Bayesian Survival Analysis of Patient Recovery Times in Primary Biliary Cirrhosis

Srivatsan Sarvesan
Department of Electrical and Computer Engineering
University of California San Diego
La Jolla, USA
ssarvesan@ucsd.edu

Karthik Nandagudi Raviprakash
Department of Electrical and Computer Engineering
University of California San Diego
La Jolla, USA
knandagudiraviprakas@ucsd.edu

***Abstract*—Primary Biliary Cirrhosis (PBC) is a chronic liver disease characterized by progressive bile duct damage and substantial heterogeneity in patient survival. Accurate modeling of time-to-event outcomes is essential for prognosis and treatment evaluation. We propose a Bayesian survival analysis approach using the Mayo Clinic PBC dataset, incorporating demographic variables, biochemical biomarkers, treatment assignments, and censoring information. A baseline parametric survival model is first developed to quantify covariate effects, including age, sex, treatment, serum bilirubin, and albumin levels. The model is then extended to a hierarchical Bayesian formulation to enable partial pooling and capture group-level variability across disease stages and treatment cohorts. Inference is performed using Markov Chain Monte Carlo methods implemented in PyMC, with model assessment based on posterior predictive checks and information criteria, including LOO-CV and WAIC. Results indicate that elevated bilirubin is associated with increased hazard, while higher albumin levels exhibit a protective effect. The hierarchical model demonstrates improved predictive performance and uncertainty quantification relative to the non-hierarchical baseline, highlighting the value of Bayesian hierarchical methods for robust survival modeling in PBC.**

***Index Terms*—Bayesian, Survival Analysis, PBC, Hierarchical Modeling, Hazard Ratios**

I. Introduction

Primary Biliary Cirrhosis (PBC) is a chronic autoimmune liver disease characterized by progressive destruction of the intrahepatic bile ducts, which can ultimately lead to cirrhosis and liver failure. Survival outcomes in PBC exhibit substantial heterogeneity across patients due to differences in demographic characteristics, biochemical markers, and treatment responses. Accurate modeling of time-to-event outcomes is therefore critical for prognosis, risk stratification, and evaluation of therapeutic efficacy.

Traditional survival analysis methods, including Cox proportional hazards models and parametric approaches, typically rely on fixed covariate effects and shared baseline hazard structures, which may inadequately capture patient-level heterogeneity and uncertainty. In contrast, Bayesian survival modeling provides a principled framework for incorporating prior information and producing full posterior distributions over model parameters and predicted survival times. This enables coherent uncertainty quantification through credible intervals and posterior predictive inference [1], [2].

This study addresses two primary research questions: (1) which demographic and biochemical covariates are most strongly associated with survival in patients with PBC, and (2) whether hierarchical Bayesian modeling can improve robustness and predictive performance by accounting for latent patient heterogeneity. The contributions of this project are threefold: (i) formulation of a Bayesian exponential survival model for PBC with right-censored data, (ii) extension to a hierarchical Bayesian framework that captures group-level variability, and (iii) empirical evaluation of model performance using posterior predictive checks and leave-one-out cross-validation (LOO-CV) [3].

II. Data Description

The analysis utilizes the Mayo Clinic Primary Biliary Cirrhosis (PBC) dataset (`pbc2`) [4], which comprises longitudinal clinical records from 312 patients, totaling 1,945 observations collected across multiple follow-up visits. The dataset used in this study was obtained via the `DynForest` R package [5].

The dataset includes time-to-event survival outcomes measured in years, along with right-censoring indicators. Demographic variables such as age and sex are recorded for each patient. Clinical and treatment-related covariates include drug assignment, biochemical biomarkers—specifically serum bilirubin, albumin, and serum glutamic-oxaloacetic transaminase (SGOT)—as well as physical indicators of disease severity, including edema, ascites, and spider angiomas.

Data preprocessing involved imputing missing numerical values using median imputation and categorical values using mode imputation. Categorical variables, including sex, drug, edema, and ascites, were encoded into numerical representations suitable for modeling. Numerical features were optionally standardized to improve model stability and convergence. To avoid redundancy due to repeated measurements, only baseline observations for each patient were retained for summary statistics and downstream model input.

