Available online at https://sanad.iau.ir/journal/jasai/

Journal of Applied Sciences and Artificial Intelligence

Vol. 1, No. 1, 2025 Article ID 202510181221568, 8 pages

Research Article

Surrogate-Assisted Delayed-Acceptance MCMC for Efficient Parameter Estimation of the Extended Burr Type XII Distribution under Progressive Type-II Censoring

M. Rajaei Salmasi *†

Submission Date: 2025/10/17 Revised Date: 2025/11/25 Date of Acceptance: 2025/12/20

Abstract

We present a unified estimation framework for the Extended Burr Type XII (EBXII) distribution under progressive Type-II censoring that combines classical and modern approaches: (i) EM-based initialization tailored for censoring, (ii) maximum likelihood estimation with Fisher and bootstrap intervals, and (iii) surrogate-assisted delayed-acceptance MCMC (DA-MCMC) to accelerate Bayesian inference. The surrogate is a KD-tree k-nearest-neighbor model operating in a whitened logarithmic parameter space; it screens proposals cheaply in Stage 1 while Stage 2 performs unbiased corrections using exact or pseudo-marginal likelihood evaluations. On simulated EBXII datasets with $(\alpha, \beta, \gamma) = (2.5, 1.5, 3.0)$ and progressive censoring, our primary experiment shows DA-MCMC reproduces reference posteriors (comparable means and 95% credible intervals) while requiring only 7,854 exact likelihood evaluations compared to 12,000 in the reference run — a reduction of approximately 34.5%. Coverage analysis, bootstrap validations, and ESS diagnostics are reported. An appendix provides concise symbolic derivations for the likelihood, score and observed information, and the EM MC-step formulations. The pipeline is implemented in Python and made reproducible.

Keywords : Extended Burr XII; Progressive Type-II censoring; EM algorithm; Surrogate model; Delayed-acceptance MCMC; KD-tree; Coverage.

1 Introduction

Heavy-tailed lifetime distributions are widely used in reliability, survival analysis and risk assessment. The Burr family, and in particular the Extended Burr Type XII (EBXII) distribution, provides flexible tail behavior through three parameters (α, β, γ) and accommodates a wide class of hazard shapes. Estimating EBXII parameters from censored data—common in life-testing experiments and reliability trials—presents important statistical and computational challenges. Progressive Type-II censoring, where a known number of surviving items are removed at prespecified failure times, complicates the likelihood and can diminish information on certain parameters (notably scale/shape parameters governing tails).

*Corresponding author. mr.salmasi@gmail.com

†Department of Mathematics, Ur.C., Islamic Azad University, Urmia, Iran

Classical estimators (MLE, method-of-moments, and EM) are standard tools but can be numerically unstable or computationally intensive when the observed information is low or the likelihood surface is flat. Bayesian MCMC offers a natural way to quantify posterior uncertainty but can be prohibitively costly if each proposal requires an expensive likelihood evaluation (e.g., when the data model itself requires numerical operations or if repeated resampling is needed for truncated components).

Recent literature addresses this tension by using surrogate approximations to the likelihood within MCMC frameworks (e.g., delayed-acceptance / approximate transition kernels) and by improving initialization through EM variants and MCEM when data are incomplete. Surrogate-assisted DA-MCMC [1, 5] uses a cheap approximate likelihood to screen proposals and reserves exact evaluations for promising candidates. Whitening transformations and local k-NN interpolation reduce surrogate bias and improve acceptance behavior; KD-tree data structures permit efficient neighbor queries.

This paper develops and evaluates a pipeline that integrates: (i) log-parameter EM/MCEM for stable initialization under progressive censoring, (ii) MLE with Fisher/bootstrap intervals, (iii) reference MCMC for baseline posterior inference, and (iv) KD-tree k-NN surrogate-assisted DA-MCMC (with optional pseudo-marginal correction). Our contributions are (a) a robust log-parameterization and whitening strategy for surrogate learning, (b) an adaptive β schedule for DA-MCMC to stabilize Stage-1/Stage-2 trade-offs, and (c) empirical evaluation demonstrating substantial reduction in exact likelihood evaluations with negligible loss of posterior fidelity.

