Fetal Weight Estimation In Case Of Missing Data

**Loc Nguyen1**\***, Thu-Hang T. Ho2**

1 Board of Directors, Sunflower Soft Company, An Giang, Vietnam

2 Board of Directors, Vinh Long General Hospital, Vinh Long, Vietnam

**Email Address**

[ng\_phloc@yahoo.com](mailto:ng_phloc@yahoo.com) (Loc Nguyen), [bshangvl2000@yahoo.com](mailto:bshangvl2000@yahoo.com) (Thu-Hang T. Ho)

\*Correspondence: [Timothy.Schmutte@yale.edu](file:///E:\文章\文章\ITS精排\ITS%2020180502%20精排文章\EM1003\Timothy.Schmutte@yale.edu)

**Received:** 2017-12-06**; Accepted:** 2017-12-07**; Published:** 2017-12-08

**Abstract:**

Fetal weight estimation before delivery is important in obstetrics, which assists doctors diagnose abnormal or diseased cases. Linear regression based on ultrasound measures such as bi-parietal diameter (*bpd*), head circumference (*hc*), abdominal circumference (*ac*), and fetal length (*fl*) is common statistical method for weight estimation. In this research, we proposed a so-called regression expectation maximization (REM) algorithm which is a combination of maximum likelihood estimation (MLE) method and expectation maximization (EM) method to construct the regression model in case that both ultrasound measures and fetal weight are missing. The special technique in REM is to build parallelly an entire regression function and many partial inverse regression functions for solving the problem of highly sparse data, in which missing values are filled in by expectations relevant to both entire regression function and inverse regression functions. Experimental results proved withstanding of REM for incomplete data, in which accuracy of REM decreases insignificantly when data sample is made sparse with high loss ratio 80%.

**Keywords:**

Fetal Weight Estimation, Regression Model, Ultrasound Measures, Expectation Maximization Algorithm, Missing Data.

1. **Introduction**

|  |  |
| --- | --- |
|  | (99) |

According to the regression approach of fetal weight estimation, without loss of generality, an estimation formula is a linear regression function *Z = α*0 *+ α*1*X*1 *+ α*2*X*2 *+ … + αnXn* where *Z* is estimated fetal weight whereas *Xi* (s) are gestational ultrasound measures such as bi-parietal diameter (*bpd*), head circumference (*hc*), abdominal circumference (*ac*), fetal length (*fl*). Variable *Z* is called response variable or dependent variable. Each *Xi* is called regression variable, regressor, predictor, regression variable, or independent variable. Each *αi* is called regression coefficient. In previous research (Nguyen & Ho, 2018) we survey many researches related to the regression approach. Hence, we focus on applying expectation maximization (EM) algorithm into constructing regression model. We proposed a so-called regression expectation maximization (REM) algorithm to learn linear regression function from incomplete data in which some values of *Z* and *Xi* are missing. Because this research is the successive one after our previous research (Nguyen & Ho, 2018), they share some common contents, but we confirm that their methods are different. The algorithm in the previous research is dual regression expectation maximization (DREM) algorithm. DREM only accepts incomplete *Z* but REM accepts both incomplete *Z* and incomplete *X*.

There is a demand to construct regression model in case of missing data because taking ultrasound examinations is a hard task and early weight estimation is necessary in some cases (Nguyen & Ho, 2018). EM algorithm is an approach to solve the problem of incomplete data in regression analysis. Here we browse some researches relevant to EM algorithm and regression model. Kokic (Kokic, 2002) proposed an excellent method to calculate expectation of errors for estimating coefficients of multivariate linear regression model. In Kokic’s method, response variable *Z* has missing values. Ghitany, Karlis, Al-Mutairi, and Al-Awadhi (Ghitany, Karlis, Al-Mutairi, & Al-Awadhi, 2012) calculated the expectation of function of mixture random variable in the expectation step of EM algorithm and then used such expectation for estimating parameters of multivariate mixed Poisson regression model in the maximization step. Anderson and Hardin (Anderson & Hardin, 2013) used reject inference technique to estimate coefficients of logistic regression model when response variable *Z* is missing but characteristic variables (regressors *Xi*) are fully observed. Anderson and Hardin replaced missing Z by its conditional expectation on regressors *Xi* where such expectation is logistic function. Zhang, Deng, and Su (Zhang, Deng, & Su, 2016) used EM algorithm to build up linear regression model for studying glycosylated hemoglobin from partial missing data. In other words, they aim to discover relationship between independent variables (predictors) and diabetes.

