محاسبه مؤلفه‌های اصلی و مقایسه آن با اندازه‌های وابستگی مبتنی بر آنتروپی

جواد قاسمیان[۱] علی عمیدی[۲]

چکیده

در این مقاله پس از معرفی اندازه‌های وابستگی تصادفی با جمع‌آوری داده‌های مربوط به شاخصهای توسعه کشاورزی ابتدا مؤلفه‌های اصلی را بر اساس ماتریس کواریانس و همبستگی به دست آورده و متغیرهای تأثیرگذار را مشخص می‌کنیم. سپس با کمینه کردن اندازه وابستگی مبتنی بر آنتروپی نیز تعداد دلخواه متغیر مؤثر را مشخص و بعد داده‌ها را کاهش می‌دهیم. نتایج مقایسه همخوانی قابل قبولی را نشان می‌دهد.

واژه‌های کلیدی : آنتروپی، نظریه اطلاع، اندازه وابستگی، خطای رده‌بندی غلط.

۱. مقدمه

اگر $X_1, X_2, \ldots, X_p$، p مؤلفه یک بردار تصادفی X و $f(x_1, x_2, \ldots, x_p)$ تابع چگالی توأم این p مؤلفه و $f_i(x_i)$ها، $i = 1,2,\ldots,p$ توابع چگالی احتمال حاشیه‌ای آنها باشند، در صورتی که این p مؤلفه بطور تصادفی مستقل باشند، آنگاه توزیع توأم، اطلاعی بیشتر از توزیعهای p مؤلفه به ما نمی‌دهد ولی در حالتی که مؤلفه‌ها وابسته باشند، اطلاعی اضافه درباره درجه وابستگی بین p مؤلفه را می‌توان بر اساس تعریف اطلاع دوطرفه که توسط کولبک[۳] (۱۹۶۸) ویل و فاییگ[۴] (۱۹۸۴)، صوفی[۵] (۱۹۹۴) به صورت زیر معرفی شد، بدست آورد.

$$D = \int \ldots \iint f(x_1,\ldots,x_p) \ln \frac{f(x_1,\ldots,x_p)}{f_1(x_1)\ldots f_p(x_p)} dx_1 \ldots dx_p \qquad (1)$$

واتاناب[۶] (۱۹۸۴) اندازه فوق را قدرت ساختار[۷] نامید. بوزدگان[۸] (۱۹۹۰) و هریس[۹] (۱۹۷۸) رابطه (۱) را برای اندازه گیری وابستگی بین متغیرهای تصادفی به کار بردند. بدین ترتیب شاخص دیگری برای اندازه گیری ضریب همبستگی ρ که کاپور و داند[۱۰] (۱۹۸۶) شش ضعف زیررا برای آن بیان کرده بودند، پیدا شد.

---

[۱] گروه آمار، دانشکده ریاضی، دانشگاه علوم پایه دامغان

[۲] گروه آمار، دانشگاه شهید بهشتی

[۳] Kullback

[۴] Theil and Fiebig

[۵] Soofi

[۶] Watanabe

[۷] Strength of Structure

[۸] Bozdogan

[۹] Harris

[۱۰] Kapur and Dhande

- اگر P متغیر تصادفی داشته باشیم واندازه‌ای تک از وابستگی بین آنها بخواهیم، تعداد ۲/ $p(p-1)$ ضریب همبستگی، چنین اندازه‌ای را به ما نمی‌دهد.
- در بعضی موارد حتی وقتی که متغیرها مستقل نباشند، ρ صفرمی‌شود.
- حتی با وجود اینکه متغیرها کاملاً همبسته باشند، شاید ρ یک نباشد.
- برای تبدیل غیر خطی از متغیرها ρ تغییر می‌کند.
- وابستگی بین دو متغیر کیفی در یک جدول توافقی را نمی‌توان بر حسب ضریب همبستگی بیان کرد.
- در بعضی موارد شاید ρ وجود نداشته باشد.

کاپور و کساوان[11] (۱۹۹۲) یک اندازه نظری اطلاع، و کیم[12] (۱۹۹۷) نیز اندازه جدید نرمال شده‌ای از وابستگی تصادفی بین مجموعه‌ای از متغیرهای تصادفی معرفی کرده‌اند.

