

Bayesian Prediction of Online Shoppers' Purchasing Intention

Liang-Jen Huang

Uppsala University

November 3, 2025

Motivation

- ▶ E-commerce platforms seek to understand whether a session will lead to a purchase.
- ▶ Accurate prediction improves targeting, recommendations, and marketing efficiency.
- ▶ Machine learning models provide point estimates, but often ignore uncertainty.
- ▶ Bayesian modeling provides posterior distributions and interpretable uncertainty quantification.

Goals

- ▶ Build a Bayesian model to predict online shoppers' purchasing intention.
- ▶ Identify key behavioral and contextual factors influencing purchase decisions.
- ▶ Evaluate and compare hierarchical structures to capture group-level differences.

Dataset Overview

- ▶ **Dataset:** Online Shoppers Purchasing Intention (Sakar & Kastro, 2018)
- ▶ 12,330 sessions from a one-year period.
- ▶ Target: *Revenue* (binary; purchase vs. no purchase).
- ▶ 17 predictors covering behavioral and contextual information.
- ▶ 84.5% non-purchase vs. 15.5% purchase → imbalanced data.

Variable Description

► Behavioral variables:

- *Administrative, Informational, ProductRelated* (and their corresponding durations)
- *BounceRates, ExitRates, PageValues, SpecialDay*

► Contextual variables:

- *OperatingSystem, Browser, Region, TrafficType, VisitorType, Weekend, and Month.*

- *Month, Region, and TrafficType* are treated as grouping factors for the hierarchical Bayesian model

Bayesian Logistic Regression Model

- ▶ Target variable: $y_i \in \{0, 1\}$

- ▶ Model:

$$y_i \sim \text{Bernoulli}(\pi_i), \quad \text{logit}(\pi_i) = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta}$$

- ▶ π_i : purchase probability for session i

- ▶ Priors:

$$\alpha \sim \mathcal{N}(0, 5^2), \quad \beta_j \sim \mathcal{N}(0, 2^2)$$

- ▶ **Implementation:**

- Sampling used the No-U-Turn Sampler (NUTS), a variant of HMC.
- Four Markov chains were run with 2,000 iterations each, including 1,000 warm-up steps.

Hierarchical Model

- ▶ Adds varying intercepts to capture group-level variation.
- ▶ Each observation belongs to a group $g[i]$:

$$y_i \sim \text{Bernoulli}(\pi_i)$$

$$\text{logit}(\pi_i) = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta} + \alpha_{g[i]}$$

$$\alpha_g \sim \mathcal{N}(0, \sigma_{\text{group}}), \quad \sigma_{\text{group}} \sim \text{Exponential}(1)$$

- ▶ Partial pooling stabilizes small-group estimates and captures heterogeneity.

Shrinkage Priors: Bayesian LASSO

- ▶ To improve generalization and perform variable selection:

$$\beta_j \sim \text{Laplace}(0, \lambda^{-1}), \quad \lambda \sim \text{Exponential}(1)$$

- ▶ Laplace prior induces sparsity — Bayesian analogue of LASSO.
- ▶ λ : global shrinkage parameter controlling regularization strength.

Model Assessment and Comparison

- ▶ Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO).
- ▶ Measures how well the model predicts unseen data:

$$\widehat{\text{elpd}}_{\text{loo}} = \sum_{i=1}^n \log \left(\frac{1}{S} \sum_{s=1}^S p(y_i \mid \theta^{(s)})^{w_i^{(s)}} \right),$$

where $w_i^{(s)}$ are Pareto-smoothed importance weights for observation i and posterior draw s .

- ▶ The effective number of parameters: p_{loo} reflects variability in pointwise log-likelihoods.
- ▶ Overall information criterion:

$$\text{LOOIC} = -2 \widehat{\text{elpd}}_{\text{loo}}.$$

- ▶ Pareto- k diagnostics ($k < 0.7$) assess the reliability of the approximation.

Model Comparison using PSIS-LOO

Table: Model comparison using PSIS-LOO. Higher elpd_{loo} indicates better predictive performance.

Model	elpd_{loo}	p_{loo}	LOOIC	Pareto-k status
Pooled	-3603.4	61.1	7206.9	all < 0.7
Hierarchical (Month)	-3604.2	60.7	7208.3	2 in (0.7, 1.0]
Hierarchical (TrafficType)	-3603.0	59.2	7206.1	all < 0.7
Hierarchical (Region)	-3600.5	55.8	7201.0	all < 0.7
Hierarchical (Region + TrafficType)	-3597.7	38.7	7195.4	all < 0.7
Hierarchical (Region + Traffic, LASSO)	-3593.4	37.6	7186.8	all < 0.7

Differences in Expected Log Predictive Density

Table: Differences in expected log predictive density (elpd_{loo}) relative to the best model.

