

استنباط درستنمایی مرکب و ملاک انتخاب مدل در مدل‌های پارامترمُبنا

حسین باغیشنسی، سید محمد‌مهدی طباطبائی
دانشگاه فردوسی مشهد

تاریخ دریافت: ۱۳۸۴/۱۱/۳ تاریخ آخرین بازنگری: ۱۳۸۵/۹/۱۹

چکیده: در مدل‌های پارامترمُبنا، مشکل اساسی تقریب درستنمایی این مدلها و در واقع برآوردهای پارامترهای مدل است. یک روش برخورد با این مشکل استفاده از درستنمایی‌های ساده‌تر مانند درستنمایی مرکب است. در این مقاله پس از معرفی مدل‌های پارامتر مُبنا و درستنمایی مرکب، به معرفی یک ملاک انتخاب مدل مبتنی بر درستنمایی مرکب پرداخته‌ایم. در ادامه با استفاده از شبیه‌سازی، توانایی درستنمایی مرکب را در استنباط و انتخاب مدل صحیح در مدل‌های پارامترمُبنا نشان داده‌ایم.

واژه‌های کلیدی: داده‌های شمارشی، مدل‌های پارامترمُبنا، الگوریتم MCEM، درستنمایی مرکب، اطلاع کولبک لیبلر، زیرنمونه‌گیری پنجره‌ای.

۱ مقدمه

کاکس (۱۹۸۱) دو رده از مدل‌های مربوط به داده‌های وابسته به زمان را دسته‌بندی کرد: مدل‌های مشاهدات مُبنا و مدل‌های پارامترمُبنا.^۱ در یک مدل مشاهدات مُبنا،

آدرس الکترونیک مسئول مقاله: حسین باغیشنسی، hbaghishani@modares.ac.ir

^۱ Parameter Driven Models

۲ استنباط درستنمایی مرکب و ملاک انتخاب مدل در مدل‌های پارامترمبا

توزیع شرطی y_t به عنوان تابعی از مشاهدات گذشته y_{t-1}, \dots, y_1 در نظر گرفته می‌شود. مدل‌های خودبرگشت^۲ برای سریهای زمانی نرمال و زنجیرهای مارکوف برای داده‌های گسته، مثالهایی از این مدلها می‌باشند. فرم کلی این مدلها برای داده‌های شمارشی بصورت زیر است (زگر و کواکیش، ۱۹۸۸):

$$\log(\mu_t) = x'_t \beta + \sum_{i=1}^p \gamma_i Y_{t-i} \quad (1)$$

در مدل‌های پارامترمبا فرض می‌شود خودهمبستگی موجود در بین داده‌ها از طریق یک فرآیند پنهان^۳ تولید می‌شود که بصورت ضربی وارد مدل می‌شود. مشکل اساسی در این مدلها، تقریب تابع درستنمایی آنها و در واقع برآورد پارامترهای مدل است. پیش‌بینی مشاهدات نیز خیلی مشکلتر از مدل‌های مشاهدات‌مبا انجام می‌گیرد (برای جزئیات بیشتر دوربین و کوپمن (۲۰۰۰) و یونگ و لیزنفلد (۲۰۰۱) را مشاهده کنید).

همانطور که بیان شد عیب اصلی مدل‌های پارامترمبا سختی محاسبه و حتی گاهی قابل محاسبه نبودن تابع درستنمایی پیچیده آنهاست. با توجه به این مشکل اساسی، روشها و رهیافت‌های مختلفی برای برآورد و استنباط در این دسته از مدلها توسط آماردانان مختلفی پیشنهاد شده است. رهیافت‌های معادلات برآورد، مدل‌های خطی تعیین یافته و الگوریتم MCEM^۴ از آن جمله‌اند.

از جمله رهیافت‌های جدید برآورد پارامتر در مدل‌های مبتنی‌مبا، استفاده از انواع دیگر درستنمایی است. در سالهای اخیر علاقه و استفاده از نوعی درستنمایی معروف به شبهدrstنمایی^۵ در حال افزایش است. این نوع درستنمایی اولین بار توسط بی‌سگ (۱۹۷۴) پیشنهاد شد. لیندنسی (۱۹۸۸) این نوع درستنمایی را با نام درستنمایی مرکب معرفی کرد. انگیزه استفاده از برآورد درستنمایی مرکب، جایگزین کردن تابع درستنمایی توسط تابعی است که ساده‌تر محاسبه و در نتیجه ماکسیمم می‌شود. این نوع درستنمایی خواص نظری خوبی دارد و در بسیاری از کاربردهای

^۲ Auto-Regressive Models

^۳ Latent Process

^۴ Monte Carlo Expectation Maximization

^۵ Pseudo Likelihood

پیچیده رفتار مناسبی دارد است. در این مقاله به کاربردهای این نوع درستنمايی در این مدلها پرداخته شده است.

