تشخیص و حذف هم‌خطی در مدل رگرسیونی با استفاده از نظریهٔ اطلاع

فائزه شکیبا[۱]، غلامرضا محتشمی برزادران[۲]

تاریخ دریافت: ۱۳۹۳/۱۲/۱۷
تاریخ پذیرش: ۱۳۹۵/۷/۱۷

چکیده:
در این مقاله نخست مسئلهٔ هم‌خطی در مدل رگرسیونی معرفی می‌شود و شیوهٔ تشخیص هم‌خطی و راه‌های برطرف کردن هم‌خطی مطرح می‌شوند. در ادامه تعاریف مقدماتی از نظریهٔ اطلاع عنوان می‌شوند که در نهایت با استفاده از نظریهٔ اطلاع، هم‌خطی در مدل رگرسیونی شناسایی می‌شود و راه حلی برای برطرف کردن آن پیشنهاد می‌شود.
واژه‌های کلیدی: رگرسیون، هم‌خطی، نظریهٔ اطلاع، توزیع پسین.

۱ مقدمه

واژهٔ رگرسیون از لحاظ لغوی در فرهنگ لغت به معنی پس‌روی، برگشت و بازگشت است؛ اما از دید آمار و ریاضیات به مفهوم بازگشت به یک مقدار متوسط یا میانگین به کار می‌رود. مسائلی از نوع رگرسیون در قرن هجدهم مورد توجه قرار گرفت. از رگرسیون برای به الگو در آوردن رابطهٔ بین متغیرهای آماری استفاده می‌شود. می‌توان رگرسیون را این‌گونه معرفی کرد که روابط نادقیق بین متغیرهای آماری را تعیین می‌کند و به تحلیل این روابط می‌پردازد. در رگرسیون، متغیر پاسخ و پیشگو داریم (متغیرهای پاسخ و پیشگو به‌ترتیب به‌عنوان متغیرهای وابسته و مستقل شناخته می‌شوند). در این مقاله متغیر پاسخ را پیوسته در نظر گرفته‌ایم؛ بنا بر این مدل‌های خطی رگرسیون را به کار می‌بریم. اگر در مدل رگرسیونی فقط یک متغیر پیشگو باشد مدل رگرسیونی، مدل رگرسیونی خطی ساده است که برابر است با $Y=\beta_{0}+\beta_{1}X+\varepsilon$. β_{0} عرض از مبدأ، β_{1} شیب خط و ε_i ها خطاها می‌باشند که دارای توزیع نرمال و همچنین مستقل هستند.

حال اگر در مدل رگرسیونی k متغیر پیشگو داشته باشیم، با مدل رگرسیونی خطی چندگانه سروکار داریم که به‌صورت $\boldsymbol{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$ است. یک مسئلهٔ جدی که می‌تواند استفاده از مدل رگرسیونی را با اشکال مواجه کند، مسئلهٔ هم‌خطی یا همبستگی خطی نزدیک بین متغیرهای پیشگو است. قابل ذکر است که اگر هیچ رابطه‌ای خطی بین متغیرهای پیشگو وجود نداشته باشد، متغیرهای پیشگو متعامد هستند و پیش‌بینی و انتخاب یک مجموعهٔ مناسب از متغیرها برای مدل به‌آسانی صورت می‌پذیرد. یکی از راه‌های تشخیص هم‌خطی، ماتریس همبستگی است (مونتگومری [۱۲]) که می‌توان برای برطرف کردن هم‌خطی از مدل رگرسیونی ستیغی استفاده کرد که توسط هورل [۶] معرفی شده است. در رگرسیون ستیغی، [۳]ضرایب رگرسیونی از رابطهٔ

$$\beta_R=(\mathbf{X}'\mathbf{X}+\lambda\boldsymbol{I})^{-1}\mathbf{X}'\boldsymbol{Y}, \qquad (۱)$$

به دست می‌آیند. در رابطهٔ (۱)، $\lambda \geq ۰$ است که توسط تحلیل‌گر انتخاب می‌شود. برآورد ضرایب مدل رگرسیونی ستیغی، برآوردی اریب با کمترین واریانس است که ضرایب رگرسیونی را برآورد می‌کند.

با اختراع تلفن، مفهوم اطلاع و انتقال، افزایش یافت که برخی از دانشوران به بررسی مشکلات انتقال اطلاع پرداختند. تحقیق در مورد نظریهٔ اطلاع، اولین بار توسط هری نیکوئست [۱۳] در مقاله‌ای به نام «عوامل خاصی که سرعت تلگراف را تحت تأثیر قرار می‌دهند» صورت گرفت. کار مهم دیگر در این زمان، انتشار مقالهٔ «انتقال اطلاع» توسط هارتلی [۵] بود که اولین پایه‌های ریاضی نظریهٔ اطلاع را بنا گذاشت. تولد واقعی نظریهٔ اطلاع را می‌توان به مقالهٔ « نظریهٔ ریاضی مخابرات» نوشتهٔ کلود شانون

[۱] دانش‌آموختهٔ کارشناسی ارشد آمار، دانشگاه فردوسی مشهد
[۲] استاد گروه آمار، دانشگاه فردوسی مشهد

[3] regression ridge

[۱۷] نسبت داد. مفهوم اطلاعی که توسط شانون مطرح شد از دید آمار و احتمال است و با مفاهیم روزمره از اطلاع، مانند دانش، تفاوت دارد.

حال می‌خواهیم هم‌خطی را با کمک نظریهٔ اطلاع برطرف کنیم. تفسیر هم‌خطی بیزی توسط لیمر [۱۴] ارائه‌شده است. در ساختار بیزی علاقه‌مند به درک مجموعه‌ای از داده‌هایی هستیم که هم‌خطی دارند، و می‌خواهیم بدانیم که هم‌خطی چه اثرهایی بر توزیع پسین می‌گذارد.

