مجله علوم آماری، پاییز و زمستان ۱۳۸۹

جلد ۴، شماره ۲، ص ۱۹۳–۲۰۹

بهبود کران بالای معیار کولبک–لیب‌لر بر اساس ترکیب محدب k مدل رقیب

عبدالرضا سیاره

گروه آمار، دانشگاه رازی کرمانشاه

تاریخ دریافت: ۱۳۸۹/۵/۱۰ تاریخ آخرین بازنگری: ۱۳۸۹/۱۲/۱۷

چکیده : یکی از مسائل اساسی در استنباط آماری، انتخاب مدل بهینه از میان مدل‌های رقیب است. در این مقاله ثابت شده است که خطای نسبی بین دو مدل دارای خاصیت زبرجمعی است و با استفاده از آن نشان داده شده است که ترکیب محدب (مدل آمیخته) k مدل رقیب از نظر معیار واگرایی کولبک–لیب‌لر، مدلی را ایجاد می‌کند که یا بهتر از تمام k مدل رقیب است و یا لااقل از دورترین مدل رقیب به مدل درست داده‌ها، بهتر است. بررسی شبیه‌سازی صحت یافته‌های نظری را تائید می‌کنند.

واژه‌های کلیدی : انتخاب مدل، مدل آمیخته، خطای نسبی، ترکیب محدب، معیار کولبک–لیب‌لر، میانگین هندسی.

آدرس الکترونیک مسئول مقاله: عبدالرضا سیاره، asayyareh@razi.ac.ir

کد موضوع‌بندی ریاضی (۲۰۱۰): ۲۶D۱۵

۱ مقدمه

انتخاب مدل آماری برای داده‌ها، یکی از موضوعات اساسی دراستنباط آماری است. یک مدل آماری توصیف پدیده‌های تصادفی را با توجه به آگاهی‌هایی در مورد آن پدیده‌ها ساده‌تر می‌سازد و مفهوم مفیدی برای معنا بخشیدن به مشاهداتی است که توسط یک یا چند متغیر تصادفی تولید شده‌اند. در واقع انتخاب مدل برآورد توانایی مدل‌های مختلف به منظور انتخاب بهترین آنها است. نمونهٔ تصادفی n تایی $Y_1, \ldots, Y_n$ را در نظر بگیرید. تحت فرض استقلال و هم‌توزیعی Y_iها، فرض کنید $H(\cdot)$ توزیع درست و $h(\cdot)$ چگالی یا مدل درست داده‌ها باشد. معمولاً چگالی مولد داده‌ها بسیار پیچیده است، به طوری که با اطلاعات موجود قابل توصیف نیست. بنابراین در یک موقعیت نااطمینانی قرار داریم و باید در مورد $h(\cdot)$ تصمیم بگیریم. دو روش معمول برای اتخاذ تصمیم در مورد $h(\cdot)$، برآورد یا تقریب آن بر پایهٔ اطلاعات موجود است. به منظور برآورد $h(\cdot)$ خانواده‌ای پارامتری به صورت

$$\mathcal{G} = \{g(\cdot, \beta) : \mathcal{R} \to \mathcal{R}^+; \beta \in B \subset \mathcal{R}^d\} = (g^\beta(\cdot))_{\beta \in B}$$

در نظر گرفته می‌شود. این گونه مدل‌های آماری بندرت بر مدل مولد داده‌ها منطبق خواهند شد، اما به دلایل مختلف مدل‌های رقیب (پیشنهادی) به کار گرفته می‌شوند. در این انتخاب این امکان وجود دارد که مدل پارامتری رقیب به مدل درست نزدیک باشد، اما بر آن منطبق نباشد. لذا اندکی اطلاعات از دست داده خواهد شد. بنابراین انتخاب و ارزیابی یک مدل رقیب بخش مهمی از استنباط آماری است. از لحاظ عملی نیاز به مدلی داریم که بتوان آن را به کار گرفت. به عبارتی این مدل باید بهترین و ساده‌ترین تقریب برای $h(\cdot)$ باشد. بهترین تقریب در واقع کمترین فاصله بین مدل درست و مدل رقیب را به ذهن متبادر می‌کند. در مباحث احتمال و نظریه اطلاع، معیار واگرایی کولبک- لیب‌لر ($\mathcal{KL}$) (کولبک-لیب‌لر، ۱۹۶۸ و ۱۹۵۱)، یک اندازهٔ واگرایی بین توزیع احتمال درست و توزیع احتمال دلخواه است. فرض کنید $p = (p_1, \ldots, p_n)$ و $q = (q_1, \ldots, q_n)$ دو تابع احتمال گسسته باشند، لذا معیار $\mathcal{KL}$

به صورت

$$\mathcal{KL}(p,q)=\sum_{i=1}^{n}p_i\log(\frac{p_i}{q_i})\geq 0 \qquad (1)$$

تعریف می‌شود، که در آن تساوی رخ خواهد داد اگر و تنها اگر به ازای هر i در مجموعه $\{1,\dots,n\}$، تساوی $p_i=q_i$ برقرار باشد. با توجه به این که $\mathcal{KL}(\cdot,\cdot)\geq 0$ است، تعیین یک کران بالای مناسب برای این معیار واگرایی سبب خواهد شد تا مدل رقیب مناسب‌تری انتخاب شود. معیار χ^2 نیز معیار دیگری برای بررسی فاصله دو مدل است که توسط پیرسن به صورت

$$\mathcal{D}_{\chi^2}=\sum_{\omega\in A(\mu)\cup A(v)}\{\frac{(\mu(\omega)-\nu(\omega))^2}{\nu(\omega)}\}$$

