مجله علوم آماری، بهار ۱۳۸۶
جلد ۱، شماره ۱، ص ۱-۱۷

استنباط درستنمایی مرکب و ملاک انتخاب مدل در مدلهای پارامترمبنا

حسین باغیشنی، سید محمدمهدی طباطبایی
دانشگاه فردوسی مشهد

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

چکیده : در مدلهای پارامترمبنا، مشکل اساسی تقریب درستنمایی این مدلها و در واقع برآورد 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 مبنا 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 پارامترمبنا نشان داده‌ایم.

واژه‌های کلیدی : د اده‌های شمارشی، مدلهای پارامترمبنا، 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 مشاهدات‌مبنا و مدلهای پارامترمبنا[۱]. در یک مدل مشاهدات مبنا،

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

۱ Parameter Driven Models

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

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

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

در مدلهای پارامترمبنا 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 مشاهدات‌مبنا انجام می‌گیرد (برای جزییات بیشتر دوربین و کوپمن (۲۰۰۰) و یونگ و لیزنفلد (۲۰۰۱) را مشاهده کنید).

همانطور که بیان شد عیب اصلی مدلهای پارامترمبنا 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 تعمیم یافته و الگوریتم MCEM[۴] از آن جمله‌اند.

از جمله رهیافتهای جدید برآورد پارامتر در مدلهای مبتنی‌مبنا، استفاده از انواع دیگر درستنمایی است. در سالهای اخیر علاقه و استفاده از نوعی درستنمایی معروف به شبه‌درستنمایی[۵] در حال افزایش است. این نوع درستنمایی اولین بار توسط بی‌سگ (۱۹۷۴) پیشنهاد شد. لیندسی (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

[Downloaded from jss.irstat.ir on 2026-11-08]
[DOR: 20.1001.1.17358183.1386.1.1.15.2]

پیچیده رفتار مناسبی داراست. در این مقاله به کاربردهای این نوع درستنمایی در این مدلها پرداخته شده است.

۲ مدلهای پارامترمبنا

مدلهای پارامترمبنا برای داده‌های شمارشی، معمولاً مبتنی بر فرض زیربنایی رگرسیون پواسون همراه با وجود یک فرآیند پنهان به منظور تولید بیش‌پراکنش[۶] و همبستگی پیاپی[۷] است (زگر، ۱۹۸۸). با این استراتژی کلی سه دسته از مدلها را می‌توان معرفی کرد. دسته اول، مبتنی بر رهیافت استاندارد بیش‌پراکنش برای داده‌های شمارشی یک متغیره است. در این روش داده‌های شمارشی پواسون آمیخته با اثرات تصادفی گاما را در نظر می‌گیرند که منجر به تولید چگالی‌های حاشیه‌ای دوجمله‌ای منفی می‌شود و بر اساس این چگالی‌ها برآورد صورت می‌گیرد. این دسته از مدلها در متون آماری به رگرسیون دوجمله‌ای منفی معروف شده‌اند (زگر، ۱۹۸۸). رهیافت دوم، بیش‌پراکنش و همبستگی پیاپی را از طریق ضرب یک فرآیند پنهان نرمال در پیشگوی خطی در داده‌ها القا می‌کند (چان و لدولتر، ۱۹۹۵). در این دسته، برآوردگر کارا وجود دارد اما نیاز به محاسبات سنگین و پیچیده عددی و نرم‌افزاری دارد و فرم توزیع‌های حاشیه‌ای و گشتاورهای آنها نیز به صورت بسته وجود ندارند. استفاده از الگوریتمهای پیچیده آماری مانند MCEM در این دسته معمول است. رهیافت سوم مبتنی بر معادلات برآورد کننده تعمیم‌یافته با مشخص کردن گشتاورهای مرتبه اول و دوم و برآورد از طریق آماره‌های مناسب نمونه‌ای است. این رهیافت شیوه‌ای استوار[۸] است ولی از کارایی لازم برخوردار نیست (زگر، ۱۹۸۸ و دیویس و همکاران، ۲۰۰۰).