۲ آنتروپی و اندازه‌های اطلاع

به‌طور کلّی نظریهٔ اطلاع، مدلی ریاضی از شرایط و عوامل مؤثر بر انتقال و پردازش داده‌ها و اطلاع فراهم می‌آورد. آنتروپی اطلاع که با نام آنتروپی شانون هم شناخته می‌شود، در واقع میزان تصادفی بودن یک رخداد را به‌صورت یک نسخهٔ ریاضی گزارش می‌کند. بنا بر این برای متغیر تصادفی پیوستهٔ X با تابع چگالی f، آنتروپی از رابطهٔ

$$H(X) = -\int_{-\infty}^{\infty} f(x)\log f(x)dx$$

به دست می‌آید. قابل ذکر است که اگر X و Y متغیرهای تصادفی پیوسته با تابع چگالی توأم f باشند، آنتروپی توأم و آنتروپی شرطی از فرمول‌های

$$H(X,Y) = -E_f(\log f(X,Y))$$

$$H(X|Y) = -E_f(\log f(X|Y=y))$$

حاصل می‌شوند.

اگر X و Y از هم مستقل باشند، آن‌گاه

$$H(X,Y) = H(X) + H(Y)$$

و اگر X و Y هم‌توزیع باشند، تساوی $H(X) = H(Y)$ برقرار است. کولبک- لیبلر [۱۱] اندازهٔ اطلاع کولبک- لیبلر را معرفی کردند که این اندازه را آنتروپی نسبی یا اطلاع واگرا هم می‌نامند. چون که این اندازه اختلاف بین دو توزیع از متغیر تصادفی را به دست می‌دهد، دارای اهمیت است. فرض کنید f و g دو تابع چگالی باشند. آنتروپی نسبی f و g به‌صورت زیر تعریف می‌شود:

$$K(f:g) = \int_{-\infty}^{\infty} f(x)\log\frac{f(x)}{g(x)}dx. \qquad (۲)$$

می‌توان ثابت کرد که $K(f:g) \geq 0$ است و تساوی برقرار خواهد بود اگر و فقط اگر $f(x) = g(x)$ باشد. همچنین قابل توجه است که نابرابری $K(f:g) \neq K(g:f)$ برقرار است؛ یعنی آنتروپی کولبک- لیبلر خاصیت تقارن ندارد (کاور و توماس [۳]). حاصل جمع $K(f:g)$ و $K(g:f)$ اطلاع جفریز [۸] را نتیجه می‌دهد که از رابطهٔ

$$J(f:g) = K(f:g) + K(g:f)$$

حاصل می‌شود.

علاوه بر این، آنتروپی کولبک- لیبلر بر حسب آنتروپی شانون به‌صورت

$$K(f:g) = -H(X) - E_f(\log g(X)) \qquad (۳)$$

خواهد بود. به‌عنوان مثال، اگر f و g دو تابع چگالی نرمال، به‌ترتیب با بردارهای میانگین و ماتریس واریانس-کوواریانس عبارت است از: $(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f)$ و $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ باشند، اطلاع کولبک- لیبلر برابر است با

$$\begin{aligned} K(f:g) =& \frac{1}{2}(tr(\boldsymbol{\Sigma}_f\boldsymbol{\Sigma}_g^{-1}) - n) - \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_f|}{|\boldsymbol{\Sigma}_g|} \\ &+ \frac{1}{2}(\boldsymbol{\mu}_f - \boldsymbol{\mu}_g)'\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\mu}_f - \boldsymbol{\mu}_g) \qquad (۴) \end{aligned}$$

که در آن tr اثر ماتریس می‌باشد.

۳ آنتروپی ماکسیمم

کاوش و توسیع بحث آنتروپی از سال ۱۹۴۸ تا کنون تحولات بسیاری را موجب شده است که بحث آنتروپی ماکسیمم یکی از آن‌هاست. آنتروپی ماکسیمم که توسط جینز [۷] مطرح شد، اشاره به این مطلب دارد که در انتخاب توزیع برای متغیر، باید فقط و فقط از اطلاعاتی استفاده کنیم که داریم. به این معنی که باید توزیعی را در نظر بگیریم که تحت اطلاع داده‌شده بیشترین عدم قطعیت یا کمترین تصادفی بودن را داشته باشد؛ چرا که اگر از توزیعی با عدم قطعیت کمتر استفاده کنیم به این معنی است که از اطلاعاتی استفاده کرده‌ایم که به ما داده نشده است، برای محاسبهٔ آنتروپی ماکسیمم شانون برای متغیر تصادفی X لازم است در ابتدا قضیهٔ ۱.۳ مطرح شود که در این قضیه می‌توان توابع f و f' (مشتق تابع چگالی) را به دست آورد.

قضیهٔ ۱.۳. اگر فرض کنیم رابطهٔ

$$I=\int_a^b F(x,f(x),f'(x))dx$$

را داشته باشیم (که در آن F یک تابع معلوم است)، به‌کمک معادلهٔ لاگرانژ اویلر، تابع $f(x)$ ای که I را ماکسیمم می‌کند از تساوی زیر به دست می‌آید (وحیدیان و بزرگ‌نیا [۱]):

$$\frac{\partial F}{\partial f(x)}-\frac{d}{dx}\left(\frac{\partial F}{\partial f'(x)}\right)=0.$$

به‌کمک قضیهٔ ۱.۳، کاگان و همکاران [۹] قضیهٔ زیر را معرفی کردند که با کمک آن می‌توان آنتروپی ماکسیمم شانون را برای متغیر تصادفی X به دست آورد.

قضیهٔ ۲.۳. هرگاه X یک متغیر تصادفی گسسته باشد که به‌ازای هر x متعلق به تکیه‌گاه D، $p(x)>0$ و همچنین توابع $r_i(x)$ برای $i=1,...,k$ طوری باشند که $\sum_{x\in D}p(x)r_i(x)=\alpha_i$ و $\sum_{x\in D}p(x)=1$ و α_i ها ثابت باشند، توزیع دارای آنتروپی ماکسیمم به‌صورت نمایی

$$p^*(x)=e^{\lambda_0+\lambda_1 r_1(x)+\cdots+\lambda_k r_k(x)}$$

است، که $\lambda_0,\lambda_1,...,\lambda_k$ به‌کمک شرایط ذکر شده به دست می‌آیند (کاگان و همکاران [۹]).

