A Comparison on Consistency of Parameter Estimation Using Optimization Methods for a Mixture

Marianne C. Turley   E. David Ford

NRCSE

Technical Report Series

NRCSE-TRS No. 050

July 17, 2000

The **NRCSE** was established in 1996 through a cooperative agreement with the United States Environmental Protection Agency which provides the Center's primary funding.

EPA

NRCSE Report 50

A COMPARISON ON CONSISTENCY OF PARAMETER ESTIMATION USING OPTIMIZATION METHODS FOR A MIXTURE MODEL

MARIANNE C. TURLEY[1] AND E. DAVID FORD[2]

[1] *Quantitative Ecology and Resource Management, University of Washington,*

Box 351720, Seattle, Washington 98195-1720 USA

[2] *Department of Forestry, University of Washington,*

Box 352100, Seattle, Washington 98195-2100 USA

ABSTRACT

Many ecological and environmental studies result in bimodal data. Such data may be modeled as a mixture of two normally distributed populations. Model parameters must be estimated using numerical optimization. We tested three optimization algorithms, expectation-maximization (EM), minimizing-sums (MS), and evolutionary computation (EC), on a data set generated from known parameters. All of the algorithms tended to be biased. Comparisons were then made on the algorithms for consistency using assessment statistics on the sample distribution of the known parameters (30 samples). We found the evolutionary computation algorithm, when optimized on multiple objective functions, to have the most strongly consistent estimators of the algorithms we tested. We also recommend this algorithm for analyzing bimodal data from ecological and environmental studies because it allows one to include additional information, such as expert knowledge on a particular parameter of interest, in the optimization process.

Keywords: mixture models, consistency, expectation-maximization, evolutionary computation, multiple objective function optimization

1. INTRODUCTION

Many ecological and environmental studies result in bimodal data: for example, plant competition for light (Ford 1975), echinoderm egg size production (Sewell and Young 1997), Atlantic salmon growth (Skilbrei, Hansen and Stefansson 1997), and female bank vole reproduction (Tkadlec and Zejda 1998). Such data are often modeled as a mixture of two normally distributed populations. The density function of this mixture model is:

$$f(\mathbf{x};\theta) = \pi \; f(\mathbf{x};\mu_1,\sigma_1) + (1-\pi) \; f(\mathbf{x};\mu_2,\sigma_2) \qquad (1)$$

where $\mathbf{x}=x_1,...,x_n$ are the n observations in the data set and $\theta = (\mu_1,\mu_2,\sigma_1,\sigma_2,\pi)$ are the model parameters. $f(\cdot)$ is the normal density function with mean, μ_i, and standard deviation, σ_i, where i=1 for the first and i=2 for the second population (Everitt and Hand 1981). π is the mixing parameter. Ideally, a likelihood equation is derived from (1) and maximized for parameter estimation. However, for this model, the likelihood equation is unbounded and may have multiple roots, so a maximum likelihood may not exist (McLachlan and Basford 1988). Instead the parameters of the mixture model may be solved for using numerical optimization.

Many parameter estimation methods for mixture models by numerical optimization have been presented and compared in the literature (e.g., Everitt 1984, Woodward et al. 1984, McLachlan and Basford 1988). Two common methods involve some variation of either the expectation maximization (EM) algorithm (Dempster, Laird and Rubin 1977) or a gradient-search algorithm such as the minimizing sums (MS) function in Splus (Chambers and Hastie 1993). In this paper, we examine evolutionary computation (EC) as an alternative optimization method (Srinivas and Deb 1995). In section 2, we discuss the three algorithms. In section 3, details are given for testing the algorithms on a single sample, a data set simulated from known parameters. In section 4, the results are given.

It has been shown that likelihood estimators are strongly consistent (Keifer and Wolfowitz 1956), but if one has only a single sample of a limited size, which algorithm produces the most consistent estimates? To assess which estimators were most consistent, the algorithms were tested on a small number of samples; 30 data sets were generated from the known parameters to create a sample distribution of the mixture model. The details of these tests are also discussed in section 3. In section 4, the estimators were evaluated by different assessment statistics of consistency; bias, standard error, mean squared error, and the ratio of bias over standard error. To improve the estimation by the EC optimization, EC allows for multiple objective function optimization (Reynolds and Ford 1999). We show the objective function optimization may improve the consistency of specific parameters of interest. We discuss the implications of these results in section 5.

