

تعیین برآوردهای انقباضی پارامترهای کیفیت در مدل‌های خطی بیزی تعمیم‌یافته فوق بالا بعد

ربابه حسین پور صمیم ممقانی

دانشجوی دکتری، گروه آمار، دانشکده آمار، ریاضی و رایانه، دانشگاه علامه طباطبایی، تهران، ایران. hosseinpour@atu.ac.ir

فرزاد اسکندری^۱

(نویسنده مسئول) استاد تمام، گروه آمار، دانشکده آمار، ریاضی و رایانه، دانشگاه علامه طباطبایی، تهران، ایران. askandari@atu.ac.ir

وحید رضایی تبار

دانشیار، گروه آمار، دانشکده آمار، ریاضی و رایانه، دانشگاه علامه طباطبایی، تهران، ایران. RezaeiTabar@atu.ac.ir

چکیده: یکی از مسائل اساسی در تجزیه و تحلیل داده‌های فوق بالا بعد، برازش مدل بهینه و برآورد پارامترهای نامعلوم کیفیت آن به گونه‌ای است که بتواند ساختار داده‌های مورد بررسی را به درستی تفسیر کند. در این مقاله به منظور انتخاب متغیر به شیوه‌های انقباضی بیزی برای مدل‌های خطی تعمیم‌یافته فوق بالا بعد^۱، به مقایسه دو ابرپیشین ناموضعی: گشتاور ضربی و گشتاور وارون ضربی در تعیین مدل بهینه هم‌زمان با برآورد پارامترهای مدل می‌پردازیم. به منظور محاسبه احتمال‌های پسین، از روش تقریب لاپلاس و جهت انتخاب مدل بهینه در فضای متراکم احتمال‌های پسین، از الگوریتم تکراری جستجوی تصادفی تفنگی ساده شده همراه با غربالگری^۲ استفاده شده است. در انتها از طریق مطالعه شبیه‌سازی و تحلیل داده‌ی واقعی، کارایی روش‌های انقباضی بیزی فوق با روش درست‌نمایی توان‌یافته‌ی اسکاد و لاسو مورد ارزیابی قرار گرفته است.

واژگان کلیدی: انتخاب متغیر، فوق بالا بعد، درست‌نمایی توان‌یافته، پارامترهای کیفیت، غربالگری قطعی لاسو، بهینه‌سازی.

۱. مقدمه

توسط پژوهشگران آماری و یادگیری ماشین جهت استنباط و ایجاد برآوردهای سازگار در مدل‌های آماری با طبیعت تنک انجام شده است.

گسترش روزافزون کاربرد مدل‌های خطی تعمیم‌یافته^۳ از یک سو و حجم زیاد ورودی‌های آن از سوی دیگر، موجب ایجاد پیچیدگی‌هایی در مراحل تعیین مدل و برآوردیابی پارامترهای مدل می‌شود. بنا بر این منطقی به نظر می‌رسد تا در حالت بالا بعد، تعداد کمی از متغیرهای ورودی وارد مدل آماری شده و در نتیجه رویکرد رگرسیون توان‌یافته وارد مباحث آماری گردد. در تعیین روش برآورد پارامترهای مدل، بحث نامتناهی شدن برآوردهای روش بیشینه درست‌نمایی و اربیی کم و واریانس زیاد روش کم‌ترین توان‌های دوم، به لزوم داشتن روشی با قابلیت هم‌زمان انتخاب متغیر و برآورد پارامترها تاکید می‌کند. تاکنون استفاده از روش‌های فراوانی‌گرای انتخاب بهترین زیرمجموعه

امروزه رشد روزافزون پژوهش‌های مختلف علمی از یک سو و پیشرفت‌های اخیر در حوزه فناوری و افزایش سرعت ثبت اطلاعات و دستیابی به داده‌های حجیم از سوی دیگر، تجزیه و تحلیل داده‌های بالا بعد را اجتناب‌ناپذیر کرده است. در سال‌های اخیر مدل‌سازی برای داده‌های بالا بعد در حالتی که بعد متغیرها بزرگ‌تر از اندازه نمونه است، مورد توجه مباحث آمار و یادگیری ماشین بوده و دست‌یابی به مدل‌های صرفه‌جو با تفسیرپذیری بالاتر، دقت پیش‌بینی بیش‌تر و پیچیدگی محاسباتی کم‌تر از جمله اهداف اصلی کار با این داده‌ها است. لذا به منظور پردازش، تجزیه و تحلیل سریع داده‌ها، مدل‌سازی و برآورد پارامترهای

تاریخ دریافت: ۱۴۰۱/۰۷/۱۳ تاریخ پذیرش: ۱۴۰۱/۱۱/۱۸

دوره ۱۲/ شماره ۲

صفحات ۱۰۳-۱۲۴

^۱ (Corresponding author): askandari@atu.ac.ir

^۲ Ultra-high dimensional

^۳ Simplified Shotgun Stochastic Search Algorithm with Screening (S5)

^۴ Generalized Linear Model (GLM)

فن و ال وی [۹]، استفاده از روش‌های غربالگری متغیر را پیشنهاد دادند و ثابت کردند که این روش‌ها قادرند با احتمال مجانبی یک، متغیرهای ورودی مهم را در مدل حفظ کنند. فن و همکاران [۱۰] هم به منظور یافتن روش‌های استوار و کارا، روش‌های غربالگری شرطی تکراری در انتخاب مدل را معرفی کردند. فن و سانگ [۱۱] نسخه‌ای جامع‌تر از یادگیری مستقل با رتبه‌بندی برآوردگرهای پیشینه درست‌نمایی حاشیه‌ای^۶ یا خود پیشینه درست‌نمایی حاشیه‌ای^۷ در مدل‌های خطی تعمیم‌یافته را پیشنهاد کرده و نشان دادند در روش غربالگری مستقل قطعی^۸ نرخ انتخاب ناصحیح متغیرها به سمت صفر میل می‌کند و با توجه به ساختار ماتریس کوواریانس ورودی‌ها و پارامترهای واقعی، بعد مدل با روش غربالگری مستقل کاهش می‌یابد.

در مدل‌های خطی با فرض این که $y_n \sim N(X_n \beta, \phi_k)$ برداری n -بعدی، X_n ماتریس طرح $n \times p$ بعدی، ϕ_k پارامتر پراکندگی، n حجم نمونه و p تعداد متغیرهای ورودی یا پیش‌گو باشد، جانسون و راسل [۱۲]، پیشین‌های موضعی یعنی تابع‌های چگالی که در مقادیر صفر پارامتر، مثبت و پیشین‌های ناموضعی یعنی تابع‌های چگالی که هر جا پارامتر مدل برابر مقدار صفرش است، به‌طور یکسان صفر می‌باشند را معرفی کرده و انواع پیشین‌های ناموضعی: گشتاور ضربی $\pi_M(\beta | \phi_k, M_k)$ و گشتاور وارون ضربی $\pi_I(\beta | \phi_k, M_k)$ را به‌صورت زیر ارائه دادند:

$$\pi_M(\beta | \phi_k, M_k) = \prod_{j \in M_k} \frac{\beta_j^2}{\tau \phi_k} N(\beta_j; 0, \tau \phi_k) \quad (1)$$

$$\pi_I(\beta | \phi_k, M_k) = \prod_{j \in M_k} \frac{1}{\sqrt{\pi} \beta_j^2} \exp \left\{ -\frac{\tau \phi_k}{\beta_j^2} \right\} \quad (2)$$

که در آن فضای مدل، M_k پارامتر مقیاس و τ مرتبه توزیع است. در شکل (۱) نمای کلی پیشین‌های موضعی و ناموضعی نشان داده شده‌اند.

مانند روش‌های درست‌نمایی توانانیده، در آمار کاربردی گسترش یافته و در انتخاب متغیرهای موثر و برآورد اثرات آن‌ها در استنباط آماری مربوط به مدل‌های بالا بعد به کار می‌روند. اما از لحاظ محاسباتی، وقت و هزینه زیادی صرف انجام این روش‌ها می‌شود. تحقیقات اخیر نشان داده که در پیش‌بینی مدل با افزایش حجم نمونه اغلب شیوه بیزی، کارا تر از شیوه‌های فراوانی‌گرا است.

تاریخچه‌ی طولانی انتخاب متغیر به شیوه بیزی در کارهای زلنر [۱] و روش‌های تسریع‌کننده زنجیر مارکوف مونت کارلویی، میانگین‌گیری مدل‌های رگرسیونی و انتخاب توزیع پیشین برای انتخاب مدل به شیوه بیزی در کارهای برگر و پری چی [۲] مشاهده می‌شود. در فرایند انتخاب مدل، آن‌ها استفاده از پیشین‌های ناسره برای پارامترها را نامناسب می‌دانستند.

تیبشیرانی [۳] بیان کرد هنگامی که ضرایب رگرسیونی دارای توزیع‌های پیشین مستقل و یکنواخت لاپلاس هستند، برآوردگر لاسو همان برآوردگر پیشینه پسین خواهد بود. بعدها رهیافت لاسوی بیزی با استفاده از پیشین‌های شبه لاپلاس مانند پیشین گامای وارون گاوسی پیشنهاد گردید و به‌منظور حل مسئله تراکم فضای نمونه ناشی از افزایش بعد، الگوریتم‌های پیشرفته مونت کارلویی توسعه یافتند. لیانگ و همکاران [۴] برای مدل‌های رگرسیون خطی، ارتباط بین رهیافت بیزی و رهیافت درست‌نمایی توانانیده را بررسی و نشان دادند با افزایش حجم نمونه، اغلب رگرسیون زیرمجموعه‌ای بیزی، در پیش‌بینی کارا تر است. برای GLM ‌های بالا بعد با شرط تنک‌بودن و توزیع برنولی مدل و پیشین گاوسی برای پارامترها، لیانگ و همکاران [۵] رگرسیون زیرمجموعه‌ای بیزی^۵ را معرفی نمودند که به‌طور تقریبی با افزایش حجم نمونه به ملاک اطلاع بیز تعمیم‌یافته همگرا است. زلنر [۶] هم برای پارامترهای مدل رگرسیونی، پیشین g -را معرفی کرد که به دلیل کارایی بالای محاسباتی آن در ارزیابی درست‌نمایی‌های حاشیه‌ای و سادگی تفسیر مدل مورد استقبال زیادی قرار گرفت. بعدها لیانگ و همکارانش [۷]، خانواده جدیدی از پیشین‌های g -را با نام ابرپیشین‌های g -پیشنهاد دادند و بوو و هلد [۸]، پیشین g -را برای مدل‌های خطی تعمیم‌یافته گسترش داده و دو ابرپیشین، زلنرسی‌او (گامای وارون) و ابرپیشین $\frac{g}{n}$ را معرفی و در این حالت‌ها سازگاری پسین را بررسی کردند.

^۶ Maximum Marginal Likelihood Estimator (MMLE)

^۷ Maximum Likelihood Estimator (MLE)

^۸ Sure Independence Screening (SIS)

^۵ Bayesian Subset Regression (BSR)

سازگاری عامل بیز آن را بررسی کردند. آن‌ها با ترکیب ماتریس اطلاع فیشر با پیشین گشتاور ضربی و پیشین گشتاور وارون ضربی جانسون-راسل و تعیین ابرپیشین روی پارامتر مقیاس، ابر پیشین‌های ناموضعی را برای مدل‌های خطی تعمیم‌یافته غیر بالابعد معرفی کردند. آن‌ها با کمک تقریب لاپلاس، عامل بیز را محاسبه و نشان دادند تحت شرایط معین نظم، عامل بیز به‌دست آمده سازگار است. در ادامه کار وو و همکاران [۱۷]، در این مقاله با به‌کارگیری شیوه‌ی انقباضی بیزی از طریق تحمیل تاوان در هسته توزیع پیشین، به مقایسه کارایی دو پیشین ناموضعی: گشتاور ضربی و گشتاور وارون

ضربی در تعیین مدل بهینه هم‌زمان با برآورد پارامترهای مدل برای مدل‌های خطی تعمیم‌یافته فوق بالابعد می‌پردازیم. بخش ۲ مقاله به مدل‌سازی مسئله می‌پردازد. در بخش ۳ قیده‌های لازم جهت برقراری شرایط پیشنهادی مقاله در انتخاب مدل به‌شیوه انقباضی بیزی با استفاده از ابر پیشین‌های ناموضعی برای مدل‌های لوجستیک فوق بالابعد، بررسی می‌شود. بخش ۴ به روش محاسباتی می‌پردازد. برای بررسی عملکرد روش معرفی شده، مطالعه‌ی شبیه‌سازی و همچنین تحلیل داده‌های واقعی به‌ترتیب در بخش‌های ۵ و ۶ فراهم آمده است. بحث و نتیجه‌گیری بخش پایانی مقاله است.

