Therefore, its expectation must be non-negative. A standard identity for the forward difference operator connects the expectation of such a polynomial to the moments $\{m_k\}$:

$$\mathbb{E}_\mu[\theta^r(1-\theta)^s] = \sum_{k=0}^{s}(-1)^k\binom{s}{k}\mathbb{E}_\mu[\theta^{r+k}] = \sum_{k=0}^{s}(-1)^k\binom{s}{k}m_{r+k} = (-1)^s\Delta^s m_r.$$

The expectation is non-negative, so $(-1)^s\Delta^s m_r \geqslant 0$. And Lemma 5 implies that the forecasts satisfy equation 17. This proves statement (ii).

(ii) $\Rightarrow$ (i) This direction is a direct application of Schoenberg [1932]'s solution to the truncated Hausdorff moment problem. Assume the conditions in (ii) hold. Define a linear functional L on the space of polynomials of degree at most T by setting $L(x^k) = m_k$ for $k = 0, \ldots, T$ and extending by linearity. The conditions in (ii) are precisely the requirement that this functional is non-negative on the cone of polynomials that are non-negative on $[0,1]$. A necessary and sufficient condition for the existence of a positive measure μ on $[0,1]$ such that $m_k = \int_0^1 x^k d\mu(x)$ for $k = 0, \ldots, T$ is that the finite sequence of moments satisfies the conditions in (ii). Since $L(1) = m_0 = 1$, the measure μ must be a probability measure. Moreover, since the mixed moments $m_{a,b} := \mathbb{E}\big[\theta^a(1-\theta)^b\big]$, $a+b \leqslant T$ are the expected values of polynomials of degree at most T, their value is completely determined by the moments $(m_i)_{i=0}^T$. Finally, since the forecast ratios were assumed to satisfy equation 17, Bayes' rule applied to this candidate μ reproduces the given conditional forecasts.

Thus μ has a CIID rationalization of (p, q) up to horizon T. $\square$

A.15 Proof of Theorem 4

Proof. We prove the equivalence in two steps.

($\Rightarrow$) Assume (p, q) has a CIID representation. Thus there exists a random $\theta \in \Delta(Y)$ with law μ such that, conditional on θ, the process $(Y_t)_{t \leqslant T}$ is i.i.d. with distribution θ, and the one–step–ahead forecasts induced by this process coincide with (p, q).

Let

$$m_\nu^* := \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta) = \int_{\Delta(Y)} \prod_{i \in Y} \theta_i^{\nu_i} \, d\mu(\theta), \qquad |\nu| \leqslant T.$$

Under the CIID model, Bayes' rule implies that after any history with count vector ν, the posterior on θ has density proportional to θ^ν with respect to μ, so the forecast of $Y_{t+1} = i$ is

$$q_i(\nu) = \mathbb{E}_\mu[\theta_i \mid \nu] = \frac{\int \theta_i \theta^\nu \, d\mu(\theta)}{\int \theta^\nu \, d\mu(\theta)} = \frac{m_{\nu+e_i}^*}{m_\nu^*}, \quad \text{whenever } m_\nu^* > 0.$$

Now compare the recursively defined sequence (m_ν) with (m_ν^*). Both satisfy

$$m_0 = m_0^* = 1, \qquad m_{\nu+e_i} = q_i(\nu) \, m_\nu = \frac{m_{\nu+e_i}^*}{m_\nu^*} \, m_\nu,$$

for all $|\nu| \leqslant T - 1$ and $i \in Y$ (interpreting the recursion trivially on nodes with $m_\nu^* = 0$, where also $m_{\nu+e_i}^* = 0$ and $q_i(\nu)$ is irrelevant). By induction on $|\nu|$ this implies $m_\nu = m_\nu^*$ for every $|\nu| \leqslant T$, that is,

$$m_\nu = \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta) \quad \text{for all } |\nu| \leqslant T.$$

