CLUSTERING VIA NORMAL MIXTURE MODELS

G.J. McLachlan, D. Peel, and P. Prado

G.J. McLachlan,
Department of Mathematics,
University of Queensland, Brisbane, Queensland 4072, AUSTRALIA

KEYWORDS: Finite mixture models; Maximum likelihood; EM algorithm; Likelihood ratio test

ABSTRACT We consider a model-based approach to clustering, whereby each observation is assumed to have arisen from an underlying mixture of a finite number of distributions. The number of components in this mixture model corresponds to the number of clusters to be imposed on the data. A common assumption is to take the component distributions to be multivariate normal with perhaps some restrictions on the component covariance matrices. The model can be fitted to the data using maximum likelihood implemented via the EM algorithm. There is a number of computational issues associated with the fitting, including the specification of initial starting points for the EM algorithm and the carrying out of tests for the number of components in the final version of the model. We shall discuss some of these problems and describe an algorithm that attempts to handle them automatically.

1. INTRODUCTION In some applications of mixture models, questions related to clustering may arise only after the mixture model has been fitted. For instance, suppose that in the first instance the reason for fitting a mixture model was to obtain a satisfactory model for the distribution of the data. If this were achieved by the fitting of, say, a three-component mixture model, then it may be of interest to consider the problem further to see if the three components can be identified with three externally existing groups, or if the clusters implied by the fitted mixture model reveal the existence of previously unrecognized or undefined subpopulations.

However, in other applications of mixture models, the clustering of the data at hand is the primary aim of the analysis. In this case, the mixture model is being used purely as a device for exposing any grouping that may underlie the data. McLachlan and Basford (1988) highlighted the usefulness of mixture models as a way of providing an effective clustering of various data sets under a variety of experimental designs. In the sequel, we focus exclusively on the latter situation.

With a mixture model-based approach to clustering, it is assumed that the data to be clustered are from a mixture of an initially specified number g of groups in some unknown proportions $\pi_1, \ldots, \pi_g$. That is, each data point is taken to be a realization of the mixture probability density function (p.d.f.),

$$f(\boldsymbol{y};\, \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i c_i(\boldsymbol{y};\, \boldsymbol{\theta}_i), \tag{1}$$

where the g components correspond to the g groups. Here the vector $\boldsymbol{\Psi}$ of unknown parameters consists of the mixing proportions π_i and the elements of the $\boldsymbol{\theta}_i$ known *a priori* to be distinct. On specifying a parametric form for each component p.d.f. $c_i(\boldsymbol{y};\, \boldsymbol{\theta}_i)$, $\boldsymbol{\Psi}$ can be estimated by maximum likelihood (or some other method). Once the mixture model has been fitted, a probabilistic clustering of the data into g clusters can be obtained in terms of the fitted posterior probabilities of component membership for the data. An outright assignment of the data into g clusters is achieved by assigning each data point to the component to which it has the highest estimated posterior probability of belonging.

Within this mixture likelihood-based framework, a fundamental question associated with a given clustering application, namely how many clusters are there in the data, can be assessed formally. A test for the smallest number of components in the mixture model compatible with the data can be formulated using the likelihood ratio statistic, although unfortunately it does not have its usual asymptotic null distribution. The mixture likelihood-based approach to clustering can obviously play a major role in any exploratory data analysis in both searching for groupings in the data and testing the validity of any cluster structure discovered; that is, testing whether the apparent clusters are due to random fluctuations or whether they reflect a real separation of the data into distinct groups.

2. NORMAL MIXTURE MODELS

A common assumption in practice is to take the component densities to be multivariate normal, and to proceed on the basis that any nonnormal features in the data are due to some underlying group structure. Further, often clusters in the data are essentially elliptical, so that it is reasonable to consider fitting mixtures of elliptically symmetric component densities. Within this class of component densities, the multivariate normal density is a convenient choice given its computational tractability. However, our algorithm to be described shortly can also fit mixtures of t-distributions with either specified or unspecified degrees of freedom. The family of t-distributions provides a heavy-tailed alternative to the normal family.