چون در آنتروپی ماکسیمم، تابع f' را نداریم، تابع زیر علامت انتگرال تابعی از x و $f(x)$ است و فقط $\frac{\partial F}{\partial f(x)}=0$ باید حل شود.

مثال ۳.۳. آنتروپی ماکسیمم n-متغیرهٔ $\mathbf{X}$ بر حسب محدودیت‌های $E(\mathbf{X})=\boldsymbol{\mu}$ و $E(\mathbf{X}-\boldsymbol{\mu})^2=\boldsymbol{\Sigma}$ به‌صورت زیر است:

$$f^*(\boldsymbol{x})=\left(\frac{1}{\sqrt{2\pi}}\right)^n|\boldsymbol{\Sigma}|^{-\frac{1}{2}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'|\boldsymbol{\Sigma}|^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}.$$

همچنین آنتروپی $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ برابر با

$$H(X)=\frac{n}{2}\log 2\pi e+\frac{1}{2}\log|\boldsymbol{\Sigma}|$$

خواهد بود، که $|\boldsymbol{\Sigma}|$ دترمینان ماتریس کوواریانس است و برآورد درست‌نمایی $\boldsymbol{\Sigma}$ با استفاده از چگالی آنتروپی ماکسیمم عبارت است از

$$\widehat{\boldsymbol{\Sigma}}=\frac{1}{n}(\mathbf{X}-\widehat{\boldsymbol{\mu}})'(\mathbf{X}-\widehat{\boldsymbol{\mu}}).$$

قابل ذکر است که $\widehat{\boldsymbol{\mu}}$ برآورد درست‌نمایی $\boldsymbol{\mu}$ در چگالی آنتروپی ماکسیمم است.

۴ تشخیص هم‌خطی در مدل رگرسیونی خطی با استفاده از نظریهٔ اطلاع

حال می‌خواهیم با کمک تعریف آنتروپی و اطلاع کولبک-لیبلر، هم‌خطی در مدل رگرسیونی را تشخیص دهیم. برای این منظور باید توزیع پسین ضرایب را داشته باشیم. اگر در مدل رگرسیونی با مسئلهٔ هم‌خطی مواجه باشیم، مدل رگرسیونی خطی را دوباره به‌صورت زیر بازنویسی می‌کنیم:

$$\boldsymbol{Y}=(\mathbf{X}\boldsymbol{\Gamma})(\boldsymbol{\Gamma}'\boldsymbol{\beta})+\varepsilon=\mathbf{W}\boldsymbol{\alpha}+\varepsilon, \qquad (5)$$

که $\boldsymbol{\Gamma}=[\gamma_1,...,\gamma_p]$ ماتریس متعامد از ویژه‌بردارهای ماتریس $\mathbf{X}'\mathbf{X}$ و $\mathbf{W}'\mathbf{W}$ ماتریس قطری از ویژه‌مقدارهای (λ_j) ماتریس $\mathbf{X}'\mathbf{X}$ است. از این‌جا به بعد، ماتریس $\mathbf{W}'\mathbf{W}$ را با نماد $\boldsymbol{\Lambda}$ نشان می‌دهیم. با فرض این‌که داده‌ها و σ^2 (واریانس خطاها) معلوم باشند، توزیع پسین پارامتر β عبارت است از

$$\Pi_{\mathbf{X}}(\boldsymbol{\beta}|y,\sigma^2)=N(\boldsymbol{b},\sigma^2(\mathbf{X}'\mathbf{X})^{-1}).$$

بنا بر این اندازهٔ عدم اطمینان یا همان آنتروپی $\boldsymbol{\beta}$ به‌وسیلهٔ توزیع پسین از رابطهٔ

$$H_{\mathbf{X}}(\boldsymbol{\beta}|\boldsymbol{y},\sigma^2)=\frac{p}{2}\log(2\pi e\sigma^2)-\log|\mathbf{X}'\mathbf{X}|^{\frac{1}{2}} \qquad (6)$$

به دست می‌آید. مقادیر بزرگ $H(\boldsymbol{\beta}|y)$ نشان‌دهندهٔ این است که داده‌ها شامل اندازهٔ کوچکی از اطلاع پارامتر β هستند. بنا بر این اثر هم‌خطی ضرایب رگرسیونی بر مبنای

$$I_X(\boldsymbol{\beta}|\sigma^2)=-H_X(\boldsymbol{\beta}|y,\sigma^2)$$

تعیین می‌شود. در نتیجه

$$I_{\mathbf{X}}(\boldsymbol{\beta}|\sigma^2)=\sum_{j=1}^{p}\log\lambda_j^{\frac{1}{2}}-\frac{p}{2}\log(2\pi e). \qquad (7)$$

اگر $\mathbf{X}^\circ$ نشان‌دهندهٔ ماتریس متعامد و

$$\Pi_{\mathbf{X}^\circ}(\boldsymbol{\beta}|y,\sigma^2)=N(b^\circ,\sigma^2\boldsymbol{I})$$

باشد، اثر هم‌خطی ضرایب رگرسیونی عبارت است از

$$I_{\mathbf{X}^\circ}(\boldsymbol{\beta}|\sigma^2)=\frac{p}{2}\log(2\pi e)+\frac{p}{2}\log(\sigma^2). \qquad (8)$$

با توجه به $I_{\mathbf{X}}(\boldsymbol{\beta}|\sigma^{۲})$ و $I_{\mathbf{X}^{\circ}}(\boldsymbol{\beta}|\sigma^{۲})$ اندازهٔ اطلاع $IL_{\mathbf{X}}(\boldsymbol{\beta}|\sigma^{۲})$ به‌صورت زیر تعریف می‌شود:

$$\begin{aligned} IL_{\mathbf{X}}(\boldsymbol{\beta}|\sigma^{۲}) &= I_{\mathbf{X}^{\circ}}(\beta|\sigma^{۲}) - I_{\mathbf{X}}(\beta|\sigma^{۲}) \\ &= -\log|\mathbf{X}'\mathbf{X}|^{\frac{۱}{۲}} \\ &= -\sum_{j=۱}^{p}\log\lambda_j^{\frac{۱}{۲}}. \end{aligned} \tag{۹}$$

