Title
Likelihood Free Inference for a Flexible Class of Bivariate Beta Distributions

Author
Crackel, Roberto Carlos

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Likelihood Free Inference for a Flexible Class of Bivariate Beta Distributions

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Roberto Carlos Crackel

March 2015

Dissertation Committee:

Dr. James Flegal , Chairperson
Dr. Barry C. Arnold
Dr. Gregory J. Palardy

The Dissertation of Roberto Carlos Crackel is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

First and foremost, I thank God for all the blessings He has given me. I thank Him for my loving and supportive family and friends, and the opportunity and ability to complete this program and the production of this dissertation.

I would like to give my sincerest thanks and appreciation to my advisor, Dr. James Flegal, for his guidance and support in seeing me through my qualifying exams and the successful completion of this dissertation. I wish him all the best with his career and family. I thank all of my committee members, Dr. Barry Arnold, for his words of wisdom, and for his role in providing the problem for this dissertation, Dr. Gregory Palardy, for his support and kindness, Dr. Subir Ghosh, for his academic advice and words of encouragement and Dr. Kurt Schwabe, for his input during my oral exam. I would also like to extend my thanks to the rest of the statistics faculty and staff for the invaluable education, experiences and words of encouragement that I received.

I would like to thank my grandmother, Ellen M. Crackel, and uncle, Lyndon D. Crackel for their prayers and support. I thank my aunt, Rosa M. Brewer for her prayers, encouragement and confidence in me. I would also like to extend my gratitude and thankfulness to my entire family, for their prayers and extensive support. I thank all the friends with whom I have made lasting friendships with during my life, and for their support.

Finally, there are three people that I can never truly thank enough, and for whom I am so blessed and grateful to have in my life. I would like to thank my father, Garland R. Crackel, my mother, Carlota Esmeralda Crackel, and sister, Tanya C. Rayfield (Crackel), for their insurmountable love, prayers, support, guidance in life, and for never doubting my ability to complete this program and dissertation.

This dissertation is dedicated to my father, $\mathscr{G}$arland $\mathscr{R}$. $\mathscr{C}$rackel, my mother, $\mathscr{C}$arlota $\mathscr{E}$smeralda $\mathscr{C}$rackel, and sister, $\mathscr{T}$anya $\mathscr{C}$. $\mathscr{R}$ayfield ($\mathscr{C}$rackel).

ABSTRACT OF THE DISSERTATION

Likelihood Free Inference for a Flexible Class of Bivariate Beta Distributions

by

Roberto Carlos Crackel

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, March 2015
Dr. James Flegal , Chairperson

Several bivariate beta distributions have been proposed in the literature. In particular, Olkin and Liu (2003) proposed a 3 parameter bivariate beta model, which Arnold and Ng (2011) extend to 5 and 8 parameter models. The 3 parameter model allows for only positive correlation, while the latter models can accommodate both positive and negative correlation. However, these come at the expense of a density that is mathematically intractable. The focus of this dissertation is on Bayesian estimation for the 5 and 8 parameter models. Since the likelihood does not exist in closed form, we apply approximate Bayesian computation, a likelihood free approach.

Chapter one briefly describes the univariate beta distribution and its properties. The 5 and 8 parameter bivariate beta distribution is defined and estimation strategies are discussed. Chapter two is dedicated to the background of approximate Bayesian computation (ABC), where the foundation and groundwork is laid. Toy examples are provided to better understand the algorithm and to study its properties. Chapter three is the application of ABC to the 5 and 8 parameter bivariate beta model. Simulation studies have been carried out for the 5 and 8 parameter cases under various priors, sample sizes, and tolerance levels. We apply the 5 parameter model to a real data

set by allowing the model to serve as a prior to correlated proportions of a bivariate beta binomial model. Results and comparisons are then discussed. Chapter four attempts to lay the ground work to modify existing ABC (accept reject) algorithms to search for maximum likelihood type estimates in the absence of the likelihood function. Examples are provided to demonstrate the relationship between maximum likelihood estimation and acceptance rates. Algorithms are proposed and applied to data sets in an attempt to search for maximum likelihood type estimates using only sufficient statistics. Results are compared to the known maximum likelihood estimates.

Contents

List of Figures

List of Tables

Chapter 1

The beta and bivariate beta distributions

In this chapter, we briefly describe the univariate beta distribution and its properties. We then provide a background discussion to the bivariate beta distribution. In particular, we mention various definitions of the bivariate beta distribution that authors have defined in the literature. Specifically, we focus on the 5 parameter bivariate beta model defined by Arnold and Ng (2011). We then discuss the properties of this model and estimation methods developed by Arnold and Ng (2011). We then briefly discuss the 8 parameter model and an extension to the k-variate beta distribution defined by Arnold and Ng (2011).

1.1 The beta distribution (of the first kind)

The beta distribution (of the first kind) is a continuous distribution, which has support on the unit interval, and is controlled by two positive shape parameters a and b. The beta distribution is used to model the random behavior of proportions such as a batters batting average or a basketball players percentage of free throws made. In Bayesian inference, the beta distribution serves as the conjugate prior distribution to the binomial, negative binomial and geometric distribution. A random variable X follows a beta distribution, if it has the following density

$$f_X(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}; 0 < x < 1, a > 0, b > 0 \, .$$

Here, $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$ is the gamma function. If X follows a beta distribution, we denote as $X \sim Beta(a, b)$. Furthermore, the first two moments and the variance, which will be useful in parameter estimation are

$$E(X) = \frac{a}{a+b}, \qquad E(X^2) = \left(\frac{a}{a+b}\right)\left(\frac{a+1}{a+b+1}\right) \tag{1.1}$$

$$Var(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

1.1.1 Related distributions

We can derive the beta distribution using the well known gamma distribution. A random variable X follows a gamma distribution, if it has the following density

$$f_X(x; a, \theta) = \frac{x^{a-1}e^{-x/\theta}}{\Gamma(a)\theta^a}; 0 < x < \infty, a > 0, \theta > 0.$$

Where a is the shape parameter and θ is the scale parameter. If X follows a gamma distribution, we denote as $X \sim \Gamma(a, \theta)$. Suppose $X_1 \sim \Gamma(a, \theta)$ and $X_2 \sim \Gamma(b, \theta)$ and are independent, then

$$\frac{X_1}{X_1 + X_2} \sim Beta(a, b).$$

This definition will play a vital role in the construction of the bivariate beta distribution, which will be discussed in the next section.

A useful transformation is known as the beta distribution of the second kind. The support for this distribution is the entire positive real line and serves as the conjugate prior for Bernoulli trials expressed in odds. The beta distribution of the second kind is derived by letting $X \sim Beta(a, b)$, and so

$$Y = \frac{X}{1-X}$$

is said to follow a beta distribution of the second kind, which we denote as $\beta'(a, b)$. The density of Y is given by

$$f_Y(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1+y)^{-(a+b)}; 0 < y < \infty, a > 0, b > 0.$$

The expectation and variance of Y is

$$E(Y) = E\left(\frac{X}{1-X}\right) = \frac{a}{b-1}, \text{ conditional on } b > 1$$

and

$$Var(Y) = \frac{a(a+b-1)}{(b-2)(b-1)^2}, \text{ conditional on } b > 2.$$

Furthermore, it is straightforward to show

$$\frac{1}{Y} \sim \beta'(b, a).$$

These properties will be exploited in one of the estimation methods for the 5 parameter bivariate beta model.

1.1.2 Estimation

If $X \sim Beta(a, b)$, then we are interested in the estimation of a and b. There are two primary methods of estimation, the first is maximum likelihood estimation (MLE) and the second is the method of moments (MOM). We will now discuss these methods in detail. The MLE for (a,b) is obtained via the 2-dimensional sufficient statistic

$$\left(\prod_{i=1}^{n} X_i, \prod_{i=1}^{n}(1-X_i)\right) \quad \text{or equivalently} \quad \left(\sum_{i=1}^{n} \log X_i, \sum_{i=1}^{n} \log(1-X_i)\right)$$

and then obtained by solving the nonlinear equations

$$\psi(a) - \psi(a+b) = \frac{1}{n}\sum_{i=1}^{n} \log x_i \quad \text{and} \quad \psi(b) - \psi(a+b) = \frac{1}{n}\sum_{i=1}^{n} \log(1-x_i)$$

where $\psi(\cdot)$ is the digamma function. We denote the MLE's of a and b as $\hat{a}$ and $\hat{b}$, respectively. The MOM estimators are obtained by setting the first two theoretical moments at (1.1) to the sample moments, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{a}{a+b} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \left(\frac{a}{a+b}\right)\left(\frac{a+1}{a+b+1}\right).$$

This yields the solutions to a and b as follows

$$\tilde{a} = \bar{X}\left(\frac{\bar{X}(1-\bar{X})}{S_X^2} - 1\right) \quad \text{and} \quad \tilde{b} = (1-\bar{X})\left(\frac{\bar{X}(1-\bar{X})}{S_X^2} - 1\right)$$

where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X}_i)^2.$$

1.2 The bivariate beta distribution

Bivariate beta distributions are becoming increasingly popular across many disciplines. Furthermore, it is common for Bayesian analysts to use them as prior distributions to correlated binomial random variables. An incomplete list of bivariate beta distributions includes use of the Dirichlet distribution, as well as those studied by Arnold and Ng (2011), Jones (2002), Olkin and Liu (2003), and Nadarajah and Kotz (2005). Furthermore, Gupta and Wong (1985) defined a bivariate beta distribution from the Morgenstern system of bivariate distributions (Morgenstern (1956)). Likewise

Ting Lee (1996) defined a bivariate beta distribution from the Sarmanov system of bivariate distributions (Sarmanov (1966)). Also, Gupta et al. (2011) considered a non-central bivariate beta model. An interested reader is directed to Balakrishnan and Lai (2009) for an extensive list of bivariate beta models along with other bivariate continuous distributions.

Unfortunately, many bivariate beta models contain parameter and correlation restrictions and hence may not be suitable in applications. For example, suppose $\boldsymbol{Z} = (Z_1, Z_2)$ defines a bivariate beta random vector. Then, it is well known that if $\boldsymbol{Z}$ follows a Dirichlet distribution, the marginals are beta distributed with $z_1+z_2 = 1$. Further, the family of bivariate distributions of Morgenstern (1956) has a limited correlation range of $(-1/3, 1/3)$, as shown by Schucany et al. (1978), and the model of Olkin and Liu (2003) only allow for positive correlation.

The focus of this dissertation is on parameter estimation for the flexible 5 and 8 parameter models of Arnold and Ng (2011), which extend the 3 parameter specification of Olkin and Liu (2003). The models of Arnold and Ng (2011), allow for both positive and negative correlation, that is, any correlation in $(-1, 1)$. The cost of this increased flexibility is a joint density unavailable in closed form, but simulating pseudo-random observations is trivial.

1.2.1 The 5 parameter model

Arnold and Ng (2011) defined the proposed 5 parameter bivariate beta distribution by letting $U_i \overset{ind}{\sim} \Gamma(\alpha_i, \theta), i = 1, ..., 5$ (without loss of generality, we can assume $\theta = 1$). The bivariate random vector $\boldsymbol{Z} = (Z_1, Z_2)$, as a function of the U_i's is defined as follows

$$Z_1 = \frac{U_1 + U_3}{U_1 + U_3 + U_4 + U_5} \text{ and } Z_2 = \frac{U_2 + U_4}{U_2 + U_3 + U_4 + U_5}.$$

Therefore, $Z_1 \sim Beta(\alpha_1 + \alpha_3, \alpha_4 + \alpha_5)$ and $Z_2 \sim Beta(\alpha_2 + \alpha_4, \alpha_3 + \alpha_5)$. We denote the 5 parameter model as $\mathcal{BB}(\alpha_1, ..., \alpha_5)$ or $\mathcal{BB}(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)'$. Furthermore,

$$W_1 = \frac{Z_1}{1 - Z_1} = \frac{U_1 + U_3}{U_4 + U_5} \sim \beta'(\alpha_1 + \alpha_3, \alpha_4 + \alpha_5)$$

and

$$W_2 = \frac{Z_2}{1 - Z_2} = \frac{U_2 + U_4}{U_3 + U_5} \sim \beta'(\alpha_2 + \alpha_4, \alpha_3 + \alpha_5).$$

It also follows that

$$\frac{1}{W_1} \sim \beta'(\alpha_4 + \alpha_5, \alpha_1 + \alpha_3) \quad \text{and} \quad \frac{1}{W_2} \sim \beta'(\alpha_3 + \alpha_5, \alpha_2 + \alpha_4).$$

Furthermore, some useful moments and expectations in parameter estimation are

$$E(Z_1) = \frac{\alpha_1 + \alpha_3}{\alpha_1 + \alpha_3 + \alpha_4 + \alpha_5}$$

$$E(Z_1^2) = \left(\frac{\alpha_1 + \alpha_3}{\alpha_1 + \alpha_3 + \alpha_4 + \alpha_5}\right)\left(\frac{\alpha_1 + \alpha_3 + 1}{\alpha_1 + \alpha_3 + \alpha_4 + \alpha_5 + 1}\right)$$

$$E(Z_2) = \frac{\alpha_2 + \alpha_4}{\alpha_2 + \alpha_3 + \alpha_4 + \alpha_5}$$

$$E(Z_2^2) = \left(\frac{\alpha_2 + \alpha_4}{\alpha_2 + \alpha_3 + \alpha_4 + \alpha_5}\right)\left(\frac{\alpha_2 + \alpha_4 + 1}{\alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + 1}\right)$$

$$\begin{aligned} E\left[\frac{(1 - Z_1)(1 - Z_2)}{Z_1 Z_2}\right] &= \left(\frac{\alpha_4}{\alpha_2 + \alpha_4}\right)\left(\frac{\alpha_3}{\alpha_1 + \alpha_3}\right) + \left(\frac{\alpha_3}{\alpha_1 + \alpha_3}\right)\left(\frac{\alpha_5}{\alpha_2 + \alpha_4 - 1}\right) \\ &+ \left(\frac{\alpha_4}{\alpha_2 + \alpha_4}\right)\left(\frac{\alpha_5}{\alpha_1 + \alpha_3 - 1}\right) \\ &+ \left(\frac{\alpha_5}{\alpha_1 + \alpha_3 - 1}\right)\left(\frac{\alpha_5 + 1}{\alpha_2 + \alpha_4 - 1}\right) \end{aligned} \tag{1.2}$$

$$E\left[\frac{1-Z_1}{Z_1}\right] = \frac{\alpha_4+\alpha_5}{\alpha_1+\alpha_3-1} \tag{1.3}$$

$$E\left[\frac{1-Z_2}{Z_2}\right] = \frac{\alpha_3+\alpha_5}{\alpha_2+\alpha_4-1} \tag{1.4}$$

$$E\left[\left(\frac{1-Z_1}{Z_1}\right)^2\right] = \left(\frac{\alpha_4+\alpha_5}{\alpha_1+\alpha_3-1}\right)\left(\frac{\alpha_4+\alpha_5+1}{\alpha_1+\alpha_3-2}\right)$$

$$E\left[\left(\frac{1-Z_2}{Z_2}\right)^2\right] = \left(\frac{\alpha_3+\alpha_5}{\alpha_2+\alpha_4-1}\right)\left(\frac{\alpha_3+\alpha_5+1}{\alpha_2+\alpha_4-2}\right).$$

The lack of a closed form density eliminates maximum likelihood estimation. Arnold and Ng (2011) derived 3 methods for the estimation of $\alpha_i, i = 1, ..., 5$. The first method, which the authors coined as modified maximum likelihood estimation (MMLE), works by obtaining the MLE's based off the marginal distributions for Z_1 and Z_2, and then use a method of moment estimate to obtain parameter estimates. The second method uses the MOM estimates as a function of the sample means and sample variances (based off the marginal distributions). The third method uses the MOM estimates based on beta distributions of the second kind. For ease of notation, suppose we have n observations from our bivariate beta model. Specifically, define $\vec{\boldsymbol{z}}_1 = (z_{11}, z_{21}, \ldots, z_{n1})'$, $\vec{\boldsymbol{z}}_2 = (z_{12}, z_{22}, \ldots, z_{n2})'$, and $\tilde{\boldsymbol{z}} = (\vec{\boldsymbol{z}}_1, \vec{\boldsymbol{z}}_2)$.

a) **Modified maximum likelihood estimation (MMLE)**

The method of modified maximum likelihood estimation obtains the MLE's based off the marginals of Z_1 and Z_2, which will yield four equations. For the fifth equation (needed because we have 5 unknowns), a method of moment estimate is used. Hence, the MLE's based off Z_1 are obtained for $a = \alpha_1+\alpha_3$ and $b = \alpha_4+\alpha_5$, and are denoted by $\hat{a}$ and $\hat{b}$, respectively. Similarly, the MLE's based off Z_2 for $c = \alpha_2 + \alpha_4$ and $d = \alpha_3 + \alpha_5$ are obtained, which we denote by $\hat{c}$ and $\hat{d}$, respectively. Furthermore, let

$$\mathcal{S}(\tilde{\boldsymbol{z}}) = \frac{1}{n}\sum_{i=1}^{n}\frac{(1-z_{i1})(1-z_{i2})}{z_{i1}z_{i2}}. \tag{1.5}$$

We set the theoretical moment at (1.2) equal to the sample moment at (1.5), and after some algebra, we can set this equation in terms of $\hat{a}$, $\hat{b}$, $\hat{c}$, $\hat{d}$, and α_5 as follows

$$\begin{aligned}\mathcal{S}(\tilde{\boldsymbol{z}}) = \frac{1}{n}\sum_{i=1}^{n}\frac{(1-z_{i1})(1-z_{i2})}{z_{i1}z_{i2}} &= \left(\frac{\hat{b}-\alpha_5}{\hat{c}}\right)\left(\frac{\hat{d}-\alpha_5}{\hat{a}}\right) \\ &+ \left(\frac{\hat{d}-\alpha_5}{\hat{a}}\right)\left(\frac{\alpha_5}{\hat{c}-1}\right) \\ &+ \left(\frac{\hat{b}-\alpha_5}{\hat{c}}\right)\left(\frac{\alpha_5}{\hat{a}-1}\right) \\ &+ \left(\frac{\alpha_5(\alpha_5+1)}{(\hat{a}-1)(\hat{c}-1)}\right)\end{aligned}$$

which yields the quadratic equation

$$\alpha_5^2 + B\alpha_5 + C = 0 \tag{1.6}$$

where

$$B = \hat{b}\hat{c} + \hat{a}\hat{c} + \hat{a}\hat{d} - \hat{b} - \hat{d}$$

and

$$C = (\hat{a}-1)(\hat{c}-1)\hat{b}\hat{d} - \frac{\hat{a}\hat{c}(\hat{a}-1)(\hat{c}-1)}{n}\sum_{i=1}^{n}\frac{(1-z_{i1})(1-z_{i2})}{z_{i1}z_{i2}}.$$

Therefore, the estimate for α_5 is the solution to (1.6). Now, it is possible that the solution to (1.6) can be negative, however, recall that MLE's cannot yield estimates outside the parameter space, therefore, applying this principle, the estimate for α_5 will be the maximum of 0 and the larger root of the quadratic equation. Once an estimate for α_5 is obtained, we can then obtain point estimates for $\alpha_i, i = 1, ..., 4$.

The MMLE's for the parameters $\alpha_i, i = 1, ..., 5$ are

$$\hat{\alpha}_5 = \max\left\{0, \frac{-B+\sqrt{B^2-4C}}{2}\right\}, \quad \hat{\alpha}_4 = \max\left\{0, \hat{b}-\hat{\alpha}_5\right\}, \quad \hat{\alpha}_3 = \max\left\{0, \hat{d}-\hat{\alpha}_5\right\},$$

$$\hat{\alpha}_2 = \max\left\{0, \hat{c}-\hat{\alpha}_4\right\}, \text{ and } \hat{\alpha}_1 = \max\left\{0, \hat{a}-\hat{\alpha}_3\right\} \ . \tag{1.7}$$

b) **Method of moments based on sample means and sample variances**

We can obtain estimates for $\alpha_i, i = 1, ..., 5$ based off the sample moments obtained from the marginals of Z_1 and Z_2, as a function of the sample means and variances. Denote the sample means of Z_1 and Z_2 by

$$\bar{Z}_1 = \frac{1}{n}\sum_{i=1}^{n} Z_{i1}, \qquad \bar{Z}_2 = \frac{1}{n}\sum_{i=1}^{n} Z_{i2}$$

and denote the sample variances of Z_1 and Z_2 by

$$S^2_{Z_1} = \frac{1}{n-1}\sum_{i=1}^{n}(Z_{i1}-\bar{Z}_1)^2, \qquad S^2_{Z_2} = \frac{1}{n-1}\sum_{i=1}^{n}(Z_{i2}-\bar{Z}_2)^2.$$

The method of moment estimates for a, b, c and d are

$$\tilde{a} = \bar{Z}_1\left(\frac{\bar{Z}_1(1-\bar{Z}_1)}{S^2_{Z_1}}-1\right), \qquad \tilde{b} = (1-\bar{Z}_1)\left(\frac{\bar{Z}_1(1-\bar{Z}_1)}{S^2_{Z_1}}-1\right)$$

$$\tilde{c} = \bar{Z}_2\left(\frac{\bar{Z}_2(1-\bar{Z}_2)}{S^2_{Z_2}}-1\right), \qquad \tilde{d} = (1-\bar{Z}_2)\left(\frac{\bar{Z}_2(1-\bar{Z}_2)}{S^2_{Z_2}}-1\right).$$

By using the quadratic equation from (1.6), we can then obtain moment based estimates for $\alpha_i, i = 1, ..., 5$ by choosing the larger root of the quadratic equation and substituting the estimates with 0, if they are negative.

c) **Method of moments based on beta distributions of the second kind**

Using (1.3) and (1.4), we can set up the following moment equations

$$\bar{Z}_1 = \frac{1}{n}\sum_{i=1}^{n} Z_{i1} = \frac{\alpha_1+\alpha_3}{\alpha_1+\alpha_3+\alpha_4+\alpha_5}$$

$$\bar{Z}_2 = \frac{1}{n}\sum_{i=1}^{n} Z_{i2} = \frac{\alpha_2 + \alpha_4}{\alpha_2 + \alpha_3 + \alpha_4 + \alpha_5}$$

$$m_{Z_1} = \frac{1}{n}\sum_{i=1}^{n} \frac{1 - Z_{i1}}{Z_{i1}} = \frac{\alpha_4 + \alpha_5}{\alpha_1 + \alpha_3 - 1}$$

$$m_{Z_2} = \frac{1}{n}\sum_{i=1}^{n} \frac{1 - Z_{i2}}{Z_{i2}} = \frac{\alpha_3 + \alpha_5}{\alpha_2 + \alpha_4 - 1}$$

and we can estimate a, b, c, and d as

$$a^\star = \frac{\bar{Z}_1 m_{Z_1}}{\bar{Z}_1 m_{Z_1} + \bar{Z}_1 - 1}, \qquad b^\star = \frac{m_{Z_1}(1 - \bar{Z}_1)}{\bar{Z}_1 m_{Z_1} + \bar{Z}_1 - 1}$$

$$c^\star = \frac{\bar{Z}_2 m_{Z_2}}{\bar{Z}_2 m_{Z_2} + \bar{Z}_2 - 1}, \qquad d^\star = \frac{m_{Z_2}(1 - \bar{Z}_2)}{\bar{Z}_2 m_{Z_2} + \bar{Z}_2 - 1}.$$

Again, by using the quadratic equation from (1.6), we can obtain estimates of $\alpha_i, i = 1, ..., 5$ by choosing the larger root of the quadratic equation and substituting the estimates with 0, if they turn out to be negative.

1.3 Caution with the MMLE

Unfortunately, $\mathcal{S}(\tilde{\boldsymbol{z}})$ at (1.5) is easily influenced by observed data points near zero. For example, in our simulation studies (discussed in detail later), a particular data set of sample size 50, denoted $\mathcal{D}$, produced the bivariate observation $(z_{43,1}, z_{43,2}) = (0.1089, 0.0038)$. Clearly $z_{43,2}$ will severely inflate $\mathcal{S}(\tilde{\boldsymbol{z}})$, thus affecting the MMLE at (1.7). Furthermore, it will have an affect on our likelihood free algorithms by using $\mathcal{S}(\tilde{\boldsymbol{z}})$ as a summary statistic (which we discuss in detail later). For illustration, Table 1.1 compares summary statistics for $\mathcal{D}$ to those of a more typical data set, denoted $\mathcal{D}'$, with no observed points near zero. Notice, that the sufficient statistics for the marginal distributions of Z_1 and Z_2 are not much affected, however there is a heavy influence on $\mathcal{S}(\tilde{\boldsymbol{z}})$.

	$\sum \frac{\log z_{i1}}{50}$	$\sum \frac{\log z_{i2}}{50}$	$\sum \frac{\log(1-z_{i1})}{50}$	$\sum \frac{\log(1-z_{i2})}{50}$	$\mathcal{S}(\tilde{z})$
$\mathcal{D}$	-0.81	-1	-1	-0.76	47.77
$\mathcal{D}'$	-0.76	-0.85	-0.76	-0.84	1.67

Table 1.1: Comparison of five summary statistics between dataset $\mathcal{D}$ and dataset $\mathcal{D}'$.

1.4 The 8 parameter model

Arnold and Ng (2011), generalized the bivariate beta model by defining an 8 parameter model by letting $U_i \overset{ind}{\sim} \Gamma(\delta_i, \theta), i = 1, ..., 8$ (without loss of generality, we can assume $\theta = 1$). Now define

$$V_1 = \frac{U_1 + U_5 + U_7}{U_3 + U_6 + U_8} \quad \text{and} \quad V_2 = \frac{U_2 + U_5 + U_8}{U_4 + U_6 + U_7}$$

and then define

$$Z_1 = \frac{V_1}{1 + V_1} \quad \text{and} \quad Z_2 = \frac{V_2}{1 + V_2}.$$

Here, the marginal distributions for (V_1, V_2) are beta distributed of the second kind and (Z_1, Z_2) have marginal distributions of the first kind. The 8 parameter model includes the 5 parameter model by setting $\delta_3 = \delta_4 = \delta_5 = 0$ and by relabeling the remaining δ_i as $\alpha_1 = \delta_1, \alpha_2 = \delta_2, \alpha_3 = \delta_7, \alpha_4 = \delta_8, \alpha_5 = \delta_6$. This model also includes the Dirichlet model, the model proposed by Jones (2002) and the 3 parameter model of Olkin and Liu (2003). Just as the 5 parameter model gained the ability to allow for both positive and negative correlation over the 3 parameter model at the price of a likelihood function that does not exist in closed form, the 8 parameter model allows for extra flexibility over the 5 parameter model but at the price of added complexity and the inability (to our knowledge) to obtain closed form estimates for $\delta_i, i = 1, ..., 8$.

1.5 The k-variate beta distribution

Arnold and Ng (2011) defined a k-variate generalization by letting $U_1, ..., U_k$ and $V_1, ..., V_k$ and W be independent gamma random variables with some common scale parameter θ. As with the 5 and 8 parameter models, we let $\theta = 1$, and so $U_i \overset{ind}{\sim} \Gamma(\delta_{U_i}, 1)$ and $V_i \overset{ind}{\sim} \Gamma(\delta_{V_i}, 1), i = 1, ..., k$ and $W \overset{ind}{\sim} \Gamma(\delta_W, 1)$. So for $i = 1, ..., k$, define

$$Z_i = \frac{U_i + V_i}{U_i + \sum_{l=1}^{k} V_l + W}. \tag{1.8}$$

The random vector $\boldsymbol{Z}_{[k]} = (Z_1, Z_2, ..., Z_k)$ follows a k variate beta distribution with $2k+1$ parameters. It may be verified that $Z_i, i = 1, ..., k$ is beta distributed of the first kind and each random vector $\boldsymbol{Z}_{[k^\star]} = (Z_{\tau_1}, Z_{\tau_2}, ..., Z_{\tau_{k^\star}}), k^\star \leq k, \tau_i \in \{1, 2, ..., k\}, \tau_i \neq \tau_j$ for $i \neq j, i, j = 1, ..., k^\star$, has a joint distribution that is a $k^\star$ variate beta distribution defined in (1.8). For example, each random vector $\boldsymbol{Z}_{[2]} = (Z_i, Z_j), i = 1, 2, ..., k, j = 1, 2, ..., k, i \neq j$ has the 5 parameter bivariate beta distribution with parameters $\alpha_1 = \delta_{U_i}, \alpha_2 = \delta_{U_j}, \alpha_3 = \delta_{V_i}, \alpha_4 = \delta_{V_j}$ and $\alpha_5 = \delta_W + \sum_{l=1, l\neq i,j}^{k} \delta_{V_l}$. However, estimation of the parameters remains an open problem.

Chapter 2

Approximate Bayesian computation

2.1 Introduction

In this chapter, we will introduce a class of likelihood free algorithms known as approximate Bayesian computation (ABC). We will provide a brief background on the fundamentals of Bayesian inference and then proceed to describe ABC. In short, ABC samples from the posterior (or approximate) distribution in the absence of the likelihood function. There have been many developments in the history of ABC and various algorithms have been proposed. The simplest algorithm is the accept reject method (ABC-AR), where the resulting outcome form an *i.i.d.* sample from the posterior (or approximate) distribution. However, this method suffers from low acceptance rates and other methods have been developed to generate higher acceptances. We will detail ABC-AR and provide toy examples to illustrate how it works and demonstrate that it samples from the posterior distribution. We then briefly discuss other methods designed to improve acceptance rates.

