Supplementary material

Annex 1 : Computation of the probability of a scenario and its confidence interval through a weighted polychotomous logistic regression

Following the same rationale that introduced the local linear regression in the posterior distributions of parameters, we perform a weighted polychotomous logistic regression to estimate the posterior probability of scenarios. As in Beaumont *et al.* (2002) for the local linear regression, this analysis is performed using the frequentist paradigm.

Consider an analysis in which we want to evaluate the probability of K scenarios. We have a reference table of N simulated data sets and for each data set, there are J summary statistics, noted S_{ij}^k ($i \in [1, N]$, $j \in [1, J]$, $k \in [1, K]$) for the j-th summary statistic of the i-th data set simulated according to scenario k. Let $S_{obs,j}$ the j-th summary statistic of the observed data set. Let $\mathbf{Y}$ a matrix of N rows and K columns such that the element $y_{ik} = 1$ if the i-th data set has been simulated with scenario k and 0 otherwise. Eventually, let $\omega_i =$ the Epanechnikov weight given to the i-th data set, the same as in the local linear regression used for parameters (Beaumont *et al.*, 2002) but normalized such that $\sum_{i=1}^{N} \omega_i = N$.
We are looking for the relationship between the dependent variable p_{ik}, the posterior probability of scenario k for the data set i and the predictive variables $x_{ij} = S_{ij}^k - S_{obs,j}$.
Using a *logit* link function, the regression model used is:

$$log(\frac{p_{ik}}{1-p_{ik}}) = \beta_{k0} + \beta_{k1}x_{i1} + ... + \beta_{kJ}x_{iJ} = \beta_k^T \mathbf{x}_i$$

with $\beta_k = (\beta_{k0}, \beta_{k1}, ..., \beta_{kJ})$ and $\mathbf{x}_i = (1, x_{i1}, ..., x_{iJ})$

Because of the constraint $\sum_{k=1}^{K} p_{ik} = 1$, we consider only the first $K-1$ scenarios, the posterior probability of the last one being deduced from the others, i.e.:

$$p_k = Pr(Scen = k|X = \mathbf{x}) = \frac{e^{\beta_k^T \mathbf{x}}}{1 + \sum_{l=1}^{K-1} e^{\beta_l^T \mathbf{x}}}$$

for $k = 1, ..., K-1$, and

$$p_K = Pr(Scen = K|X = x) = \frac{1}{1 + \sum_{l=1}^{K-1} e^{\beta_l^T \mathbf{x}}}$$

This regression provides an estimate of the probability $p_{obs,k}$ that the observed data set has been obtained with scenario k, given that the actual scenario is one of the K. Since by definition all $x_{obs,j}$ are zero, the estimates of $p_{obs,k}$ simplify to:

$$\hat{p}_{obs,k} = \frac{e^{\hat{\beta}_{k0}}}{1+\sum_{l=1}^{K-1} e^{\hat{\beta}_{l0}}} \quad \text{for } k \in [1, K-1], \text{ and} \quad \hat{p}_{obs,K} = \frac{1}{1+\sum_{l=1}^{K-1} e^{\hat{\beta}_{l0}}} \qquad (1)$$

Noting $\beta_K = (0, 0, ..., 0)$, the log-likelihood:

$$log(lik) = \sum_{i=1}^{N} \omega_i \left\{ \sum_{k=1}^{K} y_{ik} log\left(\frac{e^{\beta_k^T \mathbf{x}_i}}{1 + \sum_{l=1}^{K-1} e^{\beta_l^T \mathbf{x}}} \right) \right\}$$

Let call β the vector $(\beta_{10}, \beta_{11}, ...\beta_{1J}, \beta_{20}, ..., \beta_{2J}, ..., ..., \beta_{K-1,0}, ..., \beta_{K-1,J})$. The Newton Raphson algorithm is such that starting from an arbitrary value of β, the following equation is repeated until $\beta^{new} - \beta^{old} < \epsilon$ for some tolerance level :

$$\beta^{new} = \beta^{old} + \left(\frac{\partial^2 log(lik)}{\partial\beta_u \partial\beta_u} \right)^{-1} \times \left(\frac{\partial log(lik)}{\partial\beta_u} \right)$$

Furthermore, $\hat{\beta} \approx N(\beta, V_\beta)$ where $V_\beta = \left(\frac{\partial^2 log(lik)}{\partial\beta_u \partial\beta_u} \right)^{-1}$.
Because only $\hat{\beta}_{k0}$ are useful in formula (1), let define β^* the sub-vector $(\beta_{10}, \beta_{20}, ..., \beta_{K-1,0})$, and V^* the

corresponding sub-matrix of V_β.
Noting $d = 1 + \sum_{l=1}^{K-1} e^{\beta_l^*}$, equation (1) reduces to

$$\hat{p}_{obs,k} = \frac{e^{\beta_k^*}}{d} \qquad 1 \leq k \leq K-1 \qquad \hat{p}_{obs,K} = \frac{1}{d} \tag{2}$$

According to Slutsky's formula, $\hat{p_k} \approx N(p_k, B_k^T V^* B_k)$, with $B_k = (\frac{\partial p_k}{\partial \beta_1^*}, \frac{\partial p_k}{\partial \beta_2^*}, ..., \frac{\partial p_k}{\partial \beta_{K-1}^*})$
Elements of B_k being the following:

$$\begin{array}{ll} \frac{\partial p_k}{\partial \beta_l^*} = \frac{e^{\beta_k^*}(d - e^{\beta_k^*})}{d^2} & \text{if } l = k \text{ and } k < K \\ \frac{\partial p_k}{\partial \beta_l^*} = \frac{-e^{\beta_k^*} e^{\beta_l^*}}{d^2} & \text{if } l \neq k \text{ and } k < K \\ \frac{\partial p_K}{\partial \beta_l^*} = \frac{-e^{\beta_K^*}}{d^2} & \text{if } k = K \end{array}$$

it is straighforward to derive approximate C.I. for $p_{obs,k}$.

Annex 2 : Measures of dispersion computed in *DIYABC*

the Relative square Root of the Mean Square Error (RRMSE) : the square root of the average square difference between the point estimate (e) and the true value, divided by the true value, $\frac{1}{v}\sqrt{\frac{1}{n}\sum(e-v)^2}$

the Relative square Root of the Mean Integrated Square Error (RRMISE) : the square root of the average (over test data sets) of the integrated square error (measured on each test data set) divided by the true value, $\frac{1}{v}\sqrt{\frac{1}{n}\sum(\frac{1}{n_x}\sum(x-v)^2)}$. x is a value of the sample of the posterior distribution for a given test data set, n_x being the sample size. n is the number of test data sets and v is the true value.

the Relative Mean Absolute Deviation (RMAD) : the average (over test data sets) of the mean absolute deviation (measured on each data set), divided by the true value, $\frac{1}{v}\frac{1}{n}\sum(\frac{1}{n_x}\sum|x-v|)$

the 50 and 95% coverages : the proportion of test data sets for which the 50% and 95% credibility intervals include the true value.

the factor 2 : the proportion of test data sets for which the point estimate is at least half and at most twice the true value.

