دانشکده علوم ریاضی

استنباط آماری اندازه‌های وابستگی دمی و برخی کاربردها

پایان‌نامه

برای دریافت درجه کارشناسی‌ارشد آمارریاضی

استاد راهنما

جناب آقای دکتر محمد امینی

استاد مشاور

جناب آقای دکتر هادی جباری نوقابی

نگارش

مهلا قاسم نژاد فرسنگی

مهر ۹۰

تقدیم به

یگانه ی بی همتا، فرمانروای مطلقی که در عین بی نیازی، به بندگان کوچک و ناتوانش عشق می ورزد

و تمام عالم امکان از ازل تا ابد، وام دار سخاوت بی انتهای اوست.

تقدیم به

حریم پاک و آسمانی مولایم علی بن موسی الرضا(ع)

غریب غریب نواز

به پاس صداقتی که در تک تک سنگ فرش های خانه اش دیدم

و زلال لحظه هایی که به گل دسته های مرقد نورانی اش نگریستم.

تقدیم به پدر و مادرم

که عمریست برای تحقق رؤیای فرزندانشان اهتمام نموده اند

و به برکت وجود ارزشمندشان زندگی ام پابرجاست.

هر کس کلمه ای به من بیاموزد، مرا بنده ی خود ساخته است. امام علی (ع)

مراتب تشکر و قدردانی خود را خاضعانه تقدیم می کنم به محضر **اساتید بزرگوارم**

جناب آقای دکتر محمد امینی که حق استادی را تمام نموده و با دلسوزی، درایت و مدیریتی بی نظیر حقیقتا بیش از تعهد کاری شان

همت گماردند، بی شک قلم ناتوان یک آموزنده از عهده ی سپاس آموزگار بر نمی آید . . .

جناب آقای دکتر هادی جباری که همواره از نظرات ارزشمند ایشان بهره فراوان برده ام و تلاش و پشتکار بی وقفه در انجام امور را در

محضر این بزرگوار آموختم .

جناب آقای دکتر ابوالقاسم بزرگ نیا و **جناب آقای دکتر جعفر احمدی** که زحمت داوری این پایان نامه را تقبل نموده و با پیشنهادات

ارزنده شان به پربارتر شدن آن کمک کردند.

جناب آقای دکتر غلامرضا محتشمی که راهنمایی های پدرانه و کلام پرمهر ایشان مایه ی دلگرمی ام بوده است.

جناب آقای دکتر محسن مددی که با افکار بلند و حمایت بی دریغشان اهداف دانشجویان را پررنگ می سازند.

مدیر محترم گروه آمار ، **سرکار خانم دکتر سیمیندخت براتپور** و سایر اساتید محترم گروه ، سرکار خانم سلیمانی، جناب آقای هادی علیزاده ،

سرکار خانم زهره زمانی و سایر دوستانم که در این راه از کمک های سخاوتمندانه شان استفاده فراوان نمودم .

و **خانواده عزیزم** که همواره صمیمیت و صداقتشان روشنایی بخش راهم بوده است.

پیش‌گفتار

کاربرد مفهوم وابستگی در شاخه‌های مختلف علمی از اهمیت بالایی برخوردار است. اندازه‌گیری وابستگی بین متغیرها در علوم پزشکی، اجتماعی، سیاسی و اقتصادی بسیار متداول است. از این‌رو تا حدی تعجب‌آور است که در متون آماری، حداقل تا اواخر سال ۱۹۶۶ به اندازه‌ی کافی به مفاهیم وابستگی توجه نشده است. مفهوم همبستگی در سال ۱۸۸۵ توسط گالتون معرفی شد و تا مدت‌ها تنها اندازه‌ی وابستگی موجود بود.

در سی سال آخر قرن بیستم، شاهد تجدید حیات سریع در مطالعات مربوط به وابستگی از نقطه‌نظر آماری و احتمالی بودیم. اولین و بهترین متن ارایه شده در زمینه‌ی مفاهیم وابستگی که در حدود ۴۰۰ صفحه بود، در اواخر سال ۱۹۹۷ توسط هری جو ارایه شد.