بنابه فرض زیربنایی در مدلهای‌مبنا، خودهمبستگی در بین داده‌های شمارشی از طریق یک فرآیند پنهان تولید می‌شود. فرض کنید $\theta_t = \log \mu_t$، پارامتر کانونی برای مدل لگاریتم خطی باشد. بنابراین فرض می‌شود θ_t وابسته

---

۶ Overdispersion

۷ Serial Correlation

۸ Robust

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

به یک فرآیند اغتشاش غیرقابل مشاهده ایستای ضعیف، ϵ_t، است، یعنی $\theta_t = \theta(\epsilon_t, y_{t-1}, \ldots, y_1)$. به عنوان مثال، ممکن است y_t با شرط ϵ_t پواسون با میانگین $E(y_t \mid \epsilon_t) = \epsilon_t \exp\left(x_t'\beta\right)$ باشد. فرض کنید y_t با شرط فرآیند پنهان ϵ_t، یک دنباله از مقادیر شمارشی مستقل با میانگین و واریانس زیر باشد:

$$u_t = E(y_t \mid \epsilon_t) = \exp\left(x_t'\beta\right)\epsilon_t, \quad w_t = Var(y_t \mid \epsilon_t) = u_t \qquad (۲)$$

همچنین فرض کنید ϵ_t یک فرآیند مانای نامنفی غیرقابل مشاهده با میانگین $E(\epsilon_t) = ۱$ و $Cov(\epsilon_t, \epsilon_{t+h}) = \sigma^۲\rho_\epsilon(h)$ باشد. بنابراین گشتاورهای حاشیه‌ای y_t عبارتند از (زگر، ۱۹۸۸ و دیویس و همکاران، ۲۰۰۰):

$$\begin{aligned} \mu_t = E(y_t) &= \exp(x_t'\beta), \\ \nu_t = Var(y_t) &= \mu_t + \sigma^۲\mu_t^۲, \\ \rho_y(t,h) = Corr(y_t, y_{t+h}) &= \frac{\rho_\epsilon(h)}{\left[\{۱ + (\sigma^۲\mu_t)^{-۱}\}\{۱ + (\sigma^۲\mu_{t+h})^{-۱}\}\right]^{\frac{۱}{۲}}} \end{aligned} \qquad (۳)$$

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 میانگین فرآیند پنهان از ضریب ثابت رگرسیون است. زیرا چنانچه این شرط را قایل نشویم، میانگین فرآیند پنهان و ضریب ثابت با هم مخلوط می‌شوند و قابل تفکیک و شناسایی نیستند. در واقع با این محدودیت میانگین 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)، $Var(y_t) = \mu_t\left(۱ + \sigma^۲\mu_t\right)$. بنابراین درجه نسبی بیش‌پراکنش مدل به میانگین 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 $\sigma^۲$ کاهش یابند.

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

۳ درستنمایی مرکب

عنوان درستنمایی مرکب به رده قدرتمندی از شبه‌درستنمایی‌های مبتنی بر ترکیب مولفه‌هایی از نوع درستنمایی اشاره می‌کند. چون این روش بطور کلی ناکاراست (کاکس و رید، ۲۰۰۳)، ممکن است این سئوال مطرح شود که چرا علاقه‌مندی به استفاده از این نوع درستنمایی در حال افزایش است؟ در پاسخ به این پرسش می‌توان دو دلیل برشمرد: اول اینکه وقتی محاسبه برآورد ماکسیمم درستنمایی مشکل است، این نوع درستنمایی یک روش جانشین و ساده را برای برآورد فراهم می‌کند. دوم اینکه این نوع درستنمایی در مدل‌بندی و برآورد پارامترها گاهی مواقع خواصی که دلخواه ما هستند از جمله سازگاری برآوردگرها را حتی زمانیکه برآوردگرهای ماکسیمم درستنمایی سازگار نیستند، نتیجه می‌دهد. در واقع یکی از بزرگترین مزیت‌های درستنمایی مرکب، قدرت انعطاف‌پذیری این روش در تولید برآوردگرهای سازگار در موقعیت‌های پیچیده است (لیندسی، ۱۹۸۸ و کاکس و رید، ۲۰۰۳). تعریف زیر، تابع درستنمایی مرکب را معرفی می‌کند (وارین و ویدونی، ۲۰۰۴).

