Supporting Information

Mengersen et al. 10.1073/pnas.1208827110

SI Statistical Methods

Convergence of the Empirical Likelihood Approximation. The validation of the empirical likelihood approximation is provided by theorem 3.4 of ref. 1, which establishes an extension of Wilk's theorem to the empirical likelihood ratio. (Note that n^{-n} is the maximum of L_{el}.)

Theorem. *Let $X, Y_1, \ldots, Y_n$ be independent random vectors with common distribution F_0. For $\theta \in \Theta$, let $h(X,\theta) \in \mathbb{R}^s$. Let $\theta_0 \in \Theta$ be such that $\mathrm{Var}(h(Y_i,\theta_0))$ is finite and has rank $q>0$. If θ_0 satisfies $\mathbb{E}(h(X,\theta_0))=0$, then $-2\log\left(\frac{L_{\mathrm{el}}(\theta_0|Y_1,\ldots,Y_n)}{n^{-n}}\right) \to \chi^2_{(q)}$ in distribution when $n \to \infty$.*

We also reproduce here an illuminating comment from Art Owen (ref. 1, p 41):

> The interesting thing about Theorem 3.4 is what is not there. It includes no conditions to make $\hat{\theta}$ a good estimate of θ_0, nor even conditions to ensure a unique value for θ_0, nor even that any solution θ_0 exists. Theorem 3.4 applies in the just determined, over-determined, and under-determined cases. When we can prove that our estimating equations uniquely define θ_0, and provide a consistent estimator $\hat{\theta}$ of it, then confidence regions and tests follow almost automatically through Theorem 3.4.

Pairwise Composite Likelihoods in Population Genetics. We detail here the derivation of the composite likelihoods used for the version of the Bayesian computation with empirical likelihood ($\mathrm{BC_{el}}$) algorithm implemented in the case of the population genetics study.

Two genes from the same deme. First, we recall that we scale the time axis so that a pair of genes of the same deme coalesces at a random time with an exponential distribution with rate 1. We now consider a given locus and two microsatellite genes from our sample that come from the same deme. We denote their respective allelic states by x_1 and x_2. Their most recent common ancestor (MRCA) dates back to a time T, where $T \sim \mathcal{E}(1)$. We assume that the mutation rate, namely $\theta/2$, does not vary along the whole history of our populations. Therefore, conditioned on T, the number of mutations between x_i $(i=1,2)$ and the MRCA is distributed according to a Poisson distribution with mean $\theta T/2$. Hence, conditional on T, the number N_0 of mutations between x_1 and x_2 is a Poisson variable with mean θT and

$$\mathbb{E}\left[u^{N_0}|T\right] = \exp\{\theta T(u-1)\}.$$

Thus,

$$\mathbb{E}\left[u^{N_0}\right] = \int_0^\infty e^{-t}\exp\{\theta t(u-1)\}dt = \frac{1}{1+\theta(1-u)} = \frac{1/(1+\theta)}{1-\theta/(1+\theta)u},$$

i.e., $N_0 \sim \mathcal{G}e(\theta/(1+\theta))$, the geometric distribution with positive weight at 0. Finally, the difference between both genes is the accumulation over the N_0 mutations, i.e.,

$$x_2 - x_1 = \sum_{k=1}^{N_0} \epsilon_k,$$

where the ϵ_k's are iid Rademachers (± 1 with equal probability). Thus,

$$\mathbb{E}\left[\mathrm{e}^{i\zeta(x_2-x_1)}\right] = \mathbb{E}\left\{\mathbb{E}\left[\mathrm{e}^{i\zeta(x_2-x_1)}|N_0\right]\right\} = \mathbb{E}\left\{\prod_{k=1}^{N_0}\mathbb{E}[\exp(i\zeta\epsilon_k)]\right\} = \mathbb{E}\left[(\cos\zeta)^{N_0}\right] = \sum_{n=0}^{\infty}(1-p)(p\cos\zeta)^n, \quad \text{since } N_0 \sim \mathcal{G}e(p) \text{ with } p=\frac{\theta}{1+\theta} = \frac{1-p}{1-p\cos\zeta} = \frac{1}{1+\theta(1-\cos\zeta)} \quad [S1]$$

which proves that the pairwise likelihood is

$$\ell_2(x_1,x_2|\phi) = \frac{1}{\sqrt{1+2\theta}}\rho(\theta)^{|x_2-x_1|}, \quad [S2]$$

with $\rho(\theta) = \theta/(1+\theta+\sqrt{1+2\theta})$.

Two genes from different demes. We now consider two genes that come from two different demes that diverged from an ancestral deme at time τ in the past. We denote the allelic state of the two ancestors at time τ by x_1^0 and x_2^0, respectively. Then, $x_1-x_2 = (x_1-x_1^0)+(x_1^0-x_2^0)+(x_2^0-x_2)$, where $x_1^0-x_2^0$ follows a distribution whose Fourier transform is given by **S1**, whereas $(x_j-x_j^0)$, $j=1,2$ are iid, whose distribution is given by the difference of two allelic states separated by a fixed time τ. This distribution is derived in equation 3 of ref. 2:

$$\mathbb{P}\left(x_j - x_j^0 = \delta\right) = \mathrm{e}^{-\tau\theta/2}I_\delta(\tau\theta/2),$$

where I_δ denotes the δth-order modified Bessel function of the first kind, given by $(n \geq 0)$

$$I_{-n}(z) = I_n(z) = \sum_{k=0}^{\infty}\frac{(z/2)^{n+2k}}{k!(n+k)!}.$$

Using the independence between $(x_1-x_1^0)$ and $(x_1-x_1^0)$, we obtain

$$\mathbb{P}\left((x_1-x_1^0)+(x_1-x_1^0)=\delta\right) = \mathrm{e}^{-\tau\theta}I_\delta(\tau\theta). \quad [S3]$$