Besides EM algorithm, there are other approaches to solve the problem of incomplete data in regression analysis. Haitovsky (Haitovsky, 1968) stated that there are two main approaches to solve such problem. The first approach is to ignore missing data and to apply the least squares method into observations. The second approach is to calculate covariance matrix of regressors and then to apply such covariance matrix into constructing the system of normal equations. Robins, Rotnitzki, and Zhao (Robins, Rotnitzki, & Zhao, 1995) proposed a class of inverse probability of censoring weighted estimators for estimating coefficients of regression model. Their approach is based on the dependency of mean vector of response variable *Z* on vector of regressors *Xi* when *Z* has missing values. Robins, Rotnitzki, and Zhao (Robins, Rotnitzki, & Zhao, 1995) assumed that the probability *λit*(*α*) of existence of *Z* at time point *t* is dependent on existence of *Z* at previous time point *t*–1 but independent from *Z*. Even though *Z* is missing, the probability *λit*(*α*) is also determined and so the coefficients *β* is calculated based on the inverse of *λit*(*α*) and *Xi*. The inverse of *λit*(*α*) is considered as weight for complete case. Robins, Rotnitzki, and Zhao used additional time-dependent covariates *Vit* to determine *λit*(*α*).

In the article “Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models”, Horton and Ken P. Kleinman (Horton & Kleinman, 2007) classified 6 methods of regression analysis in case of missing data such as complete case method, ad-hoc method, multiple imputation, maximum likelihood, weighting method, and Bayesian method. EM algorithm belongs to maximum likelihood method.

In general, the ideology of applying EM algorithm into regression model is not new but our proposed REM algorithm can build up regression models in case that both response variable *Z* and regressors *Xi* have missing values. In other words, REM accepts highly sparse data. From experimental results, the accuracy of REM decreases insignificantly when data sample is made sparse with high loss ratio 80%. The special technique in REM is to build parallelly an entire regression function *Z* = *α*0 *+ α*1*X*1 *+ α*2*X*2 *+ … + αnXn* and many partial inverse regression functions *Xj* = *βj*0 *+ βj*1*Z* for solving the problem of highly sparse data, in which missing values are filled in by expectations relevant to both entire regression function and inverse regression functions. Such expectations are re-estimated by a so-called balance process until their bias is small enough.

1. **Methodology**

Suppose we estimate the linear regression model *Z* = *α*0 *+ α*1*X*1 *+ α*2*X*2 *+ … + αnXn* where *Z* is fetal weight and *Y* is fetal age whereas *Xi* (s) are gestational ultrasound measures such as *bpd*, *hc*, *ac*, and *fl*. Suppose the random variable *Z* conforms normal distribution, according to equation (1) (Lindsten, Schön, Svensson, & Wahlström, 2017, pp. 8-9). Note, *Z* is random variable whereas *X* is data in equation (1).

|  |  |
| --- | --- |
|  | (1) |

Where *α* = (*α*0, *α*1,…, *αn*)*T* is parameter vector and *X* = (1, *X*1, *X*2,…, *Xn*)*T* is data vector. The mean and variance of *Z* with regard to *P*(*Z* | *X*, *α*) are *αTX* and *σ*2, respectively. The superscript “*T*” denotes transposition operator in vector and matrix. Suppose each has an inverse linear regression model *Xj* = *βj*0 *+ βj*1*Z.* In other words, *Zj* now is considered as the random variable conforming normal distribution according to equation (2).

|  |  |
| --- | --- |
|  | (2) |

Where *βj* = (*βj*0, *βj*1)*T* is a partial parameter vector and (1, *Z*)*T* is a partial data vector. The mean and variance of each *Xj* with regard to the inverse distribution *Pj*(*Xj* | *Z*, *βj*) are *βjT*(1, *Z*)*T* and *σj*2, respectively. Of course, there are *n* inverse linear regression models.