تعریف ۱ : فرض کنید $\{f(y;\theta), y \in Y, \theta \in \Theta\}$ یک مدل آماری پارامتری باشد، بطوریکه $Y \subseteq \Re^n, \Theta \subseteq \Re^d$ و $d \geq 1, n \geq 1$. مجموعه پیشامدهای $\{A_i : A_i \in F, i \in I\}$ را در نظر بگیرید، که در آن $I \subseteq N$ و F یک سیگما میدان روی Y است. بنابراین تابع درستنمایی مرکب، تابعی از θ است که بصورت زیر تعریف می‌شود:

$$CL_f(\theta; y) = \left\{ \prod_{i \in I} f(y \in A_i; \theta) \right\}^{w_i} \qquad (۴)$$

بطوریکه: $f(y \in A_i; \theta) = f\left(\{y_j \in Y; y_j \in A_i\}; \theta\right)$ با $y = (y_1, \ldots, y_n)'$ در حالیکه $\{w_i, i \in I\}$ یک مجموعه از وزنهای مناسب است. لگاریتم درستنمایی مرکب را نیز با $\log CL_f(\theta; y)$ نشان می‌دهند.

مثال ۱ : در اینجا سه مثال مهم از لگاریتم‌های درستنمایی مرکب را بیان می‌کنیم.

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

- 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 $\{(y_j, y_k),\ j,k = ۱,\ldots,n\}$ 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:

$$\log BPL(\theta;y) = \sum_{j=۱}^{n} \log f(y_j \mid y_{(-j)};\theta) w_j$$

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

برآوردگر ماکسیمم درستنمایی مرکب از حل معادله درستنمایی مرکب $CS(\theta) = \nabla \log CL_f(\theta;y) = ۰$ بدست می‌آید که $\nabla \log CL_f(\theta;y) = \sum_{i\in I} \nabla \log f(y \in A_i;\theta) w_i$ تابع امتیاز مرکب[۱۲] و عملگر ∇، عملگر مشتق‌گیری جزیی نسبت به θ است. تحت شرایط مناسب نظم برآوردگر ماکسیمم درستنمایی مرکب $\hat{\theta}_{MCL}(y)$، 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

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

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 تعمیم واگرایی کولبک لیبلر[۱۳] معرفی کردند. در این قسمت به بررسی جزییات مربوط به این روش انتخاب مدل پرداخته‌ایم.

تعریف 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_iها همان پیشامدهای به کار رفته در تعریف تابع درستنمایی مرکب می‌باشند و امید ریاضی برحسب $g(z)$ می‌باشد. همچنین:

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

ملاک اطلاعی که در اینجا معرفی می‌شود، به عنوان برآوردگر ناااریب 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 برآورد شده متناظر آن مرتبط است.

نمونه $Y=(Y_1,\ldots,Y_n)$ 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

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

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 درستنمایی مرکب هستیم. انتخاب مدل می‌تواند براساس اطلاع کولبک لیب‌لر مرکب مورد انتظار بین تابع چگالی صحیح $g(z)$ و تابع چگالی برآورد شده $\hat{f}(z) = f\left(z;\hat{\theta}_{MCL}(Y)\right)$ تحت مدل آماری مفروض، صورت گیرد. یعنی مدلی را انتخاب می‌کنیم که $E_{g(y)}\{I_c(g,\hat{f})\}$ را می‌نیمم کند یا بطور مشابه کمیت هدف (۵) را ماکسیمم کند که به آن لگاریتم درستنمایی مرکب پیش‌بین مورد انتظار[۱۴] اطلاق می‌شود:

$$\varphi(g,f) = \sum_{i\in I} E_{g(y)}\left[E_{g(z)}\{\log f(Z\in A_i;\hat{\theta}_{MCL}(Y)\}\right] \qquad (۵)$$