این اندازهٔ اطلاع برای تشخیص هم‌خطی مفید است؛ چون بر حسب ویژه‌مقدارهای ماتریس $\mathbf{X}'\mathbf{X}$ است (بلسلی و همکاران [۲]). قابل ذکر است که ویژه‌مقدارهای کوچک نشان‌دهندهٔ این است که بین متغیرهای پیشگو وابستگی خطی زیادی وجود دارد. یکی دیگر از راه‌های تشخیص هم‌خطی با استفاده از نظریهٔ اطلاع، با کمک اطلاع کولبک- لیبلر است. برای این منظور فرض کنید

$$f_{\mathbf{X}}(\boldsymbol{b}|\boldsymbol{\beta},\sigma^{۲}) = N(\boldsymbol{\beta},\sigma^{۲}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}})$$
$$f_{\mathbf{X}^{\circ}}(\boldsymbol{b}^{\circ}|\boldsymbol{\beta},\sigma^{۲}) = N(\boldsymbol{\beta},\sigma^{۲}\boldsymbol{I}_n)$$

باشد.

آن‌گاه اطلاع کولبک-لیبلر بین $f_{\mathbf{X}}(\boldsymbol{b}|\boldsymbol{\beta},\sigma^{۲})$ و $f_{\mathbf{X}^{\circ}}(\boldsymbol{b}^{\circ}|\boldsymbol{\beta},\sigma^{۲})$ عبارت است از

$$\begin{aligned} K(\boldsymbol{b}:\boldsymbol{b}^{\circ}) &= K\{f_{\mathbf{X}}(\boldsymbol{b}|\boldsymbol{\beta},\sigma^{۲}) : f_{\mathbf{X}^{\circ}}(\boldsymbol{b}^{\circ}|\boldsymbol{\beta},\sigma^{۲})\} \\ &= \frac{۱}{۲}\{tr(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} - \log|\mathbf{X}'\mathbf{X}|^{-\mathbf{1}} - \mathbf{p}\} \\ &= \frac{۱}{۲}(\sum_{j=۱}^{p}\frac{۱}{\lambda_j} + \sum_{j=۱}^{p}\log\lambda_j - p), \end{aligned} \tag{۱۰}$$

که $tr(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}$ و $|\mathbf{X}'\mathbf{X}|$ برای تشخیص هم‌خطی مفید هستند (چون بر حسب ویژه‌مقدارهای ماتریس $\mathbf{X}'\mathbf{X}$ هستند). در نتیجه $K(\boldsymbol{b}:\boldsymbol{b}^{\circ})$ اندازهٔ زیان اطلاع ناشی از ماتریس غیر متعامد $\mathbf{X}$ است، همچنین

$$K(\boldsymbol{b}:\boldsymbol{b}^{\circ}) \geq ۰$$

است و تساوی رخ خواهد داد اگر و فقط اگر ماتریس $\mathbf{X}$ متعامد باشد. اگر $K(\boldsymbol{b}:\boldsymbol{b}^{\circ}) \to \infty$ (وقتی $\lambda_j \to \infty$)، بین متغیرهای پیشگو هم‌خطی وجود دارد. اگر در فرضیات مسئله توزیع پیشین ضرایب رگرسیونی داده شده باشد می‌توانیم اثر هم‌خطی را به دست آوریم.

حال اگر فرض کنیم توزیع پیشین ضرایب رگرسیونی برابر با $\Pi(\boldsymbol{\beta}|\boldsymbol{m},\tau^{۲}) = N(\boldsymbol{m},\tau^{۲}\boldsymbol{I}_p)$ باشد، آنتروپی پیشین ضرایب رگرسیونی به‌صورت زیر به دست می‌آید:

$$H(\boldsymbol{\beta}|\tau^{۲}) = \frac{p}{۲}\log(۲\pi e) + \frac{p}{۲}\log\tau^{۲} \tag{۱۱}$$

و همچنین فرض می‌کنیم که توزیع پسین $(\Pi(\boldsymbol{\beta}|\boldsymbol{y},\sigma^{۲},\boldsymbol{m},\tau^{۲}))$ ضرایب رگرسیونی برابر با $N(\boldsymbol{b}(\phi,\boldsymbol{m}),\sigma^{۲}(\mathbf{X}'\mathbf{X}+\phi\boldsymbol{I}_{\mathbf{p}})^{-\mathbf{1}})$ باشد، که

$$\boldsymbol{b}(\phi,\boldsymbol{m}) = (\mathbf{X}'\mathbf{X}+\phi\boldsymbol{I}_{\mathbf{p}})^{-\mathbf{1}}(\mathbf{X}'\boldsymbol{Y}+\phi\boldsymbol{m}), \tag{۱۲}$$

و $\phi = \frac{\sigma^{۲}}{\tau^{۲}}$ نسبت دقت مدل است. اگر در رابطهٔ (۱۲)، $m = ۰$ باشد، ضرایب رگرسیونی مدل ستیغی به دست خواهد آمد. در نتیجه

$$\begin{aligned} H_{\mathbf{X}}(\boldsymbol{\beta}|\boldsymbol{y},\sigma^{۲},\tau^{۲}) =& \frac{p}{۲}\log(۲e) + \frac{p}{۲}\log(\sigma^{۲}) \\ & - \log|\phi I_p + \mathbf{X}'\mathbf{X}|^{\frac{۱}{۲}}. \end{aligned} \tag{۱۳}$$

علاوه بر این، اثر هم‌خطی ضرایب رگرسیونی با استفاده از توزیع پسین و پیشین از رابطهٔ ذیل به دست می‌آید (لیندلی [۱۵]):

$$\begin{aligned} \vartheta_{\mathbf{X}}(\boldsymbol{\beta}|\phi) &= H(\boldsymbol{\beta}|\tau^{۲}) - H_{\mathbf{X}}(\boldsymbol{\beta}|\boldsymbol{y},\sigma^{۲},\tau^{۲}) \\ &= \log|\boldsymbol{I}_p + \phi^{-۱}\mathbf{X}'\mathbf{X}|^{\frac{1}{2}}, \end{aligned} \tag{۱۴}$$

که رابطهٔ (۱۴) با توجه به نسبت $\phi = \dfrac{\sigma^{۲}}{\tau^{۲}}$ حاصل می‌شود.

