The performance of annealed sequential Monte Carlo sampling as a joint variable selection and parameter estimation method in the linear (mixed) model setting

by

Quang Vuong

B.Sc., University of British Columbia, 2022

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

Declaration of Committee

Wei Lin
Lecturer, Statistics and Actuarial Science

Supervisor
Associate Professor, Statistics and Actuarial Science

Committee Member
Assistant Professor, Statistics and Actuarial Science

Examiner
Associate Professor, Statistics and Actuarial Science

Abstract

Variable selection is the statistical problem of identifying predictors that explain the variation in a response, which is challenging when the number of candidate predictors is large. Several existing frequentist and Bayesian methods can perform variable selection in high-dimensional settings with reasonable computation times. Modern Bayesian methods focus on sampling models from the posterior distribution on the model space while neglecting the estimation of model coefficients. Annealed sequential Monte Carlo (SMC) sampling is an appealing method that provides a weighted sample of models and model parameters simultaneously, thus simultaneously performing selection and estimation without further computational effort. We examine the selection and estimation performance of annealed SMC sampling for linear regression and mixed-effects models under different conditions to determine factors that impact its efficacy. We demonstrate that sample size, signal-to-noise

Dedication

I dedicate this thesis to my family, who have supported me all the way in my educational journey.

Acknowledgements

I would like to thank Professor Rachel Altman for being patient with my thesis research.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Variable selection is the statistical problem of identifying an active subset of predictors of a set of candidate predictors that explain the variation of a response variable. Variable selection methods have been developed and studied for decades and continue to be an active field of research, especially in modern, high-dimensional contexts. Stepwise selection is a conceptually straightforward method that is often used in scientific contexts, but it can be highly unstable with respect to the dataset and requires data splitting to provide coefficient estimates with known theoretical distributions, effectively reducing the sample size. Simultaneous variable selection and coefficient estimation can be done using LASSO, and much work has been dedicated to developing confidence intervals for model coefficients [10]. Inference about model coefficients using this method is conditional on the selection of the corresponding predictors. We are interested in Bayesian analogues of these methods that allow us to simultaneously identify important predictors that explain a response variable and provide interval estimates with good marginal coverage properties for model coefficients, i.e., that do not require data splitting.

Bayesian model averaging (BMA) is an approach for incorporating uncertainty in statistical modelling that introduces a prior distribution on the space of possible models and derives the posterior distribution on this space using Bayes' rule. BMA is an intuitive method for averaging posterior distributions of quantities of interest over a range of plausible models, which is useful in situations where there is considerable model uncertainty. In addition, BMA usually leads to improved predictive performance over fitted models that have not been subjected to variable selection [9]. More formally, let denote the (possibly countably infinite) set of possible models M_k, where k is an arbitrary index and a prior distribution (M_k) is given over . For each model M_k, there are parameters $\boldsymbol{\theta}_k$ and the likelihood $f(D \;\; M_k \;\; \boldsymbol{\theta}_k)$ of the data D under M_k. If Δ is a quantity of interest, then we can average its posterior distribution across models in as follows [9]:

$$(\Delta \;\; D) = \sum_{M_k \in \mathcal{M}} (\Delta \;\; M_k \;\; D) \;\; (M_k \;\; D)$$

where

$$(M_k \quad D) = \frac{(D \quad M_k) \ (M_k)}{\sum_{M_j \in \mathcal{M}} \ (D \quad M_j) \ (M_j)}$$

and

$$(D \quad M_k) = \int \ (D \quad \boldsymbol{\theta}_k \ M_k) \ (\boldsymbol{\theta}_k \quad M_k) \, d\boldsymbol{\theta}_k$$

In the context of variable selection, the space of possible models typically consists of a family of models defined by considering all possible subsets of candidate predictors that may explain the response variable. Specific predictors can be selected by imposing criteria on the posterior model probabilities $(M_k \quad D)$, such as the predictors in the model with maximum a posteriori probability or predictors whose posterior inclusion probabilities, which are defined as the sum of posterior probabilities of models containing the predictor, are above some threshold. Model coefficients are interpreted as per their usual meaning for models that include the associated predictors and as degenerate point masses at 0 for models that exclude them. Their posterior distributions account for the possibility of including other predictors in the model. From each posterior distribution, both point estimates (the posterior mean) and interval estimates (credible intervals) are available for model coefficients, making BMA an attractive framework for the current problem.

of interest primarily focus on predictors with fixed e�ects.

We justify our choice to use annealed SMC sampling for our project as follows. Initially, we explored the use of reversible jump Markov chain Monte Carlo to obtain draws of the parameters and the model jointly from the posterior distribution, using a formulation that makes the algorithm equivalent to a Gibbs sampler [2]. However, in our preliminary investiis a preferable alternative to MCMC when the latter produces chains with poor convergence. SMC sampling can be adapted to sampling a single target distribution , leading to annealed SMC. Annealed SMC is simple to implement and does not rely on convergence arguments in the same sense as do MCMC methods, which tends to reduce computation time while still maintaining accuracy of the posterior distribution. By sampling both models and parameters, the need to compute the evidence for all possible models is avoided, thereby enabling the use of annealed SMC for more complicated models and more candidate predictors. However, its theoretical justification still depends on a convergence argument as

Chapter 2

Motivating example

Chapter 3

Methods

In this chapter, we describe the annealed SMC sampling algorithm applied to a family of linear mixed models indexed by candidate predictors that are not identically zero. We describe the models first and then the algorithm.