Table I summarizes the distributions of key numerical variables by reporting the mean, standard deviation, median, minimum, and maximum values. Categorical variables were

ECE 225A: Probability and Statistics for Data Science, Fall 2025

TABLE I
SUMMARY STATISTICS OF KEY NUMERICAL VARIABLES

Variable	Mean	SD	Median	Min	Max
Age (years)	50.02	10.58	49.80	26.28	78.44
Bilirubin (mg/dL)	3.22	4.49	1.35	0.30	28.00
Albumin (g/dL)	3.52	0.42	3.55	1.96	4.64
SGOT (U/L)	122.57	56.7	114.7	26.4	457.3

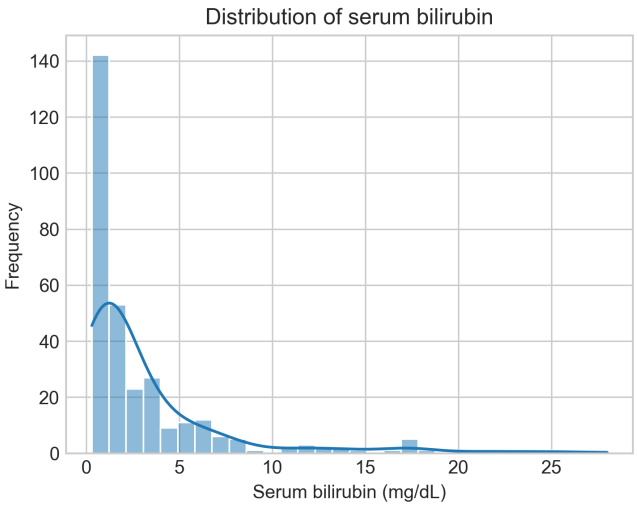

Fig. 1. Distribution of baseline serum bilirubin levels (mg/dL) across patients.

summarized using counts and percentages. Figure 1 illustrates the distribution of serum bilirubin levels, highlighting substantial inter-patient heterogeneity and its potential impact on survival outcomes. These preprocessing steps ensure a clean and analyzable dataset suitable for Bayesian survival modeling.

III. EXPLORATORY DATA ANALYSIS

We first examined patient survival patterns using Kaplan–Meier (KM) estimators [6]. Figure 2 presents KM survival curves stratified by edema status, which emerged as the most informative clinical indicator among baseline covariates. A log-rank test revealed statistically significant differences between the survival distributions ($p < 0.01$), indicating a strong association between edema and patient survival. In contrast, treatment assignment (placebo versus D-penicillamine) demonstrated minimal impact on survival, consistent with prior clinical findings.

To further investigate biomarker effects, violin plots were generated for serum bilirubin and albumin levels stratified by survival status, as shown in Figure 3. Patients with elevated serum bilirubin generally exhibited shorter survival times, whereas higher albumin levels appeared to be protective. These observations emphasize the prognostic relevance of key biochemical markers.

Overall, the exploratory analysis highlights substantial heterogeneity in survival outcomes and underscores the importance of demographic, clinical, and biochemical covariates. These findings guided covariate selection for Bayesian survival modeling and informed the hierarchical structure designed to capture patient-level variability.

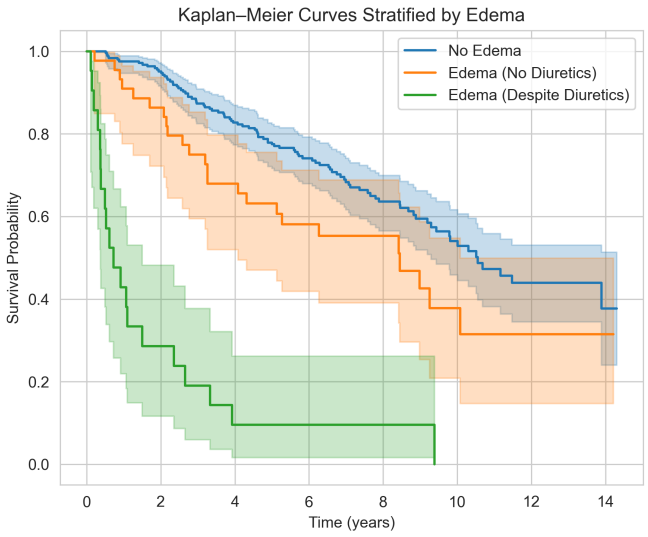

Fig. 2. Kaplan–Meier survival curves stratified by edema status, with 95% confidence intervals.

