دانشگاه رازی

دانشکده علوم
گروه آمار

پایان نامه کارشناسی ارشد آمار ریاضی

انتخاب مدل بر اساس معیار کولبک–لیبلر برای مشاهدات سانسوریده

توسط

پریسا ترکمان

استاد راهنما

دکتر عبدالرضا سیاره

استاد مشاور

۱۳۸۹

ماندگارترین ذات لایتناهی،

او که ذاتش یگانه، مهرش جاودانه و علمش بی کرانه است.

تقدیم به‌پژوهشگران عرصه علم و هنر،

تقدیم به خانواده‌ی عزیزم

و تقدیم به همه‌ی انسان هایی که به شادی دیگران شاد می شوند

و به غم آنان غمگین

انتخاب مدل بر اساس معیار کولبک—لیبلر برای مشاهدات سانسوریده

چکیده

انتخاب مدل به منظور استنباط و پیش بینی رفتار آینده جوامع تحت بررسی از اهمیت ویژه‌ای برخوردار است. از این رو جز با انتخاب مدل دسترسی به دنیای واقعی داده‌ها غیر ممکن است. هدف انتخاب مدل، انتخاب مدل بهینه بر پایه مشاهدات جامعه از میان مجموعه مدل‌های رقابتی است. مطالعات گسترده‌ایی در زمینه انتخاب مدل و آزمون فرض برای مشاهدات کامل انجام گرفته است. از جمله مشاهدات ناقص، مشاهدات سانسوریده هستند که به دلیل صرفه‌جویی در زمان و هزینه و همچنین کاربرد فراوان در آزمون‌های طول عمر، آنالیز بقا و تئوری قابلیت اعتماد، جایگاه خاصی دارند. در این پایان نامه به موضوع انتخاب یک مدل مناسب از بین مدل‌های رقابتی برای داده‌های سانسوریده پرداخته می شود. در بررسی مشاهدات سانسوریده از راست تصادفی برآوردکننده‌های درست نمایی ماکسیمم تقریبی و تقریبی جک نایف را معرفی کرده و بر اساس آن معیار انتخاب مدل برای این نوع سانسور را به دست می آوریم. با مطالعه شبیه سازی به مقایسه برآوردکننده درست نمایی ماکسیمم و این دو برآوردکننده تحت سانسور راست تصادفی به نتایج قابل توجهی دست می یابیم. مبنی بر آنکه وقتی مدل رقابتی بد—توصیف شده باشد برآوردکننده درست نمایی ماکسیمم تقریبی و تقریبی جک نایف بهتر از برآوردکننده درست نمایی ماکسیمم در مینیمم کردن معیار کولبک—لیبلر رفتار خواهد کرد. نشان داده خواهد شد که استنباط براساس داده‌های مشاهده شده و سانسوریده به طور همزمان به جای در نظر گرفتن فقط داده‌های مشاهده شده به نتایج بهتری منتهی خواهد شد و یک فاصله رد یابی مناسب که مجموعه ایی از فرض‌های پذیرفتنی تحت فرض صفر است، برای تفاضل امید ریسک‌های کولبک‌ـلیبلر مشاهدات سانسوریده از راست نوع II بدست آوردیم.

فرض کنید $X_1, \ldots, X_n$ از توزیع نامعلوم h و مدل رقابتی $f(.,\theta)$ برای برآورد توزیع نامعلوم h در نظر گرفته شده است. با استفاده از شرایط وایت (۱۹۸۲)، برآوردکنندهٔ شبه درست نمایی ماکسیمم، $\hat{\theta}_n$ از رابطهٔ زیر برای مشاهدات کامل زیر به دست می آید.

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} l_f(\theta).$$

که در آن $l_f(\theta)$، لگاریتم تابع شبه درست نمایی است و از آنجا که همگرایی در احتمال زیر برقرار است

$$\frac{1}{n} l_f(\theta) \xrightarrow{P} E_h[\log f(X;\theta)]$$

در نتیجه برآوردکنندهٔ شبه درست نمایی ماکسیمم که از ماکسیمم کردن $\frac{1}{n} l_f(\theta)$ به دست می آید در احتمال به مقداری که $E_h[\log f(X;\theta)]$ را ماکسیمم می کند همگرا خواهد شد که در آن E_h امید ریاضی تحت مدل درست است.

