دانشگاه رازی

دانشکده علوم
گروه آمار

پایان نامه کارشناسی ارشد آمار ریاضی

انتخاب مدل براساس معیار کولبک – لیبلر برای $k \geq$ ۲ مدل غیرآشیانه‌ایی

توسط

قباد برمال زن

استاد راهنما

دکتر عبدالرضا سیاره

استاد مشاور

بهمن ماه ۱۳۸۸

تقدیم به

ماندگارترین ذات لایتناهی،

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

تقدیم به پژوهشگران 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

قدردانی

سپاس خدایی که انسان را قدرت اندیشه آفرید.

بی‌تردید اثر حاضر مدیون راهنمایی اساتید گرانقدر و مساعدت و همکاری انسان‌های بزرگوار بسیاری است. اکنون در این مرحله از زندگی علمی خود، که خود را وام‌دار تمامی آنان می‌بینم، بر خود فرض می‌دانم که از زحمات ارزنده‌ی آنان صمیمانه قدردانی نمایم.

در ابتدا لازم می‌دانم از زحمات بی‌شائبه و راهنمایی‌های ارزنده‌ی استاد محترم جناب آقای دکتر عبدالرضا سیاره سپاس‌گزاری کنم که بی‌شک بدون همدلی و مساعدت ایشان، پژوهش حاضر به انجام نمی‌رسید. همچنین از تمامی دوستانی که در تدوین این پایان‌نامه مرا یاری کردند، تشکر و قدردانی می‌نمایم.

از اساتید گرانقدر آقای دکتر صادقی فر و آقای دکترخالدی سپاسگزارم که بزرگوارانه زحمت داوری این پایان‌نامه را متقبل شده و با دقت عالمانه خود نکات ظریفی را در زیبا شدن این پایان‌نامه بیان داشتند.

در پایان از پدر و مادرم صمیمانه تشکر و سپاس‌گزاری می‌نمایم که سرایت مشکلات تحصیلی اینجانب در دوران تحصیل در دانشگاه را در زندگی عادی و خانوادگی تحمل نمودند، امیدوارم با چشم‌پوشی از کوتاهی‌های اینجانب در انجام وظایف خانوادگی‌ام بر من منت گذارند.

قباد برمال زن

کرمانشاه، بهمن ۱۳۸۸

انتخاب مدل براساس معیار کولبک – لیبلر برای $k \geq ۲$ مدل غیرآشیانه‌ایی

چکیده

انتخاب مدل یک مفهوم اساسی به منظوراستنباط در مورد جوامع است که در روش کلاسیک محدود به بررسی پارامترهای جامعه می‌شود. فرض کنید یک نمونه تصادفی n تایی از یک جامعه با چگالی درست $h(.)$ در اختیار داریم. در حالت کلی h نامعلوم است و ما مدلی مانند $f(x;\beta)$ را به‌عنوان تقریبی از این چگالی بکار می‌بریم و استنباط خود را براساس $f(x;\beta)$ انجام می‌دهیم. بطوربدیهی می‌بایست $f(x;\beta)$ به چگالی درست h نزدیک باشد تا به استنباط معتبر در مورد جامعه برسیم. پیشنهاد یک مدل قطعی براساس تعداد محدودی از مشاهدات به عنوان تقریب یا برآوردی از چگالی درست h، موجب بروز ریسک بزرگی در انتخاب مدل برای جامعه خواهد شد. به همین دلیل k مدل غیرآشیانه‌ایی را انتخاب و می‌خواهیم بررسی کنیم کدام مدل به چگالی درست داده‌ها نزدیکتر است. برای انتخاب این k مدل، معیاری تعریف نشده است. ما به بررسی این سؤال اساسی در انتخاب مدل پرداخته‌ایم که چگونه می‌توان مجموعه‌ای از مدل‌های مناسب را برای برآورد چگالی درست h به دست آورد. با در اختیار داشتن این مجموعه مجاز از چگالی‌ها، جستجوی ما برای انتخاب بهترین تقریب برای چگالی درست h محدود به بررسی این مجموعه از چگالی ها خواهد شد.

