

دانشگاه تربیت مدرس

دانشکده علوم ریاضی

رساله دوره دکتری آمار

تحلیل بیزی داده‌های طولی

توسط

سجاد نوریان

استاد راهنما

دکتر

استادان مشاور

دکتر

دکتر

شهریور ۱۳۹۲

تقدیم بہ پدر بزرگوارم کہ

ہمچون اشک

از چشمم رفت

و باز نمی آید

رسیدن، به دانش است و به کردار نیک...

و بی دانش به کردار نیک هم نتوان رسید، که نیکی را بیشتر باید شناختن، آنگاه بجای آوردن.

پس منطق ناکزیر آمد بر جوینده ی رسگاری.^۱

^۱مقدمه ی رساله ی منطق دانشنامه ی علائی، شیخ الرئیس ابن سینا

سپاسگزاری...

ستایش و سپاس اولاً و بالذات مخصوص خداوندی است که منطق را فطرتاً در وجود آدمی نهاد.

سجاد نوریان
شهریور ۱۳۹۲

چکیده

در این رساله، یک چارچوب بیزی برای تحلیل داده‌های پاسخ ترتیبی طولی ارائه شده است. در تحلیل داده‌های طولی، باید همبستگی‌های بین پاسخ‌های مرتبط با هر آزمودنی در نظر گرفته شود. برای مدل‌بندی این همبستگی‌ها مدل‌های مختلفی را می‌توان به کار برد که عبارتند از مدل‌بندی حاشیه‌ای، مدل‌بندی اثرهای تصادفی و مدل‌بندی انتقالی (مارکوف). به منظور به دست آوردن پارامترها، از یک مدل رگرسیون لوزستیک تجمعی و رهیافت بیزی، استفاده شده است. برآورد بیزی پارامتر گاما در هر سطح از متغیر تبیینی به دست آمده است. سپس، حساسیت خلاصه‌های پسین نسبت به تغییر ابرپارامترهای توزیع پیشین مورد ارزیابی قرار گرفته است.

واژه‌های کلیدی: تحلیل بیزی، عامل بیزی، عرض پیشگوی شرطی.

فهرست مندرجات

ب	لیست جداول
ج	لیست اشکال
۱	۱ مفهومی‌های اولیه
۱	۱.۱ معرفی داده‌های طولی و مقایسه‌ی آن‌ها با داده‌های مقطعی
۳	۲ مدل‌هایی برای تحلیل داده‌های طولی با تأکید بر مدل‌های انتقالی
۳	۱.۲ مدل‌های مختلف تحلیل داده‌های طولی
۳	۱.۱.۲ مدل‌های حاشیه‌ای
۴	۳ تحلیل بیزی مدل‌های انتقالی
۴	۱.۳ تحلیل بیزی مدل‌های انتقالی بدون مقادیر گم‌شده
۴	۱.۱.۳ مدل‌های انتقالی و تابع درست‌نمایی
۵	۴ همگنی پارامتر پیوند در داده‌های با پاسخ‌های آمیخته‌ی ترتیبی و پیوسته
۵	۱.۴ مقدمه
۶	کتاب‌نامه
۹	آ توپولوژی‌های روی فضاهای اندازه‌ها
۹	۱.آ توپولوژیِ مبهم روی فضای اندازه‌ها

فهرست مندرجات

ب

واژه‌نامه فارسی به انگلیسی

۱۰

واژه‌نامه انگلیسی به فارسی

۱۲

لیست جداول

لیست اشکال

۱.۱.۱ داده‌های فرضی در مورد رابطه‌ی بین سن (محور افقی) و توانایی خواندن (محور عمودی) (دیگ‌ل و همکاران،