Table S1 : First example: properties of the approximate posterior distribution of parameters under scenario 1

Parameter	mean	median	mode	$Q_{0.025}$	$Q_{0.050}$	$Q_{0.950}$	$Q_{0.975}$
N (1,000)	1,684	1,425	1,050	659	729	3,422	4,320
r_1 (0,6)	0.577	0.577	0.577	0.495	0.512	0.642	0.658
r_2 (0,4)	0.499	0.500	0.547	0.107	0.162	0.828	0.880
t_1 (10)	52	52	52	4	7	95	97
t_2 (500)	742	797	979	227	303	984	992
t_3 (10,000)	12,324	10,126	6,538	5,263	5,526	27,040	32,334
t_4 (20,000)	33,068	34,186	43,836	11,414	13,867	48,402	49,244
t_5 (200,000)	315,213	323,636	335,525	95,697	120,911	481,411	490,450
$\bar{\mu}$ (0.0005)	0.00056	0.00055	0.00053	0.00018	0.00022	0.00091	0.00095
$\bar{P}$ (0.22)	0.24	0.25	0.28	0.14	0.15	0.30	0.30
θ (=$4N\bar{\mu}$)	3.21	2.99	2.83	1.73	1.91	5.17	5.93
τ_1 (=$t_1\bar{\mu}$)	0.028	0.025	0.015	0.002	0.003	0.065	0.072
τ_2 (=$t_2\bar{\mu}$)	0.40	0.39	0.41	0.10	0.14	0.71	0.77
τ_3 (=$t_3\bar{\mu}$)	6.57	5.42	4.41	1.90	2.30	14.82	18.45
τ_4 (=$t_4\bar{\mu}$)	18.2	16.7	11.3	4.7	5.9	35.4	39.0
τ_5 (=$t_5\bar{\mu}$)	173	156	111	44	54	343	380

Mean, median, mode and quantiles of the posterior distribution sample for original and composite parameters of the simulated data set. True values of original parameters (used to simulate the data set) are given between parentheses in the first column. Results have been obtained with uniform priors on parameters (see text and Figures S3 and S4).

Table S2 : First example: bias and precision of parameter estimation under scenario 1

Uniform priors

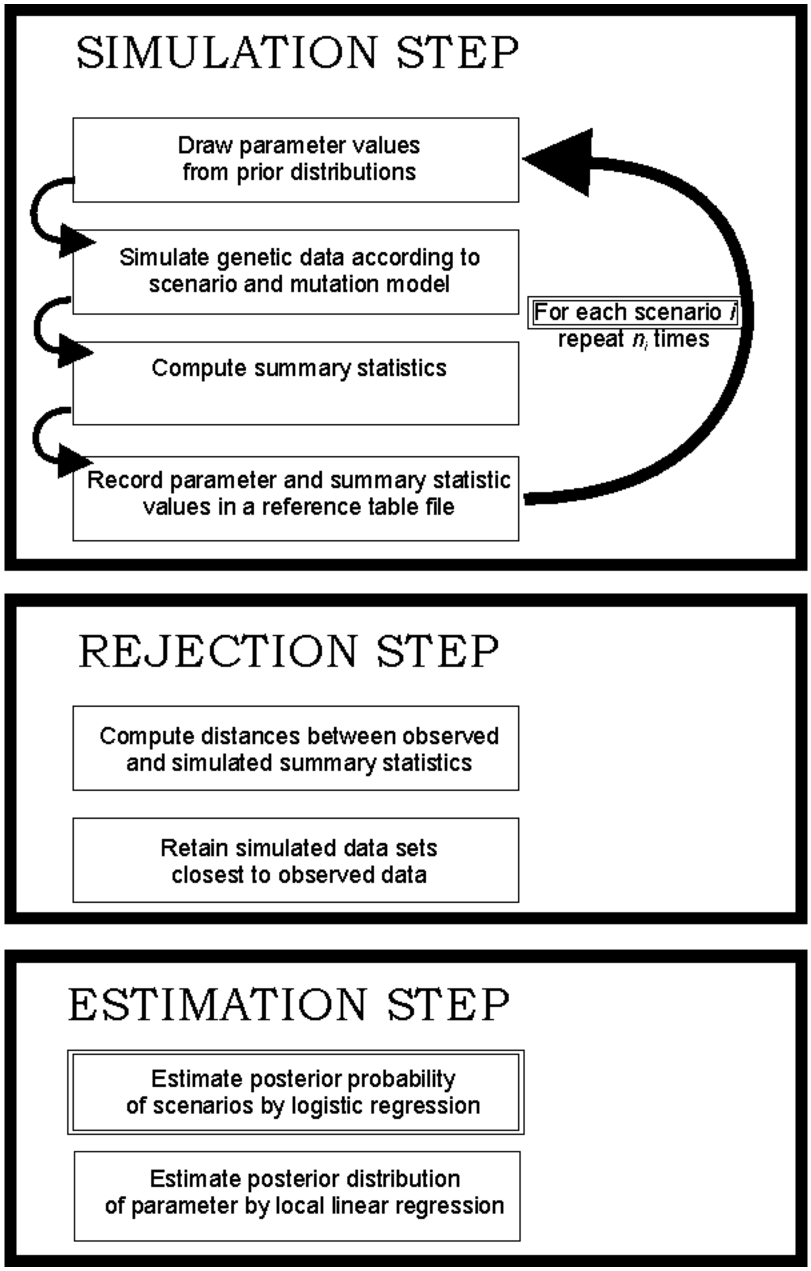
		Posterior distribution				Posterior median		
Parameter	true value	RRMISE	RMAD	50% cov.	95% cov.	ARB	RRMSE	fact2
N	1,000	1.188	0.752	0.46	0.99	0.431	0.588	0.91
r_1	0.6	0.103	0.079	0.56	0.98	-0.022	0.063	1.00
r_2	0.4	0.658	0.555	0.57	0.98	0.034	0.468	0.86
t_1	10	4.661	3.787	0.02	1.00	3.521	3.690	0.01
t_2	500	0.519	0.444	0.82	1.00	0.099	0.285	1.00
t_3	10,000	1.475	1.130	0.16	1.00	0.903	0.984	0.57
t_4	20,000	0.944	0.836	0.01	0.97	0.876	0.8886	0.82
t_5	200,000	0.765	0.635	0.56	1.00	0.424	0.514	1.00
$\bar{\mu}$	0.0005	0.459	0.393	0.73	1.00	-0.151	0.273	0.95
$\bar{P}$	0.22	0.233	0.206	0.26	1.00	0.181	0.192	1.00
θ (=$4N\bar{\mu}$)	2	0.496	0.334	0.78	1.00	0.117	0.174	1.00
τ_1 (=$t_1\bar{\mu}$)	0.005	4.687	3.339	0.12	1.00	2.489	2.811	0.09
τ_2 (=$t_2\bar{\mu}$)	0.25	0.635	0.503	0.60	1.00	-0.134	0.363	0.88
τ_3 (=$t_3\bar{\mu}$)	5	1.547	1.021	0.54	1.00	0.506	0.706	0.83
τ_4 (=$t_4\bar{\mu}$)	10	1.121	0.811	0.56	1.00	0.448	0.603	0.90
τ_5 (=$t_5\bar{\mu}$)	100	0.927	0.669	0.81	1.00	0.090	0.373	0.95