Let ***D*** = (***X***, ***z***) be collected sample in which ***X*** is a set of sample measures and ***z*** is a set of fetal weights with note that both ***X*** and ***z*** are incomplete. In other words, ***X*** and ***z*** have missing values. Now we focus on estimating *α* and *βj* based on ***D***. As a convention let *α\** and *βj\** be estimates of *α* and *βj*, respectively (Lindsten, Schön, Svensson, & Wahlström, 2017, p. 8).

|  |  |
| --- | --- |
|  | (3) |

The expectation of sufficient statistic *Z* regard to the entire linear model *P*(*Z* | *X*, *α*) is specified by equation (4).

|  |  |
| --- | --- |
|  | (4) |

The expectation of each sufficient statistic *Xj* with regard to each inverse linear model *Pj*(*Xj* | *Z*, *βj*) is specified by equation (5).

|  |  |
| --- | --- |
|  | (5) |

Please pay attention to equations (4) and (5) because *Z* and *Xj* will be estimated by these expectations later.

By applying sample ***D*** into equations (1) and (2) and using maximum likelihood estimation (MLE) method, we retrieve equation (6) to estimate *α\** and *βj\** (Lindsten, Schön, Svensson, & Wahlström, 2017, pp. 8-9).

|  |  |
| --- | --- |
|  | (6) |

Where ***X***, ***z***, ***Z***, and ***x****j* are specified in equation (3). Because ***X*** and ***Z*** are incomplete, we apply expectation maximization (EM) algorithm into estimating (*α\**, *βj\**)*T*. EM algorithm has many iterations and each iteration has expectation step (E-step) and maximization step (M-step) for estimating parameters. Given current parameter Θ(*t*) = (*α*(*t*), *βj*(*t*))*T* at the *t*th iteration, missing values *zi* and *xij* are calculated in E-step so that ***X*** and ***Z*** become complete. In M-step, the next parameter Θ(*t*+1) = (*α*(*t*+1), *βj*(*t*+1))*T* is determined by equation (6) and the complete data ***X*** and ***Z***.

The most important problem in our research is how to estimate missing values *zi* and *xij*. Recall that every missing value *zi* is estimated as the expectation based on the current parameter *α*(*t*), according to equation (4).

Let *Ui* be a set of indices of missing values *xij*. In other words, if then, *xij* is missing. The set *Ui* can be empty. The equation (4) is written:

Note, *xi*0 = 1. According to equation (5), missing value *xij* is estimated by:

Combining equation (4) and equation (5), we have:

It implies:

|  |  |
| --- | --- |
|  | (7) |

Missing values *zi* and *xij* are estimated by the balance process shown in Table 1.

**Table 1.** Balance process for estimating missing values

|  |
| --- |
| 1. Step 1: Missing values *zi* are estimated by equation (7), based on the current parameter Θ(*t*) = (*α*(*t*), *βj*(*t*))*T*.   Missing values *xij* where are estimated by equation (5) and the estimated values *zi* above, based on the current parameter Θ(*t*) = (*α*(*t*), *βj*(*t*))*T*.   1. Step 2: For balancing both *P*(*Z* | *X*, *α*) and *Pj*(*Xj* | *Z*, *βj*) in estimation, values *zi* and *xij* are re-estimated by equations (4) and (5) as new *zi*’ and *xij*’, based on the current parameter Θ(*t*) = (*α*(*t*), *βj*(*t*))*T*. 2. Step 3: If the deviation between (*zi*’, *xij*’) and (*zi*, *xij*) is smaller than a small enough threshold or the process reaches a large enough number of iterations, the process stops; at that time *zi*’ and *xij*’ are final estimated values. Otherwise, going back step 2 with assignment *xij* = *xij*’. |

In fact, the balance process is an iterative process which is a combination of equations (4), (5), and (7). The process starts to estimate missing values *zi* without use of *xij*. Conversely, the process can start to estimate missing values *xij* without use of *zi*, which is called inverse balance process.

Recall that *Ui* is the set of indices of missing values *xij*. Every missing value *xil* is estimated as the expectation based on the current parameter *βj*(*t*), according to equation (5).

According to equation (4), missing value *zi* is estimated by:

Combining equation (5) and equation (4), we have:

In other words, we have:

Where,

Suppose the cardinality of *Ui* is *k*, which means that there is *k* missing values *xij*. Derived from the combination above, missing values are solution of the following system of *k* equations.

Therefore, missing values *xij* are calculated by equation (8) according to Cramer method.

|  |  |
| --- | --- |
|  | (8) |

Where,

Table 2 shows the inverse balance process.