۲ مدلهاي پaramترمبا

مدلهاي پaramترمبا برای دادههای شمارشي، معمولاً مبتنی بر فرض زيربنائي رگرسيون پواسون همراه با وجود يك فرآيند پنهان به منظور توليد بيشپراكنش^۶ و همبستگي پياپى^۷ است (زگر، ۱۹۸۸). با اين استراتژي كلی سه دسته از مدلها را میتوان معرفی کرد. دسته اول، مبتنی بر رهيافت استاندارد بيشپراكنش برای دادههای شمارشي يك متغيره است. در اين روش دادههای شمارشي پواسون آميخته با اثرات تصادفي گاما را در نظر می گيرند که منجر به توليد چگاليهای حاشيهای دو جمله‌اي منفی می‌شود و بر اساس اين چگاليهای برآورد صورت می‌گيرد. اين دسته از مدلها در متون آماری به رگرسيون دو جمله‌اي منفی معروف شده‌اند (زگر، ۱۹۸۸). رهيافت دوم، بيشپراكنش و همبستگي پياپى را از طريقي ضرب يك فرآيند پنهان نرمال در پيشگوي خطى در دادهها القا می‌کند (چان و لدولتر، ۱۹۹۵). در اين دسته، برآوردگر کارا وجود دارد اما نياز به محاسبات سنگين و پيچيده عددی و نرمافزاری دارد و فرم توزيعهای حاشيهای و گشتاورهای آنها نيز به صورت بسته وجود ندارند. استفاده از الگوريتمهای پيچيده آماری مانند MCEM در اين دسته معمول است. رهيافت سوم مبتنی بر معادلات برآورد کننده تعليم يافته با مشخص کردن گشتاورهای مرتبه اول و دوم و برآورد از طريقي آمارههای مناسب نمونه‌اي است. اين رهيافت شيوه‌اي استوار^۸ است ولی از کارايبی لازم برخوردار نيست (زگر، ۱۹۸۸ و ديويس و همكاران، ۲۰۰۰).

بنابه فرض زيربنائي در مدلهاي مبنا، خودهمبستگي در بين دادههای شمارشي از طريقي يك فرآيند پنهان توليد می‌شود. فرض كنيد $\theta_t = \log \mu_t$ وابسته پaramتر کانوني برای مدل لگاريتم خطى باشد. بنابراین فرض می‌شود θ_t وابسته

^۶ Overdispersion

^۷ Serial Correlation

^۸ Robust

۴ استنباط درستنمايی مرکب و ملاک انتخاب مدل در مدلهاي پارامترمبا

به يك فرآيند اغتشاش غيرقابل مشاهده ايستاي ضعيف، ϵ_t ، است، يعني $(\epsilon_t, y_{t-1}, \dots, y_1)$. به عنوان مثال، ممکن است y_t با شرط ϵ_t پواسون با ميانگين $E(y_t | \epsilon_t) = \epsilon_t \exp(x_t' \beta)$ باشد. فرض كنيد y_t با شرط فرآيند پنهان ϵ_t يك دنباله از مقادير شمارشي مستقل با ميانگين و واريанс زير باشد:

$$u_t = E(y_t | \epsilon_t) = \exp(x_t' \beta) \epsilon_t, \quad w_t = \text{Var}(y_t | \epsilon_t) = u_t \quad (2)$$

همچنين فرض كنيد ϵ_t يك فرآيند ماناي نامنفي غيرقابل مشاهده با ميانگين $Cov(\epsilon_t, \epsilon_{t+h}) = \sigma^2 \rho_\epsilon(h)$ و $E(\epsilon_t) = 1$ باشد. بنابراین گشتاورهای حاشیه‌ای y_t عبارتند از (زگر، ۱۹۸۸ و دیویس و همکاران، ۲۰۰۰):

$$\begin{aligned} \mu_t &= E(y_t) = \exp(x_t' \beta), \\ \nu_t &= \text{Var}(y_t) = \mu_t + \sigma^2 \mu_t^2, \\ \rho_y(t, h) &= \text{Corr}(y_t, y_{t+h}) = \frac{\rho_\epsilon(h)}{[\{1 + (\sigma^2 \mu_t)^{-1}\} \{1 + (\sigma^2 \mu_{t+h})^{-1}\}]^{\frac{1}{2}}} \end{aligned} \quad (3)$$

بایستی دقت داشته باشید که y_t مانا نیست، لذا تابع خودهمبستگی آن از طریق μ_t به زمان وابسته است. محدودیت گشتاور اول، $E(\epsilon_t)$ ، به دلیل خاصیت قابل تفکیک شدن ميانگين فرآيند پنهان از ضریب ثابت رگرسیون است. زیرا چنانچه این شرط را قابل نشویم، ميانگين فرآيند پنهان و ضریب ثابت با هم مخلوط می‌شوند و قابل تفکیک و شناسایی نیستند. در واقع با این محدودیت ميانگين غیرشرطی، μ_t ، تنها به $x_t' \beta$ وابسته می‌شود و به گشتاورهای سری ϵ_t بستگی نخواهد داشت. با توجه به معادله دوم در روابط (۳)، $\text{Var}(y_t) = \mu_t (1 + \sigma^2 \mu_t)$. بنابراین درجه نسبی بیشپراکنش مدل به ميانگين μ_t بستگی دارد. از طرفی چون مخرج کسر در معادله سوم روابط (۳) بزرگتر از یک است، خودهمبستگی در y_t کمتر یا مساوی خودهمبستگی در ϵ_t می‌باشد. این نتیجه، تشخیص وجود یک ساختار همبستگی معنی‌دار در بین مشاهدات را بدون در نظر گرفتن فرآیند پنهان مخدوش می‌کند. درجه خودهمبستگی در y_t متناسب با ϵ_t کاهش می‌یابد، هرگاه μ_t و σ^2 کاهش یابند.

