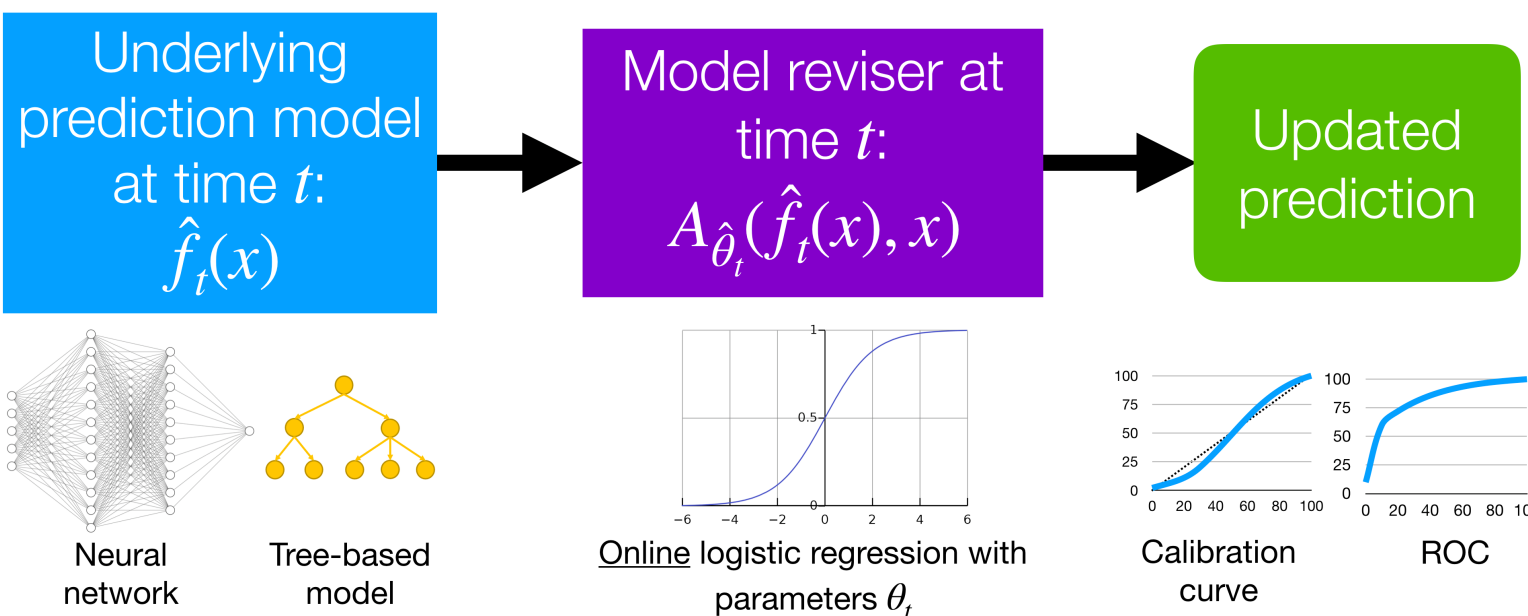


Bayesian Logistic Regression for Online Recalibration and Revision of Clinical Prediction Algorithms with Performance Guarantees

Jean Feng, Alexej Gossman, Berkman Sahiner, Romain Pirracchio

Background

- The prevailing FDA policy for machine learning (ML)-based Software as a Medical Device (SaMD) is to lock algorithms post-approval. There is now growing interest in developing a regulatory framework for **algorithms that learn continuously (or online)**.
- Online learning algorithms can outperform locked systems by protecting against changes in the environment and learning from accumulating data. Nevertheless, algorithmic modifications carry the risk of deteriorating model performance instead.
- The most popular procedure for updating clinical prediction models is **logistic recalibration and revision**.
- Can we design an **online** logistic recalibration and revision procedure that provides performance guarantees, even in the presence of arbitrary **distribution shifts**?



Contributions

- We introduce two online error quantities—**Type I and II regret**—for characterizing the long-term safety and effectiveness of an online model revision procedure.
- We design **Bayesian logistic regression** procedures for online recalibration and revision that control of Type I and II regret in the presence of arbitrary distribution shifts.
- We evaluate the methods empirically via simulation studies and a real-world case study.