2. OPTIMIZATION METHODS

2. 1. Expectation-maximization

There are many variations of the expectation-maximization (EM) optimization method (McLachlan and Krishnan 1997), but they all use some form of the method generalized by Dempster, Laird and Rubin (1977). For the mixture model, the EM computes maximum likelihood estimates (MLEs) analytically derived from the mixture model rewritten as a *complete data* problem. The likelihood equation for the complete data problem is:

$$L(\mathbf{y}\theta) = \prod_{i=1}^{n} (\pi\ f(\mathrm{x_i}, \mu_1, \sigma_1))^{z_i}\ \ ((1-\pi)\ f(\mathrm{x_i}, \mu_2, \sigma_2))^{1-z_i} \tag{2}$$

where the complete data set, **y**, is the set which includes data that are observable, **x**, and data that are not observable (or latent), **z**. θ is equivalent to the parameter vector described for (1). For the mixture model, an unobservable datum, z_i, indicates whether an observation, x_i, is from the first or the second population.

To start the optimization, the EM algorithm calculates the conditional expectations of the unobservable data using starting parameter estimates, values supplied to EM, and the observed data. These expected values are then used in the analytical solutions of the MLEs, which result in new parameter estimates. The new estimates are used in the conditional expectations to calculate new expected values for the unobservable data. This sequence of expectation and maximization continues until convergence. The algorithm converges when iterative changes in the parameter estimates fall below a tolerance level (= 0.00001) or a maximum number of iterations has been performed (= 50).

2.2. Gradient-based methods

There are also many variations of gradient-based (a.k.a. Newton-type) optimization methods, but their common feature is the utilization of first and second derivatives of the objective function to direct the search. We used a gradient-based method available in Splus, the minimizing sums (MS) function (MathSoft 1995, Chambers and Hastie 1993). To use the MS function, the objective function must be written as a sums equation to be minimized. For the mixture model, the negative log-likelihood of (1) is an appropriate objective function and is:

$$-\log L(\mathbf{x},\theta) = \log(-\sum_{i=1}^{n}(\pi \ f(x_i;\mu_1,\sigma_1) + (1-\pi) \ f(x_i;\mu_2,\sigma_2))) \quad (3)$$

where **x** and θ are defined as above.

To start the optimization, the MS algorithm calculates the value of (3) at the starting parameter estimates, values supplied to MS. It calculates the gradient from this value using quadratic approximation (or by derivatives if they are supplied) based on an algorithm by Gay (1983). The gradient directs the search toward the optimal parameter estimates. The algorithm increments the current parameter estimates by some step size change in the direction of the gradient. It calculates (3) at the new parameter estimates. If (3)'s value decreases, MS recalculates the gradient and increments the current parameter estimates by some step size along the new gradient to produce new parameter estimates. It recalculates (3) at the new parameter estimates and so on. If (3)'s value increases or does not change, MS takes a smaller step size change of the parameter estimates and repeats the calculation of (3) with these estimates. The algorithm repeats these calculations of (3) and movements along the gradients to optimal parameter estimates at convergence. MS converges when any of several measures of the optimization process are satisfied, e.g., iterative changes in (3) or the parameter estimates fall below the tolerance level (= 0.001) or a maximum number of iterations has been performed (= 50).

2.3. Evolutionary computation

Evolutionary computation is a stochastic optimization method based on ideas of evolutionary theory; populations evolve through breeding techniques of crossovers and mutations where breeder selection depends on one's fitness (Michalewicz 1992). As an optimization method, a population is a set of parameter estimate vectors. Each parameter estimate vector, such as $\theta_j = (\mu_1, \mu_2, \sigma_1, \sigma_2, \pi)_j$, is referred to as an *individual* where there are j=1,...,m individuals in a population.

