Journal of Biometrics & Biostatistics
Gunaratnam and Rai, J Biom Biostat 2019, 10:1
Research Article
Open Access

Comparing the Variability Using Louis' Method and Resampling Methods

Gunaratnam B[1]* and Rai SN[1,2]

[1]Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY 40202, USA
[2]Biostatistics Shared Facility, Department of Bioinformatics and Biostatistics, James Graham Brown Cancer Center, University of Louisville, KY 40202, USA

Abstract

There is a problem when a relatively simple analysis is changed into a complex one just because some of the information is missing. Louis showed how to estimate the standard deviation of maximum likelihood estimate (MLE) for a parameter θ using the missing information. In the meantime, the resampling method is one of the best methods to calculate the standard deviation of sample estimates. In this article, we define and compare the standard deviation of a parameter θ using complete data, incomplete data, and the EM algorithm. As an illustration, we analyze a data from Rao and compare all methods for estimating variability.

Keywords: EM algorithm; Maximum likelihood estimation; Incomplete data; Complete data; Bootstrap; Standard deviation

Introduction

Over the past several decades, the subject of missing values, incomplete data, the expectation and the maximization (EM) algorithm and observed information were discussed by many authors, see e.g., [2-5]. Hartley and Hocking [2] tried to introduce a simple taxonomy for incomplete problems and developed unified methods of analyzing incomplete data including estimating the parameters in a multivariate normal distribution when some of the responses are missing. Sundberg also studied an iterative method for the solution of the likelihood equations for incomplete data from exponential families, i.e., the data being a function of an exponential family. Efron and Hinkley [4] studied the variance estimate of the maximum likelihood estimator (MLE) to the normal distribution with one-parameter families, i.e. var $(\hat{\theta}) = 1/E_\theta[I(x)]$, where $I(x)$ is the observed information, i.e. minus the second derivative of the log-likelihood function at $\hat{\theta}$ given data x.

Dempster et al. [3] then made significant contributions with the essential ideas of EM algorithm and maximum likelihood estimates for incomplete data including examples, missing values situations, and proofs. Specially, Dempster et al. [3] showed a general approach to the iterative computation of maximum likelihood estimates when the observations can be considered as incomplete data with an example from Rao. In addition, Professor Cedric A.B. Smith made on this paper's example that dealt with the standard error of maximum likelihood estimator using the binomial approach.

The EM algorithm can be easily used to find the MLE of a parameter, say θ when the log-likelihood function of incomplete data is known. But, in most cases, the log-likelihood function of incomplete data is challenging to maximize (perhaps not even in closed form). In this case, the EM algorithm is particularly useful to calculate the MLE of θ. That is, the EM algorithm is able to calculate the MLE of θ without explicit formula of the log-likelihood function of incomplete data. Thus, the calculation of the standard deviation of θ based on likelihood ratio or approximate normality could be problematic. In the meantime, Louis [1] showed that EM algorithm could be extended to compute the curvature of the log-likelihood function of incomplete data without calculation of the log-likelihood function of incomplete data itself. In this paper, we calculate the standard deviation using several of the following methods: Louis method, bootstrap for complete data and bootstrap for incomplete data and comparing all standard deviation values for a data from Rao.

The paper is organized as follows, in section 2, we give a brief overview of the MLE and EM algorithm. Then, in section 3 we provide a more detailed, structured proof of the Louis' method. In the final section 4, we illustrate with an example from Rao's paper in which 197 animals are multinomially distributed into four categories. We describe a simulation study that was carried out to compare the results on standard deviation estimate using Louis' method and resampling methods. In the 5th and final section of the paper, we discuss the performance of all methods including Louis' method and bootstrap methods in terms of variability and then make our conclusions. These bootstrap methods are very useful to some data which are also discussed in this section.

The Maximum Likelihood Estimation and EM Algorithm

In this article, we denote the vector x represents the complete data and it is denoted as c, while the vector y represents the actual observed data or incomplete data or partial data and it is denoted as p. We let X be an n dimensional random vector with a probability density function (pdf) $f(x;\theta)$ on a sample space $\mathcal{X}$, where $x=(x_1,x_2,\ldots,x_n)^T$ and $\theta=(\theta_1,\theta_2,\ldots,\theta_d)^T$ is the vector containing the unknown d parameters in the postulated form for the pdf of X, and the corresponding likelihood function, say $Lc(x,\theta)$ Similarly, $y=(y_1,y_2,\ldots,y_m)^T$ denotes an observed random sample of size m on random vector Y with pdf g(y; θ) and the corresponding likelihood function, say $L_p(y,\theta)$. The superscript T denotes the transpose of a matrix. The objective is to find the MLE of θ say $\hat{\theta}$.

An estimate $\hat{\theta}$ can be obtained as a solution of the log-likelihood equation:

$$\frac{\partial \log L_c(x,\theta)}{\partial \theta} = 0. \quad (1)$$

***Corresponding author:** Gunaratnam B, Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY 40202, USA, Tel: 502-852-3525; E-mail: b0guna01@louisville.edu

Received January 24, 2019; **Accepted** February 15, 2019; **Published** February 20, 2019

Citation: Gunaratnam B, Rai SN (2019) Comparing the Variability Using Louis' Method and Resampling Methods. J Biom Biostat 10: 424.

In addition, $\hat{\theta}$ satisfies the following condition:

$$\frac{\partial^2 \log L_c(x,\theta)}{\partial\theta\partial\theta^T}|_{\theta=\hat{\theta}} < 0.$$

We let,

$$I_c(x,\theta) = -\frac{\partial^2 \log L_c(x,\theta)}{\partial\theta\partial\theta^T} \tag{2}$$

be the matrix of the negative of the second-order partial derivatives of the log-likelihood function with respect to the elements of θ, and it is called observed information matrix. Under regularity conditions, the expected (Fisher) information matrix $\mathcal{I}(\theta)$ is given by:

$$_X(\theta) = E\{S_c(X,\theta)S_c\ (X,\theta)\} = \quad E\{I_c(X,\theta)\}.$$

That is,

$$\mathcal{I}_X(\theta) = -E\{I_c(X,\theta)\} \tag{3}$$

where

$$S_c(x,\theta) = \frac{\partial \log L_c(x,\theta)}{\partial\theta} \tag{4}$$

is the gradient vector of the log-likelihood function; that is the score statistic.