۳ درستنمايی مرکب

عنوان درستنمايی مرکب به رده قدرتمندی از شبهه درستنمايی های مبتنی بر ترکيب مولفه هايى از نوع درستنمايی اشاره مى کند. چون اين روش بطور كلی ناكاراست (کاس و ريد، ۲۰۰۳)، ممکن است اين سؤال مطرح شود که چرا علاقه مندی به استفاده از اين نوع درستنمايی در حال افزایش است؟ در پاسخ به اين پرسش مى توان دو دليل برشمرد: اول اينكه وقتی محاسبه برآورده ماکسيمم درستنمايی مشكل است، اين نوع درستنمايی يك روش جانشين و ساده را برای برآورده فراهم مى کند. دوم اينكه اين نوع درستنمايی در مدل بندی و برآورده پارامترها گاهی موقع خواصی که دلخواه ما هستند از جمله سازگاري برآورده را حتی زمانیکه برآورده گهای ماکسيمم درستنمايی سازگار نیستند، نتیجه مى دهد. در واقع يكى از بزرگترین مزیت های درستنمايی مرکب، قدرت انعطاف پذيری اين روش در تولید برآورده گهای سازگار در موقعیت های پيچide است (ليندسى، ۱۹۸۸ و کاس و ريد، ۲۰۰۳). تعریف زير،تابع درستنمايی مرکب را معرفی مى کند (وارين و ويدونى، ۲۰۰۴).

تعريف ۱ : فرض کنيد $\{f(y; \theta), y \in Y, \theta \in \Theta\}$ يك مدل آماری پارامتری باشد، بطوریکه $\Theta \subseteq \mathbb{R}^d, Y \subseteq \mathbb{R}^n$ و $n \geq 1, d \geq 1$. مجموعه پيشامدهای $\{A_i : A_i \in F, i \in I\}$ را در نظر بگيريد، که در آن $N \subseteq I$ و F يك سيگما ميدان روی Y است. بنابراين تابع درستنمايی مرکب، تابعی از θ است که بصورت زير

تعريف مى شود:

$$CL_f(\theta; y) = \left\{ \prod_{i \in I} f(y \in A_i; \theta) \right\}^{w_i} \quad (4)$$

بطوريکه: $y = (y_1, \dots, y_n)'$ با $f(y \in A_i; \theta) = f(\{y_j \in Y; y_j \in A_i\}; \theta)$ در حالیکه $\{w_i, i \in I\}$ يك مجموعه از وزنهای مناسب است. لگاريتم درستنمايی مرکب را نيز با $\log CL_f(\theta; y)$ نشان مى دهن.

مثال ۱ : در اينجا سه مثال مهم از لگاريتم های درستنمايی مرکب را بيان مى کنيم.

۶ استنباط درستنمایی مرکب و ملاک انتخاب مدل در مدل‌های پارامترمبا

- لگاریتم درستنمایی کامل که با $\log L(\theta; y) = \log f(y; \theta)$ داده می‌شود.

- لگاریتم درستنمایی جفتی^۹ که بصورت

$$\log PL(\theta; y) = \sum_{j < k} \log f(y_j, y_k; \theta) w_{(j,k)}$$

تعریف می‌شود و عمل جمع بر روی تمام جفت‌های $\{(y_j, y_k), j, k = 1, \dots, n\}$ از مشاهدات صورت می‌گیرد. در اینجا از نماد $w_{(j,k)}$ به منظور نمایش وزن متناظر با (y_j, y_k) استفاده کرده‌ایم. بطور مشابه می‌توان لگاریتم درستنمایی سه‌تایی^{۱۰} را تعریف کرد، که در آن چگالی‌های توأم سه‌تایی‌های مشاهدات مورد استفاده قرار می‌گیرند. به همین ترتیب می‌توان لگاریتم درستنمایی‌های مرکب بیشتری را تعریف کرد.

- لگاریتم شبه درستنمایی بی‌سگ^{۱۱} که شکل آن بصورت زیر تعریف می‌شود:

$$\log BPL(\theta; y) = \sum_{j=1}^n \log f(y_j | y_{(-j)}; \theta) w_j$$

که در آن عمل جمع روی تمام پیشامدهای شرطی $\{y_{(-j)} | y_j\}$ صورت می‌گیرد، بطوریکه $y_{(-j)}$ بودار مشاهدات y بدون مولفه y_j است.