**Table 2.** Inverse balance process of missing values

|  |
| --- |
| 1. Step 1: Missing values *xij* where are estimated by equation (8), based on the current parameter Θ(*t*) = (*α*(*t*), *βj*(*t*))*T*. Missing values *zi* are estimated by equation (7), based on the current parameter Θ(*t*) = (*α*(*t*), *βj*(*t*))*T*.   Missing values *zi* are estimated by equation (4) and the estimated values *xij* above, based on the current parameter Θ(*t*) = (*α*(*t*), *βj*(*t*))*T*.   1. Step 2: For balancing both *P*(*Z* | *X*, *α*) and *Pj*(*Xj* | *Z*, *βj*) in estimation, values *xij* and *zi* are re-estimated by equations (5) and (4) as new *xij*’ and *zi*’, based on the current parameter Θ(*t*) = (*α*(*t*), *βj*(*t*))*T*. 2. Step 3: If the deviation between (*zi*’, *xij*’) and (*zi*, *xij*) is smaller than a small enough threshold or the process reaches a large enough number of iterations then, the process stops; at that time *zi*’ and *xij*’ are final estimated values. Otherwise, going back step 2 with assignment *zi* = *zi*’. |

In fact, the inverse balance process is an iterative process which is a combination of equations (4), (5), and (8). We use the balance process shown in Table 1 for experiments in this research although balance process and inverse balance process are exchangeable. As a result, EM algorithm (Dempster, Laird, & Rubin, 1977, p. 4) associated with the balance process for regression model is shown in Table 3. This is our so-called Regression Expectation Maximization (REM) algorithm.

**Table 3.** Regression Expectation Maximization (REM) Algorithm

|  |
| --- |
| 1. E-step: Missing values *zi* and *xij* are estimated by the balance process shown in Table 1 or Table 2. The balance process is the core of REM. 2. The next parameter Θ(*t*+1) = (*α*(*t*+1), *βj*(*t*+1))*T* is determined by equation (6) and the complete data ***X*** and ***Z*** fulfilled in E-step. |

The REM algorithm stops if at some *t*th iteration, we have Θ(*t*) = Θ(*t*+1) = Θ*\**. At that time, Θ*\** = (*α\**, *β\**)*T* is the optimal estimate of EM algorithm. In practice, the algorithm can stop if deviation between Θ(*t*) and Θ(*t*+1) is smaller than a small enough threshold or REM reaches a large enough number of iterations. The smaller the terminated threshold is, the more accurate REM is. REM uses both the terminated threshold *ε* = 0.1% = 0.001 and the maximum number of iterations *maximum-iteration* = 1000. The parameter *maximum-iteration* = 1000 prevents REM from running for a long time.

An technique to improve the convergence of REM is to initialize the parameter Θ(1) = (*α*(1), *β*(1))*T* at the first iteration of EM process in proper way instead of initializing Θ(1) in arbitrary way (Nguyen & Ho, 2018). As usual, Θ(1) is initialized arbitrarily. Let ***X***’ be the complete matrix of ultrasound measures, which is created by removing all rows whose values are missing from ***X***. Similarly, let ***Z***’ be the complete matrix of fetal weights, which is created by removing rows whose weights are missing from ***Z***. The advanced Θ(1) = (*α*(1), *β*(1))*T* is initialized by equation (9).

|  |  |
| --- | --- |
|  | (9) |

Where ***z***’ is the complete vector of non-missing weights and ***x****j*’ is the complete vector of non-missing measures. Equation (9) is variant of equation (6) where ***X***, ***Z***, ***x****j*, and ***z*** are replaced by ***X***’, ***Z***’, ***x****j*’, and ***z***’.

1. **Results and Discussion**

We uses the gestational sample of 127 cases in which each case includes ultrasound measures, fetus age, and fetus weight. Ultrasound measures are bi-parietal diameter (*bpd*), head circumference (*hc*), abdominal circumference (*ac*), and fetal length (*fl*). The unit of *bpd*, *hc*, *ac*, and *fl* is millimeter whereas the unit of fetal weight is gram. Ho and Phan (Ho & Phan, 2011), (Ho & Phan, 2011) collected the sample of pregnant women at Vinh Long General Hospital – Vietnam with obeying strictly all medical ethical criteria. The dataset is split into two folders and each folder owns one training dataset and one testing dataset. Later on, the training dataset is made sparse with sparse ratios 0.2, 0.4, 0.6, and 0.8, which is as same as our previous research (Nguyen & Ho, 2018). There are ten testing pairs of complete and incomplete training datasets and testing datasets according to Table 4 (Nguyen & Ho, 2018).

