Forecasts and Conditionally I.I.D. Models*

Drew Fudenberg[†] Giacomo Lanzani[‡]

February 28, 2026

Abstract

We characterize when one-step-ahead forecasts 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, symmetry (pairwise exchangeability) and reinforcement (realized outcomes become more likely) are necessary and sufficient. For two periods and arbitrary finite outcomes, forecasts admit a CIID representation if and only if a forecast-derived matrix of joint probabilities is completely positive; with at most four outcomes, complete positivity reduces to positive semidefiniteness. Two-period forecasts cannot detect beliefs in positively autocorrelated outcomes, but some negatively autocorrelated beliefs can be identified. 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 minimal-support rationalizations depend on the number of periods. With multiple periods and outcomes, CIID holds exactly when forecasts satisfy pairwise exchangeability and the associated hierarchy of moment tensors is completely positive.

*We thank Ian Ball, Simone Cerreia-Vioglio, Anujit Chakraborty, Ben Golub, Shengwu Li, Matthew Rabin, 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, with observations independent conditional on that state. This conditional independence implies that the induced probability measure over sequences of observations is exchangeable, meaning 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 believe there is 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. A different approach, developed by Shmaya and Yariv [2016], Bohren and Hauser [2024], and Molavi [2025], provides necessary and sufficient conditions under which beliefs in two-period models before and after receiving information can be rationalized as the result of Bayesian updating. This paper lies at the intersection of these approaches: As in work following de Finetti [1937], we characterize the existence of a conditionally i.i.d. model in terms of beliefs over observable events, but we do so using forecasts elicited after outcomes are realized, rather than only ex-ante distributions.

More concretely, we study when one-step-ahead forecasts are consistent with Bayesian learning about a stable but unknown i.i.d. data-generating process. In contrast to much of the belief-revision literature, we do not assume that the analyst can elicit beliefs about the data-generating process itself, but only predictions of the next outcome given the outcomes so far. Also, in contrast to the exchangeability literature, we do not assume that the analyst can elicit beliefs about arbitrary-length (or infinite) sequences. Eliciting beliefs about probability distributions over sequences can be difficult or impossible when these beliefs are not summarized by a fixed state, while predictions about next outcomes are more readily observed through surveys or revealed choices. Moreover, although an agent's initial beliefs may satisfy exchangeability, biases in updating can break this consistency. Our results characterize exactly when such violations can be detected from forecast data alone.

Our main results provide testable characterizations of consistency with conditionally i.i.d. learning. 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 develop two applications of this general characterization. First, we analyze which departures from a belief in an i.i.d. process can be detected using 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 forecast 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 next period (as in the gambler's fallacy), can be

[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.

detected, as can the beliefs of a decision maker who (correctly or not) believes they have observed a garbled version of the outcome.

Second, we characterize when beliefs in a two-player incomplete-information game admit a symmetric common prior over signals, i.e., players are certain that their types are i.i.d. draws from the same distribution and share the same belief about that distribution. This assumption is commonly made in models of common-value auctions; it collapses to the existence of a symmetric common prior if there is no residual incomplete information given the vector of players' signals. We show that this is equivalent to a cross-player consistency condition and the complete positivity of a matrix constructed solely from players' interim beliefs about the opponent type and their ex-ante beliefs about their own type.

When the analyst has elicited one-period-ahead forecasts over a longer time horizon T, the CIID representation imposes additional constraints: longer horizons identify higher-order moments of the latent distribution and impose nontrivial dynamic consistency requirements across periods. 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 minimal number of i.i.d. components needed to rationalize the forecasts 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 use results on the one-dimensional truncated moment problem from Schmüdgen [2020].

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, and Chen, Hansen, and Hansen [2020] shows how to recover a set of investor beliefs from pricing restrictions expressed in terms of conditional moments. Golub and Morris [2020] characterizes the condition for having a (possibly asymmetric) common prior over signals. Sarnoff [2025] highlights that it is more common for forecasts to violate "posterior statistical sufficiency"[2] than exchangeability. Catonini and Lanzani [2026] 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 elicited in the field in many settings, see, e.g., Weber, d'Acunto, Gorodnichenko, and Coibion [2022] for surveys on beliefs about inflation and Dominitz and Manski [2011] and Greenwood and Shleifer [2014] for stock returns.[3] More broadly, this paper is also 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

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 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

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

[3] For example, the Michigan Survey of Consumers has a rolling panel of one-period-ahead forecasts, where each agent is dropped after a small number of them.

probability measure $\mu \in \Delta(\Delta(Y))$.[4]

Definition 1. A forecast (p, q) 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 $j \in Y$ with $p_j > 0$, the posterior measure $\mu(\cdot|j)$ is defined by Bayes' rule: for any Borel set $A \subseteq \Delta(Y)$,

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

and for all $i \in Y$ 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.

In a CIID representation, a probability distribution $\theta \in \Delta(Y)$ represents the unknown but stable data-generating process: outcomes are drawn independently from θ, and μ is the agent's uncertainty about what θ is. Thus, in a CIID representation, 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 $i, j \in Y$, $p_i q_j^{(i)} = p_j q_i^{(j)}$.

This follows from exchangeability: the probability of sequence (i, j) must equal that of (j, i). Note that symmetry also implies the martingale property, i.e., that $p_i = \sum_{j \in Y} p_j q_i^{(j)}$.

The following condition also holds in any conditionally i.i.d. model.

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

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.[5] This distinguishes CIID models from other exchangeable models, such as the sampling without replacement examples discussed in Diaconis [1977], where reinforcement fails because each draw reduces the proportion of that outcome in the remaining population.

CIID models imply reinforcement because any θ that assigns relatively high probability to today's observed outcome will assign relatively high probability to that outcome 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.*

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 (or alternatively because the variance of θ_i under μ is non-negative and equals $\mathbb{E}_\mu[\theta_i^2] - \mathbb{E}_\mu[\theta_i]^2$), 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.

[5]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 $q_2^{(1)} = 5/26 > .13 = p_2$.

3 Binary outcomes

We begin by providing a simple characterization in the case of binary outcomes (i.e., 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. *For binary outcomes ($n = 2$), (p, q) has a CIID representation if and only if (p, q) satisfies symmetry and reinforcement.*

Necessity follows from Lemma 1. We prove sufficiency 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 Beta 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 Beta prior is not the unique CIID representation. Indeed, as we show in Section 4.2, 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 yields simple moment formulas. But the key point here is that *some* CIID representation exists, not that it must be Beta.

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. With more than two outcomes, the scalar second moment that suffices in the binary case is replaced by a full matrix of joint probabilities, and new constraints arise. 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 $n \times n$ *forecast 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 j in period 1 and i in period 2. Note that for any forecast (p,q), the associated matrix $M(p,q)$ automatically satisfies

$$\sum_{i\in Y}\sum_{j\in Y} m_{ij}(p,q) = 1,$$

since $p \in \Delta(Y)$ and each $q^{(j)} \in \Delta(Y)$. This will be important below.

The next lemma says that in a CIID model, this matrix must coincide with the second-moment matrix $\left(\mathbb{E}_\mu[\theta_i\theta_j]\right)_{i,j}$ 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.

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 $q^{(1)} = (0.4, 0.5, 0.1), q^{(2)} = (0.5, 0.4, 0.1)$, and $q^{(3)} = (0.1, 0.1, 0.8)$. 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 CIID representation, even though (p,q) satisfies symmetry and reinforcement.

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 Left makes Center the most likely policy next period, while observing Center makes Left the most likely outcome. Although the example satisfies symmetry and reinforcement, 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." Section A.8 elaborates on when similar cycles are compatible or incompatible with a CIID representation.

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. 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 resembles the diagonalization of a symmetric matrix, with the key difference that the b vectors are not orthogonal. Complete positivity implies that $M(p, q)$ is positive semidefinite and therefore symmetric. Since symmetry of $M(p, q)$ is equivalent to symmetry of the forecast, complete positivity implies symmetry. We will momentarily see that complete positivity also implies reinforcement. There is an extensive literature on completely positive matrices, see, e.g., Berman and Shaked-Monderer [2003].[6]

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

Theorem 2. *For a forecast (p, q), the following are equivalent:*

1. *$M(p, q)$ is completely positive.*
2. *(p, q) has a CIID representation.*
3. *(p, q) has a CIID representation that has finite support.*

[6] 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.

To prove the result, we introduce a strengthening of complete positivity called simplex complete positivity, and show that it is equivalent to complete 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 $M(p,q)$ is simplex completely positive so that $M(p,q) = \sum_{s=1}^{r} \gamma_s \pi^{(s)} {\pi^{(s)}}^{\top}$ for some $\pi^{(1)}, \ldots, \pi^{(r)} \in \Delta(Y)$ and weights $\gamma_s > 0$ with $\sum_{s=1}^{r} \gamma_s = 1$. Define the discrete measure $\mu := \sum_{s=1}^{r} \gamma_s \delta_{\pi^{(s)}}$, where δ is the Dirac measure. By construction, this measure has second moments $\int_{\Delta(Y)} \theta_i \theta_j \, d\mu = m_{ij}(p,q)$, 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.

Multiplicity The next corollary uses the characterization above to establish the multiplicity of a CIID representation in various cases. Towards this, we say that a forecast (p,q) is dogmatic if $p = q^i$ for all $i \in Y$. Dogmatic forecasts clearly admit a unique CIID representation with Dirac measure on p.

Corollary 1. *Let p be strictly positive. If there is a CIID representation for (p,q) with infinite support, then the Bayes-rationalizing model is not unique. In particular, with binary outcomes, any nondogmatic forecast with a CIID representation has at* *least two of them.*

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. *When $n \leqslant 4$, the following are equivalent:*

1. *(p,q) has a CIID representation.*

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 distribution of θ, because if forecast (p, q) admits a CIID representation μ, then

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

Thus, observing outcome i makes outcome j strictly more likely if and only if $\mathrm{Cov}_\mu(\theta_j, \theta_i) > 0$.[7] The simplex constraint $\sum_{k \in Y} \theta_k = 1$ implies

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

Since variances are nonnegative, this identity implies that the sum of all pairwise covariances must be nonpositive. Intuitively, there is a fixed "covariance budget": any positive covariances must be offset by sufficiently negative covariances elsewhere.

Definition 5. A forecast (p, q) exhibits a *full cycle* if there is an ordering of outcomes such that $q_{i+1}^{(i)} \geqslant p_{i+1}$ for all $i \in Y$, where indices are taken modulo n, and the inequality is strict for at least one i.

A full cycle requires n non-negative covariances arranged along a cyclic permutation of all outcomes with at least one covariance strictly positive. For small n, this exhausts too much of the covariance budget.

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

The impossibility of full cycles for small n has an intuitive explanation: the combination of the simplex and PSD constraints forces too many negative covariances for a full cycle when $n \leqslant 4$.