برآورده‌گر ماکسیمم درستنمایی مرکب از حل معادله درستنمایی مرکب $CS(\theta) = \nabla \log CL_f(\theta; y) = 0$ بدست می‌آید که $\nabla \log CL_f(\theta; y) = \sum_{i \in I} \nabla \log f(y \in A_i; \theta) w_i$ ، عملگر مشتق‌گیری جزیی نسبت به θ است. تحت شرایط مناسب نظم برآورده‌گر ماکسیمم درستنمایی مرکب $(y, \hat{\theta}_{MCL})$ سازگار است و بطور مجانبی دارای توزیع نرمال می‌باشد (کاکس و رید، ۲۰۰۳).

^۹ Pairwise Log-Likelihood

^{۱۰} Tripletwise Log-Likelihood

^{۱۱} Besag's Pseudo Log-Likelihood

^{۱۲} Composite Score Function

۴ ملاک انتخاب مدل مبتنی بر درستنمایی مرکب

تعمیم و بسط چهارچوب درستنمایی مرکب برای انتخاب مدل، روشن و طبیعی است ولی تاکنون در متون آماری کمتر مورد توجه قرار گرفته است. وارین و ویدونی (۲۰۰۴) یک روش انتخاب مدل را مبتنی بر تعییم واگرایی کولبک لیبلر^{۱۳} معرفی کردند. در این قسمت به بررسی جزئیات مربوط به این روش انتخاب مدل پرداخته‌ایم.

تعريف ۲ : دو تابع چگالی $g(z)$ و $h(z)$ را در نظر بگیرید. اطلاع کولبک لیبلر مرکب متناظر با این دو تابع چگالی با کمیت نامنفی زیر تعریف می‌شود:

$$\begin{aligned} I_c(g, h) &= E_{g(z)} \left\{ \log \left(\frac{CL_g(Z)}{CL_h(Z)} \right) \right\} \\ &= \sum_{i \in I} E_{g(z)} \{ \log g(Z \in A_i) - \log h(Z \in A_i) \} w_i \end{aligned}$$

که در آن A_i ها همان پیشامدهای به کار رفته در تعریف تابع درستنمایی مرکب می‌باشند و امید ریاضی بر حسب $g(z)$ می‌باشد. همچنین:

$$\log CL_g(Z) = \sum_{i \in I} w_i \log g(Z \in A_i), \quad \log CL_h(Z) = \sum_{i \in I} w_i \log h(Z \in A_i)$$

ملک اطلاعی که در اینجا معرفی می‌شود، به عنوان برآورده ناریب مرتبه اول برای کمیت هدف مرتبط با اطلاع کولبک لیبلر مرکب مورد انتظار تعریف می‌شود و با چگالی درست و نامعلوم مشاهده آینده و چگالی برآورده شده متناظر آن مرتبط است. نمونه $(Y_1, \dots, Y_n) = Y$ و مدل آماری پارامتری مشخص شده توسط خانواده توابع چگالی $\{f(y; \theta), y \in Y, \theta \in \Theta\}$ را در نظر بگیرید. چندین مدل آماری دلخواه برای Y می‌تواند وجود داشته باشد که ممکن است شامل چگالی درست $g(y)$ باشد یا نباشد. مایلیم مدلی را انتخاب کنیم که توصیف پیش‌بینی راضی‌کننده‌تری برای داده‌های مشاهده شده y داشته باشد. به عنوان مثال مجموع توانهای دوم خطای

^{۱۳} Kullback-Leibler Divergence

کوچکتری را نتیجه دهد. برای توضیح بیشتر، فرض کنید Z یک متغیر تصادفی آینده باشد که یک شکل از Y و مستقل از Y است. علاقه‌مند به انتخاب بهترین مدل برای پیش‌بینی Z با فرض مشاهده Y با استفاده از روش‌های درستنماهی مرکب هستیم. انتخاب مدل می‌تواند براساس اطلاع کولبک لیبلر مرکب مورد انتظار بین تابع چگالی صحیح $(z; g)$ و تابع چگالی برآورده شده $(z; \hat{\theta}_{MCL}(Y))$ $\hat{f}(z) = f(z; \hat{\theta}_{MCL}(Y))$ تحت مدل آماری مفروض، صورت گیرد. یعنی مدلی را انتخاب می‌کنیم که $E_{g(y)}\{I_c(g, \hat{f})\}$ را می‌نیم کند یا بطور مشابه کمیت هدف (۵) را ماکسیمم کند که به آن لگاریتم درستنماهی مرکب پیش‌بین مورد انتظار^{۱۴} اطلاق می‌شود:

$$\varphi(g, f) = \sum_{i \in I} E_{g(y)} \left[E_{g(z)} \{ \log f(Z \in A_i; \hat{\theta}_{MCL}(Y)) \} \right] \quad (5)$$

چون محاسبه (۵) نیاز به شناخت چگالی $(z; g)$ دارد که شناخت آن در واقعیت غیرممکن است، بنابراین ممکن است انتخاب مدل براساس یک آماره انتخاب صورت گیرد که به عنوان برآورده مناسب (g, f) φ تعریف می‌شود. تحت شرایط مناسب نظم، ملاک انتخاب زیر که برآورده ناریب مرتبه اول برای (g, f) φ است، مبتنی بر یک آماره انتخاب است (وارین و ویدونی، ۲۰۰۴).

