مجله علوم آماری، پاییز و زمستان ۱۳۸۹
جلد ۴، شماره ۲، ص ۱۴۹-۱۶۵

انتخاب مجموعه‌ای مجاز از k مدل رقیب غیرآشیانی

قباد برمالزن[۱]، عبدالرضا سیاره[۲]
[۱]گروه آمار، دانشگاه زابل
[۲]گروه آمار، دانشگاه رازی کرمانشاه

تاریخ دریافت: ۱۳۸۹/۲/۱۰ تاریخ آخرین بازنگری: ۱۳۸۹/۱۲/۱۸

چکیده : در تحلیل‌های آماری با یک نمونه تصادفی n تایی از یک جامعه با چگالی درست و نامعلوم $h(\cdot)$ روبرو هستیم. معمولاً مدلی مانند $f(x;\theta)$ به عنوان تقریبی از این چگالی در نظر گرفته می‌شود و استنباط براساس آن صورت می‌گیرد. بدیهی است که $f(x;\theta)$ باید به چگالی درست h نزدیک باشد تا هرگونه استنباط در مورد جامعه معتبر باشد. پیشنهاد مدلی قطعی براساس‌مشاهدات به عنوان تقریب یا برآورد h موجب بروز ریسک بزرگی در انتخاب مدل برای جامعه می‌شود. به همین دلیل k مدل غیرآشیانی $F_{\Theta(k)},\ldots,F_{\Theta(1)}$ انتخاب و بررسی می‌شود که کدام مدل به چگالی درست داده‌ها نزدیک‌تر است. در این مقاله نحوهٔ به‌دست آوردن مجموعه‌ای از مدل‌های مناسب برای برآورد چگالی درست h ارائه خواهد شد. سپس روشی پیشنهاد می‌شود که براساس ریسک کولبک-لیب‌لر در هر خانواده از مدل‌های رقیب چگالی‌هایی که از لحاظ نزدیکی به چگالی درست h معادل هستند، تعیین می‌شوند.

واژه‌های کلیدی : ا نتخاب مدل، ریسک کولبک-لیب‌لر، مدل‌های غیرآشیانی، برآوردگر شبه‌درستنمایی ماکسیمم، مدل‌های رقیب.

آدرس الکترونیک مسئول مقاله: عبدالرضا سیاره، asayyareh@razi.ac.ir
کد موضوع‌بندی ریاضی (۲۰۱۰): ۲۶D۱۵

۱ مقدمه

تحلیل مدل های آشیانی[۱] معمولاً به دو روش مجزا، استفاده از ملاک‌های انتخاب مدل و آزمون فرضیه انجام می‌گیرد. مسأله آزمون فرضیه در مدل‌های کلاسیک (آشیانی) شامل فرضیه‌هایی در مورد پارامتر مجهول جامعه است که با روش‌های متداول قابل آزمون کردن هستند. لم نیمن پیرسون و آزمون نسبت درستنمایی روش‌هایی هستند که برای آزمون کردن چنین فرضیه‌هایی به‌کار برده می‌شوند. اگر فرضیه‌ها متعلق به یک خانواده آشیانی پارامتری نباشند روش‌های کلاسیک دیگر قابل استفاده نیستند. روش آزمون کردن چنین فرضیه‌هایی، ابتدا توسط کاکس (۱۹۶۲-۱۹۶۱) مورد بررسی قرار گرفت. وونگ (۱۹۸۹) آزمون فرضیه در مورد مدل‌های غیرآشیانی[۲] را برای این فرضیه صفر که دو مدل پیشنهاد شده از لحاظ نزدیکی به توزیع درست داده‌ها معادل هستند پیشنهاد کرد.

تحلیل مدل‌های آشیانی در آمار کلاسیک به وفور مورد بررسی و مطالعه قرار گرفته است اما کمتر به تحلیل مدل‌های غیرآشیانی پرداخته شده است. تاریخچه یک مطالعه جدی و اساسی در مورد مدل‌های غیرآشیانی به کاکس (۱۹۶۲-۱۹۶۱) و وونگ (۱۹۸۹) برمی‌گردد. ملاک‌های دیگری برای انتخاب مدل معرفی شده‌اند که می‌توان به ملاک اطلاع آکائیک (AIC) (آکائیک، ۱۹۷۳)، ملاک اطلاع بیزی (BIC) (شوارتز ۱۹۷۸)، تکنیک اعتبارسنجی متقابل و ملاک اطلاع خودگردانی[۳] (EIC) ایشیگارو (۱۹۷۷) اشاره کرد.

از ملاک اطلاع آکائیک (AIC) به‌طور گسترده برای انتخاب بهترین مدل از مجموعهٔ مدل‌های رقیب پارامتری استفاده می‌شود. شیمودیرا (۱۹۹۸) با استفاده از امید ریاضی اطلاع آکائیک ($E(AIC)$) هر مدل، مجموعه‌ای مناسب از مدل ها را به جای انتخاب فقط یک مدل، تشکیل داد. این مجموعه از مدل‌ها را مجموعه اطمینان نامید که شامل مدلی است که دارای مینیمم $E(AIC)$ است و نرخ خطای آن کوچکتر از سطح معنی‌داری مشخص شده است (شیمودیرا، ۱۹۹۸).

---

۱ Nested Models

۲ Non-nested Models

۳ Bootstrap

آکائیک $AIC = -۲\sum_{t=۱}^{n} \log f(x_t; \hat{\beta}_n) + ۲p$ را به عنوان ملاکی در انتخاب مدل معرفی کرد که در آن p برابر تعداد پارامترهای بکار رفته در مدل است. لذا از مدلی که دارای مینیمم AIC است به‌عنوان یک برآورد مقدماتی از چگالی درست h استفاده می‌شود. شیمودیرا روش انتخاب مدلی را در نظر گرفت که آمارشناس را به یک درجه از اطمینان و اعتبار رهنمون می سازد. ایدهٔ اساسی در نظر گرفتن آزمون فرضیه روی $E(AIC)$ها و تشکیل یک مجموعه اطمینان از مدل‌ها است. البته باید توجه داشت که مجموعهٔ اطمینان یک رتبه‌بندی از مینیمم AIC (MAIC) مدل‌ها نیست بلکه به عنوان یک اطلاع مکمل در انتخاب مدل برای دستیابی به بهترین مدل بکار می‌رود. لذا می‌توان این مجموعه اطمینان را به عنوان یک برآورد فاصله‌ایی و MAIC را به عنوان یک برآورد نقطه‌ایی از بهترین مدل در نظر گرفت.