3.1 Linear Mixed Model

We restrict attention to the random-intercept model for simplicity. Let Y_{ij} and (x_{ij1} x_{ijp}) denote the response and vector of predictor variables, respectively, observed on the ith individual at time point j, i = 1 n, j = 1 n_i. We proceed with a joint model-parameter space in the sense of Barker & Link (2013) [2]. Let β = ($_0$ $_p$) be a column vector of regression coe□cients. We identify a model by = ($_1$ $_p$), where $_j$ 0 1 for j = 1 p as follows. Let $_1$ $_{|\zeta|}$ be the indices for which $_j$ is 1 (listed in increasing order), where indicates the L^1-norm of , and let ψ = ($_0$ $_p$) be an auxiliary parameter that is common across all models. We introduce ψ because in the annealed SMC sampling algorithm, we draw ψ as well as the model and other parameters. Consequently, we sample from a model-parameter space with constant dimension, which facilitates the construction of simple MCMC moves to be used in the algorithm. We then define

$$\boldsymbol{x}_{ij} = (1 \ x_{ij1} \quad x_{ijp})$$

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{x}_{i1} \\ \vdots \\ \boldsymbol{x}_{in_i} \end{pmatrix}$$

$$\beta_\zeta = \begin{pmatrix} _0 \\ _{m_1} \\ \vdots \\ _{m_{|\zeta|}} \end{pmatrix}$$

$$\boldsymbol{\psi}_\zeta = \begin{pmatrix} 0 \\ m_1 \\ \vdots \\ m_{|\zeta|} \end{pmatrix}$$

predictors addresses scaling concerns, and the choice of   is at the discretion of the analyst [8]. We comment on this issue further in section 5.

The interpretation of the model coefficients, the $_j$'s, when averaged across models is an important consideration when their estimates are of interest. In a single model that contains a continuous predictor x_j, $_j$ is the change in the mean response when x_j increases by one unit and other predictors are held constant. Given a prior distribution on the parameters of this model, the posterior distribution of $_j$ represents the effect of x_j after having adjusted for other predictors in the model. When the posterior distributions of $_j$ are averaged across models containing $_j$, the mode of the averaged distribution may be understood as the change in mean response when x_j increases by one unit adjusted by the remaining candidate predictors, which possibly have coefficients of zero. For models not containing x_j, $_j$ is set to be zero, meaning x_j has no effect on the mean response. Hence when the posterior distribution of $_j$ is averaged over all models, it has two components: onedn suggesting otherwise, ^{1} $^{T-1}$ are not true posterior distributions of (Θ) conditioned on y. But they are proper densities that depend on y.

The main idea is to use the intermediate distributions ^{t} to facilitate the sampling of ^{T} starting with draws from ^{0} in the following way. Let w_t

For this project, K_t is constructed to be a symmetric random walk Metropolis kernel that is ^{t}

Chapter 4

Simulation studies

The aim of our simulation studies is to examine the behaviour of the posterior distribution of the model parameters obtained using the annealed SMC method under different data generating mechanisms. In particular, we investigate the marginal posterior inclusion probabilities of candidate predictors, the coverage probabilities of 95% credible intervals for the regression coefficients, and the bias of the posterior means of the model coefficients using the Monte Carlo approximation $\hat{\ }^T$ from the last iteration of the annealed SMC algorithm.

We describe the data generating mechanisms and the results of each simulation study in the following subsections. We first perform a simulation study in a simple setting to confirm that the method can perform well. We then conduct a full factorial experiment in a more realistic linear regression setting to explore factors that influence the performance of the method. Finally, we test the method using a mixed-effects model with a random intercept.

In all simulation studies, the marginal posterior inclusion probabilities of each candidate predictor and the 95% credible intervals for the model coefficients are calculated using the Monte Carlo approximation $\hat{\ }^T$ i.e. the last distribution targeted by the annealed SMC sampling algorithm. For this project, a predictor is selected if its marginal posterior inclusion probability is at least 0.5. Other thresholds can be used, but our choice is inspired by the

4.1 Fixed-effects benchmark case

We first generate data according to a Gaussian linear model (i.e., no random effects) for n = 1000 subjects and p = 3 candidate fixed-effect predictors. Although variable selection

Table 4.1: Estimated marginal selection probabilities of candidate predictors in fixed-e□ects benchmark case.

[illegible]	[illegible]	[illegible]
x_1	0 962	0.009
x_2	0 938	0.011
x_3	0 072	0.012

Table 4.2: Estimated false and negative selection rates for the fixed-e□ects benchmark case.

[illegible]	[illegible]	[illegible]
False selection rate	0 036	0 129
Negative selection rate	0 050	0 150

Table 4.3: Estimated coverage probabilities of 95% credible intervals of model coe□cients for fixed-e□ects benchmark case.

[illegible]	[illegible]	[illegible]
$_1$	0 970	0.008
$_2$	0 944	0.010
$_3$	1 000	0.000

equal to the nominal level of 95% except in the case of $_3$. These statistics suggest that our implementation of annealed SMC is correct and is e□ective at selecting important predictors where 0_6 is the zero vector in ^{6} and

$$\Sigma = \sigma_X^2 \begin{pmatrix} 1 & \rho & & \rho \\ \rho & 1 & & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & & 1 \end{pmatrix} \quad {}_{6\times 6}$$

Let

$$Y_i = {}_0 + x_i{}^T \beta + {}_i$$

where $\beta^T = ({}_1 \quad {}_6)$, ${}_i \quad N($0 datasets. The parameters σ_X^2 and β are chosen to be similar to those in section 4.1.