IV. MODEL AND METHODOLOGY

A. Bayesian Survival Modeling Framework

We model patient survival using a Bayesian time-to-event framework that explicitly accounts for right censoring, parameter uncertainty, and inter-patient heterogeneity [1], [2]. Let

$$T_i > 0$$

denote the observed survival time for patient $i = 1, \ldots, N$, and let

$$\delta_i \in \{0, 1\}$$

be the censoring indicator, where $\delta_i = 1$ indicates an observed event (death or transplant) and $\delta_i = 0$ indicates right censoring.

Let

$$\mathbf{x}_i \in \mathbb{R}^p$$

be a vector of baseline covariates including age, sex, treatment assignment, edema status, and selected biochemical biomarkers (serum bilirubin and albumin). Continuous covariates are standardized to zero mean and unit variance prior to modeling to improve numerical stability and interpretability of regression coefficients.

We adopt an exponential baseline hazard as a parsimonious starting point, enabling closed-form likelihoods and clear interpretation of covariate effects, while acknowledging that this assumption may be restrictive [7].

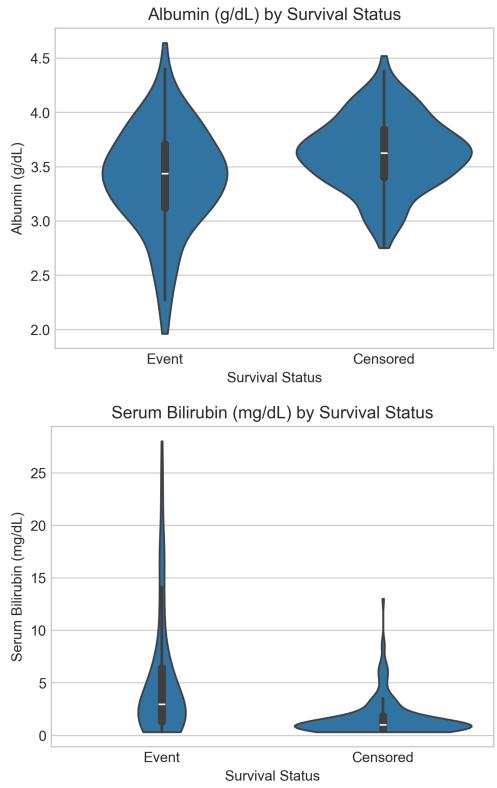

Fig. 3. Violin plots of serum bilirubin and albumin levels stratified by survival status.

B. Non-Hierarchical Exponential Survival Model

In the non-hierarchical model, the hazard function for patient i is defined as

$$\lambda_i = \exp\left(\alpha + \boldsymbol{\beta}^\top \mathbf{x}_i\right), \tag{1}$$

where α is a global intercept and $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of regression coefficients.

The corresponding survival function is

$$S(t_i \mid \alpha, \boldsymbol{\beta}, \mathbf{x}_i) = \exp(-\lambda_i t_i)\,. \tag{2}$$

Under right censoring, the log-likelihood for the observed data is

$$\log p(\mathbf{T}, \boldsymbol{\delta} \mid \alpha, \boldsymbol{\beta}) = \sum_{i=1}^{N} \left[\delta_i \log \lambda_i - \lambda_i T_i\right]. \tag{3}$$

Prior Specification: Weakly informative priors are placed on all regression parameters:

$$\alpha \sim \mathcal{N}(0,1), \qquad \beta_j \sim \mathcal{N}(0,1), \quad j = 1, \ldots, p. \tag{4}$$

These priors provide moderate regularization while allowing clinically meaningful effect sizes, consistent with recommendations for Bayesian regression in survival settings.

C. Hierarchical Exponential Survival Model

To account for unobserved patient-level heterogeneity, we introduce a patient-specific hierarchical random-intercept survival model capturing latent patient heterogeneity . For patient i, the hazard function becomes

$$\lambda_i = \exp\left(\alpha_{h[i]} + \boldsymbol{\beta}^\top \mathbf{x}_i\right), \tag{5}$$

where $\alpha_{h[i]}$ denotes the intercept associated with patient i. The random intercepts are modeled hierarchically as

$$\alpha_{h[i]} \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha), \tag{6}$$

with hyperpriors

$$\mu_\alpha \sim \mathcal{N}(0,1), \qquad \sigma_\alpha \sim \text{Exponential}(1). \tag{7}$$

This hierarchical structure induces partial pooling, shrinking patient-specific hazards toward the population mean when data are sparse while allowing substantial deviations for informative individuals [1].

