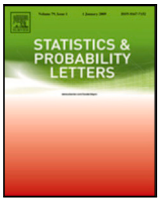
Statistics and Probability Letters 231 (2026) 110627

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro



Martingale posteriors for generative classifiers

Pier Giovanni Bissiri *, Matteo Borrotti

Department of Economics, Management and Statistics, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, Milan, 20126, Italy

ARTICLE INFO

Keywords:
Generative methods for classification
Martingale posteriors
Conditionally identically distributed sequences
Linear discriminant analysis
Quadratic discriminant analysis

ABSTRACT

Generative models for classification are a well-established method in statistics and machine learning. Martingales posteriors provide a computationally feasible method for performing prior-free Bayesian analysis. This paper aims to address the problem of uncertainty quantification through martingale posteriors for generative models for classification. To this aim, a conditionally identically distributed sequence of observations is considered. An empirical analysis is given.

1. Introduction

Fong et al. (2023) have introduced martingale posteriors, which provide a generalization of the Bayesian approach that relies on the elicitation of a sequence of predictive distributions instead of a prior distribution and a likelihood. This approach, which shares the motivation of the prequential approach of Dawid (1984), is based on the idea that the source of Bayesian uncertainty is the missing observations. If it were possible to observe the complete population, then the object of inference would be known precisely.

The existence of the martingale posterior for the object of inference is guaranteed if the sequence of observations is conditionally identically distributed (c.i.d.), a notion that has been introduced and studied in Berti et al. (2004). Letting the observations be c.i.d. amounts to relaxing the traditional condition of exchangeability and makes it easier to assess predictive distributions. As a consequence, one can conveniently sample from the martingale posterior by recursively sampling the missing observations (a procedure known as predictive resampling). This has opened the door to addressing various statistical problems through an approach that provides a prior-free, computationally feasible posterior distribution for the object of inference. In fact, martingale posteriors have been used, for example, for model-based clustering (Rodríguez et al., 2025), time series modeling (Moya and Walker, 2025), and survival analysis (Walker, 2024). For more information on Bayesian inference through c.i.d. sequences and martingale posteriors, see also Bissiri and Walker (2025) and Holmes and Walker (2023).

This paper considers classical generative classifiers (specifically LDA and QDA), which are long-standing and widely used methods in statistics and machine learning (see, for instance, James et al., 2021). We extend these classifiers to a martingale posterior framework that does not require specifying a prior distribution. Under a c.i.d. construction, we obtain closed-form posterior distributions for the class-posterior probabilities.

The outline of the paper is the following. Section 2 provides background information, Section 3 describes the proposed model, and Section 4 contains empirical analysis. In Section 5, concluding remarks are given. Proofs of the two Theorems given in Section 3 and details about the empirical analysis (Section 4) are deferred to the Supplementary file.

* Corresponding author.
E-mail addresses: pier.bissiri@unimib.it (P.G. Bissiri), matteo.borrotti@unimib.it (M. Borrotti).

https://doi.org/10.1016/j.spl.2025.110627
Received 26 October 2025; Received in revised form 16 December 2025; Accepted 17 December 2025

Available online 18 December 2025

2. Background

Before describing our proposal in Section 3, we briefly summarize martingale posteriors, introduced and studied by Fong et al. (2023), and generative models for classification.

2.1. Martingale posteriors

Let $X_{1:\infty} = (X_1, X_2, \dots)$ be the sequence of observations, where X_i is valued into $\mathbb{R}^d$ for some positive integer d and every $i = 1, 2, \dots$ We assume that $X_1, X_2, \dots$ and all other random variables considered in this paper are defined on some probability space $(\Omega, \mathscr{F}, P)$.