For each run (combination of factor levels), 200 datasets (replicates) are generated, and

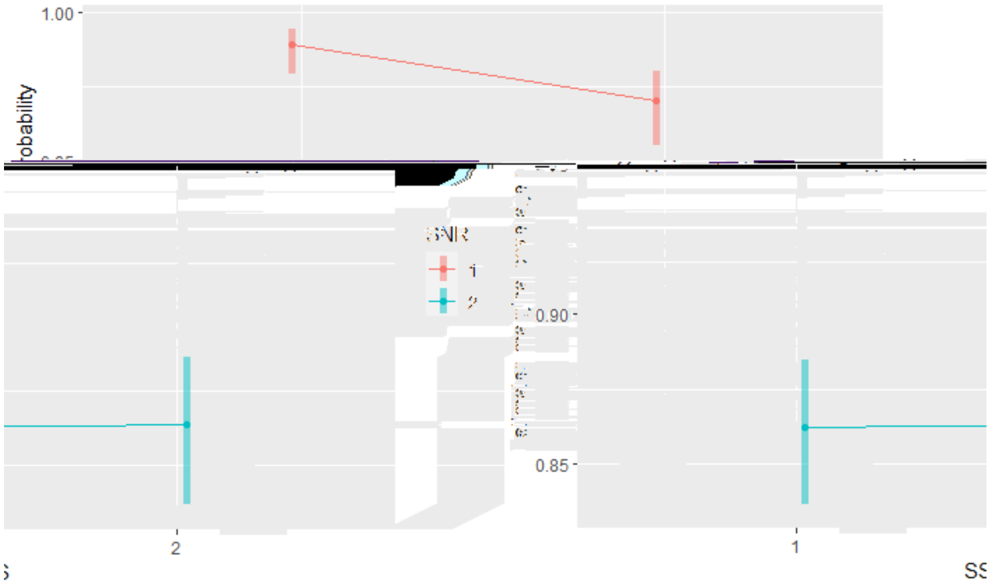

Figure 4.1: Interaction plot showing the effects of SNR and SS on the coverage probability of the 95% credible interval for β_1.

suggest that SNR and Cor have the most impact on the bias of the posterior means of the coefficients. Prop may have some effect, but the size of the impact of this factor differs for β_1 and β_2, so evidence for its impact is less conclusive. Figure 4.2 shows the effects of SNR and Cor on the bias of the posterior mean of β_1. The plot for β_4 is similar. The plots for β_2 and β_3 are also similar but are reflected about the line y = 0, which may be explained by the fact that the true values of β_1 and β_4 are set to be negative, while those of β_2 and β_3 are set to be positive. The plots then indicate that the posterior means are biased towards zero, which is expected because the prior mean of β

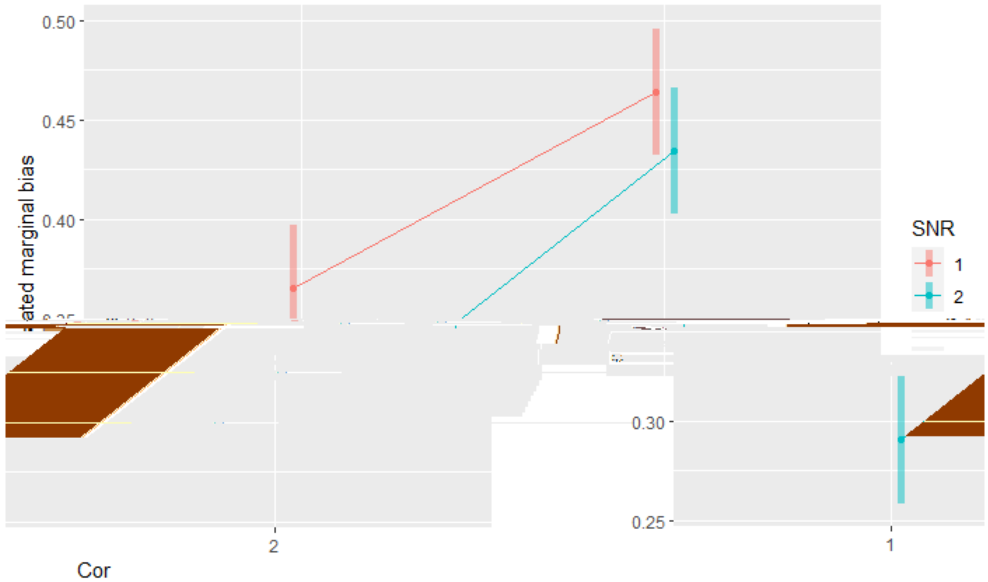

Figure 4.2: Interaction plot showing the effects of SNR and Cor on the bias of the posterior mean of $_1$ as an estimator for $_1$.

where

$$\text{logit}(p_i^{(3)}) = \boldsymbol{\eta}_i^{(3)}$$

and the $Y_i^{(3)}$'s are independent. We use analysis of deviance tables as before. We stratify analyses of x_3 and x_4 by Prop because the importance of x_3 and x_4 differs by Prop. The analysis over all datasets for these predictors would find that Prop has an outsized impact on the selection probabilities of x_3 and x_4, which is expected because x_3 and x_4 should have high selection probabilities when Prop is high and low selection probabilities when Prop is low. This effect could swamp the effects of other factors on the selection probability of these predictors.

