Applications of the EM Algorithm

Ryan Cooper

May 18, 2017

Introduction

The Expectation Maximization (EM) algorithm is a useful algorithm for maximizing a likelihood function when latent variables are present. The essential idea behind the algorithm is to alliteratively compute the expected value of the latent variables given the data and current parameter estimates and then use this conditional expectation in the likelihood function and update the maximum likelihood estimates. For a full description of the algorithm and its important properties see [1]. The aim of this report is to use the EM algorithm to find maximum likelihood estimates for parameters in models with incomplete data. In the first section we analyze continuous uni-variate data and fit a two-component normal model. Additionally, we use the bootstrap to test model assumptions and provide confidence intervals for the median. For the second section we fit both a one and two-component Poisson model to count data and test the goodness of fit for each model.

1 Two-Component Gaussian Mixture

The data in this section consists of 90 observations from a continuous random variable. Let $X' = (X_1, \ldots, X_{90})$ denote the data for this section. The histogram of the sample X shown in Figure 1 indicates that the distribution is bi-modal and the normal assumption is not appropriate. After testing the fit of X against a single normal distribution, we fit a two component Gaussian mixture model using the EM algorithm.

Testing for Normality

A natural test in this composite null hypothesis setting is the Shapiro-Wilk test for normality, which is known to be sensitive to a wide range of alternatives. The test statistic is of anova type; it's proportional to the ratio of two estimators for the population variance. For a description of the test statistic and its properties see [2]. The resulting p-value for the test is 0.0043. Hence we reject the null hypothesis that the sample was drawn from a single normal distribution at the 0.05 level. Based on the histogram in Figure 1, a more appropriate model for the data is a two-component Gaussian mixture.

Fitting a Two-Component Normal Mixture via EM

Now consider two possible models for the observation X. Suppose that $X_1, \ldots, X_{90}$ is an iid sample from

$$(1-p)N(\mu_0, \sigma_0^2) + pN(\mu_1, \sigma_1^2).$$

The first model assumes that $\sigma_0 = \sigma_1$ while the second assumes no restriction on σ_0 and σ_1. For each model we use the EM algorithm to obtain the maximum likelihood estimates for the model parameters. Tables 1 and 2 show the iterations up to three decimals of accuracy of the EM algorithm applied to each model.

Iter	p	μ_0	μ_1	σ^2
1	0.5000	2.8742	5.8535	16.2248
5	0.4646	2.7328	6.5181	1.6065
10	0.4242	2.6277	7.0213	0.7701
12	0.4237	2.6293	7.0236	0.7678

Table 1: Iterations of the EM algorithm for fitting a two-component normal mixture with $\sigma_0 = \sigma_1 = \sigma$

Iter	p	μ_0	μ_1	σ_0^2	σ_1^2
1	0.5000	2.8742	5.8535	16.2248	16.2248
5	0.5043	3.8924	5.0801	5.0237	6.6638
10	0.5106	3.4776	5.4632	3.9480	6.4492
15	0.5258	2.7759	6.0383	1.9885	4.9885
20	0.5168	2.4082	6.4388	1.0171	3.1989
25	0.5002	2.4233	6.5579	1.0097	2.8434
30	0.4909	2.4448	6.6136	1.0339	2.7115
35	0.4863	2.4567	6.6410	1.0484	2.6490
40	0.4840	2.4627	6.6544	1.0561	2.6191
45	0.4828	2.4657	6.6609	1.0600	2.6047
50	0.4823	2.4672	6.6641	1.0619	2.5978
51	0.4822	2.4674	6.6645	1.0622	2.5969

Table 2: Iterations of the EM algorithm for fitting a two-component normal mixture with σ_0 and σ_1 unconstrained.

Figure 1 shows the histogram of X overlaid with the two density estimates. The figure indicates that the unrestricted assumption on σ_0 and σ_1 produces a better candidate for the model density of X. However, note that the Kolmogorov-Smirnov goodness-of-fit test gives a p-value of less than 0.001 for both models, so we reject each hypothesis that X is an iid sample from one of the fitted models at the 0.05 level.

