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

SIMON FRASER UNIVERSITY
Spring 2024

Declaration of Committee

Name:	**Quang Vuong**
Degree:	**Master of Science**
Thesis title:	**The performance of annealed sequential Monte Carlo sampling as a joint variable selection and parameter estimation method in the linear (mixed) model setting**
Committee:	**Chair:** Wei Lin Lecturer, Statistics and Actuarial Science **Rachel Altman** Supervisor Associate Professor, Statistics and Actuarial Science **Haolun Shi** Committee Member Assistant Professor, Statistics and Actuarial Science **Liangliang Wang** 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 ratio, the proportion of important predictors, the correlation of predictors, and the inclusion of a random effect appreciably impact the performance of annealed SMC sampling.

Keywords: variable selection; parameter estimation; Bayesian model averaging; annealed SMC

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

Declaration of Committee **ii**

Abstract **iii**

Dedication **iv**

Acknowledgements **v**

Table of Contents **vi**

List of Tables **vii**

List of Figures **viii**

1 Introduction **1**

2 Motivating example **5**

3 Methods **6**

3.1 Linear Mixed Model 6

3.2 Annealed SMC 8

3.2.1 Construction 8

3.2.2 Selection of L_{t-1} and K_t 10

4 Simulation studies **12**

4.1 Fixed-effects benchmark case 13

4.2 Factors that affect the performance of annealed SMC in the fixed-effects case 14

4.3 Mixed-effects case 24

4.4 Effect of the number of particles 26

5 Discussion **28**

Bibliography **33**

Appendix A Analysis of deviance tables **35**

List of Tables

Table 4.1 Estimated marginal selection probabilities of candidate predictors in fixed-effects benchmark case. 14

Table 4.2 Estimated false and negative selection rates for the fixed-effects benchmark case. 14

Table 4.3 Estimated coverage probabilities of 95% credible intervals of model coefficients for fixed-effects benchmark case. 14

Table 4.4 Estimated marginal selection probabilities of candidate predictors in the mixed-effects case. 25

Table 4.5 Estimated false and negative selection rates for the mixed-effects case. 25

Table 4.6 Estimated coverage probabilities of 95% credible intervals of model coefficients for the mixed-effects case. 25

Table 4.7 Estimated marginal selection probabilities of candidate predictors using 6000 and 1000 particles. 27

Table 4.8 Estimated false and negative selection rates using 6000 and 1000 particles. 27

Table 4.9 Estimated coverage probabilities of 95% credible intervals of model coefficients using 6000 and 1000 particles. 27

List of Figures

Figure 4.1 Interaction plot showing the effects of SNR and SS on the coverage probability of the 95% credible interval for β_1. 18

Figure 4.2 Interaction plot showing the effects of SNR and Cor on the bias of the posterior mean of β_1 as an estimator for β_1. 19

Figure 4.3 Interaction plot showing the effects of SNR and Cor on the selection probability of x_1. 20

Figure 4.4 Interaction plot showing the effects of SNR and Cor on the selection probability of x_3 when Prop is low. 21

Figure 4.5 Interaction plot showing the effects of SNR and Prop on the selection probability of x_5. 21

Figure 4.6 Interaction plot showing the effects of SNR and Cor on the logit false selection rate when Prop is low. 22

Figure 4.7 Interaction plot showing the effects of SNR and Cor on the logit false selection rate when Prop is high. 23

Figure 4.8 Interaction plot showing the effects of SNR and Cor on the logit negative selection rate. 24

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 $\mathcal{M}$ denote the (possibly countably infinite) set of possible models $\{M_k\}$, where k is an arbitrary index and a prior distribution $\pi(M_k)$ is given over $\mathcal{M}$. For each model M_k, there are parameters θ_k and the likelihood $f(D \mid M_k, \theta_k)$ of the data D under M_k. If Δ is a quantity of interest, then we can average its posterior distribution across models in $\mathcal{M}$ as follows [9]:

$$\pi(\Delta \mid D) = \sum_{M_k \in \mathcal{M}} \pi(\Delta \mid M_k, D)\pi(M_k \mid D),$$

where

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

and

$$\pi(D \mid M_k) = \int \pi(D \mid \theta_k, M_k)\pi(\theta_k \mid M_k)\, d\theta_k.$$

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 $\pi(M_k \mid D)$, 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.

For a simple family of models for the data, the posterior model probabilities may be calculated directly for all possible models. But this approach becomes infeasible when more complicated models and more candidate predictors are involved. Specifically, the quantity $P(D \mid M_k)$, often called the evidence for the model M_k, can be difficult to calculate when θ_k is high-dimensional (which is often the case in regression contexts), and the number of posterior model probabilities to calculate grows exponentially with the number of candidate predictors. Early methods for overcoming these issues are unsatisfactory for various reasons. For example, leaps-and-bounds and Occam's window [9, 17, 13] involve excluding ill-fitting models from the model space, thus failing to fully account for model uncertainty. MCMC methods (MC3, SSVS [6]), an alternative approach, often have poor convergence and/or mixing properties [3]. Therefore, a variety of modern approaches have been developed to address the limitations of these early methods. This project specifically focuses on the application of one of these modern approaches, annealed sequential Monte Carlo (SMC) sampling, as a variable selection and parameter-estimation method in the context of linear models. We investigate the application of annealed SMC sampling to both fixed-effects and mixed-effects models. The latter are more appropriate for longitudinal data, which frequently occur in disciplines such as health and medical research. By studying the performance of the method for mixed-effects models as well, our project gives insight about the performance of the method for longitudinal models. We consider the selection only of predictors with fixed, not random, effects because in practice, variable selection questions of interest primarily focus on predictors with fixed effects.

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 investigations, we found that this method has poor convergence and is computationally inefficient, in part because the joint posterior of the parameters given the model does not have a simple form. Next, we attempted the direct calculation of the evidence of individual models using Monte Carlo integration, but we quickly found that the selection performance of this method is very sensitive to prior hyperparameter selection. Therefore, we sought more sophisticated posterior sampling methods in the literature. There are a variety of methods that efficiently sample from the posterior distribution on only the model space, but methods that draw models and parameters jointly are lacking. One example is EM-based variable selection, which applies the EM algorithm to the likelihood of the joint posterior distribution of the model and the parameters primarily to identify models with high posterior probability [15]. Another example is Bayesian adaptive sampling, which draws a sample without replacement from the model space so that the sample marginal inclusion probabilities of the predictors are close to their true posterior marginal inclusion probabilities [4]. Split-and-merge variable selection partitions the set of candidate predictors, individually selects for predictors within each partition, and then selects predictors from the pool of these previously selected predictors, which allows consistent selection of predictors even with multicollinearity [16]. Population MCMC methods, which have improved mixing properties compared to simpler MCMC methods, are also used; evolutionary Monte Carlo uses a parallel tempering approach with multiple chains' targeting different temperatures and global exchange moves between chains [3]. These methods mostly provide a sample of models from the posterior distribution on the model space, from which posterior model probabilities may be estimated. These probabilities may be paired with samples from the posterior distribution of parameters for each model in order to obtain averaged posteriors of the model coefficients. However, this approach is sub-optimal because the posterior distributions of model parameters for all sampled models must be computed, which is again prohibitive when a large number of iterations is used. Annealed SMC sampling is an alternative that permits the joint sampling of models and parameters from the posterior distribution, thereby reducing the computational demand.

Annealed SMC sampling is a modification of (regular) SMC sampling, which is a method based on repeated importance sampling that is used to draw samples from a sequence of distributions [18, 14]. It is used in settings where the joint distribution of a sequence of random variables is of interest but may be intractable or known only up to a constant. It is a preferable alternative to MCMC when the latter produces chains with poor convergence. SMC sampling can be adapted to sampling a single target distribution π, 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 the number of observations sampled, called particles, increases [14]. Ideally, a larger number of particles is used for more complex distributions, such as those induced by complex models or large model spaces, but too many particles may pose a computational burden that negates the potential time-savings of annealed SMC compared to competing methods. Thus, this project seeks to investigate the ability of a minimally tuned annealed SMC sampling scheme to select important predictors and produce valid interval estimates of model coefficients under different conditions. Using a comprehensive simulation study, our main purpose is to provide analysts with insight into of the performance of the method in practice.

The remainder of this report is organized as follows. In chapter 2, we describe the CAYACS dataset that motivates the project. In chapter 3, we describe the model of interest and the annealed SMC sampling algorithm we use. In chapter 4, we describe and present the results of the simulation studies that examine the performance of the described annealed SMC sampling algorithm, and we provide a discussion of our work in chapter 5.