Remark (Short cycles)**.** Proposition 1 rules out full cycles when $n \leqslant 4$, i.e., configurations in which there exists an ordering of all outcomes such that $q_{i+1}^{(i)} \geqslant p_{i+1}$ for every i (indices modulo n) with some strict inequality. It does not rule out shorter cycles involving only a proper subset of outcomes: the simplex constraint $\sum_{i \in Y} \theta_i = 1$ forces only the sum of all pairwise covariances to be nonpositive, so the positive co-

[7] So observing an outcome makes all the others less likely if and only if every covariance is negative.

variances required by a shorter cycle can be offset by sufficiently negative covariances involving the remaining outcomes.

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

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

The proof constructs an explicit finite-support CIID prior that generates a full cycle.

4.4 Scaled Diagonal Dominance

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

Definition 6. Forecast (p, q) satisfies *scaled diagonal dominance* if there exist weights $s \in \mathbb{R}^n_{++}$ with $\sum_{i \in Y} s_i = 1$ such that

$$p_i\, q_i^{(i)}\, s_i \;\geqslant \sum_{j \in Y \setminus \{i\}} p_j\, q_i^{(j)}\, s_j \qquad \forall i \in Y.$$

The next corollary shows that scaled diagonal dominance is sufficient for a CIID representation, and provides an easily computed criterion under which this condition holds. Let $M(p, q)$ satisfy $m_{ii}(p, q) > 0$ for all $i \in Y$, and define the *row–ratio* matrix $R(p, q)$ by

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

Corollary 3 (Sufficient conditions for CIID rationalizability)**.** *Let (p, q) be a forecast that satisfies symmetry.*

[8] 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 $\mathbb{R}^n$ for $n \geqslant 5$ is qualitatively different (cf. Berman and Shaked-Monderer, 2003).

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

(ii) If $p_i q_i^{(i)} > 0$ for all $i \in Y$ and the spectral radius of $R(p,q)$ is less than 1, then (p,q) satisfies scaled diagonal dominance and hence admits a CIID representation.

Remark. Row diagonal dominance of $M(p,q)$ requires $\sum_{j \in Y} R_{ij}(p,q) \leqslant 1$ for every row, i.e., $\|R\|_\infty \leqslant 1$. Since R is nonnegative, its spectral radius is no greater than $|R\|_\infty$, so the spectral condition is strictly weaker. See Example 6 in the Online Appendix for an example of this.

5 Non-Uniqueness of CIID Representations

When a forecast has a CIID representation, it need not be unique. Moreover, uniqueness can fail in two ways: there may be CIID representations with different support sizes, and there may be multiple CIID representations with the same support size. This follows from the fact that CIID forecasts are determined by the first two moments of the distribution of θ, and many distributions can match those 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 non-dogmatic CIID representation μ. Now let $0 \leqslant \theta_L < \theta_H \leqslant 1$, fix $\lambda \in (0,1)$ and let μ_λ be the two-point prior placing mass λ on θ_L and $1-\lambda$ on θ_H.[9] Let $h := \theta_H - \theta_L > 0$, then $\mathbb{E}_{\mu_\lambda}[\theta_1] = \mathbb{E}_\mu[\theta_1]$ is satisfied if and only if $\theta_L = p_1 - (1-\lambda)h$ and $\theta_H = p_1 + \lambda h$. Moreover, $\mathbb{E}_{\mu_\lambda}[\theta_1^2] = \mathbb{E}_\mu[\theta_1^2]$ is satisfied if and only if $h^2 = \dfrac{\mathbb{E}_\mu[\theta_1^2] - (\mathbb{E}_\mu[\theta_1])^2}{\lambda(1-\lambda)}$, so the

[9] Then $p_1 = \mathbb{E}_\mu[\theta_1]$, $q_1^{(1)} = \mathbb{E}_\mu[\theta_1^2]/\mathbb{E}_\mu[\theta_1]$, $q_1^{(2)} = \mathbb{E}_\mu[\theta_1(1-\theta_1)]/\mathbb{E}_\mu[1-\theta_1]$ and $\mathbb{E}_{\mu_\lambda}[\theta_1] = \lambda\theta_L + (1-\lambda)\theta_H, \mathbb{E}_{\mu_\lambda}[\theta_1^2] = \lambda\theta_L^2 + (1-\lambda)\theta_H^2$.

two-point priors that match $(\mathbb{E}_\mu[\theta_1], \mathbb{E}_\mu[\theta_1^2])$ are

$$\theta_L = \mathbb{E}_\mu[\theta_1] - (1-\lambda)\sqrt{\frac{\mathbb{E}_\mu[\theta_1^2] - (\mathbb{E}_\mu[\theta_1])^2}{\lambda(1-\lambda)}}, \qquad \theta_H = \mathbb{E}_\mu[\theta_1] + \lambda\sqrt{\frac{\mathbb{E}_\mu[\theta_1^2] - (\mathbb{E}_\mu[\theta_1])^2}{\lambda(1-\lambda)}}$$

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

Although there is a 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 $\mathbb{E}_{\mu_\lambda}[\theta^3] = \lambda\theta_L^3 + (1-\lambda)\theta_H^3$, 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 an $r \geqslant 2$ point support then* $\operatorname{rank} M(p,q) \leqslant r$.

Proposition 3 is proved by establishing that if (p,q) has a CIID representation with support r, then $\operatorname{cpr}(M(p,q)) \leqslant r$. 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 $\operatorname{cpr} \leqslant \operatorname{rank}(\operatorname{rank}+1)/2 - 1$.

Proposition 4. *Let forecast (p,q) be such that $M(p,q)$ is completely positive. If M has rank l, then (p,q) admits a CIID representation where μ has at most* $\max\{l(l+1)/2-1, 1\}$ *point support.*

Propositions 3 and 4 immediately imply this corollary.

Corollary 4. *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. *$M(p,q)$ is completely positive and has rank* 2.

Example 3 (A CIID model with support > rank M)**.** *Suppose* $Y = \{1, 2\}$*, and that forecast* (p, q) *admits a CIID representation* μ *with a three point support:* $\mu(\theta_1 = \frac{1}{3}) = \frac{1}{4}, \mu(\theta_1 = \frac{1}{2}) = \frac{1}{2}, \mu(\theta_1 = \frac{2}{3}) = \frac{1}{4}$*, so that* $p_1 = \mathbb{E}_\mu[\theta_1] = \frac{1}{2}$*. Then* $\mathbb{E}_\mu[\theta_1^2] = \frac{19}{72}$*,* $\mathbb{E}_\mu[\theta_1(1-\theta_1)] = 17/72$*,* $q_1^{(2)} = \frac{\mathbb{E}[\theta(1-\theta)]}{\mathbb{E}[1-\theta]} = \frac{17}{36}$*, and* $q_1^{(1)} = \frac{19}{36}$*. Thus the implied second-period conditional forecasts satisfy reinforcement, and*

$$M(p,q) = \begin{pmatrix} p_1 q_1^{(1)} & (1-p_1)q_1^{(2)} \\ p_1(1-q_1^{(1)}) & (1-p_1)(1-q_1^{(2)}) \end{pmatrix} = \begin{pmatrix} \frac{19}{72} & \frac{17}{72} \\ \frac{17}{72} & \frac{19}{72} \end{pmatrix}$$

is symmetric, entrywise nonnegative, and has $\operatorname{rank} M(p,q) = 2 < \operatorname{supp} \mu$*.*

6 Applications

6.1 Belief in Persistence or Reversal

We next apply our 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. Forecast (p, q) has a *persistent Bayesian representation* if there exists an $\alpha \in (0, 1)$ such that $(p, \hat{q})$ has a CIID representation where

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

and 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. Forecast (p, q) has a *reversing Bayesian representation* if there exists an $\alpha \in (0, 1)$ such that $(p, \hat{q})$ has a CIID representation where

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

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.[10] Note also 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.

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

Thus, in two periods, belief in persistence (positive autocorrelation) 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 $\hat{q}_1^{(1)} < p_1$, 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.[11]

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.*

We conclude by considering a perturbation that is detectable from forecasts:

[10]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.

[11]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.

belief in a CIID model with garbled observations. 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 $p_i q_j^{(\tilde{i})} \neq p_j q_i^{(\tilde{j})}$ where the tilde denotes the garbled realization of the outcome.[12]

6.2 Existence of a Symmetric Common Prior over Signals

We now show how our results can be used to verify the existence of a symmetric common prior from the players' interim beliefs and their ex-ante beliefs about their own type. Formally, consider a game of incomplete information with two players and state space $\Omega \times S_1 \times S_2$ with $S_1 = S_2 = S$ finite and $|S| = n$. Here, $\omega \in \Omega$ is an underlying fundamental in the game, and s_i corresponds to player i type, as in Dekel, Fudenberg, and Morris [2007].

A *symmetric common prior over signals* is a distribution $\pi \in \Delta(\Omega \times S_1 \times S_2)$ such that

$$\sum_{\omega \in \Omega} \pi(\omega, s_1, s_2) = \int_{\Delta(S)} \theta(s_1)\theta(s_2)\, d\mu(\theta) \quad \text{for some } \mu \in \Delta(\Delta(S)).$$

With this belief structure, players are certain that their types are i.i.d. draws from the same distribution and share the same belief about that distribution. One example is the case in which the players' signals jointly and completely reveal the state: if knowing (s_1, s_2) determines ω with certainty, then a symmetric common prior over signals pins down all uncertainty about ω as well, so a symmetric common prior over signals coincides with a symmetric common prior over the entire state space $\Omega \times S_1 \times S_2$. Another example arises in common-value auctions,[13] where ω denotes the value of the good and agents' signals are i.i.d. conditional on ω, with distribution $\hat{\pi}(\cdot \mid \omega)$. In this case, a common prior over signals exists and is given by

[12] For a concrete example, suppose the agent believes that the true outcome is 1 with probability 1/5 or 3/4 with equal probability, and that the garbling matrix is $G = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$ for some $\varepsilon > 0$. The agent's period-2 forecast is conditioned on the garbled signal rather than the true outcome, and a straightforward computation shows that symmetry fails for any $\varepsilon > 0$.

[13] Although this structure is typical in common-value auctions, the only thing that is needed is that the value of the good for agent i can be expressed as a function f_i of θ, with the two players possibly having different f_i.

$\mu(C) = \pi\big(\{\omega : \hat{\pi}(\cdot \mid \omega) \in C\}\big)$ for all $C \in \Delta(S)$.

