دانشگاه الزهرا

دانشکده علوم پایه

پایان نامه

جهت اخذ درجه کارشناسی ارشد

رشته آمار ریاضی

عنوان

آزمون های نیکویی برازش برای مدل های ارشمیدسی توابع مفصل

استاد راهنما

دکتر صدیقه شمس

دانشجو

فاطمه رحمانی

بهمن ۱۳۹۱

بسم الله الرحمن الرحيم

تقدیم به

مادر بزرگوارم که با مهر تمام و صبر و بردباری همواره حامی و پشتیبانم در تمام مسیر زندگی ام بوده است.

و پدرم که گرچه تنهایم گذاشت اما می دانم که بهتر و بیشتر از گذشته با من است. (روحش شاد و دلش آرام ان شاء الله)

و همه عزیزانی که با لطف بی دریغ و امید روح افزای خویش روشنی راه و قوت قلبم بوده اند.

تشکر و قدردانی

آدمی در مسیر رسیدن جز در سایه ی الطاف باری تعالی، گامی نمی تواند بردارد و یکی از الطاف پروردگار بی مثال، بی شک یاری و گذر از ورطه ی حوادث است و در این راه هستند انسان هایی که با وجود ارزشمند خویش چون چراغی فروزان موجب روشنی راه و گرمی دل اند. آن که این نعمت بداد، بر ما سپاس آن را نیز نهاد. از این روست که جای دارد تلاش های بی شائبه استاد گرانقدر خویش، سرکار خانم دکتر صدیقه شمس را قدر نهم و از زحمات بی دریغشان نهایت سپاس را داشته باشم و همچنین از سرکار خانم دکتر ناهید سنجری فارسی پور و سرکار خانم دکتر احترام خوشبین که داوری این پایان نامه را بر عهده گرفتند، سپاس فراوان نموده و از همراهی و یاری دوستان شفیق و مهربانم نهایت تشکر را داشته باشم. امید است هر که بر ما حقی دارد از جام حقیقت ظرفی بردارد.

چکیده

تابع مفصل یکی از اصلی‌ترین ابزارهای بیان وابستگی است. برای بررسی این که ساختار وابستگی داده‌ها به خوبی توسط تابع مفصل منتخب بیان می‌شود، روش‌های زیادی توصیه شده است. در بسیاری از روش‌ها موضوع چند متغیره ابتدا به یک متغیره کاهش یافته و سپس از آزمون‌های یک متغیره استفاده می‌شود. هم چنین روش‌هایی وجود دارند که از چند متغیره استفاده می‌کنند. در این پایان نامه هر دو دسته از روش‌ها مورد بررسی قرار می‌گیرند. از آن جایی که خانواده مفصل ارشمیدسی کاربرد فراوانی در مدل سازی دارند، آزمون‌های نیکویی برازش خاص این خانواده نیز در این پایان نامه بررسی می‌شوند. با توجه به این که در تحلیل داده‌های به دست آمده از صنعت یا پزشکی، بسیاری از مواقع با داده‌های سانسور شده مواجه هستیم، آزمون‌های نیکویی برازش خانواده مفصل ارشمیدسی در این حالت نیز بیان می‌شود. برای بررسی عملکرد این آزمون‌ها با استفاده از شبیه‌سازی، احتمال خطای نوع اول و نوع دوم آزمون‌ها محاسبه می‌شوند.
کلید واژه‌ها : تابع مفصل، تبدیل انتگرالی احتمال، تبدیل لاپلاس، توزیع کندال، خانواده مفصل ارشمیدسی، داده‌های سانسورشده، نیکویی برازش.