در این پایان‌نامه، نوعی از اندازه‌های وابستگی باعنوان اندازه‌های وابستگی دمی را بررسی می‌کنیم. معمولاً در عمل اگر بخواهیم وابستگی بین متغیرهای موجود در جوامع را به‌دست آوریم، همیشه نمی‌توان از ضرایب وابستگی معمولی مانند ρ پیرسن، τ کندال، ρ اسپیرمن و... استفاده‌کرد، زیرا همبستگی یک مفهوم خطی و گاُوسی است. به‌علاوه چون نوسانات در طول زمان تکامل می‌یابند، همبستگی می‌تواند تغییر کند.

ساختارهای وابستگی بین بازارهای بین المللی همیشه از جوانب مالی مختلف شامل انتخاب سهام، قیمت محصولات مالی و مدیریت ریسک مورد توجه بوده‌است. طبق تحقیقات امبرتس و همکاران در سال ۲۰۰۲، اندازه‌های وابستگی معمولی همیشه برای اندازه‌گیری میزان وابستگی صحیح بین بازارها مناسب نیستند. به‌عنوان مثال ρ پیرسن بنا به دلایل زیر اغلب به‌عنوان یک اندازه وابستگی، خوب عمل نمی‌کند:

۱) وابستگی‌های خطی را اندازه می‌گیرد.

۲) تحت تغییرات حاشیه‌ای پایا نیست.

۳) زیاد تحت تأثیر داده‌های پرت قرار می‌گیرد.

علاوه بر این چون نمی‌توان آن‌را بر حسب تابع مفصل بیان کرد، تنها در صورتی قابل استفاده است که توزیع جامعه را داشته باشیم. بنابراین اندازه‌های دیگر را مورد بررسی قرار می‌دهیم. اندازه‌هایی

مانندρ اسپیرمن و τ کندال را می‌توان بر حسب تابع مفصل نوشت، ولی چون این اندازه‌ها برای بررسی وابستگی توزیع‌های مقدارفرین مناسب نمی‌باشند، باز هم به یک نوع دیگر از اندازه‌ها نیاز داریم.
برای رفع این‌گونه مسایل از اندازه‌های وابستگی دمی استفاده می‌کنیم، این اندازه‌ها از این نظر مورد توجه هستند که:

۱) می‌توان آن‌ها را بر حسب تابع مفصل بیان کرد و بدون آگاهی از تابع توزیع توأم جامعه، قابل دسترسی هستند.

۲) اطلاعاتی فرا موضعی و سراسری در اختیار ما قرار داده و ما را از وجود وابستگی داخل مقادیرفرین(حدی) مجموعه داده‌ها مطلع می‌سازند.

اندازه‌های وابستگی دمی که کلین و همکاران(۲۰۱۱)، آن‌ها را با نام اندازه‌های دمی قوی [۱] معرفی کرده‌اند، وابستگی بین متغیرهارا در گوشه یک چهارم بالایی سمت راست مربع I^2 ودر گوشه یک چهارم پایین سمت چپ آن اندازه می‌گیرند، این اندازه‌ها در سال ۱۹۶۰ توسط ماسواکی سبوئیا[۲] معرفی شده و شکل متداول و امروزی‌شان را می‌توان در کتاب جو (۱۹۹۷)، یافت.
کاربرد این اندازه‌ها در موارد بسیاری از جمله در اندازه‌گیری مدیریت ریسک، پیشنهاد یک اندازه ریسک برای سهام، توصیف وابستگی داده‌ها در مالیه (انی و کاروبی، ۲۰۰۳ و سورنت، ۲۰۰۴)، بررسی پدیده شرکت انتقال (سرایت) در بازارهای بین المللی سهام(وی سان و همکاران، ۲۰۰۸)، اندازه‌گیری وابستگی پیاپی در مدل‌های ($ARCH$)، تحلیل فراوانی دومتغیره و سایرموارد محض و کاربردی حائز اهمیت است.
کولز و همکاران(۲۰۱۱)، نوع دیگری از اندازه‌های دمی به‌نام اندازه دمی باقی‌مانده[۳] را معرفی‌کردند. این اندازه که مقادیر آن در فاصله [۱, ۱−] قرار می‌گیرد، هنگامی مورد استفاده است که اندازه دمی قوی برابر با صفر باشد در حالی‌که هنوز مقداری وابستگی در دم داشته باشیم. کلین و همکاران(۲۰۱۱) این اندازه را با نام اندازه دمی ضعیف [۴] معرفی‌کرده‌اند.