We then retrieve this distribution by computing Fourier transforms in the same vein as above. First, we note that the number N_1 of mutations between x_1 and its ancestor at time τ is a Poisson variate with parameter $\tau\theta/2$. And,

$$\mathbb{E}\left[\mathrm{e}^{i\zeta\left(x_1^0-x_1\right)}\right] = \mathbb{E}\left\{\mathbb{E}\left[\mathrm{e}^{i\zeta\left(x_1^0-x_1\right)}|N_1\right]\right\} = \mathbb{E}\left(\prod_{k=1}^{N_1}\mathbb{E}\left(\exp\left(i\zeta\epsilon_k\right)\right)\right) = \mathbb{E}\left[(\cos\zeta)^{N_1}\right] = \sum_{n=0}^{\infty}\mathrm{e}^{-\tau\theta/2}\frac{(\tau\theta/2)^n}{n!}(\cos\zeta)^n \quad \text{since } N_1 \sim \mathcal{P}o(\tau\theta/2) = \exp\{\tau\theta(\cos\zeta-1)/2\}. \quad [S4]$$

Finally, the distribution of x_2-x_1 is a (discrete) convolution product between the distributions given by **S2** and **S3**, which yields

$$\ell_2(x_1,x_2|\phi)=\frac{\mathrm{e}^{-\tau\theta}}{\sqrt{1+2\theta}}\sum_{k=-\infty}^{+\infty}\rho(\theta)^{|k|}I_{|x_1-x_2|-k}(\tau\theta).$$

Quantile Estimation. Examples of quantile distributions are the three-, four-, and five-parameter Tukey's lambda distributions and their generalizations and the Burr family of distributions; particular examples include the *g*-and-*h* and *g*-and-*k* distributions (3–5, 7).

Proposed methods for estimation of quantile distributions include maximum likelihood estimation using numerical approximations to the likelihood (7–9), moment matching (10, 11), location- and scale-free shape functionals (12), percentile matching (6), quantile matching (13), and, more recently, approximate Bayesian computation (ABC) (14–16). Sequential Monte Carlo approaches for multivariate extensions of the *g*-and-*k* have also been proposed (17).

There has been a number of ABC approaches proposed for this problem. For example, ref. 15 adopted the ABC-MCMC (Markov chain Monte Carlo) algorithm of ref. 18, in which draws of θ are based on a Metropolis algorithm with a Gaussian proposal distribution, and are accepted based on the rule $\rho(S(D), S(D'))<\varepsilon)$, where D is the entire set of order statistics, ρ is the Euclidean norm, and ϵ is heuristically chosen after inspection of a histogram of $\rho(S,S')$ obtained from a preliminary run using a very large value of ε. This approach has recently been improved by ref. 16 through more sophisticated MCMC approaches, the use of regression summary statistics for D based on percentiles and their powers, and more automated choices of ε. However, they still maintain a form of distance-based measure $\rho(S,S')$ in accepting θ.

SI Results

Normal Estimation. Figs. S1–S3 evaluate the impact on the posterior distribution approximation of increasing the number of constraints in the empirical likelihood definition. Because this is a formal example, the true posterior distribution is available.

Quantile Estimation. The $\mathrm{BC_{el}}$ experiment involved evaluation of *Algorithm 1* for estimation of the parameters of the *g*-and-*k* distribution using the two values of $\theta=(A,B,g,k)$, namely $\theta_N=(0,1,0,0)$, which corresponds to the standard normal distribution, and $\theta_A=(3,2,1,0.5)$, which was chosen by ref. 15 as 'an interesting, far-from-normal distribution. The simulation experiment comprised multiple repetitions of $\mathrm{BC_{el}}$ using different combinations of sample size, $n=(100,500)$, number of iterations, $M=(1,000;5,000;10,000)$, and number of constraints ($p=3,4,4,5,9$), corresponding to percentile sets $(0.2,0.5,0.8)$, $(0.2,0.4,0.6,0.8)$, $(0.1,0.4,0.6,0.9)$, $(0.1,0.25,0.5,0.75,0.9)$, and $(0.1,0.2,\ldots,0.9)$. Two sets of priors were considered for (A,B,g,k): $U[0,5]^4$ (denoted as P_1) and $U(-5,5).U(0,5)$, $U(-5,5)$, $(-0.1,1)$ (denoted as P_2). Although the priors were set independently for each element of θ, the four elements were drawn together at each iteration of the algorithm, so that the same importance weight ω_i was attached to the values $A^{(i)}$, $B^{(i)}$, $g^{(i)}$, $k^{(i)}$ drawn in the ith iteration. The experiment was replicated 10 times with different draws of samples of size n. Posterior means and SDs were computed for each parameter, and the overall goodness of fit to the true curve was assessed by comparing the true quantiles at $(0.05,0.10,\ldots,0.95)$ with two measures: the estimated mean at each quantile [denoted by residual sum of squares for the mean (RSSm)] and the average of the estimated quantile for each importance sample [residual sum of squares for the quantile (RSSt)].