In the case of multivariate normal components, the ith group-conditional density $c_i(\boldsymbol{y};\, \boldsymbol{\theta}_i)$ is given by

$$\begin{aligned} c_i(\boldsymbol{y};\, \boldsymbol{\theta}_i) \quad = \quad & (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \\ & \exp\{-\tfrac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} \\ & (\boldsymbol{y}-\boldsymbol{\mu}_i)\}, \end{aligned} \tag{2}$$

where $\boldsymbol{\theta}_i$ denotes the parameter vector consisting of the elements of $\boldsymbol{\mu}_i$ and the $\frac{1}{2}p(p+1)$ distinct elements of $\boldsymbol{\Sigma}_i$ $(i = 1, \ldots, g)$. The vector $\boldsymbol{\Psi}$ containing all the unknown parameters is given now by

$$\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_{g-1}, \boldsymbol{\theta}^T)^T,$$

where $\boldsymbol{\theta}$ is the vector containing the elements of $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g$ known *a priori* to be distinct. For example, in fitting normal mixture models, the covariance matrices in the component densities may be taken to be equal. Less restrictive constraints can be imposed by a reparameterization of the component-covariance matrices in terms of their eigenvalue decompositions as, for example, in Banfield and Raftery (1993), Flury, Schmid, and Narayanan (1993), Celeux and Govaert (1995), Bensmail and Celeux (1996), and Bensmail et al. (1997).

Under the assumption that $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are independent realizations of the feature vector $\boldsymbol{Y}$, the log likelihood function for $\boldsymbol{\Psi}$ is given by

$$\log L(\boldsymbol{\Psi}) = \sum_{j=1}^{n} \log \sum_{i=1}^{g} \pi_i c_i(\boldsymbol{y}_j\,;\, \boldsymbol{\theta}_i). \tag{3}$$

With the maximum likelihood approach to the estimation of $\boldsymbol{\Psi}$, an estimate is provided by an appropriate root of the likelihood equation,

$$\partial \log L(\boldsymbol{\Psi}) / \partial \boldsymbol{\Psi} = \mathbf{0}. \tag{4}$$

3. APPLICATION OF EM ALGORITHM

It is straightforward to find solutions of (**??**) using the EM algorithm of Dempster, Laird, and Rubin (1977). For the purpose of the application of the EM algorithm, the observed-data vector $\boldsymbol{y}_{obs} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_n^T)^T$ is regarded as being incomplete. The component-label variables z_{ij} are consequently introduced, where z_{ij} is defined to be one or zero according as $\boldsymbol{y}_j$ did or did not arise from the ith component of the mixture model $(i = 1, \ldots, g\,;\, j = 1, \ldots, n)$. This complete-data framework in which each observation is conceptualized as having arisen from one of the components of the mixture is directly applicable in those situations where $\boldsymbol{Y}$ can be physically identified as having come from a population which is a mixture of g groups. On putting $\boldsymbol{z}_j = (z_{1j}, \ldots, z_{gj})^T$, the complete-data vector $\boldsymbol{x}_c$ is therefore given by

$$\boldsymbol{x}_c = (\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_n^T)^T,$$

where $\boldsymbol{x}_1 = (\boldsymbol{y}_1^T, \boldsymbol{z}_1^T)^T, \ldots, \boldsymbol{x}_n = (\boldsymbol{y}_n^T, \boldsymbol{z}_n^T)^T$ are taken to be independent and identically distributed with $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ being independent realizations from a multinomial distribution consisting of one draw on g categories with respective probabilities $\pi_1, \ldots, \pi_g$. That is,

$$\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n \overset{iid}{\sim} \text{Mult}_g(1, \boldsymbol{\pi}),$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_g)^T$.

For this specification, the complete-data log likelihood is

$$\log L_c(\boldsymbol{\Psi}) = \sum_{i=1}^{g} \sum_{j=1}^{n} z_{ij} \log\{\pi_i c_i(\boldsymbol{y}_j\,;\, \boldsymbol{\theta}_i)\}. \tag{5}$$

It is clear that the complete-data likelihood $L_c(\boldsymbol{\Psi})$ implies the incomplete-data likelihood $L(\boldsymbol{\Psi})$.

