

چکیده

ویژگی‌های آماری مدل‌های پیشنهادی از طریق انجام مطالعه شبیه‌سازی مورد بررسی قرار گرفت. نتایج شبیه‌سازی برتری مدل‌های پیشنهادی نسبت به مدل‌های برازش شده قبلی برای داده‌های طولی ترتیبی را که دارای ویژگی‌های بیش‌پراکنش و آماسیدگی هستند، نشان داد. مدل‌های پیشنهادی با توجه به اریبی کمتر در برآورد پارامترها و پیش‌بینی احتمالات سطوح پاسخ برآوردهای قابل اعتمادتر و دقیق‌تری را ارائه دادند.

کلید واژه‌ها: بیش‌پراکنش، آماسیدگی، داده‌های طولی، داده‌های ترتیبی، رهیافت بیز

فهرست مطالب

۱	اعتبارسنجی متقابل با خروج یک مشاهده
۲	۱.۱ انحرافات تاوانیده
۴	پیوست‌ها
۵	آ کد مربوط به مدل
۵	۱.آ اثر تصادفی آمیختگی گاما
۶	واژه نامه انگلیسی به فارسی
۸	واژه نامه فارسی به انگلیسی

فهرست جداول

فهرست تصاویر

فصل ۱

اعتبارسنجی متقابل با خروج یک مشاهده

اعتبارسنجی متقابل به شرایطی اطلاق می‌گردد که پیش‌بینی زیرمجموعه Y_s از داده‌های مشاهده شده براساس پارامترهای برآورد شده بر مبنای بخش باقی‌مانده از داده‌ها که با $Y_{[s]}$ نمایش داده می‌شود، انجام گیرد. بخشی از داده‌ها که برازش مدل و برآورد پارامترها براساس آن صورت می‌گیرد، مجموعه آموزشی و بخشی که براساس آن‌ها پیش‌بینی صورت می‌گیرد مجموعه اعتبار یا پیش‌بینی نام دارد. پیش‌بینی اعتبارسنجی متقابل ممکن است براساس نمونه‌گیری مکرر از توزیع پسین θ در زنجیره مارکف مونت‌کارلو به صورت زیر معرفی شود

$$P(Y_s|Y_{[s]}) = \int P(Y_s|\theta, Y_{[s]})P(\theta|Y_{[s]})d\theta$$

اگر $Y_{[s]} \equiv y_i$ شامل تنها یک مشاهده باشد و $Y_{[i]}$ نشان دهنده $n - 1$ مشاهده باقی‌مانده باشد، آنگاه چگالی

$$P(y_i|Y_{[i]}) = \frac{P(Y)}{P(y_i)} = \int P(y_i|\theta, Y_{[i]})P(\theta|Y_{[i]})d\theta \quad (1.1)$$

بعنوان قرارداد پیش‌بینی مشروط (CPO) برای مورد i -ام و بصورت قابلیت پیش‌بینی توزیع شرطی هر مشاهده به شرط بقیه مشاهدات تعریف می‌شود [۱]. برآورد مونت‌کارلو CPO ، در واقع بدون حذف مورد i و بوسیله میانگین هارمونیک درست‌نمایی برای مورد i به صورت

$$CPO_i = \frac{T}{\sum_t \left\{ \frac{1}{P(y_i|\theta^{(t)})} \right\}} \quad (2.1)$$

به دست می‌آید. بنابراین اعتبارسنجی متقابل براساس N برآورد مجزا می‌تواند بوسیله یک برآورد مبتنی بر نمونه کامل تقریب زده شود [۲]. مقادیر (CPO) ممکن است با تقسیم بر مقدار بیشینه مقیاس‌پذیر شوند و مقادیر کوچک برای

هر مشاهده (کمتر از ۰/۰۱) نشان‌دهنده پرت بودن آن مشاهده و عدم برازش مناسب مدل است. اگر هیچ مشاهده‌ای با مقدار (CPO) کمتر از ۰/۰۱ وجود نداشته باشد، می‌توان نتیجه گرفت مدل برازش مناسبی دارد.

گلفند^۱ در سال ۱۹۹۶ استفاده از (CPO) را برای تبیین و انتخاب مدل‌های بیزی مطرح کرد [۳]. او عنوان کرد که اعتبارسنجی متقابل براساس حذف یک مشاهده یک جایگزین برای برآورد درستنمایی حاشیه‌ای ارائه می‌دهد. بنابراین اگر y_i شامل تنها یک مشاهده باشد و $n - 1$ مشاهده باقی‌مانده برای برآورد θ استفاده شود، آنگاه یک نماینده برای درستنمایی حاشیه‌ای $P(Y)$ ، که بعنوان درستنمایی حاشیه‌ای-نما شناخته می‌شود به صورت حاصل ضرب جملات (CPO) تعریف می‌شود

$$PsML = P(\hat{Y}) = \prod_{i=1}^n CPO_i = \prod_{i=1}^n P(y_i | Y_{[i]}). \quad (3.1)$$

مقدار $P(y_i | Y_{[i]})$ را می‌توان از رابطه (۲.۱) برآورد نمود. معمولاً لگاریتم برآورد درستنمایی حاشیه‌ای-نما در رابطه (۳.۱) را بعنوان معیاری برای مقایسه مدل‌ها مورد استفاده قرار می‌دهند که به صورت زیر تعریف می‌شود

$$LPML = \sum_{i=1}^n \log CPO_i. \quad (4.1)$$

مدلی که دارای بیشترین مقدار از این معیار باشد بعنوان مدل بهتر شناخته می‌شود.