In Bayesian statistics, the typical choice is to let the distribution of $X_{1:\infty}$ be exchangeable, namely invariant under finite permutations. Under exchangeability, both the predictive distribution $P_n(\cdot) = P(X_{n+1} \in \cdot \mid X_{1:n})$ and the empirical measure $\nu_n = \sum_{i=1}^n \delta_{X_i}/n$ converge to a random probability measure P_∞ (with respect to the topology of weak convergence), as n diverges to infinity, almost surely. Berti et al. (2004) showed that the existence of such limit random measure P_∞ is preserved if the exchangeability condition is relaxed letting the observations $X_{1:\infty}$ be conditionally identically distributed. This means that the observations are identically distributed, i.e. $X_n \sim X_1$, and moreover,

$$X_{n+1} \mid X_{1:n} \sim X_{n+2} \mid X_{1:n},$$

for $n = 1, 2, \dots$, where $X_{1:n} = (X_1, \dots, X_n)$. This is a reasonable assessment, as there is often no reason to predict X_{n+2} differently from X_{n+1} once $X_{1:n}$ have been seen. If the observations $X_{1:\infty}$ are conditionally identically distributed then for every measurable subset B of $\mathbb{R}^d$, the sequence $P_n(B)$ is a bounded martingale and therefore converges almost surely.

As suggested by Fong et al. (2023), given some data $x_1, \dots, x_n$, one can consider the sequence of observations $X_{1:\infty}$ such that $X_1 = x_1, \dots, X_n = x_n$ and the sequence of future observations $X_{n+1:\infty}$ is conditionally identically distributed with given predictive distributions $P_{n:\infty}$. In this setting, the weak limit of both the empirical measures and the predictive distributions is almost surely a random probability measure whose distribution is the posterior distribution of P_∞, namely the conditional distribution of P_∞ given $X_1 = x_1, \dots, X_n = x_n$. So, by simulating multiple times the future observations $X_{n+1:m}$ for a large m, we can obtain several samples of the empirical measure ν_m or of P_m. Each sample is approximately a sample of P_∞ and in this way we have an approximation of the posterior distribution of P_∞.

We now recall the Gaussian generative classification framework to which the martingale posterior construction will be applied.

2.2. Generative models for classification

Denote by $(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_n, y_n)$ our data where $\boldsymbol{x}_i \in \mathbb{R}^p$ and $y_i \in \{1, \dots, k\}$, for $i = 1, \dots, n$. Generative models for classification usually assume that $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are realizations of a continuous p-variate random vector $\boldsymbol{X}$, $y_1, \dots, y_n$ are realizations of a discrete random variable Y valued into $\{1, \dots, k\}$, and, for $j = 1, \dots, k$,

$$\boldsymbol{X} \mid Y = j \sim \mathcal{N}_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j),$$

where $\mathcal{N}_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ denotes the p-variate Gaussian distribution with mean vector $\boldsymbol{\mu}_j$ and variance matrix $\boldsymbol{\Sigma}_j$.

If we denote by $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ the Gaussian multivariate density with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and we let $\pi_j = P(Y = j)$, for $j = 1, \dots, k$, then by Bayes' Theorem,

$$p_j(\boldsymbol{x}) = P(Y = j \mid \boldsymbol{X} = \boldsymbol{x}) = \frac{\pi_j \, \phi(\boldsymbol{x}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_{\ell=1}^k \pi_\ell \, \phi(\boldsymbol{x}; \boldsymbol{\mu}_\ell, \boldsymbol{\Sigma}_\ell)} \tag{1}$$

If $\hat{\boldsymbol{\mu}}_j$ is the sample mean vector referred to the observations within the j–th class, $\hat{\boldsymbol{\Sigma}}_j$ is a suitable estimate of $\boldsymbol{\Sigma}_j$, and $\hat{\pi}_j$ is the proportion of training observations $y_1, \dots, y_n$ belonging to the jth class, namely:

$$\hat{\pi}_j = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{j\}}(y_i), \qquad j = 1, \dots, k,$$

then the plug-in estimate of (1) is considered:

$$\hat{p}_j(\boldsymbol{x}) = \frac{\hat{\pi}_j \, \phi(\boldsymbol{x}; \hat{\boldsymbol{\mu}}_j, \hat{\boldsymbol{\Sigma}}_j)}{\sum_{\ell=1}^k \hat{\pi}_\ell \, \phi(\boldsymbol{x}; \hat{\boldsymbol{\mu}}_\ell, \hat{\boldsymbol{\Sigma}}_\ell)}, \tag{2}$$

for $j = 1, \dots, k$.