برای درک بهتر این روش، شیمودیرا (۱۹۹۸) ابتدا ساده‌ترین حالت با دو مدل غیرآشیانی $F_{\Theta(۱)} = \{f(x;\theta(۱)); \theta(۱) \in \Theta(۱) \subset \Re^{p_۱}\}$ و $F_{\Theta(۲)} = \{f(x;\theta(۲)); \theta(۲) \in \Theta(۲) \subset \Re^{p_۲}\}$ را برای مقایسه درنظر گرفت. برای $i = ۱, ۲$ منظور از $\Theta(i)$ و $\theta(i)$ به ترتیب فضای پارامترمدل iام و پارامتر مدل iام است. همچنین فرض کنید $AIC_۱$ و $AIC_۲$ به ترتیب ملاک‌های آکائیک مربوط به این دو مدل باشند. شیمودیرا ابتدا فرضیه صفر $E(AIC_۱) \leq E(AIC_۲)$ را در مقابل $E(AIC_۱) > E(AIC_۲)$ آزمون کرد. بر این اساس هرگاه فرضیه صفر $E(AIC_۱) \leq E(AIC_۲)$ پذیرفته شود، $F_{\Theta(۱)}$ عضوی از مجموعهٔ اطمینان در نظر گرفته می‌شود. سپس با عوض کردن نقش $E(AIC_۱)$ و $E(AIC_۲)$ بار دیگر آزمون را تکرار می‌شود. مجموعه اطمینان به‌دست آمده ممکن است یکی از مجموعه‌های $\{F_{\Theta(۱)}\}$ ، $\{F_{\Theta(۲)}\}$ یا $\{F_{\Theta(۱)}, F_{\Theta(۲)}\}$ باشد. از اختلاف AIC های استاندارد شده به عنوان آمارهٔ این آزمون‌ها استفاده می‌شود که تحت فرضیه $E(AIC_۱) = E(AIC_۲)$ دارای توزیع نرمال استاندارد است (لینهارت، ۱۹۸۸).

تعمیم این آزمون برای حالتی که بیشتر از دو مدل غیرآشیانی وجود داشته باشد توسط شیمودیرا انجام شده است. مجموعه مدل‌های پیشنهاد شده را به صورت $M = \left\{F_{\Theta(i)} \middle| i \in \{۱, \ldots, k\}\right\}$ در نظر بگیرید. برای هر $i \in \{۱, \ldots, k\}$ آزمون $H_{\circ i} : E(AIC_i) \leq \min_{j \neq i} E(AIC_j)$ در مقابل

$H_{۱i} : E(AIC_i) > \min_{j\neq i} E(AIC_j)$ معرفی می‌شود. در این حالت مدل $F_{\Theta(i)}$ عضوی از مجموعهٔ اطمینان به حساب می‌آید مگر این که فرضیه $H_{۰i}$ در سطح معنی‌داری مشخص شده، رد شود.

کومانژ و همکاران (۲۰۰۸) به برآورد تفاضل ریسک کولبک-لیب‌لر[۴] میان دو مدل پیشنهاد شده و فاصله ردیابی مناسب پرداخته‌اند. همچنین نشان داده‌اند که تفاضل AIC های نرمال شده، برآورد نقطه‌ایی از تفاضل ریسک کولبک-لیب‌لر دو مدل رقیب است. این برآورد مجموع وزنی لگاریتم درستنمایی دو مدل رقیب است. توزیع این آماره توسط وونگ برای مدل‌های غیرآشیانی و توسط والد (۱۹۴۳) برای مدل‌های آشیانی ارائه شده است. در این مقاله به تحلیل مدل‌های غیرآشیانی به روش استفاده از ملاک‌های انتخاب مدل پرداخته شده است.

زیان استفاده از چگالی $f(x;\theta)$ به جای چگالی درست h برای مشاهده X به صورت $\log[\frac{h(x)}{f(x;\theta)}]$ تعریف می شود. امید ریاضی این زیان تحت چگالی درست h، ریسک کولبک-لیب‌لر $f(x;\theta)$ نسبت به $h(x)$ نامیده می‌شود. به عبارت دیگر

$$KL_h[h, f(x;\theta)] = E_h\Big[\log\Big(\frac{h(X)}{f(X;\theta)}\Big)\Big],$$

که در آن E_h بیانگر امیدریاضی نسبت به چگالی درست و نامعلوم h است. این ملاک به اختصار با نماد KL نشان داده می‌شود و دارای ویژگی‌های زیر است:

الف) $KL_h[h, f(x;\theta)] \geq ۰$.

ب) $KL_h[h, f(x;\theta)] = ۰$ اگر و فقط اگر $\theta_\star$ ایی متعلق به فضای پارامتر Θ وجود داشته باشد به‌طوری که $h(x) = f(x;\theta_\star)$.

ریسک KL معمولاً به عنوان فاصله بین دو چگالی احتمال یا به طور کلی دو اندازه احتمال تفسیر می‌شود. این ملاک یک متر ریاضی نیست زیرا خاصیت تقارن متر را ندارد. از این دیدگاه، کوچک بودن ملاک KL بیانگر نزدیک بودن چگالی $f(x;\theta)$ به h است. اگر F_Θ و G_Γ دو مدل رقیب باشند، می‌گوییم F_Θ نسبت به G_Γ به‌چگالی درست h نزدیکتر است هرگاه $KL_h[h, f(x;\theta_\star)] \leq KL_h[h, g(x;\gamma_\star)]$، که در آن $\theta_\star$ و $\gamma_\star$ به ترتیب مینیمم کننده‌های ملاک KL در مدل‌های F_Θ و G_Γ هستند.