We first describe the results for x_1 and x_2, which are important in all datasets. SNR and Cor have the most impact on the selection probabilities of x_1 and x_2. Figure 4.3 shows the effects of SNR and Cor on the selection probability of x_1. The corresponding plot for x_2 is similar. Higher SNR is associated with higher selection probability for these predictors, especially when Cor is low. Higher Cor is associated with lower selection probability for these predictors, particularly when SNR is high.

Next, we describe the results for x_3 and x_4, which are unimportant when Prop is low and important when Prop is high. The results suggest that SNR and Cor have the most impact on the selection probabilities of x_3 and x_4 both when Prop is low and when Prop is high. Figure 4.4 depicts the interaction plot showing the effects of these factors on the selection probability of x_3 when Prop is low; the corresponding plot for x_4 is similar. When Prop is low, higher SNR is associated with lower selection probability of these predictors, with the difference more pronounced when Cor is high. The interaction plots showing the effects of SNR and Cor on the selection probabilities of x_3 and x_4 when Prop is high are similar to the corresponding plot for x_1. Specifically, when Prop is high, higher SNR is associated with higher selection probabilities for x_3 and x_4 with the difference being more pronounced when Cor is low, and higher Cor is associated with lower selection probabilities for these predictors, more so when SNR is high.

Finally, we describe the results for x_5 and x_6, which are unimportant in all datasets. The analysis suggests that SNR and Prop have the most impact on the selection probabilities of x_5 and x_6. Figure 4.5 shows the effects of SNR and Prop on the selection probability of x_5; the corresponding plot for x_6 is similar. We note that due to the fact that the coefficients are of the same size for both levels of Prop and to the definition of SNR, the true value of σ^2 is larger when Prop is higher (for a fixed level of SNR). Therefore, the effect of Prop is confounded with the effect of a larger value of σ^2. Generally, higher SNR is associated

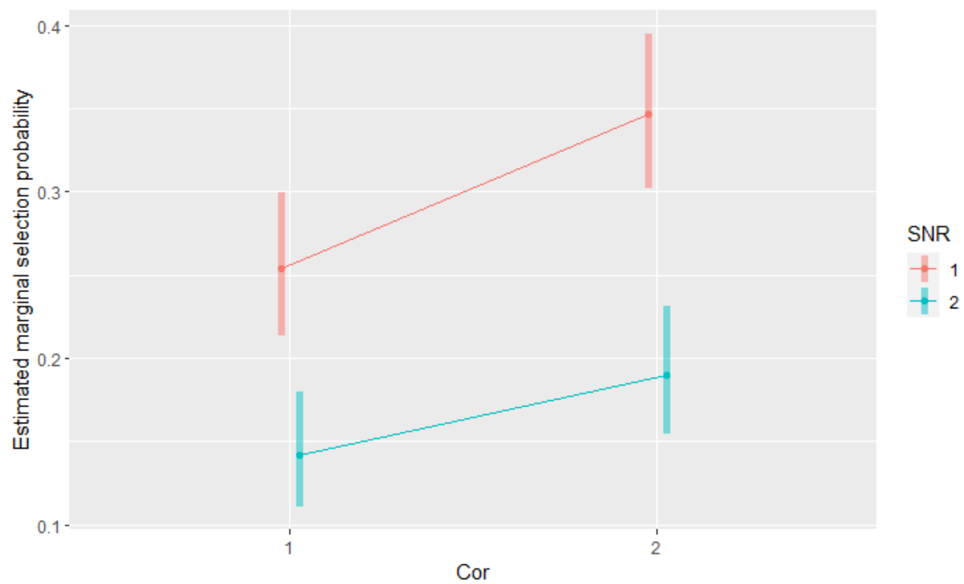

Figure 4.4: Interaction plot showing the eects of SNR and Cor on the selection probability of x_3 when Prop is low.

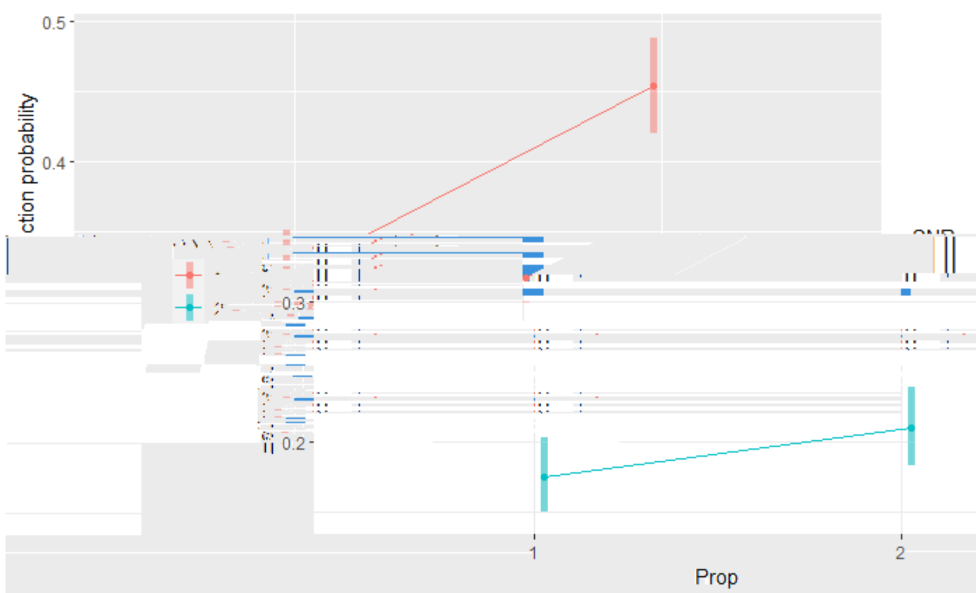

Figure 4.5: Interaction plot showing the e ects of SNR and Prop on the selection probability of x_5.