3. The model

As above, we denote by $(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_n, y_n)$ our data. We define the sequence of random vectors $(\boldsymbol{X}_1, Y_1), (\boldsymbol{X}_2, Y_2), \dots$, where for each $m = 1, 2, \dots$, the random vector $(\boldsymbol{X}_m, Y_m)$ is valued into $\mathbb{R}^p \times \{1, \dots, k\}$ by setting $\boldsymbol{X}_1 = \boldsymbol{x}_1, Y_1 = y_1, \dots, \boldsymbol{X}_n = \boldsymbol{x}_n, Y_n = y_n$ and

specifying the conditional distribution of Y_{m+1} given $\boldsymbol{X}_{1:m}, Y_{1:m}$, and of $\boldsymbol{X}_{m+1}$ given $\boldsymbol{X}_{1:m}, Y_{1:m+1}$, for $m = n, n+1, \dots$

For each $m = n, n+1, \dots$, we let Y_{m+1} and $\boldsymbol{X}_{1:m}$ be conditionally independent given $Y_{1:m}$ and the distribution of $Y_n, Y_{n+1}, \dots$ is assessed through the Bayesian bootstrap. In other words, if we let $M_{m,j}$ be the number of observations in $Y_{1:m}$ belonging to the j–th class, namely:

$$M_{m,j} = \sum_{i=1}^{m} \mathbf{1}_{\{Y_i = j\}},$$

then we let:

$$P(Y_{m+1} = j \mid \boldsymbol{X}_{1:m}, Y_{1:m}) = \frac{M_{m,j}}{m}, \tag{3}$$

for $j = 1, \dots, k$ and $m = n, n+1, \dots$ So, in particular, $M_{n,j} = n\hat{\pi}_j$, for $j = 1, \dots, k$.

At this stage, for each $m = n, n+1, \dots$, and $j = 1, \dots, k$, let $\overline{\boldsymbol{X}}_{m,j}$ be the sample mean vector referred to the observations $\boldsymbol{X}_{1:m}$ within the j–th class, namely:

$$\overline{\boldsymbol{X}}_{m,j} = \frac{1}{M_{m,j}} \sum_{i=1}^{m} \boldsymbol{X}_i \mathbf{1}_{\{Y_i = j\}}. \tag{4}$$

So, in particular, $\overline{\boldsymbol{X}}_{n,j} = \hat{\boldsymbol{\mu}}_j$, for $j = 1, \dots, k$.

For $m = n, n+1, \dots$, let

$$\boldsymbol{X}_{m+1} \mid \boldsymbol{X}_{1:m}, Y_{1:(m+1)} \sim \mathcal{N}_p(\overline{\boldsymbol{X}}_{m,Y_{m+1}}, \boldsymbol{\Sigma}_{m,Y_{m+1}}) \tag{5}$$

where, for $j = 1, \dots, k$, $\boldsymbol{\Sigma}_{n:\infty,j} = (\boldsymbol{\Sigma}_{n,j}, \boldsymbol{\Sigma}_{n+1,j}, \dots)$ is a sequence of p–variate (random) covariance matrices.

For each $j = 1, \dots, k$, it is reasonable to assess the first value of the sequence $\boldsymbol{\Sigma}_{n:\infty,j}$ based on the sample. So, we set $\boldsymbol{\Sigma}_{n,j} = \hat{\boldsymbol{\Sigma}}_j$, for $j = 1, \dots, k$. All subsequent theoretical results hold regardless of the method used to obtain the estimate $\hat{\boldsymbol{\Sigma}}_j$. The next Theorem states how we should assess the rest of the sequence $\boldsymbol{\Sigma}_{n:\infty,j}$ to ensure that the sequence of future observations is conditionally identically distributed.