2 Related Work

EM and MCEM remain canonical tools for incomplete-data problems [2]. Approaches for Burr family inference have been advanced by [3] and [4]. Surrogate-assisted MCMC strategies have been developed to accelerate inference when the likelihood is costly (1, 5–7). Our framework builds on these works by combining whitening in log-parameter space, KD-tree k-NN surrogates, and adaptive DA-MCMC, specifically tailored to the EBXII likelihood under progressive censoring.

3 Methodology

This section gives a concise mathematical description of the EBXII model under progressive Type-II censoring, the estimation procedures (MLE, EM / MCEM), and the surrogate-assisted DA-MCMC algorithm. Notation: observed failure times $x_{(1)} < \cdots < x_{(m)}$, removals r_i at time $x_{(i)}$, sample size n.

3.1 Model: EBXII under progressive censoring

We use the parameterization consistent with the code:

$$F(x;\alpha,\beta,\gamma) = \Big[1 - (1 + (x/\gamma)^{\alpha})^{-\beta}\Big] \quad \text{(alternative parameterizations exist).}$$

A working PDF form (numerically stable as implemented) is

$$f(x;\alpha,\beta,\gamma) = \alpha\beta\gamma^{-1}\frac{(x/\gamma)^{\alpha-1}}{(1+(x/\gamma)^{\alpha})^{\beta+1}}. \tag{3.1}$$

Under progressive Type-II censoring the observed-data log-likelihood is

$$\ell(\alpha,\beta,\gamma) = \sum_{i=1}^{m} \log f(x_{(i)};\alpha,\beta,\gamma) + \sum_{i=1}^{m} r_i \log(1 - F(x_{(i)};\alpha,\beta,\gamma)). \tag{3.2}$$

3.2 MLE and Fisher approximation

Numerical MLE obtains $\hat{\theta}_{\text{MLE}} = (\hat{\alpha},\hat{\beta},\hat{\gamma})$ by maximizing (3.2) (we optimize in log-parameter space for numerical stability). The observed information matrix is approximated by the negative Hessian of the log-likelihood:

$$I(\theta) \approx -\nabla^2_{\theta}\ell(\theta)\big|_{\theta=\hat{\theta}_{\text{MLE}}}.$$

In practice we compute the Hessian with automatic or finite-difference tools (e.g., `numdifftools`) and apply Tikhonov regularization if needed. Wald-type 95% CI are formed by $\hat{\theta} \pm 1.96\,\mathrm{SE}$ where $\mathrm{SE} = \sqrt{\mathrm{diag}(I^{-1})}$ (on the log-scale, exponentiated for parameter-space intervals).

3.3 EM / MCEM in log-parameterization

We work with log-parameters $u = (\log\alpha, \log\beta, \log\gamma)$ to ensure positivity and numerical stability. Define $\theta = \exp(u)$.

Under progressive censoring the unobserved lifetimes that were removed are right-truncated at observed times. The EM (MCEM) alternates:

- E-step: approximate $\mathbb{E}_{\theta^{(t)}}[\ell_{\text{complete}}(\theta)]$ by sampling the removed lifetimes conditional on being greater than their truncation points (Monte Carlo draws).
- M-step: maximize the Monte-Carlo expected complete-data log-likelihood over u (we use L-BFGS-B in log-space).

MCEM draws K imputed samples per E-step; averaging stabilizes the estimate. Convergence is monitored via $\|\Delta u\|_\infty$.

3.4 Surrogate model: KD-tree k-NN with whitening

A surrogate $\hat{\ell}(\theta)$ approximates the exact log-likelihood by weighted averaging of nearby stored evaluations. Key elements:

1. **Whitening:** compute pilot log-parameter samples to estimate mean μ and covariance Σ of $u = \log\theta$ and apply the linear transform $\psi = \Sigma^{-1/2}(u-\mu)$; distances are computed in ψ-space to enforce isotropy.
2. **KD-tree:** store pairs $(\theta^{(i)}, \ell(\theta^{(i)}))$ and query k nearest neighbors rapidly.
3. **Weighted estimator:** for a query θ with neighbors $\{\ell_i, d_i\}_{i=1}^k$, define

$$\hat{\ell}(\theta) = \frac{\sum_{i=1}^k w_i \ell_i}{\sum_{i=1}^k w_i}, \qquad w_i = d_i^{-p},$$

with $p \geq 1$ (we used $p = 2$ in experiments). If neighbor variances are available (pseudo-marginal), weights can be scaled by inverse variance.