Chapter 2

Motivating example

This project is motivated by the BC Cancer Childhood, Adolescent, and Young Adult Cancer Survivors (CAYACS) research program. As cancer treatment continues to improve, an increasing number of cancer survivors may have special healthcare needs [5]. As part of this program, clinical and demographic data of survivors of cancer occurring at a relatively young age in British Columbia, Canada were collected in order to examine late health effects and health-services-utilization patterns. Such information is critical for health care planning purposes.

The collected data are longitudinal and contain at least 30 variables, so questions about which variables have important effects on late health effects and utilization of health services are naturally of interest. For example, one question is which variables (cancer type, treatment, age at diagnosis, sex, etc.) impact the levels of prognostic, cancer-related metabolites measured on survivors over time. Such questions motivate the investigation of the performance of methods that can simultaneously do both variable selection and parameter estimation in the longitudinal data setting.

Unfortunately, the release of the CAYACS data to researchers was delayed for reasons beyond the control of BC Cancer. We therefore were unable to apply our proposed methods to real data. However, the context remains one important motivation for our work.

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, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$. We proceed with a joint model-parameter space in the sense of Barker & Link (2013) [2]. Let $\boldsymbol{\beta} = (\beta_0, ..., \beta_p)$ be a column vector of regression coefficients. We identify a model by $\zeta = (\zeta_1, ..., \zeta_p)$, where $\zeta_j \in \{0, 1\}$ for $j = 1, ..., p$ as follows. Let $m_1, ..., m_{|\zeta|}$ be the indices for which ζ_j is 1 (listed in increasing order), where $|\zeta|$ indicates the L^1-norm of ζ, and let $\boldsymbol{\psi} = (\psi_0, ..., \psi_p)$ be an auxiliary parameter that is common across all models. We introduce $\boldsymbol{\psi}$ because in the annealed SMC sampling algorithm, we draw $\boldsymbol{\psi}$ 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

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

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

$$\boldsymbol{\beta}_\zeta = \begin{pmatrix} \beta_0 \\ \beta_{m_1} \\ \vdots \\ \beta_{m_{|\zeta|}} \end{pmatrix},$$

$$\boldsymbol{\psi}_\zeta = \begin{pmatrix} \psi_0 \\ \psi_{m_1} \\ \vdots \\ \psi_{m_{|\zeta|}} \end{pmatrix},$$

and

$$\boldsymbol{Y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}.$$

Let $\boldsymbol{\beta}_{-\zeta}$ and $\boldsymbol{\psi}_{-\zeta}$ be the vectors containing the remaining β and ψ parameters not in $\boldsymbol{\beta}_\zeta$ and $\boldsymbol{\psi}_\zeta$, respectively, in increasing order of the index. We use a Gaussian linear mixed effects model for Y_{ij}, i.e.

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + b_i\mathbf{1}_{n_i} + \boldsymbol{\varepsilon}_i.$$

We use ζ to identify the models in the model space by setting $\boldsymbol{\beta}_\zeta = \boldsymbol{\psi}_\zeta$ and $\boldsymbol{\beta}_{-\zeta} = 0$ conditional on ζ. Here $b_i \sim N(0, \sigma_b^2)$, and $\mathbf{1}_{n_i}$ is the n_i-dimensional column vector of ones. The errors, $\boldsymbol{\varepsilon}_i$, $i = 1, ..., n$, are assumed to be distributed as $MVN(\mathbf{0}_{n_i}, \sigma^2\boldsymbol{I}_{n_i})$, where $\mathbf{0}_m$ denotes the m-dimensional column vector of zeroes and $\boldsymbol{I}_m$ denotes the $m \times m$ identity matrix. In addition, $b_1, \dots, b_n, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ are assumed independent. Crucially, given ζ, $\boldsymbol{\psi}$ completely determines $\boldsymbol{\beta}$. We focus on $\boldsymbol{\psi}$ instead of $\boldsymbol{\beta}$ to facilitate the construction of the chosen forward kernels in section 3.2.2.

Let $\Omega = \{(\zeta_1, .., \zeta_p) \mid \zeta_j \in \{0, 1\} \text{ for all } j\}$ represent the model space. We use a uniform prior on the model space, i.e., $P(\zeta) = \frac{1}{2^p}$ for all $\zeta \in \Omega$. This choice is appealing because it does not require prior elicitation from domain-knowledge experts and because all predictors are given equal *a priori* importance, i.e., no predictors are preferred over others. Other priors on the model space can be used; the choice has minimal impact on the details of the model selection algorithm. Given each ζ, we use the priors

$$\boldsymbol{\psi}_\zeta \sim MVN\left(\mathbf{0}_{|\zeta|+1}, \tau\begin{pmatrix} v & 0 \\ 0 & \boldsymbol{I}_{|\zeta|} \end{pmatrix}\right),$$

$$\boldsymbol{\psi}_{-\zeta} \sim MVN\left(\mathbf{0}_{p-(|\zeta|+1)}, \sigma_u^2\boldsymbol{I}_{p-(|\zeta|+1)}\right),$$

$$\sigma^2 \sim InvGamma(\alpha, \nu),$$

and

$$\sigma_b^2 \sim InvGamma(\alpha_b, \nu_b),$$

where v, τ, α, ν, α_b, and ν_b are hyperparameters to be set according to the problem. Here, v should be set to be large in order to allow a large range of plausible values for the intercept coefficient. We standardize predictors and set $\tau = 1$ for this project; standardizing predictors addresses scaling concerns, and the choice of τ is at the discretion of the analyst [8]. We comment on this issue further in section 5.

The interpretation of the model coefficients, the β_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, β_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 β_j represents the effect of x_j after having adjusted for other predictors in the model. When the posterior distributions of β_j are averaged across models containing β_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, β_j is set to be zero, meaning x_j has no effect on the mean response. Hence when the posterior distribution of β_j is averaged over all models, it has two components: one that describes the effect of x_j on the response adjusted for other candidate predictors with possibly zero effect, and a point mass at zero. Averaging over all possible models gives a non-zero value to the probability $P(\beta_j = 0)$, which is useful for variable selection, whereas fitting only the full model would make this probability 0 no matter the importance of x_j (because the posterior distribution in this case would be continuous). When a predictor is selected, the credible interval for the corresponding coefficient is largely determined by the distribution describing the effect adjusted for other predictors, but when it is not, the credible interval will be concentrated around 0, which is the desired behavior.

3.2 Annealed SMC

3.2.1 Construction

Denote $\Theta = (\boldsymbol{\psi}^T, b_1, ..., b_n, \sigma^2, \sigma_b^2)$. We seek to sample (Θ, ζ) from the posterior distribution defined by

$$\pi(\Theta, \zeta \,|\, \boldsymbol{y}) \propto f(\boldsymbol{y} \,|\, \Theta, \zeta)\pi(\Theta \,|\, \zeta)\pi(\zeta).$$

Here, $f(\boldsymbol{y} \,|\, \Theta, \zeta)$ denotes the likelihood of data given (Θ, ζ), $\pi(\Theta \,|\, \zeta)$ denotes the prior on Θ given the model ζ, and $\pi(\zeta)$ is the prior on ζ. The random effects, b_i, are sampled for computational ease, since it avoids the integration of the likelihood over the random effects. Set $\pi^0(\Theta, \zeta) \equiv \pi(\Theta \,|\, \zeta)\pi(\zeta)$ and

$$\pi^t(\Theta, \zeta \,|\, \boldsymbol{y}) \propto f(\boldsymbol{y} \,|\, \Theta, \zeta)^{\alpha_t}\pi(\Theta \,|\, \zeta)\pi(\zeta) \equiv \gamma^t(\Theta, \zeta)$$

for a strictly increasing sequence $\alpha_1, ..., \alpha_T \in (0, 1]$ with $\alpha_T = 1$, so that $\pi^T(\Theta, \zeta \,|\, \boldsymbol{y}) = \pi(\Theta, \zeta \,|\, \boldsymbol{y})$ is the target distribution from which we wish to draw. Despite the notation suggesting otherwise, $\pi^1, ..., \pi^{T-1}$ are not true posterior distributions of (Θ, ζ) conditioned on $\boldsymbol{y}$. But they are proper densities that depend on $\boldsymbol{y}$.

The main idea is to use the intermediate distributions π^t to facilitate the sampling of π^T starting with draws from π^0 in the following way. Let $\omega_t \equiv (\Theta_t, \zeta_t)$, $t = 0, ..., T$. At the t-th step, we obtain weighted samples that approximate the following joint distribution of $\omega_{0:t} = (\omega_0, ..., \omega_t)$:

$$\tilde{\pi}^t(\omega_{1:t}) \equiv \pi^t(\omega_t) \prod_{s=1}^{t} L_{s-1}(\omega_s, \omega_{s-1}).$$

L_{s-1} is called a backward kernel and is an arbitrary distribution of ω_{s-1} given ω_s. The joint distribution $\tilde{\pi}^t$ has the property that the marginal distribution of ω_t at the t-th step is π^t; in other words, the marginal distribution of ω_T at the last step is π^T, which is our target distribution. We discuss the choice of L_s in section 3.2.2.