In Bayesian inference, the posterior distribution for the parameter $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ is given by

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{x})}$$

where one's prior beliefs about the unknown parameter $\boldsymbol{\theta}$ is expressed through the prior distribution $\pi(\boldsymbol{\theta})$, but is then updated by the observed data $\mathbf{x}$, through the likelihood function $p(\mathbf{x}|\boldsymbol{\theta})$. Inference for the parameter $\boldsymbol{\theta}$ is then based on the posterior distribution. In particular, if we consider the mean squared error as our risk function, then the Baye's estimate of the unknown parameter $\boldsymbol{\theta}$ is the mean of the posterior distribution, i.e., $\hat{\boldsymbol{\theta}} = E(\theta|\mathbf{x}) = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}$. In many cases however, computing the posterior is difficult (if not impossible) since the marginal likelihood, $p(\mathbf{x}) = \int p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ is mathematically intractable. Provided the likelihood can be evaluated up to a normalizing constant, Markov Chain Monte Carlo (MCMC) methods such as the Metropolis Hastings (MH) algorithm or the Gibbs sampler allow us to sample from the posterior distribution. However, these methods require that the likelihood function be known (i.e., we can write them down). This begs the question, if we are working with a model where the likelihood function cannot be written down, is it still possible to sample from the posterior distribution?

To address this question, a class of algorithms, known as "likelihood-free computation" or "approximate Bayesian computation" (ABC) have been developed. This name refers to the circumventing of explicit evaluation of the likelihood by a simulation based approximation (Brooks et al. (2011)). The underlying idea of ABC is to consider a candidate parameter $\boldsymbol{\theta}'$ from the prior distribution and to generate an auxiliary data set $\mathbf{y}$, conditioned on $\boldsymbol{\theta}'$, i.e., $\mathbf{y}\sim p(\mathbf{y}|\boldsymbol{\theta}')$. If $\mathbf{y}$ is "close" to the observed data $\mathbf{x}$ in some manner, we accept $\boldsymbol{\theta}'$ as a likely candidate parameter to have generated the observed data $\mathbf{x}$. However, if $\mathbf{y}$ is not "close" to $\mathbf{x}$, then $\boldsymbol{\theta}'$ is unlikely to have generated $\mathbf{x}$ and so $\boldsymbol{\theta}'$ is rejected. We continue this algorithm until m candidate parameters have been accepted. The accepted parameter values form an *i.i.d.* sample from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{x})$. Hence, this is an algorithm that allows us to sample from the posterior distribution even in the absence of the likelihood function.

2.2 Likelihood free basics

Brooks et al. (2011) begins describing likelihood-free inference by augmenting the target posterior from

$$p(\boldsymbol{\theta}|\mathbf{x}) \;\propto\; p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \qquad \text{to} \qquad p_{LF}(\boldsymbol{\theta}, \mathbf{y}|\mathbf{x}) \;\propto\; p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

where the auxiliary data set $\mathbf{y}$, is generated from $p(\mathbf{y}|\boldsymbol{\theta})$ on the same space as $\mathbf{x} \in \mathcal{X}$. The distribution $p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ is chosen to weight the posterior $p(\boldsymbol{\theta}|\mathbf{y})$ with high density in regions where $\mathbf{y}$ and $\mathbf{x}$ are similar. The probability density function $p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ is assumed to be constant with respect to $\boldsymbol{\theta}$ at the point $\mathbf{y} = \mathbf{x}$, so that $p(\mathbf{x}|\mathbf{x}, \boldsymbol{\theta}) = c$, for some constant $c > 0$, with the result that the target posterior is recovered exactly at $\mathbf{y} = \mathbf{x}$, i.e., $p_{LF}(\boldsymbol{\theta}, \mathbf{x}|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$. Ultimately, interest is typically in the marginal posterior

$$p_{LF}(\boldsymbol{\theta}|\mathbf{x}) \;\propto\; \pi(\boldsymbol{\theta}) \int_{\mathcal{X}} p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})\, d\mathbf{y}$$

where we integrate out the auxiliary data set $\mathbf{y}$. The distribution then $p_{LF}(\boldsymbol{\theta}|\mathbf{x})$ becomes an approximation to $p(\boldsymbol{\theta}|\mathbf{x})$.

The likelihood free posterior distribution $p_{LF}(\boldsymbol{\theta}|\mathbf{x})$ will only recover the target posterior $p(\boldsymbol{\theta}|\mathbf{x})$ exactly when the density $p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ is precisely a point mass at $\mathbf{y} = \mathbf{x}$ and zero elsewhere. In this case

$$p_{LF}(\boldsymbol{\theta}|\mathbf{x}) \;\propto\; \pi(\boldsymbol{\theta}) \int_{\mathcal{X}} p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})\, d\mathbf{y} = p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

However, this choice for $p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ will result in a rejection sampler with an acceptance probability of zero unless the proposed auxiliary data set equals the observed data, i.e., $\mathbf{y} = \mathbf{x}$.

The first algorithm we will present is an exact algorithm, meaning that it samples from the posterior distribution (as opposed to an approximate posterior distribution). The algorithm requires that we match the auxiliary data set $\mathbf{y}$, to the observed data set $\mathbf{x}$ (or the sufficient statistics of $\mathbf{y}$ to the sufficient statistics of $\mathbf{x}$). It should be immediately clear that this algorithm will only apply to low dimensional discrete cases. For continuous models (and for high dimensional discrete cases), an adjustment will need to be made, however, it will come at the cost of no longer sampling from the true posterior distribution. The ABC-AR (exact) algorithm is described as follows

Algorithm ABC-AR exact
1. Generate $\boldsymbol{\theta}' \sim \pi(\boldsymbol{\theta})$.
2. Generate a data set $\mathbf{y}$ from the model $p(\mathbf{y}|\boldsymbol{\theta}')$.
3. Accept $\boldsymbol{\theta}'$ if $\mathbf{y} = \mathbf{x}$, otherwise discard $\boldsymbol{\theta}'$.
Continue until m observations have been accepted.

Hence, the outcome $\boldsymbol{\theta}'_1, ..., \boldsymbol{\theta}'_m$ forms an *i.i.d.* sample from $p(\boldsymbol{\theta}|\mathbf{x})$. Now, if sufficient statistics are known, we can replace the sufficient statistics in step 3, rather than $\mathbf{y} = \mathbf{x}$. In other words, accept $\boldsymbol{\theta}'$ if $\mathbf{S}(\mathbf{y}) = \mathbf{S}(\mathbf{x})$, otherwise reject $\boldsymbol{\theta}'$, where $\mathbf{S}(\cdot) = (S_1(\cdot), ..., S_p(\cdot))$ is the set of sufficient statistics and $p \geq dim(\boldsymbol{\theta})$. The benefit of using sufficient statistics, is that it will greatly reduce computational effort. Thus, our algorithm becomes

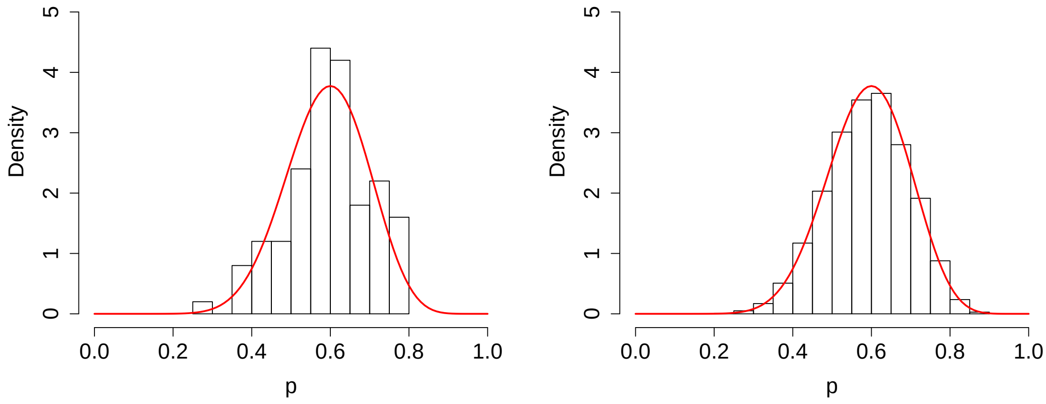
Algorithm ABC-AR exact
1. Generate $\boldsymbol{\theta}' \sim \pi(\boldsymbol{\theta})$.
2. Generate a data set $\mathbf{y}$ from the model $p(\mathbf{y}|\boldsymbol{\theta}')$.
3. Accept $\boldsymbol{\theta}'$ if $\mathbf{S}(\mathbf{y}) = \mathbf{S}(\mathbf{x})$, otherwise discard $\boldsymbol{\theta}'$.
Continue until m observations have been accepted.

2.2.1 Toy example one

We will now illustrate ABC-AR exact with a toy example. Suppose the observed data was generated from 20 Bernoulli trials with probability of success of 0.20, i.e., $\{X_i\}_{i=1}^{20} \overset{i.i.d.}{\sim} Ber(0.6)$, and we observe $\sum x_i = 12$, however p is unknown. We wish to estimate p and will take a Bayesian approach. Now suppose that the prior distribution for p is a $Beta(1,1)$ or equivalently a $Unif(0,1)$, i.e., $p \sim Beta(1,1) \equiv Unif(0,1)$. Here, we can compute the exact posterior distribution, i.e., $p|\mathbf{x}{\sim}Beta(\alpha+\sum x_i, \beta+n-\sum x_i)$ $\equiv Beta(13,9)$. Thus, using the mean squared error as our risk, the Bayes estimate is $\frac{13}{22} = 0.5909$. So after applying ABC-AR exact, we should expect to see that as the number of acceptances increases, the better the approximation to the posterior distribution, hence, also a better approximation to the true Bayes estimate. Our algorithm becomes

Algorithm ABC-AR exact
1. Generate $p' \sim Beta(1,1) \equiv Unif(0,1)$.
2. Generate $\{Y_i\}_{i=1}^{20} \overset{i.i.d.}{\sim} Ber(p')$.
3. Accept p' if $\sum y_i = 12$, otherwise discard p'.
Continue until m observations have been accepted.

The outcome $p'_1, ..., p'_m$ forms an $i.i.d.$ sample from a $Beta(13,9)$ distribution. Thus, for large m, we should obtain a good approximation to this distribution.



(a) $m = 100, \hat{p} = 0.5894$

(b) $m = 10,000, \hat{p} = 0.5915$

Figure 2.1: Plots comparing the estimated posterior for different acceptance sizes.

Figure 2.1 show the histograms for $m = 100$ and $m = 10,000$ acceptances, respectively. The red line is the density curve to the posterior, i.e., a $Beta(13, 9)$ distribution. We can see that the histogram for $m = 10,000$ acceptances is a better approximation than for $m = 100$ acceptances. Furthermore, the Bayes estimate for p under $m = 10,000$ is 0.5915, while the Bayes estimate for p under $m = 100$ is 0.5894. Recall that the true Bayes estimate is 0.5909, and so we see that as the number of acceptances increases, the approximation to the Bayes estimate also improves.

Method	Bayes	ABC-AR exact	
m	–	100	10,000
Proposals	–	2,141	213,715
$\hat{p}$	0.5909	0.5894	0.5915

Table 2.1: Comparison of acceptance sizes.

Table 2.1 shows the Bayes estimates for $m = 100$ and $m = 10,000$ compared to the true Bayes estimate, but also the number of proposals required to achieve the desired number of acceptances. Here, the number of proposals required for $m = 10,000$ is 231,715 compared to the 2,141 for $m = 100$. In this one dimensional setting, the computing time between $m = 100$ and $m = 10,000$ is insignificant, however in higher dimensional problems, the computational effort will be an issue of concern. Furthermore, while the posterior and Bayes estimate for $m = 10,000$ is better than the estimate for $m = 100$, the precision doesn't differ by much. It is on this note, that for higher dimensional problems, an increased number of acceptances may not be worth the additional computational effort.

2.3 ABC-AR for continuous random variables

Under continuous models, the probability that $\mathbf{y} = \mathbf{x}$ or $\mathbf{S}(\mathbf{y}) = \mathbf{S}(\mathbf{x})$ is zero (or approximately zero for high dimensional discrete models). Therefore, we make an adjustment by measuring the distance between $\mathbf{y} = \mathbf{x}$ or $\mathbf{S}(\mathbf{y}) = \mathbf{S}(\mathbf{x})$ through some distance function ρ. If this distance is within some tolerance level ϵ, we will accept the proposed candidate parameter. However, it does come at the cost of no longer sampling from the posterior distribution, but rather sampling from a distribution that is an approximation to the posterior (provided that ϵ is small). The algorithm becomes

Algorithm ABC-AR continuous
1. Generate $\boldsymbol{\theta}' \sim \pi(\boldsymbol{\theta})$.
2. Generate a data set $\mathbf{y}$ from the model $p(\mathbf{y}\|\boldsymbol{\theta}')$.
3. Accept $\boldsymbol{\theta}'$ if $\rho(\mathbf{S}(\mathbf{y}), \mathbf{S}(\mathbf{x})) < \epsilon$, otherwise discard $\boldsymbol{\theta}'$.
Continue until m observations have been accepted.

The outcome $\boldsymbol{\theta}'_1, ..., \boldsymbol{\theta}'_m$ then forms an *i.i.d.* sample from

$$p_\epsilon(\boldsymbol{\theta}|\mathbf{x}) = p(\boldsymbol{\theta}|\rho(\mathbf{S}(\mathbf{y}), \mathbf{S}(\mathbf{x})) < \epsilon).$$

Here, the idea being, if ϵ is small, the better the approximation to the posterior distribution. In other words, the likelihood free algorithm above samples from the marginal in $\mathbf{y}$ of the joint distribution

$$p_\epsilon(\boldsymbol{\theta}, \mathbf{y}|\mathbf{x}) = \frac{\pi(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})\mathbb{I}_{A_{\epsilon,\mathbf{x}}}(\mathbf{y})}{\int_{A_{\epsilon,\mathbf{x}}\times\boldsymbol{\theta}} \pi(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})d\mathbf{y}d\boldsymbol{\theta}}$$

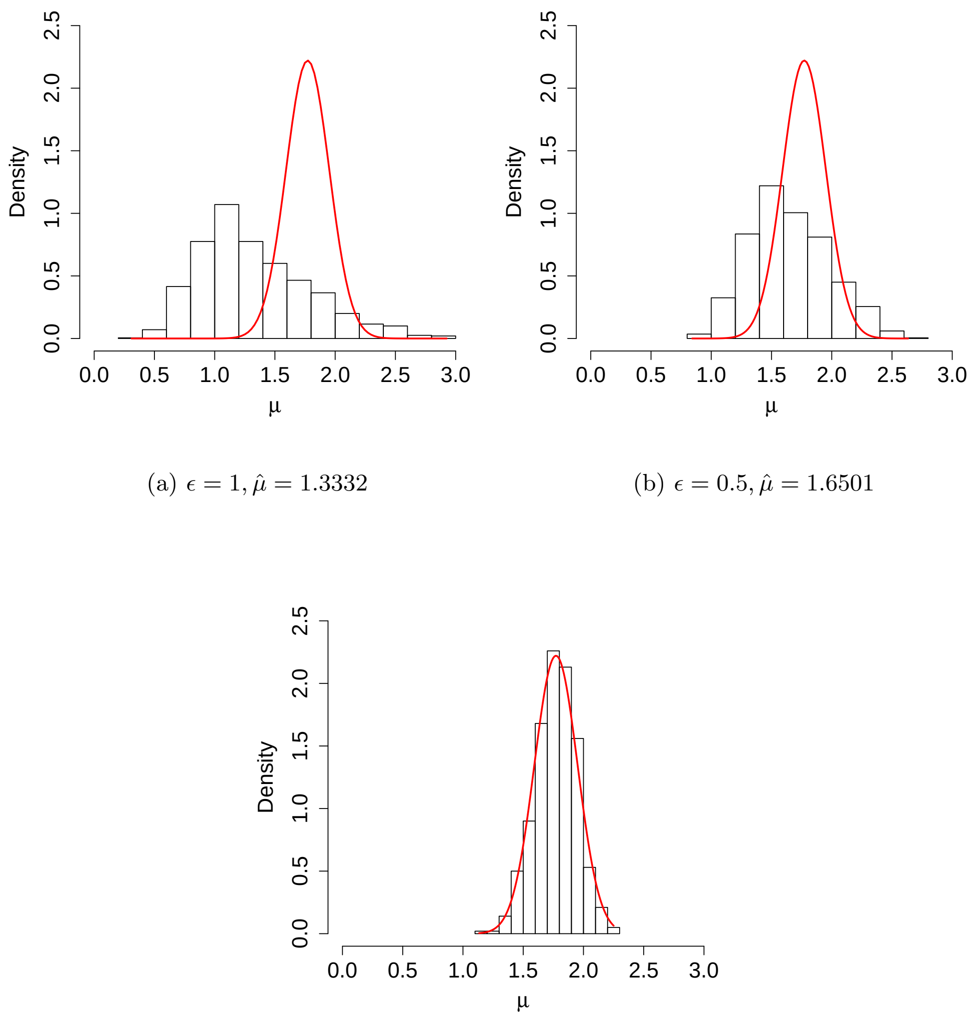
where $\mathbb{I}_{A_{\epsilon,\mathbf{x}}}(\cdot)$ denotes the indicator function of the set $A_{\epsilon,\mathbf{x}} = \{\mathbf{y}\in\mathcal{X}|, \rho(\mathbf{S}(\mathbf{y}), \mathbf{S}(\mathbf{x})) < \epsilon\}$. So, the smaller the tolerance level ϵ, the better the approximation to the posterior distribution, i.e., $p_\epsilon(\boldsymbol{\theta}|\mathbf{x}) = \int p_\epsilon(\boldsymbol{\theta}, \mathbf{y}|\mathbf{x})d\mathbf{y} \approx p(\boldsymbol{\theta}|\mathbf{x})$. Furthermore, as ϵ tends to zero, the posterior distribution is captured.

2.3.1 Toy example two

We will now illustrate ABC-AR continuous with another toy example. Suppose the observed data consisted of a sample size of 30, which was generated from a normal distribution with mean 2 and variance of 1, and suppose we observed $\sum x_i = 54.9275$, so that $\bar{x} = 1.8309$. Assume that the mean is unknown, however the variance is known and we wish to estimate μ. Suppose that the prior distribution for μ is a standard normal distribution, i.e., $\mu \sim N(0, 1)$, and so the computed posterior distribution is $\mu|\mathbf{x}{\sim}N(\frac{n\bar{x}}{n+1}, \frac{1}{n+1}) \equiv N(1.7719, 0.0323)$. Therefore, the algorithm becomes

Algorithm ABC-AR continuous
1. Generate $\mu' \sim N(0, 1)$. 2. Generate $\{Y_i\}_{i=1}^{30} \overset{i.i.d.}{\sim} N(\mu', 1)$. 3. Accept μ' if $\mid \frac{1}{n}\sum y_i - \frac{1}{n}\sum x_i \mid = \mid \frac{1}{n}\sum y_i - 1.8309 \mid < \epsilon$, otherwise discard μ'. Continue until m observations have been accepted.

The outcome $\mu'_1, ..., \mu'_m$ forms an *i.i.d.* sample from a distribution that is approximately normal, with a mean of 1.7719 and variance of 0.0323, i.e., $N_\epsilon(1.7719, 0.0323)$, provided that ϵ is small.



(a) $\epsilon = 1, \hat{\mu} = 1.3332$

(b) $\epsilon = 0.5, \hat{\mu} = 1.6501$

(c) $\epsilon = 0.01, \hat{\mu} = 1.7715$

Figure 2.2: Effect of ϵ on the posterior distribution.

Figure 2.2 shows the histograms for $\epsilon = 1$, $\epsilon = 0.5$, and $\epsilon = 0.01$. Here, we can clearly see that as ϵ is decreasing, the better the approximation to the posterior distribution. Furthermore, the true Bayes estimate for μ is 1.7719, and so we see that as ϵ is tending toward 0, we also obtain a better approximation to the true Bayes estimate. It is worth noting that for this example, we see two sources of bias. The first source of bias is coming from the $N(0,1)$ prior, which is proposing candidate values that are centered around 0. This is no surprise, since it is a fact that Bayes estimates are always biased. However, there is a second source of bias that is stemming from the value of ϵ. In histogram (a), we see a large number of proposed candidate parameters (that have small numerical values) that are being accepted due to the large ϵ, thus further weighing down our approximate Bayes estimate. However, in histogram (c), because ϵ is small, the parameters that are being accepted for $\epsilon = 1$ are now being rejected at a much higher rate, and so the bias stemming from a large ϵ vanishes.

Method	Bayes	ABC-AR continuous		
ϵ	–	1	0.5	0.01
Proposals	–	4,692	12,112	643,754
$\hat{\mu}$	1.7719	1.3332	1.6501	1.7715

Table 2.2: Comparing the effect of ϵ.

Table 2.2 shows that as ϵ is decreasing, the estimate of μ converges to the true Bayes estimate. Furthermore, we see the effect of ϵ on computational effort. For $\epsilon = 1$, 4,692 proposals were needed, however 643,754 proposals were needed for $\epsilon = 0.01$. As with the first toy example, because this is a one dimensional problem, the difference in computing time is almost insignificant, however for higher dimensional problems, the choice of ϵ will have an impact. From the first and second toy examples, we can see that computing time is a function of the number of acceptances and the size of ϵ. This is a balance we will seek when applying ABC-AR in higher dimensions.

2.4 Summary statistics

Often sufficient statistics are unknown and it is common practice to replace the set of unknown sufficient statistics with a set of summary statistics or “near” sufficient statistics. The cost of this replacement is that it will further hurt the degree of the approximation of the posterior distribution. Furthermore, as ϵ goes to zero, we will no longer capture the posterior distribution. How much precision is lost depends on the information contained within the choice of summary statistics. For our 5 and 8 parameter bivariate beta models, because the likelihood function cannot be written down, we cannot identify the sufficient statistics and so we must heuristically choose a set of summary statistics. Discussion for the choice of summary statistics will be discussed in detail later. The interested reader is directed to Burr and Skurikhin (2013), Joyce and Marjoram (2008), and Fearnhead and Prangle (2012) for more discussion on the choice of summary statistics.

2.5 ABC Metropolis Hastings

In scenarios where the prior distribution is far from the posterior distribution, it will lead to low acceptance rates (albeit, one of the advantages of the ABC-AR algorithm is that because the proposals are independent, we can use embarrassingly parallel computation to reduce computing time.). To this end, Marjoram et al. (2003) proposed embedding the Metropolis Hastings (MH) algorithm in the ABC-AR algorithm, which we will denote as ABC-MH, in order to improve acceptance rates. Before describing the ABC-MH algorithm, let us review the MH algorithm. Recall that the MH algorithm is used to obtain a sequence of dependent samples from a probability distribution for which direct sampling is nearly impossible. The MH algorithm is particularly useful in

Bayesian analysis, where calculation of the posterior distribution is difficult due to the complexity of the marginal likelihood. This sequence is then used to approximate the distribution through histograms and/or integrals. The MH algorithm (in a Bayesian context) is described as follows

Algorithm Metropolis Hastings
1. Initialize $\boldsymbol{\theta}^{(1)}$, $v = 1$. 2. Generate $\boldsymbol{\theta}' \sim q(\boldsymbol{\theta}\|\boldsymbol{\theta}^{(v)})$, where q is some proposal density. 3. Set $\boldsymbol{\theta}^{(v+1)} = \boldsymbol{\theta}'$ with probability $h = \min\left\{1, \frac{p(\mathbf{x}\|\boldsymbol{\theta}')\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}^{(v)}\|\boldsymbol{\theta}')}{p(\mathbf{x}\|\boldsymbol{\theta}^{(v)})\pi(\boldsymbol{\theta}^{(v)})q(\boldsymbol{\theta}'\|\boldsymbol{\theta}^{(v)})}\right\}$, otherwise set $\boldsymbol{\theta}^{(v+1)} = \boldsymbol{\theta}^{(v)}$. 4. Continue for N iterations.

The output of this algorithm are dependent draws from the posterior distribution, from which we can estimate various Bayes estimators. To implement the MH algorithm in the ABC-AR algorithm, we add 2 steps between the second and third step of the MH algorithm. The first step is to generate an auxiliary data set from the likelihood model and the second step is to compare the sufficient statistics from the auxiliary data set to the observed data set. The ABC-MH (exact) algorithm is as follows

Algorithm ABC-MH exact
1. Initialize $\boldsymbol{\theta}^{(1)}$, $v = 1$. 2. Generate $\boldsymbol{\theta}' \sim q(\boldsymbol{\theta}\|\boldsymbol{\theta}^{(v)})$, where q is some proposal density. 3. Generate a data set $\mathbf{y}$ from the model $p(\mathbf{y}\|\boldsymbol{\theta}')$. 4. If $\mathbf{S}(\mathbf{y}) = \mathbf{S}(\mathbf{x})$, go to step 5, otherwise remain at $\boldsymbol{\theta}^{(v)}$. 5. Set $\boldsymbol{\theta}^{(v+1)} = \boldsymbol{\theta}'$ with probability $h = \left\{1, \frac{\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}^{(v)}\|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(v)})q(\boldsymbol{\theta}'\|\boldsymbol{\theta}^{(v)})}\right\}$, otherwise set $\boldsymbol{\theta}^{(v+1)} = \boldsymbol{\theta}^{(v)}$. 6. Continue for N iterations.

The output of the ABC-MH algorithm will sample from the stationary distribution, i.e., $p(\boldsymbol{\theta}|\mathbf{x})$. We prove this as follows

Theorem 1 $p(\boldsymbol{\theta}|\mathbf{x})$ *is the stationary distribution of the chain (Marjoram et al. (2003)).*

Proof.

Denote the transition mechanism of the chain by $r(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(v)})$ and without loss of generality, choose $\boldsymbol{\theta}' \neq \boldsymbol{\theta}^{(v)}$ satisfying

$$\frac{\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}^{(v)}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(v)})q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(v)})} \leq 1. \tag{2.1}$$

Then

$$\begin{aligned}
p(\boldsymbol{\theta}^{(v)}|\mathbf{x})r(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(v)}) &= p(\boldsymbol{\theta}^{(v)}|\mathbf{x})q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(v)})p(\mathbf{x}|\boldsymbol{\theta}')h \\
&= \frac{p(\mathbf{x}|\boldsymbol{\theta}^{(v)})\pi(\boldsymbol{\theta}^{(v)})}{p(\mathbf{x})}\left\{q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(v)})p(\mathbf{x}|\boldsymbol{\theta}')\frac{\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}^{(v)}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(v)})q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(v)})}\right\} \\
&= \frac{p(\mathbf{x}|\boldsymbol{\theta}')\pi(\boldsymbol{\theta}')}{p(\mathbf{x})}\left\{q(\boldsymbol{\theta}^{(v)}|\boldsymbol{\theta}')p(\mathbf{x}|\boldsymbol{\theta}^{(v)})\right\} \\
&= p(\boldsymbol{\theta}'|\mathbf{x})q(\boldsymbol{\theta}^{(v)}|\boldsymbol{\theta}')p(\mathbf{x}|\boldsymbol{\theta}^{(v)})h \\
&= p(\boldsymbol{\theta}'|\mathbf{x})r(\boldsymbol{\theta}^{(v)}|\boldsymbol{\theta}').
\end{aligned}$$

The argument when the ratio on the left of (2.1) is > 1 is analogous. Thus, $p(\boldsymbol{\theta}|\mathbf{x})$ satisfies the detailed balance equations, which implies that indeed $p(\boldsymbol{\theta}|\mathbf{x})$ is the stationary distribution of the chain, and the proof is complete. ■

There are two special cases (Marjoram et al. (2003)):

1. If $q(\boldsymbol{\theta}^{(v)}|\boldsymbol{\theta}') = q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(v)})$ then h depends only on the prior.

2. If q is reversible with respect to π, (so that $\pi(\boldsymbol{\theta}^{(v)})q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(v)}) = \pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}^{(v)}|\boldsymbol{\theta}')$, for all $\boldsymbol{\theta}^{(v)} = \boldsymbol{\theta}'$), then $h = 1$, and the algorithm reduces to a rejection method with correlated outputs.

Just as with ABC-AR exact, the ABC-MH algorithm is valid for only low dimensional discrete models. As with ABC-AR continuous, we make extension to continuous models by comparing $\mathbf{S}(\mathbf{y})$ to $\mathbf{S}(\mathbf{x})$ through some distance function ρ. Thus, we adjust ABC-MH exact by approximating the intractable likelihood ratio by 1, if the auxiliary data set and observed data set are sufficiently close, and 0 otherwise. The ABC-MH (continuous) algorithm is described as follows

Algorithm ABC-MH continuous
1. Initialize $\boldsymbol{\theta}^{(1)}$, $v = 1$.
2. Generate $\boldsymbol{\theta}' \sim q(\boldsymbol{\theta}\vert\boldsymbol{\theta}^{(v)})$, where q is some proposal density.
3. Generate a data $\mathbf{y}$ from the model $p(\mathbf{y}\vert\boldsymbol{\theta}')$.
4. Set $\boldsymbol{\theta}^{(v+1)} = \boldsymbol{\theta}'$ with probability $h = \left\{1, \dfrac{\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}^{(v)}\vert\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(v)})q(\boldsymbol{\theta}'\vert\boldsymbol{\theta}^{(v)})}\mathbb{I}(\rho(\mathbf{S}(\mathbf{y}), \mathbf{S}(\mathbf{x})) < \epsilon)\right\}$, otherwise set $\boldsymbol{\theta}^{(v+1)} = \boldsymbol{\theta}^{(v)}$.
5. Continue for N iterations.