در بخش بعدی، هم‌خطی را با کمک آنتروپی ماکسیمم تعمیم‌یافته برطرف می‌کنیم.

۵ آنتروپی ماکسیمم تعمیم‌یافته

آنتروپی ماکسیمم تعمیم‌یافته[۴] بر اساس رویکرد آنتروپی ماکسیمم کلاسیک جینز [۷] است و واضح است که در چارچوب آنتروپی ماکسیمم کلاسیک، گشتاورها دقیق فرض شده‌اند. آنتروپی ماکسیمم تعمیم‌یافته در اقتصاد و به‌ویژه در مدل‌های خطی به کار می‌رود که با مسئلهٔ هم‌خطی روبه‌رو هستند. این رویکرد به‌وسیلهٔ گولان و همکاران [۴] توسعه پیدا کرده است. آنتروپی ماکسیمم تعمیم‌یافته بیشتر در مدل‌های پارامتری‌شده به کار می‌رود. بنا بر این می‌خواهیم $\boldsymbol{\beta}$ و ε (با فرض مستقل بودن خطاها) را از روش GME تحت مدل رگرسیونی

$$\boldsymbol{Y}_{T\times۱} = \mathbf{X}_{T\times K}\boldsymbol{\beta}_{K\times۱} + \varepsilon_{T\times۱}$$

به دست آوریم. در این مدل، $\boldsymbol{\beta}$ ضریب رگرسیونی و ε عبارت خطا می‌باشند. فرض کنید z_k برداری $M\times۱$ برای پارامتر $\boldsymbol{\beta}_k$

[۴] maximum generalized entropy

تعریف شده باشد و همچنین اعضای بردار z_k از کوچک ($z_{k۱}$) به بزرگ (z_{kM}) مرتب شده باشند. آن‌گاه برای هر پارامتر $\boldsymbol{\beta}_k$ بردار احتمال $\boldsymbol{p}_k$، ($\sum_{m=۱}^{M}\boldsymbol{p}_{km}=۱$)، وجود دارد به قسمی که

$$\boldsymbol{\beta}_k=\sum_{m=۱}^{M}\boldsymbol{p}_{km}\boldsymbol{z}_{km},\quad k=۱,...,K,\quad M\geq ۲.$$

به‌طور مشابه فرض می‌کنیم $\boldsymbol{v}_t$ برداری $J\times ۱$ برای خطاها تعریف شده باشد و برای هر ε_t بردار احتمال $\boldsymbol{w}_t$، ($\sum_{j=۱}^{J}\boldsymbol{w}_{tj}=۱$)، وجود داشته باشد. بنا بر این تساوی زیر برقرار است:

$$\varepsilon_t=\sum_{j=۱}^{J}\boldsymbol{w}_{tj}\boldsymbol{v}_j,\quad t=۱,...,T,\quad J\geq ۲.$$

قابل ذکر است که $\boldsymbol{\beta}_k$ و ε_k متغیرهای تصادفی پیوسته‌اند.

برای منابع عدم قطعیت مستقل، آنتروپی جمع‌پذیر است (برای جزئیات بیشتر به کاپور و کواسان [۱۰] مراجعه شود). با فرض این‌که بردارهای احتمال $\boldsymbol{p}_k$ و $\boldsymbol{w}_t$ مستقل از یکدیگر باشند، می‌توان با حل مسئلهٔ بهینه‌سازی آنتروپی ماکسیمم، پارامترها و خطاهای مجهول مدل رگرسیونی را برآورد کرد. بنا بر این روش آنتروپی ماکسیمم تعمیم‌یافته به‌صورت زیر است:

$$\begin{aligned}&argMax_{P,W}(H(P)+H(W))=\\&\qquad argMax_{P,W}\Big\{-\sum_{k=۱}^{K}\sum_{m=۱}^{M}\boldsymbol{p}_{km}\log\boldsymbol{p}_{km}\\&\qquad -\sum_{t=۱}^{T}\sum_{j=۱}^{J}\boldsymbol{w}_{tj}\log\boldsymbol{w}_{tj}\Big\},\end{aligned}$$

به قسمی که

$$\begin{aligned}\boldsymbol{y}_t=&\sum_{k=۱}^{K}\sum_{m=۱}^{M}\boldsymbol{z}_{km}\boldsymbol{p}_{km}\boldsymbol{x}_{tk}\\&+\sum_{j=۱}^{J}v_j\boldsymbol{w}_{tj},\quad t=۱,...,T,\end{aligned}$$

و

$$\sum_{m=۱}^{M}\boldsymbol{p}_{km}=۱,\quad k=۱,...,K,$$
$$\sum_{j=۱}^{J}\boldsymbol{w}_{tj}=۱,\quad t=۱,...,T.$$

برای تعیین کران‌های تکیه‌گاه خطاها، گولان و همکاران [۹] از قانون ۳σ که توسط پوکلس‌هایم [۱۶] مطرح شده بود استفاده کردند. آن‌ها کران پایین را -۳σ قرار داده و برای کران بالا از $+۳\sigma$ استفاده کرده‌اند. مقدار σ نیز همان انحراف معیار نمونه‌ای بردار (Y) است. در نتیجه مدل رگرسیونی به‌صورت

$$\boldsymbol{Y}=\mathbf{X}\boldsymbol{\beta}+\varepsilon=\mathbf{X}\boldsymbol{Z}\boldsymbol{P}+\boldsymbol{\nu}\boldsymbol{W},$$