We model the number of selected predictors that are not important, F_i, conditional on the total number of selected predictors, S_i, using a quasi-binomial regression approach with

$$\mathbb{E}[F_i \quad S_i] = S_i p_i^{(4)}$$

and

$$\mathrm{Var}(F_i \mid S_i) = \; S_i p_i^{(4)}(1 - p_i^{(4)})$$

where

$$\mathrm{logit}(p_i^{(4)}) = \boldsymbol{\eta}_i^{(4)}$$

and the F_i's are independent. In this model, $p_i^{(4)}$ represents the false selection rate of the annealed SMC algorithm when applied to the ith dataset. A quasi-binomial regression model is used because a binomial GLM is inappropriate: the outcomes associated with different predictors (important/unimportant) conditional on selection are not independent and do not have a common binary distribution. Note that the false selection rate will tend to be

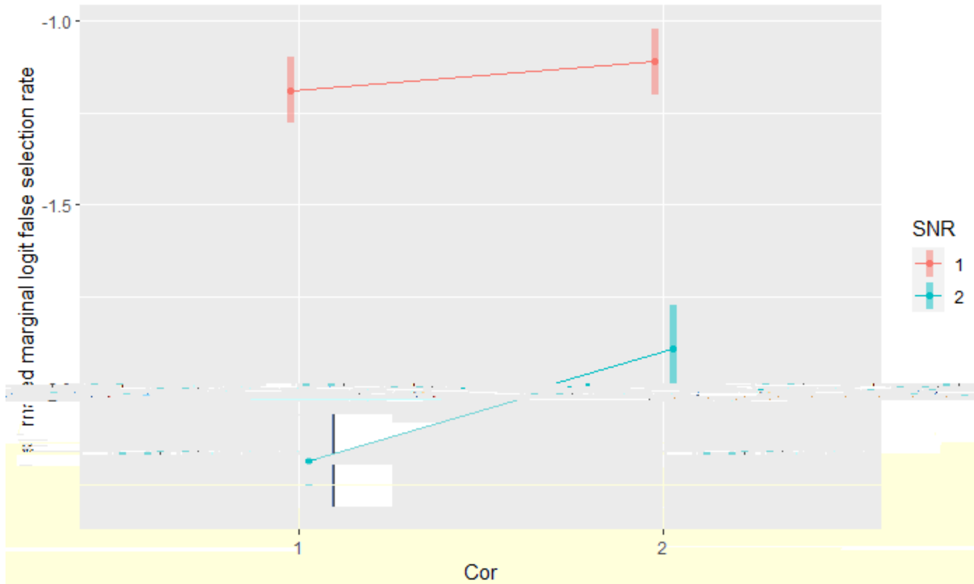

Figure 4.7: Interaction plot showing the effects of SNR and Cor on the logit false selection rate when Prop is high.

and

$$ar(M_i) = {}'t_i p_i^{(5)}(1 \quad p_i^{(5)})$$

where

$$\text{logit}(p_i^{(5)}) = \eta_i^{(5)}$$

and the $M_i's$ are independent. In this model, $p_i^{(5)}$ represents the negative selection rate of the annealed SMC algorithm when applied to the ith dataset. For the same reasons as given in the context of the false selection rate, a quasi-binomial regression approach is used, but this time incorporating Prop in the model. The analysis of deviance table shows that SNR and Cor have the most impact on negative selection rate. Figure 4.8 shows the effects of SNR and Cor on the logit of the negative selection rate. Higher SNR is associated with lower negative selection rate, with the effect more pronounced when Cor is low. Higher Cor is associated with higher negative selection rate, with the effect more pronounced when SNR is high.

In summary, under our experimental conditions, higher SNR is associated with better point estimation performance but worse interval estimation performance, while higher Cor is

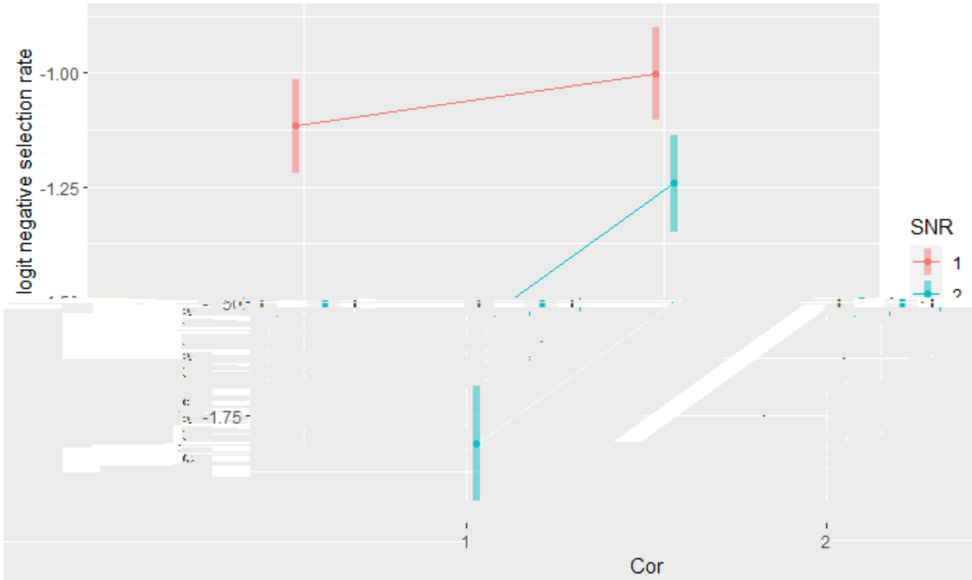

Figure 4.8: Interaction plot showing the effects of SNR and Cor on the logit negative selection rate.

individual and joint selection performance. Although higher Prop is associated with higher selection probabilities of x_5 and x_6, this effect is confounded with the effect of σ^2, so the direct effect of Prop is unclear. Overall, the annealed SMC sampling algorithm has good performance when SNR is high and Cor is low.