۲. اندازه‌های وابستگی تصادفی

اگر $h_1(x)$ و $h_2(x)$ دو تابع چگالی احتمال باشند، آنگاه اندازه کولبک – لایبلر[13] آنتروپی متقاطع(یا فاصله نامتقارن کولبک – لایبلر) ایندو به صورت زیر تعریف می‌شود:

$$D(h_1 : h_2) = \int h_1(x) \ln \frac{h_1(x)}{h_2(x)} dx \qquad (۲)$$

که $D(h_1 : h_2) \geq 0$ و صفر است هرگاه $h_1 = h_2$. $D(h_1 : h_2)$ یک تابع محدب از x است. با فرض اینکه $X_1, X_2, \dots, X_p$، p متغیر تصادفی با چگالی‌های حاشیه‌ای و چگالی توأم تعریف شده دربخش قبل بوده و $g(x_1,\dots,x_p) = \prod_{i=1}^{p} f_i(x_i)$ حاصلضرب توابع چگالی احتمال حاشیه‌ای آنها باشد، واتاناب (۱۹۶۹) اندازه وابستگی زیر را تعریف کرد؛

$$D(f : g) = \int \dots \iint f(x_1,\dots,x_p) \ln \frac{f(x_1,\dots,x_p)}{g(x_1,\dots,x_p)} d x_1 \dots d x_p \qquad (۳)$$

با فرض استقلال دامنه تغییرات متغیرهای تصادفی، رابطه (۳) را می‌توان به صورت زیر نوشت،

$$D(f : g) = -s + \sum_{i=1}^{p} s_i \qquad (۴)$$

که در آن

$$s = \int \dots \iint f(x_1,\dots,x_p) \ln f(x_1,\dots,x_p)\, d x_1 \dots d x_p .$$

آنتروپی توزیع توأم $X_1, X_2, \dots, X_p$ و $s_i = -\int f_i(x_i) \log f_i(x_i)\, d x_i$ آنتروپی توزیع حاشیه‌ای X_i $(i = 1,\dots,p)$ است.

اندازه‌های دیگر مهم وابستگی تصادفی تعریف شده توسط کاپور(۱۹۸۵ الف - ۱۹۸۵ ب) به صورت زیرند،

(۵)

$$D_1^* = \int \dots \iint f_1(x_1) f_2(x_2) \dots f_p(x_p) \times \ln \frac{f(x_1) f(x_2) \dots f_p(x_p)}{f(x_1,\dots,x_p)} dx_1 \dots dx_p$$

$$D_2^* = \frac{1}{\alpha - 1} \left\{ \int f^{\alpha}(x_1,\dots,x_p) \times \left[f_1(x_1) \dots f_p(x_p) \right]^{1-\alpha} dx_1 \dots dx_p - 1 \right\}$$
$$\alpha > 0, \alpha \neq 1 \qquad (۶)$$

$$D_3^* = \frac{1}{\alpha - 1} \ln \left(\int f^{\alpha}(x_1,\dots,x_p) \times \left[f_1(x_1) \dots f_p(x_p) \right]^{1-\alpha} dx_1 \dots dx_p \right)$$
$$\alpha > 0, \alpha \neq 1 \qquad (۷)$$

همه این اندازه‌ها غیر منفی‌اند و فقط و فقط وقتی صفر می‌شوند که متغیرها مستقل باشند. برای مطالعه اندازه‌های دیگر می‌توان به کاپور و داند(۱۹۹۰) مراجعه کرد.

۳. اندازه‌های نرمال شده وابستگی تصادفی

اندازه‌های نظری اطلاع بحث شده می‌توانند مقادیری در بازه $[0,\infty)$ بگیرد، لذا نیازمند اندازه‌های نرمال شده‌ای هستیم که مقادیری در بازه $[0,1]$ بگیرد. تعدادی از اندازه‌های نرمال شده وابستگی موجود به صورت زیرند.