Log-uniform priors

		Posterior distribution				Posterior median		
Parameter	true value	RRMISE	RMAD	50% cov.	95% cov.	ARB	RRMSE	fact2
N	1,000	0.537	0.378	0.66	1.00	-0.126	0.240	1.00
r_1	0.6	0.097	0.074	0.57	0.99	-0.003	0.058	1.00
r_2	0.4	0.676	0.571	0.50	0.99	0.071	0.497	0.84
t_1	10	2.315	1.457	0.95	1.00	0.771	0.520	0.79
t_2	500	0.514	0.453	0.51	1.00	0.312	0.388	0.77
t_3	10,000	0.833	0.562	0.98	1.00	0.063	0.187	1.00
t_4	20,000	0.701	0.569	0.80	1.00	0.312	0.372	1.00
t_5	200,000	0.620	0.520	0.82	1.00	-0.119	0.308	0.92
$\bar{\mu}$	0.0005	0.499	0.418	0.66	1.00	0.225	0.304	0.95
$\bar{P}$	0.22	0.320	0.279	0.29	1.00	-0.253	0.278	1.00
θ (=$4N\bar{\mu}$)	2	0.338	0.252	0.87	1.00	-0.010	0.127	1.00
τ_1 (=$t_1\bar{\mu}$)	0.005	2.994	1.733	0.96	1.00	0.090	0.661	0.79
τ_2 (=$t_2\bar{\mu}$)	0.25	0.630	0.506	0.56	1.00	-0.182	0.378	0.85
τ_3 (=$t_3\bar{\mu}$)	5	1.223	0.768	0.74	1.00	0.257	0.396	0.98
τ_4 (=$t_4\bar{\mu}$)	10	1.149	0.840	0.61	1.00	0.460	0.592	0.92
τ_5 (=$t_5\bar{\mu}$)	100	0.926	0.669	0.77	1.00	-0.007	0.376	0.94

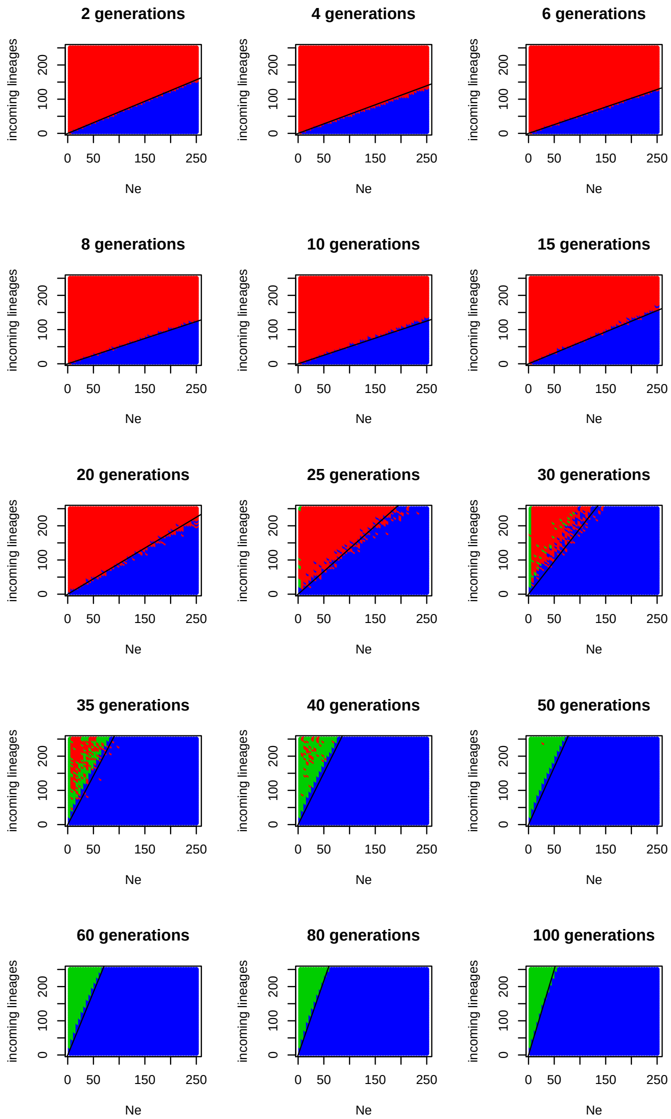
Extract of the output of *DIYABC*/option "Compute bias and mean square error" with scenario 1 and parameter values identical to those used to simulate the exemple data set. Measures relative to the posterior distribution of parameters are the relative square root of the mean integrated square error (RRMISE), the relative mean absolute deviation (RMAD) and the proportion of times where the true value is within the 50% and 95% credibility intervals (*50% cov* and *95% cov.*, respectively). Precision measures are also given for the posterior median as a point estimate : the average relative bias (ARB), the relative square root of the mean square error (RRMSE) and the proportion of times the posterior median is between the half and the double of the true value (*fact2*). Values have been obtained with 500 test data sets simulated with parameter values as shown in the second column.