تعریف ۳: نمونه تصادفی Y را همان طور که در بالا تعریف کردہ‌ایم، در نظر بگیرید. ملاک اطلاع درستنماهی مرکب، CLIC، یک مدل را با ماکسیمم کردن رابطه (۶) انتخاب می‌کند:

$$\Psi^c(Y; f) = \Psi(Y; f) + \text{tr}\{\hat{J}(Y)\hat{H}^{-1}(Y)\} \quad (6)$$

بطوریکه:

$$\Psi(Y; f) = \log CL_f(\hat{\theta}_{MCL}(Y); Y) = \sum_{i \in I} \log f(Y \in A_i; \hat{\theta}_{MCL}(Y))w_i$$

و $\hat{J}(Y)$ و $\hat{H}(Y)$ به ترتیب برآورده‌های مناسب ناریب مرتبه اول و سازگار مبتنی بر Y برای $J(\theta)$ و $H(\theta)$ هستند.

^{۱۴} Expected Predictive Composite Log-Likelihood

ذکر دو نکته در اینجا حائز اهمیت است. اولین نکته مربوط به انتخاب وزنهای درستنمایی مرکب است. در درستنمایی جفتی معمولاً وزنها بر حسب یک نقطه برش^{۱۵} در جفت مشاهدات غیرهمسايه که اطلاع کمتری دارند، انتخاب می‌شوند. یعنی با تعریف یک نقطه برش به جفت‌هایی که متناظر با این نقطه در همسایگی هم قرار می‌گیرند، وزن بیشتری نسبت به نقاط غیرهمسايه داده می‌شود. یک استراتژی ساده‌تر برای وزن دادن اینست که یک دامنه همبستگی برآورده کنیم و به همه جفت‌هایی که فاصله بیشتری از این دامنه دارند، وزن صفر اختصاص دهیم. دومین نکته مربوط به برآورده کارای $J(\theta)$ و $H(\theta)$ می‌باشد. برآورد $H(\theta)$ مشکلی را ایجاد نمی‌کند و تحت شرایط استاندارد نظم، یک برآورده‌گر سازگار عبارت است از $\hat{H}(\hat{\theta}_{MCL}(Y)) = \nabla^2 \log CL(\hat{\theta}_{MCL}(Y))$ ولی برآورده $J(\theta)$ پیچیده‌تر است، چون برآورده تجربی شهودی['] $= \nabla \log PL(\theta; y) \nabla \log PL(\theta; y)$ در نقطه $\hat{J}(\theta; y)$ ماقسیم درستنمایی مرکب، صفر می‌شود (لوملی و هگارتی، ۱۹۹۹). در عمل، برآورده $J(\theta)$ به وسیله روشهای مختلفی که وابسته به نوع درستنمایی مرکب است و بطور خاص درستنمایی جفتی در نظر گرفته می‌شود، انجام می‌گیرد. یکی از روشهای روش نمونه‌گیری دوباره استفاده شده^{۱۶} معروف به زیرنمونه‌گیری پنجره‌ای^{۱۷} است (هگارتی و لله، ۱۹۹۸ و لوملی و هگارتی، ۱۹۹۹). بطور خلاصه انجام این روش برای سری مشاهدات['] (y_1, \dots, y_n) شامل مراحل زیر است:

(۱) یک مجموعه متداول^{۱۸} از زیرسری‌های با بعد m

$$y_i^{i+m} = (y_i, \dots, y_{i+m-1})'$$

(۲) برآورده‌گر تجربی $(\hat{J}(\hat{\theta}_{MPL}; y_i^{i+m}))$ را در هر زیرسری محاسبه می‌کنیم.

(۳) میانگین این برآوردها را با یک تبدیل مقیاس مناسب برای لحاظ کردن بعد زیرسری‌ها، محاسبه کرده و برآورده زیرنمونه‌گیری پنجره‌ای زیر را ارایه

^{۱۵} Cut Point

^{۱۶} Reuse Sampling

^{۱۷} Window Subsampling

^{۱۸} Overlapping

۱۰ استنباط درستنماهی مرکب و ملاک انتخاب مدل در مدل‌های پارامترمبا

می‌دهیم:

$$\hat{J}^{(m)}(\hat{\theta}_{MPL}; y) = \frac{\sum_{i=1}^{n-m+1} \hat{J}(\hat{\theta}_{MPL}; y_i^{i+m}) \frac{m}{n}}{n-m+1}$$

هگارتی و لوملی (۲۰۰۰) نشان دادند که بعد بهینه m برای زیرپنجره‌ها، $Cn^{1/3}$ است که C یک ثابت با مقادیر پیشنهادی بین ۴ و ۸ است که به قدرت وابستگی درون داده‌ها بستگی دارد.