۲۰۰۲، ص ۲) ۲

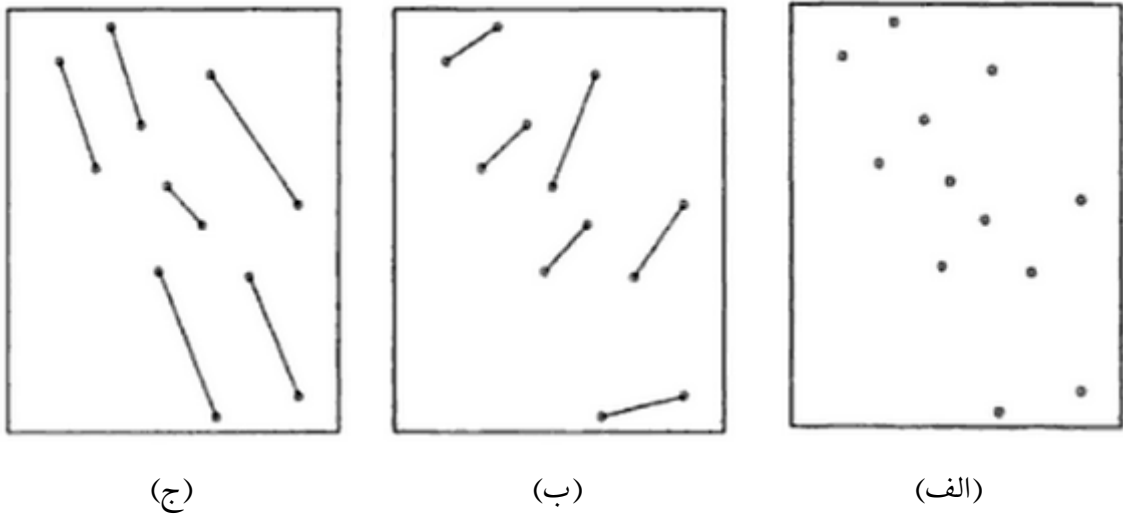
فصل ۱

مفهوم‌های اولیه

هدف از این فصل معرفی مفهوم‌های اولیه داده‌های طولی و داده‌های گم‌شده است.

۱.۱ معرفی داده‌های طولی و مقایسه‌ی آن‌ها با داده‌های مقطعی

هدف اصلی مطالعه‌های طولی، مشخص‌سازی تغییرات متغیر پاسخ در طول زمان و بررسی عامل‌های تأثیرگذار بر این تغییرات است.



شکل ۱.۱.۱: داده‌های فرضی در مورد رابطه‌ی بین سن (محور افقی) و توانایی خواندن (محور عمودی) (دیگل و همکاران، ۲۰۰۲، ص ۲)

فصل ۲

مدلهایی برای تحلیل داده‌های طولی با تأکید بر مدل‌های انتقالی

در این فصل مدل‌های انتقالی برای تحلیل داده‌های طولی مورد بررسی قرار می‌گیرند.

۱.۲ مدل‌های مختلف تحلیل داده‌های طولی

در این بخش به بررسی رهیافت‌های ممکن برای تحلیل داده‌های طولی گسسته و پیوسته می‌پردازیم.

۱.۱.۲ مدل‌های حاشیه‌ای

در یک مدل حاشیه‌ای، رابطه‌ی بین متغیر پاسخ و متغیرهای تبیینی به‌طور جداگانه (واحد) مدل‌بندی می‌شود.

فصل ۳

تحلیل بیزی مدل‌های انتقالی

در این فصل به تحلیل بیزی مدل‌های انتقالی برای داده‌های طولی می‌پردازیم. سرانجام کاربرد مدل‌های پیشنهادی در تحلیل دو مجموعه داده‌ی بی‌خواهی و فلووکسامین مورد بحث قرار خواهد گرفت.

۱.۳ تحلیل بیزی مدل‌های انتقالی بدون مقادیر گم‌شده

در این بخش تحلیل بیزی مدل‌های انتقالی مرتبه‌ی اول برای پاسخ‌های طولی ترتیبی بدون داده‌ی گم‌شده ارائه خواهد شد.

