دانشگاه رازی

دانشکده علوم
گروه آمار

پایان نامه کارشناسی ارشد آمار ریاضی

انتخاب مدل بر اساس معیار کولبک—لیبلر و کاربرد آن در مدل‌های رگرسیونی خطی غیرتودرتو

توسط

اسداله فقیهی

استاد راهنما

دکتر عبدالرضا سیاره

استاد مشاور

دکتر بهاءالدین خالدی

بهمن ماه ۱۳۸۸

انتخاب مدل بر اساس معیار کولبک–لیبلر و کاربرد آن در مدل‌های رگرسیونی خطی غیرتودرتو

چکیده

یکی از اساسی ترین مسائل بنیادی در انتخاب مدل بررسی دوری و نزدیکی مدل‌های پیشنهادی به مدل مولد داده‌ها یعنی $h(.)$ است. برای حالتی که چند مدل پیشنهادی وجود دارد، بر اساس آزمون‌ها و معیارهای موجود تصمیم می‌گیریم که آیا این مدل‌ها را به‌عنوان مدل‌های خوب در نظر بگیریم یا به عنوان مدل‌هایی که برازش خوبی به داده‌ها ندارند. در فرآیند تصمیم‌گیری برای دو مدل پیشنهادی گاه دو مدل پیشنهادی را رد می‌کنیم. روش تصمیم‌گیری آزمون کاکس، رد یا پذیرش مدل‌های پیشنهادی بر اساس تعلق $h(.)$ به خانوادهٔ چگالی‌های پیشنهادی است. سوالی که مطرح می‌شود آن است که آیا این دو مدل، که توسط آزمون کاکس رد شده‌اند، دو مدل نزدیک به $h(.)$ بوده‌اند یا دو مدل دور از $h(.)$. از طرفی آزمون وونگ براساس دوری و نزدیکی مدل‌ها به $h(.)$ طراحی شده است. بر این اساس در این پایان نامه با شبیه‌سازی مدل‌های رگرسیونی خطی غیرتودرتو با خطای نرمال به بررسی موضوعاتی از این نوع پرداخته‌ایم که از دو مدل پیشنهادی که توسط آزمون کاکس پذیرفته نشده‌اند کدام یک به $h(.)$ نزدیک‌تر بوده‌اند. از طرفی چنانچه آزمون وونگ یا معیارهای انتخاب مدل مانند BIC،AIC،KIC،$KICc$، یکی از مدل‌های پیشنهادی را برگزینند آیا آن مدل توسط آزمون کاکس به عنوان یک مدل خوب–توصیف شده برگزیده می‌شود یا خیر. با بررسی این موضوعات می‌توان نتیجه گرفت که به دلیل حساسیت زیاد آزمون کاکس به ساختار مدل‌های پیشنهادی پذیرش مدل توسط آزمون کاکس به معنی پذیرش توسط همه معیارهاست، و معادل بودن دو مدل توسط آزمون وونگ به مفهوم رد هر دو توسط آزمون کاکس است. اما در مواردی که هر دو مدل توسط آزمون کاکس رد می‌شوند، آزمون وونگ و دیگر معیارها مدل نزدیکتر را برمی‌گزینند.

پیشگفتار

یکی از مسائل اساسی در آمار بررسی یک نمونه مستقل و هم توزیع از مشاهدات است. در واقع هرگاه پدیده‌های مورد بررسی تصادفی بوده و به‌طور کامل قابل پیش‌بینی نباشند، برای ساختاربندی این پدیده‌ها از مدل‌های آماری استفاده می‌کنیم. تابعی که قادر به توصیف یک پدیده‌ی تصادفی باشد را مدل آماری می‌گویند. به‌طور کلی از مدل‌ها برای نمایش دادن ساختار پدیده‌های تصادفی، پیش بینی رفتار متغیر در آینده و کسب اطلاعات مفید از داده‌ها استفاده می‌شود. در واقع مدل آماری یک توزیع احتمالی است که با استفاده از داده‌های مشاهده شده تقریبی برای توزیع درست پدیده احتمالی ارائه می‌کند. به عبارت دیگر ایده انتخاب مدل، ساختن مدلی است که ساختار درست داده‌ها را با استفاده از مشاهدات تقریب بزند. مشکل اساسی در تحلیل‌های آماری، انتخاب یک مدل اولیه‌ی مناسب برای داده‌ها است که معمولاً بر اساس نمودار پراکنش یا هیستوگرام یک مدل اولیه مناسب پیشنهاد می‌شود. تحلیل و انتخاب مدل‌ها در آمار به دو روش انجام می‌گیرد، روش آزمون فرض و انتخاب مدل بر اساس معیارهای انتخاب مدل. برخی اوقات در تصمیم‌گیری‌ها نیاز است حداقل مدلی انتخاب شود که بدترین مدل نباشد. این روش با آزمون فرض کلاسیک برای انتخاب مدل متفاوت است.

