Scalable Multivariate Martingale Posteriors

Gregor Steiner[*1] **David Huk**[*1] **Mark F.J. Steel**[1]

[1]Department of Statistics, University of Warwick

Abstract

Martingale posteriors are a post-Bayesian approach that quantifies uncertainty by replacing the prior and likelihood with a predictive rule, imputing the unobserved population by predictive resampling. Existing constructions largely focus on univariate outcomes, and multivariate extensions factorise the joint predictive autoregressively, scaling poorly with dimension. We address this gap with a copula representation of the joint predictive rule: each coordinate evolves through its own univariate recursion, and a single copula, calibrated once to the data, carries the dependence between them. Because the marginals are decoupled from the dependence, they can be fitted and resampled in parallel giving a scheme that is both more scalable and tractable. We demonstrate the resulting substantial runtime advantage in one synthetic and one real-data application.

1 INTRODUCTION

Martingale posteriors (MPs) [Fong et al., 2023] offer an alternative to traditional Bayesian inference. Rather than placing a prior on a parameter, one specifies a sequence of one-step-ahead predictive distributions and imputes the unobserved population by forward simulation. The induced distribution over any functional of interest is the MP, obtained exactly, without MCMC, and in a manner that is trivially parallel across the independent forward-sampling chains. Most existing constructions, however, target univariate outcomes. The multivariate extension of Fong et al. [2023] factorises the joint predictive autoregressively, drawing each coordinate conditional on the preceding ones. This imposes a serial bottleneck within every chain: the cost of a forward step grows linearly with the dimension and cannot be parallelised across coordinates, which confines the method to low-dimensional problems.

We propose a more scalable forward sampling scheme based on a copula factorisation of the joint predictive, separating the marginal dynamics from the cross-dimensional dependence. Each coordinate is advanced by its own univariate update, while a coupling copula, fitted once to the observed data and subsequently held fixed, supplies the dependence among coordinates. A forward step then reduces to a single vectorised copula draw followed by marginal updates that run in parallel, in place of the sequential conditionals of the autoregressive scheme. The resulting method can be parallelised both across coordinates within a sequence and across sequences, and we show that convergence of the marginal sequences alone suffices for the joint to define a valid MP. The factorisation is also more tractable: each marginal is directly available, the dependence is given explicitly by the copula rather than left implicit, and for some copula families this yields a closed-form cumulative distribution function (CDF) — properties on which many downstream tasks rely.

In simple experiments, we demonstrate the significant runtime advantage of our method, while its predictive performance is comparable or only slightly worse.

2 PRELIMINARIES

Copula. Named after the Latin term for a link or bond, a copula is a joint probability distribution coupling univariate marginal distributions into a multivariate distribution:

Theorem 1 (Sklar [1959]) *Let P be a d-dimensional distribution on $\mathcal{Y} \subset \mathbb{R}^d$ with continuous marginal cumulative distribution functions $P^{(1)}, P^{(2)}, \ldots, P^{(d)}$. Then there exists a unique copula distribution $C : [0,1]^d \to [0,1]$ such that for all $\mathbf{y} = (y^{(1)}, y^{(2)}, \ldots, y^{(d)}) \in \mathbb{R}^d$:*

$$P(\mathbf{y}) = C(P^{(1)}(y^{(1)}), \ldots, P^{(d)}(y^{(d)})).$$

If a probability density function (PDF) $p : \mathcal{Y} \to \mathbb{R}$ is avail-

[*]Equal contribution.

able:

$$p(\mathbf{y}) = \prod_{i=1}^{d} p^{(i)}(y^{(i)}) \cdot c(P^{(1)}(y^{(1)}), \ldots, P^{(d)}(y^{(d)})),$$

where $p^{(1)}(y^{(1)}), \ldots, p^{(d)}(y^{(d)})$ *are the marginal PDFs, and* $c : [0,1]^d \to (0, \infty)$ *is the copula PDF.*

First, each marginal distribution is fitted independently. Then, the dependence structure among variables is modelled using a d-dimensional copula. This offers considerable tractability as the copula can be selected to capture the desired form of dependence while marginals can be separately tailored per dimension. See Appendix B for more details.