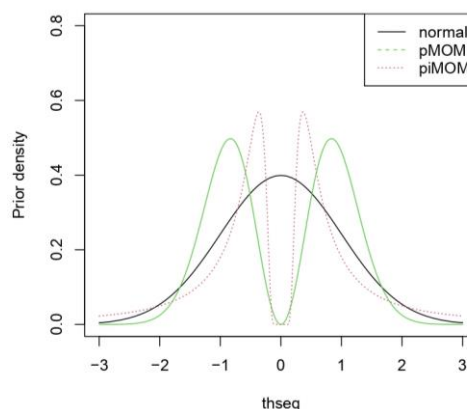
۲. مدل‌سازی مسئله

فرض کنیم $\mathbf{y}_n = (y_1, \dots, y_n)$ متغیر پاسخ و برداری $n \times 1$ بعدی، \mathbf{X}_n ماتریس طرح $n \times p$ بعدی، n حجم نمونه و p تعداد متغیرهای ورودی یا پیش‌گو، $\mathbf{k} \subseteq \{0, 1\}^p$ نماد یک مدل شامل زیرمجموعه‌ای از ستون‌های \mathbf{X}_n ، $|\mathbf{k}|$ تعداد عناصر مجموعه \mathbf{k} ، فضای مدل $\mathbf{M} = \{\mathbf{k}; \mathbf{k} \subseteq \{0, 1\}^p\}$ یعنی \mathbf{M} و $g(\cdot)$ تابع پیوند باشند، در این صورت مدل به‌صورت زیر خواهد بود:

$$g(E(y_i)) = \beta_0 + \mathbf{X}_{ki}^T \beta_k \quad (3)$$

که در آن β_0 مقدار ضریب ثابت است که با استاندارد کردن داده‌ها در مدل تاثیری نخواهد داشت، \mathbf{X}_k ماتریس طرح مدل \mathbf{k} ، \mathbf{X}_{ki} سطر i ام ماتریس \mathbf{X}_k و β_k بردار ضرایب رگرسیونی غیر صفر مدل خطی تعمیم‌یافته مورد نظر هستند. بنا بر این مدل \mathbf{k} دارای p بعد است و ما به دنبال انتخاب بهترین زیرمجموعه از ورودی‌ها یا پیش‌گوها برای پیش‌بینی متغیر پاسخی از خانواده توزیع‌های نمایی هستیم و در این مقاله متغیر پاسخ را از نوع دودویی در نظر گرفته‌ایم و در نتیجه مدل رگرسیونی موردنظر،

جلد ۱۲- شماره ۲- تابستان ۱۴۰۱



شکل ۱: نمودار مقایسه پیشین موضعی نرمال با پیشین‌های ناموضعی گشتاور ضربی و گشتاور وارون ضربی به‌زای پارامتر مقیاسین (۳ و -۳)

جانسون و راسل [۱۲]، نشان دادند روش‌های انتخاب مدل مبتنی بر پیشین‌های ناموضعی در مقایسه با روش‌های فراوانی‌گرایی درست‌نمایی تاوانیده در تعیین مدل صحیح با میزان خطای پیش‌بینی کم‌تر، توانا تر هستند. روش‌های انتخاب مدل مبتنی بر پیشین ناموضعی، به‌طور کارا مدل‌های رگرسیونی شامل متغیرهای کمکی غیرلازم را حذف می‌کنند که این ویژگی برای سازگاری احتمالات مدل پسین لازم است. راسل و تسکا [۱۳]، با به‌کارگیری پیشین ناموضعی و نمونه‌گیری از توزیع پسین به شیوه گیبس با استفاده از میانگین‌گیری بیزی، برآورد پارامترهای مدل‌های خطی بالابعد و نرخ‌های همگرایی عامل بیز در مدل‌های خطی بالابعد را محاسبه کردند.

شین و همکاران [۱۴]، رفتار پیشین‌های ناموضعی در انتخاب متغیر در مدل‌های خطی بالابعد را بررسی و برای اولین بار الگوریتم جستجوی تصادفی تفنگی ساده شده همراه با غربالگری، S_5 را به‌منظور کشف فضای متراکم مدل پیشنهاد دادند. نیکونژاد و همکاران [۱۵]، انتخاب متغیر به شیوه بیزی با کمک پیشین‌های ناموضعی در حالت بالابعد در مدل‌های لوجستیک را بررسی کردند و به‌منظور نمونه‌گیری از توزیع پسین روی فضای مدل، از روش زنجیر مارکوف مونت کارلویی و الگوریتم متروپلیس هستینگس استفاده کردند. نیکونژاد و همکاران [۱۶]، انتخاب متغیر به شیوه بیزی با کمک پیشین‌های ناموضعی در حالت بالابعد در مدل‌های بقا را مطالعه کرده و از روش جستجوی تصادفی به‌منظور کشف فضای مدل و برای برآورد از میانگین‌گیری بیزی استفاده کردند.

وو و همکاران [۱۷]، روش انتخاب متغیر بیزی تحت پیشین‌های g را در مدل‌های خطی تعمیم‌یافته مطالعه و

نشریه مهندسی و مدیریت کیفیت

$$\pi(\mathbf{k}) \propto I(|\mathbf{k}| \leq m_n) \quad (7)$$

که در آن $m_n = \left(\frac{n}{\log p}\right)^\alpha$ برای $0 < \alpha < 1$ ، m_n عدد صحیح مثبتی است که اندازه بزرگترین مدل را کراندار می کند یعنی حداکثر تعداد متغیرهای داخل مدل m_n است. از طرفی ابرپارامتر مقیاس τ در توزیع پیشین ناموضعی، پراکندگی پیشین را در اطراف صفر نشان می دهد و در واقع اندازه ضرایب رگرسیونی که به صفر منقبض می شوند را نشان می دهد لذا برخلاف کارهای جانسون و راسل [۱۸]، جانسون و راسل [۱۲]، نیکونژاد [۱۵]، نیکونژاد [۱۶] و شین و همکاران [۱۴] که همگی برای τ در کلیه مراحل محاسباتی یک مقدار ثابت در نظر گرفته اند، در این مقاله برای آن یک مقدار متغیر $0 < \tau \leq \log p$ در نظر می گیریم که دارای توزیع گاما یا گامای وارون است.

مقاله روبرت و همکاران [۱۹]، شرح می دهد که از توزیع پسین می توان برای توصیف ویژگی های پارامترهای مدل استفاده کرد. با توجه به روش بیز سلسله مراتبی و وجود تابع درست نمایی $L(\boldsymbol{\beta}|\mathbf{y})$ ، ابرپیشین های ناموضعی گشتاور ضربی و وارون ضربی و توزیع پیشین یکنواخت فضای مدل، احتمال پسین عبارت است از:

$$\begin{aligned} \pi(\boldsymbol{\beta}_k|\mathbf{y}_n) &= \frac{L(\boldsymbol{\beta}_k)\pi(\boldsymbol{\beta}_k|k)\pi(k)}{\sum_{k \in M} L(\boldsymbol{\beta}_k)\pi(\boldsymbol{\beta}_k|k)\pi(k)} \\ &= \frac{m_k(\mathbf{y}_n)\pi(k)}{\sum_{j \in M} m_j(\mathbf{y}_n)\pi(j)} \propto m_k(\mathbf{y}_n)\pi(k) \end{aligned} \quad (8)$$

بنا بر این پسین تحت ابرپیشین های ناموضعی گشتاور ضربی برای پارامتر مدل و توزیع پیشین برای پارامتر مقیاس $\tau \sim \Gamma(a, b)$ عبارت است از:

$$\begin{aligned} \pi_M(\boldsymbol{\beta}_k|\mathbf{y}_n) &\propto \frac{\exp\left\{\sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta}\right\}}{\prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})]} \\ &\quad \times (\tau)^{-r|\mathbf{k}|/2} |I_{k0}|^{r/2} \\ &\quad \times N(\boldsymbol{\beta}_k; 0, (\tau)(I_{k0})^{-1}) \\ &\quad \times \prod_{i=1}^{|\mathbf{k}|} |\beta_{ki}|^{(r+1)} \times I(|\mathbf{k}| \leq m_n) \end{aligned} \quad (9)$$

و پسین تحت ابرپیشین های ناموضعی گشتاور وارون ضربی برای

مدل لوجستیک است. فرض بر آن است که مدل واقعی $\mathbf{t} \subseteq \{1, 2, \dots, p\}$ و برداری از مقادیر ثابت و کوچکترین مدل موجود در فضای مدل است که داده های حاضر از طریق توزیع این مدل واقعی تولید شده است. فرض کنیم مدل \mathbf{k} و $g(\boldsymbol{\mu}) = \log\left(\frac{\boldsymbol{\mu}}{1-\boldsymbol{\mu}}\right)$ تابع پیوند لوجیت باشد. از طرفی در GLM ها، توزیع \mathbf{y} به شرط \mathbf{x} ترکیبی خطی از $\mathbf{x}^T \boldsymbol{\beta}$ و به شکل $\mathbf{y}_n = f(\mathbf{y}|\mathbf{x}^T \boldsymbol{\beta})$ است، لذا تابع درست نمایی $\mathbf{y}_n = (y_1, \dots, y_n)$ به صورت زیر است:

$$\begin{aligned} (\boldsymbol{\beta}|\mathbf{y}) &= \prod_{i=1}^n \left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})}\right)^{y_i} \left(\frac{1}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})}\right)^{1-y_i} \\ &= \frac{\exp\left\{\sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta}\right\}}{\prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})]} \end{aligned} \quad (4)$$

وو و همکاران [۱۷]، برای مدل های GLM در حالت غیر بالابعد ($p < n$)، دو ابرپیشین ناموضعی گشتاور ضربی^۱ و گشتاور وارون ضربی^۱ را از ترکیب ماتریس اطلاع فیشر $\text{diag}(I_{k0}) = I_{k0}$ پیشین های جانسون و راسل و در نظر گرفتن توزیع پیشین گامای وارون برای پارامتر مقیاس τ درگشتاور ضربی و گاما برای پارامتر مقیاس τ درگشتاور وارون ضربی، به صورت زیر معرفی کردند:

$$\begin{aligned} \pi_M(\boldsymbol{\beta}_k|k) &= (\tau)^{-\frac{r|\mathbf{k}|}{2}} |I_{k0}|^{\frac{r}{2}} \times N(\boldsymbol{\beta}_k; 0, (\tau)(I_{k0})^{-1}) \\ &\quad \times \prod_{i=1}^{|\mathbf{k}|} |\beta_{ki}|^{(r+1)} \end{aligned} \quad (5)$$

$$\begin{aligned} \pi_I(\boldsymbol{\beta}_k|k) &= \frac{(\tau)^{r|\mathbf{k}|/2} |I_{k0}|^{-r/2}}{(\Gamma(\frac{r}{2}))^{|\mathbf{k}|}} \times \exp\{-\tau(\boldsymbol{\beta}_k^T I_{k0} \boldsymbol{\beta}_k)^{-1}\} \\ &\quad \times \prod_{i=1}^{|\mathbf{k}|} |\beta_{ki}|^{-(r+1)} \end{aligned} \quad (6)$$

