Theoretical Population Biology 99 (2015) 31–42

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

AABC: Approximate approximate Bayesian computation for inference in population-genetic models

Erkan O. Buzbas [a,b,*], Noah A. Rosenberg [a]

[a] Department of Biology, Stanford University, Stanford, CA 94305-5020, USA
[b] Department of Statistical Science, University of Idaho, Moscow, ID 84844-1104, USA

ARTICLE INFO

Article history:
Received 13 April 2014
Available online 26 September 2014

Keywords:
Approximate Bayesian computation
Likelihood-free methods
Population genetics
Posterior distribution

ABSTRACT

Approximate Bayesian computation (ABC) methods perform inference on model-specific parameters of mechanistically motivated parametric models when evaluating likelihoods is difficult. Central to the success of ABC methods, which have been used frequently in biology, is computationally inexpensive simulation of data sets from the parametric model of interest. However, when simulating data sets from a model is so computationally expensive that the posterior distribution of parameters cannot be adequately sampled by ABC, inference is not straightforward. We present "approximate approximate Bayesian computation" (AABC), a class of computationally fast inference methods that extends ABC to models in which simulating data is expensive. In AABC, we first simulate a number of data sets small enough to be computationally feasible to simulate from the parametric model. Conditional on these data sets, we use a statistical model that approximates the correct parametric model and enables efficient simulation of a large number of data sets. We show that under mild assumptions, the posterior distribution obtained by AABC converges to the posterior distribution obtained by ABC, as the number of data sets simulated from the parametric model and the sample size of the observed data set increase. We demonstrate the performance of AABC on a population-genetic model of natural selection, as well as on a model of the admixture history of hybrid populations. This latter example illustrates how, in population genetics, AABC is of particular utility in scenarios that rely on conceptually straightforward but potentially slow forward-in-time simulations.

1. Introduction

Stochastic processes motivated by mechanistic considerations enable investigators to capture salient phenomena in modeling biological systems. Statistical models resulting from these stochastic processes are often parametric, and estimating model-specific parameters – which often have a biological interpretation – is a major aim of data analysis. Contemporary mechanistic models tend to involve complex stochastic processes, however, and parametric statistical models resulting from these processes lead to computationally intractable likelihood functions. When likelihood functions are computationally intractable, likelihood-based inference is a challenging problem that has received considerable attention in the literature (Robert and Casella, 2004; Liu, 2008).

Approximate Bayesian computation (ABC) methods (Beaumont et al., 2002; Marjoram et al., 2003) use data sets simulated from the model to assess parameter likelihoods without explicit evaluation of likelihood functions, and thereby facilitate sampling an *approximate* posterior distribution of the parameters. Intuitively, parameter values producing simulated data sets similar to the observed data set arise in approximate proportion to their likelihood, and hence, when weighted by prior probabilities, to their posterior probabilities.

1.1. The ABC literature

ABC methods have been based on rejection algorithms (Tavaré et al., 1997; Beaumont et al., 2002; Blum and François, 2010), Markov chain Monte Carlo (Beaumont, 2003; Marjoram et al., 2003; Wegmann et al., 2009), and sequential Monte Carlo (Sisson et al., 2007, 2009; Beaumont et al., 2009). Model selection using ABC (Pritchard et al., 1999; Fagundes et al., 2007; Grelaud et al., 2009; Blum and Jakobsson, 2010; Robert et al., 2011), the choice of summary statistics when the likelihood is based on summaries instead of the full data (Joyce and Marjoram, 2008; Nunes and Balding, 2010; Fearnhead and Prangle, 2012), and the equivalence of

* Correspondence to: Department of Statistical Science, 415 Brink Hall, PO Box 441104, Moscow, ID 83844-1104, USA.
E-mail addresses: erkanb@uidaho.edu (E.O. Buzbas), noahr@stanford.edu (N.A. Rosenberg).

http://dx.doi.org/10.1016/j.tpb.2014.09.002
0040-5809/

posterior distributions targeted in different ABC methods (Wilkinson, 2008; Sisson et al., 2010) have also been investigated.

ABC methods have been widely used for model-based inference in disciplines that rely on genetic data, particularly data shaped by diverse evolutionary, demographic, and environmental forces. Example applications have included problems in the demographic history of populations (Pritchard et al., 1999; François et al., 2008; Verdu et al., 2009; Blum and Jakobsson, 2010) and species (Estoup et al., 2004; Plagnol and Tavaré, 2004; Becquet and Przeworski, 2007; Fagundes et al., 2007; Wilkinson et al., 2010), as well as problems in the evolution of cancer cell lineages (Tavaré, 2005; Siegmund et al., 2008), the evolution of protein networks (Ratmann et al., 2009) and the study of dynamic molecular networks in systems biology (Bonassi et al., 2011).

1.2. A limitation of ABC methods and our contribution

Adequately sampling a posterior distribution of a parameter by ABC requires many random realizations from the sampling distribution of the data. However, the computational cost of simulating a data set increases quickly with the complexity and number of stochastic processes involved in a model. When only a small number of data sets can be simulated from the model, likelihoods cannot be accurately assessed using ABC, and hence, the posterior distribution of parameters cannot be adequately sampled.

In this article, we introduce *approximate* approximate Bayesian computation (AABC), a class of fast computational statistical methods that perform inference on model-specific parameters when standard ABC methods are computationally infeasible to apply. AABC methods overcome the computational intractability associated with simulating many data sets under the model by making approximations on the *parameter space* and the *model space*, in addition to standard ABC approximations on the data space (Fig. 1).

Our approach is to condition on a small number of data sets that can be feasibly simulated from the model and to employ a non-mechanistic statistical model to simulate a large number of data sets. The data values from the small number of simulated data sets are used to construct new random data sets, thereby rendering the simulation of a large number of data sets inexpensive in AABC. Intuitively, the information conditioned upon by the non-mechanistic model increases with the number of data sets simulated from the mechanistic model, and the expected accuracy of inference obtained by AABC methods increases. We formalize this intuition by showing that the posterior distribution of parameters obtained by AABC converges to the corresponding posterior distribution obtained by standard ABC, as the sample size of the observed data set and the number of data sets simulated from the model increase. Next, we briefly review a standard ABC-by-rejection algorithm.

2. A standard ABC algorithm by rejection sampling

To set up the class of problems in which ABC methods are useful, we assume that a parametric model generates (possibly multivariate) observations conditional on parameter $\theta \in \Theta \equiv \mathbf{R}^{\ell},\ \ell \geq 1$. We denote a random data set of n independent and identically distributed (IID) observations by $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}$, where $\mathcal{X}$ is the space in which the data set sits, and the observed data set by $\mathbf{x}_o$. In the population genetics context, a data point x_i might be a vector denoting the allelic types of a genetic locus at genomic position i in a group of individuals; the data matrix $\mathbf{x}$ might contain genotypes from these individuals in a sample of n genetic loci.

ABC methods make two approximations on the likelihood $p(\mathbf{x}_o|\theta)$ of the parameters given the observed data set. First, the observed data set $\mathbf{x}_o$ and any simulated data set $\mathbf{x}_i$ are substituted by $\mathbf{s}_o$ and $\mathbf{s}_i$, respectively. Second, the likelihood function of the data, $p(\mathbf{x}|\theta)$, is substituted with an approximate likelihood function $p(\|\mathbf{s}-\mathbf{s}_o\| < \epsilon|\theta)$, for an appropriate distance $\|\cdot\|$ and a tolerance parameter ϵ. A standard ABC algorithm by rejection sampling is as follows (Pritchard et al., 1999).

Algorithm 1: ABC by rejection sampling.

1. Simulate $\theta_i \sim \pi(\theta)$.
2. Simulate $\mathbf{x}_i \sim p(\mathbf{x}|\theta_i)$.
3. Calculate the summary statistics $\mathbf{s}_i$ from $\mathbf{x}_i$.
4. If $\|\mathbf{s}_o - \mathbf{s}_i\| < \epsilon$, output θ_i.

AABC methods utilize the established machinery of ABC methods in sampling the posterior distribution of the parameters. Therefore, standard approximations on the data space involved in an ABC method – features of the distance function, tolerance parameter, and weighting of simulated data that are sufficiently close to the observed data – apply to AABC methods as well. We assume that these standard ABC approximations work reasonably well, and we focus on introducing new modeling approximations on the parameter and model spaces (Table 1).

3. Approximate approximate Bayesian computation (AABC)

Algorithm 1 returns an adequate sample size from the posterior distribution of a parameter if it is iterated a large number of times, M. The set of realizations simulated from the joint distribution of the parameter and the data by steps 1 and 2 of Algorithm 1 is then $\{(\mathbf{x}_1, \theta_1), (\mathbf{x}_2, \theta_2), \ldots, (\mathbf{x}_M, \theta_M)\}$. Our interest in this article is inference when simulating M data sets from $p(\mathbf{x}|\theta)$ is computationally infeasible. We thus assume that only a small number m of data sets $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ can be obtained by step 2 of Algorithm 1 ($m \ll M$). We denote the set of realizations simulated from the joint distribution of the parameter and the data by $\mathcal{Z}_{n,m} = \{(\mathbf{x}_1, \theta_1), (\mathbf{x}_2, \theta_2), \ldots, (\mathbf{x}_m, \theta_m)\}$, where each data set $\mathbf{x}_i$ of n observations is simulated from the model $p(\mathbf{x}|\theta_i)$. See Table 2.

3.1. AABC algorithms

An AABC algorithm has three parts:

I. Simulating a limited number of realizations from the prior distribution of the parameter and the distribution of the data.
II. Simulating a new parameter value θ^* from its prior distribution and a data set from a statistical model $q(\mathbf{x}|\theta^*, \mathcal{Z}_{n,m})$.
III. Comparing the summary statistics $\mathbf{s}^*$ calculated from the simulated data set $\mathbf{x}^*$ with the summary statistics $\mathbf{s}_o$ calculated from the observed data set $\mathbf{x}_o$, to accept or reject the parameter value θ^*.

Part I involves the application of steps 1 and 2 from Algorithm 1 only for m iterations, and obtains the set $\mathcal{Z}_{n,m}$. The novelty of AABC is constructing the statistical model $q(\mathbf{x}|\theta^*, \mathcal{Z}_{n,m})$ used in Part II, and we describe the details of this model in the next section. The calculation and comparison of summary statistics in Part III follow steps 3 and 4 of Algorithm 1.