The EM algorithm is easy to program and proceeds iteratively in two steps, E (for expectation) and M (for maximization); see McLachlan and Krishnan (1997) for a recent account of the EM algorithm in a general context. On the $(k+1)$th iteration, the E-step requires the calculation of

$$Q(\boldsymbol{\Psi}\,;\, \boldsymbol{\Psi}^{(k)}) = E_{\boldsymbol{\Psi}^{(k)}}\{\log L_c(\boldsymbol{\Psi}) \mid \boldsymbol{y}_{obs}\},$$

the conditional expectation of the complete-data log likelihood $\log L_c(\boldsymbol{\Psi})$, given the observed data $\boldsymbol{y}_{obs}$, using the current fit $\boldsymbol{\Psi}^{(k)}$ for $\boldsymbol{\Psi}$. Since $\log L_c(\boldsymbol{\Psi})$ is a linear function of the unobservable component-label variables z_{ij}, the E-step is effected simply by replacing z_{ij} by its conditional expectation given $\boldsymbol{y}_j$, using $\boldsymbol{\Psi}^{(k)}$ for $\boldsymbol{\Psi}$.

On the M-step on the $(k+1)$th iteration, the intent is to choose the value of $\boldsymbol{\Psi}$, say $\boldsymbol{\Psi}^{(k+1)}$, that maximizes

$$Q(\boldsymbol{\Psi}\,;\,\boldsymbol{\Psi}^{(k)}) = \sum_{i=1}^{g}\sum_{j=1}^{n} \tau_i(\boldsymbol{y}_j\,;\,\boldsymbol{\Psi}^{(k)}) \log\{\pi_i c_i(\boldsymbol{x}_j\,;\,\boldsymbol{\theta}_i)\}.$$

One nice feature of the EM algorithm is that the solution to the M-step often exists in closed form, as for the present case of normal mixtures. Another nice feature of the EM algorithm is that the mixture likelihood $L(\boldsymbol{\Psi})$ can never be decreased after the EM sequence. Hence

$$L(\boldsymbol{\Psi}^{(k+1)}) \geq L(\boldsymbol{\Psi}^{(k)}),$$

which implies that $L(\boldsymbol{\Psi}^{(k)})$ converges to some L^* for a sequence of likelihood values bounded above.

Let $\hat{\boldsymbol{\Psi}}$ be the chosen solution of the likelihood equation. The likelihood function $L(\boldsymbol{\Psi})$ tends to have multiple local maxima for normal mixture models. In this case of unrestricted component covariance matrices, $L(\boldsymbol{\Psi})$ is unbounded, as each data point gives rise to a singularity on the edge of the parameter space; see, for example, McLachlan and Basford (1988, Chapter 2). In practice, however, consideration has to be given to the problem of relatively large local maxima that occur as a consequence of a fitted component having a very small (but nonzero) variance for univariate data or generalized variance (the determinant of the covariance matrix) for multivariate data. Such a component corresponds to a cluster containing a few data points either relatively close together or almost lying in a lower dimensional subspace in the case of multivariate data. There is thus a need to monitor the relative size of the fitted mixing proportions and of the component variances for univariate observations and of the generalized component variances for multivariate data in an attempt to identify these spurious local maximizers. Hathaway (1983) has considered a constrained formulation of this problem in order to avoid singularities and to reduce the number of spurious local maximizers. There is also a need to monitor the Euclidean distances between the fitted component means to see if the implied clusters represent a real separation between the means or whether they arise because one or more of the clusters fall almost in a subspace of the original feature space.