Boxplots of the posterior means and SDs are shown in Figs. S4 and S5, respectively, for the four parameters, based on θ_A, prior P_2 and the 20 replicates for $M=(5,000;10,000)$, for 10 of the trials: $p=3,4,4,5,9$ for $n=100$ (trials 1–5) and $n=500$ (trials 6–10). Boxplots for the corresponding overall goodness of fit measures (RSSm, RSSt) are given in Fig. S6.

Superposition of Point Processes. An alternative to ABC is discussed in ref. 19 using fractional design and linear interpolation. Although their purpose is the non-Bayesian processing of models with intractable likelihood functions, they propose as their main example a model consisting of the superposition of N renewal processes with waiting times $\tau_{ij}(i=1,\ldots,M),j=1,\ldots)$ distributed as $\mathcal{G}(\alpha,\beta)$ variables, when N is unknown. The renewal processes are thus

$$\zeta_{i1}=\tau_{i1},\zeta_{i2}=\zeta_{i1}+\tau_{i2},\ldots,$$

and the observations are made of the first n values of the ζ_{ij}'s,

$$z_1=\min\{\zeta_{ij}\},z_2=\min\{\zeta_{ij};\zeta_{ij}>z_1\},\ldots,$$

ending with

$$z_n=\min\{\zeta_{ij};\zeta_{ij}>z_{n-1}\}.$$

This model offers an interesting testing ground for $\mathrm{BC_{el}}$ in that the data points z_t are neither iid nor Markov. It is however possible to recover and exploit an iid structure in this case by first simulating a pseudodataset, $(z_1^\star,\ldots,z_n^\star)$, as in ABC settings, and then deriving a sequence of renewal processes indicators $(\nu_1,\ldots,\nu_n)$, as

$$z_1^\star=\zeta_{\nu_1 1},z_2^\star=\zeta_{\nu_2 j_2},\ldots,z_n^\star=\zeta_{\nu_n j_n}.$$

These indicators are thus distributed from the prior distribution on the ν_t's and an iid sample of $\mathcal{G}(\alpha,\beta)$ variables can be derived from those indicators and the genuine data, leading to an associated empirical likelihood. As shown in Fig. S7, when applied to a simulated dataset (as in ref. 19), the empirical likelihood approximation produces a better approximation than the corresponding ABC solution based on the same statistics as ref. 19 (for exactly the same computational cost).

Time Gains in Population Genetic Models. Both population genetic experiments conducted in this paper analyze datasets with a large number of loci (100). Thus, ABC, which requires simulations of all loci to produce a simulated dataset, is quite time-consuming and particularly so when the evolutionary scenario is more complex than the one in the first experiment. We compare here the computing times required by our implementation of the $\mathrm{BC_{el}}$–AMIS sampler and by DIYABC (20) on an Intel Xeon W3680 Plateform with GNU/Linux. Both methods were parallelized over five among the six cores of this central processing unit with the OpenMP application program interface. Table S1 exhibits computation time averages on 10 replicates of the estimation.

1. Owen AB (2001) *Empirical Likelihood* (Chapman & Hall, London).
2. Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics* 150(1):499–510.
3. Gilchrist W (2000) *Staztistical Modelling with Quantile Functions* (Chapman and Hall, London).
4. Gilchrist W (2007) Modelling and fitting quantile distributions and regressions. *Am J Math Manage Sci* 27(4):401–439.
5. Haynes M, MacGillivray H, Mengersen K (1997) Robustness of ranking and selection rules using generalized *g*-and *k* distributions. *J Stat Plann Inference* 65(1): 45–66.

6. Fournier B, et al. (2007) Estimating the parameters of a generalized lambda distribution. *Comput Stat Data Anal* 51(6):2813–2835.
7. Rayner G, MacGillivray H (2002) Numerical maximum likelihood estimation for the *g*-and-*k* and generalized *g*-and-*h* distributions. *Stat Comput* 12(1):55–75.
8. Su S (2007) Numerical maximum log-likelihood estimation for generalized lambda distributions. *Comput Stat Data Anal* 51(8):3983–3998.
9. Haynes M, Mengersen K (2005) Bayesian estimation of g-and-k distributions using MCMC. *Comput Stat* 20(1):7–30.
10. Asquith W (2007) L-moments and tl-moments of the generalized lambda distribution. *Comput Stat Data Anal* 51(9):484–4496.
11. Karvanen J, Nuutinen A (2008) Characterizing the generalized lambda distribution by l-moments. *Comput Stat Data Anal* 52(4):1971–1983.
12. King R, MacGillivray H (1999) A starship fitting method for the generalized lambda distribution. *Aust N Z J. Stat* 41:353–374.
13. Rayner G, MacGillivray H (2002) Weighted quantile-based estimation for a class of transformation distributions. *Comput Stat Data Anal* 39(4):401–433.
14. Peters G, Sisson S (2006) Bayesian inference, Monte Carlo sampling and operational risk. *J Oper Risk* 1(3):27–50.
15. Allingham D, King R, Mengersen K (2009) Bayesian estimation of quantile distributions. *Stat Comput* 19(2):189–201.
16. Prangle D (2011) Summary statistics and sequential methods for approximate Bayesian computation. PhD thesis (Department of Statistics, Lancaster University, United Kingdom).
17. Drovandi C, Pettitt A (2011) Likelihood-free Bayesian estimation of multivariate quantile distributions. *Comput Stat Data Anal* 55(9):2541–2556.
18. Marjoram P, Molitor J, Plagnol V, Tavaré S (2003) Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci USA* 100(26):15324–15328.
19. Cox D, Kartsonaki C (2012) The fitting of complex parametric models. *Biometrika* 99 (3):741–747.
20. Cornuet JM, et al. (2008) Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24(23): 2713–2719.