Theorem 1. *Under the construction described above, the random sequence* $(\boldsymbol{X}_{n+1}, Y_{n+1}), (\boldsymbol{X}_{n+2}, Y_{n+2}), \dots$ *is conditionally identically distributed if and only if*

$$\boldsymbol{\Sigma}_{m+1,j} = \boldsymbol{\Sigma}_{m,j} \mathbf{1}_{\{Y_{m+1} \neq j\}} + \left(1 - \frac{1}{(1 + M_{m,j})^2}\right) \boldsymbol{\Sigma}_{m,j} \mathbf{1}_{\{Y_{m+1} = j\}}, \tag{6}$$

for $j = 1, \dots, k$.

The complete proof of the Theorem is deferred to the Supplementary Material, but it is worth to highlight here how one can easily verify that (6) is necessary for the c.i.d. condition. Indeed, if the future observations are c.i.d., then in particular, for $j = 1, \dots, k$,

$$\begin{aligned} &\operatorname{cov}(X_{m+2,h}, X_{m+2,h'} \mid \boldsymbol{X}_{1:m}, Y_{1:m}, Y_{m+2} = j) \\ &\qquad = \operatorname{cov}(X_{m+1,h}, X_{m+1,h'} \mid \boldsymbol{X}_{1:m}, Y_{1:m}, Y_{m+1} = j), \end{aligned} \tag{7}$$

where $X_{m,h}$ denotes the hth element of the random vector $\boldsymbol{X}_m$, for every $m = n, n+1, \dots$ and every $h, h' = 1, \dots, p$. By the covariance decomposition formula and (5),

$$\begin{aligned} &\operatorname{cov}(X_{m+2,h}, X_{m+2,h'} \mid \boldsymbol{X}_{1:m}, Y_{1:m}, Y_{m+2}) \\ &= \operatorname{cov}(X_{m+2,h}, X_{m+2,h'} \mid \boldsymbol{X}_{1:m}, Y_{1:m+2}) \\ &= \mathbb{E}(\operatorname{cov}(X_{m+2,h}, X_{m+2,h'} \mid \boldsymbol{X}_{1:m+1}, Y_{1:m+2}) \mid \boldsymbol{X}_{1:m}, Y_{1:m+2}) \\ &\quad + \operatorname{cov}\left(\frac{X_{m+1,h}}{M_{m+1,Y_{m+1}}}, \frac{X_{m+1,h'}}{M_{m+1,Y_{m+1}}} \mid \boldsymbol{X}_{1:m}, Y_{1:m+2}\right) \mathbf{1}_{\{Y_{m+1} = Y_{m+2}\}} \end{aligned} \tag{8}$$

Combining (7) with (8) one obtains (6).

By Theorem 1, the sequence of future observations is c.i.d. As explained in Section 2, this ensures convergence of the predictive distributions. In particular, the martingale posterior of $p_j(\boldsymbol{x})$ (for $x \in \mathbb{R}^p$ and $j = 1, \dots, k$) is given by the distribution of the almost sure limit (as $m \to \infty$) of

$$P(Y_{m+1} = j \mid \boldsymbol{X}_{1:m}, Y_{1:m}, \boldsymbol{X}_{m+1} = \boldsymbol{x}) = \frac{M_{m,j}\, \phi(\boldsymbol{x}; \overline{\boldsymbol{X}}_{m,j}, \boldsymbol{\Sigma}_{m,j})}{\sum_{\ell=1}^{k} M_{m,\ell}\, \phi(\boldsymbol{x}; \overline{\boldsymbol{X}}_{m,\ell}, \boldsymbol{\Sigma}_{m,\ell})}.$$

The following Theorem provides a closed-form for this martingale posterior (and therefore predictive resampling is not needed here).

Theorem 2. *For each* $\boldsymbol{x} \in \mathbb{R}^p$, *the martingale posterior of* $(p_1(\boldsymbol{x}), \dots, p_k(\boldsymbol{x}))$ *is the distribution of the following random vector:*

$$\left(\frac{\Pi_1\, \phi(\boldsymbol{x}; \overline{\boldsymbol{X}}_{\infty,1}, \boldsymbol{\Sigma}_{\infty,1})}{\sum_{\ell=1}^{k} \Pi_\ell\, \phi(\boldsymbol{x}; \overline{\boldsymbol{X}}_{\infty,\ell}, \boldsymbol{\Sigma}_{\infty,\ell})}, \dots, \frac{\Pi_k\, \phi(\boldsymbol{x}; \overline{\boldsymbol{X}}_{\infty,k}, \boldsymbol{\Sigma}_{\infty,k})}{\sum_{\ell=1}^{k} \Pi_\ell\, \phi(\boldsymbol{x}; \overline{\boldsymbol{X}}_{\infty,\ell}, \boldsymbol{\Sigma}_{\infty,\ell})}\right) \tag{9}$$