۱.۱ انحرافات تاوانیده

یک معیار شناخته شده و مرسوم براساس انحرافات تاوانیده توسط اشپیگل‌هالتر^۲ و همکاران در سال ۲۰۰۲ معرفی گردید [۴]. این معیار که معیار اطلاعاتی انحراف نام دارد و به اختصار با DIC نشان داده می‌شود، نسخه بیزی معادل با معیار اطلاعاتی آکائیک (AIC) است. این معیار براساس کمیت انحراف تعیین می‌گردد. این کمیت که با نماد D نشان داده می‌شود، به صورت

$$D(\theta) = -2 \log f(y|\theta) \quad (5.1)$$

¹Gelfand

²Spiegelhalter

محاسبه می‌گردد. هرچه مقدار این کمیت کوچک‌تر باشد، نشانگر بهتر بودن مدل برازش داده شده است. نکته اصلی در ارائه انحرافات تاوانیده مقداری است که بعنوان جریمه لحاظ می‌شود. مقدار جریمه مرتبط با تعداد پارامترهای مؤثر در مدل و معیاری از پیچیدگی آن است. یک مشکل این است که تعداد پارامترهای مدل های پیچیده همانند مدل های سلسله مراتبی با اثرات تصادفی در واقع تعریف نشده است. اسپیکل هالتر و همکاران در واقع رویکردی برای برآورد پارامترهای مؤثر در مدل که با p_D نمایش می‌دهند، ارائه دادند. برای داده‌های y و پارامترهای θ ، p_D بوسیله رابطه

$$p_D = E[D|y, \theta] - D(\bar{\theta}|y)$$

برآورد می‌شود. در این رابطه $E[D|y, \theta]$ انحراف مورد انتظار است که بوسیله میانگین پسین نمونه‌های انحرافات تولید شده در زنجیر مارکف مونت کارلو اندازه‌گیری می‌شود. در واقع بیانگر برآورد بیز $D(\theta)$ است و بعنوان معیاری از میزان برازش مدل است. انحراف $D(\bar{\theta}|y)$ در واقع انحراف میانگین پسین پارامترها است. یعنی مقدار کمیت انحراف به ازای برآورد بیز پارامتر θ ($\bar{\theta}$) است. بنابراین معیار اطلاعاتی انحراف یا DIC بصورت

$$DIC = D(\bar{\theta}|y) + 2p_D, \quad (6.1)$$

تعریف می‌شود. در سنجش مناسب مدل هرچه مقدار DIC کمتر باشد، مدل برازش داده شده مناسب‌تر خواهد بود [۵، ۶].

کتاب نامه

- .1 Geisser S, Eddy WF. A Predictive Approach to Model Selection. Journal of the American Statistical Association. 1979 mar;74(365):153160-. Available from: <http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1979.10481632>.
- .2 Congdon P. Bayesian models for categorical data. John Wiley & Sons; .2005 .
- .3 Gelfand AE. Model determination using sampling-based methods. Markov chain Monte Carlo in practice. 1996;p. .161–145
- .4 Sarrafzadegan N, Talaei M, Sadeghi M, Kelishadi R, Oveisgharan S, Mohammadifard N, et al. The Isfahan cohort study: rationale, methods and main findings. Journal of human hypertension. .553–545:(9)25;2011
- .5 Hassanzadeh F, Kazemi I. Analysis of over-dispersed count data with extra zeros using the Poisson log-skew-normal distribution. Journal of Statistical Computation and Simulation. 2016 sep;86(13):26442662-. Available from: <http://dx.doi.org/10.1080/00949655.2015.1117086>.
- .6 Vahabi N, Kazemnejad A, Datta S. A joint overdispersed marginalized random-effects model for analyzing two or more longitudinal ordinal responses. Statistical methods in medical research. 2017 jan;p. .962280217714616

پیوست آ

کد مربوط به مدل

آ.۱ اثر تصادفی آمیختگی گاما

```
model{  
  for(i in ۱ : Ntot){  
    eta[i] ← b[۱] * Age[i] + b[۲] * Sex[i] + b[۳] * SES[i] + b[۴] * logPhysical.act[i] + b[۵] *  
    Smoking[i] + b[۶] * BMI[i]}
```

واژه‌نامه انگلیسی به فارسی

A

Akaike information criterion معیار اطلاعاتی آکائیک

C

Conditional predictive ordinate قرارداد پیش‌بینی مشروط

Cross validation اعتبارسنجی متقابل

D

Deviance information criterion معیار اطلاعاتی انحراف

H

Harmonic mean میانگین هارمونیک

M

Monte Carlo estimate برآورد مونت‌کارلو

P

Penalized deviances انحرافات تاوانیده

Pseudo-marginal likelihood درست‌نمایی حاشیه‌ای-نما

T

Training set مجموعه آموزشی

V

Validation/Prediction set مجموعه اعتبار یا پیش‌بینی

واژه‌نامه فارسی به انگلیسی

ا

Cross validation اعتبارسنجی متقابل

Penalized deviances انحرافات تاوانیده

ب

Monte Carlo estimate برآورد مونت‌کارلو

د

Pseudo-marginal likelihood درست‌نمایی حاشیه‌ای-نما

ق

Conditional predictive ordinate قرارداد پیش‌بینی مشروط

م

Training set مجموعه آموزشی

Validation/Prediction set مجموعه اعتبار یا پیش‌بینی

Akaike information criterion معیار اطلاعاتی آکائیک

Deviance information criterion معیار اطلاعاتی انحراف

Harmonic mean میانگین هارمونیک