تعریف شده است. برای مطالعه بیشتر در مورد این معیار می‌توان به لیز و وجدا (۱۹۸۷) مراجعه نمود. کوور و توماس (۱۹۹۱) کران پایین معیار $\mathcal{KL}$ را مورد بررسی قرار داده و نشان دادند که این معیار بزرگ‌تر یا مساوی معیار تغییرات کل است و کران پایین معیار $\mathcal{KL}$ را بهبود بخشیدند. دراگومایر و گلوژویک (۲۰۰۱) نامساوی‌هایی را برای معیارهای $\mathcal{KL}$ و χ^2 گسترش داده و برای بررسی اطلاع توأم به‌کار بردند. دراگومایر (۲۰۰۸) به بررسی خاصیت زبرجمعی[1] تابعی از $\mathcal{KL}(p,\cdot)$ به شکل $\exp[-\mathcal{KL}(p,\cdot)]$ پرداخت.

یکی از مسائل حل نشده در نظریه انتخاب مدل، انتخاب از بین k مدل رقیب غیرآشیانی است. برای $k=2$ به وونگ (۱۹۸۹) و کومانز و همکاران (۲۰۰۸) مراجعه شود. سیاره و ترکمان (۱۳۸۸) به بررسی برآورد تفاضل مخاطره‌های کولبک-لیبلر برای مشاهدات سانسوریده در مدل‌های غیرآشیانی پرداختند. سیاره و همکاران (۲۰۱۱) و سیاره (۲۰۱۱) معیارهای منتج از معیار کولبک- لیبلر و آزمون‌های انتخاب مدل را در مدل‌های کلاسیک و خطی مورد بررسی قرار دادند. در این مقاله سعی شده است تا نشان داده شود از نظر معیار کولبک-لیبلر تحت شرایطی مدل آمیخته حاصل از مدل‌های رقیب بهتر از هر یک از k مدل رقیب به کار برده شده در ترکیب محدب است. لزوم بررسی این مسئله از آن جهت حائز

[1] Superadditive

اهمیت است که گاهی اوقات مدل‌های رقیب را به دلایلی نمی توان کنار گذاشت. زیرا ممکن است هر یک از آن‌ها به دلایل نظری یا عملی از ویژگی‌های خاصی برخوردار باشند و ترجیح یکی بر دیگری از دلایل کافی برخوردار نباشد. لذا در این مقاله مدل آمیخته مدل‌های رقیب پیشنهاد شده است و به صورت نظری و به کمک شبیه‌سازی این ایده مورد بررسی قرار گرفته است. البته این که کدام k مدل به عنوان مدل‌های رقیب انتخاب شوند خود یک مسئله قابل بررسی است.

در بخش ۲ تعاریف و قضایای مورد نیاز ارائه شده‌اند. در بخش ۳ خاصیت زیرجمعی $\exp[-\mathcal{KL}(p,\cdot)]$ که نقش اساسی در سایر بخش‌های این مقاله دارد آورده شده است و به بررسی خطای نسبی و رابطه آن با $\exp[-\mathcal{KL}(p,\cdot)]$ پرداخته شده است. در بخش ۴ مدل آمیخته حاصل از ۲ و k مدل رقیب در نظر گرفته شده و نشان داده شده است که این مدل براساس معیار کولبک-لیبلر یا از هر یک از مدل‌های رقیب در مدل آمیخته بهتر است و یا لااقل از دورترین مدل رقیب به مدل درست نزدیک‌تر است. همچنین نشان داده شده است که در حالت اخیر مدل آمیخته را می‌توان به عنوان یک انتخاب مینیماکس در نظر گرفت. در بخش ۵ مطالعه شبیه‌سازی برای بررسی نتایج نظری آورده شده است.

۲ تعاریف و قضایا

در این بخش به بررسی تعاریف و قضایای مورد نیاز پرداخته شده است. برای مطالعه بیشتر در مورد این قضایا به دراگومایر (۲۰۰۸) مراجعه شود.

تعریف ۱ : میانگین هندسی وزنی: برای nـتایی از اعداد حقیقی غیر منفی $a=(a_1,\ldots,a_n)$ و توزیع احتمال $p=(p_1,\ldots,p_n)$ میانگین هندسی وزنی به صورت زیر تعریف می‌شود.

$$G_n(p,a)=\prod_{i=1}^{n} {a_i}^{p_i}.$$

فرع ۱ : خاصیت زبرجمعی میانگین هندسی: اگر میانگین هندسی وزنی $G_n(p,\cdot)$ را تابعی از متغیر دوم در نظر بگیریم، آنگاه

$$G_n(p, a+b) \geq G_n(p,a) + G_n(p,b).$$

فرع ۲ : خاصیت یکنوایی میانگین هندسی وزنی: برای $b \geq a$ داریم

$$G_n(p,b) \geq G_n(p,a).$$

فرع ۳ : برای هر مقدار دلخواه m

$$G_n(p, ma) = \prod_{i=1}^{n} (ma_i)^{p_i} = m \prod_{i=1}^{n} {a_i}^{p_i} = mG_n(p,a) \qquad (2)$$