By Lemma B.1, the existence of such a measure μ with $m_\nu = \int \theta^\nu d\mu$ for all $|\nu| \leqslant T$ implies that, for each $k \leqslant T$, the tensor $M^{(k)}$ is simplex completely positive: there exist an integer $r \geqslant 1$, weights $\gamma_1, \ldots, \gamma_r > 0$ with $\sum_s \gamma_s = 1$, and points $\pi^{(1)}, \ldots, \pi^{(r)} \in \Delta(Y)$ such that $M_{i_1 \cdots i_k}^{(k)} = m_{\nu(i_1, \ldots, i_k)} = \sum_{s=1}^r \gamma_s \, \pi_{i_1}^{(s)} \cdots \pi_{i_k}^{(s)}$. Equivalently, $m_\nu = \sum_{s=1}^r \gamma_s \, (\pi^{(s)})^\nu$ for all $|\nu| \leqslant T$, so $\{M^{(k)}\}_{k=1}^T$ is SCP.

($\Leftarrow$) Now assume that the count moment tensors $\{M^{(k)}\}_{k=1}^T$ built from (p, q) as above are simplex completely positive. By construction of m_ν we have $m_0 = 1$, $m_\nu \geqslant 0$, and, for every $|\nu| \leqslant T-1$, $\sum_{i\in Y} m_{\nu+e_i} = \sum_{i\in Y} q_i(\nu)\, m_\nu = m_\nu \sum_{i\in Y} q_i(\nu) = m_\nu$, so the linear consistency identities hold. Thus the truncated array $\{m_\nu\}_{|\nu|\leqslant T}$ satisfies: $m_\nu \geqslant 0, m_0 = 1, \sum_{i\in Y} m_{\nu+e_i} = m_\nu$ for all $|\nu| \leqslant T-1$, and each order-k tensor $M^{(k)}$ is SCP in the sense of the definition.

By Lemma B.1, these conditions are sufficient for the existence of a probability measure μ on $\Delta(Y)$ such that $m_\nu = \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta)$ for all $|\nu| \leqslant T$. Define a process by first drawing $\theta \sim \mu$ and then, conditional on θ, drawing $Y_1, \dots, Y_T$ i.i.d. with distribution θ. Let $\tilde{p}, \tilde{q}$ denote the corresponding one–step–ahead forecasts. As in the first part of the proof, Bayes' rule and the moment representation yield $\tilde{p}_i = \Pr(Y_1 = i) = \int \theta_i \, d\mu(\theta) = m_{e_i}$, and, whenever $m_\nu > 0$, $\tilde{q}_i(\nu) = \mathbb{E}_\mu[\theta_i \mid \nu] = \frac{\int \theta_i \theta^\nu \, d\mu(\theta)}{\int \theta^\nu \, d\mu(\theta)} = \frac{m_{\nu+e_i}}{m_\nu}$.

But by the way we defined m_ν from the original forecast system (p, q) we also have, for all $|\nu| \leqslant T-1$ with $m_\nu > 0$, $q_i(\nu) = \frac{m_{\nu+e_i}}{m_\nu}$. Therefore $\tilde{p}_i = p_i$ and $\tilde{q}_i(\nu) = q_i(\nu)$ at every node with $m_\nu > 0$, and on nodes with $m_\nu = 0$ the values of $q(\nu)$ are irrelevant for the induced law. Hence the CIID process constructed from μ rationalizes the original forecast system (p, q) up to horizon T.

This shows that (p, q) has a CIID representation if and only if the associated count moment tensors $\{M^{(k)}\}_{k=1}^T$ are simplex completely positive, and establishes the stated identities between (p, q) and the moments. □

B For Online Publication

B.1 Proof of Lemma 5

Proof. Let $m_r = \mathbb{E}[\theta^r]$ for $r \geqslant 1$.

Base Case (k=1): The initial forecast for outcome 1 is the prior expectation of θ:

$$p_1 = \mathbb{P}(X_1 = 1) = \mathbb{E}[\mathbb{P}(X_1 = 1 \mid \theta)] = \mathbb{E}[\theta] = m_1.$$

Thus, m_1 is directly identified by the initial forecast.

Inductive Step: Assume that the moments $\{m_1, \ldots, m_k\}$ are uniquely identified for some $k < T$. We will show that m_{k+1} is also uniquely identified.

