بسم الله الرحمن الرحیم

دانشگاه رازی
دانشکده علوم پایه
گروه آمار

پایان نامه جهت اخذ درجه کارشناسی ارشد
رشته آمار ریاضی

عنوان پایان‌نامه:

آزمون نیکویی برازش برای توابع مفصل

استاد راهنما:
دکتر عبدالرضا سیاره

نگارش:
بهاره قاسمی

اسفند ماه ۱۳۹۱

چکیــده

یکی از مسائل اساسی در آمار، مدل‌بندی پدیده‌های تصادفی است. به طور کلی از مدل‌های آماری برای نشان دادن ساختارهای تصادفی، پیش‌بینی رفتار متغیرها در آینده، استنباط و استخراج اطلاعات از داده‌ها استفاده می‌شود. در این میان تابع مفصل به عنوان یک مدل برای مشاهدات چند متغیره و وابسته توانسته است در مطالعات اخیر توجه بسیاری از کاربران آمار را به خود جلب نماید. در حقیقت این تابع به آماردان کمک می‌کند تا بتواند وابستگی بین متغیرهای تصادفی را صرف نظر از نوع توزیع‌های حاشیه‌ای آن‌ها، مدل‌بندی کند. بدین منظور آشنایی با توابع مفصل و در اختیار داشتن یک مجموعه بزرگ از این توابع در بحث مدل‌بندی آماری امری ضروری است. از طرفی اگر از یک تابع مفصل خاص برای مدل‌بندی ساختار وابستگی داده‌ها استفاده شود، ممکن است این سوال به وجود آید که آیا این تابع می‌تواند داده‌ها را به خوبی برازش دهد یا نه؟

بر این اساس در این پایان‌نامه سعی شده است توابع مفصل و ویژگی‌های آن‌ها به همراه چند خانواده مهم از این توابع مطرح شوند. در پایان نیز دو روش آزمون نیکویی برازش برای توابع مفصل بر اساس تبدیل انتگرال احتمال شرطی ارائه شده و توان آن‌ها مورد مقایسه قرار گرفته است.

واژه‌های کلیدی:

آزمون نیکویی برازش، آماره اندرسون—دارلینگ، تابع توزیع تجربی، تابع مفصل، تابع مفصل ارشمیدسی، وابستگی دمی.

فهرست مطالب

عنوان | صفحه

لیست اشکال

لیست جداول

پیشگفتار

در علم آمار و احتمال تا مدت زمان طولانی مدل‌بندی‌ها تنها بر اساس فرضیات ساده انجام می‌گرفتند. در اکثر مطالعات از فرض نرمال بودن و وجود استقلال بین داده‌ها در مدل‌بندی‌ها استفاده می‌شد و مطالعه توزیع‌های چند متغیره تنها به توزیع‌های نرمال محدود شده بود. اما در دنیای واقعی با مسائلی روبرو هستیم که فرضیات نرمال تقریب خوبی برای مجموعه داده‌ها نیست و همچنین بین داده‌ها وابستگی وجود دارد. در چنین مواردی باید از مدل‌هایی استفاده نمود که برای بسیاری از توزیع‌ها به راحتی قابل استفاده باشند، به آماردان در تجزیه و تحلیل داده‌های وابسته کمک نمایند، در نمونه‌های چند متغیره بتوانند چگونگی ارتباط بین متغیرها را نشان دهند و اندازه همبستگی بین متغیرها را تعیین نمایند.

بر این اساس بحث مدل‌بندی روابط وابستگی بین متغیرهای تصادفی مطرح شد. این مبحث یکی از موضوعاتی است که به طور گسترده در تئوری آمار و احتمال مورد مطالعه قرار گرفته است. روش معمول برای مطالعه ساختار وابستگی، استفاده از گشتاورهای مرتبه دوم متغیرهای تصادفی است. اما با این روش فقط می‌توان روابط وابستگی خطی را تعیین نمود. به علاوه این روش فقط برای کلاس خاصی از توزیع‌ها مانند توزیع‌های نرمال به راحتی قابل استفاده است. به همین دلیل این سوال به ذهن می‌رسد که آیا امکان به دست آوردن ساختار وابستگی بین متغیرهای تصادفی در حالت چند بعدی بدون هیچ اثر نگران کننده ناشی از اثرات توابع توزیع حاشیه‌ای وجود دارد؟ در پاسخ به این سوال می‌توان گفت که تابع مفصل، تابعی است که چنین امکانی را برای آماردان به راحتی فراهم می‌نماید.

