

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده علوم ریاضی

پایان نامه کارشناسی ارشد آمار ریاضی

آزمون‌های ناپارامتری برای وابستگی ربعی مثبت

نگارنده: زهرا صیدآبادی

استاد راهنما

دکتر نگار اقبال

استاد مشاور

دکتر حسین باغیشنی

بهمن ۱۳۹۹

پروردگارا

نه می‌توانم موهایشان را که در راه عزت من سفید شد، سیاه کنم و نه برای دست‌های پینه‌بسته آن‌ها که ثمره تلاش برای افتخار من است، مرهمی دارم. پس توفیقم ده که هر لحظه شکرگزارشان باشم و ثانیه‌های عمرم را در عصای دست بودنشان بگذرانم. شکر و سپاس خدا را که بزرگ‌ترین امید و یاور در لحظه‌لحظه‌ی زندگیست.

سپاس‌گزاری ...

سپاس بی‌کران پروردگار یکتا را که هستی‌ام بخشید.
به طریق علم و دانش رهنمونم شد.
به همنشینی رهروان علم و دانش مفتخرم نمود.
خوشه‌چینی از علم و معرفت را روزی‌ام ساخت.
تشکر و سپاس از اساتید دانشمند و پرمایه‌ام.
سرکار خانم دکتر نگار اقبال و جناب آقای دکتر حسین باغی‌شینی که از محضر پرفیضشان بهره بردم.
و تشکر از خانواده‌ام که به حمایتشان دلگرمم و تشکر از دوست عزیزم که درهای موفقیت را پیش رویم گشود.

زهرا صیدآبادی
بهمن ۱۳۹۹

تعهد نامه

اینجانب زهرا صیدآبادی دانشجوی کارشناسی ارشد رشته آمار ریاضی دانشکده علوم ریاضی دانشگاه صنعتی شاهرود، نویسنده پایان نامه با عنوان **آزمون های ناپارامتری برای وابستگی ربعی مثبت**، تحت راهنمایی **نگار اقبال** متعهد می شوم:

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش های دیگر پژوهش گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود تعلق دارد، و مقالات مستخرج با نام “دانشگاه صنعتی شاهرود” یا “Shahrood University of Technology” به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آوردن نتایج اصلی پایان نامه تاثیرگذار بوده اند، در مقالات مستخرج از پایان نامه رعایت می گردد.
- در تمام مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافت های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

زهرا صیدآبادی

بهمن ۱۳۹۹

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان نامه بدون ذکر منبع مجاز نمی باشد.

چکیده

وابستگی ربعی مثبت (PQD) به رفتار توأم متغیرهای تصادفی، زمانی که مشاهده همزمان مقادیر کوچک (یا مقادیر بزرگ) برای آن‌ها محتمل است، مرتبط می‌شود. آگاهی از حضور این ساختار وابستگی در داده‌ها می‌تواند منجر به بهبود محاسبات و تصمیم‌گیری شود. بنابراین، در این پژوهش آزمون استقلال دو متغیر تصادفی در مقابل اینکه دارای وابستگی PQD هستند، مورد بررسی قرار گرفته است. با توجه به اینکه توزیع واقعی متغیرهای تصادفی در عمل نامعلوم هستند، از نسخه‌های (توأم و حاشیه‌ای) تجربی آن‌ها در رهیافت ناپارامتری مبتنی بر تابع مفصل استفاده خواهیم کرد.

نتایج شبیه‌سازی شده نشان می‌دهد که آزمون درست‌نمایی تجربی عملکرد خوبی دارد که نشان‌دهنده این است که (X, Y) ، PQD هستند.

کلمات کلیدی: وابستگی ربعی مثبت، آزمون استقلال، تابع مفصل، آزمون درست‌نمایی تجربی

فهرست مطالب

۳	فهرست تصاویر
۵	فهرست جداول
۷	پیشگفتار
۱	۱ مفاهیم اولیه
۱	۱.۱ مقدمه
۵	۲.۱ مفصل‌ها
۱۱	۳.۱ مفصل و متغیرهای تصادفی
۱۴	۴.۱ مفصل بقا
۱۶	۵.۱ ترتیب‌های غیرتصادفی
۱۹	۲ وابستگی ربعی مثبت و ویژگی‌های آن
۱۹	۱.۲ مقدمه
۱۹	۲.۲ وابستگی ربعی مثبت
۲۱	۳.۲ وابستگی رگرسیونی
۲۱	۴.۲ وابستگی ربعی مثبت و تابع مفصل
۲۲	۵.۲ یکنوایی دمی
۲۷	۳ آزمون فرض وابستگی ربعی مثبت
۲۷	۱.۳ مقدمه
۲۹	۲.۳ مراحل آزمون
۲۹	۱.۲.۳ فرمول درست‌نمایی تجربی
۳۶	۲.۲.۳ آزمون استقلال
۳۷	۳.۲.۳ آزمون فرضیه مبتنی بر احتمال تجربی
۳۸	۴.۲.۳ آزمایش تقارن

۳۹	استنباط براساس تابع توزیع	۳.۳
۴۱	استنباط بر اساس مفصل	۴.۳
۴۴	روش آزمایش	۵.۳
۴۵	آزمون‌های مفصل مبتنی بر مسافت	۶.۳
۴۷		مطالعه شبیه‌سازی و مثال‌های کاربردی	۴
۴۹	ارزیابی آزمون‌ها با شبیه‌سازی	۱.۴
۵۱	مثال‌های واقعی	۲.۴
۵۷		مراجع	
۶۱		نمایه	

فهرست تصاویر

۱.۴	داده‌های دوقلوها: دستمزد ساعتی (به دلار آمریکا) برای $n = 149$ جفت از
۵۲	دوقلوهای همسان. نمودار سمت راست شبه-مشاهدات هستند.
۲.۴	مجموعه داده تحصیلات: نمودار پراکنش مبلغ هزینه‌شده برای هر دانشجو
	(به دلار آمریکا) در سال تحصیلی ۲۰۱۵-۲۰۱۶ در مقابل نرخ فارغ‌التحصیلی
	دبیرستان در سال ۲۰۱۶ (سمت چپ) و نمودار پراکنش شبه-مشاهدات
۵۳	(سمت راست)
۳.۴	نمودارهای پراکنش جفتی داده‌های سهام (بالا) و نمودارهای شبه-مشاهدات
۵۶	متناظر (پایین)

فهرست جداول

۵۰	..	اندازه و توان آزمون برآوردشده شش آزمون پیشنهادی برای $n = 100$	۱.۴
۵۲	داده‌های دوقلوها: آماره آزمون، مقادیر بحرانی و P -مقدار	۲.۴
۵۳	داده‌های تحصیلات: آماره آزمون، مقادیر بحرانی و P -مقدار	۳.۴
۵۴	مجموعه داده سهام: آماره آزمون، مقادیر بحرانی و P -مقدار	۴.۴

پیشگفتار

ما احتمال تجربی موضعی را برای آزمایش استقلال در متغیر تصادفی (X, Y) در مقابل اینکه (X, Y) وابسته به ربع مثبت (PQD) هستند را بررسی می‌کنیم. برقراری این نوع رابطه بین (X, Y) مورد توجه بسیاری از جمله امور مالی، بیمه، مهندسی و ... قرار گرفته است. در مهندسی اگر (X, Y) طول عمر دو مؤلفه را مشخص کند، اینکه بگوییم (X, Y) مستقل هستند، نادرست است (در حالی که آن‌ها PQD هستند). منجر به پایین آمدن قابلیت اطمینان سیستم می‌شود. بنابراین آگاهی از این نوع ساختار وابستگی می‌تواند به بهبود محاسبات کمک کند.

هدف اصلی ما ارزیابی روش تجربی (EL) برای آزمایش H_0 در مقابل $H_1 - H_0$ است که:

$$\begin{cases} H_0 : H(x, y) = F(x)G(y) & (x, y) \in R^2 \\ H_1 : H(x, y) \geq F(x)G(y) & (x, y) \in R^2 \end{cases}$$

به‌طور کلی آزمون مبتنی بر EL ترکیبی از انعطاف‌پذیری روش‌های غیرپارامتری با کارایی استنباط نسبت احتمال دارد.

محتوای مطالب این پایان‌نامه به صورت زیر تدوین شده است:

- در فصل اول به بیان برخی از تعاریف و مفاهیم اولیه که در اثبات نتایج اصلی به آن‌ها نیاز داریم، می‌پردازیم.
- در فصل دوم مفهوم وابستگی ربعی مثبت (PQD) را معرفی می‌کنیم.
- در فصل سوم، آزمون EL را توسعه می‌دهیم و یک مطالعه شبیه‌سازی که آزمون EL را با آزمون فاصله و رویکردهای موجود مقایسه می‌کند، می‌پردازیم.
- در فصل چهارم برای نشان دادن روش‌های خود از سه مجموعه داده برای شبیه‌سازی استفاده می‌کنیم.

فصل ۱

مفاهیم اولیه

۱.۱ مقدمه

وابستگی رُبعی مثبت به رفتار توأم دو متغیر تصادفی زمانی که مشاهده همزمان مقادیر کوچک (یا مقادیر بزرگ) برای آن‌ها محتمل است، مرتبط می‌شود. این نوع وابستگی کاربردهای مختلفی در امور مالی، بیمه و مهندسی پیدا کرده است. به عنوان مثال اگر X و Y بازده سهام دو شرکت در یک سبد سهام باشد، یک سرمایه‌گذار محافظه‌کار ترجیح می‌دهد این دو متغیر مستقل باشند تا وابسته از نوع PQD ، زیرا احتمال دو زیان بزرگ همزمان برای دو سهام PQD بزرگ‌تر از احتمال متناظر برای سهام‌های مستقل است (مالورن و سورنت ۲۰۰۶) به عنوان مثالی دیگر، در مهندسی اگر X و Y طول عمر دو مؤلفه باشد، به اشتباه مستقل فرض کردن آن‌ها (زمانی که PQD هستند) باعث کم‌برآورد کردن قابلیت اعتماد سیستم در یک سیستم سری و بیش‌برآورد کردن آن در یک سیستم موازی می‌شود (لای و ژی ۲۰۰۶). با این انگیزه، آگاهی از حضور این ساختار وابستگی در داده‌ها می‌تواند منجر به بهبود محاسبات و تصمیم‌گیری شود. بنابراین، آزمون استقلال دو متغیر تصادفی در مقابل اینکه دارای وابستگی PQD هستند، جالب توجه و کاربردی است.

در این فصل به برخی از تعاریف و قضایایی که در فصل‌های بعدی به آن احتیاج داریم

می‌پردازیم.

لازم به ذکر است که منبع استفاده شده در این فصل، کتاب نلسن (۱۹۹۴) است.

در ابتدا برخی از علائم را معرفی می‌کنیم:

R را روی دامنه‌ی $(-\infty, +\infty)$ و \tilde{R} را روی دامنه‌ی $[-\infty, +\infty]$ و R^2 را به صورت $\tilde{R} \times \tilde{R}$ تعریف می‌کنیم. مستطیل R^2 یک مکان دوبعدی به صورت $B = [x_1, x_2] \times [y_1, y_2]$ و رأس‌های آن به صورت $(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2)$ است. مربع واحد I^2 به صورت $I \times I$ تعریف می‌شود که $I = [0, 1]$.

تابع دوبعدی H تابعی است که دامنه آن را با $DOM H$ نمایش می‌دهیم که زیرمجموعه‌ای از R^2 است؛ به طوری که بُرد آن $(RAN H)$ یک زیرمجموعه از R است.

تعریف ۱.۱.۱. فرض کنید S_1 و S_2 زیر مجموعه‌هایی ناتهی از R باشند و H یک تابع دو بُعدی، به طوری که: $DOM H = S_1 \times S_2$.

همچنین $B = [x_1, x_2] \times [y_1, y_2]$ است که همه‌ی مربع مستطیل‌هایی است که رأس‌های آن $DOM H$ است. مقدار H - حجم^۱ مجموعه B از رابطه زیر به دست می‌آید:

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1)$$

توجه داشته باشید که تفاضل مرتبه‌ی اول H روی مستطیل B به صورت زیر تعریف می‌شود:

$$\Delta_{x_1}^{x_2} H(x, y) = H(x_2, y) - H(x_1, y)$$

$$\Delta_{y_1}^{y_2} H(x, y) = H(x, y_2) - H(x, y_1)$$

مقدار H مربع مستطیل‌های تفاضل مرتبه‌ی دوم H روی B است؛ یعنی

$$V_H(B) = \Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x, y)$$

چون

$$\begin{aligned} \Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x, y) &= \Delta_{y_1}^{y_2} (H(x_2, y) - H(x_1, y)) \\ &= (H(x_2, y_2) - H(x_1, y_2)) - (H(x_2, y_1) - H(x_1, y_1)) \\ &= V_H(B) \end{aligned}$$

تعریف ۲.۱.۱. اگر برای هر $B = [x_1, x_2] \times [y_1, y_2] \subseteq S_1 \times S_2$ داشته باشیم $V_H(B) \geq 0$ آنگاه تابع H را دو صعودی^۲ می‌نامند. توجه داشته باشید که دو صعودی بودن، معادل با غیرنزولی بودن آن نیست.

^۱H-value

^۲2-Increasing

مثال ۱.۱.۱. فرض کنید تابع H روی I^2 با ضابطه $H(x, y) = \max(x, y)$ باشد، در این صورت H تابعی نانزولی از x و y است ولی دوصعودی نیست چون:

$$V_H(I^2) = \max(1, 1) - \max(1, \circ) - \max(\circ, 1) + \max(\circ, \circ) = -1 < \circ$$

لم ۱.۱.۱. S_1, S_2 را مجموعه‌هایی ناتهی از R و H را یک تابع دوبعدی با دامنه‌ی $S_1 \times S_2$ در نظر بگیرید و همچنین x_1 و x_2 در S_1 به طوری که $x_1 \leq x_2$ است و y_1 و y_2 را در S_2 به طوری که $y_1 \leq y_2$ پس تابع $H(t, y_2) - H(t, y_1)$ به ازای هر $t \in S_1$ نانزولی است و تابع $H(x_2, t) - H(x_1, t)$ به ازای هر $t \in S_2$ نیز نانزولی است.

اثبات:

$V_H(B) \geq \circ$ تابعی دوبعدی و صعودی است

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) \geq \circ$$

برای $x_1 \leq x_2$:

$$= (H(x_2, y_2) - H(x_2, y_1)) - (H(x_1, y_2) - H(x_1, y_1)) \geq \circ$$

$$\Rightarrow H(x_2, y_2) - H(x_2, y_1) \geq H(x_1, y_2) - H(x_1, y_1)$$

پس تابع $H(t, y_2) - H(t, y_1)$ به ازای $t \in S_1$ نانزولی است.

فرض کنیم $y_1 \leq y_2$ باشد، مشابه قسمت قبل

$$V_H(B) \geq \circ$$

$$\Rightarrow (H(x_2, y_2) - H(x_1, y_2)) - (H(x_2, y_1) - H(x_1, y_1)) \geq \circ$$

$$\Rightarrow H(x_2, y_2) - H(x_1, y_2) \geq H(x_2, y_1) - H(x_1, y_1)$$

پس تابع $H(x_2, t) - H(x_1, t)$ نسبت به مؤلفه‌ی t نانزولی است.