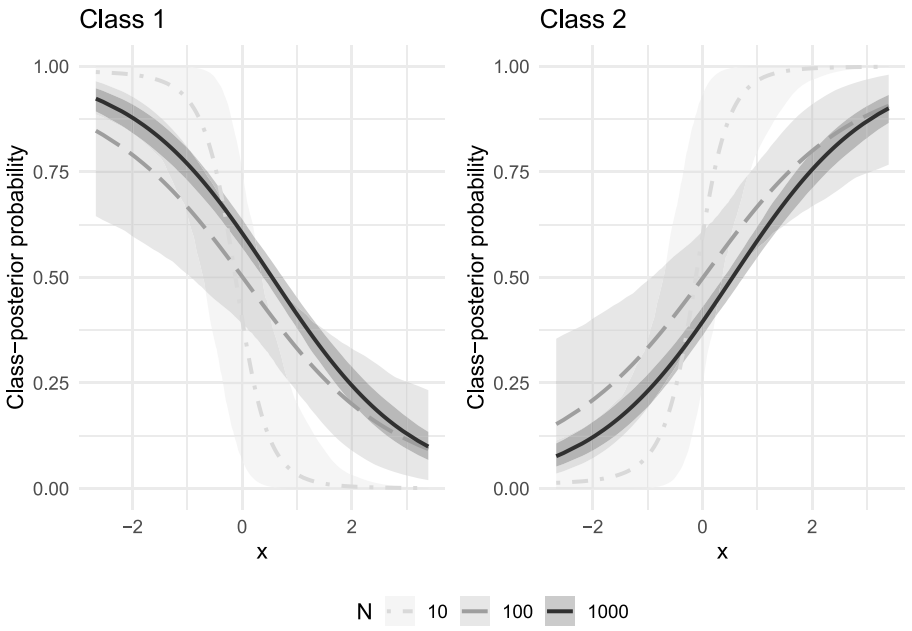


Fig. 1. MP-LDA ($p = 1$): 95% marginal credible intervals for class-posterior probabilities at $n \in \{10, 100, 1000\}$.

where

$$\Pi_j = \lim_{m\to\infty} \frac{M_{m,j}}{m}, \overline{\boldsymbol{X}}_{\infty,j} = \lim_{m\to\infty} \overline{\boldsymbol{X}}_{m,j}, \Sigma_{\infty,j} = \lim_{m\to\infty} \Sigma_{m,j},$$

for $j = 1, \dots, k$, *a.s.,* $(\Pi_1, \dots, \Pi_{k-1})$, $\overline{\boldsymbol{X}}_{\infty,1}, \dots, \overline{\boldsymbol{X}}_{\infty,k}$ *are independent, and:*

$$(\Pi_1, \dots, \Pi_{k-1}) \sim \text{Dirichlet}(n\hat{\pi}_1, \dots, n\hat{\pi}_k), \quad \Pi_k = 1 - \sum_{\ell=1}^{k-1} \Pi_\ell \tag{10}$$

$$\overline{\boldsymbol{X}}_{\infty,j} \sim \mathcal{N}_p\left(\hat{\boldsymbol{\mu}}_j, \frac{1}{n\hat{\pi}_j + 1}\hat{\boldsymbol{\Sigma}}_j\right), \quad j = 1, \dots, k, \tag{11}$$

$$\boldsymbol{\Sigma}_{\infty,j} = \frac{n\hat{\pi}_j}{n\hat{\pi}_j + 1}\hat{\boldsymbol{\Sigma}}_j, \quad j = 1, \dots, k. \tag{12}$$

The martingale posterior given in Theorem 2 is obtained as a function of the future observations under the c.i.d. construction and provides a prior-free alternative to a conjugate Bayesian posterior. Uncertainty here arises from unobserved future data rather than from prior elicitation.