واژه‌های کلیدی : آزمون فرض، انتخاب مدل، تابع شبه درستنمایی ماکسیمم، توزیع مجانبی، ریسک کولبک – لیبلر، مدل‌های آشیانه‌ایی، مدل‌های غیرآشیانه‌ایی، معیار آکائیک، معیار اطلاع.

فهرست مندرجات

لیست جداول

پیشگفتار

مدل های آماری موضوع مهمی در تحلیل علمی داده‌ها هستند. مدل‌ها برای نشان دادن ساختارهای تصادفی، پیش بینی رفتارهای آینده، استنباط و استخراج اطلاعات با استفاده از داده‌ها بکار می‌روند. مدل‌ها نقش اساسی در تحلیل‌های آماری ایفا می‌کنند. زمانیکه مدل مشخص شده باشد می‌توان از آن برای پیش‌بینی، کنترل و استخراج اطلاعات و تصمیم‌گیری استفاده نمود. بنابراین کلید حل مسائل پیچیده دنیای واقعی در توسعه و ساختار یک مدل مناسب قرار دارد.

یک مساله مهم در آمار در رابطه با یک نمونه n تایی مستقل و همتوزیع از مشاهدات، بررسی این موضوع است که آیا این مشاهدات از توزیع خاصی آمده‌اند یا خیر. در چنین موقعیت هایی بر اساس مشاهدات مدل‌های آماری ساخته می‌شود و سپس به تحلیل و آنالیز این مدل‌ها پرداخته می شود. در این پایان نامه تحلیل مدل‌ها به دو روش مجزا مورد بررسی قرار گرفته‌اند: آزمون فرض برای انتخاب مدل و معیارهای انتخاب مدل. دو مدل ممکن است نسبت به همدیگر آشیانه‌ایی، غیرآشیانه‌ایی و یا حتی آشیانه‌ایی جزئی باشند. تحلیل مدل‌های آشیانه‌ایی در آمار کلاسیک به وفور مورد بررسی و مطالعه قرار گرفته است اما کمتر به تحلیل مدل‌های غیرآشیانه‌ایی پرداخته شده است. تاریخچه یک مطالعه جدی و اساسی در مورد مدل‌های غیرآشیانه‌ایی به کاکس(۱۹۶۲ — ۱۹۶۱) و وونگ(۱۹۸۹) برمی‌گردد. از شباهت‌ها یا تفاوت‌های بین آزمون کاکس و آزمون وونگ می توان گفت که آزمون وونگ تعمیم آزمون کاکس است. هر دو این آزمون‌ها بطور کلی تعمیمی از آزمون نسبت درستنمایی باهدف‌های متفاوت هستند. نکته مهم در اینجا این است که زمانی که یک فرضیه رد می‌شود به این معنی نیست که آن فرضیه به نفع فرضیه‌ی دیگری رد شده است. برای مثال رد هر دو مدل نشان می دهد که هیچکدام از مدل ها نمی‌توانند مدل درست را توصیف کنند. بنابراین نتیجه می‌گیریم که هر دو مدل بد — توصیف شده‌اند. ممکن است یک راه حل این مشکل این باشد که از ایدهٔ انتخاب مدل استفاده کرده و مدلی را که به چگالی درست داده‌ها نزدیکتر است انتخاب کنیم. از دیدگاه وونگ بهترین مدل، مدلی است که جمله دوم کولبک — لیبلر(۱۹۵۱) را ماکسیمم کند.

معیارهای دیگری برای انتخاب مدل معرفی شده‌اند که می‌توان آکائیک (AIC) (۱۹۷۳)، معیار

شوارتز(۱۹۷۸) (BIC)، تکنیک اعتبار سنجی متقابل و معیار اطلاع بوت استراپ ایشیگارو(EIC) (۱۹۷۷) اشاره نمود.

در این پایان‌نامه ضمن بررسی روش‌های مختلف انتخاب مدل‌های غیرآشیانه‌ایی، به بررسی روشی پرداخته‌ایم که از بین $k \geq 2$ مدل پیشنهاد شده، می‌توان یک مجموعه مجاز از مدل‌ها را انتخاب و سپس از بین آنها بهترین را انتخاب کرد.