تفاوت کار این مقاله با کار وو و همکاران [۱۷]، آن است که ما در وضعیت GLM های فوق بالابعد مسئله را بررسی می کنیم. بنا بر این برای آن که ماتریس گرام تحت مدل k یعنی $\left(\frac{\mathbf{x}_k^T \mathbf{x}_k}{n}\right)$ نامنفرد شود، نیازمندیم تا اندازه بزرگترین مدل را کراندار کنیم. به این منظور توزیع پیشین یکنواخت را برای فضای مدل در نظر می گیریم:

تقریب انتگرال‌های موجود در این دو کمیت به دلیل پیچیدگی توزیع پسین ناشی از انتخاب نوع پیشین و ساختار مدل‌های GLM از روش‌های عددی محاسبه تقریبی مونت کارلویی یا تقریب لاپلاس استفاده می‌شود. در این مقاله چون توزیع درست‌نمایی حاشیه‌ای $m_k(y_n)$ فرم بسته‌ای ندارد لذا لازم است تا از تقریب عددی برای محاسبه مقدار عددی آن استفاده نمود. تاینری و کدین [۲۰]، تقریب لاپلاس را برای محاسبه عددی مقدار توزیع درست‌نمایی حاشیه‌ای بادقت $O(\frac{1}{n})$ به صورت زیر معرفی کردند:

$$\hat{m}(y|k) = (2\pi)^{\frac{|k|}{2}} |\hat{\mathbf{H}}_k|^{-\frac{1}{2}} \exp[\ell(\tilde{\boldsymbol{\beta}}_k|y)] \pi(\tilde{\boldsymbol{\beta}}_k|k) \quad (15)$$

که در آن $\ell(\tilde{\boldsymbol{\beta}}_k)$ لگاریتم تابع درست‌نمایی، $\tilde{\boldsymbol{\beta}}_k$ مدهای پسین تحت مدل \mathbf{k} و $\hat{\mathbf{H}}_k$ ماتریس هیشین لگاریتم پسین در نقاط $\tilde{\boldsymbol{\beta}}_k$ می‌باشد. بنا بر این برای هر مدل موجود در فضای مدل‌ها، از تقریب لاپلاس برای محاسبه احتمالات پسین تقریبی استفاده می‌شود. از طرفی برای محاسبه تقریب لاپلاس، نیاز به محاسبه مدهای پسین و واریانس پارامترهای مدل یا همان ماتریس هیشین است. در این مقاله از الگوریتم بهینه‌سازی ($L - BFGS$) که برای مسائل دارای بیش از ۱۰۰۰ متغیر مفید هستند، استفاده می‌کنیم. در این روش برای یافتن مدهای پسین و مقدار عددی ماتریس هیشین، نخست لازم است تا برای پارامترهای مدل، مقادیر اولیه که برآوردهای بیشینه درست‌نمایی هستند و همچنین تابع امتیاز توزیع پسین مربوط به رگرسیون لوجستیک، محاسبه گردد. به‌طور کلی تابع امتیاز توزیع پسین شامل دو قسمت: مشتق مرتبه اول لگاریتم تابع درست‌نمایی و توزیع پیشین نسبت به پارامترهای $\boldsymbol{\beta}_k$ می‌باشد.

در مدل‌های GLM ، با فرض آن که \mathbf{W} ماتریس قطری وزن‌های کاری و \mathbf{M} ماتریس قطری مشتقات تابع پیوند باشد آنگاه مشتق مرتبه اول لگاریتم تابع درست‌نمایی نسبت به پارامتر $\boldsymbol{\beta}_k$ یا همان تابع امتیاز مدل GLM به صورت زیر است:

$$\mathbf{U} = \frac{1}{\phi} \mathbf{X}^T \mathbf{W} \mathbf{M} (\mathbf{Y} - \boldsymbol{\mu}) \quad (16)$$

لذا وو و همکاران [۱۷]، تابع امتیاز برای توزیع‌های پسین تحت برپیشین‌های ناموضعی گشتاور ضربی برای پارامتر مدل و توزیع

پارامتر مدل و توزیع پیشین برای پارامتر مقیاس $\tau \sim \Gamma(a, b)$ برابر است با:

$$\begin{aligned} \pi_I(\boldsymbol{\beta}_k|y_n) & \propto \frac{\exp\left\{\sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta}\right\}}{\prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})]} \\ & \times \frac{(\tau)^{r|k|/2} |I_{k0}|^{-r/2}}{(\Gamma(\frac{\tau}{2}))^{|k|}} \\ & \times \exp\{-\tau(\boldsymbol{\beta}_k^T I_{k0} \boldsymbol{\beta}_k)^{-1}\} \\ & \times \prod_{i=1}^{|k|} |\beta_{ki}|^{-(r+1)} \times I(|\mathbf{k}| \leq m_n) \end{aligned} \quad (10)$$

روبرت و همکاران [۱۹] بیان کردند که باتوجه به نوع توزیع نمونه، توزیع پیشین و تابع زبان، می‌توان بهترین برآوردگر را انتخاب کرد. یکی از انواع تابع‌های زبان، تابع زبان ۱-۰ است که باتوجه به کمینه کردن امید ریاضی تابع زبان تحت توزیع پسین، مدهای توزیع پسین^{۱۱} به‌عنوان برآوردهای بهینه پارامترها تعیین می‌شوند.

$$\hat{\boldsymbol{\beta}}_k = \operatorname{argmax}_{\boldsymbol{\beta}} \pi(\boldsymbol{\beta}_k|y_n) \quad (11)$$

راسل و تسکا [۱۳]، برای برآورد پارامترها از روش میانگین‌گیری بیزی به شیوه زیر استفاده کردند:

$$E(\boldsymbol{\beta}|y_n) = \sum_{k=1}^K E(\boldsymbol{\beta}|k, y_n) \pi(k|y_n) \quad (12)$$

که در آن:

$$\pi(k|y_n) \propto m_k(y_n) \pi(k) \quad (13)$$

و درست‌نمایی حاشیه‌ای تحت مدل عبارت است از:

$$m_k(y_n) = \int f(y_n|\boldsymbol{\beta}_k) \pi(\boldsymbol{\beta}_k|k) \pi(k) d\boldsymbol{\beta}_k \quad (14)$$

روش میانگین‌گیری بیزی با اختصاص $\pi(k|y_n)$ کوچک از طریق به‌کارگیری پیشین‌های ناموضعی $(\pi(\boldsymbol{\beta}|\phi_k, k))$ ، وزن‌های کم‌تری به مدل‌های پیچیده اختصاص داده و برآوردها را منقبض می‌کند.

دو کمیت ارزیابی مدل یا فرضیه‌ها در مباحث استنباط به شیوه بیزی، احتمال پسین و عامل بیز هستند و معمولاً برای محاسبه

^{۱۱} Maximum A Posteriori (MAP)

$$-H(\beta_k) = \left(\frac{\partial^2 \ell(\beta_k)}{\partial \beta_k \partial \beta_k^T} \right) = X_k^T \Sigma_k X_k \quad (22)$$

که در آن:

$$\Sigma_k = \text{diag}(\sigma_1^2(\beta_k), \dots, \sigma_n^2(\beta_k)) \quad (23)$$

$$\sigma_i^2(\beta_k) = \frac{\mu_i(\beta_k)}{1 - \mu_i(\beta_k)} = \frac{\exp(x_{ik}^T \beta_k)}{1 + \exp(x_{ik}^T \beta_k)} \quad (24)$$

$$= \frac{\exp(x_{ik}^T \beta_k)}{1 - \frac{\exp(x_{ik}^T \beta_k)}{1 + \exp(x_{ik}^T \beta_k)}}$$

و ماتریس اطلاع فیشر استاندارد شده با حجم نمونه n ، که همان امید ریاضی منفی هیشین می باشد، برابر است با:

$$E(-H(\beta_k)) = I(\beta_{0,k}) = \frac{1}{n} X_k^T \Sigma_k X \quad (25)$$

برای برقراری ویژگی های نظری مسئله، شرایط نظم زیر را در نظر می گیریم:

شرط ۱) برای هر $0 < \alpha < 1$ ، به منظور کارا بودن روش پیشنهادی مقاله در حالت فوق بالا بعد، یعنی هنگامی که تعداد پیشگوها نسبت به حجم نمونه رابطه نمایی دارد، شرط زیر را داریم:

$$\log p = O(n^\alpha) \quad (26)$$

از طرفی به منظور تعیین مدل های داخل فضای مدل، این فضا را مقید به توزیع پیشین یکنواخت و حداکثر تعداد متغیرهای داخل مدل را به عدد صحیح مثبت m_n محدود کرده و داریم:

$$m_n = \left(\frac{n}{\log p} \right)^\alpha \quad (27)$$

از طرفی ابرپارامتر مقیاس τ در توزیع پیشین ناموضعی، نشان دهنده پراکندگی پیشین در اطراف صفر و در واقع اندازه ضرایب رگرسیونی که به صفر منقبض می شوند را نشان می دهد لذا در این مقاله برای آن محدوده زیر را در نظر می گیریم:

$$0 < \tau \leq \log p \quad (28)$$

شرط ۲) برای راحتی محاسبات به ازای $k = 1, \dots, m_n$ متغیرهای پیش گو را استاندارد می کنیم:

$$\max_{i,k} |x_{ik}| \leq 1 \quad (29)$$

با فرض آن که پارامترهای مدل واقعی β_t باشد، به منظور ایجاد تنگی در مدل واقعی شرط زیر را داریم:

پیشین برای پارامتر مقیاس $\tau \sim \Gamma(a, b)$ را به صورت:

$$S_{hpmGLM} = \frac{1}{\phi} X^T W M (Y - \mu) - \frac{(r|k| + \frac{|k|}{2} + \frac{a}{2})(2I_{k0} \beta)}{(b + \beta^T I_{k0} \beta)} + 2r1_{|k|} \otimes \hat{\beta}^{-1} \quad (17)$$

و تابع امتیاز برای توزیع های پسین تحت ابرپیشین های ناموضعی گشتاور وارون ضریبی برای پارامتر مدل و توزیع پیشین برای پارامتر مقیاس $\tau \sim \Gamma(a, b)$ را به صورت زیر محاسبه کردند:

$$S_{hpmGLM} = \frac{1}{\phi} X^T W M (Y - \mu) + \frac{(r|k| + 2a)[\text{diag}(\hat{\beta}^{-3} \beta^T) I_{k0}^{-1}] \beta^{-1}}{(b + \text{tr}((\text{diag}(\beta \beta^T) I_{k0})^{-1}))} - (r+1)1_{|k|} \otimes \hat{\beta}^{-1} \quad (18)$$

که در آن \otimes ضرب هادامارد و $|k|$ برداری از یکها به طول $|k|$ می باشد. بنا بر این مدل بهینه مدلی است که احتمال پسین یعنی مدل:

$$\pi(k|y_n) = \frac{\pi(k)m_k(y_n)}{\sum_{j \in M} \pi(j)m_j(y_n)} \quad (19)$$

و بیشترین مقدار باشد به عبارت دیگر:

$$\hat{k} = \text{argmax}_k \pi(k|y_n) \quad (20)$$

۳. قیدهای لازم جهت برقراری شرایط انتخاب مدل به شیوه انقباضی بیزی با استفاده از ابرپیشین های ناموضعی برای مدل های لوجستیک فوق بالا بعد

فرض کنید $t \subseteq \{1, 2, \dots, p\}$ مدل واقعی باشد که برداری از مقادیر ثابت است. لگاریتم تابع درست نمایی یک مدل لوجستیک عبارت است از:

$$\ell(\beta_k|y) = \log \left(\prod_{i=1}^n \left(\frac{\exp(x_{ik}^T \beta_k)}{1 + \exp(x_{ik}^T \beta_k)} \right)^{y_i} \times \left(\frac{1}{1 + \exp(x_{ik}^T \beta_k)} \right)^{1-y_i} \right) \quad (21)$$

منفی هیشین $\ell(\beta|y)$ که همان مشتق دوم لگاریتم تابع درست نمایی می باشد، برابر است با:

زنجر مارکوف مونت کارلو^{۱۲} کارایی لازم را ندارد. در این مقاله جهت محاسبه احتمال‌های پسین، الگوریتم ۱ اجرا شده است:

الگوریتم ۱. محاسبه احتمال‌های پسین مدل k

فرض کنیم M فضای مدل باشد، سپس گام‌های زیرین تا رسیدن k به جواب نهایی تکرار می‌شوند:

(۱) یک مدل اولیه $k \in M$ را انتخاب کنید. نگاه با توجه به تابع درست‌نمایی توزیع نمونه‌ها و پیشین پارامترها مقادیر لگاریتم تابع درست‌نمایی $\ell(\beta_k)$ ، پیشین پارامتر $\pi(\beta_k)$ و توزیع اندازه مدل $\pi(k)$ موجودند.