Figure S1



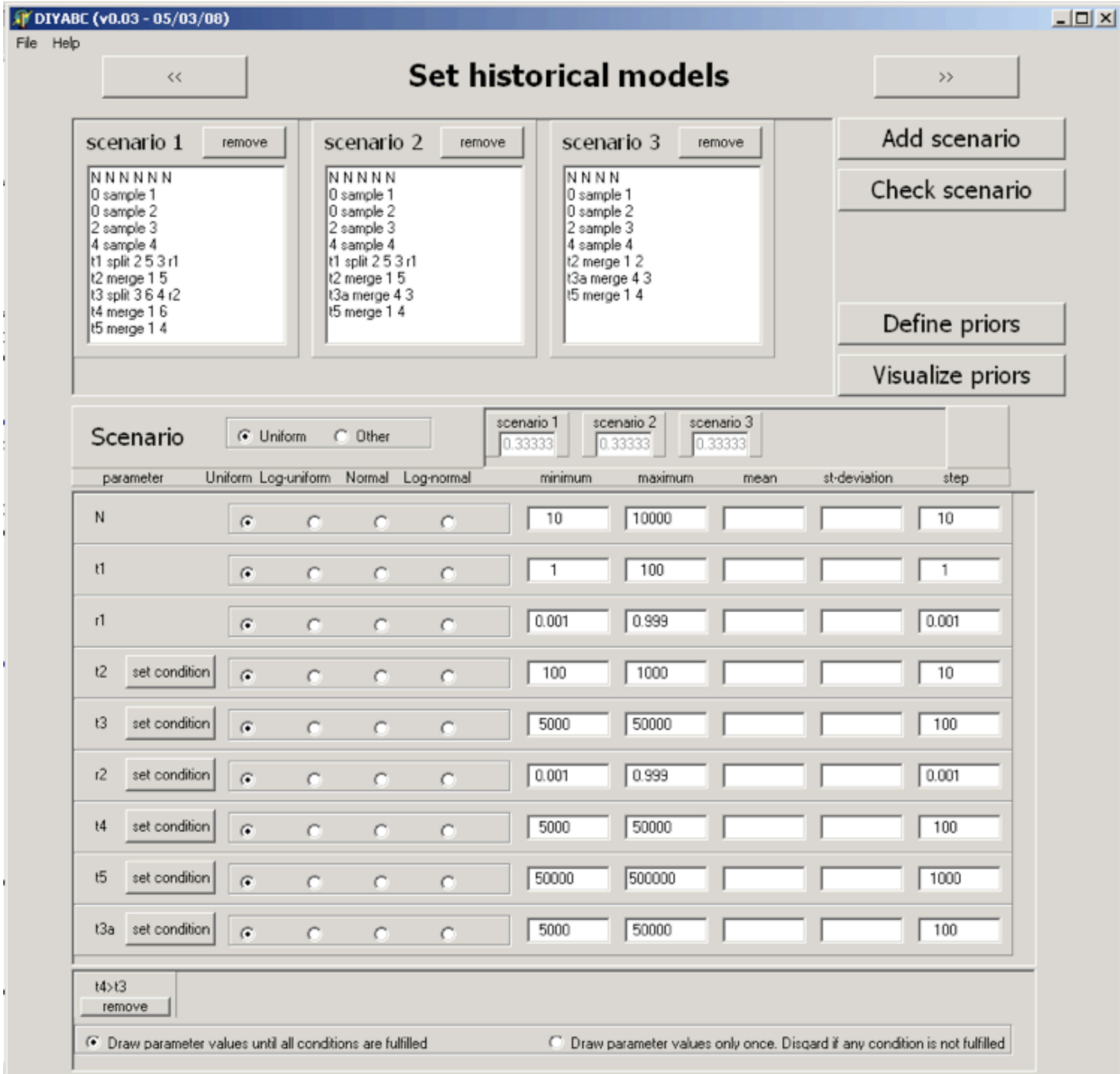
The three steps of an ABC analysis. The two boxes with a double line correspond to the case where there is more than one scenario.

Figure S2



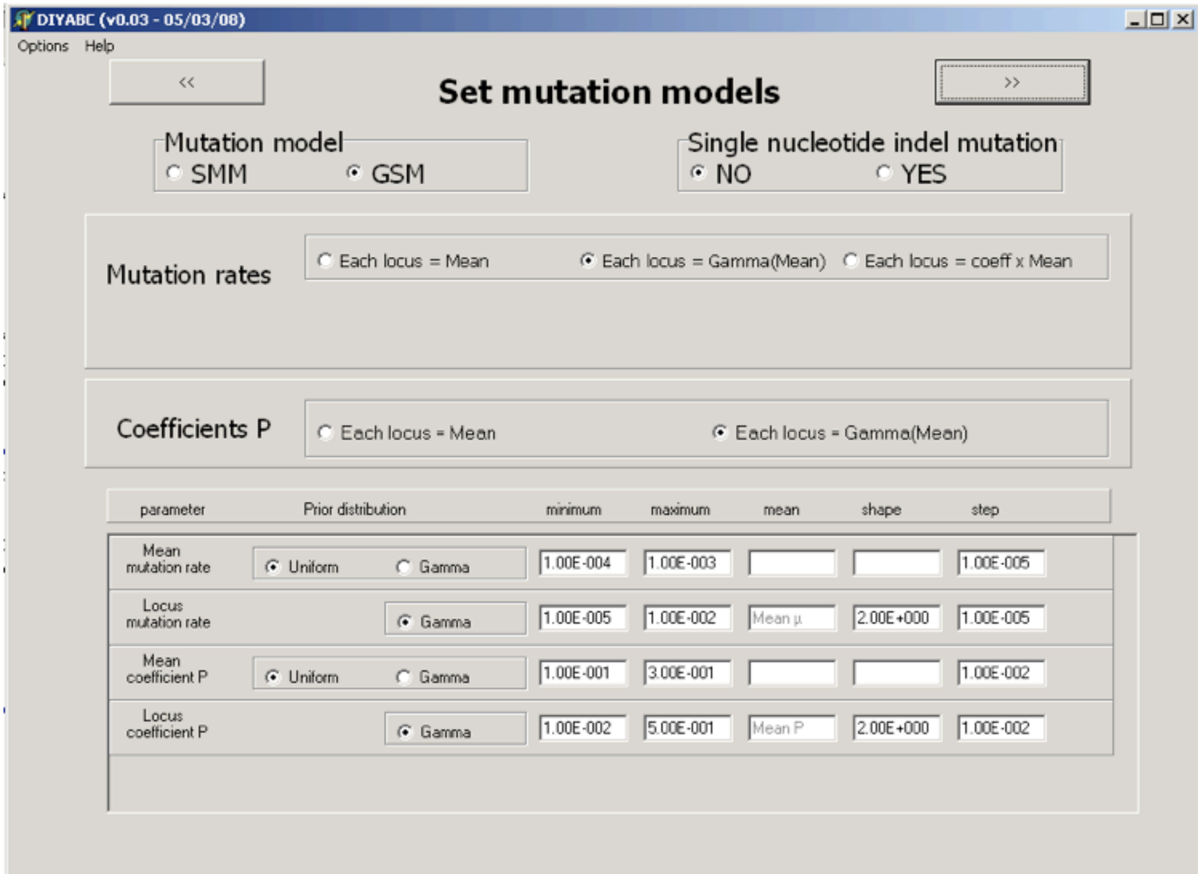
Graphs indicate in green the area of the plane for which the generation by generation (GbG) algorithm is faster than the continuous time (CT) algorithm, in red the area for which the CT algorithm produces significantly (5%) less coalescences than the GbG algorithm and in blue the area for which the CT algorithm produces the same number of coalescences than the GbG algorithm (with tolerance=5%) and is faster. Limits between areas are almost linear. The black line (intercept=0) has a slope taken as $0.0031g^2 - 0.053g + 0.7197$ for $g \leq 30$, $0.033g + 1.7$ for $30 < g \leq 100$ and 5 when $100 < g$, g being the duration of the coalescence module in number of generations. N_e is the diploid effective population size.