To start the algorithm, a population is randomly chosen from the parameter space. An objective function such as (3) is evaluated with each individual. The individuals are ranked in order of their objective function values and assigned a fitness value. For example, for the optimization goal of minimizing (3), an individual receives a high fitness value for a low value of (3). Individuals are randomly selected with replacement to breed the next population of individuals, or *offspring*. One's chance for selection is based on their fitness and some individuals may be selected more than once. The selected individuals are called *parents*. Once the parents are selected, they are randomly assigned to receive a crossover, a mutation or both.

For a crossover, two parents exchange portions of their parameter vectors with each other to create two new individuals. For example, two parents, (12, 35, 1.0, 3.0, 0.8) and (7, 23, 2.0, 2.5, 0.3), when crossed, may produce (12, 23, 2.0, 3.0, 0.8) and (7, 35, 1.0, 2.5, 0.3). Crossovers occur on one or more parameters, but not on all parameters. For a mutation, small, random increments of change within the parameter space are applied to a selected parent to produce a new individual. For example, a parent, (11, 25, 1.5, 3.0, 0.8), might be mutated into (12, 28, 1.5, 3.0, 0.6). Mutations occur on at least one and up to all parameters. Mutations and crossovers may both occur on a parent. The offspring population is of a size equal to the current population's size.

The EC algorithm then evaluates the offspring similarly with regard to (3). They, too, are ranked, assigned fitness measures, selected to breed, and crossed and/or mutated. The algorithm repeats these steps of breeding and selection until convergence. EC converges when iterative changes in (3) or in individuals fall below a tolerance level (= 0.00001) or a maximum number of iterations has been performed (= 1000).

The EC algorithm also has another feature, which allows it to optimize on multiple objective functions simultaneously (Reynolds and Ford 1999). EC uses Pareto Optimality to compare how well individuals do in terms of all functions simultaneously. To calculate Pareto Optimality, Reynolds and Ford (1999) say:

> "Given a set of vectors from which you wish to select the optimum, compare the vectors and remove all [the] dominated ones, where vector X dominates vector Y if and only if: X is at least as good as Y with respect to all criteria and there is a criterion for which X is strictly better than Y. The Pareto Optimal Set consists of those vectors left after all dominated ones have been removed."

In the mixture model context, the vectors are the multiple objective function values for each individual of the population evaluated per iteration of an EC search and are called *assessment vectors*. As an example, suppose three objective functions, $\mathbf{f} = (f_1, f_2, f_3)$, are to be minimized simultaneously. Two individuals have assessment vectors, $\mathbf{f}_1 = (378, 25, 0.75)$ and $\mathbf{f}_2 = (378, 18, 1.78)$ and are compared using Pareto Optimality. Both do equally well for f_1, yet individual *1* does better for f_3 and worse for f_2 than individual *2*. Individual *2* does better for f_2, but worse for f_3 than individual *1*. The Pareto Optimal set would include both individuals since neither individual dominates the other.

We used the multiple objective optimization feature with the mixture model to improve the consistency of a particular parameter of interest. Since π is a parameter of particular importance for mixture models (Woodward et al. 1984, McLachlan and Basford 1988), a criterion to optimize that parameter was added to the algorithm: a function of the absolute difference between the parameter estimate, $\hat{\pi}$, and the mean of the conditional expectations of π. The mixture data were then optimized to find the parameter estimates that simultaneously minimized (3) and this additional objective function, the mixing parameter function. This search is referred to in the paper as EC-P.

3. SIMULATION METHODS

3.1. Data

For the single sample examination, a sample of size n = 100 was generated as test data for the mixture model given by (1). To generate an observation, a random value was chosen from a uniform (0,1) distribution and compared to the mixing parameter, $\pi = 0.4$. If the value was less than or equal to π, a random value was generated from the first normal density, which had parameters μ_1 = 18.0 and σ_1 = 2.0. If the value was greater than π, a random value was generated from the second normal density, which had parameters μ_2 = 31.0 and σ_2 = 5.0. For the consistency tests, 30 samples were generated similarly to above.