As the log-likelihood function of X is unobservable, we will not be able to find the MLE of θ, $I_c(x,\theta)$ and $\mathcal{I}_X(\theta)$. So, instead of observing $x\in\mathcal{X}$, we can find the MLE of θ using the observed data y. The algorithm operates as follows:

Then, we can write the log-likelihood function of observed data $L_p(y,\theta)$ as:

$$\log L_p(y,\theta) = \log\left(f_Y(y\,|\,\theta)\right) = \log\left(\int_R f_X(x\,|\,\theta)d\mu(x)\right) \tag{5}$$

where $R=\{x{:}y(x)=y\}$, and $d\mu(x)$ is a measure.

The EM algorithm approaches the problem of solving the observed data log-likelihood eqn. (5) indirectly by proceeding iteratively in terms of the complete data log-likelihood function $L_c(x,\theta)$. As it is unobservable, it is replaced by its conditional expectation given y, using the current fit for θ. The definitions of the EM algorithm for complete data was defined by Dempster et al. [3], as follows:

Suppose first that $f(x|\theta)$ has the regular exponential family form:

$$f(x\,|\,\theta) = b(x)\exp(c(\theta)^T t(x)) / a(\theta) \tag{6}$$

where θ denotes a $1\times d$ vector parameter, $t(x)$ corresponds $1\times d$ vector of *complete-data* sufficient statistics, $b(x)$denotes a function of x, $a(\theta)$ represents a function of θ. In addition, let y denote the "observed" or "incomplete data" from the sample, and $\theta^{(k)}$ represents the current value of θ after k cycles of the EM algorithm. The next cycle of the algorithm consists of the following steps:

E-Step: Estimate the *complete-data* sufficient statistics, $t(x)$, by evaluating:

$$t^{(k)} = E\left[t(X)\,|\,Y\right]|_{\theta=\theta^{(k)}}, \tag{7}$$

the conditional expectation of $t(x)$ given the observed data y at the current value of θ. In the case of the exponential family, the properties of the estimator are well established. However, if the distribution is non-exponential family such as Cauchy, then mean does not exist and thus the properties (self-sufficiency and monotone convergence) need to be established case by case basis.

M-Step: Determine $\theta^{(+1k)}$, an updated value of θ, as the solution in θ of the equation:

$$E\left[t(X)\right]|_{\theta=\theta^{(k+1)}} = t^{(k)} \tag{8}$$

Alternatively, the above equation can be expressed as:

$$\frac{\partial \log a(\theta)}{\partial c(\theta)}|_{\theta=\theta^{(k+1)}} = t^{(k)}, \tag{9}$$

see more details the paper Dempster et al. [3] and Louis [1].

Louis' Method

To see how to compute the observed information in the EM, let $S_c(x,\theta)$ and $S_p(y,\theta)$ be the gradient vectors of log-likelihood functions log $L_c(x,\theta)$ and log $L_p(y,\theta)$ for the complete data and observed data respectively and $I_c(x,\theta)$ and $I_p(y,\theta)$ be the negatives of the associated second derivative matrices. Then, Louis [1] proved that the following statements are true.

$$(i)\ \ S_p(y,\theta) = E\left[S_c(X,\theta)\,|\,X\in R\right] \tag{10}$$

$$(ii)\ \ \mathcal{I}_Y(\hat{\theta}) = E\left[I_c(X,\theta)\,|\,X\in R\right] - E\left[S_c(X,\theta)S_c^T(X,\theta)\,|\,X\in R\right]|_{\theta=\hat{\theta}}. \tag{11}$$

In the eqn. (11), the first term is the conditional expected full data observed information matrix while the second term produces the expected information for the conditional distribution of X given $X\in R$. That is, we can write using a simplified notation as below:

$$\mathcal{I}_Y(\hat{\theta}) = \mathcal{I}_X(\hat{\theta}) - \mathcal{I}_{X|Y}(\hat{\theta}) \tag{12}$$

which is an application of the missing information principle [6] to the observed information. It has the following appealing interpretation:

Observed Information=Complete Information-Missing Information.

The eqn. (12) can be written as:

$$\mathcal{I}_Y(\hat{\theta}) = E\left[-\frac{\partial^2}{\partial\theta^2}\ln[X\,|\,\theta\,]\,|\,X\in R\right] - Cov\left[\frac{\partial}{\partial\theta}\ln[X\,|\,\theta\,]\,|\,X\in R\right]|_{\theta=\hat{\theta}}.$$

The first term of the right side of the above equation (i.e., complete information) is the conditional expected information matrix of the complete data X and is typically easy to compute. However, it may be computationally intractable to calculate in some situations. Tanner [7] suggested a Monte Carlo approach to Louis' method by replacing the expectations with a Monte Carlo estimate. Louis [1] proved that the second term of the right side (Missing Information) is the expected information for the conditional distribution of X given $X\in R$ (Appendix 1 for proving the Louis method).

This eqn. (12) is an application of the missing information principle [6] to the observed information. Notice that all of these conditional expectations can be computed in the EM algorithm using only S and I, which are the gradient and curvature for a complete-data problem. Of course, they need to be evaluated only on the last iteration of the EM procedure, where $S_p(y,\theta)$ is zero.

Examples

Estimating maximum likelihood estimator (MLE)

The following example is based on the lead example used in Rao and is derived from a model for recombination in gene mapping studies. In this data, 197 animals are distributed multinomially into four categories, so the observed data consists of $y=(y_1,y_2,y_3,y_4)=(125,18,20,34)$. A gene model for the population specifies cell probabilities with $\{(\frac{1}{2}+\frac{1}{4}\theta),\frac{1}{4}(1-\theta),\frac{1}{4}(1-\theta),\frac{1}{4}\theta\}$, $0 \le \theta \le 1$. With y_1,y_2,y_3,y_4 as the frequencies, let $Y_1=X_2+X_2$, $Y_2=X_3+X_3$, $Y_3=X_4+X_4$, where X is a multinomial with parameters $\{\frac{1}{2},\frac{1}{4}\theta,\frac{1}{4}(1-\theta),\frac{1}{4}(1-\theta),\frac{1}{4}\theta\}$. Here, $Y=(Y_1,Y_2,Y_3,Y_4)$ and $X=(X_1,X_2,X_3,X_4,X_5)$ are called incomplete data and complete data respectively. In this example, we will show:

(i) How to find the MLE of θ from incomplete data?