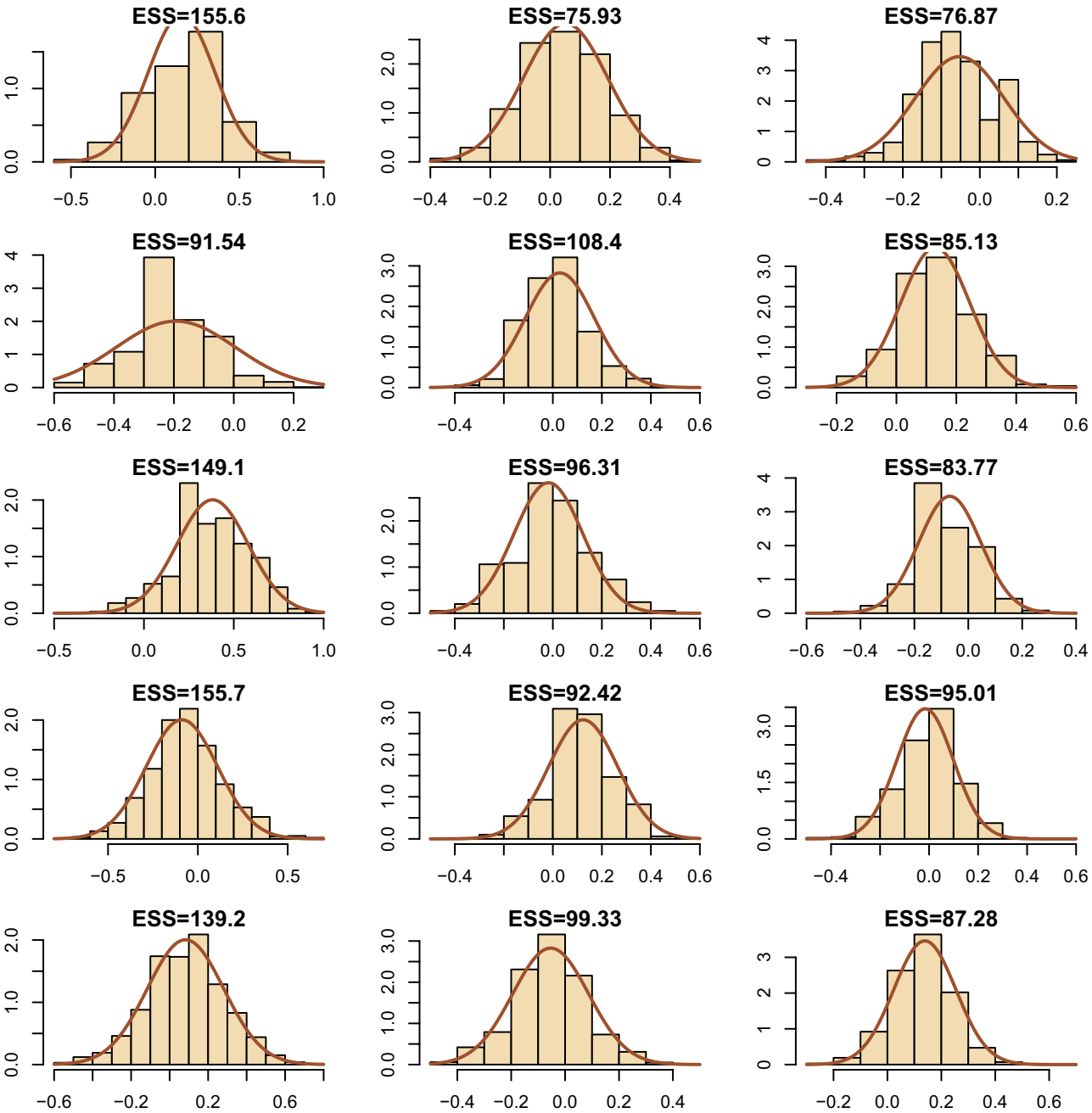

Fig. S1. Comparison of the true posterior on the normal mean (solid lines) with the empirical distribution of weighted simulations resulting from *Algorithm* BC_{el}. The normal sample sizes are 25 (column 1), 50 (column 2), and 75 (column 3), the number of simulated θ's is 10^3, and the effective sample size (ESS) M^{ESS} are given on top of each histogram. The constraint is on the first moment of the dataset.