۳ خاصیت زبرجمعی $\exp[-\mathcal{KL}(p,\cdot)]$

اگر سه متغیر تصادفی گسسته X، Y و Z با توزیع احتمال‌های $p = (p_1, \ldots, p_n)$، $q = (q_1, \ldots, q_n)$ و $r = (r_1, \ldots, r_n)$ مفروض باشند، میانگین هندسی وزنی $\frac{q}{p} = (\frac{q_1}{p_1}, \ldots, \frac{q_n}{p_n})$ با وزن‌های $p = (p_1, \ldots, p_n)$ به طوری که به ازای هر $i \in \{1, \ldots, n\}$، $p_i \neq 0$ باشد را به صورت

$$\begin{aligned} G_n(p, \frac{q}{p}) &= \prod_{i=1}^{n} (\frac{q_i}{p_i})^{p_i} \\ &= \exp[\log(\prod_{i=1}^{n} (\frac{q_i}{p_i})^{p_i})] \\ &= \exp[\sum_{i=1}^{n} p_i \log(\frac{q_i}{p_i})] \\ &= \exp[-\mathcal{KL}(p,q)], \end{aligned}$$

در نظر بگیرید. به طور مشابه می‌توان نشان داد

$$G_n(p, \frac{r}{p}) = \exp[-\mathcal{KL}(p, r)].$$

فرع ۴ : میانگین هندسی وزنی $(\frac{q_1+r_1}{p_1}, \ldots, \frac{q_n+r_n}{p_n}) = \frac{q}{p} + \frac{r}{p}$ با وزن‌های $p = (p_1, \ldots, p_n)$ عبارت است از

$$\begin{aligned} G_n(p, \frac{q+r}{p}) &= \prod_{i=1}^{n} (\frac{q_i + r_i}{p_i})^{p_i} \\ &= \exp[\sum_{i=1}^{n} p_i \log(\frac{q_i + r_i}{p_i})] \\ &= \exp[-\mathcal{KL}(p, q + r)]. \end{aligned} \tag{۳}$$

فرع ۵ : با توجه به زبرجمعی بودن میانگین هندسی وزنی، زبرجمعی بودن $\exp[-\mathcal{KL}(p; \cdot)]$ را نیز می توان نتیجه گرفت. بنابراین می توان نوشت

$$\exp[-\mathcal{KL}(p, q + r)] \geq \exp[-\mathcal{KL}(p, q)] + \exp[-\mathcal{KL}(p, r)].$$

فرع ۶ : اگر مقادیر مثبت m و M که $m < M$ موجود باشند به‌طوری که

$$0 < m \leq \frac{r_i}{q_i} \leq M \;\; ,\forall i \in \{1, \ldots, n\},$$

آنگاه

$$G_n(p, m\frac{q}{p}) \leq G_n(p, \frac{r}{p}) \leq G_n(p, M\frac{q}{p}), \tag{۴}$$

با توجه به فرع ۲، $G_n(p, \frac{mr}{p}) = m \exp[-\mathcal{KL}(p; r)]$، بنابراین

$$m \exp[-\mathcal{KL}(p, q)] \leq \exp[-\mathcal{KL}(p, r)] \leq M \exp[-\mathcal{KL}(p, q)]. \tag{۵}$$

برای مطالعه بیشتر در این زمینه و خصوصاً در مورد نامساوی (۵) به دراگومایر (۲۰۰۸) مراجعه شود.

۱.۳ خطای نسبی

تابع مخاطره کولبک-لیبلر مقادیری در فاصله $(\circ,\infty)$ اختیار می‌کند. اما در عمل مخاطرات به طور نسبی و با مقدار کمتر از یک مورد توجه هستند. اگر وقوع پیشامد c تحت دو مدل درست h و مدل رقیب g تحت بررسی باشد، خطای نسبی به صورت $r_c(P_g(c),P_h(c)) = \frac{P_h(c)-P_g(c)}{P_h(c)}$ در نظر گرفته می‌شود. حال اگر پیشامد c به شکل $c=\{x: g(x)<h(x)\}$ تعریف شود می‌توان نشان داد $r_c(P_g(c),P_h(c)) = \sqrt{۱-\exp\{-۲\mathcal{KL}(g,h)\}}$ و در نتیجه $\exp\{-\mathcal{KL}(g,h)\} = \sqrt{۱-r_c^۲(P_g(c),P_h(c))}$. لذا تابعی از خطای نسبی حاصل از استفاده از یک مدل رقیب به جای مدل درست نیز دارای خاصیت زیرجمعی است. از این خاصیت خطای نسبی استفاده می‌شود تا معیاری برای انتخاب مدل به دست آید. تحلیل دقیقی از خطای نسبی در کومانژ و همکاران (۲۰۰۸) آمده است.