تعریف ۳.۱.۱. فرض کنید کم‌ترین مقدار a_1 ، S_1 و کم‌ترین مقدار a_2 ، S_2 باشد. می‌گوییم

تابع H از $S_1 \times S_2$ به R زمینی^۳ است؛ اگر به ازای همه‌ی (x, y) در $S_1 \times S_2$:

$$H(a_1, y) = H(x, a_2) = \circ.$$

تعریف ۴.۱.۱. فرض کنید S_1 دارای بزرگ‌ترین عنصر b_1 و S_2 بزرگ‌ترین عنصر b_2 باشد.

تابع $H : S_1 \times S_2 \rightarrow R$ دارای توابع حاشیه‌ای F و G به صورت زیر می‌باشد:

برای هر $x \in S_1$ ، $F(x) = H(x, b_2)$ که $DOM F = S_1$ است و برای هر $y \in S_2$ ،

$G(y) = H(b_1, y)$ که $DOM G = S_2$ است.

³Grounded

مثال ۲.۱.۱. H را یک تابع با دامنه‌ی $[-1, 1] \times [0, \infty)$ و $H(x, y) = \frac{(x+1)(e^y - 1)}{x + 2e^y - 1}$ در نظر بگیرید. H یک تابع زمینی است زیرا $H(x, 0) = 0$ و $H(-1, y) = 0$ و H دارای توابع حاشیه‌ای $F(x)$ و $G(y)$ است که به صورت زیر تعریف می‌شوند:

$$F(x) = H(x, \infty) = (x+1)/2$$

$$G(y) = H(1, y) = 1 - e^{-y}$$

لم ۲.۱.۱. S_1 و S_2 زیرمجموعه‌های ناتهی از R و H تابعی دو-صعودی و زمینی با حاشیه‌های F و G و دامنه $S_1 \times S_2$ در نظر بگیرید. برای (x_1, y_1) و (x_2, y_2) هر نقطه در S_1 و S_2 داریم:

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)| \quad (1.1)$$

اثبات: با اضافه و کم کردن مقدار $H(x_1, y_2)$ و با استفاده از نامساوی مثلث داریم

$$\begin{aligned} |H(x_2, y_2) - H(x_1, y_1)| &= |H(x_2, y_2) - H(x_1, y_2) + H(x_1, y_2) - H(x_1, y_1)| \\ &\leq |H(x_2, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, y_1)| \end{aligned}$$

اکنون فرض کنید $b_1 = \sup\{x : x \in S_1\}$ و $b_2 = \sup\{y : y \in S_2\}$. دو حالت در نظر بگیرید:

در حالت اول $x_1 \leq x_2$ در نظر گرفته، طبق فرض دو-صعودی و زمینی بودن H می‌توان گفت تابع H نسبت به هر مولفه نانزولی است و طبق تعریف $h_2(y_2)$ داریم

$$h_2(y_2) = H(x_2, y_2) - H(x_1, y_2) \geq 0$$

همچنین با توجه به تعریف b_2 و نانزولی بودن h_2 خواهیم داشت

$$0 \leq H(x_2, y_2) - H(x_1, y_2) \leq H(x_2, b_2) - H(x_1, b_2) = F(x_2) - F(x_1). \quad (2.1)$$

در حالت دوم $x_1 \geq x_2$ در نظر گرفته و لذا طبق حالت اول، خواهیم داشت

$$H(x_2, y_2) - H(x_1, y_2) = -h_2(y_2) \geq -h_2(b_2) = F(x_1) - F(x_2). \quad (3.1)$$

بنابراین طبق خاصیت قدرمطلق و نامساوی‌های بالا داریم

$$|H(x_2, y_2) - H(x_1, y_2)| \leq |F(x_2) - F(x_1)|. \quad (4.1)$$

مفصل‌ها ۵

با استدلالی مشابه و در نظر گرفتن $b_1, h_1(x_1)$ و دو حالت $y_1 \leq y_2$ و $y_1 \geq y_2$ در محاسبات فوق، می‌توان به‌دست آورد

$$|H(x_1, y_2) - H(x_1, y_1)| \leq |G(y_2) - G(y_1)| \quad (۵.۱)$$

در نتیجه با توجه به محاسبات فوق بخصوص نامساوی‌های (۳.۱) و (۴.۱) خواهیم داشت

$$\forall x_1, x_2 \in S_1, \forall y_1, y_2 \in S_2; |H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|$$

در این صورت حکم ثابت می‌شود.

۲.۱ مفصل‌ها

تعریف ۱.۲.۱. تابع C' یک زیرمفصل^۴ دو بُعدی است، اگر:

- (۱) $DOM C' = S_1 \times S_2$ به‌طوری‌که S_1 و S_2 زیرمجموعه‌هایی از $I = [0, 1]$ باشند.
- (۲) C' یک تابع زمینی و دو بُعدی و افزایشی باشد.
- (۳) برای هر u در S_1 و v در S_2 :

$$C'(u, 1) = u \quad C'(1, v) = v.$$

تعریف ۲.۲.۱. اگر تابع دو بُعدی زیرمفصل C از I به I^2 شرایط زیر را داشته باشد به آن مفصل^۵ می‌گوییم.

(۱) برای هر u و v در I :

$$C(0, v) = C(u, 0) = 0$$

$$C(u, 1) = u \quad \text{و} \quad C(1, v) = v$$

(۲) برای هر u_1, u_2, v_1, v_2 در I که $u_1 \leq u_2$ و $v_1 \leq v_2$ داریم:

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

شرایط (۲) بیانگر این مطلب است که مفصل C برای هر مستطیل

$B = [c, u] \times [0, v]; u, v \in I$ عدد $C(u, v)$ را نسبت می‌دهد که مقداری در I است.

⁴Subcopula

⁵Copula

ملاحظه می‌کنید که تفاوت میان زیرمفصل و مفصل در دامنه‌ی آن‌ها است و واضح است که هر مفصل یک زیرمفصل می‌باشد ولی عکس آن برقرار نیست. لازم به ذکر است که تمامی خواص مهم برای زیرمفصل نیز برقرار هستند.

مثال ۱.۲.۱. توابع $C(u, v) = \max(u + v - 1, c)$ و $C(u, v) = \min(u, v)$ مفصل هستند که به‌طور جداگانه مفصل بودن آن‌ها را بررسی می‌کنیم:
از آنجا که برای هر $u, v \in I$:

$$u - 1 \leq \circ \quad \text{و} \quad v - 1 \leq \circ$$

پس

$$C(\circ, v) = \max(v - 1, \circ) = \circ = C(u, \circ) = \max(u - 1, \circ)$$

لذا تابع C زمینی است، همچنین:

$$C(1, v) = \max(v, \circ) = v \quad \text{و} \quad C(u, 1) = \max(u, \circ) = u$$

پس ویژگی (۱) برقرار است. حال فرض کنید $u_1, u_2, v_1, v_2 \in I$ به‌طوری که $v_1 \leq v_2$ و $u_1 \leq u_2$:

$$\begin{aligned} h(t) &= \max(t + v_2 - 1, \circ) - \max(t + v_1 - 1, \circ) \\ &= \begin{cases} (t + v_2 - 1) - (t + v_1 - 1) & t + v_2 \geq t + v_1 \geq 1 \\ (t + v_2) - \circ & t + v_2 \geq 1 \geq t + v_1 \\ \circ - \circ & 1 \geq t + v_2 \geq t + v_1 \end{cases} \\ &= \begin{cases} v_2 - v_1 & t + v_2 \geq t + v_1 \geq 1 \\ t + v_2 & t + v_2 \geq 1 \geq t + v_1 \\ \circ & 1 \geq t + v_2 \geq t + v_1 \end{cases} \end{aligned}$$

واضح است که تابع $h(t)$ یک تابع صعودی نسبت به t است پس اگر $u_2 \leq u_1$:

$$\begin{aligned} h(u_2) &\leq h(u_1) \\ \Rightarrow \max(u_2 + v_2 - 1, \circ) - \max(u_2 + v_1 - 1, \circ) \\ &\geq \max(u_1 + v_2 - 1, \circ) - \max(u_1 + v_1 - 1, \circ) \\ \Rightarrow \max(u_2 + v_2 - 1, \circ) - \max(u_2 + v_1 - 1, \circ) \\ &- \max(u_1 + v_2 - 1, \circ) + \max(u_1 + v_1 - 1, \circ) \geq \circ \end{aligned}$$

$$V_C([u_1, u_2] \times [v_1, v_2]) \geq \circ$$

برای $M(u, v) = \min(u, v)$ به صورت مشابه برای هر $u, v \in I$ داریم:

$$C(\circ, v) = \min(\circ, v) = \circ = C(u, \circ) = \min(u, \circ)$$

$$C(\mathbb{1}, v) = \min(\mathbb{1}, v) = v \quad \text{و} \quad C(u, \mathbb{1}) = \min(u, \mathbb{1}) = u$$

پس خاصیت اول برقرار است و برای خاصیت دوم داریم:

$$h(t) = \min(u_2, t) - \min(u_1, t)$$

$$= \begin{cases} t - t & t \leq u_1 \leq u_2 \\ t - u_1 & u_1 \leq t \leq u_2 \\ u_2 - u_1 & u_1 \leq u_2 \leq t \end{cases}$$

$$= \begin{cases} \circ & t \leq u_1 \leq u_2 \\ t - u_1 & u_1 \leq t \leq u_2 \\ u_2 - u_1 & u_1 \leq u_2 \leq t \end{cases}$$

با توجه به ضابطه تابع h ، $h(t)$ یک تابع صعودی نسبت به t است پس اگر $v_1 \leq v_2$

$$h(v_2) \geq h(v_1)$$

$$\Rightarrow \min(u_2, v_2) - \min(u_1, v_2) \geq \min(u_2, v_1) - \min(u_1, v_1)$$

$$\Rightarrow \min(u_2, v_2) - \min(u_1, v_2) - \min(u_2, v_1) + \min(u_1, v_1) \geq \circ$$

که با توجه به فرض $u_1 \leq u_2$ و $v_1 \leq v_2$ داریم:

$$\min(u_2, v_2) - \min(u_1, v_2) \geq \min(u_2, v_1) - \min(u_1, v_1)$$

لذا

$$V_C([u_1, u_2] \times [v_1, v_2]) \geq \circ.$$

قضیه ۱.۲.۱. فرض کنید C' یک زیرمفصل است پس برای هر $(u, v) \in \text{DOM } C'$ داریم:

$$\max(u + v - \mathbb{1}, \circ) \leq C'(u, v) \leq \min(u, v) \quad (1.1)$$

اثبات: فرض کنید (u, v) یک نقطه‌ی دلخواه در $\text{DOM } C'$ باشد، حال

$$C'(u, v) \leq C'(u, \mathbb{1}) = u$$

$$C'(u, v) \leq C'(\mathbb{1}, v) = v$$

$$C'(u, v) \leq \min(u, v)$$

همچنین

$$\begin{aligned} V_{C'}([u, 1] \times [v, 1]) &\geq 0 \\ \Rightarrow C'(1, 1) - C'(u, 1) - C'(1, v) + C'(u, v) &\geq 0 \\ \Rightarrow 1 - u - v + C'(u, v) &\geq 0 \end{aligned}$$

$$C'(u, v) \geq u + v - 1$$

و

$$C'(u, v) \geq 0$$

در نتیجه

$$C'(u, v) \geq \max(u + v - 1, 0)$$

از آنجا که هر مفصل یک زیرمفصل است، بنابراین قضیه‌ی فوق برای همه‌ی مفصل‌ها صدق می‌کند. در واقع توابع زیر مفصل هستند و معمولاً با

$$M(u, v) = \min(u, v)$$

$$W(u, v) = \max(u + v - 1, 0)$$

نمایش داده می‌شوند. بنابراین برای هر مفصل C و هر $(u, v) \in I^2$ داریم:

$$W(u, v) \leq C(u, v) \leq M(u, v)$$

به M کران بالای فرچت هافدینگ^۶ و به W کران پایین آن گفته می‌شود. یکی از توابع مفصل مهم، مفصل حاصل ضرب $\Pi(u, v) = uv$ است.

تعریف ۳.۲.۱. C را یک مفصل و a را یک عدد در I در نظر بگیرید. بخش افقی C در نقطه a یک تابع از I به I است که به وسیله‌ی $C(t, a)$ نشان داده می‌شود. بخش عمودی C در نقطه a تابعی از I به I است که با $C(a, t)$ و بخش قطری C (مرکزی C) تابعی از δ_C است که از I به I که به صورت زیر تعریف می‌شود:

$$\delta_C(t) = C(t, t)$$

که برای بخش عمودی مؤلفه‌ی اول و برای بخش افقی مؤلفه‌ی دوم را ثابت نگه می‌داریم.

تعریف ۴.۲.۱. تابع توزیع یک تابع مانند F با دامنه \bar{R} است، به طوری که:

$$F(-\infty) = 0, \quad F(\infty) = 1 \quad (1)$$

^۶Frechet-Hoffding

(۲) F نانزولی باشد.

مثال ۲.۲.۱. برای هر عدد a در R ، تابع توزیع ε_a تباهیده در a است که به صورت زیر تعریف می‌شود:

$$\varepsilon_a(a) = \begin{cases} 0 & x \in [-\infty, a) \\ 1 & x \in [a, \infty] \end{cases}$$

و برای هر تعداد از a و b در R به طوری که $a < b$ باشد، توزیع مشابه آن روی $[a, b]$ تابع توزیع یکنواخت U_{ab} به صورت زیر است:

$$U_{ab}(x) = \begin{cases} 0 & x \in [-\infty, a) \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x \in (b, \infty] \end{cases}$$

تعریف ۵.۲.۱. تابع توزیع توأم تابع H با دامنه‌ی \bar{R}^2 است، به طوری که:

۱. H دو ضابطه‌ای باشد.

$$2. H(x, -\infty) = H(-\infty, y) = 0 \text{ و } H(\infty, \infty) = 1$$

که همان شرط تابع مفصل است. بنابراین H یک تابع زمینی است؛ زیرا $\text{DOM } H = \bar{R}^2$ و H دارای حاشیه‌های F و G که به صورت $F(x) = H(x, \infty)$ و $G(y) = H(\infty, y)$ تعریف می‌شوند. بنابراین F و G توابع توزیع هستند.

مثال ۳.۲.۱. H را یک تابع با دامنه‌ی \bar{R}^2 در نظر بگیرید که به صورت زیر تعریف می‌شود:

$$H(x, y) = \begin{cases} \frac{(x+1)(e^y-1)}{x+1e^y-1} & (x, y) \in [-1, 1] \times [0, \infty] \\ 1 - e^{-y} & (x, y) \in (1, \infty] \times (0, \infty] \\ 0 & \text{سایر جاها} \end{cases}$$

H یک تابع دوبعدی و زمینی است و $H(\infty, \infty) = 1$. از آنجایی که H یک تابع توزیع توأم است حاشیه‌های H توابع توزیع F و G است که به صورت زیر تعریف می‌شوند:

$$F = U_{-1,1}$$

$$G(y) = \begin{cases} 0 & y \in [-\infty, 0) \\ 1 - e^{-y} & y \in [0, \infty] \end{cases}$$

که شرط تابع توزیع را دارد.

این قضیه مهم‌ترین قضیه‌ی این فصل است که توسط اسکالر در سال ۱۹۵۹ بیان شده است که پیوندی بین تابع توزیع توأم و تابع مفصل و توابع توزیع حاشیه‌ای است که به صورت زیر بیان می‌شود،

قضیه ۲.۲.۱. قضیه‌ی اسکالر: اگر H یک تابع توزیع دوبعدی با حاشیه‌های F ، G باشد، آنگاه یک مفصل C وجود دارد به طوری که برای همه‌ی x, y ها در R داریم:

$$H(x, y) = C(F(x), G(y))$$

نکته ۱.۲.۱. اگر F و G پیوسته باشند، C یکتاست. به عبارت دیگر C به طور منحصر به فرد روی $\text{DOM } F \times \text{DOM } G$ مشخص می‌شود اگر C یک مفصل و F و G توابع توزیع حاشیه‌ای باشند تابع H یک تابع توزیع توأم با حاشیه‌های F و G است.