در حالت سانسور راست نوع II برآوردکننده شبه درست نمایی ماکسیمم که از حل رابطه

$$\frac{\partial}{\partial\theta} l_{fc}(\theta) = \sum_{i=1}^{r} \frac{\partial}{\partial\theta} \log f(y_i;\theta) + (n-r)\frac{\partial}{\partial\theta} \log \bar{F}(y_r;\theta) = 0$$

به دست می‌آید. با توجه به قضیه دوم باتاچاریا (۱۹۸۵) $\frac{1}{n}$ جمله اول و جملهٔ دوم رابطه فوق به ترتیب به $pE_{\frac{h}{H(\zeta)}}[\log f(Y;\theta)]$ و $(1-p)\log \bar{F}(\zeta;\theta)$ همگرا خواهد بود به طوریکه $\frac{h}{H(\zeta)}$ امید ریاضی تحت مدل درست بریده شده از راست است. بنابراین برآوردکنندهٔ شبه درست نمایی ماکسیمم در احتمال به مقدار شبه درست پارامتری همگرا می‌شود که $pE_{\frac{h}{H(\zeta)}}[\log f(Y;\theta)] + (1-p)\log \bar{F}(\zeta;\theta)$ را ماکسیمم کند که آن را با θ_{*c} نشان داده و رابطه زیر را داریم:

$$\theta_{*c} = \arg\max_{\theta\in\Theta}\{pE_{\frac{h}{H(\zeta)}}[\log f(Y;\theta)] + (1-p)\log \bar{F}(\zeta;\theta)\}.$$

واژه‌های کلیدی : برآوردکننده شبه درست نمایی ماکسیمم، داده‌های سانسوریده، فاصله ردیابی، مدل‌های غیر آشیانه‌ایی، معیار آکائیک، معیار اطلاع کولبک-لیبلر.

فهرست مندرجات

فصل ۱

مفاهیم و تعاریف اولیه

۱—۱ مقدمه

دراین فصل به بیان تعاریف و مفاهیم مقدماتی و قضایای مورد نیاز فصل‌های بعدی می‌پردازیم. تابع درست نمایی و خصوصیات مجانبی برآوردکننده درست نمایی ماکسیمم در دو حالت که مدل رقابتی چگالی درست داده‌ها را شامل باشد و یا نباشد، مورد مطالعه قرار می‌گیرد. در پایان فصل به بررسی معیار اطلاع کولبک—لیبلر[۱] به عنوان معیاری برای بررسی میزان واگرایی مدل درست از مدل رقابتی پرداخته شد. معیار اطلاع آکائیک [۲] به عنوان برآوردگر نااریبی برای معیار اطلاع کولبک—لیبلر و گسترشی از آن تحت عنوان KIC معرفی می‌شود.

فرض کنید $X^n = \{X_n\}_{n\geq ۱}$ یک دنباله از متغیرهای تصادفی $i.i.d$ باشد.

تعریف ۱.۱ دنباله X^n تقریباً همه جا[۳] به متغیر تصادفی X همگرا است، اگر به ازای هر $\varepsilon > ۰$:

$$P(\lim_{n\longrightarrow+\infty} |X_n - X| < \varepsilon) = ۱,$$

و با نماد $X_n \xrightarrow{a.s} X$ نشان داده می شود.

تعریف ۲.۱ دنباله X^n در احتمال به متغیر تصادفی X همگرا است، اگر به ازای هر $\varepsilon > ۰$:

$$\lim_{n\longrightarrow+\infty} P(|X_n - X| < \varepsilon) = ۱,$$

در این صورت می‌نویسیم، $X_n \xrightarrow{P} X$.

تعریف ۳.۱ دنباله X^n در توزیع به متغیر تصادفی X همگرا است، اگر به ازای تمام نقاط پیوستگی F:

$$\lim_{n\longrightarrow+\infty} F_{X_n}(x) = F_X(x),$$

و با نماد $X_n \xrightarrow{L} X$ نشان داده می شود.

---

۱ $Kullback - Leibler\ information$

۲ $Akiake$

۳ $Almost\ surely$

تعریف ۴.۱ $o_p(1)$ نشان دهنده کمیتی است که در احتمال به صفر همگرا است و $O_p(1)$ نشان دهنده کمیتی است که در احتمال کراندار است وقتی n به سمت بی نهایت میل می‌کند.