در نظریه آمار و احتمال از توابع مفصل در مدل‌بندی وابستگی بین متغیرهای تصادفی استفاده می‌شود. اخیراً این توابع به دلیل ان که می‌توانند ساختار وابستگی بین متغیرها را صرف نظر از نوع توزیع‌های حاشیه‌ای مورد مطالعه قرار دهند، مورد توجه بسیاری از آماردانان قرار گرفته‌اند. خصوصاً توابع مفصل در علم اقتصاد و بیمه به منظور مدل‌بندی وابستگی بین متغیرهای ریسک تا حد زیادی مورد توجه اقتصاددانان قرار گرفته‌اند.

با هدف آشنایی با توابع مفصل و استفاده از آن‌ها در مدل‌بندی‌های آماری در این پایان‌نامه، فصل اول به معرفی توابع مفصل در حالت دو بعدی و ویژگی‌های آن‌ها

اختصاص یافته است. در فصل دوم مفاهیم وابستگی، ارتباط آن‌ها با توابع مفصل و روش‌های برآورد پارامترهای این توابع، که این پارامترها وابستگی بین متغیرها را نشان می‌دهند، بیان شده است. فصل سوم نیز به بیان روش‌های ساخت توابع مفصل، معرفی یک کلاس مهم از این توابع با نام کلاس مفاصل ارشمیدسی به همراه چند خانواده دیگر از توابع مفصل و ویژگی‌های آن‌ها پرداخته است. در ادامه این فصل توابع مفصل در حالت چند بعدی مطرح شده‌اند. در فصل چهارم ابتدا آزمون‌های نیکویی برازش بر اساس توابع توزیع تجربی مطرح شده و در ادامه دو آزمون نیکویی برازش برای توابع مفصل بر اساس تبدیل انتگرال احتمال شرطی ارائه شده و توان آن‌ها مورد مقایسه قرار گرفته است.

فصل ۱

معرفی توابع مفصل

۱.۱ مقدمه

از دیدگاه استنباط آماری پدیده‌های طبیعی اغلب با چندین متغیر تصادفی مشخص می‌شوند. بررسی وابستگی میان این متغیرهای تصادفی نیازمند ایجاد ارتباط بین تابع توزیع توأم و توابع حاشیه‌ای یک بعدی آن‌ها است. به منظور تجزیه و تحلیل چنین پدیده‌هایی محققان در تلاش بوده‌اند که بتوانند به نحوی بین توابع توزیع چند متغیره و حاشیه‌ای‌های دلخواه ارتباط برقرار کنند. مطالعات زیادی در این زمینه انجام شده است. از این میان می‌توان به هافدینگ[۱] (۱۹۴۰) وفرچه[۲](۱۹۵۱) اشاره کرد. اما سرانجام اسکلار[۳](۱۹۵۹) توانست با استفاده از تابع مفصل[۴] صرف نظر از نوع توزیع‌های حاشیه‌ای این ارتباط را برقرار کند. واژه مفصل به معنای اتصال و ارتباط، برای اولین بار توسط اسکلار و به منظور تأکید بر نقش آن‌ها معرفی شد. این توابع اخیراً به دو دلیل عمده مورد توجه آماردانان قرار گرفته‌اند: اول به عنوان راهی برای مطالعه وابستگی بین متغیرها به صورت ناپارامتری و دوم به عنوان نقطه شروعی برای ساخت توابع توزیع چند متغیره بر اساس توزیع‌های حاشیه‌ای یک بعدی.
در این فصل مفاصل و ویژگی‌های مهم آن‌ها به طور کامل معرفی خواهند شد.