3.2. Approximations on the parameter and model spaces due to replacing mechanistic model $p(\mathbf{x}|\theta)$ with statistical model $q(\mathbf{x}|\theta, \mathcal{Z}_{n,m})$

We use a statistical model $q(\mathbf{x}|\theta, \mathcal{Z}_{n,m})$ as a surrogate for the mechanistic model $p(\mathbf{x}|\theta)$ to simulate new data sets. For a parameter value θ^* under which we want to simulate a new data set, we first calculate the Euclidean distances $\|\theta^*-\theta_i\|$ for all $\theta_i \in \mathcal{Z}_{n,m}$. We then assign weights ω_i to data sets $\mathbf{x}_i$ simulated under θ_i according to an Epanechnikov kernel: $\omega_i = (3/4)[1/(\theta^*-\theta_{(k+1)})][1-\|(\theta^*-\theta_i)/(\theta^*-\theta_{(k+1)})\|^2]\mathbf{I}_{\{\|\theta^*-\theta_i\|<\|\theta^*-\theta_{(k+1)}\|\}}$, where $\theta_{(k+1)}$ is the parameter value with the $(k+1)$th smallest distance to θ^*. Here, the ω_i

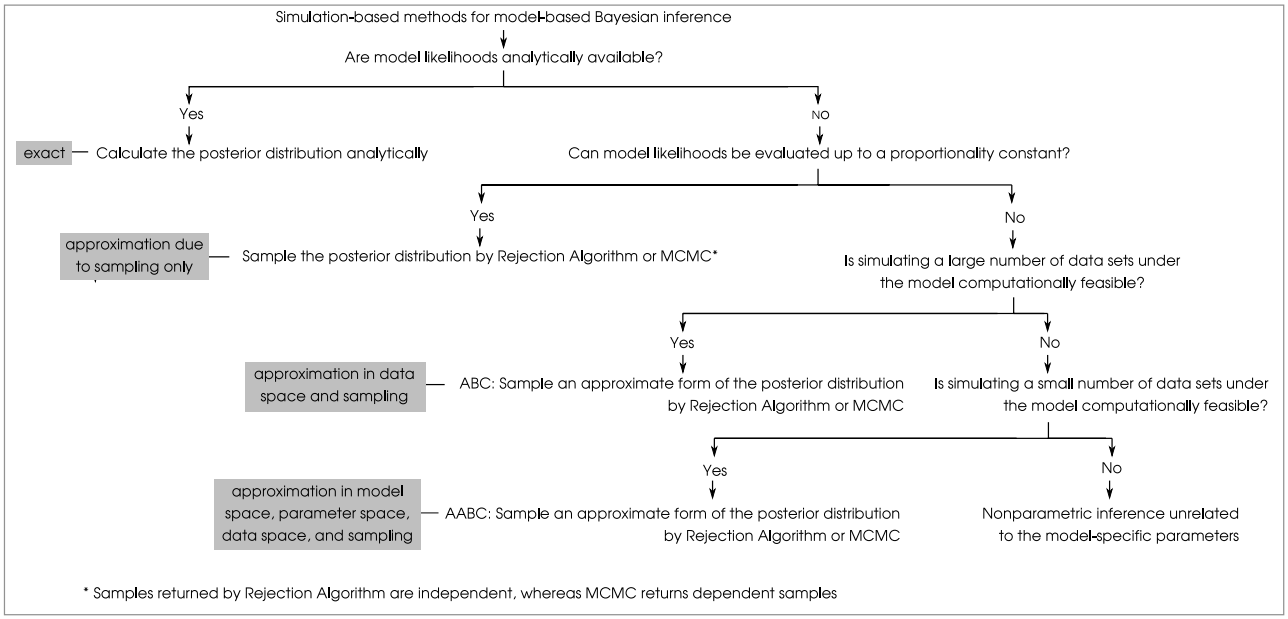

Fig. 1. Applicability of simulation-based inference methods in relation to the information available about the likelihood function.

Table 1
Approximations and errors involved in simulation-based ABC inference methods. Likelihood functions of the full data and the summary statistics are denoted respectively by $p(\mathbf{x}|\theta)$ and $p(\mathbf{s}|\theta)$. Exact ABC with full data involves only the Monte Carlo approximation due to sampling and thus is equivalent to a standard rejection algorithm. Summary statistics $\mathbf{s}$ are assumed not to be sufficient, so that dimension reduction from $\mathbf{x}_o$ to $\mathbf{s}_o$ results in an approximation.

True quantity	Approximated by	Space involved	Source of error	Method employing approximation			
				Exact ABC with full data	Exact ABC with summary statistics	ABC	AABC
$\int_{\mathcal{X}} p(\mathbf{x}\|\theta)d\mathbf{x}$	$\sum_{i=1}^{M} p(\mathbf{x}_i\|\theta)$	Data	Monte Carlo	Yes	Yes	Yes	Yes
$p(\mathbf{x}\|\theta)$	$p(\mathbf{s}\|\theta)$	Data	Dimension reduction	No	Yes	Yes	Yes
$p(\mathbf{s}_o\|\theta)$	$p(\|\mathbf{s}_o - \mathbf{s}_i\| < \epsilon\|\theta)$	Summary statistics	Tolerance, Kernel, Distance	No	No	Yes	Yes
$p(\mathbf{s}_o\|\theta)$	$p(\mathbf{s}_o\|\tilde{\theta})$	Parameter	Tolerance, Kernel, Distance	No	No	No	Yes
$p(\mathbf{x}\|\theta)$	$\hat{p}(\mathbf{x}\|\theta)$	Model	Empirical distribution	No	No	No	Yes

decrease with the squared distance of θ_i from θ^*, and a zero weight is assigned to all θ_i that are not among the first k closest parameter values to θ^*. We denote the k values that get a positive weight by $\tilde{\theta}_i,\ \ i = 1, 2, \ldots, k$, and the data sets simulated under these parameter values by $\tilde{\mathbf{x}}_i$. Our model $q(\mathbf{x}|\theta, \mathcal{Z}_{n,m})$ is a k-dimensional mixture distribution, where the support of this distribution is the set $\cup_{i=1}^{k} \tilde{\mathbf{x}}_i$ and the mixing weights are ω_i.

In a data set $\tilde{\mathbf{x}}_i$, we assume that all n data points $(\tilde{x}_{1i}, \tilde{x}_{2i}, \ldots, \tilde{x}_{ni})$ are equally likely and that the weight for each data point j is $\omega_{ji} = \omega_i/n$. We denote the probability that a random data value x in the new data set is equal to a specific data value $\tilde{x}_{ji}$ observed in the set $\cup_{i=1}^{k} \tilde{\mathbf{x}}_i$ by ϕ_{ji}. We let $\boldsymbol{\phi} = \{\phi_{ji}\}$ and $\boldsymbol{\omega} = \{\omega_{ji}\},\ j = 1, 2, \ldots, n,\ i = 1, 2, \ldots, k$, and we place a Dirichlet prior distribution

$$\pi(\boldsymbol{\phi}|\boldsymbol{\omega}) \propto \prod_{i=1}^{k}\prod_{j=1}^{n} \phi_{ji}^{\omega_{ji}-1}, \tag{1}$$

on the $(kn-1)$-dimensional simplex Φ. A new data set $\mathbf{x}$ under θ is simulated by first drawing $\boldsymbol{\phi}$ from this prior distribution and then simulating n IID observations, where the probability of an observation to take the value $\tilde{x}_{ji}$ is the simulated value of ϕ_{ji}.

We clarify the simulation of a new data set in AABC with a numerical toy example using $m = 3,\ \ k = 2$, and $n = 2$. We assume that the data are generated from an exponential distribution, $x \sim \text{Exp}(\theta)$, with the prior $\theta \sim \text{Unif}(0, 1)$. Because $m = 3$, our reference set has three θ values, and from the uniform prior we sample (0.08, 0.19, 0.76). Using $x \sim \text{Exp}(0.08)$, with $n = 2$, we get a data set (1.36, 3.65), and similarly, using Exp(0.19) and Exp(0.76), we get the data sets (16.25, 1.93) and (0.62, 0.12), respectively. A new parameter value simulated from the prior is 0.34, under which we desire to simulate a data set. The Euclidean distances between each of the three parameter values and the new parameter value are $\sqrt{(0.34-0.08)^2} \approx 0.26,\ \sqrt{(0.34-0.19)^2} \approx 0.15$, and $\sqrt{(0.34-0.76)^2} \approx 0.42$. Because $k = 2$, only the first two data sets, which are simulated under parameter values closest to 0.34, are considered for resampling. Using the Epanechnikov kernel, the weights for the first and second data sets are $(3/4)(1/0.42)[1-(0.26/0.42)] \approx 0.68$, and $(3/4)(1/0.42)[1-(0.15/0.42)] \approx 1.14$, respectively, where 0.42 scales the kernel so that the data set simulated with parameter value 0.76 gets a weight of zero. We now simulate the probabilities (ϕ_1, ϕ_2) from the Dirichlet distribution Dir(0.68, 1.14) to get $\phi_1 = 0.43,\ \ \phi_2 = 0.57$. Resampling of the data points within each data set is performed with equal probability, so we have the following resampling distribution: $P(x = 1.36) = 0.43/2,\ \ P(x = 3.65) = 0.43/2,\ \ P(x = 16.25) = 0.57/2,\ \ P(x = 1.93) = 0.57/2$. A new sample under the parameter value 0.34 is obtained by simulating a sample of size $n = 2$ from this distribution on a set of {1.36, 3.65, 16.25, 1.93}, where the observations are generated independently from each other.

The joint sampling distribution of a new data set $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ for our model is given by

$$q(\mathbf{x}|\theta, \mathcal{Z}_{n,m}) = \int_{\Phi} q'(\mathbf{x}|\boldsymbol{\phi}, \mathcal{Z}_{n,m})\pi(\boldsymbol{\phi}|\boldsymbol{\omega})\, d\boldsymbol{\phi}\ , \tag{2}$$

Table 2
Notation used in the text and algorithms.

Random value	Realized value	Method	Description
θ	θ_i	ABC/AABC	Parameter value
	θ_i^*	AABC	Parameter value
$\tilde{\theta}$	$\tilde{\theta}_i$	AABC	Parameter value in the set $\mathcal{Z}_{n,m}$
$\boldsymbol{\phi}$	$\boldsymbol{\phi}_i$	AABC	Probability vector for θ or θ_i^*
	ϕ_{ij}	AABC	jth element of $\boldsymbol{\phi}_i$
$\mathbf{x}$	$\mathbf{x}_i$	ABC	Data set of size n simulated from $p(\mathbf{x}\|\theta)$ or $p(\mathbf{x}\|\theta_i)$
	$\mathbf{x}_i^*$	AABC	Data set of size n simulated from $q(\mathbf{x}\|\theta_i^*)$
$\tilde{\mathbf{x}}$	$\tilde{\mathbf{x}}_i$	AABC	Data set of size n in $\mathcal{Z}_{n,m}$ simulated from $p(\mathbf{x}\|\tilde{\theta})$ or $p(\mathbf{x}\|\tilde{\theta}_i)$
x_j	x_{ij}	ABC	jth data point in $\mathbf{x}_i$
	x_{ij}^*	AABC	jth data point in $\mathbf{x}_i^*$
$\tilde{x}_j$	$\tilde{x}_{ij}$	AABC	jth data point in $\tilde{x}_i$

where $q'(\mathbf{x}|\boldsymbol{\phi}, \mathcal{Z}_{n,m})$ is a multinomial distribution with number of trials n and event probabilities ϕ_{ji} for events of observing $x = \tilde{x}_{ji}$.