Martingale Posteriors. The MP framework [Fong et al., 2023] recasts Bayesian inference as a predictive problem. Let the parameter of interest be a functional $\theta = \theta(P)$ of the unknown data distribution $\mathbf{y} \sim P$. Given the observed sample $\mathbf{y}_{1:n}$, posterior uncertainty about θ is no longer derived from a prior and likelihood, but from the uncertainty in imputing the missing observations $\mathbf{y}_{n+1:\infty}$. This imputation proceeds by forward simulating $\mathbf{y}_{n+1}, \mathbf{y}_{n+2}, \ldots$ from a sequence of one-step-ahead predictive distributions $p(\mathbf{y}_{m+1} \mid \mathbf{y}_{1:m})$, yielding a completed dataset whose empirical measure converges to a random limit P. Repeating this forward simulation independently produces draws of $\theta(P)$, whose distribution is the MP. The terminology reflects the sufficient condition that the predictive sequence forms a martingale, which ensures the limit exists almost surely and that the construction is probabilistically coherent. When the predictive rule coincides with the posterior predictive of a traditional Bayesian model, the MP recovers the conventional posterior as a special case.

Predictive Updates. A crucial choice is the predictive rule that generates the forward samples. The simplest is the Bayesian bootstrap [Rubin, 1981]. Forward-sampling each $\mathbf{y}_{m+1}$ from the empirical distribution of $\mathbf{y}_{1:m}$ yields a limiting measure supported only on the observed points. Its discreteness is restrictive, motivating updates that spread mass continuously. Hahn et al. [2018] show that any univariate Bayesian predictive sequence admits a recursive update

$$p_m(y) = c_m\{P_{m-1}(y), P_{m-1}(y_m)\}\, p_{m-1}(y),$$

where c_m is a bivariate copula density and P_{m-1} the current predictive distribution function. However, the exact copula is intractable for most predictives. Motivated by a Dirichlet process mixture, they propose the following *copula update*

$$\begin{aligned} p_m(y) &= (1 - \alpha_m)\, p_{m-1}(y) \\ &\quad + \alpha_m\, c_\rho\{P_{m-1}(y), P_{m-1}(y_m)\}\, p_{m-1}(y), \end{aligned}$$

with c_ρ the bivariate Gaussian copula density and $\alpha_m = O(m^{-1})$. This predictive rule is extended by Fong et al. [2023] to the multivariate, regression and classification settings with later works improving its flexibility [Ghalebikesabi et al., 2023, Huk et al., 2024]. Other predictive rules include Newton's algorithm [Newton et al., 1998], as well as MPs using quantile regression [Fong and Yiu, 2025], a pretrained foundation model [Ng et al., 2025], and log-concave densities [Cui and Walker, 2026].

Importantly, the multivariate copula update of Fong et al. [2023] factorises the joint predictive autoregressively. This means that for predictive resampling, a single imputed observation is drawn one coordinate at a time from the conditionals $P_{m-1}(y^{(1)})$, $P_{m-1}(y^{(2)} \mid y^{(1)}), \ldots, P_{m-1}(y^{(d)} \mid y^{(1:d-1)})$. It inherits the autoregressive bottleneck: the i^{th} coordinate cannot be sampled from $P_{m-1}(y^{(i)} \mid y^{(1:i-1)})$ until the preceding $i-1$ univariate samples are in hand, so the d dimensions are processed strictly in sequence with a growing conditioning set. Thus, the cost of a single forward simulation of the joint scales linearly with d and admits no parallelism across dimensions, rendering predictive resampling cumbersome for even moderate dimensions.