This construction permits the usage of the following SMC algorithm, which provides a collection of weighted samples of $\omega_{1:t}$ at the t-th step that can be used to form a Monte Carlo approximation of $\tilde{\pi}^t$:

1. Let $t = 1$ and draw $\omega_0^i = (\Theta_0^i, \zeta_0^i) \sim \pi^0$ independently for $i = 1, ..., P$, where P is a pre-determined positive integer. The draw ω_0^i is called the i^{th} particle obtained at step 0, so P is the number of particles sampled. Let $W_0^1 = ... = W_0^P = \frac{1}{P}$ be the initial importance weights of the particles.

2. At step t:

   (a) Draw $\omega_t^i = (\Theta_t^i, \zeta_t^i) \sim K_t((\Theta_{t-1}^i, \zeta_{t-1}^i), \cdot) = K_t(\omega_{t-1}^i, \cdot)$ independently for $i = 1, ..., P$ for some distribution $K_t(\omega_{t-1}^i, \cdot)$ dependent on ω_{t-1}^i. K_t is called a forward kernel; we discuss possible choices in section 3.2.2.

   (b) Let
   $$w_t^i = \frac{\gamma^t(\omega_t^i) L_{t-1}(\omega_t^i, \omega_{t-1}^i)}{\gamma^{t-1}(\omega_{t-1}^i) K_t(\omega_{t-1}^i, \omega_t^i)}$$
   be the incremental weight at step t of particle i, and let
   $$W_t^i = \frac{W_{t-1}^i w_t^i}{\sum_{j=1}^P W_{t-1}^j w_t^j},$$
   be the (normalized) importance weight of particle i at step t.

   (c) Let
   $$ESS = \frac{\left(\sum_{j=1}^P W_{t-1}^j w_t^j\right)^2}{\sum_{j=1}^P (W_{t-1}^j)^2 (w_t^j)^2}$$
   be the effective sample size at step t. If $ESS < \frac{P}{2}$, resample the particles according to probabilities W_t^i. Then let $W_t^i = \frac{1}{P}$ for all $i = 1, ..., P$.

(d) Increment t by 1.

3. Repeat step 2 until $t = T + 1$.

To see how this algorithm provides a Monte Carlo approximation to $\tilde{\pi}^t$ at step t, ignoring for the moment the resampling in step 2(c), note that step t provides particles $\omega^i_{1:t}$ with weights

$$W^i_t = \frac{\gamma^t(\omega^i_t)\prod_{s=1}^t L_{s-1}(\omega^i_s, \omega^i_{s-1})}{\gamma^0(\omega^i_0)\prod_{s=1}^t K_s(\omega^i_{s-1}, \omega^i_s)} \propto \frac{\tilde{\pi}^t(\omega_{1:t})}{\tilde{\eta}^t(\omega_{1:t})}$$

where $\tilde{\eta}^t(\omega_{1:t})$ denotes the distribution of $\omega_{1:t}$ as drawn in the above algorithm. Then the above algorithm is equivalent to importance sampling targeting $\tilde{\pi}^t$ with the proposal $\tilde{\eta}^t$ at step t. Accordingly, the Monte Carlo approximations to $\tilde{\pi}^t$ and π^t at step t are given by

$$\hat{\tilde{\pi}}^t = \sum_{i=1}^P W^i_t \delta_{\omega^i_{1:t}}$$

and

$$\hat{\pi}^t = \sum_{i=1}^P W^i_t \delta_{\omega^i_t},$$

respectively, where δ_a denotes the Dirac delta measure supported on a. It can be shown that these approximations still behave well when resampling occurs as per step 2(c) [14].

3.2.2 Selection of L_{t-1} and K_t

Theoretically, the backward kernels L_{t-1} are arbitrary, but selecting them optimally according to K_t will lead to importance weights with lower variance; see the discussion in Del Moral, Doucet & Jasra (2006) [14]. A convenient choice for K_t is a π^t-invariant kernel, i.e., a one-step transition density whose stationary distribution is π^t. Then let

$$L_{t-1}(\omega_t, \omega_{t-1}) = \frac{\pi^t(\omega_{t-1})K_t(\omega_{t-1}, \omega_t)}{\pi^t(\omega_t)},$$

which will produce good Monte Carlo approximations of the target distribution if the α_t's are close to each other. The number and values of the α_t's are tuning parameters that are chosen manually according to the model being fitted. Often, more complicated models require higher T and more closely spaced α_t's, and the α_t's may be spaced in many different ways (linearly, quadratically, exponentially, etc.). The α_t's may also be chosen adaptively [18], but this option is not pursued in this project due to time constraints. The incremental weights become

$$w^i_t = \frac{\gamma^t(\omega^i_{t-1})}{\gamma^{t-1}(\omega^i_{t-1})} = f(\boldsymbol{y} \mid \Theta, \zeta)^{\alpha_t - \alpha_{t-1}},$$

with $\alpha_0 = 0$.

For this project, K_t is constructed to be a symmetric random walk Metropolis kernel that is π^t-invariant. Recall that in section 3.2.1, we draw samples $\omega_t = (\Theta_t, \zeta_t)$ from the distribution π^t at step t. Since Θ is defined using $\boldsymbol{\psi}$ instead of $\boldsymbol{\beta}$, the joint model-parameter space $\{\omega\}$ has fixed dimension. Therefore, Metropolis-Hastings (MH) kernels used in MCMC can be used on $\{\omega_t\}$ and may be constructed to be π^t-invariant. An MH kernel that is π^t-invariant draws a proposal for the next sample $\omega^* \sim q(\omega, \cdot)$ conditional on the previous sample ω and moves from ω to ω^* with probability

$$\min\left\{1, \frac{\pi^t(\omega^*)q(\omega^*, \omega)}{\pi^t(\omega)q(\omega, \omega^*)}\right\},$$

staying at ω otherwise. If q is symmetric, i.e. $q(\omega^*, \omega) = q(\omega, \omega^*)$, then the acceptance probability involves only the ratio of target densities at ω^* and ω. For this project, the proposal adds or removes a predictor from M_{t-1} with uniform probability and samples

1. $\boldsymbol{\psi}_t \sim MVN(\boldsymbol{\psi}_{t-1}, \sigma_1^2 \boldsymbol{I}_{p+1})$, where $\boldsymbol{\psi}_t$ is the value of $\boldsymbol{\psi}$ sampled at step t,

2. $b_{t,i} \sim N(b_{t-1,i}, \sigma_2^2)$, where $b_{t,i}$ is the random effect for subject i sampled at step t,

3. $\sigma_t^2 = \sigma_{t-1}^2 e^{A_1}$, where $A_1 \sim N(0, \sigma_3^2)$, where σ_t^2 is the value of σ^2 sampled at step t, and

4. $\sigma_{t,b}^2 = \sigma_{t-1,b}^2 e^{A_2}$, where $A_2 \sim N(0, \sigma_4^2)$, where $\sigma_{t,b}^2$ is the value of σ_b^2 sampled at step t.

Here, σ_1^2, σ_2^2, σ_3^2, and σ_4^2 are step-size parameters and are subject to tuning. If this set of forward and backward kernels is used, the incremental weights at step t are known before propagating the particles at step $t-1$, and so step 2(a) in the algorithm should be executed after steps 2(b) and (c) to allow more diversity of proposals propagated from highly-weighted particles [14]. To see the benefit of changing the order of steps in this way, when resampling is done before propagation, the collection of particles is repopulated with highly-weighted particles, which are those in high-probability regions of π^t. The propagation of these particles then efficiently explores the joint model-parameter space according to the distribution of π^t. When resampling is done after propagation, some low-weighted particles are propagated and then discarded by resampling, and there is some redundancy among resampled particles that are propagated previously. Therefore, performing resampling before propagation when possible allows more efficient exploration of the joint model-parameter space and leads to better approximations of the target distribution.

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 $\hat{\pi}^T$ 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 $\hat{\pi}^T$ 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 theoretical result that says the model with the best predictive probability is the one with predictors with marginal posterior inclusion probabilities of at least 0.5 [1]. We summarize our results by computing performance measures such as estimates of the marginal selection rate of each predictor, the false selection rate (the proportion of selected predictors that are not important), the negative selection rate (proportion of important predictors that are not selected), and the coverage probabilities of the credible intervals. We note that these measures are estimates of the marginal probabilities of interest (averaged across all possible datasets) and therefore depend on the joint distribution of the response and the predictors. The algorithm is parallelized to speed up computation in all simulation studies.