Figure S3



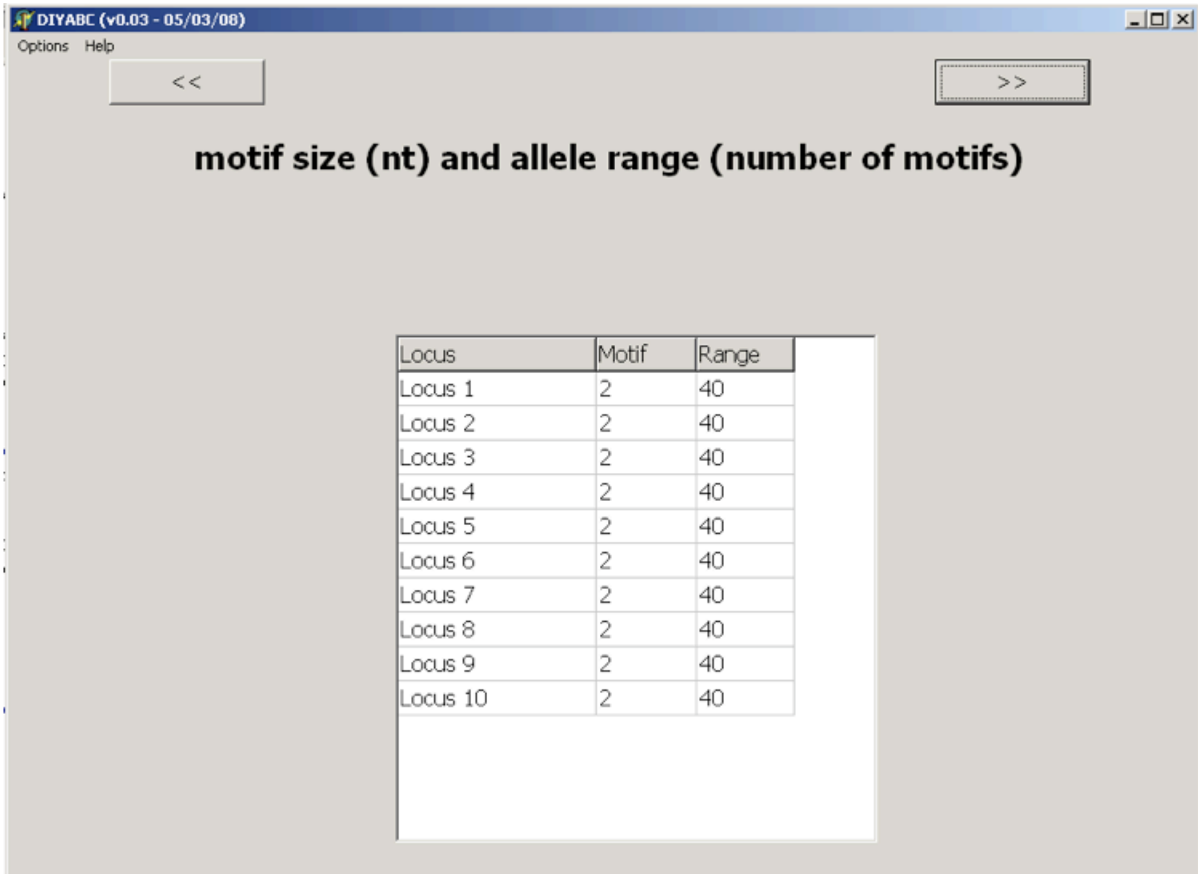
Screenshot from *DIYABC*. This screen is used to input coded scenarios and prior distributions of historic and demographic parameters. The first line of each scenario gives the effective size of all (sampled and unsampled) populations. The following lines all start with the time of the event, the nature of the event and additional data depending on the nature of the event. Default prior distributions are Uniform but other prior probabilities or distributions are possible, for scenarios and parameters respectively. When there are more than two effective sizes or times of event, it is possible to set a binary condition between them. For instance here, the condition $t_4 > t_3$ has been set and only the draws in which this condition is fulfilled will be considered when silmulating data sets. This, of course, will change the shape of the prior distributions for both parameters (see Figure S8).

Figure S4



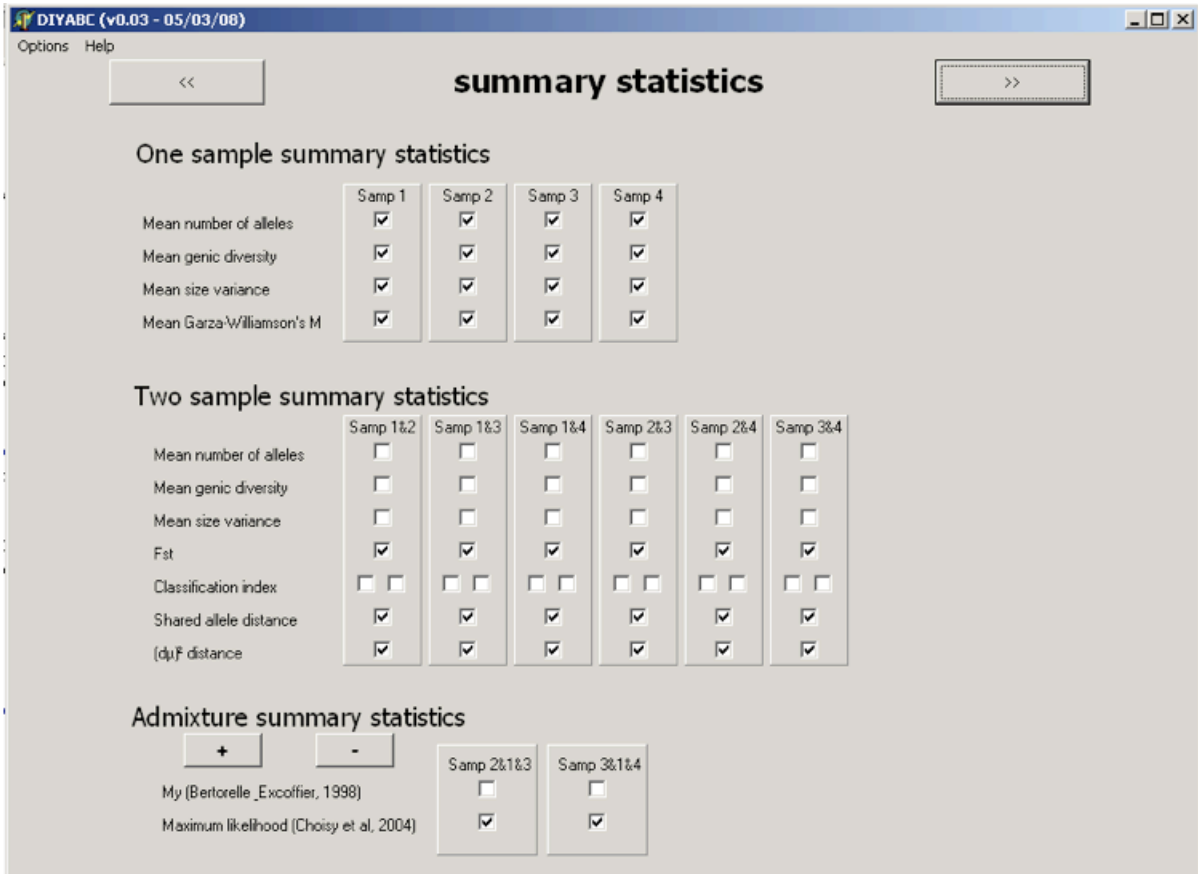
Screenshot from *DIYABC*. This screen is used to input the mutation model and the prior distributions of mutation parameters. The Single nucleotide indel mutation allows to consider possible allele sizes that do not fit with a unique scale of repeated motifs. For instance, suppose a dinucleotide microsatellite locus in which some allele lengths in bp are odd and some are even. This can be explained only if there has been an indel mutation which length is not a multiple of the motif length. To cope with not so rare sitauation, the single nucleotide indel mutation has been implemented in *DIYABC*. But for a simulated data set as in our example, this option is meaningless.