(۲) احتمال پسین برای β_k را به صورت

$$\pi(\beta_k | y_n) \propto \ell(\beta_k) \pi(\beta_k | k) \pi(k)$$

محاسبه کنید.

(۳) برای مدل k ، به منظور بهینه کردن توزیع پسین، برآوردهای بیشینه درست‌نمایی را به عنوان مقادیر اولیه پارامترهای مدل قرار دهید.

(۴) مدهای پسین مدل

$$\tilde{\beta}_k = \operatorname{argmax}_{\beta} \pi(\beta_k | y_n)$$

(۵) را با کمک الگوریتم بهینه سازی (L-BFGS) بیابید.

(۶) ماتریس هیشین عددی لگاریتم پسین را بیابید.

(۷) توزیع حاشیه‌ای درست‌نمایی را با کمک تقریب لاپلاس به صورت زیر محاسبه کنید:

$$\tilde{m}_k(y_n) = (2\pi)^{\frac{|k|}{2}} |\tilde{H}_k|^{-\frac{1}{2}} \exp[\ell(\tilde{\beta}_k)] \pi(\tilde{\beta}_k | k)$$

(۸) احتمال پسین تقریبی را محاسبه کنید:

$$\tilde{\pi}(k | y_n) \propto \frac{\pi(k) \tilde{m}_k(y_n)}{\sum_{j \in M} \pi(j) \tilde{m}_j(y_n)}$$

(۹) برای هر $k \in M$ ، مقدار تقریب پسین $\tilde{\pi}(k | y_n)$ را به عنوان ملاک انتخاب متغیر به شیوه بیز در نظر بگیرید.

$$\lim_{p \rightarrow \infty} \sum_{k=1}^p |\beta_t| < \infty \quad (30)$$

از طرفی حداکثر اندازه پارامتر مدل واقعی با شرط زیر تعیین می‌شود:

$$\|\beta_t\|_2^2 = O((\log p)^\alpha) \quad (31)$$

شرط (۳) برای هر $\lambda > 0$ و مقادیر ویژه مرتب‌شده ماتریس اطلاع فیشر به صورت: $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{\max}$ و ماتریس گرام مربوط به مدل k به صورت $\left(\frac{X_k^T X_k}{n}\right)$ ، شرط زیر را داریم:

$$0 < \lambda \leq \min_{k: |k| \leq m_n} \lambda_1(I_{(0,k)}) \leq \max_{k: |k| \leq m_n} \lambda_{\max} \left(\frac{X_k^T X_k}{n} \right) \leq (\log p)^\alpha \quad (32)$$

ناریستی و همکاران [۲۱]، شرط کرانداری مقادیر ویژه را به منظور تضمین انحنای تابع هدف در حالت بالابعد لازم و ضروری دانستند.

شرط (۴) با فرض آن که مدل واقعی t باشد، برای هر مقدار ثابت C داریم:

$$\min_{k \in t} \beta_k^2 \geq C \max \left\{ \frac{t \max_{k: |k| \leq t} \lambda_{\max} \left(\frac{X_k^T X_k}{n} \right) \log p}{n}, \frac{1}{\log p} \right\} \quad (33)$$

این شرط برای پارامترهای غیر صفر مدل، کران پایین در نظر گرفته که برای حفظ پارامترهای غیر صفر در مدل ضروری است. با برقراری شرط‌های تعیین شده فوق، بحث تئوری مسئله تضمین شده خواهد بود و می‌توان محاسبات عددی را با وجود قیدهای تعیین شده انجام داد.

۴. روش محاسباتی

چنانچه در بخش ۲ بیان شد، به جز در حالت‌های خاص ساده، معادلات درست‌نمایی مدل‌های GLM شکل‌های بسته‌ای نداشته و از طرفی پیشین‌های ناموضعی هم پیچیدگی محاسبه پسین مدل‌ها را افزایش می‌دهد لذا نیاز به الگوریتم‌های خاص عددی جهت محاسبه احتمالات پسین و انتخاب مدل بهینه است. در وضعیت فوق بالابعد نمونه‌گیری از پسین با کمک الگوریتم‌های

^{۱۲} Markov Chain Monte Carlo (MCMC)

- ۳) به منظور افزایش تعداد مدل‌های برخوردی، ITER را به عنوان تعداد تکرارها در هر آستانه در نظر بگیرید.
- ۴) یک مدل اولیه $k^{(1:l)}$ را انتخاب کنید و بر اساس این مدل اولیه، مجموعه متغیرهای بعد از غربالگری را $S_k^{(1:l)}$ بنامید.
- ۵) برای تکرار الگوریتم S5، یک عدد اولیه C_0 انتخاب کنید.
- ۶) برای $l = 1, \dots, S$
- برای $i = 1, \dots, ITER - 1$
- ✓ برای همه $k \in \{\Gamma_{scr}^+, \Gamma^-\}$ مقادیر $\pi(k | y)$ را بیابید.
- ✓ از Γ_{scr}^+ و Γ^- نمونه‌های k^+ و k^- را با احتمالات متناسب با $\frac{1}{t^i} \pi(k | y)$ انتخاب کنید.
- ✓ نمونه $k^{(i+1,l)}$ را از $\{k^+, k^-\}$ با احتمال متناسب با $\left\{ \pi(k^+ | y)^{\frac{1}{t^i}}, \pi(k^- | y)^{\frac{1}{t^i}} \right\}$ انتخاب کنید.
- ✓ مجموعه متغیرهای در نظر گرفته شده $S_k^{(i+1:l)}$ را به‌هنگام کنید به‌قسمی که اجتماع همه متغیرها $k^{(i+1,l)}$ و $d = 2[\log p]$ متغیر با بیشترین درست‌نمایی حاشیه‌ای باشند.
- پایان حلقه تکرار.

۴. مطالعه شبیه‌سازی

در این بخش، با انجام مطالعه شبیه‌سازی، عملکرد روش‌های ارائه شده را بررسی کرده و با تحلیل داده‌های واقعی، کارآمدی روش‌های پیشنهادی را نشان می‌دهیم. فرایند برآورد پارامترهای مدل و انتخاب متغیر به‌صورت عددی ارزیابی می‌شود. ابتدا مجموعه‌ای از داده‌های شبیه‌سازی شده توسط روش‌های پیشنهادی مقاله یعنی شیوه انقباضی بیزی در دو حالت استفاده از ابرپیشین‌های ناموضعی گشتاور ضربی و گشتاور وارون ضربی در مدل‌های خطی تعمیم‌یافته فوق بالا بعد مورد تحلیل قرار می‌گیرد. سپس به‌منظور بررسی عملکرد روش‌ها در مسئله هم‌زمانی برآورد و انتخاب متغیر، نتایج حاصل از تحلیل داده‌ها با روش‌های درست‌نمایی تاوانیده‌ی اسکاد و لاسو مقایسه می‌شود.

در وضعیت فوق بالا بعد برای کشف مدل در فضای متراکم، از الگوریتم جستجوی تصادفی تفنگی ساده شده همراه با غربالگری (S5) برای GLM ها استفاده می‌کنیم. روش S5 را اولین بار شین و همکاران [۱۴] برای کشف مدل در فضای متراکم برای مدل‌های خطی بالا بعد به‌کار بردند که یک نوع روش جستجوی تصادفی است که در هر گام، متغیرهای ورودی یا پیشگو را غربال می‌کند. این غربالگری در مدل‌های خطی بر اساس همبستگی بین متغیرهای خارج از مدل کنونی با مانده‌های مدل رگرسیونی کنونی است در حالی که برای GLM ها در این مقاله از ایده فن و سانگ [۱۱]، یعنی رتبه‌بندی بیشینه درست‌نمایی حاشیه‌ای استفاده می‌کنیم.

فرض کنیم k مدل کنونی و k^c ستون‌هایی از ماتریس طرح هستند که در مدل کنونی وجود ندارند. در هر مرحله به تعداد $d = 2[\log p]$ متغیر پیشگو با بیشترین درست‌نمایی، کاندید وارد شدن به مدل کنونی می‌شوند که با نماد Γ_{scr}^+ و با خارج نمودن یک متغیر از مدل کنونی، مدل را با Γ^- نمایش می‌دهیم. می‌توان از مدل کنونی متناسب با احتمال‌های حاشیه‌ای به هر یک از همسایگی‌های Γ_{scr}^+ یا Γ^- حرکت نمود. برای اجتناب از درگیری با بیشینه موضعی، احتمال‌های مدلی که در S5 به‌کار می‌روند به توان $\left(\frac{1}{t^i}\right)$ می‌رسند که در آن l امین آستانه در برنامه حرارت را با t^i نشان می‌دهیم که آستانه‌ها کاهش می‌دهند. به منظور افزایش تعداد مدل‌های برخوردی، در هر آستانه تعداد معینی از تکرارها انجام می‌شود بنا بر این در انتهای کار، مدلی که بالاترین احتمال پسین^{۱۳} را در بین کلیه مدل‌های برخورد شده دارد، به عنوان مدل بهینه انتخاب می‌شود. نحوه اجرای الگوریتم S5 برای GLM ها در الگوریتم ۲ شرح داده شده است.

الگوریتم ۲. جستجوی تصادفی تفنگی ساده شده همراه با غربالگری (S5) برای GLM ها

- (۱) به منظور غربالگری در بین متغیرها، مقدار اولیه‌ی S را انتخاب کنید.
- (۲) برای اجتناب از برخورد با بیشینه موضعی، برنامه‌ای آستانه‌ای را در نظر بگیرید.

^{۱۳} Highest Posterior Probability Model (HPPM)

فرض کنیم:

X و $p = \{1000, 2000, 3000, 4000\}$ و $n = \{200, 400\}$ ماتریس طرح از توزیع نرمال با میانگین صفر و ماتریس واریانس کوواریانس Σ با همبستگی اتورگرسیو $0.5^{|i-j|} = \Sigma_{ij}$ برای هر $1 \leq i \leq j \leq p$ و متغیر پاسخ برداری از نمونه‌های برنولی با احتمال موفقیت π_i لوجیت:

$$\pi_i = \frac{e^{X_{ik}^T \beta_k}}{1 + e^{X_{ik}^T \beta_k}} \quad (34)$$

از مدل واقعی و ثابت $t = (1, 2, 3, 4, 5, 6)$ و ضرایب $\beta_0^t = (-4, -3, -2, 4, 2, 2)$ باشد. روش انتخاب متغیر در تمام الگوریتم ۵۰ بار تکرار می‌شود. با در نظر گرفتن میانگین توان دوم خطای پیش‌بینی ($MSPE$) و میانگین توان دوم خطای احتمال موفقیت (MSE) به صورت زیر:

$$MSPE = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (\hat{y}_i - y_{test,i})^2 \quad (35)$$

$$MSE(\hat{\pi}) = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (\hat{\pi}_i - \pi_i)^2 \quad (36)$$

نتایج بدست آمده برای این مطالعه شبیه‌سازی در جدول (۱) و (۲) خلاصه شده است.