Let $\pi_i^{-i}(\cdot|s_i) \in \Delta(S_{-i})$ denote the interim belief of a player i of type s_i about the opponent type. Also, let $\pi_i^{\text{eao}} \in \Delta(S_i)$ denote the ex-ante belief about their own private signal. Our results help us link the existence of a symmetric common prior to the matrices $M_i(\pi) \in [0,1]^{S_{-i} \times S_i}$, $i \in \{1,2\}$ defined by:

$$m_i(\pi)_{s_{-i},s_i} = \pi_i^{\text{eao}}(s_i)\pi_i^{-i}(s_{-i}|s_i).$$

Corollary 5. *Given π_i^{-i} and π_i^{eao} there exists a symmetric common prior over signals consistent with those beliefs if and only if $M_1(\pi) = M_2(\pi)$ and is completely positive.*

The result follows from the fact that, by Theorem 2, the complete positivity of $M_i(\pi)$ is equivalent to the existence of a prior over signals μ_i that is CIID. Then the equality $M_1(\pi) = M_2(\pi)$ guarantees that this prior is common to both players.

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, because longer horizons identify higher-order moments of the latent distribution, and imposes tighter necessary conditions. This means fewer forecasts will have CIID representations, and those that do will have fewer.

In the binary case, observing one-step-ahead forecasts up to horizon T identifies the first T moments of the latent Bernoulli parameter θ. We use this to connect to the *Hausdorff truncated moment problem*. 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 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 completely positive and satisfy pairwise exchangeability and reinforcement; 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 $M(p, q)$. When forecasts are observed over longer histories this invariance breaks down, as the finite-horizon CIID conditions of Theorem 4 converge to exchangeability: a forecast system satisfying them for every finite horizon generates an exchangeable measure on infinite sequences, recovering the classical characterizations of de Finetti [1937] and Hewitt and Savage [1955]. Additional periods therefore 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 $T > 2$

As in the two-period case, we begin with some intuitive necessary conditions that any CIID representation must satisfy. Fix a horizon $T > 2$. For $t \in \{0, ..., T-1\}$, a history of length t is $h_t = (y_1, \ldots, y_t) \in Y^t$, with $\varnothing$ denoting the empty history $(t = 0)$. For any history $h_t \in Y^t$ with $0 \leqslant t < T$, let $q^{(h_t)} \in \Delta(Y)$ denote the period-t forecast of the period-$(t+1)$ outcome, and define $\mathbf{q} = \{q^{(h_t)}\}_{t \in \{0,...,T-1\}, h_t \in Y^t}$

Definition 9. Forecast $\mathbf{q}$ has a *CIID representation* if there exists a probability measure $\mu \in \Delta(\Delta(Y))$ such that:

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

$$q_i^{(\varnothing)} = \int_{\Delta(Y)} \theta_i \, d\mu(\theta). \tag{11}$$

(ii) For all $t \in \{1, ..., T-1\}, h_t = (y_1, ..., y_t) \in Y^t$ such that $q_{y_\tau}^{(\varnothing)} > 0$ for all $\tau \in \{1, ..., t\}$, the posterior measure $\mu(\cdot | h_t)$ is defined by Bayes' rule: for any Borel set $A \subseteq \Delta(Y)$,

$$\mu(A|h_t) = \frac{\int_{\theta \in A} \prod_{s=1}^{t} \theta_{y_s} \, d\mu(\theta)}{\int_{\theta \in \Delta(Y)} \prod_{s=1}^{t} \theta_{y_s} \, d\mu(\theta)} \tag{12}$$

and for all $i \in Y$ the conditional forecast satisfies

$$q_i^{(h_t)} = \int_{\Delta(Y)} \theta_i \, d\mu(\theta|h_t). \tag{13}$$

(iii) For all $t \in \{1, ..., T-1\}$, $h_t = (y_1, ..., y_t) \in Y^t$ such that $q_{y_\tau}^{(\varnothing)} = 0$ for some $\tau \in \{1, ..., t\}$, the value of $q^{(h_t)}$ is unrestricted.

Point (iii) leaves the belief after zero probability events completely unrestricted.[14]

Perhaps the most basic necessary condition for CIID forecasts is that it must depend only on the count; this was vacuously satisfied in the two-period setting. Define ν by $\nu_i(h_t) = \sum_{s=1}^{t} \mathbf{1}\{y_s = i\}, i \in Y$, for all $t \in \{0, ..., T-1\}$ and $h_t \in Y^t$. Note that when outcome i is observed $\nu((h_t, i)) = \nu(h_t) + e_i$, where e_i is the i-th unit vector. Let $\mathcal{N}(\mathbf{q})$ denote the set of count vectors ν with $\|\nu\|_1 \leqslant T-1$ such that $\mathbf{q}_i^{(\varnothing)} > 0$ for every $i \in Y$ with $\nu_i > 0$.

Definition 10 (Count sufficiency)**.** Forecast $\mathbf{q}$ satisfies *count sufficiency* if for any two histories h_t, h_t' with $\nu(h_t) = \nu(h_t') \in \mathcal{N}(\mathbf{q})$,

$$q^{(h_t)} = q^{(h_t')} =: q^{(\nu)}. \tag{14}$$

We will assume throughout that count sufficiency is satisfied, and use it to simplify the statements of other conditions. We also restrict attention to histories with ex-ante positive probability.

For $T > 2$, count sufficiency breaks the observational equivalence between persistence and CIID that holds when $T = 2$. Indeed, except in degenerate cases, an agent with a persistent Bayesian representation who observes two histories h_t, h_t' with $\nu(h_t) = \nu(h_t')$ but where outcome i is the most recent realization in h_t but not in h_t' will forecast $q_i^{(h_t)} > q_i^{(h_t')}$, violating count sufficiency and hence incompatible with a

[14]One could alternatively require that beliefs are updated using Bayes rule whenever possible, as in perfect extended-Bayesian equilibrium (Fudenberg and Tirole, 1991); the results would be essentially unchanged but slightly more cumbersome to state.

CIID representation.

In addition, CIID representations imply generalizations of the symmetry and reinforcement conditions. The next condition generalizes symmetry.

Definition 11 (Pairwise Exchangeability)**.** Forecast $\mathbf{q}$ satisfies *pairwise exchangeability* if for every $\nu \in \mathcal{N}(\mathbf{q})$ and $i, j \in Y$ such that $\nu + e_i, \nu + e_j \in \mathcal{N}(\mathbf{q})$,

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

Because any permutation of $\{1, \ldots, T\}$ can be written as a composition of adjacent transpositions, under count invariance pairwise exchangeability up to period T implies invariance under all permutations of length T, which is the definition of exchangeability.

Definition 12 (Reinforcement)**.** Forecast $\mathbf{q}$ satisfies *reinforcement* if for every $\nu \in \mathcal{N}(\mathbf{q})$ and $i \in Y$ such that $\nu + e_i \in \mathcal{N}(\mathbf{q})$,

$$q_i^{(\nu+e_i)} \; \geqslant \; q_i^{(\nu)}. \tag{15}$$

Lemma 3. *If forecast* $\mathbf{q}$ *has a CIID representation, it satisfies pairwise exchangeability and reinforcement.*

Definition 13 (Martingale property)**.** Forecast $\mathbf{q}$ satisfies the *martingale property* if for every $\nu \in \mathcal{N}(\mathbf{q})$ such that $\nu + e_j \in \mathcal{N}(\mathbf{q})$ for all $j \in Y$,

$$q^{(\nu)} \; = \; \sum_{j \in Y} q_j^{(\nu)} \, q^{(\nu+e_j)}. \tag{16}$$

That is, the agent's current one–step–ahead forecast equals the expectation of their next-period forecast, where expectations are taken using the agent's own current beliefs.

Lemma 4. *If forecast* $\mathbf{q}$ *satisfies pairwise exchangeability then it satisfies the martingale property.*

Note that, despite the probabilistic language, this lemma is not a statement about a probability distribution per se but only about the reported forecasts.

Moments and Forecasts As in the case $T = 2$, the sufficient conditions for CIID developed below apply to the moments associated with the forecasts.

Definition 14 (Implied Moments)**.** The *implied moments* $\{m_\nu\}_{||\nu||_1 \leqslant T}$ of forecast $\mathbf{q}$ are $m_{\mathbf{0}} = 1, m_{\nu+e_i} = q_i^{(\nu)} m_\nu$ if $\nu \in \mathcal{N}(\mathbf{q})\backslash\mathbf{0}$ and $m_\nu = 0$ otherwise.

With an arbitrary forecast, reversing the order of two outcomes might yield different values for the same moment, but under pairwise exchangeability this cannot occur.

Lemma 5. *1. If forecast* $\mathbf{q}$ *is pairwise exchangeable, its implied moments are well-defined:* m_ν *does not depend on the order in which outcomes occur to reach* ν*.*

2. If there is a measure $\mu \in \Delta(\Delta(Y))$ *that generates the implied moments of a forecast sequence, the forecasts are consistent with the conditional probabilities generated by* μ*.*

7.2 Binary Outcomes and More than Two Periods

Let $Y = \{1, 2\}$. In this case, if there is a CIID representation we can work with a one-dimensional latent state θ, and count vectors take the form $(j, t-j)$ where ν has j 1's and $t - j$ 2's. In this case there is an easily-checked characterization of when such a measure exists. To state the sufficient condition, define the *Hankel matrices* $\underline{H}$ and $\overline{H}$ as follows: If T is even,

$$\underline{h}_{i,j} = m_{(i+j,0)} \qquad \forall i, j \in \{0, ..., T/2\},$$

$$\overline{h}_{i,j} = -\Delta m_{(i+j+1,0)} \qquad \forall i, j \in \{0, ..., T/2 - 1\}.$$

If T is odd,

$$\underline{h}_{i,j} = m_{(i+j+1,0)} \qquad \forall i, j \in \{0, ..., (T-1)/2\}$$

$$\overline{h}_{i,j} = -\Delta m_{(i+j,0)} \qquad \forall i, j \in \{0, ..., (T-1)/2\},$$

where $\Delta m_{(r,0)} := m_{(r+1,0)} - m_{(r,0)}$.

Theorem A (Schmüdgen, 2017 Theorems 10.1 and 10.2)**.** Let $n = 2$ and let $(l_0, ..., l_T) \in [0, 1]^{T+1}$. There is a probability measure μ on $[0, 1]$ such that $\mathbb{E}_\mu[\theta^k] = l_k$

for all $k \in \{0, ..., T\}$ if and only if the the associated Hankel matrices $\underline{H}$ and $\overline{H}$ are positive semidefinite.

When applied to the implied moments of a forecast system, Theorem A only constrains the behavior of the "pure" moments that arise when only outcome 1 is observed. The next example shows why this is not enough.