(ii) How to find the MLE of θ from complete data?

(iii) How to find the standard deviation of MLE using:

- Louis' Method

- Second Derivative w.r.t θ for Complete and Incomplete Data

- Binomial Approach

- Bootstrap.

(i) Using incomplete data

The likelihood function of θ is:

$$L_p(y;\theta)=\frac{(y_1+y_2+y_3+y_4)!}{y_1!y_2!y_3!y_4!}\left(\frac{2+\theta}{4}\right)^{y_1}\left(\frac{1-\theta}{4}\right)^{y_2+y_3}\left(\frac{\theta}{4}\right)^{y_4} \quad (13)$$

If we derivative log-likelihood function with respect to θ, and set the derivative equal to zero then we will get the quadratic equation of θ, as follows:

$$\frac{\partial}{\partial\theta}\log\left(L_p(y;\theta)\right)=0+\frac{y_1}{\theta+2}-\frac{(y_2+y_3)}{(1-\theta)}+\frac{y_4}{\theta}$$

$$=0 \;\; if \;\; \theta=\hat{\theta}$$

That is,

$$\left(y_1+y_2+y_3+y_4\right)\hat{\theta}^2+\left(-y_1+2y_2+2y_3+y_4\right)\hat{\theta}-2y_4=0. \quad (14)$$

If we substitute of (y_1,y_2,y_3,y_4)=(125,18,20,34) then we will get the value of $\hat{\theta}$ is 0.6268 as $\hat{\theta}>0$.

If we do the second derivative with respect to θ, then

$$\frac{\partial}{\partial\theta^2}\log\left(L_p(y;\theta)\right)=-\frac{y_1}{(\theta+2)^2}-\frac{(y_2+y_3)}{(1-\theta)^2}-\frac{y_4}{\theta^2}=-\left[\frac{y_1}{(\theta+2)^2}+\frac{(y_2+y_3)}{(1-\theta)^2}+\frac{y_4}{\theta}\right]$$

$$<0 \; if \; \theta=\hat{\theta}$$

Since the sign of the second derivative is negative when $\theta=\hat{\theta}$, we can say that the MLE of θ is $\hat{\theta}=0.6268$. But, the most of the cases, the log-likelihood function of incomplete data is challenging to maximize as we mentioned in the introduction. If we are unable to use the incomplete data to find the MLE, then we can use the complete data to find the MLE via EM algorithm, as shown in the paper Dempster et al. [3].

Remark 4.1: Finding the MLE using Tweedie Equation [8].

$$L_p(y;\theta)=\frac{(y_1+y_2+y_3+y_4)!}{y_1!y_2!y_3!y_4!}\left(\frac{2+\theta}{4}\right)^{y_1}\left(\frac{1-\theta}{4}\right)^{y_2}\left(\frac{1-\theta}{4}\right)^{y_3}\left(\frac{\theta}{4}\right)^{y_4}$$

If we derivative with respect to θ,

$$\frac{\partial}{\partial\theta}\log\left(L_p(y;\theta)\right)=0+\frac{y_1}{\theta+2}-\frac{y_2}{(1-\theta)}-\frac{y_3}{(1-\theta)}+\frac{y_4}{\theta}$$

$$=0 \; if \; \theta=\hat{\theta}$$

The maximum likelihood eqn. (13) for the incomplete data is:

$$\frac{y_1}{\theta+2}-\frac{y_2}{(1-\theta)}-\frac{y_3}{(1-\theta)}+\frac{y_4}{\theta}=0.$$

As Tweedie [8] suggested, we can easily solve the linear equation for MLE by replacing each term by its reciprocal, as follows:

$$\frac{\theta+2}{y_1}-\frac{(1-\theta)}{y_2}-\frac{(1-\theta)}{y_3}+\frac{\theta}{y_4}=0.$$

That is,

$$\theta\left[\frac{1}{y_1}+\frac{1}{y_2}+\frac{1}{y_3}+\frac{1}{y_4}\right]=\left[\frac{1}{y_2}+\frac{1}{y_3}-\frac{2}{y_1}\right].$$

If we substitute (y_1,y_2,y_3,y_4)=(125,18,20,34) then the solution of θ is 0.6264, which is MLE from Tweedie equation.

(ii) Using complete data

The likelihood function of θ is:

$$L_c(x;\theta)=\frac{(x_1+x_2+x_3+x_4+x_5)!}{x_1!x_2!x_3!x_4!x_5!}\left(\frac{1}{2}\right)^{x_1}\left(\frac{1}{4}\theta\right)^{x_2+x_5}\left(\frac{1}{4}(1-\theta)\right)^{x_3+x_4} \quad (15)$$

As we have done for the incomplete data, we derivative log-likelihood function with respect to θ, and set the derivative equal to zero then we will get:

$$\frac{\partial}{\partial\theta}\log\left(L_c(x;\theta)\right)=0+\left(x_2+x_5\right)\frac{1}{\theta}+\left(x_3+x_4\right)\frac{1}{(1-\theta)}(-1)$$

$$=0 \; if \; \theta=\hat{\theta}$$

Therefore

$$\hat{\theta}=\frac{x_2+x_5}{x_2+x_3+x_4+x_5}.$$

If we do the second derivative with respect to θ,

$$\frac{\partial}{\partial\theta^2}\log\left(L_c(x;\theta)\right)=-\left(x_2+x_5\right)\frac{1}{\theta^2}-\left(x_3+x_4\right)\frac{1}{(1-\theta)^2}$$

$$=-\left[\frac{(x_2+x_5)}{\theta^2}+\frac{(x_3+x_4)}{(1-\theta)^2}\right]$$

$$<0 \;\; if \; \theta=\hat{\theta}$$

Since the sign of the second derivative is negative when $\theta=\hat{\theta}$, we can say that the MLE of θ is $\hat{\theta}$.

In this case, we are unable to estimate the MLE value of θ as x_2 is unobservable. However, it can be estimated using the EM algorithm as described in section 2. But in order to estimate, first we have to find the *complete-data* $t(x)$ of this distribution.