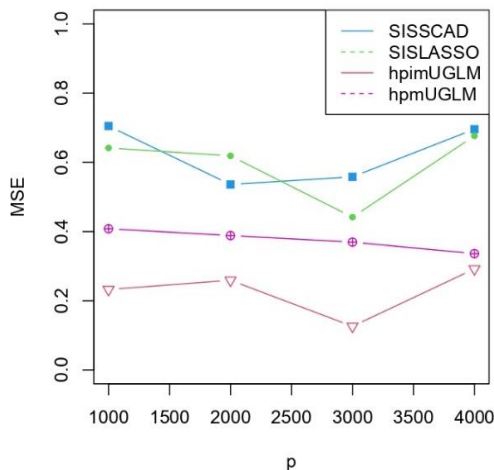
جدول ۱- میانگین توان دوم خطای پیش‌بینی برای ارزیابی عملکرد انتخاب متغیر در چهار روش hpimUGLM، hpmUGL، SIS-LASSO و SCAD.

For n=200	hpimUGLM	hpmUGLM	SISSCAD	SISLASSO
p=1000	0.214360	0.5266667	0.103400	0.110218
p=2000	0.223594	0.5266667	0.126842	0.083746
p=3000	0.161210	0.5266667	0.111738	0.126522
p=4000	0.230350	0.2044667	0.102818	0.098780

جدول ۲- میانگین توان دوم خطای احتمال موفقیت برای ارزیابی عملکرد انتخاب متغیر در چهار روش hpimUGLM، hpmUGL، SIS-LASSO و SIS-SCAD.

For n=200	hpimUGLM	hpmUGLM	SISSCAD	SISLASSO
p=1000	0.232600	0.408200	0.7049667	0.6416223
p=2000	0.259300	0.2888233	0.5361667	0.6188667
p=3000	0.126200	0.3695233	0.558200	0.441800
p=4000	0.2921667	0.3362667	0.695900	0.6765223

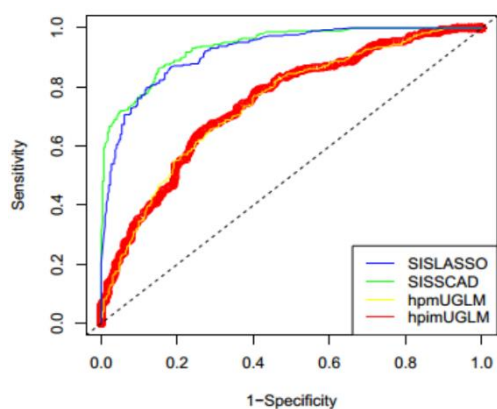
چنان‌که از نتایج درج شده در جدول ۱ و ۲ دیده می‌شود، میانگین توان دوم خطای پیش‌بینی در انتخاب متغیر به شیوه درست‌نمایی تاوانیده لاسو و اسکاد نسبت به روش انقباضی بی‌زی کم‌تر است در حالی‌که میانگین توان دوم خطای احتمال موفقیت خلاف نتیجه بالا است یعنی در دو حالت استفاده از ابر پیشین‌های ناموضعی گشتاور ضربی و گشتاور وارون ضربی، احتمال موفقیت مدل لوجیت خطای کم‌تری را نسبت به مدل‌های درست‌نمایی تاوانیده لاسو و اسکاد دارد و از طرفی استفاده از ابر پیشین‌های ناموضعی گشتاور وارون ضربی، این خطا را کمینه می‌کند که در ارزیابی عملکرد روش‌های به کار رفته، نشانه‌ی مناسب‌تر بودن شیوه انقباضی بی‌زی در دو حالت استفاده از ابر پیشین‌های ناموضعی گشتاور ضربی و گشتاور وارون ضربی در برآورد مقدار حقیقی متغیر پاسخ است.



شکل ۲. نمودار مقایسه‌ی میانگین توان دوم خطای احتمال موفقیت در چهار روش hpimUGLM، hpmUGL، SIS-LASSO و SCAD برای اندازه‌های مختلف p.

در شکل ۲، میزان میانگین توان دوم خطای احتمال موفقیت در چهار روش hpimUGLM، hpmUGL، SIS-SCAD و SIS-LASSO برای اندازه‌های مختلف p مقایسه شده است. ملاحظه می‌شود که انتخاب متغیر به شیوه انقباضی بی‌زی در دو حالت استفاده از ابر پیشین‌های ناموضعی گشتاور ضربی و گشتاور وارون ضربی، در تعیین احتمال موفقیت مدل لوجیت خطای کم‌تری را نسبت به مدل‌های درست‌نمایی تاوانیده لاسو

منحنی‌ها مشخصه عملکرد دریافت‌کننده ROC، معمولا برای ارزیابی مدل‌های رگرسیونی با متغیرهای پاسخ دودویی به کار



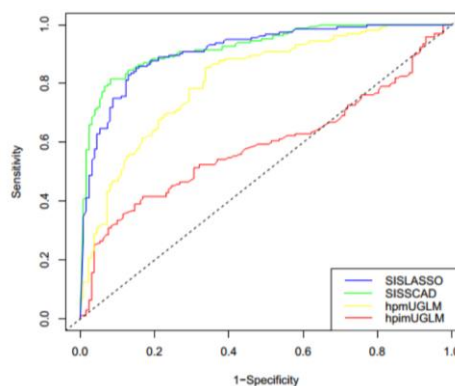
نمودار ۳. میانگین منحنی ROC با ۵۰ بار تکرار داده‌ها در چهار روش hpmUGLM، hpimUGLM، SIS-SCAD و SIS-LASSO برای اندازه‌های مختلف n و p .

می‌روند. هرچه سطح زیر منحنی بزرگ‌تر باشد نشانه مناسب‌تر بودن مدل در برآورد مقدار حقیقی متغیر پاسخ است. چنان‌که از نمودارها نمایان است، سطح زیر منحنی ROC در روش‌های انتخاب متغیر به شیوه درست‌نمایی توان‌یافته لاسو و اسکاد نسبت به روش‌های انقباضی بیزی بیش‌تر است که می‌تواند ناشی از ناتراز بودن تعداد صفر و یک‌های مربوط به متغیر پاسخ باشد زیرا یکی از ایرادهای منحنی ROC در حالت‌های ناتراز آن است که این منحنی‌ها در این حالت سطح منحنی را به غلط بزرگ‌تر نشان می‌دهند.

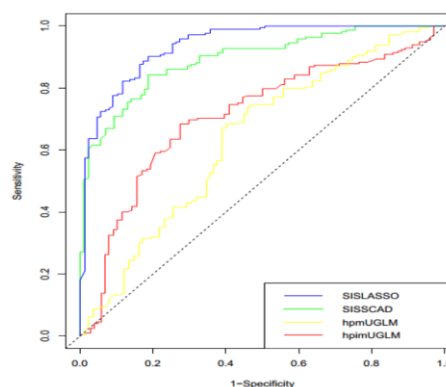
۵. مثال کاربردی

در این بخش با استفاده از داده‌های فوق بالابعد مربوط به بیماری سرطان خون و داده‌های غیر بالابعد مربوط به کارایی موتورهای تولیدشده توسط گروه صنعتی برنز تولیدکننده موتور خودرو، رفتار روش‌های پیشنهادی مطالعه می‌شود. گلوب و همکاران [۲۲]، برای گروه‌بندی سرطان بر اساس نوع ژن موجود در دی ان ای افراد، روی بیماران دارای سرطان خون مطالعه کردند. هدف ما استفاده از داده‌های گلوب و همکاران و تشخیص بین دو نوع

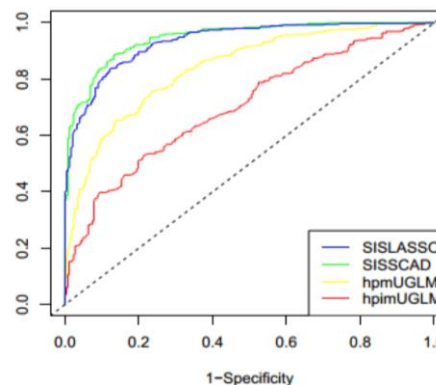
و اسکاد دارد و از طرفی استفاده از ابر پیشین‌های ناموضعی گشتاور وارون ضریبی، این خطا را کمینه می‌کند که نشانه‌ای مناسب‌تر بودن روش در برآورد مقدار حقیقی متغیر پاسخ است.



(الف) $P=1000, n=200$



(ب) $P=2000, n=200$



(ج) $P=3000, n=400$

(د) $P=4000, n=400$

در مثالی دیگر، کارایی روش‌های انتخاب مدل به شیوه بیزی پیشنهادی با داده‌های صنعتی مربوط به ۱۶ عامل تاثیرگذار در کارایی بهینه موتورهای خودرو تولید شده در شرکت قطعه‌سازی برنز و برای ۱۸۷ نمونه بررسی شده و نتایج انتخاب مدل به شیوه بیزی با استفاده از ابر پیشین‌های گشتاور ضربی، ابر پیشین‌های گشتاور وارون ضربی، اسکاد و لاسو در جدول ۴ آورده شده است.

جدول ۴. میانگین توان دوم خطای پیش‌بینی و خطای احتمال موفقیت در انتخاب عوامل موثر در کارایی بهینه موتور خودرو به چهار روش $hpmUGLM$ ، $hpimUGLM$ ، $SIS-SCAD$ و $SIS-LASSO$.

	$hpmUGLM$	$hpimUGLM$	$SIS-SCAD$	$SIS-LASSO$
MSPE	۰.۱۷۷۴	۰.۱۲۲۹	۰.۱۲۷۶	۰.۱۲۷۶
MSE	۰.۲۴۷۳۵۳۸	۰.۲۲۲۴۱۵۵	۰.۲۴۶۶۷۸۳	۰.۲۴۶۶۷۸۳

در ارزیابی نتایج درج شده در جدول ۴ ملاحظه می‌شود که میانگین توان دوم خطای پیش‌بینی و میانگین توان دوم خطای احتمال موفقیت در انتخاب متغیر به شیوه انقباضی بیزی با ابر پیشین‌های ناموضعی گشتاور ضربی نسبت به روش‌های درست‌نمایی تاوانیده لاسو و اسکاد کم‌تر است یعنی در حالت استفاده از ابر پیشین‌های ناموضعی گشتاور ضربی، احتمال موفقیت مدل لوجیت خطای کم‌تری را نسبت به مدل‌های درست‌نمایی تاوانیده لاسو و اسکاد و حتی استفاده از ابر پیشین‌های ناموضعی گشتاور وارون ضربی دارد که نشانه‌ی مناسب‌تر بودن شیوه انقباضی بیزی در حالت استفاده از ابر پیشین‌های ناموضعی گشتاور ضربی در نوع انتخاب مدل در تشخیص بین عوامل موثر در کارایی بهینه موتور خودرو است. با مقایسه دو مثال کاربردی سرطان خون و کارایی بهینه موتور خودرو ملاحظه می‌شود که روش‌های انتخاب متغیر به شیوه بیزی از روش‌های درست‌نمایی تاوانیده عملکرد بهتری دارند و از سوی دیگر برای داده‌های فوق بالابعد، استفاده از ابر پیشین‌های ناموضعی گشتاور وارون ضربی و در داده‌های غیر بالا عد استفاده از ابر پیشین‌های ناموضعی گشتاور ضربی کارایی مناسب‌تری دارند.