قضیه ۱.۱ (**قانون ضعیف اعداد بزرگ**[۴]): فرض کنید X^n یک دنباله از متغیرهای تصادفی $i.i.d$ با $E(X_i) = \mu$ و $Var(X_i) = \sigma^2 < \infty$ باشد و $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$، آنگاه به ازای هر $\varepsilon > 0$:

$$\lim_{n \longrightarrow +\infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1,$$

به این معنا که $\bar{X}_n$ در احتمال به μ همگرا است. به عبارت دیگر وقتی $n \longrightarrow +\infty$ آنگاه:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu.$$

قضیه ۲.۱ (**قانون قوی اعداد بزرگ**[۵]): فرض کنید X^n یک دنباله از متغیرهای تصادفی $i.i.d$ با $E(X_i) = \mu$ و $Var(X_i) = \sigma^2 < \infty$ باشد و $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$، آنگاه به ازای هر $\varepsilon > 0$:

$$P(\lim_{n \longrightarrow +\infty} |\bar{X}_n - \mu| < \varepsilon) = 1,$$

به این معنا که $\bar{X}_n$ تقریباً همه جا به μ همگرا است. به عبارت دیگر وقتی $n \longrightarrow +\infty$ آنگاه:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s} \mu.$$

قضیه ۳.۱ (**قانون قوی اعداد بزرگ یکنواخت**[۶]): فرض کنید $f(x,\theta)$ یک تابع پیوسته از $\theta \in \Theta$ باشد. برای θ ثابت، $f(X_1,\theta), f(X_2,\theta), \ldots$ یک دنباله از متغیرهای تصادفی مستقل و همتوزیع است، به طوری که میانگین این دنباله در احتمال همگرا به $E[f(X,\theta)]$ خواهد بود اگر شرایط زیر برقرار باشد:

الف) Θ فشرده است.

---

۴ *Weak Law of Large Numbers*

۵ *Strong Law of Large Numbers*

۶ *Uniform strong Law of Large Numbers*

ب) یک تابع مانند $d(x)$ وجود دارد به طوری که $E[d(X)] < \infty$ و $|f(x,\theta)| < d(x)$ برای $\theta \in \Theta$، بنابراین $E[f(X,\theta)]$ در θ پیوسته است و رابطه همگرایی:

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n} f(x_i,\theta) - E[f(X,\theta)]\right| \xrightarrow{P} 0,$$

را خواهیم داشت.

قضیه ۴.۱ (قضیه اسلوتسکی [۷]): فرض کنید X^n و Y^n دو دنباله از متغیرهای تصادفی $i.i.d$ باشند. اگر $X_n \xrightarrow{L} X$، $Y_n \xrightarrow{P} c$ و c یک مقدار ثابت باشد آنگاه:

الف) $X_nY_n \xrightarrow{L} cX$

ب) $X_n + Y_n \xrightarrow{L} X + c$

پ) $\frac{X_n}{Y_n} \xrightarrow{L} \frac{X}{c}$ به شرط آنکه $c \neq 0$ و $P(Y_n = 0) = 0$

قضیه ۵.۱ (قضیه حد مرکزی چند متغیره [۸]): فرض کنید $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$ یک دنباله از بردارهای تصادفی k–بعدی $i.i.d$ از بردار تصادفی k–بعدی بابردار $E(\underline{X}_i) = \mu$ و ماتریس واریانس–کوواریانس $\Sigma = E[(\underline{X}_i - \mu)(\underline{X}_i - \mu)']$ باشد و $\bar{\underline{X}}_n = \frac{1}{n}\sum_{i=1}^{n}\underline{X}_i$ آنگاه وقتی $n \longrightarrow +\infty$:

$$\sqrt{n}(\bar{\underline{X}}_n - \mu) \xrightarrow{L} N_k(\underline{0}, \Sigma).$$

قضیه ۶.۱ (قضیه حد مرکزی لیندبرگ–فلر [۹]): فرض کنید X^n یک دنباله از متغیرهای تصادفی مستقل باشد، به طوری که $E(X_i) = \mu_i$ و $VarX_i = \sigma_i^2 < \infty$ و نیز $T_n = \sum_{i=1}^{n} X_i$، $E(T_n) = \sum_{i=1}^{n} E(X_i)$ و $s_n^2 = \sum_{i=1}^{n}\sigma_i^2$ باشد، آنگاه $Z_n = \frac{T_n - E(T_n)}{s_n}$ دارای توزیع نرمال با میانگین صفر و واریانس ۱ است، هنگامی که برای $n \to \infty$ شرط زیر برقرار است:

$$\forall\varepsilon > 0 \qquad \frac{1}{s_n^2}\sum_{i=1}^{n} E\left\{(X_i - \mu_i)^2 I_{|X_i-\mu_i|>\varepsilon s_n}\right\} \longrightarrow 0.$$