4.3 Mixed-effects case

We originally planned to conduct a comprehensive simulation study with a design similar to that used in section 4.2 in the linear mixed-effects model setting, but preliminary studies showed that the annealed SMC method struggles, in general, to have good selection and estimation performance in this context, which suggests that such a study would not be informative. Instead, we consider only the setting described in section 4.1, modified to include a random intercept. The purpose is simply to demonstrate the extent of the deterioration of the method's performance when it is applied in the mixed-effects context (relative to the benchmark established in the fixed-effects context).

More specifically, we generate the responses from the model

$$Y_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + b_i + \epsilon_{ij}$$

$j = 1, \ldots, 4$, $i = 1, \ldots, n$, where $\epsilon_{ij} \sim N(0, \sigma^2)$ and $b_i \sim N(0, \sigma_b^2)$ with $\sigma^2 = \sigma_b^2 = 5$ and $n = 1000$. The ϵ_{ij}'s and b_i's are generated independently. A smaller σ^2 compared to that used in section 4.1 is chosen to keep the signal-to-noise ratio the same. The annealed SMC algorithm is applied as in section 4.1 but with [illegible]$_b$ = 200, $_b$ = 2 1 and σ_4^2 = 0 25. We repeated the procedure only 100 times due to time constraints. The computation time was 3.5 hours. In contrast, the time taken to conduct the analogous simulation study for fixed-e□ects models was about 70 minutes.

Table 4.4: Estimated marginal selection probabilities of candidate predictors in the mixed-e□ects case.

[illegible]	[illegible]	[illegible]
x_1	0 830	0.038
x_2	0 810	0.039
x_3	0 230	0.042

Table 4.5: Estimated false and negative selection rates for the mixed-e□ects case.

[illegible]	[illegible]	[illegible]
False selection rate	0 121	0 227
Negative selection rate	0 180	0 261

Table 4.6: Estimated coverage probabilities of 95% credible intervals of model coe□cients for the mixed-e□ects case.

[illegible]	[illegible]	[illegible]
1	0 840	0.037
2	0 800	0.040
3	1 000	0.000

The estimated marginal selection probabilities, coverage probabilities, and false and negative selection rates are lower than those in section 4.1, indicating that the annealed SMC

4.4 Effect of the number of particles

Chapter 5

Discussion

The benchmark case indicates that annealed SMC can have satisfactory selection and estimation performance in simple cases. However, its performance deteriorates when more candidate predictors are added and a random intercept is introduced, as evidenced by the lower selection probabilities for important predictors, the higher selection probabilities for unimportant predictors, and the lower coverage probabilities of the 95% credible intervals for the model coefficients.

In practical applications, the number of candidate predictors will likely far exceed six (the maximum number we considered in our simulation study)—for example, the CAYACS data may have at least 30 candidate predictors. Although the number of candidate predictors in our simulation studies does not accurately reflect the size of the motivating problem, our work was constrained by available computing time. Moreover, the nature of this project is primarily exploratory with respect to the performance of the annealed SMC algorithm and the factors affecting it. Future work should include comprehensive simulation studies similar to those done for this project but with a larger number of candidate predictors, larger sample sizes, and data generated using mixed effects models (so that the generated datasets are more similar to the CAYACS data). In addition, the number of intermediate distributions and the number of particles should be considered as factors in the studies. The main challenge of such work is that significantly more computational resources will be required.

In the meantime, we have several ideas about how the performance of annealed SMC sampling could be improved for larger, more complicated problems. In our simulation studies, the algorithm is tuned minimally; it is not tuned for optimal performance for each setting due to time constraints. For example, it is common to use hundreds or thousands of values α_t, chosen non-linearly or adaptively, for more complicated models [18], whereas we used only ten evenly spaced values for our simulation studies. Furthermore, suppose that we wish to estimate the expectation of $\varphi(\omega)$, $= \mathbb{E}[\varphi(\omega)]$ where ω ^{T} and φ is a function of interest, using P particles $\omega_1^T \quad \omega_P^T$ produced by the annealed SMC algorithm, with the estimator $\hat{\ } = \sum_{i=1}^{P} W_T^i \varphi(\omega_i^T)$. Then one can show that

$$\frac{1}{\sqrt{N}}(\hat{\ } \qquad) \to_d N(0 \ (\sigma^*)^2)$$

for some asymptotic variance $(\sigma^*)^2$ in the limit of the number of particles [14]. By choosing $\ _A(x) = I(x \in A)$ for (measurable) subsets A of the sample space, this result may be interpreted as demonstrating the pointwise convergence of the distribution of weighted observations to the target distribution as the number of particles increases. Therefore, using more particles may improve performance up to a certain point, after which we would expect performance gains to be minimal. The simulation study of the six candidate predictor case

section 4.2 and a half-Cauchy prior for the variance, via the `rstanarm::stan_glm` **function [7]. The resulting credible intervals are more than 30 times shorter than those computed using annealed SMC sampling applied to Bayesian model averaging. Hence we find that the credible intervals computed by annealed SMC in our simulations are useful in determining**

Bibliography

[1] MM Barbieri and JO Berger. Optimal predictive model selection. The Annals of Statistics, 32(3):870–897, 2004.