سرطان خون حاد: مغز استخوان^{۱۴} و لنف^{۱۵} است. ماتریس طرح حاوی آرایه‌های دی ان ای از نمونه‌های مغز استخوان و شامل ۷۲ نمونه و ۷۱۲۹ ژن می‌باشد. داده‌ها را به دو بخش آموزشی و آزمودنی تفکیک کرده به قسمی که داده‌های آزمودنی حاوی اطلاعات ۳۴ بیمار سرطان خون حاد شامل ۲۰ بیمار از نوع مغز استخوان و ۱۴ بیمار از نوع لنف هستند و داده‌های آموزشی حاوی اطلاعات ۳۸ بیمار سرطان خون حاد شامل ۲۷ بیمار از نوع مغز استخوان و ۱۱ بیمار از نوع لنف هستند. نتایج اجرای انتخاب مدل به شیوه بیزی با استفاده از ابر پیشین‌های گشتاور ضربی و ابر پیشین‌های گشتاور وارون ضربی و اسکاد و لاسو در جدول ۳ آورده شده است.

جدول ۳. میانگین توان دوم خطای پیش‌بینی و خطای احتمال موفقیت در انتخاب متغیر مربوط به داده‌های سرطان خون به چهار روش $hpmUGLM$ ، $hpimUGLM$ ، $SIS-SCAD$ و $SIS-LASSO$.

	$hpmUGLM$	$hpimUGLM$	$SIS-SCAD$	$SIS-LASSO$
MSPE	۰.۳۵۲۹۰۰۰	۰.۴۴۱۲۰۰۰	۰.۱۱۷۶۰۰۰	۰.۱۴۷۰۰۰
MSE	۰.۱۰۲۰۰۰۰	۰.۲۴۶۹۰۰۰	۰.۱۵۹۱۰۰۰	۰.۲۴۱۲۰۰۰

در ارزیابی به‌کارگیری روش‌های انتخاب متغیر برای داده‌های مربوط به سرطان خون، از نتایج درج شده در جدول ۳ ملاحظه می‌شود که میانگین توان دوم خطای پیش‌بینی در انتخاب متغیر به شیوه درست‌نمایی تاوانیده لاسو و اسکاد نسبت به روش انقباضی بیزی کم‌تر است در حالی که میانگین توان دوم خطای احتمال موفقیت خلاف آن را نشان می‌دهد یعنی در حالت استفاده از ابر پیشین‌های ناموضعی گشتاور وارون ضربی، احتمال موفقیت مدل لوجیت خطای کم‌تری را نسبت به مدل‌های درست‌نمایی تاوانیده لاسو و اسکاد و حتی استفاده از ابر پیشین‌های ناموضعی گشتاور ضربی دارد که نشانه‌ی مناسب‌تر بودن شیوه انقباضی بیزی در حالت استفاده از ابر پیشین‌های ناموضعی گشتاور وارون ضربی در برآورد تاثیر نوع انتخاب مدل در تشخیص بین دو نوع سرطان خون حاد: مغز استخوان و لنف است.

^{۱۵} Acute Lymphoblastic Leukemia (ALL)

^{۱۴} Acute Myeloid Leukemia (AML)

۶. بحث و نتیجه‌گیری

۸- منبع

- [1] Zellner, A. (1971), *An introduction to Bayesian inference in econometrics*, John Wiley & Sons Inc.
- [2] Berger, J. O., Pericchi, L. R., Ghosh, J., Samanta, T., & De Santis, F. (2001), "Objective Bayesian methods for model selection: Introduction and comparison," *Lecture Notes Monograph Series*, 135-207.
- [3] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
- [4] Liang, F., Truong, Y. K., & Wong, W. H. (2001). Automatic Bayesian model averaging for linear regression and applications in Bayesian curve fitting. *Statistica Sinica*, 1005-1029.
- [5] Liang, F., Song, Q., & Yu, K. (2013). Bayesian subset modeling for high-dimensional generalized linear models. *Journal of the American Statistical Association*, 108, 589-606.
- [6] Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques*.
- [7] Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103, 410-423.
- [8] Bové, D. S., & Held, L. (2011). Hyper-g priors for generalized linear models. *Bayesian Analysis*, 6, 387-410.
- [9] Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70, 849-911.
- [10] Fan, J., Samworth, R., & Wu, Y. (2009). Ultrahigh dimensional feature selection: beyond the linear model. *The Journal of Machine Learning Research*, 10, 2013-2038.
- [11] Fan, J., & Song, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *The Annals of Statistics*, 38(6), 3567-3604.

در این مقاله به مقایسه عملکرد دو پیشین ناموضعی گشتاور ضربی و گشتاور وارون ضربی در انتخاب متغیر همزمان با برآورد پارامترها در مدل‌های خطی تعمیم‌یافته فوق بالا بعد به شیوه انقباضی بیزی پرداخته شده است. جهت برقراری شرایط محاسباتی در روش پیشنهادی، قیدهایی در نظر گرفته، سپس با فرض برقراری قیدهای مسئله، به دلیل پیچیدگی محاسباتی احتمالات پسین ناشی از به‌کارگیری پیشین‌های ناموضعی و مدل‌های خطی تعمیم‌یافته، از تقریب لاپلاس برای محاسبه مقدار عددی درست‌نمایی حاشیه‌ای مخرج پسین استفاده شده و در نهایت مقدار تقریبی برای احتمالات پسین کلیه مدل‌های موجود در فضای مدل محاسبه گردیده است. از طرفی جهت انتخاب مدل بهینه در فضای متراکم مدل، الگوریتم جستجوی تصادفی تفنگی ساده شده همراه با غربالگری (S5) برای GLMها اجرا شده است. در نهایت، عملکرد روش پیشنهادی با مطالعه شبیه‌سازی و تحلیل یک مجموعه داده واقعی مورد ارزیابی قرار گرفته است. نتایج نشان می‌دهند انتخاب متغیر به شیوه انقباضی بیزی در دو حالت استفاده از ابر پیشین‌های ناموضعی گشتاور ضربی و گشتاور وارون ضربی، با توجه به آن که در تعیین احتمال موفقیت مدل لوجیت خطای کم‌تری را نسبت به مدل‌های درست‌نمایی توان‌یافته لاسو و اسکاد دارد لذا در برآورد ضرایب غیر صفر و انتخاب مدل عملکرد بهتری دارد و از طرفی استفاده از ابر پیشین‌های ناموضعی گشتاور وارون ضربی، این خطا را کمینه می‌کند که نشانه‌ی مناسب‌تر بودن این پیشین در برآورد مقدار حقیقی متغیر پاسخ است.

۷. تقدیر و تشکر

نویسندگان از سردبیر، داوران و ویراستار محترم مجله که نظرات ارزشمند ایشان باعث بهبود مطالب ارائه شده در این مقاله گردید، کمال تشکر و قدردانی را دارند.

- [12] Johnson, V. E., & Rossell, D. (2012). Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association*, 107, 649-660.
- [13] Rossell, D., & Telesca, D. (2017). Nonlocal priors for high-dimensional estimation. *Journal of the American Statistical Association*, 112, 254-265.
- [14] Shin, M., Bhattacharya, A., & Johnson, V. E. (2018). Scalable Bayesian variable selection using nonlocal prior densities in ultrahigh-dimensional settings. *Statistica Sinica*, 28(2), 1053.
- [15] Nikooienejad, A., Wang, W., & Johnson, V. E. (2016). Bayesian variable selection for binary outcomes in high-dimensional genomic studies using non-local priors. *Bioinformatics*, 32, 1338-1345.
- [16] Nikooienejad, A., Wang, W., & Johnson, V. E. (2020). Bayesian variable selection for survival data using inverse moment priors. *The annals of applied statistics*, 14, 809.
- [17] Wu, H. H., Ferreira, M. A., Elkhoully, M., & Ji, T. (2020). Hyper nonlocal priors for variable selection in generalized linear models. *Sankhya A*, 82, 147-185.
- [18] Johnson, V. E., & Rossell, D. (2010). On the use of non-local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, 143-170.
- [19] Robert, C. P. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*, Vol. 2. New York: Springer.
- [20] Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81, 82-86.
- [21] Narisetty, N. N., Shen, J., & He, X. (2018). Skinny gibbs: A consistent and scalable gibbs sampler for model selection. *Journal of the American Statistical Association*.
- [22] Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., et al. (1999), "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *science*, 286, 531-537.

Bayesian Shrinkage Variable Selection with Non-Local Priors in Ultrahigh-Dimensional Logistic Generalized Linear Models.

Robabeh Hosseinpour Samim Mamaghani

Ph.D. Student, Department of Statistics, Faculty of Statistics, Mathematics and Computer,
Allameh Tabataba'i University, Tehran, Iran. hosseinpour@atu.ac.ir

Farzad Eskandari

Full Professor, Department of Statistics, Faculty of Statistics, Mathematics and Computer,
Allameh Tabataba'i University, Tehran, Iran. askandari@atu.ac.ir¹⁶

Vahid RezaeiTabar

Professor, Department of Statistics, Faculty of Statistics, Mathematics and Computer, Allameh
Tabataba'i University, Tehran, Iran. RezaeiTabar@atu.ac.ir

Abstract: One of the basic issues in Ultrahigh-dimensional data analysis is fitting the optimal model and estimating its unknown parameters in such a way that it can correctly interpret the structure of the investigated data. In this article, we compare two non-local hyper priors: hyper product moment and hyper product inverse moment priors in determining the optimal model at the same time as estimating the parameters in variable selection using Bayesian Shrinkage in ultrahigh-dimensional generalized linear models. In order to compute the posterior probabilities, the Laplace approximation method was used, and to select the optimal model in the model space of posterior probabilities, Simplified Shotgun Stochastic Search algorithm with Screening (S5) for GLMs was used along with screening. Finally, through the study of simulation and real data analysis, the effectiveness of the above Bayesian Shrinkage methods has been evaluated with the ISIS-LASSO and ISIS-SCAD method.

Keywords: Variable Selection, Ultrahigh dimensional, Penalized likelihood, Quality parameters, ISIS –LASSO, Optimization.

1. Introduction

When the number of covariates grows at a sub-exponential rate of n , variable selection will be the first step for dimension reduction to estimate the parameters in regression. Our objective is to fit a GLM by efficiently estimating regression coefficient and use it for subsequent inference. The increasing use of generalized linear models on the one hand and the large volume of its inputs on the other hand, causes complications in the stages of fitting model and estimation of model parameters, so it seems logical in the high-dimensional state, a small number of input variables entered the statistical model. In determining the estimation method of the model parameters, the discussion of the infiniteness of the estimations of the maximum likelihood method and the low skewness and high variance of the least squares method confirms the necessity of having a method with the ability to select the variable and estimate the parameters at the same time. Many common methods have recently been done for variable selection from both frequentist and Bayesian perspectives. Most of the frequentist methods can be interpreted from a Bayesian perspective because they share the basic desire of shrinkage toward sparse models. Local priors (LPs) put a positive probability on the null value of the parameter whereas nonlocal priors (NLPs) put zero probability on the null value.

¹⁶ Corresponding author: askandari@atu.ac.ir

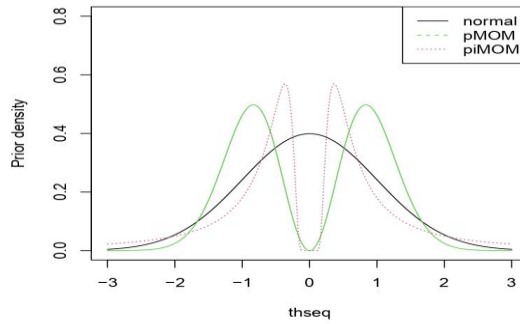


Figure1. A depiction of normal local prior and Nonlocal priors pMOM and piMOM according to the scale parameter between 3 and -3.

Thus, nonlocal priors consider a clear separation between the null hypothesis that the regression coefficient is equal to zero and the alternative hypothesis that the coefficient is different from zero. (Johnson & Rossell) [4] can refer to a detailed discussion on properties of local and nonlocal priors in the context of Bayesian testing. Nonlocal priors lead to faster accumulation of evidence in favor of a true null hypothesis. This latter property was demonstrated by (Johnson & Rossell) [5] for a Gaussian linear model.