۵ شبیه‌سازی

در این قسمت با انجام شبیه‌سازی‌هایی، دو کاربرد روش‌های درستنماهی مرکب با تأکید بر جنبه‌های استنباطی مبتنی بر این درستنماهی در مدل‌های پارامترمبا نشان داده شده است. این دو کاربرد، برآورد مدل و انتخاب مدل را شامل می‌شوند. برای تشریح عملکرد ملاک انتخاب درستنماهی مرکب، CLIC، همچنین نمایش قدرت این نوع درستنماهی در برآورد مدل، دو مدل پارامترمبا در نظر گرفته شده است. در هر دو مدل، متغیر تبیینی و حتی ضریب ثابت وجود ندارد. اولین مدل، مدل ساده پواسون با فرآیند پنهان (۱) $AR(1)$ است. یعنی:

$$Y_t | \alpha_t \sim Poisson(\exp \alpha_t), \quad t = 1, \dots, n$$
$$\alpha_t = \phi \alpha_{t-1} + \epsilon_t, \quad t = 2, \dots, n$$

بطوریکه $\epsilon_t \sim N(0, \sigma^2)$. همچنین برای اینکه فرآیند پنهان مانا باشد، فرض شده است $1 < |\phi|$ و در نتیجه $\alpha_t \sim N(0, \frac{\sigma^2}{(1-\phi^2)})$.

مدل دوم با استفاده از اولین رهیافت یعنی مبتنی بر رهیافت استاندارد بیش‌پراکنش برای داده‌های شمارشی، شبیه‌سازی شده است. در اینجا فرض شده است که $Y_t | \alpha_t, t = 1, \dots, n$ دارای توزیع دوجمله‌ای منفی با میانگین $\mu_t = \exp(\alpha_t)$ و پارامتر مقیاس $\sigma > 0$ است. پس در نتیجه برای $t = 1, \dots, n$

داريم:

$$f(y_t | \alpha_t; k) = \frac{\Gamma(k^{-1} + y_t)}{\Gamma(k^{-1})y_t!} \left(\frac{k\mu_t}{1 + k\mu_t} \right)^{y_t} \left(\frac{1}{1 + k\mu_t} \right)^{1/k}, \quad y_t = 0, 1, \dots$$

توجه داشته باشيد دو مدلی که در نظر گرفته شده اند، لانهای هستند زیرا چنانچه $\rightarrow k$ ، مدل دوجمله‌ای منفی به مدل پواسون می‌کند. مقایسه این دو مدل از نظر عملی جالب توجه است، زیرا حضور فرآیند پنهان (۱) در مدل پواسونی ممکن است توصیف کاملی از بیش‌پراکنش داده‌ها ارایه ندهد. تحلیل این دو مدل با استفاده از درستنمایی‌های استاندارد و روش‌های انتخاب مدل مشکل است، زیرا برای محاسبه تابع درستنمایی باید انتگرال n گانه زیر را حل کnim که کار خیلی مشکلی است:

$$L(\theta; y) = \int \prod_{t=1}^n f(y_t | \alpha_t; \theta) f(\alpha_t | \alpha_{t-1}) d\alpha_1 \dots d\alpha_n$$

که در آن $f(\alpha_1 | \alpha_0; \theta) = f(\alpha_1; \theta)$. برای تقریب انتگرال بالایی در شبیه‌سازی‌ها به منظور سادگی بیشتر از درستنمایی جفتی استفاده شده است. در واقع درستنمایی جفتی با استفاده از $(1 - n)$ مشاهده جفتی، بصورت زیر محاسبه می‌شود:

$$PL(\theta; y) = \prod_{t=1}^n \int \int f(y_t | \alpha_t; \theta) f(y_{t-1} | \alpha_{t-1}; \theta) f(\alpha_t | \alpha_{t-1}; \theta) f(\alpha_{t-1}; \theta) d\alpha_t d\alpha_{t-1}$$

محاسبه $(1 - n)$ انتگرال دوگانه فوق که درستنمایی جفتی را تعریف می‌کند، خیلی ساده‌تر از محاسبه درستنمایی کامل است. برای محاسبه $(1 - n)$ انتگرال دوگانه درستنمایی جفتی از روش تربیع گاوس هرمیت^{۱۹} استفاده شده است. این روش برای تقریب انتگرهای با بعد پایین بسیار مناسب است اما در ابعاد بالا ضعیف عمل می‌کند (شان و مک کالاک، ۱۹۹۵).

در اولین شبیه‌سازی که به منظور بررسی نحوه برآورد مدل صورت گرفته است، ۵۰۰ مجموعه داده (تعداد تکرار) با حجم نمونه ۳۰۰ از مدل پواسون با فرآیند پنهان (۱) $AR(1)$ تولید شده که در آن $\sigma = 1/\sqrt{35}$ و $\phi = 0.299$. انتگرال دوگانه با استفاده از تربیع گاوس هرمیت دوگانه با 10^0 گره^{۲۰} برای هر بعد و در نتیجه 10^0 گره برای هر

^{۱۹} Gauss-Hermit Quadrature

^{۲۰} Node

۱۲ استنباط درستنماهی مرکب و ملاک انتخاب مدل در مدل‌های پارامترمبا

دو بعد، تقریب زده شده‌اند. همین فرآیند شبیه‌سازی برای $\phi = 0/5$ و $\sigma = 1/2$ تکرار شده است.