**Table 4.** Ten testing pairs

|  |  |  |  |
| --- | --- | --- | --- |
| Pair | Training dataset | Testing dataset | Sparse ratio |
| 1 | *sample1.base* | *sample1.test* | 0% |
| 2 | *sample2.base* | *sample2.test* | 0% |
| 3 | *sample1.base.0.2.miss* | *sample1.test* | 20% |
| 4 | *sample2.base.0.2.miss* | *sample2.test* | 20% |
| 5 | *sample1.base.0.4.miss* | *sample1.test* | 40% |
| 6 | *sample2.base.0.4.miss* | *sample2.test* | 40% |
| 7 | *sample1.base.0.6.miss* | *sample1.test* | 60% |
| 8 | *sample2.base.0.6.miss* | *sample2.test* | 60% |
| 9 | *sample1.base.0.8.miss* | *sample1.test* | 80% |
| 10 | *sample2.base.0.8.miss* | *sample2.test* | 80% |

The 1st and 2nd pairs are called completed pairs whereas 3rd, 4th, 5th, 6th, 7th, 8th, 9th, and 10th are called incomplete pairs. Experimental results from incomplete pairs are compared together and are aligned with experimental results from complete pairs in order to evaluate withstanding of REM for missing values. Table 5 shows ten regression models corresponding to ten testing pairs.

**Table 5.** Ten resulted regression models

|  |  |
| --- | --- |
| Pair | Regression model |
| 1 | weight = -5633.9620 + 44.6266\**bpd* + 1.8354\**hc* + 15.7559\**fl* + 9.2693\**ac* |
| 2 | weight = -5814.4685 + 48.2633\**bpd* + 2.0894\**hc* + 17.6452\**fl* + 8.1373\**ac* |
| 3 | weight = -5749.4868 + 40.8380\**bpd* + 2.0846\**hc* + 24.3397\**fl* + 8.6234\**ac* |
| 4 | weight = -5839.8749 + 42.2349\**bpd* + 2.1478\**hc* + 23.4239\**fl* + 8.6419\**ac* |
| 5 | weight = -6129.6996 + 47.2796\**bpd* + 3.7401\**hc* + 23.9699\**fl* + 6.5586\**ac* |
| 6 | weight = -5916.1902 + 39.3985\**bpd* + 1.6670\**hc* + 29.7328\**fl* + 8.8414\**ac* |
| 7 | weight = -5910.2794 + 51.1290\**bpd* + 0.1551\**hc* + 24.9170\**fl* + 8.0808\**ac* |
| 8 | weight = -5878.4885 + 30.1106\**bpd* + 3.5594\**hc* + 38.2013\**fl* + 7.6653\**ac* |
| 9 | weight = -5730.8654 + 35.1776\**bpd* + 1.4169\**hc* + 47.2740\**fl* + 5.9599\**ac* |
| 10 | weight = -6058.3298 - 2.3765\**bpd* + 17.6105\**hc* + 58.8048\**fl* - 1.0858\**ac* |

Now we assess such ten regression models with subject to two typical metrics such as mean absolute error (MAE) and correlation coefficient (R). Let *W* = {*w*1, *w*2,…, *wK*} and *V* = {*v*1, *v*2,…, *vK*} be sets of actual weights and estimated weights, respectively. Equation (10) specifies the MAE metric.

|  |  |
| --- | --- |
|  | (10) |

The smaller the MAE is, the more accurate the DREM is. Table 6 shows MAE metric which evaluates the ten models.

**Table 6.** MAE of ten models

|  |  |
| --- | --- |
| Pair | MAE |
| 1 | 164.3777 |
| 2 | 171.5367 |
| 3 | 166.7447 |
| 4 | 173.4081 |
| 5 | 169.3976 |
| 6 | 178.1556 |
| 7 | 165.2540 |
| 8 | 184.2218 |
| 9 | 176.5944 |
| 10 | 265.0925 |
| Average | 181.4783 |