Example 4. *Consider the following forecast system with horizon $T = 3$.*

$$\begin{aligned} q_1^{\nu=(0,0)} &= \frac{1}{2} \\ q_1^{\nu=(1,0)} &= \frac{5}{8}, q_1^{\nu=(0,1)} = \frac{3}{8} \\ q_1^{\nu=(2,0)} &= \frac{7}{10}, q_1^{\nu=(0,2)} = \frac{3}{10}, q_1^{\nu=(1,1)} = \frac{5}{8}. \end{aligned}$$

This forecast satisfies reinforcement and pairwise exchangeability. Its pure moments are consistent with the binary support CIID representation $\mu\,(3/4, 1/4) = \mu(1/4, 3/4) = \frac{1}{2}$, so by Theorem A, the associated Hankel matrices are PSD. However, the mixed moment $m^{(2,1)}$ implied by the forecast is obtained by

$$m_{(1,1)} = m_{(1,0)} q_2^{\nu=(1,0)} = \frac{1}{2}\frac{3}{8} = \frac{3}{16}$$

so that

$$m_{(2,1)} = m_{(1,1)} q_1^{\nu=(1,1)} = \frac{3}{16}\frac{5}{8} = \frac{15}{128}.$$

While the one derived by μ has

$$m_{(2,1)} = \mathbb{E}_\mu\left[\theta_1\theta_2\theta_1\right] = \frac{1}{2}\left[\frac{3}{4}\frac{3}{4}\frac{1}{4} + \frac{1}{4}\frac{1}{4}\frac{3}{4}\right] = \frac{3}{32}.$$

The next condition constrains forecasts so that the implied moments at mixed counts align with those at extreme counts. It is motivated by the fact that if the moments $\{m_{r,0}\}_{r=0}^T$ arise from some distribution of θ, then for every pair a, b with $a + b \leqslant T$ there is a uniquely determined mixed moment $m_{a,b} := \mathbb{E}\big[\theta^a(1-\theta)^b\big]$, since $\theta^a(1-\theta)^b$ is a polynomial of degree $a + b$ in θ.

Definition 15 (Polynomial consistency)**.** Forecast $\mathbf{q} \in \mathcal{N}$ satisfies *polynomial con-*

sistency if

$$m_{a,b} = \sum_{i=0}^{b} \binom{b}{i} (-1)^{b-i} m_{i+a,0} \quad \forall a, b \geqslant 0, a + b \leqslant T.$$

The characterization of CIID for binary outcomes and arbitrary horizon then follows by combining Lemma 5 (which shows when the forecasts are pairwise exchangeable, that a measure that matches all of the implied moments matches the forecasts), Theorem A, which shows there is a measure that matches the pure moments, and the polynomial consistency condition.

Theorem 3. *Let $n = 2$ and $T \geqslant 2$. Forecast $\mathbf{q}$ has a CIID representation if and only if it is pairwise exchangeable, satisfies polynomial consistency, and its Hankel matrices are PSD.*

7.2.1 Parity

A striking feature is a parity effect: for forecasts whose Hankel matrices are not only PSD but positive definite (PD), an odd number of forecast periods uniquely determines the minimal-support prior, whereas an even number yields a continuum of minimal-support priors.[15]

Proposition 7 (Parity effect for binary CIID representations)**.** *Suppose that $n = 2$, $T \geqslant 2$, and that forecast $\mathbf{q}$ has a CIID representation and positive definite Hankel matrices.*

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 smaller than or equal to 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.15. The proposition implies a parity effect: When $T = 2k - 1$, the truncated moment sequence determines a unique

[15] To gain intuition for the PD condition, consider the case $T = 2$. Then $\underline{H} = \begin{pmatrix} 1 & m_{(1,0)} \\ m_{(1,0)} & m_{(2,0)} \end{pmatrix}$, which is PD if and only if $m_{(2,0)} > m_{(1,0)}^2$, i.e., if the variance of the underlying θ is strictly positive.

minimal-support prior with k atoms; while when $T = 2k$, minimal-support representations have $k+1$ atoms and form a one-parameter family. For example, $T = 3$ yields a unique two-point prior, whereas $T = 4$ admits a continuum of distinct three-point priors generating the same forecasts. Any forecast generated by a Beta prior with both parameters strictly positive has positive definite Hankel matrices, so the parity effect applies: three observed periods yield a unique two-point rationalization, while four observed periods yield a one-parameter family of three-point rationalizations.

This distinction reflects the geometry of the truncated moment problem: a k-point distribution on $[0, 1]$ has $2k - 1$ free parameters (support locations and probabilities). Thus $2k - 1$ moments generically identify it uniquely, while $2k$ moments leave one remaining degree of freedom.

7.3 Many Outcomes and Many Periods

7.3.1 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.

The forecasts provided by an agent are the one-step-ahead conditional probabilities. If these forecasts depend only on the outcome counts, as required by a CIID model, we can express them directly in terms of the count moments. The key object for the multi-period characterization is the collection of order-k moment tensors, which generalize the second-moment matrix $M(p, q)$ from the two-period case.

Definition 16 (Completely Positive Tensors and Simplex Completely Positive Tensors)**.** An order-k tensor S is *completely positive (CP)* if there exist $r \in \mathbb{N}$ and nonnegative vectors $(b^{(\ell)})_{\ell=1}^r \in (\mathbb{R}^Y_{\geqslant 0})^r$ such that

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

It is *simplex completely positive (SCP)* if the vectors can be chosen in the simplex: there exist $r \in \mathbb{N}$, weights $(\gamma_\ell)_{\ell=1}^r \in \mathbb{R}^r_+$ with $\sum_{\ell=1}^r \gamma_\ell = 1$, and vectors $(\pi^{(\ell)})_{\ell=1}^r \in \Delta(Y)^r$ such that

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

The simplex restriction in the definition of SCP tensors reflects the fact that the latent vectors correspond to probability distributions rather than arbitrary nonnegative vectors. In general, simplex complete positivity is a stronger condition than complete positivity; Lemma B.4 shows that for count moment tensors the two conditions are equivalent.

Definition 17 (Order-k Count Moment Tensor)**.** For each $k \in \{1, \ldots, T\}$, define the *order-k count moment tensor* $M^{(k)}(\mathbf{q})$ by

$$M^{(k)}_{i_1\cdots i_k}(\mathbf{q}) := m_{\nu(i_1,\ldots,i_k)}(\mathbf{q}) \qquad \text{for all count vectors } \nu \text{ with } ||\nu||_1 = k.$$

Example 5. *If $|Y| = 3$ and $T = 3$, the order-3 count moment tensor can be viewed as a cube formed by stacking three 3×3 matrices, where each matrix varies the first and second period outcomes holding the 3rd outcome fixed If μ is Dirichlet $(1, 1, 1)$, then $M^{(3)}_\nu$ consists of the 3 matrices*

$$M^{(3)}_{\cdot\cdot 1} = \begin{bmatrix} \frac{1}{10} & \frac{1}{30} & \frac{1}{30} \\ \frac{1}{30} & \frac{1}{30} & \frac{1}{60} \\ \frac{1}{30} & \frac{1}{60} & \frac{1}{30} \end{bmatrix} \quad M^{(3)}_{\cdot\cdot 2} = \begin{bmatrix} \frac{1}{30} & \frac{1}{30} & \frac{1}{60} \\ \frac{1}{30} & \frac{1}{10} & \frac{1}{30} \\ \frac{1}{60} & \frac{1}{30} & \frac{1}{30} \end{bmatrix} \quad M^{(3)}_{\cdot\cdot 3} = \begin{bmatrix} \frac{1}{30} & \frac{1}{60} & \frac{1}{30} \\ \frac{1}{60} & \frac{1}{30} & \frac{1}{30} \\ \frac{1}{30} & \frac{1}{30} & \frac{1}{10}. \end{bmatrix}$$

The main result of this section generalizes the CIID characterization of Theorem 2 to more than two periods. The characterization requires that the associated moment tensor is simplex completely positive.

Theorem 4. *Fix a horizon $T \geqslant 2$. Forecast $\mathbf{q}$ admits a CIID representation if and only if the following conditions hold:*

1. Pairwise exchangeability (Definition 11).

2. Reinforcement (Definition 12).

3. Complete positivity: *The order-T count moment tensor $M^{(T)}(\mathbf{q})$ is completely positive.*

The proof of sufficiency proceeds as follows. First, Lemma B.4 shows that if a moment tensor is completely positive, it is simplex completely positive, and Lemma 5 shows that pairwise exchangeability implies the implied moments are well defined and satisfy the identities implied by the simplex constraint $\sum_i \theta_i = 1$. Next, Lemma B.1 uses the martingale property of the forecasts to show that all lower-order tensors $M^{(k)}$, $k < T$, are simplex completely positive and can all be written as a product of probability vectors with the weights. Carathéodory's theorem and the martingale identities imply that the same weights and atoms generate all lower–order tensors.

Theorem 4 thus clarifies the content of CIID rationalizability. Once forecasts satisfy pairwise exchangeability and reinforcement, no further dynamic restrictions remain: CIID rationalizability reduces to a moment-feasibility problem, namely whether the implied hierarchy of moment tensors can be jointly represented as moments of a single probability measure on the simplex.

This perspective clarifies the informational content of longer horizons. As T increases, the moment constraints tighten and converge to exchangeability.

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. This characterization clarifies the boundary between CIID models and more general random-intensity models that satisfy dynamic coherence and complete positivity but fail the simplex restriction.

Our 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) shows that the agent's updating rule is inconsistent with any CIID representation.

Conversely, our persistence result highlights the limits of one-step-ahead data. An agent who believes in spurious positive autocorrelation (persistence) will generate forecasts that are indistinguishable from those of 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, 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 (2026). "A Dutch-book trap for misspecification." *mimeo.*

Chen, X., L. P. Hansen, and P. G. Hansen (2020). "Robust identification of investor beliefs." *Proceedings of the National Academy of Sciences*, 117, 33130–33140.

Cheney, W. (2013). *Analysis for Applied Mathematics*. Springer Science & Business Media.

Dekel, E., D. Fudenberg, and S. Morris (2007). "Interim correlated rationalizability." *Theoretical Economics*, 2, 15–40.

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

Dominitz, J. and C. F. Manski (2011). "Measuring and interpreting expectations of equity returns." *Journal of applied econometrics*, 26, 352–370.

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.

Fudenberg, D. and J. Tirole (1991). "Perfect Bayesian equilibrium and sequential equilibrium." *journal of Economic Theory*, 53, 236–260.

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.

Golub, B. and S. Morris (2020). "Expectations, networks, and conventions." *arXiv preprint arXiv:2009.13802*.

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.

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

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

Schmüdgen, K. (2017). *The moment problem*. Springer.

— (2020). "Ten lectures on the moment problem." *arXiv preprint arXiv:2008.12698*.

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 Suppose $1 > p_1 > q_1^{(2)} > 0$ and take the Beta parameters from (4). Then

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

The period-2 forecasts satisfy

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

Since

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

we obtain

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

By symmetry $p_1(1-q_1^{(1)}) = (1-p_1)q_1^{(2)}$, hence

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

completing the proof for 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 $\mu = p_1\delta_{(1,0)} + (1-p_1)\delta_{(0,1)}$ recovers the specified (p, q).

Case d If $p_1 = 1$, then by reinforcement $q_1^{(1)} = 1$ and $q^{(2)}$ is arbitrary, so specifying that μ is a point mass on $\theta = (1, 0)$ recovers the specified (p, q).