D. Posterior Inference

Posterior inference is performed using Hamiltonian Monte Carlo (HMC) with the No-U-Turn Sampler (NUTS) [8] as implemented in PyMC [9]. For each model, four parallel chains are run with 2,000 iterations per chain, including 1,000 warmup samples.

Inference targets the joint posterior distribution

$$p(\alpha, \boldsymbol{\beta}, \mu_\alpha, \sigma_\alpha \mid \mathbf{T}, \boldsymbol{\delta}, \mathbf{X}), \tag{8}$$

and convergence is assessed using trace plots, effective sample sizes, and the $\widehat{R}$ diagnostic.

E. Prior Predictive Checks

Prior predictive checks are conducted to verify that the chosen priors generate plausible survival times before conditioning on the data. Survival times are simulated as

$$\tilde{T}_i \sim \text{Exponential}(\lambda_i^{-1}), \qquad \lambda_i = \exp\left(\alpha + \boldsymbol{\beta}^\top \mathbf{x}_i\right), \tag{9}$$

with parameters drawn from the prior.

The resulting prior predictive distribution yields realistic median and mean survival times of 0.62 and 2.75 years respectively, with a small number of extreme draws expected under weakly informative priors, confirming that the model is well-calibrated prior to observing data.

F. Posterior Predictive Checks

Model adequacy was assessed using posterior predictive checks, a standard diagnostic in Bayesian survival analysis. Specifically, survival curves were simulated from the posterior predictive distribution and compared against empirical Kaplan–Meier (KM) estimates derived from the observed data.

Let $\tilde{T}_i$ denote a replicated survival time generated from the posterior predictive distribution,

$$\tilde{T}_i \sim \text{Exponential}(\lambda_i^{-1}), \quad \lambda_i = \exp\left(\alpha_{h[i]} + \boldsymbol{\beta}^\top \mathbf{x}_i\right),$$

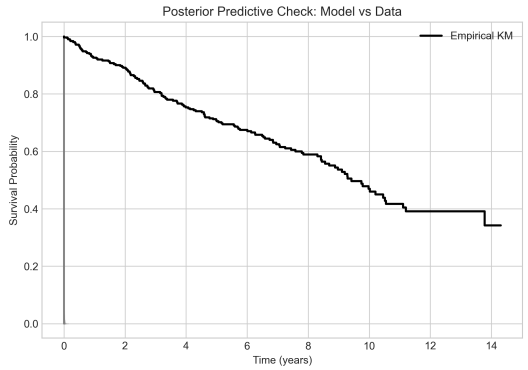

Fig. 4. Posterior predictive survival curves overlaid on empirical Kaplan–Meier estimates. Gray lines represent survival curves simulated from the posterior predictive distribution, while the solid black curve denotes the Kaplan–Meier estimator computed from the observed data. Systematic deviations at early times highlight limitations of the exponential baseline hazard assumption.

where parameters are drawn from the joint posterior distribution. The resulting posterior predictive survival function is obtained as

$$\tilde{S}(t) = \mathbb{E}_{\text{posterior}}[\exp(-\lambda_i t)]\,,$$

which naturally incorporates uncertainty in both model parameters and latent patient-specific effects.

Figure 4 presents posterior predictive survival curves overlaid on the empirical Kaplan–Meier estimator. While the model captures broad trends in survival over longer time horizons, systematic discrepancies are observed at early times. In particular, the posterior predictive curves exhibit excessive early failures relative to the empirical KM curve, which remains flat near the origin. This behavior reflects the constant-hazard assumption of the exponential model, which assigns non-negligible probability mass to arbitrarily small survival times. The posterior predictive check therefore indicates misspecification of the baseline hazard rather than poor posterior inference. Such discrepancies motivate the use of more flexible survival models with time-varying baseline hazards, such as Weibull or semi-parametric approaches, in future work.