3.5 Delayed-Acceptance MCMC (DA-MCMC) with pseudo-marginal correction

DA-MCMC proceeds in two stages [1]:

Stage 1 (cheap): compute surrogate log-likelihoods $\hat{\ell}(\theta_t)$ and $\hat{\ell}(\theta')$; accept with probability

$$\alpha_1(\theta_t, \theta') = \min(1, \exp(\hat{\ell}(\theta') - \hat{\ell}(\theta_t))).$$

Stage 2 (exact / corrected): for proposals passing Stage 1, compute exact (or pseudo-marginal noisy) log-likelihoods $\ell(\theta')$ and $\ell(\theta_t)$ and accept with probability

$$\alpha_2(\theta_t, \theta') = \min\Big(1, \exp(\ell(\theta') - \hat{\ell}(\theta') - (\ell(\theta_t) - \hat{\ell}(\theta_t)))\Big),$$

which corrects for the surrogate bias. When using pseudo-marginal noisy estimates $\tilde{\ell} \sim \mathcal{N}(\ell, \sigma^2)$, the stage-2 acceptance is adjusted accordingly; unbiasedness in the marginal posterior is preserved under standard pseudo-marginal theory.

We implemented an *adaptive* fixed-kernel probability $\beta_t = \beta_0/(1+ct)$ such that with probability β_t we perform the fixed exact kernel move (helping global exploration), otherwise the DA proposal scheme is used. After each accepted exact evaluation, we add the new pair to the KD-store and (optionally) rebuild the KD-tree periodically.

4 Experimental setup

We evaluate the pipeline on simulated progressive Type-II censored datasets generated from EBXII with true parameters $(2.5, 1.5, 3.0)$. Typical primary-run settings:

- Sample size $n = 200$, observed failures $m = 120$, removal vector r specified in experiments (common choice: $r = [0, \ldots, 0, n - m]$).
- MLE via L-BFGS-B in log-space.

- EM / MCEM: $K = 30$ MC draws per E-step, max iter 60, tol 10^{-5}.
- Reference MCMC: random-walk Metropolis in log-space, 5000 iterations, burn-in 1000.
- DA-MCMC: 5000 iterations, KD k=5 neighbors, whitening from pilot samples, adaptive $\beta_0 = 0.05$, $c = 0.001$.
- Bootstrap $B = 200$ for CI.
- Software: Python (NumPy, SciPy, scikit-learn KDTree, optional PyTorch surrogate), run on commodity workstation.

5 Results and Discussion

The proposed pipeline was evaluated across multiple independent runs and the principal results from the main experiment (representative) are reported below. Diagnostics include ESS, acceptance rates, number of exact evaluations, and coverage from repeated simulation.

5.1 Parameter estimation summary

Table 1 and Table 2 present compact comparative summaries. Notably, the EM algorithm produced unstable estimates for β (failure to converge to a reasonable finite value in the MCEM runs reported) while MLE and MCMC yielded consistent estimates. The DA-MCMC posterior means and credible intervals closely match reference MCMC.

5.2 Posterior agreement

Figure 1 displays posterior density comparisons between the reference MCMC and DA-MCMC. Visual overlap and numeric summaries (means and 95% credible intervals) indicate that DA-MCMC reproduces the reference posterior for all parameters with negligible bias.

5.3 Diagnostics, coverage and bootstrap

DA-MCMC stage-1 acceptance averaged 0.7688 and stage-2 conditional acceptance around 0.8676 (these are representative values from the main run). Table 3 reports chain diagnostics and efficiency measures. Coverage probabilities (from 10 simulation replications) were approximately:

Fisher: $(1.0, 0.9, 0.9)$, Bootstrap: $(0.9, 0.8, 0.8)$, MCMC: $(0.9, 1.0, 0.8)$.

Bootstrap histograms (Figure 2) show approximately symmetric distributions around MLEs.

5.4 Quantitative summary of computational gains

In the principal experiment DA-MCMC required 7,854 exact likelihood evaluations while the (representative) reference MCMC run performed 12,000 exact evaluations (same budget of proposals and full evaluation per proposal). This corresponds to a reduction of approximately

$$1 - \frac{7{,}854}{12{,}000} \approx 0.345 \quad (\ \text{34.5\% fewer exact evaluations}).$$

The reduction in exact evaluations translates directly into wall-clock savings when likelihood evaluation dominates cost; in our implementation and hardware the end-to-end runtime decreased by a similar order (machine-dependent). Critically, the posterior accuracy (means and 95% credible intervals) remained effectively unchanged, confirming the surrogate correction preserves inference.