Since reinforcement requires $p_1 \geqslant q_1^{(2)}$, every forecast satisfying reinforcement falls into exactly one of cases a–d.

□

A.2 Proof of Lemma 2

Proof. *Only if.* Observe that by equations (11) and (13) if $\mu \in \Delta(\Delta(Y))$ is a CIID representation for (p, q) then for all $i, j \in Y$

$$\begin{aligned} m_{ij}(p,q) &= \int_{\Delta(Y)} \theta_j d\mu(\theta) \int_{\Delta(Y)} \theta_i d\mu(\theta|j) \\ &= \int_{\Delta(Y)} \theta_j d\mu(\theta) \int_{\Delta(Y)} \frac{\theta_i\theta_j}{\int_{\Delta(Y)} \theta_j d\mu(\theta)} d\mu(\theta) = \int_{\Delta(Y)} \theta_i\theta_j d\mu(\theta) \end{aligned} \tag{17}$$

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 $p_j = \sum_{i=1}^n m_{ij}(p,q)$.

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 $\mu(\cdot|j)$ arbitrary for j such that $p_j = 0$. Moreover, $M(p, q)$ is positive semidefinite because it is a mixture of rank-one positive semidefinite matrices, the outer products $\theta\theta^\top$. $\square$

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

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

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. $\square$

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

$$M(p,q) = \sum_{s=1}^{r+1} \gamma_s\, \pi^{(s)} \pi^{(s)\top}, \text{ where all } \gamma_s > 0, \sum_{s=1}^{r+1} \gamma_s = 1, \text{ and all } \pi^{(s)} \in \Delta(Y) \tag{18}$$

establishing complete positivity.

(1) ⇒ (3). Since $M(p,q)$ is a forecast matrix, its entries sum to one (as shown in the text), so complete positivity implies simplex complete positivity by Claim 1. Given the representation $M(p,q) = \sum_{s=1}^{r} \gamma_s\, \pi^{(s)} \pi^{(s)\top}$, define the discrete measure $\mu = \sum_{s=1}^{r} \gamma_s\, \delta_{\pi^{(s)}}$. Then

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

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

(3) ⇒ (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 CIID representation with finite support. For the second part, suppose that (p,q) has a CIID representation. By Theorem 2 (p,q) has a CIID representation with finite support. By Lemma 1 (p,q) satisfies symmetry and reinforcement, so the proof of Theorem 1 case a., and the fact that with binary outcomes a nondogmatic forecast has $p \neq q^{(2)}$, 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) ⇒ (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 nonnegative entries is completely positive. (See Theorem 2.4 in Berman and Shaked-Monderer [2003].)

□

A.7 Proof of Proposition 1

Proof. By equation (7), a full cycle requires $\mathrm{Cov}_\mu(\theta_{i+1}, \theta_i) \geqslant 0$ for all i along the cycle, with at least one strict inequality.

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

Case n=3: From equation (8):

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

Since variances are nonnegative, if the sum of the 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

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

A full 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

$$2H + (n-2)L \;=\; 1. \tag{19}$$

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 (19),

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

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),

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

We need $q_{i+1}^{(i)} > p_{i+1} = 1/n$. 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

$$f_n(L) = \frac{1}{4} - \frac{n-4}{2}L + \frac{n(n-4)}{4}L^2,$$

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 (19) yields $q_{i+1}^{(i)} > 1/n = p_{i+1}$ for every $i \in Y$, which constitutes a full cycle. $\square$

A.9 Proof of Corollary 3

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

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

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

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

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.

Part (ii). The inequalities $m_{ii}(p,q)s_i \geqslant \sum_{j\neq i} m_{ij}(p,q)s_j$ for all $i \in Y$ are equivalent to $s \geqslant R(p,q)s$. If $\rho(R) < 1$ then by the Neumann series lemma (see, e.g., Cheney, 2013) $s := (I-R)^{-1}\mathbf{1} \;=\; \sum_{k=0}^{\infty} R^k\mathbf{1}$ is well defined and strictly positive. Moreover, $(I-R)s = \mathbf{1}$ implies $s - Rs = \mathbf{1} \geqslant 0$, i.e. $s \geqslant Rs$, and rescaling s to be a probability distribution shows that (p,q) satisfies scaled diagonal dominance. $\square$

A.10 Proof of Proposition 3

Proof. 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 be defined by $m_{ij} = \langle f_i, f_j \rangle$. 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

$$q_j^{(i)} = \sum_{k=1}^{r} \frac{\phi(k)\,\theta_i(k)}{\sum_{\kappa=1}^{r}\phi(\kappa)\,\theta_i(\kappa)}\theta_j(k) \tag{20}$$

and

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

where the third equality follows from equation (20). Therefore the CP-rank of M is no more than r and by Proposition 3.2 in Berman and Plemmons [1994], $\operatorname{rank} M \leqslant \operatorname{cpr}(M)$. $\square$

A.11 Proof of Proposition 4

Proof. Since a forecast (p,q) has $\operatorname{rank} M(p,q) = 1$ only if it is dogmatic, the statement immediately follows for $l = 1$. So in what follows, suppose $l \geqslant 2$. Since $M(p,q)$ is completely positive, $m_{ij}(p,q) = \sum_{l=1}^{r} x_i(l)x_j(l)$ with $x(l) \in \mathbb{R}_+^n$ for each l. Moreover, by Theorem 3.5 in Berman and Plemmons [1994], when $\operatorname{rank} M = l \geqslant 2$ we can pick these $(x(l))_{l=1}^{r}$ such that $r \leqslant l(l+1)/2 - 1$.

Without loss of generality, we may assume that each $x(l)$ is nonzero and that no two vectors $x(k)$ and $x(l)$ are proportional: Any zero vector contributes nothing to the sum and may be omitted, and if $x(k) = \lambda x(l)$ for some $\lambda > 0$, then

$$x(k)x(k)^{\top} + x(l)x(l)^{\top} = (\lambda^2 + 1)\, x(l)x(l)^{\top},$$

so the two terms can be merged into a single rank-one term without changing $M(p,q)$.

Let

$$S_k = \sum_{i \in Y} x_i(k) \qquad \forall k \in \{1, ..., r\}.$$

We have $1 = \sum_{i,j \in Y} m_{i,j}(p,q) = \sum_{i,j \in Y} \sum_{k=1}^{r} x_i(k)\, x_j(k) = \sum_{k=1}^{r} (\sum_{i \in Y} x_i(k))(\sum_{j \in Y} x_j(k)) = \sum_{k=1}^{r} S_k^2$. 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 $\sum_{i \in Y} x_i(k) = S_k$, each $\theta(k)$ are probability vectors in Δ^{n-1}. 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:

$$p_j = \sum_{i\in Y} m_{ij} = \sum_{i\in Y}\sum_{k=1}^{r} x_i(k)\, x_j(k) = \sum_{k=1}^{r} x_j(k) \sum_{i\in Y} x_i(k) = \sum_{k=1}^{r} x_j(k)\, S_k$$
$$= \sum_{k=1}^{r} x_j(k)\sqrt{\phi(k)} = \sum_{k=1}^{r} \theta_j(k)\sqrt{\phi(k)}\sqrt{\phi(k)} = \sum_{k=1}^{r} \theta_j(k)\,\phi(k)\,.$$

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 μ. □

A.12 Proof of Proposition 5

Proof. Suppose that q and $\hat{q}$ are related as in equation (9). Then

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

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. □

A.13 Proof of Proposition 6

Proof. Define $\alpha = 1 - \max_{i\in Y} q_i^{(i)}$. 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 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. □

A.14 Proof of Lemma 3

Proof. Suppose (p, q) has a CIID representation with measure μ. We will show that the two stated properties hold.

Pairwise exchangeability. Let $\nu \in \mathcal{N}(\mathbf{q})$ and $i, j \in Y$ be such that $\nu+e_i, \nu+e_j \in \mathcal{N}(\mathbf{q})$, and let $t = \|\|\nu\|\|_1$, so that ν arises after t observations.

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.

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 $\Pr(Y_{t+1} = i, Y_{t+2} = i \mid \nu) = \mathbb{E}_\mu[\theta_i^2 \mid \nu]$ and $\Pr(Y_{t+1} = i \mid \nu) = \mathbb{E}_\mu[\theta_i \mid \nu]$, so $q_i^{(\nu+e_i)} = \dfrac{\mathbb{E}_\mu[\theta_i^2 \mid \nu]}{\mathbb{E}_\mu[\theta_i \mid \nu]}$.

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

$$\frac{\mathbb{E}_\mu[\theta_i^2 \mid \nu]}{\mathbb{E}_\mu[\theta_i \mid \nu]} \geqslant E_\mu[\theta_i \mid \nu].$$

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

A.15 Uniqueness and Multiplicity

Let $\mathrm{Pos}([0,1])_T$ denote the cone of real polynomials of degree at most T that are nonnegative on the interval $[0,1]$, and for any sequence $(m_0, \ldots, m_T) \in \mathbb{R}^{T+1}$ let L_m be the linear functional on $\mathbb{R}[x]_{\leqslant T}$ defined by $L_m(x^j) = m_j$ for all $j \in \{0, ..., T\}$. We will say that sequence $(m_0, \ldots, m_T)$ is *interior* if $L_m(p) > 0$ for every nonzero $p \in \mathrm{Pos}([0,1])_T$. Theorem 10.7 in Schmüdgen [2017] implies that when the Hankel matrices of the implied moments are (strictly) positive definite, as is assumed in Proposition 7, the implied moment sequence $(m_{(0,0)}, m_{(1,0)}, \ldots, m_{(T,0)})$ is interior.

Theorem 5 (Unique Minimal Support Prior with Odd Moments)**.** *Let $T = 2j - 1$ for an integer $j \geqslant 1$. If the sequence $(m_0, \ldots, m_T) \in \mathbb{R}^{T+1}$ is interior, there exists a unique discrete probability measure $\mu \in \Delta([0,1])$ with exactly j support points such that $\mathbb{E}_\mu[\theta^t] = m_t$.*

Theorem 6 (Non-uniqueness of minimal-support prior with even moments)**.** *Let $T = 2k$ for an integer $k \geqslant 1$. If the associated moment sequence is interior, there is no rationalizing prior supported on k or fewer points. Moreover, there exists a one-parameter family of distinct rationalizing priors, each supported on exactly $k+1$ points, that all represent the same truncated moment sequence $(m_0, \ldots, m_{2k})$.*

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 open interval (a,b).

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

Let $\{p_1, \dots, p_j\}$ be the roots of the j-th orthogonal polynomial $P_j(x)$. Then there exist unique positive weights $\{\lambda_1, \dots, \lambda_j\}$ 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 (Schmüdgen, 2017 Theorem 10.7)**.** Let $k \in \mathbb{N}$ and let $m = (m_0, \dots, m_{2k})$ with $m_0 = 1$. The following are equivalent:

(i) There exists $\mu \in \Delta([0,1])$ such that

$$m_j = \int_0^1 x^j \, d\mu(x), \qquad \forall j \in \{0, \dots, 2k\}. \tag{21}$$

(ii) $L_m(p) \geqslant 0$ for every $p \in \mathrm{Pos}([0,1])_{2k}$.