3 COPULA-COUPLED PREDICTIVE RESAMPLING

We forego the autoregressive factorisation in favour of a copula based decomposition for better scalability. We first prescribe an update rule $P_m^{(i)} \leftarrow \psi^{(i)}(P_{m-1}^{(i)}, y_m^{(i)})$ to each dimension separately to carry its own univariate predictive distribution function $P_m^{(i)}$, $i = 1, \ldots, d$. These can be learned via one of the aforementioned univariate updates. We then tie the dimensions together with a copula C on $[0,1]^d$, fitted once to $\{P_n^{(i)}(y_m^{(i)})\}_{m=1}^n$, $i = 1, \ldots, d$ and held fixed thereafter. By Sklar's Theorem [Sklar, 1959], the joint predictive has the following form:

$$P_m(\mathbf{y}) = C\Big\{P_m^{(1)}\left(y^{(1)}\right), \ldots, P_m^{(d)}\left(y^{(d)}\right)\Big\}. \quad (1)$$

Forward Sampling. Each forward step draws one vector $\mathbf{u}_m = (u_m^{(1)}, \ldots, u_m^{(d)}) \sim C$ and imputes the coordinates by inverse-CDF sampling, $y_m^{(i)} = (P_{m-1}^{(i)})^{-1}(u_m^{(i)})$, before updating the marginals. Since C has uniform margins, each $y_m^{(i)}$ is a valid draw from $P_{m-1}^{(i)}$, with the dependence supplied by C and held fixed throughout. Both steps depend only on $\mathbf{u}_m$ and each coordinate's current marginal, and can therefore be parallelised. Algorithm 1 summarises the procedure and contrasts it with the autoregressive construction.

Computational speed. Our forward step scales better with d: the d inverse-CDF evaluations and d marginal updates carry no cross-coordinate dependencies and so have $\mathcal{O}(1)$ depth, whereas the autoregressive method has $\mathcal{O}(d)$ depth and admits no parallelism across dimensions. We

Algorithm 1 Forward sampling for a single martingale-posterior draw: our copula-coupled scheme (a) versus the joint autoregressive scheme of Fong et al. [2023] (b). Blue steps avoid serial cross-coordinate dependence: line 1 draws the whole copula pool in one batched call, and the inner loop runs independently over the d dimensions. In (b) coordinate i is sampled conditional on $y_m^{(1:i-1)}$, so the inner loop is sequential. The outer loop over m is sequential in both.

(a) Copula-coupled (ours)

Require: copula C, marginals $\{P_n^{(i)}\}_{i=1}^d$, update $\psi^{(i)}$, N

1: $\{\mathbf{u}_m\}_{m=n+1}^N \sim C$
2: **for** $m = n+1$ **to** N **do**
3:     **for** $i = 1$ **to** d **in parallel do**
4:         $y_m^{(i)} \leftarrow (P_{m-1}^{(i)})^{-1}(u_m^{(i)})$
5:         $P_m^{(i)} \leftarrow \psi^{(i)}(P_{m-1}^{(i)}, y_m^{(i)})$
6:     **end for**
7: **end for**
8: $P_N(\mathbf{y}) \leftarrow C\{P_N^{(1)}(y^{(1)}), \ldots, P_N^{(d)}(y^{(d)})\}$
9: **return** P_N

(b) Autoregressive [Fong et al., 2023]

Require: joint predictive P_n, joint update Ψ, N

1: **for** $m = n+1$ **to** N **do**
2:     **for** $i = 1$ **to** d **do**
3:         $y_m^{(i)} \sim P_{m-1}(\cdot \mid y_m^{(1:i-1)})$
4:     **end for**
5:     $P_m \leftarrow \Psi(P_{m-1}, \mathbf{y}_m)$
6: **end for**
7: **return** P_N

draw the $N-n$ copula samples in a single batch, amortising the sampling cost through vectorisation, and assign each sequence a different permutation. The recursions then run in parallel across both dimensions and sequences.

Marginal recursion. The marginal update function $\psi^{(i)}$ can be chosen separately for each dimension. In this work, to enhance tractability, we focus on the univariate Gaussian-copula recursion with explicit form

$$P_m^{(i)}(y) = (1-\alpha_m)\, P_{m-1}^{(i)}(y) + \alpha_m\, H_{\rho^{(i)}}\left\{P_{m-1}^{(i)}(y),\, u_m^{(i)}\right\},$$

where H_ρ is the conditional Gaussian copula and $\rho^{(i)}$ a per-dimension bandwidth. When the update $\psi^{(i)}$ acts directly on CDF values $u_m^{(i)} = P_{m-1}^{(i)}(y_m^{(i)})$ rather than on observables $y_m^{(i)}$ — as in the copula update above — the coordinates need not be imputed explicitly. Instead, we recurse directly using the appropriate dimension of the copula sample $\mathbf{u}_m$.