Quantifying safety and effectiveness for online model revision

Type I Regret (“Safety”)

How much worse is the model reviser $\hat{\theta}_t$ compared to locking the original model \hat{f}_1 , with respect to loss function ℓ ?

$$\frac{1}{T} \sum_{t=1}^T \ell(y_t, \hat{A}_t(\hat{f}_t(x_t), x_t)) - \frac{1}{T} \sum_{t=1}^T \ell(y_t, \hat{A}_1(\hat{f}_1(x_t), x_t))$$

Loss of online model reviser Loss of locked model

Type II Regret (“Effectiveness”)

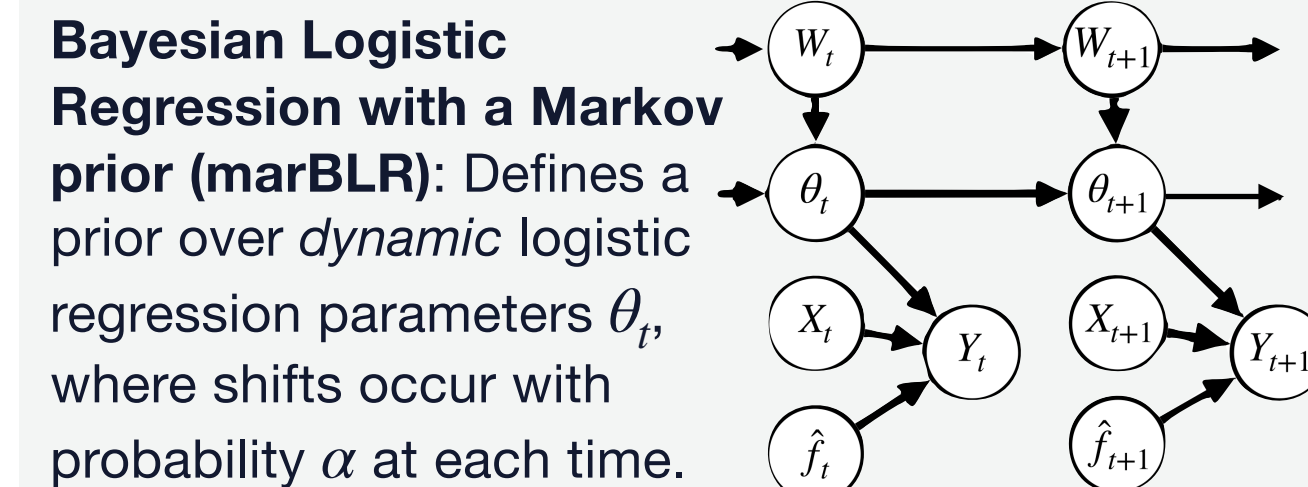
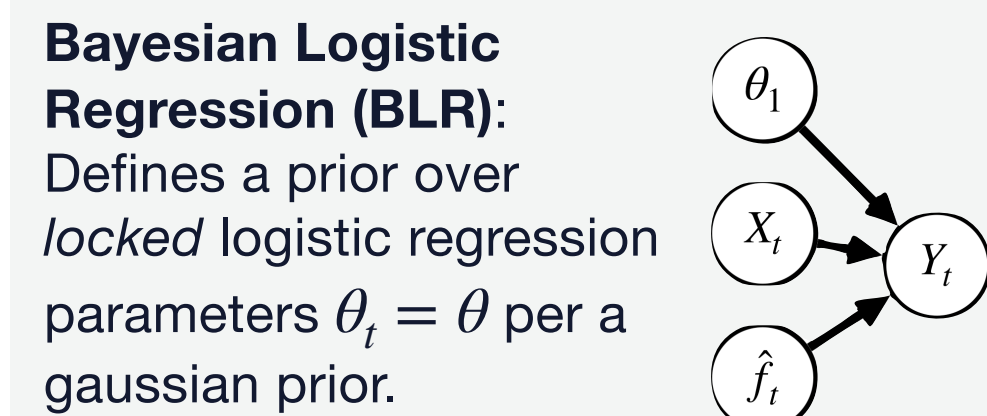
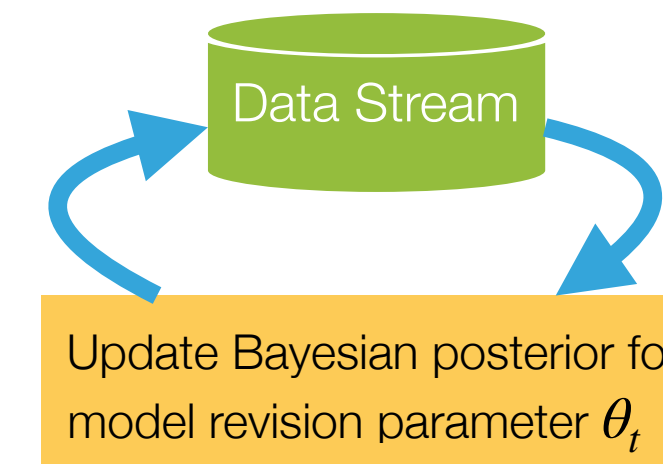
How much worse is the model reviser $\hat{\theta}_t$ compared to the oracle model reviser θ_t^* , with respect to loss function ℓ ?

$$\frac{1}{T} \sum_{t=1}^T \ell(y_t, \hat{A}_t(\hat{f}_t(x_t), x_t)) - \frac{1}{T} \sum_{t=1}^T \ell(y_t, A_t^*(\hat{f}_t(x_t), x_t))$$

Loss of online model reviser Loss of dynamic oracle

Bayesian model revision

- The Bayesian framework is naturally designed for online learning.
- At each time t :
 - Predict the posterior mean for y_t given variables x_t
 - Observe the true outcome y_t and update the Bayesian posterior for θ_t .