There are two approximations involved in replacing $p(\mathbf{x}|\theta)$ with $q(\mathbf{x}|\theta, \mathcal{Z}_{n,m})$. The first approximation is due to replacing the sampling distribution of the data set $p(\mathbf{x}|\theta)$ with a k-dimensional mixture distribution $\sum_{i=1}^{k} \omega_i p(\mathbf{x}|\tilde{\theta}_i)$. Accordingly, we call this an approximation on the parameter space, because in a sense we use a combination of parameter values $\tilde{\theta}_i$ to approximate the value θ. This parameter space approximation alone is not helpful for simulating data sets efficiently, since it is still computationally infeasible to simulate data sets under models $p(\mathbf{x}|\tilde{\theta}_i)$. The second approximation is due to replacing each model $p(\mathbf{x}|\tilde{\theta}_i)$ that contributes to the mixture distribution with its estimate $\hat{p}(\mathbf{x}|\tilde{\theta}_i)$, the empirical distribution of the data set $\tilde{\mathbf{x}}_i$. We call this an approximation on the model space, because the desired model is replaced with an estimate obtained from a data set. This approximation is implicitly formulated in the Dirichlet prior distribution $\pi(\boldsymbol{\phi}|\boldsymbol{\omega})$, because $\pi(\boldsymbol{\phi}|\boldsymbol{\omega})$ assigns probability mass only on the kn data values in the set $\cup_{i=1}^{k} \cup_{j=1}^{n} \tilde{x}_{ji}$. Thus, it only uses the empirical distribution of data sets $\tilde{\mathbf{x}}_i$ and not all possible values in the support of $p(\mathbf{x}|\tilde{\theta}_i)$.

An AABC algorithm by rejection is as follows:

Algorithm 2: AABC by rejection sampling.

Initialization.
i. For $i = 1, 2, \ldots, m$, simulate $\theta_i \sim \pi(\theta)$.
ii. For $i = 1, 2, \ldots, m$, simulate $\mathbf{x}_i \sim p(\mathbf{x}|\theta_i)$.

1. Simulate $\theta^* \sim \pi(\theta)$.
2. Calculate $\omega_i = (3/4)[1/\|\theta^* - \theta_{(k+1)}\|][1 - \|(\theta^* - \theta_i)/(\theta^* - \theta_{(k+1)})\|^2]\mathbf{I}_{\{\|\theta^* - \theta_i\| < \|\theta^* - \theta_{(k+1)}\|\}}$, for $i = 1, 2, \ldots, k$.
3. Find $(\tilde{\mathbf{x}}_i, \tilde{\theta}_i) \in \{(\mathbf{x}_1, \theta_1), \ldots, (\mathbf{x}_m, \theta_m)\}$, $i = 1, 2, \ldots, k$ for which $\omega_i > 0$.
4. Simulate $\boldsymbol{\phi} \sim \pi(\boldsymbol{\phi}|\boldsymbol{\omega}) \propto \prod_{i=1}^{k} \prod_{j=1}^{n} \phi_{ji}^{\omega_{ji}-1}$, $\boldsymbol{\omega} = \{\omega_{ji} = \omega_i/n\}$.
5. Simulate n new data points x_ℓ^*, $\ell = 1, 2, \ldots, n$, each with $P(x_\ell^* = \tilde{x}_{ji}) = \phi_{ji}$, and set $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$.
6. Calculate the summary statistics $\mathbf{s}^*$ from $\mathbf{x}^*$.
7. If $\|\mathbf{s}_o - \mathbf{s}^*\| < \epsilon$, output θ^*.

In Algorithm 2, we call steps i and ii initialization steps, because these steps are run only once. The information obtained by initialization steps in AABC is used to bypass a large number M of simulations from the model that are required in a standard ABC approach.

3.3. The posterior distribution of θ sampled by AABC

In sampling the approximate posterior distribution of θ by AABC methods, we use the two ABC approximations described in Section 2. In steps 6 and 7 of Algorithm 2, each data instance $\mathbf{x}$ is substituted with summary statistics $\mathbf{s}$, and an acceptance condition with tolerance ϵ is used to measure the proximity of the summary statistics calculated from the observed and simulated data by the Euclidean distance. Substituting $p(\mathbf{x}|\theta)$ with $q(\mathbf{x}|\theta, \mathcal{Z}_{n,m})$ gives the approximate posterior distribution sampled by an AABC method as

$$\pi_{AABC}(\theta|\mathbf{x}_o) = \frac{1}{C_q} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s}-\mathbf{s}_o\|<\epsilon\}} \times \left[\int_{\Phi} q'(\mathbf{x}|\boldsymbol{\phi}, \mathcal{Z}_{n,m}) \pi(\boldsymbol{\phi}|\boldsymbol{\omega})\, d\boldsymbol{\phi} \right] \pi(\theta)\, d\mathbf{x}, \tag{3}$$

where $C_q = \int_{\Theta} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s}-\mathbf{s}_o\|<\epsilon\}} \left[\int_{\Phi} q'(\mathbf{x}|\boldsymbol{\phi}, \mathcal{Z}_{n,m}) \pi(\boldsymbol{\phi}|\boldsymbol{\omega})\, d\boldsymbol{\phi} \right] \pi(\theta)\, d\mathbf{x}\, d\theta$ is the normalizing constant.

The probability of sampling a parameter value θ^* in Algorithm 2 is proportional to

$$\sum_{\mathbf{x}} \mathbf{I}_{\{\|\mathbf{s}-\mathbf{s}_o\|<\epsilon\}} \sum_{\boldsymbol{\phi}} q'(\mathbf{x}|\boldsymbol{\phi}, \mathcal{Z}_{n,m}) \pi(\boldsymbol{\phi}|\boldsymbol{\omega}) \pi(\theta^*). \tag{4}$$

Expression (4) is equal to the finite sampling version for the posterior distribution in Eq. (3), except that it is missing the normalizing constant $1/C_q$. Therefore, Algorithm 2 samples any parameter value θ^* in proportion to its correct posterior probability $\pi_{AABC}(\theta^*|\mathbf{x}_o)$ given by Eq. (3).

The AABC approach is sensible in that as a model increasingly permits a larger number of simulated data sets, for large sample sizes, the posterior distribution obtained by an AABC method approaches the same distribution as the posterior distribution obtained by an ABC method. We codify this claim with a theorem.

Theorem. *Let $\pi(\theta)$ be a bounded prior on θ, and let $\mathbf{x}_o = (x_{1o}, x_{2o}, \ldots, x_{no})$ be a data set of size n. Let $\pi_{ABC}(\theta|\mathbf{x}_o)$ and $\pi_{AABC}(\theta|\mathbf{x}_o)$, be the posterior distributions sampled by a standard ABC method and an AABC method, respectively. Then*

$$\lim_{m\to\infty} \lim_{n\to\infty} \pi_{AABC}(\theta|\mathbf{x}_o) = \lim_{n\to\infty} \pi_{ABC}(\theta|\mathbf{x}_o). \tag{5}$$

A proof of the theorem appears in the Appendix. The convergence of the posterior distribution sampled by AABC is a consequence of the fact that for each value of θ, the sampling distribution $\int_{\Phi} q'(\mathbf{x}|\boldsymbol{\phi}, \mathcal{Z}_{n,m}) \pi(\boldsymbol{\phi}|\boldsymbol{\omega})\, d\boldsymbol{\phi}$ converges to the true sampling distribution $p(\mathbf{x}|\theta)$ as the sample size n and the number of simulated samples m from $p(\mathbf{x}|\theta)$ increase.

At first glance, our theorem might seem not to be very useful in practice, since it does not quantify the behavior of the posterior distribution obtained by AABC for small m, and when m is large, AABC is not needed. However, the theorem is important because it guarantees that the approximate model $q(\mathbf{x}|\theta, \mathcal{Z}_{n,m})$ used in AABC is a legitimate statistical model for convergence to the posterior distribution obtained by ABC.

The double limit in Eq. (5) is required because the standard notion of a distribution converging to a point in the parameter space as n increases does not apply to $\pi_{AABC}(\theta|\mathbf{x}_o)$. The posterior distribution $\pi_{AABC}(\theta|\mathbf{x}_o)$ depends not only on the sample size n, but also on the number m of simulated data sets from $p(\mathbf{x}|\theta)$. Hence, for convergence of the posterior distribution based on the likelihood $q(\mathbf{x}|\theta, \mathcal{Z}_{n,m})$, the requirement is that both $n \to \infty$ and $m \to \infty$. As $n \to \infty$, the distribution of a data set $\tilde{\mathbf{x}}_i$ converges to $p(\mathbf{x}|\tilde{\theta}_i)$, the correct sampling distribution with the incorrect parameter value $\tilde{\theta}_i$. As $m \to \infty$, the distance between the parameter value θ under which we want to simulate a new data set and the parameter values $\tilde{\theta}_i \in \mathcal{Z}_{n,m}$ closest to θ approaches zero. Therefore, taking both limits results in convergence to the correct sampling distribution $p(\mathbf{x}|\theta)$.

3.4. Computational performance of AABC methods

In practice, a number of problem-specific factors, including the availability of computational resources and the level of accuracy desired in the results, can affect the choice of computational method for a given problem. Hence, providing generic recommendations on when to choose AABC methods over ABC methods is not simple. However, we present an analysis to gain insight into the computational time complexity of an AABC algorithm as compared with an ABC algorithm.

The calculation of summary statistics from a data set, the comparison of simulated with observed summary statistics, and the assessment of the rejection condition have the same computational cost in AABC and ABC algorithms. The computational cost of simulating data sets, however, differs in AABC and ABC.

Let the computation time required to simulate a data set from distribution $p(\mathbf{x}|\theta)$ and to draw a parameter value θ from its prior distribution $\pi(\theta)$ be c_d and c_p respectively. Simulating data sets in M iterations of an ABC algorithm requires time Mc_d, because all data sets in ABC are simulated from the distribution $p(\mathbf{x}|\theta)$. The total time to simulate parameter and data set pairs in ABC is $M(c_d + c_p)$.

In AABC, building the set $\mathcal{Z}_{n,m}$ requires simulating m parameter values from the prior and m data sets from the model $p(\mathbf{x}|\theta)$, and thus has computation time $m(c_d + c_p)$. For a scalar θ, finding the distances between each element θ_i of $\mathcal{Z}_{n,m}$ and the parameter value θ under which we want to simulate a new data set requires m calculations, and thus, for an ℓ-dimensional parameter, $m\ell$ calculations. Sorting these distances to find the k values $\tilde{\theta}_i$ closest to the parameter value θ has a cost on the order of $m \log m$. Once the appropriate $\tilde{\theta}_i$ are found, the data sets $\tilde{\mathbf{x}}_i$ simulated under $\tilde{\theta}_i$ are accessed in a negligible constant time. Finally, we simulate a $(kn-1)$-dimensional Dirichlet variable $\boldsymbol{\phi}$, and we draw n IID uniform random variables to sample the distribution given by the probabilities $\boldsymbol{\phi}$ on the support $\cup_{i=1}^{k} \tilde{\mathbf{x}}_i$. The computational cost of these two steps is linear in n, or $O(n)$. Hence, given $\mathcal{Z}_{n,m}$, simulating data sets in M iterations of an AABC algorithm requires time $Mm(\ell + \log m) + M[O(n)]$. Therefore, the computational time difference between an ABC algorithm and an AABC algorithm is

$$M(c_d + c_p) - \{m(c_p + c_d) + Mm(\ell + \log m) + M[O(n)]\}. \quad (6)$$

Simulating from the prior distribution $\pi(\theta)$ and the Dirichlet distribution $\pi(\boldsymbol{\phi}|\boldsymbol{\omega})$ is often fast. Therefore, $O(n)$ and c_p are relatively small in Eq. (6), which gives the computational time difference between an ABC and an AABC algorithm roughly as $(M-m)c_d - Mm(\ell + \log m)$. Because the motivation for use of AABC is that c_d is large, this computation clarifies that AABC is substantially faster than ABC when $m \ll M$.