Testing $\sigma_0 = \sigma_1$ vs $\sigma_0 \neq \sigma_1$

In this section we test the null hypothesis that $\sigma_0 = \sigma_1$ versus $\sigma_0 \neq \sigma_1$. The test is given by using the EM algorithm to create a bootstrap 95% confidence interval for the difference
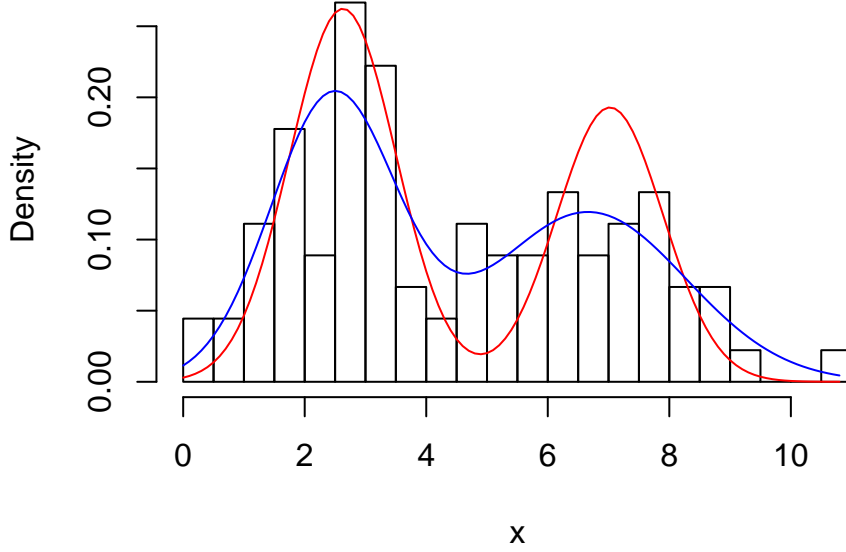
Figure 1: Histogram with density estimate for each of the normal mixture models (red: $\sigma_0 = \sigma_1$, blue: σ_0 and σ_1 unrestricted).

$\sigma_0 - \sigma_1$ and rejecting the null hypothesis if 0 is not in the interval. The 95% confidence interval for $\sigma_0 - \sigma_1$ is $(-1.483, 0.025)$. Since 0 is in the interval we fail to reject the null hypothesis at level 0.05.

Median

In this section we construct two 95% confidence intervals for the median of the distribution. The sample median is 3.813. Our first confidence interval for the median is gotten by using the bootstrap. The pivotal bootstrap confidence interval for the median is $(2.524, 4.401)$. The second confidence interval is also nonparametric and is obtained by noting that

$$\mathbb{P}(X_{(j)} \leq \xi_q \leq X_{(k)}) = \sum_{i=j}^{k-1} \binom{n}{i} q^i (1-q)^{n-i}$$

where $X_{(j)}$ is the jth order statistic and ξ_q is the qth quantile. Using this fact we obtain $(3.138, 5.102)$ as an approximate 95% confidence interval for the median. The length bootstrap confidence interval is 1.876, while the length of the second confidence interval is 1.964. Hence both methods produce confidence intervals of the same approximate length. Note that the second confidence interval obtained through the order statistics is more simpler to implement and thus a more convenient interval with the same approximate coverage as the bootstrap interval. Lastly, a third possible confidence interval can be obtained by using the

asymptotic normality of the median. However, this method requires an estimator of the density at the median, and although we have estimated the density via the EM algorithm, it's lack of fit may result in an unreliable confidence interval.

2   Two Component Poisson Mixture

The data in this section comes from the London Times and is a count of the number of deaths per day among women 80 years and older. A total of 1096 deaths were recorded with a range of 0 to 9 deaths per day. Let Y_i denote the number of deaths on day i for $i = 1, \ldots, n$. It is reasonable to assume that the deaths are independent from one another, so we can model Y_i with a Poisson distribution with rate λ_i.

First we test the hypothesis that the observations are a sample from a single component Poisson distribution. Formally, we are testing H_0: there exists $\lambda > 0$ such that $\lambda_i = \lambda$ for all i versus H_1: λ_i are unconstrained. In this setting the likelihood ratio is approximately equal to the dispersion statistic

$$T = \frac{1}{\overline{Y}} \sum (Y_i - \overline{Y})^2.$$

Under the null hypothesis T is approximately chi-square distributed with $n - 1$ degrees of freedom. The resulting p-value for the sample is 5.1513947×10^{-7}. Thus we reject H_0 at the 0.05 level.

Next we fit a two-component Poisson mixture model to the data. Suppose $Y_1, \ldots, Y_n$ is an iid sample from

$$(1 - p)\text{Pois}(\lambda_0) + p\text{Pois}(\lambda_1).$$

Using the EM algorithm the maximum likelihood estimates for p, λ_0 and λ_1 are 0.2871, 2.5762 and 1.1158, respectively. To test how well the distribution fits the data, we apply the chi-square goodness-of-fit test. The resulting p-value is 0.9846, so we accept the null hypothesis that the count proportions are determined by a two-component Poisson mixture model at the 0.05 level. The table shows the frequency of deaths for the observed data, the expected one-component Poisson model and two-component Poisson model. Note that the lack of fit for the one-component Poisson model appears most prominently in 0, 2, 3 and 5 death counts. The table also shows that the two-component Poisson model enjoys a significant improvement in the fit to the data.

	0	1	2	3	4	5	6	7+
Observed	162.0	267.0	271.0	185.0	111.0	61.0	27.0	12.0
Expected 1-component	126.8	273.5	294.9	212.0	114.3	49.3	17.7	7.4
Expected 2-component	162.5	268.1	261.4	193.2	115.7	57.7	24.4	12.9

References

[1] William Navidi, *A graphical illustration of the EM algorithm*, The American Statistician, Vol. 51, No. 1 (1997), 29-31

[2] Samuel S. Shapiro and Martin B. Wilk, *An analysis of variance test for normality (complete samples)*, Biometrika, Vol. 52, No. 3/4 (1965), pp. 591–611