کامارگو و اسرائیل[14] (۱۹۵۶)

$$\overline{r}_0 = \frac{s_1 + s_2 - s}{s_1 + s_2}$$

راجسکی[15] (۱۹۶۱)

$$\overline{D} = \frac{s_1 + s_2 - s}{s}$$

---

[11] Kapur and Kesavan

[12] Kim

[13] Kullback and Leibler

[14] Camargo and Israel

کاپور (۱۹۸۶)

$$\overline{D}_1 = \frac{\sum_{i=1}^{p} s_i - s}{(p-1)s}$$

$$\overline{D}_2 = \frac{m}{m-1} \times \frac{\sum_{i=1}^{p} s_i - s}{\sum_{i=1}^{p} s_i}$$

این اندازه‌ها بین صفر و یک واقع‌اند . یک برای وابستگی کامل و صفر برای استقلال کامل در نظر گرفته می‌شود .برای اهداف کاربردی اندازه وابستگی $D(f:g)$ را به صورتهای مختلف نرمال می‌کنند یکی از مهمترین این اندازه‌ها بر اساس الگوی کاپور و داند (۱۹۹۰) توسط کیم (۱۹۹۷) به صورت زیر معرفی شد،

$$D_3 = 1 - e^{-D(f:g)}$$

که آن را با $DI(f:g)$ [16] نشان می‌دهیم.

٤. اندازه وابستگی برای توزیع نرمال چند متغیره

آنتروپی متقاطع یا اطلاع دوطرفه برای m زیرمجموعه یک بردار تصادفی با p مؤلفه که زیرمجموعه i ام ، p_i متغیر دارد به صورت زیر داده می‌شود،

$$D(f:g) = \frac{1}{2}\ln\frac{\Pi_i|\Sigma_{ii}|}{|\Sigma|} \quad (8)$$

$$= -\frac{1}{2}\ln\frac{|R|}{\prod_{i=1}^{m}|R_{ii}|} \quad (9)$$

که در آن $f(x\,|\,\mu,\Sigma)$ چگالی توأم و $g(x\,|\,\mu,\Sigma) = \prod_i g_i(x_i\,|\,\mu_i,\Sigma_{ii})$ حاصلضرب چگالی‌های کناری X_i ، $\Sigma_{ii} = Var(X_i)$ ، $i = 1,2,\ldots,m$ و R ماتریس همبستگی است .(جانسون و ویچرن[17] را ببینید.)

بر این اساس شاخص وابستگی به صورت زیر تعریف می شود.

$$DI = 1 - e^{-D} = 1 - \left[\frac{|R|}{\prod_{i=1}^{m}|R_{ii}|}\right]^{\frac{1}{2}} \qquad R_{ii} > 0 \,. \quad (10)$$

یکی از خواص جالب DI در نرمال چندمتغیره پایا بودن آن نسبت به تبدیل خطی از هر مجموعه از متغیرهاست.

۵. کاربرد اندازه وابستگی

فرض کنید یک آزمودنی را بتوان با تعداد زیادی از مشخصه‌های $X_1,\ldots,X_n$ که بابردار

$$X = (X_1, X_2, \ldots, X_n)^T \quad (11)$$

نشان داده می‌شوند توصیف کرد .

متغیر تصادفی X را به عنوان بردار الگو می‌شناسیم وفرض این است که توزیع احتمال آن را می‌دانیم . در عمل وقتی بردار X دارای ۲۰ مؤلفه یا بیشتر باشد، به دلیل حجم زیاد، کار با داده‌ها خاصه در بحث رده‌بندی[18] مشکل است، لذا ابتدا تبدیلی از بردار تصادفی X در یک فضای n بعدی به بردار Y در یک فضای m بعدی را در نظر می گیریم، به طوری که $m << n$.

می‌توان یک تبدیل خطی به صورت

$$Y = AX \quad (12)$$

در نظر گرفت،که $X = (X_1,\ldots,X_n)^T$, $Y = (Y_1,\ldots,Y_m)^T$ و A ماتریسی از مرتبه $m \times n$ است.

به دلیل اینکه $m << n$ است، لذا کار با بردار Y ساده‌تر از کار با بردار X است . ولی این ساده کردن برای ما هزینه دارد و به واسطه تبدیل تکین (۱۲)، بعضی از اطلاعات را از دست می‌دهیم و در نتیجه توان تمایز بردارهایی را نیز که می‌توانند منجر به افزایش احتمال رده‌بندی غلط (PEM) شوند، از دست می‌دهیم.