ساختار این پایان‌نامه به صورت زیر است: در فصل اول به بررسی معیار کولبک — لیبلر به عنوان معیاری برای ارزیابی مدل‌های آماری پرداخته و آن را به عنوان ریسک در انتخاب مدل مورد توجه قرار داده ایم. در فصل دوم ابتدا روش آزمون فرض های کلاسیک را یادآوری کرده سپس آزمون وونگ را بیان می‌کنیم همچنین برای k مدل پیشنهاد شدهٔ غیرآشیانه‌ایی آزمون می کنیم که کدام مدل ها از لحاظ نزدیکی به توزیع درست داده ها از یکدیگر اختلاف معنی داری ندارند. مجموعه این مدل ها ما را قادر خواهد ساخت که مدل صحیح را با دقت بیشتری توسط یکی از مدل های عضو مجموعه برآورد کنیم. محاسبه اریبی لگاریتم تابع درستنمایی و برآورد تفاضل ریسک KL میان دو مدل پیشنهاد شده و فاصلهٔ ردیابی مناسب که توسط کومانژ و سیاره (۲۰۰۸) ارائه شده است در فصل سوم آمده است و سرانجام در فصل چهارم که فصل نتیجه‌گیری است ما به بررسی این سؤال اساسی در انتخاب مدل پرداخته‌ایم که چگونه می‌توان مجموعه‌ای از مدل‌های مناسب را برای برآورد چگالی درست h بدست آورد. روشی پیشنهاد می‌کنیم تا نشان دهیم که براساس ریسک کولبک — لیبلر در هر خانواده از مدل‌های پیشنهاد شده کدامیک از چگالی‌ها از لحاظ نزدیکی به چگالی درست h یکسان هستند. مجموعه تمام عضوهایی از این خانواده ها را که از لحاظ نزدیکی به چگالی درست h یکسان هستند مجموعه مجاز می‌نامیم. با در اختیار داشتن این مجموعه مجاز از چگالی‌ها، جستجوی ما برای انتخاب بهترین تقریب برای چگالی درست h، محدود به بررسی این مجموعه از چگالی‌ها خواهد شد.

فصل ۱

تعاریف و قضایا

مقدمه:

در این فصل به بررسی مدل های آماری و معیار کولبک – لیبلر[1] به عنوان معیاری برای ارزیابی مدل‌های آماری پرداخته و آن را به عنوان ریسک در انتخاب مدل مورد توجه قرار داده ایم. خصوصیات مجانبی برآوردگرهای درستنمایی ماکسیمم را بیان کرده و در پایان نیز معیار اطلاع آکائیک[2] را به عنوان برآوردکننده‌ایی از معیار کولبک – لیبلر مورد بررسی قرار خواهیم داد.

۱–۱ مدل های آماری

مدل های آماری موضوع مهمی در تحلیل علمی داده‌ها هستند. مدل‌ها برای نشان دادن ساختارهای تصادفی، پیش بینی رفتارهای آینده، استنباط و استخراج اطلاعات با استاده از داده‌ها بکار می‌روند. مدل‌ها نقش اساسی در تحلیل‌های آماری ایفا می‌کنند. زمانیکه مدل مشخص شده باشد می‌توان از آن برای پیش‌بینی، کنترل و استخراج اطلاعات و تصمیم‌گیری استفاده نمود. بنابراین کلید حل مسائل پیچیده دنیای واقعی در توسعه و ساختار یک مدل مناسب قرار دارد.

فرض کنید $X_1, X_2, ..., X_n$ یک نمونه تصادفی (i.i.d) از توزیعی با تابع چگالی (تابع جرم احتمال) درست $h(.)$ باشد که در آن h معمولاً نامعلوم است.

تعریف ۱–۱–۱ گوییم مدل $F_B = \{f(x;\beta); \beta \in B \subset \Re^p\}$ خوب – توصیف شده[3] است هرگاه $\beta_\circ$ متعلق به B چنان وجود داشته باشد بطوریکه $h \in F_B$. یا به عبارت دیگر $h(x) = f(x;\beta_\circ)$.