۱.۱.۳ مدل‌های انتقالی و تابع درستنمایی

فصل ۴

همگنی پارامتر پیوند در داده‌های با پاسخ‌های آمیخته‌ی ترتیبی و پیوسته

۱.۴ مقدمه

در برخی موارد، داده‌های اندازه‌گیری شده و متغیرهای موجود، صرفاً دارای یک مقیاس نیستند. به عبارتی دیگر، در یک بررسی ممکن است که برخی از داده‌ها دارای مقیاس فاصله‌ای، و برخی دیگر دارای مقیاس ترتیبی یا اسمی باشند.

کتاب نامه

- [1] Aerts, M., Claeskens, G., Hens, N. and Molenberghs, G., (2002). Local Multiple Imputation, *Biometrika*, **89**, 375-388.
- [2] Agresti, A., (1984). *Analysis of Ordinal Categorical Data*, John Wiley, New York.
- [3] Agresti, A., (2002). *Analysis of Categorical Data*, John Wiley and Sons, New York.
- [4] Agresti A., (2010). *Analysis of Ordinal Categorical Data*, (2nd edition), John Wiley and Sons, New York.
- [5] Agresti, A. and Hitchcock, D. B., (2005). Bayesian Inference for Categorical Data Analysis, *Statistical Methods and Applications*, **14**(3), 297-330.
- [6] Anderson, T. W. and Goodman, L. A., (1957). Statistical Inference about Markov Chains, *The Annals of Mathematical Statistics*, **28**(1), 89-110.
- [7] Bahrami Samani, E. and Ganjali, M., (2008). A Multivariate Latent Variable Model for Mixed Continuous and Ordinal Responses, *World Applied Sciences Journal*, **3**(2), 294-299.
- [8] Bahrami Samani, E., Ganjali, M. and Eftekhari Mahabadi, S., (2010). A Latent Variable Model for Mixed Continuous and Ordinal Responses with Nonignorable Missing Responses, *Sankhya*, **72**(B), 38-57.
- [9] Bartholomew, D. J., (1987). *Latent Variable Models and Factor Analysis*, Oxford University Press, New York.
- [10] Breitner, S., Peters, A., Kuuchenhoff, H., Ibal-Mulli, A. and Wichmann, H. E., (2003). Association between Air Pollution and Health. Statistical Analysis of a Longitudinal Study with a Binary Outcome, *18th International Workshop on Statistical Modelling*, July 7-11, Leuven, Belgium.

- [11] Catalano, P. and Ryan, L. M. (1992), Bivariate Latent Variable Models for Clustered Discrete and Continuous Outcomes, *Journal of the American Statistical Association*, **87**, 651-658.
- [12] Chan, J. S. K. and Chau. V. K. K., (2003). Informative Drop-out Model for Longitudinal Binary Data using Bayesian Approach, *18th International Workshop on Statistical Modelling*, July 7-11, Leuven, Belgium.
- [13] Schafer, J. L., (1997). *Analysis of Incomplete Multivariate Data*, Chapman and Hall, New York.
- [14] Shih, W. J., (1992). On Informative and Random Dropouts in Longitudinal Studies, *Biometrics*, **48**, 970-972.
- [15] Sommer, A., (1982). *Nutritional Blindness*, Oxford University Press, New York.
- [16] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. V. D., (2002). Bayesian Measures of Model Complexity and Fit (with discussion), *Journal of the Royal Statistical Society, Series B (Methodological)*, **64**(4), 583-639.
- [17] Stefanski, L. A. and Carroll, R. J., (1985). Covariate Measurement Error in Logistic Regression, *The Annals of Statistics*, **13**, 1335-1351.
- [18] Sung, M., Soyer, R. and Nhan, N., (2007). Bayesian Analysis of Non-homogeneous Markov Chains: Application to Mental Health Data, *Statistics in Medicine*, **26**, 3000-3017.
- [19] Tsay, R., (1984). Regression Models With Time Series Errors, *Journal of the American Statistical Association*, **79**, 118-124.
- [20] Tutz, G., (2005). Modelling of Repeated Ordered Measurements by Isotonic Sequential Regression, *Statistical Modelling*, **5**(4), 269-287.
- [21] Zeger, S. L. and Liang, K. Y., (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes, *Biometrics*, **42**(1), 121-130.
- [22] Zeger, S. L. and Liang, K. Y., (1992). An Overview of Methods for the Analysis of Longitudinal Data, *Statistics in Medicine*, **11**, 1825-1839.
- [23] Zeger, S. L., Liang, K. Y. and Albert, P. S., (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach, *Biometrics*, **44**(4), 1049-1060.
- [24] Zeger, S. L., Liang, K. Y. and Heagerty, P., (1993). Regression Models for Discrete Longitudinal Responses, *Statistical Science*, **8**(3), 304-306.