---

۷ *Slutesky Theorem*

۸ *Multivariate Central Limit Theorem*

۹ *Lindeberg − Feller Central Limit Theorem*

۱—۲ تابع درست نمایی

روش درست نمایی ماکسیمم که به اختصار با ML نمایش داده می‌شود یکی از قدیمی ترین و پر اهمیت‌ترین روش ها در نظریه برآورد پارامترها است. این روش متداول توسط گوس (۱۸۲۱) و بعدها توسط فیشر (۱۹۲۵) به صورت گسترده‌تری مورد بررسی قرار گرفت. روش ML برای بدست آوردن برآوردکننده درست نمایی ماکسیمم است که مبتنی بر تابع درست نمایی است.

تعریف ۵.۱ (تابع درست نمایی) : فرض کنید X^n یک دنباله از متغیرهای تصادفی با تابع چگالی احتمال $f(x,\theta)$ و پارامتر $\theta \in \Theta \subset R^p$ باشند. برای هر مقدار داده شده $X = x$، تابع درست نمایی X را تابع چگالی احتمال $f(X,\theta)$ تعریف کرده که به صورت تابعی از θ در نظر گرفته می‌شود. لذا

$$L_f(\theta) = \prod_{i=1}^{n} f(x_i,\theta),$$

را به عنوان تابع درست نمایی تعریف می‌کنیم. معمولاً لگاریتم درست نمایی $L_f(\theta)$ یعنی $l_f(\theta) = \log L_f(\theta)$ مورد استفاده قرار می‌گیرد.

تعریف ۶.۱ (برآوردکننده درست نمایی ماکسیمم MLE)[۱۰] : فرض کنید X^n یک دنباله از متغیرهای تصادفی از توزیعی با خانواده چگالی‌های $\{f(.,\theta), \theta \in \Theta \subset R^p\}$ باشد، آنگاه $\hat{\theta}_n$ برآوردکننده درست نمایی ماکسیمم است اگر در شرط

$$l_f(\hat{\theta}_n) = \sup_{\theta\in\Theta} l_f(\theta),$$

صدق کند، به طوری که برآوردکننده درست نمایی ماکسیمم دارای دو خصوصیت زیر است:

$$P(\hat{\theta}_n \in \Theta) = 1 \qquad \forall\theta \in \Theta \qquad (1$$

$$L_f(\hat{\theta}_n) \geq L_f(\theta) \qquad \forall\theta \in \Theta \qquad (2$$

خواص خوب مجانبی برآوردکننده درست نمایی ماکسیمم، این برآوردکننده را نسبت به برآوردکننده‌های دیگر ممتاز کرده است. تحت شرایط نظم[۱۱] این برآوردکننده سازگار است. شرایط

[۱۰] *Maximum Likelihood Estimator*

[۱۱] *Regularity Conditions*

نظم که در بررسی خواص مجانبی برآوردکننده درست نمایی ماکسیمم لازم است، در زیر آورده شده است:

شرایط نظم برای تابع چگالی $f(x;\theta)$ را بصورت زیر در نظر می‌گیریم.

شرط الف۱) تابع چگالی $f(x;\theta)$ (در واقع $\log f(x;\theta)$) نسبت به پارامتر θ دارای مشتق اول و دوم پیوسته است.

شرط الف۲) $\theta_\circ \in \Theta$ یک جواب معادلهٔ

$$\int f(x;\theta)\frac{\partial \log f(x;\theta)}{\partial\theta}dx = \circ,$$

باشد.

شرط الف۳) توابع انتگرال پذیر $F_۱(x)$ و $F_۲(x)$ روی اعداد حقیقی R وجود دارند به طوری که نامساوی های

$$|\frac{\partial \log f(x,\theta)}{\partial\theta}| < F_۱(x), \qquad |\frac{\partial^۲ \log f(x;\theta)}{\partial\theta\partial\theta'}| < F_۲(x).$$

به ازای تمام θ های عضو Θ برقرار باشند.