دو مدل ممکن است تودرتو، غیرتودرتو و یا اشتراک جزئی داشته باشند. آزمون فرض‌های غیرتودرتو در آمار توسط کاکس (۱۹۶۲،۱۹۶۱) و اتکینسن (۱۹۷۰) ارائه گردید. پسران (۱۹۷۴) و پسران و دیتون (۱۹۷۸) آن را در مدل‌های اقتصادی به‌کار بردند. آنالیز مدل‌های رگرسیونی غیرتودرتو در سال (۱۹۸۱) به‌وسیله دیویدسون و ماکینون انجام گرفت. فیشر و مک‌آلر(۱۹۸۱)، دستور(۱۹۸۳)، دیتون(۱۹۸۲) و ساویر(۱۹۸۳) از جمله کسانی هستند که به گسترش این تئوری پرداخته‌اند.

روش‌های دیگر آزمون فرض‌های غیرتودرتو از جمله روش .E.p[۱] و مدل‌های .C.h[۲] به‌وسیله میزون و ریچارد(۱۹۸۶)، گوریو[۳] ، مونفورت[۴] (۱۹۹۵) و اسمیت[5] (۱۹۹۳) مورد بحث قرار گرفت.

---

۱ Encompassing

۲ Comprehensive

۳ Gourieroux

۴ Monfort

۵ Smith

در آزمون فرض‌های غیرتودرتو معیار کولبک–لیبلر(۱۹۵۱،۱۹٦۸) نقش محوری دارد و اکثر آزمون‌ها و معیارهای انتخاب مدل به نوعی این معیار یا برآورد آن را به کار می‌گیرند. ساختار آزمون کاکس در واقع براساس تعمیم آماره لگاریتم نسبت درست نمایی است. آزمون وونگ براساس اختلاف معیارهای کولبک–لیبلر دو مدل پیشنهادی ساخته شده است و محاسبه‌ی آماره‌ی آن بسیار ساده‌تر از محاسبه‌ی آماره آزمون کاکس است. محاسبه‌ی آماره آزمون کاکس در بسیاری از موارد با تکنیک‌های آنالیزی مشکل است به همین دلیل پسران و پسران (۱۹۹۳) روش شبیه‌سازی تصادفی را ارائه کردند، آماره‌ای که آن‌ها به کار گرفتند برآورد کننده‌ای سازگار برای معیار کولبک–لیبلر است. از معیارهای انتخاب مدل ارائه شده، می‌توان معیار اطلاع آکائیک(AIC) را به عنوان یک معیار بسیار کاربردی نام برد که از آن برای انتخاب یک مدل از بین مدل‌های پیشنهادی استفاده می‌شود. از معیارهای دیگر، معیار اطلاع بیزی (BIC) (شوارتز(۱۹۷۸))، روش اعتبارسنجی متقابل[٦] معیار اطلاع ایشی گورو(EIC)(ایشی‌گورو و همکاران (۱۹۹۷)) و $ICOMP$ که توسط بزدگان (۲۰۰۰) ارائه شد. همچنین معیارهایی که بر اساس معیار کولبک متقارن ساخته می‌شوند، مانند KIC، به عنوان معیار انتخاب مدل به‌کارمی‌روند.

اخیراً توزیع مجانبی AIC در مدل‌های رگرسیونی خطی و تصحیح اریبی این آماره توسط یاناگی‌هارا[۷] و اوهوموتو[۸] (۲۰۰۵) مورد بحث قرارگرفته‌اند. همچنین کومانژ[۹] و همکاران (۲۰۰۸) اختلاف نرمال شده AIC را به عنوان برآورد اختلاف کولبک–لیبلرها بین دو مدل مورد بررسی قرار داده‌اند.

جینیوس و استرازرا (۲۰۰۲) نتایج دو آزمون کاکس، وونگ و معیار AIC را برای مدل‌های گسسته همزمان مورد بررسی قرار داده‌اند. ما علاقه‌مند به مقایسه بین نتایج آزمون‌های کاکس و وونگ و معیارهای AIC، KIC، $KICc$ و BIC برای مدل‌های رگرسیونی خطی با خطای نرمال هستیم. در این رابطه سوال‌هایی مطرح است. به عنوان مثال؛ مفهوم معادل بودن دو مدل توسط آزمون وونگ چیست؟ آیا این دو مدل، به مدل درست نزدیک و معادل هستند یا دور و معادل هستند؟ نتیجه‌ی آزمون کاکس در مورد آن‌ها چگونه است؟ هرگاه دو مدل توسط آزمون کاکس رد شوند، آزمون وونگ و معیارهای دیگر چه نتایجی به بار می‌آورند؟ برای پاسخ به این گونه سوالات به برآورد آماره‌ها و شبیه‌سازی مدل‌های رگرسیونی خطی غیرتودرتو در فصل ۴ پرداخته‌ایم.