4. Applications

In this section, we demonstrate the inferential performance of the AABC approach with two examples.

4.1. Example 1: The strength of balancing selection in a multi-locus K-allele model

Here, we consider inference from the stationary distribution of allele frequencies in the diffusion approximation to a Wright–Fisher model with symmetric balancing selection and mutation (Wright, 1949). If we let $a_i > 0$, with $i = 1, 2, \ldots, K$, and $\sum_{i=1}^{K} a_i = 1$, denote the frequency of allelic type i in the population at a genetic locus, then the joint probability density function of allele frequencies $x = (a_1, a_2, \ldots, a_K)$ is

$$f(x|\sigma, \mu) = c(\sigma, \mu)^{-1} \exp\left(-\sigma \sum_{i=1}^{K} a_i^2\right) \prod_{i=1}^{K} a_i^{\mu/K-1}. \quad (7)$$

Parameters σ and μ determine the population-scaled strength of balancing selection and the mutation rate, respectively. We assume that a random sample of n draws from the population approximates the allele frequencies in the population.

In our example, we assume that the data are generated with the same true parameter values over 50 loci, each with $K = 4$, and that the allele frequencies at each locus are independent of the allele frequencies at other loci. Thus, the joint probability density function of allele frequencies for all loci is equal to the product of probability density functions across loci.

In Eq. (7), the likelihood function $f(x|\sigma, \mu)$ is hard to compute, as a consequence of difficulty in calculating the normalizing constants $c(\sigma, \mu)$. Therefore, performing likelihood-based inference on σ and μ by standard computational approaches such as MCMC is difficult. Fortunately, a numerical integration method specifically designed to calculate $c(\sigma, \mu)$ allows us to determine the likelihoods (Genz and Joyce, 2003). We will use this method to sample the posterior distribution of the parameters by a standard MCMC algorithm. The distribution sampled by a standard MCMC algorithm can be regarded as the "true" posterior distribution and is not an approximate posterior distribution as in ABC or AABC, because MCMC does not involve the approximations used in ABC and AABC. Therefore, we will use the posterior sample obtained by the MCMC approach for comparing the accuracy of the posterior samples obtained by ABC and AABC. The numerical integration method to obtain $c(\sigma, \mu)$ is computationally feasible only for small values of μ and σ, and thus, in our example we restrict our attention to uniform prior distributions, on $(1, 10)$ for the mutation rate (μ), and on $(1, 50)$ for the selection parameter (σ).

ABC and AABC methods are well-suited for inference from this model because the statistics $\sum_{j=1}^{K} a_j^2$ and $-\sum_{j=1}^{K} \log a_j$ are jointly sufficient for parameters σ and μ, and no information is lost in dimension reduction to the summary statistics. We used a method specifically designed for simulating data sets from $f(x|\sigma, \mu)$ (Joyce et al., 2012).

We designed our simulation study as follows. First, we simulated μ_i from a uniform distribution on $(1, 10)$ and σ_i from a uniform distribution on $(1, 50)$, for $i = 1, 2, \ldots, 10^6$. Next, we simulated a data set $\mathbf{x}_i$ that consists of the allele frequencies from 50 loci, where the allele frequencies at each locus are simulated independently from the distribution given in Eq. (7) under the parameter values (μ_i, σ_i). This process creates a reference set with $M = 10^6$, consisting of $\{(\mathbf{x}_1, \mu_1, \sigma_1), (\mathbf{x}_2, \mu_2, \sigma_2), \ldots, (\mathbf{x}_{10^6}, \mu_{10^6}, \sigma_{10^6})\}$. We built the sets $\mathcal{Z}_{n,m}$, with $m = 5 \times 10^2, 10^3, 5 \times 10^3, 10^4, 5 \times 10^4, 10^5$ independently from each other, by sampling m triplets $(\mathbf{x}_i, \mu_i, \sigma_i)$ from the reference set, uniformly at random without replacement. We also selected 10^3 test cases $(\mathbf{x}_i, \mu_i, \sigma_i)$ from the reference set, independently from each other and uniformly at random without replacement. For each test case, (μ_i, σ_i) is the "true" parameter vector, and $\mathbf{x}_i$ is the "observed" test data set. Given a test data set, we performed AABC by rejection sampling (Algorithm 2) and ABC by rejection sampling (Algorithm 1).

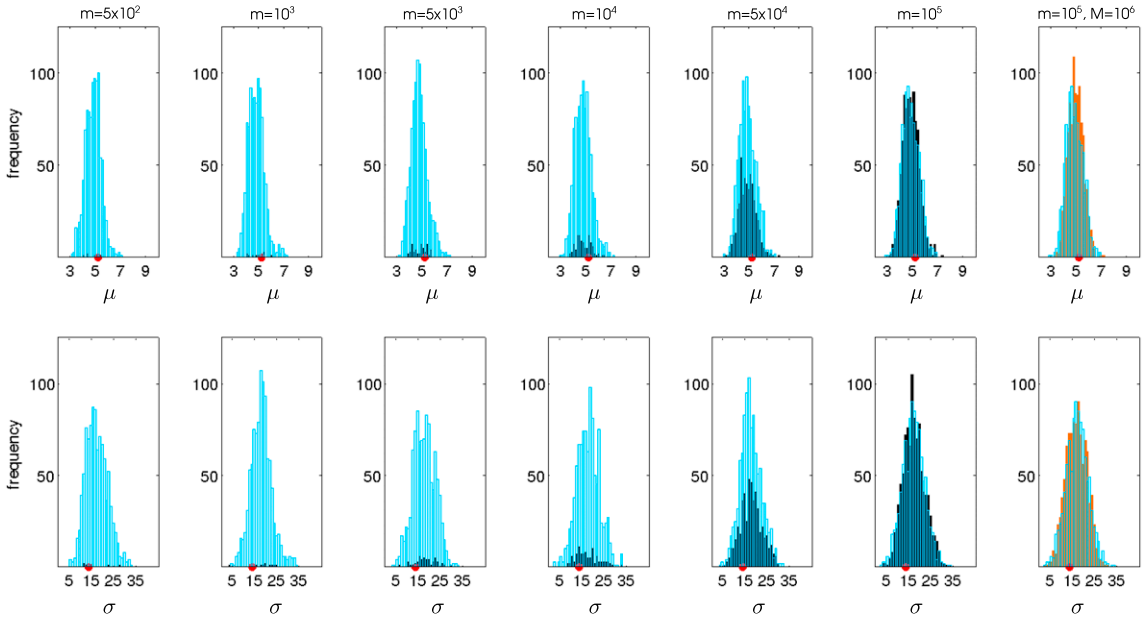

Fig. 2. Inference on the strength of balancing selection. The figure shows the marginal posterior distributions of parameters μ (top), and σ (bottom) of example 1 obtained with: ABC by rejection (black), AABC by rejection (blue), and MCMC (orange, the last column). The number m of data sets simulated from the mechanistic model for each analysis performed by AABC and ABC appears at the top of each column. For the MCMC algorithm that samples the true posterior distribution, the full reference set is used, and thus, the total number of proposed moves is $M = 10^6$. The red dot on the x-axis is the true value of the parameter, equal in all plots.

We implemented MCMC by a Metropolis–Hastings (MH) algorithm, where the proposal distribution is chosen as the prior distribution of the parameters. For the moves proposed in the MH algorithm, we used the 10^6 parameter values (μ_i, σ_i), $i = 1, 2, \ldots, 10^6$, already simulated from their prior distributions in the reference set. Each proposed pair of parameter values (μ_i, σ_i) was accepted according to the standard MH acceptance rule. Thus, if the new parameter values $(\mu_{(i+1)}, \sigma_{(i+1)})$ increase the likelihood, accept the new parameter values. Otherwise, accept the new parameter values with probability proportional to the ratio of the likelihood under the proposed values to the likelihood under the current values, $f(\mathbf{x}_o|\sigma_{(i+1)}, \mu_{(i+1)})/f(\mathbf{x}_o|\sigma_i, \mu_i)$, where $\mathbf{x}_o$ is the observed data set. We started to sample the chain after 10^3 burn-in steps to decrease the effect of the starting point (μ_1, σ_1). After initiation of sampling, to decrease the correlation in the sampled values, we used thinning by treating every 999th sampled value as a draw from the posterior distribution of the parameters. This procedure resulted in a sample of size 10^3 from the posterior distribution of the parameters (μ, σ).

We compared the performance of the ABC (Algorithm 1) and AABC (Algorithm 2) approaches when only m data sets can be simulated under the model, as given in $\mathcal{Z}_{n,m}$. In both the ABC and AABC approaches, we accepted the top 0.1 percentile of m simulated parameter values as a posterior sample. The simulated data sets used in Algorithm 1 for ABC and Algorithm 2 for AABC are different, and because we fix the posterior sample as the top 0.1 percentile of simulated values, we will have different ϵ values in the two algorithms. Therefore, we provide a comparison of the empirical tolerance values ϵ in ABC and AABC.

We assessed the error in the posterior samples for μ and σ separately using the root sum of squared error in the posterior sample and we report the root mean squared error (RMSE), averaged over 10,000 test cases (see Blum et al., 2013).

Results. Samples from the posterior distribution of parameters (σ, μ) obtained by ABC and AABC using a single typical data set of true parameter values are given in Fig. 2. In analyses with $m = 5 \times 10^2$, 10^3, 5×10^3 or 10^4 simulated data sets, few samples are accepted with ABC, and thus, little mass is observed in ABC histograms (black). For small m, because of our use of only the 0.1% top closest simulated parameter values, ABC does not produce an adequate sample size from the posterior distribution of parameters. AABC, however, produces a posterior sample of size 10^3 for any m, because 10^6 data sets are simulated from the non-mechanistic model and the top 0.1 percentile are accepted as belonging to the approximate posterior distribution. The posterior samples obtained by AABC recover the true value well even for small m (Fig. 2).

In Fig. 3, we present pairwise plots for the posterior means of the 10^3 test cases, each with different true parameter values μ_i and σ_i, obtained by an AABC analysis using $m = 5 \times 10^2$, an ABC analysis using $M = 10^6$, and an MCMC analysis. The comparison between AABC and ABC (last column in Fig. 3) shows that posterior means obtained by AABC with small $m = 5 \times 10^2$ can be as accurate as posterior means obtained by ABC with a larger $M = 10^6$. Further, we see that AABC with m as low as 5×10^2 performs well against MCMC, which samples the "true" posterior distribution (middle column in Fig. 3).