Let *rMAEi* be the bias ratio of MAE between the pair *i*th and the pair 1th if *i* odd or the pair 2th if *i* even. For example, we have (Nguyen & Ho, 2018):

|  |  |
| --- | --- |
|  | (11) |

From equation (11), these bias ratios indicate withstanding of REM for incomplete data. For instance, the value *rMAE*3 = 0.0144 implies that the accuracy of dual REM decreases 1.44% when the completion of training dataset of the 3rd pair decreases 20%. The value *rMAE*4 = 0.0109 implies that the accuracy of REM decreases 1.09% when the completion of training dataset of the 4th pair decreases 20%. The bias ratios of the pairs 3rd (20% missing values), 5th (40% missing values), 7th (60% missing values), and 9th (80% missing values) are 1.44%, 3.05%, 0.53%, and 7.43%. It is concluded that such bias ratios are much smaller than percentages of missing values and so the withstanding of REM for missing values is significant. Like our previous research (Nguyen & Ho, 2018), we make a one-way paired t-test of *X* = {20%, 40%, 60%, 80%} and *Y* = {1.44%, 3.05%, 0.53%, 7.43%}. Given significant level 95%, the statistic *t*0 is calculated by equation (12) (Montgomery & Runger, 2010, p. 376).

|  |  |
| --- | --- |
|  | (12) |

Where,

Note that = 0.4689 and *sD* = 0.2394 are sample mean and sample standard deviation of *D*. Because the *t*0 is larger than the percentage point *t*0.05, 3 = 2.353, difference between the percentage of missing values and the percentage of decrease in accuracy of DREM is significant with odd pairs (3rd, 5th, 7th, 9th). Table 7 shows paired t-tests, given MAE metric and significant level 95%. We use odd pairs (even pairs) in a same group which is compared with the 1st pair (the 2nd pair) because odd pairs (even pairs) share the same testing dataset *sample1.test* (*sample2.test*).

**Table 7.** Paired t-tests given MAE metric where *t*0.05, 3 = 2.353

|  |  |  |
| --- | --- | --- |
|  | *t*0 | Difference |
| Odd pairs | 3.9173 | Significant |
| Even pairs | 4.5241 | Significant |

From paired t-tests in Table 7, it is asserted that the withstanding of REM for missing values with regard to MAE metric is significant because the bias ratios with regard to MAE metric are much smaller than percentages of missing values.

We continue to assess such ten regression models with subject to R metric.

|  |  |
| --- | --- |
|  | (9) |

The *R* reflects adequacy of a given formula. The larger the *R* is, the better the formula is. Table 8 shows R metric which evaluates our models.

**Table 8.** R metric of ten models

|  |  |
| --- | --- |
| Pair | R |
| 1 | 0.9595 |
| 2 | 0.9588 |
| 3 | 0.9592 |
| 4 | 0.9588 |
| 5 | 0.9590 |
| 6 | 0.9583 |
| 7 | 0.9591 |
| 8 | 0.9561 |
| 9 | 0.9551 |
| 10 | 0.9018 |
| Average | 0.9526 |

Table 9 shows paired t-tests given R metric.

**Table 9.** Paired t-tests given R metric where *t*0.05, 3 = 2.353

|  |  |  |
| --- | --- | --- |
|  | *t*0 | Difference |
| Odd pairs | 3.8599 | Significant |
| Even pairs | 3.6563 | Significant |

From paired t-tests in Table 9, it is asserted that the withstanding of REM for missing values with regard to R metric is significant because the bias ratios with regard to R metric are much smaller than percentages of missing values.

1. **Conclusions**

In general, from experimental results on two typical evaluation metrics such as MAE and R, we conclude that REM solves totally the problem in which fetal weight, fetal ages, and ultrasound measures can be missing. This problem was raised in our previous research (Nguyen & Ho, 2018). As a result, practitioners will have a lot of benefits when they will not be stressful in taking ultrasound examinations. In other words, it is acceptable for practitioners to make unintentional mistakes when taking ultrasound examinations. Of course, early weight estimation is achieved because ultrasound examination can be taken at any time of gestational period because it is not mandatory to know fetal weights. When the withstanding of REM for missing values is proved, we will improve REM with prior distribution of coefficients (*α*, *βj*) and compare REM with other algorithms for further research.

**Conflicts of Interest**

The authors Loc Nguyen and Thu-Hang T. Ho declare that there is no conflict of interest regarding the publication of this article. Because this research is the successive one after our previous research “Early Fetal Weight Estimation with Expectation Maximization Algorithm” published in Experimental Medicine (EM) Journal of International Technology and Science Publications (ITS) on 7th May 2018, they share some common contents, but we confirm that their methods are different.