در دومین شبیه‌سازی، مسئله انتخاب مدل مورد بررسی قرار گرفته است. در این بررسی 100 مجموعه داده با حجم نمونه 300 از مدل دوجمله‌ای منفی تولید شده که در آن فرآیند پنهان بصورت یک فرآیند حالت $(1) AR$ با $\phi = 0/5$ و $\sigma = 0/5$ تعریف شده است. برای پارامتر مقیاس k چندین مقدار در نظر گرفته شده، یعنی $0, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, k = 1$. دقت کنید که در حالت $0, k = 1$ ، مدل به مدل پواسون با فرآیند $(1) AR$ برای α_t تبدیل می‌شود. ماتریس $J(\theta)$ نیز با یک روش زیرنمونه‌گیری پنجره‌ای شامل 251 زیرسری متداول از مشاهدات با بعد $m = 50$ برآورد شده است. برای همگرایی نیز از قاعده $\hat{\delta}^{(d+1)} - \hat{\delta}^{(d)}$ برای توقف الگوریتم برآورد استفاده شده است که در آن $\hat{\delta}^{(d)}$ برآورد پارامتر در مرحله d ام الگوریتم می‌باشد.

نتایج حاصل از شبیه‌سازی اول را می‌توانید در جدول ۱ ببینید. در این جدول مقدار برآورد پارامترها، متوسط برآوردهای درستنماهی جفتی شبیه‌سازی شده در تکرارها است و انحراف معیارهای برآوردهای پارامترها نیز متوسط انحراف معیارهای برآوردهای پارامترها در تکرارها است. همان طور که از نتایج جدول ملاحظه می‌کنید، در هر دو حالت برآوردهای دوجمله‌ای جفتی تمایل ناچیز به کم برآورد کردن σ از خود نشان می‌دهند. اما مقدار برآوردها خیلی نزدیک به مقادیر واقعی پارامترها در فرآیند پنهان هستند.

در شبیه‌سازی دوم، هدف مقایسه مدل‌های پواسون و دوجمله‌ای منفی در حضور فرآیند پنهان $(1) AR$ با استفاده از $CLIC$ است. جدول ۲ فراوانی انتخاب‌های درست مدل را در 100 مجموعه داده شبیه‌سازی شده برای مقادیر مختلف k نشان می‌دهد. دقت داشته باشید که $k = 0$ معادل مدل پواسون است. در این حالت برای فرآیند پنهان یک بار $\phi = 0/25$ و $\sigma = 0/5$ در نظر گرفته شده است و بار دوم $\phi = 0/6$ و $\sigma = 0/7$. نتایج برای حالتی که $k > 0$ ، گزارش نشده‌اند، زیرا $CLIC$ همیشه مدل درست که همان مدل دوجمله‌ای منفی است را انتخاب می‌کند. در هر دو فرآیند در نظر گرفته شده، نتایج مشابهی حاصل شده‌اند. این نتایج یک توجیه اولیه را برای استفاده از روش درستنماهی مرکب به منظور اهداف انتخاب

مدل نشان می‌دهند. با توجه به نتایج بدست آمده، همان طور که انتظار داشتیم وقتی که سمت صفر میل می‌کند، $CLIC$ بیشتر به انتخاب مدل پواسون متمایل می‌شود. وقتی که k کمتر از 25° است، اغلب مقادیر مشاهده شده آماره انتخاب مبتنی بر درستنمايی جفتی خيلي به هم نزديك می‌شوند. اين نتایج توسط نمودارهاي جعبه‌اي در شكلهاي ۱ و ۲ برای دو مدل شبیه‌سازی شده نشان داده شده‌اند. در اين نمودارها اختلافهای مشاهده شده بین آماره‌های انتخاب برای دو مدل پواسون و دو جمله‌ای منفی به نمایش درآمده‌اند. به ویژه برای $CLIC$ $k = \frac{1}{\lambda}, {}^{\circ}$ ، اختلافها اغلب ناچيزند.

جدول ۱: برآورد پارامترها با استفاده از درستنمايی جفتی

پارامتر	مقدار صحیح	مقدار برآورد	انحراف معیار
ϕ	$0/35$	$0/34996$	$0/1128$
	$-0/5$	$-0/4922$	$0/0671$
σ	۱	$0/982$	$0/0831$
	$1/2$	$1/1397$	$0/0897$

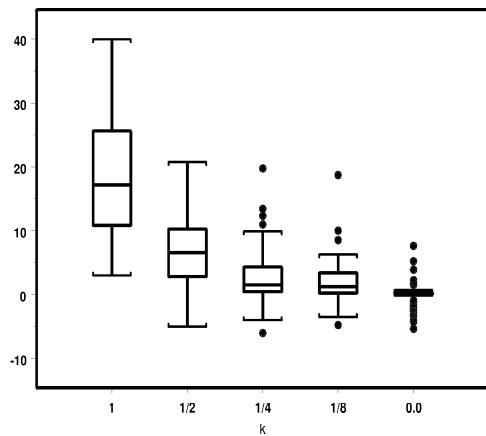
جدول ۲: فراوانی انتخاب درست مدل با فرآیند پنهان (۱) $AR(1)$

		k				
ϕ	σ	۱	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	${}^{\circ}$
$0/35$	$0/5$	۱۰۰	۹۱	۶۳	۵۳	۵۱
$-0/6$	$0/7$	۹۷	۸۷	۵۹	۴۲	۴۹