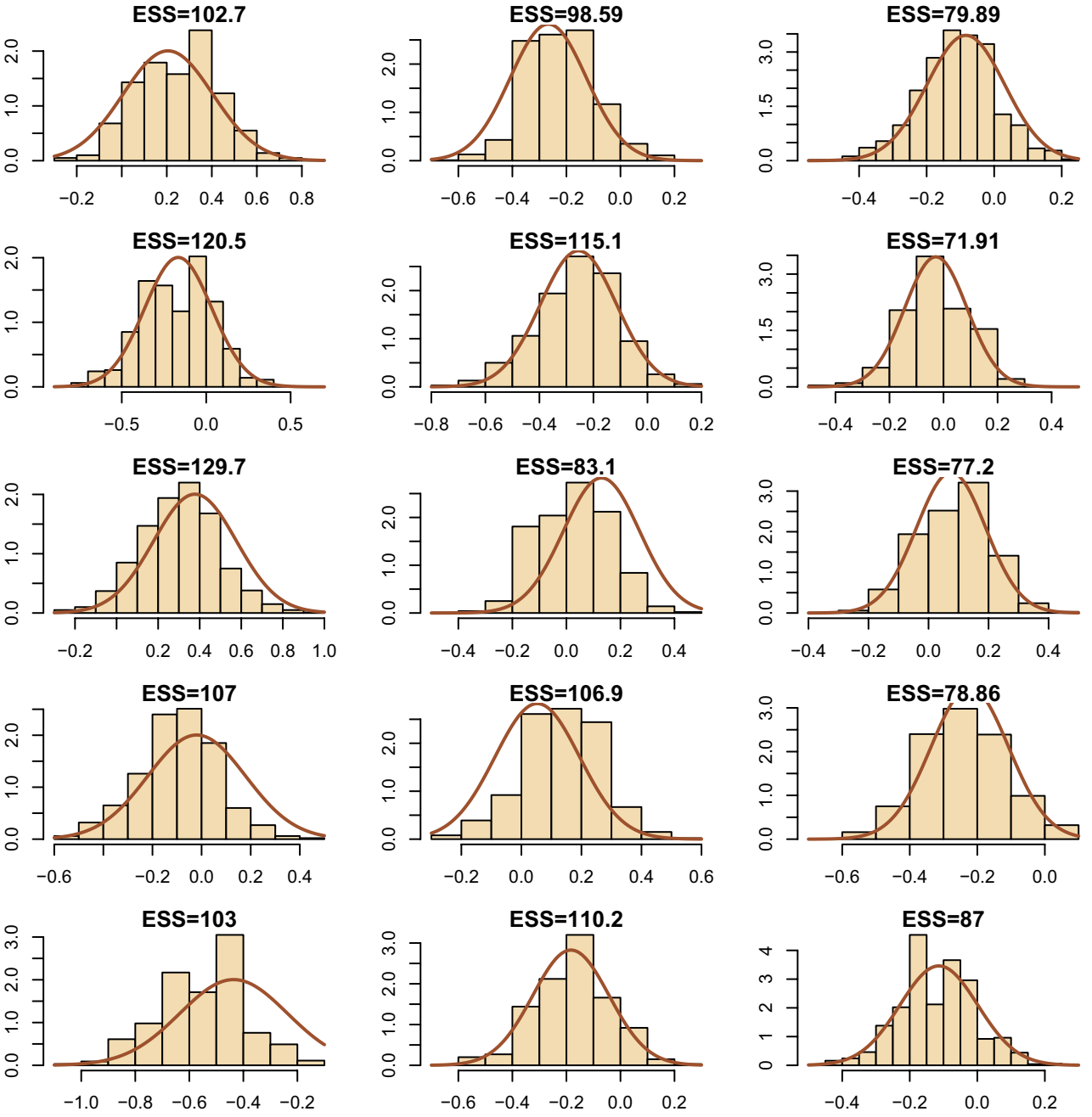

Fig. S2. Comparison of the true posterior on the normal mean (solid lines) with the empirical distribution of weighted simulations resulting from *Algorithm* BC_{el}. The normal sample sizes are 25 (column 1), 50 (column 2), and 75 (column 3), the number of simulated θ's is 10^3, and the ESS M^{ESS} are given on top of each histogram. The constraint is on the first two moments of the dataset.