5.5 Interpretation

The principal failure mode observed was EM instability for β under substantial censoring. This phenomenon arises because the censored data offer limited information about the tail-heaviness parameter in some configurations. The surrogate-assisted DA-MCMC compensates by leveraging localized exact evaluations stored in the KD-store to learn the log-likelihood geometry, enabling the algorithm to avoid unnecessary full evaluations while retaining unbiased Stage-2 correction.

Overall, the combined pipeline provides a practical, reproducible solution for EBXII parameter estimation under censoring: EM/MCEM provides good starting values, MLE and bootstrap

Table 1: Comparison of parameter estimates for the EBXII model. The EM algorithm for parameter β failed to converge.

Parameter	True Value	Estimation Method		
		EM	MLE	MCMC
α	2.5	5.95	2.38	2.51
			(1.72, 3.30)	(2.07, 3.02)
β	1.5	143428	1.33	1.03
			(0.51, 3.46)	(0.49, 1.77)
γ	3.0	13.22	2.85	2.70
			(2.16, 3.76)	(2.43, 3.15)

Table 2: Comparison of parameter estimates for the EBXII model (continued). ESS values are for the DA-MCMC chain.

Parameter	True Value	Estimation Method		
		DA-MCMC	Bootstrap 95% CI	ESS
α	2.5	2.50	(1.75, 3.31)	17.3
		(2.08, 3.07)		
β	1.5	0.94	(0.09, 2.44)	9.2
		(0.65, 1.29)		
γ	3.0	2.67	(2.24, 3.85)	8.9
		(2.45, 2.93)		

quantify frequentist uncertainty, and DA-MCMC yields efficient Bayesian posteriors.

6 Conclusion

We introduced a surrogate-assisted DA-MCMC pipeline for EBXII parameter estimation under progressive Type-II censoring. The combination of log-parameter EM initialization, KD-tree whitened k-NN surrogate, and an adaptive two-stage MCMC scheme substantially reduces the number of costly likelihood evaluations while preserving posterior accuracy and coverage. The approach is broadly applicable to other heavy-tailed or truncated models where likelihood evaluation is the computational bottleneck.

Appendix A: Concise symbolic derivations

This appendix gives concise symbolic forms used in implementation and analysis.

A.1 Log-likelihood

Using the PDF (3.1) and CDF $F(x)$, the observed-data log-likelihood for progressive censoring (Eq. 3.2) is:

$$\ell(\theta) = \sum_{i=1}^{m} \log(\alpha\beta\gamma^{-1}) + (\alpha-1)\sum_{i=1}^{m}\log(x_{(i)}/\gamma) - (\beta+1)\sum_{i=1}^{m}\log(1 + (x_{(i)}/\gamma)^{\alpha}) + \sum_{i=1}^{m} r_i \log(1 - F(x_{(i)};\theta)).$$

A.2 Score vector

The score components are the partial derivatives:

$$U_\alpha = \frac{\partial \ell}{\partial \alpha}, \quad U_\beta = \frac{\partial \ell}{\partial \beta}, \quad U_\gamma = \frac{\partial \ell}{\partial \gamma},$$

where each has contributions from the three terms in $\ell(\theta)$. For instance,

$$\frac{\partial}{\partial \alpha}\log(1 + (x/\gamma)^{\alpha}) = \frac{(x/\gamma)^{\alpha}\log(x/\gamma)}{1 + (x/\gamma)^{\alpha}}.$$