- [25] Wu, M. C. and Carroll, R. J., (1988). Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process, *Biometrics*, **44**, 175-188.
- [26] Yang, Y., Kang, J., Mao, K. and Zhang, J., (2007). Regression Models for Mixed Poisson and Continuous Longitudinal Data, *Statistics in Medicine*, **26**, 3782-3800.
- [۲۷] رضایی قهرودی، ز.، (۱۳۸۷). کاربردهای مدل‌های انتقالی برای تحلیل داده‌های طولی با پاسخ رسته‌ای با و بدون مقادیر گم‌شده، رساله‌ی دکتری آمار، دانشگاه شهید بهشتی.

پیوست آ

توپولوژی‌های روی فضاهای اندازه‌ها

۱.آ توپولوژی مبهم روی فضای اندازه‌ها

چندین توپولوژی وجود دارد که می‌توان آنها را برای مجموعه اندازه‌ها انتخاب کرد. یک شرط قابل قبول و حداقلی این است که اگر تور $(m_i)_{i \in I}$ به m همگرا باشد آنگاه باید در \mathbb{R} داشته باشیم $\int dm_i \rightarrow \int f dm$.

واژه‌نامه فارسی به انگلیسی

Cross-Over Trials.....	آزمایه‌های متقاطع.....
Peabody Individual Achievement Test (PIAT).....	آزمون انفرادی پیشرفت بی‌بادی.....
Mastitis Data	آماس غده‌ها
Nelder and Mead Simplex Algorithm	الگوریتم سادگی نلدر و مید
Insomnia	بی‌خوابی
Parsimonious Parametrization	پارامتریدن ممسك
Link Function.....	تابع ربط
Exchangeable.....	تبادل‌پذیر.....
Xerophthalmia	رمد چشم
Absorbing Markov chain	زنجیر مارکوف جاذب
Non-homogeneous Markov Chain.....	زنجیر مارکوف ناهمگن.....
Identifiability.....	شناسایی‌پذیری.....
Conditional Predictive Ordinate (CPO).....	عرض پیشگوی شرطی.....
Irreversible Process.....	فرایند برگشت‌ناپذیر.....
Serially Correlated Process	فرایند همبسته‌ی پیاپی
Quadrature Formula	فرمول تربیعی
Fluvoxamine.....	فلووکسامین.....
Discriminant Analysis Applications.....	کاربردهای تحلیل تشخیصی.....
Latent Variable	متغیر پنهان
Bayesian Generalized Additive Mixed Model	مدل آمیخته جمع‌ی تعمیم‌یافته بیزی
Random-Coefficient Pattern-Mixture Models.....	مدل‌های الگو آمیخته‌ی ضرایب تصادفی.....