---
[۱] *Strong tail dependence*
[۲] *Masaaki Sibuya*
[۳] *Residual tail dependence*
[۴] *Weak tail dependence*

لازم‌به‌ذکر است که اندازه‌های دمی نیز مانند سایر اندازه‌های وابستگی، با وجود خواص مفید و منحصربه‌فردی که دارند، از وجود برخی معایب مصون نیستند. با این حال چون در حال حاضر تنها اندازه‌های موجود برای تعیین میزان وابستگی در پیشامدهای فرین محسوب می‌شوند، تا کنون جای‌گزینی برای آن‌ها معرفی نشده است.

این پایان‌نامه مشتمل بر پنج فصل به‌شرح زیر است:

- در فصل ۱، به بیان کلیات و مفاهیم مورد نیاز پرداخته‌ایم.
- معرفی اندازه‌های وابستگی دمی و نمایش آن‌ها از طریق نمودارهای پراکنش برای چند خانواده‌ی مهم از مفصل‌ها در فصل ۲ بررسی می‌شود.
- فصل ۳، شامل معرفی میانگین‌های موزون حسابی، هندسی، هارمونیک و توانی دو مفصل و تعیین اندازه‌های وابستگی دمی برای هر یک از این میانگین‌ها است.
- در فصل ۴، ضمن ارایه‌ی سه نوع برآوردگر ناپارامتری برای اندازه‌ی دمی پایین که توسط دوبریک و اسمیت(۲۰۰۵) معرفی‌شده‌اند، سه برآوردگر متناظر را برای اندازه‌ی دمی بالا به‌دست آوردیم و معیارهای بهینگی این برآوردگرها به‌روش مونت‌کارلو بررسی شد.
- آزمون فرضیه وابستگی دمی بالا در توزیع‌های دومتغیره فرین در فصل ۵ مطالعه شده است.

مطالب جدید موجود در این پایان‌نامه را در نتیجه‌گیری انتهای هر فصل مشخص کرده‌ایم.

مهلا قاسم‌نژاد فرسنگی
مهر ۱۳۹۰

نمادها و علائم اختصاری

مفصل بقا	$\hat{C}$
تابع توزیع بقا	$\bar{H}$
اندازه وابستگی دمی قوی پایین	λ_l
اندازه وابستگی دمی قوی بالا	λ_u
اندازه وابستگی دمی ضعیف پایین	χ_l
اندازه وابستگی دمی ضعیف بالا	χ_u
مفصل حاصل‌ضرب	$\prod$
مفصل کران بالای فرشه	M
مفصل کران پایین فرشه	W
خانواده مفصل‌های ارشمیدسی با تابع مولد φ	C_φ
خانواده مفصل‌های ارشمیدسی با تبدیل لاپلاس ψ	C_ψ
خانواده مفصل‌های کوادراس	CA
خانواده مفصل‌های گامبل	GB
خانواده مفصل‌های علی—میخائیل —حق	AMH
خانواده مفصل‌های فارلی—گامبل—مورجنسترن	FGM
میانگین موزون توانی دو مفصل	$\bar{C}_r$
ترکیب خطی محدب دو مفصل	$\mathrm{C}_{1\theta}$
میانگین موزون هندسی دو مفصل	$\mathrm{C}_{2\theta}$
میانگین موزون هارمونیک دو مفصل	$\mathrm{C}_{3\theta}$
خانواده مفصل‌های سری توانی	Ps_2
بسط طیفی تابع توزیع H	$\mathrm{H}_z(c)$
نرخ استقلال دمی	ϱ