۴ ترکیب محدب مدل‌های رقیب

با توجه به مقدمات، در این بخش مدل آمیخته‌ای از مدل‌های رقیب معرفی می‌شود که براساس معیار کولبک-لیبلر یا بهتر از هر یک از اجزا تشکیل دهنده مدل آمیخته است و یا از دورترین مدل رقیب به مدل درست نزدیک تر است. دو تابع احتمال گسسته p و r را در نظر بگیرید. اگر $\alpha+\beta=۱$، $۱\leq\alpha\leq\circ$ و $۱\leq\beta\leq\circ$، باتوجه به (۳) نامساوی به صورت

$$\exp[-\mathcal{KL}(p,\alpha q+\beta r)] \geq \exp[-\mathcal{KL}(p,\alpha q)] + \exp[-\mathcal{KL}(p,\beta r)],$$

به دست خواهد آمد. همچنین براساس (۲) می‌توان نامساوی اخیر را به‌صورت

$$\exp[-\mathcal{KL}(p,\alpha q+\beta r)] \geq \alpha\exp[-\mathcal{KL}(p,q)] + \beta\exp[-\mathcal{KL}(p,r)]. \qquad (۶)$$

نوشت. اگر از طرفین (۶) لگاریتم گرفته و در ۱− ضرب شود، نامساوی به‌صورت

$$\mathcal{KL}(p,\alpha q+\beta r) \leq -\log[\alpha\exp[-\mathcal{KL}(p,q)] + \beta\exp[-\mathcal{KL}(p,r)]].$$

به‌دست خواهد آمد. چون $-\log(\cdot)$ تابعی محدب است با توجه به این نکته که اگر $f:R\to R$ تابعی محدب و $x=(x_۱,\dots,x_n)$ یک بردار n- تایی از اعداد حقیقی و $p=(p_۱,\dots,p_n)$ توزیع احتمال باشد، آنگاه

$$\sum_{i=۱}^{n} p_i f(x_i) \geq f(\sum_{i=۱}^{n} p_i x_i).$$

لذا نامساوی به‌صورت

$$-\log\{\alpha\exp[-\mathcal{KL}(p,q)]+\beta\exp[-\mathcal{KL}(p,r)]\} \leq \alpha\mathcal{KL}(p,q)+\beta\mathcal{KL}(p,r). \quad (۷)$$

خواهیم داشت. با استفاده از (۷) نامساوی

$$\mathcal{KL}(p,\alpha q+\beta r) \leq \alpha\mathcal{KL}(p,q)+\beta\mathcal{KL}(p,r), \quad (۸)$$

به‌دست می‌آید. با قرار دادن $(۱-\alpha)$ به جای β در طرف دوم نامساوی (۸) تساوی به صورت

$$\begin{aligned}\alpha\mathcal{KL}(p,q)+(۱-\alpha)\mathcal{KL}(p,r) &= \alpha\sum_{i=۱}^{n} p_i\log(\frac{p_i}{q_i})+(۱-\alpha)\sum_{i=۱}^{n} p_i\log(\frac{p_i}{r_i})\\ &= \sum_{i=۱}^{n} p_i\log(\frac{p_i}{r_i})+\alpha\sum_{i=۱}^{n} p_i\log(\frac{r_i}{q_i}),\end{aligned}$$

خواهیم داشت. بنابراین

$$\mathcal{KL}(p,\alpha q+\beta r) \leq \mathcal{KL}(p,r)+\alpha\sum_{i=۱}^{n} p_i\log(\frac{r_i}{q_i}). \quad (۹)$$

به همین ترتیب می‌توان نشان داد که

$$\mathcal{KL}(p,\alpha q+\beta r) \leq \mathcal{KL}(p,q)+\beta\sum_{i=۱}^{n} p_i log(\frac{q_i}{r_i}) \quad (۱۰)$$

با توجه به برابری $E_p\{\log\frac{q}{r}\}=-E_p\{\log\frac{r}{q}\}$ یکی از این دو امید ریاضی منفی خواهد بود. اگر $E_p\{\log\frac{r}{q}\}$ منفی باشد بنابر (۹)

$$\mathcal{KL}(p,\alpha q+\beta r) \leq KL(p,r) \quad (۱۱)$$

و $E_p\{\log \frac{r}{q}\} > E_p\{\log \frac{q}{r}\}$. این نامساوی ایجاب می‌کنده $\mathcal{KL}(p,q) < \mathcal{KL}(p,r)$.
بنابراین دو نامساوی به‌صورت

$$\mathcal{KL}(p, \alpha q + \beta r) < \mathcal{KL}(p,q) < \mathcal{KL}(p,r) \tag{۱۲}$$

$$\mathcal{KL}(p,q) < \mathcal{KL}(p, \alpha q + \beta r) < \mathcal{KL}(p,r) \tag{۱۳}$$

به‌دست می‌آیند. از (۱۲) نتیجه می‌شود

$$\mathcal{KL}(p, \alpha q + \beta r) < \min \{\mathcal{KL}(p,q), \mathcal{KL}(p,r)\} \tag{۱۴}$$

رابطهٔ (۱۳) نشان می‌دهد که مدل آمیخته حاصل از مدل‌های رقیب مدلی را تولید می‌کند که از مدل دورتر از مدل درست داده‌ها، بهتر است. نتایج حاصل در لم زیر خلاصه شده است.