The probability density function of x given θ is:

$$f(x\mid\theta)=\frac{(x_1+x_2+x_3+x_4+x_5)}{x_1!x_2!x_3!x_4!x_5!}\left(\frac{1}{2}\right)^{x_1}\left(\frac{\theta}{4}\right)^{x_2+x_5}\left(\frac{1-\theta}{4}\right)^{x_3+x_4} \quad (16)$$

This density can be written as an exponential-family form, as follows:

$$f(x\mid\theta)=\left[\frac{(x_1+x_2+x_3+x_4+x_5)}{x_1!x_2!x_3!x_4!x_5!}\left(\frac{1}{2}\right)^{x_1}\right]\exp\left[ln\left(\frac{\theta}{4}\right)x_2\right]\left(\frac{\theta}{4}\right)^{x_5}\left(\frac{1-\theta}{4}\right)^{x_3+x_4}$$

$$=b(x)\exp\left(c(\theta)t(x)\right)/a(\theta),$$

where

$$b(x)=\frac{(x_1+x_2+x_3+x_4+x_5)}{x_1!x_2!x_3!x_4!x_5!}\left(\frac{1}{2}\right)^{x_1}, \; c(\theta)=ln\left(\frac{\theta}{4}\right), \; t(x)=x_2,$$

$$\text{and } a(\theta)=\left(\frac{\theta}{4}\right)^{-x_5}\left(\frac{1-\theta}{4}\right)^{-(x_3+x_4)}.$$

E-Step: If we assume that the initial value of θ is $\theta^{(0)}$, then the first iteration of the EM algorithm, the E-step requires the computation of the conditional expectation of *complete-data* given incomplete data y at $\theta=\theta^{(0)}$. It can be written as:

$$t^{(0)}=E\left[t(X)\mid Y\right]|_{\theta=\theta^{(0)}}$$

$$=E\left[X_2\mid Y\right]|_{\theta=\theta^{(0)}}$$

$$=E\left[X_2\mid Y_1\right]|_{\theta=\theta^{(0)}}.$$

Remark 4.2: *Prove that* $E_{X_1/Y}(x_1)=E_{X_1/Y_1}(x_1)$

Consider:

$$f_{X_1/Y}(x_1)=\frac{f_{X_1Y}(x_1,y)}{f_Y(y)}$$

$$=\frac{f_{X_1Y}(x_1,y_1,y^*)}{f_Y(y_1,y^*)}\ \text{where } y=(y_1,y^*) \text{ and } y^*=(y_2,y_3,y_4)$$

$$=\frac{f_{X_1Y}((x_1,y_1),y^*)}{f_Y(y_1,y^*)}$$

$$=\frac{f_{X_1Y_1}(x_1,y_1).f_{Y^*}(y^*)}{f_{Y_1}(y_1)f_{Y^*}(y^*)}\ \text{as } Y^* \text{ does not depend on } X_1 \text{ and } Y_1$$

$$=\frac{f_{X_1Y_1}(x_1,y_1)}{f_{Y_1}(y_1)}$$

$$=f_{X_1/Y_1}(x_1).$$

Therefore we can write $E_{X_1/Y}(x_1)=E_{X_1/Y_1}(x_1)$ by the definition of conditional expectation.

Let's say the distribution of X_1 given Y_1 is $Bin(n_1,p_1)$, and the distribution of X_2 given Y_1 is $Bin(n_2,p_2)$ where $n_1=y_1$, $n_2=y_1$ and p_1 and p_2 can be calculated as follows:

$$p_1=P(X_1\mid Y_1)=\frac{P(X_1\cap Y_1)}{p(Y_1)}=\frac{1/2}{\left(1/2+\theta^{(0)}/4\right)}=\frac{2}{2+\theta^{(0)}}. \quad (17)$$

$$p_2=P(X_2\mid Y_1)=\frac{P(X_2\cap Y_1)}{p(Y_1)}=\frac{\theta^{(0)}/4}{\left(1/2+\theta^{(0)}/4\right)}=\frac{\theta^{(0)}}{2+\theta^{(0)}}. \quad (18)$$

The expected value of X_1 given Y_1 and θ is:

$$x_1^{(0)}=E\left(X_1\mid Y_1\right)=n_1p_1=y_1\frac{1/2}{\left(1/2+\theta^{(0)}/4\right)}=\frac{2y_1}{2+\theta^{(0)}}. \quad (19)$$

Also, we can find the value of X_2 given Y_2 and θ is:

$$x_2^{(0)}=y_1-y_1\frac{1/2}{\left(1/2+\theta^{(0)}/4\right)}=y_1\frac{\theta^{(0)}/4}{\left(1/2+\theta^{(0)}/4\right)}=\frac{y_1\theta^{(0)}}{2+\theta^{(0)}}. \quad (20)$$

Therefore,

$$t^{(0)}=E\left[X_2\mid Y_1\right]|_{\theta=\theta^{(0)}}=\frac{y_1\theta^{(0)}}{2+\theta^{(0)}}=x_2^{(0)}$$

$$t^{(0)}=x_2^{(0)} \quad (21)$$

This completes the E-step on the first iteration.

M-Step: The M-step then takes the estimated complete data $\left(x_1^{(0)},x_2^{(0)},18,20,34\right)$ and estimate θ on the first iteration by choosing $\theta^{(1)}$ to be the value of θ. $\theta^{(1)}$ is obtained from the following equation.

$$t^{(0)}=\frac{\partial\log a(\theta)}{\partial c(\theta)}|_{\theta=\theta^{(1)}}$$

$$\log a(\theta)=-x_5\log\left(\frac{\theta}{4}\right)-(x_3+x_4)\log\left(\frac{1-\theta}{4}\right)$$

Therefore,

$$t^{(0)}=\frac{\partial\log a(\theta)}{\partial c(\theta)}|_{\theta=\theta^{(1)}}$$

$$=\frac{\partial\log a(\theta)}{\partial\theta}\frac{\partial\theta}{\partial c(\theta)}|_{\theta=\theta^{(1)}}$$

$$=\left[\frac{x_3+x_4}{(1-\theta)}-\frac{x_5}{\theta}\right](\theta)|_{\theta=\theta^{(1)}}$$

If we substitute $x_2^{(0)}$ instead of $t^{(0)}$ and $\theta=\theta^{(1)}$ then

$$\frac{x_2^{(0)}}{\theta^{(1)}}=\frac{x_3+x_4}{(1-\theta^{(1)})}-\frac{x_5}{\theta^{(1)}}$$