است. بنا بر این احتمالات برآورد شدهٔ برای $\boldsymbol{\beta}$ و ε به‌صورت زیر خواهد بود:

$$\widehat{\boldsymbol{p}}_{km}=\frac{exp(-\boldsymbol{z}_{km}\sum_{t=۱}^{T}\boldsymbol{x}_{tk}\widehat{\boldsymbol{\lambda}}_t)}{\sum_{m=۱}^{M}exp(-\boldsymbol{z}_{km}\sum_{t=۱}^{T}\boldsymbol{x}_{tk}\widehat{\boldsymbol{\lambda}}_t)},$$
$$m=۱,...,M,\quad k=۱,...,K,$$

$$\widehat{\boldsymbol{w}}_{tj}=\frac{exp(-\boldsymbol{v}_{tj}\widehat{\boldsymbol{\lambda}}_t)}{\sum_{j=۱}^{J}exp(-\boldsymbol{v}_j\widehat{\boldsymbol{\lambda}}_t)},\quad j=۱,...,J,\quad t=۱,...,T.$$

در نهایت، مقادیر برآورد شدهٔ $\boldsymbol{\beta}$ و ε برابر خواهد بود با

$$\widehat{\boldsymbol{\beta}}_k\equiv\sum_{m=۱}^{M}\boldsymbol{z}_{km}\widehat{\boldsymbol{p}}_{km},$$
$$\widehat{\varepsilon}_t\equiv\sum_{j=۱}^{J}\boldsymbol{v}_j\widehat{\boldsymbol{w}}_{tj}.$$

همچنین λ_tها ضرایب لاگرانژ می‌باشند که از رابطهٔ زیر به دست می‌آیند:

$$\begin{aligned}\lambda_t=&Min\Big\{\sum_{t=۱}^{T}y_t\boldsymbol{\lambda}_t+\sum_{k=۱}^{K}\log\Big(\sum_{m=۱}^{M}exp(-\boldsymbol{z}_{km}\\&\times\sum_{t=۱}^{T}\boldsymbol{\lambda}_t\boldsymbol{x}_{tk})\Big)+\sum_{t=۱}^{T}\log\Big(\sum_{j=۱}^{J}exp(-\boldsymbol{\lambda}_t\boldsymbol{v}_j)\Big)\Big\}.\end{aligned}$$

همان‌طور که ملاحظه می‌کنید، برای محاسبهٔ $\widehat{\boldsymbol{\beta}}_k$، $\widehat{\boldsymbol{w}}_j$، $\widehat{\boldsymbol{p}}_{km}$ و $\widehat{\varepsilon}_t$ نیاز به محاسبهٔ $\boldsymbol{\lambda}_t$ ها داریم. اما برای $\boldsymbol{\lambda}_t$ فرم بسته‌ای نداریم؛ بنا بر این برای محاسبهٔ $\widehat{\boldsymbol{\beta}}_k$، $\widehat{\boldsymbol{w}}_j$، $\widehat{\boldsymbol{p}}_{km}$ و $\widehat{\varepsilon}_t$ از روش عددی استفاده می‌کنیم.

آنتروپی ماکسیمم تعمیم‌یافته بر اساس آنتروپی ماکسیمم کلاسیک به دست می‌آید؛ با این تفاوت که هیچ محدودیتی بر ضرایب و جملهٔ خطای مدل رگرسیونی چندگانه وجود ندارد و هر دو پارامتر را با توجه به جمع‌پذیر بودن آنتروپی به‌روش عددی محاسبه می‌کنند. در روش آنتروپی ماکسیمم، محدودیت‌ها به‌طور دقیق، تنها برای یک پارامتر ذکر می‌شوند. اگر در روش آنتروپی ماکسیمم تعمیم‌یافته جداگانه برای هر یک از پارامترها محدودیت‌هایی مشخص شود، آنتروپی ماکسیمم ضرایب و خطای مدل رگرسیونی به دست می‌آیند.

مثال ۱.۵. داده‌های جداول ۱ و ۲ مربوط به میزان محصول (بر حسب کیلوگرم) زمین‌های زراعی در سه سال متفاوت در فصل

بهار است. هدف از جمع‌آوری این داده‌ها یافتن رابطه بین عواملی مانند میزان بارندگی ($X_۱$)، دمای هوا ($X_۲$)، دمای خاک ($X_۳$) و رطوبت هوا ($X_۴$)است، که کشاورزان در آن دخیل نیستند.

در این مثال، مدل رگرسیونی عبارت است از

$$\begin{aligned} Y =& ۱۳۵/۹۸۰۵۶ + ۲۳/۲۰۶۲۸X_۱ \\ &+ ۷۲/۱۸۷۸۷X_۲ + ۱۵/۲۳۲۹۸X_۳ \\ &- ۲۱/۱۸۶۴۷X_۴. \end{aligned} \tag{۱۵}$$

با استفاده از نرم‌افزار SAS مسئلهٔ هم‌خطی بین داده‌ها را بررسی کردیم و جدول همبستگی را به دست آوردیم (جدول ۳).

در جدول ۳ می‌بینیم که بین $X_۲$ و $X_۴$ یک رابطهٔ کاملاً خطی وجود دارد؛ یعنی همبستگی بین $X_۲$ و $X_۴$ برابر با یک است، که این نشان‌دهندهٔ وجود هم‌خطی می‌باشد. بنا بر این با توجه به جدول ضرایب همبستگی می‌فهمیم که دمای خاک، تابع خطی از دمای هوا است. همچنین $K(\boldsymbol{b} : \boldsymbol{b}^{\circ}) = ۶/۸۹$ می‌باشد (با استفاده از نرم‌افزار R). می‌خواهیم با توجه به نظریهٔ اطلاع، مدل رگرسیونی چندگانه را برای این داده‌ها به دست آوریم. اکنون با کمک آنتروپی ماکسیمم تعمیم‌یافته و با استفاده از نرم‌افزار SAS، ضرایب مدل رگرسیونی را برآورد می‌کنیم.

