PROGRAMMA DEL CICLO DI LEZIONI:
MODELLI PER TABELLE DI CONTINGENZA

Roberto Colombi
colombi@unibg.it

- **Modelli di regressione per variabili casuali dicotome e modelli log-lineari per tabelle a due e tre entrate**

  Partendo da concetti di base mutuati dall'ambito del modello di regressione lineare classico si introducono ad un livello elementare concetti e metodi che sono rilevanti nell'ambito della modellistica per mutabili. Vengono discussi due modelli di regressione per variabili dicotome la cui importanza nell'ambito, oggetto del corso, è paragonabile a quella che riveste il modello di regressione lineare nella modellistica per variabili continue. In un contesto elementare vengono inoltre introdotti i modelli log-lineari. Questa parte vuole essere una introduzione "soft" agli argomenti successivi.

  - approccio di Pearson e approccio di Yule nella specificazione di modelli di regressione per variabili dicotome
  - modelli logit
  - modelli probit
  - estensione dei modelli logit e probit al caso di variabili politome
  - inferenza sui parametri dei modelli Logit e Probit basata sul metodo di massima verosimiglianza
  - introduzione ai modelli log-lineari

- **La parametrizzazione log-lineare per tabelle di contingenza**

  L' analisi classica delle tabelle di contingenza è basata su una classe di modelli chiamati log-lineari. Questa parte si propone di descrivere in modo rigoroso e generale la costruzione di tali modelli partendo da una conoscenza, data per acquisita, degli stessi nel caso semplice di tabelle a due e tre entrate. Viene inoltre prestata molta attenzione al problema della interpretazione dei parametri del modello log-lineare e alla espressione in termini matriciali dei modelli; quest'ultima parte è importante sia per la comprensione degli algoritmi di calcolo usati nei problemi di stima MV sia per accedere alla letteratura inerente recenti generalizzazioni di questi modelli.

  - terminologia e concetti di base
  - spazi fattoriali e spazi di interazione
  - operatore di sostituzione e operatori di proiezione su spazi fattoriali
  - scomposizione di un vettore di logaritmi di probabilità congiunte in termini di proiezioni su spazi di interazione

  - teorema di inversione di Mobius e modello log-lineare saturo
  - modello log-lineare gerarchico associato ad uno spazio fattoriale
  - parametri di interazione e loro interpretazione
  - rappresentazione matriciale di un modello log-lineare

- **Grafi di indipendenza condizionale e interpretazione dei modelli log-lineari**

  L'interpretazione dei modelli lg-lineari è stata enormemente semplificata da recenti risultati che hanno mostrato come a certi modelli log-lineari sia associato un grafo e come alle proprietà algebriche di tale grafo corrispondano delle proprietà probabilistiche della funzione di probabilità congiunta. I risultati sono estremamente rilevanti sia a fini operativi che teorici e costituiscono uno dei più importanti risultati degli ultimi anni.

  - richiami di teoria dei grafi
  - funzioni di probabilità markoviane rispetto ad un grafo: markovianità a coppie, markovianità locale e markovianità globale.
  - teorema di equivalenza tra markovianità a coppie, markovianità locale e markovianità globale
  - teorema di fattorizzazione
  - grafo associato a un modello log-lineare
  - modelli log-lineari grafici
  - teorema di scomposizione

- **Inferenza sui parametri dei modelli log-lineari basata sul metodo di massima verosimiglianza**

  La teoria degli stimatori MV viene particolarizzata all'ambito dei modelli log-lineari dando rilevanza sia alla peculiarità del problema sia agli aspetti computazionali connessi al calcolo delle stime MV

  - verosimiglianza nell'ipotesi di campionamento multinomiale
  - verosimiglianza nell'ipotesi di frequenze Poissoniane
  - scores functions e matrice di Fisher
  - proprietà asintotiche degli stimatori MV dei parametri di un modello log-lineare
  - verifica di ipotesi su parametri di modelli log-lineari.
  - cenno sugli algoritmi di ottimizzazione utilizzati nel calcolo delle stime MV.

- **Generalizzazioni dei modelli log-lineari**

  La classe dei modelli log-lineari pur essendo ampia ed estremamente flessibile gode di alcuni limiti il più imprtante dei quali è l'impossibilità di modellare in modo semplice funzioni di probabilità marginali. Vengono quì discusse le più promettenti alternative avanzate negli ultimi anni.

- i rapporti incrociati generalizzati
- i modelli Logit Multivariati
- i modelli log-lineari generalizzati

- **Modelli per serie storiche di variabili categoriche: modelli retti da catene di Markov non osservabili**

  Il problema dell'analisi di serie storiche di variabili categoriche è ben lungi dall'avere la sistematizzazione raggiunta nel caso di variabili continue. Qui viene descritto un approccio basato sull'idea degli Hidden Markov Models.

  - hidden markov models
  - hidden markov models per mutabili
  - hidden markov models per tabelle di contingenza
  - associazione condizionata a una realizzazione della catena di Markov e associazione marginale
  - algoritmi di filtering, forecasting, smoothing e cxalcolo della verosimiglianza
  - massimizzazione della funzione di verosimiglianza e algoritmo EM
  - proprietà asintotiche degli stimatori MV dei parametri di modelli retti da catene di Markov non osservabili

Prerequisiti

Viene dato per acquisito il contenuto dei capitoli 3 e 5 del libro di A. Agresti (1990): Categorical data analysis, Wiley, New York.

Lo studio individuale del capitolo quattro dello stesso libro è pure raccomandato

References

[1] M.B. Birch (1963): A new prof of the Pearson Fisher Theorem, Annals of mathematical statistics, 35, pp. 818-824.

[2] C. Cox (1984): An Elementary Introduction to Maximum likelihood estimation for multinomial models: Birch's theorem and the delta method, The American Statistician ,58, pp. 283-287.

[3] Cowell-Dawid-Lauritzen-Spiegelhalter (1999): Probabilistic Networks and Expert Systems, Springer, New York., cap. 3,4,5.

[4] Darroch-Speed (1983): Additive and Multiplicative models and interactions, The Annals of Statistics, pp. 724-738.

[5] Fahrmeir-Kaufmann (1985): Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models, Annals of Statistics, vol. 13, pp. 342-368.

[6] Kaufmann (1987): regression models for nonstationary categorical time series: asymptotic estimation theory.

[7] Krolzig (1996): Markov-switching vector autoregression, Springer, New York, cap. 1,2,5,6.

[8] Lang-McDonald-Smith (1999). Association-Marginal Modeling of Multivariate Categorical Responses: A Maximum Likelihood Approach, J.A.S.A., 94, pp. 1161-1171.

[9] MacDonald-Zucchini (1998): Hidden Markov and Other Models for Discrete Valued Time Series, Chapman-Hall, cap.3.

[10] Stefen L. Lauritzen (1996): Graphical Models, Oxford Science Publications, cap. 1,2,3,4.