Coupling copula. C_ϕ may itself carry hyperparameters ϕ (e.g. a correlation matrix for a Gaussian or t-copula). We propose choosing ϕ by maximising the joint log-score of the observed data $\mathbf{y}_{1:n}$ under the fitted predictive (1). By Theorem 1, this reduces to a standard copula MLE problem on the pseudo-observations $\mathbf{u}_j = \left(P_n^{(1)}(y_j^{(1)}), \ldots, P_n^{(d)}(y_j^{(d)})\right)$

$$\hat{\phi} = \arg\max_{\phi} \sum_{j=1}^{n} \log c_\phi(\mathbf{u}_j).$$

We use the tractable pair-copula construction of Czado and Nagler [2022] in this paper, and leave the exploration of more flexible copula methods [Janke et al., 2021, Huk et al., 2025, Huk and Damoulas, 2026] for future work.

Regression. The construction adapts readily to regression tasks. Consider a response y with covariates $\mathbf{x} = (x^{(1)}, \ldots, x^{(d)})$. We learn a univariate predictive for each of the $d+1$ coordinates and couple them with a single copula over $(y, \mathbf{x})$ with density c. Since the joint predictive factorises into its marginals and c, the covariate marginals cancel such that

$$p_m(y \mid \mathbf{x}) = \frac{c\left\{P_m^{(y)}(y),\, P_m^{(1)}(x^{(1)}), \ldots, P_m^{(d)}(x^{(d)})\right\}}{c_{\mathbf{x}}\left\{P_m^{(1)}(x^{(1)}), \ldots, P_m^{(d)}(x^{(d)})\right\}}\, p_m(y),$$

where $c_{\mathbf{x}}$ is the covariate sub-copula. For an elliptical copula, $c_{\mathbf{x}}$ is the sub-copula indexed by the covariate block, so a single fit suffices. Alternatively, fitting numerator and denominator copulas separately allows a more flexible model.

3.1 FORWARD SAMPLING CONVERGENCE

To theoretically justify our algorithm, we show that the copula-coupled joint predictive converges to a well-defined limit if the marginals converge. This guarantees that the imputed sequence $\mathbf{y}_{n+1}, \mathbf{y}_{n+2}, \ldots$ is asymptotically exchangeable [Berti et al., 2004] as required for a valid MP [Fong et al., 2023]. We make the following assumptions:

(A1) $P_m^{(i)} \Rightarrow P^{(i)}$ almost surely for each i, where $P^{(i)}$ is strictly increasing and almost surely continuous on $\mathbb{R}$.

(A2) Given observed $y_{1:n}$, C is fixed, and for $u_m \sim C$, $u_m \perp\!\!\!\perp \{P_{m-1}^{(i)}\}_{i=1}^d$ for all resampling steps $m > n$.

Assumption (A1) holds for the univariate copula recursion and many other predictive rules, while (A2) ensures the copula stays fixed once fitted on observables. The following result, proved in Appendix A, gives the desired convergence.

Theorem 2 (Copula-coupled predictive convergence) *Under (A1) and (A2), let the forward samples for $m > n$ be drawn as $y_m^{(i)} = (P_{m-1}^{(i)})^{-1}(u_m^{(i)})$ with $\mathbf{u}_m \sim C$. Then the joint predictive P_m converges weakly to $P(\mathbf{y}) = C\{P^{(1)}(y^{(1)}), \ldots, P^{(d)}(y^{(d)})\}$, almost surely, as $m \to \infty$.*

4 EXPERIMENTS

Additional details on the experiments are available in Appendix C and code to reproduce our findings can be found at `https://github.com/gregorsteiner/MMP`.

4.1 GAUSSIAN MIXTURE

To illustrate the copula-coupled predictive resampling scheme, we consider $n = 100$ observations drawn from a two-dimensional Gaussian mixture with well-separated modes, and the two coordinates are positively dependent both within and across mixture components.