Online performance guarantees for Bayesian model revision

We derive Type I and II regret bounds in terms of Bayesian model hyperparameters *in the presence of arbitrary distribution shifts*, which provides a recipe for selecting model priors.

Type I Regret $\approx O\left(d \frac{\log T}{T}\right) +$

Prior probability assigned to the initial model

+ Only for MarBLR: Prior probability assigned to locked models

Type II Regret $\approx O\left(d \frac{\log T}{T}\right) +$

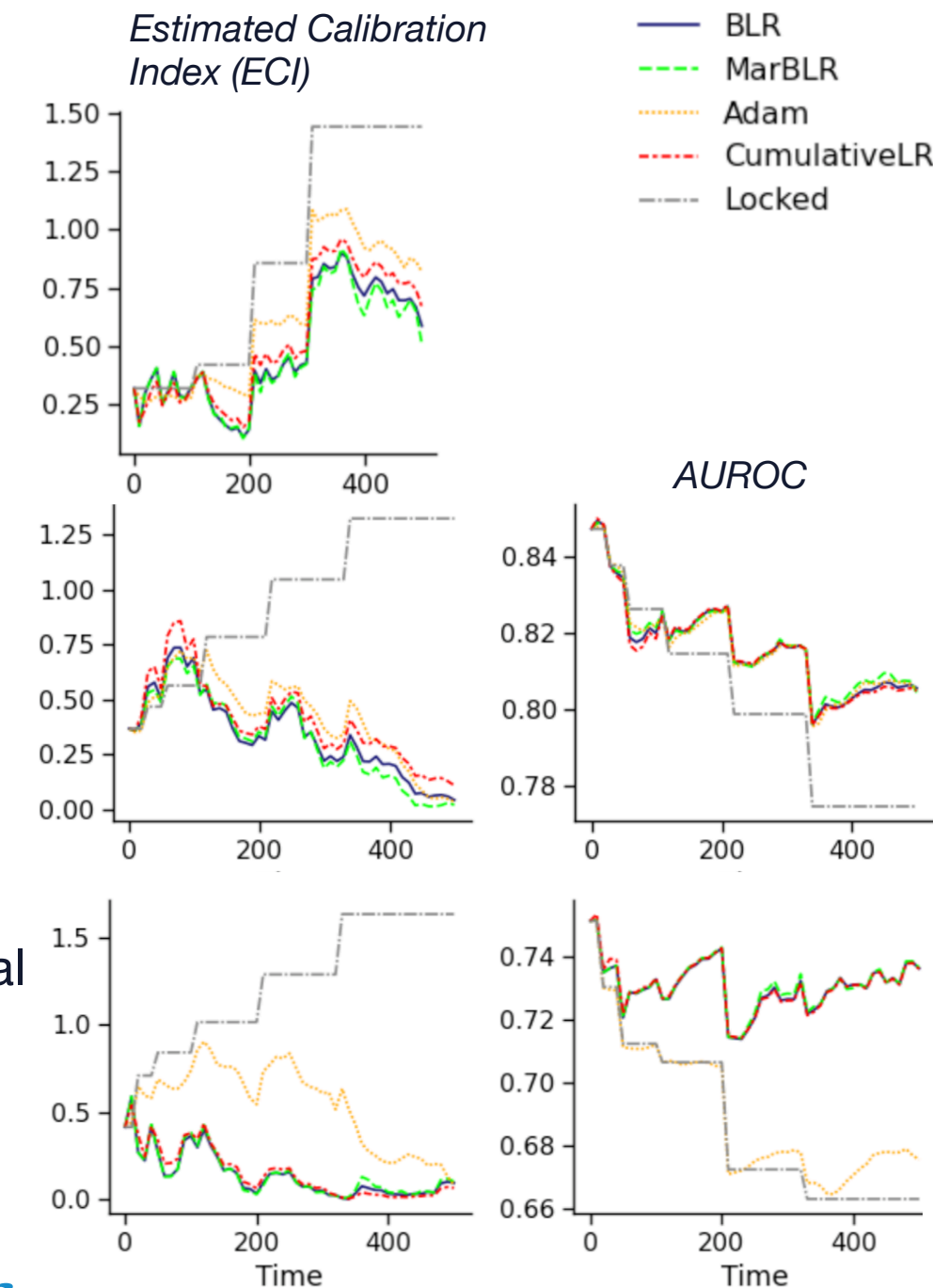
Prior probability assigned to the oracle