---

[۴] Kullback - Leibler Risk

ملاک KL می‌تواند به صورت

$$KL_h[h, f(x;\theta)] = E_h\left[\log\left(\frac{h(X)}{f(X;\theta)}\right)\right] = E_h[\log h(X)] - E_h[\log f(X;\theta)].$$

تجزیه شود. چون جمله اول سمت راست این رابطه، مقداری ثابت است و فقط به چگالی درست h بستگی دارد، واضح است که برای مقایسهٔ مدل‌های متفاوت، کافیست فقط جمله دوم سمت راست در نظر گرفته شود. بزرگ بودن $E_h[\log f(X;\theta)]$ نزدیک بودن چگالی $f(x;\theta)$ به چگالی درست h و کوچک بودن $E_h[\log f(X;\theta)]$ دور بودن چگالی $f(x;\theta)$ از چگالی درست h را نشان می‌دهد. اما مقدار $E_h[\log f(X;\theta)]$ مجهول است، زیرا به چگالی نامعلوم h بستگی دارد. با این وجود، با استفاده از تابع توزیع تجربی، قانون اعداد بزرگ و برآوردکننده‌های شبه درستنمایی ماکسیمم، زمانی که n به سمت بینهایت میل کند، از برآوردگر مجانبی آن به منظور مقایسه بین مدل های متفاوت استفاده می‌شود (کونیشی و کیتاگاوا، ۱۹۹۶).

در بخش ۲ این مقاله، تعریف مسأله و قضایای مورد نیاز برای ساختن مجموعه مجاز از مدل‌ها آورده شده است. معرفی مجموعه مجاز از مدل‌ها در بخش ۳ آورده شده است و سرانجام در بخش ۴ با روش‌های شبیه‌سازی نشان داده شده است که ملاک معرفی شده قادر است چگالی‌های رقیب مناسب را انتخاب نماید.

۲ تعریف مسأله و قضایای مورد نیاز

فرض کنید یک نمونه تصادفی $X_1, \ldots, X_n$ از یک جامعه با چگالی درست $h(\cdot)$ را در اختیار داریم. در حالت کلی h نامعلوم است. برای تقریب این چگالی نامعلوم، k مدل رقیب غیرآشیانی $F_{\Theta(1)}, \ldots, F_{\Theta(k)}$ از مدل‌های پارامتری را که هر کدام خانواده‌ایی از چگالی‌ها هستند در اختیار داریم. مسأله مورد توجه در این مقاله بررسی این موضوع است که برای $i = 1, \ldots, k$، در هر خانواده پارامتری $F_{\Theta(i)}$ به صورت $F_{\Theta(i)} = \{f(x, \theta(i)); \theta(i) \in \Theta(i) \subset \Re^{p_i}\}$ کدام چگالی‌ها تقریب مناسب‌تری برای چگالی درست h هستند و در حالت ایده‌آل بهترین آن‌ها انتخاب شود. وقتی که مجموعه‌ای از چگالی‌های مناسب از لحاظ نزدیکی به چگالی درست h به دست آمد جستجوی ما برای انتخاب بهترین چگالی، محدود به بررسی این

مجموعه از چگالی‌ها خواهد بود. حال این سؤال پیش می‌آید که این مجموعه را چگونه می‌توان ساخت؟ در این مقاله ملاکی معرفی شده است که با استفاده از آن، چگالی‌های مناسب را در هر خانواده از مدل‌های رقیب می‌توان پیدا کرد.

یکی از مسائل اساسی در انتخاب مدل برای داده‌ها، یافتن مجموعه‌ای از مدل‌های مناسب است. لذا در عمل با مدل‌های پارامتری رقیب $F_{\Theta(k)}, \ldots, F_{\Theta(1)}$ روبرو هستیم.

تعریف ۱ : فرض کنید $F_{\Theta(m)}$ و $F_{\Theta(l)}$ دو مدل رقیب باشند. دو مدل $F_{\Theta(m)}$ و $F_{\Theta(l)}$ نسبت به هم غیرآشیانی هستند اگر و فقط اگر $F_{\Theta(m)} \cap F_{\Theta(l)} = \emptyset$.

تعریف ۲ : مدل $F_{\Theta(m)}$ خوب - توصیف شده[۵] است، اگر و فقط اگر $\theta_\star$ای در فضای پارامتری $\Theta(m)$ وجود داشته باشد، به طوری که $h(x) = f(x;\theta_\star)$، یعنی $h \in F_{\Theta(m)}$. در غیر این‌صورت مدل $F_{\Theta(m)}$ بد - توصیف شده[۶] است یا به عبارت دیگر $h \notin F_{\Theta(m)}$.

تعریف ۳ : فرض کنید $X_n, \ldots, X_1$ نمونه‌ای تصادفی از چگالی درست h باشد. اگر $f(x;\theta)$ یک مدل رقیب برای تقریب h باشد، آنگاه $L(\theta) = \sum_{t=1}^{n} \log f(x_t;\theta)$ تابع شبه‌درستنمایی نامیده می‌شود. در این حالت $\hat{\theta}_n$ برآوردگر شبه‌درستنمایی ماکسیمم[۷] (QMLE) نامیده می‌شود، هرگاه در شرط $L(\hat{\theta}_n) = \sup_{\theta\in\Theta} L(\theta)$ صدق کند.