تعریف ۱–۱–۲ گوییم مدل F_B بد – توصیف شده[4] است هرگاه $\beta_\circ$ متعلق به B وجود نداشته باشد بطوریکه $h \in F_B$. یا به عبارت دیگر $h(x) \neq f(x;\beta_\circ)$.

تعریف ۱–۱–۳ فرض کنید F_B و $G_\Gamma = \{g(x;\gamma); \gamma \in \Gamma \subset \Re^q\}$ دو مدل پیشنهادی برای تقریب h باشند.

الف) گوئیم مدل G_Γ در مدل F_B آشیانه‌ایی[5] است هرگاه:

$$G_\Gamma \subset F_B.$$

---

۱ Kullback-Leibler

۲ Akaiek

۳ Well Specified

۴ Misspecified

۵ Nested

ب) گوئیم دو مدل F_B و G_Γ نسبت به هم غیرآشیانه‌ایی [6] هستند اگر و فقط اگر:

$$F_B \cap G_\Gamma = \emptyset.$$

پ) گوئیم دو مدل F_B و G_Γ نسبت به هم آشیانه‌ایی جزئی [7] هستند هرگاه $F_B \not\subset G_\Gamma$ و $G_\Gamma \not\subset F_B$.

۱—۲ معیار کولبک — لیبلر (KL)

۱—۲—۱ تعاریف و ویژگی ها

فرض کنید $X_۱, X_۲, ..., X_n$ یک نمونه تصادفی (i.i.d) از توزیعی با تابع چگالی (تابع جرم احتمال) درست h باشد که در آن h معمولاً نامعلوم است. همچنین فرض کنید $F_B = \{f(x;\beta); \beta \in B \subset \Re^p\}$ یک مدل پیشنهاد شده برای h باشد که سعی می شود به گونه‌ایی انتخاب شود که برازش مناسبی به داده ها داشته باشد.

زیان [8] استفاده از چگالی f_β به جای چگالی درست h ، برای مشاهده X را می توان به شکل $\log(\frac{h(X)}{f(X;\beta)})$ تعریف کرد. امید ریاضی این زیان تحت چگالی درست h، ریسک کولبک — لیبلر نامیده می‌شود. از این به بعد، ریسک کولبک — لیبلر را با نماد اختصاری KL نشان داده و به صورت زیر تعریف می کنیم:

$$KL_h\left(h, f_\beta\right) = E_h\left[\log\left(\frac{h(X)}{f(X;\beta)}\right)\right]$$

$$= \int \left[\log\left(\frac{h(x)}{f(x;\beta)}\right)\right] dH(x).$$

که در آن E_h بیانگر امید ریاضی نسبت به چگالی درست h و H تابع توزیع X است.

ریسک KL دارای ویژگی های زیر است:

الف) $KL_h\left(h, f_\beta\right) \geq ۰$.

ب) $KL_h\left(h, f_\beta\right) = ۰$ اگر و فقط اگر $\beta_۰$ متعلق به B چنان وجود داشته باشد که $h(x) = f(x;\beta_۰)$.

از این دیدگاه، هر چه مقدار کمیت KL کوچکتر باشد، مدل پیشنهاد شده به مدل درست نزدیک تر و در

---

۶ Nonnested

۷ Partially Nonested

۸ Loss

نتیجه بهتر است.

برای بررسی ویژگی های (الف) و (ب) تابع $K(t) = \log t - t + ۱$ را که برای $t > ۰$ تعریف شده است در نظر بگیرید. در این حالت با مشتق گرفتن از $K(t)$ داریم، $\frac{\partial K(t)}{\partial t} = \frac{۱}{t} - ۱$ ، $\left.\frac{\partial K(t)}{\partial t}\right|_{t=۱} = ۰$ و $\frac{\partial^۲ K(t)}{\partial t^۲} < ۰$. در نتیجه تابع $K(t)$ ماکسیمم مقدار خود را در نقطه یک اختیار می‌کند که برابر صفر است. بدین معنی که به ازای همه $t > ۰$ مقدار تابع $K(t)$ کوچکتر یا مساوی صفر است. به عبارت دیگر:

$$\log t \leq t - ۱$$

و تساوی برقرار است اگر و فقط اگر $t = ۱$ باشد. در حالت پیوسته با جایگذاری $t = f(x;\beta)/h(x)$ داریم:

$$\log \frac{f(x;\beta)}{h(x)} \leq \frac{f(x;\beta)}{h(x)} - ۱.$$

با ضرب کردن دو طرف نامساوی بالا در $h(x)$ و انتگرال گرفتن از طرفین می‌توان نوشت:

$$\int \log\left(\frac{f(x;\beta)}{h(x)}\right) h(x)dx \leq \int \left(\frac{f(x;\beta)}{h(x)} - ۱\right) h(x)dx = ۰,$$

و در نتیجه:

$$\int \log\left(\frac{h(x)}{f(x;\beta)}\right) h(x)dx = -\int \log\left(\frac{f(x;\beta)}{h(x)}\right) h(x)dx \geq ۰.$$

بنابراین رابطه (الف) نتیجه می شود. به وضوح رابطه (ب) برقرار است هرگاه $h(x) = f(x;\beta_۰)$. برای حالت گسسته، کافیست تابع چگالی‌های $h(x)$ و $f(x;\beta)$ را به ترتیب با تابع جرم احتمال‌های $h(x_i)$ و $p(x_i)$ جایگذاری کرده و به جای انتگرال از سیگما استفاده کنیم.

توجه : برای اثبات نامساوی بالا، همچنین می‌توان از نامساوی جنسن[۹] استفاده کرد.

معیار KL درحقیقت امید ریاضی یک تابع زیان است. به همین دلیل آن را ریسک KL نیز می‌گویند و معمولاً به عنوان معیاری از نزدیکی دو تابع چگالی احتمال (تابع جرم احتمال) در نظر گرفته می‌شود. معیار KL یک متر ریاضی نیست، زیرا خاصیت تقارن را ندارد. ویژگی های (الف) و (ب) را در مثال های زیر بررسی می کنیم:

---

۹ Jensen Inquality

۱—۲—۲ مثال هایی از ریسک KL

مثال ۱: (**ریسک KL در حالت پیوسته**) فرض کنید X دارای چگالی درست نرمال استاندارد باشد که معمولاً از آن بی خبر هستیم و مدل پیشنهادی نرمال با پارامترهای $\beta = (\mu, \sigma^2)$ است. به عبارت دیگر:

$$f(x;\beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\},$$

و

$$h(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}.$$

همچنین داریم:

$$E_h\left[\log f(X;\beta)\right] = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1+\mu^2}{2\sigma^2},$$

و

$$E_h\left[\log h(X)\right] = -\frac{1}{2}\log(2\pi) - \frac{1}{2},$$

بنابراین ریسک $f(x;\beta)$ نسبت به چگالی درست h برای $\beta = (\mu, \sigma^2)$ به صورت زیر است:

$$KL_h\left(h, f_\beta\right) = \frac{1}{2}\left[\log\sigma^2 + \frac{1+\mu^2}{\sigma^2} - 1\right].$$

در حالت خاص، اگر $\mu = 0$ و $\sigma^2 = 1$ آنگاه بدیهی است که $KL_h\left(h, f_\beta\right) = 0$ در این حالت $\beta_0 = (0, 1) \in B$.

مثال ۲: (**ریسک KL در حالت گسسته**) می دانیم در پرتاب یک تاس سالم، احتمال آمدن هر یک از خال‌ها برابر ۱/۶ است. لذا تابع جرم احتمال درست به صورت $h = p_i = \{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}$ است. دو تابع جرم احتمال برای مشاهدهٔ خال های این تاس به شکل زیر پیشنهاد شده است:

$$P_a = \{0/20, 0/12, 0/18, 0/12, 0/20, 0/18\},$$

و

$$P_b = \{0/18, 0/12, 0/14, 0/19, 0/22, 0/15\}.$$

می‌خواهیم بررسی کنیم کدام مدل برای توصیف نتایج حاصل از پرتاب یک تاس سالم مناسب تر است. با استفاده از تعریف ریسک KL برای حالت گسسته داریم:

$$KL_h\left(h, P_a\right) = \sum_{i=1}^{6} p_i \log\frac{p_i}{p_{ai}} = 0/023,$$

و

$$KL_h\left(h, P_b\right) = \sum_{i=۱}^{٦} p_i \log \frac{p_i}{p_{bi}} = ۰/۰۲۰.$$