The observable forecast q^*_{k+1} is the conditional probability of a success at time $k+1$ given k prior successes. By the law of total expectation and the definition of a CIID model:

$$q^*_{k+1} = \mathbb{P}(X_{k+1} = 1 \mid X_1 = \cdots = X_k = 1) = \frac{\mathbb{P}(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 1)}{\mathbb{P}(X_1 = 1, \ldots, X_k = 1)}.$$

The numerator is the unconditional probability of $k+1$ successes. In a CIID model, this is:

$$\mathbb{P}(X_1 = \cdots = X_{k+1} = 1) = \mathbb{E}[\mathbb{P}(X_1 = \cdots = X_{k+1} = 1 \mid \theta)] = \mathbb{E}[\theta^{k+1}] = m_{k+1}.$$

Similarly, the denominator is the unconditional probability of k successes:

$$\mathbb{P}(X_1 = \cdots = X_k = 1) = \mathbb{E}[\mathbb{P}(X_1 = \cdots = X_k = 1 \mid \theta)] = \mathbb{E}[\theta^k] = m_k.$$

Substituting these into the expression for the forecast gives:

$$q^*_{k+1} = \frac{m_{k+1}}{m_k}.$$

By the inductive hypothesis, m_k is known and identified. Since q^*_{k+1} is an observable forecast, we can uniquely identify m_{k+1} as $m_{k+1} = m_k q^*_{k+1}$. By induction, the moments $\{m_1, \ldots, m_T\}$ are uniquely identified by the sequence of forecasts $\{p_1, q^*_2, \ldots, q^*_T\}$. □

B.2 Proof of Lemma 7

Proof. The proof shows each direction of the equivalence.

(1) ⇒ (2) Assume there exists a probability measure μ on $\Delta(Y)$ such that $m_\nu = \mathbb{E}_\mu[\theta^\nu]$ for all $|\nu| \leqslant T$. We must show that the consistency identities hold and that each tensor $M^{(k)}$ is simplex completely positive.

a) Consistency Identities: Let ν be a count vector with $|\nu| \leqslant T - 1$. We examine the sum $\sum_{j \in Y} m_{\nu+e_j}$. Using the assumption and the linearity of expectation:

$$\sum_{j=1}^{n} m_{\nu+e_j} = \sum_{j=1}^{n} \mathbb{E}_\mu[\theta^{\nu+e_j}] = \mathbb{E}_\mu\left[\sum_{j=1}^{n} \theta^{\nu+e_j}\right] = \mathbb{E}_\mu\left[\theta^\nu \sum_{j=1}^{n} \theta_j\right].$$

Since $\theta \in \Delta(Y)$, its components sum to one: $\sum_{j=1}^{n} \theta_j = 1$, so $\sum_{j=1}^{n} m_{\nu+e_j} = \mathbb{E}_\mu[\theta^\nu \cdot 1] = \mathbb{E}_\mu[\theta^\nu] = m_\nu$. This verifies the consistency identities.

b) Simplex Complete Positivity: Fix an integer $k \in \{1, \ldots, T\}$. Let $M^{(k)}$ be the symmetric order-k tensor whose components in the count basis are $\{m_\nu\}_{|\nu|=k}$. By assumption, for any such ν: $m_\nu = \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta)$. Let C_k be the set of all rank-one tensors formed by outer products of vectors from the simplex: $C_k = \{\pi^{\otimes k} \mid \pi \in \Delta(Y)\}$. In the count basis, a tensor in C_k has components $(\pi^{\otimes k})_\nu = \pi^\nu$ for $|\nu| = k$.

The set of simplex completely positive (SCP) tensors of order k is, by definition, the convex hull of C_k, denoted $\operatorname{conv}(C_k)$. The integral representation $m_\nu = \int_{\Delta(Y)} \theta^\nu d\mu(\theta)$ implies that the tensor $M^{(k)}$ lies in the *closed convex hull* of C_k, denoted $\overline{\operatorname{conv}}(C_k)$.

However, the set C_k is the image of the compact set $\Delta(Y)$ under the continuous map $\pi \mapsto \pi^{\otimes k}$. Therefore, C_k is a compact set in the finite-dimensional space of order-k tensors. In a finite-dimensional vector space, the convex hull of a compact set is also compact (a consequence of Carathéodory's theorem). A compact set is always closed, which means $\operatorname{conv}(C_k)$ is a closed set.