3.2. Optimization methods

We wrote code in C, including a pseudo-random number generator, to generate the data samples. We wrote code for Splus to run the mixture model with EM and with MS. We wrote code in C to run the mixture model with EC. Version 1.0 (alpha-test-stage) software is available at http://faculty.washington.edu/edford/research. Our software is generic and may be modified to optimize small to medium sized nonlinear statistical and mathematical process models.

Starting parameter values were supplied to EM and MS: μ_1=25.0, μ_2=30.0, σ_1=3.0, σ_2=3.0, and π=0.23. EM and MS were both sensitive to starting values, but were not a focus of interest in this paper. See Woodward et al. (1984) for an analysis on the effects of starting values and mode separation on the EM algorithm. Parameter search ranges were supplied to EC: $\mu_1 \in$(15.0, 27.0), $\mu_2 \in$(27.0, 40.0), $\sigma_1 \in$(0.75, 10.0), $\sigma_2 \in$(0.75, 10.0), and $\pi \in$(0.20, 0.60). EC was insensitive to starting values since it used a population of randomly selected individuals from within the parameter search ranges.

For the consistency investigations, each algorithm was optimized on the same 30 samples. Since EC is a stochastic algorithm, each sample was optimized 20 times. Each sample's EC results were summarized by the optimal set, which was the parameter estimate vector of the 20 searches with the lowest objective function value. Previous analyses showed that the optimal set had the lowest bias and smallest variance as compared with other summary statistics such as mean and median. In the multiple objective case, the EC searches resulted in Pareto Optimal sets. Since the Pareto Optimal sets contained multiple, optimal parameter estimate vectors, the results were summarized by taking the parameter estimate vector with the lowest objective function value for the likelihood function, then for the lowest mixing parameter function.

3.3. Estimator assessment

For the single sample investigation, the parameter estimates were found using the three algorithms described above. Their standard errors were calculated with the covariance matrix, which was formulated using Fisher's information matrix (Louis 1982). The estimates on the single sample investigation were informally assessed for their consistency using 95% confidence intervals.

In the sample distribution investigation, several assessment statistics (Lunneborg 2000) were used to judge which algorithm produced the most strongly consistent estimators. Bias measured how close an estimator was on average to the true parameter. Since the parameters were known, bias was calculated by:

$$Bias(\hat{\theta} \mid \theta) = \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}_j - \theta \quad (4)$$

where j=1,...,m=30 samples, $\hat{\theta} \in (\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\pi})$, and $\theta \in (\mu_1, \mu_2, \sigma_1, \sigma_2, \pi)$.

Bias assesses how accurate an estimator is, but another characteristic of a good estimator is precision. The standard error of an estimator quantified the variability of an estimator and did not depend on the true parameter. It was calculated by:

$$SE(\hat{\theta} \mid X) = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(\hat{\theta}_j - \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}_j\right)^2} \quad (5)$$

where X was the population distribution of the data and $\hat{\theta}$ was defined as above. A small standard error measure is desirable as it would indicate that there is little sample variability. Any sample would be as good as another for parameter estimation.

Since accuracy and precision are both important qualities for an estimator, some assessment statistics include both the bias and the standard error in a single statistic such as the mean squared error or a ratio of the bias over the standard error. Some authors consider the mean squared error as the best criterion for comparing biased estimators (Arnold 1992, p.266). Mean squared error was calculated by:

$$MSE(\hat{\theta} \mid X, \theta) = \sqrt{Bias(\hat{\theta} \mid \theta)^2 + SE(\hat{\theta} \mid X)^2} \quad (6)$$

A low mean squared error indicates that an estimator has both low bias and low variance, but it is arguable as to whether accuracy and precision deserve equal weighting in the assessment of an estimator.

The ratio of the bias over the standard error was also calculated to assess the estimators' consistency. This statistic implies that the error of bias should be small relative to the variability of the estimator. If bias is large and standard error small, there is little chance any sample will be close to the true parameter. If both bias and standard error are large, there is better chance that a sample will be close to the true parameter. If the ratio is below 0.10, the estimator's bias is considered not too problematic (Cochran 1977). Even a ratio of 0.20 is still not too much bias for an estimator.