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

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

$$\Psi^c(Y;f) = \Psi(Y;f) + tr\{\hat{J}(Y)\hat{H}^{-1}(Y)\} \qquad (۶)$$

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 برآوردگرهای مناسب نااریب مرتبه اول و سازگار مبتنی بر Y برای $J(\theta)$ و $H(\theta)$ هستند.

[۱۴] Expected Predictive Composite Log-Likelihood

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

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

۱) یک مجموعه متداخل[۱۸] از زیرسری‌های با بعد m، $y_i^{i+m} = (y_i, \ldots, y_{i+m-1})'$، را اختیار می‌کنیم.

۲) برآوردگر 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 برآورد زیرنمونه‌گیری پنجره‌ای زیر را ارایه

---

۱۵ Cut Point

۱۶ Reuse Sampling

۱۷ Window Subsampling

۱۸ Overlapping

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

می‌دهیم:

$$\hat{J}^{(m)}(\hat{\theta}_{MPL};y)=\frac{\sum_{i=۱}^{n-m+۱}\hat{J}(\hat{\theta}_{MPL};y_i^{i+m})\frac{m}{n}}{n-m+۱}$$

هگارتی و لوملی (۲۰۰۰) نشان دادند که بعد بهینه m برای زیرپنجره‌ها، $Cn^{۱/۳}$ 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 درستنمایی مرکب با تأکید بر جنبه‌های روشهای استنباطی مبتنی بر این درستنمایی در مدلهای پارامترمبنا نشان داده شده است. این دو کاربرد، برآورد مدل و انتخاب مدل را شامل می‌شوند. برای تشریح عملکرد ملاک انتخاب درستنمایی مرکب، CLIC، و همچنین نمایش قدرت این نوع درستنمایی در برآورد مدل، دو مدل پارامترمبنا 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 $AR(۱)$ است. یعنی:

$$\begin{aligned} Y_t \mid \alpha_t &\sim Poisson(\exp\alpha_t), \quad t=۱,\ldots,n \\ \alpha_t &= \phi\alpha_{t-۱}+\epsilon_t, \qquad\quad t=۲,\ldots,n \end{aligned}$$

بطوریکه $\epsilon_t \overset{iid}{\sim} N(۰,\sigma^۲)$. همچنین برای اینکه فرآیند پنهان مانا باشد، فرض شده است $|\phi|<۱$ و در نتیجه $\alpha_t \sim N(۰,\frac{\sigma^۲}{(۱-\phi^۲)})$.

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 $Y_t \mid \alpha_t, t=۱,\ldots,n$ دارای توزیع دوجمله‌ای منفی با میانگین $\mu_t=\exp(\alpha_t)$ و پارامتر مقیاس $k>۰$ است. پس در نتیجه برای $t=۱,\ldots,n$،

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

داریم:

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

توجه داشته باشید دو مدلی که در نظر گرفته شده‌اند، لانه‌ای هستند زیرا چنانچه $k \longrightarrow 0$، مدل دوجمله‌ای منفی به مدل پواسون میل می‌کند. مقایسه این دو مدل از نظر عملی جالب توجه است، زیرا حضور فرآیند پنهان $AR(1)$ 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 nگانه زیر را حل کنیم که کار خیلی مشکلی است:

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

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

$$PL(\theta; y) = \prod_{t=2}^{n} \int \int f(y_t \mid \alpha_t; \theta) f(y_{t-1} \mid \alpha_{t-1}; \theta) f(\alpha_t \mid \alpha_{t-1}; \theta) f(\alpha_{t-1}; \theta) d\alpha_t d\alpha_{t-1}$$