Since $\operatorname{conv}(C_k)$ is closed, its closed convex hull is simply itself: $\overline{\operatorname{conv}}(C_k) = \operatorname{conv}(C_k)$. Thus, $M^{(k)}$ must lie in $\operatorname{conv}(C_k)$. By definition, this means $M^{(k)}$ is simplex completely positive. This holds for each $k \in \{1, \ldots, T\}$, completing the proof of (1) ⇒ (2).

(2) ⇒ (1) Assume the consistency identities hold and that for each $k \in \{1, \ldots, T\}$, the tensor $M^{(k)}$ with count moments $\{m_\nu\}_{|\nu|=k}$ is simplex completely positive. We must construct a single probability measure μ that represents all moments $\{m_\nu\}_{|\nu| \leqslant T}$.

a) Representation at Horizon T: Consider the tensor for the highest horizon, $M^{(T)}$. By assumption (2), $M^{(T)}$ is simplex completely positive. By definition of SCP, this means there exists a finite set of points $\{\pi^{(\ell)}\}_{\ell=1}^r \subset \Delta(Y)$ and non-negative weights $\{\gamma_\ell\}_{\ell=1}^r$ with $\sum_{\ell=1}^r \gamma_\ell = 1$ such that for all count vectors ν with $|\nu| = T$:

$$m_\nu = \sum_{\ell=1}^r \gamma_\ell (\pi^{(\ell)})^\nu.$$

This is an expectation with respect to a discrete probability measure μ on $\Delta(Y)$ defined by $\mu = \sum_{\ell=1}^r \gamma_\ell \delta_{\pi^{(\ell)}}$, where $\delta_{\pi^{(\ell)}}$ is the Dirac measure at point $\pi^{(\ell)}$. Thus, we have found a measure μ such that $m_\nu = \mathbb{E}_\mu[\theta^\nu]$ for all $|\nu| = T$.

b) Extending the Representation to Lower Orders: We now show that this *same measure* μ correctly represents the moments for all lower orders, i.e., for all $|\nu| < T$. Define a new set of moments, $\{m'_\nu\}_{|\nu| \leqslant T}$, generated by μ:

$$m'_\nu := \mathbb{E}_\mu[\theta^\nu] = \sum_{\ell=1}^r \gamma_\ell (\pi^{(\ell)})^\nu \quad \text{for all } |\nu| \leqslant T.$$

By construction, we know that $m'_\nu = m_\nu$ for all ν with $|\nu| = T$. Our goal is to show that $m'_\nu = m_\nu$ for all $|\nu| < T$.

c) Using the Consistency Identities: FPick an arbitrary count vector ν with $|\nu| = k < T$. By repeatedly applying the consistency identity $\sum_j m_{\eta+e_j} = m_\eta$, we can express m_ν in terms of moments of order T:

$$\begin{aligned} m_\nu &= \sum_{j_1 \in Y} m_{\nu+e_{j_1}} \\ &= \sum_{j_1 \in Y} \left(\sum_{j_2 \in Y} m_{\nu+e_{j_1}+e_{j_2}} \right) = \sum_{j_1, j_2 \in Y} m_{\nu+e_{j_1}+e_{j_2}} \\ &\vdots \\ &= \sum_{j_1,\ldots,j_{T-k} \in Y} m_{\nu+e_{j_1}+\cdots+e_{j_{T-k}}}. \end{aligned}$$

Let $\eta = \nu + e_{j_1} + \cdots + e_{j_{T-k}}$. The size of this count vector is $|\eta| = |\nu| + (T-k) = k + T - k = T$.