۲.۱ تعاریف و مفاهیم اولیه

در این بخش برخی تعارف و قضایای مورد نیاز عنوان خواهند شد. اما قبل از آن لازم است چند نماد که در سراسر این پایان‌نامه به کار خواهند رفت، معرفی شوند.

نمادگذاری

اگر $R = (-\infty, \infty)$ نشان دهنده خط حقیقی باشد، آن‌گاه $\bar{R} = [-\infty, \infty]$ نماد خط حقیقی

---

۱ $Hoeffding$

۲ $Frechet$

۳ $Sklar$

۴ $Copula$

توسعه یافته[5] و $\bar{R}^{۲} = \bar{R} \times \bar{R}$ نماد صفحه حقیقی توسعه یافته خواهند بود. در این صورت $B = [x_۱, x_۲] \times [y_۱, y_۲]$ نماد یک مستطیل در $\bar{R}^۲$ است. رئوس این مستطیل نقاط $(x_۱, y_۱)$، $(x_۱, y_۲)$، $(x_۲, y_۱)$ و $(x_۲, y_۲)$ است که همگی در $\bar{R}^۲$ قرار دارند. مربع واحد $I^۲$ نیز ضرب خارجی $I \times I$ است که در آن $I = [۰, ۱]$ است.

همچنین دامنه و برد تابعی مانند H به ترتیب با نماد $DomH$ و $RanH$ نشان داده می‌شوند. در این صورت اگر H یک تابع حقیقی دو بعدی باشد، $DomH$ زیرمجموعه‌ای از $\bar{R}^۲$ و $RanH$ زیرمجموعه‌ای از R خواهد بود.

تعریف ۱.۲.۱ (همگرایی قریب به یقین).

دنباله متغیرهای تصادفی $\{X_n\}_{n=۱}^{\infty}$ به طور قریب به یقین به متغیر تصادفی X همگرا است، $X_n \xrightarrow{a.s} X$، هرگاه

$$\epsilon > ۰ \qquad \lim_{n\to\infty} P\{\sup |X_n - X| \geq \epsilon\} = ۰.$$

تعریف ۲.۲.۱(همگرایی در توزیع).

فرض کنید $\{X_n\}_{n=۱}^{\infty}$ دنباله‌ای از متغیرهای تصادفی به ترتیب با توابع توزیع $F_{X_۱}, F_{X_۲}, ...$ باشد. این دنباله در توزیع به متغیر تصادفی X با تابع توزیع $F_X(x)$ همگرا است، $X_n \xrightarrow{D} X$، هر گاه به ازای تمام نقاط پیوستگی F

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

قضیه ۱.۲.۱ (قضیه حد مرکزی).

فرض کنید $\{X_n\}_{n=۱}^{\infty}$ یک دنباله از متغیرهای تصادفی مستقل و هم‌توزیع با میانگین μ و واریانس متناهی $\sigma^۲$ باشد، وقتی $n \to +\infty$ آن‌گاه

$$\sqrt{n}\frac{(\overline{X}_n - \mu)}{\sigma} \xrightarrow{D} N(۰, ۱).$$

[5] Extended Real line

قضیه ۲.۲.۱ (قانون قوی اعداد بزرگ).
فرض کنید $\{X_n\}_{n=1}^{\infty}$ یک دنباله از متغیرهای تصادفی مستقل و هم‌توزیع با میانگین متناهی $E(X_j) = \mu$ باشد. اگر $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$، آن‌گاه

$$\forall \epsilon > 0 \qquad P(\lim_{n\to\infty} |\overline{X}_n - \mu| < \epsilon) = 1.$$

به عبارت دیگر وقتی $n \to \infty$، آن‌گاه

$$\overline{X}_n \xrightarrow{a.s} \mu.$$

تعریف ۳.۲.۱ (نامساوی لیپ‌شیتس[6]).
فرض کنید (X, d_X) و (Y, d_Y) دو فضای متریک باشند، به طوری که d_X نماد متر روی مجموعه X و d_Y نماد متر روی مجموعه Y است. اگر f تابعی از X به Y باشد، آن‌گاه به ازای هر x_1 و x_2 در X نامساوی

$$d_Y\left(f(x_1), f(x_2)\right) \leq K d_X(x_1, x_2)$$

به ازای ثابت حقیقی $K \geq 0$ برقرار خواهد بود. این نامساوی معروف به نامساوی لیپ‌شیتس است.