قضیه ۳.۲.۱. فرض کنید C' یک زیرمفصل باشد، آنگاه برای هر (u_1, u_2) و (v_1, v_2) عضو $\text{DOM } C'$ ، رابطه زیر که به نامساوی لیپشیتز^۷ مشهور است، همواره برقرار است:

$$|C'(u_2, v_2) - C'(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|$$

از این رو C' روی دامنه‌اش به طور یکنواخت پیوسته است.

اثبات: با توجه به لم ۳.۰.۰.۱ بدیهی به نظر می‌رسد.

تعریف ۶.۲.۱. اگر F یک تابع توزیع باشد، شبه معکوس^۸ تابع F ، $F^{(-1)}$ با دامنه I است که

(۱) اگر t در دامنه F باشد، $F^{(-1)}(t)$ هر نقطه x در \tilde{R} است که $F(x) = t$

$$F(F^{(-1)}(t)) = t$$

(۲) اگر t در دامنه F نباشد، پس:

$$F^{(-1)}(t) = \inf\{x | F(x) \geq t\} = \sup\{x | F(x) \leq t\}$$

شبه معکوس هر تابع را با علامت F^{-1} نشان می‌دهند.

مثال ۴.۲.۱. شبه معکوس ε_a که در مثال ۲.۲.۱ بود به صورت زیر است:

$$\varepsilon_a^{(-1)}(t) = \begin{cases} a_0 & t = 0 \\ a & t \in (0, 1) \\ a_1 & t = 1 \end{cases}$$

از آنجا که $a_0 < a \leq a_1$ اعدادی در \tilde{R} هستند، بنابراین

⁷Lipshitz

⁸Quasi-Inverses

نتیجه ۱.۲.۱. فرض کنید H تابع توزیع توأم با توابع توزیع حاشیه‌های F و G و زیرمفصل C' به طوری که $\text{DOM } C' = \text{RAN } F \times \text{RAN } G$ و برای هر x و y عضو \bar{R} داشته باشیم $H(x, y) = C'(F(x), G(y))$ و همچنین فرض کنید $F^{(-1)}$ و $G^{(-1)}$ به ترتیب شبه‌وارون‌های F و G باشند، آنگاه برای هر (u, v) عضو $\text{DOM } C'$ داریم:

$$C'(u, v) = H(F^{(-1)}(u), G^{(-1)}(v))$$

مثال ۵.۲.۱. برای تابع توزیع H در مثال ۳.۲.۱ حاشیه‌های F و G به صورت زیر به دست می‌آیند:

$$F(x) = \begin{cases} 0 & x < -1 \\ (x+1)/2 & x \in [-1, 1] \\ 1 & x > 1 \end{cases}$$

$$G(y) = \begin{cases} 0 & y < 0 \\ 1 - e^{-y} & y \geq 0 \end{cases}$$

شبه‌معکوس‌های F و G به صورت

$$F^{(-1)}(y) = 2y - 1$$

$$G^{(-1)}(v) = -\ln(1 - v)$$

برای همه $(u, v) \in I$ است زیرا $\text{DOM } F = \text{DOM } G = I$.

$$C(u, v) = \frac{uv}{u + v - uv}$$

۳.۱ مفصل و متغیرهای تصادفی

قضیه ۱.۳.۱. فرض کنید X و Y متغیرهای تصادفی پیوسته به ترتیب با توابع توزیع F و G و تابع توزیع توأم H هستند. اگر F و G پیوسته باشند، مفصل C که در قضیه اسکالر صدق می‌کند، C یکتاست.

مفصل C در این قضیه یک مفصلی از X و Y نامیده می‌شود که به صورت C_{XY} نوشته می‌شود.

قضیه‌ی زیر، مفصل $\Pi(u, v) = uv$ را در متغیرهای تصادفی مستقل هنگامی که تابع توزیع آن پیوسته است را نشان می‌دهد که اثبات آن از قضیه قبل و اینکه X و Y مستقل هستند اگر و فقط اگر

$$\forall x, y \in \bar{R} \quad H(x, y) = F(x)G(y).$$

قضیه ۲.۳.۱. X و Y متغیرهای تصادفی پیوسته هستند. می‌گوییم X و Y مستقل اند، اگر و فقط اگر:

$$C_{XY} = \Pi$$

C به عنوان مفصل X و Y نامیده می‌شود که با متغیرهای تصادفی X و Y شناخته می‌شوند و با C_{XY} نمایش داده می‌شود.

مثال ۱.۳.۱. فرض کنید (X, Y) دارای توزیع گامبل با پارامتر Θ و H_Θ تابع توزیع توأم (X, Y) به صورت زیر باشد:

$$H_\Theta = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y+\Theta xy)} & x, y \geq 0 \\ 0 & \text{جاهای دیگر} \end{cases}$$

به طوری که پارامتر Θ مقداری در $[0, 1]$ است و توابع توزیع حاشیه‌ای $(EXP(1))$ می‌باشند که دارای معکوسی به صورت

$$G^{(-1)}(v) = -\ln(1-v) \quad \text{و} \quad F^{(-1)}(u) = -\ln(1-u)$$

برای هر $u, v \in I$ هستند از این رو می‌توان طبق قضیه اسکالر گفت که مفصل به شکل

$$C_\Theta(u, v) = H_\Theta(F^{(-1)}(u), G^{(-1)}(v))$$

$$C_\Theta(u, v) = u + v - 1 + (1-u)(1-v)e^{-\Theta \ln(1-u) \ln(1-v)}$$

خواهد بود.

قضیه ۳.۳.۱. X و Y متغیرهای تصادفی پیوسته با تابع مفصل C_{XY} را نظر بگیرید و α و β صعودی اکید به ترتیب روی دامنه‌ی $(DOM X)X$ و دامنه‌ی $(DOM Y)Y$ است. در این صورت:

(۱) اگر α صعودی اکید و β نزولی اکید باشند، پس:

$$C_{\alpha(X)\beta(Y)}(u, v) = u - C_{XY}(u, 1-v)$$

(۲) اگر α نزولی اکید و β صعودی اکید باشند، پس:

$$C_{\alpha(X)\beta(Y)}(u, v) = v - C_{XY}(1-u, v)$$

(۳) اگر α و β هر دو نزولی باشند، پس:

$$C_{\alpha(X)\beta(Y)}(u, v) = u + v - 1 + C_{XY}(1-u, 1-v)$$

اثبات: فرض کنید F_Y و G_Y به ترتیب توابع توزیع $\alpha(X)$ و $\beta(Y)$ باشند و F_X و G_X به ترتیب توابع توزیع X و Y باشند. در این صورت برای قسمت اول با توجه به اینکه α صعودی اکید و β نزولی اکید است، داریم:

$$G_Y(y) = P(\beta(Y) \leq y) = P(Y \geq \beta^{-1}(y)) = 1 - G_X(\beta^{-1}(y)) \quad (1.1)$$

$$F_Y(x) = P(\alpha(X) \leq x) = P(X \leq \alpha^{-1}(x)) = F_X(\alpha^{-1}(x))$$

بنابراین

$$\begin{aligned} C_{\alpha(X)\beta(Y)}[F_Y(x), G_Y(y)] &= P[\alpha(X) \leq x, \beta(Y) \leq y] \\ &= P[X \leq \alpha^{-1}(x), Y \geq \beta^{-1}(y)] \end{aligned}$$

حال با توجه به قاعده $P(A \cap B') = P(A) - P(A \cap B)$ داریم:

$$C_{\alpha(X)\beta(Y)}[F_Y(x), G_Y(y)] = P[X \leq \alpha^{-1}(x)] - P[X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)]$$

با توجه به (1.1) داریم:

$$\begin{aligned} C_{\alpha(X)\beta(Y)}[F_Y(x), G_Y(y)] &= F_X(\alpha^{-1}(x)) - H[\alpha^{-1}(x), \beta^{-1}(y)] \\ &= F_Y(x) - C_{XY}[F_X(\alpha^{-1}(X)), G_X(\beta^{-1}(y))] \\ &= F_Y(x) - C_{XY}[F_Y(x), 1 - G_Y(y)] \end{aligned}$$

از آنجا که X و Y پیوسته هستند و $\text{RAN } F = \text{RAN } G_Y = I$ بنابراین روی I^2 با قرار دادن $u = F_Y(x)$ و $v = G_Y(y)$ خواهیم داشت:

$$\forall u, v \in I \quad C_{\alpha(X)\beta(Y)}(u, v) = u - C_{XY}(u, 1 - v)$$

لذا برای قسمت اول حکم ثابت می‌شود. دو قسمت دیگر به‌طور مشابه قابل اثبات هستند.

لم ۱.۳.۱. X و Y متغیرهای تصادفی با توزیع توأم H را در نظر بگیرید، پس H

کران بالا فرچت هافدینگ دارد اگر و فقط اگر برای هر (X, Y) در \tilde{R}^2 داریم:

$$P[X > x, Y \leq y] = 0$$

$$P[X \leq x, Y > y] = 0$$

اثبات: F و G توابع حاشیه‌ای برای H هستند، پس:

$$\begin{aligned} F(x) &= P[X \leq x] = P[X \leq x, Y \leq y] + P[X \leq x, Y > y] \\ &= H(x, y) + P[X \leq x, Y > y] \end{aligned}$$

$$\begin{aligned} G(y) &= P[Y \leq y] = P[X \leq x, Y \leq y] + P[X > x, Y \leq y] \\ &= H(x, y) + P[X > x, Y \leq y] \end{aligned}$$

رابطه زیر برقرار است

$$H(x, y) = M(F(x), G(y)) = \min(F(x), G(y))$$

اگر و تنها اگر

$$P[X > X, Y \leq Y] = 0$$

۴.۱ مفصل بقا

در بسیاری از موارد، متغیرهای تصادفی مورد علاقه نشان‌دهنده‌ی طول عمر افراد یا اشیاست. امکان اینکه هر فرد در زمان x زنده باشد توسط تابع بقا (تابع قابلیت اطمینان) نشان داده می‌شود.

$$\bar{F}(x) = P[X > x] = 1 - F(x)$$

که در آن مانند قبل F نشان‌دهنده‌ی تابع توزیع X است. هنگام برخورد با طول عمر یک متغیر تصادفی است که اغلب از $[0, \infty)$ است. به هر حال از اصطلاح «تابع بقا» برای $P[X > x]$ استفاده می‌کنیم حتی زمانی که دامنه \bar{R} باشد. برای هر جفت از (x, y) متغیرهای تصادفی با تابع توزیع توأم H ، و تابع بقای توأم با $\bar{H}(x, y) = P[X > x, Y > y]$ و حاشیه‌های \bar{H} توابع $\bar{H}(x, -\infty)$ و $\bar{H}(-\infty, y)$ هستند که توابع بقای یک متغیره‌ی آن به ترتیب \bar{F} و \bar{G} است.

فرض کنید که مفصل X و Y ، C است. بنابراین:

$$\begin{aligned} \bar{H}(x, y) &= 1 - F(x) - G(y) + H(x, y) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + C(F(x), G(y)) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + C(1 - \bar{F}(x), 1 - \bar{G}(y)) \end{aligned}$$

بنابراین ما تابع \hat{C} را از I^2 به I به صورت زیر تعریف می‌کنیم:

$$\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$$

و داریم:

$$\bar{H}(x, y) = \hat{C}(\bar{F}(x), \bar{G}(y)) \quad (1.1)$$

مثال ۱.۴.۱. یک تابع توزیع دوتایی پارتو (هاچینسون و لای، ۱۹۹۰) X و Y متغیرهای تصادفی با تابع بقا توأم به صورت زیر در نظر بگیرید:

$$\bar{H}_\Theta(x, y) = \begin{cases} (1 + x + y)^{-\Theta} & x \geq 0, y \geq 0 \\ (1 + x)^{-\Theta} & x \geq 0, y < 0 \\ (1 + y)^{-\Theta} & x < 0, y \geq 0 \\ 1 & x < 0, y < 0 \end{cases}$$

از آنجا که $\Theta > 0$ بنابراین تابع بقا حاشیه‌ای \bar{F} و \bar{G} به صورت زیر است:

$$\bar{F}(x) = \begin{cases} (1 + x)^{-\Theta} & x \geq 0 \\ 1 & x < 0 \end{cases} \quad \bar{G}(y) = \begin{cases} (1 + y)^{-\Theta} & y \geq 0 \\ 1 & y < 0 \end{cases}$$

به طوری که X و Y توزیع پارتو یکسانی دارند. معکوس کردن تابع بقا و بکارگیری تابع بقا از نتیجه ۱.۲.۱ مفصل بقا به دست می‌آید:

$$\hat{C}_\Theta(u, v) = (u^{-1/\Theta} + v^{-1/\Theta} - 1)^{-\Theta}$$

نکته ۱.۴.۱. اگر X یک متغیر تصادفی و a یک عدد باشد و توابع توزیع متغیرهای تصادفی $X - a$ و $a - X$ یکسان باشند، یعنی برای هر $x \in R$ رابطه‌ی زیر برقرار باشد:

$$P[X - a \leq x] = P[a - X \leq x]$$

وقتی که X دارای تابع توزیع پیوسته F باشد، در این صورت:

$$F(a + x) = \bar{F}(a - x) \quad (2.1)$$

که به نقطه‌ی تقارن می‌گوییم.

تعریف ۱.۴.۱. فرض کنید X و Y متغیرهای تصادفی پیوسته با تابع توزیع H و توابع حاشیه‌ای F و G و همچنین (a, b) یک نقطه در R^2 باشد. (x, y) متقارن شعاعی^۹ روی (a, b) است اگر و تنها اگر

$$\forall x, y \in R^2 \quad H(a + x, b + y) = \bar{H}(a - x, b - y)$$

^۹Radially symmetric

قضیه ۱.۴.۱. X و Y متغیرهای تصادفی پیوسته با تابع توزیع توأم H و توابع توزیع حاشیه‌ای F و G و مفصل C را در نظر بگیرید. X و Y به ترتیب روی a و b متقارن هستند، پس X و Y متقارن شعاعی روی (a, b) هستند اگر و تنها اگر $C = \hat{C}$ و C برای همه $(u, v) \in I^2$ برابر است با:

$$C(u, v) = u + v - 1 + C(1 - u, 1 - v)$$

اثبات: براساس رابطه‌ی (۱.۱) و (۲.۱) داریم:

$$\begin{aligned} H(a+x, b+y) &= \bar{H}(a-x, b-y) & (x, y) \in R^2 \\ \iff C(F(a+x), G(b+y)) &= \hat{C}(\bar{F}(a-x), \bar{G}(b-y)) & (x, y) \in R^2 \\ \iff C(F(a+x), G(b+y)) &= \hat{C}(F(a-x), G(b-y)) & (x, y) \in R^2 \\ C(u, v) &= \hat{C}(u, v) & (u, v) \in I^2 \end{aligned}$$

۵.۱ ترتیب‌های غیرتصادفی

ترتیب‌های غیرتصادفی را با o و O نشان می‌دهند که نمادهایی برای توصیف مرتبه‌ی مجانبی کمیت‌های غیرتصادفی می‌باشند. ون در وارت (۱۹۹۸) تعاریفی از ترتیب‌های غیرتصادفی با توجه به ترتیب همگرایی در احتمال به همراه نمادهای مناسب ارائه داد که به شرح زیر می‌باشد.