Now do the same for the moments $\{m'_\nu\}$. As shown in the (1) $\Rightarrow$ (2) part of the proof, any set of moments generated by a measure on the simplex automatically satisfies the consistency identities. Therefore:

$$m'_\nu = \sum_{j_1,\ldots,j_{T-k} \in Y} m'_{\nu+e_{j_1}+\cdots+e_{j_{T-k}}}.$$

Let's compare the two expressions. For any multi-index $(j_1, \ldots, j_{T-k})$, let $\eta = \nu + e_{j_1} + \cdots + e_{j_{T-k}}$. Since $|\eta| = T$, we know from step (a) that $m_\eta = m'_\eta$. This means that the sums are equal term-by-term:

$$m_\nu = \sum_{j_1,\ldots,j_{T-k} \in Y} m_{\nu+e_{j_1}+\cdots+e_{j_{T-k}}} = \sum_{j_1,\ldots,j_{T-k} \in Y} m'_{\nu+e_{j_1}+\cdots+e_{j_{T-k}}} = m'_\nu.$$

The equality holds for any ν with $|\nu| = k$. Since we chose k to be any integer less than T (including $k = 0$), this shows that $m_\nu = m'_\nu = \mathbb{E}_\mu[\theta^\nu]$ for all $|\nu| \leqslant T$.

We have constructed a single probability measure μ on $\Delta(Y)$ that represents all the moments $\{m_\nu\}_{|\nu| \leqslant T}$, which completes the proof of (2) $\Rightarrow$ (1). □

Lemma B.1 (Tensor Characterization up to T)**.** *Let $\{m_\nu\}_{|\nu| \leqslant T}$ be a collection of non-negative numbers with $m_{\mathbf{0}} = 1$. The following are equivalent:*

(1) There exists a probability measure μ on the simplex $\Delta(Y)$ such that the moments

are given by the expectation

$$m_\nu = \mathbb{E}_\mu[\theta^\nu] = \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta) \quad \textit{for all } |\nu| \leqslant T.$$

(2) The numbers $\{m_\nu\}$ satisfy the linear consistency identities

$$\sum_{j \in Y} m_{\nu+e_j} = m_\nu \quad \textit{for all } |\nu| \leqslant T-1,$$

and for each $k \in \{1, \ldots, T\}$, the symmetric order-k tensor $M^{(k)}$ defined by the count moments $\{m_\nu\}_{|\nu|=k}$ is simplex completely positive.

Proof. (1) $\Rightarrow$ (2). Assume there exists a probability measure μ on $\Delta(Y)$ such that

$$m_\nu \;=\; \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta) \qquad \text{for all } |\nu| \leqslant T.$$

Linear consistency. Fix ν with $|\nu| \leqslant T-1$. Then

$$\sum_{j \in Y} m_{\nu+e_j} \;=\; \sum_{j \in Y} \int_{\Delta(Y)} \theta^{\nu+e_j} \, d\mu(\theta) \;=\; \int_{\Delta(Y)} \theta^\nu \Big(\sum_{j \in Y} \theta_j \Big) \, d\mu(\theta).$$

Since $\theta \in \Delta(Y)$, we have $\sum_{j \in Y} \theta_j = 1$, so

$$\sum_{j \in Y} m_{\nu+e_j} \;=\; \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta) \;=\; m_\nu,$$

establishing the linear consistency identities.

Simplex complete positivity of $M^{(k)}$. For each $k \in \{1, \ldots, T\}$, define the order–k tensor $M^{(k)}$ by

$$M^{(k)}_{i_1 \cdots i_k} \;:=\; m_{e_{i_1} + \cdots + e_{i_k}} \;=\; \int_{\Delta(Y)} \theta_{i_1} \cdots \theta_{i_k} \, d\mu(\theta) \qquad (i_1, \ldots, i_k \in Y).$$

Equivalently,

$$M^{(k)} \;=\; \int_{\Delta(Y)} \theta^{\otimes k} \, d\mu(\theta).$$

$\{\theta^{\otimes k} : \theta \in \Delta(Y)\}$.

The map $\theta \mapsto \theta^{\otimes k}$ is continuous and $\Delta(Y)$ is compact, so

$$M^{(k)} \in \operatorname{conv}\{\theta^{\otimes k} : \theta \in \Delta(Y)\}.$$

By Carathéodory's theorem there exist points $\pi^{(1)}, \ldots, \pi^{(r)} \in \Delta(Y)$ and weights $\gamma_1, \ldots, \gamma_r > 0$ with $\sum_{s=1}^{r} \gamma_s = 1$ such that

$$M^{(k)} = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)\otimes k}.$$

This is exactly the definition of simplex complete positivity of $M^{(k)}$, so (2) holds.