That is,

$$\theta^{(1)}=\frac{x_2^{(0)}+x_5}{x_2^{(0)}+x_3+x_4+x_5} \quad (22)$$

The new fit $\theta^{(1)}$ for θ is then updated for θ into the right sides of (eqn 19) and then (eqn 20) to produce updated values of $x_2^{(1)}$ and $t^{(1)}$ respectively. Note that we do not need to maximize the θ again and again since the MLE of θ has a closed form. Now if we substitute $x_2^{(1)}$ instead of $x_2^{(0)}$ in (eqn 22), it leads new fit $\theta^{(2)}$ for θ, and so on. It follows on so alternating the E- and M-steps on the (k+1)th iteration of the EM algorithm that

$$\theta^{(k+1)}=\frac{x_2^{(k)}+x_5}{x_2^{(k)}+x_3+x_4+x_5} \quad (23)$$

where

$$x_2^{(k)}=y_1-x_1^{(k)}=y_1\frac{\theta^{(k)}/4}{\left(1/2+\theta^{(k)}/4\right)}=\frac{y_1\theta^{(k)}}{2+\theta^{(k)}}=t^{(k)}. \quad (24)$$

Remark 4.3: The value of θ must be determined by the rules of thump, which is defined as follows: The ratio of successive deviations, say r^k,

$$r^k=\left(\theta^{(k+1)}-\theta^{(k)}\right)/\left(\theta^{(k)}-\theta^{(k-1)}\right)$$

is essentially constant or the log-likelihood function $\log(L(\theta^{(k)}))$ is a non-decreasing function.

In eqn. (24), if we substitute instead of $x_2^{(k)}=\frac{y_1\theta^{(k)}}{2+\theta^{(k)}}$, then we will get

$$\theta^{(k+1)}=\frac{\frac{y_1\theta^{(k)}}{2+\theta^{(k)}}+x_5}{\frac{y_1.\theta^{(k)}}{2+\theta^{(k)}}+x_3+x_4+x_5}$$

$$=\frac{y_1\theta^{(k)}+x_5(2+\theta^{(k)})}{y_1\theta^{(k)}+(x_3+x_4+x_5)(2+\theta^{(k)})}$$

$$= \frac{y_1\theta^{(k)} + y_4(2+\theta^{(k)})}{y_1\theta^{(k)} + (y_2+y_3+y_4)(2+\theta^{(k)})}.$$

The formula of $\theta^{(k+1)}$ in terms of y's is:

$$\theta^{(k+1)} = \frac{y_1\theta^{(k)} + y_4(2+\theta^{(k)})}{y_1\theta^{(k)} + (y_2+y_3+y_4)(2+\theta^{(k)})}. \tag{25}$$

We define a function of $\theta^{(k+1)}$ such that $\theta^{(k+1)} = M\left(\theta^{(k)}\right), \text{ for } k = 0,1,2,....,$ where the function $M(\theta^{(k)})$ is defined as:

$$M\left(\theta^{(k)}\right) = \frac{y_1\theta^{(k)} + y_4(2+\theta^{(k)})}{y_1\theta^{(k)} + (y_2+y_3+y_4)(2+\theta^{(k)})}$$

$$= \frac{159\theta^{(k)} + 68}{197\theta^{(k)} + 144} \quad if \ \left(y_1, y_2, y_3, y_4\right) = \left(125,18,20,34\right).$$

Since $\theta^{(k+1)}$ converges to some point, say $\theta^{(*)}$ and $M(\theta^{(k)})$ is monotonically increasing and continuous function, then $\theta^{(*)}$ must satisfy the condition $\theta^{(*)}=M(\theta^{(*)})$, [3].

On putting $\theta^{(k+1)}={}^{(k)}=\theta^{(*)}$ in the equation $\theta^{(k+1)}=M(\theta^{(k)})$, we can explicitly solve the resulting quadratic equation in $\theta^{(*)}$ to confirm that the sequence of EM iterates $\{\theta^{(k)}$, irrespective of the starting value $\theta^{(0)}$.

That is,

$$\theta^{(*)} = \frac{159\theta^{(*)} + 68}{197\theta^{(*)} + 144}.$$

The quadratic equation in $\theta^{(*)}$ is $197\theta^{(*)2}$-$15\theta^{(*)}$-68=0. Since $0<\theta^{(*)}<1$, the value of $\theta^{(*)}$ is 0.6268, which is MLE of θ.

(iii) Estimating the standard deviation of MLE

In this subsection, we are concentrated on estimating the standard deviation of MLE using simulation. There are different methods we used in order to estimate the standard deviation. For example, we used Louis' method, the binomial approach, the second derivative of complete data and incomplete data and bootstrap methods as well. In addition, we have calculated the absolute relative percentage (ARP) based on Louis' method and then compare those values to find out which method would give the best estimate of standard deviation for the MLE based on Louis' method. The ARP is defined by:

$$APR = \frac{|\,Louis'\ Method - Method\,|}{Louis'\ Method} * 100\%.$$

However, there are several other criteria to evaluate our methods such as mean absolute error (MSE), root mean squared error (RMSE) and bias. Taking the square root of the average squared errors has some interesting implications for RMSE. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE should be more useful when large errors are particularly undesirable. In order to use Louis' method, we have to calculate the observed information $\mathcal{I}_Y(\hat{\theta})$ using the formula as mentioned in section 3. Since we know the log-likelihood function of θ from eqn. (16), we can directly calculate the $S_c(x, \theta)$ and $B_c(x, \theta)$, and then find $\mathcal{I}_X$ and $\mathcal{I}_{X|Y}$ as follows:

$$S_c(x,\theta) = \frac{\partial \log(L_c(x,\theta))}{\partial\theta} = \frac{x_2+x_5}{\theta} - \frac{x_3+x_4}{(1-\theta)}$$

$$B_c(x,\theta) = -\frac{\partial S_c(x,\theta))}{\partial\theta} = \frac{x_2+x_5}{\theta^2} + \frac{x_3+x_4}{(1-\theta)^2}$$

$$\mathcal{I}_X(\hat{\theta}) = E[B_c(X,\theta)\,|\,\theta=\hat{\theta},Y]$$

$$= E\left[\frac{X_2+X_5}{\theta^2} + \frac{X_3+X_4}{(1-\theta)^2}\right]|_{\theta=\hat{\theta}}$$