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 methods are typically not applied in a three-predictor setting, this simple case allows us to evaluate the correctness of our annealed SMC implementation. The predictors are i.i.d. generated according to the distribution

$$\begin{pmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{pmatrix} \sim MVN(\mathbf{0}_3, \sigma_X^2 \boldsymbol{I}_3),$$

with $\sigma_X^2 = 0.01$. The responses are generated as

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i,$$

where $\beta_0 = 2, \beta_1 = -10, \beta_2 = 10, \text{and } \beta_3 = 0$ so that x_1 and x_2 are the important predictors in explaining Y. Here $\varepsilon_i \sim N(0, \sigma^2)$ are i.i.d. with $\sigma^2 = 15$. With reference to the model and algorithm specification in sections 3.1 and 3.2.2, we also set $v = 100$, $\tau = 1$, $\alpha = 2.1, \nu = 100, \alpha_b = 0, \nu_b = 0, \sigma_u^2 = 5$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0.25$, and $\sigma_4^2 = 0$. We generate 500 datasets from this model, each time applying annealed SMC as described in section 3.2 with 1000 particles and $\alpha_t = \frac{t}{10}$ for $t = 1, ..., 10$. The parameters in the simulation and for the algorithm are initially chosen arbitrarily and then slightly modified in order to produce a case under which the algorithm performs well. In particular, σ_X^2 is chosen to be different from 1 so that the candidate predictors are non-trivially standardized (i.e., scaled by a factor substantially different from 1) as part of the simulation, and the true model coefficients β_i are chosen in conjunction with σ_X^2 so that the coefficients on the standardized scale are not too large. We also set v to be an arbitrary large value and select α and ν so that an uninformative prior for σ^2 is obtained. The step size parameters and number of particles are chosen mostly arbitrarily. The α_t's are selected with two considerations. First, the α_t's should be sufficiently numerous and close to each other so that draws in each step provide decent approximations to the target distribution in the next step. Second, the α_t's should not be so numerous that computing time is prohibitive. We determine that using only ten intermediate distributions is sufficient for an initial exploration of the performance of the method while remaining computationally feasible. Tables 4.1–4.3 show the performance metrics of interest. The computing time for this study was about five hours.

The important predictors, x_1 and x_2, are selected with high probability, whereas x_3 is selected with low probability. The estimated false and negative selection rates are also low (at most 0.05). The estimated coverage probabilities of the credible intervals are approximately

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

Candidate predictor	**Estimated marginal selection probability**	**Standard error**
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-effects benchmark case.

Selection error metric	**Estimated mean**	**Estimated standard deviation**
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 coefficients for fixed-effects benchmark case.

Model coefficient	**Estimated coverage probability**	**Standard error**
β_1	0.970	0.008
β_2	0.944	0.010
β_3	1.000	0.000

equal to the nominal level of 95% except in the case of β_3. These statistics suggest that our implementation of annealed SMC is correct and is effective at selecting important predictors and estimating the coefficients of important predictors in a simple context.

4.2 Factors that affect the performance of annealed SMC in the fixed-effects case

To assess the selection and estimation performance of annealed SMC under different variants of the fixed-effects case, we consider the following model. Define

$$\boldsymbol{x}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{i6} \end{pmatrix} \sim MVN(\mathbf{0}_6, \Sigma),$$

where $\mathbf{0}_6$ is the zero vector in $\mathbb{R}^6$ and

$$\Sigma = \sigma_X^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{6\times 6}.$$

Let

$$Y_i = \beta_0 + \boldsymbol{x_i}^T\boldsymbol{\beta} + \varepsilon_i,$$

where $\boldsymbol{\beta}^T = \begin{pmatrix}\beta_1 & \cdots & \beta_6\end{pmatrix}$, $\varepsilon_i \sim N(0, \sigma^2),\ i = 1, ..., n$, and the ϵ_i's are independent. Here, we incorporate six candidate predictors to test the method in a more realistic case (i.e., a case where variable selection might realistically be of interest) while avoiding a prohibitive computational demand.

We use a full factorial experiment to investigate the effect of various factors on the performance of the method. The factors and their levels are as follows:

- Sample size (`SS`): low ($n = 500$) and high ($n = 1000$)
- Signal-to-noise ratio (`SNR`): low ($\texttt{SNR} = 0.02$) and high ($\texttt{SNR} = 2$), where

$$\texttt{SNR} \equiv \frac{\sigma_X^2\|\boldsymbol{\beta}\|_2^2}{\sigma^2}$$

- Proportion of important variables (`Prop`): few ($\boldsymbol{\beta}^T = (-10, 10, 0, 0, 0, 0)$) and many ($\boldsymbol{\beta}^T = (-10, 10, 10, -10, 0, 0)$)
- Pairwise correlation of predictors (`Cor`): low ($\rho = 0$) and high ($\rho = 0.5$).

The signal-to-noise ratio is important to investigate because preliminary simulations suggest that the relative sizes of $\boldsymbol{\beta}$ and σ^2 influence the performance of the algorithm. The factor levels are chosen to be similar to section 4.1; recall that in the benchmark case, $n = 1000$ and $\texttt{SNR} = 0.133$. A smaller sample size is chosen for the low level of `SS` to reduce the computational burden. On the other hand, we are not limited by computation time when exploring the effect of `SNR` because changing `SNR` changes only the variance of the responses while keeping the number of intermediate distributions, the number of particles, and the dimensions of the dataset the same. Therefore, we choose two well-separated levels to make the effect of `SNR` clearer. The levels of `Prop` are chosen so that situations where most variables are important and where most variables are unimportant are simulated. The levels of `Cor` are chosen to simulate uncorrelated predictors and moderately correlated predictors, as the algorithm is almost certain to have poorer performance when predictors are highly correlated. The parameters β_0 and σ_X^2 are set to be 2 and 0.01, respectively, for all datasets. The parameters σ_X^2 and $\boldsymbol{\beta}$ are chosen to be similar to those in section 4.1.

For each run (combination of factor levels), 200 datasets (replicates) are generated, and the annealed SMC algorithm is applied using the same priors and tuning parameters as in section 4.1. The same performance measures are recorded for each replicate as well: coverage of the 95% credible interval (yes/no), marginal posterior inclusion probabilities of each candidate predictor, and the estimated false and negative selection rates. As before, a predictor is considered selected if its marginal posterior inclusion probability is at least 0.5. The time needed to complete this simulation study was about 16 hours.

To determine factors that impact the estimation performance of the annealed SMC algorithm, we define two response variables of interest. For the i-th dataset, let $Y_i^{(1)}$ be the binary indicator for the coverage of the 95% credible interval for a chosen model coefficient and $Y_i^{(2)}$ be the estimated bias of the posterior mean of the model coefficient. For the i-th dataset and the k-th response, we also define the linear predictor

$$\begin{aligned}
\eta_i^{(k)} = \alpha_0^{(k)} &+ (1 - 2I(\texttt{SS}_i = \text{low}))\alpha_1^{(k)} \\
&+ (1 - 2I(\texttt{SNR}_i = \text{low}))\alpha_2^{(k)} \\
&+ (1 - 2I(\texttt{Prop}_i = \text{low}))\alpha_3^{(k)} \\
&+ (1 - 2I(\texttt{Cor}_i = \text{low}))\alpha_4^{(k)} \\
&+ (1 - 2I(\texttt{SS}_i = \text{low})I(\texttt{SNR}_i = \text{low}))\alpha_5^{(k)} \\
&+ (1 - 2I(\texttt{SS}_i = \text{low})I(\texttt{Prop}_i = \text{low}))\alpha_6^{(k)} \\
&+ (1 - 2I(\texttt{SS}_i = \text{low})I(\texttt{Cor}_i = \text{low}))\alpha_7^{(k)} \\
&+ (1 - 2I(\texttt{SNR}_i = \text{low})I(\texttt{Prop}_i = \text{low}))\alpha_8^{(k)} \\
&+ (1 - 2I(\texttt{SNR}_i = \text{low})I(\texttt{Cor}_i = \text{low}))\alpha_9^{(k)} \\
&+ (1 - 2I(\texttt{Prop}_i = \text{low})I(\texttt{Cor}_i = \text{low}))\alpha_{10}^{(k)}.
\end{aligned}$$

We model our first response, $Y_i^{(1)}$, using a binary generalized linear model (GLM) with sum contrasts for the factors:

$$Y_i^{(1)} \sim Bern(p_i^{(1)}),$$

where

$$\text{logit}(p_i^{(1)}) = \eta_i^{(1)}$$

and the $Y_i^{(1)}$'s are independent. In this model, $p_i^{(1)}$ represents the probability that the 95% credible interval for a particular model coefficient constructed using the i^{th} dataset contains the true value of the model coefficient. We then construct an analysis of deviance table using type III changes of deviance. By looking at the magnitude of the change of deviance for each main and interaction effect, we can informally determine which factors have the most impact on coverage probability. We analyze the coverage probabilities of 95% credible intervals of β_1 and β_2 in this way. We perform a similar analysis for β_3 and β_4 but using only observations generated at the high level of `Prop` (removing the effects involving `Prop` from the linear predictor). The reason is that all credible intervals for β_3 and β_4 calculated for datasets with low `Prop` contain zero, the true value of these coefficients when `Prop` is low. Therefore, we can't estimate interaction effects involving Prop. We do not analyze the coverage probabilities of the credible intervals for β_5 and β_6 because they all contain zero, their true values.