مسأله انتخاب A برای کمینه کردن (PEM) را، مسأله استخراج کیفیت خاص[19] یا مسأله کاهش بعد می گویند. به بردارهای Y ، بردارهای با کیفیت خاص[20] گویند. برای استخراج کیفیت خاص، باید (PEM) را به عنوان تابعی از A بیان کنیم، که کار مشکلی است. بنابراین به جای کمینه

---

[15] Rajski

[16] Dependence Index

[17] Johnson and Wichern

[18] Classification

[19] Problem of feature extraction

[20] Feature vectors

کردن (PEM) ، توابع دیگری را که با (PEM) رابطه نزدیک دارند و می‌توان آنها را به راحتی بر حسب A ردیابی کرد کمینه می‌کنیم.

مخصوصاً می‌خواهیم اندازه وابستگی بین $Y_1,...,Y_m$ را کمینه کنیم، زیرا استقلال بیشتر این متغیرها افزایش کمتر (PEM) را موجب خواهد شد یا A را به قسمی انتخاب می‌کنیم که از دست رفتن اطلاعات کمینه باشد. فرض کنید $f(y_1,...,y_m)$ تابع چگالی توأم بردارهای $(Y_1,...,Y_m)$ و $g_i(y_i)$ برای $i=1,...,m$ تابع چگالی حاشیه‌ای هر بردار Y_i باشد، پس

(۱۳)

$$\begin{aligned} D(f:g) &= \int\cdots\iint f(y_1,...,y_m)\ln\frac{f(y_1,...,y_m)}{g_1(y_1)...g_m(y_m)}dy_1...dy_m \\ &= \int\cdots\iint f(y_1,...,y_m)\ln f(y_1,...,y_m)dy_1...dy_m - \\ &\int g_1(y_1)\ln g_1(y_1)dy_1 - \cdots - \int g_m(y_m)\ln g_m(y_m)dy_m \\ &= -s+\sum_{i=1}^{m}s_i \end{aligned}$$

که در آن s و s_i در رابطه (٤) صدق می‌کنند.

سپس A را به نحوی انتخاب می‌کنیم که D را کمینه کند. می‌توان نشان داد که ماتریس زیر چنین ویژگی را دارد(کاپور و کساوان ۱۹۹۲ را ببینید).

$$(W_1,W_2,...,W_m)^T \qquad (١٤)$$

که $W_1,W_2,...,W_m$ ویژه بردارهای یکه متناظر با m مقدار بزرگ ویژه مقدار ماتریس همبستگی یا کواریانس بردار الگوی X هستند.

توضیح با یک مثال:

داده‌های مربوط به شاخصهای کشاورزی برای ۲٦ استان را در سال ۱۳۷۵ با استفاده از داده های خام اعلام شده توسط مرکز آمار ایران و وزارت جهاد کشاورزی محاسبه کرده‌ایم[21]. متغیرهای مورد نیاز به صورت زیرند،

x_1 : محصول گندم (هزار تن)

x_2 : درصد شاغلین بخش کشاورزی

x_3 : سرانه تولید محصولات کشاورزی (تن)

x_4 : درصد سطح زیر کشت ناحیه آبی به سطح کل آبی و دایمی

x_5 : مقدار محصول در هر هکتار (عملکرد) (تن)

x_6 : درصد سطح زیر کشت محصولات صنعتی (پنبه _ توتون _ تنباکو _ چغندر _ دانه های روغنی و نیشکر)

فرض کنید $X^T=(x_1,...,x_6)$ بردار الگوی شاخصهای توسعه کشاورزی باشد ، ویژه مقدارهای محاسبه شده برای ماتریسهای کواریانس S و ضریب همبستگی R به صورت زیرند،

$$L_R=\begin{pmatrix} 2/84 \\ 1/19 \\ 0/82 \\ 0/55 \\ 0/35 \\ 0/24 \end{pmatrix} \qquad L_S=\begin{pmatrix} 145673 \\ 857 \\ 80 \\ 33 \\ 24 \\ 8 \end{pmatrix}$$

اگر این شاخصها به صورت دسته‌جمعی مورد استفاده قرار گیرند، نمی‌توان تصور قابل درکی از تشخیص کیفیت خاص بین استانهای مختلف به دست آورد. برای مطالعه راحت‌تر ، این داده‌ها را به دو صورت زیر با بعد ($m=2$) تلخیص می‌کنیم؛

الف) به دست آوردن بردارهای کیفیت خاص به دو روش با

۱- استفاده از ماتریس کواریانس که افت اطلاعات مینیمم می‌شود.