تعریف ۱.۵.۱. دو دنباله از اعداد ثابت $\{a_n; n \geq 1\}$ و $\{b_n; n \geq 1\}$ را در نظر بگیرید. گوییم $b_n = O(a_n)$ هرگاه برای هر $\varepsilon > 0$ ، $K(\varepsilon) > 0$ و عدد صحیح $N(\varepsilon)$ وجود داشته باشند به طوری که برای هر $n \geq N$ ، آن‌گاه

$$|b_n| < K(\varepsilon)|a_n|$$

تعریف ۲.۵.۱. دو دنباله از اعداد ثابت $\{a_n; n \geq 1\}$ و $\{b_n; n \geq 1\}$ را در نظر بگیرید. گوییم $b_n = o(a_n)$ هرگاه

$$\lim_{n \rightarrow \infty} \left| \frac{b_n}{a_n} \right| = 0$$

با استفاده از تعاریفی که در بالا به آن اشاره شد، برای هر ثابت حقیقی مثل C ، ترتیب‌های $O(a_n)$ و $o(a_n)$ که به ترتیب معادل با $Ca_n O(1)$ ، $Ca_n o(1)$ هم‌چنین اگر $X_n = O(n^C)$

آن‌گاه $X_n = O(n^{C+1})$ قرار می‌گیرد اما عکس رابطه برقرار نیست.

فصل ۲

وابستگی ربعی مثبت و ویژگی‌های آن

۱.۲ مقدمه

در این بخش به معرفی یک نوع وابستگی می‌پردازیم که در سال (۱۹۶۶) توسط لهن صورت گرفت و یکی از مهمترین و قوی‌ترین وابستگی‌های مثبت است. در ادامه به بررسی ویژگی‌های این نوع وابستگی اشاره می‌کنیم.

۲.۲ وابستگی ربعی مثبت

تعریف ۱.۲.۲. گوییم بردار تصادفی (X, Y) یا توزیع توام آن وابسته ربعی مثبت (PQD) ^۱ است اگر برای هر $x, y \in \mathbb{R}$

$$P(X \leq x, Y \leq y) \geq P(X \leq x)P(Y \leq y) \quad (۱.۲)$$

وابستگی اکید است اگر نابرابری برای حداقل یک زوج برقرار باشد.

خانواده همه توزیع‌های F که در رابطه (۱.۲) صدق کند را با \mathcal{F}_1 نشان می‌دهیم.

^۱Positively Quadrant Dependent

تعریف ۲.۲.۲. می‌خواهیم مفهوم وابستگی به ربع مثبت را به بیش از دو متغیر تصادفی گسترش دهیم. اگر X و Y و Z سه متغیر تصادفی باشند می‌توان گفت وابسته به اورتان مثبت $(PLOD)^2$ و اگر:

$$\forall x, y, z \quad P(X \leq x, Y \leq y, Z \leq z) \geq P(Y \leq y)P(Z \leq z)$$

و می‌گوییم (X, Y, Z) به‌طور مثبت وابسته به اورتان فوقانی $(PUOD)^3$ هستند اگر:

$$\forall x, y, z \quad P(X > x, Y \geq y, Z \geq z) \geq P(X \geq x)P(Y \leq y)P(Z \geq z)$$

این دو مفهوم توسط لئون و پروشان (۱۹۷۸) مورد بررسی قرار گرفته است.

مثال ۱.۲.۲. بارلو و پروشان (۱۹۸۱)

اگرچه در بسیاری از مطالعات مربوط به قابلیت اطمینان فرض می‌شود که اجزا دارای طول عمر مستقل هستند اما فرض نوعی وابستگی بین اجزا واقع‌گرایانه‌تر است. به عنوان مثال، یک سیستم ممکن است اجزایی داشته باشد که تحت آسیب باشد یا خرابی یک قطعه منجر به افزایش بار کاری اجزای زنده مانده شود. در چنین سیستم دو مؤلفه‌ای با طول عمر X و Y ، از مدلی استفاده می‌کنیم که مقادیر کوچک X با مقادیر بزرگ X متفاوت باشد. (صرف‌نظر از اشکال در توزیع حاشیه‌ای X و Y). یعنی در این مدل X و Y ، PQD هستند.

به طور مشابه، (X, Y) یا تابع توزیع توام آن‌ها، F ، وابسته ربعی منفی $(NQD)^4$ است اگر جهت نابرابری (۱.۲) برعکس شود. مجموعه کلیه توزیع‌های وابسته ربعی منفی را با G_1 نمایش می‌دهیم. برای سادگی $(X, Y) \in \mathcal{F}$ به معنی این است که توزیع (X, Y) متعلق به کلاس \mathcal{F} است.

لم ۱.۲.۲. (لهمن؛ ۱۹۶۶)

(۱) برای هر متغیر تصادفی مثل X ، $(X, X) \in \mathcal{F}_1$.

(۲) اگر $(X, Y) \in \mathcal{F}_1$ اگر و تنها اگر $(X, -Y) \in G_1$.

(۳) اگر $(X, Y) \in \mathcal{F}_1$ و r و S توابع نانزولی (ناصعودی) دلخواه باشند آنگاه $(r(X), S(Y)) \in \mathcal{F}_1$.

(۴) می‌توان یکی یا هر دو نامساوی رابطه (۱.۲) را به طور اکید نوشت یعنی $Y \leq y$ یا $X \leq x$ را به‌صورت $Y < y$ یا $X < x$ تغییر داد.

²Positively Lower Orthant Dependent

³Positively Upper Orthant Dependent

⁴Negatively Q uadrant Dependent

(۵) رابطه (۱.۲) با هر یک از نابرابری‌های زیر معادل است.

$$P(X \leq x, Y \geq y) \leq P(X \leq x)P(Y \geq y)$$

$$P(X \geq x, Y \leq y) \leq P(X \geq x)P(Y \leq y)$$

$$P(X \geq x, Y \geq y) \geq P(X \geq x)P(Y \geq y).$$

لم ۲.۲.۲. (لهمن؛ ۱۹۶۶) اگر $(X, Y) \in \mathcal{F}_1$ آنگاه به شرط وجود امید ریاضی می‌توان نوشت

$$E(XY) \geq E(X)E(Y)$$

نتیجه ۱.۲.۲. اگر $(X, Y) \in \mathcal{F}_1$ آنگاه

$$COV(X, Y) \geq 0$$

۳.۲ وابستگی رگرسیونی

رابطه (۱.۲) را می‌توان به صورت زیر بازنویسی کرد

$$\forall x, y \in \mathbb{R} \quad P(Y \leq y | X \leq x) \geq P(Y \leq y)$$

یعنی اطلاع از کوچک بودن X ، احتمال کوچک بودن Y را افزایش می‌دهد. مفهوم شهودی وابستگی مثبت با شرط قوی‌تر زیر بهتر درک می‌شود. برای هر $x < x'$ و تمام $y \in \mathbb{R}$

$$P(Y \leq y | X \leq x) \geq P(Y \leq y | X \leq x')$$

به جای آن می‌توان شرط قوی‌تر زیر را جایگزین کرد

$$P(X \leq y | X = x)$$

تابعی غیرصعودی نسبت به x است. فرض کنید رابطه قبل تابعی صعودی نسبت به x باشد، گوییم Y وابسته رگرسیونی مثبت^۵ نسبت به X است.

۴.۲ وابستگی ربعی مثبت و تابع مفصل

فرض کنید X و Y دارای تابع توزیع توام H و به ترتیب با حاشیه‌های پیوسته F و G و مفصل C باشند، اگر برای هر $(x, y) \in \mathbb{R}^2$

^۵Positively Regression Dependent

$$H(X, Y) \geq F(X)G(Y) \quad (1.2)$$

یا به طور معادل برای $(u, v) \in I^2$

$$C(u, v) \geq uv \quad (2.2)$$

آن‌گاه متغیرهای تصادفی X و Y ، PQD هستند.

نکته ۱.۴.۲. توجه داشته باشید که مانند استقلال، وابستگی ربعی مثبت (منفی) یک ویژگی از مجموعه متغیرهای تصادفی پیوسته است که تحت تغییرات افزایشی از متغیرهای تصادفی ثابت است.

زمانی که متغیرهای تصادفی X و Y ، PQD هستند، می‌توان گفت تابع توزیع توأم آن‌ها (H) و مفصل آن‌ها (C)، PQD هستند.

توجه داشته باشید که مانند استقلال وابستگی ربعی مثبت یا منفی از خصوصیات مجموعه متغیرهای تصادفی پیوسته است در نتیجه تحت تغییرات کاملاً فزاینده متغیرهای تصادفی ثابت است. همچنین تفسیرهای دیگری در مورد (۲.۲) وجود دارد. اول اینکه اگر X و Y ، PQD باشند، شکل مفصل X و Y ، خطی رو یا بالای مثال مفصل Π مستقل است. دوم اینکه رابطه‌ی (۲.۲) همان $C \succ \Pi$ است یعنی C بزرگ‌تر از Π است پس C ، PQD بیشتری دارد. بسیاری از توابع مفصل تک پارامتری که در فصل ۲ و ۳ (نلسن، ۲۰۰۷) ارائه شد عضو خانواده F_1 و G_1 هستند. برای مثال اگر C_θ عضوی از خانواده ماردیا، خانواده FGM ، خانواده علی میخائیل حق، یا خانواده فرانک، باشد، پس برای $\theta \geq 0$ ، PQD و برای $\theta \leq 0$ ، NQD است. زیرا همه‌ی خانواده‌ها مثبت هستند و $C_\theta = \Pi$ است.

۵.۲ یکنوایی دمی

برای وابستگی ربعی مثبت می‌توان رابطه‌ی زیر را نوشت:

$$P[Y \leq y | X \leq x] \geq P[Y \leq y]$$

$$P[Y \leq y | X \leq x] \geq P[Y \leq y | X \leq \infty]$$

شرط قوی‌تر این است که برای هر $y \in R$ تابع $P[Y < y | X \leq x]$ یک تابع نزولی نسبت به x است. اگر X و Y نشان‌دهنده طول عمر اجزا در یک قابلیت اطمینان

باشد، می‌گوییم احتمال اینکه Y عمر کوتاهی داشته باشد کاهش می‌یابد (به‌طور دقیق‌تر افزایش نمی‌یابد). و طول عمر X افزایش می‌یابد. این رفتار براساس دم چپ توزیع X و Y شبیه رفتار دم راست براساس روابط قبلی است.

تعریف ۱.۵.۲. X و Y را متغیرهای تصادفی در نظر بگیرید.

۱. Y کاهش‌ی دمی چپ^۶ نسبت به X است و می‌نویسیم $LTD(Y|X)$ اگر

$$P(Y \leq y | X \leq x)$$

که یک تابع صعودی نسبت به x برای همه y هاست.

۲. X کاهش‌ی دمی چپ نسبت به Y است و می‌نویسیم $LTD(X|Y)$ اگر

$$P(X \leq x | Y \leq y)$$

یک تابع صعودی نسبت به y برای همه x هاست.

۳. Y افزایش‌ی دمی راست^۷ نسبت به X است و می‌نویسیم $RTI(Y|X)$ اگر

$$P(Y > y | X > x)$$

که یک تابع نزولی نسبت به x برای همه y هاست.

۴. Y افزایش‌ی دمی راست نسبت به X است و می‌نویسیم $RTI(X|Y)$ اگر

$$P(X > x | Y > y)$$

که یک تابع نزولی نسبت به y برای همه x هاست.

البته اگر توزیع توأم X و Y ، H باشد با حاشیه‌های F و G پس ما می‌توانیم بنویسیم $H(X, Y)/F(X)$ وقتی که $F(x) > 0$. نسبت به $P[Y \leq y | X \leq x]$ طبق روابط قبلی برقرار است.

به اصطلاح می‌گوییم «کاهش دم چپ» و «افزایش دم راست» که در اغلب موارد «کاهش» به معنای «غیر صعودی» و «افزایش» به معنای «غیر نزولی» است. خصوصیات وابستگی منفی مشابهی وجود دارد که به عنوان افزایش دمی چپ و کاهش دمی راست شناخته می‌شوند.

هریک از چهار شرط یکنواختی دم وابستگی ربع مثبت را نشان می‌دهد. به عنوان مثال، اگر $LTD(Y|X)$ برقرار باشد پس

$$P[Y \leq y | X \leq x] \geq P[Y \leq y | X \leq \infty] = P[Y \leq y]$$

^۶Left Tail Decreasing

^۷Right Tail Increasing

و از آنجا که

$$\begin{aligned} P[X \leq x, Y \leq y] &\geq P[X \leq x]P[Y \leq y|X \leq x] \\ &= P[X \leq x]P[Y \leq y] \end{aligned}$$

اگر $RTI(X|Y)$ باشند، پس:

$$P[Y > y|X > x] \geq P[Y > y|X > \infty] = P[Y > y] \quad (1.2)$$

و از آنجا که:

$$\begin{aligned} P[X > x, Y > y] &= P[X > x]P[Y > y|X > -\infty] \\ &\geq P[X > x]P[Y > y] \end{aligned}$$

بنابراین (X, Y) ، PQD هستند.

لم ۱.۵.۲. (نسلن، ۲۰۰۷) X و Y متغیرهای تصادفی پیوسته با مفصل C است. پس چهار نتیجه زیر حاصل می‌شود:

۱.

$$LTD(X|Y) \Leftrightarrow v \in I \quad \text{و} \quad C(u, v)/u$$

غیر صعودی نسبت به u است.

۲.

$$LTD(X|Y) \Leftrightarrow u \in I \quad \text{و} \quad C(u, v)/v$$

غیر صعودی نسبت به v است.

۳.

$$RTI(Y|X) \Leftrightarrow v \in I \quad \text{و} \quad [1 - u - v + C(u, v)]/(1 - u)$$

غیر نزولی نسبت به u است و یا

$$[v - C(u, v)]/(1 - u)$$

غیر صعودی نسبت به u است.

۴.

$$RTI(X|Y) \Leftrightarrow \forall u \in I \quad \text{و} \quad [1 - u - v + C(u, v)]/(1 - v)$$

غیر نزولی نسبت به v است و یا

$$[u - C(u, v)]/(1 - v)$$

غیر صعودی نسبت به v است.

تعریف ۲.۵.۲. برای متغیرهای تصادفی X و Y :

۱. Y صعودی تصادفی^۸ در X است و می‌نویسیم $SI(Y|X)$ اگر

$$P[Y > y|X = x]$$

برای هر $y \in \mathbb{R}$ تابعی غیرنزولی نسبت به x باشد.

۲. X صعودی تصادفی در Y است و می‌نویسیم $SI(X|Y)$ اگر

$$P(X > x|Y = y)$$

برای هر $x \in \mathbb{R}$ تابعی غیرنزولی نسبت به y باشد.

مثال ۱.۵.۲. (نلسن ۲۰۰۷)، فرض کنید X و Y متغیرهای تصادفی با توزیع متغیر نمایی مارشال اولکین با پارامترهای λ_1 ، λ_2 و λ_{12} باشد. احتمال بقا مشروط $P[Y > y|X = x]$ برابر است با:

$$P[Y > y|X = x] = \begin{cases} \frac{\lambda_1}{\lambda_1 + \lambda_{12}} \exp(-\lambda_{12}(y-x) - \lambda_2 y) & x \leq y \\ \exp(-\lambda_2 y) & x > y \end{cases}$$

چون تابع فوق نسبت به x غیرنزولی است پس $SI(Y|X)$ است.