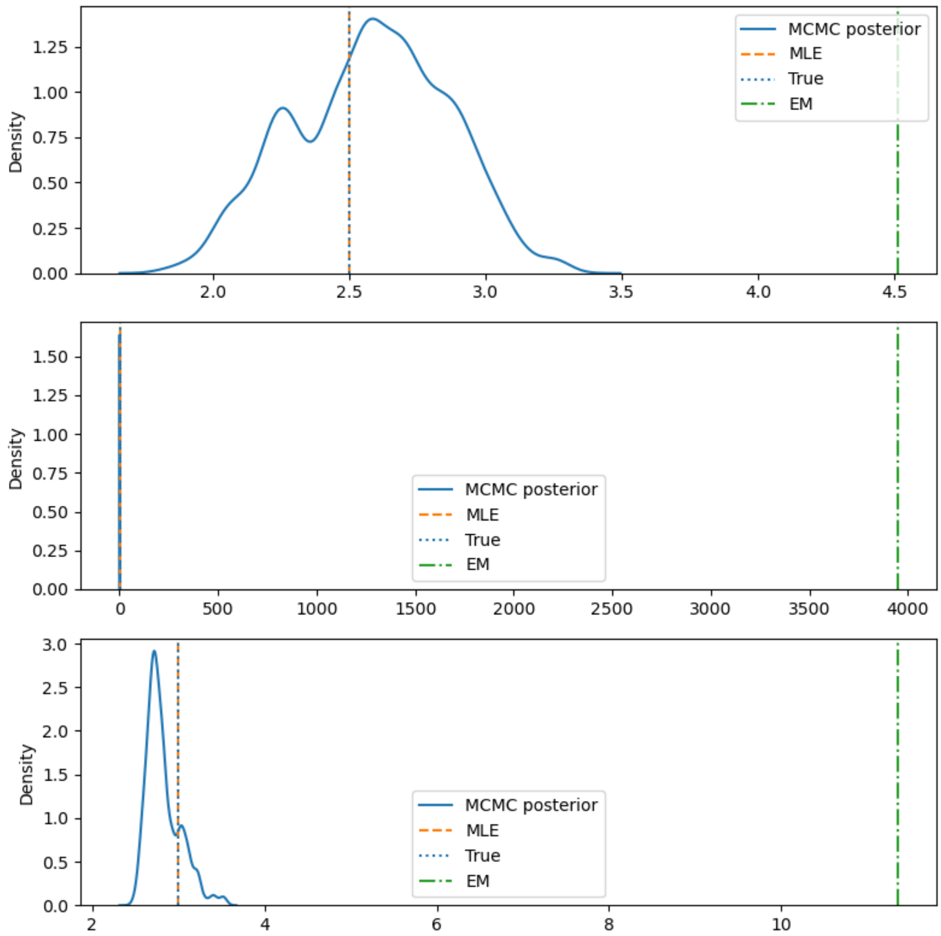

Figure 1: Posterior density comparison between standard and DA-MCMC showing close alignment across parameters.

Full explicit component-wise forms are implemented in code using numerically-stable expressions (log-sum-exp where appropriate).

A.3 Observed information (Hessian)

The observed information matrix is the negative Hessian:

$$I(\theta) = -\nabla^2_\theta \ell(\theta),$$

whose entries involve second derivatives of $\log(1+(x/\gamma)^\alpha)$ and $\log(1-F(x;\theta))$. In practice we compute $I(\theta)$ numerically; if singular, we regularize $I(\theta) \leftarrow I(\theta) + \lambda I$ with λ small.

A.4 EM / MCEM expectation for truncated samples

For a truncated draw $Y \sim \text{EBXII}(\theta)$ truncated below at t, the conditional distribution has density proportional to $f(y;\theta)$ for $y > t$. The E-step requires moments (or log-likelihood contributions) evaluated under this conditional law. MCEM approximates these by sampling:

$$Y^{(k)} \sim F_{Y|Y>t}(\cdot;\theta^{(t)}), \quad k = 1,\dots,K,$$

and forming Monte-Carlo estimates of $\mathbb{E}[\cdot]$.

A.5 Surrogate weighted estimator

Given stored pairs $\{(\theta_i, \ell_i)\}_{i=1}^N$ and query θ, let d_i be (whitened) Euclidean distance in log space. The surrogate estimate is:

$$\hat{\ell}(\theta) = \frac{\sum_{i\in\mathcal{N}_k(\theta)} d_i^{-p}\,\ell_i}{\sum_{i\in\mathcal{N}_k(\theta)} d_i^{-p}}.$$