One of the major problems of the ABC-MH continuous algorithm, is that it generates a sequence of serially and highly correlated samples from $p(\boldsymbol{\theta}\vert\rho(\mathbf{S}(\mathbf{y}), \mathbf{S}(\mathbf{x})) < \epsilon)$. Determination of the chain length, N, is therefore obtained through a careful assessment of convergence and consideration of the chain's ability to explore the parameter space, i.e., chain mixing (Sisson et al. (2007)).

2.6 Other ABC methods

To improve upon the inefficiency of ABC-AR and ABC-MH, there have been a large number of proposed methods that incorporate sequential Monte Carlo (SMC) techniques. Sisson et al. (2007) was one of the first to make use of SMC methodology, proposing coupling SMC with partial rejection control and a biased approximation of the posterior distribution. Beaumont et al. (2009) further proposed utilizing population Monte Carlo methods of Cappé et al. (2004). Similarly, Toni et al. (2009) proposed an algorithm derived from the framework of sequential importance sampling. Peters et al.

(2012) embed the partial rejection control mechanism of Liu (2001), which incorporates a mutation and correction step within the standard SMC sampler algorithm. This incorporation of a mutation kernel reduces the variability of the importance weights when compared to more standard SMC algorithms.

Another class of algorithms that improve ABC-AR is to incorporate regression methodology. The pioneer of this approach was Beaumont et al. (2002), who assumed that the conditional density can be described by a regression model. The idea was to weight the parameters by comparing the auxiliary summary statistics with the observed summary statistics. An interested reader is directed to Blum and François (2010) and Leuenberger and Wegmann (2010) for extensions of this approach. Unfortunately, these methods focus on univariate settings, though some authors comment that an extension using multivariate regression is straightforward.

Chapter 3

Application of ABC to the bivariate beta model

3.1 Introduction

In this chapter, we will apply approximate Bayesian computation (ABC) to the 5 and 8 parameter models proposed by Arnold and Ng (2011). In order to apply ABC, we will need to choose prior distributions and so we will begin by first describing the selected prior distributions used in our simulation studies. We then discuss simulation settings, i.e., true parameter settings, sample sizes, summary statistics, and tolerance levels. Our work considers both the ABC-AR and ABC-MH algorithms. The bivariate beta distributions are continuous models, so we will be applying the continuous versions of the algorithms. For simplicity, we will refer to ABC-AR continuous simply as ABC-AR, and likewise call ABC-MH continuous as ABC-MH. Given the ease in which one can simulate bivariate beta random vectors, we did not consider SMC or regression methodology in our study. We then close the chapter by discussing simulation results. Much of this chapter is the work of Crackel and Flegal (2014).

3.2 Prior distributions

We consider 4 prior distributions in our simulation study. Since $\alpha_i > 0$, we select prior distributions that reflect this support. A natural prior to consider is a gamma prior. In our study, $\alpha_i, i = 1, ..., 5$ will be independent and identically distributed, i.e., $\alpha_i \overset{i.i.d.}{\sim} \Gamma(\lambda, \beta)$, where λ and β are the hyperparameters. Specifically, our simulations consider $\Gamma(2.5, 0.52)$ and $\Gamma(2.5, 1.04)$, denoted $\mathcal{G}1$ and $\mathcal{G}2$, respectively.

We also consider a modified uniform distribution. For the modified uniform distribution, the density curve is uniform on the interval $(0, \mu)$. At $\alpha = \mu$, the density curve then tails off, i.e., $f(\alpha)$ approaches 0 as α goes to ∞ (see Figure 3.1). In other words, we can choose μ and p such that $P(\alpha \in (0, \mu)) = p$ and $P(\alpha \in (\mu, \infty)) = 1 - p$. The density function is

$$f(\alpha|\mu, p) = \begin{cases} \frac{p}{\mu} & \text{if } \alpha \in (0, \mu) \\ \frac{p}{\mu} exp\left(\frac{-p(\alpha-\mu)}{\mu(1-p)}\right) & \text{if } \alpha \in (\mu, \infty). \end{cases}$$

The motivation for the modified uniform is to reflect a lack of information on the interval $(0, \mu)$. The tail is added to cover the entire support to maintain a proper prior. We denote the modified uniform as $\mathcal{U}_p(0, \mu)$. Our simulations consider $\mathcal{U}_{0.8}(0, 2)$ and $\mathcal{U}_{0.8}(0, 4)$, which we denote as $\mathcal{U}1$ and $\mathcal{U}2$ respectively.

To compare the gamma and modified uniform priors, we have selected hyperparameters resulting in the same mean and variance for $\mathcal{G}1$ and $\mathcal{U}1$, and for $\mathcal{G}2$ and $\mathcal{U}2$. Since the α_i's are $i.i.d.$, for the 5 parameter case, we have $\pi(\boldsymbol{\alpha}) = \prod_{i=1}^{5} \pi_i(\alpha_i)$. Likewise, for the 8 parameter case, we assume the δ_i's are $i.i.d.$, and so we have $\pi(\boldsymbol{\delta}) = \prod_{i=1}^{8} \pi_i(\delta_i)$, where $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7, \delta_8)'$.

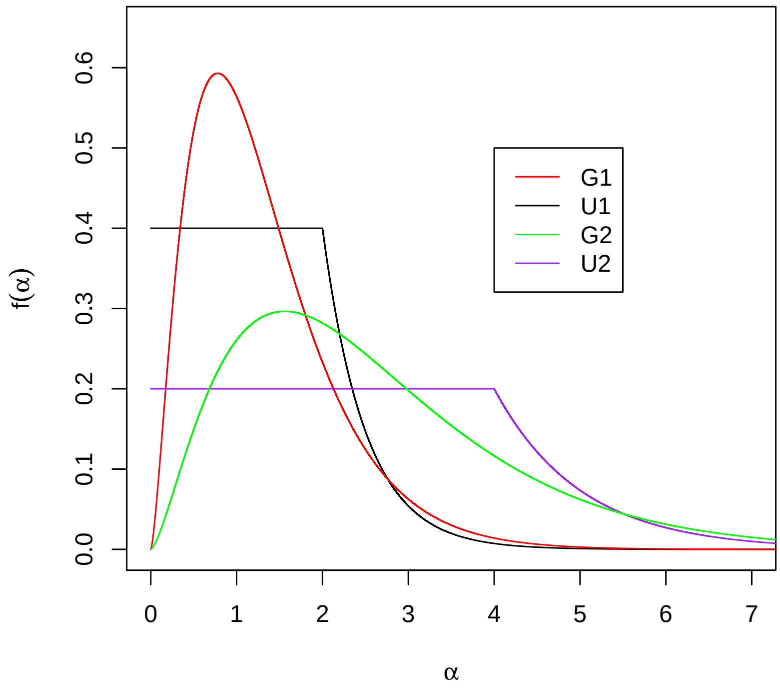


Figure 3.1: Plot of priors, $\mathcal{G}1$ in red, $\mathcal{U}1$ in black, $\mathcal{G}2$ in green, and $\mathcal{U}2$ in purple.

There are a number of other potential priors including the uniform, triangle, Epanechnikov or Gaussian (truncated on $\mathbb{R}^+$) kernels. In applications, these may increase acceptance rates for the ABC-AR algorithm, which is a direction of future research.

3.3 Summary statistics and distance function

Given the joint likelihood is unavailable in closed form, sufficient statistics cannot be determined. Thus, we are forced to choose informative summary or near sufficient statistics. First, consider the 5 parameter model where we will use 5 summary statistics. Since the marginals of Z_1 and Z_2 are distributed as beta random variables, we choose the corresponding univariate sufficient statistics. Specifically, we have summary statistics $\mathcal{S}_1(\tilde{\boldsymbol{z}}) = \frac{1}{n}\sum \log z_{i1}$, $\mathcal{S}_2(\tilde{\boldsymbol{z}}) = \frac{1}{n}\sum \log z_{i2}$, $\mathcal{S}_3(\tilde{\boldsymbol{z}}) = \frac{1}{n}\sum \log(1 - z_{i1})$, and $\mathcal{S}_4(\tilde{\boldsymbol{z}}) = \frac{1}{n}\sum \log(1 - z_{i2})$.

For our fifth summary statistic, we first considered $\mathcal{S}(\tilde{\boldsymbol{z}})$ at (1.5) used in MMLE. As we have illustrated in Table 1.1, $\mathcal{S}(\tilde{\boldsymbol{z}})$ is influenced by small observed values. The implications in a preliminary simulation study showed that a small portion of simulated datasets contained severely inflated values of $\mathcal{S}(\tilde{\boldsymbol{z}})$, one of which is illustrated in Table 1.1. Thus, there would be difficulty generating an auxiliary dataset $\tilde{\boldsymbol{y}}$, where $\mathcal{S}(\tilde{\boldsymbol{y}})$ is close to that of the observed data set, $\mathcal{S}(\tilde{\boldsymbol{z}})$, and further resulting in extremely low acceptance rates.

Given the problems with $\mathcal{S}(\tilde{\boldsymbol{z}})$ as a summary statistic, we require an alternative to capture the correlation that offers more stability for any observed values. To this end, we will use the Pearson correlation between $\vec{\boldsymbol{z}}_1$ and $\vec{\boldsymbol{z}}_2$, that is

$$\mathcal{S}_5(\tilde{\boldsymbol{z}}) = \frac{\sum(z_{i1} - \bar{z}_1)(z_{i2} - \bar{z}_2)}{\sqrt{\sum(z_{i1} - \bar{z}_1)^2 \sum(z_{i2} - \bar{z}_2)^2}}. \tag{3.1}$$

Thus, $\mathbf{S}(\tilde{\boldsymbol{z}}) = (\mathcal{S}_1(\tilde{\boldsymbol{z}}), \mathcal{S}_2(\tilde{\boldsymbol{z}}), \mathcal{S}_3(\tilde{\boldsymbol{z}}), \mathcal{S}_4(\tilde{\boldsymbol{z}}), S_5(\tilde{\boldsymbol{z}}))$ is the vector of summary statistics used for ABC-AR in the 5 parameter model. The use of $\mathcal{S}_5(\tilde{\boldsymbol{z}})$ instead of $\mathcal{S}(\tilde{\boldsymbol{z}})$ vastly improved acceptance rates, which will be further discussed in section 3.5.5.

Finally, we considered the distance function

$$\rho(\mathbf{S}(\tilde{\boldsymbol{y}}), \mathbf{S}(\tilde{\boldsymbol{z}})) = \sum_{i=1}^{5} |\mathcal{S}_i(\tilde{\boldsymbol{y}}) - \mathcal{S}_i(\tilde{\boldsymbol{z}})|\,.$$

Investigation of alternative distance functions is outside the scope of this dissertation. Preliminary simulations showed the 5 summary statistics had approximately equal variability for a variety of distance cutoff values, hence, there is no need to consider weights or a scale adjustment in the distance function.

Next, consider the 8 parameter model where we will use 8 summary statistics. We begin by including the 5 summary statistics from the 5 parameter sub model. In order to capture additional dependency between Z_1 and Z_2, we added the Spearman rank correlation and Kendall correlation, that is

$$\mathcal{S}_6(\tilde{\boldsymbol{z}}) = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \text{ where } d_i = z_{i1} - z_{i2}$$

and

$$\mathcal{S}_7(\tilde{\boldsymbol{z}}) = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\frac{1}{2}n(n-1)} .$$

Finally, we consider $\mathcal{S}_8(\tilde{\boldsymbol{z}}) = \frac{1}{n}\sum\sqrt{z_{i1}z_{i2}}$ as our eight summary statistic. Our distance function is, $\rho(\mathbf{S}(\tilde{\boldsymbol{y}}), \mathbf{S}(\tilde{\boldsymbol{z}})) = \sum_{i=1}^{8} |\mathcal{S}_i(\tilde{\boldsymbol{y}}) - \mathcal{S}_i(\tilde{\boldsymbol{z}})|$.

3.4 Simulations

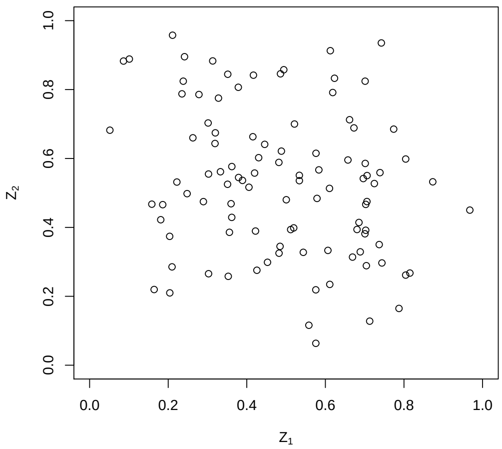
For the 5 parameter model, we considered 3 parameter settings. Within each parameter setting, we used each of the 4 priors described above, $\mathcal{G}1$, $\mathcal{G}2$, $\mathcal{U}1$, and $\mathcal{U}2$. For each prior, we considered 2 sample sizes, $n \in \{50, 100\}$, and for each sample size, we considered 4 tolerance levels, $\epsilon \in \{0.2, 0.4, 0.6, 0.8\}$. All together, this will account for a total of 96 settings. To draw comparisons between the ABC-AR algorithm to MMLE, each setting was repeated across 200 datasets to obtain an estimate of the bias and MSE. For each data set, the ABC-AR algorithm ran for 1,000 acceptances or 15e6 proposals, whichever came first. Furthermore, we wish to compare the behavior of each of the priors and the precision of estimation as the tolerance level decreases. For ABC-AR, simulations were ran for both the old $\mathcal{S}_5$, ($\mathcal{S}(\tilde{\boldsymbol{z}})$ at (1.5)), and the new $\mathcal{S}_5$ (the Pearson correlation at (3.1)). For ABC-MH, simulations were only ran using the old $\mathcal{S}_5$, 1e6 iterations, and $\epsilon \in \{0.4, 0.6, 0.8\}$.

Similarly, for the 8 parameter model, we considered 2 parameter settings, the 4 prior distributions, a sample size of $n = 100$, and the 4 tolerance levels as with the 5 parameter case. This totaled 32 settings, and we ran each setting across 200 datasets to estimate the bias and MSE, and the same stopping rule of 1,000 acceptances or 15e6 proposals was used. The results were not compared to a baseline, since we know of no other method applicable in this setting.

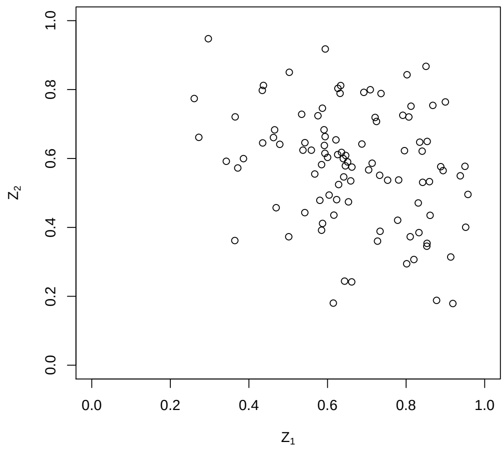
Results of the simulations are shown in the appendix. For Tables A.1 $-$ A.12 and Tables A.19 $-$ A.20, the average number of proposals required to reach the desired $m = 1,000$ acceptances or 15e6 proposals are reported along with the standard deviation (in parenthesis). For Tables A.13 $-$ A.18, the average number of acceptances are reported along with the standard deviation (in parenthesis). Furthermore, for Tables A.1 $-$ A.12, the results for $\epsilon = 0.6$ and for Tables A.19 $-$ A.20, the results for $\epsilon = 0.4$ and $\epsilon = 0.6$ are not shown to preserve space.

3.5 5 parameter model

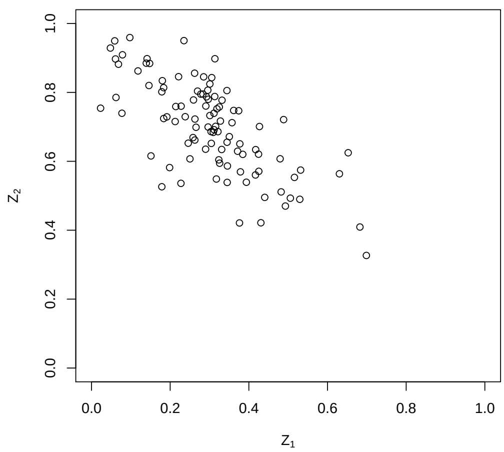
We considered the following three parameter settings, $A_1 = (1, 1, 1, 1, 1)'$, $A_2 = (3, 2.5, 2, 1.5, 1)'$, and $A_3 = (1, 1, 2, 6, 1)'$. The first two settings were used by Arnold and Ng (2011), while the third was added of our own accord. Figure 3.2 below shows a scatter plot for a data set (when $n = 100$), for each parameter setting. Each of these parameter settings have a negative correlation.



(a) A_1, r=-0.2384



(b) A_2, r=-0.2859



(c) A_3, r=-0.7068

Figure 3.2: Scatterplots of Z_1 and Z_2 and the estimated correlation.

3.5.1 Results using the old $\mathcal{S}_5$

Table A.1 shows the results for A_1 and $n = 100$. We can see that as we decrease the tolerance level from $\epsilon = 0.8$ to $\epsilon = 0.2$, the bias and MSE decrease substantially. Compared to the MMLE, $\mathcal{G}1$ and $\mathcal{U}1$ had a smaller MSE for each $\hat{\alpha}_i, i = 1, ..., 5$ while there was a decrease in the MSE for $\hat{\alpha}_3$ and $\hat{\alpha}_4$ under $\mathcal{G}2$ and $\mathcal{U}2$. For each prior, as ϵ is decreasing, we start seeing a dramatic increase in the required number of proposals. For example, when $\epsilon = 0.2$, the $\mathcal{G}1$ prior required an average of slightly more than 5.6e6 proposals while $\mathcal{U}2$ required a little more than 6.5e6 proposals.

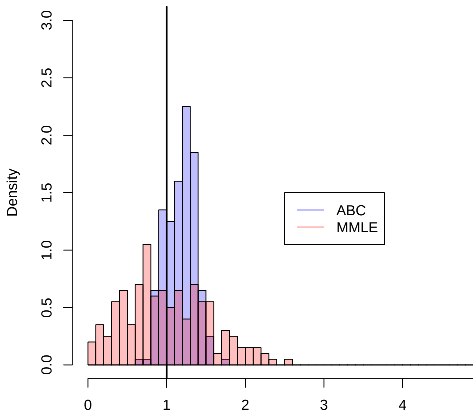
Table A.2 shows the results for A_2 and $n = 100$. For $\epsilon = 0.2$, only the $\mathcal{G}1$ prior had a smaller MSE for each $\hat{\alpha}_i, i = 1, ..., 5$ compared to the MMLE, whereas the other 3 priors had a decrease in the MSE for certain $\hat{\alpha}_i$. The results for A_3 and $n = 100$ are shown in Table A.3. For $\epsilon = 0.2$, $\mathcal{G}1$ and $\mathcal{U}1$ reduced the MSE for $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}_5$, when compared to the MMLE, while for $\mathcal{G}2$ and $\mathcal{U}2$, there was a reduction in MSE for certain $\hat{\alpha}_i$. Furthermore, we see the affect of $\alpha_4 = 6$, by observing larger MSE's and required number of proposals for the $\mathcal{G}1$ and $\mathcal{U}1$ priors. The MSE improves substantially for $\hat{\alpha}_4$ when using the $\mathcal{G}2$ and $\mathcal{U}2$ priors, while also reducing the overall computational effort. This is due to $\mathcal{G}1$ and $\mathcal{U}1$ primarily proposing values in low posterior regions.

In summary, for each of the parameter settings, the bias and MSE decreases as ϵ decreases. Reduction in MSE for $\hat{\alpha}_i$ is dependent on the parameter setting and prior distribution. For example, under the A_1 and A_2 setting, the $\mathcal{G}1$ prior performed the best with respect to MSE, when compared to the MMLE, and in particular under the A_1 setting, it performed exceptionally well. Furthermore, under the A_1 setting, the $\mathcal{G}1$ prior required the fewest number of proposals, while under the A_2 and A_3 setting, the $\mathcal{G}2$ prior required the fewest proposals. Results for $n = 50$ are similar and are shown in Tables A.4 $-$ A.6.

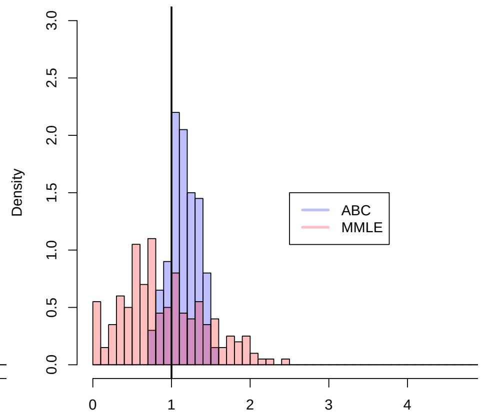
3.5.2 Results using the new $\mathcal{S}_5$

Table A.7 shows the results for A_1 when $n = 100$. For $\epsilon = 0.2$, compared to the MMLE, $\mathcal{G}1$, $\mathcal{U}1$, and $\mathcal{G}_2$ had a smaller MSE for each $\hat{\alpha}_i, i = 1, ..., 5$ while there was a decrease in the MSE under $\mathcal{U}2$ for $\hat{\alpha}_i, i = 1, ..., 3$. We observe however, that the $\mathcal{G}1$ prior clearly performs the best. Table A.8 shows the results for A_2 when $n = 100$. For $\epsilon = 0.2$, compared to the MMLE, $\mathcal{G}1$ and $\mathcal{U}1$ had a smaller MSE for each $\hat{\alpha}_i, i = 1, ..., 5$ while there was a decrease in the MSE under $\mathcal{G}2$ and $\mathcal{U}2$ for certain $\hat{\alpha}_i$. The results for A_3 and $n = 100$ are shown in Table A.9. For $\epsilon = 0.2, \mathcal{U}2$ decreased the MSE for $\hat{\alpha}_1$, $\hat{\alpha}_3$, $\hat{\alpha}_4$, and $\hat{\alpha}_5$, whereas the other 3 priors had a decrease in the MSE for certain $\hat{\alpha}_i$, when compared to the MMLE. As with the old $\mathcal{S}_5$, we see the affect of $\alpha_4 = 6$ by observing the large MSE's and number of proposals required for the $\mathcal{G}1$ and $\mathcal{U}1$ priors but the reduction in MSE and computing time by using the $\mathcal{G}2$ and $\mathcal{U}2$ priors.

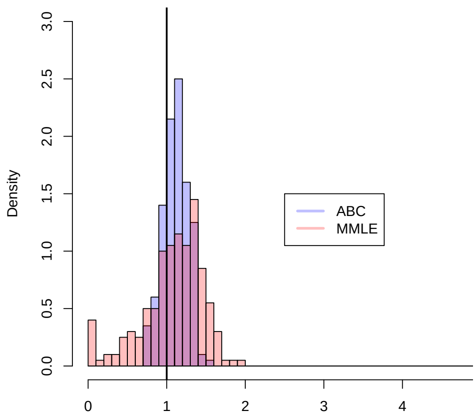
Figure 3.3 below shows the histograms for $\hat{\alpha}_i, i = 1, ...5$ using the ABC-AR algorithm and the $\mathcal{G}1$ prior, superimposed on the histogram using the MMLE. The histograms were generated under the A_1 setting for $n = 100$ and $\epsilon = 0.2$. The black vertical line represents the true parameter values, i.e., $\alpha_i = 1, i = 1, ..., 5$. Here, we can clearly see the bias that is present using the ABC-AR algorithm, however, we see larger variability using the MMLE. As a whole, there has been a reduction in the MSE using the ABC-AR algorithm as opposed to the MMLE (for this particular setting).



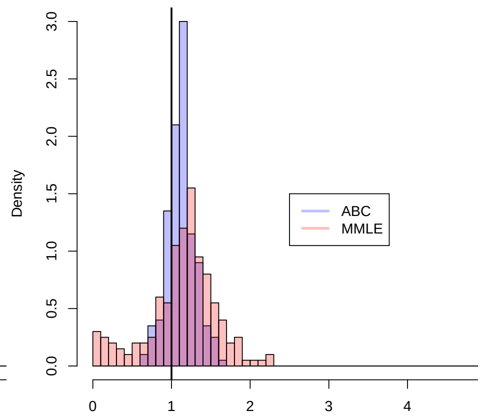
(a) $\hat{\alpha}_1$



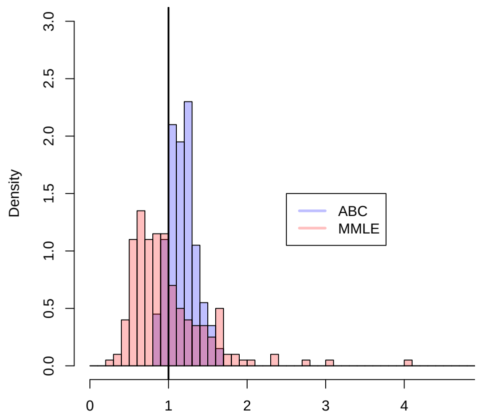
(b) $\hat{\alpha}_2$



(c) $\hat{\alpha}_3$



(d) $\hat{\alpha}_4$



(e) $\hat{\alpha}_5$

Figure 3.3: Histograms comparing ABC-AR and the $\mathcal{G}1$ prior to the MMLE, under A_1, $n = 100$, and $\epsilon = 0.2$ for $\hat{\alpha}_i, i = 1, ..., 5$.

In summary, as ϵ decreases, the precision of estimation improves. Under the A_1 setting, we see a decrease in the MSE for almost every $\hat{\alpha}_i$ using each of the priors. For the A_2 and A_3 settings, the reduction in the MSE for $\hat{\alpha}_i$, when compared to the MMLE, is dependent on the prior. Furthermore, under the A_1 setting, the $\mathcal{G}1$ prior required the fewest number of proposals, while under the A_2 and A_3 setting, the $\mathcal{G}2$ prior required the fewest proposals. Results for $n = 50$ are similar and are shown in Tables A.10 $-$ A.12.

Comparing computational effort between the old and new $\mathcal{S}_5$, under the A_1 setting, the Pearson correlation required significantly fewer proposals using $\mathcal{G}1$, $\mathcal{G}2$, and $\mathcal{U}1$, but the old $\mathcal{S}_5$ required fewer proposals for $\mathcal{U}_2$ (which is consistent with the results described in subsection 3.5.5). For the A_2 setting, the old $\mathcal{S}_5$ tended to require fewer proposals, while for the A_3 setting, the required amount of proposals favored the new $\mathcal{S}_5$. In general, replacing the old $\mathcal{S}_5$ with the Pearson correlation improved on acceptance rates. Recall that Tables A.1 $-$ A.12 and Tables A.19 $-$ A.20 report the average number of proposals required to reach the desired $m = 1,000$ acceptances or 15e6 proposals and says nothing about the percentage of data sets that reached (or did not reach) the $m = 1,000$ acceptances. Subsection 3.5.5 discusses these percentages between the old and new $\mathcal{S}_5$, under the A_1 setting, $n = 100$ and $\epsilon = 0.2$.

3.5.3 ABC-MH results

Table A.13 shows the results for A_1 and $n = 100$. We see a slight decrease in the bias and MSE as the tolerance level decreases from $\epsilon = 0.8$ to $\epsilon = 0.4$ for each prior. For $\epsilon = 0.4$, the $\mathcal{G}1$ prior had a smaller MSE for $\hat{\alpha}_i, i = 1, ..., 5$ while $\mathcal{U}1$ offered and a decrease for $\hat{\alpha}_i, i = 1, ..., 4$ when compared to the MMLE. The $\mathcal{G}2$ and $\mathcal{U}2$ priors decreased the MSE for $\hat{\alpha}_3$ and $\hat{\alpha}_4$ when compared to the MMLE. For $\epsilon = 0.8$, the

$\mathcal{G}1$ and $\mathcal{U}1$ priors averaged 3,517 and 3,370 acceptances, respectively, while $\mathcal{G}2$ and $\mathcal{U}2$ averaged 6,777 and 6,637 acceptances, respectively. Here, we see that $\mathcal{G}2$ and $\mathcal{U}2$ had a higher acceptance rate than $\mathcal{G}1$ and $\mathcal{U}1$. However, for $\epsilon = 0.4$, $\mathcal{G}1$ and $\mathcal{U}1$ averaged 403 and 389 acceptances, respectively, while $\mathcal{G}2$ and $\mathcal{U}2$ averaged 437 and 467 acceptances, respectively. So the decreasing of ϵ appeared to even out the number of acceptances between the 4 priors.

The results for A_2 and $n = 100$ are shown in Table A.14. For $\epsilon = 0.4$, the $\mathcal{G}1$ prior had a smaller MSE for $\hat{\alpha}_i, i = 2, ..., 5$, while $\mathcal{U}1$, $\mathcal{G}2$, and $\mathcal{U}2$ had a smaller MSE for certain $\hat{\alpha}_i$, when compared to the MMLE. For $\epsilon = 0.4$, the $\mathcal{G}1$ and $\mathcal{U}1$ priors averaged 3,568 and 2,844 acceptances, respectively, while the $\mathcal{G}2$ and $\mathcal{U}2$ prior averaged 13,412 and 12,236 acceptances, respectively. In this case, the $\mathcal{G}2$ and $\mathcal{U}2$ priors clearly had better acceptance rates. The results for A_3 and $n = 100$ are shown in Table A.15. For $\epsilon = 0.4$, when compared to the MMLE, the $\mathcal{G}1$ and $\mathcal{U}1$ priors had a smaller MSE for $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}_5$, while the $\mathcal{G}2$ and $\mathcal{U}2$ priors had a smaller MSE for only $\hat{\alpha}_4$. Under $\epsilon = 0.4$, the $\mathcal{G}1$ and $\mathcal{U}1$ priors averaged 1,535 and 1,156 acceptances, respectively, while the $\mathcal{G}2$ and $\mathcal{U}2$ priors averaged 7,762 and 6,314 acceptances, respectively. As with the A_2 setting, the $\mathcal{G}2$ and $\mathcal{U}2$ priors had better acceptance rates. Results for $n = 50$ are similar and are shown in Tables A.16 $-$ A.18.