Moreover, m lies in the interior of the truncated moment cone on $[0,1]$ if and only if $L_m(p) > 0$ for every nonzero $p \in \mathrm{Pos}([0,1])_{2k}$.

For a discrete probability measure μ supported on $[0,1]$, define its *index* by

$$\mathrm{ind}(\mu) \;:=\; 2\,\#\big(\mathrm{supp}(\mu) \cap (0,1)\big) \;+\; \mathbf{1}\{0 \in \mathrm{supp}(\mu)\} \;+\; \mathbf{1}\{1 \in \mathrm{supp}(\mu)\}.$$

Definition 18 (Canonical representing measure [Schmüdgen, 2020, Definition 3.8])**.** Let $m = (m_0, \dots, m_{2k})$ be a truncated moment sequence on $[0,1]$ that admits a representing measure. A representing measure μ for m is called *canonical* if its index satisfies $\mathrm{ind}(\mu) \leqslant 2k + 2$, where the index of a discrete measure μ supported on $[0,1]$ is defined by

$$\mathrm{ind}(\mu) \;:=\; 2\,\#\big(\mathrm{supp}(\mu) \cap (0,1)\big) \;+\; \mathbf{1}\{0 \in \mathrm{supp}(\mu)\} \;+\; \mathbf{1}\{1 \in \mathrm{supp}(\mu)\}.$$

Theorem D (Canonical representing measures [Schmüdgen, 2020, Theorem 3.9])**.** Let $m = (m_0, \dots, m_{2k})$ be an interior truncated moment sequence on $[0,1]$. Then the following hold.

(i) For every $\xi \in (0,1)$ there exists a *unique* canonical representing measure μ_ξ for m such that $\xi \in \mathrm{supp}(\mu_\xi)$.

(ii) Distinct points $\xi_1 \neq \xi_2$ in $(0,1)$ generate distinct canonical measures: $\mu_{\xi_1} \neq \mu_{\xi_2}$.

(iii) Every canonical representing measure for m has index at most $2k+2$.

A.15.1 Proof of Theorem 5

Proof. By Theorem C, interiority implies the existence of a representing measure $\mu \in \Delta([0,1])$ such that equation (21) is satisfied. Use μ to define the bilinear form on $\mathbb{R}[x]_{\leqslant T}$ that maps polynomials f, g into $\langle f, g\rangle := \int_0^1 f(x)g(x)\,d\mu(x)$. Moreover, interiority implies positive definiteness.

Claim 2. *There exists a unique monic polynomial $P_j(x)$ of degree j that is orthogonal to all polynomials of degree at most $j-1$.*

Proof. We first establish by induction on $l \in \{0, ..., j\}$ that there exists a unique monic polynomial of degree l that is orthogonal to all polynomials of degree at most $l-1$.

Base Case: For $l = 0$, the monic polynomial is trivially $P_0(x) = 1$.

Inductive Step: Assume there exists a set of mutually orthogonal monic polynomials $\{P_0, P_1, \ldots, P_{l-1}\}$ where each P_k has degree exactly k. By the basis theorem, these form an orthogonal basis for $\mathbb{R}[x]_{\leqslant l-1}$. We construct $P_l(x)$ by subtracting the projection of x^l onto $\mathbb{R}[x]_{\leqslant l-1}$:

$$P_l(x) = x^l - \sum_{k=0}^{l-1} c_k P_k(x)$$

To enforce orthogonality, we require $\langle P_j, P_i\rangle = 0$ for all $0 \leqslant i < l$:

$$0 = \left\langle x^l - \sum_{k=0}^{j-1} c_k P_k, P_i \right\rangle$$

By the inductive hypothesis, $\langle P_k, P_i\rangle = 0$ for $k \neq i$, so the sum collapses:

$$0 = \langle x^l, P_i\rangle - c_i\langle P_i, P_i\rangle \implies c_i = \frac{\langle x^l, P_i\rangle}{\langle P_i, P_i\rangle}$$

Because P_i is monic (and thus non-zero) and $P_iP_i \in \mathrm{Pos}([0,1])_{2i}$, the interiority property guarantees $\langle P_i, P_i\rangle > 0$. The coefficients c_i are well-defined, and $P_l(x)$ exists and is monic.

We next establish uniqueness. Suppose there are two monic polynomials of degree j, $P_j(x)$ and $\tilde{P}_j(x)$, both orthogonal to $\mathbb{R}[x]_{\leqslant j-1}$. Let $D(x) = P_j(x) - \tilde{P}_j(x)$.

Since both are monic of degree j, the x^j terms cancel, meaning the degree of $D(x)$ is at most $j-1$. Thus, $D \in \mathbb{R}[x]_{\leqslant j-1}$.

Because P_j and $\tilde{P}_j$ are orthogonal to all polynomials in $\mathbb{R}[x]_{\leqslant j-1}$, they are both orthogonal to $D(x)$:

$$\langle D, D\rangle = \langle P_j - \tilde{P}_j, D\rangle = \langle P_j, D\rangle - \langle \tilde{P}_j, D\rangle = 0 - 0 = 0$$

Since $DD \in \mathrm{Pos}([0,1])_{2j}$,by the interiority condition $\langle D, D\rangle = 0$ implies $D(x) = 0$. Therefore, $P_j(x) = \tilde{P}_j(x)$, proving uniqueness. □

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 roots 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}).

To prove uniqueness, assume there is a different j-point distribution with support $\{q_k\}$ and weights $\{w_k\}$ that also generates the moments $m_0, \ldots, m_{2j-1}$. Construct the monic polynomial $Q(x) = \prod(x - q_k)$ and note that since its 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 $Q(x) = P_j(x)$ which implies their roots are identical, so $\{q_k\} = \{p_k\}$. This contradicts the assumption that the distributions were different, so their support points must be the same. The uniqueness of the weights follows from the unique solution to the invertible Vandermonde system defined by these points. □

A.15.2 Proof of Theorem 6

Proof of Theorem 6. Let $T = 2k$ and let $m = (m_0, \ldots, m_{2k})$ be an interior truncated moment sequence on $[0,1]$, with associated linear functional L_m. As above, interiority implies that m admits at least one representing probability measure supported on $[0,1]$.

Step 1: No k-atomic rationalization. Suppose, toward a contradiction, that there exists a rationalizing prior ν for m supported on $\{t_1, \ldots, t_r\} \subset [0,1]$ with $r \leqslant k$. Define the polynomial $q(x) := \prod_{j=1}^{r}(x - t_j)$, which has degree $r \leqslant k$ and vanishes at every support point of ν. Then q^2 is a nonzero polynomial of degree at most $2k$ that is nonnegative on $[0,1]$, and

$$L_m(q^2) = \int_0^1 q(x)^2 \, d\nu(x) = 0,$$

contradicting interiority. Hence no rationalizing prior with k or fewer support points exists.

Step 2: Existence of multiple $(k+1)$-atomic solutions. If m is interior, Theorem D guarantees the existence of canonical representing measures. Moreover Schmüdgen Theorem 3.9 shows that for every $\xi \in (0,1)$ there is a distinct and unique canonical representing measure μ_ξ that has ξ as an atom. Choose two distinct points $\xi_1 \neq \xi_2$ in $(0,1)$. By definition of canonical measures (Definition 3.8) every canonical measure for truncation order $2k$ satisfies $\mathrm{ind}(\mu) \leqslant 2k+2$. If μ_ξ has r atoms in $(0,1)$ (and possibly atoms at the endpoints), then

$$\mathrm{ind}(\mu_\xi) = 2r + \mathbf{1}\{0 \in \mathrm{supp}(\mu_\xi)\} + \mathbf{1}\{1 \in \mathrm{supp}(\mu_\xi)\} \geqslant 2r,$$

so $2r \leqslant 2k+2$ and hence $r \leqslant k+1$. Thus each μ_{ξ_j} has at most $k+1$ atoms.

Combining this upper bound with Step 1 (which rules out any representing measure with $\leqslant k$ atoms) yields that each μ_{ξ_j} has exactly $k+1$ atoms. Since each $\xi \in (0,1)$ has a distinct canonical measure, we have a 1-parameter family of $(k+1)$-point representing measures for m. □

A.16 Proof of Theorem 4

Proof. We prove the equivalence in two steps.

($\Rightarrow$) Assume $\mathbf{q}$ 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 $\mathbf{q}$.

Let

$$m^*_\nu := \int_{\Delta(Y)} \prod_{i\in Y} \theta_i^{\nu_i} \, d\mu(\theta), \qquad ||\nu||_1 \leqslant T.$$

Under the CIID model, Bayes' rule implies that after any history with count vector ν the forecast of $Y_{||\nu||_1+1} = i$ is

$$q_i(\nu) = \mathbb{E}_\mu[\theta_i \mid \nu] = \frac{\int_{\Delta(Y)} \theta_i \prod_{j\in Y} \theta_j^{\nu_j} \, d\mu(\theta)}{\int_{\Delta(Y)} \prod_{j\in Y} \theta_j^{\nu_j}, d\mu(\theta)} = \frac{m^*_{\nu+e_i}}{m^*_\nu}, \quad \text{whenever } m^*_\nu > 0.$$

With this, it is immediate that $m_\nu(\mathbf{q}) = m^*_\nu = \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta)$ for every $\nu \in \mathcal{N}(\mathbf{q})$.

By Lemma B.2, the existence of such a measure μ with $m_\nu = \int \theta^\nu d\mu$ for all $||\nu||_1 \leqslant T$ 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 $M^{(k)}_{i_1\cdots i_k} = m_{\nu(i_1,\ldots,i_k)} = \sum_{s=1}^r \gamma_s \, \pi^{(s)}_{i_1} \cdots \pi^{(s)}_{i_k}$. [16] Pairwise exchangeability and reinforcement follow from Lemma 3.