4. ASSESSING THE NUMBER OF COMPONENTS A major concern in cluster analysis is deciding on the number of clusters that are present in the data, where typically no *a priori* knowledge exists on the group-structure of the data. In the context of clustering via a finite mixture model, the problem can be approached by testing for the smallest number of components in the mixture model compatible with the data. For an account of this problem, the reader is referred to McLachlan and Basford (1988, Section 1.10). More recent references may be found, for example, in Celeux and Soromenho (1996). The algorithm MIXFIT to be described in the next section has the option to assess the most appropriate number of clusters via the resampling approach as discussed in McLachlan (1987). This option is based on the MMRESAMP program of McLachlan et al. (1995); see also McLachlan and Peel (1997).

5. MIXFIT ALGORITHM It follows from above that care must be taken in the choice of the root of the likelihood equation in the case of unrestricted covariance matrices where $L(\boldsymbol{\Psi})$ is unbounded. In order to fit a mixture model using the EM algorithm, an initial value has to be specified for the vector $\boldsymbol{\Psi}$ of unknown parameters for use on the E-step on the first iteration of the EM algorithm. Equivalently, initial values must be specified for the posterior probabilities of component membership of the mixture, $\tau_1(\boldsymbol{y}_j;\ \boldsymbol{\Psi}^{(0)}), \ldots, \tau_g(\boldsymbol{y}_j;\ \boldsymbol{\Psi}^{(0)})$, for each $\boldsymbol{y}_j$ $(j = 1, \ldots, n)$ for use on commencing the EM algorithm on the M-step the first time through. The latter posterior probabilities can be specified as zero-one values, corresponding to an outright partition of the data into g groups. In this case, it suffices to specify the initial partition of the data. In a cluster analysis context it is usually more appropriate to do this rather than specifying an initial value for $\boldsymbol{\Psi}$. Typically with the fitting of mixture models to multivariate data, the likelihood equation will have multiple roots, and so the choice of starting values of the EM algorithm requires careful consideration.

We now describe an algorithm called MIXFIT, which automatically provides a selection of starting values for this purpose if the user does not provide any. The MIXFIT algorithm automatically provides starting values for the application of the EM algorithm by considering a selection obtained from three sources (a) random starts (b) hierarchical clustering-based starts, and (c) k-means clustering-based starts. Concerning (b), the user has the option of using in either standardized or unstandardized form, the results from seven hierarchical methods (nearest neighbour, farthest neighbour, group average, median, centroid, flexible sorting, and Ward's method). There are several algorithm parameters that the user can optionally specify; alternatively, default values are used. The program fits the normal mixture model for each of the initial grouping specified from the three sources (a) to (c). All these computations are automatically carried out by the program. The user only has to provide the data set, the extent of the selection of the initial groupings to be used to determine starting values, and the number of components that are to be fitted. Summary information is automatically given as output for the final fit, which is taken to be the fit corresponding to the largest of the local maxima located. However, the summary information can be recovered for any distinct fit.

As well as the options pertaining to the automatic provision of starting values covered above, several other options are available, including the provision of standard errors for the fitted parameters in the mixture model, and the bootstrapping of the likelihood ratio statistic λ for testing $g = g_0$ versus $g = g_0 + 1$ components in the mixture model, where the value g_0 is specified by the user. With the latter option, the bootstrap samples are generated parametrically from the g_0-component normal mixture model with $\boldsymbol{\Psi}$ set equal to the fit $\hat{\boldsymbol{\Psi}}_{g_0}$ for $\boldsymbol{\Psi}$ under the null hypothesis of g_0 components.

6. EXAMPLE We consider now the well-known set of *Iris* data as originally collected by Anderson (1935) and first analysed by Fisher (1936). It consists of measurements of the length and width of both sepals and petals of 50 plants for each of three types of Iris species *setosa, versicolor*, and *virginica*. As pointed out by Wilson (1982), the *Iris* data were collected originally by Anderson (1935) with the view to seeing whether there was "evidence of continuing evolution in any group of plants". Her aural approach to data analysis suggested that both the *versicolor* and *virginica* species should each be split into two subspecies.