2. Problem modeling

For generalized linear regression with the $p \leq n$ setting, (Wu, Ferreira, Elkhoully, & Ji)[8] propose hyper nonlocal priors for variable selection in generalized linear models. They combine the Fisher information matrix with the Johnson and Rossell moment and inverse moment priors and assign hyper priors to the scale parameters. Let $\mathbf{y}_n = (y_1, \dots, y_n)$ be an n -dimensional response vector and \mathbf{X}_n be a $n \times p$ design matrix, where n is the sample size and p is the total number of covariates. Suppose \mathbf{M} denotes the model space that collects all the model indices \mathbf{k} ; i.e., $\mathbf{M} = \{\mathbf{k}: \mathbf{k} \subseteq \{0,1\}^p\}$ and $|\mathbf{k}|$ is cardinal of \mathbf{k} set. We assume that the true model exists, and is defined as the smallest model in the model space \mathbf{M} that contains the true data-generating distribution. Consequently, the problem of selecting the best subset is now equivalent to the problem of identifying the true model in \mathbf{M} . Let $\mathbf{t} \subseteq \{1,2, \dots, p\}$ be the true model. When the link function is logit $asg(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$, then the likelihood function of $\mathbf{y}|\boldsymbol{\beta} \sim f(\mathbf{y}|\mathbf{x}^T \boldsymbol{\beta})$ will be

$$\begin{aligned}
 L(\boldsymbol{\beta}|\mathbf{y}) &= \prod_{i=1}^n \left(\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right)^{y_i} \left(\frac{1}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right)^{1-y_i} \\
 &= \exp\left\{ \sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta} \right\} / \prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})]
 \end{aligned}
 \tag{1}$$

If $diag(\mathbf{I}_{k0}) = \mathbf{I}_{k0}$ is unit Fisher information matrix evaluated at $\boldsymbol{\beta} = \mathbf{0}$, moment and inverse moment (piMOM) priors will be

$$\begin{aligned}
 \pi_M(\boldsymbol{\beta}_k|k) &= (\tau)^{-r|k|/2} |\mathbf{I}_{k0}|^{r/2} \times N(\boldsymbol{\beta}_k; \mathbf{0}, (\tau)(\mathbf{I}_{k0})^{-1}) \times \prod_{i=1}^{|\mathbf{k}|} |\beta_{ki}|^{(r+1)} \\
 \pi_I(\boldsymbol{\beta}_k|k) &= \frac{(\tau)^{r|k|/2} |\mathbf{I}_{k0}|^{-r/2}}{\Gamma\left(\frac{\tau}{2}\right)^{|\mathbf{k}|}} \times \exp\{-\tau(\boldsymbol{\beta}_k^T \mathbf{I}_{k0} \boldsymbol{\beta}_k)^{-1}\} \times \prod_{i=1}^{|\mathbf{k}|} |\beta_{ki}|^{-(r+1)}
 \end{aligned}
 \tag{2}$$

The scale parameter τ in the nonlocal prior density reflects the dispersion of the nonlocal

prior density around zero, and determines the size of the regression coefficients that will be shrunk to zero. We suppose τ has the inverse gamma hyper prior in pmGLM and a gamma hyper prior in pimGLM that are able to learn about the prior scale parameter from data and provide robust inferential results. Also the model space prior is assumed as follows

$$\pi(\mathbf{k}) \propto I(|\mathbf{k}| \leq m_n) \quad (3)$$

The $m_n = O(\min\left\{\log p, \left(\frac{n}{\log p}\right)^\alpha\right\})$ for $0 < \alpha < 1$, is a positive integer restricting the size of the largest model, and a uniform prior is placed on the model space restricting our analysis to models having size less than or equal to m_n .

By the hierarchical Bayesian model (1) to (3) and Bayes' rule, the resulting posterior probability for moment (pMOM) and inverse moment (piMOM) priors will be

$$\begin{aligned} \pi_M(\boldsymbol{\beta}_k | \mathbf{y}_n) &\propto \frac{\exp\{\sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta}\}}{\prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})]} \times (\tau)^{-r|\mathbf{k}|/2} |I_{k_0}|^{r/2} \times N(\boldsymbol{\beta}_k; 0, (\tau)(I_{k_0})^{-1}) \\ &\times \prod_{i=1}^{|\mathbf{k}|} |\beta_{ki}|^{(r+1)} \times I(|\mathbf{k}| \leq m_n) \\ \pi_I(\boldsymbol{\beta}_k | \mathbf{y}_n) &\propto \frac{\exp\{\sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta}\}}{\prod_{i=1}^n [1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})]} \times \frac{(\tau)^{r|\mathbf{k}|/2} |I_{k_0}|^{-r/2}}{(\Gamma(\frac{r}{2}))^{|\mathbf{k}|}} \times \exp\{-(\tau)(\boldsymbol{\beta}_k^T I_{k_0} \boldsymbol{\beta}_k)^{-1}\} \times \prod_{i=1}^{|\mathbf{k}|} |\beta_{ki}|^{-(r+1)} \\ &\times I(|\mathbf{k}| \leq m_n) \end{aligned} \quad (4)$$

Then with the 0-1 loss function, a possible estimator of $\boldsymbol{\beta}_k$ based on posterior is the maximum a posteriori (MAP) estimator and it is defined by

$$\hat{\boldsymbol{\beta}}_k = \operatorname{argmax}_{\boldsymbol{\beta}} \pi(\boldsymbol{\beta}_k | \mathbf{y}_n) \quad (5)$$

Of course, the closed form of posterior probabilities (4) cannot be obtained due to not only the nature of GLMs but also the structure of hyper nonlocal prior. Therefore, special efforts need to be devoted to computational strategy. Laplace approximation maximizes the logarithm of the unnormalized joint posterior density with one of several optimization algorithms and the goal is to estimate the posterior mode and variance of each parameter. In variable selection perspective, the essence is to force the estimated model to be sparse by penalizing dense models. The resulting posterior probability for model \mathbf{k} is denoted by,

$$\pi(\mathbf{k} | \mathbf{y}_n) = \frac{\pi(\mathbf{k}) m_{\mathbf{k}}(\mathbf{y}_n)}{\sum_{\mathbf{j} \in M} \pi(\mathbf{j}) m_{\mathbf{j}}(\mathbf{y}_n)} \quad (6)$$

$m_{\mathbf{k}}(\mathbf{y}_n)$ is the marginal density of \mathbf{y}_n under model \mathbf{k} given by

$$m_{\mathbf{k}}(\mathbf{y}_n) = \int \exp\{\ell(\boldsymbol{\beta}_k | \mathbf{y}_n)\} \pi(\boldsymbol{\beta}_k | \mathbf{k}) d\boldsymbol{\beta}_k \quad (7)$$

where log likelihood function is

$$\ell(\boldsymbol{\beta}_k | \mathbf{y}) = \log \left(\prod_{i=1}^n \left(\frac{\exp(\mathbf{x}_{ik}^T \boldsymbol{\beta}_k)}{1 + \exp(\mathbf{x}_{ik}^T \boldsymbol{\beta}_k)} \right)^{y_i} \left(\frac{1}{1 + \exp(\mathbf{x}_{ik}^T \boldsymbol{\beta}_k)} \right)^{1-y_i} \right) \quad (8)$$

In particular, these posterior probabilities can be used to select a model by computing the posterior mode defined by

$$\hat{\mathbf{k}} = \operatorname{argmax}_{\mathbf{k}} \pi(\mathbf{k} | \mathbf{y}_n) \quad (9)$$

The marginal density $m_{\mathbf{k}}(\mathbf{y}_n)$ does not have a closed form for GLMs, then according to Tierney & Kadane [7], Laplace approximation to the marginal likelihood of model \mathbf{k} observing \mathbf{y}_n is given by

$$\tilde{m}_{\mathbf{k}}(\mathbf{y}_n) = (2\pi)^{\frac{|\mathbf{k}|}{2}} |\hat{\mathbf{H}}_{\mathbf{k}}|^{-\frac{1}{2}} \exp[\ell(\tilde{\boldsymbol{\beta}}_{\mathbf{k}}|\mathbf{y}_n)] \pi(\tilde{\boldsymbol{\beta}}_{\mathbf{k}}|\mathbf{k}) \quad (10)$$

,where $\ell(\tilde{\boldsymbol{\beta}}_{\mathbf{k}}|\mathbf{y}_n)$ denotes the log likelihood function, $\tilde{\boldsymbol{\beta}}_{\mathbf{k}}$ denotes the posterior mode of $\boldsymbol{\beta}_{\mathbf{k}}$ under model \mathbf{k} , and $\hat{\mathbf{H}}_{\mathbf{k}}$ denotes the Hessian matrix of the logarithm of the posterior of $\boldsymbol{\beta}_{\mathbf{k}}$ evaluated at $\tilde{\boldsymbol{\beta}}_{\mathbf{k}}$. Finally, we replace $m_{\mathbf{k}}(\mathbf{y}_n)$ in (6) with $\tilde{m}_{\mathbf{k}}(\mathbf{y}_n)$ for any $\mathbf{k} \in M$, to obtain $\tilde{\pi}(\mathbf{k}|\mathbf{y}_n)$, the approximate posterior probability of model \mathbf{k} , and then employ $\tilde{\pi}(\mathbf{k}|\mathbf{y}_n)$ as the criterion for Bayesian variable selection. We consider the following regularity conditions for theoretical properties of our posterior:

Condition (A1): For some $0 < \alpha < 1$,

$$\begin{aligned} \log p &= O(n^\alpha) \\ m_n &= \left(\frac{n}{\log p}\right)^\alpha \\ 0 < \tau &\leq \log p \end{aligned} \quad (11)$$

Condition (A2): For some $0 < \alpha < 1$,

$$\begin{aligned} \max_{i,k} |x_{ik}| &\leq 1 \\ \lim_{p \rightarrow \infty} \sum_{k=1}^p |\beta_k^*| &< \infty \\ \|\beta_{0,t}\|_2^2 &= O((\log p)^\alpha) \end{aligned} \quad (12)$$

Condition (A3): For some $\lambda > 0$, suppose the ordered eigenvalues $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{\max}$ of unit Fisher information matrix $\mathbf{I}(\boldsymbol{\beta}_{0,k})$ and the Gram matrix

$\left(\frac{\mathbf{X}_{\mathbf{k}}^T \mathbf{X}_{\mathbf{k}}}{n}\right)$ over model \mathbf{k} , then,

$$0 < \lambda \leq \min_{\mathbf{k}:|\mathbf{k}|\leq m_n} \lambda_1(\mathbf{I}(\boldsymbol{\beta}_{0,k})) \leq \max_{\mathbf{k}:|\mathbf{k}|\leq m_n} \lambda_{\max}\left(\frac{\mathbf{X}_{\mathbf{k}}^T \mathbf{X}_{\mathbf{k}}}{n}\right) \leq (\log p)^\alpha \quad (13)$$

Condition (A4) (Beta-min condition): Let $\mathbf{t} \subseteq \{1, 2, \dots, p\}$ be the true model, for some constant $C > 0$,

$$\min_{\mathbf{k} \in t} \beta_{\mathbf{k}}^2 \geq C \max \left\{ \frac{|t| \max_{\mathbf{k}: |\mathbf{k}| \leq t} \lambda \left(\frac{X_{\mathbf{k}}^T X_{\mathbf{k}}}{n} \right) \log p_{\max}}{n \frac{1}{\log p}} \right\} \quad (14)$$

3. Computational Strategy

After using the Laplace approximations for the marginal probabilities in (10), we need method to explore interesting regions of the resulting high-dimensional model spaces and quickly identify regions of high posterior probability over models. For computational purposes, we use another stochastic algorithm to search the model space for rapidly identifying regions with high posterior probability and finding the maximum a posteriori (MAP) model. The Simplified shotgun stochastic search algorithm with screening is a stochastic search method that screens covariates at each step to explore the enormous model space. To increase the efficiency of exploring the model space, we use the S5 algorithm proposed by (Shin, Bhattacharya, & Johnson)[6] for variable selection in linear regression problems. It is a stochastic search method that screens covariates at each step. The concept of screening covariates for GLMs response data is proposed in (Fan & Song) [2] and defined a more general version of the independent learning with ranking the maximum marginal likelihood estimator (MMLE) or the maximum marginal likelihood itself. Suppose that the current model is \mathbf{k} . Let \mathbf{k}^c be the complement of set \mathbf{k} containing columns of the design matrix that are not presented in the current model. The S5 algorithm for GLM data works as follows: At each step the $d = 2[\log p]$ covariates with highest maximum marginal likelihood are candidates to be added to the current model \mathbf{k} and comprise the addition set, Γ_{scr}^+ . The deletion set, Γ^- contains the current model, except that one variable is removed. From the current model \mathbf{k} , we consider moves to each of its neighbors in Γ_{scr}^+ and Γ^- with a probability proportional to the marginal probabilities of these neighboring models. To avoid local maxima, the model probabilities used in S5 are raised to the power of $1/t^l$, where t^l is the l th temperature in an annealing schedule in which “temperatures” decrease. To increase the number of visited models, a specified number of iterations are performed at each temperature. At the end of the procedure, the model with the highest posterior probability of visited models is identified as the HPPM.