فهرست مطالب

لیست جداول

لیست تصاویر

فصل ۱

معرفی توابع مفصل و مدل‌های ارشمیدسی این توابع

مقدمه و پیشینه مطالعه توابع مفصل

تابع مفصل و کاربردهای آن در آمار از جمله مباحث مهم و نسبتا جدید می‌باشد. نظریه مفصل اولین بار توسط اسکلار در سال ۱۹۵۹ مطرح شد و اصطلاح تابع مفصل را به این توابع اختصاص داد.[۴۵] این نظریه به دنبال پاسخ به چگونگی ارتباط یک تابع توزیع چند متغیره با توزیع‌های حاشیه‌ای یک بعدی که توسط فرشه مطرح شده بود، به وجود آمد. اولین مقاله مرتبط با توابع مفصل که مطالعه‌ای در زمینه وابستگی بین متغیرهای تصادفی بود، در سال ۱۹۸۱ توسط اسچیوایزر و ولف ارائه شد و در سال ۱۹۹۷، جو و در سال ۱۹۹۹، نلسن در جهت شناخت و معرفی توابع مفصل کتاب و مقالاتی ارائه کردند.[۳۳] هم اکنون دپارتمان‌های علمی زیادی بر روی چگونگی و کاربرد توابع مفصل در مدیریت ریسک مطالعه و بررسی‌های علمی و کاربردی انجام می‌دهند. بین سال‌های ۱۹۵۹ تا ۱۹۷۶ بیشتر نتایج و مطالعات به دست آمده در مورد توابع مفصل به وسیله فضاهای احتمال متریک نشان داده شده و توسعه یافتند.[۸] ساختار وابستگی بین متغیرهای تصادفی یکی از گسترده‌ترین موضوعات مورد مطالعه در احتمال و آمار است.

در مورد ضرورت مطالعه توابع مفصل دلایل متعددی وجود دارد که مهم‌ترین آن مفید بودن این توابع در ساخت و بررسی توزیع‌های چند متغیره است. ساخت توزیع‌های توأم و بررسی ویژگی‌های آن‌ها با توزیع‌های حاشیه‌ای معلوم در صورت عدم استقلال متغیرها کار پیچیده‌ای است. تابع مفصل برای ساخت یا بررسی توزیع‌های چند متغیره وقتی که توزیع‌های حاشیه‌ای یک متغیره و ساختار وابستگی آن‌ها داده شده است، بسیار مفید است. تابع مفصل ابزار مفیدی برای ساختار وابستگی می‌باشد.

۱.۱ تعریف تابع مفصل

تابع مفصل ابزاری است که ارتباط بین یک تابع توزیع چند متغیره با توزیع‌های حاشیه‌ای یک بعدی آن و ساختار وابستگی آن را نشان می‌دهد. در بیشتر مقالات و کتاب‌های آماری دو نوع تعریف برای تابع مفصل ارائه شده است هر دو تعریف برای حالت دو متغیره بیان می‌شود، یک تابع مفصل دو متغیره تابع توزیع بر روی $[0,1] \times [0,1]$ با توزیع‌های حاشیه‌ای یکنواخت می‌باشد این تعریف ساده و غیر رسمی از توزیع و تابع مفصل است.

اگر (X,Y) یک جفت متغیر تصادفی پیوسته با تابع توزیع توأم $H(X,Y)$ و توزیع‌های حاشیه‌ای $F_X(x)$ و $F_Y(y)$ باشد. با توجه به این که $V = F_Y(y) \sim U(0,1)$ و $U = F_X(x) \sim U(0,1)$ ، در این صورت تابع توزیع $C(u,v)$ تعریف شده به صورت زیر است:

$$C(u,v) = p(U \leq u, V \leq v) = p(X \leq F_X^{-1}(u), Y \leq F_Y^{-1}(v))$$
$$C(u,v) = H(F_X^{-1}(u), F_Y^{-1}(v))$$

این یک تعریف غیر رسمی ولی متعارف تابع مفصل می‌باشد.

تعریف رسمی تایع مفصل

C مفصل است اگر: $C : [0,1]^2 \longrightarrow [0,1]$ با شرایط ذیل:

۱. $C(u,0) = C(0,v) = 0$

۲. $C(u,1) = u, C(1,v) = v$

۳. اگر $u_2 \geq u_1, v_2 \geq v_1 \quad ; v_1, v_2, u_1, u_2 \in [0,1]$ آن‌گاه
$C(u_2,v_2) - C(u_1,v_2) - C(u_2,v_1) + C(u_1,v_1) \geq 0$

که شرط سوم نشان می‌دهد تابع مفصل تابعی غیر نزولی است.