هر تابعی که در نامساوی لیپ‌شیتس صدق کند، به طور مطلق پیوسته[7] خواهد بود. بنابراین می‌توان نتیجه گرفت که این تابع:
۱. به طور یکنواخت پیوسته[8] است.
۲. تقریباً همه جا به جز روی مجموعه‌ای با اندازه لبگ[9] صفر مشتق پذیر[10] است.

۶ $Lipschitz \quad inequality$
۷ $Absolutely \quad Continuous$
۸ $Uniformly \quad Continuous$
۹ $Lebesgue \quad Measure$
۱۰ $differentiable$

تعریف ۴.۲.۱ (تابع شبه معکوس[۱۱]).
فرض کنید F یک تابع توزیع باشد. در این صورت شبه معکوس F، تابعی با دامنه I خواهد بود که با نماد $F^{(-1)}$ نشان داده می‌شود به طوری که:

۱. اگر $t \in RanF$، آن‌گاه $F^{(-1)}(t)$ عددی مانند x در $\bar{R}$ خواهد بود به طوری که $F(x) = t$. یعنی

$$\forall t \in RanF \quad F\left(F^{(-1)}(t)\right) = t.$$

۲. اگر $t \notin RanF$، آن‌گاه

$$F^{(-1)}(t) = \inf_{x \in \bar{R}}\{x \mid F(x) \geq t\} = \sup_{x \in \bar{R}}\{x \mid F(x) \leq t\}.$$

اگر F اکیداً صعودی باشد، فقط یک شبه معکوس خواهد داشت که همان تابع معکوس معمولی F^{-1} خواهد بود.

لم ۱.۲.۱ اگر U یک متغیر تصادفی یکنواخت روی $(0, 1)$ و F یک تابع توزیع تجمعی پیوسته باشد، آن‌گاه متغیر تصادفی $F^{-1}(U)$ هم‌توزیع با F خواهد بود.

قضیه ۳.۲.۱
فرض کنید $\{X_n\}_{n=1}^{\infty}$ یک دنباله از متغیرهای تصادفی مستقل و هم‌توزیع با تابع توزیع پیوسته F باشد. اگر برای $n = 1, 2, \cdots$، $X_1, \cdots, X_n$ نشان دهنده طول عمر n قطعه یک دستگاه الکتریکی باشد که آماره مرتب آن به صورت

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

نمادگذاری شده است، آن‌گاه احتمال آن که به ازای $i = 1, \cdots, n$، قطعه $(n - i + 1)$–اُم بیشتر از زمان t عمر کند، به شرط آن که قطعه $(n - i)$–اُم تا زمان s عمر کرده باشد به صورت

$$P(X_{(n-i+1)} > t | X_{(n-i)} = s) = \left\{\frac{1 - F(t)}{1 - F(s)}\right\}^i, \quad t > s, X_{(0)} = 0, F(s) < 1$$

[۱۱] $Quasi - Inverse$

محاسبه خواهد شد.

تعریف ۵.۲.۱ (تابع معکوس نما[۱۲])

فرض کنید φ یک تابع پیوسته و اکیداً نزولی از I به $[\circ, \infty]$ باشد به طوری که $\varphi(\text{۱}) = \circ$. در این صورت معکوس نما تابع φ با نماد $\varphi^{[-\text{۱}]}$ نشان داده شده و به صورت تابعی از $[\circ, \infty]$ به I با ضابطه

$$\varphi^{[-\text{۱}]}(t) = \begin{cases} \varphi^{-\text{۱}}(t), & \circ \leq t \leq \varphi(\circ) \\ \circ, & \varphi(\circ) < t \leq \infty. \end{cases}$$

معرفی خواهد شد. توجه کنید که:

۱. اگر $\varphi(\circ) = \infty$، آن‌گاه $\varphi^{[-\text{۱}]} = \varphi^{-\text{۱}}$.