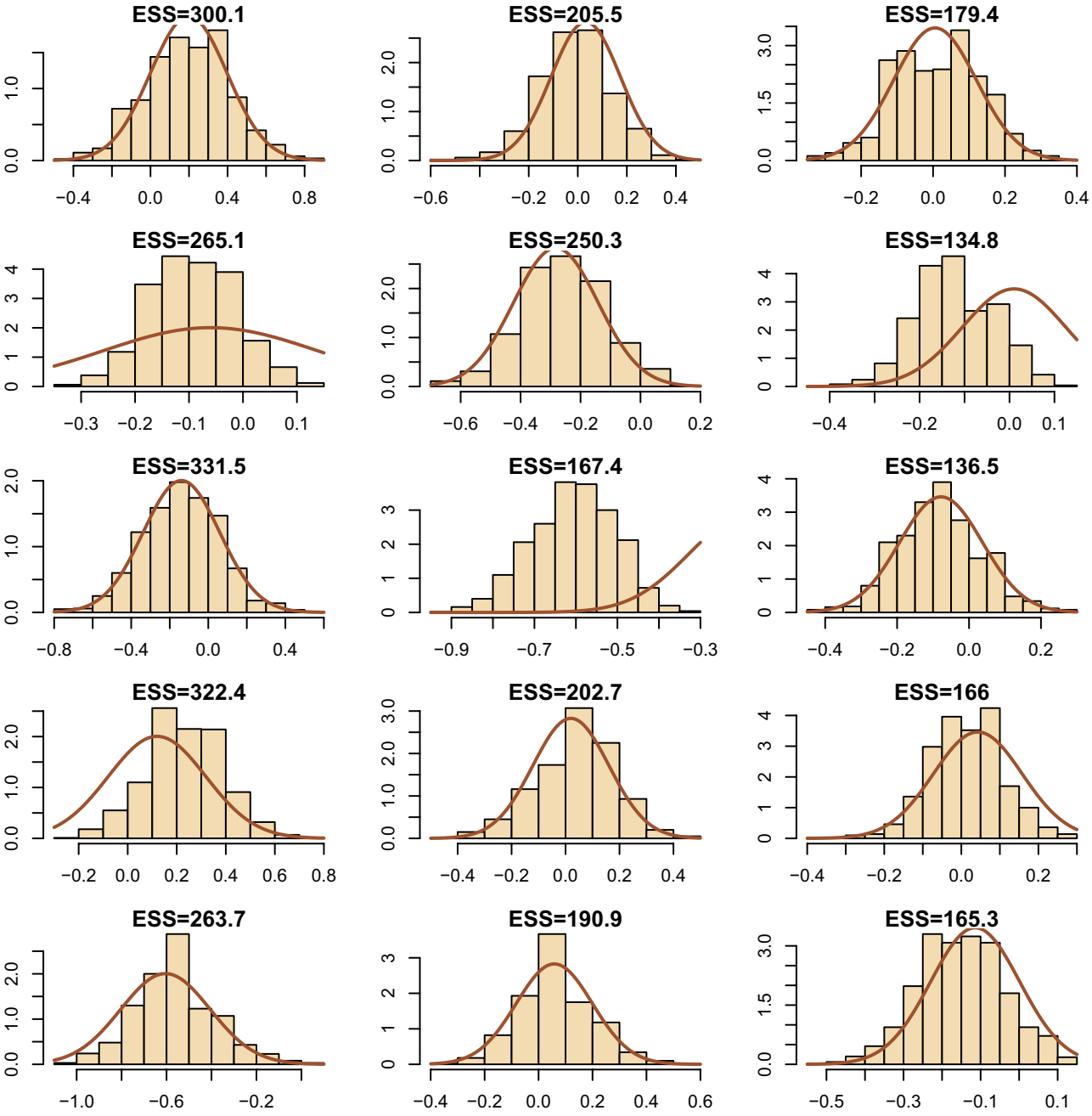

Fig. S3. Comparison of the true posterior on the normal mean (solid lines) with the empirical distribution of weighted simulations resulting from *Algorithm* BC_{el}. The normal sample sizes are 25 (column 1), 50 (column 2), and 75 (column 3), the number of simulated θ's is 10^3, and the ESS M^{ESS} are given on top of each histogram. The constraint is on the first three moments of the dataset.

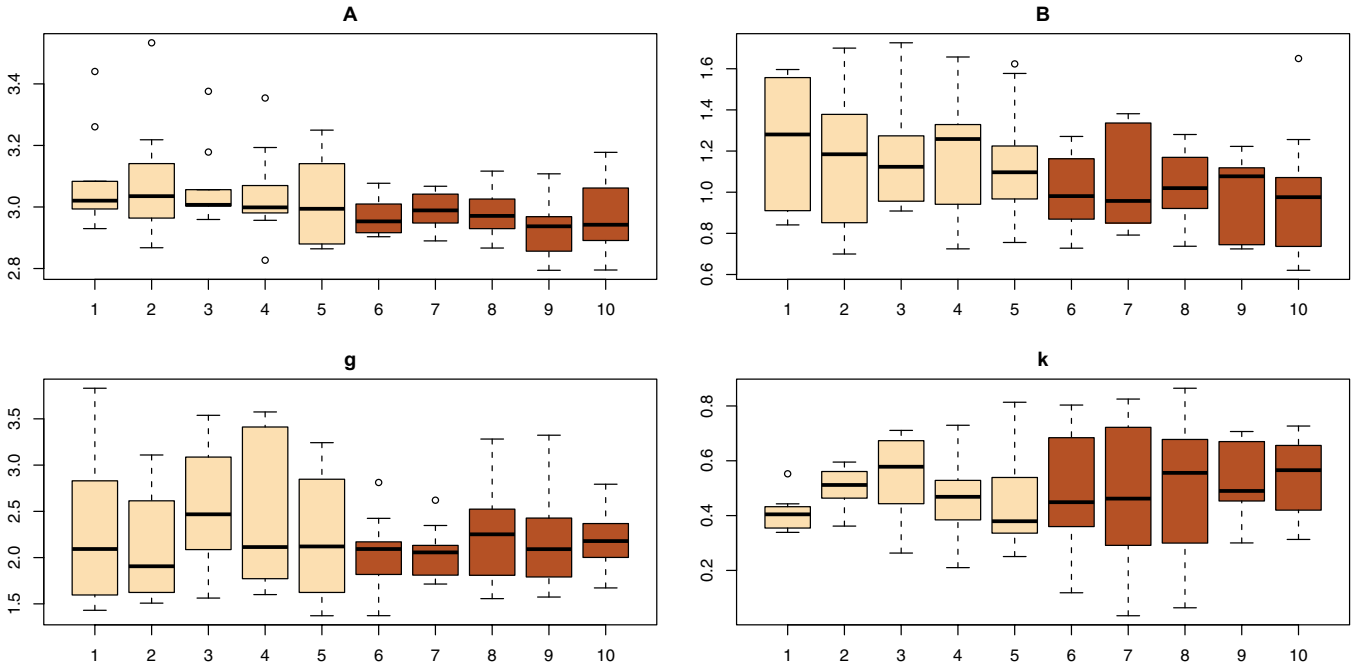

Fig. S4. Boxplots of the variations of the posterior means of the four parameters of the *g*-and-*k* distribution, based on BC_{el} approximations, for $n = 100$ observations (1–5) and $n = 500$ observations (6–10), based on $M = 10^4$ simulations and 10 replications. The moment conditions used in the BC_{el} algorithm are quantiles of order $(0.2, 0.5, 0.8)$ (1, 6), $(0.2, 0.4, 0.6, 0.8)$ (2, 7), $(0.1, 0.4, 0.6, 0.9)$ (3, 8), $(0.1, 0.25, 0.5, 0.75, 0.9)$ (4, 9), and $(0.1, 0.2, \ldots .0.9)$ (5, 10).