(2) $\Rightarrow$ (1). Assume now that:

- $m_\nu \geqslant 0$ for all $|\nu| \leqslant T$ and $m_{\mathbf{0}} = 1$;
- for every $|\nu| \leqslant T - 1$, $\sum_{j \in Y} m_{\nu+e_j} = m_\nu$;
- for each $k \in \{1, \ldots, T\}$, the order–k tensor $M^{(k)}$ with entries $M^{(k)}_{i_1 \cdots i_k} = m_{e_{i_1} + \cdots + e_{i_k}}$ is simplex completely positive.

We will show there exists a probability measure μ on $\Delta(Y)$ such that $m_\nu = \int \theta^\nu \, d\mu(\theta)$ for all $|\nu| \leqslant T$.

Step 1: A linear functional on polynomials. Let $\mathcal{A}_T$ be the real vector space of all polynomials in $(\theta_1, \ldots, \theta_n)$ of total degree at most T. Define a linear functional $L : \mathcal{A}_T \to \mathbb{R}$ by

$$L(\theta^\nu) := m_\nu \qquad \text{for all } \nu \text{ with } |\nu| \leqslant T,$$

and extend linearly to arbitrary polynomials $p(\theta) = \sum_{|\nu| \leqslant T} a_\nu \theta^\nu$ by $L(p) := \sum_{|\nu| \leqslant T} a_\nu m_\nu$.

Let

$$S(\theta) := \sum_{i \in Y} \theta_i.$$

Step 2: L treats S as 1. We claim that for every polynomial $p \in \mathcal{A}_T$ with $\deg p \leqslant T - 1$,

$$L(S\, p) = L(p). \tag{23}$$

It suffices to check this on monomials and extend by linearity.

Fix ν with $|\nu| \leqslant T-1$. Then

$$S(\theta)\,\theta^{\nu} = \Big(\sum_{j\in Y}\theta_j\Big)\theta^{\nu} = \sum_{j\in Y}\theta^{\nu+e_j},$$

so

$$L(S\,\theta^{\nu}) = \sum_{j\in Y} L(\theta^{\nu+e_j}) = \sum_{j\in Y} m_{\nu+e_j}.$$

By the linear consistency identities, $\sum_{j\in Y} m_{\nu+e_j} = m_\nu = L(\theta^\nu)$, so $L(S\,\theta^\nu) = L(\theta^\nu)$. By linearity, (23) holds for all p with $\deg p \leqslant T-1$.

Iterating this identity, we obtain

$$L(S^k p) = L(p) \quad \text{whenever } \deg p + k \leqslant T. \tag{24}$$

Step 3: L is nonnegative on polynomials nonnegative on the simplex. Let $K := \Delta(Y)$ and let $p \in \mathcal{A}_T$ satisfy $p(\theta) \geqslant 0$ for all $\theta \in K$. Let $d = \deg p \leqslant T$ and set $k := T - d \geqslant 0$. Define the homogeneous polynomial

$$q(\theta) := S(\theta)^k\, p(\theta).$$

Then $\deg q = d + k = T$, and for every $\theta \in K$ we have $S(\theta) = 1$, so

$$q(\theta) = p(\theta) \geqslant 0 \qquad \forall\, \theta \in K.$$

Write q as

$$q(\theta) = \sum_{|\nu|=T} a_\nu\, \theta^\nu,$$

so that

$$L(q) = \sum_{|\nu|=T} a_\nu\, m_\nu.$$

Let $M^{(T)}$ denote the order–T tensor with entries $M^{(T)}_{i_1\cdots i_T} = m_{e_{i_1}+\cdots+e_{i_T}}$. By simplex complete positivity, there exist $r \in \mathbb{N}$, weights $\gamma_1, \ldots, \gamma_r \geqslant 0$ with $\sum_{s=1}^{r}\gamma_s = 1$, and points $\pi^{(1)}, \ldots, \pi^{(r)} \in \Delta(Y)$ such that

$$M^{(T)} = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)\otimes T},$$

i.e.,

$$M^{(T)}_{i_1 \cdots i_T} = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)}_{i_1} \cdots \pi^{(s)}_{i_T}.$$