$$= \frac{\left(\frac{y_1\theta}{2+\hat{\theta}}\right)+X_5}{\theta^2} + \frac{X_3+X_4}{(1-\theta)^2}|_{\theta=\hat{\theta}} = 435.296 \ \ if \ \hat{\theta} = 0.6268$$

and

$$\mathcal{I}_{X|Y}(\hat{\theta}) = Cov[S_c(X,\theta)\,|\,Y]|_{\theta=\hat{\theta}}$$

$$= \frac{Var(X_2)}{\theta^2}|_{\theta=\hat{\theta}}$$

$$= \frac{1}{\theta^2}\frac{y_1\theta/4}{\left(1/2+\theta/4\right)}\frac{1/2}{\left(1/2+\theta/4\right)}|_{\theta=\hat{\theta}}$$

$$= \frac{1}{\theta^2}\frac{2y_1\theta}{\left(2+\theta\right)^2}|_{\theta=\hat{\theta}} = 57.804 \ \ if \ \hat{\theta} = 0.6268$$

Therefore the $\mathcal{I}_Y(\hat{\theta})$ can be calculated using the formula (3.3),

$$\mathcal{I}_Y(\hat{\theta}) = \mathcal{I}_X(\hat{\theta}) - \mathcal{I}_{X|Y}(\hat{\theta}) = 435.296 - 57.804 = 377.4917.$$

The MLE $\hat{\theta}$ is asymptotically distributed as:

$$\hat{\theta} \sim N\left(\theta,\ \frac{1}{\mathcal{I}_Y(\theta)}\right), \tag{26}$$

as $\hat{\theta}$ satisfies the following conditions.

(i) the dimension of the parameter space does not change with n

(ii) the distribution is exponential family

(iii) the range of Y does not depend on θ.

Where $\mathcal{I}_Y(\theta)$denotes the expected Fisher information from one observation [9] and it is defined as:

$$\mathcal{I}_Y(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\ln[Y\,|\,\theta\,]\,|\,Y \in R\right] \tag{27}$$

Because θ is unknown, we can plug in $\hat{\theta}$ to obtain an estimate the standard deviation, $SD(\hat{\theta}) \approx \sqrt{1/\mathcal{I}_Y(\hat{\theta})}$. Therefore the approximate standard deviation value for MLE is $1/\sqrt{377.4917} = 0.05147$.

Remark 4.4: *In this example, we have shown that*

$$\mathcal{I}_Y(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\ln[Y\,|\,\theta\,]\,|\,Y \in R\right] = \mathcal{I}_X(\theta) - \mathcal{I}_{X|Y}(\theta)$$

The first derivative of log-likelihood function w.r.t. θ is,

$$\frac{\partial \log\left(L_p(y;\theta)\right)}{\partial\theta} = \frac{y_1}{\theta+2} - \frac{(y_2+y_3)}{(1-\theta)} + \frac{y_4}{\theta}$$

and the expected value of the second derivative is:

$$-E\left[\frac{\partial^2 \log\left(L_p(y;\theta)\right)}{\partial\theta^2}\,|\,Y \in R\right] = \frac{y_1}{(\theta+2)^2} - \frac{(y_2+y_3)}{(1-\theta)^2} - \frac{y_4}{\theta^2}$$

$$= \left[\left(\frac{y_1}{\theta(2+\theta)} - \frac{2y_1\theta}{\theta^2(2+\theta)^2}\right) + \frac{(y_2+y_3)}{(1-\theta)^2} + \frac{y_4}{\theta^2}\right]$$

$$= \left[\left(\frac{y_1}{\theta(2+\theta)} + \frac{y_4}{\theta^2} + \frac{(y_2+y_3)}{(1-\theta)^2}\right) - \frac{2y_1\theta}{\theta^2(2+\theta)^2}\right]$$

$$= \left[\mathcal{I}_X(\theta) - \mathcal{I}_{X|Y}(\theta)\right].$$

There was another method to estimate the standard derivation of MLE in the discussion on the paper Dempster et al. [3] by Professor Cedric A. B. Smith. The standard deviation of MLE, $\hat{\theta}$ is:

$$SD(\hat{\theta}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n^*(1-\lambda)}} \quad (28)$$

where n^* can be estimated as:

$$n^* = \hat{X}_2 + Y_2 + Y_3 + Y_4 = Y1\left(\frac{\hat{\theta}}{2+\hat{\theta}}\right) + Y_2 + Y_3 + Y_4 \quad (29)$$

and λ is calculated as follows:

For each iterations in the EM algorithms find the ratio as shown in the Table 1. After a few iterations, we can see that the ratio does not change, which is denoted as λ.

Since the EM algorithm provides non-decreasing maximum likelihood function, the convergence criteria can be subjected. Relative change should be computed in the magnitude of 10^{-3},10^{-4},10^{-5}, etc. If the consecutive values provide same estimate, then the middle value should be used as a criteria for convergence. In our example, we expand with 10^{-3},10^{-4} and 10^{-5} and we recommend the magnitude of 10^{-3}.

The following Tables 2 and 3 show the results from our simulation studies for the θ values of 0.3, 0.4, 0.5, 0.6, 0.6268, 0.7, 0.8 and 0.9 to identify which the method would give the best estimate for SD of MLE as good as Louis' estimates. In other words, we have estimated the SD of MLE using several methods for all θ (Table 2) and find the ARP values (Table 3) for each approach based on Louis' method. As we can see the simulation results in the tables, the methods binomial approach (Method I), bootstrap for incomplete data (Method II) and the second derivative of incomplete data w.r.t θ (Method III) give best estimates while all methods from complete data (Method IV-IX) give worst estimates. However, the Method I requires the distribution of $X_2|Y_1$ is known and MLE from the incomplete data exists, and the Method III requires the second derivative of incomplete data w.r.t θ exists.

Iteration (I)	$\theta^{(I)}$	$\left(\theta^{(I+1)}-\hat{\theta}\right)/\left(\theta^{(I)}-\hat{\theta}\right)$
1	0.300	0.175
2	0.570	0.139
3	0.619	0.134
4	0.626	0.133
5	0.627	0.133

Table 1: The EM iterations for the example and λ value.

Although in this example, computing time were very smaller (and non-issue due to close form solution). In other situations, EM algorithm may take more time. However, this is not a issue due to very advance computational method.