بسیاری از ملاک‌های انتخاب مدل، براساس توابع درستنمایی ماکسیمم طراحی شده‌اند. دلیل خوبی این ملاک‌ها ایدهٔ استفاده از اصل درستنمایی ذکر شده است. در قضیه زیر نشان داده می شود که چگونه ایدهٔ درستنمایی ماکسیمم کمک می‌کند تا از اعضای یک خانوادهٔ چگالی‌ها، با استفاده از ملاک KL به بهترین تبیین کننده داده‌ها دست یافت. در واقع نشان داده می‌شود که اگر $f(x;\theta_\star(i))$ نزدیک‌ترین عضو خانواده چگالی‌های رقیب iام ($i = 1, \ldots, k$) به چگالی درست h باشد، آنگاه

---

۵ Well Specified

۶ Misspecified

۷ Quasi Maximum Likelihood Estimator

فاصله تابع درستنمایی ماکسیمم شده تا $f(x;\theta_\star(i))$ بیشتر از فاصله چگالی درست h تا $f(x;\theta_\star(i))$ نخواهد شد. بنابراین اگر $f(x;\theta_\star(i))$ انتخابی مناسب و مجهول، به عنوان چگالی رقیب برای چگالی h باشد، این اطمینان حاصل می‌شود که $f(x;\hat{\theta}(i))$ بهترین جایگزین $f(x;\theta_\star(i))$ است، که در آن $\hat{\theta}(i)$ برآوردگر شبه‌درستنمایی ماکسیمم $\theta(i)$ است.

قضیه ۱ : فرض کنید $M=\{F_{\Theta(i)};\theta(i)\in\Theta(i)\subset\Re^{p_i},i=1,\ldots,k\}$ کلاس تمام خانواده‌های رقیب برای چگالی درست داده‌ها یعنی h باشد. اگر $\hat{\theta}(i)$ برآوردگر شبه‌درستنمایی ماکسیمم برای $\theta(i)$ در خانواده مدل رقیب iام و چنانچه $f(x;\theta_\star(i))$ نزدیکترین عضو خانوادهٔ چگالی‌های رقیب iام برای $i=1,\ldots,k$ به چگالی درست h باشد، آنگاه

$$KL_h\left[h,f\left(x;\theta_\star(i)\right)\right]>KL_{f(x;\hat{\theta}(i))}\left[f\left(x;\hat{\theta}(i)\right),f(x;\theta_\star(i)\right)\Big].$$

برهان بنابه تعریف ملاک KL داریم

$$\begin{aligned}
KL_h[h,f(x;\theta_\star(i))] &= E_h\Big[\log\Big(\frac{h(X)}{f(X;\theta_\star(i))}\Big)\Big]\\
&= E_h\Big[\log\Big(\frac{h(X)}{f(X;\hat{\theta}(i))}\Big)\Big]+E_h\Big[\log\Big(\frac{f(X;\hat{\theta}(i))}{f(X;\theta_\star(i))}\Big)\Big]\\
&> E_h\Big[\log\Big(\frac{f(X;\hat{\theta}(i))}{f(X;\theta_\star(i))}\Big)\Big]\\
&\geq E_{f(x;\hat{\theta}(i))}\Big[\log\Big(\frac{f(X;\hat{\theta}(i))}{f(X;\theta_\star(i))}\Big)\Big]\\
&= KL_{f(x;\hat{\theta}(i))}\Big[f(x;\hat{\theta}(i)),f(x;\theta_\star(i))\Big].
\end{aligned}$$

در اولین نامساوی از نامنفی بودن KL و در نامساوی بعدی از این نکته استفاده شده است که

$$\begin{aligned}
E_h\Big[\log\Big(\frac{f(X;\hat{\theta}(i))}{f(X;\theta_\star(i))}\Big)\Big] &= \int\Big(\log\frac{f(x;\hat{\theta}(i))}{f(x;\theta_\star(i))}\Big)dH(x)\\
&> \int\Big(\log(\frac{f(x;\hat{\theta}(i))}{f(x;\theta_\star(i))}\Big)dF(x;\hat{\theta}(i))
\end{aligned}$$

$$= \quad KL_{f(x;\hat{\theta}(i))}\Big[f(x;\hat{\theta}(i)), f(x;\theta_\star(i))\Big].$$

قضیه ۱ نشان می‌دهد که استفاده از تابع درستنمایی، ملاک مناسبی برای انتخاب مدل در مجموعه M است. یعنی برای $i = ۱,\ldots,k$ فاصله $f(x;\hat{\theta}(i))$ از $f(x;\theta_\star(i))$ بیشتر از فاصله چگالی درست h از $f(x;\theta_\star(i))$ نخواهد بود. اما موضوع اریبی که در اثر استفاده از $f(x;\hat{\theta}(i))$ به جای چگالی درست h به‌وجود می‌آید باید به نحوی در نظر گرفته شود.

به منظور بررسی هر چه بیشتر اریبی‌هایی که در ملاک‌های انتخاب مدل وجود دارند، بسط $KL_h\left[h, f\left(x;\hat{\theta}(i)\right)\right]$ را در نظر بگیرید. لینهارت و زوکچینی (۱۹۸۶) نشان دادند

$$KL_h\Big[h, f\Big(x;\hat{\theta}(i)\Big)\Big] = KL_h\Big[h, f\Big(x;\theta_\star(i)\Big)\Big] + \frac{tr(J^{-۱}_{F_{\Theta(i)}} I_{F_{\Theta(i)}})}{۲n} + o(\frac{۱}{n}). \quad (۱)$$

که در آن

$$I_{F_{\Theta(i)}} = E_h\left[(\frac{\partial \log f(X;\theta(i))}{\partial\theta(i))}\Big|_{\theta_\star})(\frac{\partial \log f(X;\theta(i))}{\partial\theta(i)}\Big|_{\theta_\star})^T\right],$$

و

$$J_{F_{\Theta(i)}} = -E_h\Big[\frac{\partial^۲ \log f(X;\theta(i))}{\partial\theta(i)\partial\theta(i)^T}\Big|_{\theta_\star}\Big],$$

و $tr(J^{-۱}_{F_{\Theta(i)}} I_{F_{\Theta(i)}})$ بیانگر اثر[۸] ماتریس $(J^{-۱}_{F_{\Theta(i)}} I_{F_{\Theta(i)}})$ و $f(x;\theta_\star(i))$ بیانگر نزدیکترین عضو خانواده مدلiام به چگالی درست h است. بنابراین ریسک $KL_h\left[h, f(x;\hat{\theta}(i))\right]$ به عنوان مجموع ریسک بد - توصیف شدگی $KL_h\left[h, f(x;\theta(i))\right]$ و ریسک آماری $tr(J^{-۱}_{F_{\Theta(i)}} I_{F_{\Theta(i)}})$ تفسیر می‌شود.