For each multi–index ν with $|\nu| = T$, we have

$$m_\nu = M^{(T)}_{i_1 \cdots i_T} \quad \text{whenever} \quad \nu = e_{i_1} + \cdots + e_{i_T}.$$

Thus

$$L(q) = \sum_{|\nu|=T} a_\nu m_\nu = \sum_{|\nu|=T} a_\nu \sum_{s=1}^{r} \gamma_s (\pi^{(s)})^\nu = \sum_{s=1}^{r} \gamma_s \sum_{|\nu|=T} a_\nu (\pi^{(s)})^\nu = \sum_{s=1}^{r} \gamma_s \, q\big(\pi^{(s)}\big).$$

Each $\pi^{(s)}$ lies in $\Delta(Y)$, so $q(\pi^{(s)}) = p(\pi^{(s)}) \geqslant 0$, and hence $L(q) \geqslant 0$.

Using (24) with this p and $k = T - d$, we have

$$L(p) = L(S^k p) = L(q) \ \geqslant \ 0.$$

Therefore

$$L(p) \ \geqslant \ 0 \quad \text{whenever } p \in \mathcal{A}_T \text{ and } p(\theta) \geqslant 0 \ \forall \, \theta \in \Delta(Y). \tag{25}$$

Step 4: m lies in the convex hull of truncated moment vectors. Let $\mathcal{I} := \{\nu \in \mathbb{N}^n : |\nu| \leqslant T\}$, and let $m := (m_\nu)_{\nu \in \mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$. For each $\theta \in \Delta(Y)$, define the truncated moment vector

$$\phi(\theta) := (\theta^\nu)_{\nu \in \mathcal{I}} \in \mathbb{R}^{\mathcal{I}}.$$

Let

$$C := \operatorname{conv}\{\phi(\theta) : \theta \in \Delta(Y)\}$$

be the convex hull of all such vectors. We claim $m \in C$.

Suppose, to the contrary, that $m \notin C$. Since C is a compact convex subset of the finite–dimensional space $\mathbb{R}^{\mathcal{I}}$, the separating hyperplane theorem implies that there exist a nonzero vector $a = (a_\nu)_{\nu\in\mathcal{I}}$ and a scalar α such that

$$\sum_{\nu\in\mathcal{I}} a_\nu m_\nu \;<\; \alpha \quad \text{and} \quad \sum_{\nu\in\mathcal{I}} a_\nu \phi(\theta)_\nu \;\geqslant\; \alpha \quad \forall\, \theta \in \Delta(Y).$$

Define the polynomial

$$p(\theta) := \sum_{\nu\in\mathcal{I}} a_\nu\, \theta^\nu - \alpha.$$

Then $p(\theta) \geqslant 0$ for all $\theta \in \Delta(Y)$ by construction, while

$$L(p) = \sum_{\nu} a_\nu m_\nu - \alpha \;<\; 0.$$

This contradicts (25). Hence $m \in C$.

Step 5: Constructing a representing measure. Since $m \in C$, there exist $\theta^{(1)}, \ldots, \theta^{(r)} \in \Delta(Y)$ and weights $\lambda_1, \ldots, \lambda_r \geqslant 0$ with $\sum_{s=1}^r \lambda_s = 1$ such that

$$m_\nu \;=\; \sum_{s=1}^r \lambda_s\, (\theta^{(s)})^\nu \qquad \forall\, \nu \in \mathcal{I}.$$

Define a probability measure μ on $\Delta(Y)$ by

$$\mu := \sum_{s=1}^r \lambda_s\, \delta_{\theta^{(s)}},$$

where $\delta_{\theta^{(s)}}$ is the Dirac measure at $\theta^{(s)}$. Then, for all $|\nu| \leqslant T$,

$$\int_{\Delta(Y)} \theta^\nu \, d\mu(\theta) = \sum_{s=1}^r \lambda_s\, (\theta^{(s)})^\nu = m_\nu.$$

Thus μ is a probability measure on $\Delta(Y)$ with the required moments, establishing (1).

This completes the proof of the equivalence of (1) and (2). $\square$