For each dimension $d = 1, 2$ we fit a univariate copula-based predictive $P_n^{(d)}$ using the recursive Gaussian-copula update, with the bandwidth parameter $\rho^{(d)}$ chosen by maximising the in-sample log-score. The two marginal sequences are then coupled via a transformation local likelihood copula [Geenens et al., 2017]. We run $B = 100$ independent forward-sampling sequences of length $T = 5000$. Figure 1 shows the posterior mean and standard deviations of the resulting joint densities along with those obtained by the AR method, overlaid on the observed data. Both methods recover the underlying structure well, with the highest mean PDFs at the modes and uncertainty concentrated in the same regions. Copula-coupled forward sampling takes only 2.9 seconds here, versus 37.2 for the autoregressive method.

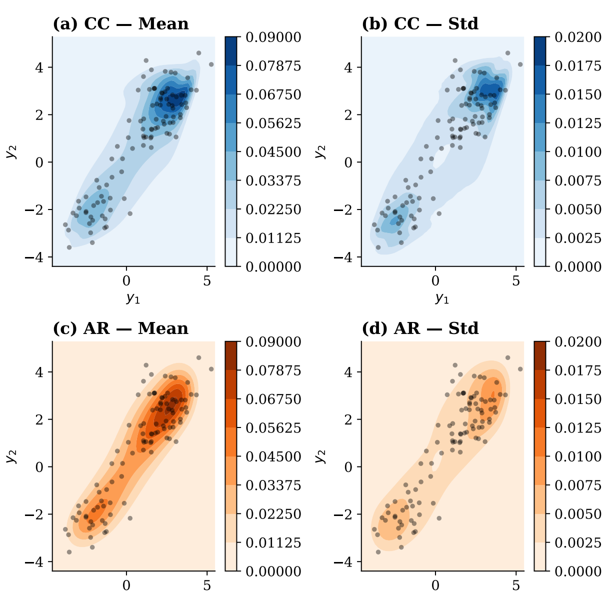

Figure 1: **Posterior uncertainty.** We show the posterior mean (left) and standard deviation (right) of the joint predictive PDF for the copula-coupled (CC, blue) and the autoregressive (AR, red) method with the observed data (black). Both approaches result in similar uncertainty estimates.

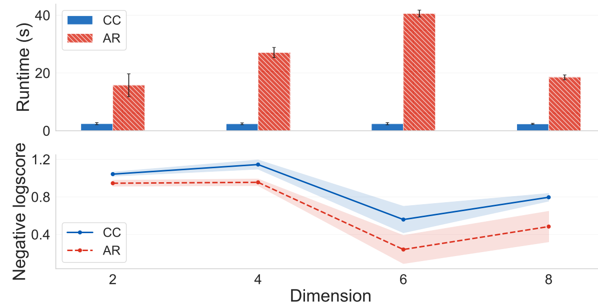

Figure 2: Runtime (in seconds) and negative log-score on the test data for the copula-coupled (CC) and the autoregressive (AR) resampling on the `abalone` data. We show the mean and standard deviation across three random splits.

4.2 ABALONE

We evaluate copula-coupled predictive resampling using a vine copula [Czado and Nagler, 2022], against the autoregressive method of Fong et al. [2023] on the `abalone` dataset [Nash et al., 1994], using the first $d \in \{2, 4, 6, 8\}$ variables. We use a 50/50 train-test split repeated across three random seeds, with $B = 50$ sequences of length $T = 2000$, and report the negative log-score at the posterior mean on the test data. Figure 2 displays the results. The forward sampling time of our method remains essentially *constant with the dimension*, whereas the autoregressive scales substantially worse. The decrease in runtime from $d = 6$ to $d = 8$ for the autoregressive method is worth exploring in future work. In terms of predictive performance, the autoregressive method achieves a lower negative log-score across all dimensions. However, the gap remains modest, suggesting that our approach quantifies uncertainty reasonably well at a fraction of the computational cost.