In summary, for all 3 parameter settings, there was some decrease in the bias and MSE as ϵ decreased. Due to the low probability of acceptance, we did not attempt $\epsilon = 0.2$. As a whole, between the 4 priors, the $\mathcal{G}1$ prior performed the best with respect to reducing the MSE when compared to the MMLE, while the $\mathcal{G}2$ and $\mathcal{U}2$ priors tended to have a higher number of acceptances. Again, the drawback of ABC-MH is that the draws are highly correlated as auto correlation function plots showed (not produced in this dissertation). Nevertheless, there was some improvement over the

MMLE (with respect to MSE), depending on the parameter setting, prior distribution and the parameter being estimated. However, due to the highly correlated draws, it is suggested to use ABC-MH with caution.

3.5.4 Summary of simulations

The use of a Bayesian approach introduced significant bias in most settings, however, in general, there tended to be a reduction in the MSE relative to the MMLE. Under certain parameter settings and prior distributions, there was a noticeable improvement in the estimation compared to the MMLE, and there were settings where there was little or no improvement. Furthermore, there were scenarios where the MSE was smaller using MMLE than ABC-AR. Also, there were circumstances where a particular $\hat{\alpha}_i$ had a small MSE under a particular parameter setting and prior, and a large MSE under a different parameter setting and the same prior. For example, under A_1, $n = 100$, using the Pearson correlation and the $\mathcal{G}1$ prior, the MSE for $\hat{\alpha}_4$ was 0.047, while under the A_3 setting, $n = 100$, using the Pearson correlation and the $\mathcal{G}1$ prior, the MSE for $\hat{\alpha}_4$ was 4.463. Generally, the gamma priors tended to perform better with respect to reducing MSE and computational effort, relative to the modified uniform priors. For this reason, we suggest use of a gamma prior in conjunction with a "small" ϵ.

3.5.5 Comparison of acceptances between old and new $\mathcal{S}_5$

Table 3.1 shows the acceptances between the old and new $\mathcal{S}_5$ under the A_1 setting, $n = 100$ and $\epsilon = 0.2$ for each prior distribution. Recall that we ran each simulation setting across 200 data sets, and for each data set, we ran it for 1,000 acceptances or 15e6 proposals.

	$\mathcal{G}1$		$\mathcal{U}1$	
Acceptances	Old S_5	New S_5	Old S_5	New S_5
≤ 10	0	0	1	0
11-250	10	0	10	0
251-500	3	0	8	0
501-750	3	0	13	0
750-999	6	0	11	0
1000	178	200	157	200
Total	200	200	200	200

	$\mathcal{G}2$		$\mathcal{U}2$	
Acceptances	Old S_5	New S_5	Old S_5	New S_5
≤ 10	13	0	2	0
11-250	100	17	23	49
251-500	32	21	7	69
501-750	19	22	10	29
750-999	9	25	5	11
1000	27	115	153	42
Total	200	200	200	200

Table 3.1: Comparing acceptance rates between the old and new S_5, under A_1, $n = 100$, and $\epsilon = 0.2$.

For the $\mathcal{G}1$ and $\mathcal{U}1$ priors, using the old $\mathcal{S}_5$, we can see that there were 10 and 11 data sets, respectively, that had less than 250 acceptances. In fact, under the $\mathcal{U}1$ prior, there was 1 data set that had no more than 10 acceptances. Furthermore, there were 178 and 157 data sets that reached the desired 1,000 acceptances. Contrast this with the Pearson correlation, where we see that all 200 data sets achieved the desired 1,000 acceptances for both the $\mathcal{G}1$ and $\mathcal{U}1$ priors.

For the $\mathcal{G}2$ prior, using the old S_5, there were 13 data sets that had no more than 10 acceptances, 132 data sets that had between 11 and 500 acceptances, and only 27 data sets that achieved the full 1,000 acceptances. Using the Pearson correlation, there were no data sets that had no more than 10 acceptances, 38 data sets that had between 11 and 500 acceptances and 115 data sets that achieved the full 1,000 acceptances. Thus, for the $\mathcal{G}2$ prior, using the Pearson correlation, while the acceptance rates suffered relative to the $\mathcal{G}1$ and $\mathcal{U}1$ priors, the acceptance rates were still significantly better than using the old $\mathcal{S}_5$.

For the $\mathcal{U}2$ prior, using the old $\mathcal{S}_5$, there were 2 data sets that had no more than 10 acceptances, compared to 0 data sets using the Pearson correlation. However, for the old $\mathcal{S}_5$, there were 30 data sets that had between 11 and 500 acceptances and 153 data sets that achieved the full 1,000 acceptances, while for the Pearson correlation, there were 118 data sets that had between 11 and 500 acceptances and only 42 data sets that achieved the full 1,000 acceptances. So for the $\mathcal{U}2$ prior, the old $\mathcal{S}_5$ had better acceptance rates than the Pearson correlation, which was an unexpected result.

Overall, we can see that using the Pearson correlation improved on the acceptance rates, however, this improvement is dependent on the prior. Under the $\mathcal{G}1$ and $\mathcal{U}1$ priors, there was vast improvement in acceptance rates, while there was considerable improvement under the $\mathcal{G}2$ prior. However, we do note that for the $\mathcal{U}2$ prior, the acceptance rate was better using the old $\mathcal{S}_5$, so we suggest using careful attention when selecting the prior and summary statistics. Comparisons for other settings are not produced in this dissertation.

3.6 8 parameter model

For the 8 parameter model, we consider settings $A_4 = (2, 1, 1, 2, 4, 6, 2, 1)'$ and $A_5 = (3.5, 2, 1.5, 4, 1, 2.5, 3, 4.5)'$, both for a sample size $n = 100$. Table A.19 displays the results for the A_4 setting. As with the 5 parameter model, the bias and MSE decreases as ϵ decreases from 0.8 to 0.2. We notice that there is a significant amount of bias and large MSE for $\hat{\alpha}_5$ and $\hat{\alpha}_6$ under the $\mathcal{G}1$ and $\mathcal{U}1$ priors, however, the situation improves significantly under the $\mathcal{G}2$ and $\mathcal{U}2$ priors. This is due to $\mathcal{G}1$ and $\mathcal{U}1$ primarily proposing values in low posterior regions. Table A.20 shows the results for A_5. Here, we can see that there is a significant amount of bias and MSE for $\hat{\alpha}_1$, $\hat{\alpha}_4$, and $\hat{\alpha}_8$ under the $\mathcal{G}1$ and $\mathcal{U}1$ priors, but is not present for the $\mathcal{G}2$ and $\mathcal{U}2$ priors. Due to the complexity of the 8 parameter model and not having a baseline to compare results, we did not invest much effort into the inference of this model.

3.7 Application to correlated binomial random variables

To motivate our application of the 5 parameter bivariate beta distribution, let us consider the example from Arnold and Ng (2011), in which they describe the use of the 5 parameter model as the prior distribution to correlated binomial random variables. They describe the ABC-AR algorithm to draw samples from the posterior distribution. We will demonstrate this example by letting $X_1 \sim Bin(15, p_1)$ and $X_2 \sim Bin(15, p_2)$, where p_1 and p_2 are unknown, and further assume that X_1 and X_2 are correlated. Table 3.2 shows the outcome of this experiment.

	X_1		
X_2	0	1	Total
0	3	4	7
1	2	6	8
Total	5	10	15

Table 3.2: Bivariate binomial counts.

For convenience, call Table 3.2 as $\mathbb{T}$. We wish to estimate (p_1, p_2) using a Bayesian approach. Since p_1 and p_2 are correlated, we can use the 5 parameter bivariate beta model as the prior distribution. Since this is a low dimensional discrete model, we can simulate from the posterior distribution. However, we wish to make comparisons between drawing from the posterior and the approximate posterior distribution. To do this, we compare the generated auxiliary table $\mathbb{T}'$ to the observed table $\mathbb{T}$. If $\mathbb{T}'$ is sufficiently close to $\mathbb{T}$, we accept the proposed candidate parameter. Here, we must define what it means for $\mathbb{T}' \approx \mathbb{T}$. While there are numerous ways of defining "close", we simply consider the absolute difference between the cells of $\mathbb{T}'$ and $\mathbb{T}$. In other words, let $\mathbb{T} = \{a_{lj}\}$ and $\mathbb{T}' = \{b_{lj}\}$, and the distance function be $\rho = \sum_{l=1}^{2}\sum_{j=1}^{2}|a_{lj} - b_{lj}|$. Thus, to draw from the posterior, we accept the candidate parameter if $\mathbb{T}' = \mathbb{T}$. For the case of drawing from the approximate posterior, we accept the candidate parameter if $\rho = \sum_{l=1}^{2}\sum_{j=1}^{2}|a_{lj} - b_{lj}| < \epsilon$. We consider $\epsilon \in \{2, 6, 10\}$, and let $\boldsymbol{\alpha} = (1, 1, 2, 4, 1)'$ be the hyperparameters and run the algorithm for $m = 1,000$ acceptances. The algorithm for the latter case is described as follows

Algorithm ABC-AR
1. Generate $(p_1, p_2)'|\boldsymbol{\alpha} \sim \mathcal{BB}(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (1, 1, 2, 4, 1)'$.
2. Generate $X_1|p_1' \sim Bin(15, p_1')$ and $X_2|p_2' \sim Bin(15, p_2')$, i.e., this will generate an auxiliary table $\mathbb{T}'$.
3. Accept $(p_1, p_2)'$ if $\rho = \sum_{l=1}^{2}\sum_{j=1}^{2}|a_{lj} - b_{lj}| < \epsilon$, otherwise discard $(p_1, p_2)'$.
Continue until 1,000 observations have been accepted.

The outcome $(p_1, p_2)'_1, ..., (p_1, p_2)'_{1,000}$ is an *i.i.d.* sample from the approximate posterior distribution, conditioned on ϵ. Table 3.3 below, compares the results of drawing from the posterior distribution to an approximate posterior distribution.

Method	Exact	Approximate		
ϵ	–	10	6	2
Proposals	488,111	2,692	10,521	481,152
$\widehat{(p_1, p_2)}$	(0.5571, 0.5173)	(0.4911, 0.5535)	(0.5344, 0.5335)	(0.5611, 0.5181)

Table 3.3: Results comparing draws from the posterior distribution to an approximate posterior distribution.

We can see that $\widehat{(p_1, p_2)}$ is approximately the same for $\epsilon = 2$ compared to sampling from the (true) posterior, and that the computational effort is also approximately equal. For $\epsilon = 6$, we see that $\widehat{(p_1, p_2)}$ is comparable to sampling from the (true) posterior, however it requires roughly 2% of the computational effort. For $\epsilon = 10$, we notice that the precision of estimation suffers, but are still reasonable, however, it requires roughly 0.5% of the computational effort. Hence, as we saw in section 2.3.1, we conclude that for the sake of computing time, we can consider different tolerance levels without much sacrifice in the precision of our estimates, and so this is a balance that we will seek when selecting our tolerance levels.

3.8 Bacon and eggs

In this section, we apply the 5 parameter bivariate beta model of Arnold and Ng (2011), to a real data example previously analyzed by Danaher and Hardie (2005). The objective of the study was to observe the behavior of households and their grocery store habits. In particular, we study the probabilities and correlation of purchasing bacon and eggs on a single shopping trip. In the study, a sample of 548 independent households were taken and details of what the household purchased at the market were recorded over 4 consecutive trips. For each trip, it was recorded whether or not the household purchased bacon or eggs or both, see Table 3.4. We will refer to Table 3.4 as

$\mathcal{T}$, and note that the correlation found in $\mathcal{T}$ is 0.23.

	Eggs					
Bacon	0	1	2	3	4	Total
0	254	115	42	13	6	430
1	34	29	16	6	1	86
2	8	8	3	3	1	23
3	0	0	4	1	1	6
4	1	1	1	0	0	3
Total	297	153	66	23	9	548

Table 3.4: Bivariate binomial counts describing bacon and egg purchases.

Let X_{kb} and X_{ke} represent the number of times the k^{th} customer purchased bacon and eggs over the course of the 4 trips, respectively. Clearly, X_{kb} and X_{ke} are correlated, and so Danaher and Hardie (2005) proposed a bivariate beta binomial model to capture the over dispersion and correlation. Let p_b and p_e denote the probability of purchasing bacon and eggs, respectively. In this model, (p_b, p_e) is a bivariate random vector, where the requirement is that it follow some bivariate joint density, where the marginals are beta distributed. Thus, $(X_{kb}, X_{ke})|(p_b, p_e) \sim \mathcal{B}iv\mathcal{B}in(4, p_b, p_e)$ where $\mathcal{B}iv\mathcal{B}in$ is the notation we use to denote a bivariate binomial distribution. Furthermore, X_{kb} and X_{ke} are conditionally independent given (p_b, p_e), i.e., $X_{kb}|p_b{\sim}Bin(4, p_b)$ and $X_{ke}|p_e{\sim}Bin(4, p_e)$. The unconditional correlation between X_{kb} and X_{ke} is introduced through the bivariate distribution of (p_b, p_e).

Danaher and Hardie (2005) proposed using the bivariate beta model from the Sarmanov system of bivariate distributions (Sarmanov (1966)), which can be described as $g(p_b, p_e) = f_b(p_b)f_e(p_e)[1 + \omega\phi_b(p_b)\phi_e(p_e)]$, where $\phi_b(p_b)$ is a bounded non-constant "mixing" function, such that $\int \phi_b(l)f_b(l)dl{=}0$ (and similarly for "eggs"). The parameter ω determines the correlation between p_b and p_e and must satisfy the condition $1 +$

$\omega\phi_b(p_b)\phi_e(p_e) > 0$, for all p_b and p_e to be a valid joint density function. Furthermore, the marginals are beta distributed, i.e., $p_b \sim Beta(\alpha_b, \beta_b)$ and $p_e \sim Beta(\alpha_e, \beta_e)$. Letting $\phi_b(p_b) = p_b - \mu_b$, where

$$\mu_b = E(p_b) = \frac{\alpha_b}{\alpha_b + \alpha_b} \tag{3.2}$$

and likewise for "eggs," yields a closed form likelihood enabling estimation via maximum likelihood. In our analysis, we define p_{kb} and p_{ke} as the probability that the k^{th} household will purchase bacon and eggs, respectively. Therefore, we have $p_{kb} \sim Beta(\alpha_b, \beta_b)$ and $p_{ke} \sim Beta(\alpha_e, \beta_e)$.

3.8.1 Simulation setup

As a competitor to the bivariate beta distribution used by Danaher and Hardie (2005), we propose use of the 5 parameter bivariate beta model of Arnold and Ng (2011), as the prior distribution to (p_{kb}, p_{ke}), that is

$$(p_{kb}, p_{ke}) \sim \mathcal{BB}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5).$$

Therefore,

$$p_{kb} \sim Beta(\alpha_1 + \alpha_3, \alpha_4 + \alpha_5) \equiv Beta(\alpha_b, \beta_b)$$

and

$$p_{ke} \sim Beta(\alpha_2 + \alpha_4, \alpha_3 + \alpha_5) \equiv Beta(\alpha_e, \beta_e)\ .$$

Furthermore, we introduce an additional hierarchical step, meaning, we place a prior distribution on $\boldsymbol{\alpha}$. Thus, after generating a candidate parameter $\boldsymbol{\alpha}' \sim \pi(\cdot)$, we generate $(p_{kb}, p_{ke}) \sim \mathcal{BB}(\boldsymbol{\alpha}')$, $k = 1, ..., 548$. For notational convenience, let $\vec{\boldsymbol{p}}_b = (p_{1b}, p_{2b}, \ldots, p_{548b})'$ and $\vec{\boldsymbol{p}}_e = (p_{1e}, p_{2e}, \ldots, p_{548e})'$ be the generated proportions and let

$\tilde{\boldsymbol{p}} = (\vec{\boldsymbol{p}}_b, \vec{\boldsymbol{p}}_e)$. As we've already seen, the likelihood cannot be written down, and so we apply ABC-AR in the estimation of $\alpha_i, i = 1, ..., 5$. In this context, since we are simulating 548 bivariate binomial random variables, it is theoretically possible to simulate $\mathcal{T}$ exactly. However, due to the large dimensionality, the probability of this event is almost 0, therefore, we compare $\mathcal{T}$ to $\mathcal{T}'$. Extending the distance function defined in section 3.7, let $\mathcal{T} = \{a_{lj}\}$ and $\mathcal{T}' = \{b_{lj}\}$, and the distance function becomes $\rho = \sum_{l=1}^{5} \sum_{j=1}^{5} |a_{lj} - b_{lj}|$.

We consider both the ABC-AR and ABC-MH algorithms with $\epsilon = 100$ and $\epsilon = 40$. For the ABC-AR algorithm, the simulation was run for 500 acceptances and is described as follows

Algorithm ABC-AR
1. Generate $\boldsymbol{\alpha}' \sim \pi(\cdot)$.
2. Generate $(p_{kb}, p_{ke})\|\boldsymbol{\alpha}' \sim \mathcal{BB}(\boldsymbol{\alpha}')$ for $k = 1, ..., 548$.
3. Generate an auxiliary table $\mathcal{T}'\| \ \tilde{\boldsymbol{p}}$.
4. Accept $\boldsymbol{\alpha}'$ if $\rho = \sum_{l=1}^{5} \sum_{j=1}^{5} \|a_{lj} - b_{lj}\| < \epsilon$, otherwise discard.
Continue until 500 observations have been accepted.

Selection of prior distributions will be discussed in subsection 3.8.2. The ABC-MH algorithm is described as follows

Algorithm ABC-MH
1. Initialize $\boldsymbol{\alpha}^{(1)}$, $v = 1$.
2. Generate a candidate parameter $\alpha_i' \sim N(\alpha_i^{(v)}, \sigma_i^2), i = 1, ..., 5$.
3. Generate an auxiliary table $\mathcal{T}'\| \ \tilde{\boldsymbol{p}}$.
4. Accept $\boldsymbol{\alpha}'$ if $\rho = \sum_{l=1}^{5} \sum_{j=1}^{5} \|a_{lj} - b_{lj}\| < \epsilon$, otherwise discard.
5. Set $\boldsymbol{\alpha}^{(v+1)} = \boldsymbol{\alpha}'$ with probability
$h = \left\{ 1, \dfrac{\pi(\boldsymbol{\alpha}')}{\pi(\boldsymbol{\alpha}^{(v)})} \mathbb{I}(\rho = \displaystyle\sum_{l=1}^{5} \sum_{j=1}^{5} \|a_{lj} - b_{lj}\| < \epsilon) \right\}$, otherwise set $\boldsymbol{\alpha}^{(v+1)} = \boldsymbol{\alpha}^{(v)}$.
6. Continue for $5e6$ iterations.

For ABC-MH, the simulation was run for 5e6 iterations using a random walk (with a normally distributed proposal). The component standard deviations (σ_i) selected were 0.10, 0.10, 0.001, 0.001, and 0.2 for $i = 1, ..., 5$, respectively. Since we know $\alpha_i > 0$, any draws that yielded negative values resulted in an automatic rejection.

3.8.2 Prior distributions / Empirical Bayes

For the ABC-AR algorithm, we consider gamma priors using an empirical Bayes approach, in other words, we use the data in the selection of the prior distributions. Specifically, we computed marginal MLE's under the beta binomial model and used this as a guide. Using the beta binomial distribution family function within the `VGAM` package in `R`, we have the following MLE's, $\hat{\alpha}_b = 0.3571$, $\hat{\beta}_b = 4.4552$, $\hat{\alpha}_e = 0.8592$, and $\hat{\beta}_e = 3.9593$. Linking our bivariate beta parameters to these estimates, we have

$$\begin{aligned} \widehat{\alpha_1 + \alpha_3} &= \hat{\alpha}_b = 0.3571 & \widehat{\alpha_4 + \alpha_5} &= \hat{\beta}_b = 4.4552 \\ \widehat{\alpha_2 + \alpha_4} &= \hat{\alpha}_e = 0.8592 & \widehat{\alpha_3 + \alpha_5} &= \hat{\beta}_e = 3.9593\ . \end{aligned} \tag{3.3}$$

We wish to choose values for $\hat{\alpha}_i$ such that α_i will center around $\hat{\alpha}_i$, and have a variance of 1, i.e., $\alpha_i \sim \Gamma(\hat{\alpha}_i^2, 1/\hat{\alpha}_i), i = 1, ..., 5$. (Note that in this context (section 3.8.2), $\hat{\alpha}_i$ denotes the values of the hyperparameters. However, in the next section, $\hat{\alpha}_i$ will denote the posterior estimates). Furthermore, we choose $\hat{\alpha}_i, i = 1, ..., 5$ such that the Monte Carlo correlation estimate for $\mathcal{BB}(\hat{\alpha}_1, ..., \hat{\alpha}_5)$ is close to the correlation found in Table 3.4. We are presented with the problem of having 4 equations and 5 unknowns. Here, we simply choose $\hat{\alpha}_i, i = 1, ..., 5$ close to the constraints in (3.3), subject to a correlation of near 0.23. Our chosen values are $\hat{\alpha}_1 = 1.0487$, $\hat{\alpha}_2 = 1.6649$, $\hat{\alpha}_3 = 0.1012$, $\hat{\alpha}_4 = 0.1128$, $\hat{\alpha}_5 = 3.7697$, where the Monte Carlo correlation is 0.2336. As we will see, this choice allows for exploration of the parameter space with reasonable computational effort.

3.8.3 Results for the bacon and egg analysis

Table 3.5 shows the estimated posterior means and correlations for ABC-AR and ABC-MH, where we can see that each simulation setting yielded similar results. However, there was a vast computational difference between $\epsilon = 100$ and $\epsilon = 40$. Specifically, for the ABC-AR algorithm, the number of proposals required were 47,219 and more than 6.18e8 (not shown in Table 3.5) for $\epsilon = 100$ and $\epsilon = 40$, respectively. For ABC-MH, from the $5e6$ iterations, there were 242,729 and 14 acceptances (not shown in Table 3.5) in the chain for $\epsilon = 100$ and $\epsilon = 40$, respectively. Given the similarity of the estimates, it appears that use of $\epsilon = 40$ may be too computationally demanding and the larger ϵ value will suffice. Furthermore, the estimates of $\hat{\alpha}_3$ and $\hat{\alpha}_4$ are near 0, which suggests that the 3 parameter model of Olkin and Liu (2003) may be more appropriate.

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	r
ABC-AR ($\epsilon = 100$)	0.3295	0.8760	0.0029	0.0047	4.4429	0.1178
ABC-AR ($\epsilon = 40$)	0.3525	0.9256	0.0015	0.0026	4.5349	0.1224
ABC-MH ($\epsilon = 100$)	0.3067	0.9186	0.0321	0.0054	4.6023	0.1062
ABC-MH ($\epsilon = 40$)	0.3421	0.8345	0.0004	0.0005	4.0853	0.1271

Table 3.5: Comparison of results between ABC-AR and ABC-MH algorithms.

Table 3.6 compares our results to the model proposed by Danaher and Hardie (2005), where we can see the estimates are similar. The obvious difference is between the estimated correlations, when compared to the observed table correlation of 0.23. In short, our model underestimates this correlation (where as Danaher and Hardie (2005) overestimates the table correlation), but still provides a better fit than that of Danaher and Hardie (2005).

	$\hat{\alpha}_b$	$\hat{\beta}_b$	$\hat{\alpha}_e$	$\hat{\beta}_e$	r
D&H (2005)	0.3571	4.4551	0.8592	3.9593	0.4300
ABC-AR ($\epsilon = 100$)	0.3324	4.4476	0.8807	4.4458	0.1178
ABC-AR ($\epsilon = 40$)	0.3540	4.5375	0.9282	4.5364	0.1224
ABC-MH ($\epsilon = 100$)	0.3388	4.6077	0.9240	4.6344	0.1062
ABC-MH ($\epsilon = 40$)	0.3425	4.0858	0.8350	4.0857	0.1271

Table 3.6: Comparison of results between $\mathcal{BB}(\boldsymbol{\alpha})$ and Danaher and Hardie (2005).

Table 3.7 shows observed average cell counts based on accepted parameters for ABC-AR. Comparing Table 3.7 to Table 3.4, we can see there is no apparent pattern of bias between the tables. However, we do see a reduction in bias as ϵ is decreased from 100 to 40. Hence, it is clear the observed bias stems from the ABC-AR approximation rather than the prior.

	$\epsilon = 100$, Eggs					
Bacon	0	1	2	3	4	Total
0	253.48	116.00	48.10	17.04	3.98	438.60
1	41.01	21.67	10.02	3.58	0.88	77.16
2	12.00	6.57	3.41	1.42	0.42	23.82
3	3.29	1.86	1.14	0.53	0.19	7.01
4	0.57	0.42	0.25	0.11	0.05	1.40
Total	310.35	146.52	62.92	22.68	5.52	548

	$\epsilon = 40$, Eggs					
Bacon	0	1	2	3	4	Total
0	254.03	115.64	43.87	14.73	4.66	432.93
1	36.27	27.66	13.80	4.82	0.93	83.48
2	10.16	7.72	3.30	1.92	0.48	23.58
3	2.30	1.45	1.78	0.59	0.23	6.35
4	0.72	0.47	0.33	0.11	0.03	1.66
Total	303.48	152.94	63.08	22.17	6.33	548

Table 3.7: Average cell counts based on the 500 accepted parameters of the ABC-AR algorithm.

3.8.4 Alternate analysis

In this section, we briefly consider an alternative analysis of the bacon and eggs data with negative correlation. In this case, the model of Olkin and Liu (2003) would be inappropriate, since it only allows for positive correlation. To this end, we consider a partial transpose of the data as shown in Table 3.8 below, where the observed correlation is -0.23.

	Eggs^c					
Bacon	0	1	2	3	4	Total
0	6	13	42	115	254	430
1	1	6	16	29	34	86
2	1	3	3	8	8	23
3	1	1	4	0	0	6
4	0	0	1	1	1	3
Total	9	23	66	153	297	548

Table 3.8: Transposed data to illustrate a negative correlation.

We apply ABC-AR and ABC-MH. Under the new table, the MLE's are $\hat{\alpha}_b = 0.3571$, $\hat{\beta}_b = 4.4552$, $\hat{\alpha}_e = 3.9593$, and $\hat{\beta}_e = 0.8592$. Linking our bivariate beta model, we have the following

$$\begin{aligned} \widehat{\alpha_1 + \alpha_3} &= \hat{\alpha}_b = 0.3571 & \widehat{\alpha_4 + \alpha_5} &= \hat{\beta}_b = 4.4552 \\ \widehat{\alpha_2 + \alpha_4} &= \hat{\alpha}_e = 3.9593 & \widehat{\alpha_3 + \alpha_5} &= \hat{\beta}_e = 0.8592 \ . \end{aligned} \tag{3.4}$$

Therefore, $\hat{\alpha}_i, i = 1, ..., 5$ are chosen close to the constraints in (3.4), subject to a correlation of near -0.23 (again, in this context, $\hat{\alpha}_i$ denotes the values of the hyperparameters). The chosen values are $\hat{\alpha}_1 = 0.3697$, $\hat{\alpha}_2 = 0.2714$, $\hat{\alpha}_3 = 0.1892$, $\hat{\alpha}_4 = 3.6822$, and $\hat{\alpha}_5 = 0.7661$, where the Monte Carlo correlation is -0.2338.

As with before, we ran the ABC-AR and ABC-MH algorithms for $\epsilon = 100$ and $\epsilon = 40$, and the results are summarized in Table 3.9. The results from each method are similar, though there is more variability than in the original analysis. Here, perhaps a 4 parameter model would suffice, since $\hat{\alpha}_1$ is near zero for most settings. In general, the Monte Carlo correlation is considerably smaller (approximately -0.55) for three of the settings than the observed correlation in Table 3.8. However, the estimated correlation under ABC-AR and $\epsilon = 40$ is -0.2679, which is close to the table correlation of -0.23. One explanation as to why the results vary significantly for ABC-AR and $\epsilon = 40$ could be due to the fact that we are choosing values for $\hat{\alpha}_i, i = 1, ...5$ from the 4 equations in (3.4), and so there may not exist unique solutions to $\hat{\alpha}_i, i = 1, ...5$.

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	r
ABC-AR ($\epsilon = 100$)	0.1906	0.2509	0.1642	4.0214	0.7247	-0.2679
ABC-AR ($\epsilon = 40$)	0.0245	0.1265	0.3655	4.2917	0.5761	-0.5647
ABC-MH ($\epsilon = 100$)	0.0261	0.1618	0.3313	4.1398	0.6108	-0.5312
ABC-MH ($\epsilon = 40$)	0.0094	0.5725	0.3872	4.2807	0.6235	-0.5617

Table 3.9: Comparison of results for ABC-AR and ABC-MH using the partially transposed data.

Table 3.10 below shows the posterior estimates for α_b, β_b, α_e, α_e, and the correlation under the 4 settings. The results are similar, where the only exception is in the estimated correlation under ABC-AR and $\epsilon = 40$.