As explained in detail in the Supplementary Material, the random vectors $\overline{\boldsymbol{X}}_{\infty,1}, \dots, \overline{\boldsymbol{X}}_{\infty,k}$ are independent as a consequence of (5). Intuitively, future observations that fall in a given class do not bring information about the future observations that fall in the other classes. Instead, the independence of $(\overline{\boldsymbol{X}}_{\infty,1}, \dots, \overline{\boldsymbol{X}}_{\infty,k})$ and $(\Pi_1, \dots, \Pi_{k-1})$ follows from the stability property of the Gaussian distribution. The distribution of $(\Pi_1, \dots, \Pi_{k-1})$ is Dirichlet since the distribution of $Y_n, Y_{n+1}, \dots$ is assessed through the Bayesian bootstrap.

4. Empirical analysis

In this empirical study, we apply Linear and Quadratic Discriminant Analysis (LDA, QDA) for classification, and then compare their martingale posterior extensions (MP-LDA and MP-QDA) to evaluate overall performance and potential advantages. A practical advantage of MP-LDA and MP-QDA is that they provide full posteriors for class probabilities, enabling credible intervals at prediction time, a quantification of uncertainty that LDA and QDA cannot offer. Credible intervals are computed from Monte Carlo posterior draws of the predictive class probability for each class separately.

Fig. 1 shows 95% marginal credible intervals for class-posterior probabilities under a homoscedastic Gaussian data-generating process (details in the Supplementary file, Section S2). We analyze MP-LDA with one covariate ($p = 1$) and vary the sample size $n \in 10, 100, 1000$. For each x, we plot the mean predicted probability and its 95% credible interval.

In the one-dimensional, two-class case, the mean posterior curve has a sigmoidal discriminant shape with a single decision threshold. MP-LDA (and similarly MP-QDA) provides credible intervals for $\Pr(Y = j \mid X = x)$, and those credible intervals shrink

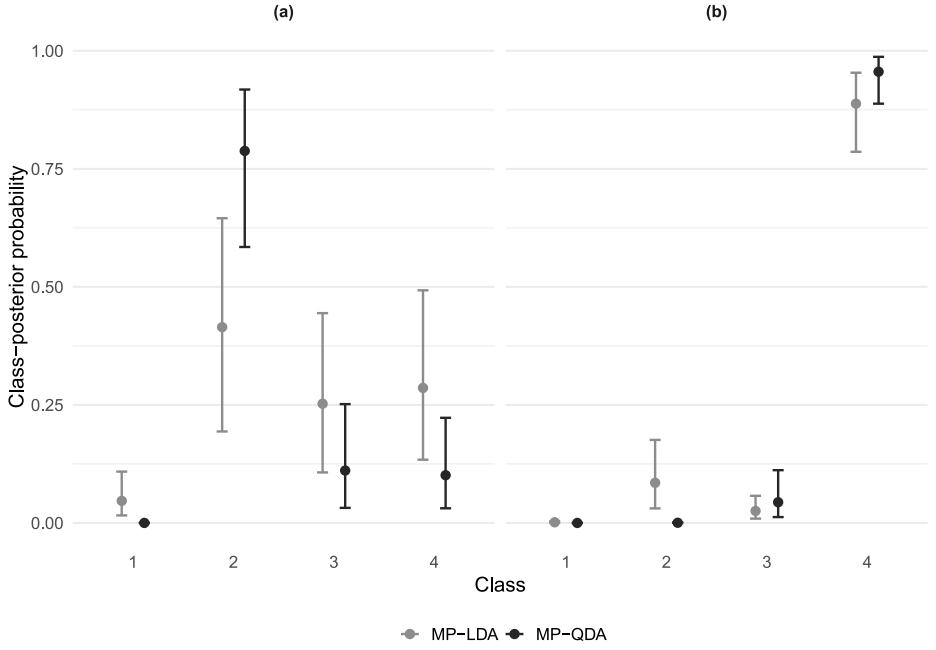


Fig. 2. 95% marginal credible intervals for class-posterior probabilities for two test observations in the Vehicle Silhouettes dataset.