Figure S5



Screenshot from *DIYABC*. This screen is used to input the motif size and allele range of microsatellite loci.

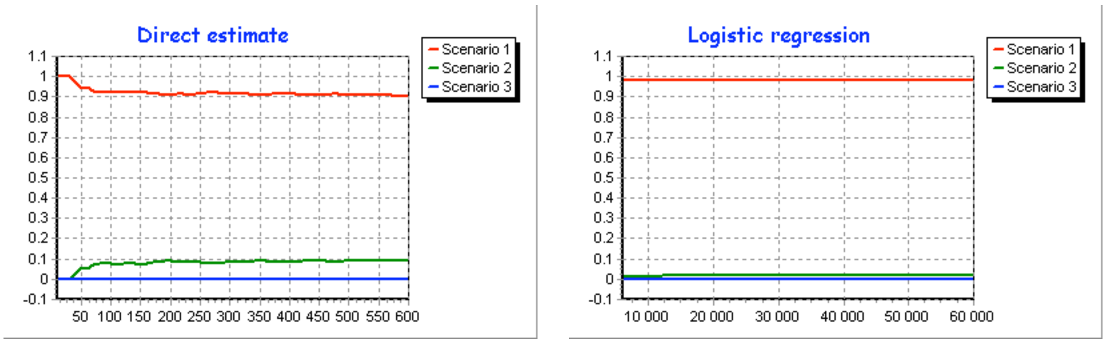
Figure S6



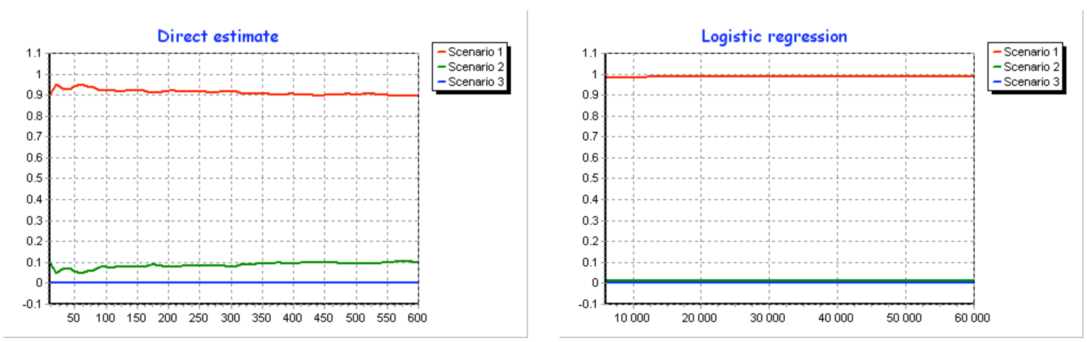
Screenshot from *DIYABC*. This screen is used to choose which summary statistics are chosen for the reference table. In this example, 36 summary statistics will be computed for each set.