۲- استفاده از ماتریس همبستگی به منظور کمینه شدن وابستگی متغیرهای جدید.

۱- در این روش A را برای کمینه کردن PEM انتخاب می‌کنیم که این کار یک مسأله استخراج کیفیت خاص یا کاهش بعد در شناخت الگوست. ابتدا دو مقدار بزرگتر ویژه مقدارهای ماتریس کواریانس را انتخاب و ویژه بردارهای متعامد متناظر را به دست آورده و A را با این دو بردار به صورت سطری به دست می‌آوریم. بنابراین

$$A=\begin{bmatrix} -1 & -0/013 \\ -0/004 & 0/149 \\ -0/004 & -0/108 \\ 0/013 & -0/970 \\ 0/003 & -0/124 \\ -0/001 & -0/098 \end{bmatrix}$$

لذا بردارهای کیفیت خاص که افت اطلاعات را کمینه می‌کنند به صورت زیرند،

$$\begin{aligned} y_1 &= -x_1-0/004x_2-0/004x_3+0/013x_4 \\ &\quad +0/003x_5-0/001x_6 \\ y_2 &= -0/013x_1+0/149x_2-0/108x_3-0/970x_4 \\ &\quad -0/124x_5-0/098x_6 \end{aligned}$$

---

[21] در صورت نیاز به اطلاعات با نویسنده اول تماس بگیرید.

Ghasemian@dus.ac.ir

۲- به طور مشابه A را به وسیله اختیار کردن ویژه بردارهای متعامد متناظر با دو ویژه مقدار بزرگتر ماتریس همبستگی به صورت سطری به دست می‌آوریم. بنابراین

$$A=\begin{bmatrix} -۰/۰۸۵ & ۰/۸۶۵ \\ -۰/۴۶۷ & ۰/۰۶۹ \\ ۰/۳۹۶ & ۰/۴۲ \\ ۰/۴۷۳ & -۰/۱۳۲ \\ ۰/۵۲۸ & -۰/۱۲۳ \\ ۰/۳۴ & ۰/۱۹۵ \end{bmatrix}$$

و بردارهای کیفیت خاص که وابستگی تعمیم یافته را کمینه می‌کنند به صورت زیرند،

$$z_۱=-۰/۰۸۵x_۱-۰/۴۶۷x_۲+۰/۳۹۶x_۳+۰/۴۷۳x_۴ +۰/۵۲۸x_۵+۰/۳۴x_۶$$

$$z_۲=۰/۸۶۵x_۱+۰/۰۶۹x_۲+۰/۴۲x_۳-۰/۱۳۲x_۴ -۰/۱۲۳x_۵+۰/۱۹۵x_۶$$

حال ممکن است این سؤال پیش آید که کدامیک از این جفت بردارها مناسب‌ترند؟ جواب این است که وقتی شاخصها در بخشهای متفاوت و با واحدهای گوناگون اندازه‌گیری می‌شوند استفاده از ماتریس همبستگی بر ماتریس کواریانس ارجحیت دارد. مقادیر مربوط به استانهای مختلف را می‌توان بر اساس $y_۱, y_۲, z_۱, z_۲$ محاسبه کرد.

ب) در این روش با فرض اینکه $X=(X_۱,...,X_۶)$ یک بردار تصادفی نرمال باشد، دوتا از شش خاصیت را می‌توانیم به ۱۵ طریق مختلف انتخاب کنیم. آنتروپی $H(X)$ و آنتروپی‌های $H(Y_i)$، $i=۱,...,۱۵$ را برآورد کرده و افت اطلاع یعنی $H(X)-H(Y_i)$ را محاسبه می‌کنیم. بعد از محاسبه افت اطلاع جفتی را که افت را کمینه کند انتخاب می‌کنیم. در این حالت شاخصهای $x_۱, x_۴$ کمترین افت اطلاع را دارند.