For $m = 5 \times 10^2$ to $m = 10^6$ simulated data sets, the RMSE value for parameter μ decreases from 0.0422 to 0.0414 in AABC (Table 3). These values slightly underestimate the variability of the posterior distribution as determined by MCMC, but they are comparable to the RMSE value of 0.0420 in the standard ABC analysis using $M = 10^6$ simulated data sets from the mechanistic model. The error in the posterior sample is a function of the tolerance condition ϵ in the ABC and AABC approaches. In our ABC and AABC analyses in Fig. 3, the values of ϵ considered are different because we use different simulated data sets in the two procedures, and we accepted a fixed sample size of 10^3 values in the posterior samples with both methods. To assess the magnitude of the error for the 10^3 test cases, we calculated the relative error by $r_t = (\epsilon_{AABC} - \epsilon_{ABC})/\epsilon_{ABC}$, where ϵ_{AABC} and ϵ_{ABC} are tolerances for a posterior sample of size 10^3 in the AABC and ABC analyses, respectively. As m increases, the number of test cases that have $\epsilon_{AABC} < \epsilon_{ABC}$, and thus $r_t < 0$, increases, indicating that the error due to tolerance approximations in the AABC approach is smaller than the error in the ABC approach for the accepted values in the posterior distribution (Table 4). For

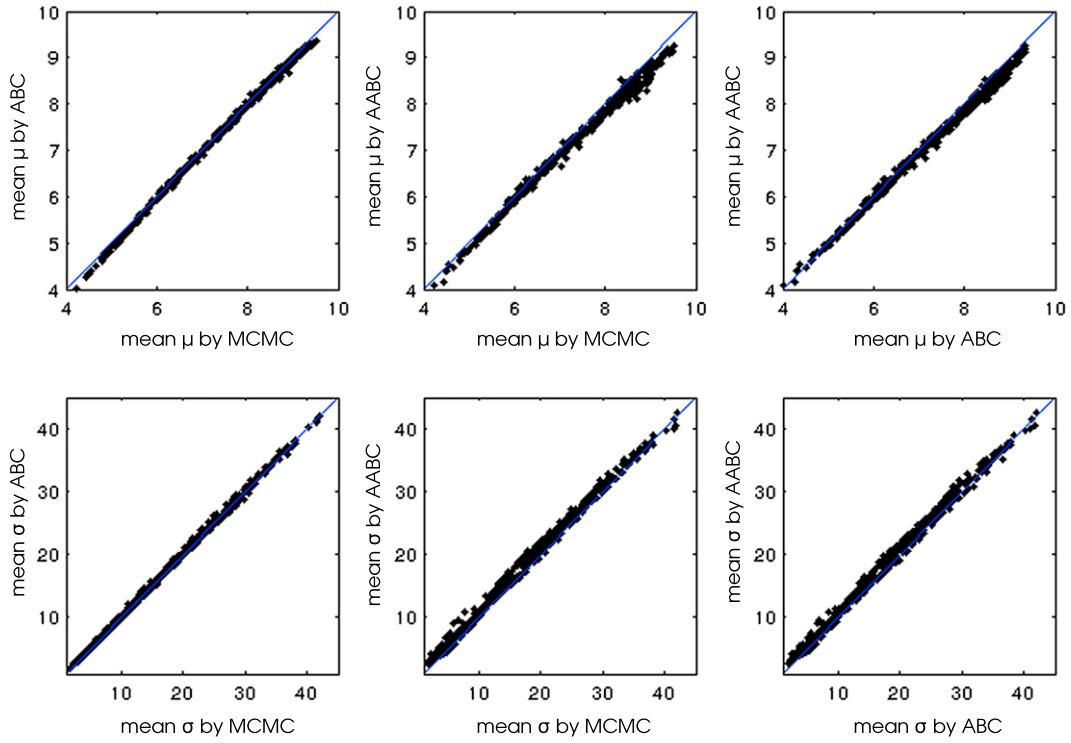

Fig. 3. Comparisons of posterior means of μ and σ obtained by three methods: AABC with $m = 500$, ABC with $M = 10^6$, and MCMC using 10^3 "true" data sets in example 1. Each mean is taken across all values in the appropriate posterior distribution. The means of posterior samples obtained by AABC with a small number of simulated data sets such as 500 show almost perfect correlation with means of posterior samples obtained by ABC and MCMC methods, indicating that the means in AABC are comparably accurate.

$m = 5 \times 10^3$, in 770 cases among 10^3, samples from the AABC posterior have a smaller error than the samples from the ABC posterior, and there are no cases in which the error in the AABC approach is larger than twice the error in the ABC approach ($\epsilon_{AABC} > 2\epsilon_{ABC}$).

4.2. Example 2: Admixture rates in hybrid populations

Models in which hybrid populations are founded by, and receive genetic contributions from, multiple source populations are of interest in describing the demographic history of admixture. Stochastic models including admixture often result in likelihoods that are difficult to calculate, and statistical methods capable of performing inference on admixture rates have received much attention for their implications on topics ranging from human evolution to conservation ecology (Falush et al., 2003; Tang et al., 2005; Buerkle and Lexer, 2008). Here, we consider inference on admixture rates from a mechanistic model of Verdu and Rosenberg (2011). We use reported estimates of individual admixture as data.

We consider a model of admixture for a diploid hybrid population of constant size N, founded at some known t generations in the past with contributions from source populations A and B. We follow the distribution of admixture fractions of individuals in the hybrid population at a given genetic locus. Each generation, the admixture fraction for each individual in the hybrid population is obtained as the mean of the admixture fractions of its parents. The parents are chosen independently of each other, from source population A, source population B, or the hybrid population of the previous generation with probabilities p_A, p_B, and p_H, respectively ($p_A + p_B + p_H = 1$). In the special case of the founding generation, $p_H = 0$, and we assume $p_A = p_B = 0.5$. Individuals from source populations A and B are assigned admixture fractions of 1 and 0, respectively. For example, if both parents of an individual in the hybrid population of the founding generation are from source population A, then that individual has admixture fraction $(1 + 1)/2 = 1$. If both parents are from population B, the admixture fraction is $(0 + 0)/2 = 0$, and if one parent is from population A and the other is from population B, then the admixture fraction is $(1 + 0)/2 = 0.5$. The distribution of the admixture fraction in the hybrid population is propagated in this manner for t generations until the present, in which a sample of n individuals is obtained from the resulting distribution (Fig. 4). Our goal is to estimate the admixture rates (p_A, p_B, p_H), given the individual admixture fractions estimated from observed genetic data.

We apply the AABC approach using individual admixture fractions from $n = 604$ individuals from Central African Pygmy populations reported by Verdu et al. (2009), with an assumed constant population size of $N = 2 \times 10^4$. This assumption differs slightly from the original model in Verdu and Rosenberg (2011) in that a finite population size is assumed, so that only $2 \times 10^4 + 1$ admixture fraction values are allowed in the population at any given generation. The admixture fractions from Verdu et al. (2009) were computed using STRUCTURE (Pritchard et al., 2000), applied to microsatellite data, and we treat the estimates as parametric values.

We assume that an admixture event with contributions from two ancestral source populations started at the mean estimate of $t = 771$ generations ago (Verdu et al., 2009) with a generation time of 25 years, and that it continued until the present. Source population A refers to an ancestral Pygmy population, and source population B refers to an ancestral non-Pygmy population. The feature of this model relevant to our method is the computational intractability of simulating data sets. For each set of parameter values (p_A, p_B, p_H) simulated from the priors, the distribution of admixture fractions is discrete on a support of a number of admixture fraction values that doubles each generation, and this distribution evolves for 771 generations. A random sample of admixture fraction values comparable to the values calculated from the observed data set is obtained from the distribution of the present generation. Simulating a large number of data sets under this model with such a large number of generations is computationally infeasible,

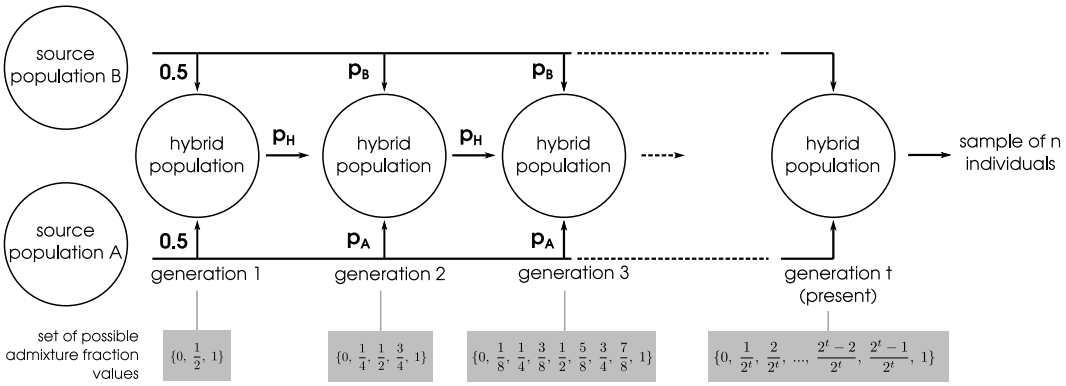

Fig. 4. The admixture model of example 2. Each generation after the founding, the parents of an individual are chosen independently of each other, from source population A, source population B, or the hybrid population of the previous generation, with probabilities p_A, p_B, and p_H, respectively ($p_A + p_B + p_H = 1$).

Table 3
RMSE values based on 10^3 data sets, for parameters μ and σ in a balancing selection model, obtained by three methods: AABC, ABC, and MCMC. M^* indicates the number of total proposed values in the MCMC algorithm. The RMSE values obtained by AABC are relatively constant for both parameters. The differences between RMSE values obtained by AABC and true RMSE values obtained by MCMC are small, indicating that the posterior sample obtained by AABC is close to a sample obtained by sampling the true posterior distribution of parameters.

	AABC						ABC	MCMC
	$m = 5 \times 10^2$	$m = 10^3$	$m = 5 \times 10^3$	$m = 10^4$	$m = 5 \times 10^4$	$m = 10^5$	$M = 10^6$	$M^* = 10^6$
RMSE(μ)	0.0421	0.0422	0.0415	0.0415	0.0415	0.0414	0.0420	0.0422
RMSE(σ)	0.0421	0.0425	0.0419	0.0420	0.0420	0.0420	0.0429	0.0432

Table 4
Error in the posterior sample obtained by AABC relative to ABC. Each column under m is a histogram for the number of data sets for which the relative error falls within the interval in the first column. A total of 10^3 "true" data sets are analyzed. We measured the error by $r_t = (\epsilon_{AABC} - \epsilon_{ABC})/\epsilon_{ABC}$, where ϵ_{AABC} and ϵ_{ABC} are tolerances for a posterior sample of size 10^3 in AABC and ABC respectively. As m increases, the number of data sets that have smaller ϵ_{AABC} increases ($r_t < 0$), indicating that the error in AABC is smaller than the error in ABC for the accepted values in the posterior distribution.

Relative tolerance (r_t)	m					
	5×10^2	10^3	5×10^3	10^4	5×10^4	10^5
$r_t \leq 0$	285	380	770	842	988	996
$0 < r_t \leq 0.10$	185	305	171	144	12	4
$0.10 < r_t \leq 0.25$	268	199	51	14	0	0
$0.25 < r_t \leq 1.00$	226	90	8	0	0	0
$1.00 < r_t \leq 10.0$	30	26	0	0	0	0
$10.0 < r_t \leq 100$	6	0	0	0	0	0

and standard ABC is impractical. We thus perform AABC by rejection (Algorithm 2) using $m = 10^4$ realizations from this model. We assume a Dirichlet prior with hyperparameters (1, 1, 1) on parameters (p_A, p_B, p_H).