در نتیجه مدل رگرسیونی چندگانه به‌صورت زیر است:

$$\begin{aligned} Y =& ۰/۴۷۴۵۶۸ + ۰/۹۰۶۵۳۱X_۱ \\ &+ ۸۱/۳۳۷۷۶X_۲ + ۰/۱۴۷۴۵۸X_۳ \\ &+ ۴۷/۹۸۸۲۲. \end{aligned}$$

در جدول ۴ (جدول برآورد پارامترها)، ضرایب و خطاهای مدل رگرسیونی برآورد شده و p-مقدار هر یک از ضرایب رگرسیونی، کمتر از ۰.۰۵ است، که نتیجه می‌شود، همهٔ ضرایب، معنی‌دار هستند.

۶ نتیجه‌گیری

مسئلهٔ هم‌خطی یک مشکل جدی در مبحث رگرسیون است؛ زیرا با وجود هم‌خطی در مدل رگرسیونی، ضرایب مدل به‌درستی برآورد نمی‌شوند و دقت مدل کاهش می‌یابد. یکی از راه‌های برطرف کردن مشکل هم‌خطی، به کارگیری مدل رگرسیونی ستیغی می‌باشد. همچنین می‌توان با کمک نظریهٔ اطلاع، از جمله آنتروپی ماکسیمم تعمیم‌یافته، مشکل هم‌خطی را برطرف کرد. در این روش، خطاها و ضرایب رگرسیونی به‌روش‌های عددی به دست می‌آیند.

جدول ۱. داده‌های میزان محصول زمین‌های زراعی

شمارهٔ مشاهده	X_1	X_2	X_3	X_4	Y
۱	۲۲/۶۸	۱۶/۱۷	۷۱/۹۱	۲۴/۲۵	۲۴۱۹/۳۲
۲	۱۴/۴۶	۱۴/۳۴	۷۶/۰۸	۲۱/۵۱	۲۲۰۴/۵۴
۳	۱۲/۷۷	۱۵/۱۸	۶۹/۰۵	۲۲/۷۷	۲۰۹۲/۶۲
۴	۲۷/۷۳	۱۵/۳۷	۷۶/۸۶	۲۳/۰۵	۲۵۶۶/۰۶
۵	۱۲/۹۷	۱۵/۵۰	۷۴/۲۵	۲۳/۲۵	۲۲۰۱/۴۸
۶	۲۷/۷۴	۱۵/۴۸	۷۷/۳۱	۲۳/۲۲	۲۶۱۷/۵۸
۷	۱۶/۱۴	۱۴/۱۹	۷۳/۲۱	۲۱/۲۹	۲۱۷۸/۳۲
۸	۱۹/۵۲	۱۶/۵۷	۸۰/۳۳	۲۴/۸۵	۲۴۸۳/۶۹
۹	۲۵/۹۵	۱۴/۹۱	۷۵/۷۴	۲۲/۳۶	۲۴۷۸/۸۹
۱۰	۱۹/۳۳	۱۴/۹۲	۷۲/۲۲	۲۲/۳۷	۲۲۸۱/۹۲
۱۱	۲۱/۴۳	۱۴/۶۰	۷۳/۱۹	۲۱/۸۹	۲۳۲۳/۰۲
۱۲	۳۰/۵۴	۱۵/۹۵	۷۲/۶۸	۲۳/۹۳	۲۵۹۴/۰۲
۱۳	۱۵/۸۶	۱۵/۲۳	۷۵/۵۶	۲۲/۸۳	۲۲۸۵/۱۸
۱۴	۳۳/۲۷	۱۵/۴۷	۷۴/۲۰	۲۳/۲۱	۲۶۷۴/۷۳
۱۵	۳۳/۸۷	۱۶/۰۱	۷۶/۱۹	۲۴/۰۱	۲۷۳۰/۳۳
۱۶	۲۵/۷۳	۱۴/۸۱	۷۴/۸۶	۲۲/۲۱	۲۴۹۱/۷۷
۱۷	۱۹/۱۳	۱۴/۷۲	۷۳/۰۹	۲۲/۰۷	۲۲۸۹/۲۳
۱۸	۲۷/۸۵	۱۵/۶۶	۷۷/۱۸	۲۳/۴۸	۲۵۸۷/۹۵
۱۹	۱۲/۷۳	۱۵/۲۲	۷۳/۱۲	۲۲/۸۳	۲۱۶۰/۶۴
۲۰	۲۴/۴۶	۱۵/۳۶	۷۴/۵۶	۲۳/۰۳	۲۴۶۰/۷۴
۲۱	۲۰/۹۷	۱۶/۱۲	۷۵/۴۷	۲۴/۱۸	۲۴۱۹/۹۶
۲۲	۲۰/۴۲	۱۵/۰۵	۷۱/۲۲	۲۲/۵۸	۲۳۰۰/۲۰
۲۳	۱۶/۶۹	۱۵/۲۸	۷۲/۶۰	۲۲/۹۲	۲۲۴۶/۱۷
۲۴	۲۳/۸۰	۱۵/۶۹	۷۱/۷۵	۲۳/۵۴	۲۴۲۲/۹۰
۲۵	۲۴/۲۶	۱۵/۳۵	۷۵/۵۱	۲۳/۰۱	۲۴۷۸/۶۴
۲۶	۲۷/۵۷	۱۵/۶۴	۷۴/۲۸	۲۳/۴۷	۲۵۲۵/۹۷
۲۷	۲۴/۶۸	۱۵/۳۵	۷۰/۹۷	۲۳/۰۲	۲۴۱۴/۲۶
۲۸	۳۱/۷۷	۱۷/۳۵	۷۱/۵۶	۲۶/۰۲	۲۶۵۲/۸۰
۲۹	۲۰/۴۳	۱۴/۹۲	۷۰/۱۴	۲۲/۳۷	۲۲۷۹/۹۹
۳۰	۲۳/۷۱	۱۴/۸۹	۷۵/۵۳	۲۲/۳۵	۲۴۳۴/۷۱
۳۱	۱۷/۹۳	۱۵/۱۳	۷۳/۹۳	۲۲/۷۱	۲۲۹۳/۴۰