تابع مفصل برای حالت nمتغیره

تعریف: یک تابع مفصل چند بعدی C با شرایط و مشخصات زیر تعریف می‌شود:

$$C : [0,1]^n \longrightarrow [0,1]$$

به شرطی که:

۱. برای هر $u=(u_1,...,u_n)$ از $[0,1]^n$ ، $C(\underline{u})=0$ اگر حداقل یکی از u_i ها صفر باشد وقتی $i=1,2,...,n$.

۲. C یک تابع صعودی است یعنی برای هر $\underline{u}$ از $[0,1]^n$ و $\underline{v}$ از $[0,1]^n$ به طوری که $\underline{u}\le\underline{v}$ ، آن گاه حجم فضای $[\underline{u},\underline{v}]$ از جعبه $[\underline{u},\underline{v}]$ نامنفی باشد.

۳. برای هر $u_i\in[0,1]$ به ازای $i=1,...,n$ ، $C(1,...,1,u_i,1,...,1)=u_i$ برقرار است.

یک تعریف غیر رسمی برای تابع مفصل چند متغیره

فرض کنید $X_1,...,X_n$ متغیرهای تصادفی پیوسته با تابع توزیع توأم $H(x_1,...,x_n)$ و توزیع‌های حاشیه‌ای $F_{X_1},...,F_{X_n}$ برای هر یک از متغیرهای تصادفی $X_1,...,X_n$ به ترتیب می‌باشد. برای هر داده $(x_1,...,x_n)$ از $(-\infty,1]^n$ نقطه‌ای در $[0,1]^{n+1}$ وجود دارد که $(F_{X_1}(x_1),...,F_{X_n}(x_n),H(x_1,...,x_n))$ که این تابعی از $[0,1]^n\longrightarrow[0,1]$ است و به آن تابع چند متغیره مفصل [1] می‌گویند.

هر دو تعریف رسمی و غیر رسمی تابع مفصل براساس قضیه مهم اسکلار به هم ارتباط پیدا می‌کند. این قضیه چگونگی ارتباط بین تابع توزیع چند متغیره با توزیع‌های حاشیه‌ای یک بعدی را نشان می‌دهد.

قضیه اسکلار : فرض کنید H تابع توزیع n بعدی با حاشیه‌های $F_{X_1},...,F_{Xn}$ باشد و هم چنین یک تابع مفصل n متغیره به نام C وجود دارد که برای همه $(x_1,...x_n)$ بر روی $(-\infty,1]^n$ داشته باشیم:

$$H(x_1,...,x_n)=C(F_{X_1}(x_1),...,F_{X_n}(x_n)) \tag{1.1}$$

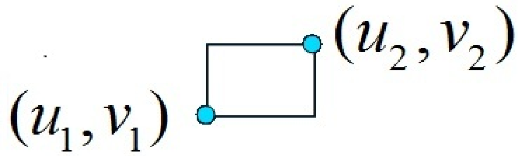
بنابراین اگر C یک تابع مفصل n بعدی و $F_{X_1},...,F_{X_n}$ تابع توزیع‌ها باشند و تابع H با (۱.۱) تعریف شده باشد، آن گاه H یک تابع توزیع با حاشیه‌ای‌های $F_{X_1},...,F_{X_n}$ است اگر این توزیع‌های حاشیه‌ای پیوسته باشند، C یکتاست در غیر این صورت C به طور یکنواخت بر روی $RankF_{X_1}\times...\times RankF_{X_n}$ تعیین شده است.

تابع مفصل یکی از ابزارهای بسیار مهم و مفید برای کاهش ابعاد توزیع‌های چند متغیره به حاشیه‌های یکنواخت شناخته شده است.

یک تابع توزیع توأم مانند H با تابع توزیع حاشیه‌های پیوسته $F_{X_1},...,F_{X_n}$ طبق قضیه اسکلار به سادگی تبدیل به ساختار تابع مفصل می‌شود. در واقع اگر C تابع مفصل باشد داریم:

$$C(u_1,...,u_n)=H(F_{X_1}^{-1}(u_1),...,F_{X_n}^{-1}(u_n))$$

[1] nCopula



شکل ۱.۱: H–حجم

که $F_{X_i}^{-1}$ تابع معکوس تابع توزیع F_{X_i} است و داریم:

$$F_{X_i}^{-1}(u_i) = sup\Big\{x_i | F_{X_i}(x_i) \leq x_i\Big\} \qquad i = 1, ..., n$$

توجه داشته باشید اگر $X_1, ... X_n$ همه متغیرهای پیوسته تصادفی با تابع توزیع فوق باشند آن گاه C تابع توزیع توأم برای متغیرهای تصادفی $U_i = F_{X_i}(x_i)$ (که U_i ها دارای توزیع یکنواخت هستند) می‌باشد. این براساس یک قضیه برای متغیرهای تصادفی پیوسته که در احتمال ثابت شده است، بیان می‌شود.