در فصل ۱ به برخی تعاریف مورد نیاز به صورت کلی پرداخته شده است. در فصل ۲ تابع و برآورد درست نمایی و قضیه‌های مربوط به آن آورده شده است. در فصل ۳ آزمون فرض‌های تودرتو و

٦ Cross Validation

۷ Yanagihara

۸ Ohomoto

۹ Commenges

غیرتودرتو و همچنین معیارهای انتخاب مدل که بر پایه‌ی معیار اطلاع کولبک—لیبلر ساخته می‌شوند مورد بحث قرار گرفته‌اند، در فصل ۴ آماره آزمون‌های کاکس و وونگ و بقیه‌ی معیارها برای مدل‌های رگرسیونی خطی غیرتودرتو محاسبه شده و آنگاه به وسیله‌ی شبیه سازی به مقایسه‌ی نتایج آن‌ها و پاسخ‌گویی به سوالات فوق پرداخته‌ایم.

هرگاه در این پایان‌نامه از $f(.;\theta)$ به عنوان یک مدل نام برده‌ایم منظور $f(.;\theta)$ برای $\theta \in \Theta$ است.

فهرست مندرجات

لیست جداول

فصل ۱

تعاریف و قضایا

۱—۱ مقدمه

در این فصل به بیان قضایا و تعریف‌هایی پرداخته شده است که در فصل‌های بعد مورد استفاده قرار می‌گیرند.

در بخش (۲—۲) به معرفی اجمالی تابع درست نمایی و برآورد درست نمایی در آمار کلاسیک پرداخته شده است. بخش (۲—۳) شامل قضایای مجانبی و فرض‌های لازم در بحث همگرایی و شرایط نظم است. در بخش (۲—۴) قضیه‌های لازم از جمله قضایای مورد استفاده در بحث رگرسیونی در فصل (۴) آورده شده است. مفاهیم مورد استفاده در انتخاب مدل شامل مفاهیمی است که در آمار کلاسیک کمتر یا اصلاً به آنها پرداخته نمی‌شود در بخش (۲—۵) مورد بررسی قرار گرفته است. توزیع مجانبی برآورد کننده شبه درست نمایی ماکسیمم (QMLE) در بخش (۲—٦) مورد بررسی قرار گرفته است. در بخش (۲—۷) به معرفی معیار اطلاع کولبک—لیبلر و مدل‌های غیر تودرتو و رابطه آن‌ها با یکدیگر پرداخته شده است.

۱—۲ تابع و برآورد درست نمایی در آمار کلاسیک

روش درست نمایی ماکسیمم[۱] (ML) روشی شناخته شده در استنباط آماری به منظور برآورد پارامترها است. این روش اولین بار توسط گوس[۲] (۱۸۲۱) به کار گرفته شد و پس از آن به صورت گسترده‌تری در سال (۱۹۲۵) توسط فیشر[۳] مورد استفاده قرار گرفت. یکی از مهم‌ترین اصول در آمار که به اصل

[۱] Maximum Likelihood

[۲] Gauss

[۳] Fisher

درست نمایی مشهور است، بیان می‌کند که تابع درست نمایی شامل تمام اطلاعات موجود در نمونه درباره پارامتر یا پارامترهای نامعلوم است. روش ML عبارت است از دستورالعملی برای به دست آوردن برآوردگری به نام برآوردگر درست نمایی ماکسیمم که به اختصار با MLE نشان داده خواهد شد و مبتنی بر تابع درست نمایی است. خواص مجانبی خوب این برآوردکننده و اطلاع کاملی که از نمونه می‌دهد باعث شده است که توجه ما به معیار اطلاع کولبک—لیبلر که ارتباط نزدیکی با برآورد کننده‌های ML دارد، معطوف شود ساخته شوند.

در این بخش اصول تابع و برآورد درست نمایی مورد بررسی قرار گرفته است.

تعریف ۱.۱ (تابع درست نمایی) فرض کنید X_1،X_2،...،X_n متغیرهای تصادفی با تابع چگالی احتمال توأم (تابع احتمال توأم) $f(\mathrm{x};\theta)$ به طوری که $\theta \in \Theta \subseteq \Re^p$ باشند. در این صورت تابع درست نمایی عبارت است از $L_f(\theta) = f(x_1, \dots, x_n;\theta)$ که تابعی از θ است. برای نمونه‌های تصادفی و مستقل X_1،X_2،...،X_n ، تابع درست نمایی $L_f(\theta)$ تابعی با مقادیر حقیقی است که بر روی فضای پارامتر Θ تعریف می‌شود.

در تعریف بعد، روش برآورد درست نمایی ماکسیمم آورده شده است.