[2] RJ Barker and WA Link. Bayesian multimodel inference by rjmcmc: A gibbs sampling approach. The American Statistician, 67(3):150–156, 2013.

[3] L Bottolo and S Richardson. Evolutionary stochastic search for bayesian model exploration. Bayesian Analysis, 5(3):583–618, 2010.

[4] MA Clyde, J Ghosh, and ML Littman. Bayesian adaptive sampling for variable selection and model averaging. Journal of Computational and Graphical Statistics, 20(1):80–101, 2011.

[5] ML McBride et al. Childhood, adolescent, and young adult cancer survivors research program of british columbia: Objectives, study design, and cohort characteristics. Pediatric Blood and Cancer, 55(2):324–330, 2010.

[6] EI George and RE McCulloch. Variable selection via gibbs' sampling. Journal of the American Statistical Society, 88(423):881–889, 1993.

[7] Ben Goodrich, Jonah Gabry, Imad Ali, and Sam Brilleman. rstanarm: Bayesian applied regression modeling via Stan., 2024. R package version 2.32.1.

[8] C Hans, A Dobra, and M West. Shotgun stochastic search for "large p" regression. Journal of the American Statistical Association, 102(478):507–516, 2007.

[9] JA Hoeting, D Madigan, AE Raftery, and CT Volinsky. Bayesian model averaging: A tutorial. Statistical Science, 14(4):382–401, 1999.

[10] JD Lee, DL Sun, Y Sun, and JE Taylor. Exact post-selection inference, with application to the lasso. The Annals of Statistics, 44(3):907–927, 2016.

[11] Russell V. Lenth. emmeans: Estimated Marginal Means, aka Least-Squares Means, 2024. R package version 1.10.0.

[12] F Liang, R Paulo, G Molina, MA Clyde, and JO Berger. Mixture of g priors for bayesian variable selection. Journal of the American Statistical Association, 103(481):410–423, 2008.

[13] D Madigan and AE Raftery. Model selection and accounting for model uncertainty in graphical models using occam's window. Journal of the American Statistical Association, 89(428):1535–1546, 1994.

[14] P Del Moral, A Doucet, and A Jasra. Sequential monte carlo samplers. Journal of the Royal Statistical Society Series B, 68(3):411–436, 2006.

[15] V Ročková and EI George. Emvs: The em approach to bayesian variable selection. Journal of the American Statistical Association, 109(506):828–846, 2014.

[16] Q Song and F Liang. A split-and-merge bayesian variable selection approach for ultrahigh dimensional regression. Journal of the Royal Statistical Society Series B, 77(5):947–972, 2015.

[17] CT Volinsky, D Madigan, AE Raftery, and RA Kronmal. Bayesian model averaging in proportional hazard models: Assessing the risk of a stroke. Journal of the Royal Statistical Society Series C, 46(4):433–448, 1997.

[18] Y Zhou, AM Johansen, and JAD Aston. Toward automatic model comparison: An adaptive sequential monte carlo approach. Journal of Computational and Graphical

Appendix A

Analysis of deviance tables

Table A.1: Analysis of deviance table for the estimated coverage probability of 95% credible intervals for $_1$.

Effect	Change in deviance	d.f.	p-value
SS	6.29	1.00	0.01
SNR	167.20	1.00	0.00
Prop	3.13	1.00	0.08

Table A.2: Analysis of deviance table for the estimated coverage probability of 95% credible intervals for $_{2}$.

Effect	Change in deviance	d.f.	p-value
SS	12.52	1.00	0.00
SNR	176.24	1.00	0.00
Prop	2.19	1.00	0.14
Cor	1.36	1.00	0.24
SS:SNR	11.46	1.00	0.00
SS:Prop	4.74	1.00	0.03
SS:Cor	0.00	1.00	0.99
SNR:Prop	0.80	1.00	0.37
SNR:Cor	1.83	1.00	0.18
Prop:Cor	0.31	1.00	0.58

Table A.3: Analysis of deviance table for the estimated coverage probability of 95% credible intervals for $_{3}$ when Prop is high.

Effect	Change in deviance	d.f.	p-value
SS	6.50	1.00	0.01
SNR	55.33	1.00	0.00
Cor	1.58	1.00	0.21
SS:SNR	0.52	1.00	0.47
SS:Cor	0.78	1.00	0.38
SNR:Cor	1.39	1.00	0.24

Table A.4: Analysis of deviance table for the estimated coverage probability of 95% credible intervals for $_{4}$ when Prop is high.

Effect	Change in deviance	d.f.	p-value
SS	11.63	1.00	0.00
SNR	102.71	1.00	0.00
Cor	0.14	1.00	0.71
SS:SNR	13.79	1.00	0.00
SS:Cor	1.02	1.00	0.31
SNR:Cor	2.55	1.00	0.11

Table A.5: Analysis of variance table for the estimated bias of the estimated posterior mean of $_1$.