در حالت خاص، اگر مدل $F_{\Theta(i)}$ خوب - توصیف شده باشد، آنگاه $KL_h\left[h, f(x;\theta_\star(i))\right] = ۰$ و $I_{F_{\Theta(i)}} = J_{F_{\Theta(i)}}$ و $tr(J^{-۱}_{F_{\Theta(i)}} I_{F_{\Theta(i)}}) = \frac{p_i}{n}$ در نتیجه $KL_h\left[h, f(x;\hat{\theta}(i))\right] = \frac{p_i}{n} + o(\frac{۱}{n})$ است. همچنین

$$KL_h\Big[h, f(x;\hat{\theta}(i))\Big] \quad = \quad E_h\left[\log h(X)\right] - E_h\Big[\frac{۱}{n}\sum_{t=۱}^{n} \log f(X_t;\hat{\theta}(i))\Big]$$

---

[۸] Trace

$$+ \quad \frac{tr(J^{-1}_{F_{\Theta(i)}} I_{F_{\Theta(i)}})}{n} + o_p(\frac{1}{n}). \qquad (۲)$$

در رابطه (۱) مقدار $E_h\left[\log f(X;\theta_\star(i))\right]$ با $E_h\left[\frac{1}{n}\sum_{t=1}^{n}\log f(X_t;\hat{\theta}(i))\right]$ در رابطه (۲) برآورد شده است. اما به دلیل بیش برآورد، ضریب ۱/۲ در رابطه (۲) ظاهر نشده است. از رابطه (۱) می‌توان نتیجه گرفت

$$KL_h\left[h, f(x;\hat{\theta}(i))\right] \geq KL_h\left[h, f(x;\theta_\star(i))\right]$$

یا به طور معادل $E_h\left[\log f(x;\hat{\theta}(i))\right] \leq E_h\left[\log f(x;\theta_\star(i))\right]$ برای $i = 1,\ldots,k$ همواره برقرار است (برای تحلیل بیشتر دو بسط فوق کومانژ و همکاران (۲۰۰۸) را ببینید). لذا $E_h\left[\log f(x;\hat{\theta}(i))\right] - E_h\left[\log f(x;\theta_\star(i))\right] \leq 0$ که این تفاوت ناشی از استفاده $\hat{\theta}(i)$ به جای $\theta_\star(i)$ است. یکی از مسائل اساسی در انتخاب مدل تقلیل این گونه اریبی‌هااست.

۳ تعیین مجموعه مجاز از مدل‌ها

فرض کنید k مدل رقیب غیرآشیانی $F_{\Theta(k)},\ldots,F_{\Theta(1)}$ از مدل‌های پارامتری را که هر کدام خانواده‌ای از چگالی‌ها هستند در اختیار داریم. برای انتخاب بهترین عضوها از هر مدل پارامتری و در نتیجه ساخت مجموعه مجاز[۹] از ایدهٔ مینیمم کولبک-لیبلر یعنی $\min_i KL_h[h, f(x;\hat{\theta}(i))]$ استفاده خواهد شد.

تعریف ۴ : مجموعه همه عضوهای $f(x,\theta(ij))$ برای $i = 1,\ldots,k$ و $j = 1,\ldots,D_{\theta(i)}$ از مدل‌های رقیب را که در شرط

$$KL_h\left[h, f(x;\theta(ij))\right] \leq \min_i KL_h\left[h, f(x;\hat{\theta}(i))\right]$$

صدق کنند، مجموعه مجاز از چگالی ها نامیده می‌شود.

برای ساختن این مجموعه به صورت زیر عمل می شود:

الف) ابتدا در هر مدل پارامتری $F_{\Theta(i)}$ برای $i = 1,\ldots,k$ مقدار $KL_h[h, f(x;\hat{\theta}(i)]$

[۹] Admissible set

محاسبه می‌شود.

ب) در بین k مقدار به‌دست آمده از قسمت (الف)، مقدار مینیمم KL را پیدا کرده و ρ می‌نامیم.

ج) مجموعه مجاز از چگالی‌ها را τ_ρ نامیده و ملاک پیشنهادی به‌صورت زیر تعریف می‌شود

$$\tau_\rho = \left\{ f(x;\theta(ij)) \in F_{\Theta(i)} \middle| KL_h[h, f\,(x;\theta(ij))] \leq \min_i KL_h\left[h, f(x;\hat{\theta}(i))\right] = \rho \right\}.$$

شاید این سوال پیش آید که به دلیل مجهول بودن KL و در نتیجه مجهول بودن ρ ساختن این مجموعه عملی نباشد. اما باید توجه داشت که رابطه $KL_h[h, f(x;\theta(ij))] \leq \min_i KL_h\left[h, f(x;\hat{\theta}(i))\right]$ برقرار است اگر و فقط اگر

$$E_h[\log h(X)] - E_h[\log f(X;\theta(ij))] \leq \min_i \left[E_h[\log h(X)] - E_h[\log f(X;\hat{\theta}(i))]\right],$$

یا به‌طور معادل

$$E_h[\log h(X)] - E_h[\log f(X;\theta(ij))] \leq E_h[\log h(X)] - \max_i E_h\left[\log f(X;\hat{\theta}(i))\right],$$

حال با فرض متناهی بودن $E_h[\log h(X)]$ و حذف آن از طرفین داریم

$$E_h[\log f(X;\theta(ij))] \geq \max_i E_h\left[\log f(X;\hat{\theta}(i))\right].$$