تعریف ۲.۱ (برآورد درست نمایی ماکسیمم) برای نمونه تصادفی X_1،X_2،...،X_n، $\hat{\theta}_n$ را برآوردگر درست نمایی ماکسیمم (MLE) برای θ گوئیم هرگاه

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

که در آن $\hat{\theta}_n \in \Theta$ است.

استفاده از زیرنویس n برای آماره $\hat{\theta}_n$ جهت نشان دادن وابستگی برآوردگربه تعداد نمونه مشاهده شده است. با این فرض که n بدون هیچ محدودیتی می‌تواند افزایش یابد.

یکی از مفیدترین خواص برآوردگرهای درست نمایی ماکسیمم خاصیت پایایی آن‌ها است. اگر $\rho = q(\theta)$ و q تابعی دلخواه و $\hat{\theta}_n$ برآوردگر درست نمایی ماکسیمم پارامتر θ باشند، آنگاه تابع $q(\hat{\theta}_n)$ برآوردگر ML پارامتر ρ است، زهنا[۴] (۱۹٦٦).

به طور کلی دو روش برای به دست آوردن MLE وجود دارد، وقتی که تکیه گاه داده‌ها به پارامتر وابسته است می‌توان با توجه به مقادیر مشاهده شده X_1،X_2،...،X_n که حدود آن‌ها به θ وابسته است، مشخص کرد که چه مقداری از θ، $L_f(\theta)$ را، ماکسیمم می‌کند. در حالتی که تکیه گاه داده‌ها به پارامتر وابسته نیست، فضای پارامتر ،Θ، یک مجموعه باز است. تابع درست نمایی نسبت به θ روی Θ مشتق

۴ Zehna

پذیر است و MLE پارامتر θ عبارت است از $\hat{\theta}_n$ که از حل معادله $\nabla_\theta \log L_f(\theta) = 0$ حاصل می‌شود. این معادله را معادله درست نمایی و $\log L_f(\theta)$ را تابع لگاریتم درست نمایی می نامیم. ∇_θ نشان دهنده عملگر گرادیان نسبت به θ است.

۱—۳ قضایای مجانبی برای MLE

تحت شرایط نظم[5] برآوردگرهای MLE سازگار و دارای توزیع مجانبی نرمال هستند. شرایط نظم در زیر آورده شده است.

(فرض الف ۱): فرض کنید X_1،X_2،...،X_n متغیرهای تصادفی .i.i.d با تابع چگالی احتمال (تابع احتمال) $f(x;\theta)$ هستند به طوری که $\theta \in \Theta$ است.

(فرض الف ۲): پارامترها شناسایی پذیرند[6] به این معنی که اگر $\theta_2 \neq \theta_1$ آنگاه $f(x;\theta_1) \neq f(x;\theta_2)$.

(فرض الف ۳): تکیه گاه متغیر تصادفی X، یعنی S، به θ بستگی نداشته باشد و $f(x;\theta)$ در θ مشتق پذیر باشد.

(فرض الف ۴): فضای پارامتر Θ شامل یک مجموعه باز است که مقدار صحیح پارامتر، θ_0، یک نقطهٔ داخلی آن است.

(فرض الف ۵): برای هر $x \in S$ چگالی $f(x;\theta)$ به طور متوالی سه بار نسبت به θ مشتق پذیر بوده و $\int_S f(x;\theta)dx$ نیز تحت انتگرال سه بار مشتق پذیر است.

(فرض الف ٦): برای $\theta_0 \in \Theta$ تابع $M(x)$ و مقدار مثبت C (ممکن است هر دو به θ_0 وابسته باشند) وجود دارند به طوری که

$$|\nabla_\theta^{(3)} \log f(x;\theta)| \leq M(x) \quad \forall x \in S,$$

به‌طوری‌که

$$E_{\theta_0}\{M(x)\} < \infty$$

و

$$|\theta_0 - \theta| < C\,.$$

لی کم[7] (۱۹۵۳) نشان داد تحت شرایط نظم برای همه nها، $\hat{\theta}_n$ یکتا موجود است.

۵ Regularity Condition

٦ Identifiable

۷ LeCam

قضیه ۱.۱ فرض کنید $X_1، X_2، \ldots، X_n$ متغیرهای تصادفی .i.i.d با چگالی $f(x;\theta)$ و $L_f(\theta) = \prod_{i=1}^n f(x_i;\theta)$ تابع درست نمایی و $\hat{\theta}_n$ نشان دهنده MLE پارامتر θ باشند. تحت شرایت نظم

$$\lim_{n\to\infty} P_\theta(|\hat{\theta}_n - \theta_0| \geq \varepsilon) = 0 \quad \forall \varepsilon > 0 \text{ و } \forall\, \theta \in \Theta$$

به عبارت ساده‌تر $\hat{\theta}_n \xrightarrow{p} \theta_0$.