Our analysis suggests that, for important variables, `SNR` has by far the most impact on the coverage probability of credible intervals for their coefficients; see tables A.1 and A.2. Figure 4.1 shows the effect of the interaction between `SNR` and `SS`, the (distant) second-most important factor, on the coverage probability of the 95% credible interval for β_1. This interaction plot and the others that follow are constructed using the `emmeans::emmip` function [11], which estimates marginal mean responses at specified levels of a subset of factors by averaging the predicted response on the scale of the link function over all combinations of levels of other factors and then back-transforming using the inverse link function. (The plots for β_2, β_3, and β_4 look similar and are thus omitted.) The coverage probability of the credible intervals exceeds the nominal level of 95% when `SNR` is low, which indicates that the intervals are valid but conservative in this setting. However, the coverage probability is much lower than 95% when `SNR` is high.

We model our second response, the estimated bias of the posterior mean of a model coefficient, using a linear regression model:

$$Y_i^{(2)} = \eta_i^{(2)} + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, (\sigma')^2)$ and the ε_i's are independent. We use ANOVA tables with type III errors as above to analyze the bias of the estimators of β_1 and β_2. We analyze β_3 and β_4 using only observations generated at the high level of `Prop` (removing the effects with `Prop` from the model). The reason is that the estimated biases of the posterior means of β_3 and β_4 are all very close to zero when `Prop` is low, while the importance of x_3 and x_4 changes with `Prop`. Therefore, the analysis for β_3 and β_4 using all datasets would find that `Prop` has a very strong impact on the bias of the posterior means, which is already explained by the fact that x_3 and x_4 are unimportant when `Prop` is low and important when `Prop` is high. This effect could mask insights about the effects of other factors on the bias of the posterior means of these coefficients. We also omit the analyses for β_5 and β_6 because the biases of the estimated posterior means for these coefficients are all very close to zero. The analyses

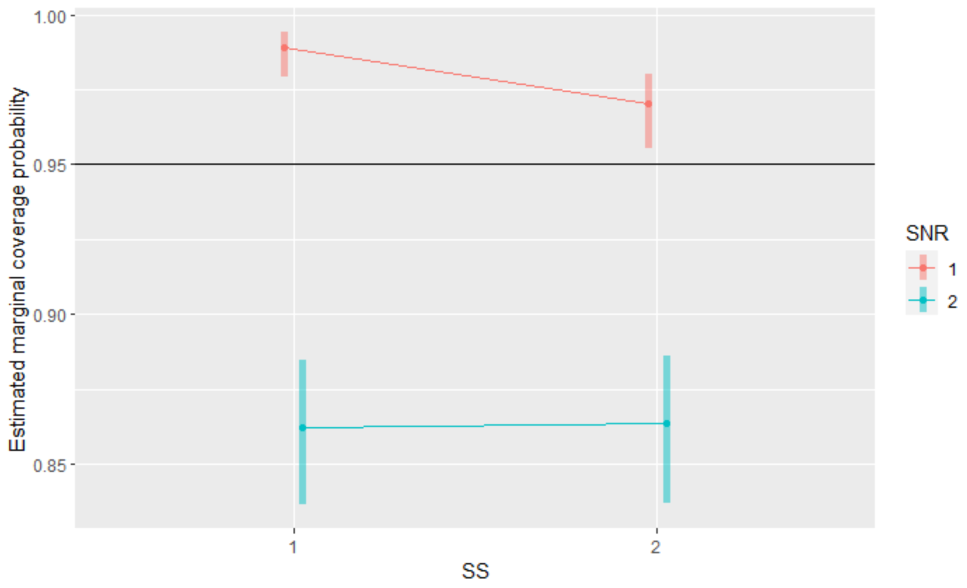

Figure 4.1: Interaction plot showing the effects of SNR and SS on the coverage probability of the 95% credible interval for β_1.

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 $\boldsymbol{\beta}$ is set to be zero in the algorithm. In general, higher SNR is associated with less bias (particularly when Cor is low), and higher Cor is associated with more bias.

We now determine factors that impact the selection performance of the annealed SMC algorithm. To this end, we define the following three responses of interest. For the i-th dataset, let $Y_i^{(3)}$ be the binary selection indicator for a chosen candidate predictor, F_i be the number of selected predictors that are not important, and M_i be the number of important predictors that are not selected. Additionally, let S_i be the number of predictors that are selected, and let t_i be the number of important predictors (so that $t_i = 2$ when Prop is low and $t_i = 4$ when Prop is high).

We model the binary indicator for the selection of an individual predictor $Y_i^{(3)}$ using a GLM:

$$Y_i^{(3)} \sim Bern(p_i^{(3)}),$$

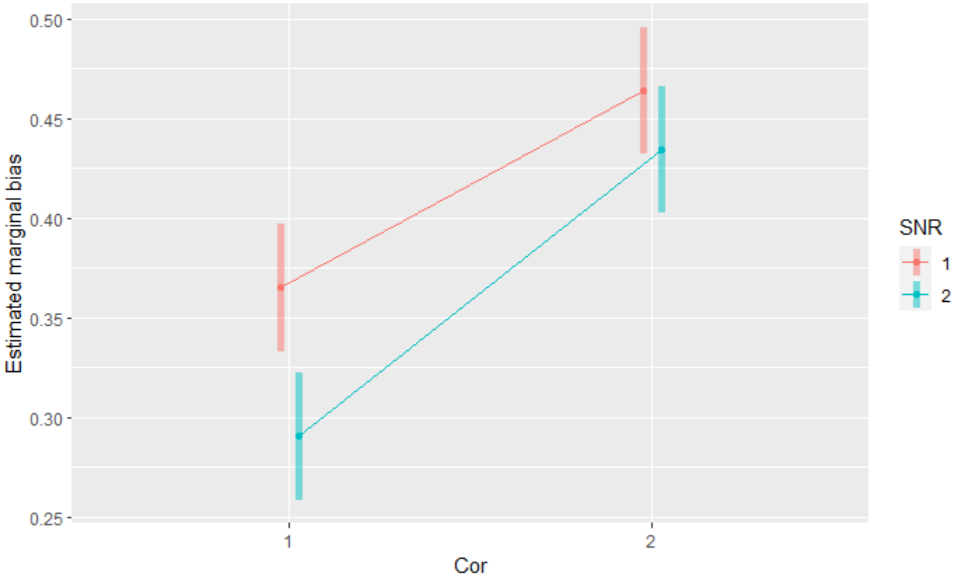

Figure 4.2: Interaction plot showing the effects of SNR and Cor on the bias of the posterior mean of β_1 as an estimator for β_1.

where

$$\operatorname{logit}(p_i^{(3)}) = \eta_i^{(3)}$$

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 with lower selection probabilities of these predictors, where the effect is more pronounced when `Prop` is high. Higher `Prop` is associated with higher selection probabilities of these predictors, especially when `SNR` is low.

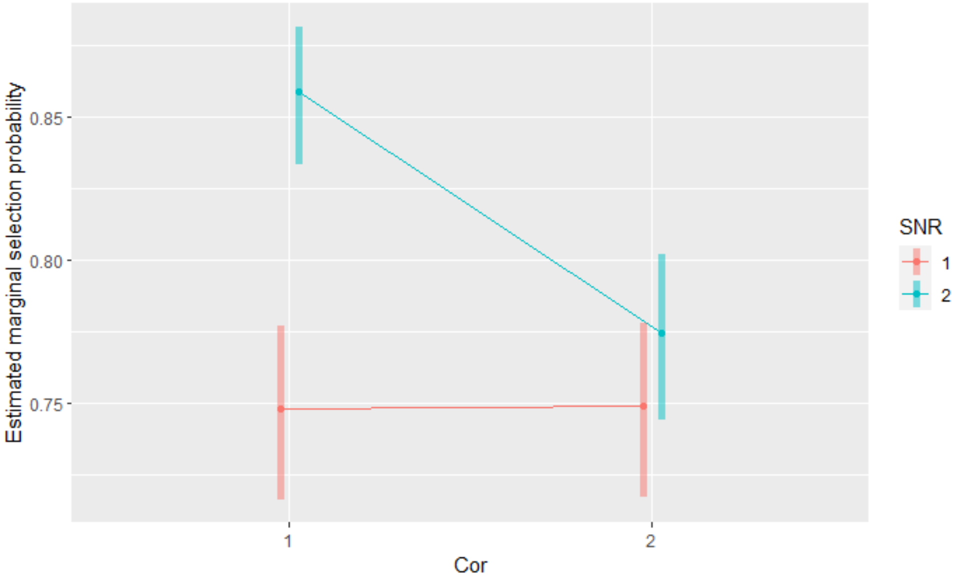

Figure 4.3: Interaction plot showing the effects of `SNR` and `Cor` on the selection probability of x_1.