Note that in this example, the likelihood based on incomplete data and complete data is straight forward. Therefore, it is easier to compute difference variance estimates. When incomplete data likelihood function is not closed form, this computation may not be possible.

Discussions and Conclusions

Generally, the Louis' method is one of the best methods to estimate the standard deviation of MLE when the observations can be viewed as incomplete data. If the bootstrap approach for incomplete data (Method II) is properly done as we described, then this method would give the best estimate for the standard deviation of MLE and be as good as Louis' method estimates. We also calculated the estimate of standard deviation of MLE using nine different possible methods to show that Method II is the best approach to estimate the standard deviation

Theta	Simulation	Method I: Binomial Approach	Method II: Bootstrap for Incomplete Data Y	Method III: Using Second Derivative of Y: theta is not fixed	Method IV: Complete Data using V(E())+E(V())	Method V: Bootstrap Complete Data X (SD)	Method VI: Using Second Derivative of X: theta is not fixed	Method VII: Complete Data using E(V(X))	Method VIII: Bootstrap for Complete Data X (MLE)	Method IX: Using Second Derivative of X: theta is fixed
0.3	5,000	0.73%	0.32%	1.06%	8.40%	17.07%	18.67%	30.27%	30.27%	31.25%
	10,000	0.66%	0.60%	0.91%	8.14%	16.93%	18.55%	30.84%	30.84%	31.2 1%
	20,000	0.82%	0.45%	0.95%	8.61%	16.92%	18.53%	30.12%	30.12%	31.17%
0.4	5,000	0.43%	0.30%	0.64%	14.40%	13.69%	22.54%	24.96%	24.96%	27.58%
	10,000	0.29%	0.82%	0.68%	14.70%	13.70%	22.49%	24.49%	24.49%	27.51%
	20,000	0.34%	0.25%	0.65%	14.40%	13.66%	22.47%	25.00%	25.00%	27.51%
0.5	5,000	0.10%	0.44%	0.59%	19.40%	10.95%	27.19%	20.50%	20.50%	23.98%
	10,000	0.18%	1.52%	0.56%	20.44%	10.95%	27.18%	18.95%	18.95%	23.99%
	20,000	0.05%	0.44%	0.63%	19.89%	10.94%	27.17%	19.76%	19.76%	23.96%
0.6	5,000	0.14%	0.03%	0.79%	24.52%	8.65%	32.00%	15.42%	15.42%	20.13%
	10,000	0.22%	0.05%	0.86%	24.40%	8.71%	32.09%	15.54%	15.54%	20.18%
	20,000	0.32%	0.14%	0.74%	24.56%	8.63%	32.01%	15.37%	15.37%	20.15%
0.5268	*5,000*	0.05%	0.67%	0.74%	25.32%	8.01%	33.25%	14.95%	14.95%	19.06%
	10,000	0.32%	0.29%	0.92%	25.45%	8.15%	33.40%	14.61%	14.61%	19.12%
	20,000	0.34%	0.56%	0.85%	25.31%	8.11%	33.39%	14.85%	14.85%	19.13%
0.7	5,000	0.18%	1.01%	1.23%	28.15%	6.87%	37.2 1%	12.06%	12.06%	16.20%
	10,000	0.13%	1.08%	1.13%	28.17%	6.79%	37.13%	12.12%	12.12%	16.17%
	20,000	0.19%	0.47%	1.12%	28.55%	6.79%	37.13%	11.57%	11.57%	16.18%
0.8	5,000	1.18%	0.96%	2.29%	31.68%	5.90%	42.46%	8.15%	8.15%	11.81%
	10,000	0.84%	2.39%	2.07%	30.89%	5.71%	42.38%	9.47%	9.47%	11.80%
	20,000	1.08%	0.90%	2.01%	31.84%	5.67%	42.34%	8.15%	11.80%	11.80%
0.9	5,000	1.10%	3.83%	4.11%	32.51%	5.97%	47.89%	7.48%	7.48%	6.86%
	10,000	0.78%	4.83%	3.43%	32.27%	5.33%	47.60%	8.46%	8.46%	6.82%
	20,000	0.78%	3.72%	3.52%	32.93%	5.41%	47.64%	7.43%	7.43%	6.82%

Table 2: Results of Louis' method and bootstrap for the example on estimation of standard deviation.