لم ۲.۵.۲. (نلسن، ۲۰۰۷) اگر X و Y متغیرهای تصادفی پیوسته با تابع مفصل C است، پس:

۱. $SI(Y|X)$ اگر و تنها اگر برای هر $v \in I$ و برای اغلب u ها $\partial C(u, v)/\partial u$ تابعی غیرصعودی نسبت به u است.

۲. $SI(X|Y)$ اگر و تنها اگر برای $u \in I$ و برای اغلب v ها $\partial C(u, v)/\partial v$ تابعی غیرصعودی نسبت به v است.

نتیجه ۱.۵.۲. (X, Y) متغیرهای تصادفی پیوسته با مفصل C هستند، پس:

$$۱) SI(Y|X) \Leftrightarrow v \in I$$

و $C(u, v)$ یک تابع مقعر برای u است.

$$۲) SI(X|Y) \Leftrightarrow u \in I$$

و $C(u, v)$ یک تابع مقعر برای v است.

⁸Stochastically Increasing

قضیه ۱.۵.۲. (نلسن، ۲۰۰۷) فرض کنید (X, Y) متغیرهای تصادفی پیوسته با مفصل C هستند. پس:

۱. اگر $SI(Y|X)$ باشد، پس $LTD(Y|X)$ و $RTI(Y|X)$

۲. اگر $SI(X|Y)$ باشد، پس $LTD(X|Y)$ و $RTI(X|Y)$

فصل ۳

آزمون فرض وابستگی ربعی مثبت

در این فصل به معرفی آزمون برای تشخیص وابستگی ربعی مثبت می‌پردازیم و مراحل و آزمون و آزمون‌های مفصل‌های مبتنی بر مسافت را بررسی می‌کنیم. مطالب این فصل از تانگ و همکارانش (۲۰۱۹) و دنویت (۲۰۰۴) و عینمل (۲۰۰۳) آورده شده‌اند.

۱.۳ مقدمه

وابستگی ربعی مثبت (PQD) به رفتار توأم متغیرهای تصادفی، زمانی که مشاهده همزمان مقادیر کوچک (یا مقادیر بزرگ) برای آن‌ها محتمل است، مرتبط می‌شود. آگاهی از حضور این ساختار وابستگی در داده‌ها می‌تواند منجر به بهبود محاسبات و تصمیم‌گیری شود.

فرض کنید (X, Y) یک بردار تصادفی پیوسته با تابع توزیع تجمعی توأم H باشد و F و G به ترتیب توابع توزیع حاشیه‌ای X و Y را نشان می‌دهند. توجه داشته باشید که وابستگی ربعی مثبت دلالت بر همبستگی دارد اما برعکس آن صحت ندارد. وابستگی ربعی مثبت در زمینه‌های مالی، بیمه، مهندسی و ... کاربرد دارد. به‌طور مثال اگر X و Y بازده دو سهام در یک نمونه کار باشند، ممکن است سرمایه‌گذاران محتاط X و Y مستقل را به PQD ترجیح دهند زیرا احتمال ضرر همزمان برای هر دارایی PQD

بزرگ‌تر است (تحت استقلال). بنابراین آگاهی از این نوع ساختار وابستگی می‌تواند به بهبود محاسبات مورد نیاز برای ایجاد برنامه کمک کند. استقلال X و Y را در مقابل اینکه X و Y ، PQD هستند، یعنی H_0 را در مقابل $H_1 - H_0$ به‌طوری‌که:

$$\begin{cases} H_0 : H(x, y) = F(x)G(y) & (x, y) \in \mathbb{R}^2 \\ H_1 : H(x, y) \geq F(x)G(y) & (x, y) \in \mathbb{R}^2 \end{cases}$$

در اینجا لازم به ذکر است که F و G ناشناخته هستند و H_0 و H_1 هر دو مرکب هستند. علاوه بر این جایگزین مرتب شده است که باعث ایجاد مشکلات کلی مرتبط با استنباط محدود می‌شود.

هدف اصلی ما ارزیابی یک روش تجربی (EL) برای آزمایش H_0 در مقابل $H_1 - H_0$ است. روش EL توسط اون (۱۹۸۸) و اون (۱۹۹۰) محبوبیت یافت تا فاصله اطمینان را برای پارامترها در معادلات بسازد.

آزمون مبتنی بر EL ، انعطاف‌پذیری روش‌های غیرپارامتری را با کارایی استنباط مبتنی بر نسبت احتمال ترکیب می‌کند. روش EL به راحتی قابل اجراست و تحت انواع ساختار وابستگی عملکرد خوبی دارد.

آزمون درست‌نمایی تجربی^۱

آزمون درست‌نمایی یکی از روش‌های آزمون فرض آماری است که بین درستی یک مجموعه و متمم آن تصمیم می‌گیرد یعنی $H_0 : \theta \in \Theta_0$ در برابر متمم آن یعنی $H_1 : \theta \in \Theta_0^c$ را بررسی می‌کند. نسبت درست‌نمایی به صورت زیر تعریف می‌شود:

$$R(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta|x)}{\sup_{\theta \in \Theta_0^c} L(\theta|x)}$$

که در آن $L(\theta|x)$ درست‌نمایی داده است. یک آزمون نسبت درست‌نمایی آزمونی است که فرض H_0 را در صورت $R(x) \leq C$ رد کند.

معیاری از ارتباط کلی بین دو متغیر تصادفی (X, Y) ، کواریانس بین هر جفت از توابع حقیقی نزولی f و g است که به صورت زیر تعریف می‌شود:

$$cov[f(X), g(Y)] \geq 0$$

اگر (X, Y) ، PQD باشند، آنگاه:

$$cov(X, Y) \geq 0$$

¹Empirical Likelihood Test

اگر f و g یک جفت از توابع نزولی حقیقی باشند و (X, Y) ، PQD باشند بیانگر این است که $[f(X), g(Y)]$ ، PQD است که به نوبه خود بیانگر این است که

$$cov(f(X), g(Y)) \geq 0$$

۲.۳ مراحل آزمون

۱.۲.۳ فرمول درست‌نمایی تجربی

فرض کنید $\{(X_i, Y_i)\}_{i=1}^n$ نمونه‌ای مستقل و یکنواخت (iid) از توزیع H باشند که ما می‌خواهیم H_0 را در مقابل H_1 که در بالا تعریف شده است آزمون کنیم. $\Theta^{x,y}$ مجموعه‌ای از همه‌ی توابع توزیع توأم دو متغیره را مشخص می‌کند که $\tilde{H}(x, y) = \tilde{F}(x)\tilde{\Theta}(y)$ و $\Theta^{x,y}$ یک مجموعه مشابه به‌طوری که $\tilde{H}(x, y) \geq \tilde{F}(x)\tilde{\Theta}(y)$ است. پس نسبت درست‌نمایی تجربی (موضعی) درست‌نمایی در (X, Y) به‌صورت زیر است:

$$R_n(x, y) = \frac{\sup\{L(\tilde{H}); \tilde{H} \in \Theta^{x,y}\}}{\sup\{L(H); H \in \Theta^{x,y}\}} \quad (1.3)$$

به‌طوری که $L(\tilde{H}) = \prod_{i=1}^n \tilde{P}_i$ یک تابع EL و \tilde{P}_i احتمال اختصاص داده شده به (X_i, Y_i) برای $i = 1, 2, \dots, n$ با تابع \tilde{H} و حاشیه‌های F و G است، چون $L(\tilde{H}) = 0$ موقعی که \tilde{H} پیوسته است پس فقط کافی است که توزیع‌های گسسته \tilde{H} با $\tilde{P}_i > 0$ موقعی که ماکسیمم $L(\tilde{H})$ است را در نظر بگیریم. برای مثال توابع توزیع تجربی توأم

$$H_n(a, b) = n^{-1} \sum_{i=1}^n I(X_i \leq a, Y_i \leq b)$$

با احتمال $\frac{1}{n}$ برای هر (X_i, Y_i) و ماکسیمم کردن $L(\tilde{H})$ برای همه‌ی توابع توزیع تجمعی است.

لازم است که تابع $L(\tilde{H})$ را برای $\Theta^{x,y}$ و $\Theta^{x,y}$ ماکسیمم کنیم. چهار مجموعه زیر را در نظر بگیرید:

$$\begin{aligned} A_{11} &= (-\infty, X] \times (-\infty, Y] & A_{12} &= (-\infty, X] \times (Y, \infty) \\ A_{21} &= (X, \infty) \times (-\infty, Y] & A_{22} &= (X, \infty) \times (Y, \infty) \end{aligned}$$

پس $\tilde{\Theta}_i = \tilde{P}_i / \tilde{\Phi}_{rs}$ به‌طوری که $(X_i, Y_i) \in A_{rs}$ به‌طوری که

$$\begin{aligned} \tilde{\Phi}_{rs} &= \tilde{\Phi}_{rs}(x, y) = P_{\tilde{H}}(x, y) \in A_{rs} \\ &= P_{\tilde{H}}(A_{rs}) \quad \forall r, s \in \{1, 2\} \end{aligned}$$

در اصطلاح پارامتری جدید $\tilde{\Theta}_i$ و $\tilde{\Phi}_{rs}$ ، $L(\tilde{H})$ را می‌توانیم به صورت زیر بنویسیم:

$$L(\tilde{H}) = \prod_{i=1}^n \tilde{\Theta}_i \prod_{r=1}^2 \prod_{s=1}^2 \tilde{\Phi}_{rs} n \hat{P}_H(A_{rs}) \quad (۲.۳)$$

به طوری که $\hat{P}_H(B) = n^{-1} \sum_{i=1}^n I((X_i, Y_i) \in B)$ برای هر مجموعه $B \subset R^2$ و \hat{P}_H اندازه احتمال تجربی مطابق با H است. توجه داشته باشید که برای سادگی معادله $L(\tilde{H})$ را روی (x, y) در معادله (۲.۳) فرض کرده‌ایم. در اصطلاح $\tilde{\Theta}_i$ و $\tilde{\Phi}_{rs}$ در فرضیه $H_0^{x,y}$ برابر است با:

$$\begin{aligned} 0 < \tilde{\Theta}_i &\leq 1, & \sum_{(X_i, Y_i) \in A_{rs}} \tilde{\Theta}_i &= 1 \\ \tilde{\Phi}_{rs} &= \tilde{\Phi}_{r+} \tilde{\Phi}_{+s}; & r, s &\in \{1, 2\} \end{aligned}$$

به طوری که

$$\begin{aligned} \tilde{\Phi}_{1+} &= \tilde{F}(x) \\ \tilde{\Phi}_{2+} &= 1 - \tilde{F}(x) \\ \tilde{\Phi}_{+1} &= \tilde{G}(y) \\ \tilde{\Phi}_{+2} &= 1 - \tilde{G}(y) \end{aligned}$$

و به طور مشابه برای فرضیه $H_1^{x,y}$ داریم:

$$0 < \tilde{\Theta}_i \leq 1 \quad \text{و} \quad \sum_{(X_i, Y_i) \in A_{rs}} \tilde{\Theta}_i = 1$$

$$\begin{cases} \tilde{\Phi}_{11} \geq \tilde{\Phi}_{1+} \tilde{\Phi}_{+1} & \tilde{\Phi}_{12} \leq \tilde{\Phi}_{1+} \tilde{\Phi}_{+2} \\ \tilde{\Phi}_{21} \leq \tilde{\Phi}_{2+} \tilde{\Phi}_{+1} & \tilde{\Phi}_{22} \geq \tilde{\Phi}_{2+} \tilde{\Phi}_{+2} \end{cases}$$

با بررسی معادله (۱.۳) و (۲.۳) فواید اندازه‌گیری مجدد نشان داده می‌شود. برای مثال در فرضیه $H_0^{x,y}$ داریم:

$$\sup\{L(\tilde{H}) : \tilde{H}(x, y) = \tilde{F}(x)\tilde{G}(y)\} = \sup_{\tilde{\Theta}_1, \dots, \tilde{\Theta}_n} \prod_{i=1}^n \tilde{\Theta}_i \sup_{\tilde{\Phi} \in H_0^{x,y}} \prod_{r=1}^2 \prod_{s=1}^2 \tilde{\Phi}_{rs} n \hat{P}_H(A_{rs})$$

به طوری که:

$$\tilde{\Phi} = (\tilde{\Phi}_{11}, \tilde{\Phi}_{12}, \tilde{\Phi}_{21}, \tilde{\Phi}_{22})$$

به طور مشابه در فرضیه $H_1^{x,y}$ داریم:

$$\sup\{L(\tilde{H}) : \tilde{H}(x, y) \geq \tilde{F}(x)\tilde{G}(y)\} = \sup \prod_{i=1}^n \tilde{\Theta}_i \sup_{\tilde{\Phi} \in H_1^{x,y}} \prod_{r=1}^2 \prod_{s=1}^2 \tilde{\Phi}_{rs} n \hat{P}_H(A_{rs})$$

توجه داشته باشید که محدودیت‌های اولیه در معادله (۲.۳) یعنی $\prod_{i=1}^n \tilde{\Theta}_i$ در فرضیه‌های $H_1^{x,y}$ و $H_0^{x,y}$ یکسان هستند. بنابراین هنگام محاسبه EL محلی نسبت به $R_n(x, y)$ ، $\sup_{\tilde{\Theta}_1, \dots, \tilde{\Theta}_n}$ حذف می‌شود و هیچ سهمی ندارد پس:

$$\frac{\sup_{\tilde{\Phi} \in H_0^{x,y}} \prod_{r=1}^{\gamma} \prod_{s=1}^{\gamma} \tilde{\Phi}_{rs}^{n\hat{P}_H(A_{rs})}}{\sup_{\tilde{\Phi} \in H_1^{x,y}} \prod_{r=1}^{\gamma} \prod_{s=1}^{\gamma} \tilde{\Phi}_{rs}^{n\hat{P}_H(A_{rs})}}$$

یعنی نسبت دو احتمال چندجمله‌ای (بدون در نظر گرفتن ثابت‌ها) $n\hat{P}_H(A_{rs})$ و $r, s \in \{1, 2\}$ شمارش می‌شود مطابقت دارد.