لم ۱ : اگر $E_p\{\log \frac{r}{q}\} < ۰$ باشد و $\alpha \in [۰,۱]$، $\beta \in [۰,۱]$ و $\alpha + \beta = ۱$ باشد، آنگاه

$$\mathcal{KL}(p, \alpha q + \beta r) \leq \min\{\mathcal{KL}(p,q), \mathcal{KL}(p,r)\}.$$

در غیر این صورت براساس (۱۳) مدل آمیخته بهتر از دورترین مدل از مدل درست است. به عبارتی

$$\mathcal{KL}(p, \alpha q + (۱-\alpha) r) \leq \max\{\mathcal{KL}(p,q), \mathcal{KL}(p,r)\}$$

برهان : اثبات از نتایجی که تاکنون حاصل شده‌اند بدیهی است.

۱.۴ ترکیب محدب k مدل رقیب

در این بخش مدل آمیخته حاصل از k مدل رقیب در نظر گرفته شده است و نشان داده خواهد شد که این مدل از نظر معیار کولبک–لیب‌لر یا از مدلی با کمترین فاصله

از مدل درست بهتر است یا بهتر از دورترین مدل از مدل درست در مجموعه k مدل رقیب است.

لم ۲ : اگر توزیع های احتمال $q_j = \{q_{j1}, \ldots, q_{jn}\}$ برای $j \in \{1, \ldots, k\}$ مفروض باشند، فرض کنید به ازای هر $j \in \{1, \ldots, k\}$، $0 \leq \alpha_j \leq 1$ و $j \neq \ell$، مدل رقیب q_ℓ وجود دارد به طوری که $E_p\{\log \frac{q_\ell}{q_j} < 0\}$ است. آنگاه

$$\mathcal{KL}(p, \sum_{j=1}^{k} \alpha_j q_j) \leq \min\{\mathcal{KL}(p, q_j)\}.$$

در غیر این صورت مدل آمیخته از دورترین مدل به مدل درست داده‌ها، بهتر است.

برهان : ابتدا برقراری نامساوی

$$\mathcal{KL}(p, \sum_{j=1}^{k} \alpha_j q_j) \leq \sum_{j=1}^{k} \alpha_j \mathcal{KL}(p, q_j) \tag{15}$$

را بررسی می‌کنیم. بنا بر (۸)

$$\begin{aligned}
\mathcal{KL}(p, \sum_{j=1}^{k} \alpha_j\, q_j) &= \mathcal{KL}(p, \sum_{j=1}^{k-1} \alpha_j q_j + \alpha_k\, q_k) \\
&\leq \mathcal{KL}(p, \sum_{j=1}^{k-1} \alpha_j\, q_j) + \alpha_k \mathcal{KL}(p, q_k) \\
&\leq \mathcal{KL}(p, \sum_{j=1}^{k-2} \alpha_j\, q_j) + \alpha_{k-1} \mathcal{KL}(p, q_{k-1}) + \alpha_k \mathcal{KL}(p, q_k) \\
&\ \ \vdots \\
&\leq \sum_{j=1}^{k} \alpha_j \mathcal{KL}(p, q_j).
\end{aligned}$$

با بسط طرف دوم نامساوی (۱۵) برای $j \neq \ell$ خواهیم داشت

$$\sum_{j=1}^{k} \alpha_j \mathcal{KL}(p, q_j) = \alpha_\ell \mathcal{KL}(p, q_\ell) + \sum_{j=1}^{k} \alpha_j \mathcal{KL}(p, q_j)$$

$$
\begin{aligned}
&= \alpha_\ell \sum_{i=1}^{n} p_i \log(\frac{p_i}{q_{\ell i}}) + \sum_{j=1}^{k} \alpha_j \sum_{i=1}^{n} p_i \log(\frac{p_i}{q_{ji}}) \\
&= \alpha_\ell \sum_{i=1}^{n} p_i \log(p_i) - \alpha_\ell \sum_{i=1}^{n} p_i \log(q_{\ell i}) + \sum_{j=1}^{k} \alpha_j \sum_{i=1}^{n} p_i \log(p_i) \\
&- \sum_{j=1}^{k} \alpha_j \sum_{i=1}^{n} p_i \log(q_{ji}) \\
&= \alpha_\ell \sum_{i=1}^{n} p_i \log(p_i) - (1 - \sum_{j=1}^{k} \alpha_j) \sum_{i=1}^{n} p_i \log(q_{\ell i}) \\
&+ (1 - \alpha_\ell) \sum_{i=1}^{n} p_i \log(p_i) - \sum_{j=1}^{k} \alpha_j \sum_{i=1}^{n} p_i \log(q_{ji}) \\
&= \alpha_\ell \sum_{i=1}^{n} p_i \log(p_i) - \sum_{i=1}^{n} p_i \log(q_{\ell i}) + \sum_{j=1}^{k} \alpha_j \sum_{i=1}^{n} p_i \log(q_{\ell i}) \\
&+ \sum_{i=1}^{n} p_i \log(p_i) - \alpha_\ell \sum_{i=1}^{n} p_i \log(p_i) - \sum_{j=1}^{k} \alpha_j \sum_{i=1}^{n} p_i \log(q_{ji}) \\
&= \sum_{i=1}^{n} p_i \log(\frac{p_i}{q_{\ell i}}) + \sum_{j=1}^{k} \alpha_j \sum_{i=1}^{n} p_i \log(\frac{q_{\ell i}}{q_{ji}})
\end{aligned}
$$