محاسبه $(n-1)$ انتگرال دوگانه فوق که درستنمایی جفتی را تعریف می‌کند، خیلی ساده‌تر از محاسبه درستنمایی کامل است. برای محاسبه $(n-1)$ انتگرال دوگانه درستنمایی جفتی از روش تربیع گاوس هرمیت[۱۹] استفاده شده است. این روش برای تقریب انتگرالهای 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 $AR(1)$ تولید شده که در آن $\phi = 0/35$ و $\sigma = 1$. ۲۹۹ انتگرال دوگانه با استفاده از تربیع گاوس هرمیت دوگانه با ۱۰ گره[۲۰] برای هر بعد و در نتیجه ۱۰۰ گره برای هر

[۱۹] Gauss-Hermit Quadrature

[۲۰] Node

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

دو بعد، تقریب زده شده‌اند. همین فرآیند شبیه‌سازی برای $\phi = -۰/۵$ و $\sigma = ۱/۲$، تکرار شده است.

در دومین شبیه‌سازی، مسأله انتخاب مدل مورد بررسی قرار گرفته است. در این بررسی ۱۰۰ مجموعه داده با حجم نمونه ۳۰۰ از مدل دوجمله‌ای منفی تولید شده که در آن فرآیند پنهان بصورت یک فرآیند حالت $AR(۱)$ با $\phi = ۰/۳۵$ و $\sigma = ۰/۵$، تعریف شده است. برای پارامتر مقیاس k چندین مقدار در نظر گرفته شده، یعنی $k = ۱, \frac{۱}{۲}, \frac{۱}{۴}, \frac{۱}{۸}, ۰$. دقت کنید که در حالت $k = ۰$، مدل به مدل پواسون با فرآیند $AR(۱)$ برای α_t تبدیل می‌شود. ماتریس $J(\theta)$ نیز با یک روش زیرنمونه‌گیری پنجره‌ای شامل ۲۵۱ زیرسری متداخل از مشاهدات با بعد $m = ۵۰$، برآورد شده است. برای همگرایی نیز از قاعده $|\frac{\hat{\delta}^{(d+۱)} - \hat{\delta}^{(d)}}{\hat{\delta}^{(d)}}|$ برای توقف الگوریتم برآورد استفاده شده است که در آن $\hat{\delta}^{(d)}$ برآورد پارامتر در مرحله dام الگوریتم می‌باشد.

نتایج حاصل از شبیه‌سازی اول را می‌توانید در جدول ۱ ببینید. در این جدول مقدار برآورد پارامترها، متوسط برآوردهای درستنمایی 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 برآوردگرهای درستنمایی جفتی تمایل ناچیز به کم‌برآورد 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 $AR(۱)$ با استفاده از $CLIC$ است. جدول ۲ فراوانی انتخاب‌های درست مدل را در ۱۰۰ مجموعه داده شبیه‌سازی شده برای مقادیر مختلف k، نشان می‌دهد. دقت داشته باشید که $k = ۰$ معادل مدل پواسون است. در این حالت برای فرآیند پنهان یک بار $\phi = ۰/۳۵$ و $\sigma = ۰/۵$ در نظر گرفته شده است و بار دوم $\phi = -۰/۶$ و $\sigma = ۰/۷$. نتایج برای حالتی که $k > ۰$، 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 درستنمایی مرکب به منظور اهداف انتخاب

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

مدل نشان می‌دهند. با توجه به نتایج بدست آمده، همان طور که انتظار داشتیم وقتی که k به سمت صفر میل می‌کند، $CLIC$ بیشتر به انتخاب مدل پواسون متمایل می‌شود. وقتی که k کمتر از ۰/۲۵ است، اغلب مقادیر مشاهده شده آماره انتخاب مبتنی بر درستنمایی جفتی خیلی به هم نزدیک می‌شوند. این نتایج توسط نمودارهای جعبه‌ای در شکلهای ۱ و ۲ برای دو مدل شبیه‌سازی شده نشان داده شده‌اند. در این نمودارها اختلافهای مشاهده شده بین آماره‌های انتخاب $CLIC$ برای دو مدل پواسون و دوجمله‌ای منفی به نمایش درآمده‌اند. به ویژه برای $k = \frac{۱}{۸}, ۰$، اختلافهااغلب ناچیزند.