**Acknowledgments**

All acknowledgments (if any) may include supporting grants, presentations, and so forth. Keep acknowledgements brief, naming those who helped with your research; contributors, or suppliers who provided free materials. You should also disclose any financial or other substantive conflict of interest that could be seen to influence your results or interpretations.

**References**

Anderson, B., & Hardin, M. J. (2013, January 1). Modified logistic regression using the EM algorithm for reject inference. *International Journal of Data Analysis Techniques and Strategies, 5*(4), 359-373. doi:10.1504/IJDATS.2013.058582

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. (M. Stone, Ed.) *Journal of the Royal Statistical Society, Series B (Methodological), 39*(1), 1-38.

Ghitany, M. E., Karlis, D., Al-Mutairi, D. K., & Al-Awadhi, F. (2012). An EM Algorithm for Multivariate Mixed Poisson Regression Models and its Application. *Applied Mathematical Sciences, 6*(137), 6843-6856. Retrieved from http://www.m-hikari.com/ams/ams-2012/ams-137-140-2012/ghitanyAMS137-140-2012.pdf

Haitovsky, Y. (1968, January 1). Missing Data in Regression Analysis. *Journal of the Royal Statistical Society: Series B (Methodological), 30*(1), 67-82. Retrieved from https://www.jstor.org/stable/2984459

Ho, T. T., & Phan, D. T. (2011, December). Ước lượng cân nặng của thai từ 37 – 42 tuần bằng siêu âm 2 chiều. (D. Thai, Ed.) *Journal of Practical Medicine, 12*(797), 8-9.

Ho, T.-H. T., & Phan, D. T. (2011, December). Ước lượng tuổi thai qua các số đo thể tích cánh tay bằng siêu âm 3 chiều và các số đo bằng siêu âm 2 chiều. (D. Thai, Ed.) *Journal of Practical Medicine, 12*(798), 12-15.

Kokic, P. (2002). *The EM Algorithm for a Multivariate Regression Model: including its applications to a non-parametric regression model and a multivariate time series model.* Frankfurt: Qantaris GmbH. Retrieved from https://www.cs.york.ac.uk/euredit/\_temp/The%20Euredit%20Software/NAG%20Prototype%20platform/WorkingPaper4.pdf

Lindsten, F., Schön, T. B., Svensson, A., & Wahlström, N. (2017). *Probabilistic modeling – linear regression & Gaussian processes.* Uppsala University, Department of Information Technology. Uppsala: Uppsala University. Retrieved January 24, 2018, from http://www.it.uu.se/edu/course/homepage/sml/literature/probabilistic\_modeling\_compendium.pdf

Montgomery, D. C., & Runger, G. C. (2010). *Applied Statistics and Probability for Engineers* (5th ed.). Hoboken, New Jersey, USA: John Wiley & Sons. Retrieved from https://books.google.com.vn/books?id=\_f4KrEcNAfEC

Nguyen, L., & Ho, T.-H. T. (2018, May 7). Early Fetal Weight Estimation with Expectation Maximization Algorithm. (T. Schmutte, Ed.) *Experimental Medicine (EM), 1*(1), 12-30. doi:10.31058/j.em.2018.11002

|  |  |
| --- | --- |
| F:\期刊所需资料\文章template\已准备的资料\完成文档\图片\ccby4.0.jpg | © 2017 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/) |

1. **Introduction (Heading 1)**

The introduction of your article is organized as a funnel that begins with a definition of why the experiment is being performed and ends with a specific statement of your research approach. And it highlights controversial and diverging hypotheses when necessary.

1. **Materials and Methods (Heading 2)**

This section should contain sufficient details so that methods can be appropriately cited and readers can assess whether the materials and methods justify the conclusions or not. It can be divided into subsections if several other methods need to be described. You need explain how you studied the topic, identify the procedures you followed, and structure this information as logically as possible.

2.1 Abbreviations and Acronyms (Sub-Heading 2.1)

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

2.2 Links and Bookmarks (Sub-Heading 2.2)

All hypertext links and section bookmarks will be removed from articles during the processing of articles for publication. If you need to refer to an Internet email address or URL in your articles, you must type out the address or URL fully in Regular font.