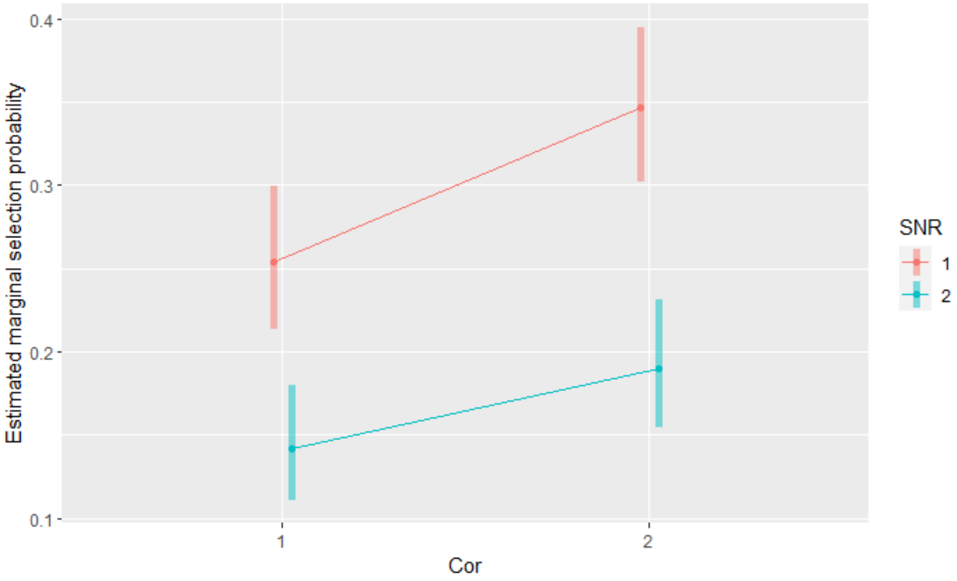

Figure 4.4: Interaction plot showing the effects of `SNR` and `Cor` on the selection probability of x_3 when `Prop` is low.

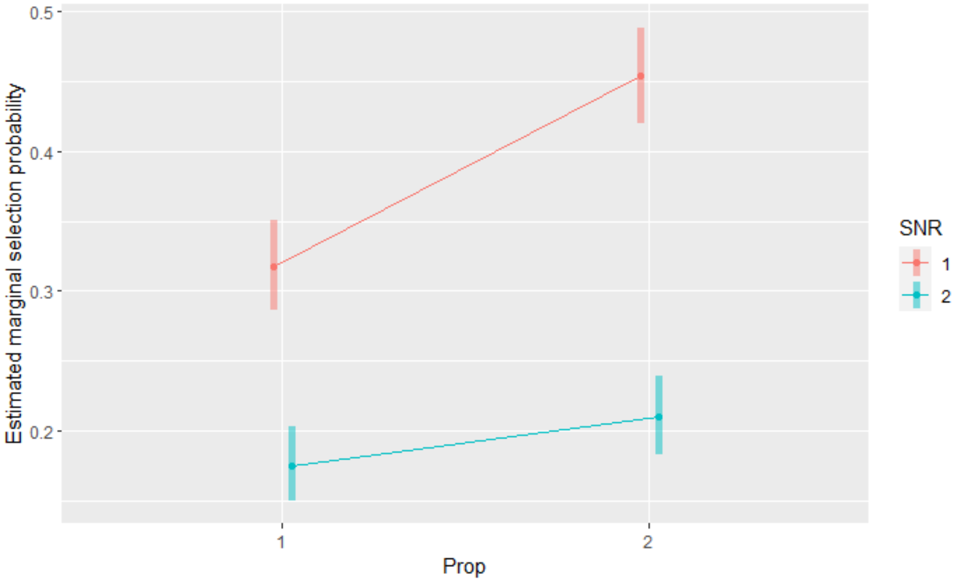

Figure 4.5: Interaction plot showing the effects of `SNR` and `Prop` on the selection probability of x_5.

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

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

and

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

where

$$\mathrm{logit}(p_i^{(4)}) = \eta_i^{(4)}$$

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 lower when a greater percentage of variables are important because the chance of selecting an unimportant variable is inherently lower. Therefore, this analysis is stratified by `Prop` with the effects including `Prop` removed from the linear predictor. Analysis of deviance tables are used in a similar manner as before; `SNR` and `Cor` have the most impact on false selection rate. Figures 4.6 and 4.7 show the effects of these factors on the logit of the false selection rate when `Prop` is low and high, respectively. Higher `SNR` is associated with lower false selection rate, with the difference being more pronounced when `Cor` is low, and higher `Cor` is associated with higher false selection rate, especially when `SNR` is high.

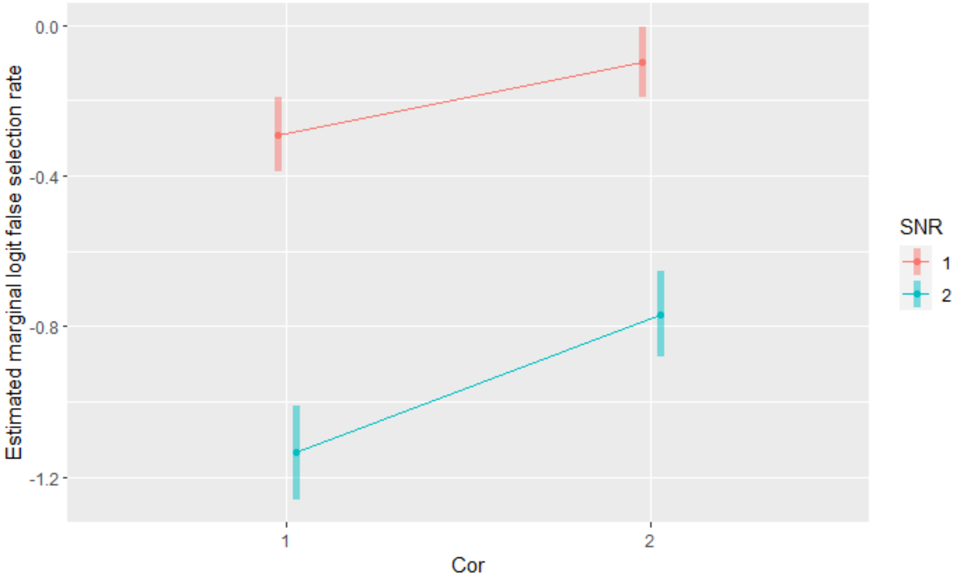

Figure 4.6: Interaction plot showing the effects of `SNR` and `Cor` on the logit false selection rate when `Prop` is low.

We similarly model the number of important predictors that are not selected, M_i, using a quasi-binomial regression approach:

$$\mathbb{E}M_i = t_i p_i^{(5)}$$

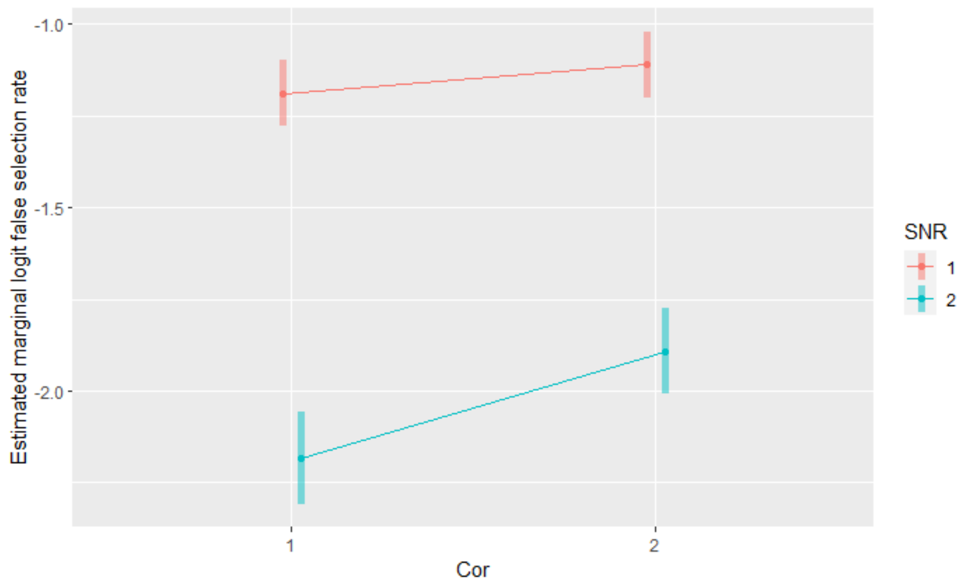

Figure 4.7: Interaction plot showing the effects of SNR and Cor on the logit false selection rate when Prop is high.