شرط الف۴) نامساوی

$$\circ < \int f(x;\theta)\frac{\partial \log f(x;\theta)}{\partial\theta}\frac{\partial \log f(x;\theta)}{\partial\theta'}dx < \infty,$$

به ازای تمام θ های عضو Θ برقرار است.

با استفاده از قضیه حد مرکزی و شرایط نظم بیان شده، قضیه زیر را بیان می‌کنیم.

قضیه ۷.۱ برآوردکننده درست نمایی ماکسیمم $\hat{\theta}_n$، در احتمال به مقدار $\theta_\circ \in \Theta$ همگرا و دارای توزیع مجانبی

$$n^{۱/۲}(\hat{\theta}_n - \theta_\circ) \xrightarrow{L} N(\circ, I^{-۱}(\theta_\circ)),$$

است که در آن $\theta_\circ$ پارامتر چگالی درست داده‌ها و $I(\theta)$ ماتریس اطلاع فیشر به صورت

$$I(\theta) = E_f[\frac{\partial \log f(X;\theta)}{\partial\theta}\frac{\partial \log f(X;\theta)}{\partial\theta'}],$$

است.

در روش‌های کلاسیک در یک خانواده پارامتری به دنبال عضوی از این خانواده هستیم که داده‌ها از آن تولید شده‌اند. در این حالت $\theta \in \Theta$ وجود دارد به طوری که مشاهدات از $f(x,\theta)$ آمده‌اند که آن را با $\theta_{\circ}$ نشان داده‌ایم. حال اگر مدل رقابتی مدل درست داده‌ها را شامل نشود، آیا برآوردکننده درست نمایی ماکسیمم وجود دارد؟ آیا در صورت وجود خواص مجانبی خوبی از خود نشان می‌دهد؟ قبل از بررسی این سؤالات لازم است با تعاریف زیر آشنا شویم.

فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی $i.i.d$ از توزیعی با تابع چگالی احتمال درست و نامعلوم $h(.)$ باشد.

تعریف ۷.۱ مدل $F_\theta = \{f(.,\theta), \theta \in \Theta \subseteq R^p\}$ را خوب—توصیف شده [۱۲] گوییم هرگاه $\theta_{\circ}$ متعلق به Θ وجود داشته باشد به طوری که $h(x) = f(x, \theta_{\circ})$.

تعریف ۸.۱ مدل $F_\theta = \{f(.,\theta), \theta \in \Theta \subseteq R^p\}$ را بد—توصیف شده [۱۳] گوییم هرگاه $\theta_{\circ}$ متعلق به Θ وجود نداشته باشد به طوری که $h(x) = f(x, \theta_{\circ})$.

تعریف ۹.۱ فرض کنید دو خانواده $F_\theta = \{f(.,\theta), \theta \in \Theta \subseteq R^p\}$ و $G_\gamma = \{g(.,\gamma), \gamma \in \Gamma \subseteq R^q\}$ برای تخمین چگالی درست نامعلوم h پیشنهاد شده اند. این دو خانوادع نسبنت به هم غیر آشیانه‌ایی هستند اگر $F_\theta \cap G_\gamma = \emptyset$، در غیر این صورت آشیانه ایی نامیده می‌شوند.

وایت (۱۹۸۲) برای حالتی که مدل رقابتی بد—توصیف شده باشد، برآوردکننده شبه درست نمایی ماکسیمم [۱۴] را بر اساس فرضیاتی محاسبه و خواص مجانبی آن را مورد بررسی قرار داد. اگر مدل بد—توصیف شده باشد چون مشاهدات $x_1, x_2, \ldots, x_n$ از توزیع واقعی آمده و در تابع لگاریتم درست نمایی و برآوردکننده درست نمایی ماکسیمم مدل رقابتی مورد استفاده قرار می‌گیرند در نتیجه با تعاریف، تابع شبه لگاریتم درست نمایی[۱۵] و برآوردکننده شبه درست نمایی ماکسیمم [۱۶] روبه رو خواهیم بود.