با توجه به این که θ_0 حد $\hat{\theta}_n$ است بنابراین $(\hat{\theta}_n - \theta_0) = O_p(1)$ به این معنی که $(\hat{\theta}_n - \theta_0)$ در احتمال کران دار است و می توان تابع توزیع مجانبی آن را به دست آورد.

نماد $o_p(1)$ نشان دهنده کمیتی است که در احتمال همگرا به صفر است. همچنین نماد $O_p(1)$ نشان دهنده کمیتی است که با افزایش n در احتمال کراندار است.

قضیه ۲.۱ فرض کنید $X_1، X_2، \ldots، X_n$ متغیرهای تصادفی .i.i.d با چگالی $\theta \in \Theta$ $f(x;\theta)$ هستند. آنگاه تحت شرایط نظم

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L} N(0, I^{-1}(\theta_0)),$$

که در آن $I(\theta)$ نشان دهنده اطلاع فیشر است.

همان طور که گفته شد در حالت کلاسیک همیشه یادآوری می‌شود که $X_1، X_2، \ldots، X_n$ نمونه‌ای تصادفی از تابع چگالی احتمال (یاتابع احتمال) $f(x;\theta)$ هستند. در واقع $f(x;\theta)$ مدل درست[۸] یا مدل درست–توصیف شده[۹] است. حال این سوال مطرح می‌شود که اگر مدل $f(x;\theta)$ به عنوان مدل پیشنهادی درست–توصیف شده یا مدل درست فرض نشود، خواص MLE که بر شمرده شد هنوز برقرار هستند؟ آیا برآوردگرها در این حالت به مقدار صحیح خود همگرا هستند؟ توزیع آن‌ها چیست؟ و آیا با حالت کلاسیک تفاوتی دارد؟ وایت[۱۰] (۱۹۸۲) در مقاله خود بد–توصیف شدگی[۱۱] را بررسی و به این سوال‌ها پاسخ داد.

در بخش بعد تعریف‌ها و قضیه‌های لازم در همگرایی و همچنین قضیه‌هایی که در بخش رگرسیونی مورد استفاده قرار گرفته است، آورده شده‌اند.

---

۸ True Model

۹ Correctly Specified

۱۰ White

۱۱ Mis-Specification

۱—۴ مروری بر قضیه‌های لازم

قضیه ۳.۱ اگر $\{X_n, Y_n\}$ دنباله‌ای از متغیرهای تصادفی، C یک عدد ثابت و g تابعی پیوسته در نظر گرفته شود داریم

۱) اگر $X_n \xrightarrow{L} X$ و $Y_n \xrightarrow{P} 0$ آنگاه $X_nY_n \xrightarrow{P} 0$

۲) اگر $X_n \xrightarrow{L} X$ و $Y_n \xrightarrow{P} C$ آنگاه $X_n + Y_n \xrightarrow{L} X + C$، $X_nY_n \xrightarrow{L} XC$ و $X_n/Y_n \xrightarrow{L} X/C$ اگر $C \neq 0$.

۳) اگر $X_n \xrightarrow{P} X$ آنگاه $g(X_n) \xrightarrow{P} g(X)$،

۴) اگر $X_n \xrightarrow{P} X$ و $Y_n \xrightarrow{L} Y$ آنگاه $g(X_n) - g(Y_n) \xrightarrow{P} g(X) - g(Y)$.

جنریخ[۱۲] (۱۹۶۹) قضیه مقدار میانگین را برای متغیرهای تصادفی به کار برد که در زیر آورده شده است

لم ۱.۱ فرض کنید $f(x;\theta)$ تابع چگالی احتمال (تابع احتمال) است که برای هر x در S تابعی مشتق پذیر نسبت به θ است. همچنین θ_1 و θ_2 توابعی از S به Θ هستند، بنابراین تابع $\bar{\theta}_n$ از S به Θ وجود دارد به طوری که

۱) $f(x;\theta_1) - f(x;\theta_2) = \nabla_\theta f(x;\bar{\theta}_n)(\theta_1 - \theta_2)$

۲) $\bar{\theta}_n$ برای هر x در $\Re$ در فاصله θ_1 و θ_2 قرار می گیرد.

در ادامه قضایایی که نقش اساسی در این پایان نامه دارند، آورده شده است.

قضیه ۴.۱ (قضیه حد مرکزی)

فرض کنید X_1،X_2،...،X_n متغیرهای تصادفی i.i.d. با میانگین μ و واریانس متناهی σ^2 باشند، بنابراین

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{L} N(0, \sigma^2).$$

---

۱۲ Jennrich

قضیه ۵.۱ (قانون ضعیف اعداد بزرگ)

اگر X_1،X_2،...،X_n متغیرهای تصادفی .i.i.d با میانگین μ و واریانس متناهی σ^2 باشند، داریم

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

قضیه ۶.۱ (قانون قوی اعداد بزرگ)

اگر X_1،X_2،...،X_n متغیرهای تصادفی .i.i.d با میانگین μ و واریانس متناهی σ^2 باشند، آنگاه

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

به دلیل استفاده از مدل‌های خطی جهت شبیه سازی به قضایای زیر در فصل چهارم نیازمندیم.