در نتیجه

$$\sum_{j=1}^{k} \alpha_j \mathcal{KL}(p, q_j) = \mathcal{KL}(p, q_\ell) + \sum_{j=1}^{k} \alpha_j \sum_{i=1}^{n} p_i \log(\frac{q_{\ell i}}{q_{ji}}).$$

باجایگذاری در رابطه (۱۵) نامساوی

$$\mathcal{KL}(p, \sum_{j=1}^{k} \alpha_j q_j) \leq \mathcal{KL}(p, q_\ell) + \sum_{j=1}^{k} \alpha_j \sum_{i=1}^{n} p_i \log(\frac{q_{\ell i}}{q_{ji}}), \qquad (۱۶)$$

به دست خواهد آمد. از طرفی چون $k(k-1)/2$ جمله از مجموع‌های سمت راست نامساوی (۱۶) منفی هستند و $(k-1)$ جمله از آن‌ها دارای صورت کسرهای مساوی‌اند، می توان نتیجه گرفت که $\ell \in \{1, \dots, k\}$ وجود دارد به طوری که

$$\mathcal{KL}(p, \sum_{j=1}^{k} \alpha_j q_j) \leq \mathcal{KL}(p, q_\ell).$$

چون $\circ < \sum_{i=1}^{n} p_i \log(\frac{q_{\ell i}}{q_{ji}})$ $= E_p\{\log \frac{q_\ell}{q_j}\}$ پس $\mathcal{KL}(p,q_j) < \mathcal{KL}(p,q_\ell)$ و لذا یکی از نامساوی‌های

$$\mathcal{KL}(p,\sum_{j=1}^{k}\alpha_j q_j) < \min\{\mathcal{KL}(p,q_j)\}$$

یا

$$\mathcal{KL}(p,q_j) < \mathcal{KL}(p,\sum_{j=1}^{k}\alpha_j q_j) < \mathcal{KL}(p,q_\ell). \qquad (۱۷)$$

را خواهیم داشت. به عبارتی در حالت اخیر

$$\mathcal{KL}(p,\sum_{j=1}^{k}\alpha_j q_j) \leq \max\{\mathcal{KL}(p,q_j)\}$$

و اثبات قضیه به اتمام می‌رسد.

مثال ۱ : سه مدل رقیب q_1، q_2 و q_3 را به صورت

$$\begin{aligned} q_1 &= (۰/۲۰, ۰/۱۲, ۰/۱۸, ۰/۱۲, ۰/۲۰, ۰/۱۸) \\ q_2 &= (۰/۱۸, ۰/۱۲, ۰/۱۴, ۰/۱۹, ۰/۲۲, ۰/۱۵) \\ q_3 &= (۰/۱۴, ۰/۱۵, ۰/۲۰, ۰/۱۱, ۰/۲۰, ۰/۲۰) \end{aligned}$$

و مدل درست $p = (\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6})$ را در نظر بگیرید. در این مثال $E_p\{\log \frac{q_2}{q_1}\} = -۰/۰۰۱۶ < ۰$، $E_p\{\log \frac{q_3}{q_2}\} = -۰/۰۰۴۳ < ۰$ و مقادیر کولبک-لیب‌لر برای سه مدل رقیب به ترتیب عبارت‌اند از ۰/۰۲۳۱، ۰/۰۲۰۴ و ۰/۰۲۴۷. برای سه مجموعه از مقادیر $\{\alpha_1,\alpha_2,\alpha_3\}$ مانند $\{۰/۴, ۰/۳, ۰/۳\}$، $\{۰/۶, ۰/۳, ۰/۱\}$ و $\{۰/۲, ۰/۷, ۰/۱\}$ مقادیر $\mathcal{KL}(p,\sum_{j=1}^{k}\alpha_j q_j)$ به ترتیب مساوی ۰/۰۱۲۶، ۰/۰۱۱۸ و ۰/۰۱۳۸ است، که در هر حالت به صورت زیر است.

$$\mathcal{KL}(p,\sum_{j=1}^{k}\alpha_j q_j) \leq \min\{\mathcal{KL}(p,q_j)\}$$

۵ مطالعه شبیه‌سازی

برای بررسی نتایج نظری سه مدل رقیب لگ‌نرمال LN، گاما Γ و وایبول W، به‌صورت

$$
\begin{aligned}
LN(\alpha_۱, \alpha_۲) &: \quad ۱/x\sqrt{۲\pi\alpha_۲}\exp(\frac{۱}{۲\alpha_۲}(\log x-\mu)^۲) \\
\Gamma(\lambda_۱, \lambda_۲) &: \quad \frac{۱}{{\lambda_۲}^{\lambda_۱}}x^{\lambda_۱-۱}\exp(x/\lambda_۲) \\
W(\beta_۱, \beta_۲) &: \quad \beta_۱\beta_۲ x^{\beta_۱-۱}\exp(\beta_۲ x^{\beta_۱})
\end{aligned}
$$