+ Only for BLR: Misspecification error from assuming a locked oracle

Simulation Results

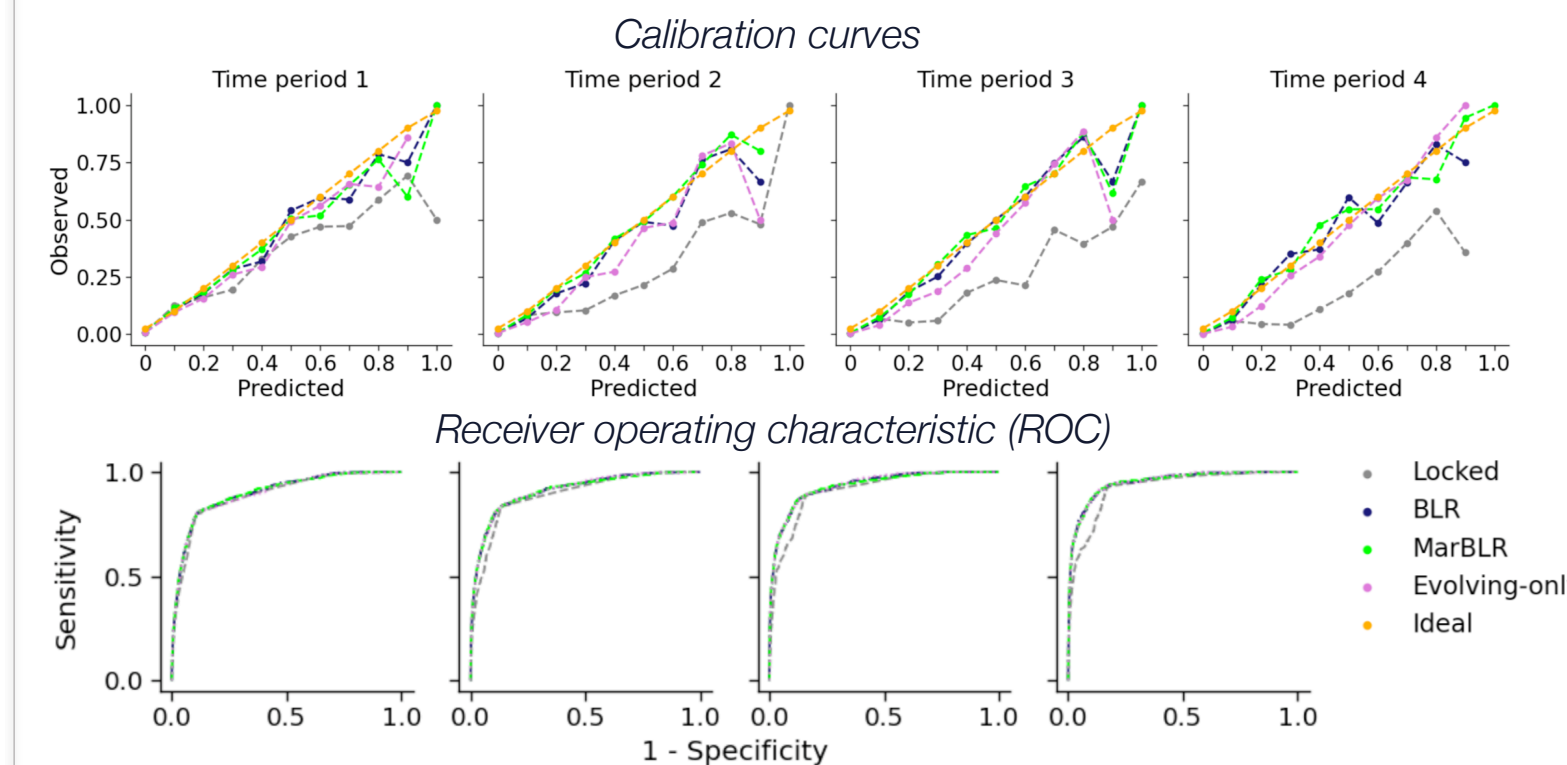
We simulate distribution shifts and perform:

- Online logistic recalibration of the original model
- Online logistic revision of the original model, adjusting for additional variables
- Online revision that combines the original model with an evolving black-box model



Real-world shifts

Online revision of a COPD risk prediction model that integrates predictions from an evolving gradient-boosted tree. Applied retrospectively to EHR data from UCSF from 2012 to 2020.



Acknowledgments: This publication was supported by the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) as part of a financial assistance award Center of Excellence in Regulatory Science and Innovation grant to University of California, San Francisco (UCSF) and Stanford University, U01FD005978. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by FDA/HHS, or the U.S. Government.