Figure S7

Uniform priors

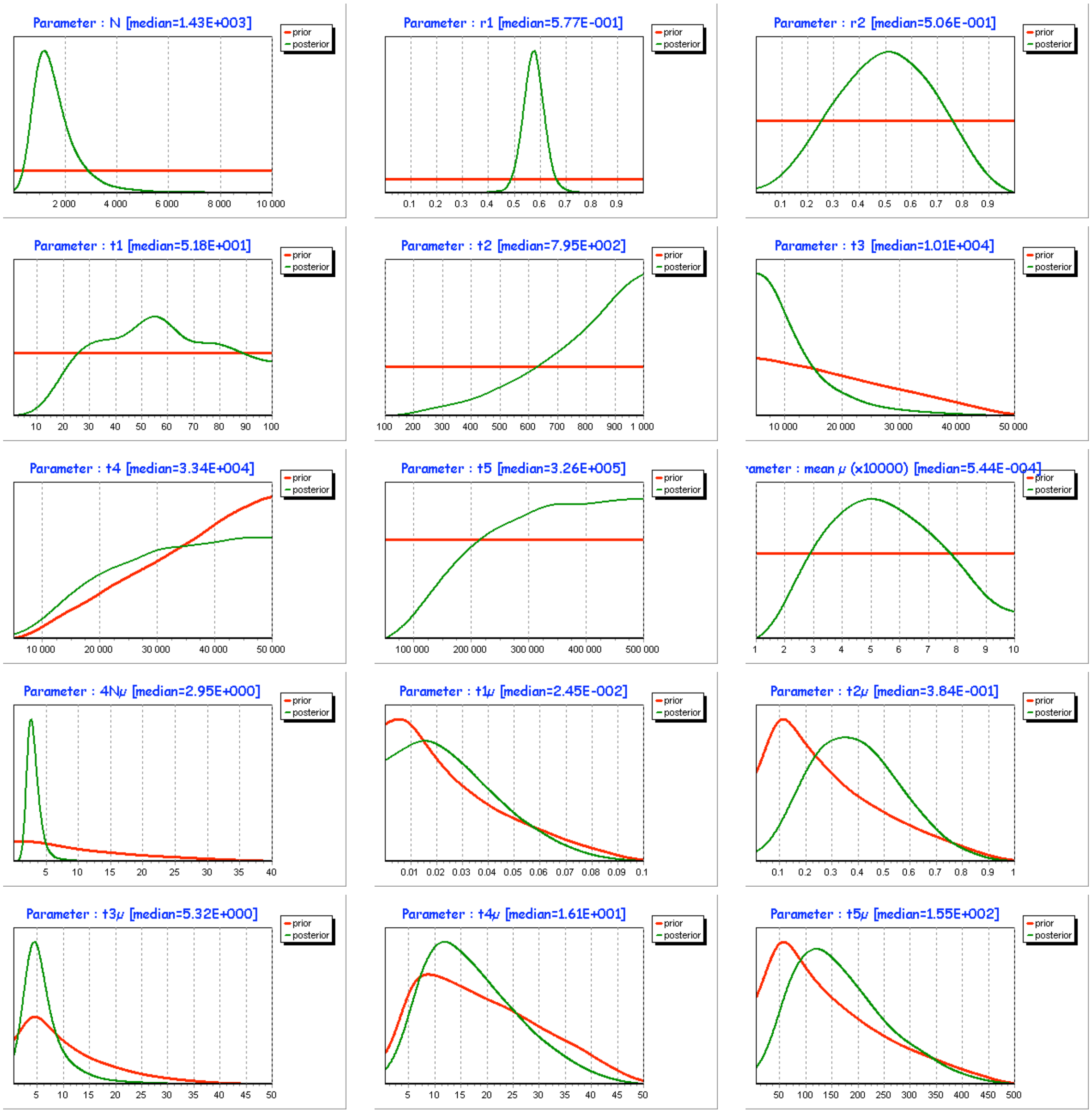


Log-uniform priors



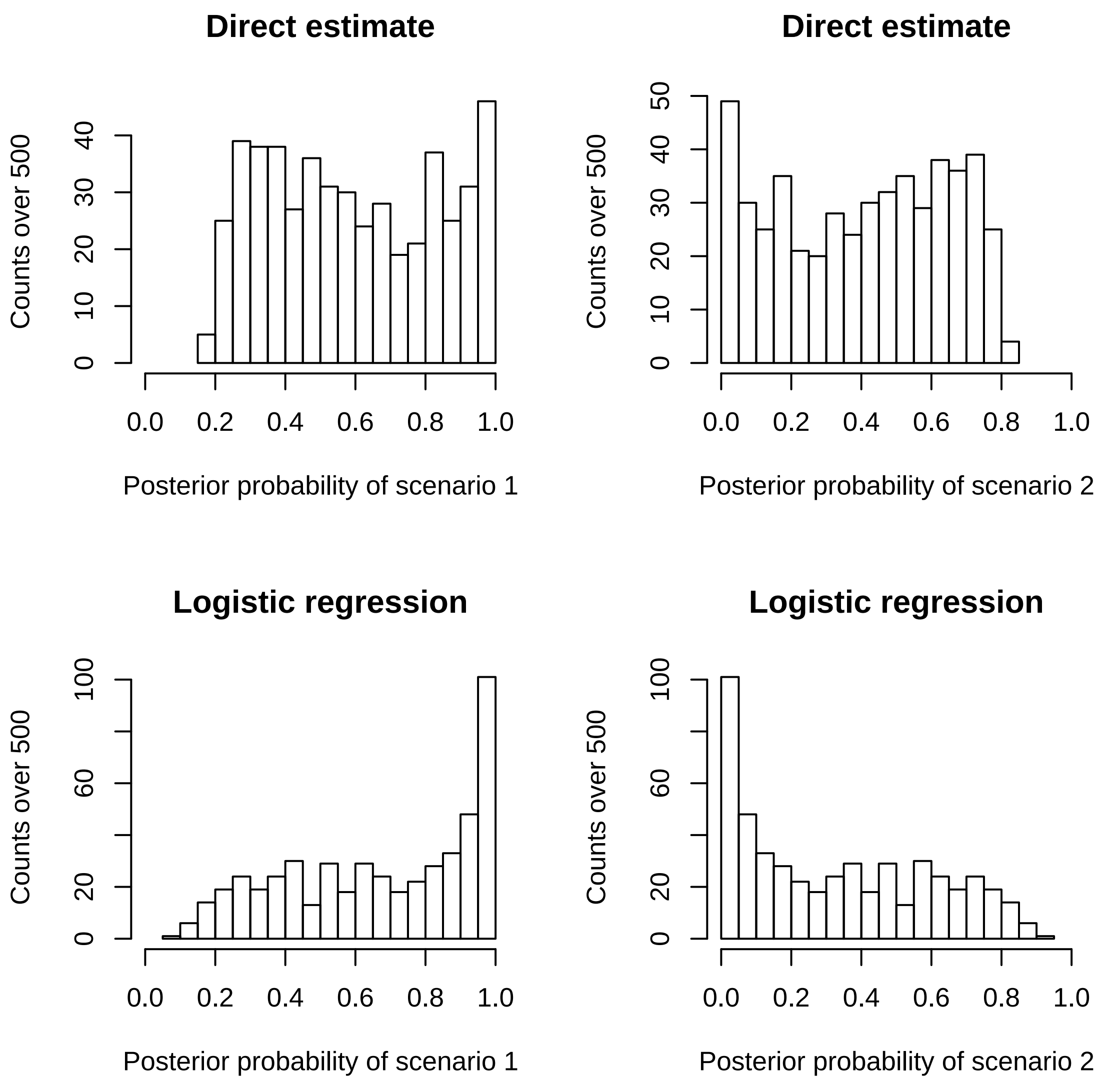
First example: plots output by DIYABC showing the posterior probability (y-axis) of the three scenarios through the *direct* (left) and the *logistic* (right) approaches as output by *DIYABC*. The x-axis correspond to the different n_δ values used in the computations. On upper row, results have been obtained with Uniform priors on parameters (as described in the text and in Figures S3 and S4). On bottom row, all effective size and time of event parameters (N, t_1, t_2, t_3, t_{3a} and t_4) have been given Log-Uniform priors, keeping the same bounds.

Figure S8



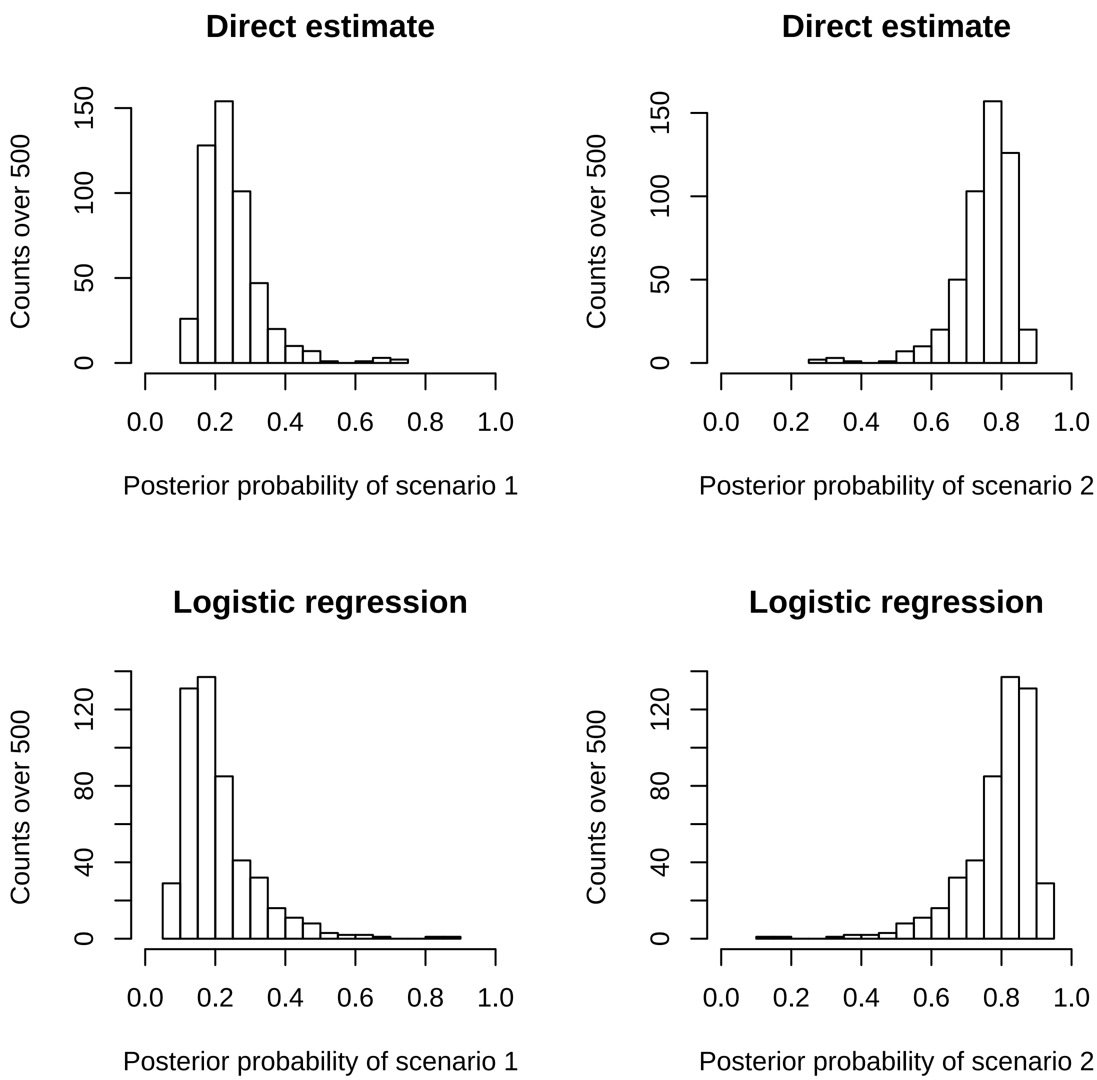
First example: estimation of parameters under scenario 1. This is the graphical output obtained with *DIYABC*. It shows prior (in red) and posterior (in green) distributions of original and composite parameters. Note the result of setting condition $t_4 > t_3$ on the prior distributions of these two parameters. All posterior distributions as well as prior distributions of composite or "conditionned" (here t_3 and t_4) are computed using a Gaussian kernel density.