لذا نامنفی بودن KL در این مثال به‌وضوح دیده می شود. همچنین با توجه به مقادیر بدست آمده برای ریسک های KL، احتمال های متناظر با مدل b به مدل صحیح نزدیکتر است. بنابراین مدل b به مدل a ترجیح داده می شود.

۱—۳ برخی قضایای همگرایی

در این بخش قضایا و تعاریفی را که در ادامه به آن ها نیاز داریم بیان می‌کنیم .

تعریف ۱—۳—۱ یک دنباله از متغیرهای تصادفی $X_۱, X_۲, ...$ در توزیع به متغیر تصادفی X همگراست هرگاه به ازای تمام نقاط پیوستگی تابع توزیع $F_X(x)$ داشته باشیم:

$$lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

در اینجا از نماد $X_n \xrightarrow{D} X$ استفاده می کنیم.

تعریف ۱—۳—۲ یک دنباله از متغیرهای تصادفی $X_۱, X_۲, ...$ در احتمال به متغیر تصادفی X همگراست، $X_n \xrightarrow{P} X$، هرگاه به ازای هر $\varepsilon > ۰$:

$$lim_{n\to\infty} P\left(\Big|X_n - X\Big| \geq \varepsilon\right) = ۰,$$

یا به طور معادل:

$$lim_{n\to\infty} P\left(\Big|X_n - X\Big| < \varepsilon\right) = ۱.$$

قضیه ۱—۳—۱ (قضیه اسلوتسکی[۱۰]) اگر $X_n \xrightarrow{D} X$، $Y_n \xrightarrow{P} a$ و a یک مقدار ثابت باشد، آنگاه:

الف— $X_n Y_n \xrightarrow{D} aX$.

ب— $X_n + Y_n \xrightarrow{D} X + a$.

پ— $\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{a}$ مشروط به آنکه $a \neq ۰$ و $p(Y_n = ۰) = ۰$.

قضیه ۱—۳—۲ (قانون ضعیف اعداد بزرگ[۱۱]) فرض کنید $X_۱, X_۲, ...$ یک دنباله $i.i.d$ از متغیرهای

[۱۰] Slutesky Theorem

[۱۱] Weak Low of Large Numbers

تصادفی با $E(X_t) = \mu$ و $Var(X_t) = \sigma^۲ < \infty$ باشد. اگر $\bar{X}_n = \frac{۱}{n}\sum_{t=۱}^{n} X_t$، آنگاه به ازای هر $\varepsilon > ۰$:

$$lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| < \varepsilon\right) = ۱,$$

بدین معنی که $\bar{X}_n$، در احتمال به μ همگراست. به عبارت دیگر اگر $n \to \infty$، آنگاه:

$$\frac{۱}{n}\sum_{t=۱}^{n} X_t \xrightarrow{P} \mu.$$

تعریف ۱–۳–۳ یک دنباله $i.i.d$ از متغیرهای تصادفی $X_۱, X_۲, ...$ تقریباً مطمئن[۱۲] به متغیر تصادفی X همگراست، $X_n \xrightarrow{a.s} X$، هرگاه به ازای هر $\varepsilon > ۰$:

$$P\left(\lim_{n\to\infty} \left|X_n - X\right| < \varepsilon\right) = ۱.$$

قضیه ۱–۳–۳ (قانون قوی اعداد بزرگ[۱۳]) فرض کنید $X_۱, X_۲, ...$ یک دنباله $i.i.d$ از متغیرهای تصادفی با $E(X_t) = \mu$ و $Var(X_t) = \sigma^۲ < \infty$ باشد. اگر $\bar{X}_n = \frac{۱}{n}\sum_{t=۱}^{n} X_t$، آنگاه به ازای هر $\varepsilon > ۰$:

$$P\left(\lim_{n\to\infty} \left|\bar{X}_n - \mu\right| < \varepsilon\right) = ۱.$$

