STAT878 Modern Computational Statistical Methods

Semester 1, 2005

Time: Lectures & Pracs: Tuesday 6 – 9 pm E5A 118

Lecturers: Dr Jun Ma (weeks 1 – 7)
Assoc. Prof. Andrzej Kozek (weeks 8 – 13)

Unit Home Page: http://www.stat.mq.edu.au/units/stat878/index.htm

Brief Description:

This unit considers computational statistical methods for modern data analysis. It starts with introduction to the maximum likelihood (ML) and Bayesian estimation methods and their associated computational procedure. Then it develops to the concept of missing data, a common problem in data analysis. The Expectation-Maximization (EM) algorithm is discussed as a method for single data imputation followed by a complete-case analysis. The EM algorithm is then extended as a general optimization algorithm for statistical inference based on incomplete data. Multiple imputation, an alternative to single data imputation, is studied with examples. The unit completes with topics on nonparametric density estimation and nonparametric regression.

Topics:

1. **Maximum likelihood estimation**
   - Likelihood and maximum likelihood estimates
   - Some iterative methods for computing maximum likelihood estimates
2. **Bayesian estimation**
   - Prior and posterior distributions
   - Bayesian estimates
   - Relationship with maximum penalized likelihood
   - Ridge regression
3. **Normal based inference**
   - Asymptotic distribution of MLE
   - The δ - method
4. **Missing observations and EM algorithm**
   - Missing data mechanism
   - Complete data and incomplete data
   - Inference based on incomplete data
   - The EM algorithm
   - Example of the EM algorithm
5. **Handling missing data**
   - Complete case analysis
   - Single imputation

- Multiple imputation

6. **Multiple imputation details**

   - Regression method
   - Propensity score method
   - Markov Chain Monte Carlo (MCMC) method
   - Combining inferences from imputed data sets

7. **Nonparametric density function**

   - Histogram
   - Kernel Density Estimators
   - Bandwidth selection

8. **Nonparametric regression**
   - Nonparametric Regression
   - Bandwidth selection
   - Local parametric estimators.

9. **NoLoEss - local regression quantiles**

Reference books

1. *Smoothing Techniques*, Hardle, W. Springer-Verlag. QA278.H348 1991.
2. *Local regression and likelihood*, Loader, C. Springer-Verlag, 1999. QA276.8 .L6/1999.
3. *Nonlinear time series*, Fan, J. and Yao, Q. Springer-Verlag, 2003.
4. *Statistical analysis with missing data*, Little, R. and Rubin, D. Wiley, 2002.
5. *Analysis of incomplete multivariate data*, Schafer, J. L. Chapman & Hall, 1997.

Assessment:

- Two assignments 20% each
- Two mini take home exams 30% each

Note: *A student must have satisfactory performance in both the final examination and the coursework to pass the unit. The final grade depends on overall interpretation of the data rather than strict numerical accumulation.*

STAT878 Modern Computational Statistical Methods

Semester 1, 2005

Time: Lectures & Pracs: Tuesday 6 – 9 pm E5A 118

Lecturers: Dr Jun Ma (weeks 1 – 7)
Assoc. Prof. Andrzej Kozek (weeks 8 – 13)

Unit Home Page: http://www.stat.mq.edu.au/units/stat878/index.htm

Brief Description:

This unit considers computational statistical methods for modern data analysis. It starts with introduction to the maximum likelihood (ML) and Bayesian estimation methods and their associated computational procedure. Then it develops to the concept of missing data, a common problem in data analysis. The Expectation-Maximization (EM) algorithm is discussed as a method for single data imputation followed by a complete-case analysis. The EM algorithm is then extended as a general optimization algorithm for statistical inference based on incomplete data. Multiple imputation, an alternative to single data imputation, is studied with examples. The unit completes with topics on nonparametric density estimation and nonparametric regression.