۱—۴.۱ قضیه‌های مورد استفاده در مدل‌های خطی

در این بخش به معرفی شکل درجه دوم[۱۳] ماتریس‌های مربعی و محاسبه امید و توزیع آن‌ها پرداخته شده است. برای مطالعه بیشتر می‌توان کریستنسن[۱۴] (۱۹۹۶) را ملاحظه کرد.

تعریف ۳.۱ (شکل درجه دوم ماتریس)

Y بردار تصادفی n—بعدی و A ماتریس $n \times n$ فرض شده است. شکل درجه دوم یک ماتریس خود متغیر تصادفی است که به صورت $Y'AY$ تعریف می‌شود. باید توجه داشت که $Y'AY$ اسکالر است.

قضیه ۷.۱ با استفاده از تعریف (۳.۱) اگر $E(Y) = \mu$ و $Cov(Y) = V$ باشد بنابراین

$$E(Y'AY) = tr(AV) + \mu' A\mu$$

$tr(.)$ نشان دهنده اثر ماتریس است.

۱۳ Quadratic Form

۱۴ Christensen

قضیه ۸.۱ اگر $Y \sim N(\mu, V)$ باشد، بنابراین $Y'AY \sim \chi^{۲}(r(AV), \mu' A\mu)$ است اگریکی از حالت‌های زیر برقرار باشد.

۱) $VAVAV = VAV$ ۲) $\mu' AVA\mu = \mu' A\mu$ ۳) $VAVA\mu = VA\mu$.

$r(.)$ نشان دهنده رتبه ماتریس است.

۱—۵ تعاریف مورد استفاده در انتخاب مدل

فرض کنید $X_۱$،$X_۲$،...،X_n یک نمونه تصادفی .i.i.d از توزیعی با تابع چگالی احتمال (تابع احتمال) درست $h(.)$ باشد که در آن $h(.)$ نامعلوم است.

تعریف ۴.۱ (مدل درست—توصیف شده)

مدل $\mathcal{F}_\theta = \{f(.;\theta); \theta \in \Theta \subset \Re^p\}$ را درست—توصیف شده گوییم هرگاه $\theta_\circ \in \Theta$ وجود داشته باشد به طوری که $h \in \mathcal{F}_\theta$، به عبارت دیگر مدل $\mathcal{F}_\theta$ را درست—توصیف شده گوییم اگر $h(.) = f(.;\theta_\circ)$.

تعریف ۵.۱ (مدل بد—توصیف شده)

اگر $\theta_\circ \in \Theta$ وجود نداشته باشد به طوری که $h(.) = f(.;\theta_\circ)$، مدل $\mathcal{F}_\theta$ را بد—توصیف شده گویند و لذا $h \notin \mathcal{F}_\theta$.

در این حالت توابع شبه—درست نمایی ماکسیمم[۱۵] تعریف می‌شود که در ادامه به خواص آن‌ها خواهیم پرداخت.

۱—۵.۱ سازگاری

تعریف ۶.۱ (تابع شبه—لگاریتم درست نمایی)

فرض کنید $X_۱$،$X_۲$،...،X_n متغیرهای تصادفی .i.i.d از تابع چگالی نامعلوم $h(.)$ هستند. اگر $f(.;\theta)$ تابع چگالی احتمال (تابع احتمال)، پیشنهادی باشد، آنگاه

$$\frac{۱}{n}\log L_f(\theta) = \frac{۱}{n}\sum_{i=۱}^{n}\log f(X_i;\theta)$$

را تابع شبه—لگاریتم درست نمایی می‌نامند.

[۱۵] Quasi-Maximum Likelihood

تعریف ۷.۱ (برآوردگر شبه—درست نمایی ماکسیمم(QMLE))
فرض کنید X_1،X_2،...،X_n متغیرهای تصادفی .i.i.d باشند که توسط چگالی نامعلوم $h(.)$ تولید شده‌اند. $\hat{\theta}_n$ را بردار برآوردگر شبه—درست نمایی ماکسیمم (QMLE) می‌نامند هرگاه تابع $L_f(\theta)$ در $\hat{\theta}_n$ روی $\theta \in \Theta$ ماکسیمم شود.