We also separately assessed the contribution of the approximations on the parameter and model spaces in the AABC approach to the RMSE, using a simulation study with a small number of generations ($t = 30$), for which simulating data sets from the mechanistic model is feasible. Here, we used a reference set with $M = 10^5$, built by following the same steps as in Section 4.1, where 10^5 parameter vectors (p_A, p_B, p_H) are simulated from their prior distributions and then a data set of admixture fraction values is simulated under each of these parameter vectors. We selected the test cases by sampling 10^3 data sets and parameter pairs from the reference set, uniformly at random without replacement. Again, we compared the performance of the ABC and AABC by rejection using the sets $\mathcal{Z}_{n,m}$. Simulating data under the model of example 2 is computationally intensive, however, due to a long evolutionary history involved in the model. Therefore, we compared the performance of the ABC and AABC approaches in a version of the model that involved a shorter evolutionary history than our stipulated full model. We used the full model with long-term evolutionary history to analyze a real data set from Central African Pygmy populations. The accepted parameter values represent the top 1 percentile of $M = 10^5$ parameter values in the reference set.

First, we performed AABC with rejection as in Algorithm 2 with 10^3 "true" data sets using $m = 5 \times 10^2$, 10^3, 5×10^3, 10^4, 5×10^4, and 10^5 realizations from the model. We calculated the RMSE for p_A, p_B, and p_H using 10^3 "true" data sets. This AABC analysis includes error due to approximations on the parameter space and on the model space. Second, we performed an AABC analysis with the same set of m realizations, by including the error only due to the approximation on the parameter space. We eliminated the error due to the approximation on the model space by running Algorithm 2 up through step 5, and then simulating data sets from the mechanistic model by substituting steps 6 and 7 of Algorithm 2 with step 2 of Algorithm 1, the standard ABC approach by rejection. By this substitution, all data sets are simulated from the mechanistic model, but each data set is obtained using a parameter vector ($\tilde{p}_A$, $\tilde{p}_B$, $\tilde{p}_H$) found in step 5 of Algorithm 2. In this procedure, the error due to the approximation on the model space is eliminated, because data sets are simulated from the correct mechanistic model and not by resampling from the available realizations in $\mathcal{Z}_{n,m}$. However, this procedure includes error due to the approximation on the parameter space, because each data set is simulated not under the correct proposed parameter value, but under the parameter value ($\tilde{p}_A$, $\tilde{p}_B$, $\tilde{p}_H$), the closest value to the correct proposed value that can be found in $\mathcal{Z}_{n,m}$. We compared the RMSE of the AABC procedure involving the approximation on both the parameter and model spaces and the RMSE of the AABC procedure involving only the approximation on the parameter space to the RMSE obtained from a standard ABC approach. For these two AABC procedures, we also compared the percent excess in RMSE, defined as the ratio of the absolute difference in RMSE of the AABC and standard ABC approaches to the RMSE of the standard ABC approach, expressed as a percent.

Results. The individual admixture fractions calculated from the Pygmy data carry substantial information about the admixture parameters p_A, p_B, and p_H, since the joint posterior distribution is concentrated in a relatively small region of the 3-dimensional unit

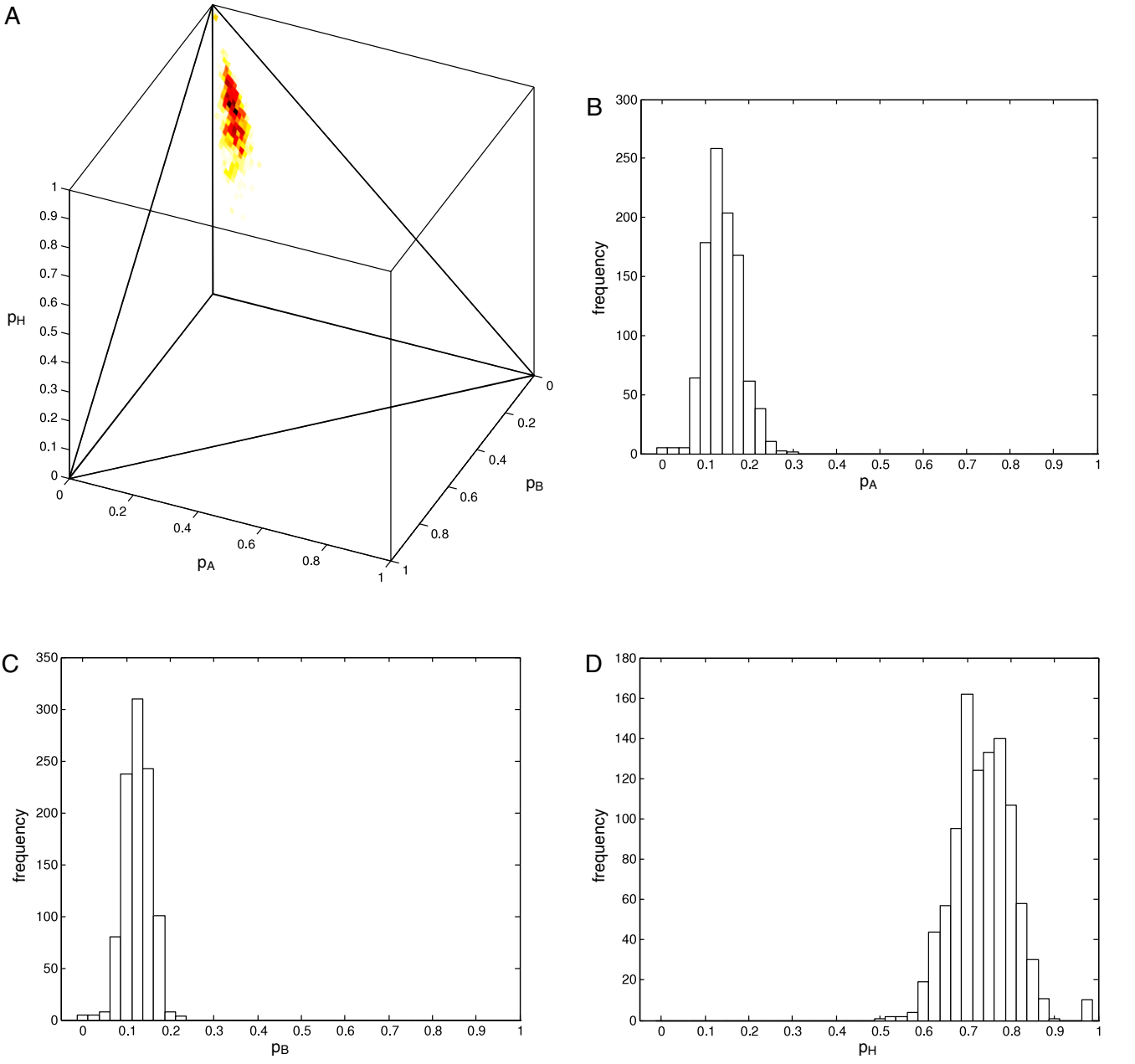

Fig. 5. AABC analysis on the Pygmy data of example 2 with $m = 10^4$ realizations under the mechanistic model. (A) The joint distribution on the unit simplex, with probability mass increasing from white to dark red. (B)–(D) Marginal distributions of p_A, p_B, and p_H.

simplex on which (p_A, p_B, p_H) sits (Fig. 5A). The marginal posterior distributions (Fig. 5B–D) have means $p_A = 0.139$, $p_B = 0.125$, and $p_H = 0.735$. These values are interpreted as a contribution of genetic material of 13.9% from the ancestral Pygmy population (source population A), 12.5% from the ancestral Non-Pygmy population (source population B), and 73.5% from the hybrid population to itself at each generation, over 771 generations of constant admixture.

For the simulation study with $t = 30$ generations and 10^4 "true" data sets, the percent excess in RMSE values from AABC analyses decreases with increasing m (Fig. 6). Further, as m increases, the percent excess in RMSE due to the approximation on the parameter space decreases, due to the fact that for large m, the difference decreases between the closest parameter value chosen at step 5 of Algorithm 2 and the correct parameter value under which we want to simulate a data set. The percent excess in RMSE values from the three analyses – AABC with $m = 10^5$ realizations and approximation only on the parameter space, AABC with $m = 10^5$ realizations and approximation on the parameter and model spaces, and the standard ABC with $M = 10^5$ realizations – are virtually indistinguishable (Fig. 6, red stars). For $m = 5 \times 10^3$, the AABC analysis with approximations on the parameter and model spaces has a small percent excess in RMSE of 0.67, 0.20, 0.97, for p_A, p_B, and p_H respectively, whereas the AABC analysis including only the approximation on the parameter space has percent excess RMSE values of 0.26, 0.25, 0.42. We note that at $m = 5 \times 10^3$, the percent excess in RMSE values is small in the AABC approach in relation to the standard ABC analysis, showing that the AABC posterior is a reasonable approximation to the ABC posterior.

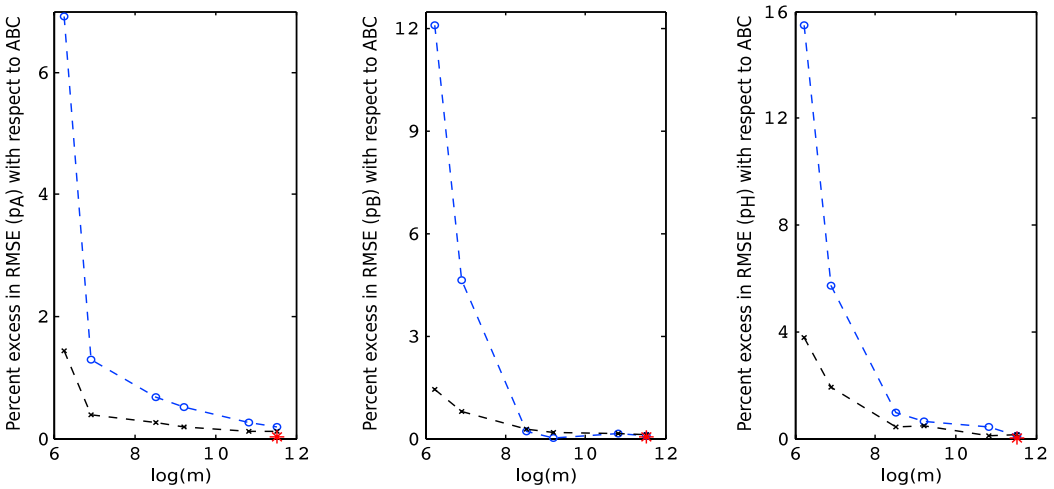

Fig. 6. Percent excess in RMSE of each parameter with respect to ABC in the admixture model. The values on the y-axis are calculated by $\left(1-\frac{\text{RMSE in AABC}}{\text{RMSE in ABC}}\right)\times 100$. The decrease in percent excess RMSE is shown for parameters p_A, p_B, and p_H with increasing m (on natural logarithmic scale), the number of simulated samples from the mechanistic admixture model in AABC. The percent excess in RMSE values for AABC analysis performed with an approximation only on the parameter space (black), and with an approximation on both the parameter space and the model space (blue) are shown. The red star in each plot represents RMSE obtained by a standard ABC analysis performed with $M = 10^5$ simulated values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Discussion

Performing likelihood-based inference from statistical models involving complex stochastic processes is often challenging due to computationally intractable likelihoods. ABC methods use data simulated from the model to assess the parameter likelihoods. To deliver an adequate sample from the posterior distribution of the parameters, however, ABC requires a large number of simulated data sets, and it might not perform well when only a small number of data sets can be simulated.