Theta	Simulation	Louis' Method	Using Second Derivative of Y: theta is fixed	Method I: Bionomial Approach	Method II: Bootstrap for Incomplete Data Y	Method III: Using Secondl Derivative of Y: theta is not fixed	Method IV : Complete Data u.sing V(E())+E(V())	Method V: Bootstrap Complete Data X (SD)	Method VI: Using Second Derivative of X: theta is not fixed	Method VII: Complete Data using E(V(X))	Method VIII: Bootstrap for Complete Data X(MLE)	Method IX : Using Second Derivative of X: theta is fixed
0.3	5,000	0.055720	0.055720	0.055315	0.055896	0.055130	0.060402	0.0462 10	0.045316	0.038854	0.038854	0.038308
	10,000	0.055644	0.055644	0.055275	0.055313	0.055140	0.060176	0.046222	0.045322	0.038485	0.038485	0.038280
	20,000	0.0556 15	0.0556 15	0.055157	0.055865	0.055089	0.060404	0.046205	0.045307	0.038862	0.038863	0.038281
0.4	5,000	0.057228	0.057228	0.056980	0.057399	0.056861	0.065469	0.04939 1	0.044327	0.042946	0.042946	0.041442
	10,000	0.057162	0.057162	0.056994	0.057632	0.056775	0.065567	0.049333	0.044306	0.043162	0.043162	0.041438
	20,000	0.057149	0.057149	0.056956	0.057292	0.056775	0.065378	0.049344	0.044306	0.042861	0.042861	0.041429
0.5	5,000	0.056484	0.056484	0.056430	0.056233	0.056150	0.067440	0.050300	0.041127	0.044905	0.044905	0.042937
	10,000	0.056491	0.056491	0.05639 1	0.057347	0.056174	0.068035	0.050306	0.041135	0.045786	0.045786	0.042938
	20,000	0.056455	0.056455	0.056424	0.056701	0.056101	0.067685	0.050276	0.041115	0.045298	0.045298	0.042930
0.6	5,000	0.053832	0.053832	0.053755	0.053818	0.053407	0.067032	0.049174	0.036604	0.045532	0.045532	0.042993
	10,000	0.053880	0.053880	0.054000	0.053853	0.053418	0.067026	0.049189	0.036592	0.045507	0.045507	0.043008
	20,000	0.053847	0.053847	0.05402 1	0.053923	0.053449	0.067074	0.049198	0.036613	0.045568	0.045568	0.042995
0.6268	*5,000*	*0.052772*	*0.052772*	*0.052800*	*0.052419*	*0.052383*	*0.066132*	*0.048545*	*0.035227*	*0.044884*	*0.044884*	*0.042711*
	10,000	*0.052846*	*0.052846*	*0.052676*	*0.052694*	*0.052361*	*0.066294*	*0.048541*	*0.035195*	*0.045127*	*0.045127*	*0.042741*
	20,000	*0.052849*	*0.052849*	*0.052668*	*0.052555*	*0.052401*	*0.066226*	*0.048562*	*0.035202*	*0.045002*	*0.045002*	*0.042742*
0.7	5,000	0.049328	0.049328	0.049241	0.048831	0.048722	0.0632 15	0.045937	0.030972	0.043380	0.043380	0.041336
	10,000	0.049290	0.049290	0.049224	0.048758	0.048735	0.063173	0.045941	0.030990	0.043315	0.043316	0.041318
	20,000	0.049290	0.049290	0.049197	0.049057	0.048738	0.063362	0.045941	0.030989	0.043588	0.043588	0.041317
0.8	5,000	0.042413	0.042413	0.041914	0.042006	0.041443	0.055851	0.0399 11	0.024405	0.038958	0.038958	0.037404
	10,000	0.042384	0.042294	0.042029	0.041372	0.041505	0.055477	0.039963	0.024423	0.038369	0.038369	0.037385
	20,000	0.042376	0.042376	0.041920	0.041995	0.041523	0.055868	0.039973	0.024432	0.03892 1	0.037376	0.037376
0.9	5,000	0.031703	0.031703	0.031356	0.030489	0.030401	0.042009	0.029809	0.01652 1	0.029332	0.029332	0.029528
	10,000	0.031610	0.031610	0.031364	0.030083	0.030527	0.041810	0.029925	0.016564	0.028935	0.028935	0.029454
	20,000	0.031622	0.031622	0.031376	0.030447	0.030509	0.042033	0.029909	0.016558	0.029272	0.029272	0.029464

Table 3: ARP values of Louis' method and bootstrap for the example on estimation of standard deviation (ARP: Absolute Relative Percentage).

of MLE. In addition, we can evaluate our method using bootstrap confidence interval [10]. Both results will not be contradicted. But, Louis' method requires the conditions $S_p(y, \theta)=0$ and the second derivative for complete data w.r.t θ to exist. In other words, if MLE of θ is not an unbiased estimator, then Louis' method can lead to large mean squared error (MSE). In this situation, we can generate an empirical distribution of $\hat{\theta}$ using a bootstrap approach, and one can estimate the better confidence interval for $\hat{\theta}$. Specially, in the design of clinical trials, we want to compute the confidence interval for $\hat{\theta}$ under the alternative hypothesis to validate for type II error and under null hypothesis to validate for type I error.

If y is a binary variable, then there are many situations in which likelihood function has no unique maximum, in which case we can say that the MLE does not uniquely exist. As a result, we will have a problem with separation, i.e., specificity and sensitivity.

Our proposed approach (Method II) has wider application of automating variance and confidence interval estimation for many clinical studies with missing data, for example, solve the missing values issues in clinical trials [11], interval censor data [12], regression with missing predictors [13], machine learning-based missing [14], and missing data in longitudinal studies [15-19].

Acknowledgement

This work was partially supported by the National Institute of Environmental Health (P42ES023716-PI: Srivastava) and the National Institute of Health (1P20 GM113226-PI: McClain). Dr. SN Rai received additional support from the Wendell Cherry Chair in Clinical Trial Research.

References

1. Louis TA (1982) Finding the observed information matrix when using the EM algorithm. J R Statist Soc 44: 226-233.
2. Hartley HO, Hocking RR (1971) The analysis of incomplete data. Biometrics 14: 174-194.
3. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood estimation from incomplete data via the EM algorithm. J R Statist Soc 39: 1-38.
4. Efron B, Hinkley DV (1978) The observed versus expected information. Biometrika 65: 457-487.
5. Meng XL, Rubin DB (1991) Using EM algorithm to obtain asymptotic variance-covariance matrices; the SEM algorithm. J Am Statist Ass 86: 899-909.
6. Woodbury MA (1971) Discussion of paper by Hartley and Hocking. Biometrics 27: 808-817.
7. Tanner MA (1993) Tools for statistical inference (2ndedn), Springer verlag, New York.
8. Tweedie MCK (1945) Inverse Statistical Variates. Nature 155: 453.
9. Millar RB (2011) Maximum likelihood estimation and inference, A John Wiley & Sons, Inc., United Kingdom.
10. Efron B, Tipshirani R (1986) Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. Stat Sci 1: 54-75.
11. Jane CL, Louise MR (1998) Tutorial in Biostatistics Methods for Interval Censored Data. Statistics in Medicine 17: 219-238.
12. Rai SN, Matthews DE, Krewski DR (2000) Mixed-Scale Models for Survival/ Sacrifice Experiments. Can J Stat 1: 65-80.
13. Little RJA, Rubin DB (1987) Statistical Analysis with Missing Data, New York: Wiley & Sons, Inc.
14. Mostafizur M, Davis DN (2013) Machine Learning-Based Missing Value

Imputation Method for Clinical Datasets. IAENG Transactions on Engineering Technologies pp: 245-257.

15. Laird NM (1988) Missing Data in Longitudinal Studies. Statistics in Medicine 7: 303-315.
16. Little RJA (1982) Regression with missing X's: a review. J Am Assoc 87: 1227-1237.
17. McLachlan GF, Krishnan T (2008) The EM Algorithms and Extensions (2ndedn), A John Wiley & Sons, Inc., New Jersey.
18. Smith ABC (1969) Biomathematics, Hafner Publishing Company, New York.
19. Sunderberg R (1974) Maximum likelihood theory for incomplete data from an exponential family. Scand J Statist 1: 49-58.