4. Simulation Studies

We applied our method to both simulated data and real data, to investigate the performance of the proposed method. Iterative Sure independence screening (ISIS) was introduced by (Fan & Lv)[1] to reduce the computation in ultra-high dimensional variable selection. It refers to ranking features according to marginal utility, namely, each feature is used independently as a predictor to decide its usefulness for predicting the response. In the ISIS-SCAD/LASSO method, first the Iterative Sure Independence Screening for different variants implements, and then fits the final regression model using the SCAD/LASSO regularized log likelihood for the variables picked by ISIS. In simulation studies, Let $n = \{200, 400\}$ and $p = \{1000, 2000, 3000, 4000\}$, X be the design matrix and for a true model \mathbf{k} , the response vector represents a sequence of Bernoulli samples with probability of success

$$\pi_i = \frac{e^{x_{ik}^T \beta_k}}{1 + e^{x_{ik}^T \beta_k}} \tag{15}$$

Elements of the design matrix X were sampled from a multivariate normal distribution with mean 0 and covariance matrix Σ , under the Autoregressive correlated design, where $\Sigma_{ij} = 0.5^{|i-j|}$, for all $1 \leq i \leq j \leq p$. With the fixed true model $t = (1,2,3,4,5,6)$ and coefficient $\beta_t^0 = (2, -2,4,3, -3, -4)$. The variable selection procedure in all algorithms was run 50 times. In each trial, true and false positive values for hpmUGLM, hpimUGLM and ISIS-SCAD/LASSO were counted by comparing the selected model with the true one. To evaluate the performance of variable selection, mean-squared prediction error (MSPE) and mean-squared error (MSE) were conducted. The criteria are defined as

$$MSPE = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (\hat{y}_i - y_{test,i})^2 \tag{16}$$

$$MSE(\hat{\pi}) = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (\hat{\pi}_i - \pi_i)^2 \tag{17}$$

Following tables summarize the results of applying hpmUGLM , hpimUGLM, ISIS-SCAD and ISIS-LASSO approaches to the simulation data

Table 1. The mean-squared prediction error (MSPE) to evaluate the performance of variable selection based on each method.

For n=200	hpimUGLM	hpmUGLM	SISSCAD	SISLASSO
p=1000	0.214360	0.5266667	0.103400	0.110218
p=2000	0.223594	0.5266667	0.126842	0.083746
p=3000	0.161210	0.5266667	0.111738	0.126532
p=4000	0.230350	0.2044667	0.102818	0.098780

Table 2. The mean-squared error (MSE) for success probability to evaluate the performance of variable selection based on each method.

For n=200	hpimUGLM	hpmUGLM	SISSCAD	SISLASSO
p=1000	0.2326000	0.4082000	0.7049667	0.6416333
p=2000	0.2593000	0.3885333	0.5361667	0.6188667
p=3000	0.1262000	0.3695333	0.558200	0.4418000
p=4000	0.2921667	0.3362667	0.6959000	0.6765333

As can be seen from the results listed in tables 1 and 2, The mean-squared prediction error in the selection of variables using the LASSO and SCAD methods are lower than the Bayesian methods hpmUGLM and hpimUGLM, while The mean-squared error (MSE) for success probability of the logit model for the Bayesian method hpmUGLM and hpimUGLM, have lower error than the LASSO and SCAD, which is a sign that the Bayesian methods are more appropriate in the evaluation of the performance of variable selection methods.

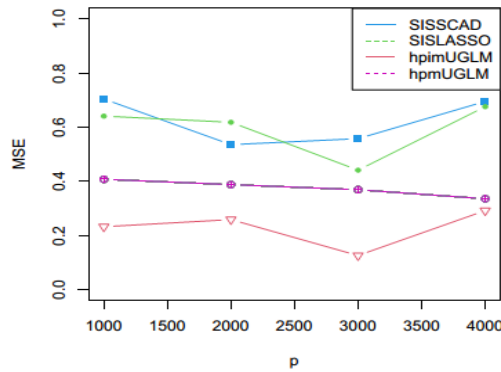


Figure2. A depiction of the mean square error of the probability of success in four methods hpmUGLM, hpmUGLM, SIS-SCAD and SIS-LASSO for different sizes of p.

It can be seen that the selection of the variable by the Bayesian methods in the two cases of using the non-local priors moment, in determining the probability of success of the logit model, has a lower error than the compensated probability models of LASSO and SCAD, on the other hand, the use of non-local priors of inverse moment minimizes this error, which is a more suitable sign of the method in estimating the true value of the response variable.

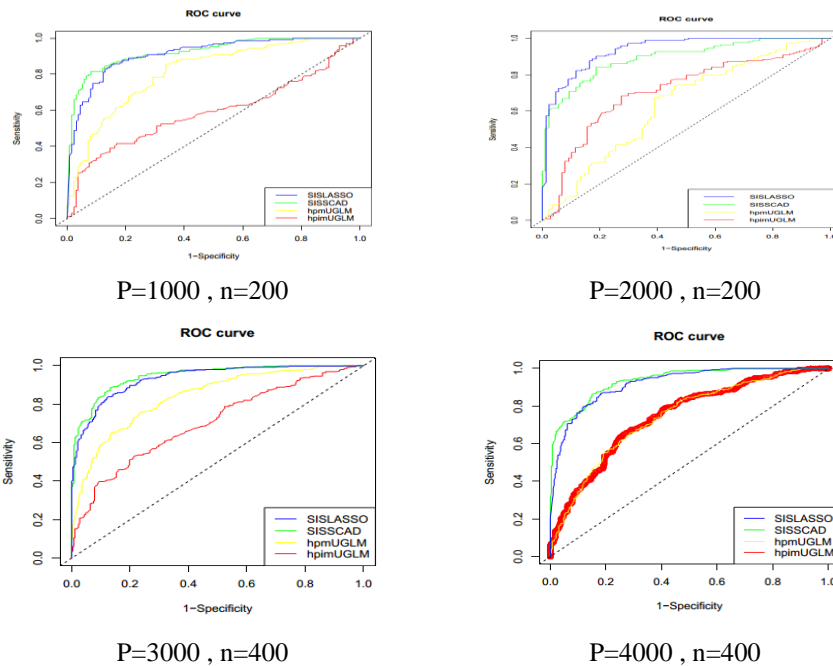


Figure3. ROC curve with 50 repetitions of data in four methods hpmUGLM, hpmUGLM, SIS-SCAD and SIS-LASSO for different sizes of p.

The larger the area under the ROC curve is, the more suitable the model is in estimating the true value of the response variable. As can be seen from the Figure3, the area under the ROC curve in the variable selection methods with the LASSO and SCAD is more than the Bayesian contraction methods, which can be caused by the imbalance of the number of zeros and ones. related to the response variable, because one of the defects of the ROC curve in non-aligned states is that these curves in this state show the surface of

the curve with a bigger error.

4. Real data analysis

In this section, the behavior of the proposed methods is studied by using the data related to the blood cancer disease and efficiency of electric motor data. (Golub, Slonim, Tamayo, & Huard)[3] studied patients with leukemia to group cancer based on the type of gene present in DNA. Our goal is to use the data of Golub et al. and distinguish between two types of acute leukemia: bone marrow 1 and lymph 2. The design matrix contains DNA arrays from bone marrow samples and includes 72 samples and 7129 genes. We split the data into training and test sets so that the test data contains the information of 34 acute leukemia patients, including 20 bone marrow type patients and 14 lymph type patients, and the training data contains the information of 38 cancer patients. Acute blood includes 27 patients of bone marrow type and 11 patients of lymph type.

Table 3. The results of performing variable selection to the blood cancer disease data using four methods hpmUGLM, hpimUGLM, SIS-SCAD and SIS-LASSO.

For Real data	hpmUGLM	hpimUGLM	SISSCAD	SISLASSO
MSPE	0.3529000	0.44120000	0.1176000	0.1471000
MSE	0.1020000	0.24690000	0.1591000	0.2413000

Table 4. The results of performing variable selection in efficiency of electric motor data using four methods hpmUGLM, hpimUGLM, SIS-SCAD and SIS-LASSO.

For Real data	hpmUGLM	hpimUGLM	SISSCAD	SISLASSO
MSPE	0.1774	0.1229	0.1276	0.1276
MSE	0.2473538	0.2224155	0.2464783	0.2464783

In evaluating the use of variable selection methods for blood cancer data, it can be seen from the results listed in Table 3 that the mean squared error of the Bayesian methods is lower, while the Mean square error of the probability of success shows the opposite result, that is, in the case of using hyper-nonlocal priors, the probability of success of the logit model has a lower error than LASSO and SCAD in the diagnosis between two types of acute leukemia: bone marrow and lymph.

5. Discussion and conclusion

In this article, the performance of non-local hyper priors: hyper product moment and hyper product inverse moment priors, have been compared in variable selection at the same time as parameter estimation in high-dimensional generalized linear models using Bayesian Shrinkage Variable Selection methods in order to establish the computational conditions in the proposed method, constraints are considered, then assuming the constraints of the problem, due to the computational complexity of the posterior probabilities caused by the use of nonlocal priors and generalized linear models, from the Laplace approximation for The calculation of the numerical value of the marginal probability of the posterior denominator was used and finally the approximate value for the posterior probabilities of all the models in the model space was calculated. On the other hand, in order to select the optimal model in the huge model space, the simplified shotgun stochastic search algorithm with screening (S5) for GLMs has been implemented. Finally, the performance of the proposed methods has been evaluated by simulation study and analysis of a real data set. The results show that the selection of the variable by Bayesian Shrinkage Variable Selection methods by using the hyper-nonlocal priors of

the moment and the inverse moment priors, considering that in determining the probability of success of the logit model, there is less error than the models It has Lasso and Scad, so it performs better in estimating non-zero coefficients and model selection, and on the other hand, the use of non-local priors of inverse moment priors minimizes this error, which is a more suitable sign of this method.

6. Reference

- [1] Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society*, 849–911.
- [2] Fan, J., & Song, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *The Annals of Statistics*, 3567-3604.
- [3] Golub, T., Slonim, D., Tamayo, P., & Huard, C. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *science*, 531–537.
- [4] Johnson, V., & Rossell, D. (2010). On the use of non-local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 143–170.
- [5] Johnson, V., & Rossell, D. (2012). Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association*, 107, 649-660.
- [6] Shin, M., Bhattacharya, A., & Johnson, V. (2018). Scalable Bayesian variable selection using nonlocal prior densities in ultrahigh-dimensional settings. *Statistica Sinica*, 1053.
- [7] Tierney, L., & Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association*, 81-86.
- [8] Wu, H., Ferreira, M., Elkhoully, M., & Ji, T. (2020). Hyper nonlocal priors for variable selection in generalized linear models. *Sankhya A*, 147-185.