تعریف ۱۰.۱ تابع شبه لگاریتم درست نمایی: فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی $i.i.d$ از توزیعی با تابع چگالی احتمال درست و نامعلوم $h(.)$ باشد. اگر $f(x,\theta) \in F_\theta$ مدل رقابتی باشد آنگاه

۱۲ $Well\ Specified$

۱۳ $Mis\text{-}Specified$

۱۴ $Quasi\ Maximum\ Likelihood\ Estimator$

۱۵ $Quasi\ Log\ Likelihood\ Function$

۱۶ $Quasi\ Maximum\ Likelihood\ Estimator$

تابع شبه لگاریتم درست نمایی به صورت زیر تعریف می‌شود.

$$l_f(\theta) = \sum_{i=۱}^{n} \log f(x_i, \theta).$$

تعریف ۱۱.۱ برآوردکننده شبه درست نمایی ماکسیمم: فرض کنید $X_n, \ldots, X_۲, X_۱$ یک نمونه تصادفی $i.i.d$ از توزیعی با تابع چگالی احتمال درست و نامعلوم $h(.)$ باشد. $\hat{\theta}_n$ برآوردکننده شبه درست نمایی ماکسیمم است اگر در شرط

$$l_f(\hat{\theta}_n) = \sup_{\theta \in \Theta} l_f(\theta),$$

صدق کند.

در حالتی که مدل رقابتی خوب–توصیف شده باشد، برآوردکننده شبه درست نمایی ماکسیمم با برآوردکننده درست نمایی ماکسیمم یکسان است. برآوردکننده شبه درست نمایی در احتمال همگرا به مقدار شبه درست پارامتر است.

تعریف ۱۲.۱ مقدار شبه درست پارامتر [۱۷] : مقدار شبه درست پارامتر برای مدل رقابتی $f(x; \theta)$ با θ_* نشان داده می‌شود و از رابطه

$$\theta_* = \arg\max_{\theta \in \Theta} E_h\{\log f(X; \theta)\},$$

به دست می آید. چنانچه مدل رقابتی خوب–توصیف شده باشد $\theta_* = \theta_۰$.

برای به دست آوردن توزیع مجانبی برآوردکننده شبه درست نمایی ماکسیمم نیاز به شرایط وایت است که در زیر آورده شده است.

شرایط وایت (۱۹۸۲):

شرط ب۱) برای هر $\theta \in \Theta$، $|\log f(x; \theta)| \leq m(x)$ که در آن $m(.)$ تابعی انتگرال پذیر نسبت به h است و $E_h[\log f(X; \theta)]$ دارای ماکسیمم یکتا در $\theta_* \in \Theta$ است.

شرط ب۲) $\frac{\partial \log f(x;\theta)}{\partial \theta}$ تابعی اندازه پذیر از x برای هر $\theta \in \Theta$ و تابعی پیوسته و مشتق پذیر از θ برای همه x های متعلق به X است که در آن X فضای متغیرهای تصادفی است.

شرط ب۳) $|\frac{\partial^۲ \log f(x;\theta)}{\partial \theta \partial \theta'}| \leq m(x)$ و $|\frac{\partial \log f(x;\theta)}{\partial \theta} \frac{\partial \log f(x;\theta)}{\partial \theta'}| \leq m'(x)$ با شرط آنکه $m(.)$ و $m'(.)$ توابع

[۱۷] *Pesudo True Value of a Parameter*

انتگرال پذیر نسبت به h برای همهٔ x ها و $\theta \in \Theta$ ها است.

شرط ب۴) θ_* یک نقطه درونی Θ است و $J(\theta)$ و$I(\theta)$ ماتریس های وارون پذیری هستند.

شرط (ب۱) نشان می‌دهد که $E_h[\log f(X;\theta)]$ خوش تعریف است و $\hat{\theta}_n$ برآورد سراسری θ است.

قضیه ۸.۱ فرض کنید $X_1, X_2, \ldots, X_n$ متغیرهای تصادفی $i.i.d$ از چگالی درست و نامعلوم h و $f(x;\theta)$ مدل رقابتی باشد. تحت شرایط وایت برای n به اندازه کافی بزرگ،

$$\hat{\theta}_n \xrightarrow{a.s} \theta_*,$$

و توزیع مجانبی این آماره به صورت

$$n^{1/2}(\hat{\theta}_n - \theta_*) \xrightarrow{L} N_p(0, J^{-1}(\theta_*)I(\theta_*)J^{-1}(\theta_*)),$$