۱.۱.۱ ویژگی‌های توابع مفصل

قبل از معرفی ویژگی‌ها بعضی مفاهیم تعریف می‌شود؛ در تعاریف زیر S_1, S_2 زیر مجموعه‌های ناتهی از $R = (-\infty, \infty)$ هستند.

۱. تابع دو بعدی حقیقی H ، تابعی است که دامنه‌اش زیر مجموعه‌ی $\mathcal{R}^2$ و برد آن زیر مجموعه $\mathcal{R}$ باشد.

$$H : \mathcal{R}^2 = \mathcal{R} \times \mathcal{R} \longrightarrow \mathcal{R} = [-\infty, \infty]$$

۲. H حجم: اگر H تابع دو بعدی حقیقی با دامنه $S_1 \times S_2$ و B یک مستطیل به صورت $B = [u_1, u_2] \times [v_1, v_2]$ باشد که رئوس آن در دامنه‌ی H قرار دارد، H –حجم B با نماد $V_H(B)$ نشان داده شده و به صورت

$$V_H(B) = H(u_1, v_2) - H(u_1, v_2) + H(u_2, v_2)$$

تعریف می‌شود. (شکل۱.۱)

۳. تابع دو بعدی حقیقی H تابع دو صعودی [۲] نامیده می‌شود. اگر به ازای هر مستطیل مانند B که رئوس آن در دامنه H قرار دارد، آن گاه $V_H(B) \geqslant 0$ باشد.

لم ۱.۱.۱. اگر H تابع ۲-صعودی روی $S_1 \times S_2$ باشد به ازای
$$\begin{cases} u_1, u_2 \in S_1, \quad u_1 \leq u_2 \\ v_1, v_2 \in S_2, \quad v_1 \leq v_2 \end{cases}$$
آن‌گاه تابع

$$t \longrightarrow H(t, v_2) - H(t, v_1)$$

روی S_1 و تابع

$$t \longrightarrow H(t, u_2) - H(t, u_1)$$

روی S_2 غیر نزولی هستند. این لم با توجه به این که تابع H یک تابع ۲-صعودی است واضح است. [۳۳]

تابع مفصل یک C -حجم در محدوده $[0, u] \times [0, v]$ می‌باشد. طبق تعریف می‌توان به صورت زیر نوشت:

$$C(u, v) = V_c([0, u] \times [0, v])$$

در واقع تابع مفصل عددی در مستطیل $[0, 1] \times [0, 1]$ و نامنفی است.

$$C(u, v) = V_c([0, u], [0, v]) = V_c(A) + V_c(B) + V_{c(C)} + V_c(D)$$

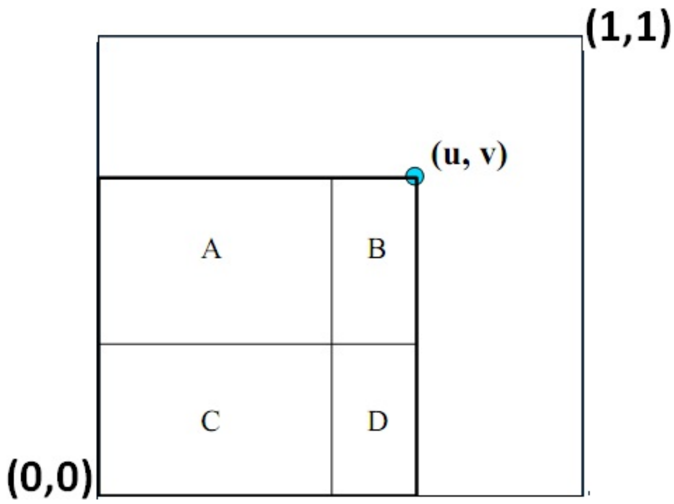
مثال ۲.۱.۱. تابع مفصل مستقل که به صورت زیر می‌باشد.