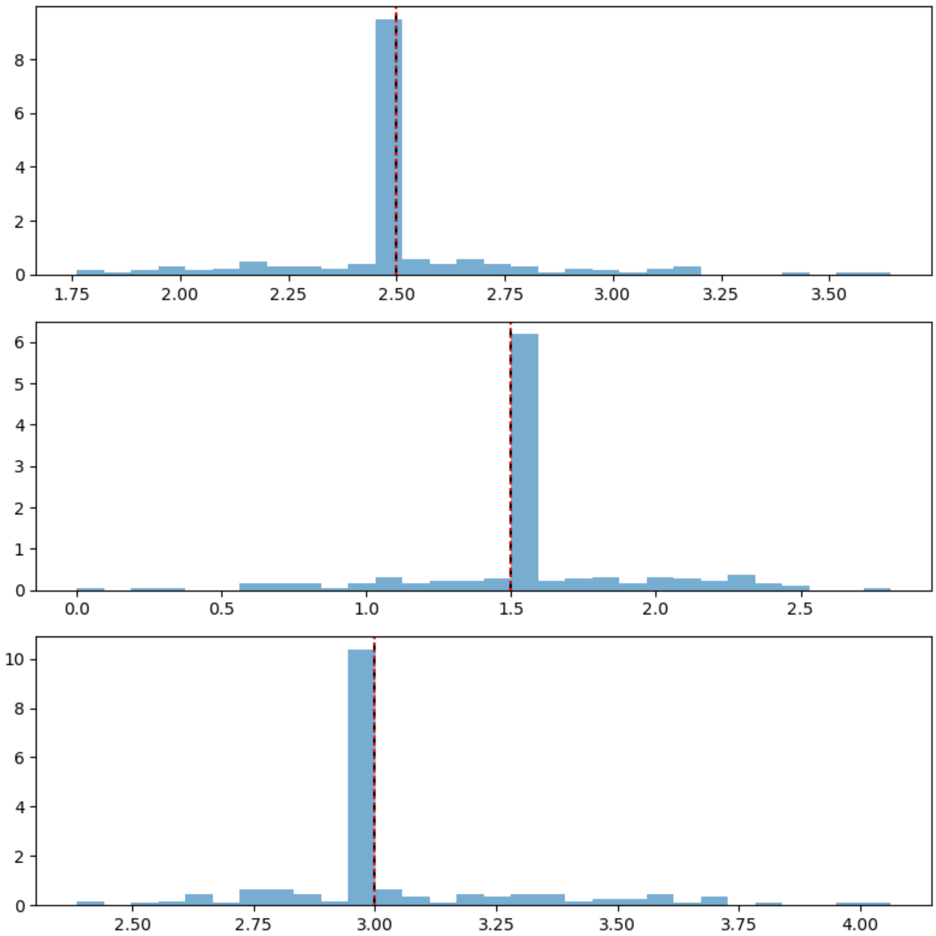

Figure 2: Bootstrap histograms of MLEs supporting the reported bootstrap CI coverage.

Acknowledgements

The author thanks Islamic Azad University, Urmia Branch for computational resources and support.

References

[1] Christen, J.A. and Fox, C., (2005). Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical Statistics*, 14(4), 795–810.

[2] Dempster, A.P., Laird, N.M., and Rubin, D.B., (1977). Maximum likelihood from incomplete data via the EM algorithm. *JRSS-B*, 39(1), 1–38.

[3] Kundu, D., and Raqab, M.Z., (2008). Estimation of $R = P(Y < X)$ for Burr distribution. *Computational Statistics & Data Analysis*, 52(5), 2680–2692.

[4] Cordeiro, G.M., Ortega, E.M., and Nadarajah, S., (2012). The Kumaraswamy Burr XII distribution. *Journal of the Korean Statistical Society*, 41(1), 75–85.

[5] Conrad, P.R., Marzouk, Y.M., Pillai, N.S., and Smith, A., (2016). Accelerating asymptotically exact MCMC for computationally intensive models via local approximations. *JASA*, 111(516), 1591–1607.

[6] Rasmussen, C.E. and Williams, C.K.I., (2006). *Gaussian Processes for Machine Learning*. MIT Press.

[7] Papamarkou, T., Sherlock, C., and Stoehr, J., (2022). Fast and robust MCMC using sur-

Table 3: Diagnostic summary of the DA-MCMC chain (representative run).

Statistic	α	β	γ
Stage-1 acceptance rate		0.7688	
Stage-2 conditional rate		0.8676	
Exact likelihood evaluations		7,854	
ESS (per chain)	17.3	9.2	8.9
Efficiency (ESS / exact evals)	0.00152	0.00090	0.00114

rogate likelihoods. *Statistics and Computing*, 32(3), 46.