4. RESULTS

Parameter estimates on the single sample were found using each optimization algorithm (Table 1). The MS, EC, and EC-P searches had similar estimates and standard errors. The EM search had estimates the furthest from the true parameters and its standard errors were larger, too. Confidence intervals on the parameter estimates showed that none of the algorithms contained the true parameter for σ_1 at 95% confidence (Table 2). The EM intervals also did not contain the true parameter for π. These results suggested that all of the algorithms were biased.

To assess which algorithm produced the most consistent estimators and in particular, least biased, the algorithms were tested on the sample distribution using four assessment statistics. The bias of the estimators were calculated (Table 3). All of the algorithms produce biased estimators, but some were less biased than others. EC had the most estimators with the least bias (μ_1,σ_1,σ_2). Interestingly, EC-P was the least biased on the other parameters (μ_2, π) and it was the second least biased on μ_1,σ_1, and σ_2. In terms of accuracy across all parameters, the EC-P algorithm did the best.

Another estimator assessment statistic was the standard error (Table 4). EM had the most estimators with the smallest standard errors, so in terms of precision across all parameters, it did the best. Having the greatest precision came at the cost of slow convergence of the EM. EC and EC-P generally had larger standard errors than the others since their results included the uncertainty of the optimization routine. Since EM and MS are deterministic algorithms, their results did not include any quantitative uncertainty of the optimization search in the standard errors. To investigate their optimization uncertainty, a user had to repeat the search some number of times with different starting values to be sure the algorithm had converged to a maximum (McLachlan and Basford 1988).

The mean squared errors were also calculated (Table 5). Since this estimator used the standard error, the results were similar to the standard error results. MS had the most estimators with the smallest and/or second smallest mean-squared errors, so it was judged the best for this assessment statistic. The EC-P did the worse on all estimators, as expected. Since it optimized the mixture model on two objective functions simultaneously, its standard error estimate quantified the additional uncertainty associated with optimizing on a plane instead of just a vector.

Another criterion for judging the estimators was the ratio of the bias over the standard error (Table 6). EC-P had the most estimators with the smallest ratio values and all were below 0.20. EC also had small ratios values, all also under 0.20.

5. Discussion

When ecological and environmental studies result in bimodal data, the data may be analyzed using a mixture model of two normally distributed populations, where the model parameters are estimated using numerical optimization. There are many methods of numerical optimization; here we investigated three, the expectation-maximization (EM) algorithm, the minimizing sums (MS) function, and the Evolutionary Computation (EC) algorithm. Each algorithm was tested on a single sample, a data set generated from known parameters, and the results showed that some algorithms tended to be more consistent than others.

To investigate the consistency of the algorithms, each was assessed on four statistics on a sample distribution, i.e., 30 data sets generated from the known parameters. The statistics -- bias, standard error, mean-squared error, and a ratio of bias over standard error -- tested the estimators for accuracy and precision in different ways. The bias and the ratio of bias over the standard error were judged to be the most useful statistics for estimator assessment of consistency. The standard error was considered less useful since it did not evaluate the deterministic and the stochastic algorithms equally. That is, the standard errors did not include the optimization uncertainty of the deterministic algorithms. The mean squared error was also considered less useful since it assumed that accuracy and precision were equally important. The ratio of bias over the standard error was helpful because it allowed for relative comparisons of the estimators across and within the algorithms.

We concluded that the Evolutionary Computation algorithm, when optimized on multiple objective functions simultaneously (the EC-P case), produced the most strongly consistent estimators of the algorithms investigated in this paper. The estimators from the EC-P search tended to have the smallest or nearly smallest biases on all parameters and had the smallest bias on the parameter of interest, π. The EC-P search also resulted in the most estimators with the smallest values for the ratio of bias over standard error statistics. Its standard errors were the largest since they included the optimization uncertainty, but the were generally of the same magnitude as the standard errors for the other algorithms.

Using evolutionary computation on even a single objective function (the EC case) also produced more strongly consistent estimators than the gradient search (MS) and the expectation-maximization (EM) algorithms. Its estimators had smaller biases and ratios than the MS or EM algorithms, and its standard errors were generally of similar magnitude as the others.