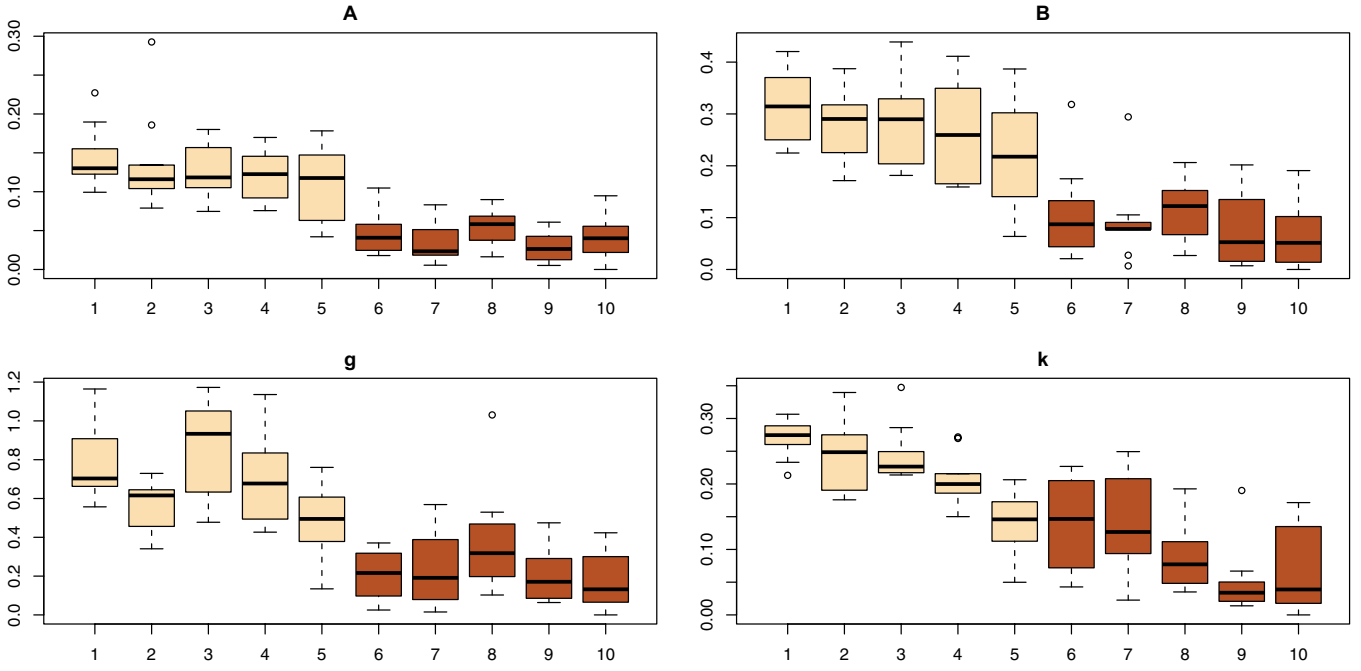

Fig. S5. Same graph as Fig. S4 for the SDs.