	$\hat{\alpha}_b$	$\hat{\beta}_b$	$\hat{\alpha}_e$	$\hat{\beta}_e$	r
ABC-AR ($\epsilon = 100$)	0.3548	4.7461	4.2723	0.8889	-0.2679
ABC-AR ($\epsilon = 40$)	0.3900	4.8678	4.4182	0.9416	-0.5647
ABC-MH ($\epsilon = 100$)	0.3574	4.7506	4.3016	0.9421	-0.5312
ABC-MH ($\epsilon = 40$)	0.3966	4.9042	4.8532	1.0107	-0.5617

Table 3.10: Comparison of results for ABC-AR and ABC-MH using the partially transposed data.

Chapter 4

Proposed methods for improvement over ABC-AR

4.1 Introduction

The motivation for this chapter stems from the low acceptance rates of the ABC-AR algorithm. Here, the ideas presented is to develop new approaches that are two fold, first, to search for maximum likelihood type estimates in the absence of the likelihood function and second, to reduce the computational effort that is required of ABC-AR (and other existing ABC methods without the cost of highly correlated draws). The ideas presented in this chapter are founded on the fact that maximum likelihood estimates are viewed as the parameter value(s) that are most likely to have produced the observed data. In this chapter, the proposed algorithms use steps within the ABC-AR algorithm to allow exploration of the parameter space, to identify parameter values that are unlikely to have generated the observed data, and thus sequentially "move" to the parameter region that is highly likely to have generated the observed data.

4.2 Motivation to modify ABC-AR

As discussed (and seen) in previous chapters, the ABC-AR algorithm can produce low acceptance rates. The acceptance rates are a function of the tolerance level and prior distributions. If a prior distribution is chosen that leads to a high percentage of "bad" proposals, then there will be a low rate of acceptances. This is the issue that we wish to address. Consider toy example two from section 2.3.1. Recall that the data was generated from a sample size of 30 from a normal distribution with unknown mean 2 and known variance of 1 and we observed $\sum x_i = 54.9275$ (or $\bar{x} = 1.8309$). The prior for μ followed a standard normal distribution, i.e., $\mu \sim N(0,1)$, and the ABC-AR algorithm is

Algorithm ABC-AR continuous
1. Generate $\mu' \sim N(0,1)$.
2. Generate $\{Y_i\}_{i=1}^{30} \overset{i.i.d.}{\sim} N(\mu',1)$.
3. Accept μ' if $\mid \frac{1}{n}\sum y_i - \frac{1}{n}\sum x_i \mid = \mid \frac{1}{n}\sum y_i - 1.8309 \mid < \epsilon$, otherwise discard μ'.
Continue until m observations have been accepted.

The outcome $\mu'_1, ..., \mu'_m$ is an *i.i.d.* sample from $N_\epsilon(1.7719, 0.0323)$. The results for $\epsilon \in \{0.1, 0.5, 1\}$ and $m = 1,000$ acceptances are presented again in Table 4.1.

Method	Bayes	ABC-AR continuous		
ϵ	–	1	0.5	0.01
Proposals	–	4,692	12,112	643,754
$\hat{\mu}$	1.7719	1.3332	1.6501	1.7715

Table 4.1: Results from toy example two using a $N(0,1)$ prior.

From Table 4.1, since the prior distribution is proposing values in high acceptance regions, the computational effort is not too intensive. Here, we observed 4,692 proposals for $\epsilon = 1$ and 643,754 for $\epsilon = 0.01$. Now, let us consider the computational effort when we choose a prior distribution that is far from the posterior (or in low acceptance regions), say, let $\mu \sim N(6, 1)$. The results are summarized in Table 4.2.

Method	ABC-AR continuous		
ϵ	1	0.5	0.01
Proposals	1,153,090	6,727,694	559,975,842
$\hat{\mu}$	2.6642	2.2265	1.9689

Table 4.2: Results from toy example two using a $N(6, 1)$ prior.

From Table 4.2, we can see the effect of choosing a prior distribution in low acceptance regions. For $\epsilon = 1$, the number of proposals required is more than 1.5e6 and slightly less than 6e8 for $\epsilon = 0.01$. This is a drastic change compared to the $N(0, 1)$ prior. Now, let us consider a prior that is centered far from the posterior, but has a large enough variance so that it will still propose plausible values to have generated the data, say $\mu \sim N(8, 2)$. The results are summarized in Table 4.3.

Method	ABC-AR continuous		
ϵ	1	0.5	0.01
Proposals	203,629	526,167	29,711,780
$\hat{\mu}$	2.3360	1.9887	1.8917

Table 4.3: Results from toy example two using a $N(8, 2)$ prior.

From Table 4.3, for $\epsilon = 1$, the number of proposals required was 203,629 and slightly less than 3e7 for $\epsilon = 0.01$. Therefore, we can see that the acceptance rates are better than using the $N(6, 1)$ prior even though the mean is at 8 (as opposed to 6), which is due to the larger variance, which allows the prior to propose values in higher acceptance regions with a higher frequency. Now, for illustration, let's examine the behavior for when the prior distribution is centered around the MLE, which is $\bar{x} = 1.8309$, and with a variance 1, i.e., let $\mu \sim N(1.8309, 1)$. The results are

Method	ABC-AR continuous		
ϵ	1	0.5	0.01
Proposals	1,445	2,665	126,415
$\hat{\mu}$	1.8412	1.8344	1.8271

Table 4.4: Results from toy example two using a $N(1.8309, 1)$ prior.

From Table 4.4, we see that the number of proposals for $\epsilon = 1$ is 1,445 and the number of proposals for $\epsilon = 0.01$ is 126,415. This is a dramatic decrease in the acceptance rates found in Table 4.2. Therefore, the higher rate of proposals that are "close" to the MLE, the higher the acceptance rate. This observation motivates our proposed work in the next section.

4.3 Maximum likelihood

If the sample is representative of a population and we assume that the population belongs to a family of distributions with unknown parameters, then the maximum likelihood estimate is the parameter value that will maximize the likelihood function. In other words, the maximum likelihood estimate seeks to find the parameter value that is most likely to have generated the data under the assumption that the data follows some parametric model.

Definition 1: Let $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ be a random vector with PDF (PMF) $f(x_1, x_2, ..., x_n; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The function

$$L(\boldsymbol{\theta}; x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n; \boldsymbol{\theta}),$$

considered as a function of $\boldsymbol{\theta}$, is called the likelihood function.

Definition 2: The principle of maximum likelihood estimation consists of choosing as an estimator of $\boldsymbol{\theta}$, a $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ that maximizes $L(\boldsymbol{\theta}; x_1, x_2, ..., x_n)$, that is, to find a mapping $\hat{\boldsymbol{\theta}}$ of $\mathcal{R}_n \mapsto \mathcal{R}_k$ that satisfies

$$L(\hat{\boldsymbol{\theta}}; x_1, x_2, ..., x_n) = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}; x_1, x_2, ..., x_n). \tag{4.1}$$

If a $\hat{\boldsymbol{\theta}}$ exists in (4.1), we call it a maximum likelihood estimator (MLE).

We will now provide some examples illustrating the relationship between the MLE and acceptance rates.

4.3.1 Toy example one

Let $X \sim Bin(20, 0.6)$, where $p = 0.6$ is unknown. Suppose we observed $\sum x = 13$, and so $\hat{p}_{mle} = \frac{13}{20} = 0.65$. Therefore, if we were to simulate $Y \sim Bin(20, p)$ for every $p \in (0, 1)$ and observe the rate for which $\sum y = 13$, we would expect $p = 0.65$ to produce the highest acceptance rate. We propose the following algorithm to validate our claim

Algorithm MLE exact
1. Let $P = \{0.01, 0.02, ..., 0.99\}$.
2. For each $p \in P$, generate $Y \sim Bin(20, p)$. Now repeat 50,000 times.
3. Evaluate the proportion of times where $\sum y = 13$.

Figure 4.1 shows the acceptance rate for each $p \in P$ (in increments of 0.01). We see that the acceptance rate is zero for $p \in (0, 0.3)$, then starts to gradually increase,

and reaches its maximum around $p = 0.65$ and then starts to tail off. This picture confirms our expectation.

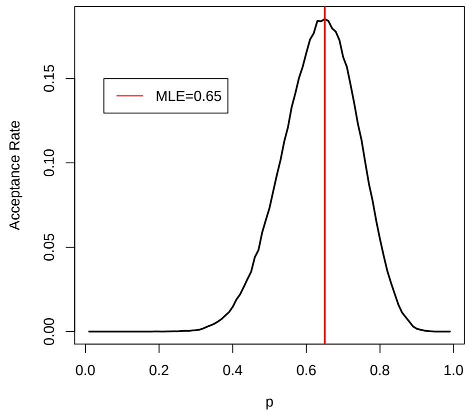


Figure 4.1: Values near $p = 0.65$, generate the highest proportion of acceptances.

This example was a univariate, discrete model, and so there is a positive probability that we can match sufficient statistics, however, this is not the case for continuous models, so we adjust our algorithm by allowing sufficient statistics to be "close" to each other (similar to the ABC-AR continuous algorithm from section 2.3).

4.3.2 Toy example two

Suppose $X_i \sim N(10, 16), i = 1, ..., 25$, where $\mu = 10$ is unknown and $\sigma^2 = 16$ is known and suppose we observe $\sum x = 260.5526$, so that $\bar{x} = 10.4221$. If we were to simulate $Y_i \sim N(10, 16), i = 1, ..., 25$, for every μ and observe the rate for which $|\frac{1}{25}\sum y - \frac{1}{25}\sum x| = |\frac{1}{25}\sum y - 10.4221| < 0.01$, we would expect $\mu = 10.4221$ to produce the highest acceptance rate. Now, it is impossible to simulate every value for μ (since the parameter space for μ is the entire real line), so we'll restrict our range to $\mu \in (7, 13)$, which contains the MLE of $\bar{x} = 10.4221$. Our algorithm is

Algorithm MLE continuous
1. Consider $P = \{7, 6.1, ..., 12.9, 13\}$.
2. For each $p \in P$ generate $Y_i \sim N(a, 16), i = 1, ..., 25$. Now repeat 50,000 times.
3. Evaluate the proportion of times where $\lvert\frac{1}{25}\sum y - 10.4221\rvert < 0.01$.

From Figure 4.2 below, we see that the acceptance rates increase in the neighborhood of the MLE ($\bar{x} = 10.4221$), and indeed the value that produced the highest acceptance rate was near $\mu = 10.4221$.

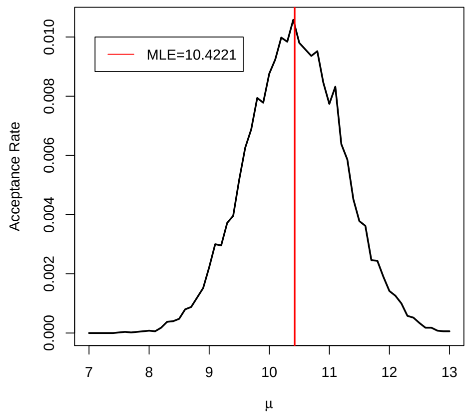


Figure 4.2: Values near $\mu = 10.4221$, generate the highest proportion of acceptances.

4.3.3 Toy example three

Let us consider another example for the continuous case. Suppose $X_i \sim \mathcal{E}xp(2), i = 1, ..., 30$, where $\lambda = 2$ is unknown. Suppose we observe $\sum x = 54.3492$ so that $\hat{\lambda} = \bar{x} = 1.8116$. If we were to simulate $Y_i \sim \mathcal{E}xp(2), i = 1, ..., 30$ for every λ (here we will consider $\lambda \in (0, 6)$) and observe the rate for which $\lvert\frac{1}{30}\sum y - \frac{1}{30}\sum x\rvert = \lvert\frac{1}{30}\sum y - 1.8116\rvert < 0.01$, we would expect $\lambda = 1.8116$ to produce the highest acceptance rate. Our algorithm is

Algorithm MLE continuous
1. Consider $P = \{0.1, 0.2, ..., 5.8, 6\}$.
2. For each $p \in P$, generate $Y_i \sim \mathcal{E}xp(a), i = 1, ..., 30$. Now repeat 50,000 times.
3. Evaluate the proportion of times where $\lvert \frac{1}{30} \sum y - 1.8116 \rvert < 0.01$.

From Figure 4.3 below, we see that the value that produced the highest acceptance rate was near $\lambda = 1.8116$, thus concluding our expectations.

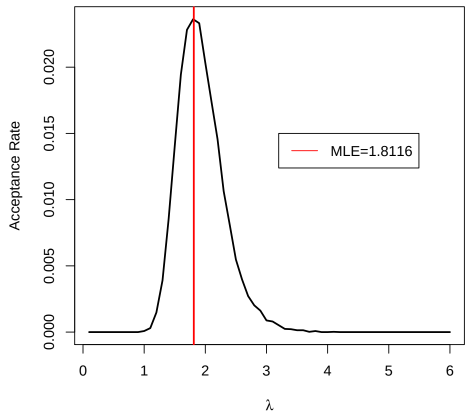


Figure 4.3: Values near $\lambda = 1.8116$, generate the highest proportion of acceptances.

4.4 Proposed method one

In our first toy example, we considered a discrete model and demonstrated through simulation efforts that there is a correspondence between the MLE and the proportion of times that the simulated data matches the observed data. For the second and third examples, we used continuous models to make the same establishment, however the difference in the continuous case is that we needed to impose a small tolerance distance between the simulated and the observed data.

Now that we've heuristically shown that the MLE is the most likely parameter value to reproduce the data, lets try and embed this idea into an ABC context. The proposed algorithm, which we'll call "Algorithm proposed method one" (Algorithm PM one for short) is described as follows

Algorithm PM one
1. Generate a large set of candidate parameters based off the prior distributions. 2. Generate an auxiliary data set (or multiple data sets) from each of the candidate parameter values and compare the average of auxiliary sufficient statistics to that of the observed sufficient statistics through some distance function. 3. Keep the d parameter values that generated the average of auxiliary sufficient statistics that were nearest to that of our observed data set. Now use these d parameter values to "update" our prior distributions, where the center is at the average of the d accepted values. 4. Generate a new set of candidate parameters from the "updated" priors. 5. Evaluate how much the previous "updated" parameters have changed. If this change is within some tolerance level, we stop, otherwise we repeat steps 2-4.

To estimate the unknown parameters, we consider 3 options. The first option is to take the average of all the values within the chain. The second option is to take the average of the previous d values, and the third option is to use the last value in the chain. We will apply 2 settings of this general algorithm to a data set. Under the first setting, we will generate 1 auxiliary data set for each candidate parameter. For the second setting, we will generate multiple auxiliary data sets for each candidate parameter.

4.4.1 Iron intake example

Consider the data set taken from the food and nutrition board of the national academy of sciences taken from Mendenhall et al. (2012). The data consists of iron intakes, in milligrams, which were obtained during a 24 hour period for 45 randomly selected adult females under the age of 51. The data is

15.0	18.1	14.4	14.6	10.9	18.1	18.2	18.3	15
16.0	12.6	16.6	20.7	19.8	11.6	12.8	15.6	11
15.3	9.4	19.5	18.3	14.5	16.6	11.5	16.4	12.5
14.6	11.9	12.5	18.6	13.1	12.1	10.7	17.3	12.4
17.0	6.3	16.8	12.5	16.3	14.7	12.7	16.3	11.5

For this data set, we will assume that $X_i \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, $i = 1, ..., 45$, where μ and σ^2 are unknown. The Shapiro Wilk test for normality yielded a p-value of 0.5008, so the assumption of normality is valid. The observed summary statistics are $\mathbf{S} = (\sum x, \sum x^2) = (660.6, 10115.88)$ and so $(\bar{x}, \overline{x^2}) = (14.68, 224.7973)$. Applying PM one, we have

---

Algorithm PM one

---

1. Generate $\boldsymbol{\theta}'_i \overset{i.i.d.}{\sim} \pi^{(1)}(\cdot), i = 1, ..., m$ where $\boldsymbol{\theta}'_i = (\mu_i, \sigma_i^2)'$. Here, let $\mu_i \overset{i.i.d.}{\sim} \pi_\mu^{(1)}(\cdot)$ and $\sigma_i^2 \overset{i.i.d.}{\sim} \pi_{\sigma^2}^{(1)}(\cdot)$, where $\pi^{(1)}(\cdot)$ denotes the initial prior distribution.
2. For each candidate parameter $\boldsymbol{\theta}'_i$, we will generate r auxiliary data sets, i.e., $\mathbf{y}_{ij}$ $(j = 1, ..., r)$. Let $\mathbf{S}_{ij}(\cdot) = (\sum y, \sum y^2)_{ij}$ be the summary statistics of $\mathbf{y}_{ij}$. Take the average of summary statistics generated by $\boldsymbol{\theta}'_i$, i.e., $\frac{(\sum y, \sum y^2)_{i1} + ... + (\sum y, \sum y^2)_{ir}}{r} = (\overline{\sum y}, \overline{\sum y^2})_i$. Now compare this to the observed summary statistics $\mathbf{S}$, via the distance function $\rho_i = |\overline{\sum y} - \sum x|_i + |\overline{\sum y^2} - \sum x^2|_i$. Let $\boldsymbol{\rho} = (\rho_1, ..., \rho_m)$ be the set of computed distances.
3. Let $\boldsymbol{\rho}_s = (\rho_{(1)}, ..., \rho_{(m)})$ be the ordered set of $\boldsymbol{\rho}$. Keep only $\rho_{(1)}$ and $\rho_{(2)}$ and consider the $\boldsymbol{\theta}'_i$ that generated $\rho_{(1)}$ and $\rho_{(2)}$ and denote these candidate values as $\boldsymbol{\theta}'_{(1)}$ and $\boldsymbol{\theta}'_{(2)}$, respectively. Taking the average of these values, we have $\frac{\boldsymbol{\theta}'_{(1)} + \boldsymbol{\theta}'_{(2)}}{2} = ((\frac{\mu_1 + \mu_2}{2}), (\frac{\sigma_1^2 + \sigma_2^2}{2}))' \equiv (\bar{\mu}_v, \bar{\sigma}_v^2)' \equiv \boldsymbol{\delta}^{(v)}$. Here, v denotes the v^{th} iteration.
4. Generate a new set of candidate values, $\boldsymbol{\theta}^\star_i \overset{i.i.d.}{\sim} \pi^{(v)}(\cdot), i = 1, ..., m$ from an "updated" prior, i.e., $\mu_i \overset{i.i.d.}{\sim} \pi_\mu^{(v)}(\cdot)$ and $\sigma_i^2 \overset{i.i.d.}{\sim} \pi_{\sigma^2}^{(v)}(\cdot)$. Here, both μ_i and σ_i^2 center around $\bar{\mu}_v$ and $\bar{\sigma}_v^2$ respectively.
5. To monitor the change in "updated" values, we look at the previous 6 sets of $\boldsymbol{\delta}^{(v)}$, i.e., $\boldsymbol{\delta}^{(v)}, \boldsymbol{\delta}^{(v-1)}, ..., \boldsymbol{\delta}^{(v-5)}$ and measure the change in these distances between $\hat{\mu}$ and $\hat{\sigma}^2$, i.e., let $q = |\bar{\mu}_d - \bar{\mu}_{(d-1)}| + |\bar{\mu}_{(d-1)} - \bar{\mu}_{(d-2)}| + ... + |\bar{\mu}_{(d-4)} - \bar{\mu}_{(d-5)}| + |\bar{\sigma}_d^2 - \bar{\sigma}_{(d-1)}^2| + |\bar{\sigma}_{(d-1)}^2 - \bar{\sigma}_{(d-2)}^2| + ... + |\bar{\sigma}_{(d-4)}^2 - \bar{\sigma}_{(d-5)}^2|$.

If $q < 3$ then stop, otherwise, repeat steps 2-4.

We now apply two settings under this algorithm. For the first setting, we drew $m = 100,000$ candidate parameters and for each parameter, generated only one data set, i.e., $r = 1$. For the second setting we drew $m = 1,000$ candidate parameters, and for each parameter, we generated $r = 100$ data sets. Furthermore, we also repeated the simulations using $\mathbf{S} = (\sum x, \sqrt{\sum x^2})$ to draw comparisons.

Table 4.5 below shows the initial prior distributions used. Both μ_i and σ_i^2 have a mean of 6 and a variance of 4 and are independent. For the updated priors, both μ_i and σ_i^2 center around $\bar{\mu}_l$ and $\bar{\sigma}_l^2$, respectively, with a variance of 4.

Prior	Distribution
$\pi_{\mu}^{(1)}(\cdot)$	$N(6, 4)$
$\pi_{\mu}^{(v)}(\cdot)$	$N(\bar{\mu}_l, 4)$
$\pi_{\sigma^2}^{(1)}(\cdot)$	$\Gamma(6^2/4, 4/6)$
$\pi_{\sigma^2}^{(v)}(\cdot)$	$\Gamma(((\bar{\sigma}_l^2)^2)/4, 4/\bar{\sigma}_l^2)$

Table 4.5: Prior distributions for PM one and the iron intake data.

Table 4.6 displays the results for the two settings. Again, μ and σ^2 are estimated by taking the average of all the values in the chain, the average of the previous 5 values in the chain, and the last value in the chain. These estimates are presented in the third, fourth, and fifth columns, respectively.

$m = 100,000,\ r = 1$				
$\mathbf{S}(\cdot)$	Iterations	$(\hat{\mu}, \hat{\sigma}^2)_{All}$	$(\hat{\mu}, \hat{\sigma}^2)_5$	$(\hat{\mu}, \hat{\sigma}^2)_1$
$(\sum x, \sum x^2)$	70	(14.62,9.62)	(14.36,8.25)	(14.49,7.60)
$(\sum x, \sqrt{\sum x^2})$	93	(14.66,9.36)	(14.70,7.89)	(14.74,8.32)

$m = 1,000,\ r = 100$				
$\mathbf{S}(\cdot)$	Iterations	$(\hat{\mu}, \hat{\sigma}^2)_{All}$	$(\hat{\mu}, \hat{\sigma}^2)_5$	$(\hat{\mu}, \hat{\sigma}^2)_1$
$(\sum x, \sum x^2)$	24	(14.57,8.88)	(14.65,10.81)	(14.64,11.13)
$(\sum x, \sqrt{\sum x^2})$	8	(14.37,8.62)	(14.67,9.09)	(14.67,8.85)

Table 4.6: Results using PM one and the iron intake data.

The MLE's for our data set are $(\hat{\mu}, \hat{\sigma}^2) = (14.68, 9.29)$. We can see that under both settings and regardless if the estimate was calculated using the entire chain, the average of the previous 5 values in the chain, or the last value in the chain, that the estimate for μ is close to the MLE. The estimate for σ^2 had some variation, but were all still comparable. The first setting ($m = 100,000$ and $r = 1$), under both $\mathbf{S} = (\sum x, \sum x^2)$ and $\mathbf{S} = (\sum x, \sqrt{\sum x^2})$ and using the average of the entire chain, yielded the closest estimates to the MLE. Furthermore, the first setting required more iterations than for the second setting ($m = 1,000$ and $r = 100$) until convergence.

4.5 Proposed method two

The idea behind our second proposed method is to propose candidate parameters within a region of the parameter space and conditioned off this parameter, simulate a data set (based off a large sample size) and use the information to decide whether or not the candidate parameter is likely to have generated the data. For example, think of the observed sufficient statistics, $\mathbf{s} = (\frac{s_1}{n}, ..., \frac{s_p}{n})$ as the "true" theoretical moments. Now, generate an auxiliary data set $\mathbf{y}$ (based off a large sample size m), and suppose we observed auxiliary sufficient statistics, $\mathbf{s}^\star = (\frac{s_1^\star}{m}, ..., \frac{s_p^\star}{m})$ such that

$$\frac{s_i^\star}{m} \overset{p.w.}{\rightarrow} E[f_i(Y)], i = 1, ..., p \quad \text{as} \quad m \rightarrow \infty.$$

From the auxiliary data $\mathbf{y}$, we can construct a confidence interval for $E[f_i(Y)], i = 1, ..., p$. If the observed sufficient statistics, $\frac{s_i}{n}, i = 1, ..., p$ are within each of the confidence intervals, this is evidence that the proposed candidate value generated the data.

First, a "general" algorithm will be proposed (called Algorithm PM two) and then we will describe the algorithm applied to normal data (μ and σ^2 are unknown) and then make application to the iron intake data.

Algorithm PM two
1. Generate $\boldsymbol{\theta}' = (\theta_1, ..., \theta_p)' \sim \pi(\bullet)$. 2. Generate a data set $\mathbf{y}$ (based off a "large" sample size m) from the model $p(\mathbf{y}\|\boldsymbol{\theta}')$. 3. Construct a confidence interval for $E[f_i(Y)], i = 1, ..., p$, and call it $\mathcal{C}_i$. **Note**: If the exact distribution of $S_i(\cdot)$ is not known, we can use Monte Carlo methods to estimate $\mathcal{C}_i$. 4. Accept $\boldsymbol{\theta}'$ if $\frac{s_i}{n} \in \mathcal{C}_i$ for all $i = 1, ..., p$. Repeat until k acceptances.

Take the mean of accepted values $\boldsymbol{\theta}'_1, ..., \boldsymbol{\theta}'_k$ to be the estimate of $\boldsymbol{\theta}$. As a side note, up until this now we have used m to represent the number of acceptances, but we will now use m to represent the "large" sample size and k as the number of acceptances.

4.5.1 Application to normally distributed data

We make application of PM two to normally distributed data as follows

Algorithm PM two
1. Generate $\boldsymbol{\theta'} \overset{i.i.d.}{\sim} \pi(\cdot)$, where $\boldsymbol{\theta'} = (\mu, \sigma^2)'$, and $\mu \sim N(\lambda, \beta)$ and $\sigma^2 \sim \Gamma(\alpha, \delta)$ and observe $(\mu', \sigma^{2'})$. 2. Generate $Y_i \overset{i.i.d.}{\sim} N(\mu', \sigma^{2'})$, $i = 1, ..., m$ where m is sufficiently large. 3. Construct confidence intervals for $E(Y)$ and $E(Y^2)$, and call them $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. 4. Accept $\boldsymbol{\theta'} = (\mu, \sigma^2)'$, if $\bar{x}_n \in \mathcal{C}_1$ and $\overline{x_n^2} \in \mathcal{C}_2$, where $\bar{x}_n = \frac{\sum x}{n}$ and $\overline{x_n^2} = \frac{\sum x^2}{n}$. Repeat until k acceptances.

Take the mean of accepted values, $(\mu, \sigma^2)'_1, ..., (\mu, \sigma^2)'_k$ to be the estimate of (μ, σ^2). Since m is large, we can rely on asymptotic theory for sample moments to construct confidence intervals for $\bar{y}$ and $\overline{y^2}$. We use the following result

$$\frac{\sum y^k}{m} \overset{\circ}{\sim} N\left(E(Y^k), \frac{Var(Y^k)}{m}\right).$$

Thus, in our normal example, we have

$$\bar{y}_m \sim N\left(\mu, \frac{\sigma^2}{m}\right) \quad \text{and} \quad \overline{y_m^2} \overset{\circ}{\sim} N\left(\sigma^2 + \mu, \frac{2\sigma^2}{m}(\sigma^2 + 2\mu^2)\right).$$

So, an approximate $100(1-\alpha)\%$ confidence interval for $E(Y)$ and $E(Y^2)$ is

$$\left(\frac{\sum y}{m} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{m}}, \frac{\sum y}{m} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{m}}\right)$$

and

$$\left(\frac{\sum y^2}{m} - z_{\frac{\alpha}{2}} \frac{\sqrt{2s^2(s^2 + 2\bar{x}^2)}}{\sqrt{m}}, \frac{\sum y^2}{m} + z_{\frac{\alpha}{2}} \frac{\sqrt{2s^2(s^2 + 2\bar{x}^2)}}{\sqrt{m}}\right),$$

respectively.

4.5.2 Iron intake example revisited

We apply PM two with a sample size of $m = 10,000$ and $k = 100$ acceptances. Here, we considered 4 "priors", where each of the priors have a variance of 4 (for both μ and σ^2). The first prior was centered near the MLE's, while the other 3 priors were centered away from the MLE's to compare computing time. The 4 prior distributions are described in Table 4.7 below.

$\pi(\cdot)$	μ	σ^2
$\mathcal{P}_1$	$N(14, 4)$	$\Gamma(9^2/4, 4/9)$
$\mathcal{P}_2$	$N(8, 4)$	$\Gamma(6^2/4, 4/6)$
$\mathcal{P}_3$	$N(10, 4)$	$\Gamma(12^2/4, 4/12)$
$\mathcal{P}_4$	$N(20, 4)$	$\Gamma(4, 1)$

Table 4.7: Prior distributions for PM two and the iron intake data.

Table 4.8 below, shows that regardless of the prior, the estimates are close to the MLE. Even the $\mathcal{P}_4$ prior (which required the most proposals), where the centering values are significantly far from the MLE's, still yielded estimates close to the MLE. The $\mathcal{P}_1$ prior required the fewest proposals (no surprise) and yielded the closest estimates to the MLE.

$\pi(\cdot)$	Proposals	$\hat{\mu}$	$\hat{\sigma}^2$
$\mathcal{P}_1$	16,284	14.6738	9.5248
$\mathcal{P}_2$	210,652	14.6842	8.5534
$\mathcal{P}_3$	37,098	14.6557	10.6451
$\mathcal{P}_4$	467,247	14.6940	8.4712
MLE's	–	14.6800	9.2949

Table 4.8: Results using PM two and the iron intake data.

4.6 Proposed method three

In this section, we extend the above proposed method (and call it Algorithm PM three) to allow for an update in the prior distribution. Here, the idea is to run PM two, and based on the accepted values, we then update the prior distributions based on the average and variability of the accepted values. We then run PM two again for k

acceptances (but now using the updated priors). Additionally, after each iteration of k acceptances, we have the option of increasing m. The thought being, that during the first iteration of k acceptances, we gain some insight of plausible candidate parameters to have generated the data, and thus, as to where to propose values. Also, after the second iteration, we might consider increasing m to further discard "implausible" values, and thus get "closer" to the MLE. We continue this process until the distance between the average of accepted sufficient statistics and the observed sufficient statistics are sufficiently small.