Model	elpd_diff (SE)
Hierarchical (Region + Traffic)+ LASSO	0.0 (0.0)
Hierarchical (Region + TrafficType)	-4.3 (1.8)
Hierarchical (Region)	-7.1 (4.7)
Hierarchical (TrafficType)	-9.6 (4.9)
Pooled	-10.0 (5.2)
Hierarchical (Month)	-10.8 (5.2)

Final Model

Hierarchical Bayesian Logistic Regression

(with Region and TrafficType group effects + Bayesian LASSO shrinkage)

Model:

$$y_i \sim \text{Bernoulli}(p_i),$$

$$\text{logit}(p_i) = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta} + \alpha_{\text{region}[i]}^{(1)} + \alpha_{\text{traffic}[i]}^{(2)}.$$

Priors:

$$\alpha \sim \mathcal{N}(0, 5),$$

$$\lambda \sim \text{Exponential}(1),$$

$$\beta_j \sim \text{Laplace}(0, \lambda^{-1}), \quad j = 1, \dots, K,$$

$$\sigma_{\text{region}}, \sigma_{\text{traffic}} \sim \text{Exponential}(1),$$

$$\alpha_{\text{region}[r]}^{(1)} \sim \mathcal{N}(0, \sigma_{\text{region}}),$$

$$\alpha_{\text{traffic}[t]}^{(2)} \sim \mathcal{N}(0, \sigma_{\text{traffic}}).$$

Posterior Summaries of Key Parameters

Parameter	Mean	SD	95% CI
α (Global intercept)	-1.86	0.21	$[-2.26, -1.45]$
β_7 (BounceRates)	-1.5	1.6	$[-5.3, 0.8]$
β_8 (ExitRates)	-12.7	1.9	$[-16.5, -8.9]$
β_9 (PageValues)	0.0827	0.0024	$[0.0780, 0.0880]$
β_{11} (Month Dec)	-0.60	0.17	$[-0.94, -0.27]$
β_{12} (Month Feb)	-1.4	0.6	$[-2.6, -0.4]$
β_{13} (Month Nov)	0.54	0.15	$[0.26, 0.84]$
σ_{region}	0.06	0.05	$[0.00, 0.18]$
σ_{traffic}	0.30	0.10	$[0.15, 0.53]$
λ^{-1} (Shrinkage)	0.86	0.20	$[0.55, 1.30]$

- ▶ **ExitRates** and **BounceRates** show strong negative effects on purchase probability.
- ▶ **Month effects** indicate higher purchasing tendency in November and lower in December and February.
- ▶ Traffic source contributes more to uncertainty.
- ▶ Shrinkage parameter $\lambda^{-1} = 0.86$ suggests moderate regularization.

Conclusions and Future Work

Conclusions

- ▶ The hierarchical Bayesian logistic regression with *Region* and *TrafficType* effects plus Bayesian LASSO achieved the best performance.
- ▶ Partial pooling and shrinkage enabled the model to capture group-level variation while maintaining generalization.
- ▶ The hierarchical Bayesian framework provided reliable uncertainty estimates and meaningful insights for data-driven strategies in e-commerce.

Future Work

- ▶ Explore Bayesian machine learning methods (e.g., BART).
- ▶ Compare with traditional ML for predictive accuracy and interpretability.

Convergence Diagnostics using Traceplots

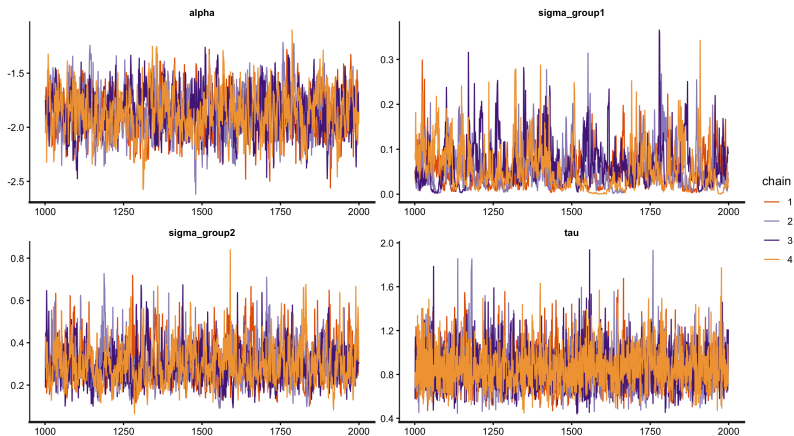


Figure: Traceplots of key parameters (α , σ_{region} , σ_{traffic} , and λ) showing good chain mixing and stable convergence.