5 DISCUSSION

We have introduced a copula-coupled scheme for multivariate MPs, in which each coordinate is advanced by its own predictive recursion while a single copula carries the dependence. This separation makes the scheme both scalable through parallelisation and tractable through directly available marginals and an explicit dependence structure. This matters, for instance, for causal inference, which reasons about marginal and conditional distributions separately, and for safety-critical applications that require auditing uncertainty per variable as well as through the joint. Since the marginals and the dependence can be inspected and tested independently, our construction is well-suited to tractable probabilistic modelling. In future work, we will explore the selection of the dependence structure for different data types, as well as alternative choices for marginal updates.

References

Patrizia Berti, Luca Pratelli, and Pietro Rigo. Limit theorems for a class of identically distributed random variables. *The Annals of Probability*, 32(3A):2029–2052, 2004.

Fuheng Cui and Stephen G Walker. Martingale posterior distributions for log-concave density functions. *Journal of Computational and Graphical Statistics*, pages 1–11, 2026.

Claudia Czado and Thomas Nagler. Vine copula based modeling. *Annual Review of Statistics and Its Application*, 9:453–477, 2022.

Edwin Fong and Andrew Yiu. Bayesian quantile estimation and regression with martingale posteriors. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2025.

Edwin Fong, Chris Holmes, and Stephen G Walker. Martingale posterior distributions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(5): 1357–1391, 2023.

Gery Geenens, Arthur Charpentier, and Davy Paindaveine. Probit transformation for nonparametric kernel estimation of the copula density. *Bernoulli*, pages 1848–1873, 2017.

Sahra Ghalebikesabi, Chris C Holmes, Edwin Fong, and Brieuc Lehmann. Quasi-bayesian nonparametric density estimation via autoregressive predictive updates. In *Uncertainty in Artificial Intelligence*, pages 658–668. PMLR, 2023.

P Richard Hahn, Ryan Martin, and Stephen G Walker. On recursive Bayesian predictive distributions. *Journal of the American Statistical Association*, 113(523):1085–1093, 2018.

David Huk and Theodoros Damoulas. Diffusion and flow-based copulas: Forgetting and remembering dependencies. In *The Fourteenth International Conference on Learning Representations*, 2026.

David Huk, Yuanhe Zhang, Mark Steel, and Ritabrata Dutta. Quasi-bayes meets vines. *Advances in Neural Information Processing Systems*, 37:40359–40392, 2024.

David Huk, Mark Steel, and Ritabrata Dutta. Your copula is a classifier in disguise: classification-based copula density estimation. In *International Conference on Artificial Intelligence and Statistics*, pages 3790–3798. PMLR, 2025.

Tim Janke, Mohamed Ghanmi, and Florian Steinke. Implicit generative copulas. *Advances in Neural Information Processing Systems*, 34:26028–26039, 2021.

Warwick Nash, Tracy Sellers, Simon Talbot, Andrew Cawthorn, and Wes Ford. Abalone. UCI Machine Learning Repository, 1994.

Michael A Newton, Fernando A Quintana, and Yunlei Zhang. Nonparametric bayes methods using predictive updating. In *Practical nonparametric and semiparametric Bayesian statistics*, pages 45–61. Springer, 1998.

Kenyon Ng, Edwin Fong, David T Frazier, Jeremias Knoblauch, and Susan Wei. Tabmgp: Martingale posterior with tabpfn. *arXiv preprint arXiv:2510.25154*, 2025.

Donald B Rubin. The bayesian bootstrap. *The Annals of Statistics*, 9(1):130–134, 1981.

M Sklar. Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP*, volume 8, pages 229–231, 1959.

Scalable Multivariate Martingale Posteriors (Supplementary Material)

Gregor Steiner^{*1} **David Huk**^{*1} **Mark F.J. Steel**[1]

[1]Department of Statistics, University of Warwick

A THEOREM

We consider the random variables $y_m^{(i)} := (P_{m-1}^{(i)})^{-1}(u_m^{(i)}), \quad i = 1, \ldots, d$, where each $P_{m-1}^{(i)}$ is a random cumulative distribution function and $\mathbf{u}_m = (u_m^{(1)}, \ldots, u_m^{(d)})$ has copula C on $(0,1)^d$, independent of $\{P_{m-1}^{(i)}\}_{i=1}^d$. We impose the following assumptions.