فهرست مندرجات

۸

فهرست جداول و نمودارها

فصل ۱

مفاهیم مقدماتی

۲–۱ مقدمه

۲–۲–۱ مفصل بقا

۲–۳ روش‌های ساخت مفصل

۲–۳–۱ روش معکوس

۲–۳–۲ روش مفصل‌های ارشمیدسی

۲–۳–۳ روش مفصل‌های مقدار فرین

۲–۳–۴ روش مفصل‌های تغییر شکل

۲–۴ مفصل تجربی

۱ — ۱ مقدمه

مطالعه مفصل‌ها و کاربرد آن‌ها در آمار، یک پدیده نسبتاً مدرن است. از یک نقطه نظر مفصل‌ها توابعی هستند که توابع توزیع چند متغیره را به توابع توزیع حاشیه‌ای آن‌ها متصل می کنند و از دیدگاه دیگر مفصل‌ها را می توان به صورت توابع توزیع چند متغیره‌ای تعریف کرد که توابع توزیع حاشیه‌ای آن‌ها به صورت یکنواخت روی فاصله $(\circ, ۱)$ توزیع شده است.
معمولاً در نوشته‌های آماری، مفصل را با واژه $Copula$ می شناسیم، واژه $Copula$ یک اسم لاتین است که به معنای «یک پیوند، گره و یا زنجیر»[۱] می‌باشد.
این واژه اولین بار در سال ۱۹۵۹ توسط اسکلار در یک قضیه به کار رفت. در این قضیه از $Copula$ برای توصیف توابعی که توابع توزیع تک بعدی آن‌ها برای ساخت تابع توزیع چند متغیره به هم متصل می شوند، استفاده شد.
طبق مطالعات فیشر : « مفصل ها به دو دلیل مورد علاقه آمارداناں می باشند:
۱) چون روشی برای مطالعه انواعی از اندازه‌های وابستگی هستند که به مقیاس بستگی ندارند[۲].
۲) چون بعضی مواقع با استفاده از شبیه سازی‌ها مشاهده می شود که مفصل‌ها یک نقطه شروع برای ساختن خانواده توزیع‌های دو متغیره هستند.»
طبق مطالعه هافدینگ (۱۹۴۱ — ۱۹۴۰) می توان توزیع‌های دو متغیره‌ای را یافت که تکیه‌گاه آن‌ها در مربع $[-\frac{1}{2}, \frac{1}{2}]^2$ واقع شده و توابع توزیع حاشیه‌ای آن‌ها به‌صورت یکنواخت در فاصله $[-\frac{1}{2}, \frac{1}{2}]$ پخش شده اند، طبق نظر شوایزر در سال ۱۹۹۱: «اگر هافدینگ برای نرمال‌سازی‌اش از مربع $[\circ, 1]^2$ به جای $[\frac{-1}{2}, \frac{1}{2}]^2$ استفاده می کرد، اینک او کاشف مفصل‌ها بود.»
هافدینگ بهترین کران‌های ممکن را نیز برای این توابع به دست آورد، توزیع‌های متناظر با این کران‌ها را تعیین نموده و اندازه‌های وابستگی را که پایای مقیاسی (پایا تحت تبدیلات اکیداً صعودی) هستند را مورد مطالعه قرار داد.
پس از هافدینگ، فرشه و اسکلار، توابعی که آن‌ها را امروزه تحت عنوان تابع مفصل می‌شناسیم،

[۱] $Cassell's\ Latin\ Dictionary$
[۲] $scale - free$

توسط دانشمندان دیگر مانند، کیملدورف و سمپسون (۱۹۷۵)، (تحت عنوان نمایش‌های یکنواخت) و گالامبوس (۱۹۷۸) و دی‌هیولز(۱۹۷۸)، (تحت عنوان توابع وابستگی) نیز به‌دست آمدند.

۱—۲ تابع مفصل

در قضیه زیر نشان می دهیم در صورتی‌که توابع توزیع حاشیه‌ای تک متغیره در دسترس باشند، می توان از توابع مفصل به عنوان ابزاری مفید برای دست‌یابی به توزیع‌های چندمتغیره استفاده کرد.