نمونه iid با $a = 5^\circ$ از H با حاشیه‌های F و G است. احتمال واقعی مرتبط با این چهار مجموعه به صورت زیر است:

$$\begin{aligned} \Phi_{11} &= P_H(A_{11}) = H(x, y) & \Phi_{12} &= P_H(A_{12}) = F(x) - H(x, y) \\ \Phi_{21} &= P_H(A_{21}) = G(y) - H(x, y) & \Phi_{22} &= P_H(A_{22}) = 1 - F(x) - G(y) + H(x, y) \end{aligned}$$

ماکسیم چندجمله‌ای‌ها تحت فرضیه‌های $H_1^{x,y}$ و $H_0^{x,y}$ مانند آنچه که در بالا تعریف شد خواهد بود یعنی ماکسیم $\Phi_{rs} = P_H(A_{rs})$ تحت فرضیه $H_0^{x,y} = H(x, y) = F(x)G(y)$ برابر است با:

$$\begin{aligned} \hat{\Phi}_{11}^{(0)} &= F_n(x)G_n(y) \\ \hat{\Phi}_{11}^{(1)} &= \{1 - F_n(x)\}G_n(y) \\ \hat{\Phi}_{12}^{(0)} &= F_n(x)\{1 - G_n(y)\} \\ \hat{\Phi}_{12}^{(1)} &= \{1 - F_n(x)\}\{1 - G_n(y)\} \end{aligned}$$

به طوری که

$$\begin{aligned} F_n(a) &= n^{-1} \sum_{i=1}^n I(X_i \leq a) \\ G_n(b) &= n^{-1} \sum_{i=1}^n I(Y_i \leq b) \end{aligned}$$

که حاشیه‌های تابع توزیع تجمعی هستند. ماکسیم تحت فرضیه $H_1^{x,y} : H(x, y) \geq F(x)G(y)$ برابر است با:

$$\begin{aligned} \hat{\Phi}_{11}^{(1)} &= \max\{\hat{\Phi}_{11}, \hat{\Phi}_{11}^{(0)}\} \\ \hat{\Phi}_{21}^{(1)} &= \min\{\hat{\Phi}_{21}, \hat{\Phi}_{21}^{(0)}\} \\ \hat{\Phi}_{12}^{(1)} &= \min\{\hat{\Phi}_{12}, \hat{\Phi}_{12}^{(0)}\} \\ \hat{\Phi}_{22}^{(1)} &= \max\{\hat{\Phi}_{22}, \hat{\Phi}_{22}^{(0)}\} \end{aligned}$$

به طوری که:

$$\hat{\Phi}_{rs} = \hat{P}(A_{rs}) \quad \forall r, s \in \{1, 2\}$$

برآورد ماکسیمم احتمال چندجمله‌ای‌ها $\Phi_{rs} = P_H(A_{rs})$ است. با ترکیب کردن هر دو مجموعه ماکسیمم‌کننده‌ها داریم:

$$R_n(x, y) = \frac{\prod_{r=1}^2 \prod_{s=1}^2 \hat{\Phi}_{rs}^{(0)} n^{\hat{\Phi}_{rs}}}{\prod_{r=1}^2 \prod_{s=1}^2 \hat{\Phi}_{rs}^{(1)} n^{\hat{\Phi}_{rs}}} = \prod_{r=1}^2 \prod_{s=1}^2 \left(\frac{\hat{\Phi}_{rs}^{(0)}}{\hat{\Phi}_{rs}^{(1)}} \right)^{n^{\hat{\Phi}_{rs}}}$$

که پس از انجام جبر ابتدایی به صورت زیر نوشته می‌شود: (I تابع نشانگر است)

$$R_n(x, y) = I(\hat{\Phi}_{11} \leq \hat{\Phi}_{11}^{(0)}) + 2 I(\hat{\Phi}_{11} > \hat{\Phi}_{11}^{(0)}) \prod_{r=1}^2 \prod_{s=1}^2 \left(\frac{\hat{\Phi}_{rs}^{(0)}}{\hat{\Phi}_{rs}^{(1)}} \right)^{n^{\hat{\Phi}_{rs}}}$$

در محاسبه‌ی $R_n(x, y)$ نسبت ماکسیمم تابع EL برای یافتن برآوردگر حداکثر احتمال (MLE) ، پارامتر چندجمله‌ای $\Phi = (\Phi_{11}, \Phi_{22}, \Phi_{21}, \Phi_{12})$ کاهش می‌یابد، دو راه وجود دارد:

۱. شامل محدودیت نیست $H(x, y) = F(x)G(y)$

۲. شامل محدودیت است $H(x, y) \geq F(x)G(y)$

MLE تحت فرضیه $H^{x,y}$ به صورت زیر است:

$$MLE(\Phi_{11} = H(x, y))$$

برابر با:

$$\hat{\Phi}_{11}^{(0)} = F_n(x)G_n(b)$$

به طوری که:

$$F_n(a) = n^{-1} \sum_{i=1}^n I(X_i \leq a)$$

$$G_n(b) = n^{-1} \sum_{i=1}^n I(Y_i \leq b)$$

که توابع توزیع تجربی حاشیه‌ای هستند.

احتمال چندجمله‌ای‌های باقی‌مانده $\Phi_{rs} = P_H(A_{rs})$ تحت $H^{(x,y)}$ به صورت زیر محاسبه می‌شود:

$$\hat{\Phi}_{11}^{(0)} = F_n(x)G_n(y)$$

$$\hat{\Phi}_{12}^{(0)} = F_n(x)\{1 - G_n(y)\}$$

$$\hat{\Phi}_{21}^{(0)} = \{1 - F_n(x)\}G_n(y)$$

$$\hat{\Phi}_{22}^{(0)} = \{1 - F_n(x)\}\{1 - G_n(y)\}$$

و تحت فرضیه $H_1^{(x,y)}$ به طوری که $\max \Phi_{rs} = P_H(A_{rs})$ داریم:

$$\hat{\Phi}_{11}^{(1)} = \max\{\hat{\Phi}_{11}, \hat{\Phi}_{11}^{(0)}\}$$

$$\hat{\Phi}_{12}^{(1)} = \min\{\hat{\Phi}_{12}, \hat{\Phi}_{12}^{(0)}\}$$

$$\hat{\Phi}_{21}^{(1)} = \min\{\hat{\Phi}_{21}, \hat{\Phi}_{21}^{(0)}\}$$

$$\hat{\Phi}_{22}^{(1)} = \max\{\hat{\Phi}_{22}, \hat{\Phi}_{22}^{(0)}\}$$

به طوری که $\hat{\Phi}_{rs} = \hat{P}_H(A_{rs})$ ، MLE پارامترهای Φ_{rs} برای $r, s \in \{1, 2\}$ با هر دو مجموعه تخمین زده شده چند مجموع نسبت احتمال تجربی درست‌نمایی آماری $R_n(x, y)$ به صورت زیر محاسبه می‌شود:

$$\begin{aligned} R_n(x, y) &= \prod_{r=1}^2 \prod_{s=1}^2 \left(\frac{\hat{\Phi}_{rs}^{(0)}}{\hat{\Phi}_{rs}^{(1)}} \right)^{n\hat{\Phi}_{rs}} \\ &= I(\hat{\Phi}_{11} \leq \hat{\Phi}_{11}^{(0)}) + I(\hat{\Phi}_{11} > \hat{\Phi}_{11}^{(0)}) \prod_{r=1}^2 \prod_{s=1}^2 \left(\frac{\hat{\Phi}_{rs}^{(0)}}{\hat{\Phi}_{rs}^{(1)}} \right)^{n\hat{\Phi}_{rs}} \end{aligned}$$

که $R_n(x, y)$ با مقادیر کوچک باعث رد $H_0(x, y)$ می‌شود.

اگر MLE بدون محدودیت برای $\tilde{\Phi} = H_n(x, y)$ خیلی کوچک باشد یعنی کمتر از $\tilde{\Phi}_{11}^{(0)} = F_n(x)G_n(y)$ باشد، این امر تقریباً هیچ مدرکی علیه استقلال ندارد و از آنجاکه $R_n(X, Y) = 1$ است پس $H_0^{x,y}$ نمی‌تواند رد شود و فقط وقتی ممکن است که $\tilde{\Phi}_{11} > \tilde{\Phi}_{11}^{(0)}$ باشد. البته رد شدن $H_0^{x,y}$ موقعی که $R_n(x, y)$ خیلی کوچک است معادل است با $H_0^{x,y}$ موقعی که $-2 \ln R_n(x, y)$ بزرگ باشد. آزمون EL ، H_0 در مقابل $H_1 - H_0$ با همه آزمون‌های محلی که منحصر به آماره آزمون می‌شود شکل گرفته است.

$$\begin{aligned} EL_n &= \int_{R^2} -2 \ln R_n(x, y) dF_n(x) dG_n(y) \\ &= \frac{-2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \ln R_n(X_i, Y_j) \end{aligned}$$

و به موارد خاص تحت فرضیه H_0 داریم:

$$EL_n \xrightarrow{d} \int_{[0,1]^2} \frac{[\{\mathcal{B}(u, v) - u\mathcal{B}(1, v) - v\mathcal{B}(y, 1)\} + 1]^2}{u(1-u)v(1-v)} du dv$$

اگر $n \rightarrow \infty$ ، $\mathcal{B}(0, 0)$ یک پل بروانی استاندارد است؛ یعنی یک روند متوسط صفر گاوسی با کواریانس زیر است:

$$Cov(\mathcal{B}(u_1, v_1), \mathcal{B}(u_2, v_2)) = \min\{u_1, u_2\} \min\{v_1, v_2\} - u_1 v_1 u_2 v_2$$

داریم:

$$\forall a \in \mathbb{R} \quad a_+ = \max\{a, 0\}$$

حال نتیجه توزیع مجانبی برای آماره EL_n تحت فرضیه $H_0^{x,y}$ را اثبات می‌کنیم؛ یعنی:

$$EL_n \xrightarrow{d} \int_{[0,1]^2} \frac{[B(u,v) - uB(\cdot,v) - vG(u,\cdot)]^2}{u(1-u)v(1-v)} dx dv$$

فرض کنید $\frac{1}{4} < \epsilon < 1$ و x_ϵ و $x_{1-\epsilon}$ را تعریف کنید به طوری که:

$$F(x_\epsilon) = \epsilon$$

$$F(x_{1-\epsilon}) = 1 - \epsilon$$

به طور مشابه y_ϵ و $y_{1-\epsilon}$ را تعریف کنید به طوری که:

$$G(y_\epsilon) = \epsilon$$

$$G(y_{1-\epsilon}) = 1 - \epsilon$$

و قرار دهید $EL_n = EL_{n_1} + EL_{n_2}$ به طوری که:

$$EL_{n_1} = \int_{y_\epsilon}^{y_{1-\epsilon}} \int_{x_\epsilon}^{x_{1-\epsilon}} -2 Ln R_n(x,y) dF_n(x) dG_n(y)$$

و نشان می‌دهیم که:

$$EL_{n_1} \xrightarrow{d} \int_{y_\epsilon}^{y_{1-\epsilon}} \int_{x_\epsilon}^{x_{1-\epsilon}} \frac{[B(u,v) - uB(\cdot,v) - vB(u,\cdot)]^2}{u(1-u)v(1-v)} du dv \quad (3.3)$$

و $EL_{n_2} = O_P(\sqrt{\epsilon})$ به طور یکنواخت در ϵ است. تعریف کنید:

$$A_\epsilon = \{(x,y); \epsilon \leq F(x) \leq 1 - \epsilon \text{ و } \epsilon \leq G(y) \leq 1 - \epsilon\}$$

به طوری که

$$EL_{n_1} = \int_{A_\epsilon} -2 Ln R_n(x,y) dF_n(x) dG_n(y)$$

ما ابتدا تقریب $Ln R_n(x,y)$ را به طور یکنواخت روی A_ϵ بررسی می‌کنیم زیرا

$$|Ln(1+a) - a + a^2/2| \leq |a|^3/3$$

برای هر $|a| \leq 1/2$ داریم:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{(x,y) \in A_\epsilon} \left| Ln R_n(x,y) + \frac{n}{2} \sum_{1 \leq r,s \leq 2} \frac{\{\hat{\Phi}_{rs}^{(o)}(x,y) - \hat{\Phi}_{rs}(x,y)\}^2}{\hat{\Phi}_{rs}(x,y)} I\{\hat{\Phi}_{11}(x,y) \geq \hat{\Phi}_{11}^{n(o)}(x,y)\} \right| \\ & \leq \lim_{n \rightarrow \infty} \sup_{(x,y) \in A_\epsilon} \frac{n}{3} \sum_{1 \leq r,s \leq 2} \frac{|\hat{\Phi}_{rs}(x,y) - \hat{\Phi}_{rs}(x,y)|^3}{\hat{\Phi}_{rs}^2(x,y)} I\{\hat{\Phi}_{11}(x,y) \geq \hat{\Phi}_{11}^{(o)}(x,y)\} = 0 \end{aligned}$$

حال EL_{n_1} را حساب می‌کنیم؛ زیرا

$$\{\hat{\Phi}_{rs}^{(o)}(x,y) - \hat{\Phi}_{rs}(x,y)\}^2 = \{F_n(x)G_n(y) - H_n(x,y)\}^2$$

برای همه $(x, y) \in R^2$ و $1 \leq r$ و $s \leq 2$ داریم:

$$\begin{aligned} EL_{n_1} &= \int_{A_\epsilon} -2 \text{Ln} R_n(x, y) dF_n(x) dG_n(y) \\ &= \int_{A_\epsilon} n \{ \hat{\Phi}_{11}(x, y) - \hat{\Phi}_{11}^{(\circ)}(x, y) \}^2 \left| \sum_{1 \leq r, s \leq 2} \frac{I\{ \hat{\Phi}_{11}(x, y) \geq \hat{\Phi}_{11}^{(\circ)}(x, y) \}}{\hat{\Phi}_{rs}(x, y)} \right| dF_n(x) dG_n(y) + O(1) \\ &= \int_{A_\epsilon} n \{ H_n(x, y) - F_n(x)G_n(y) \}^2 \left\{ \sum_{1 \leq r, s \leq 2} \frac{1}{\Phi_{rs}(x, y)} \right\} dF_n(x) dG_n(y) + O_P(1) \end{aligned}$$

تحت فرضیه $H^{x,y}$ و توجه داشته باشید که:

$$\sum_{1 \leq r, s \leq 2} \frac{1}{\Phi_{rs}(x, y)} = \frac{1}{F(x)\{1 - F(x)\}G(y)\{1 - G(y)\}}$$

بنابراین به بررسی رفتار مجانبی $n \{ H_n(x, y) - F_n(x)G_n(y) \}^2$ می پردازیم و تعریف کنید:

$$\begin{aligned} B_n(x, y) &= \sqrt{n} \{ H_n(x, y) - H(x, y) \} \\ B_{n_1}(x) &= \sqrt{n} \{ F_n(x) - F(x) \} \\ B_{n_2}(y) &= \sqrt{n} \{ G_n(y) - G(y) \} \end{aligned}$$

نتیجه می گیریم که $B_n(x, y)$ نسبت به $B(F(x), G(y))$ در R^2 و از آنجا که $B_n(x)$ و $B_{n_2}(y)$ در R به ترتیب $B(F(x), 1)$ و $B(1, G(y))$ است؛ بنابراین تحت فرضیه $H^{x,y}$ داریم:

$$\begin{aligned} n \{ H_n(x, y) - F_n(x)G_n(y) \}^2 &= [\{ B_n(x, y) - F(x)B_{n_2}(y) - G(y)B_{n_1}(x) \}^2] \\ &\xrightarrow{w} [\{ B(F(x), G(y)) - F(x)B(1, G(y)) - G(y)B(F(x), 1) \}^2] \\ &\xrightarrow{w} \text{به معنای همگرایی ضعیف است.} \end{aligned}$$

اکنون قضیه چندمتغیره هلی (Helly-Bray) را استفاده کردیم تا نتیجه زیر را بگیریم:

$$\begin{aligned} EL_{n_1} &\xrightarrow{d} \int_{A_\epsilon} \frac{[\{ B(F(x), G(y)) - F(x)B(1, G(y)) - G(y)B(F(x), 1) \}^2]}{F(x)\{1 - F(x)\}G(y)\{1 - G(y)\}} dF(x) dG(y) \\ &= \int_\epsilon^{1-\epsilon} \int_\epsilon^{1-\epsilon} \frac{[\{ B(u, v) - uB(1, v) - vB(u, 1) \}^2]}{u(1-u)v(1-v)} du dv \end{aligned}$$

که همان معادله (۳.۳) را نشان می دهد. برای نشان دادن $EL_{n_2} = O_P(\sqrt{\epsilon})$ توجه داشته باشید که $-2 \text{Ln} R_n(x, y) \leq -2 \text{Ln} R_{EM}(x, y) \leq 0$ به طوری که $R_{EM}(x, y)$ آماره آزمون محلی که تعریف شده است برای $H(x, y) = F(x)G(y)$ در مقابل $H(x, y) \neq F(x)G(y)$ است، زیرا

$$\int_{A_\epsilon^c} -2 \text{Ln} R_{EM}(x, y) dF_n(x) dG_n(y) = O_P(\sqrt{\epsilon})$$

پس اثبات کامل می شود.

