

Politecnico di Torino - Dipartimento di Scienze Matematiche
Didattica di eccellenza nel dottorato di Matematica per le Scienze dell'Ingegneria

Categorical Data Analysis

Professor **Kaisheng Song**
University of North Texas
ksong@unt.edu

Lectures will take place in aula Buzano of DISMA, third floor of Politecnico di Torino in Corso Duca degli Abruzzi 24, Torino, November 11,12,13,18,19 and 20, 2013.

For further information and emergencies call Prof. Mauro Gasparini, 011 0907546,
gasparini@calvino.polito.it

Monday 11/11/13, 09:00-11:00

1. Introduction to Categorical Data

We start with a brief review of some well-known and less-known discrete probability distributions for categorical data, including the binomial, multinomial and Poisson distributions. We then introduce the concept of likelihood function and discuss the general method of maximum likelihood for parameter estimation. By exploiting the large sample properties of ML estimators, we develop the Wald, likelihood ratio, and score tests for drawing statistical inference for the parameters of these distributions.

Monday 11/11/13, 15:00-17:00

2. Contingency Tables and Their Inference

We describe contingency tables for categorical variables, discuss their probability structure, and compare proportions. Odds ratio and parameters of association are defined. We also introduce the delta method and use it to derive the standard errors of various estimators. Pearson and likelihood ratio chi-squared tests are discussed. Small-sample tests of independence (Fisher's exact test) are developed.

Tuesday 12/11/13, 09:00-11:00

3. Generalized Linear Models

We describe the various components of a generalized linear model (GLM). We develop the maximum likelihood and maximum quasi-likelihood methods of estimation. We discuss the inference and model checking for fitting generalized linear models.

Tuesday 12/11/13, 15:00-17:00

4. Logistic and Probit Regression

We consider fitting GLMs for binary data, in particular, fitting logistic and probit regression models. How to interpret logistic regression is presented. Logit models with categorical predictor and multiple logistic regression are also described.

Wednesday 13/11/13, 09:00-11:00

5. Building and Applying Logistic Regression Models

We discuss strategies in model selection. We conduct statistical inference and model checking for logistic regression, Effects of sparse data are presented.

Wednesday 13/11/13, 15:00-17:00

6. Logit Models for Multinomial Responses

We describe multcategory logit models. We consider logit models for nominal responses and cumulative logit model for ordinal responses. Discrete-choice multinomial logit models are also discussed.

Monday 18/11/13, 09:00-11:00

7. Loglinear Models for Contingency Tables

We discuss loglinear models for two-way tables, independence and interaction in three-way tables. Loglinear model fitting and inference are described. Loglinear models for higher dimensions and the loglinear-logit model connection are also discussed.

Monday 18/11/13, 15:00-17:00

8. Models for Matched Pairs

We consider conditional logistic regression for binary matched pairs. We describe methods for comparing dependent proportions and for measuring agreement between observers.

Tuesday 19/11/13, 09:00-11:00

9. Modeling Clustered Responses (Repeated Measures)

We consider modeling and analyzing repeated categorical response data by comparing marginal distributions. We describe maximum likelihood approach and generalized estimating equations (GEE) approach for marginal modeling. We also compare marginal models with conditional models.

Tuesday 19/11/13, 15:00-17:00

10. Generalized Linear Mixed Models for Categorical Responses

We describe random effects modeling for clustered categorical data. Random effects models for both binary and multinomial data are discussed. Multivariate random effects models for binary data are also introduced.

Wednesday 20/11/13, 15:00-17:00

11. Alternative Estimation Methods for Categorical Data

We describe and discuss some alternative methods for analyzing categorical data. These methods include smoothing, regularization, and Bayesian methods.

Wednesday 20/11/13, 15:00-17:00

12. Spectral Analysis for Categorical Time Series

We describe the frequency domain analysis of categorical time series and discuss the scaling of categorical time series. We introduce and discuss the spectral envelope for categorical time series.