قضیه ۱.۱ (اسکلار، ۱۹۵۹)

اگر F یک تابع توزیع n بعدی با توابع توزیع حاشیه‌ای $F_1, ..., F_n$ باشد، آن گاه مفصل n بعدی C وجود دارد به طوری‌که برای هر $x \in \bar{\Re}$ داریم:

$$F(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)) \qquad (۱—۱)$$

که در آن اگر تمام توابع توزیع حاشیه‌ای $F_1, ..., F_n$ پیوسته باشند، مفصل متناظر C منحصربفرد (یکتا) است و آن را می‌توان به صورت زیر نوشت:

$$C(u_1, ..., u_n) = F(F_1^{-1}(u_1), ..., F_n^{-1}(u_n)) \quad ; \quad (u_1, ..., u_n) \in [0, 1]^n \qquad (۲—۱)$$

برعکس اگر C یک مفصل n بعدی باشد و $F_1, ..., F_n$ همگی تابع توزیع باشند، آن گاه F یک تابع توزیع با توابع توزیع حاشیه‌ای $F_1, ..., F_n$ می‌باشد.

در واقع در حالت n بعدی مفصل C یک تابع توزیع است که روی مکعب واحد $[0, 1]^n$ با توابع توزیع حاشیه‌ای تک بعدی یکنواخت تعریف شده‌است.

تعریف ۱.۱ تابع $C(u_1, u_2, ..., u_n) : [0, 1]^n \to [0, 1]$ را یک تابع مفصل (در حالت nبعدی) گوییم، اگر و فقط اگر

۱) مفصل C زمین‌خورده[۳] باشد:

یعنی برای هر $(u_1, ..., u_n) \in [0, 1]^n$ حداقل یک $j \in \{1, ..., n\}$ وجود داشته باشد به‌طوری‌که اگر $u_j = 0$، آن گاه،

$$C(u_1, ..., u_n) = 0.$$

۲) مفصل C، nصعودی باشد:

یعنی برای هریک از دو بردار nبعدی u و v که

$$\begin{aligned} u &= (u_1, u_2, ..., u_n) \in [0, 1]^n \\ v &= (v_1, v_2, ..., v_n) \in [0, 1]^n \end{aligned}$$

و

$$v_j \leq u_j \ , \ \ j \in \{1, ..., n\}$$

رابطه‌ی زیر برقرار باشد

$$V_c(B) = D_{u_n}^{v_n} ... D_{u_1}^{v_1} C(t_1, ..., t_n) \geq 0,$$

که در آن

$$B \ = \ ([u_1, v_1] \times ... \times [u_n, v_n]) \ , \ \ t \in [0, 1]^n,$$

و

$$D_{u_j}^{v_j} C(t_1, ..., t_n) \ = \ C(t_1, ..., t_{j-1}, v_j, t_{j+1}, ..., t_n) - C(t_1, ..., t_{j-1}, u_j, t_{j+1}, ..., t_n).$$

۳) مفصل C دارای توابع مفصل حاشیه‌ای $C_k \ , \ \ k \in \{1, ..., n\}$ می باشد که،

$$C_k(u) = u \ \ , \ \ \ u \in [0, 1]$$

---

[۳] Grounded

۱−۲−۱ مفصل بقا

برای زوج (X,Y) از متغیرهای تصادفی با تابع توزیع توأم H، تابع بقای توأم عبارت‌است از

$$\bar{H}(x,y) = P[X > x, Y > y],$$

اگر توابع توزیع حاشیه‌ای $\bar{H}$ را با $\bar{F}$ و $\bar{G}$ نشان دهیم، آن گاه،

$$\bar{F} = \bar{H}(x,-\infty) \quad , \quad \bar{G} = \bar{H}(-\infty,y).$$

سؤالی که در این‌جامطرح می شود این‌است که آیا رابطه‌ای مشابه با قضیه اسکلار بین توابع بقای توأم و تک متغیره برقراراست؟
برای پاسخ به این سؤال فرض می کنیم که X و Y دو متغیر تصادفی با تابع مفصل متناظر C باشند، آن‌گاه،

$$\begin{aligned}\bar{H}(x,y) &= ۱ - F(x) - G(y) + H(x,y)\\ &= \bar{F}(x) + \bar{G}(y) - ۱ + C(F(x),G(y))\\ &= \bar{F}(x) + \bar{G}(y) - ۱ + C(۱ - \bar{F}(x), ۱ - \bar{G}(y)).\end{aligned}$$

بنابراین تابع مفصل بقای $\hat{C}$ به صورت زیر تعریف می‌شود:

تعریف ۲.۱ اگر C تابع مفصل موجود بین متغیرهای X و Y باشد، تابع مفصل بقای $\hat{C} : I^۲ \to I$ به صورت زیراست:

$$\hat{C}(u,v) = u + v - ۱ + C(۱ - u, ۱ - v), \qquad (۳−۱)$$

همچنین بنا به قضیه اسکلار داریم،

$$\bar{H}(x,y) = \hat{C}(\bar{F}(x), \bar{G}(y)).$$

در ادامه طریقه‌ی به‌دست آوردن مفصل‌های بقا بنا به تعریف فوق با استفاده از دو مثال بیان می شود. این مثال‌ها توسط نلسن (۲۰۰٦) ارایه شده‌اند.