In this article, we have introduced AABC, a class of computationally feasible methods that extends ABC to model spaces in which only a small number of data sets can be simulated from the model. In addition to ABC approximations, the AABC approach requires approximations on the parameter and model spaces, and thus, the error in posterior samples obtained by our approach will be larger than in ABC. Therefore, AABC is not meant to be an alternative to ABC when ABC is computationally feasible, but rather, a complementary method to perform inference when ABC is computationally not feasible. The strength of AABC is that it can deliver a posterior sample from the joint distribution of parameters for a small number of simulated data sets. Therefore, a researcher can fix m and thus the computation time *a priori*, to simulate data from the mechanistic model to obtain a reasonable inference by AABC; other ABC methods may fail to produce an adequate posterior sample in equivalent computation time. In our example, for moderate values of m (e.g., 5×10^3) for which standard ABC was unsatisfactory, AABC adequately sampled an approximate posterior distribution. However, AABC has the limitation that when m is too small, its posterior sample can have a large error and give a distorted representation of the true posterior distribution.

AABC relies on two statistical approximations. In our approximation on the parameter space, we set the distribution of a data set under a parameter value to be a k-dimensional mixture distribution, where the k components of this mixture are chosen from the set $\mathcal{Z}_{n,m}$ and mixture weights are chosen according to an Epanechnikov kernel. Kernel approximations on the data space have an operational role in implementing ABC methods, and kernel weighting for mixture components extends this role to the parameter space in AABC methods. In our approximation on the model space, we modeled the uncertainty associated with model $p(\mathbf{x}|\theta)$ using Dirichlet prior probabilities on kn points of k data sets, each with n observations simulated from $p(\mathbf{x}|\theta)$. In an AABC algorithm, each draw from the Dirichlet distribution produces a probability model on kn data points, and a data set of size n from this probability model is obtained in each iteration. Our approach in handling the model uncertainty has some resemblance to statistical "emulators" (Kennedy and O'Hagan, 2001), approximative methods used to express the model uncertainty when simulating data under a mechanistic model is computationally intensive. However, emulators are often motivated in the context of Gaussian processes, where the uncertainty in the model space can be reasonably well modeled by a normal distribution. Because the assumption of normality may be questionable in many population-genetic contexts, we have avoided making this assumption in AABC.

Our approach of using a non-mechanistic statistical model to help perform inference on model-specific parameters of a mechanistic model is a fundamental difference between AABC and existing ABC methods. ABC performs inference on model-specific parameters of a mechanistic model using a likelihood based purely on that model. AABC instead performs inference on the same model-specific parameters of the mechanistic model as ABC, using a likelihood based on a non-mechanistic model that incorporates a small number of data sets simulated from the mechanistic model. Consequently, the model likelihoods used in ABC and AABC are not exactly the same, and the posterior distributions targeted by the two classes of methods are not exactly equivalent for finite sample sizes. The advantage of AABC methods in contrast to pure non-mechanistic modeling approaches (e.g., nonparametric methods) is that AABC can perform inference on the quantities of interest—the model-specific parameters of the mechanistic model.

Currently, AABC is best suited for models that have relatively few parameters and for which the stochastic process used in simulating data is computationally intensive. Forward simulation models in population genetics, such as the admixture model we have examined, often fall into this category. Due to its generality as a computational method, we expect AABC to be useful outside its immediate applications in population genetics, for example, in spatial and temporal models in ecology, evolution, and other fields.

Acknowledgments

The authors thank Paul Verdu for helpful discussions on the genetics of Central African Pygmy populations. Support for this research was partially provided by NIH grant R01 HG005855, National Science Foundation grant DBI-1146722, and National Institute of General Medical Sciences of the NIH grant P30 RR033376.

Appendix

Here, we show that the posterior distribution sampled by AABC converges to the posterior distribution sampled by ABC, as the number of simulated data sets m and the sample size n increase. The model $q(\mathbf{x}|\theta, \mathcal{Z}_{n,m})$ used in AABC is based on a mixture distribution of k components as described in Section 3.2. For notational simplicity, we set $k = 1$ and prove the theorem for a mixture distribution with a single component. Extension to $k = k_o$ for any fixed $k_o > 1$ is straightforward for the following reason. For $k = 1$, only the n data points in a data set $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ that is simulated under θ^*, the closest parameter value to θ, get positive weights. In this case, the weight vector $\boldsymbol{\omega}$ places positive weights on n data points. For $k = k_o$, the number of data points on which a positive weight is placed is equal to $n \times k_o$. Here, there are n points in each of k_o data sets. In this case, the weight vector $\boldsymbol{\omega}$ places positive weights on $n \times k_o$ data points. Therefore, considering $k = k_o$ increases the dimension of the weight vector $\boldsymbol{\omega}$, but otherwise the claim of the theorem is the same. Because the weight vector $\boldsymbol{\omega}$ always sums to 1, the following proof of the theorem is the same for any k_o.

We let $u \leq n$ be the number of distinct values $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_u$ in the set $\tilde{\mathbf{x}}$, and denote the number of observed $\tilde{x}_i$ by $\tilde{n}_i$, where $n = \sum_{i=1}^{u} \tilde{n}_i$. From Eq. (1), we recall that the prior distribution on $\boldsymbol{\phi}$ is the Dirichlet distribution $\pi(\boldsymbol{\phi}|\boldsymbol{\omega}) \propto \prod_{i=1}^{k} \prod_{j=1}^{n} \phi_{ji}^{\omega_{ji}-1}$.

Then the prior distribution for the probabilities of an AABC replicate data set based on the ABC simulated data set $\tilde{\mathbf{x}}$ is the Dirichlet distribution

$$\pi(\boldsymbol{\phi}|\boldsymbol{\omega}) = \left[\Gamma\left(\sum_{i=1}^{u} \tilde{n}_i\right) \Big/ \prod_{i=1}^{u} \Gamma(\tilde{n}_i)\right] \prod_{i=1}^{k} \phi^{\tilde{n}_i - 1}$$

with parameters $\tilde{n}_1, \tilde{n}_2, \ldots, \tilde{n}_k$, where we now have explicitly written the normalizing constant $[\Gamma(\sum_{i=1}^{u} \tilde{n}_i)/\prod_{i=1}^{u} \Gamma(\tilde{n}_i)]$ for the Dirichlet distribution. Our goal is to show that $\lim_{m\to\infty} \lim_{n\to\infty} \pi_{AABC}(\theta|\mathbf{x}_o) = \lim_{n\to\infty} \pi_{ABC}(\theta|\mathbf{x}_o)$.

Recalling Eq. (3), we have

$$\lim_{m\to\infty} \lim_{n\to\infty} \pi_{AABC}(\theta|\mathbf{x}_o) = \lim_{m\to\infty} \lim_{n\to\infty} \frac{1}{C_q} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s}-\mathbf{s}_o\|<\epsilon\}} \times \left[\int_{\Phi} q'(\mathbf{x}|\boldsymbol{\phi}, \mathcal{Z}_{n,m}) \pi(\boldsymbol{\phi}|\boldsymbol{\omega})\, d\boldsymbol{\phi}\right] \pi(\theta)\, d\mathbf{x}. \tag{A.1}$$

The integral in the brackets is the expectation of $q'(\mathbf{x}|\boldsymbol{\phi}, \mathcal{Z}_{n,m})$, with respect to the prior $\pi(\boldsymbol{\phi}|\boldsymbol{\omega})$. We let $C = \binom{n}{n_1\, n_2 \cdots n_u}$, and using the definition of $q'(\mathbf{x}|\boldsymbol{\phi}, \mathcal{Z}_{n,m}) = C \prod_{j=1}^{n} \prod_{i=1}^{n} \phi_i^{\mathbf{I}_{\{x_j=\tilde{x}_i\}}}$ in Section 3.2, and $\pi(\boldsymbol{\phi}|\boldsymbol{\omega}) = [\Gamma(\sum_{i=1}^{k} \tilde{n}_i)/\prod_{i=1}^{k} \Gamma(\tilde{n}_i)] \prod_{i=1}^{k} \phi^{\tilde{n}_i - 1}$, we get

$$\int_{\Phi} q'(\mathbf{x}|\boldsymbol{\phi}, \mathcal{Z}_{n,m}) \pi(\boldsymbol{\phi}|\boldsymbol{\omega})\, d\boldsymbol{\phi} = C \frac{\Gamma\left(\sum_{i=1}^{u} \tilde{n}_i\right)}{\prod_{i=1}^{u} \Gamma(\tilde{n}_i)} \prod_{j=1}^{n} \int_{\Phi} \left(\prod_{i=1}^{n} \phi_i^{\mathbf{I}_{\{x_j=\tilde{x}_i\}}}\right) \left(\prod_{i=1}^{u} \phi_i^{\tilde{n}_i - 1}\right) d\boldsymbol{\phi}.$$

Here, we have exchanged the order of the product over j with the integral, because the expectation of the product of n IID observations in sample $\mathbf{x}$ is equal to the product of the expectations of observations x_j. We label the realized value of the jth data point x_j by (j) such that $\prod_{i=1}^{n} \phi_i^{\mathbf{I}_{\{x_j=\tilde{x}_i\}}} = \phi_{(j)}$, and write

$$\int_{\Phi} q(\mathbf{x}|\boldsymbol{\phi}, \tilde{\mathbf{x}}) \pi(\boldsymbol{\phi}|\boldsymbol{\omega})\, d\boldsymbol{\phi} = C \frac{\Gamma\left(\sum_{i=1}^{k} \tilde{n}_i\right)}{\prod_{i=1}^{u} \Gamma(\tilde{n}_i)} \prod_{j=1}^{n} \int_{\Phi} \left(\prod_{\substack{i=1 \\ i\neq(j)}}^{u} \phi_i^{\tilde{n}_i - 1}\right) \phi_{(j)}^{\tilde{n}_{(j)}}\, d\boldsymbol{\phi}. \tag{A.2}$$

Using $\int_{\Phi} \frac{\Gamma[(\sum_{i=1,i\neq(j)}^{k} \tilde{n}_i) + \tilde{n}_{(j)} + 1]}{[\prod_{i=1,i\neq(j)}^{k} \Gamma(\tilde{n}_i)]\Gamma(\tilde{n}_{(j)}+1)} \left(\prod_{i=1,i\neq(j)}^{k} \phi_i^{\tilde{n}_i - 1}\right) \phi_{(j)}^{\tilde{n}_{(j)}}\, d\boldsymbol{\phi} = 1$ (p. 487, Kotz et al., 2000), we substitute the integral in Eq. (A.2) with the ratio of the gamma functions to get

$$\begin{aligned}
&\int_{\Phi} q(\mathbf{x}|\boldsymbol{\phi}, \tilde{\mathbf{x}}) \pi(\boldsymbol{\phi})\, d\boldsymbol{\phi} \\
&= C \frac{\Gamma\left(\sum_{i=1}^{k} \tilde{n}_i\right)}{\prod_{i=1}^{k} \Gamma(\tilde{n}_i)} \prod_{j=1}^{n} \frac{\left[\prod_{i=1,i\neq(j)}^{k} \Gamma(\tilde{n}_i)\right] \Gamma(\tilde{n}_{(j)}+1)}{\Gamma\left[\left(\sum_{i=1,i\neq(j)}^{k} \tilde{n}_i\right) + \tilde{n}_{(j)} + 1\right]} \\
&= C \prod_{j=1}^{n} \frac{\Gamma(n)}{\Gamma(\tilde{n}_{(j)})} \frac{\Gamma(\tilde{n}_{(j)}+1)}{\Gamma(n+1)} = C \prod_{j=1}^{n} \left(\frac{\tilde{n}_{(j)}}{n}\right).
\end{aligned}$$