جدول ۱: برآورد پارامترها با استفاده از درستنمایی جفتی

پارامتر	مقدار صحیح	مقدار برآورد	انحراف معیار
ϕ	۰/۳۵	۰/۳۴۹۹۶	۰/۱۱۲۸
	−۰/۵	−۰/۴۹۲۲	۰/۰۶۷۱
σ	۱	۰/۹۸۲	۰/۰۸۳۱
	۱/۲	۱/۱۳۹۷	۰/۰۸۹۷

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

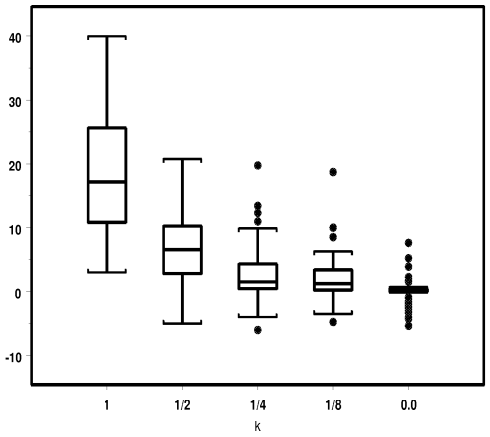
		k				
ϕ	σ	۱	$\frac{۱}{۲}$	$\frac{۱}{۴}$	$\frac{۱}{۸}$	۰
۰/۳۵	۰/۵	۱۰۰	۹۱	۶۳	۵۳	۵۱
−۰/۶	۰/۷	۹۷	۸۷	۵۹	۴۲	۴۹

۶ بحث و نتیجه‌گیری

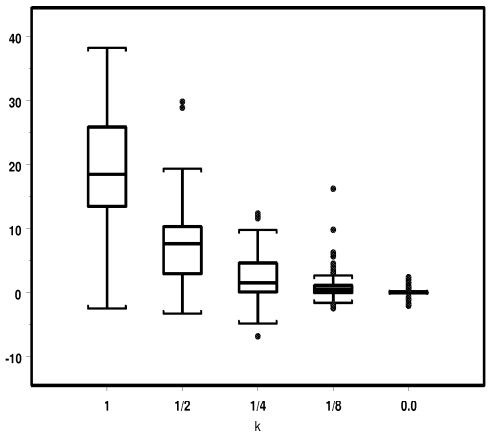
با اینکه درستنمایی مرکب بطور کلی ناکاراست اما به دلیل تولید برآوردگرهای سازگار و سادگی محاسبات مربوط به آن، مورد توجه زیادی قرار گرفته است.

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]



شکل ۱: اختلاف بین آماره‌های انتخاب در فرآیند اول



شکل ۲: اختلاف بین آماره‌های انتخاب در فرآیند دوم

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

استفاده از این نوع درستنمایی در مدلهای پارامترمبنا به منظور برآورد و استنباط، مشکلات محاسبات سنگین و پیچیده تابع درستنمایی در آنها را برطرف کرده است. نتایج شبیه‌سازی نشان دادند که نتایج برآورد و استنباط با استفاده از این نوع درستنمایی در مدلهای پارامترمبنا و همچنین مسأله انتخاب مدل درست با نتایج نظری توافق زیادی دارند.

تقدیر و تشکر

از داوران محترم این مقاله بخاطر نظرات سازنده‌شان در بهبود این اثر، تشکر و قدردانی می‌شود.

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/resarch/html.

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

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

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]

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.

[Downloaded from jss.irstat.ir on 2026-11-08]

[DOR: 20.1001.1.17358183.1386.1.1.15.2]