Random-Coefficient Selection Models	مدل‌های گزینش ضرایب تصادفی
Likelihood Based Mixture Parameter Model	مدل پارامتر آمیخته مبتنی بر درست‌نمایی
Subject-Specific Model	مدل خاص آزمودنی
General Location Model	مدل مکانی عام
Population-Averaged Model	مدل میانگین-جامعه
Generalized Heckman Model	مدل هکمن تعمیم‌یافته
Indonesian Children's Health Study	مطالعه سلامت کودکان اندونزیایی
Score Equations	معادله‌های امتیاز
Generalized Estimating Equations	معادله‌های برآوردساز تعمیم‌یافته
Cut-Point	نقطه‌ی برش
Change Point	نقطه‌ی تغییر
Isotonic	هم‌نوا
m-Dependence	m-وابستگی

واژه‌نامه انگلیسی به فارسی

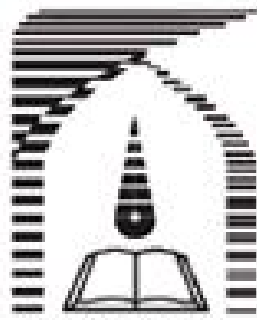
Absorbing Markov chain	زنجیر مارکوف جاذب
Bayesian Generalized Additive Mixed Model	مدل آمیخته جمعی تعمیم‌یافته بیزی
Change Point	نقطه‌ی تغییر
Conditional Predictive Ordinate (CPO)	عرض پیشگوی شرطی
Cross-Over Trials	آزمایه‌های متقاطع
Cut-Point	نقطه‌ی برش
Discriminant Analysis Applications	کاربردهای تحلیل تشخیصی
Exchangeable	تبادل‌پذیر
Fluvoxamine	فلووکسامین
General Location Model	مدل مکانی عام
Generalized Estimating Equations	معادله‌های برآوردساز تعمیم‌یافته
Generalized Heckman Model	مدل هکمن تعمیم‌یافته
Identifiability	شناسایی‌پذیری
Indonesian Children's Health Study	مطالعه سلامت کودکان اندونزیایی
Insomnia	بی‌خوابی
Isotonic	هم‌نوا
Irreversible Process	فرایند برگشت‌ناپذیر
Latent Variable	متغیر پنهان
Likelihood Based Mixture Parameter Model	مدل پارامتر آمیخته مبتنی بر درست‌نمایی
Link Function	تابع ربط
Mastitis Data	آماس غده‌ها

m-Dependence	m-وابستگی
Nelder and Mead Simplex Algorithm	الگوریتم سادکی نلدر و مید
Non-homogeneous Markov Chain.....	زنجیر مارکوف ناهمگن
Parsimonious Parametrization	پارامتریدن ممسک
Peabody Individual Achievement Test (PIAT).....	آزمون انفرادی پیشرفت پی‌بادی
Population-Averaged Model.....	مدل میانگین-جامعه
Quadrature Formula	فرمول تربیعی
Random-Coefficient Pattern-Mixture Models.....	مدل‌های الگو آمیخته‌ی ضرایب تصادفی
Random-Coefficient Selection Models.....	مدل‌های گزینش ضرایب تصادفی
Score Equations.....	معادله‌های امتیاز
Serially Correlated Process	فرایند همبسته‌ی پی‌پی
Subject-Specific Model.....	مدل خاص آزمودنی
Xerophthalmia	رمد چشم

Abstract

In this thesis, a Bayesian framework for analyzing longitudinal ordinal response data is presented. In analyzing longitudinal data, the possibility of correlations between responses given by the same individual needs to be taken into account. Various models can be used to handle such correlations such as marginal modeling, random effect modeling and transition (Markov) modeling. Here a transition modeling is used and a Bayesian approach is presented for analyzing longitudinal data. A cumulative logistic regression model and the Bayesian method, using MCMC, are implemented for obtaining the parameters estimates. the Bayesian estimate for measure of association is obtained in each level of explanatory variables. Then, the sensitivity of posterior summaries to changes of prior hyperparameters is investigated.

Key Words: Bayesian Analysis; Bayes Factor; Conditional Predictive Ordinate.



T. M. U.

Bayesian Analysis of Longitudinal Data

A Thesis Presented for the Degree of
Doctor of Philosophy in Statistics

School of Mathematical Sciences

Tarbiat Modares University

by
Sajad Noorian

Supervisor
Prof.

Advisors
Dr.
Dr.

September 2013