Topics:

1. **Maximum likelihood estimation**
   - Likelihood and maximum likelihood estimates
   - Some iterative methods for computing maximum likelihood estimates
2. **Bayesian estimation**
   - Prior and posterior distributions
   - Bayesian estimates
   - Relationship with maximum penalized likelihood
   - Ridge regression
3. **Normal based inference**
   - Asymptotic distribution of MLE
   - The δ - method
4. **Missing observations and EM algorithm**
   - Missing data mechanism
   - Complete data and incomplete data
   - Inference based on incomplete data
   - The EM algorithm
   - Example of the EM algorithm
5. **Handling missing data**
   - Complete case analysis
   - Single imputation

- Multiple imputation

6. **Multiple imputation details**

   - Regression method
   - Propensity score method
   - Markov Chain Monte Carlo (MCMC) method
   - Combining inferences from imputed data sets

7. **Nonparametric density function**

   - Histogram
   - Kernel Density Estimators
   - Bandwidth selection

8. **Nonparametric regression**
   - Nonparametric Regression
   - Bandwidth selection
   - Local parametric estimators.

9. **NoLoEss - local regression quantiles**

Reference books

1. *Smoothing Techniques*, Hardle, W. Springer-Verlag. QA278.H348 1991.
2. *Local regression and likelihood*, Loader, C. Springer-Verlag, 1999. QA276.8 .L6/1999.
3. *Nonlinear time series*, Fan, J. and Yao, Q. Springer-Verlag, 2003.
4. *Statistical analysis with missing data*, Little, R. and Rubin, D. Wiley, 2002.
5. *Analysis of incomplete multivariate data*, Schafer, J. L. Chapman & Hall, 1997.

Assessment:

- Two assignments 20% each
- Two mini take home exams 30% each

Note: *A student must have satisfactory performance in both the final examination and the coursework to pass the unit. The final grade depends on overall interpretation of the data rather than strict numerical accumulation.*

STAT878 Modern Computational Statistical Methods

Semester 1, 2005

Time: Lectures & Pracs: Tuesday 6 – 9 pm E5A 118

Lecturers: Dr Jun Ma (weeks 1 – 7)
Assoc. Prof. Andrzej Kozek (weeks 8 – 13)

Unit Home Page: http://www.stat.mq.edu.au/units/stat878/index.htm

Brief Description:

This unit considers computational statistical methods for modern data analysis. It starts with introduction to the maximum likelihood (ML) and Bayesian estimation methods and their associated computational procedure. Then it develops to the concept of missing data, a common problem in data analysis. The Expectation-Maximization (EM) algorithm is discussed as a method for single data imputation followed by a complete-case analysis. The EM algorithm is then extended as a general optimization algorithm for statistical inference based on incomplete data. Multiple imputation, an alternative to single data imputation, is studied with examples. The unit completes with topics on nonparametric density estimation and nonparametric regression.

Topics:

1. **Maximum likelihood estimation**
   - Likelihood and maximum likelihood estimates
   - Some iterative methods for computing maximum likelihood estimates
2. **Bayesian estimation**
   - Prior and posterior distributions
   - Bayesian estimates
   - Relationship with maximum penalized likelihood
   - Ridge regression
3. **Normal based inference**
   - Asymptotic distribution of MLE
   - The δ - method
4. **Missing observations and EM algorithm**
   - Missing data mechanism
   - Complete data and incomplete data
   - Inference based on incomplete data
   - The EM algorithm
   - Example of the EM algorithm
5. **Handling missing data**
   - Complete case analysis
   - Single imputation

- Multiple imputation

6. Multiple imputation details

- Regression method
- Propensity score method
- Markov Chain Monte Carlo (MCMC) method
- Combining inferences from imputed data sets

7. Nonparametric density function

- Histogram
- Kernel Density Estimators
- Bandwidth selection

8. Nonparametric regression

- Nonparametric Regression
- Bandwidth selection
- Local parametric estimators.

9. NoLoEss - local regression quantiles

Reference books

1. *Smoothing Techniques*, Hardle, W. Springer-Verlag. QA278.H348 1991.
2. *Local regression and likelihood*, Loader, C. Springer-Verlag, 1999. QA276.8 .L6/1999.
3. *Nonlinear time series*, Fan, J. and Yao, Q. Springer-Verlag, 2003.
4. *Statistical analysis with missing data*, Little, R. and Rubin, D. Wiley, 2002.
5. *Analysis of incomplete multivariate data*, Schafer, J. L. Chapman & Hall, 1997.

Assessment:

- Two assignments 20% each
- Two mini take home exams 30% each

Note: *A student must have satisfactory performance in both the final examination and the coursework to pass the unit. The final grade depends on overall interpretation of the data rather than strict numerical accumulation.*