۲. چون φ تابعی پیوسته و اکیداً نزولی روی I است، بنابراین می‌توان نتیجه گرفت که $\varphi^{[-\text{۱}]}$ روی $[\circ, \infty]$ تابعی پیوسته و غیرصعودی است. همچنین این تابع روی $[\circ, \varphi(\circ)]$ تابعی اکیداً نزولی است.

۳. به ازای هر u در I، $\varphi^{[-\text{۱}]}\left(\varphi(u)\right) = u$.

۴. به ازای هر t در $[\circ, \infty]$، رابطه

$$\varphi(\varphi^{[-\text{۱}]}(t)) = \begin{cases} t, & \circ \leq t \leq \varphi(\circ) \\ \varphi(\circ), & \varphi(\circ) < t \leq \infty \end{cases} = \min(t, \varphi(\circ)).$$

برقرار است.

تعریف ۶.۲.۱ (تابع محدب[۱۳])

تابع $f : R \to R$، تابع محدب نامیده می‌شود، اگر برای هر $\circ \leq \lambda \leq \text{۱}$ و x و y موجود در R

$$f\left(\lambda x + (\text{۱} - \lambda)y\right) \leq \lambda f(x) + (\text{۱} - \lambda) f(y).$$

همچنین تابع $f : R \to R$، نیمه محدب[۱۴] خواهد بود، اگر رابطه فوق به ازای $\lambda = \frac{\text{۱}}{\text{۲}}$ برقرار باشد.

---

۱۲ $Pseudo - Inverse$

۱۳ $Convex \quad Function$

۱۴ $Midconvex$

۳.۱ تابع مفصل

قبل از معرفی تابع مفصل لازم است تعاریف زیر بیان شوند.

تعریف ۱.۳.۱

فرض کنید $S_۱$ و $S_۲$ زیر مجموعه‌های ناتهی از $\bar{R}$ ، H تابعی با دامنه $S_۱ \times S_۲$ و $B = [x_۱, x_۲] \times [y_۱, y_۲]$ یک مستطیل در دامنه H باشد. در این صورت H–حجم[۱۵] B به صورت

$$V_H(B) = H(x_۲, y_۲) - H(x_۲, y_۱) - H(x_۱, y_۲) + H(x_۱, y_۱)$$

تعریف می‌شود.

تعریف ۲.۳.۱

تابع حقیقی دو متغیره H با دامنه $S_۱ \times S_۲$ دو–صعودی[۱۶] نامیده می‌شود، اگر برای هر مستطیل B که رئوس آن در دامنه H قرار دارد، $V_H(B) \geq ۰$ باشد.

توجه کنید که دو–صعودی بودن تابع H به معنای غیر نزولی بودن آن نسبت به هر مؤلفه‌اش نیست.

مثال ۱.۱ فرض کنید تابع H روی $I^۲ = [۰, ۱]^۲$ به صورت

$$H(x, y) = (۲x - ۱)(۲y - ۱) \quad , \quad x, y \in I$$

تعریف شده باشد. برای هر مستطیل $B = [x_۱, x_۲] \times [y_۱, y_۲]$ در $I^۲$

$$\begin{aligned} V_H(B) &= (۲x_۲ - ۱)[۲y_۲ - ۱ - ۲y_۱ + ۱] + (۲x_۱ - ۱)[۲y_۱ - ۱ - ۲y_۲ + ۱] \\ &= ۲(y_۱ - y_۲)[۲x_۱ - ۱ - ۲x_۲ + ۱] \\ &= ۴(y_۱ - y_۲)(x_۱ - x_۲) \geq ۰. \end{aligned}$$

---

۱۵ H-Volume

۱۶ Two-Increasing

بنابراین H تابعی دو—صعودی خواهد بود. اما این تابع به ازای هر $y \in (0, \frac{1}{2})$ تابعی نزولی از x و به ازای هر $x \in (0, \frac{1}{2})$ تابعی نزولی از y است.