We recommend the evolutionary computation algorithm optimized on multiple objective functions over the others not only for its consistency, but also because it allows one to include additional information in the optimization process. This feature is quite important for ecological and environmental studies where the modeling of data could be improved by including such information as expert knowledge on particular parameters of interest. The additional information is expressed by objective functions that the model must simultaneously satisfy during optimization. The difficulty is writing the objective functions to express that knowledge. Optimizing models on multiple objective functions should open up new paths for investigating models, including such simple models as the mixture model.

ACKOWLEDGEMENTS

"Although the research described in this article has been funded in part by the United States Environmental Protection Agency through agreement CR825173-01-0 to the University of Washington, it has as not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred." Marianne Turley was also partially funded by the University of Washington Presidential Fellowship.

LITERATURE CITED

Arnold, S. F., 1990. *Mathematical Statistics*, Prentice-Hall, Englewood Cliffs, New Jersey, pp. 636.

Chambers, J. M., and Hastie, T. J., 1993. *Statistical Models in S*, Chapman and Hall, New York, pp. 608.

Cochran, W. G., 1977. *Sampling Techniques*, John Wiley & Sons, Incorporated, New York, pp. 428.

Dempster, A. P., Laird, N. M., and Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society B*, **39**: 1-38.

Everitt, B. S., 1984. Maximum likelihood estimation of the parameters in a mixture of two univariate normal distributions; a comparison of different algorithms, *The Statistician*, **33**(2): 205-215.

Everitt, B. S., and Hand, D. J., 1981. *Finite Mixture Distributions*. Chapman and Hall, London, pp. 143.

Gay, D. M ., 1983. Algorithm 611: subroutines for unconstrained minimization using a model/trust-region approach. *ACM Trans. Math. Software.* **9**: 503-524.

Keifer, J., and Wolfowitz, J., 1956. Consistency of the maximum likelihood estimates in the presents of infinitely many incidental parameters, *Annual Mathematical Statistics*, **27**: 887-906.

Louis, T. A., 1982. Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, **44**(2): 226-233.

Lunneborg, C. E., 2000. *Data Analysis by Resampling: Concepts and Applications*, Duxbury Press, Pacific Grove, California, pp. 568.

MathSoft, Incorporated, 1995. *Splus Guide to Statistical and Mathematical Analysis*, Massachusetts Institute of Technology, Boston, Massachusetts, pp. 629.

McLachlan, G. J., and Basford, K. E., 1988. *Mixture Models*, Marcel Dekker, Incorporated, New York, pp.208.

McLachlan, G. J., and Krishnan, T., 1997. *The EM Algorithm and Extensions*, John Wiley & Sons, Incorporated, New York, pp. 274.

Michalewicz, Z., 1992. *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, New York, pp. 251.

Reynolds, J. H., and Ford, E. D., 1999. Multi-criteria assessment of ecological process models, *Ecology*, **80**(2): 538-553.

Sewell, M. A., and Young, C. M., 1997. Are echinoderm egg size distributions bimodal? *Biological Bulletin*, **193**(3): 297-305.

Skilbrei, O. T., Hansen, T., and Stefansson, S. O., 1997. Effects of decreases in photoperiod on growth and bimodality in Atlantic salmon *Salmo salar* L., *Aquaculture Research*, **28**: 43-49.

Srinivas, N., and Deb, K., 1995. Multiobjective optimization using nondominated sorting in genetic algorithms, *Evolutionary Computation*, **2**(3): 221-248.

Tkadlec, E., and Zejda, J., 1998. Density-dependent life histories in female bank voles from fluctuating populations, *Journal of Animal Ecology*, **67**(6): 863-873.

Woodward, W. A., Parr, W. C., Schucany, W. R., and Lindsey, H., 1984. A comparison of minimum distance and maximum likelihood estimation of a mixture proportion, *Journal of the American Statistical Association*, **79**(387): 590-598.