and

$$Var(M_i) = \phi' t_i p_i^{(5)}(1 - p_i^{(5)}),$$

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 associated with worse point and interval estimation performance. In particular, the posterior means of the model coefficients have larger biases when SNR is low and Cor is high, while the credible intervals are invalid when SNR is high. Additionally, higher SNR is associated with improved individual and joint selection performance (i.e., higher probabilities of selecting important predictors, lower probabilities of selecting unimportant predictors, lower false selection rate, and lower negative selection rate), while higher Cor is associated with worse

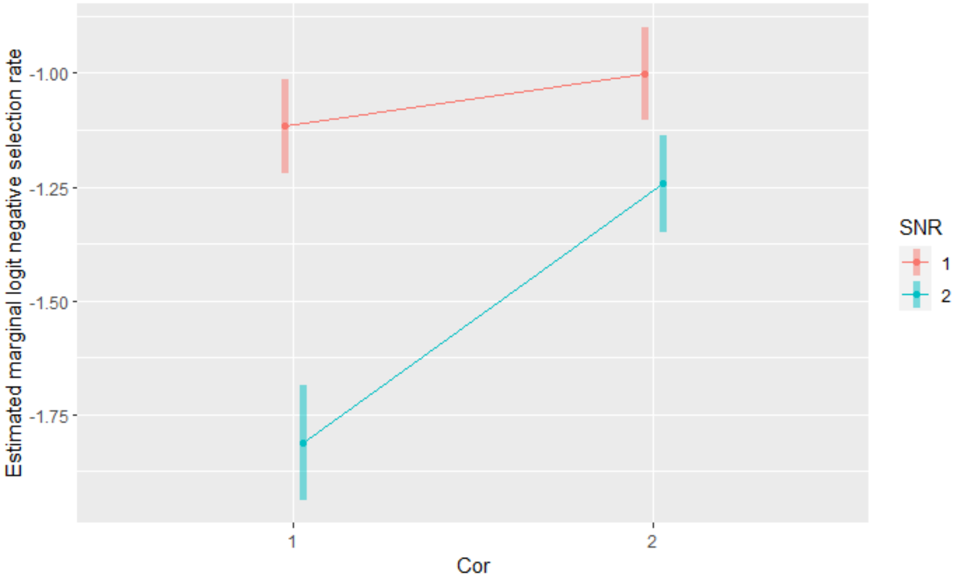

Figure 4.8: Interaction plot showing the effects of SNR and Cor on the logit negative selection rate.

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

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

$j = 1, ..., 4$, $i = 1, ..., n$, where $\varepsilon_{ij} \sim N(0, \sigma^2)$ 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 $\alpha_b = 200$, $\nu_b = 2.1$ and $\sigma_4^2 = 0.25$. 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-effects models was about 70 minutes.

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

Candidate predictor	**Estimated marginal selection probability**	**Standard error**
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-effects case.

Selection error metric	**Estimate**	**Estimated standard deviation**
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 coefficients for the mixed-effects case.

Model coefficient	**Estimated coverage probability**	**Standard error**
β_1	0.840	0.037
β_2	0.800	0.040
β_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 algorithm with similar settings to section 4.1 has deteriorated performance when applied to a linear mixed-effects model. Although the estimated marginal selection probabilities, false selection rate, and negative select rate might be acceptable in some applications, credible intervals for the effects of the important predictors are not even close to valid. In other words, as a parameter estimation method, annealed SMC is unreliable even in this very simple longitudinal setting. Exploring its performance in more complex longitudinal settings therefore seemed unnecessary, and we abandoned our original goal of conducting a comprehensive exploration in this context.

4.4 Effect of the number of particles

Although the annealed SMC algorithm performs very well in our simple benchmark setting (section 4.1), its performance is poorer when there are more candidate predictors (section 4.2) and when there is a random intercept (section 4.3). However, the algorithm may require more particles in these more complex settings to achieve excellent performance. Hence a small, additional simulation study is conducted where 200 datasets are generated according to the model described in section 4.2 for high levels of `SS` and `SNR` and low levels of `Prop` and `Cor` (the combination of levels of the factors that led to the best selection and estimation performance). The annealed SMC algorithm is applied to the datasets with the same settings as in sections 4.1 and 4.2, but with 6000 particles instead. The computing time was about five hours for 200 datasets – the same amount of time required to analyze 500 datasets using only 1000 particles.

Tables 4.7–4.9 show the resulting summary statistics of the variable selection and parameter estimation measures. For one replicate, no variables are selected, in which case the estimated false selection rate is undefined. We omit this replicate for the purpose of estimating the false selection rate.

The estimated selection probabilities for the important predictors, x_1 and x_2, and the estimated negative selection rate improved compared to those reported in section 4.2. Most importantly, the coverage probabilities of the credible intervals for the important predictors are much closer to the nominal 95% level. On the other hand, the selection performance for the unimportant predictors is worse, as reflected by the higher selection probabilities of the unimportant predictors, x_3, x_4, x_5, and x_6, and the higher estimated false selection rate; the reason for these results are unclear. This finding shows that the coverage probability of credible intervals produced by the annealed SMC method may be improved via tuning. We note that if 6000 particles were to be used in the mixed-effects case, the computation time would increase immensely because (1) the random intercepts need to be sampled as part of the algorithm, therefore increasing the time required to propagate a particle, and (2) the time complexity of the algorithm scales linearly with the number of particles.

Table 4.7: Estimated marginal selection probabilities of candidate predictors using 6000 and 1000 particles.

	Estimated marginal selection probability (Standard error)	
Candidate predictor	**6000 particles**	**1000 particles**
x_1	0.910 (0.020)	0.860 (0.025)
x_2	0.945 (0.016)	0.855 (0.025)
x_3	0.170 (0.027)	0.115 (0.023)
x_4	0.140 (0.025)	0.085 (0.020)
x_5	0.190 (0.028)	0.160 (0.026)
x_6	0.155 (0.026)	0.145 (0.025)

Table 4.8: Estimated false and negative selection rates using 6000 and 1000 particles.

	Estimated mean (Estimated SD)	
Selection error metric	**6000 particles**	**1000 particles**
False selection rate	0.215 (0.206)	0.186 (0.227)
Negative selection rate	0.073 (0.190)	0.1425 (0.242)

Table 4.9: Estimated coverage probabilities of 95% credible intervals of model coefficients using 6000 and 1000 particles.

	Estimated coverage probability (Standard error)	
Model coefficient	**6000 particles**	**1000 particles**
β_1	0.910 (0.020)	0.900 (0.021)
β_2	0.960 (0.014)	0.895 (0.022)
β_3	1.000 (0.000)	1.000 (0.000)
β_3	1.000 (0.000)	1.000 (0.000)
β_3	1.000 (0.000)	1.000 (0.000)
β_3	1.000 (0.000)	1.000 (0.000)

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 $\varphi(\omega)$, $\psi = \mathbb{E}[\varphi(\omega)]$ where $\omega \sim \pi^T$ and φ is a function of interest, using P particles $\omega_1^T, ..., \omega_P^T$ produced by the annealed SMC algorithm, with the estimator $\hat{\psi} = \sum_{i=1}^{P} W_T^i \varphi(\omega_i^T)$. Then one can show that

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

for some asymptotic variance $(\sigma^*)^2$ in the limit of the number of particles [14]. By choosing $\phi_A(x) = I(x \in A)$ 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 using 6000 particles described in section 4.4 led to estimated coverage probabilities of the credible intervals that are considerably closer to the nominal 95% level, suggesting that the performance of the method may be improved via tuning. Note that despite the parallel structure of the annealed SMC algorithm, its time complexity is linear in the number of particles. For example, the computation time would increase six-fold for the simulation studies described in sections 4.1–4.3 had 6000 particles been used. Again, we limit the number of particles used in this project due to the sheer number of simulated datasets analyzed.

Another potential avenue for improvement is the choice of the π^t-invariant forward kernel K_t. The particular kernel described in section 3.2.2 has step-size parameters that are subject to tuning. Furthermore, the kernels may be constructed using a different formulation entirely (the kernels in section 3.2.2 are essentially RJMCMC kernels, which may perform poorly when between-model moves are not carefully constructed [18]). While the performance of annealed SMC is robust to the choice of forward kernels, choosing well-mixing forward kernels will optimize the performance of the method [18].