(A1) **Pathwise weak convergence of marginals.** For each i, almost surely, $P_m^{(i)} \Rightarrow P^{(i)}$, where $P^{(i)}$ is a strictly increasing probability distribution function almost surely continuous on $\mathbb{R}$.

(A2) **Independence and copula structure.** Conditional on observations $y_{1:n}$, C is a fixed copula on $[0,1]^d$, and for each $m > n$, the random vector $\mathbf{u}_m$ is independent of $\{P_{m-1}^{(i)}\}_{i=1}^d$, where $\mathbf{u}_m \sim C$.

Conditional law representation. Conditionally on $\{P_{m-1}^{(i)}\}_{i=1}^d$, the vector $\mathbf{y}_m$ is generated by applying the inverse marginal transforms to a copula-distributed vector. Hence, for any bounded measurable function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbb{E}\Big[f(\mathbf{y}_m) \mid \{P_{m-1}^{(i)}\}_{i=1}^d\Big] = \int_{(0,1)^d} f\big(P_{m-1}^{-1}(\mathbf{u})\big)\, dC(\mathbf{u}),$$

where we use $P_{m-1}^{-1}(\mathbf{u}) := \big((P_{m-1}^{(1)})^{-1}(u^{(1)}), \ldots, (P_{m-1}^{(d)})^{-1}(u^{(d)})\big)$ for the component-wise inverse. Equivalently, the conditional distribution of $\mathbf{y}_m$ given $\{P_{m-1}^{(i)}\}_{i=1}^d$ is the pushforward measure $\mathcal{L}(\mathbf{y}_m \mid \{P_{m-1}^{(i)}\}_{i=1}^d) = (P_{m-1}^{-1})_{\#}C$. Define similarly $\mathcal{L}(\mathbf{y} \mid \{P^{(i)}\}_{i=1}^d) := (P^{-1})_{\#}C$. We show that conditional on $y_{1:n}$, $\mathcal{L}(\mathbf{y}_m \mid \{P_{m-1}^{(i)}\}_{i=1}^d) \Rightarrow \mathcal{L}(\mathbf{y} \mid \{P^{(i)}\}_{i=1}^d)$ almost surely.

Proof

Fix ω in the almost sure event where (A1) holds. Then, for each i, the functions $P_{m-1}^{(i)}(\omega)$ are deterministic distribution functions satisfying $P_{m-1}^{(i)}(\omega) \Rightarrow P^{(i)}(\omega)$, where $P^{(i)}(\omega)$ is continuous. By the quantile convergence theorem, this implies pointwise convergence of the corresponding generalised inverse functions:

$$(P_{m-1}^{(i)}(\omega))^{-1}(u^{(i)}) \to (P^{(i)}(\omega))^{-1}(u^{(i)}), \quad \forall u^{(i)} \in (0,1). \tag{2}$$

Let $f : \mathbb{R}^d \to \mathbb{R}$ be continuous and bounded. Then

$$\int f\, d\mathcal{L}(\mathbf{y}_m \mid \{P_{m-1}^{(i)}\}_{i=1}^d)(\omega) = \mathbb{E}\Big[f(\mathbf{y}_m) \mid \{P_{m-1}^{(i)}\}_{i=1}^d\Big](\omega) \tag{3}$$

$$= \int_{(0,1)^d} f\big((P_{m-1}^{-1}(\omega))(\mathbf{u})\big)\, dC(\mathbf{u}). \tag{4}$$