اگر R ماتریس همبستگی شاخصها باشد آنگاه

$$D=-\frac{۱}{۲}\ln|R|$$

D را برای ۱۵ انتخاب محاسبه و زوجی که D را کمینه می‌کند، برگزیده‌ایم. شاخصهای $x_۱, x_۶$ حداقل اندازه وابستگی D را نتیجه داده‌اند.

۶. نتیجه‌گیری

- بردارهای کیفیت خاص $y_۱, y_۲$ وزنهای بیشتر برای $x_۱, x_۴$ دارند و لذا به سمت شاخصهای با واریانس بزرگ کشیده می‌شوند. بنابراین تعجبی ندارد که $x_۱, x_۴$ را به عنوان شاخصهای اصلی که افت اطلاع را کمینه می‌کنند به دست آوریم.
- بردارهای کیفیت خاص $z_۱, z_۲$ به ترتیب با $x_۵, x_۱$ به دست می‌آیند. ولی شاخصهای $x_۱, x_۶$ حداقل وابستگی با D را نتیجه می‌دهند.

مراجع

[1] Bickel, P.J. and Doksum, K.A., 1977, *Mathematical Statistics,* San Francisco, Holden day Inc.

[2] Bozdogan, H., 1990, *On Information Based Measure of Covariance Complexity and its Application to the Evaluation of Multivariate Linear Models,* Communications in Statistics- Theory and Methods, 19, pp. 221-278.

[3] Camargo, C.M. and Israel, 1956, *Logarithmic Index of Correlation,* Annals Real Soc. Espan, Fits Y, quim, 52A, 117.

[4] Harris, C.J., 1978, *An Information Theoric Approach to Estimation,* In recent Theorical Developments in Control, ed. by M.J. Gregson, London, Academic Press.

[5] Johnson, R.A. and Wichren, D.W., 1988, *Applied Multivariate Statistical Analysis,* New Jersey, Prentice Hall Inc.

[6] Kapur, J.N., 1985 a, *New Measures of Stochastic Dependence,* IIT/Kapur Res.rep., No. 243.

[7] Kapur, J.N., 1985 b, *Normalized Measures of Stochastic Dependence,* IIT/Kapur Res.rep., No. 279.

[8] Kapur, J.N. and Dhande, M., 1986, *On the Entropic Measures of Stochastic Dependence,* Indian Journal of Pure and Applied Mathematics, 17(5), pp. 581-595.

[9] Kapur, J.N. and Dhande, M., 1990, *On a Family of Normalized Measures of Interdependence,* Acta Ciendica, Vol. XVI,M, 2, pp. 193-198.

[10] Kapur, J.N. and Kesavan, H.J., 1992, *Entropy Optimization Principles with Applications,* New York, Academic Press.
[11] Kim, Hea-Jung, 1997, *On Information Theoretic Index for Measuring the Stochastic Dependence Among Sets of Varieties,* Journal of Korean Stat. Society, Vol.26(1), pp. 131-146.
[12] Kullback, S., 1968, *Information Theory and Staistics,* New York, Dover Publications.
[13] Rajski, C., 1961, *On the Normed Information Rate of Discrete random Variables,* Trans. Third Pague Conf. Inf. Th. Stat. Dec. Funcs. Random. Proc. pp. 583-585
[14] Soofi, E.S.,1994, *Capturing the Intangible Concept of Information,* Journal of American Statistical of Association, 89, pp. 1243-1254.
[15] Theil, H. and Fiebig, D.G., 1984, *Exploiting Continuty; Maxmimum Entropy Estimation of Continuous distribution,* Massachuset, Bellinger Publishing Company, Cambrige.
[16] Watanabe, S., 1969, *Knowing and Guessing,* New York, John Wiley & Sons.
[16] Watanabe, S., 1985, *Pattern Recognition: Human and Mechanical,* New York, John Wiley & Sons.