Hence we focus on the clustering of the 50 observations in the *Iris virginica* set. We considered a clustering of this data set into two clusters C_1 and C_2 by fitting a mixture of two heteroscedastic normal components. The membership of the smaller sized cluster (C_1) is reported in Table 1 for the clustering implied by each of eleven solutions of the likelihood equation. These solutions correspond to the largest eleven local maxima of $L(\boldsymbol{\Psi})$ as found using the MIXFIT algorithm. Also listed in Table 1 for each of these local maximizers is the value of the determinant of each of the two fitted covariance matrices $|\hat{\boldsymbol{\Sigma}}_1|$ and $|\hat{\boldsymbol{\Sigma}}_2|$, the value of the log likelihood, and the value of $-2\log\lambda$ and its assessed P-value for the test of $g = 1$ versus $g = 2$, as obtained by resampling on the basis of 99 replications. Since $g = 1$ under the null hypothesis, the bootstrap replications of $-2\log\lambda$ are actual applications of this test statistic.

The clustering implied by the first solution S_1 listed in Table 1, which had been obtained previously by Wilson (1982), has nine observations in the first cluster. It was found that these nine points in C_1 lie in the lower portion of the scatter plot of the first two projection pursuit variates obtained by the procedure of Friedman (1987), and can be separated by a hyperplane from the other observations, excluding the observation numbered 36.

The ten other solutions S_2 to S_{11} in Table 1 correspond to the largest ten local maxima located. The smaller sized clusters implied by six of these solutions have only five members. Hence given that the data are of dimension $p = 4$, it is not surprising that the fitted covariance matrix for the first component of the mixture is nearly singular for each of them, with a generalized variance equal to only 7.6×10^{-8} or smaller. The clusterings implied by solutions S_2, S_4, and S_6 are similar and they have at least four members of their first cluster C_1 in common with the first cluster implied by S_1. But the first clusters C_1 implied by the solutions S_8 to S_{10} have no members in common with that of S_1. Further, it can be confirmed from scatter plots that solutions S_3 and S_7 to S_{11} do not provide as much separation between the means of the implied clusters. Thus these solutions in particular would appear to be more spurious in nature rather than representing a genuine grouping. If a lower bound were placed on each of $|\hat{\boldsymbol{\Sigma}}_1|$ and $|\hat{\boldsymbol{\Sigma}}_2|$ as discussed in Section 3, then only solution S_1 would be retained. However, it is not suggested that the clustering of a data set should be based solely on a single solution of the likelihood equation, but rather on the various solutions considered collectively as above.

Another way of proceeding to reduce the prevalence of solutions corresponding to artificially small values of the generalized fitted variances is to restrict the covariance matrices to be the same or at least the correlation matrices to be equal. The latter restriction is a reasonable one to impose in many situations in biology. Under homoscedasticity, the first cluster implied by the maximum likelihood solution (assuming it is the global maximizer), contains the union of all members of the first clusters implied by the heteroscedastic solutions S_1, S_2, and S_4 to S_6, along with observations 9 and 36. The solution corresponding to the largest maximum located under the less restrictive assumption of equal correlation matrices, gives almost the same clustering as S_4, but with the observation numbered 31 moved to the larger-sized cluster. The solution corresponding to the second largest of the local maxima located under the assumption of equal correlation matrices gives the same clustering as S_1.

On the question of whether there are signs of continuing evolution in the *virginica* species, it can be seen from the P-values reported in Table 1 that the *virginica* species can still be regarded as being homogeneous. The only clusterings in Table 1 that have an assessed P-value that would imply significance of a two-group structure at a conventional level are the last few which we have perceived to be of a spurious nature. It should be pointed out too that all the solutions in Table 1, apart from S_1, S_3, and S_4, were found only after using the stochastic version of the EM algorithm (Celeux and Diebolt, 1985). This version allows the EM algorithm to have a chance to escape from the current EM sequence. But evidently in this example, such escapes often led to convergence in the end to what we have concluded to be spurious local maximizers.

REFERENCES

Anderson, E. (1935). The irises of the Gaspé Peninsula. *Bulletin of the American Iris Society* **59**, 2–5.

Banfield, J.D. and Raftery, A.E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics* **49**, 803–821.