۲.۲.۳ آزمون استقلال

در اینجا آزمایشی را برای استقلال دو متغیر تصادفی در نظر می‌گیریم. فرض کنید X_1, \dots, X_n بردارهای تصادفی (*iid*) با تابع توزیع توأم H و توابع حاشیه‌ای F و G باشد. می‌خواهیم فرضیه صفر استقلال را آزمایش کنیم.

$$H_0 : H(x, y) = F(x)G(y) \quad x, y \in R$$

نسبت درست‌نمایی آماره آزمون به صورت زیر تعریف می‌شود:

$$R_{EM}(x, y) = \frac{\sup\{L(\tilde{H}) : \tilde{H}(x, y) = \tilde{F}(x)\tilde{G}(y)\}}{\sup\{L(\tilde{H})\}}$$

برای $(x, y) \in R$ با $L(\tilde{H}) = \prod_{i=1}^n \tilde{P}(\{X_i\})$ به طوری که \tilde{P} اندازه احتمال مربوط به \tilde{H} است پس

$$\begin{aligned} \log R_{EM}(x, y) &= nP_n(A_{11}) \log \frac{F_n(x)G_n(y)}{P(A_{11})} \\ &+ nP_n(A_{12}) \log \frac{F_n(x)(1 - G_n(y))}{P_n(A_{12})} + nP_n(A_{21}) \log \frac{(1 - F_n(x))G_n(y)}{P_n(A_{21})} \\ &+ nP_n(A_{22}) \log \frac{(1 - F_n(x))(1 - G_n(y))}{P_n(A_{22})} \end{aligned}$$

که در آن P_n اندازه تجربی است و F_n و G_n توابع توزیع حاشیه‌ای هستند و $A_{11}, A_{12}, A_{21}, A_{22}$ طبق قبل تعریف می‌شوند و آمار آزمون را به صورت زیر در نظر بگیرید:

$$T_n = -2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \log R(x, y) dF_n(x) dG_n(y)$$

واضح است T_n بدون توزیع است. حد توزیع T_n در زیر بیان می‌شود:

نتیجه ۱.۲.۳. F و G توابع توزیع حاشیه‌ای تحت فرض H_0 هستند.

$$T_n \xrightarrow{D} \int_0^1 \int_0^1 \frac{w_0^2(u, v)}{u(1-u)v(1-v)} du dv$$

که w_0 یک فرآیند وینر^۲ چهارطرفه روی $[0, 1]^2$ است.

فرآیند وینر یک فرآیند تصادفی پیوسته در ریاضیات است که به افتخار نوربرت وینر نامگذاری شده است. این فرآیند به اسم حرکت براونی استاندارد هم شناخته می‌شود. فرآیند w_t ، فرآیند وینر خوانده می‌شود در صورتی که سه خصوصیت زیر را داشته باشد:

$$w_0 = 0 \quad ۱.$$

^۲Wiener

۲. تماماً پیوسته باشد. w_t

۳. w_t فرآیندی با نمو مانا و مستقل باشد و نمو آن از توزیع نرمال و به صورت زیر است:

$$w_t - w_s \sim N(0, t - s) \quad 0 \leq t < s$$

۳.۲.۳ آزمون فرضیه مبتنی بر احتمال تجربی

این روش برای به دست آوردن مناطق اطمینان غیرپارامتری و نیمپارامتری عمدتاً نامطلوب است.

فرض کنید X_1, \dots, X_n نمونه تصادفی و مستقل (*iid*) با تابع توزیع F و فرضیه $H_0 = F = F_0$ به طوری که F_0 یک تابع توزیع مشخص و نسبت احتمال درستنمایی تجربی به صورت زیر است:

$$R(x) = \frac{\sup\{L(\tilde{F}) : \tilde{F}(x) = F_0(x)\}}{\sup\{L(\tilde{F})\}}$$

به طوری که

$$L(\tilde{F}) = \prod_{i=1}^n (\tilde{F}(X_i) - \tilde{F}(X_i - 1))$$

تابع توزیع تجربی F_n به وجه برتری در مخرج می‌رسد و برتری در تعداد $F_0(x)/(nF_n(x))$ بر روی هر مشاهده فراتر از X است و

$$(1 - F_0(x)) / (n(1 - F_n(x)))$$

پس به راحتی:

$$\log R(x) = nF_n(x) \log \frac{F(x)}{F_n(x)} + n(1 - F_n(x)) \log \frac{1 - F_0(x)}{1 - F_n(x)} \quad 0 < F_0(x) < 1$$

$$-2 \log R(x) = \frac{n(F_n(x) - F_0(x))^2}{F_n(x)(1 - F_0(x))} + OP(1) \xrightarrow{D} \chi_1^2 \quad (4.3)$$

تحت فرضیه H_0 برقرار است. ما باید $-2 \log R(x)$ به طور همزمان در محدوده‌ای از مقادیر x بررسی کنیم گرفتن انتگرال با توجه به F_0 منجر به آماره زیر می‌شود:

$$T_n = -2 \int_{-\infty}^{+\infty} \log R(x) dF_0(x)$$

اگر به جای ادغام در T_n ، سوپریمم همه X ها را به دست آوریم، آماره‌ای را به دست می‌آوریم که از هر آماره وزنی کلموگروف اسمیرنوف کارآمدتر است. در این مورد برای

پارامتر Θ ، $R(x) = R_{\Theta}(x)$ است. واضح است که T_n توزیع آزاد است. علاوه بر این طبق (۱.۳) و استفاده از نظریه فرآیند تجربی می‌توان نشان داد که:

$$T_n \xrightarrow{D} \int_0^1 \frac{B^{\vee}(t)}{t(1-t)} dt$$

تحت فرضیه H_0 به طوری که B یک پل استاندارد براونی است و T_n به صورت مجانبی معادل اندرسون-دارلینگ است:

$$A_n^{\vee} = n \int_{-\infty}^{+\infty} \frac{(F_n(x) - F_0(x))^{\vee}}{F_0(x)(1 - F_0(x))} dF_0(x)$$

ما آماره T_n را برای انواع فرضیه‌های غیرپارامتری فراتر از فرضیه صفر ساده بررسی می‌کنیم. آزمایش تقارن در حدود صفر را می‌توان با استفاده از

$$F(-x) = 1 - F(x-)$$

و محلی‌سازی در X انجام داد.

۴.۲.۳ آزمایش تقارن

درباره آزمایش تقارن در یک توزیع متقارن شناخته شده یا ناشناخته مطالب زیادی نوشته شده است که برخی از آن‌ها عبارتست از: احمد ولی (۱۹۹۷)، نیکیتین (۱۹۹۶)، باتلر (۱۹۶۰) و هیل درائو (۱۹۷۷) و ...

فرض می‌کنیم که نقطه تقارن مشخص است. و X_1, \dots, X_n نمونه‌ای تصادفی و مستقل (*iid*) با تابع توزیع توأم F باشد. فرضیه صفر متقارن برابر است با:

$$H_0 = F(-x) = 1 - F(x-) \quad x > 0$$

آماره نسبت احتمال محلی به صورت زیر تعریف می‌شود:

$$R(x) = \frac{\sup\{L(\tilde{F}) : \tilde{F}(-x) = 1 - \tilde{F}(x-)\}}{\sup\{L(\tilde{F})\}} \quad x > 0.$$

احتمال درست‌نمایی در مخرج با تنظیم تابع توزیع تجربی $\tilde{F} = F_n$ به حداکثر می‌رسد (ماکسیمم می‌شود).

سوپریمم را می‌توان با استفاده از \tilde{F} به عنوان تابع یافت که $0 \leq P \leq 1$ به طوری که \tilde{F} جرم $P/2$ را در فاصله $(-\infty, -x]$ و جرم $P/2$ در فاصله $[x, \infty)$ و جرم $1 - P$ در فاصله $(-x, x)$ است که این جرم‌ها در فواصل مربوط، به طور مساوی بین مشاهدات تقسیم می‌شوند. یعنی جرم‌های مشاهده‌های منفرد در فواصل مربوط به صورت زیر است:

$$\frac{P/2}{n\hat{P}_1}, \frac{P/2}{n\hat{P}_2}, \frac{1-P}{n(1-\hat{P})}$$

به طوری که

$$\begin{aligned}\hat{P} &= \hat{P}_1 + \hat{P}_2 \\ \hat{P}_1 &= F_n(-x) \\ \hat{P}_2 &= 1 - F_n(x-)\end{aligned}$$

بنابراین $R(x)$ ماکسیمم مقدار

$$\left(\frac{P/2}{n\hat{P}_1}\right)^{n\hat{P}_1} \left(\frac{P/2}{n\hat{P}_2}\right)^{n\hat{P}_2} \left(\frac{1-P}{n(1-\hat{P})}\right)^{n(1-\hat{P})}$$

است که به راحتی $P = \hat{P}$ به دست می آید. بنابراین:

$$\begin{aligned}\log R(x) &= n\hat{P}_1 \log \frac{\hat{P}}{2\hat{P}_1} + n\hat{P}_2 \log \frac{\hat{P}}{2\hat{P}_2} = nF_n(-x) \log \frac{F_n(-x) + 1 - F_n(x-)}{2F_n(-x)} \\ &+ n(1 - F_n(x-)) \log \frac{F_n(-x) + 1 - F_n(x-)}{2(1 - F_n(x-))}\end{aligned}$$

به طوری که $\log(a|^\circ) = \circ$. آماره آزمون به صورت زیر به دست می آید:

$$\begin{aligned}T_n &= -2 \int_0^\infty \log R(x) d\{F_n(x) - F_n(-x)\} \\ &= -2 \int_0^\infty \log R(x) dG_n(x)\end{aligned}$$

به طوری که G_n تابع توزیع تجربی $|X_i|$ است.

متناوباً ممکن است به صورت زیر بنویسیم:

$$T_n = -\frac{2}{n} \sum_{i=1}^n \log R(|X_i|)$$

واضح است که T_n بدون توزیع است. حد توزیع T_n با نتیجه زیر داده می شود.

نتیجه ۲.۲.۳. F تابع توزیع توأم است، پس تحت فرضیه H_0 داریم:

$$T_n \xrightarrow{D} \int_0^1 \frac{w^2(t)}{t} dt$$

که W فرآیند وینر استاندارد است.

۳.۳ استنباط براساس تابع توزیع

ما از PQD و $PLOD$ که در گذشته تعریف شده بود استفاده می کنیم و فقط تعداد مشخصی از نقاط مجزا را در نظر می گیریم مثلاً نقاط d . از تعریف $PLOD$ و توابع توزیع تجمعی و نقاط d و $y_i = (y_{i1}, \dots, y_{in})' \in \mathbb{R}$ استفاده می کنیم و

$$D_F^i = F(y_i) - \prod_{j=1}^n F_j(y_{ij}) \quad \text{و} \quad D_F = (D_F^1, \dots, D_F^d)'$$

را تعریف می‌کنیم. فرضیه صفر آزمون برای $PLOD$ به صورت زیر نوشته می‌شود:

$$H_0 : \{D_F : D_F \geq 0\}$$

و فرضیه مخالف به صورت زیر است:

$$H_1 : \{D_F : D_F \text{ بدون محدودیت است}\}$$

همچنین ممکن است آزمایشی را برای غیر $PLOD$ براساس فرضیه صفر

$$\bar{H}_0 : \{D_F : D_F^i \leq 0, \text{ها } i\}$$

و فرضیه مخالف

$$\bar{H}_1 : \{D_F : D_F^i > 0, \text{ها } i\}$$

در نظر بگیریم. توزیع تجربی و توزیع توأم به صورت زیر است:

$$\hat{F}(y_i) = \frac{1}{T} \sum_{t=1}^T \prod_{j=1}^n I[Y_{jt} \leq y_{ij}], \quad i = 1, \dots, d$$

$$\hat{F}_j(y_{ij}) = \frac{1}{T} \sum_{t=1}^T I[Y_{jt} \leq y_{ij}], \quad i = 1, \dots, d, \quad j = 1, \dots, n$$

فرض کنید $y_{k \wedge L} = (y_{k1} \wedge y_{L1}, \dots, y_{kn} \wedge y_{Ln})'$ به طوری که $a \wedge b = \min(a, b)$ و گزاره زیر توزیع جانبی \hat{D}_F را ارائه می‌دهد.

گزاره ۱.۳.۳. بردارهای تصادفی $\sqrt{T}(\hat{D}_F - D_F)$ در توزیع به یک متغیر تصادفی نرمال d بعدی با میانگین صفر و کواریانس v_F که عناصر اصلی آن است، همگرا می‌شود.

$$v_{F,KL} = F(y_{K \wedge L}) - F(y_K)F(y_L), \quad K, L = 1, \dots, d$$

تخمین ثابت \hat{v}_F از v_F را می‌توان با جایگزینی توزیع ناشناخته F توسط همتای تجربی آن یعنی \hat{F} به دست آورد.

اثبات: از ویژگی‌های استاندارد و توزیع تجربی، اولین تفاوت

$$\hat{D}_F^i = \hat{F}(y_i) - \prod_{j=1}^n \hat{F}_j(y_{ij})$$

از نظر $T^{-1/2}$ است در حالی که دومین تفاوت شامل یک محصول منظم $T^{-n/2}$ است. این بدان معنی است که تنها اولین واژه در توزیع جانبی نقش دارد. نتیجه اعلام شده زیر از قضیه حد مرکزی و یک محاسبه ساده کواریانس جانبی است:

$$\lim_{T \rightarrow \infty} T \text{ cov} \left(\hat{F}(y_K), \hat{F}(y_L) \right) = F(y_{K_1} \wedge y_{L_1}, \dots, y_{K_n} \wedge y_{L_n}) - F(y_K)F(y_L).$$

قضیه حد مرکزی

قضیه حد مرکزی مرتبط با توزیع مجموع متغیرهای تصادفی مستقل و هم‌توزیع و واریانس متناهی است به این ترتیب که اگر بدانیم X_1, \dots, X_n متغیرهای مستقل و هم‌توزیع هستند. می‌توانیم توزیع احتمال مجموع آن‌ها را زمانی که n به اندازه کافی بزرگ باشد، نرمال در نظر بگیریم. نکته جالب در این قضیه عدم اطلاع از توزیع متغیر تصادفی است. امید ریاضی توزیع X ها برابر با μ و واریانس آن σ^2 است. در این صورت میانگین نمونه‌ای S_n که به صورت زیر معرفی می‌شود به طور مجانبی (با افزایش n) دارای توزیع نرمال است که به آن میانگین نمونه‌ای می‌گویند:

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$$

می‌دانیم میانگین نمونه‌ای به میانگین توزیع میل خواهد کرد، یعنی

$$\lim_{n \rightarrow \infty} \bar{X}_n = \mu \quad a.s.$$

قضیه حد مرکزی در مورد توزیع \bar{X}_n نیز توضیح می‌دهد:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

رابطه بالا می‌تواند به شکل تقریبی و مجانبی تابع احتمال میانگین توزیع یعنی \bar{X}_n را مشخص کند:

$$\bar{X}_n \xrightarrow{d} N\left(\mu, \frac{\sigma^2}{n}\right)$$

۴.۳ استنباط بر اساس مفصل

از مقادیر مشابه نقاط d و

$$u_i = (u_{i1}, \dots, u_{in})', \quad u_{ij} \in (0, 1) \quad i = 1, \dots, d, \quad j = 1, \dots, n$$

استفاده می‌کنیم و همچنین تعریف می‌کنیم:

$$D_C^i = C(u_i) - \prod_{j=1}^n u_{ij}$$

و

$$D_C = (D_C^1, \dots, D_C^d)'$$

فرضیه صفر را به صورت زیر در نظر می‌گیریم:

$$H_0 : \{D_C : D_C \geq 0\}$$

و فرضیه مخالف به صورت زیر است:

$$H_1 : \{D_C : D_C \text{ محدودیت است}\}$$

در حالی که آزمون، $PLOD$ نباشد براساس فرضیه صفر زیر است:

$$\bar{H}_0 : \{D_C : D_C^i \leq 0, \text{ برای برخی } i\text{ها}\}$$

و فرضیه مخالف:

$$\bar{H}_1 : \{D_C : D_C^i > 0, \text{ برای همه } i\text{ها}\}$$

ما از این پس فرض می‌کنیم که همه توابع توزیع پیوسته هستند و $F_j(y) = u_{ij}$ یک راه حل منحصر به فرد را نشان می‌دهد که ζ_{ij} ، $i = 1, \dots, d$ و $j = 1, \dots, n$ در حالی که $f_j(\zeta_{ij}) > 0$ در هر ζ_{ij} کمی است.