است که در آن

$$\begin{aligned} J(\theta) &= -E_h\Big[\frac{\partial^2 \log f(X;\theta)}{\partial\theta\partial\theta'}\Big] \\ &= -\int h(x)\frac{\partial^2 \log f(x;\theta)}{\partial\theta\partial\theta'}dx, \end{aligned}$$

و

$$\begin{aligned} I(\theta) &= E_h\Big[\frac{\partial \log f(X;\theta)}{\partial\theta}\frac{\partial \log f(X;\theta)}{\partial\theta'}\Big] \\ &= \int h(x)\frac{\partial \log f(x;\theta)}{\partial\theta}\frac{\partial \log f(x;\theta)}{\partial\theta'}dx. \end{aligned}$$

برهان. با توجه به این که $\frac{\partial l_f(\theta)}{\partial\theta}|_{\theta=\hat{\theta}_n} = 0$، بسط تیلور $\frac{\partial l_f(\theta)}{\partial\theta}|_{\theta=\hat{\theta}_n}$ حول نقطه θ_* به صورت:

$$0 = \frac{\partial l_f(\theta)}{\partial\theta}|_{\theta=\hat{\theta}_n} = \frac{\partial l_f(\theta)}{\partial\theta}|_{\theta=\theta_*} + (\hat{\theta}_n - \theta_*)\frac{\partial^2 l_f(\theta)}{\partial\theta\partial\theta'}|_{\theta=\theta_*} + o_p(1),$$

است. از قضیه حد مرکزی داریم که:

$$n^{-1/2}\frac{\partial l_f(\theta)}{\partial\theta}|_{\theta=\theta_*} \xrightarrow{L} N(0, I(\theta_*)),$$

و با استفاده قانون ضعیف اعداد بزرگ

$$-\frac{1}{n}\frac{\partial^2 l_f(\theta)}{\partial\theta\partial\theta'}|_{\theta=\theta_*} \xrightarrow{P} J(\theta_*),$$

لذا:

$$n^{1/2}J(\theta_*)(\hat{\theta}_n - \theta_*) = n^{-1/2}\frac{\partial l_f(\theta_*)}{\partial\theta},$$

در نتیجه $n^{1/2}(\hat{\theta}_n - \theta_*)$ دارای توزیع مجانبی نرمال p—متغیره با میانگین صفر و ماتریس واریانس—کوواریانس $J^{-1}(\theta_*)I(\theta_*)J^{-1}(\theta_*)$ خواهد بود. برآوردکننده شبه درست نمایی ماکسیمم یک برآوردکننده سازگار برای θ_* است. اگر مدل رقابتی خوب—توصیف شده باشد، $\theta_* = \theta_0$ و ماتریس واریانس—کوواریانس بالا به $I^{-1}(\theta_0)$ تقلیل می‌یابد.

بین $J(\theta_*)$ و $I(\theta_*)$ رابطه زیر برقرار است.

$$
\begin{aligned}
-J(\theta) = \int h(x)\frac{\partial^2 \log f(x;\theta)}{\partial\theta\partial\theta'}dx &= \int h(x)\frac{\partial}{\partial\theta}\{\frac{\partial}{\partial\theta}\log f(x;\theta)\}dx \\
&= \int h(x)\frac{\partial}{\partial\theta}\{\frac{1}{f(x;\theta)}\frac{\partial}{\partial\theta}f(x;\theta)\}dx
\end{aligned}
$$

$$
= \int h(x)\frac{1}{f(x;\theta)}\frac{\partial^2 f(x;\theta)}{\partial\theta\partial\theta'}dx - \int h(x)\frac{1}{f^2(x;\theta)}\frac{\partial f(x;\theta)}{\partial\theta}\frac{\partial f(x;\theta)}{\partial\theta'}dx, \qquad (1.2.1)
$$

اگر مدل رقابتی شامل چگالی درست داده‌ها باشد به این معنا که θ_0ی متعلق به Θ وجود دارد به طوری که، $h(x) = f(x,\theta_0)$، بنابراین

$$
\begin{aligned}
\int h(x)\frac{1}{f(x;\theta_0)}\frac{\partial^2 f(x;\theta_0)}{\partial\theta\partial\theta'}dx &= \int \frac{\partial^2 f(x;\theta_0)}{\partial\theta\partial\theta'}dx. \\
&= \frac{\partial^2}{\partial\theta\partial\theta'}\int f(x;\theta_0)dx = 0.
\end{aligned}
$$

بنابراین جمله اول سمت راست رابطه (۱.۲.۱) صفر می‌شود و خواهیم داشت:

$$
\int h(x)\frac{\partial^2 \log f(x;\theta_0)}{\partial\theta\partial\theta'}dx = -\int h(x)\frac{\partial \log f(x;\theta_0)}{\partial\theta}\frac{\partial \log f(x;\theta_0)}{\partial\theta'}dx,
$$

در نتیجه $J(\theta_0) = I(\theta_0)$.