در نظر گرفته شده‌اند. چگالی لگ‌نرمال با پارامتر های (۲, ۱/۵) به‌عنوان چگالی مولد داده‌ها فرض شده است. پارامترهای هر سه مدل رقیب با استفاده از مشاهدات برآورد شده‌اند. مدل لگ‌نرمال در مجموعه مدل‌های رقیب خوب–توصیف شده و دو مدل دیگر بد–توصیف شده هستند. برای تشکیل مدل آمیخته از مدل‌های رقیب، باید وزن مؤلفه‌ها در ترکیب محدب معلوم باشد. در انتخاب وزن این نکته در نظر گرفته شده است که هر چه مدل رقیب به مدل مولد نزدیک‌تر باشد باید وزن بیشتری به آن اختصاص داده شود. برآورد ماکسیمم درستنمایی پارامترهای هر سه مدل رقیب با استفاده از مشاهدات محاسبه می‌شود. این مقادیر برای هر یک از مدل‌های رقیب به صورت زیر است.

$$
\begin{aligned}
(\hat{\alpha}_۱, \hat{\alpha}_۲) &= (۲/۰۲۴۸, ۱/۵۲۴۸) \\
(\hat{\lambda}_۱, \hat{\lambda}_۲) &= (۰/۷۹۹۷, ۰/۰۵۱۳) \\
(\hat{\beta}_۱, \hat{\beta}_۲) &= (۰/۸۰۳۹, ۱۳/۴۴۶۲)
\end{aligned}
$$

مقادیر ۰/۹۶، ۰/۰۱ و ۰/۰۳ به عنوان وزن در ترکیب محدب مدل‌ها انتخاب شده‌اند. اندازه واگرایی کولبک–لیبلر برای مدل مولد داده‌ها و ترکیب محدب مدل‌های رقیب با وزن‌های به دست آمده برابر با ۰/۰۳۶۰ است. میزان اطلاع کولبک–لیبلر چگالی مولد داده‌ها و هر یک از مدل‌های رقیب به ازای پارامترهای مختلف در جدول ۱ آورده شده‌اند.

جدول ۱: مقایسه میزان اطلاع کولبک-لیبلر برای مدل خوب-توصیف شده لگ‌نرمال و مدل‌های بد-توصیف شده گاما و وایبول با مدل درست $\Gamma(۲, ۱/۵)$

لگ‌نرمال	$LN(۱/۵۰, ۱/۲۱)$	$LN(۱/۸۰, ۱/۹۶)$	$LN(۱/۹۰, ۱/۸۰)$	$LN(۲, ۱/۹۶)$
KL	۰/۱۲۰	۰/۰۲۹	۰/۰۱۲	۰/۰۱۷
گاما	$\Gamma(۰/۰۹, ۶)$	$\Gamma(۰/۳۰, ۳)$	$\Gamma(۰/۴۰, ۲)$	$\Gamma(۰/۵۰, ۱)$
KL	۳/۳۲۹	۴/۴۰۹	۶/۴۵۹	۱۳/۵۶۰
وایبول	$W(۰/۵۰, ۱۰)$	$W(۰/۶۰, ۱۱)$	$W(۰/۷۰, ۱۲)$	$W(۰/۸۰, ۱۳)$
KL	۲۸/۷۰۲	۴۳/۳۴۰	۶۵/۴۱۳	۹۸/۸۱۴

مقادیر به‌دست آمده در جدول ۱ نشان می‌دهند که مدل لگ‌نرمال به ازای مقادیری از پارامتر مدل مانند $(۲, ۱/۹۶)$ و $(۱/۹۰, ۱/۸۰)$ و $(۱/۸۰, ۱/۹۶)$ مقدار KL بزرگتری نسبت به مدل آمیخته دارند. همچنین مدل آمیخته به وضوح KL کوچک‌تری نسبت به دو مدل بد-توصیف شده به ازای مقادیر مختلف پارامترهای مدل دارد. با حذف مدل خوب-توصیف شده از مجموعه مدل‌های رقیب نیز نتایج مشابهی به‌دست می‌آید. همچنین با تغییر در وزن‌های مدل آمیخته مشخص شد که نامساوی‌های بخش ۴ برقرار هستند. لذا در موارد عملی که توجیح کافی برای حذف برخی از مدل‌ها از مجموعه مدل‌های رقیب وجود ندارد، استفاده از مدل آمیخته پیشنهاد می‌شود.