برای بیان تفاوت QMLE و MLE باید به این نکته توجه داشت که وقتی مدل پیشنهادی درست—توصیف شده است برآورد θ یعنی $\hat{\theta}_n$، برآوردگر ML است ولی زمانی که مدل بد—توصیف شده است $\hat{\theta}_n$ برآوردگر QML است. در واقع به دلیل نامعلوم بودن مدل درست (مدل مولد داده‌ها) و استفاده‌از مدل پیشنهادی $f(.;\theta)$، از QMLE به جای MLE استفاده می‌شود. لی کم[۱۶] (۱۹۵۳) وجود QMLE را ثابت می‌کند اما در مورد یکتایی آن بحثی نمی‌کند.

قضیه ۹.۱ (وجود)(لم ۳ از لی‌کم (۱۹۵۳))
اگر X_1،X_2،...،X_n متغیرهای تصادفی .i.i.d تولید شده از $h(.)$ و $f(x;\theta)$ تابع چگالی پیشنهادی باشند، برای هر n، QMLE وجود دارد. برای اطلاع از چگونگی اثبات، لی کم (۱۹۵۳) مطالعه شود.

تعریف ۸.۱ (مقدار شبه—درست پارامتر)
مقدار شبه—درست[۱۷] پارامتر θ را با θ_* نشان می‌دهیم و به صورت زیر تعریف می‌شود

$$\theta_* = \arg\max_{\theta\in\Theta} E_h\{\frac{1}{n}\log L_f(\theta)\}.$$

وقتی که مدل بد—توصیف شده‌است برای تعریف θ_* باید شرایط نظم وجود داشته باشد. جهت برآورد θ_* با $\hat{\theta}_n$ نیاز به فرض‌هایی است که در ادامه آورده شده است.

(فرض ب ۱): الف) برای هر $\theta \in \Theta$، $|\log f(x;\theta)| \leq m(x)$ که در آن $m(.)$ تابعی انتگرال پذیر نسبت به h است.

ب) $E_h\{\log f(X;\theta)\}$ دارای ماکسیممی یکتا در $\theta_* \in \Theta$ است.

قسمت الف (فرض ب ۱) جهت اطمینان از خوش تعریف[۱۸] بودن $E_h\{\log f(X;\theta)\}$ است. برای مثال

۱۶ LeCam

۱۷ Pseudo-true

۱۸ Well define

اگر $f(.;\theta)$ و $h(.)$ دارای توزیع نرمال با واریانس درست $\sigma_{\circ}^{۲}$ متناهی باشد، پس مقدار $\sigma^{۲} = \circ$ در Θ نیست و قسمت الف قسمت دوم(فرض ب ۱) باعث می‌شود $\hat{\theta}_n$ برآورد سراسری[۱۹] (فراموضعی) θ باشد. وایت(۱۹۸۲) با استفاده از فرض‌های (ب ۱) و (ب ۲) سازگاری QMLE را ثابت می‌کند که در قضیه زیر آورده شده است.

قضیه ۱۰.۱ (سازگاری)

فرض کنید X_1،X_2،...،X_n متغیرهای تصادفی .i.i.d از چگالی درست $h(.)$ و $f(.;\theta)$ مدل پیشنهادی باشد. اگر فرض (ب ۱) نیز برقرار باشد، آنگاه برای هر n به اندازه کافی بزرگ $(n \longrightarrow \infty)$ $\hat{\theta}_n \xrightarrow{a.s} \theta_*$

با توجه به قضیه (۱۰.۱)، QMLE یک برآوردگر سازگار برای θ_* است و اگر مدل پیشنهادی درست—توصیف شده باشد آنگاه $E_h\{\log f(X;\theta)\}$ دارای ماکسیمم یکتایی در $\theta_* = \theta_{\circ}$ است. بنابراین $\hat{\theta}_n$ برآوردگر سازگاری برای پارامتر درست $\theta_{\circ}$ است.

چون QMLE یک برآوردگر سازگار برای پارامتر θ است بنابراین می‌توان تابع توزیع مجانبی آن را به دست آورد.

۱—۶ توزیع نرمال مجانبی QMLE

با فرض وجود مشتق‌های جزئی ماتریس‌های $I_n(\theta)$ و $J_n(\theta)$ به صورت زیر تعریف می‌شوند

$$\begin{aligned} I_n(\theta) &= \frac{۱}{n}\sum_{i=۱}^{n} \nabla_\theta^{(۲)} \log f(X_i;\theta), \\ J_n(\theta) &= \frac{۱}{n}\sum_{i=۱}^{n} (\nabla_\theta \log f(X_i;\theta))(\nabla_\theta \log f(X_i;\theta))', \end{aligned}$$

برای $\nabla_\theta^{(i)}$ $i = ۱, ۲, \ldots, n$ عملگر گرادیان نسبت به θ و ∇_{θ_j} مشتق جزئی نسبت به θ_j و $I(\theta)$ و $J(\theta)$ به صورت زیر تعریف می‌شوند

$$\begin{aligned} I(\theta) &= E_h\{\nabla_\theta^{(۲)} \log f(X;\theta)\}, \\ J(\theta) &= E_h\{(\nabla_\theta \log f(X;\theta))(\nabla_\theta \log f(X;\theta))'\}. \end{aligned}$$