The choice of prior distributions also impacts the performance of the annealed SMC algorithm. Selecting uninformative priors for the intercept and variance parameters is fairly uncontroversial. However, selecting τ, the prior variance of $\boldsymbol{\beta}$, is more delicate. For this project, τ is set to be 1 following recommendations in the literature [8]. This prior distribution may be too informative and may lead to poor estimation performance when the true model coefficients are large on the scale of the standardized predictors. However, setting τ to be large also presents some issues. In particular, Bartlett's paradox refers to the phenomenon when setting τ to be large causes the posterior probability of the empty model to dominate the posterior distribution on the model space [12]. An alternative is to impose a hyperprior on τ [3]. Preliminary testing of this method shows that this approach may have good selection performance while maintaining good estimation performance, but the prior on τ makes interpreting the prior distribution of the model coefficients challenging.

Therefore, setting τ remains at the discretion of the analyst.

The results of the experiment we described in section 4.2 are mostly as expected. Specifically, the selection and estimation performances of annealed SMC generally improve with larger `SNR`, and they generally worsen when the predictors are correlated. These associations are demonstrated by the predictor selection probabilities, the false and negative selection rates, and the bias of the estimated posterior mean across levels of these factors. A notable exception is the coverage probabilities of the credible intervals, which drop below the nominal level when `SNR` is high. We note that these results should be interpreted with some caution. As demonstrated previously, the estimation performance of the method may be improved via tuning, so the unexpected association between `SNR` and coverage probability may occur because the method is not optimally tuned when `SNR` is high.

Also, the analysis finds that `Prop` has a strong impact on some selection and estimation performance metrics. We note that σ^2 is larger for models with more important predictors at the same `SNR` due to how `SNR` is defined and how the true coefficients are set. Therefore, the apparent effect of the proportion of important predictors might, in fact, be due to the effect of a larger σ^2. For this project, we decided to set levels of `SNR` instead of σ^2 because preliminary testing showed that the relative sizes of the coefficients and σ^2 impact the performance of the algorithm, and `SNR` is a useful summary of the relative sizes of these parameters. Further studies that examine the effects of particular values of σ^2 and the true coefficients would advance understanding of this observation.

Since `SNR` has such a pronounced impact on the performance of annealed SMC, in practice, determining the approximate value of `SNR` (by fitting the full model) may be useful before attempting variable selection. In particular, knowledge of `SNR` will provide information about the likely performance of the method. Further study of variable selection and estimation methods that work well regardless of the magnitude of `SNR` are also of interest.

The relative impact of the factors may depend on the levels at which the factors are set, which may result in unfair comparisons of the impact of different factors. For example, the large sample size level is merely double that of the small sample size level, while the large `SNR` level is 100 times larger than the small `SNR` level. A future simulation study may choose more levels to study the impacts of these factors in more depth.

We remark here that in the simulation studies, the computed credible intervals for the coefficients of selected predictors are rather wide. For example, for the datasets generated using the settings described in section 4.4, we fit the model containing only the important predictors using a Bayesian approach, with similar priors for the coefficients as described in 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 the direction of the effect of the predictors, but are not as useful in determining the size of the effect. The relatively larger width of the credible intervals is expected because the posterior distributions of coefficients are averaged across different models, while the posterior distributions of coefficients given each model are concentrated around different peaks. But the degree of discrepancy in widths is perhaps surprising.

Annealed SMC sampling targets the posterior distribution of models and parameters. In other words, annealed SMC sampling can perform no better than the direct (if practically infeasible) method of explicitly computing the posterior model probabilities and the averaged posterior distributions of the model coefficients, selecting predictors with marginal inclusion probability exceeding some threshold, and then computing 95% credible intervals for the model coefficients. However, not all variable selection methods solely target the posterior distribution of models and parameters. For example, the split-and-merge method does not directly use the posterior model probabilities to select predictors; its performance may exceed that of methods that explicitly compute posterior model probabilities. On the other hand, without posterior model probabilities, credible intervals for the model coefficients cannot be constructed using the original dataset; new data are required to fit the model that includes only the selected predictors. Therefore, a future research direction of interest would be extensions of methods such as split-and-merge that provide both good selection performance and valid interval estimates of the model coefficients.

We recommend that this method be used to analyze real data when the signal-to-noise ratio is high and the correlation of predictors is low. The simulation studies in this project suggest that the method can reliably select important predictors and produce valid (if not very informative) credible intervals for model coefficients under these conditions. (Under other conditions, the performance of the method may be poor.) The signal-to-noise ratio may be estimated by fitting the model with all candidate predictors to the data. The pairwise correlation of predictors may be estimated by calculating the empirical variance-covariance matrix of the candidate predictors.

When analyzing a single dataset, we encourage practitioners to use much larger numbers of intermediate distributions and particles than we considered in our work. To determine reasonable choices for these numbers, a simulation study could be done akin to that in section 4.1. First, a preliminary analysis of the real dataset should be done using annealed SMC sampling with arbitrary tuning parameters to obtain a preliminary set of selected predictors. Then, the model containing only the preliminary selected predictors should be fit to the real dataset, and simulated datasets should be generated using this fitted model. Finally, the method should be applied to the simulated datasets to examine its selection and estimation performance. The factors of interest in this study would be the number (and/or spacing) of intermediate distributions and the number of particles. While the performance of the method generally improves with more intermediate distributions and particles, it should stabilize when sufficiently many intermediate distributions and particles are used. The practitioner should find the smallest number of intermediate distributions and the smallest number of particles such that the performance of the method does not improve significantly beyond them and use this combination for the final analysis of the real dataset.

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 Statistics*, 25(3):701–726, 2016.

Appendix A

Analysis of deviance tables

Table A.1: Analysis of deviance table for the estimated coverage probability of 95% credible intervals for β_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
`Cor`	9.76	1.00	0.00
`SS:SNR`	6.84	1.00	0.01
`SS:Prop`	3.71	1.00	0.05
`SS:Cor`	1.01	1.00	0.32
`SNR:Prop`	0.23	1.00	0.63
`SNR:Cor`	0.01	1.00	0.91
`Prop:Cor`	0.47	1.00	0.50

Table A.2: Analysis of deviance table for the estimated coverage probability of 95% credible intervals for β_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 β_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 β_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 β_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 β_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 β_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 β_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
Residual	368.65	1593.00		

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`	1.23	1.00	0.27
`SNR:Cor`	0.15	1.00	0.70

Table A.12: Analysis of deviance table for the estimated selection probability of x_3 when `Prop` is high.

Effect	Change in deviance	d.f.	p-value
`SS`	0.38	1.00	0.54
`SNR`	47.02	1.00	0.00
`Cor`	18.24	1.00	0.00
`SS:SNR`	6.68	1.00	0.01
`SS:Cor`	0.84	1.00	0.36
`SNR:Cor`	4.41	1.00	0.04

Table A.13: Analysis of deviance table for the estimated selection probability of x_4 when `Prop` is low.

	Change in deviance	d.f.	p-value
`SS`	1.03	1.00	0.31
`SNR`	89.69	1.00	0.00
`Cor`	10.55	1.00	0.00
`SS:SNR`	0.94	1.00	0.33
`SS:Cor`	1.75	1.00	0.19
`SNR:Cor`	2.62	1.00	0.11

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

Effect	Change in deviance	d.f.	p-value
SS	1.27	1.00	0.26
SNR	17.97	1.00	0.00
Cor	17.77	1.00	0.00
SS:SNR	0.07	1.00	0.79
SS:Cor	0.85	1.00	0.36
SNR:Cor	3.88	1.00	0.05

Table A.15: Analysis of deviance table for the estimated selection probability of x_5.

Effect	Change in deviance	d.f.	p-value
SS	2.89	1.00	0.09
SNR	142.50	1.00	0.00
Prop	23.85	1.00	0.00
Cor	0.85	1.00	0.36
SS:SNR	2.33	1.00	0.13
SS:Prop	1.48	1.00	0.22
SS:Cor	0.00	1.00	0.98
SNR:Prop	4.71	1.00	0.03
SNR:Cor	1.94	1.00	0.16
Prop:Cor	0.78	1.00	0.38

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.17: Analysis of deviance table for the estimated false selection rate when Prop is low.

Effect	Change in deviance	d.f.	p-value
SS	12.38	1.00	0.00
SNR	195.13	1.00	0.00
Cor	25.98	1.00	0.00
SS:SNR	6.03	1.00	0.01
SS:Cor	0.05	1.00	0.82
SNR:Cor	2.53	1.00	0.11

Table A.18: Analysis of deviance table for the estimated false selection rate when Prop is high.

Effect	Change in deviance	d.f.	p-value
SS	1.11	1.00	0.29
SNR	276.74	1.00	0.00
Cor	11.28	1.00	0.00
SS:SNR	2.35	1.00	0.13
SS:Cor	0.18	1.00	0.68
SNR:Cor	3.75	1.00	0.05

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