بنابراین مجموعه مجاز τ_ρ از چگالی‌ها، برای تقریب چگالی درست h را می‌توان به‌صورت

$$\tau_\rho = \left\{ f(x;\theta(ij)) \in F_{\Theta(i)} \middle| E_h[\log f(X;\theta(ij))] \geq \max_i E_h\left[\log f(X;\hat{\theta}(i))\right] \right\}$$

نیز بیان کرد. ملاک ساخت مجموعهٔ مجاز به چگالی مجهول h بستگی دارد و در عمل باید مقادیر $E_h[\log f(X;\hat{\theta}(i))]$ و $E_h[\log f(X;\theta(ij))]$ برآورد شوند. ابتدا برآوردگر $E_h[\log f(X;\theta(ij))]$ را به‌دست آورده می‌شود. همواره

$$\begin{aligned} b_1 &= E_h\left[\frac{1}{n}\sum_{t=1}^{n} \log f(X_t;\theta(ij)) - E_h[\log f(X;\theta(ij))]\right] \\ &= \frac{1}{n}\sum_{t=1}^{n} E_h\left[\log f\,(X_t;\theta(ij))\right] - E_h[\log f(X;\theta(ij))] = 0. \end{aligned}$$

چون مقدار b_1 برابر صفر است، لذا $\frac{1}{n}\sum_{t=1}^{n}\log f(X_t;\theta(ij))]$ یک برآوردگر ناااریب برای $E_h[\log f(X;\theta(ij))]$ است. از طرف دیگر از (۱) و (۲) به سادگی می‌توان نتیجه گرفت

$$E_h\Big[\frac{1}{n}\sum_{t=1}^{n}\log f(X_t;\hat{\theta}(i)) - E_h[\log f(X;\hat{\theta}(n,i))]\Big] \simeq \frac{tr(J^{-1}_{F_{\Theta(i)}} I_{F_{\Theta(i)}})}{n}.$$

بنابراین

$$E_h[\log f(X;\hat{\theta}(i)] \simeq E_h[\frac{1}{n}\sum_{t=1}^{n}\log f(X_t;\hat{\theta}(i)) - \frac{tr(J^{-1}_{F_{\Theta(i)}} I_{F_{\Theta(i)}})}{n}]$$

در نتیجه یک برآوردگر ناااریب مجانبی برای $E_h\left[\log f\left(X;\hat{\theta}(i)\right)\right]$ به صورت $\frac{1}{n}\sum_{t=1}^{n}\log f\left(X_t;\hat{\theta}(i)\right) - n^{-1}tr(J^{-1}_{F_{\Theta(i)}} I_{F_{\Theta(i)}})$ است. با جایگذاری این برآوردگرها در ملاک τ_ρ مجموعه مجاز برای $i=1,\ldots,k, j=1,\ldots,D_{\Theta(i)}$ به‌صورت

$$\begin{aligned}\hat{\tau}_\rho &= \Big\{f\Big|\frac{1}{n}\sum_{t=1}^{n}\log f(x_t;\theta(ij)) \\ &\geq \max_i\Big[\frac{1}{n}\sum_{t=1}^{n}\log f(x_t;\hat{\theta}(i)) - \frac{tr(J^{-1}_{F_{\Theta(i)}} I_{F_{\Theta(i)}})}{n}\Big]\Big\}.\end{aligned}$$

به‌دست می‌آید. بنابراین $\frac{1}{n}\sum_{t=1}^{n}\log f(X_t;\hat{\theta}(i))$ برآوردگری برای $E_h[\log f(X;\hat{\theta}(i))]$ با اریبی $n^{-1}tr(J^{-1}_{F_{\Theta(i)}} I_{F_{\Theta(i)}})$ است.

باید توجه داشت که استفاده از $f(x;\hat{\theta}(i))$ به عنوان تقریبی از چگالی درست h موجب اریبی در انتخاب مدل می‌شود. یعنی باعث می‌شود که $f(x;\hat{\theta}(i))$ بهتر از آنچه که هست ظاهر شود. این اریبی باعث می‌شود که $f(x;\hat{\theta}(i))$ به چگالی درست h نزدیک شود. لذا امکان انتخاب برخی از مدل‌های خوب را از ما سلب خواهد کرد. به همین دلیل باید این اریبی به طریقی تصحیح شود. در اینجا منظور از یک مدل خوب، می‌تواند یک مدل ساده باشد که برازش آن به داده‌ها تقریباً همانند برازش یک مدل پیچیده به داده‌ها است.

به‌طور کلی استفاده از توابع درستنمایی تاوانیده[۱۰]، باعث کاهش و حذف این‌گونه اریبی‌ها شده و امکان انتخاب مدل‌های ساده را میسر می‌سازد. این موضوع باعث قرار گرفتن مدل‌های بیشتری در داخل مجموعه مجاز از مدل‌ها می‌شود و ممکن است مدل ساده‌ای در داخل مجموعه قرار گیرد که به خوبی مدل مناسب پیچیده باشد.

با افزایش تعداد پارامتر، مدل رقیب بهتر می‌شود. بهتر شدن به معنای نزدیک شدن مدل رقیب به چگالی درست h است. این کار باعث می‌شود که یک مدل پیچیده با تعداد زیادی پارامتر را انتخاب کنیم در حالی که در عمل تمایل به کار کردن با مدل‌های ساده تر را داریم. از طرفی این نگرانی وجود دارد که این مدل، واقعاً مدل خوبی نبوده اما افزایش تعداد پارامتر آن را به مدلی خوب، تبدیل نموده باشد که در عمل غیر مفید است. لذا باید این اریبی به کمک تصحیحات بیان شده، کم اثر شود.