($\Leftarrow$)Now assume that the count moment tensor $M^{(T)}$ is completely positive. By condition 1 (pairwise exchangeability) and Lemma 5(1), the implied moments $m_\nu(\mathbf{q})$ are well-defined, i.e., m_ν does not depend on the order in which outcomes are accumulated to reach ν. From the construction of m_ν we have $m_{\mathbf{0}} = 1$, $m_\nu \geqslant 0$, and, for every $||\nu||_1 \leqslant T-1$,

$$\sum_{i\in Y} m_{\nu+e_i} = \sum_{i\in Y} q_i^{(\nu)} \, m_\nu = m_\nu \sum_{i\in Y} q_i^{(\nu)} = m_\nu,$$

since each $q^{(\nu)} \in \Delta(Y)$. These are the linear consistency identities (22). By Lemma B.4 $M^{(T)}$ is simplex completely positive, and since the linear consistency conditions hold Lemma B.1 applies: the linear consistency identities together with simplex complete positivity of $M^{(T)}$ imply that every lower-order tensor $M^{(k)}$, $k = 1, \ldots, T-1$, is also simplex completely positive, with the same weights and atoms. Hence all

[16]The map $\theta \mapsto \theta^{\otimes k}$ is continuous on the compact set $\Delta(Y)$, so its image is compact, and its convex hull is compact and closed. Carathéodory's theorem therefore guarantees that each $M^{(k)}$ admits a finite simplex completely positive decomposition.

tensors $\{M^{(k)}\}_{k=1}^{T}$ are simplex completely positive, so the hypotheses of Lemma B.2 are satisfied. That lemma then implies there exists a probability measure μ on $\Delta(Y)$ such that

$$m_\nu \;=\; \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta) \qquad \text{for all } ||\nu||_1 \leqslant T.$$

Define a stochastic process by first drawing $\theta \sim \mu$ and then, conditional on θ, drawing $Y_1, \ldots, Y_T$ i.i.d. with distribution θ. Let $(\tilde{p}, \tilde{q})$ denote the associated one–step–ahead forecasts. By construction, $\tilde{p}_i = \Pr(Y_1 = i) = \int_{\Delta(Y)} \theta_i \, d\mu(\theta) = m_{e_i}$. After observing a history with count vector ν (with $||\nu||_1 < T$), Bayes' rule implies that the posterior on θ has density proportional to $\theta^\nu d\mu(\theta)$. Hence the next-step forecast satisfies

$$\tilde{q}_i(\nu) = \mathbb{E}_\mu[\theta_i \mid \nu] = \frac{\int_{\Delta(Y)} \theta_i \theta^\nu \, d\mu(\theta)}{\int_{\Delta(Y)} \theta^\nu \, d\mu(\theta)} = \frac{m_{\nu+e_i}}{m_\nu}$$

from $m_\nu = \int \theta^\nu d\mu(\theta)$. Since the forecast-defined moments satisfy $q_i(\nu) = \frac{m_{\nu+e_i}}{m_\nu}$, the CIID process generated by μ reproduces the original forecast system at all histories with $m_\nu > 0$. When $m_\nu = 0$, the value of $q(\nu)$ does not affect the induced distribution. Hence the forecast system admits a CIID representation. $\square$

B For Online Publication

Example 6. *Suppose $n = 5$ and that the forecasts are $p = (3/109, 53/218, 53/218, 53/218, 53/218)$, $q^{(1)} = (\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$, and $q_1^{(i)} = 1/53$, $q_i^{(i)} = 101/106$ and $q_j^{(i)} = 1/106$ for all $i, j \in Y \backslash \{1\}$ with $i \neq j$. Then*

$$M(p,q) = \frac{1}{436}\begin{pmatrix} 4 & 2 & 2 & 2 & 2 \\ 2 & 101 & 1 & 1 & 1 \\ 2 & 1 & 101 & 1 & 1 \\ 2 & 1 & 1 & 101 & 1 \\ 2 & 1 & 1 & 1 & 101 \end{pmatrix}$$

which is not diagonal dominant. The associated row-ratio matrix is

$$R(p,q) = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{2}{101} & 0 & \frac{1}{101} & \frac{1}{101} & \frac{1}{101} \\ \frac{2}{101} & \frac{1}{101} & 0 & \frac{1}{101} & \frac{1}{101} \\ \frac{2}{101} & \frac{1}{101} & \frac{1}{101} & 0 & \frac{1}{101} \\ \frac{2}{101} & \frac{1}{101} & \frac{1}{101} & \frac{1}{101} & 0 \end{pmatrix}$$

Since the eigenvalues of $R(p,q)$ are $\lambda_1 \approx 0.2144$, $\lambda_2 \approx -0.1847$ and $\lambda = -\frac{1}{101} \approx -0.0099$ (with multiplicity 3), the spectral radius of $R(p,q)$ is

$$\rho(R(p,q)) = \approx 0.2144 < 1.$$

Proof of Lemma 4. Consider $\nu \in \mathcal{N}(\mathbf{q})$ such that $\nu + e_j \in \mathcal{N}(\mathbf{q})$ for all $j \in Y$. For every $i, j \in Y$ by pairwise exchangeability, we have

$$q_i^{(\nu)} q_j^{(\nu+e_i)} = q_j^{(\nu)} q_i^{(\nu+e_j)}.$$

For any $i \in Y$,

$$q_i^{(\nu)} = \sum_{j \in Y} q_i^{(\nu)} q_j^{(\nu+e_i)} = \sum_{j \in Y} q_j^{(\nu)} q_i^{(\nu+e_j)}$$

which is the martingale property. □

Proof of Lemma 5. (1) Let ν be any multi-index with $||\nu||_1 \leqslant T - 1$ and consider two sequences of indices $(i_1, \ldots, i_{||\nu||_1})$ and $(j_1, \ldots, j_{||\nu||_1})$ that produce the same final multi-index ν. Any permutation transforming one ordering into the other can be written as a finite composition of adjacent transpositions. Starting from any m_ν, the moment recursion implies $m_{\nu+e_i+e_j} = q_j^{(\nu+e_i)}\, q_i^{(\nu)}\, m_\nu$ and $m_{\nu+e_j+e_i} = q_i^{(\nu+e_j)}\, q_j^{(\nu)}\, m_\nu$, which are equal from pairwise exchangeability.

(2) Suppose there exists a probability measure $\mu \in \Delta(\Delta(Y))$ such that for every multi-index ν with $||\nu||_1 \leqslant T$,

$$m_\nu = \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta).$$

Fix a count vector ν with $||\nu||_1 < T$ and suppose $m_\nu > 0$. By the definition of implied moments, $q_i^{(\nu)} = \frac{m_{\nu+e_i}}{m_\nu}$. Since μ generates the moments $m_{\nu+e_i} = \int_{\Delta(Y)} \theta_i \theta^\nu \, d\mu(\theta)$ and $m_\nu = \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta)$. Hence

$$q_i^{(\nu)} = \frac{\int_{\Delta(Y)} \theta_i \theta^\nu \, d\mu(\theta)}{\int_{\Delta(Y)} \theta^\nu \, d\mu(\theta)} = \mathbb{E}_\mu[\theta_i \mid \nu] = \int_{\Delta(Y)} \theta_i \, d\mu(\theta \mid \nu),$$

which is precisely the one-step-ahead forecast generated by the CIID model with prior μ.

If $m_\nu = 0$, $\nu \notin \mathcal{N}(\mathbf{q})$, so the conclusion still holds by point (iii) of the definition of a CIID representation.

This proves part (2).

$\square$

Lemma B.1. *Let $T \geqslant 2$, $Y = \{1, \ldots, n\}$, and let $\mathbf{q}$ be a forecast satisfying count sufficiency. Suppose that the implied moments satisfy the simplex identities*

$$\sum_{i \in Y} m_{\nu+e_i} = m_\nu \qquad \textit{for all } ||\nu||_1 \leqslant T-1, \tag{22}$$

and that the order-T count-moment tensor $M^{(T)}$ is simplex completely positive: there exist $r \in \mathbb{N}$, weights $\gamma_s > 0$ with $\sum_{s=1}^r \gamma_s = 1$, and vectors $\pi^{(s)} \in \Delta(Y)$ such that

$$M^{(T)}_{i_1 \cdots i_T} = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)}_{i_1} \cdots \pi^{(s)}_{i_T} \qquad \forall (i_1, \ldots, i_T) \in Y^T, \tag{23}$$

where for each $(i_1, \ldots, i_T)$ the entry $M^{(T)}_{i_1 \cdots i_T}$ equals $m_{\nu(i_1, \ldots, i_T)}$ with $\nu(i_1, \ldots, i_T)$ the *corresponding count vector. Then, for every $k \in \{1, \ldots, T-1\}$, the order-k tensor $M^{(k)}$ is also simplex completely positive with the same factors and weights:*

$$M^{(k)}_{i_1 \cdots i_k} = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)}_{i_1} \cdots \pi^{(s)}_{i_k} \qquad \forall (i_1, \ldots, i_k) \in Y^k.$$

Proof. The simplex complete positivity assumption (23) at order T means exactly that there is a discrete probability measure μ on $\Delta(Y)$ of the form

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

such that, for all $(i_1, \ldots, i_T) \in Y^T$,

$$M^{(T)}_{i_1 \cdots i_T} = \int_{\Delta(Y)} \theta_{i_1} \cdots \theta_{i_T} \, d\mu(\theta).$$

Equivalently, for every count vector ν with $||\nu||_1 = T$,

$$m_\nu = \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta), \tag{24}$$

where $\theta^\nu := \prod_{i \in Y} \theta_i^{\nu_i}$.

We claim that the same representation holds for all lower orders:

$$m_\nu = \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta) \qquad \text{for all } \nu \in \mathbb{N}^n \text{ with } ||\nu||_1 \leqslant T. \tag{25}$$

We prove this by backward induction on $||\nu||_1$.

For $||\nu||_1 = T$ this is exactly (24). Suppose now that (25) holds for all η with $||\eta||_1 = k + 1$, where $1 \leqslant k \leqslant T - 1$, and fix any ν with $||\nu||_1 = k$. By the simplex identity (22),

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

By the induction hypothesis, for each $i \in Y$ we have

$$m_{\nu + e_i} = \int_{\Delta(Y)} \theta^{\nu + e_i} \, d\mu(\theta) = \int_{\Delta(Y)} \theta_i \, \theta^\nu \, d\mu(\theta).$$

Summing over i yields

$$m_\nu = \sum_{i\in Y} m_{\nu+e_i} = \int_{\Delta(Y)} \Big(\sum_{i\in Y}\theta_i\Big)\theta^\nu \, d\mu(\theta).$$

Because each $\pi^{(s)} \in \Delta(Y)$, every θ in the support of μ satisfies $\sum_{i\in Y}\theta_i = 1$. Hence

$$\sum_{i\in Y}\theta_i = 1 \quad \mu\text{-a.s.},$$

and therefore

$$m_\nu = \int_{\Delta(Y)} \theta^\nu \, d\mu(\theta),$$

which is (25) for $||\nu||_1 = k$. This completes the backward induction and proves (25) for all $||\nu||_1 \leqslant T$.