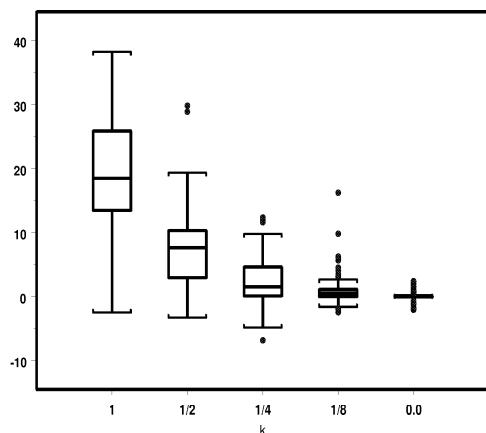
۶ بحث و نتیجه‌گیری

با اينکه درستنمايی مرکب بطور کلى ناكاراست اما به دليل تولید برآوردهای سازگار و سادگی محاسبات مربوط به آن، مورد توجه زیادي قرار گرفته است.

۱۴ استنباط درستنایی مرکب و ملاک انتخاب مدل در مدل‌های پارامترمبا



شکل ۱: اختلاف بین آماره‌های انتخاب در فرآیند اول



شکل ۲: اختلاف بین آماره‌های انتخاب در فرآیند دوم

استفاده از این نوع درستنمايی در مدلهاي پارامتر مبنا به منظور برآورد و استنباط، مشکلات محاسبات سنگين و پيچide تابع درستنمايی در آنها را بروز کرده است. نتایج شبیه سازی نشان دادند که نتایج برآورد و استنباط با استفاده از این نوع درستنمايی در مدلهاي پارامتر مبنا و همچنین مسئله انتخاب مدل درست با نتایج نظری توافق زیادي دارند.

تقدیر و تشکر

از داوران محترم این مقاله بخاطر نظرات سازنده شان در بهبود این اثر، تشکر و قدردانی می شود.

مراجع

- Besag, J. E. (1974), Spatial Interaction and the Statistical Analysis of Lattice Systems (with Discussion), *Journal of the Royal Statistical Society, B*, **36(2)**, 192-236.
- Chun, K. S. and Ledolter, J. (1995), Monte Carlo EM Estimation for Time Series Models Involving Counts, *Journal of the American Statistical Association*, **90(429)**, 242-252.
- Cox, D. R. (1981), Statistical Analysis of Time Series: Some Recent Developments, *Scandinavian Journal of Statistics*,
- Cox, D. R. and Reid, N. (2003), A Note on Pseudolikelihood Constructed from Marginal Densities, *Technical Report*, **8**, 93-115.
URL: <http://www.ustat.toronto.edu/reid/research/html>.

۱۶ استنباط درستنماهی مرکب و ملاک انتخاب مدل در مدل‌های پارامتر مبنا

Davis, R. A., Dunsmuir, W. T. M. and Wang, Y. (2000), On Autocorrelation in a Poisson Regression Model, *Biometrika*, **87(3)**, 491-506.

Durbin, J. and Koopman, S. J. (2000), Time Series Analysis of Non-Gaussian Observation Based on State Space Models from both Classical and Bayesian Perspectives, *Journal of the Royal Statistical Society, B*, **62(1)**, 3-56.

Heagerty, P. J. and Lele, S. R. (1998), A Composite Likelihood Approach to Binary Spatial Data, *Journal of the American Statistical Association*, **93(443)**, 1099-1111.

Heagerty, P. J. and Lumley, T. (2000), Window Subsampling of Estimating Functions with Application to Regression Models, *Journal of the American Statistical Association*, **95(449)**, 197-211.

Jung, R. C. and Liesenfeld, R. (2001), Estimating Time Series Models for Count Data Using Efficient Importance Sampling, *Allgemeines Statistisches Archive*, **85(4)**, 387-407.

Lindsay, B. G. (1988). Composite Likelihood Methods, *Contemporary Mathematics*, **80**, 221-239.

Lumley, T. and Heagerty, P. (1999), Weighted Empirical Adaptive Variance Estimators for Correlated Data Regression, *Journal of the Royal Statistical Society, B*, **61(2)**, 459-477.

Shun, Z. and McCullagh, C. E. (1995), Laplace Approximation of High-Dimensional Integrals, *Journal of the Royal Statistical Society, B*, **57(4)**, 749-760.

Varin, C. and Vidoni, P. (2004), A Note on Composite Likelihood Inference and Model Selection, *Working Paper*, URL:<http://www.stat.unipd.it>.

Zeger, S. L. (1988), A Regression Model for Time Series of Counts, *Biometrika*, **75(4)**, 621-629.

Zeger, S. L. and Qaqish, B. (1988), Markov Regression Models for Time Series: A QuasiLikelihood Approach, *Biometrics*, **44(4)**, 1019-1031.