به نظر می رسد یک مدل خوب به وسیله روش درستنمایی ماکسیمم مشخص می‌شود، یعنی می‌توان $\ell(\theta) = \frac{1}{n}\sum_{t=1}^{n} \log f(x_t;\hat{\theta}_n)$ را به عنوان یک ملاک نیکویی مدل $f(x;\theta)$ در نظر گرفت. اما این روش نمی‌تواند یک مقایسه عادلانه از مدل‌ها باشد زیرا در نظر گرفتن $\ell(\hat{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n} \log f(x_t;\hat{\theta}_n)$ به عنوان یک برآوردگر برای $E_h[\log f(X;\theta_\star)]$ دارای مقداری اریبی است که مقدار این اریبی به بعد فضای پارامتر بستگی دارد. دو بار استفاده از داده‌های یکسان، در برآورد پارامتر θ و در ارزیابی مدل، می‌تواند علت این اریبی باشد.

یک روش دیگر برای کاهش اریبی لگاریتم درستنمایی وزنی، استفاده از روش اعتبارسنجی متقابل مرتبهٔ m[۱۱] است. به این ترتیب که از m تا از داده‌های نمونه تصادفی برای برآورد پارامتر θ و از $n-m$ تای دیگر برای ارزیابی مدل استفاده می‌شود. با این روش تا اندازه‌ای اریبی لگاریتم درستنمایی وزنی کاهش خواهد یافت. این ملاک که روشی برای کاهش اریبی لگاریتم درستنمایی موزون است، به

---

[۱۰] Penalized Likelihood

[۱۱] Cross validation m folded

صورت

$$L_{CV,i} = \frac{1}{n-m} \sum_{t=1}^{n-m} \log f\left(x_t; \hat{\theta}_{i,-(n-m)}\right)$$

تعریف می شود، که در آن $\hat{\theta}_{i,-(n-m)}$ بیانگر برآورد شبه‌درستنمایی ماکسیمم مدل iام است و در آن فقط از m مشاهده برای برآورد پارامتر θ استفاده شده است و از $n-m$ دادهٔ دیگر در برآورد پارامتر θ صرف نظر گردیده است.

۴ شبیه‌سازی

فرض کنید یک نمونه تصادفی $n = ۵۰$ تایی از چگالی نرمال استاندارد (چگالی درست) تولید شده باشد. برای تقریب این چگالی سه مدل رقیب نرمال با پارامتر $\Theta(۱) = (\mu, \sigma)$، کوشی با پارامتر $\Theta(۲) = (\alpha, \beta)$ و لاپلاس با پارامتر $\Theta(۳) = (\eta, \gamma)$ پیشنهاد شده‌اند. ابتدا با روش شبه‌درستنمایی ماکسیمم، برآورد پارامترها را برای هر مدل رقیب پیدا کرده و سپس با توجه به برآوردهای به دست آمده، چند مقدار دلخواه از فضای پارامتر را که نزدیک‌ترین عددها به برآوردگرهای شبه درستنمایی ماکسیمم هستند را در نظر گرفته و چگالی‌های متناظر با این پارامترها به عنوان چگالی‌هایی از مدل‌های رقیب انتخاب می‌شوند. از نرم افزار S-PLUS برای پیدا کردن برآوردها و محاسبات استفاده شده و نتایج در جدول‌های ۱ تا ۳ ارائه شده‌اند، که در آن‌ها ستون اول شامل چگالی‌های رقیب، ستون دوم شامل لگاریتم درستنمایی موزون و ستون سوم لگاریتم درستنمایی تاوانیده به وسیله $n^{-۱} tr(J^{-۱}_{F_{\Theta(i)}} I_{F_{\Theta(i)}}) \simeq n^{-۱} p_i$ است. p_i بیانگر تعداد پارامترهای مدل iام است. در ستون چهارم ملاک اعتبارسنجی متقابل مرتبه دهم ارائه شده است. با استفاده از روش تقریب مونت کارلو این آزمایش ۵۰ بار تکرار شده است.

مجموعه مجاز از چگالی‌ها برای برآورد h، با استفاده از ملاک‌های بیان شده عبارت خواهند بود از:

$$\tau_{\rho,unadjust} = \{\},$$

$$\tau_{\rho,\frac{p}{n}} = \{N(۰, ۱), N(\hat{\mu}, \hat{\sigma}^۲), Laplace(\hat{\eta}, \hat{\gamma})\},$$

جدول ۱: نتایج به‌دست آمده برای چگالی‌های رقیب نرمال

چگالی‌های رقیب	$\ell(\theta)$	$\ell(\theta)-\frac{p}{n}$	L_{CV}
$N(۰/۵, ۱)$	$-۱/۵۴۲۸$	–	–
$N(۱, ۱)$	$-۱/۹۱۵۳$	–	–
$N(-۱, ۱)$	$-۱/۹۲۵۰$	–	–
$N(۰, ۱)$	$-۱/۴۲۰۲$	–	–
$N(۰/۵, ۰/۵)$	$-۱/۸۲۰۰$	–	–
$N(-۰/۵, ۰/۵)$	$-۲/۷۴۰۶$	–	–
$N(\hat{\mu}, \hat{\sigma})$	$-۱/۴۱۱۳$	$-۱/۵۴۱۳$	$-۱/۹۵۲۳$

جدول ۲: نتایج به‌دست آمده برای چگالی‌های رقیب کوشی

چگالی های رقیب	$\ell(\theta)$	$\ell(\theta)-\frac{p}{n}$	L_{CV}
$C(۰, ۱)$	$-۱/۶۸۵۰$	–	–
$C(۰, ۰/۵)$	$-۱/۶۳۰۱$	–	–
$C(-۰/۵, ۱)$	$-۱/۷۶۶۵$	–	–
$C(۰/۵, ۱)$	$-۱/۷۶۷۴$	–	–
$C(۱, ۱)$	$-۱/۹۹۳۴$	–	–
$C(۰/۰۲, ۱)$	$-۱/۶۸۵۲$	–	–
$C(\hat{\alpha}, \hat{\beta})$	$-۱/۵۸۹۷$	$-۱/۶۲۹۷$	$-۲/۲۶۱$