Effect	**Sum of squares**	**d.f.**	F	**p-value**
SS	**0.13**	**1.00**	**0.61**	**0.43**
SNR	**2.18**	**1.00**	**10.38**	**0.00**
Prop	**2.33**	**1.00**	**11.08**	**0.00**
Cor	**11.77**	**1.00**	**55.90**	**0.00**
SS:SNR	**0.60**	**1.00**	**2.84**	**0.09**
SS:Prop	**0.46**	**1.00**	**2.16**	**0.14**
SS:Cor	**0.00**	**1.00**	**0.00**	**1.00**
SNR:Prop	**0.48**	**1.00**	**2.30**	**0.13**
SNR:Cor	**0.40**	**1.00**	**1.90**	**0.17**
Prop:Cor	**0.11**	**1.00**	**0.53**	**0.46**
Residual	**671.29**	**3189.00**		

Table A.6: Analysis of variance table for the estimated bias of the estimated posterior mean of $_2$.

Effect	**Sum of squares**	**d.f.**	F	**p-value**
SS	**1.03**	**1.00**	**4.73**	**0.03**
SNR	**2.85**	**1.00**	**13.03**	**0.00**
Prop	**0.32**	**1.00**	**1.47**	**0.22**
Cor	**7.49**	**1.00**	**34.32**	**0.00**
SS:SNR	**1.05**	**1.00**	**4.81**	**0.03**
SS:Prop	**0.92**	**1.00**	**4.22**	**0.04**
SS:Cor	**0.11**	**1.00**	**0.52**	**0.47**
SNR:Prop	**0.64**	**1.00**	**2.91**	**0.09**
SNR:Cor	**1.11**	**1.00**	**5.08**	**0.02**
Prop:Cor	**0.27**	**1.00**	**1.25**	**0.26**
Residual	**696.06**	**3189.00**		

Table A.7: Analysis of variance table for the estimated bias of the estimated posterior mean of $_{3}$ when `Prop` is high.

Effect	Sum of squares	d.f.	F	p-value
SS	0.00	1.00	0.00	1.00
SNR	4.46	1.00	20.16	0.00
Cor	6.36	1.00	28.74	0.00
SS:SNR	0.72	1.00	3.26	0.07
SS:Cor	0.08	1.00	0.36	0.55
SNR:Cor	0.63	1.00	2.85	0.09
Residual	352.42	1593.00		

Table A.8: Analysis of variance table for the estimated bias of the estimated posterior mean of $_{4}$ when Prop is high.

Effect	Sum of squares	d.f.	F	p-value
SS	0.18	1.00	0.78	0.38
SNR	2.43	1.00	10.51	0.00
Cor	6.00	1.00	25.94	0.00
SS:SNR	0.01	1.00	0.05	0.82
SS:Cor	0.00	1.00	0.00	0.97
SNR:Cor	0.04	1.00	0.19	0.66

Table A.9: Analysis of deviance table for the estimated selection probability of x_1.

Effect	Change in deviance	d.f.	p-value
SS	1.63	1.00	0.20
SNR	24.16	1.00	0.00
Prop	9.19	1.00	0.00
Cor	10.50	1.00	0.00
SS:SNR	3.37	1.00	0.07
SS:Prop	0.67	1.00	0.41
SS:Cor	0.02	1.00	0.88
SNR:Prop	0.69	1.00	0.41
SNR:Cor	10.93	1.00	0.00
Prop:Cor	0.36	1.00	0.55

Table A.10: Analysis of deviance table for the estimated selection probability of x_2.

Effect	Change in deviance	d.f.	p-value
SS	3.19	1.00	0.07
SNR	28.89	1.00	0.00
Prop	4.57	1.00	0.03
Cor	11.95	1.00	0.00
SS:SNR	7.67	1.00	0.01
SS:Prop	9.75	1.00	0.00
SS:Cor	0.43	1.00	0.51
SNR:Prop	5.14	1.00	0.02
SNR:Cor	2.65	1.00	0.10
Prop:Cor	0.15	1.00	0.70

Table A.11: Analysis of deviance table for the estimated selection probability of x_3 when Prop **is low.**

Effect	Change in deviance	d.f.	p-value
SS	**3.94**	**1.00**	**0.05**
SNR	**39.87**	**1.00**	**0.00**
Cor	**10.33**	**1.00**	**0.00**
SS:SNR	**2.03**	**1.00**	**0.15**
SS:Cor			

Table A.14: Analysis of deviance table for the estimated selection probability of x_4 when `Prop` **is high.**

Table A.16: Analysis of deviance table for the estimated selection probability of x_6.

	Change in deviance	d.f.	p-value
SS	2.04	1.00	0.15
SNR	150.08	1.00	0.00
Prop	32.23	1.00	0.00
Cor	9.97	1.00	0.00
SS:SNR	1.61	1.00	0.20
SS:Prop	0.08	1.00	0.78
SS:Cor	0.15	1.00	0.70
SNR:Prop	0.53	1.00	0.47
SNR:Cor	0.00	1.00	0.97
Prop:Cor	0.00	1.00	0.96

Table A.19: Analysis of deviance table for the estimated negative selection rate.

Effect	Change in deviance	d.f.	p-value
SS	7.22	1.00	0.01
SNR	68.13	1.00	0.00
Prop	12.48	1.00	0.00
Cor	36.34	1.00	0.00
SS:SNR	12.67	1.00	0.00
SS:Prop	6.98	1.00	0.01
SS:Cor	0.22	1.00	0.64
SNR:Prop	7.27	1.00	0.01
SNR:Cor	18.23	1.00	0.00
Prop:Cor	1.31	1.00	0.25