بدین معنی که $\bar{X}_n$ تقریباً مطمئن به μ همگراست. به عبارت دیگر اگر $n \to \infty$ آنگاه:

$$\frac{۱}{n}\sum_{t=۱}^{n} X_t \xrightarrow{a.s} \mu.$$

قضیه ۱–۳–۴ (قضیه حد مرکزی چند متغیره[۱۴]) فرض کنید $\underline{X}_۱, \underline{X}_۲, ...$ یک دنباله $i.i.d$ از بردارهای تصادفی p – بعدی با $E[\underline{X}_t] = \underline{\mu}$ و ماتریس واریانس – کوواریانس

$$\Sigma = E\left[(\underline{X}_t - \underline{\mu})(\underline{X}_t - \underline{\mu})^T\right]$$

باشند. آنگاه اگر n به سمت بینهایت میل کند داریم:

$$\frac{۱}{\sqrt{n}}\sum_{t=۱}^{n}(\underline{\bar{X}}_t - \underline{\mu}) = \sqrt{n}(\underline{\bar{X}} - \underline{\mu}) \xrightarrow{D} N_p(\underline{۰}, \Sigma).$$

---

۱۲ Almost Surely

۱۳ Strong Low of Large Numbers

۱۴ Multivariate Centrel Limit Theorem

۱—۴ امید لگاریتم درستنمایی

در بخش ۱—۲ به معرفی ریسک کولبک — لیبلر پرداختیم. چون این معیار شامل توزیع نامعلوم h است به طور مستقیم نمی‌تواند محاسبه شود.

ریسک KL می‌تواند به صورت زیر تجزیه شود:

$$KL_h\left(h, f_\beta\right) = E_h\left[\log\left(\frac{h(X)}{f(X;\beta)}\right)\right]$$

$$= E_h\left[\log h(X)\right] - E_h\left[\log f(X;\beta)\right].$$

به‌وضوح دیده می‌شود که $KL_h\left(h, f_\beta\right)$ تابعی از β است. در بحث انتخاب مدل، علاقمند به پیدا کردن $\beta \in B$ هستیم که مقدار $KL_h\left(h, f_\beta\right)$ را مینیمم کند. به طور کلی چون جمله اول سمت راست یک مقدار ثابت است و فقط بستگی به چگالی درست h دارد، بنابراین برای پیدا کردن β ایی که مقدار $KL_h\left(h, f_\beta\right)$ را مینیمم کند، کافیست β ایی را پیدا کنیم که مقدار $E_h\left[\log f(X;\beta)\right]$ را ماکسیمم کند. این جمله در واقع امید لگاریتم درستنمایی[۱۵] است. در مقایسهٔ مدل های پیشنهادی، مدلی را بر سایر مدل های پیشنهادی ترجیح می دهیم که امید لگاریتم درستنمایی بزرگتری داشته باشد.

تعریف ۱—۴—۱ فرض کنید $F_B = \{f(x;\beta); \beta \in B \subset \Re^p\}$ مدل پیشنهاد شده باشد. پارامتر β_* ایی که ریسک KL را مینیمم کند، مقدار شبه پارامتر درست[۱۶] نامیده می‌شود.

چون مقدار امید لگاریتم درستنمایی بستگی به چگالی نامعلوم h و پارامتر β دارد، بنابراین یک کمیت نامعلوم است و باید برآورد شود تا بتوان از آن برای مقایسهٔ مدل ها استفاده کرد که در فصل سوم به برآورد آن می پردازیم.