جدول ۳: نتایج به‌دست آمده برای چگالی‌های رقیب لاپلاس

چگالی های رقیب	$\ell(\theta)$	$\ell(\theta)-\frac{p}{n}$	L_{CV}
$Lap(۰, ۱)$	$-۱/۴۹۸۹$	–	–
$Lap(-۱, ۱)$	$-۱/۸۶۰۷$	–	–
$Lap(۱, ۲)$	$-۱/۹۶۸۶$	–	–
$Lap(۰/۵, ۰/۵)$	$-۱/۸۰۳۰$	–	–
$Lap(۱, ۰/۱)$	$-۱۰/۳۸$	–	–
$Lap(۱, ۰/۰۱)$	$-۱۱۲/۵۶$	–	–
$Lap(\hat{\eta}, \hat{\gamma})$	$-۱/۴۴۵۶$	$-۱/۴۹۳۰$	$-۱/۶۱۰۰$

$$\tau_{\rho,CV} = \{N(0,1), N(\hat{\mu},\hat{\sigma}^2), Laplace(0,1), Laplace(\hat{\eta},\hat{\gamma})\}.$$

با توجه به نتایج بدست آمده در شبیه‌سازی ملاحظه می‌شود که براساس ملاک لگاریتم درستنمایی، هیچکدام از مدل‌های رقیب در داخل مجموعه مجاز قرار نمی‌گیرند، اما با تصحیح بیش اریبی، مدل‌های مناسبی وارد مجموعه مجاز می‌شوند. مجموعه های $\tau_{\rho,\frac{p}{n}}$ و $\tau_{\rho,CV}$ چگالی‌های مناسب برای چگالی درست h را نشان می‌دهند.

بحث و نتیجه‌گیری

دو بار استفاده از داده‌های یکسان، باعث ایجاد اریبی در لگاریتم تابع درستنمایی وزنی می‌شود که این اریبی به‌وسیلهٔ $n^{-1}tr(J^{-1}_{F_{\Theta(i)}} I_{F_{\Theta(i)}}) \simeq n^{-1}p_i$ تصحیح می‌شود. یک تکنیک دیگر برای بهبود بخشیدن لگاریتم تابع درستنمایی موزون، استفاده از ملاک اعتبارسنجی متقابل است. در هنگام استفاده از لگاریتم تابع درستنمایی موزون تاوانیده، باید حجم نمونه بزرگ باشد اما در ملاک اعتبارسنجی متقابل، نیازی به‌چنین شرطی نیست و حتی برای نمونه‌های کوچک نیز کاربرد دارد. در هنگام استفاده از این ملاک‌ها برای ساخت مجموعه مجاز هیچ محدودیتی در تعداد مدل‌های رقیب وجود ندارد. روش پیشنهاد شده برای ساخت مجموعه مجاز می‌تواند مبنایی برای تحقیق در مورد روش‌های دقیق‌تر قرار بگیرد. یافتن چنین مجموعه‌ای از مدل‌ها به محققین کمک خواهد کرد تا تحلیل بهتری از داده‌های جمع آوری شده داشته باشند و به استنباط‌های دقیق‌تری در مورد جامعه هدف دست یابند.

تقدیر و تشکر

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

مراجع

Akaike, H. (1971), Information Theory and an Extention of Maximum Likelihood Principle, *Proceedings of Second International Sympusium on Information Theory*, Tsahkadsor, 267-281.

Commenges, D. Sayyareh, A. Letenneur, L. Guedj, J. and Bar-Hen, A. (2008). *The Annal of Applied Statistics*, **2**, 1123-1142.

Cox, D. R. (1961), Tests of Separate Families of Hypothesis, *Proceedings of Fourth Berkely Symposium on Mathematical Statistics and Probability*, **1**, University of Colifornia Press, Berkeley, CA, 105-123.

Cox, D. R. (1962), Further Results on Tests of Separate Families of Hypothesis, *Royal Statistical Society*, B, **24**, 406-424.

Ishigaro, M., Sakamoti, Y., and Kitagawa, G. (1997), Bootstrapping Log-likelihood and EIC, an Extension of AIC, *Annals of the Institute of Statistical Mathematics*, **49**, 411-434.

Konishi, S. and Kitagava, G. (1996), *Information Criteria and Statistical Modeling*, Springer, Germany.

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

Linhart, H. and Zucchini, W. (1986), Model Selection, Wiley, New York.

Schwarz, G. (1978), Estimating the Dimension of a Model, Annals of Statistics, **6**, 461-464.

Shimodaira, H. (2001), Multiple Comparisons Of Log-Likelihood and Combining Non-nested Models With Applications to Phylogenetic Tree Selection, Communication in Statistics, **30**, 1751-1772.

Shimodaira, H. (1998), An Application of Multiple Comparison Techniques to Model Selection, *Annals of the Institute of Statistical Mathematics*, **50**, 1-13.

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

Wald, A. (1943), Tests of Statistical Hypotheses Concerning Serval Parameters When the Number of Observation is Large, *Transactions of the American Mathematical Society*, **54**, 426-482.

Journal of Statistical Sciences, Autumn & Winter 2010 **274**

Vol. 4, No. 2, pp 149-165.

The Choice of an Admissible Set of k Non-nested Models

Barmalzan, G. and Sayyareh, A.

Department of Statistics, University of Zabol, Zabol, Iran.

Department of Statistics, Razi University, Kermanshah, Iran.

Abstract: Suppose we have a random sample of size n of a population with true density $h(\cdot)$. In general, h is unknown and we use the model $f(x;\theta)$ as an approximation of this density function. We do inference based on $f(x;\theta)$. Clearly, $f(x;\theta)$ must be close to the true density h, to reach a valid inference about the population. The suggestion of an absolute model based on a few observations, as an approximation or estimation of the true density, h, results a great risk in the model selection. For this reason, we choose k non-nested models and investigate the model which is closer to the true density. In this paper, we investigate this main question in the model selection that how is it possible to gain a collection of appropriate models for the estimation of the true density function h, based on Kullback-Leibler risk.

Keywords: Candidate Model, KullbacK Leibler Risk; Model Selection, Non-nested Models, Quasi Maximum Likelihood Estimator.

Mathematics Subject Classification (2000): 26D15