Finally, fix any $k \in \{1,\dots,T-1\}$ and any $(i_1,\dots,i_k) \in Y^k$ with induced count vector $\nu(i_1,\dots,i_k)$. Then

$$M^{(k)}_{i_1\cdots i_k} = m_{\nu(i_1,\dots,i_k)} = \int_{\Delta(Y)} \theta_{i_1}\cdots\theta_{i_k}\, d\mu(\theta) = \sum_{s=1}^r \gamma_s\, \pi^{(s)}_{i_1}\cdots\pi^{(s)}_{i_k},$$

which shows that $M^{(k)}$ is simplex completely positive, with the same factors $\{\pi^{(s)}\}$ and weights $\{\gamma_s\}$ as $M^{(T)}$. $\square$

Lemma B.2 (Tensor Characterization up to T). *Let $\{m_\nu\}_{||\nu||_1\leqslant T}$ be a collection of nonnegative numbers with $m_{\mathbf{0}} = 1$. The following are equivalent:*

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

$$m_\nu = \mathbb{E}_\mu\Big[\prod_{i\in Y}\theta_i^{\nu_i}\Big] = \int_{\Delta(Y)} \prod_{i\in Y}\theta_i^{\nu_i}\, d\mu(\theta) \qquad \forall ||\nu||_1 \leqslant T.$$

2. *The symmetric order-k tensor $M^{(k)}$ defined by*

$$M^{(k)}_{i_1\cdots i_k} \;:=\; m_{e_{i_1}+\cdots+e_{i_k}}$$

is simplex completely positive for each $k \in \{1,\dots,T\}$, and the numbers $\{m_\nu\}$

satisfy the linear consistency identities

$$\sum_{j \in Y} m_{\nu+e_j} = m_\nu \qquad \forall ||\nu||_1 \leqslant T-1.$$

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

$$m_\nu \;=\; \int_{\Delta(Y)} \prod_{i \in Y} \theta_i^{\nu_i} d\mu(\theta) \qquad \forall ||\nu||_1 \leqslant T.$$

Linear consistency. Fix ν with $||\nu||_1 \leqslant T-1$. Then

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

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

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

establishing the linear consistency identities.

Simplex complete positivity of $M^{(k)}$. For each $k \in \{1, \ldots, T\}$, we have

$$M^{(k)}_{i_1 \cdots i_k} \;=\; \int_{\Delta(Y)} \prod_{j=1}^{k} \theta_{i_j} \, d\mu(\theta) \qquad \forall (i_1, \ldots, i_k) \in Y^k.$$

Equivalently,

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

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 that for every $||\nu||_1 \leqslant T-1$, $\sum_{j\in Y} m_{\nu+e_j} = m_\nu$; 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 $m_\nu = \int_{\Delta(Y)} \prod_{i\in Y} \theta_i^{\nu_i}\, d\mu(\theta)$ for all $||\nu||_1 \leqslant T$.

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

$$L\left(\prod_{i\in Y} \theta_i^{\nu_i}\right) := m_\nu \qquad \text{for all } \nu \text{ with } ||\nu||_1 \leqslant T,$$

and extend linearly to arbitrary polynomials $p(\theta) = \sum_{||\nu||_1\leqslant T} a_\nu \prod_{i\in Y} \theta_i^{\nu_i}$ by $L(p) := \sum_{||\nu||_1\leqslant T} a_\nu m_\nu$.

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 $\deg p \leqslant T-1$,

$$L(S\,p) = L(p). \tag{26}$$

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

Fix ν with $||\nu||_1 \leqslant T-1$. Then

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

$$L(S\prod_{i\in Y}\theta_i^{\nu_i}) = \sum_{j\in Y} L(\prod_{i\in Y}\theta_i^{(\nu+e_j)_i}) = \sum_{j\in Y} m_{\nu+e_j}.$$

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, (26) holds for all p with $\deg p \leqslant T-1$.

Iterating this identity, we obtain

$$L(S^k p) = L(p) \quad \text{whenever } \deg p + k \leqslant T. \tag{27}$$

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

$$\tilde{p}(\theta) := S(\theta)^k \, p(\theta),$$

which has degree at most T. By (27), $L(\tilde{p}) = L(S^k p) = L(p)$, so it suffices to show $L(\tilde{p}) \geqslant 0$. Write $\tilde{p}(\theta) = \sum_{||\nu||_1 \leqslant T} a_\nu \, \theta^\nu$. Applying (27) to each monomial θ^ν with $||\nu||_1 < T$ gives $L(\theta^\nu) = L\big(S^{T-||\nu||_1}\theta^\nu\big)$. Expanding $S^{T-||\nu||_1}\theta^\nu = \left(\sum_{i \in Y} \theta_i\right)^{T-||\nu||_1} \theta^\nu$ using the multinomial theorem yields

$$S^{T-||\nu||_1}\theta^\nu = \sum_{|\beta| = T - ||\nu||_1} \binom{T - ||\nu||_1}{\beta} \theta^{\nu+\beta} = \sum_{||\eta||_1 = T} c_\eta^{(\nu)} \, \theta^\eta,$$

where $c_\eta^{(\nu)} = \binom{T-||\nu||_1}{\eta-\nu}$ if $\eta \geqslant \nu$ componentwise and $c_\eta^{(\nu)} = 0$ otherwise. Applying L gives $L(\theta^\nu) = \sum_{||\eta||_1 = T} c_\eta^{(\nu)} \, m_\eta$, which expresses $L(\theta^\nu)$ entirely in terms of the degree-T moments $\{m_\eta\}_{||\eta||_1 = T}$. Therefore, by linearity of L,

$$L(\tilde{p}) = \sum_{||\nu||_1 \leqslant T} a_\nu \, L(\theta^\nu) = \sum_{||\eta||_1 = T} b_\eta \, m_\eta = \sum_{||\eta||_1 = T} b_\eta \sum_{s=1}^{r} \gamma_s \, (\pi^{(s)})^\eta = \sum_{s=1}^{r} \gamma_s \sum_{||\eta||_1 = T} b_\eta \, (\pi^{(s)})^\eta = \sum_{s=1}^{r} \gamma_s \, \tilde{p}\big(\pi^{(s)}\big),$$

where $b_\eta := \sum_{||\nu||_1 \leqslant T} a_\nu \, c_\eta^{(\nu)}$, and the third equality uses simplex complete positivity of $M^{(T)}$. The last equality holds because the reduction via $S = 1$ ensures that $\sum_{||\eta||_1 = T} b_\eta \, (\pi^{(s)})^\eta = \tilde{p}(\pi^{(s)})$ for each $\pi^{(s)} \in \Delta(Y)$. Since each $\pi^{(s)} \in \Delta(Y)$, we have $S(\pi^{(s)}) = 1$ and hence $\tilde{p}(\pi^{(s)}) = p(\pi^{(s)}) \geqslant 0$. Therefore $L(\tilde{p}) \geqslant 0$, and thus

$$L(p) \geqslant 0 \quad \text{whenever } p \in \mathcal{A}_T \text{ and } p(\theta) \geqslant 0 \; \forall \, \theta \in \Delta(Y). \tag{28}$$

Step 4: m lies in the convex hull of truncated moment vectors. Let $\mathcal{I} := \{\nu \in \mathbb{N}^n : ||\nu||_1 \leqslant T\}$, 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 [17] 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 (28). 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 $||\nu||_1 \leqslant T$,

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

[17]Since $\Delta(Y)$ is compact and the map $\theta \mapsto \phi(\theta) = (\theta^\nu)_{\nu \in \mathcal{I}}$ is continuous, its image is compact, and therefore C is compact as a convex hull of a compact set in the finite-dimensional space $\mathbb{R}^{\mathcal{I}}$.

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). $\square$

Lemma B.3. *For any horizon $T \geqslant 0$, $\sum_{(i_1,\dots,i_T)\in Y^T} M^{(T)}_{i_1\cdots i_T}(\mathbf{q}) = 1$.*

Proof. We proceed by induction on T.

Base case ($T = 0$). By definition $M^{(0)}(\mathbf{q}) = m_{\mathbf{0}} = 1$.

Inductive step. Suppose the claim holds for T. Let $\nu = e_{i_1} + \cdots + e_{i_T}$. Then

$$\begin{aligned}
\sum_{(i_1,\dots,i_{T+1})\in Y^{T+1}} M^{(T+1)}_{i_1\cdots i_{T+1}}(\mathbf{q}) &= \sum_{(i_1,\dots,i_{T+1})\in Y^{T+1}} m_{e_{i_1}+\cdots+e_{i_{T+1}}} = \sum_{(i_1,\dots,i_T)\in Y^T} \sum_{i_{T+1}\in Y} m_{\nu+e_{i_{T+1}}} \\
&= \sum_{(i_1,\dots,i_T)\in Y^T} \sum_{i_{T+1}\in Y} q^{(\nu)}_{i_{T+1}}\, m_\nu = \sum_{(i_1,\dots,i_T)\in Y^T} m_\nu \underbrace{\sum_{i_{T+1}\in Y} q^{(\nu)}_{i_{T+1}}}_{=1} \\
&= \sum_{(i_1,\dots,i_T)\in Y^T} M^{(T)}_{i_1\cdots i_T}(\mathbf{q}) \;=\; 1.
\end{aligned}$$

where the first and fifth equalities use the definition of the count moment tensor, the third uses the recursion $m_{\nu+e_i} = q^{(\nu)}_i m_\nu$, the fourth uses $q^{(\nu)} \in \Delta(Y)$, and the last is the inductive hypothesis. $\square$

Lemma B.4. *If the count moment tensor $M^{(T)}(\mathbf{q})$ is completely positive, then it is simplex completely positive.*

Proof. Since $M^{(T)}(\mathbf{q})$ is completely positive, there exist a positive integer r and nonnegative vectors $\bar{\pi}^{(1)}, \dots, \bar{\pi}^{(r)} \in \mathbb{R}^Y_{\geqslant 0}$ such that

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

We may assume each $\bar{\pi}^{(s)}$ is nonzero. For each s, define

$$Z_s := \sum_{i\in Y} \bar{\pi}^{(s)}_i \;>\; 0, \qquad \pi^{(s)}_i := \frac{\bar{\pi}^{(s)}_i}{Z_s}, \qquad \gamma_s := Z_s^T.$$

Then $\pi^{(s)} \in \Delta(Y)$ since $\pi^{(s)}_i \geqslant 0$ and $\sum_{i\in Y} \pi^{(s)}_i = 1$. Substituting $\bar{\pi}^{(s)}_i = Z_s \pi^{(s)}_i$ gives

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

It remains to verify $\sum_{s=1}^{r} \gamma_s = 1$. Summing the decomposition over all $(i_1, \ldots, i_T) \in Y^T$ and using $\pi^{(s)} \in \Delta(Y)$:

$$\sum_{(i_1,\ldots,i_T)\in Y^T} M^{(T)}_{i_1\cdots i_T}(\mathbf{q}) = \sum_{s=1}^{r} \gamma_s \underbrace{\sum_{(i_1,\ldots,i_T)\in Y^T} \pi^{(s)}_{i_1} \cdots \pi^{(s)}_{i_T}}_{=\left(\sum_{i\in Y} \pi^{(s)}_i\right)^T = 1} = \sum_{s=1}^{r} \gamma_s.$$

By Lemma B.3, the left-hand side equals 1, so $\sum_{s=1}^{r} \gamma_s = 1$. Therefore $M^{(T)}(\mathbf{q})$ is simplex completely positive. $\square$