1. **Results and Discussion (Heading 3)**

This section should present your findings objectively, explaining them largely, concisely, precisely in the text. You should show how your results contribute to the scientific knowledge in academic community clearly and logically. Results and Discussion may be divided by subheadings, concluding equations, figures, tables, etc.

3.1 Equations (Sub-Heading 3.1)

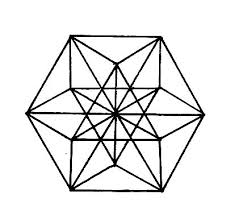
Equations and some mathematical expressions need to be provided in the main text of the article. Equations which are referred to in the text are identified by parenthetical numbers, such as (1), and are referred to in the manuscript as "equation (1)".The equations have to be numbered sequentially and to be centered, the number put in parentheses at the right-hand edge of the text, such as:

c2 = a2 + b2（1）

3.2 Figures and Tables (Sub-Heading 3.2)

Each figure and table must have a caption to describe what is shown in each panel, in sequence and the symbols used. The caption should identify the figure or the table in bold (i.e., Figure 1 or Table 1)

Captions with figure numbers must be placed after their associated figures, as shown in Figure 1.



**Figure 1.** This is a figure.

In a table caption, it includes statistical analysis of data to describe their standards of error analysis and ranges.

**Table 1.** All tables are to be numbered using Arabic numerals.

|  |  |  |
| --- | --- | --- |
| **Title 1** | **Title 2** | **Title 3** |
| data | data | data |
| data | data | data 1 |

1. **Conclusions(Heading 4)**

It should clearly explain the main conclusions of the work highlighting its importance and relevance. This is where you describe the meaning of your results, especially in the context of what was already known about the subject. You can present general and specific conclusions, but take care not to summarize your article – that’s what the abstract is for.

**Conflicts of Interest**

Authors must declare all potential interests, whether or not they actually had an influence in a ‘Conflicts of Interest’ section, which should explain why the interest may be a conflict. If there are no Conflicts of Interest, the authors should state “The author(s) declare(s) that there is no conflict of interest regarding the publication of this article.”

**Acknowledgments**

All acknowledgments (if any) may include supporting grants, presentations, and so forth. Keep acknowledgements brief, naming those who helped with your research; contributors, or suppliers who provided free materials. You should also disclose any financial or other substantive conflict of interest that could be seen to influence your results or interpretations.

**References**

References should be numbered sequentially and citations of references in text should be identified using numbers in square brackets (e.g., “as pointed by Black[1]”; “as discussed where [9, 10]”).

Journal Articles:

[1] Author 1, A.B.; Author 2, C.D. Title of the article. Abbreviated Journal Name Year, Volume, page range, DOI. Available online: URL (accessed on Day Month Year).

Books and Book Chapters:

[2] Author 1, A.; Author 2, B. Book Title, 3rd ed.; Publisher: Publisher Location, Country, Year; pp. 154–196; ISBN.

[3] Author 1, A.; Author 2, B. Title of the chapter. In Book Title, 2nd ed.; Editor 1, A.; Editor 2, B., Eds.; Publisher: Publisher Location, Country, Year; Volume 3, pp. 154–196; ISBN.

Unpublished work, submitted work, personal communication:

[4] Author 1, A.B.; Author 2, C. Title of Unpublished Work. status (unpublished; manuscript in preparation).

[5] Author 1, A.B.; Author 2, C. Title of Unpublished Work. Abbreviated Journal Name stage of publication (under review; accepted; in press).

[6] Author 1, A.B. (University, City, State, Country); Author 2, C. (Institute, City, State, Country).Personal communication, Year.

Conference Proceedings:

[7] Author 1, A.B.; Author 2, C.D.; Author 3, E.F. Title of Presentation. In Title of the Collected Work (if available), Proceedings of the Name of the Conference, Location of Conference, Country, Date of Conference; Editor 1, Editor 2, Eds. (if available); Publisher: City, Country, Year (if available); Abstract Number (optional), Pagination (optional).

Thesis:

[8] Author 1, A.B. Title of Thesis. Level of Thesis, Degree-Granting University, Location of University, Date of Completion.

Websites:

[9] Title of Site. Available online: URL (accessed on Day Month Year).

[10] Title of Site. URL (archived on Day Month Year).

|  |  |
| --- | --- |
| F:\期刊所需资料\文章template\已准备的资料\完成文档\图片\ccby4.0.jpg | © 2017 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/) |