که در آن X متغییر تصادفی مستقل از X_1،X_2،...،X_n و هم توزیع با آن‌هاست. با استفاده از تعاریف بالا $C_n(\theta)$ و $C(\theta)$ به صورت زیر تعریف می‌شود

$$C_n(\theta) = I_n^{-۱}(\theta) J_n(\theta) I_n^{-۱}(\theta),$$

[۱۹] Global

$$C(\theta) \quad = \quad I^{-1}(\theta)J(\theta)I^{-1}(\theta).$$

(فرض ب ۲): برای $i = 1, \ldots, p$، $\nabla_{\theta_i} \log f(x_t;\theta)$ تابع‌های اندازه‌پذیر از x_t برای هر $\theta \in \Theta$ و تابع‌هایی با پیوسته مشتق‌پذیر از θ برای همه x_tها هستند.

(فرض ب ۳): $|\nabla^{(2)}_{\theta_i\theta_j} \log f(x_t;\theta)| \leq m(x)$ و $|\nabla_{\theta_i} \log f(x_t;\theta)\nabla_{\theta_j} \log f(x_t;\theta)| \leq m'(x)$ برای $i,j = 1, 2, \ldots, p$، وقتی $m(.)$ و $m'(.)$ تابع‌هایی انتگرال‌پذیر نسبت به h برای همه x_tها و θها در Θ هستند.

(فرض ب ۴):

۱) θ_* نقطه داخلی Θ است.

۲) $I(\theta_*)$ و $J(\theta_*)$ ماتریس‌هایی غیرتکین هستند.

فرض (ب ۲) بیان می‌کند که مشتق‌های اول نسبت به θ وجود دارد و در x_t اندازه‌پذیر هستند.

فرض (ب ۳) ما را مطمئن می‌سازد که $I(\theta)$ و $J(\theta)$ در θ پیوسته هستند و قانون ضعیف اعداد بزرگ را می‌توان برای $I_n(\theta)$ و $J_n(\theta)$ به کار برد.

فرض (ب ۴) وجود θ_* در Θ و معکوس‌پذیری $I(\theta_*)$ و $J(\theta_*)$ را تضمین می‌کند علاوه بر آن θ_* را به‌طور موضعی شناسای پذیر می سازد. به این معنی که برای یک همسایگی باز Θ، $E_h\{\log f(X;\theta)\}$ در θ_* ماکسیمم یکتا دارند.

به دلیل استفاده فراوان از لم بعد که توسط وایت (۱۹۸۲) اثبات شده است آن را همراه با اثبات می آوریم.

لم ۲.۱ $Q_n(x;\theta)$ را تابعی اندازه‌پذیر از x در نظر بگیرید، که برای همه xها یک تابع پیوسته از θ در مجموعه فشرده Θ است. همچنین فرض کنید برای همه xها و θها وقتی که $m(x)$ تابعی انتگرال‌پذیر نسبت به h است، داریم $Q_n(x;\theta) \leq m(x)$ همچنین $Q_n(\theta)$ ($Q_n(\theta_*)$) را تابعی پیوسته از $\theta(\theta_*)$ در مجموعه فشرده Θ در نظر بگیرید. حال اگر برای هر $\theta \in \Theta$، $|Q_n(x;\theta) - Q_n(\theta)| \xrightarrow{a.s} 0$ و اگر $\hat{\theta}_n \xrightarrow{a.s} \theta_*$ بنابراین $|Q_n(x;\hat{\theta}_n) - Q_n(\theta_*)| \xrightarrow{a.s} 0$ یعنی $Q_n(x;\hat{\theta}_n) \xrightarrow{a.s} Q_n(\theta_*)$

برهان. به واسطه همگرایی یکنواخت، $\delta > 0$ را در نظر بگیرید، آنگاه به طور قریب به یقین برای nهای به اندازه کافی بزرگ $|Q_n(x;\hat{\theta}_n) - Q_n(\hat{\theta}_n)| < \delta/2$ و با توجه به پیوستگی $Q_n(\theta)$ و به طور قریب به یقین برای nهای به اندازه کافی بزرگ $|Q_n(\hat{\theta}_n) - Q_n(\theta_*)| < \delta/2$ و نظر به این که $\hat{\theta}_n \xrightarrow{a.s} \theta_*$ با به کار بردن نامساوی مثلثی به طور قریب به یقین برای nهای به اندازه کافی بزرگ $|Q_n(x;\hat{\theta}_n) - Q_n(\theta_*)| < \delta$ و چون δ به طور دلخواه انتخاب می‌شود نتایج برقرار است و لم ثابت شد.