Since f is continuous and bounded, and $(P_{m-1}^{-1}(\omega))(\mathbf{u}) \to (P^{-1}(\omega))(\mathbf{u})$ pointwise in $\mathbf{u}$, with C being a fixed measure conditional on $y_{1:n}$ invariant to $m > n$ due to (A2), the dominated convergence theorem yields

$$\int_{(0,1)^d} f\big((P_{m-1}^{-1}(\omega))(\mathbf{u})\big)\, dC(\mathbf{u}) \to \int_{(0,1)^d} f\big((P^{-1}(\omega))(\mathbf{u})\big)\, dC(\mathbf{u}), \tag{5}$$

$$\Leftrightarrow \int f\, d\mathcal{L}(\mathbf{y}_m \mid \{P_{m-1}^{(i)}\}_{i=1}^d)(\omega) \to \int f\, d\mathcal{L}(\mathbf{y} \mid \{P^{(i)}\}_{i=1}^d)(\omega), \tag{6}$$

which by the Portmanteau theorem implies $\mathcal{L}(\mathbf{y}_m \mid \{P_{m-1}^{(i)}\}_{i=1}^d) \Rightarrow \mathcal{L}(\mathbf{y} \mid \{P^{(i)}\}_{i=1}^d)$ almost surely. Because $\mathcal{L}(\mathbf{y}_m \mid \{P_{m-1}^{(i)}\}_{i=1}^d)$ and $\mathcal{L}(\mathbf{y} \mid \{P^{(i)}\}_{i=1}^d)$ are the probability measures uniquely corresponding to the joint cumulative distribution functions $P_{m-1}(\mathbf{y})$ and $P(\mathbf{y}) = C\big\{P^{(1)}(y^{(1)}), \ldots, P^{(d)}(y^{(d)})\big\}$, this weak convergence of measures equivalently establishes that $P_{m-1} \Rightarrow P$ almost surely.

■

B COPULA MODELS

A simple yet widely used parametric copula family is the Gaussian copula $c : [0,1]^d \to (0, \infty)$. It assumes the dependence matches that of a standard multivariate Gaussian distribution with zero mean and correlation matrix Σ:

$$c_\Sigma(u^{(1)}, \ldots, u^{(d)}) = \frac{\mathcal{N}_d(\Phi^{-1}(u^{(1)}), \ldots, \Phi^{-1}(u^{(d)}); \mathbf{0}, \Sigma)}{\prod_{i=1}^d \mathcal{N}(\Phi^{-1}(u^{(i)}); 0, 1)}.$$

In the bivariate case ($d = 2$), fitting the Gaussian copula reduces to selecting a correlation coefficient ρ.

C ADDITIONAL DETAILS ON THE EXPERIMENTS

C.1 IMPLEMENTATION OF THE AUTOREGRESSIVE METHOD

In implementing the autoregressive method, we follow Fong et al. [2023] exactly based on the code available at `https://github.com/edfong/MP`. The copula hyperparameters are chosen by optimising the log-score on the observed data, the initial distribution is a standard Gaussian, and the initial fit is averaged across 10 permutations of the observed data. The autoregressive method depends on the ordering of the d variables. However, Fong et al. [2023] find that the ordering makes little difference in practice. Thus, we select the default ordering in both examples.

C.2 ABALONE

The abalone dataset contains physical measurements of abalone shellfish, with the goal of predicting age from morphological features. After dropping the discrete sex variable, we retain 8 continuous variables: shell length, diameter, and height (all in mm), whole weight, shucked weight, viscera weight, and shell weight (all in grams), and the number of rings, which determines age (age in years = rings + 1.5). The dataset contains $4,177$ observations. The experiments with $d \in \{2, 4, 6, 8\}$ use the first d variables in the order above, so that each larger problem strictly contains the smaller ones. Table 1 displays the negative log-score and runtime results in table form.

	Neg. log-score		Runtime (s)	
d	CC	AR	CC	AR
2	1.04 ± 0.02	0.95 ± 0.03	2.72 ± 0.36	14.00 ± 0.90
4	1.14 ± 0.05	0.96 ± 0.04	2.58 ± 0.14	28.10 ± 1.69
6	0.56 ± 0.14	0.24 ± 0.15	2.57 ± 0.18	41.11 ± 1.26
8	0.80 ± 0.04	0.48 ± 0.16	2.47 ± 0.28	20.28 ± 1.30

Table 1: Neg. log-score and forward sampling runtime (mean $\pm$ standard deviation across splits) for copula-coupled (CC) and autoregressive (AR) predictive resampling on the abalone dataset.