مثال ۱ .۲ فرض کنید تابع H با ضابطه

$$H(x,y) = max(x,y)$$

روی I^2 تعریف شده باشد. در این صورت تابع H نسبت به x و y غیرنزولی است، اما چون $V_H(I^2) = -1$ است، این تابع دو—صعودی نخواهد بود.

لم ۱.۳.۱ فرض کنید S_1 و S_2 زیر مجموعه‌های ناتهی از $\bar{R}$ باشند و H یک تابع دو—صعودی با دامنه $S_1 \times S_2$ باشد. همچنین فرض کنید

$$x_1, x_2 \in S_1 \quad s.t \quad x_1 \leq x_2$$

$$y_1, y_2 \in S_2 \quad s.t \quad y_1 \leq y_2$$

در این صورت نگاشت $t \rightarrow H(t, y_2) - H(t, y_1)$ روی S_1 و نگاشت $t \rightarrow H(x_2, t) - H(x_1, t)$ روی S_2 غیر نزولی خواهند بود.

اثبات:

اگر تابع $f(t)$ به صورت $f(t) = H(t, y_2) - H(t, y_1)$ تعریف شود، آن‌گاه طبق خاصیت دو—صعودی بودن H به ازای هر t_1 و t_2 عضو S_1 به طوری که $t_1 \leq t_2$، رابطه

$$f(t_2) - f(t_1) = H(t_2, y_2) - H(t_2, y_1) - H(t_1, y_2) + H(t_1, y_1) \geq 0$$

برقرار خواهد بود. در این صورت می‌توان نتیجه گرفت که f تابعی غیرنزولی است. به طریق مشابه نیز قسمت دوم این لم ثابت می‌شود.

با استفاده از این لم و تعریف زیر به عنوان یک فرض اضافی، می‌توان نشان داد که یک تابع دو—صعودی نسبت به هر مؤلفه‌اش غیر نزولی خواهد بود.

تعریف ۳.۳.۱

فرض کنید S_1 و S_2 زیر مجموعه‌های ناتهی از $\bar{R}$ باشند. اگر a_1 و a_2 به ترتیب کوچکترین عضوهای S_1 و S_2 باشند، آن‌گاه تابع H از $S_1 \times S_2$ به R یک تابع اساسی[۱۷] نامیده می‌شود، هرگاه

$$\forall (x,y) \in S_1 \times S_2 \qquad H(x,a_2) = 0 = H(a_1,y).$$

لم ۲.۳.۱ اگر S_1 و S_2 زیر مجموعه‌های ناتهی از $\bar{R}$ باشند و H یک تابع دو—صعودی و اساسی با دامنه $S_1 \times S_2$ باشد.آن‌گاه تابع H نسبت به هردو مؤلفه خود غیر نزولی خواهد بود.

اثبات:

فرض کنید a_1 و a_2 به ترتیب کوچکترین مؤلفه‌های S_1 و S_2 باشند. با قرار دادن $x_1 = a_1$ و $x_2 = a_2$ در لم ۱.۳.۱ حکم به راحتی اثبات می‌شود.

تعریف ۴.۳.۱

تابع H با دامنه $\bar{R}^2$ یک تابع توزیع توأم دو متغیره خواهد بود، اگر در شرایط زیر صدق کند:

۱. دو—صعودی باشد.

۲. $H(x,-\infty) = H(-\infty,y) = 0$ و $H(\infty,\infty) = 1$.

با توجه به این تعریف می توان نتیجه گرفت که H تابعی اساسی است و دارای توابع حاشیه‌ای

$$F(x) = H(x,\infty) \qquad , \qquad G(y) = H(\infty,y)$$

خواهد بود.

تعریف ۵.۳.۱

فرض کنید S_1 و S_2 زیر مجموعه‌های ناتهی از $\bar{R}$ باشند. اگر b_1 بزرگترین عضو S_1 و b_2 بزرگترین

---

۱۷ Grounded