Figure S9



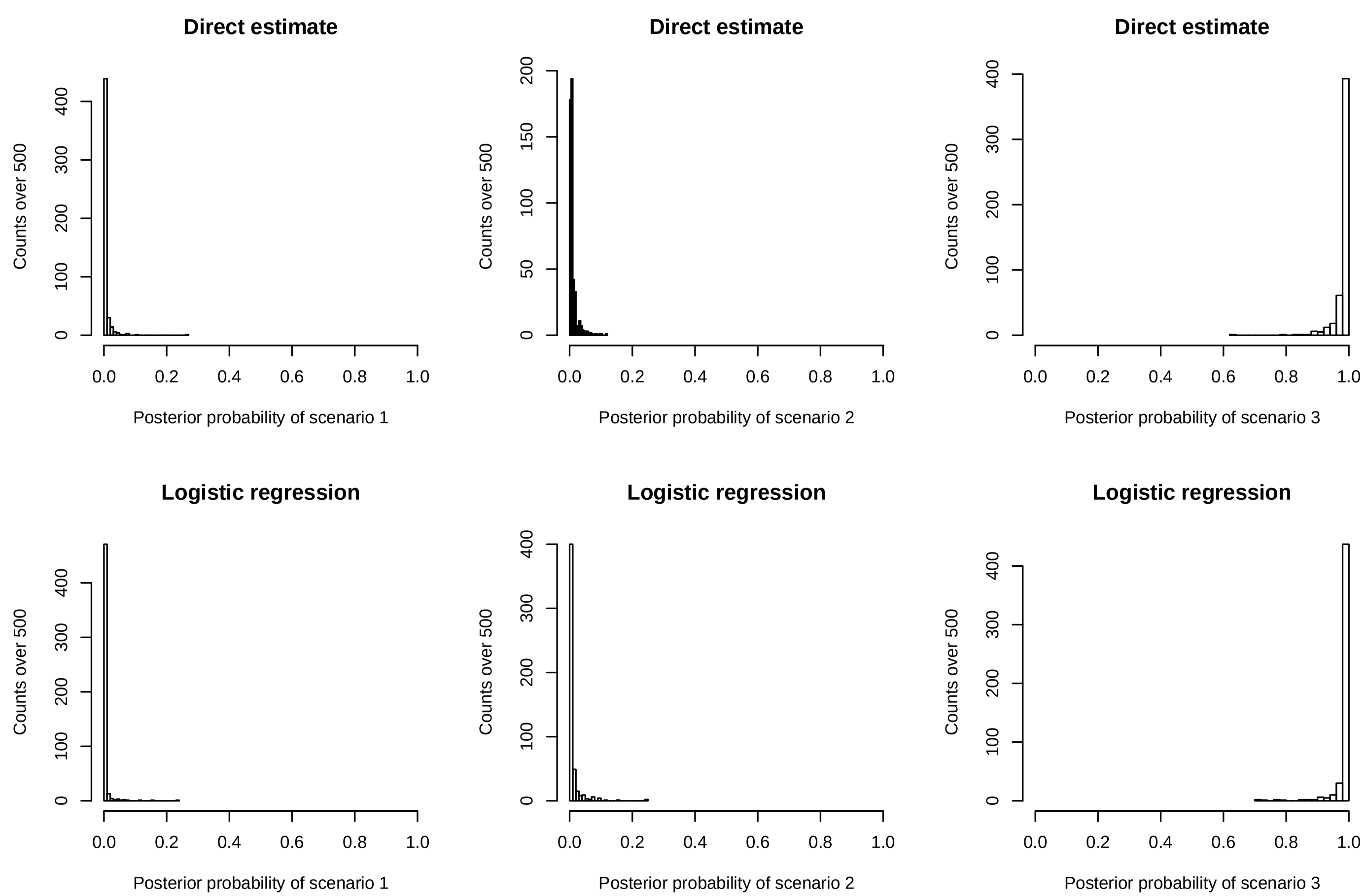
First example: Distribution of posterior probabilities of scenarios 1 and 2 obtained using 500 test data sets simulated with scenario 1 and known values of parameters (i.e. those used to produce the original data set and given between parenthesis in table S1). The histograms for scenario 3 are not shown because the posterior probability of this scenario is very rarely above 0 (a single time above 0.005 with the direct approach and always below 10^{-5} with the logistic regression). Very similar histograms have been obtained, using log-uniform distributions with the same bounds for effective size and time of event parameters when building the reference table.

Figure S10



First example: Distribution of posterior probabilities of scenarios 1 and 2 obtained using 500 test data sets simulated with scenario 2 and known values of parameters (i.e. those used to produce the original data set). The histograms for scenario 3 are not shown because the posterior probability of this scenario is very rarely above 0 (maximum values of 0.01 and 3 10^{-6} with the direct approach and the logistic regression respectively). Very similar histograms have been obtained, using log-uniform distributions for effective size and time of event parameters when building the reference table.

Figure S11



First example: Distribution of posterior probabilities of scenarios 1, 2 and 3 obtained using 500 test data sets simulated with scenario 3 and known values of parameters (i.e. those used to produce the original data set). Very similar histograms have been obtained, using log-uniform distributions for effective size and time of event parameters when building the reference table.