4.6.1 Application to normally distributed data

Let us apply PM three to a population where the observations are normally distributed. To better describe the proposed algorithm in the context of normal data, let us define the following notation, let

$$P^{(v)} = \begin{pmatrix} \mu_1' & \sigma_1^{2'} \\ \mu_2' & \sigma_2^{2'} \\ \vdots & \vdots \\ \mu_k' & \sigma_k^{2'} \end{pmatrix}^{(v)} = \begin{pmatrix} \vec{\boldsymbol{\mu}} & \vec{\boldsymbol{\sigma}}^{\mathbf{2}} \end{pmatrix}^{(v)}$$

be the matrix of accepted parameters for the v^{th} iteration. Here, denote the mean and variance based on $\vec{\boldsymbol{\mu}}$ as $\mathcal{M}_{\mu}^{(v)}$ and $\mathcal{V}_{\mu}^{(v)}$, respectively and denote the mean and variance based on $\vec{\boldsymbol{\sigma}}^{\mathbf{2}}$ as $\mathcal{M}_{\sigma^2}^{(v)}$ and $\mathcal{V}_{\sigma^2}^{(v)}$, respectively. Further, denote the accepted sample moments as

$$S^{(v)} = \begin{pmatrix} \overline{y}_1 & \overline{y_1^2} \\ \overline{y}_2 & \overline{y_2^2} \\ \vdots & \vdots \\ \overline{y}_k & \overline{y_k^2} \end{pmatrix}^{(v)}$$

and now take the average of the accepted sample moments, i.e.,

$$S^{\star(v)} = \begin{pmatrix} \frac{\overline{y}_1+\overline{y}_2+\cdots+\overline{y}_k}{k} \\ \frac{\overline{y_1^2}+\overline{y_2^2}+\cdots+\overline{y_k^2}}{k} \end{pmatrix}^{(v)} = \begin{pmatrix} \overline{\overline{y}} \\ \overline{\overline{y}^2} \end{pmatrix}^{(v)}.$$

We now apply PM three to normally distributed data as follows

Algorithm PM three
1. Propose $\mu'\sim N(\lambda,\beta)$ and $\sigma^{2'}\sim\Gamma(\alpha,\delta)$ and observe $(\mu',\sigma^{2'})$.
2. Generate $Y_i \overset{i.i.d.}{\sim} N(\mu',\sigma^{2'})$, $i=1,...,m$.
3. Construct confidence intervals for $E(Y)$ and $E(Y^2)$, and call them $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively.
4. Accept $\boldsymbol{\theta'} = (\mu',\sigma^{2'})$ if $\overline{x}\in\mathcal{C}_1$ and $\overline{x^2}\in\mathcal{C}_2$.
Repeat until k acceptances.
5. Compute $\mathcal{M}_\mu^{(v)}$, $\mathcal{V}_\mu^{(v)}$, $\mathcal{M}_{\sigma^2}^{(v)}$ and $\mathcal{V}_{\sigma^2}^{(v)}$.
6. Compute $d = \lvert\overline{\overline{y}}-\bar{x}\rvert + \lvert\overline{\overline{y}^2}-\overline{x^2}\rvert$.
7. If $d<\epsilon$ then stop, else repeat steps 1-7 by using $\mu'\sim N(\mathcal{M}_\mu^{(v)},\mathcal{V}_\mu^{(v)})$ and $\sigma^{2'}\sim\Gamma(\frac{\mathcal{M}_{\sigma^2}^{2(v)}}{\mathcal{V}_{\sigma^2}^{(v)}},\frac{\mathcal{V}_{\sigma^2}^{(v)}}{\mathcal{M}_{\sigma^2}^{2(v)}})$.
Note: To prevent the variance of the prior from getting too small, we take the max $\left\{\mathcal{V}_{\sigma^2}^{(v)},b_0\right\}$ or max $\left\{\mathcal{V}_\mu^{(v)},b_0\right\}$ where b_0 is some positive real number.
Optional: Increase m to obtain better precision.

4.6.2 Iron intake example revisited

We ran PM three using 4 (initial) priors, which is representative of different centers and variability. We also use large variances to allow for the proposal of plausible candidate parameters. For each iteration, we ran it for $k = 500$ acceptances. For the first iteration, we used $m = 100$, for the second iteration, we used $m = 1,000$, and for each iteration afterward, we used $m = 10,000$. Furthermore, we used $\epsilon = 0.02$ and $b_0 = 0.5$. The 4 prior distributions are described in Table 4.9, and the results are summarized in Table 4.10.

$\pi(\cdot)$	μ	σ^2
$\mathcal{P}_1$	$N(3, 14)$	$\Gamma(3^2/12, 12/3)$
$\mathcal{P}_2$	$N(6, 14)$	$\Gamma(4^2/9, 9/4)$
$\mathcal{P}_3$	$N(16, 10)$	$\Gamma(14^2/3, 3/14)$
$\mathcal{P}_4$	$N(22, 8)$	$\Gamma(16^2/6, 6/16)$

Table 4.9: Prior distributions for PM three and the iron intake data.

$\mathcal{P}_i$	Iteration	Proposals	Distance	$\hat{\mu}$	$\hat{\sigma}^2$
	1	1,513,301	3.0422	14.5885	6.8530
	2	3,842	0.0274	14.6789	9.3641
	3	15,497	0.1361	14.6769	9.5872
$\mathcal{P}_1$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	20	6,357	0.1012	14.6770	9.4510
	21	6,345	0.0297	14.6751	9.4021
	22	6,411	0.0046	14.6779	9.3785

	Iteration	Proposals	Distance	$\hat{\mu}$	$\hat{\sigma}^2$
	1	126,846	2.5852	14.6551	6.1970
$\mathcal{P}_2$	2	4,022	0.4154	14.6980	8.4177
	3	13,313	0.0018	14.6796	9.3359

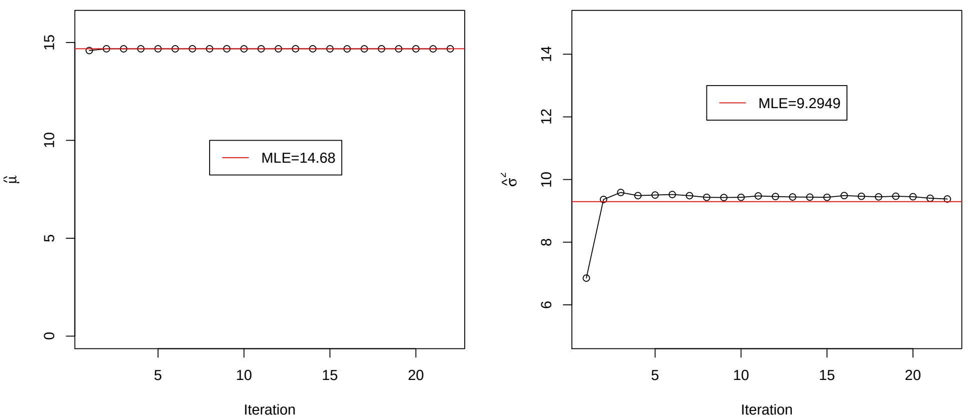
	Iteration	Proposals	Distance	$\hat{\mu}$	$\hat{\sigma}^2$
	1	3,050	4.0252	14.6663	14.2421
	2	2,274	2.5301	14.6094	13.9981
	3	48,969	1.2380	14.6367	11.7381
$\mathcal{P}_3$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	30	6,209	0.0919	14.6755	9.4123
	31	6,119	0.0741	14.6780	9.4115
	32	6,080	0.0144	14.6779	9.4161

	Iteration	Proposals	Distance	$\hat{\mu}$	$\hat{\sigma}^2$
	1	70,475	8.8433	14.9024	15.9534
	2	2,932	3.4648	14.5997	15.6505
	3	112,957	1.2879	14.6401	11.8187
$\mathcal{P}_4$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	60	6,226	0.0466	14.6809	9.3578
	61	5,935	0.0453	14.6741	9.3771
	62	6,527	0.0062	14.6777	9.3774

Table 4.10: Results for PM three and the iron intake data.

From Table 4.10, we can see that regardless of the prior distribution, the parameter values are "moving" closer to the MLE's of $(\hat{\mu}, \hat{\sigma}^2) = (14.68, 9.2949)$, and that the estimate (the final parameter value in the chain), between each of the priors are all comparable. The major difference being the number of iterations required to "move" sufficiently close to the MLE. The number of iterations required were 22, 3, 32, and 62 for $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$, and $\mathcal{P}_4$, respectively. Notice that $\hat{\mu}$ is sufficiently close to the MLE (within the first iteration under all 4 prior distributions) and that it is $\hat{\sigma}^2$ that is slow to be within sufficient distance of the MLE. This is probably explained by the fact that $\sum X^2$ has a large variance (compared to the variance of $\sum X$).

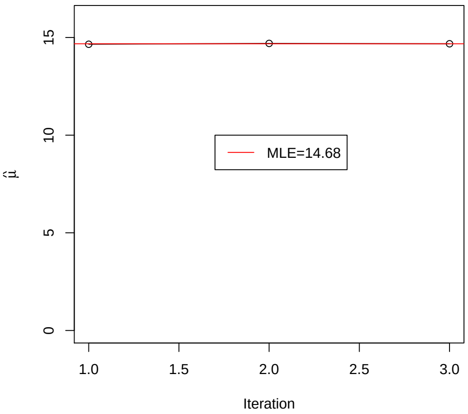
Figures 4.4-4.7 below show the trace plots for $\hat{\mu}$ and $\hat{\sigma}^2$ for P_i, $i = 1, 2, 3, 4$, where the red line represents the MLE. From the figures, we see the quick convergence of $\hat{\mu}$ and eventual convergence of $\hat{\sigma}^2$. Notice that even for the initial priors in low acceptance regions (provided that there is a large variance), this algorithm is “moving” the parameter estimates close to the MLE's.



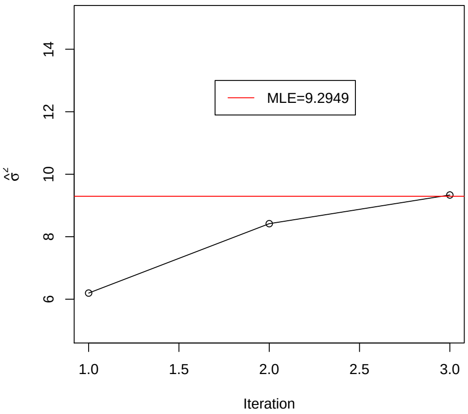
(a) Trace plot for $\hat{\mu}$

(b) Trace plot for $\hat{\sigma}^2$

Figure 4.4: Trace plots for $\hat{\mu}$ and $\hat{\sigma}^2$ under $\mathcal{P}_1$.

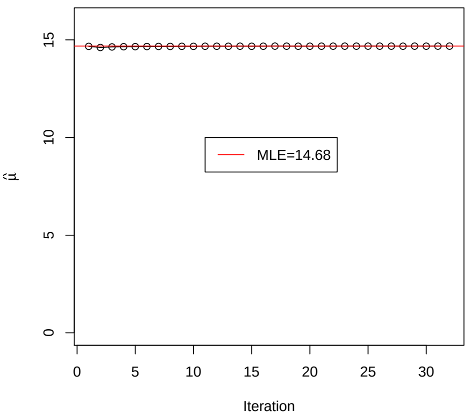


(a) Trace plot for $\hat{\mu}$

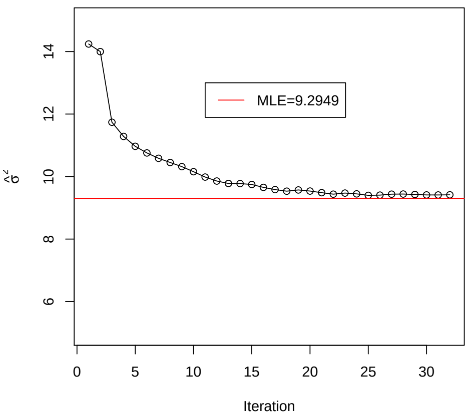


(b) Trace plot for $\hat{\sigma}^2$

Figure 4.5: Trace plots for $\hat{\mu}$ and $\hat{\sigma}^2$ under $\mathcal{P}_2$.

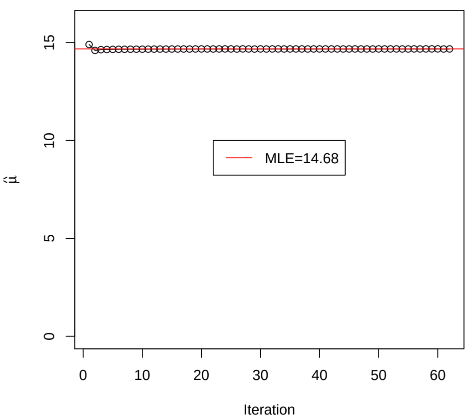


(a) Trace plot for $\hat{\mu}$

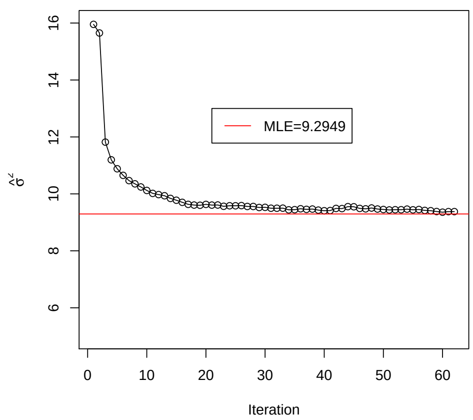


(b) Trace plot for $\hat{\sigma}^2$

Figure 4.6: Trace plots for $\hat{\mu}$ and $\hat{\sigma}^2$ under $\mathcal{P}_3$.



(a) Trace plot for $\hat{\mu}$



(b) Trace plot for $\hat{\sigma}^2$

Figure 4.7: Trace plots for $\hat{\mu}$ and $\hat{\sigma}^2$ under $\mathcal{P}_4$.

4.7 Logistic regression

Now we wish to apply PM two and PM three to a logistic regression problem and then propose an extension of PM two. Consider the following data set, which consists of 40 people who are asked whether or not they would subscribe to a new newspaper. Gender, age, and whether or not they would subscribe to the newspaper were recorded (0=No and 1=Yes). The following data set was taken from the SAS support site demonstrating an example of logistic regression and is summarized in Table 4.11.

Gender	Age	Subscription	Gender	Age	Subscription
Female	35	0	Male	44	0
Male	45	1	Female	47	1
Female	51	0	Female	47	0
Male	54	1	Male	47	1
Female	35	0	Female	34	0
Female	48	0	Female	56	1
Male	46	1	Female	59	1
Female	46	1	Male	59	1
Male	38	1	Female	39	0
Male	49	1	Male	42	1
Male	50	1	Female	45	0
Female	47	0	Female	30	1
Female	39	0	Female	51	0
Female	45	0	Female	43	1
Male	39	1	Male	31	0
Female	39	0	Male	34	0
Female	52	1	Female	46	0
Male	58	1	Female	50	1
Female	32	0	Female	52	1
Female	35	0	Female	51	0

Table 4.11: Data describing newspaper subscription behavior.

The fitted logistic regression model to this data is

$$P(Y_i = 1) = p_i = \frac{e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}}}{1+e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}}}, i = 1, ..., 40.$$

Here β_1 represents the age effect and β_2 is the gender effect.

4.7.1 Newspaper subscription example using PM two

Since the observations are not identically distributed, we cannot choose an arbitrary sample size m and so we need to make a slight adjustment and instead generate an m number of auxiliary data sets. Let $D_j = \{Z_{ij}\}_{i=1}^{40} \overset{ind}{\sim} Ber(p_i)$, and so, if m is large, we have

$$\frac{\sum z}{nm} \overset{p.w.}{\rightarrow} E(Z)$$

$$\frac{\sum zx_1}{nm} \overset{p.w.}{\rightarrow} E(Zx_1)$$

$$\frac{\sum zx_2}{nm} \overset{p.w.}{\rightarrow} E(Zx_2)$$

as $m \rightarrow \infty$.

Thus our algorithm becomes

Algorithm PM two
1. Generate $\beta_0' \sim N(\lambda_0, \delta_0)$, $\beta_1' \sim N(\lambda_1, \delta_1)$, and $\beta_2' \sim N(\lambda_2, \delta_2)$ and observe $(\beta_0', \beta_1', \beta_2')$. 2. Generate D_j based on $(\beta_0', \beta_1', \beta_2'), j = 1, ..., m$ where m is sufficiently large. 3. Construct confidence intervals for $E(Z)$, $E(Zx_1)$, and $E(Zx_2)$ and call them $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$ respectively. 4. Accept $(\beta_0', \beta_1', \beta_2')$ if $\frac{\sum y}{n} \in \mathcal{C}_1$, $\frac{\sum yx_1}{n} \in \mathcal{C}_2$, and $\frac{\sum yx_2}{n} \in \mathcal{C}_3$. Repeat until k acceptances.

Take the mean of accepted values $(\beta_0', \beta_1', \beta_2')_1, ..., (\beta_0', \beta_1', \beta_2')_k$ to be the estimate of $(\beta_0, \beta_1, \beta_2)$. Since the distribution for the sufficient statistics are unknown, we used Monte Carlo simulation to estimate $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$.

Table 4.12 shows the prior distribution ($\mathcal{P}$) that was used. Here, we allowed for a relatively large variance for each parameter. Furthermore, we used $k = 100$ acceptances and $m = 100$.

$\pi(\cdot)$	β_0	β_1	β_1
$\mathcal{P}$	$N(-3, 4)$	$N(0.2, 4)$	$N(-1, 4)$

Table 4.12: Prior distributions for PM two.

Table 4.13 below shows the results for PM two under $\mathcal{P}$. Here, we see that the estimates for β_1 and β_2 are comparable to the MLE's, and that the estimate for β_0

is a little off. Interestingly enough, it is the parameter estimates for the main effects that are comparable. It is also worth noting, that due to the extra complexity of the model (the observations are not identically distributed) that $k = 100$ and $m = 100$ are perhaps a little small and that by increasing k and m, estimation may very well improve.

$\pi(\cdot)$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
$\mathcal{P}$	-4.2512	0.1302	-2.3196
MLE's	-5.762	0.1649	-2.4224

Table 4.13: Results for PM two and the newspaper subscription data.

4.7.2 Newspaper subscription example revisited using PM three

We will now apply PM three. To better describe PM three in the context of a logistic regression, let us define the following notation. Let

$$P^{(v)} = \begin{pmatrix} \beta'_{01} & \beta'_{11} & \beta'_{21} \\ \beta'_{02} & \beta'_{12} & \beta'_{22} \\ \vdots & \vdots & \vdots \\ \beta'_{0k} & \beta'_{1k} & \beta'_{2k} \end{pmatrix}^{(v)} = \begin{pmatrix} \vec{\boldsymbol{\beta}}_0 & \vec{\boldsymbol{\beta}}_1 & \vec{\boldsymbol{\beta}}_2 \end{pmatrix}^{(v)}$$

be the matrix of accepted parameter values for the v^{th} iteration. Here, denote the mean and variance based on $\vec{\boldsymbol{\beta}}_0$ as $\mathcal{M}_{\beta_0}^{(v)}$ and $\mathcal{V}_{\beta_0}^{(v)}$, respectively, denote the mean and variance based on $\vec{\boldsymbol{\beta}}_1$ as $\mathcal{M}_{\beta_1}^{(v)}$ and $\mathcal{V}_{\beta_1}^{(v)}$, respectively, and denote the mean and variance based on $\vec{\boldsymbol{\beta}}_2$ as $\mathcal{M}_{\beta_2}^{(v)}$ and $\mathcal{V}_{\beta_2}^{(v)}$, respectively. Further, denote accepted sample moments as

$$S^{(v)} = \begin{pmatrix} \left(\frac{\sum z}{nm}\right)_1 & \left(\frac{\sum zx_1}{nm}\right)_1 & \left(\frac{\sum zx_2}{nm}\right))_1 \\ \left(\frac{\sum z}{nm}\right)_2 & \left(\frac{\sum zx_1}{nm}\right)_2 & \left(\frac{\sum zx_2}{nm}\right)_2 \\ \vdots & \vdots & \vdots \\ \left(\frac{\sum z}{nm}\right)_k & \left(\frac{\sum zx_1}{nm}\right)_k & \left(\frac{\sum zx_2}{nm}\right)_k \end{pmatrix}^{(v)}$$

and then take the average of the accepted sample moments

$$S^{\star(v)} = \begin{pmatrix} \frac{\left(\frac{\sum z}{nm}\right)_1 + \left(\frac{\sum z}{nm}\right)_2 + \cdots + \left(\frac{\sum z}{nm}\right)_k}{k} \\ \frac{\left(\frac{\sum zx_1}{nm}\right)_1 + \left(\frac{\sum zx_1}{nm}\right)_2 + \cdots + \left(\frac{\sum zx_1}{nm}\right)_k}{k} \\ \frac{\left(\frac{\sum zx_2}{nm}\right)_1 + \left(\frac{\sum zx_2}{nm}\right)_2 + \cdots + \left(\frac{\sum zx_2}{nm}\right)_k}{k} \end{pmatrix}^{(v)} = \begin{pmatrix} \overline{\frac{\sum z}{nm}} \\ \overline{\frac{\sum zx_1}{nm}} \\ \overline{\frac{\sum zx_2}{nm}} \end{pmatrix}^{(v)} .$$

As with PM two, we used Monte Carlo simulation to estimate $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$. Using the newspaper subscription data, the algorithm is as follows

---

Algorithm PM three

---

1. Propose $\beta_0' \sim N(\lambda_0, \delta_0)$, $\beta_1' \sim N(\lambda_1, \delta_1)$, and $\beta_2' \sim N(\lambda_2, \delta_2)$ and observe $(\beta_0', \beta_1', \beta_2')$.
2. Generate D_j based on $(\beta_0', \beta_1', \beta_2'), j = 1, ..., m$ where m is sufficiently large.
3. Construct confidence intervals for $E(Z)$, $E(Zx_1)$, and $E(Zx_2)$ and call them $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$ respectively.
4. Accept $(\beta_0', \beta_1', \beta_2')$ if $\frac{\sum y}{n} \in \mathcal{C}_1$, $\frac{\sum yx_1}{n} \in \mathcal{C}_2$, and $\frac{\sum yx_2}{n} \in \mathcal{C}_3$. Repeat until k acceptances.
5. Compute $d = |\overline{\frac{\sum z}{nm}} - \frac{\sum y}{n}| + |\overline{\frac{\sum zx_1}{nm}} - \frac{\sum yx_1}{n}| + |\overline{\frac{\sum zx_2}{nm}} - \frac{\sum yx_2}{n}|$.
6. Compute $\mathcal{M}_{\beta_0}^{(v)}$, $\mathcal{V}_{\beta_0}^{(v)}$, $\mathcal{M}_{\beta_1}^{(v)}$, $\mathcal{V}_{\beta_1}^{(v)}$, $\mathcal{M}_{\beta_2}^{(v)}$ and $\mathcal{V}_{\beta_2}^{(v)}$.
7. If $d < \epsilon$ then stop, else repeat steps 1-7 by using $\beta_0' \sim N(\mathcal{M}_{\beta_0}^{(v)}, \mathcal{V}_{\beta_0}^{(v)})$, $\beta_1' \sim N(\mathcal{M}_{\beta_1}^{(v)}, \mathcal{V}_{\beta_1}^{(v)})$, and $\beta_2' \sim N(\mathcal{M}_{\beta_2}^{(v)}, \mathcal{V}_{\beta_2}^{(v)})$.

Note: To prevent the variance of the prior from getting too small, we take the max $\left\{\mathcal{V}_{\beta_0}^{(v)}, b_0\right\}$, max $\left\{\mathcal{V}_{\beta_1}^{(v)}, b_0\right\}$, and max $\left\{\mathcal{V}_{\beta_2}^{(v)}, b_0\right\}$ where b_0 is some positive real number.

Optional: Increase m to obtain better precision.

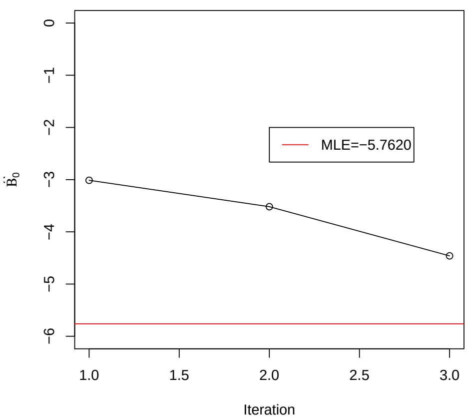
4.7.3 Results of newspaper subscription using PM three

We used the same initial prior distributions as those used in Table 4.12. For each iteration, we ran it for $k = 15$ acceptances and $m = 100$. Furthermore, we used $\epsilon = 0.2$ and $b_0 = 0.5$.

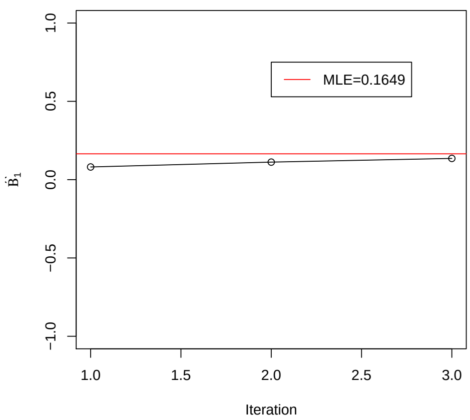
$\mathcal{P}$	Iteration	Proposals	Distance	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
	1	4,807	0.2464	-3.0122	0.0809	-0.8603
	2	200,479	0.2025	-3.5180	0.1120	-2.1868
	3	11,668	0.0402	-4.4598	0.1356	-2.3576

Table 4.14: Results for PM three and the newspaper subscription data.

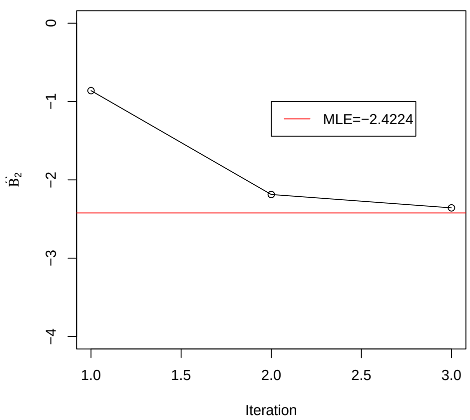
Table 4.14 shows the results for PM three. The algorithm converged after 3 iterations. As with PM two, we see that the estimates for β_1 and β_2 are comparable to the MLE while the estimate for β_0 is a little distant to the MLE (but still reasonable). Figures 4.8 below show the trace plots for $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ under P, where the red line represents the MLE. Again, we see that $\hat{\beta}_1$ and $\hat{\beta}_2$ are close to the MLE, however $\hat{\beta}_0$ is a little distant, but appears to be converging. If perhaps a smaller ϵ was chosen and/or a larger k and m was used, $\hat{\beta}_0$ would converge.



(a) Trace plot for $\hat{\beta}_0$



(b) Trace plot for $\hat{\beta}_1$



(c) Trace plot for $\hat{\beta}_2$

Figure 4.8: Trace plots for $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$.

4.8 Proposed method four

The final proposed algorithm (which we will call PM four), is an extension of PM two. The first step will apply PM two for k acceptances, i.e., $(\beta'_{0i}, \beta'_{1i}, \beta'_{2i}), i = 1, ..., k$ based on size m. For step 2, take each of the $(\beta'_{0i}, \beta'_{1i}, \beta'_{2i}), i = 1, ..., k$ and apply PM two, where the center for the prior for β'_{0i}, β'_{1i}, and β'_{2i}, are at β'_{0i}, β'_{1i}, and β'_{2i}, respectively, coupled with a small variance (say σ^2), and run it for k_1 acceptances, i.e., $(\beta'_{0i}, \beta'_{1i}, \beta'_{2i})$ will generate $(\beta'_{0ij}, \beta'_{1ij}, \beta'_{2ij}), j = 1, ..., k_1$ (based on size m_1).

For step 3, for each of the accepted k candidate parameters (from step 1), take each of the k_1 acceptances (that each of the k parameters produced), and measure the distance between the auxiliary sufficient statistics to the observed sufficient statistics. For example, consider the first accepted parameter in step 1, say (β'_{01}, β'_{11}, β'_{21}). Now, (β'_{01}, β'_{11}, β'_{21}) will have generated a set of k_1 acceptances, i.e., (β'_{011}, β'_{111}, β'_{211}). Further, (β'_{011}, β'_{111}, β'_{211}) will have generated m_1 data sets of size 40. From these m_1 data sets, compute the distance between the auxiliary sufficient statistics to the observed sufficient statistics, i.e.,

$$d_{ij} = \left| \left(\frac{\sum z}{nm_1} \right)_{ij} - \frac{\sum y}{n} \right| + \left| \left(\frac{\sum zx_1}{nm_1} \right)_{ij} - \frac{\sum yx_1}{n} \right| + \left| \left(\frac{\sum zx_2}{nm_1} \right)_{ij} - \frac{\sum yx_2}{n} \right|$$

and take the average of these distances, i.e.,

$$\bar{d}_i = \frac{\sum_{j=1}^{k_1} d_{ij}}{k_1}, i = 1, ..., k$$

(the idea being that the smaller the distance, the closer we are to the MLE). For step 4, order and keep the c_1 (where $c_1 < k$) parameter values that produced the smallest c_1 distances, and call them $(\beta_0^{(t)}, \beta_1^{(t)}, \beta_2^{(t)}), t = 1, ..., c_1$ (here, in this context, the superscript (t) represents the ordered value and not the iteration). Step 5 is to update our prior distribution by centering the prior for $\beta_0^{(t)}$,$\beta_1^{(t)}$, and $\beta_2^{(t)}$ at the average of the k_1 parameters (that $(\beta_0^{(t)}, \beta_1^{(t)}, \beta_2^{(t)})$ generated), which we denote as, $\mathcal{M}_{\beta_0}^{(t)}$, $\mathcal{M}_{\beta_1}^{(t)}$, $\mathcal{M}_{\beta_2}^{(t)}$, respectively, (and again use a "small" variance). Now repeat steps 2-4 (now for k_2 and m_2) and keep the c_2 parameter values that produced the smallest c_2 distances (where $c_2 < c_1$). Keep repeating this cycle until we only keep the 1 candidate parameter that produced the smallest distance and take the average of the parameters (that this 1 parameter generated) as our estimate for $(\beta_0, \beta_1, \beta_2)$. Note that at each cycle, the size of candidate parameters reduces from k to c_1 to c_2 and so forth until 1.