مثال ۱.۱ اگر دو متغیر تصادفی X و Y دارای تابع توزیع

$$H_\theta(x,y) = F(x)G(y)[1+\theta\bar{F}(x)\bar{G}(y)]$$

باشند، تابع مفصل موجود بین دو متغیر X و Y از نوع FGM[۴] می باشد. این مفصل توسط مورجنسترن (۱۹۵٦)، گامبل (۱۹۵۸) و فارلی (۱۹٦۰) معرفی شده‌است و مفصل بقای متناظر با آن به صورت زیر به‌دست می آید:

$$C_\theta(u,v) = uv[1+\theta(1-u)(1-v)]\ ,\ \ \theta\in[-1,1]$$

و

$$\begin{aligned}\hat{C}_\theta(u,v) &= u+v-1+C_\theta(1-u,1-v)\\ &= u+v-1+(1-u)(1-v)[1+\theta uv]\\ &= uv[1+\theta(1-u)(1-v)] = C_\theta(u,v).\end{aligned}$$

مثال ۲.۱ مفصل متناظر با دو متغیر تصادفی X و Y که دارای تابع توزیع نمایی دو متغیره H_θ با رابطه‌ی زیر هستند،

$$H_\theta(x,y) = \begin{cases} 1-e^{-x}-e^{-y}+e^{-(x+y+\theta xy)} & ;x,y\geq 0\ ,\ \ \theta\in[0,1]\\ 0; & \text{در غیر این صورت}\end{cases}$$

عبارت‌است از:

$$C_G(u,v) = u+v-1+(1-u)(1-v)e^{-\theta\log(1-u)\log(1-v)};$$

---

[۴] *Farlie Gumbel Morgenstern family*

که در آن

$$u = F(x) = 1 - e^{-x} \ \ , \ \ v = G(y) = 1 - e^{-y}.$$

مفصل C_G که برای اولین بار توسط گامبل (۱۹۶۰) معرفی شده‌است، مفصل گامبل نام دارد و مفصل بقای متناظر با آن به صورت زیر به‌دست می آید:

$$\begin{aligned} \hat{C_G}(u,v) &= u + v - 1 + C_G(1-u, 1-v) \\ &= u + v - 1 + (1-u) + (1-v) - 1 + uve^{-\theta \log u \log v} \\ &= uve^{-\theta \log u \log v}. \end{aligned}$$

۱—۳ روش‌های ساخت توابع مفصل

با توجه به نقش مهمی که توابع مفصل در شاخه‌های مختلف علوم از جمله بیمه و امور مالی دارند، همچنین بنا به اهمیت آن‌ها در تئوری و کاربرد، روش ساخت این توابع تا کنون مورد توجه بسیاری از محققین از جمله مارشال و الکین(۱۹۸۸) و جو(۱۹۹۳) و پیکند(۱۹۸۱)، قرار گرفته‌است. در این بخش به معرفی چند نمونه از این روش ها که چند خانواده مهم از توابع مفصل را تولید می کنند، می‌پردازیم.

۱—۳—۱ روش معکوس

فرض کنید X و Y دو متغیر تصادفی با تابع توزیع توأم H و توابع توزیع حاشیه ای F و G باشند. آن‌گاه مفصل C را می‌توان از طریق قضیه اسکلار به‌صورت زیر به‌دست‌آورد:

$$H(x,y) = C(F(x), G(y)) \ \ , \ \forall x, y \in \Re$$

در نتیجه

$$C(u,v) = H(F^{-1}(u), G^{-1}(v)) \ \ , \ u = F(x), v = G(y).$$