Substituting $C \prod_{j=1}^{n} \left(\frac{\tilde{n}_{(j)}}{n}\right)$ for the integral in brackets in Eq. (A.1), we have

$$\begin{aligned}
&\lim_{m\to\infty} \lim_{n\to\infty} \pi_{\epsilon}(\theta|\mathbf{x}_o, q(\mathbf{x}|\theta, \mathcal{Z}_{n,m})) \\
&= \lim_{m\to\infty} \lim_{n\to\infty} \frac{1}{C_q} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s}-\mathbf{s}_o\|<\epsilon\}}\, C \prod_{j=1}^{n} \left(\frac{\tilde{n}_{(j)}}{n}\right) \pi(\theta)\, d\mathbf{x} \\
&= \frac{\lim_{m\to\infty} \lim_{n\to\infty} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s}-\mathbf{s}_o\|<\epsilon\}}\, C \prod_{j=1}^{n} \left(\frac{\tilde{n}_{(j)}}{n}\right) \pi(\theta)\, d\mathbf{x}}{\lim_{m\to\infty} \lim_{n\to\infty} C_q}.
\end{aligned} \tag{A.3}$$

We apply the dominated convergence theorem to exchange the limits in n and the integrals in the numerator and denominator of Eq. (A.3). The assumptions of the theorem are satisfied as follows: (1) The integrand in Eq. (A.3) is bounded: The indicator functions are bounded by 1, the ratios $(\tilde{n}_{(j)}/n)$, where $n_{(j)} \leq n$, are bounded by 1, and the prior $\pi(\theta)$ is bounded by assumption. (2) $\lim_{n\to\infty} (\tilde{n}_{(j)}/n)$ converges pointwise to the probability of $x_{(j)}$ under $\tilde{\theta}$ given by $p(x_{(j)}|\tilde{\theta})$, by the frequency interpretation of probability. Exchanging the limits in n and the integrals, and using $\lim_{n\to\infty} (\tilde{n}_{(j)}/n) = p(x_{(j)}|\tilde{\theta})$,

$$\begin{aligned}
&\lim_{m\to\infty} \lim_{n\to\infty} \pi_{\epsilon}(\theta|\mathbf{x}_o, q(\mathbf{x}|\theta, \mathcal{Z}_{n,m})) \\
&= \frac{\lim_{m\to\infty} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s}-\mathbf{s}_o\|<\epsilon\}} \prod_{j=1}^{u} \left[p(x_{(j)}|\tilde{\theta})\right]^{n_{(j)}} \pi(\theta)\, d\mathbf{x}}{\lim_{m\to\infty} C_p} \\
&= \frac{\lim_{m\to\infty} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s}-\mathbf{s}_o\|<\epsilon\}} p(\mathbf{x}|\tilde{\theta})\, \pi(\theta)\, d\mathbf{x}}{\lim_{m\to\infty} C_p},
\end{aligned} \tag{A.4}$$

where (A.4) follows by the definition of the joint distribution $p(\mathbf{x}|\tilde{\theta}) = \prod_{j=1}^{u} \left[p(x_{(j)}|\tilde{\theta})\right]^{n_{(j)}}$.

We now apply the dominated convergence theorem for a second time to exchange the limits in m and the integrals on $\mathcal{X}$. Again, the assumptions of the dominated convergence theorem are satisfied because the integrand in (A.4) is a sequence in m of bounded functions, and as $m \to \infty$, $\tilde{\theta} \to \theta$, and $p(\mathbf{x}|\tilde{\theta}) \to p(\mathbf{x}|\theta)$. We get

$$\lim_{m\to\infty} \lim_{n\to\infty} \pi_{\epsilon}(\theta|\mathbf{x}_o, q(\mathbf{x}|\theta, \mathcal{Z}_{n,m}))$$

$$= \frac{1}{C_p} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s}-\mathbf{s}_o\|<\epsilon\}} p(\mathbf{x}|\theta)\pi(\theta)\, d\mathbf{x} = \lim_{n\to\infty} \pi_\epsilon(\theta|\mathbf{x}_o, p(\mathbf{x}|\theta)),$$

showing that the AABC posterior converges to the ABC posterior as the sample size n and the number of simulated data sets m increase.

References

Beaumont, M.A., 2003. Estimation of population growth or decline in genetically monitored populations. Genetics 164, 1139–1160.

Beaumont, M.A., Cornuet, J.-M., Marin, J.-M., Robert, C.P., 2009. Adaptive approximate Bayesian computation. Biometrika 96, 983–990.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. Genetics 162, 2025–2035.

Becquet, C., Przeworski, M., 2007. A new approach to estimate parameters of speciation models with application to apes. Genome Res. 17, 1505–1519.

Blum, M.G.B., François, O., 2010. Non-linear regression models for approximate Bayesian computation. Stat. Comput. 20, 63–73.

Blum, M.G.B., Jakobsson, M., 2010. Deep divergences of human gene trees and models of human origins. Mol. Biol. Evol. 28, 889–898.

Blum, M.G.B., Nunes, M.A., Prangle, D., Sisson, S.A., 2013. A comparative review of dimension reduction methods in approximate Bayesian computation. Statist. Sci. 28, 189–208.

Bonassi, F.V., Lingchong, Y., West, M., 2011. Bayesian learning from marginal data in bionetwork models. Stat. Appl. Genet. Mol. Biol. 10, Article 1.

Buerkle, C.A., Lexer, C., 2008. Admixture as the basis for genetic mapping. Trends Ecol. Evol. 23, 686–694.

Estoup, A., Beaumont, M.A., Sennedot, F., Moritz, C., Cornuet, J.-M., 2004. Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. Evolution 58, 2021–2036.

Fagundes, N.J.R., Ray, N., Beaumont, M.A., Neuenschwander, S., Salzano, F.M., Bonatto, S.L., Excoffier, L., 2007. Statistical evaluation of alternative models of human evolution. Proc. Natl. Acad. Sci. 104, 17614–17619.

Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multilocus genetic data: linked loci and correlated allele frequencies. Genetics 164, 1567–1587.

Fearnhead, P., Prangle, D., 2012. Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. J. R. Stat. Soc. Ser. B Stat. Methodol. 74, 1–28.

François, O., Blum, M.G.B., Jakobsson, M., Rosenberg, N.A., 2008. Demographic history of European populations *Arabidopsis thaliana*. PLoS Genet. 4, e1000075.

Genz, A., Joyce, P., 2003. Computation of the normalizing constant for exponentially weighted Dirichlet distribution integrals. Comput. Sci. Statist. 35, 557–563.

Grelaud, A., Robert, C.P., Marin, J.-M., Rodolphe, F., Taly, J.-F., 2009. ABC likelihood-free methods for model choice in Gibbs random fields. Bayesian Anal. 4, 317–336.

Joyce, P., Genz, A., Buzbas, E.O., 2012. Efficient simulation and likelihood methods for non-neutral multi-allele models. J. Comput. Biol. 19, 650–661.

Joyce, P., Marjoram, P., 2008. Approximately sufficient statistics and Bayesian computation. Stat. Appl. Genet. Mol. Biol. 7, Article 26.

Kennedy, M.C., O'Hagan, A., 2001. Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B Stat. Methodol. 63, 425–464.

Kotz, S., Balakrishnan, N., Johnson, N.L., 2000. Continuous Multivariate Distributions, second ed. Wiley–Interscience, New York.

Liu, J.S., 2008. Monte Carlo Strategies in Scientific Computing. Springer, New York.

Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S., 2003. Markov chain Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. 100, 15324–15328.

Nunes, M.A., Balding, D.J., 2010. On optimal selection of summary statistics for approximate Bayesian computation. Stat. Appl. Genet. Mol. Biol. 34, Article 34.

Plagnol, V., Tavaré, S., 2004. Approximate Bayesian computation and MCMC. In: Niederreiter, H. (Ed.), Monte Carlo and Quasi-Monte Carlo Methods. Springer-Verlag, pp. 99–114.

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A., Feldman, M.W., 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. 16, 1791–1798.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference on population structure using multilocus genotype data. Genetics 155, 945–959.

Ratmann, O., Andrieu, C., Wiuf, C., Richardson, S., 2009. Model criticism based on likelihood-free inference, with an application to protein network evolution. Proc. Natl. Acad. Sci. 106, 10576–10581.

Robert, C.P., Casella, G., 2004. Monte Carlo Statistical Methods, second ed. Springer, New York.

Robert, C.P., Cornuet, J.-M., Marin, J.-M., Pillai, N.S., 2011. Lack of confidence in approximate Bayesian computation model choice. Proc. Natl. Acad. Sci. 108, 15112–15117.

Siegmund, K.D., Marjoram, P., Shibata, D., 2008. Modeling DNA methylation in a population of cancer cells. Stat. Appl. Genet. Mol. Biol. 7, Article 18.

Sisson, S.A., Fan, Y., Tanaka, M.M., 2007. Sequential Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. 104, 1760–1765.

Sisson, S.A., Fan, Y., Tanaka, M.M., 2009. Correction for Sisson et al., Sequential Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. 106, 16889.

Sisson, S.A., Peters, G.W., Briers, M., Fan, Y., 2010. A note on target distribution ambiguity of likelihood-free samplers, arXiv.1005.5201.

Tang, H., Peng, J., Wang, P., Risch, N.J., 2005. Estimation of individual admixture: analytical and study design considerations. Genet. Epidemiol. 28, 289–301.

Tavaré, S., 2005. Ancestral inference for branching processes. In: Haccou, P., Jagers, P., Vatutin, V. (Eds.), Branching Processes in Biology: Variation, Growth, Extinction. Cambridge University Press, Cambridge, pp. 208–217.

Tavaré, S., Balding, D.J., Griffiths, R.C., Donnelly, P., 1997. Inferring coalescence times from DNA sequence data. Genetics 145, 505–518.

Verdu, P., Austerlitz, F., Estoup, A., Vitalis, R., Georges, M., Théry, S., Froment, A., Le Bomin, S., Gessain, A., Hombert, J.-M., Van der Veen, L., Quintana-Murci, L., Bahuchet, S., Heyer, E., 2009. Origins and genetic diversity of Pygmy hunter-gatherers from western Central Africa. Curr. Biol. 19, 312–318.

Verdu, P., Rosenberg, N.A., 2011. A general mechanistic model for admixture histories of hybrid populations. Genetics 189, 1413–1426.

Wegmann, D., Leuenberger, C., Excoffier, L., 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. Genetics 182, 1207–1218.

Wilkinson, R.D., 2008. Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. Stat. Appl. Genet. Mol. Biol. 12, 129–141.

Wilkinson, R.D., Steiper, M., Soligo, C., Martin, R., Yang, Z., Tavaré, S., 2010. Dating primate divergences through an integrated analysis of palaeontological and molecular data. Syst. Biol. 60, 16–31.

Wright, S., 1949. Adaptation and selection. In: Jepson, G.L., Simpson, G.G., Mayr., E. (Eds.), Genetics, Paleontology, and Evolution. Princeton University Press, Princeton, NJ.