جدول ۲ (ادامه). داده‌های میزان محصول زمین‌های زراعی

شمارهٔ مشاهده	$X_۱$	$X_۲$	$X_۳$	$X_۴$	Y
۳۲	۲۷/۲۲	۱۵/۳۷	۷۸/۴۵	۲۳/۰۵	۲۵۶۵/۸۳
۳۳	۲۲/۴۶	۱۶/۳۰	۷۳/۳۵	۲۴/۴۶	۲۴۳۹/۰۹
۳۴	۱۸/۶۸	۱۵/۱۶	۷۴/۴۵	۲۲/۷۴	۲۳۱۶/۶۹
۳۵	۱۹/۷۶	۱۶/۲۷	۷۷/۲۹	۲۴/۴۲	۲۴۲۲/۹۰
۳۶	۱۵/۲۱	۱۶/۰۸	۷۵/۴۷	۲۴/۱۱	۲۲۹۰/۱۸
۳۷	۱۸/۳۶	۱۵/۱۵	۷۳/۷۳	۲۲/۷۳	۲۳۱۸/۰۸
۳۸	۲۸/۶۹	۱۶/۲۲	۷۳/۰۱	۲۴/۳۳	۲۵۷۷/۶۲
۳۹	۲۳/۶۲	۱۶/۱۰	۷۲/۴۴	۲۴/۱۴	۲۴۳۹/۰۰
۴۰	۲۲/۸۱	۱۵/۸۲	۷۴/۱۴	۲۳/۷۳	۲۴۲۶/۸۰
۴۱	۲۸/۵۰	۱۶/۵۴	۶۹/۲۵	۲۴/۸۰	۲۵۱۶/۸۷
۴۲	۱۶/۳۳	۱۶/۳۶	۷۹/۵۹	۲۴/۵۳	۲۳۷۷/۸۹
۴۳	۱۶/۲۷	۱۵/۸۲	۷۶/۱۵	۲۳/۷۲	۲۳۲۲/۹۹
۴۴	۲۴/۱۲	۱۵/۷۳	۷۲/۰۸	۲۳/۶۰	۲۴۲۶/۸۱
۴۵	۲۹/۰۶	۱۷/۱۴	۷۷/۷۸	۲۵/۷۰	۲۶۷۹/۲۶

جدول ۳. همبستگی مثال ۱

متغیر	$X_۱$	$X_۲$	$X_۳$	$X_۴$	Y
$X_۱$	۱/۰۰۰۰	۰/۳۸۵۹	۰/۰۵۰۴	۰/۳۸۶۱	۰/۹۴۲۸
$X_۲$	۰/۳۸۵۹	۱/۰۰۰۰	۰/۱۵۸۸	۱/۰۰۰۰	۰/۵۵۴۰
$X_۳$	۰/۰۵۰۴	۰/۱۵۸۸	۱/۰۰۰۰	۰/۱۵۸۲	۰/۳۲۸۰
$X_۴$	۰/۳۸۶۱	۱/۰۰۰۰	۰/۱۵۸۲	۱/۰۰۰۰	۰/۵۵۴۰
Y	۰/۹۴۲۸	۰/۵۵۴۰	۰/۳۲۸۰	۰/۵۵۴۰	۱/۰۰۰۰

جدول ۴. برآورد پارامترهای مثال ۱

متغیر	برآورد پارامتر	انحراف معیار	مقدار t	$Pr > \|t\|$
$X_۱$	۰/۹۰۶۵۳۱	۰/۱۳۵۳	۶/۷۰	$<$ ۰/۰۰۰۱
$X_۲$	۸۱/۳۳۷۷۶	۰/۶۱۶۸	۱۳۱/۸۸	$<$ ۰/۰۰۰۱
$X_۳$	۰/۱۴۷۴۵۸	۰/۰۱۷۹	۸/۲۴	$<$ ۰/۰۰۰۱
$X_۴$	۴۷/۹۸۸۲۲	۰/۳۶۴۰	۱۳۱/۸۳	$<$ ۰/۰۰۰۱
عرض از مبدأ	۰/۴۷۴۵۶۸	۰/۰۴۵۳	۱۰/۴۷	$<$ ۰/۰۰۰۱

مراجع

[۱] وحیدیان و بزرگ‌نیا (۱۳۷۲). نظریهٔ کنترل و کنترل بهینه. دانشگاه فردوسی مشهد.

[2] Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Observation and Sources of Collinearity*. Wiley, New York.

[3] Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley, New York.

[4] Golan, A., Judge, G., and Miller, J. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Wiley, New York.

[5] Hartley, R. V. L. (1928). Transmission of information. *Bell System Technical Journal*, **7**, (**3**), 535-563.

[6] Horel, A. E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, **58**, 54-59.

[7] Jaynes, E. T. (1968). On the rational of maximum entropy methods. *Proceedings of Institute of Electrical and Electronics Engineers*, **70**, 939-952.

[8] Jeffreys, H. (1948). An invariant form for the prior probability in estimating problems. *Proceedings of the Royal Society of London*, **186**, 453-461.

[9] Kagan, A. M., Linnik, Y. V., and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York.

[10] Kupur, J. N. and Kesavan, H. K. (1992). *Entropy Optimization Principles with Applications*. Academic Press, London.

[11] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79-86.

[12] Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001). *Introduction to Linear Regression Analysis*. Wiley, New York.

[13] Nyquist, H. (1924). Certain factors affecting telegragh speed. *Bell System Technical Journal*, **3**, 126-147.

[14] Leamer, E. E. (1973). Multicollinearity: A Bayesian interpretation. *Review of Economics and Statistics*, **55**, 371-380.

[15] Lindley, D. V. (1956). On a measure of information provided by an experiment. *Annals of Mathematical Statistics*, **27**, 986-1005.

[16] Pukelsheim, F. (1994). The three-sigma rule. *The American Statistician*, **4**, 88-91.

[17] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-423.