TABLES

Table 1. Parameter estimates for the single sample investigation were found by optimizing likelihood functions with the algorithms. The searches were EM = expectation-maximization, MS = gradient search with minimizing sums function, EC = evolutionary computation on a single objective function, and EC-P = evolutionary computation on two objective functions The standard errors were calculated by the covariance matrix.

source	μ_1	*standard error*	μ_2	*standard error*	σ_1	*standard error*	σ_2	*standard error*	π	*standard error*
true	18.000		31.000		2.000		5.000		0.400	
EM	18.952	0.592	32.304	0.839	4.353	0.418	5.684	0.593	0.541	0.0498
MS	18.117	0.240	31.916	0.705	1.550	0.170	5.392	0.499	0.416	0.0493
EC	18.096	0.234	31.907	0.715	1.499	0.166	5.501	0.506	0.409	0.0492
EC-P	18.170	0.247	32.020	0.678	1.594	0.174	5.178	0.480	0.417	0.0493

Table 2. The 95% confidence intervals were calculated on the parameter estimates for the single sample investigation. None of the algorithms had the true parameter for σ_1 (=2.0) within their intervals and EM also did not have the true parameter for π (=0.40) within its interval.

source	μ_1	μ_2	σ_1	σ_2	π
EM	17.792, 20.112	30.660, 33.948	3.534, 5.172	4.521, 6.847	0.4433, 0.6386
MS	17.647, 18.587	30.534, 33.298	1.217, 1.883	4.414, 6.370	0.3194, 0.5126
EC	17.637, 18.555	30.506, 33.308	1.174, 1.824	4.509, 6.493	0.3126, 0.5054
EC - P	17.686, 18.654	30.691, 33.349	1.253, 1.935	4.237, 6.119	0.3204, 0.5136

Table 3. The absolute bias values were calculated for the parameter estimates resulting from the sample distribution investigation. The * indicates the estimate with the smallest bias for that parameter. The + indicates the estimate with the second smallest bias for that parameter. The EC-P did the best overall.

source	μ_1	μ_2	σ_1	σ_2	π
EM	0.4950	0.0399^{+}	2.3004	0.5638	0.1946
MS	0.0659	0.0474	0.1023	0.1294	0.0082^{+}
EC	0.0018*	0.0730	0.0486*	0.0692*	0.0115
EC - P	0.0640^{+}	0.0160*	0.0700^{+}	0.1110^{+}	0.0030*

Table 4. The standard error values were calculated for the parameter estimates resulting from the sample distribution investigation. The * indicates the smallest variance estimate for that parameter. The + indicates the second smallest variance estimate for that parameter. The EM did the best overall.

source	μ_1	μ_2	σ_1	σ_2	π
EM	0.3225*	0.7514+	0.0375*	0.0676*	0.0581*
MS	0.3842+	0.7282*	0.2843+	0.9826	0.0630
EC	0.4020	1.0490	0.3140	0.7990+	0.0590+
EC - P	0.4760	1.1570	0.4840	0.8190	0.0670

Table 5. The mean squared error (MSE) values were calculated for the parameter estimates resulting from the sample distribution investigation. The * indicates the smallest MSE for that parameter. The + indicates the second smallest MSE for that parameter. The EM did the best overall.

source	μ_1	μ_2	σ_1	σ_2	π
EM	0.5908	0.7525+	2.3007	0.5678*	0.2031
MS	0.3898*	0.7297*	0.3021*	0.9911	0.0635+
EC	0.4020+	1.0515	0.3177+	0.8020+	0.0601*
EC - P	0.4801	1.1571	0.4895	0.8263	0.0673

Table 6. The bias/standard error ratio values were calculated for the parameter estimates resulting from the sample distribution investigation. They are reported here as percentages times 100. The * indicates the smallest ratio for that parameter. The + indicates the second smallest ratio for that parameter. The EC-P did the best overall.

source	μ_1	μ_2	σ_1	σ_2	π
EM	1.535	0.0531+	61.350	8.346	3.348
MS	0.1717	0.0651	0.3598	0.1317+	0.1305+
EC	0.0044*	0.0697	0.1550+	0.0866*	0.1954
EC - P	0.1353+	0.0140*	0.1443*	0.1360	0.0382*