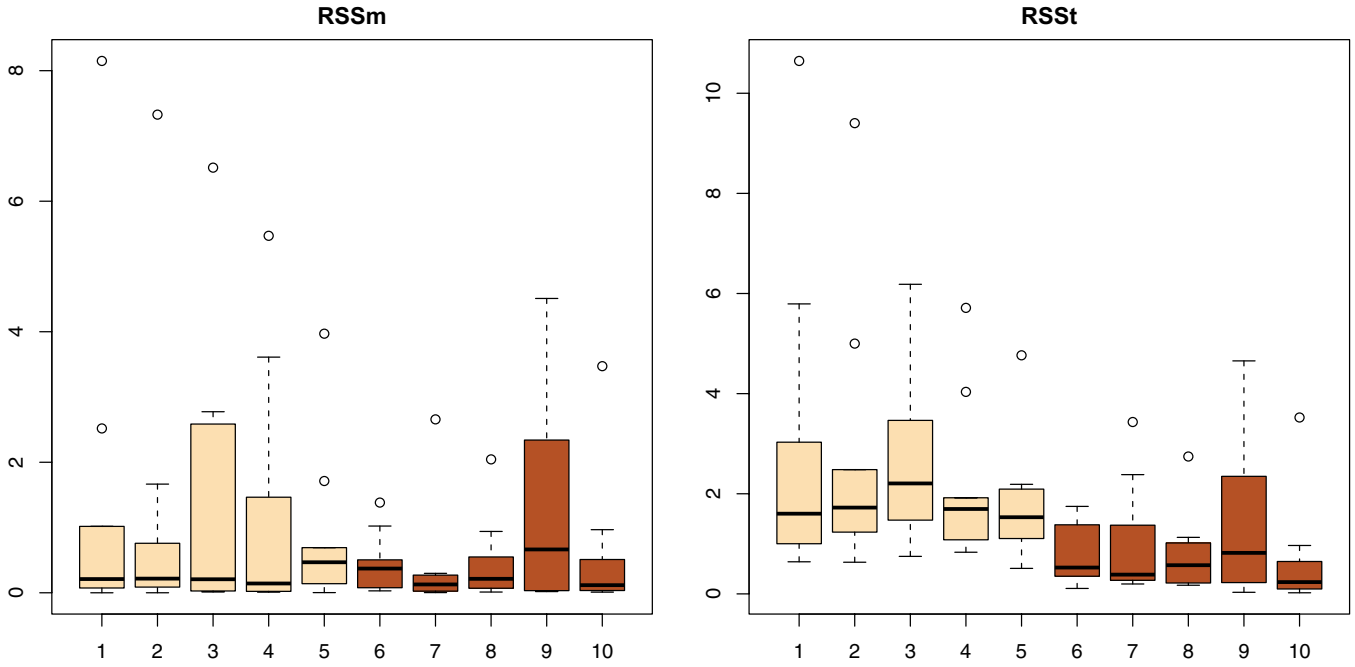

Fig. S6. Goodness-of-fit measures (RSSm, RSSt) for the same experiment as in Fig. S4.

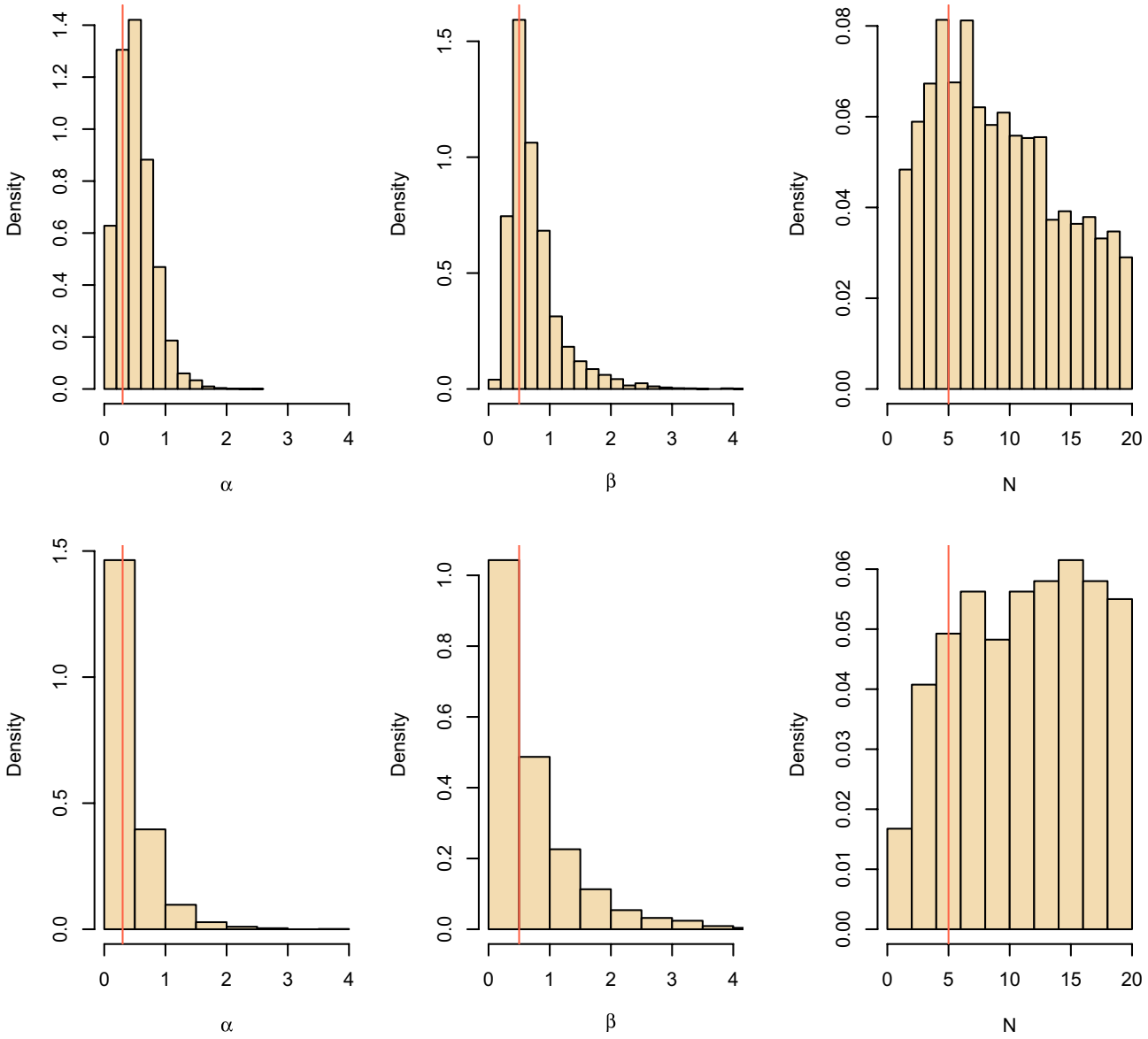

Fig. S7. Approximate posterior distributions of the parameters (α, β, N) for the superposition gamma process model, using a simulated dataset of 90 observations, with $\alpha = 0.5$, $\beta = 0.8$, and $n = 5$. (*Upper*) BC_{el} output; (*Lower*) ABC output.

Table S1. Computing times for DIYABC and BC_{el} in both population genetics experiments

	ABC	BC_{el}
Experiment	DIYABC software	BC_{el}–AMIS code
1	21 min	24 s
2	16 h	55 s