Bensmail, H. and Celeux, G. (1996). Regularized Gaussian discriminant analysis through eigenvalue decomposition. *Journal of the American statistical Association* **91**, 1743–1748.

Bensmail, H., Celeux, G., Raftery, A.E., and Robert, C.P. (1997). Inference in model-based cluster analysis. *Statistics and Computing* **7**, 1–10.

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

Celeux, G. and Diebolt, J. (1985). The SEM Algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly.* **2**, 73–92.

Celeux, G. and Govaert, G. (1995). Gaussian parsimonious clustering model. *Pattern Recognition* **28**, 781–793.

Celeux, G. and Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification* **13**, 195–212.

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics* **7**, 179–188.

Flury, B.W., Schmid, M.J., and Narayanan, A. (1993). Error rates in quadratic discrimination with constraints on the covariance matrices. *Journal of Classification* **11**, 101–120.

Friedman, J.H. (1987). Exploratory projection pursuit. *Journal of the American Statistical Association* **82**, 249–266.

Hathaway, R.J. (1983). Constrained maximum likelihood estimation for normal mixtures. In *Computer Science and Statistics: The Interface*, J.E. Gentle (Ed.). Amsterdam: North-Holland, pp. 263–267.

McLachlan, G.J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics* **36**, 318–324.

McLachlan, G.J. and Basford, K.E. (1988). *Mixture Models: Inference and Applications to Clustering.* New York: Marcel Dekker.

McLachlan, G.J. and Krishnan, T. (1997). *The EM Algorithm and Extensions.* New York: Wiley.

McLachlan, G.J. and Peel, D. (1997). On a resampling approach to choosing the number of components in normal mixture models. *Proceedings of Interface 96, 28th Symposium on the Interface*, Sydney, August 1996, pp. 260–266. Fairfax, Virginia: Interface Foundation of North America.

McLachlan, G.J., Peel D., Adams, P., and Basford, K.E. (1995). An algorithm for assessing by resampling the P-value of the likelihood ratio test on the number of components in normal mixture models. *Research Report No. 31*. Brisbane: Centre for Statistics, The University of Queensland.

Wilson, S.R. (1982). Sound and exploratory data analysis *Compstat 1982, Proc. International Association for Statistical Computing.* Vienna: Physica-Verlag, 447–450.

Table 1
Results of fitting a mixture of $g = 2$ heteroscedastic normal components to data on *Iris virginica* species

Solution No.	Cluster C_1	$\log L$	$\lvert\hat{\boldsymbol{\Sigma}}_1\rvert$	$\lvert\hat{\boldsymbol{\Sigma}}_2\rvert$	$-2\log\lambda$ (P-value)
1	6,8,18,19,23, 26,30,31,32	-36.994	1.4×10^{-6}	3.7×10^{-5}	43.2 (40%)
2	6,18,19,23,32	-36.987	7.6×10^{-8}	5.2×10^{-5}	43.2 (40%)
3	10,13,17,21,26 30,32,40,41,42 44,46,47	-35.622	2.9×10^{-7}	7.3×10^{-5}	46.0 (29%)
4	6,18,19,23,31	-35.406	6.8×10^{-9}	6.3×10^{-5}	46.4 (29%)
5	5,18,21,32,35, 40,41,42,44,46	-34.427	6.2×10^{-8}	7.2×10^{-5}	48.3 (29%)
6	6,18,19,32,35	-34.063	1.5×10^{-8}	5.5×10^{-5}	49.1 (29%)
7	2,14,17,20,30, 32,36,43	-33.690	3.6×10^{-9}	9.7×10^{-5}	49.8 (27%)
8	1,37,41,42,49	-30.374	2.9×10^{-11}	9.3×10^{-5}	56.4 (13%)
9	20,35,42,46,47	-29.756	8.0×10^{-11}	8.0×10^{-5}	57.7 (8%)
10	2,7,17,40,42, 43,48	-28.581	7.6×10^{-11}	1.2×10^{-4}	60.0 (6%)
11	8,19,23,28,39	-25.071	1.3×10^{-11}	8.0×10^{-5}	67.0 (3%)