۱—۳ معیار اطلاع کولبک—لیبلر

مدل‌های آماری برای بیان ساختار تصادفی، پیش بینی، استنباط و استخراج اطلاعات با کمک داده‌های مشاهده شده به کار گرفته شده‌اند. در حقیقت مدل آماری یک توزیع احتمالی است که با استفاده از داده‌ها، تقریبی برای توزیع درست پدیده احتمالی ارائه می‌کند. با مشخص شدن مدل می‌توان از آن برای کنترل و تصمیم گیری در مورد رفتار آینده جامعه استفاده کرد و راهگشای مسائلی در مورد جامعه

شد که تنها با ساخت مدل، می‌توان به شناخت درستی از جامعه رسید. در بررسی نزدیکی مدل رقابتی به مدل درست داده‌ها به دنبال مدلی هستیم که بهترین انتخاب برای مدل درست و نامعلوم باشد.

یکی از معیارهای انتخاب مدل، معیار اطلاع کولبک–لیبلر (۱۹۵۱) است که بر پایه تابع مخاطره بیان شده است. فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی $i.i.d$ از توزیعی با تابع چگالی درست نامعلوم h و مدل رقابتی $f(.,\theta)$ برای h در نظر گرفته شده باشد. زیان استفاده از چگالی $f(.,\theta)$ به جای چگالی درست h برای مشاهده X را بصورت $\log \frac{h(X)}{f(X;\theta)}$ نشان می‌دهیم. امید ریاضی این تابع زیان تحت چگالی درست h مخاطره کولبک–لیبلر نامیده می‌شود.

این مخاطره را با نماد اختصاری KL نشان داده و به صورت

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

$$= E_h[\log h(X)] - E_h[\log f(X;\theta)],$$

تعریف می‌کنیم. که در آن E_h بیانگر امید ریاضی نسبت به چگالی درست h است. به دلیل آن که این معیار شامل توزیع نامعلوم h است، معیار کولبک–لیبلر قابل محاسبه نیست. لذا به منظور دستیابی به مدل مطلوب در میان مدل‌های رقابتی و همچنین کوچک کردن معیار کولبک–لیبلر در جستجوی $\theta \in \Theta$ هستیم که جمله دوم کولبک–لیبلر را ماکسیمم کند که در نهایت معیار کولبک–لیبلر مینیمم شود. جمله دوم همان امید لگاریتم تابع درست نمایی است. اگر مدل رقابتی خوب–توصیف شده باشد برآوردکننده درست نمایی ماکسیمم $\hat{\theta}_n$ به سمت θ_0 همگرا و اگر مدل رقابتی بد–توصیف شده باشد برآوردکننده شبه درست نمایی ماکسیمم به سمت θ_* همگرا خواهد بود که در هر دو حالت ریسک کولبک–لیبلر را مینیمم خواهند کرد. چون جمله دوم کولبک–لیبلر امید ریاضی نسبت به چگالی درست داده‌ها است، باید برآورد شود. لذا آکائیک (۱۹۷۴) برآوردکننده نااریبی تحت عنوان معیار اطلاع آکائیک برای این ریسک معرفی کرد که به بیان آن خواهیم پرداخت. اما پیش از آن به دو خصوصیت اساسی معیار کولبک–لیبلر اشاره می‌کنیم.

معیارکولبک–لیبلر دارای دو خصوصیت زیر است:

۱) $KL(h, f_\theta) \geq 0$ و تساوی رخ خواهد داد، اگر وتنها اگر $\theta_0 \in \Theta$ وجود داشته باشد به طوری که $h(x) = f(x, \theta_0)$.

۲) $KL(h, f_\theta) \neq KL(f_\theta, h)$.

در بررسی خصوصیت (۱) تابع $K(t) = \log t - t + 1$ را که برای $t > 0$ تعریف شده است در نظر