A few comments are in order. First, it's up the individual at what rate they wish to decrease the size of candidate parameters, and second, as described above, the number of acceptances and the size of the generated data sets can increase between each iteration. PM four is described as follows

Algorithm PM four
1. Run PM two for k acceptances and size m. Hence, the output is, $P = \{\boldsymbol{\theta}'_1, ..., \boldsymbol{\theta}'_k\}$ where $\boldsymbol{\theta}'_i = (\beta'_{0i}, \beta'_{1i}, \beta'_{2i})$.
2. For each $\boldsymbol{\theta}'_i \in P$, run PM two, i.e., $\beta'_{0i} \sim N(\beta'_{0i}, \sigma^2)$, $\beta'_{1i} \sim N(\beta'_{1i}, \sigma^2)$, and $\beta'_{2i} \sim N(\beta'_{2i}, \sigma^2)$ for k_1 acceptances and size m_1.
3. For each $(\beta'_{0i}, \beta'_{1i}, \beta'_{2i}), i = 1, ..., k$, compute $\bar{d}_i$.
4. Order and keep the c_1 parameter values that produced the smallest c_1 distances.
5. Update the prior distributions that center around $\mathcal{M}_{\beta_0}^{(t)}$, $\mathcal{M}_{\beta_1}^{(t)}$, and $\mathcal{M}_{\beta_2}^{(t)}$ respectively.
Repeat steps 2-4 until we only keep the 1 parameter value that produced the smallest distance and take the average of the parameters that this 1 parameter generated as our estimate for $(\beta_0, \beta_1, \beta_2)$.

We now apply PM four to our logistic regression problem. For the initial prior distributions, we use the same priors that were used in Table 4.12. For step 1, we will run PM two for $k = 100$ acceptances and size $m = 100$. For step 2, we will use $k_1 = 100$ and $m_1 = 100$. For step 4, we keep the 10 parameters that generated the 10 smallest (i.e., $c_1 = 10$) distances, and we update our priors. We repeat by using $k_2 = 1,000$ and $m_2 = 100$, and keep the parameter that produced the smallest distance and take the average of the parameters that this 1 parameter generated as our estimate for $(\beta_0, \beta_1, \beta_2)$. We repeat the simulation twice using $\sigma^2 = 0.01$ and $\sigma^2 = 0.04$.

4.8.1 Results of newspaper subscription using PM four

σ	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_1$
0.1	-3.9327	0.1267	-2.5102
0.2	-3.6211	0.1171	-2.3533

Table 4.15: Results for PM four and the newspaper subscription data.

Table 4.15 shows the results after applying PM four. The estimates for β_1 and β_2 are similar to PM two and PM three, however, the estimate for β_0 did not perform quite as well as with PM two and PM three. We note that the implementation of PM two, PM three, and PM four on this example was very limited, in the sense that the simulations were not exhaustive. We feel confident that with extensive simulation studies and perhaps with appropriate tweaking, these algorithms can accomplish the end goal.

4.9 Summary

In summary of chapter 4, we constructed algorithms in an attempt to provide estimates that are comparable to the MLE. In the algorithms that we proposed, we made application to a data set where the observations were assumed to be normally distributed and to another data set where logistic regression was used. For the normal example, each of the algorithms seemed to work well (in particular PM two and PM three), however we keep in mind that it is a rather simple and low dimensional problem. The logistic regression example was more challenging, in that the observations were not identically distributed, and so more computational effort was required. Given this, we are pleased the estimates for β_1 and β_2 were close to the MLE and β_0 was still rather reasonable. We are confident that heavier computational effort will improve on the convergence of $\hat{\beta}_0$ to the MLE. We feel that we laid some groundwork for finding a way to obtain MLE type estimates in the absence of the likelihood. At the start of this work, one of the goals was to improve on the acceptance rates over ABC-AR (and perhaps over other ABC methods without the cost of highly correlated draws), however, time did not allow for this development. Nonetheless, we feel that one can implement the ideas that allow for an "update" in the prior distributions and develop ways to improve on the low acceptance rates.

We don't believe that PM one has much value since it was not founded on any theoretical justification, however, we believe that the PM two, PM three, and PM four algorithms do have promise because it is founded on large sample asymptotic theory. We feel that with more tweaking and refining, an algorithm can be developed that can work well in models where sufficient statistics are known. Once this has been established, one can apply these ideas to models where sufficient statistics are unknown and replaced with summary statistics.

Appendix A

Simulation tables

$\boldsymbol{\alpha} = (1,1,1,1,1)'$, $N = 200, n = 100$											
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	–		92,759(318,929)		127,845(478,956)		891,239(2,950,074)		974,948(2,999,789)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.008	0.291	0.248	0.070	0.353	0.132	1.028	1.139	1.286	1.744
	$\hat{\alpha}_2$	-0.095	0.292	0.244	0.072	0.351	0.133	1.023	1.138	1.284	1.756
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.081	0.156	0.213	0.082	0.306	0.154	0.716	0.718	0.725	0.925
	$\hat{\alpha}_4$	0.139	0.214	0.218	0.089	0.312	0.167	0.724	0.747	0.730	0.962
	$\hat{\alpha}_5$	-0.014	0.233	0.431	0.199	0.630	0.410	1.779	3.323	2.140	4.789
	Proposals	–		881,984(2,415,832)		1,136,716(2,661,383)		4,685,922(5,068,450)		5,730,297(5,089,028)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.008	0.291	0.248	0.071	0.406	0.172	0.795	0.698	1.115	1.367
	$\hat{\alpha}_2$	-0.095	0.292	0.241	0.069	0.399	0.167	0.794	0.701	1.107	1.344
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.081	0.156	0.070	0.039	0.078	0.071	0.156	0.078	-0.035	0.064
	$\hat{\alpha}_4$	0.139	0.214	0.078	0.046	0.084	0.079	0.159	0.090	-0.026	0.079
	$\hat{\alpha}_5$	-0.014	0.233	0.313	0.109	0.505	0.265	0.966	1.005	1.285	1.786
	Proposals	–		5,619,398(4,416,106)		8,059,370(4,806,564)		14,174,693(2,490,184)		6,572,853(5,440,669)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.008	0.291	0.210	0.066	0.367	0.172	0.529	0.343	1.069	1.282
	$\hat{\alpha}_2$	-0.095	0.292	0.187	0.055	0.336	0.147	0.489	0.298	1.051	1.257
$\epsilon = 0.2$	$\hat{\alpha}_3$	0.081	0.156	0.011	0.029	-0.030	0.056	0.013	0.032	-0.047	0.061
	$\hat{\alpha}_4$	0.139	0.214	0.025	0.038	-0.012	0.071	0.038	0.041	-0.034	0.073
	$\hat{\alpha}_5$	-0.014	0.233	0.218	0.064	0.378	0.171	0.527	0.323	1.216	1.641

Table A.1: ABC-AR Bias and MSE comparisons for $A_1, n = 100$ and old S_5.

$\boldsymbol{\alpha} = (3, 2.5, 2, 1.5, 1)'$, $N = 200, n = 100$											
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	–		22,874(3,292)		22,123(2,596)		12,182(1,616)		13,506(1,534)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.174	0.833	-1.067	1.142	-1.220	1.490	0.655	0.445	0.469	0.226
	$\hat{\alpha}_2$	0.075	0.598	-0.695	0.489	-0.819	0.673	0.815	0.684	0.774	0.608
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.032	0.240	-0.332	0.116	-0.160	0.029	0.909	0.847	1.286	1.671
	$\hat{\alpha}_4$	0.040	0.178	-0.225	0.054	-0.047	0.007	0.675	0.468	0.984	0.988
	$\hat{\alpha}_5$	0.047	0.122	-0.327	0.109	-0.455	0.211	0.381	0.153	0.210	0.062
	Proposals	–		131,960(27,586)		112,302(18,459)		57,470(25,056)		69,909(41,571)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.174	0.833	-0.956	0.920	-1.135	1.291	0.740	0.571	0.511	0.269
	$\hat{\alpha}_2$	0.075	0.598	-0.584	0.349	-0.775	0.605	0.863	0.785	0.856	0.744
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.032	0.240	-0.165	0.040	-0.025	0.008	0.779	0.666	1.182	1.487
	$\hat{\alpha}_4$	0.040	0.178	-0.132	0.027	0.028	0.012	0.493	0.280	0.776	0.702
	$\hat{\alpha}_5$	0.047	0.122	-0.334	0.114	-0.484	0.242	0.339	0.124	0.168	0.060
	Proposals	–		1,205,033(343,381)		1,043,427(424,560)		553,499(607,256)		887,980(835,825)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.174	0.833	-0.784	0.625	-0.943	0.895	0.633	0.448	0.431	0.209
	$\hat{\alpha}_2$	0.075	0.598	-0.418	0.189	-0.587	0.356	0.724	0.588	0.936	0.900
$\epsilon = 0.2$	$\hat{\alpha}_3$	0.032	0.240	-0.049	0.045	0.026	0.031	0.327	0.245	0.517	0.508
	$\hat{\alpha}_4$	0.040	0.178	-0.067	0.036	0.033	0.038	0.139	0.100	0.145	0.207
	$\hat{\alpha}_5$	0.047	0.122	-0.276	0.085	-0.398	0.177	0.342	0.128	0.331	0.159

Table A.2: ABC-AR Bias and MSE comparisons for $A_2, n = 100$ and old S_5.

		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
		$\boldsymbol{\alpha} = (1,1,2,6,1)'$, $N = 200, n = 100$									
	Proposals	–		83,646(18,287)		127,609(34,582)		38,101(11,350)		52,373(14,330)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\epsilon = 0.8$	$\hat{\alpha}_1$	0.012	0.252	0.079	0.009	0.093	0.011	0.810	0.668	1.000	1.015
	$\hat{\alpha}_2$	0.187	1.160	0.680	0.464	0.772	0.597	2.259	5.110	2.363	5.586
	$\hat{\alpha}_3$	0.052	0.144	-0.962	0.928	-1.009	1.023	-0.449	0.210	-0.645	0.428
	$\hat{\alpha}_4$	0.142	1.136	-3.140	9.872	-3.360	11.298	-1.169	1.402	-1.474	2.183
	$\hat{\alpha}_5$	0.027	0.181	-0.038	0.005	-0.006	0.007	1.004	1.023	1.154	1.357
	Proposals	–		934,642(276,086)		3,470,468(4,091,710)		212,263(136,276)		313,768(149,704)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\epsilon = 0.4$	$\hat{\alpha}_1$	0.012	0.252	-0.021	0.003	-0.142	0.025	0.747	0.569	0.857	0.753
	$\hat{\alpha}_2$	0.187	1.160	0.613	0.378	0.715	0.520	2.224	4.951	2.431	5.915
	$\hat{\alpha}_3$	0.052	0.144	-0.750	0.571	-0.701	0.512	-0.402	0.176	-0.594	0.368
	$\hat{\alpha}_4$	0.142	1.136	-2.440	5.988	-2.590	6.867	-0.606	0.457	-1.036	1.115
	$\hat{\alpha}_5$	0.027	0.181	-0.013	0.007	-0.045	0.016	1.021	1.059	1.130	1.307
	Proposals	–		11,816,381(3,328,399)		12,096,917(3,254,128)		2,191,733(2,939,107)		8,253,041(5,422,754)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\epsilon = 0.2$	$\hat{\alpha}_1$	0.012	0.252	-0.115	0.017	-0.124	0.022	0.628	0.444	0.177	0.176
	$\hat{\alpha}_2$	0.187	1.160	0.541	0.299	0.554	0.314	2.014	4.224	1.279	2.467
	$\hat{\alpha}_3$	0.052	0.144	-0.485	0.258	-0.485	0.258	-0.248	0.108	-0.417	0.220
	$\hat{\alpha}_4$	0.142	1.136	-1.771	3.208	-1.795	3.301	-0.173	0.448	-1.281	2.087
	$\hat{\alpha}_5$	0.027	0.181	-0.004	0.012	-0.006	0.014	0.947	0.986	0.409	0.445

Table A.3: ABC-AR Bias and MSE comparisons for $A_3, n = 100$ and old S_5.

$\boldsymbol{\alpha} = (1,1,1,1,1)'$, $N = 200, n = 50$											
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	–		188,196(1,109,806)		228,092(1,166,361)		1,037,310(3,322,266)		1,449,447(3,819,222)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-0.001	0.512	0.266	0.087	0.370	0.151	1.128	1.381	1.354	1.963
	$\hat{\alpha}_2$	-0.147	0.413	0.256	0.082	0.360	0.144	1.100	1.331	1.315	1.871
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.171	0.285	0.257	0.116	0.355	0.203	0.897	1.078	0.924	1.383
	$\hat{\alpha}_4$	0.231	0.292	0.247	0.114	0.343	0.200	0.879	1.046	0.902	1.343
	$\hat{\alpha}_5$	-0.032	0.450	0.407	0.186	0.591	0.373	1.787	3.360	2.097	4.663
	Proposals	–		1,205,237(3,205,741)		1,426,153(3,378,301)		4,337,631(5,390,737)		8,376,889(6,028,584)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-0.001	0.512	0.271	0.090	0.408	0.184	0.926	0.955	1.019	1.188
	$\hat{\alpha}_2$	-0.147	0.413	0.250	0.078	0.398	0.170	0.898	0.907	0.988	1.100
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.171	0.285	0.130	0.068	0.170	0.113	0.320	0.205	-0.102	0.090
	$\hat{\alpha}_4$	0.231	0.292	0.125	0.069	0.158	0.116	0.309	0.195	-0.111	0.078
	$\hat{\alpha}_5$	-0.032	0.450	0.309	0.113	0.486	0.257	1.058	1.220	1.149	1.466
	Proposals	–		6,310,226(5,036,628)		8,092,595(5,279,942)		13,486,863(3,324,257)		10,355,097(5,562,866)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-0.001	0.512	0.218	0.083	0.354	0.183	0.598	0.472	0.879	0.934
	$\hat{\alpha}_2$	-0.147	0.413	0.188	0.063	0.314	0.143	0.571	0.428	0.867	0.867
$\epsilon = 0.2$	$\hat{\alpha}_3$	0.171	0.285	0.072	0.049	0.068	0.085	0.144	0.080	-0.186	0.078
	$\hat{\alpha}_4$	0.231	0.292	0.073	0.050	0.070	0.086	0.133	0.084	-0.194	0.070
	$\hat{\alpha}_5$	-0.032	0.450	0.210	0.072	0.349	0.164	0.590	0.435	0.999	1.141

Table A.4: ABC-AR Bias and MSE comparisons for $A_1, n = 50$ and old S_5.

		$\boldsymbol{\alpha} = (3, 2.5, 2, 1.5, 1)'$, $N = 200, n = 50$									
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	–		22,403(4,495)		21,901(3,510)		12,440(2,535)		13,761(2,437)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.178	1.846	-1.096	1.208	-1.229	1.512	0.638	0.433	0.461	0.221
	$\hat{\alpha}_2$	0.271	1.221	-0.720	0.527	-0.821	0.679	0.815	0.699	0.787	0.639
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.117	0.466	-0.365	0.143	-0.186	0.041	0.891	0.828	1.258	1.608
	$\hat{\alpha}_4$	0.061	0.301	-0.233	0.061	-0.054	0.010	0.689	0.500	0.985	1.007
	$\hat{\alpha}_5$	0.095	0.278	-0.312	0.100	-0.432	0.194	0.396	0.171	0.235	0.086
	Proposals	–		133,119(39,111)		118,037(31,813)		65,390(48,039)		80,257(72,943)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.178	1.846	-0.983	0.978	-1.145	1.313	0.716	0.556	0.500	0.264
	$\hat{\alpha}_2$	0.271	1.221	-0.607	0.383	-0.767	0.596	0.857	0.799	0.852	0.748
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.117	0.466	-0.220	0.069	-0.070	0.019	0.737	0.640	1.121	1.406
	$\hat{\alpha}_4$	0.061	0.301	-0.156	0.041	0.008	0.021	0.494	0.311	0.762	0.739
	$\hat{\alpha}_5$	0.095	0.278	-0.310	0.101	-0.454	0.219	0.352	0.140	0.189	0.088
	Proposals	–		1,355,351(710,315)		1,263,545(941,384)		757,854(1,163,120)		1,140,260(1,477,687)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.178	1.846	-0.835	0.718	-0.982	0.974	0.625	0.475	0.425	0.234
	$\hat{\alpha}_2$	0.271	1.221	-0.444	0.225	-0.588	0.366	0.746	0.672	0.934	0.917
$\epsilon = 0.2$	$\hat{\alpha}_3$	0.117	0.466	-0.122	0.075	-0.033	0.046	0.348	0.308	0.557	0.648
	$\hat{\alpha}_4$	0.061	0.301	-0.108	0.057	0.001	0.055	0.179	0.144	0.213	0.305
	$\hat{\alpha}_5$	0.095	0.278	-0.255	0.078	-0.372	0.166	0.348	0.142	0.321	0.176

Table A.5: ABC-AR Bias and MSE comparisons for $A_2, n = 50$ and old S_5.

		$\boldsymbol{\alpha} = (1,1,2,6,1)'$, $N = 200, n = 50$									
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	–		80,432(27,556)		120,445(50,883)		39,237(13,640)		53,395(19,204)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.041	0.483	0.060	0.008	0.088	0.012	0.816	0.683	0.991	1.010
	$\hat{\alpha}_2$	0.280	2.218	0.668	0.448	0.753	0.567	2.254	5.088	2.360	5.572
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.114	0.247	-0.981	0.966	-1.030	1.066	-0.445	0.212	-0.630	0.416
	$\hat{\alpha}_4$	0.365	2.103	-3.228	10.435	-3.434	11.798	-1.199	1.496	-1.496	2.255
	$\hat{\alpha}_5$	0.027	0.277	-0.025	0.008	0.020	0.013	1.014	1.061	1.159	1.396
	Proposals	–		857,922(404,548)		6,355,496(6,390,860)		215,409(85,888)		322,089(147,109)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.041	0.483	-0.027	0.005	-0.064	0.027	0.742	0.570	0.841	0.738
	$\hat{\alpha}_2$	0.280	2.218	0.612	0.377	0.530	0.382	2.226	4.961	2.405	5.790
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.114	0.247	-0.796	0.643	-0.972	1.007	-0.402	0.180	-0.584	0.363
	$\hat{\alpha}_4$	0.365	2.103	-2.604	6.824	-3.738	15.294	-0.670	0.571	-1.087	1.228
	$\hat{\alpha}_5$	0.027	0.277	0.007	0.012	0.031	0.024	1.028	1.094	1.134	1.349
	Proposals	–		14,486,738(1,713,293)		14,922,411(619,037)		3,919,741(5,051,840)		9,601,868(6,125,378)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.041	0.483	0.066	0.025	0.067	0.030	0.571	0.387	0.302	0.177
	$\hat{\alpha}_2$	0.280	2.218	0.161	0.075	0.150	0.070	1.776	3.724	1.102	2.504
$\epsilon = 0.2$	$\hat{\alpha}_3$	0.114	0.247	-1.198	1.500	-1.235	1.569	-0.417	0.375	-0.869	1.002
	$\hat{\alpha}_4$	0.365	2.103	-4.864	24.915	-5.088	26.458	-1.017	5.033	-3.246	15.897
	$\hat{\alpha}_5$	0.027	0.277	0.096	0.021	0.099	0.022	0.856	0.886	0.513	0.517

Table A.6: ABC-AR Bias and MSE comparisons for $A_3, n = 50$ and old S_5.

$\boldsymbol{\alpha} = (1,1,1,1,1)'$, $N = 200, n = 100$											
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	–		6,320(544)		6,940(619)		5,857(751)		6,720(915)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\epsilon = 0.8$	$\hat{\alpha}_1$	0.008	0.291	0.309	0.101	0.447	0.206	1.572	2.501	1.865	3.511
	$\hat{\alpha}_2$	-0.095	0.292	0.311	0.103	0.449	0.209	1.576	2.516	1.856	3.487
	$\hat{\alpha}_3$	0.081	0.156	0.354	0.136	0.524	0.287	1.600	2.618	1.961	3.915
	$\hat{\alpha}_4$	0.139	0.214	0.356	0.139	0.527	0.293	1.605	2.639	1.970	3.961
	$\hat{\alpha}_5$	-0.014	0.233	0.336	0.125	0.522	0.286	1.629	2.710	2.013	4.116
	Proposals	–		50,369(10,637)		61,438(20,344)		166,248(168,771)		377,764(471,901)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\epsilon = 0.4$	$\hat{\alpha}_1$	0.008	0.291	0.290	0.102	0.464	0.236	1.132	1.399	1.365	2.075
	$\hat{\alpha}_2$	-0.095	0.292	0.288	0.100	0.463	0.236	1.133	1.400	1.365	2.080
	$\hat{\alpha}_3$	0.081	0.156	0.270	0.097	0.456	0.244	0.995	1.112	1.207	1.701
	$\hat{\alpha}_4$	0.139	0.214	0.273	0.101	0.460	0.249	0.998	1.119	1.209	1.713
	$\hat{\alpha}_5$	-0.014	0.233	0.306	0.118	0.499	0.278	1.147	1.452	1.406	2.234
	Proposals	–		926,581(427,404)		1,989,341(1,444,414)		10,678,542(4,732,615)		14,000,000(2,424,044)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\epsilon = 0.2$	$\hat{\alpha}_1$	0.008	0.291	0.175	0.065	0.270	0.130	0.407	0.237	0.410	0.306
	$\hat{\alpha}_2$	-0.095	0.292	0.152	0.056	0.243	0.114	0.384	0.215	0.364	0.255
	$\hat{\alpha}_3$	0.081	0.156	0.110	0.037	0.186	0.075	0.263	0.113	0.232	0.110
	$\hat{\alpha}_4$	0.139	0.214	0.125	0.047	0.201	0.091	0.277	0.132	0.257	0.142
	$\hat{\alpha}_5$	-0.014	0.233	0.175	0.061	0.273	0.117	0.398	0.219	0.391	0.233

Table A.7: ABC-AR Bias and MSE comparisons for $A_1, n = 100$ and new S_5.

	$\boldsymbol{\alpha} = (3, 2.5, 2, 1.5, 1)'$, $N = 200, n = 100$										
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	–		14,135(2,374)		18,419(3,705)		11,898(2,138)		15,696(3,184)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.174	0.833	-1.190	1.424	-1.187	1.413	0.582	0.359	0.554	0.318
	$\hat{\alpha}_2$	0.075	0.598	-0.831	0.696	-0.739	0.554	0.749	0.582	0.925	0.882
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.032	0.240	-0.546	0.310	-0.350	0.134	0.746	0.601	1.108	1.277
	$\hat{\alpha}_4$	0.040	0.178	-0.404	0.174	-0.265	0.089	0.520	0.308	0.759	0.652
	$\hat{\alpha}_5$	0.047	0.122	-0.212	0.050	-0.229	0.063	0.454	0.222	0.396	0.190
	Proposals	–		137,776(34,307)		188,337(64,245)		86,778(19,653)		135,038(36,313)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.174	0.833	-0.981	0.977	-0.999	1.005	0.714	0.545	0.646	0.430
	$\hat{\alpha}_2$	0.075	0.598	-0.599	0.373	-0.524	0.293	0.874	0.799	1.150	1.363
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.032	0.240	-0.438	0.214	-0.302	0.115	0.666	0.515	0.981	1.062
	$\hat{\alpha}_4$	0.040	0.178	-0.370	0.158	-0.288	0.116	0.395	0.219	0.521	0.395
	$\hat{\alpha}_5$	0.047	0.122	-0.210	0.058	-0.227	0.070	0.320	0.137	0.309	0.150
	Proposals	–		2,183,264(796,921)		2,702,002(1,362,655)		920,612(255,152)		1,595,734(514,779)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.174	0.833	-0.733	0.564	-0.781	0.623	0.643	0.479	0.563	0.357
	$\hat{\alpha}_2$	0.075	0.598	-0.343	0.149	-0.341	0.148	0.789	0.689	1.100	1.275
$\epsilon = 0.2$	$\hat{\alpha}_3$	0.032	0.240	-0.261	0.103	-0.200	0.068	0.542	0.379	0.801	0.760
	$\hat{\alpha}_4$	0.040	0.178	-0.264	0.102	-0.223	0.089	0.298	0.160	0.361	0.254
	$\hat{\alpha}_5$	0.047	0.122	-0.150	0.044	-0.176	0.055	0.257	0.112	0.271	0.130

Table A.8: ABC-AR Bias and MSE comparisons for $A_2, n = 100$ and new S_5.

		$\boldsymbol{\alpha} = (1,1,2,6,1)'$, $N = 200, n = 100$									
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	–		37,178(10,215)		56,157(15,393)		30,860(8,389)		45,335(12,043)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.012	0.252	-0.218	0.050	-0.347	0.127	0.464	0.223	0.201	0.063
	$\hat{\alpha}_2$	0.187	1.160	0.409	0.170	0.523	0.275	1.820	3.322	1.970	3.887
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.052	0.144	-0.960	0.925	-0.914	0.842	-0.180	0.046	-0.137	0.043
	$\hat{\alpha}_4$	0.142	1.136	-3.315	11.002	-3.509	12.317	-0.838	0.739	-1.282	1.652
	$\hat{\alpha}_5$	0.027	0.181	-0.071	0.009	-0.100	0.020	0.770	0.607	0.680	0.500
	Proposals	–		466,171(191,619)		800,330(277,493)		282,577(119,264)		420,149(132,200)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.012	0.252	-0.328	0.113	-0.521	0.281	0.170	0.045	-0.204	0.068
	$\hat{\alpha}_2$	0.187	1.160	0.321	0.110	0.552	0.318	1.519	2.336	1.813	3.328
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.052	0.144	-0.730	0.545	-0.699	0.501	0.068	0.039	0.060	0.040
	$\hat{\alpha}_4$	0.142	1.136	-2.727	7.458	-3.105	9.646	-0.176	0.095	-0.976	0.968
	$\hat{\alpha}_5$	0.027	0.181	-0.161	0.034	-0.220	0.063	0.508	0.284	0.348	0.165
	Proposals	–		7,918,996(3,124,887)		13,432,403(2,305,505)		2,931,482(1,668,552)		4,136,395(1,419,785)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.012	0.252	-0.324	0.118	-0.504	0.272	0.032	0.030	-0.324	0.136
	$\hat{\alpha}_2$	0.187	1.160	0.352	0.143	0.664	0.463	1.344	1.867	1.783	3.269
$\epsilon = 0.2$	$\hat{\alpha}_3$	0.052	0.144	-0.507	0.277	-0.480	0.249	0.210	0.093	0.172	0.075
	$\hat{\alpha}_4$	0.142	1.136	-2.105	4.463	-2.497	6.255	0.131	0.109	-0.768	0.628
	$\hat{\alpha}_5$	0.027	0.181	-0.139	0.036	-0.173	0.052	0.385	0.190	0.261	0.121

Table A.9: ABC-AR Bias and MSE comparisons for $A_3, n = 100$ and new S_5.

	$\boldsymbol{\alpha} = (1,1,1,1,1)'$, $N = 200, n = 50$										
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	–		7,204(1,112)		7,950(1,261)		6,774(1,511)		7,796(1,823)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-0.001	0.512	0.307	0.108	0.435	0.205	1.564	2.513	1.825	3.412
	$\hat{\alpha}_2$	-0.147	0.413	0.298	0.099	0.430	0.198	1.544	2.427	1.817	3.362
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.171	0.285	0.381	0.167	0.544	0.319	1.656	2.845	2.003	4.131
	$\hat{\alpha}_4$	0.231	0.292	0.368	0.163	0.526	0.308	1.626	2.777	1.961	4.011
	$\hat{\alpha}_5$	-0.032	0.450	0.319	0.124	0.489	0.266	1.581	2.607	1.931	3.859
	Proposals	–		66,424(30,483)		84,445(53,382)		331,499(838,688)		670,047(1,476,491)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-0.001	0.512	0.280	0.115	0.425	0.226	1.146	1.511	1.359	2.161
	$\hat{\alpha}_2$	-0.147	0.413	0.267	0.099	0.421	0.216	1.120	1.422	1.346	2.107
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.171	0.285	0.301	0.129	0.466	0.266	1.060	1.310	1.271	1.957
	$\hat{\alpha}_4$	0.231	0.292	0.288	0.130	0.446	0.265	1.037	1.286	1.236	1.922
	$\hat{\alpha}_5$	-0.032	0.450	0.277	0.119	0.446	0.254	1.124	1.479	1.366	2.234
	Proposals	–		1,489,410(1,343,551)		2,873,275(2,931,269)		10,561,131(4,962,914)		13,519,876(3,164,221)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-0.001	0.512	0.179	0.092	0.274	0.166	0.497	0.384	0.511	0.522
	$\hat{\alpha}_2$	-0.147	0.413	0.148	0.069	0.245	0.135	0.451	0.314	0.426	0.389
$\epsilon = 0.2$	$\hat{\alpha}_3$	0.171	0.285	0.151	0.067	0.233	0.120	0.349	0.205	0.306	0.214
	$\hat{\alpha}_4$	0.231	0.292	0.148	0.073	0.229	0.130	0.348	0.211	0.324	0.231
	$\hat{\alpha}_5$	-0.032	0.450	0.162	0.077	0.267	0.141	0.455	0.317	0.454	0.366

Table A.10: ABC-AR Bias and MSE comparisons for $A_1, n = 50$ and new S_5.