$$C(u, v) = u \times v, \quad \boldsymbol{u} \in [0, 1]^2$$

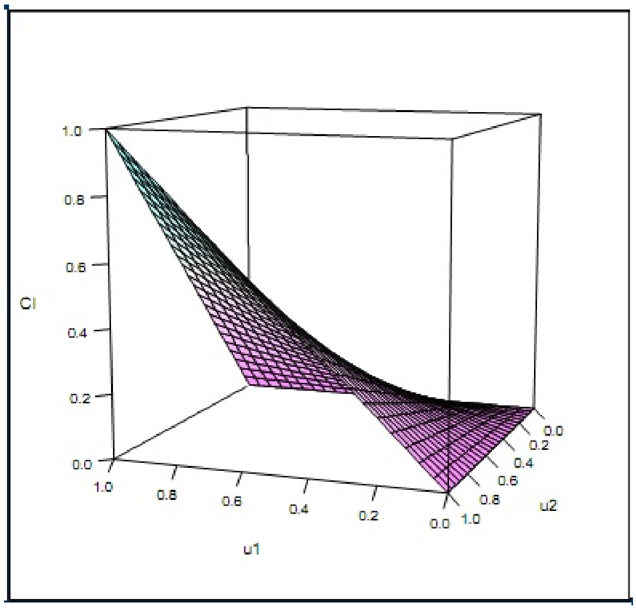
و با توجه به شکل تابع مفصل مستقل، به این مفصل، مفصل ضربی هم می‌گویند. (شکل۳.۱)

قضیه ۳.۱.۱. یک مفصل دو بعدی مانند C تابعی از $[0, 1]^2 \longrightarrow [0, 1]$ با ویژگی‌های زیر است:

[۲] 2-increasing



شکل ۲.۱: تابع مفصل c–حجم



شکل ۳.۱: تابع مفصل مستقل

۱) (شرط کرانداری) برای هر u ، z در $[0, 1]$ رابطه زیر برقرار باشد:

$$C(u, 0) = C(0, z) = 0, \qquad C(u, 1) = u, \quad C(1, z) = z$$

۲) (شرط یکنوایی) برای هر مستطیل $[u_1, u_2] \times [z_1, z_2]$ در $[0, 1]$ که $u_1 \leq u_2$ و $z_1 \leq z_2$ رابطه زیر برقرار باشد:

$$C(u_2, z_2) - C(u_2, z_1) - C(u_1, z_2) + C(u_1, z_1) \geqslant 0$$

برهان. به کتاب نلسن [۳] (۱۹۹۹) مراجعه شود.[۳۳] □

قضیه ۴.۱.۱. توابع مفصل در نامساوی زیر صدق می‌کنند:

$$W(u, z) = max(u + z - 1, 0) \leqslant C(u, z) \leqslant min(u, z) = M(u, z)$$

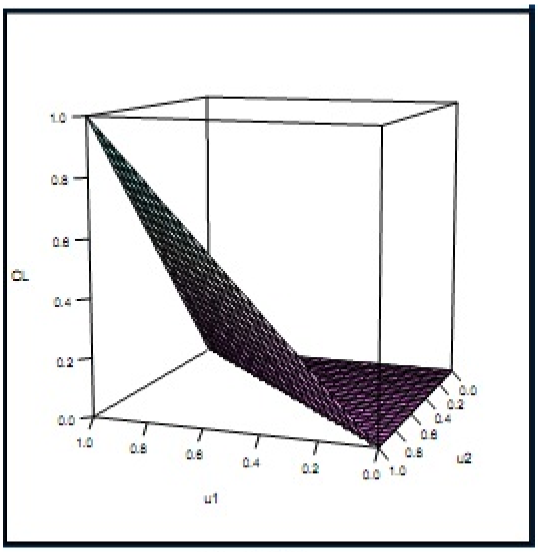
برای هر $(u, z) \in [0, 1]$. این قضیه برای توابع مفصل k بعدی هم قابل تعمیم می‌باشد. کران پایین نامساوی فوق را با C^- و کران بالا را با C^+ نشان می‌دهند.