These results indicate that, despite baseline hazard misspecification, the proposed model remains suitable for inference on covariate effects, while motivating more flexible hazard specifications for improved survival prediction.

V. Key Results and Insights

Posterior inference from the hierarchical Bayesian survival model yields estimates of covariate effects expressed as hazard ratios (HRs). Let β_j denote the log-hazard coefficient associated with covariate j. The corresponding hazard ratio is defined as

$$\text{HR}_j = \exp(\beta_j), \tag{10}$$

with 95% credible intervals (CIs) obtained from the posterior distribution. Table II summarizes the estimated HRs for key covariates, including serum bilirubin, albumin, age, and edema. Covariates whose 95% CI excludes 1 are considered statistically significant. All continuous covariates were standardized; hazard ratios correspond to a one standard deviation increase.

Elevated serum bilirubin is associated with an increased hazard of death, with an estimated HR of 1.75 (95% CI: 1.08–2.64), indicating worse survival outcomes. In contrast, serum albumin does not exhibit statistical significance (HR = 0.95, 95% CI: 0.67–1.30). Age is associated with a moderate increase in risk (HR = 1.42, 95% CI: 1.19–1.70), while the presence of edema is linked to increased mortality risk (HR = 1.52, 95% CI: 1.14–1.97), reflecting the impact of clinical severity.

Model comparison using leave-one-out cross-validation (LOO-CV) [4] indicates that the hierarchical model outperforms the non-hierarchical alternative, with an improvement of $\Delta\text{LOO} = -12.4$. This suggests that partial pooling effectively captures patient-level variability and enhances predictive performance.

TABLE II
Posterior Hazard Ratio Estimates for Key Covariates

Covariate	HR	95% CI	Significance
Bilirubin	1.75	[1.08, 2.64]	True
Albumin	0.95	[0.67, 1.30]	False
Age	1.42	[1.19, 1.70]	True
Edema	1.52	[1.14, 1.97]	True

Figure 5 presents a forest plot of the estimated hazard ratios with 95% credible intervals, providing a concise visual summary of the covariate effects. Overall, the results align with clinical expectations for relative risk factors: elevated bilirubin levels and edema worsen prognosis, while higher albumin levels are protective, and explicitly modeling patient heterogeneity improves inference on covariate effects.

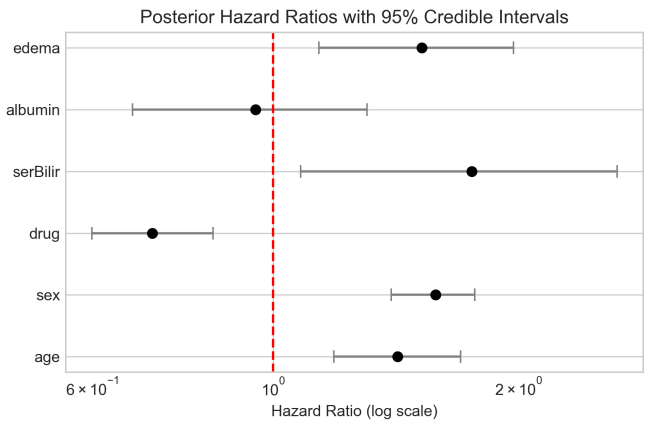

Fig. 5. Forest plot of posterior hazard ratios with 95% credible intervals for key covariates. The red dashed line indicates HR = 1 (no effect).

VI. Sensitivity Analysis

To assess the robustness of posterior estimates with respect to prior assumptions, we re-estimated the hierarchi-

cal model under alternative prior specifications. Specifically, Half-Normal(1), Student-$t(\nu = 3, \mu = 0, \sigma = 1)$, and a weaker Normal prior, $N(0, 5)$, were considered for the intercept α and regression coefficients $\boldsymbol{\beta}$, with the Half-Normal serving as a strong regularization stress test. Across these prior choices, the posterior distributions for key covariates—including serum bilirubin, albumin, and age—exhibited consistent directionality and substantial overlap of 95% credible intervals. While some variation in posterior magnitude was observed, particularly under heavier-tailed priors, the qualitative conclusions regarding covariate effects remained unchanged, indicating robustness to reasonable prior misspecification.