۱—۵ یادآوری برآوردگرهای درستنمایی ماکسیمم و خواص آن

تعریف ۱—۵—۱ فرض کنید $X_1, X_2, ..., X_n$ یک نمونه تصادفی از تابع چگالی احتمال (تابع جرم احتمال) $f(x;\beta)$ باشد که در آن $\beta = (\beta_1, ..., \beta_p)^T$ است. در این حالت، با در دست داشتن مشاهدات $x_1, ..., x_n$، لگاریتم تابع درستنمایی این مشاهدات عبارت است از:

$$L_f(\beta) = \sum_{t=1}^{n} \log f(x_t;\beta)$$

---

۱۵ Expected Log-Likelihood

۱۶ Psedo-true Value of a Parameter

که تابعی حقیقی مقدار است و روی فضای پارامتر B تعریف می‌شود. یک برآوردگر طبیعی از پارامتر مدل، β ایی از فضای پارامتر B است که $L_f(\beta)$ را ماکسیمم کند. چنین β ایی را $\hat{\beta}_n$ می نامیم که در برابری $L_f(\hat{\beta}_n) = \sup_{\beta\in B} L_f(\beta)$ صدق می کند. $\hat{\beta}_n$ برآوردگر درستنمایی ماکسیمم پارامتر β بر اساس n مشاهده نامیده می‌شود.

دراینجا به بررسی ویژگی های مجانبی برآوردگرهای درستنمایی ماکسیمم در مدل پیوستهٔ پارامتری $F_B = \{f(x;\beta); \beta \in B \subset \Re^p\}$ که در آن $\beta = (\beta_1, ..., \beta_p)^T$ است می‌پردازیم.

فرض کنید شرایط نظم برای $f(x;\beta)$ به صورت زیر برقرار باشد:

فرض ۱ چگالی $f(x;\beta)$ دارای مشتق اول و مشتق دوم پیوسته نسبت به β باشد.

فرض ۲ تابع های انتگرال‌پذیر $F_1(x)$ و $F_2(x)$ روی اعداد حقیقی $\Re$ چنان وجود داشته باشند که نامساوی های زیر به ازای تمام $\beta \in B$ برقرار باشند:

$$\left|\frac{\partial \log f(x;\beta)}{\partial \beta_i}\right| < F_1(x), \qquad \left|\frac{\partial^2 \log f(x;\beta)}{\partial \beta_i \partial \beta_j}\right| < F_2(x) \qquad i,j = 1,2,...,p.$$

فرض ۳ نامساوی زیر به ازای هر مقدار دلخواه $\beta \in B$ برقرار باشد:

$$0 < \int_{-\infty}^{+\infty} f(x;\beta) \frac{\partial \log f(x;\beta)}{\partial \beta_i} \frac{\partial \log f(x;\beta)}{\partial \beta_j}\, dx < \infty, \qquad i,j = 1,2,...,p.$$

فرض ۴ فرض کنید β_0 یک جواب معادلهٔ

$$\int_{-\infty}^{+\infty} f(x;\beta) \frac{\partial \log f(x;\beta)}{\partial \beta}\, dx = 0$$

و $\hat{\beta}_n$ برآوردگر درستنمایی ماکسیمم براساس n مشاهده، از تابع چگالی $f(x;\beta)$ باشد.

قضیه ۱.۵.۱ هرگاه $n \to \infty$ و فرضیات ۱ تا ۴ برقرار باشد آنگاه:

الف) $\hat{\beta}_n$ در احتمال به β_0 همگراست.

ب) $\hat{\beta}_n$ مجانباً نرمال است. یا به عبارت دیگر:

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} N_p(\underline{0}, I^{-1}(\beta_0))$$

که در آن $I(\beta_0)$ مقدار ماتریس $I(\beta)$ در $\beta = \beta_0$ است که از رابطه زیر بدست می آید:

$$I(\beta) = \int f(x;\beta) \frac{\partial \log f(x;\beta)}{\partial \beta} . \frac{\partial \log f(x;\beta)}{\partial \beta^T}\, dx.$$

این ماتریس، ماتریس اطلاع فیشر نامیده می‌شود.

سؤال‌هایی که در اینجا مطرح می‌شوند این‌ها هستند که اگر مدلی را که به عنوان مدل درست انتخاب می‌کنیم مدلی بد — توصیف شده باشد آیا ویژگی‌های برآوردگرهای درستنمایی ماکسیمم حفظ