برهان. به کتاب نلسن مراجعه شود.[۳۳] □

به این کران‌های تابع مفصل کران پایین فرشه C_L و کران بالای فرشه C_U نیز گفته می‌شود.که در نمودارهای زیر نشان داده می‌شود.(شکل ۴.۱ الف و ب) هر تابع مفصل را می‌توان با استفاده از مفصل‌های کران بالا و پایین فرشه، به صورت زیر کران بندی کرد:

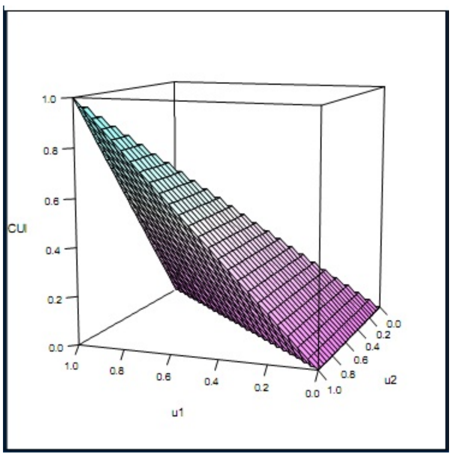
$$C_L(u_1, u_2) \leq C(u_1, u_2) \leq C_U(u_1, u_2), \qquad \forall \boldsymbol{u} \in [0, 1]^2$$

[۳] Nelsen



(الف) (ب)

شکل ۴.۱: شکل (الف)مفصل کران پایین فرشه و شکل(ب) مفصل بالای فرشه می‌باشد.



شکل ۵.۱: کران بندی تابع مفصل

۲.۱ خانواده‌هایی از توابع مفصل

در این بخش چند خانواده از توابع مفصل به همراه تابع چگالی آن‌ها معرفی شده و برخی از اندازه‌های وابستگی حاصل از آن‌ها ارائه می‌شود.
از معروف ترین مفاصل، خانواده مفاصل بیضوی [۴] و ارشمیدسی [۵] هستند. خانواده ارشمیدسی دارای شکل ریاضی ساده و خاصیت شرکت‌پذیری است. در صورتی که خانواده مفاصل بیضوی مانند گاوسی (نرمال) و t -استودنت دارای ساختار پیچیده‌تری‌اند. خانواده توابع بیضوی نظیر تابع مفصل گاوسی در اقتصادسنجی و مطالعات روش‌های مالی اهمیت ویژه دارد.

مفصل گاوسی [۶] (نرمال)

خانواده تابع مفصل گاوسی به صورت

$$C^{Ga}(u,v;\rho)=\Phi_\rho(\Phi^{-۱}(u),\Phi^{-۱}(v)) \qquad (۱.۲)$$

است ، که در آن Φ_ρ تابع توزیع نرمال استاندارد دو متغیره با ضریب همبستگی ρ است. با توجه به رابطه (۱.۲) تابع توزیع توأم به صورت

$$\Phi_\rho(\Phi^{-۱}(u),\Phi^{-۱}(v))=\int_{-\infty}^{\Phi^{-۱}(u)}\int_{-\infty}^{\Phi^{-۱}(v)}\frac{۱}{۲\pi\sqrt{۱-\rho^۲}}exp\left\{\frac{۲uv\rho-u^۲-v^۲}{۱-\rho^۲}\right\}dudv$$

ساده می‌شود.
تابع مفصل چگالی توأم خانواده گاوسی نیز به صورت

$$c^{Ga}(u,v;\rho)=\frac{۱}{۱-\rho^۲}exp\{\frac{u^۲+v^۲}{۲}+\frac{۲uv\rho-u^۲-v^۲}{۱-\rho^۲}\}$$

است.
مفصل نرمال (گاوسی) چند متغیره فرم زیر را دارد:

$$C(u_۱,...,u_k,...,u_K)=\phi_G[\phi^{-۱}(u_۱),...,\phi^{-۱}(u_k),...,\phi^{-۱}(u_k);\theta) \qquad (۱.۳)$$

---

[۴] *Eliptical Copula*

[۵] *Archimedean Copula*

[۶] *Gaussian Copula*