		$\boldsymbol{\alpha} = (3, 2.5, 2, 1.5, 1)'$, $N = 200, n = 50$									
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	–		15,741(4,358)		20,116(6,485)		13,100(3,747)		16,980(5,258)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.178	1.846	-1.197	1.445	-1.193	1.429	0.581	0.376	0.529	0.300
	$\hat{\alpha}_2$	0.271	1.221	-0.824	0.691	-0.748	0.575	0.746	0.596	0.915	0.880
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.117	0.466	-0.548	0.319	-0.366	0.154	0.730	0.601	1.085	1.265
	$\hat{\alpha}_4$	0.061	0.301	-0.393	0.171	-0.257	0.095	0.542	0.352	0.765	0.707
	$\hat{\alpha}_5$	0.095	0.278	-0.203	0.050	-0.224	0.068	0.459	0.239	0.399	0.221
	Proposals	–		161,614(67,261)		214,361(118,233)		101,273(35,005)		149,624(58,793)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.178	1.846	-1.003	1.028	-1.028	1.069	0.706	0.565	0.633	0.426
	$\hat{\alpha}_2$	0.271	1.221	-0.616	0.403	-0.556	0.336	0.881	0.834	1.113	1.302
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.117	0.466	-0.450	0.236	-0.313	0.133	0.657	0.538	0.970	1.101
	$\hat{\alpha}_4$	0.061	0.301	-0.360	0.159	-0.268	0.121	0.416	0.262	0.562	0.502
	$\hat{\alpha}_5$	0.095	0.278	-0.207	0.063	-0.230	0.083	0.335	0.168	0.306	0.185
	Proposals	–		2,692,399(1,596,205)		3,281,541(2,276,044)		1,173,015(516,978)		1,932,295(947,567)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.178	1.846	-0.798	0.687	-0.852	0.754	0.621	0.521	0.538	0.388
	$\hat{\alpha}_2$	0.271	1.221	-0.390	0.202	-0.392	0.199	0.807	0.785	1.072	1.268
$\epsilon = 0.2$	$\hat{\alpha}_3$	0.117	0.466	-0.289	0.133	-0.206	0.083	0.529	0.406	0.793	0.808
	$\hat{\alpha}_4$	0.061	0.301	-0.264	0.112	-0.201	0.097	0.308	0.194	0.394	0.340
	$\hat{\alpha}_5$	0.095	0.278	-0.158	0.056	-0.195	0.076	0.275	0.148	0.269	0.169

Table A.11: ABC-AR Bias and MSE comparisons for $A_2, n = 50$ and new S_5.

	$\boldsymbol{\alpha} = (1,1,2,6,1)'$, $N = 200, n = 50$										
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	–		39,487(15,158)		58,300(23,082)		33,280(13,799)		47,682(19,257)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.041	0.483	-0.205	0.046	-0.315	0.110	0.483	0.249	0.247	0.100
	$\hat{\alpha}_2$	0.280	2.218	0.418	0.178	0.524	0.278	1.834	3.378	1.975	3.915
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.114	0.247	-0.968	0.946	-0.926	0.871	-0.187	0.062	-0.154	0.069
	$\hat{\alpha}_4$	0.365	2.103	-3.360	11.310	-3.542	12.552	-0.885	0.854	-1.314	1.741
	$\hat{\alpha}_5$	0.027	0.277	-0.051	0.011	-0.069	0.026	0.802	0.676	0.730	0.616
	Proposals	–		502,221(278,106)		822,906(401,474)		317,519(202,732)		450,929(216,350)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.041	0.483	-0.294	0.096	-0.466	0.237	0.226	0.079	-0.121	0.069
	$\hat{\alpha}_2$	0.280	2.218	0.355	0.139	0.550	0.322	1.580	2.547	1.836	3.441
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.114	0.247	-0.770	0.614	-0.738	0.568	0.014	0.060	0.012	0.072
	$\hat{\alpha}_4$	0.365	2.103	-2.826	8.020	-3.168	10.049	-0.291	0.193	-1.025	1.076
	$\hat{\alpha}_5$	0.027	0.277	-0.120	0.033	-0.172	0.063	0.577	0.394	0.429	0.291
	Proposals	–		8,327,213(3,846,393)		12,688,679(3,139,416)		3,504,625(2,898,214)		4,697,328(2,551,502)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.041	0.483	-0.281	0.100	-0.453	0.239	0.117	0.064	-0.226	0.120
	$\hat{\alpha}_2$	0.280	2.218	0.372	0.168	0.633	0.440	1.431	2.155	1.789	3.364
$\epsilon = 0.2$	$\hat{\alpha}_3$	0.114	0.247	-0.572	0.362	-0.548	0.335	0.126	0.103	0.107	0.096
	$\hat{\alpha}_4$	0.365	2.103	-2.259	5.158	-2.647	7.040	-0.022	0.170	-0.833	0.762
	$\hat{\alpha}_5$	0.027	0.277	-0.094	0.043	-0.132	0.067	0.478	0.322	0.345	0.247

Table A.12: ABC-AR Bias and MSE comparisons for $A_3, n = 50$ and new S_5.

$\boldsymbol{\alpha} = (1,1,1,1,1)'$, $N = 200, n = 100$											
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Acceptances	–		3,517(2,213)		3,370(2,077)		6,777(6,507)		6,637(6,188)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.008	0.291	0.244	0.070	0.355	0.135	1.023	1.129	1.282	1.745
	$\hat{\alpha}_2$	-0.095	0.292	0.247	0.074	0.350	0.135	1.014	1.131	1.278	1.758
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.081	0.156	0.215	0.082	0.305	0.158	0.714	0.721	0.722	0.918
	$\hat{\alpha}_4$	0.139	0.214	0.219	0.088	0.310	0.166	0.719	0.743	0.729	0.955
	$\hat{\alpha}_5$	-0.014	0.233	0.432	0.200	0.630	0.411	1.762	3.270	2.133	4.781
	Acceptances	–		1,520(1,100)		1,441(1,010)		2,134(2,268)		2,152(2,123)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.008	0.291	0.244	0.074	0.382	0.157	0.910	0.923	1.248	1.683
	$\hat{\alpha}_2$	-0.095	0.292	0.249	0.074	0.373	0.159	0.915	0.949	1.222	1.655
$\epsilon = 0.6$	$\hat{\alpha}_3$	0.081	0.156	0.148	0.058	0.204	0.118	0.405	0.308	0.282	0.321
	$\hat{\alpha}_4$	0.139	0.214	0.154	0.065	0.212	0.122	0.408	0.319	0.299	0.345
	$\hat{\alpha}_5$	-0.014	0.233	0.379	0.158	0.582	0.352	1.354	1.978	1.719	3.164
	Acceptances	–		403(323)		389(298)		437(444)		467(441)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.008	0.291	0.236	0.075	0.393	0.176	0.770	0.685	1.078	1.360
	$\hat{\alpha}_2$	-0.095	0.292	0.237	0.075	0.396	0.184	0.769	0.694	1.085	1.351
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.081	0.156	0.082	0.037	0.087	0.074	0.155	0.087	-0.021	0.072
	$\hat{\alpha}_4$	0.139	0.214	0.091	0.045	0.096	0.082	0.162	0.094	-0.012	0.083
	$\hat{\alpha}_5$	-0.014	0.233	0.318	0.115	0.508	0.276	0.948	0.986	1.264	1.759

Table A.13: ABC-MH Bias and MSE comparisons for $A_1, n = 100$ and old S_5.

$\boldsymbol{\alpha} = (3, 2.5, 2, 1.5, 1)'$, $N = 200, n = 100$											
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Acceptances	–		13,088(471)		11,016(494)		50,140(1,763)		46,104(1,852)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.174	0.833	-1.068	1.145	-1.220	1.490	0.655	0.442	0.472	0.226
	$\hat{\alpha}_2$	0.075	0.598	-0.692	0.484	-0.819	0.674	0.812	0.679	0.776	0.611
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.032	0.240	-0.332	0.116	-0.160	0.029	0.911	0.850	1.281	1.655
	$\hat{\alpha}_4$	0.040	0.178	-0.226	0.054	-0.047	0.007	0.677	0.469	0.981	0.981
	$\hat{\alpha}_5$	0.047	0.122	-0.326	0.108	-0.455	0.211	0.379	0.151	0.212	0.062
	Acceptances	–		7,915(341)		6,562(356)		31,222(2,261)		28,797(2,133)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.174	0.833	-1.022	1.048	-1.191	1.420	0.706	0.514	0.500	0.254
	$\hat{\alpha}_2$	0.075	0.598	-0.650	0.428	-0.812	0.662	0.847	0.742	0.799	0.649
$\epsilon = 0.6$	$\hat{\alpha}_3$	0.032	0.240	-0.268	0.079	-0.100	0.014	0.892	0.822	1.303	1.719
	$\hat{\alpha}_4$	0.040	0.178	-0.195	0.042	-0.015	0.006	0.618	0.398	0.943	0.921
	$\hat{\alpha}_5$	0.047	0.122	-0.334	0.113	-0.477	0.232	0.350	0.130	0.165	0.046
	Acceptances	–		3,568(282)		2,844(232)		13,412(2,530)		12,236(2,379)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.174	0.833	-0.951	0.911	-1.137	1.296	0.740	0.571	0.520	0.278
	$\hat{\alpha}_2$	0.075	0.598	-0.581	0.347	-0.775	0.605	0.861	0.778	0.856	0.745
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.032	0.240	-0.165	0.040	-0.023	0.008	0.779	0.663	1.182	1.484
	$\hat{\alpha}_4$	0.040	0.178	-0.132	0.026	0.028	0.012	0.496	0.284	0.778	0.703
	$\hat{\alpha}_5$	0.047	0.122	-0.332	0.113	-0.484	0.242	0.338	0.123	0.170	0.060

Table A.14: ABC-MH Bias and MSE comparisons for $A_2, n = 100$ and old S_5.

$\boldsymbol{\alpha} = (1,1,2,6,1)'$, $N = 200, n = 100$											
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Acceptances	–		6,186(717)		4,648(607)		27,210(3,053)		22,630(2,751)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.012	0.252	0.078	0.009	0.093	0.011	0.810	0.666	1.000	1.016
	$\hat{\alpha}_2$	0.187	1.160	0.679	0.462	0.771	0.595	2.254	5.086	2.369	5.612
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.052	0.144	-0.961	0.926	-1.009	1.022	-0.452	0.213	-0.644	0.425
	$\hat{\alpha}_4$	0.142	1.136	-3.137	9.854	-3.360	11.293	-1.175	1.413	-1.469	2.168
	$\hat{\alpha}_5$	0.027	0.181	-0.038	0.005	-0.005	0.007	1.003	1.021	1.155	1.360
	Acceptances	–		3,536(431)		2,646(365)		16,633(2,016)		13,731(1,756)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.012	0.252	0.037	0.004	-0.009	0.004	0.772	0.607	0.941	0.901
	$\hat{\alpha}_2$	0.187	1.160	0.660	0.437	0.779	0.609	2.261	5.114	2.394	5.732
$\epsilon = 0.6$	$\hat{\alpha}_3$	0.052	0.144	-0.887	0.792	-0.912	0.839	-0.456	0.218	-0.659	0.446
	$\hat{\alpha}_4$	0.142	1.136	-2.855	8.167	-3.118	9.729	-0.956	0.964	-1.303	1.716
	$\hat{\alpha}_5$	0.027	0.181	-0.026	0.005	-0.021	0.011	1.007	1.030	1.140	1.327
	Acceptances	–		1,535(202)		1,156(173)		7,762(1,051)		6,314(888)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.012	0.252	-0.022	0.003	-0.144	0.029	0.746	0.568	0.854	0.747
	$\hat{\alpha}_2$	0.187	1.160	0.617	0.384	0.745	0.564	2.238	5.014	2.421	5.865
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.052	0.144	-0.748	0.569	-0.740	0.560	-0.402	0.177	-0.592	0.366
	$\hat{\alpha}_4$	0.142	1.136	-2.442	5.997	-2.734	7.496	-0.615	0.463	-1.031	1.105
	$\hat{\alpha}_5$	0.027	0.181	-0.012	0.007	-0.050	0.018	1.024	1.065	1.128	1.301

Table A.15: ABC-MH Bias and MSE comparisons for $A_3, n = 100$ and old S_5.

$\boldsymbol{\alpha} = (1,1,1,1,1)'$, $N = 200, n = 50$											
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Acceptances	–		4,030(2,600)		3,823(2,410)		9,145(8,382)		8,881(8,037)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-0.001	0.512	0.262	0.086	0.370	0.156	1.112	1.370	1.362	2.002
	$\hat{\alpha}_2$	-0.147	0.413	0.254	0.080	0.365	0.147	1.105	1.337	1.331	1.925
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.171	0.285	0.259	0.117	0.356	0.205	0.903	1.076	0.967	1.429
	$\hat{\alpha}_4$	0.231	0.292	0.250	0.115	0.343	0.202	0.879	1.043	0.948	1.389
	$\hat{\alpha}_5$	-0.032	0.450	0.405	0.185	0.597	0.379	1.786	3.351	2.138	4.800
	Acceptances	–		1,778(1,318)		1,675(1,219)		3,004(3,286)		2,938(3,082)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-0.001	0.512	0.258	0.086	0.385	0.168	1.039	1.216	1.338	1.962
	$\hat{\alpha}_2$	-0.147	0.413	0.254	0.082	0.383	0.164	1.018	1.174	1.328	1.935
$\epsilon = 0.6$	$\hat{\alpha}_3$	0.171	0.285	0.205	0.095	0.284	0.173	0.599	0.566	0.554	0.676
	$\hat{\alpha}_4$	0.231	0.292	0.197	0.094	0.269	0.164	0.589	0.546	0.537	0.654
	$\hat{\alpha}_5$	-0.032	0.450	0.369	0.155	0.553	0.325	1.438	2.227	1.809	3.498
	Acceptances	–		472(402)		445(364)		565(631)		578(597)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-0.001	0.512	0.249	0.099	0.406	0.204	0.882	0.964	1.192	1.675
	$\hat{\alpha}_2$	-0.147	0.413	0.258	0.089	0.404	0.191	0.870	0.893	1.182	1.638
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.171	0.285	0.143	0.069	0.178	0.118	0.330	0.209	0.173	0.192
	$\hat{\alpha}_4$	0.231	0.292	0.137	0.068	0.169	0.116	0.320	0.205	0.157	0.176
	$\hat{\alpha}_5$	-0.032	0.450	0.320	0.123	0.498	0.274	1.037	1.197	1.352	2.051

Table A.16: ABC-MH Bias and MSE comparisons for $A_1, n = 50$ and old S_5.

$\boldsymbol{\alpha} = (3, 2.5, 2, 1.5, 1)'$, $N = 200, n = 50$											
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Acceptances	–		12,503(658)		10,672(683)		48,821(2,695)		44,972(2,783)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.178	1.846	-1.097	1.209	-1.231	1.516	0.633	0.424	0.463	0.220
	$\hat{\alpha}_2$	0.271	1.221	-0.716	0.522	-0.822	0.680	0.809	0.686	0.781	0.624
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.117	0.466	-0.362	0.140	-0.188	0.041	0.891	0.826	1.257	1.603
	$\hat{\alpha}_4$	0.061	0.301	-0.233	0.061	-0.057	0.011	0.687	0.494	0.988	1.011
	$\hat{\alpha}_5$	0.095	0.278	-0.311	0.099	-0.432	0.193	0.392	0.167	0.232	0.085
	Acceptances	–		7,453(505)		6,269(482)		29,998(3,350)		27,698(3,203)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.178	1.846	-1.051	1.112	-1.202	1.447	0.686	0.499	0.492	0.249
	$\hat{\alpha}_2$	0.271	1.221	-0.674	0.465	-0.810	0.663	0.842	0.750	0.809	0.673
$\epsilon = 0.6$	$\hat{\alpha}_3$	0.117	0.466	-0.306	0.105	-0.131	0.024	0.863	0.788	1.262	1.632
	$\hat{\alpha}_4$	0.061	0.301	-0.206	0.052	-0.025	0.011	0.624	0.420	0.937	0.939
	$\hat{\alpha}_5$	0.095	0.278	-0.316	0.103	-0.453	0.214	0.363	0.147	0.190	0.073
	Acceptances	–		3,246(405)		2,654(341)		12,647(3,162)		11,577(2,928)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.178	1.846	-0.977	0.965	-1.141	1.305	0.721	0.558	0.510	0.275
	$\hat{\alpha}_2$	0.271	1.221	-0.607	0.384	-0.759	0.584	0.849	0.786	0.852	0.750
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.117	0.466	-0.222	0.070	-0.071	0.019	0.736	0.639	1.119	1.399
	$\hat{\alpha}_4$	0.061	0.301	-0.155	0.040	0.006	0.022	0.496	0.312	0.764	0.743
	$\hat{\alpha}_5$	0.095	0.278	-0.310	0.101	-0.449	0.215	0.351	0.139	0.192	0.088

Table A.17: ABC-MH Bias and MSE comparisons for $A_2, n = 50$ and old S_5.

	$\boldsymbol{\alpha} = (1,1,2,6,1)'$, $N = 200, n = 50$										
		MMLE		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Acceptances	–		5,921(938)		4,464(781)		26,666(3,785)		22,147(3,446)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.041	0.483	0.061	0.008	0.080	0.012	0.813	0.678	0.992	1.010
	$\hat{\alpha}_2$	0.280	2.218	0.667	0.447	0.746	0.558	2.257	5.099	2.360	5.571
$\epsilon = 0.8$	$\hat{\alpha}_3$	0.114	0.247	-0.982	0.968	-1.027	1.060	-0.444	0.210	-0.627	0.412
	$\hat{\alpha}_4$	0.365	2.103	-3.230	10.451	-3.432	11.787	-1.201	1.499	-1.494	2.247
	$\hat{\alpha}_5$	0.027	0.277	-0.024	0.008	0.013	0.014	1.012	1.056	1.159	1.394
	Acceptances	–		3,346(539)		2,464(449)		16,144(2,417)		13,272(2,193)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.041	0.483	0.024	0.005	-0.005	0.008	0.776	0.620	0.929	0.891
	$\hat{\alpha}_2$	0.280	2.218	0.649	0.423	0.759	0.579	2.255	5.090	2.378	5.657
$\epsilon = 0.6$	$\hat{\alpha}_3$	0.114	0.247	-0.917	0.846	-0.951	0.912	-0.450	0.216	-0.643	0.430
	$\hat{\alpha}_4$	0.365	2.103	-2.971	8.857	-3.231	10.451	-0.985	1.043	-1.333	1.798
	$\hat{\alpha}_5$	0.027	0.277	-0.010	0.009	0.009	0.018	1.017	1.069	1.145	1.364
	Acceptances	–		1,397(238)		1,025(195)		7,387(1,250)		5,987(1,085)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	0.041	0.483	-0.030	0.006	-0.130	0.030	0.741	0.568	0.845	0.748
	$\hat{\alpha}_2$	0.280	2.218	0.618	0.387	0.741	0.560	2.221	4.941	2.401	5.773
$\epsilon = 0.4$	$\hat{\alpha}_3$	0.114	0.247	-0.798	0.647	-0.798	0.652	-0.397	0.176	-0.583	0.361
	$\hat{\alpha}_4$	0.365	2.103	-2.611	6.865	-2.908	8.489	-0.657	0.557	-1.074	1.201
	$\hat{\alpha}_5$	0.027	0.277	0.007	0.012	-0.014	0.027	1.027	1.092	1.136	1.352

Table A.18: ABC-MH Bias and MSE comparisons for $A_3, n = 50$ and old S_5.

$\boldsymbol{\alpha} = (2,1,1,2,4,6,2,1)',\ N = 200, n = 100$									
		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	88,062(84,908)		69,023(53,108)		73,257(69,436)		57,608(42,752)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-0.696	0.487	-0.721	0.523	0.580	0.346	0.532	0.297
	$\hat{\alpha}_2$	-0.039	0.005	-0.049	0.010	0.879	0.786	0.856	0.764
	$\hat{\alpha}_3$	0.186	0.039	0.313	0.103	1.347	1.831	1.590	2.549
$\epsilon = 0.8$	$\hat{\alpha}_4$	-0.472	0.225	-0.450	0.204	1.059	1.132	1.102	1.222
	$\hat{\alpha}_5$	-2.287	5.262	-2.268	5.159	-0.733	0.649	-0.675	0.515
	$\hat{\alpha}_6$	-3.672	13.523	-3.814	14.557	-1.462	2.286	-1.717	2.985
	$\hat{\alpha}_7$	-0.981	0.975	-0.984	0.999	0.012	0.051	0.009	0.122
	$\hat{\alpha}_8$	-0.378	0.149	-0.491	0.254	0.207	0.068	-0.022	0.050
	Proposals	11,893,421(3,788,548)		11,555,679(4,002,650)		6,519,920(4,768,656)		5,124,180(3,841,150)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-0.534	0.291	-0.612	0.378	0.637	0.421	0.512	0.280
	$\hat{\alpha}_2$	-0.049	0.014	-0.034	0.017	0.664	0.473	0.620	0.450
	$\hat{\alpha}_3$	0.204	0.055	0.431	0.202	1.164	1.397	1.562	2.496
$\epsilon = 0.2$	$\hat{\alpha}_4$	-0.287	0.087	-0.274	0.078	1.184	1.417	1.194	1.433
	$\hat{\alpha}_5$	-1.832	3.419	-1.914	3.694	-0.476	0.419	-0.507	0.339
	$\hat{\alpha}_6$	-2.966	8.873	-3.280	10.786	-1.012	1.247	-1.449	2.143
	$\hat{\alpha}_7$	-0.799	0.672	-0.822	0.741	-0.037	0.080	-0.109	0.216
	$\hat{\alpha}_8$	-0.395	0.170	-0.522	0.293	-0.012	0.032	-0.242	0.106

Table A.19: ABC-AR Bias and MSE comparisons for $A_4, n = 100$.

$\boldsymbol{\alpha} = (3.5, 2, 1.5, 4, 1, 2.5, 3, 4.5)'$, $N = 200, n = 100$									
		$\mathcal{G}1$		$\mathcal{U}1$		$\mathcal{G}2$		$\mathcal{U}2$	
	Proposals	18,098(8,861)		17,639(7,187)		14,997(7,165)		14,867(5,797)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-2.272	5.165	-2.192	4.807	-1.092	1.202	-0.921	0.857
	$\hat{\alpha}_2$	-0.825	0.682	-0.759	0.579	0.299	0.098	0.434	0.199
	$\hat{\alpha}_3$	-0.105	0.013	-0.084	0.009	1.267	1.613	1.310	1.724
$\epsilon = 0.8$	$\hat{\alpha}_4$	-2.553	6.521	-2.526	6.385	-1.130	1.282	-1.078	1.168
	$\hat{\alpha}_5$	-0.153	0.034	-0.151	0.049	0.639	0.448	0.637	0.503
	$\hat{\alpha}_6$	-1.331	1.788	-1.225	1.532	-0.181	0.098	0.031	0.117
	$\hat{\alpha}_7$	-1.238	1.564	-1.166	1.373	0.427	0.298	0.591	0.405
	$\hat{\alpha}_8$	-2.865	8.241	-2.778	7.732	-1.340	1.901	-1.140	1.359
	Proposals	4,197,067(2,437,381)		3,593,032(2,444,162)		1,170,664(738,735)		1,090,612(568,018)	
	Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
	$\hat{\alpha}_1$	-2.171	4.719	-2.071	4.292	-1.133	1.303	-0.894	0.814
	$\hat{\alpha}_2$	-0.727	0.534	-0.670	0.451	0.247	0.079	0.417	0.190
	$\hat{\alpha}_3$	0.066	0.007	0.031	0.003	1.336	1.795	1.330	1.778
$\epsilon = 0.2$	$\hat{\alpha}_4$	-2.370	5.622	-2.383	5.679	-1.032	1.074	-1.006	1.018
	$\hat{\alpha}_5$	-0.054	0.029	-0.086	0.057	0.510	0.320	0.466	0.358
	$\hat{\alpha}_6$	-1.059	1.164	-0.985	1.024	-0.167	0.126	0.060	0.208
	$\hat{\alpha}_7$	-0.731	0.595	-0.795	0.659	0.656	0.602	0.789	0.687
	$\hat{\alpha}_8$	-2.421	5.915	-2.447	6.015	-1.160	1.495	-0.990	1.048

Table A.20: ABC-AR Bias and MSE comparisons for $A_5, n = 100$.

Bibliography

Arnold, B. C. and Ng, H. K. T. (2011). Flexible bivariate beta distributions. *Journal of Multivariate Analysis*, 102(8):1194–1202.

Balakrishnan, N. and Lai, C.-D. (2009). *Continuous bivariate distributions*. Springer.

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika*, 96(4):983–990.

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035.

Blum, M. G. and François, O. (2010). Non-linear regression models for approximate Bayesian computation. *Statistics and Computing*, 20(1):63–73.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. Taylor & Francis US.

Burr, T. and Skurikhin, A. (2013). Selecting summary statistics in approximate bayesian computation for calibrating stochastic models. *BioMed research international*, 2013.

Cappé, O., Guillin, A., Marin, J.-M., and Robert, C. P. (2004). Population Monte Carlo. *Journal of Computational and Graphical Statistics*, 13(4).

Crackel, R. C. and Flegal, J. M. (2014). Approximate bayesian computation for a flexible class of bivariate beta distributions. *arXiv preprint arXiv:1402.1782*.

Danaher, P. J. and Hardie, B. G. S. (2005). Bacon with your eggs? applications of a new bivariate beta-binomial distribution. *The American Statistician*, 59(4):282–286.

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.

Del Moral, P., Doucet, A., and Jasra, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing*, 22(5):1009–1020.

Drovandi, C. C. and Pettitt, A. N. (2011). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics*, 67(1):225–233.

Fearnhead, P. and Prangle, D. (2012). Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474.

Gupta, A. and Wong, C. (1985). On three and five parameter bivariate beta distributions. *Metrika*, 32(1):85–91.

Gupta, A. K., Orozco-Castañeda, J. M., and Nagar, D. K. (2011). Non-central bivariate beta distribution. *Statistical papers*, 52(1):139–152.

Jasra, A., Singh, S. S., Martin, J. S., and McCoy, E. (2012). Filtering via approximate Bayesian computation. *Statistics and Computing*, 22(6):1223–1237.

Jones, M. (2002). Multivariate t and beta distributions associated with the multivariate F distribution. *Metrika*, 54(3):215–231.

Joyce, P. and Marjoram, P. (2008). Approximately sufficient statistics and bayesian computation. *Statistical Applications in Genetics and Molecular Biology*, 7(1).

Leuenberger, C. and Wegmann, D. (2010). Bayesian computation and model selection without likelihoods. *Genetics*, 184(1):243–252.

Liu, J. S. (2001). Monte carlo strategies in scientific computing. *NY: Springer*.

Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328.

Mendenhall, W., Beaver, R., and Beaver, B. (2012). *Introduction to Probability and Statistics*. Textbooks Available with Cengage Youbook. Cengage Learning.

Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitt. Math. Statist*, 8(1):234–235.

Nadarajah, S. and Kotz, S. (2005). Some bivariate beta distributions. *Statistics*, 39(5):457–466.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

Olkin, I. and Liu, R. (2003). A bivariate beta distribution. *Statistics & Probability Letters*, 62(4):407–412.

Olkin, I. and Trikalinos, T. A. (2015). Constructions for a bivariate beta distribution. *Statistics & Probability Letters*, 96:54–60.

Peters, G. W., Fan, Y., and Sisson, S. A. (2012). On sequential Monte Carlo, partial rejection control and approximate Bayesian computation. *Statistics and Computing*, 22(6):1209–1222.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798.

Ratmann, O., Andrieu, C., Wiuf, C., and Richardson, S. (2009). Model criticism based on likelihood-free inference, with an application to protein network evolution. *Proceedings of the National Academy of Sciences*, 106(26):10576–10581.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, 12(4):1151–1172.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392.

Sarmanov, O. V. (1966). Generalized normal correlation and two-dimensional Frechet classes. *Doklady(Soviet Mathematics)*, 168:596–599.

Schucany, W. R., Parr, W. C., and Boyer, J. E. (1978). Correlation structure in Farlie-Gumbel-Morgenstern distributions. *Biometrika*, 65(3):650–653.

Sisson, S., Fan, Y., and Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765.

Sisson, S., Fan, Y., and Tanaka, M. M. (2009). Correction: Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 106(39):16889–16890.

Tavare, S., Balding, D. J., Griffiths, R., and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, 145(2):505–518.

Ting Lee, M.-L. (1996). Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics-Theory and Methods*, 25(6):1207–1222.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372.

Wilkinson, R. D. (2013). Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. *Statistical applications in genetics and molecular biology*, 12(2):129–141.