۱.۵ قاعده مینیماکس برای انتخاب مدل

یکی از مسائل مهم در انتخاب مدل، تشکیل کلاسی از مدل‌های رقیب هم‌ارز است. کلاس‌هایی از مدل‌های هم ارز $\mathcal{M}_۱, \mathcal{M}_۲, \ldots, \mathcal{M}_k$ را در نظر بگیرید. براساس روابط (۱۳) و (۱۷) در هر کلاس $\mathcal{M}_j$ عضوی وجود دارد که مدل آمیخته لااقل از آن بهتر است. در هر کلاس مدل حاصل از ترکیب محدب مدل‌های کلاس را $\mathcal{T}$ می‌نامیم. در این صورت $\min_j \mathcal{KL}(p, \mathcal{T}_j)$ انتخاب مینیماکس از بین مدل‌های آمیخته است. خاصیت سازگاری برآوردکننده‌های پارامترهای یک مدل آمیخته و خطای

منتج از این برآوردها روابط(۱۳) و (۱۷) را تفسیر می‌کنند. استفاده از این نامساوی‌ها این تضمین را ایجاب می‌کند که انتخاب مدل آمیخته از بروز بیشینه خطا در انتخاب مدل جلوگیری می‌کند و لذا مدل‌های آمیخته را می‌توان به عنوان مدل مرجع برای تعیین مدل‌های رقیب مناسب در نظر گرفت.

بحث و نتیجه‌گیری

یافتن مدلی که از نظر معیار واگرایی کولبک- لیبلر بهترین مدل برای تخمین مدل درست باشد کاری دشوار است، در این مقاله ترکیب محدب مدل‌های رقیب در نظر گرفته شده‌اند. این مدل آمیخته تحت شرایطی از نزدیک‌ترین مدل در مجموعه مدل‌های رقیب به مدل درست نزدیک‌تر و یا حداقل از بدترین مدل در میان k مدل رقیب به مدل درست نزدیکتر است. لذا به کمک تکرار این فرایند می توان به مدل آمیخته‌ای با تعداد جملات کمتر از k دست یافت. این فرایند تا حصول شرایط مطلوب می‌تواند ادامه یابد. البته این تصمیم‌گیری محدود به k مدلی است که ما انتخاب کرده‌ایم. در عمل ممکن است مدلی پیشنهاد شود که از مدل حاصل از آمیختن k مدل مورد نظر بهتر باشد. یافتن چنین مدلی مستلزم آزمون و خطا است و ممکن است فرایند یافتن مدل بهینه موفق نباشد. به عنوان یک موضوع باز در این زمینه می‌توان به شرایط لازم جهت ساختن مدل آمیخته‌ای اشاره کرد که همواره از تمام k مدل رقیب به مدل درست نزدیک‌تر باشد.

تقدیر و تشکر

نویسندگان از پیشنهادات ارزنده داوران محترم مجله که موجب بهبود مقاله گردید، کمال تشکر و قدردانی را دارند.

مراجع

سیاره، ع.، ترکمان، پ. (۱۳۸۸)، برآورد مخاطره‌های کولبک-لیبلر برای مشاهدات سانسوریده از راست نوع II تحت مدل‌های غیر آشیانه‌ای، جلد **۳**، شماره ۱، ۷۸-۵۹.

Commenges, D. Sayyareh, A. Letenneur, L. Guedj, J and Bar-Hen, A. (2008), Estimating a Difference of Kullback-Leibler Risks Using a Normalized Difference of AIC, *The Annals of Applied Statistics*, **2**, 1123-1142.

Cover, T. M. and Thomas, A. (1991), Elements of Information Theory, *John Wiley and Sons Inc.*

Dragomir, S. C. and Gluscevic, V. (2001), Some Inequalities for the Kullback-Leibler and $\chi^۲$ Distances in Information Theory and Applications, *Tamsui Oxford of Mathematical Sciences*, **17**, 97-111.

Dragomir, S. C. (2008), Some Properties for the Exponential of the Kullback-Leibler Divergence, *Tamsui Oxford of Mathematical Sciences* **24**, 141-151.

Kullback, S. (1968), Information Theory and Statistics, *Dover, New York.*

Kullback, S. and Leibler, R. A. (1951), On Information and Sufficiency, *Annals of Mathematical Statistics*, **22**, 79-86.

Liese, F. and Vajda, I. (1987), Convex Statistical Distance, B. G. Teubner Verlagsgesellschaft, Leipzig.

Sayyareh, A. (2011), Inference after Separated Hypotheses Testing: An Empirical investigation for Linear Models, Journal of Statistical Computation and Simulation. In Press. DOI: 10.1080/00949655.2011.575783.

Sayyareh, A., Obeidi, R. and Bar-Hen, A. (2011), Empirical Comparison between Some Model Selection Criteria, *Communications in Statistics-Simulation and Computation*, **89**, 72-86.

Vuong, Q. H. (1989), Likelihood Ratio Tests for Model Selection and Nonnested Hypotheses, *Econometrica*, **57**, 307-333.

271 Journal of Statistical Sciences, Autumn & Winter 2010
Vol. 4, No. 2, pp 193-209.

Improved Kullback-Leibler Upper Bound Based on Convex Combination of k Rival Models

Sayyareh, A.

Deptartment of Statistics, Razi University, Kermanshah, Iran.

Abstract: In this paper we have established for the Kullback-Leibler divergence that the relative error is supperadditive. It shows that a mixture of k rival models gives a better upper bound for Kullback-Leibler divergence to model selection. In fact, it is shown that the mixed model introduce a model which is better than of the all rival models in the mixture or a model which is better than the worst rival model in the mixture.

Keywords: Convex Combination, Geometric Mean, Kullback-Leibler Risk, Mixture of Models, Model Selection, Relative Error.

Mathematics Subject Classification (2000): 26D15