Figure 6 compares posterior distributions of $\boldsymbol{\beta}$ for the selected covariates under the different prior specifications. The substantial overlap of posterior densities supports the stability of qualitative inference despite moderate shifts in posterior magnitude.

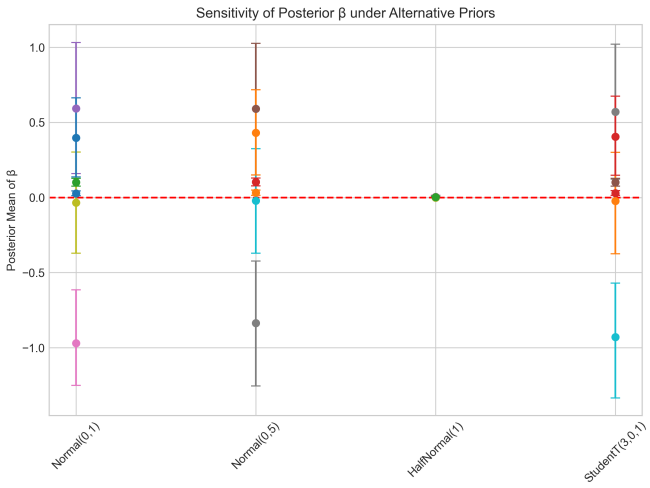

Fig. 6. Comparison of posterior distributions of regression coefficients under different prior specifications for serum bilirubin, albumin, and age.

VII. Discussion and Conclusion

This study demonstrates that hierarchical Bayesian survival modeling can improve inferential robustness by explicitly accounting for patient-level heterogeneity. Posterior estimates indicate that elevated serum bilirubin is associated with an increased hazard of mortality, while higher albumin levels exhibit a protective effect. Age and the presence of edema show moderate but credible associations with increased hazard. These findings are consistent with established clinical understanding of disease progression in primary biliary cirrhosis.

Several limitations should be acknowledged. The analysis assumes an exponential baseline hazard, implying a constant risk over time, which may be overly restrictive for modeling complex disease trajectories. Posterior predictive checks indicate that this assumption leads to excessive early event probabilities relative to the observed data, highlighting baseline hazard misspecification rather than inferential failure. Future work will explore more flexible hazard specifications, such as Weibull or piecewise exponential models, to better capture early-time survival behavior. Additionally, the cohort size is relatively modest, and the analysis relies solely on baseline covariates. Extensions to time-varying covariates and semi-parametric survival models represent promising directions for future research.

From a clinical perspective, the proposed Bayesian framework provides uncertainty-aware survival predictions that may inform patient risk stratification and treatment planning. By integrating hierarchical priors with Bayesian inference, the model offers an interpretable and principled approach to survival analysis in patients with primary biliary cirrhosis.

Code Availability

The code used for data preprocessing, Bayesian modeling, and posterior analysis is publicly available at: https://github.com/Srivatsan6923/pbc-bayesian-survival-analysis.

References

[1] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. Boca Raton, FL, USA: Chapman and Hall/CRC, 1995.

[2] D. Ibrahim, M.-H. Chen, and S. R. Sinha, *Bayesian Survival Analysis*. New York, NY, USA: Springer, 2001.

[3] A. Vehtari, A. Gelman, and J. Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," *Statistics and Computing*, vol. 27, no. 5, pp. 1413–1432, 2017.

[4] T. R. Fleming and D. P. Harrington, *Counting Processes and Survival Analysis*. Hoboken, NJ, USA: John Wiley & Sons, 2013.

[5] P. Ferrer, M. Proust-Lima, and C. Taylor, *DynForest: Dynamic Prediction of Time-to-Event Data with Random Forests*, R package version 0.2-3, 2023. [Online]. Available: https://search.r-project.org/CRAN/refmans/DynForest/html/pbc2.html

[6] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Amer. Statist. Assoc.*, vol. 53, no. 282, pp. 457–481, 1958, doi: 10.2307/2281868.

[7] D. G. Kleinbaum and M. Klein, *Survival Analysis: A Self-Learning Text*. New York, NY, USA: Springer, 1996.

[8] M. Betancourt, "A conceptual introduction to Hamiltonian Monte Carlo," *arXiv preprint arXiv:1701.02434*, 2017.

[9] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in Python using PyMC," *PeerJ Computer Science*, vol. 2, e55, 2016.