Table 1
Results on Vehicle Silhouettes dataset. In this dataset, the QDA performs poorly on log-loss. LDA and MP-LDA are more robust. MP-QDA has higher performance with respect to the other approaches.

Method	Accuracy	Log-loss
LDA	0.756	0.525
QDA	0.818	1.280
MP-LDA	0.756	0.519
MP-QDA	0.842	0.564

with increasing sample size and widen in regions of class overlap. More results are available in the Supplementary Materials, Section S2.

To stress this point, we evaluate the proposed methods on the Statlog Vehicle Silhouettes dataset, which includes 846 observations and 19 variables: 18 scale-independent numeric shape features and a four-class label. Details on preprocessing and analysis are in the Supplementary file (Section S3).

Table 1 reports accuracy and log-loss metric, formulated for multi-class classification. MP generative classifiers yield more calibrated probabilities with competitive accuracy: MP-LDA matches LDA on accuracy (0.756) while slightly reducing log-loss (0.519 vs 0.525). For quadratic models, classical QDA is overconfident (log-loss 1.28), whereas MP-QDA improves accuracy (0.842) and cuts log-loss by more than half (0.564). Fig. 2 reports 95% marginal credible intervals for class-posterior probabilities for two test observations, comparing MP-LDA and MP-QDA. In both panels, the predicted class concentrates most of the posterior mass within a tight interval; when classes are ambiguous, the 95% intervals widen accordingly, quantifying residual uncertainty that classical LDA/QDA do not reveal.

Empirically, MP-LDA and MP-QDA retain the predictive strength of LDA/QDA while improving calibration and, crucially, deliver credible intervals that contract with sample size and expand in ambiguous regions.

5. Concluding remarks

In this work, we introduce martingale posterior (MP) generative classifiers as an extension for Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), retaining the interpretability of generative models while providing posterior uncertainty without eliciting priors. On synthetic data and in a case study, Martingale Posterior-LDA (MP-LDA) and Martingale Posterior-QDA (MP-QDA) match the predictive performance of their classical counterparts and also provide calibrated credible intervals that are useful when decision boundaries are more complex (e.g., class imbalance).

Future research directions make this work particularly promising. In Statistical Learning (SL), MP methods can move beyond the Gaussian setting by coupling the martingale update with nonparametric modeling, considering for instance kernel-based density estimators.

In conclusion, we view our work as a first step toward a unified prior-free Bayesian framework for modern learning tasks, with clear gains in interpretability and room for nonparametric extensions.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2025.110627.

Data availability

The link to the dataset is given in the supplementary file.

References

Berti, P., Pratelli, L., Rigo, P., 2004. Limit theorems for a class of identically distributed random variables. Ann. Probab. 32 (3), 2029–2052.
Bissiri, P.G., Walker, S.G., 2025. Bayesian analysis with conditionally identically distributed sequences. Electron. J. Stat. 19 (1), 1609–1632.
Dawid, A.P., 1984. Present position and potential developments: Some personal views: Statistical theory: The prequential approach. J. R. Stat. Soc. Ser. A (General) 147 (2), 278–292.
Fong, E., Holmes, C., Walker, S.G., 2023. Martingale posterior distributions. J. R. Stat. Soc. Ser. B. Stat. Methodol. 85 (5), 1357–1391.
Holmes, C.C., Walker, S.G., 2023. Statistical inference with exchangeability and martingales. Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 381 (2247), 20220143.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An introduction to statistical learning—with applications in R, second ed. In: Springer Texts in Statistics, Springer, New York.
Moya, B., Walker, S.G., 2025. Martingale Posterior Distributions for Time-Series Models. Statist. Sci. 40 (1), 68–80.
Rodríguez, C.E., Mena, R.H., Walker, S.G., 2025. Martingale posterior inference for finite mixture models and clustering. J. Comput. Graph. Statist. 34 (4), 1253–1262.
Walker, S.G., 2024. Martingale posterior distributions for cumulative hazard functions. Scand. J. Stat. 51 (3), 936–955.