ممکن است به فکر برآورد $C(u_i) = F(\zeta_i)$ با $\hat{C}(u_i) = \hat{F}(\hat{\zeta}_i)$ باشیم به طوری که $\hat{\zeta}_i = (\hat{\zeta}_{i1}, \dots, \hat{\zeta}_{in})'$ از تعداد متغیر تجربی یکسان $\hat{\zeta}_{ij}$ ساخته شده است.

فرض کنید

$$\zeta_{k \wedge j^L} = (\zeta_{k1}, \dots, \zeta_{kj} \wedge \zeta_{Lj}, \dots, \zeta_{kn})'$$

و

$$u_{k \wedge L, j} = (u_{kj} \wedge u_{Lj})$$

و

$$F_{j_1 j_2}(\zeta_{k \wedge L}) = P[Y_{j_1} \leq (\zeta_{kj_1} \wedge \zeta_{Lj_1})$$

$$Y_{j_2} \leq (\zeta_{kj_2} \wedge \zeta_{Lj_2})], \quad j_1, j_2 = 1, \dots, n, \quad j_1 \neq j_2$$

گزاره زیر توابع توزیع مجانبی \hat{D}_C را ارائه می‌دهد.

گزاره ۱.۴.۳. بردارهای تصادفی $\sqrt{T}(\hat{D}_C - D_C)$ در توزیع به یک متغیر تصادفی نرمال d بعدی با ماتریس میانگین صفر و کوارینانس V_C که عناصر اصلی آن نرمال است، همگرا می‌شود:

$$v_{C,kl} = b'_k A_{kl} b_l, \quad k, l = 1, \dots, d$$

به طوری که

$$b_i = \left(\begin{array}{c} -\frac{\partial F(\zeta_i)}{\partial x_1} \quad \dots \quad -\frac{\partial F(\zeta_i)}{\partial x_n} \\ 1 \quad \frac{f_1(\zeta_{i1})}{f_1(\zeta_{i1})} \quad \dots \quad \frac{f_n(\zeta_{in})}{f_n(\zeta_{in})} \end{array} \right)' \quad i = 1, \dots, d$$

$$A_{kl} = \begin{pmatrix} F(\zeta_{k \wedge l}) - F(\zeta_k)F(\zeta_l) & F(\zeta_{k \wedge l}) - F(\zeta_k u_{l1}) & \cdots & F(\zeta_{k \wedge n l}) - F(\zeta_k u_{ln}) \\ F(\zeta_{l \wedge k}) - F(\zeta_l)u_{k1} & u_{k \wedge l 1} - u_{k1}u_{l1} & \cdots & F_{\setminus n}(\zeta_{k \wedge l}) - u_{kl}u_{ln} \\ \cdots & \cdots & \vdots & \cdots \\ F(\zeta_{l \wedge m k}) - F(\zeta_l)u_{kn} & F_{\setminus n}(\zeta_{k \wedge l}) - u_{l1}u_{kn} & \cdots & u_{k \wedge l 1 n} - u_{kn}u_{ln} \end{pmatrix}$$

ماتریس کواریانس مجانبی V_C از F مشتق گرفته شده و مقادیر تک متغیره f_j را می توان با روش های استاندارد کرنل به منظور برآورد ثابت \hat{V}_c از V_C تخمین زد (اسکات (۱۹۹۲)). به عنوان مثال ممکن است a کرنل گوسی و مقادیر متفاوت h_j در هر بعد را نشان دهد که منجر به:

$$\frac{\partial \hat{F}(\hat{\zeta}_i)}{\partial x_i} = (Th_j)^{-1} \sum_{t=1}^T \varphi \left(\frac{Y_{jt} - \hat{\zeta}_{ij}}{h_j} \right) \prod_{l \neq j} \Phi \left(\frac{Y_{lt} - \hat{\zeta}_{il}}{h_l} \right)$$

$$\hat{f}_j(\hat{\zeta}_{ij}) = (Th_j)^{-1} \sum_{t=1}^T \varphi \left(\frac{Y_{jt} - \hat{\zeta}_{ij}}{h_j} \right)$$

شود به طوری که φ و Φ ، تابع چگالی و تابع توزیع یک متغیر استاندارد گوسی است.

اثبات: فرض کنید

$$\{I[0 \leq x_1] \cdots I[0 \leq x_n] : x_j \in R, j = 1, \dots, n\}$$

از آنجا که M شرایط آنتروپی پولار^۳ برخی از ثابت ها در نظر می گیرد: دنباله

$$\{\hat{F}(x) = T^{-1} \sum_{t=1}^T \prod_{j=1}^n I[Y_{jt} \leq x_j] : T \geq 1\}$$

که از نظر تصادفی ζ_i با بردار $(d \times 1)$ مشتق تصادفی $D\hat{F}(\zeta_i)$ قابل تغییر است. پس تقریب زیر را داریم:

$$\hat{F}(\hat{\zeta}_i) = \hat{F}(\zeta_i) + D\hat{F}(\zeta_i)'(\hat{\zeta}_i - \zeta_i) + O_P(T^{-1/2})$$

به طوری که $\bar{\zeta}_i$ میانگین مقادیر بین $(\hat{\zeta}_i)$ و ζ_i است. به طور مشابه تقریبها را به دست می آوریم:

$$\hat{F}(\hat{\zeta}_i) = \hat{F}(\zeta_i) - D\hat{F}(\hat{\zeta}_i)' \text{diag} S_i + O_P(T^{-1/2})$$

$$(\hat{F}(\zeta_{ij}) - u_{ij}) / D\hat{F}_j(\hat{\zeta}_{ij}), \quad j = 1, \dots, n$$

³Entropy Pollard's

$diag(S_i)$ عناصر روی قطر اصلی ماتریس را نشان می‌دهد. با استفاده از همگرایی احتمال $D_j \hat{F}(\bar{\zeta}_i)$ به $\partial F(\zeta_j)/\partial x_j$ و $j = 1, \dots, n$ و $D_j \hat{F}_j(\bar{\zeta}_{ij})$ به $f_j(\zeta_{ij})$ و نتیجه بیان شده از قضیه حد مرکزی و محاسبه کواریانس داریم:

$$\lim_{T \rightarrow \infty} T \text{cov}[\hat{F}(\zeta_k), \hat{F}_j(\zeta_{ij})] = F(\zeta_{k1}, \dots, \zeta_{kj} \wedge \zeta_{lj}, \dots, \zeta_{kn}) - F(\zeta_k)F_j(\zeta_{lj})$$

$$\lim_{T \rightarrow \infty} T \text{cov}[\hat{F}_j(\zeta_{kj}), \hat{F}_j(\zeta_{ij})] = F_j(\zeta_{ij} \wedge \zeta_{lj}) - F_j(\zeta_{kj})F_j(\zeta_{ij})$$

$$\lim_{T \rightarrow \infty} T \text{cov}[\hat{F}_{j_1}(\zeta_{kj_1}), \hat{F}_{j_2}(\zeta_{lj_2})] = F_{j_1 j_2}(\zeta_{kj_1} \wedge \zeta_{lj_1}, \zeta_{kj_2} \wedge \zeta_{lj_2}) - F_{j_1}(\zeta_{kj_1})F_{j_2}(\zeta_{lj_2}).$$

۵.۳ روش آزمایش

نتایج استنباط براساس تابع توزیع و مفصل عناصر سازنده روش‌های آزمایش هستند. اولین روش آزمون H_0^F (مربوط به H_0^C) در مقابل H_1^F (مربوط به H_1^C) از آزمون‌های فاصله استفاده می‌کند. منفی بودن یک یا چند از \hat{D}_K مهم است (در چنین حالتی شخص می‌خواهد بداند که آیا $PLOD$ برقرار است یا نه) فرض کنید $K = F, C$ و \tilde{D}_K است. حل مسئله با محدود کردن معادله درجه دوم به صورت زیر است:

$$\inf_D T(D - \hat{D}_K)' \hat{V}_K^{-1} (D - \hat{D}_K), \quad D \geq 0$$

که در آن \hat{V}_K یک تخمین ثابت از V_K است و قرار دهید:

$$\hat{\epsilon}_K = T(\tilde{D}_K - \hat{D}_K)' \hat{V}_K^{-1} (\tilde{D}_K - \hat{D}_K)$$

به‌طور تقریبی \tilde{D}_K نزدیک‌ترین نقطه به \hat{D}_K در زیر صفر و فاصله اندازه‌گیری شده در \hat{V}_K متغیر است و آماره آزمون $\hat{\epsilon}_K$ فاصله بین \tilde{D}_K و \hat{D}_K است. وقتی این فاصله بیش از حد بزرگ باشد H_0^K رد می‌شود. توزیع مجانبی $\hat{\epsilon}_K$ تحت فرضیه صفر برای هر x مثبت داریم:

$$P[\hat{\epsilon}_K \geq x] = \sum_{i=1}^d P[\chi_i^2 \geq x] w^K(d, d-i, \hat{V}_K)$$

که در آن وزن $w(d, d-i, \hat{V}_K)$ احتمال \hat{D}_K دقیقاً $d-i$ عناصر مثبت است. محاسبه راه‌حل \hat{D}_K می‌تواند توسط یک روال بهینه‌سازی عددی برای مشکلات محدود برنامه‌نویسی درجه دوم موجود در اکثر نرم‌افزارهای آماری انجام شود. در واقع کافی است تعداد قابل توجهی از توزیع نرمال چندمتغیره با میانگین صفر و ماتریس کواریانس \hat{V}_K را در نظر بگیریم. پس از این توزیع‌ها به عنوان \hat{D}_K در فرآیند آزمون استفاده می‌شود. \tilde{D}_K و تعداد عناصر بردار بزرگ‌تر از صفر را محاسبه کنید. نسبت رسم‌هایی که \hat{D}_K دقیقاً دارای عناصر $d-i$ بزرگ‌تر از صفر باشد، تخمین مونت کارلو از $w(d, d-i, \hat{V}_K)$ را می‌دهد.

اکنون توجه خود را به روش آزمایش دوم معطوف کنیم که هدف آن آزمایش \bar{H}^F مربوط به H_0^C در مقابل \bar{H}_1^F مربوط به H_1^C با تکیه بر اصل تقاطع است. با استفاده از مثبت بودن تمام اجزای \hat{D}_K ، سؤال این است که آیا اطمینان از $PLOD$ وجود دارد؟!

فرض کنید $\hat{\gamma}_K^i = \sqrt{T} \hat{D}_K^i / \sqrt{\hat{V}_{k,ii}}$ و $K = F, C$ تحت فرضیه \bar{H}_K^0 حد $P[\inf \hat{\gamma}_K^i > Z_{1-\alpha}]$ کمتر یا برابر با α باشد و اگر $D_K^i = 0$ برای هر i داده شده و $D_K^L > 0$ برای $L \neq i$ در حالی که حد آن تحت \bar{H}_K^1 است. در این صورت آزمون رد می‌شود موقعی که $\inf \hat{\gamma}_K^i$ بالاتر از $(1 - \alpha)$ در $Z_{1-\alpha}$ از یک توزیع نرمال استاندارد دارای حد بالای α روی اندازه مجانبی است.

۶.۳ آزمون‌های مفصل مبتنی بر مسافت

در آزمون‌های مناسب برای PQD نظرات اساسی و متفاوتی وجود دارد یعنی آزمون کنیم که X و Y ، PQD هست در مقابل اینکه X و Y مکمل که PQD نیستند. در واقع یک آزمایش ثابت شده برای H_0 در مقابل $H_1 - H_2$ است که در آن هیچ محدودیتی برای وابستگی به توابع توزیع مشترک H و توابع توزیع حاشیه‌ای F و G وجود ندارد. نتایج به دست آمده نشان می‌دهد که فرضیه H_1 رد نمی‌شود. البته رد نشدن H_1 به فرد اجازه نمی‌دهد که استقلال X و Y را نتیجه بگیرد که به‌طور قطعی PQD باشد. اکثر روش‌های غیرپارامتری آزمون H_1 را در مقابل $H_1 - H_2$ ، چارچوب مسئله را از نظر تابع مفصل مورد بررسی قرار می‌دهد.

از قضیه‌ی اسکالر مفصل C مثل $C(F(x), G(y)) = H(x, y)$ است، به‌طوری که برای همه‌ی $(x, y) \in R$ وقتی که F و G حاشیه‌ای هستند، یعنی C یک توزیع مشترک روی $[0, 1]^2$ متناسب با $\Pi(u, v) = uv$ به عنوان یک مفصل مستقل است. آزمون H_1 در مقابل $H_1 - H_2$ معادل آزمون $C(u, v) \geq uv$ برای همه‌ی $(u, v) \in [0, 1]^2$ در مقابل $C(u, v) < uv$ برای تعدادی از $(u, v) \in [0, 1]^2$ است. مشاهدات به صورت زیر تعریف می‌شود:

$$V_i = nG_n(y_i)/(n+1) \quad i = 1, \dots, n$$

$$U_i = nF_n(x_i)/(n+1) \quad i = 1, \dots, n$$

مقیاس‌گذاری توابع توزیع تجربی $n/(n+1)$ با مشکلات احتمالی تخمین در نزدیکی مرزها $[0, 1]^2$ را کاهش می‌دهد موقعی که H_0 درست است مقادیر (u_i, v_i) باید تقریباً به‌طور یکنواخت در مربع واحد توزیع می‌شوند که یک برآوردگر طبیعی از C است:

$$C_n(u, v) = n^{-1} \sum_{i=1}^n I(U_i \leq v, V_i \leq u)$$

و آزمون‌های مبتنی بر مسافت می‌تواند با مقایسه برآوردگر تجربی $C_n(u, v)$ مفصل مستقل، تحت یک زیرمسافت معین تشکیل شود.

$$C(u, v) = uv$$

KS_n ، $C_v M_n$ و AD_n به بررسی فرض نرمال بودن داده‌ها می‌پردازد که به صورت زیر محاسبه می‌شوند:

$$KS_n = \sqrt{n} \sup_{uv \in [0, 1]^2} \{C_n(u, v) - uv\}_+$$

$$C_v M_n = n \int_{[0, 1]^2} [\{C_n(u, v) - uv\}_+] dC_n(u, v)$$

$$AD_n = n \int_{[0, 1]^2} \frac{[\{C_n(u, v) - uv\}_+]^2}{u(1-u)v(1-v)} dC_n(u, v)$$

واضح است که مقادیر زیادی از هر آماره X و Y موجود است که به شدت PQD هستند. ما دریافتیم که استفاده از مقادیر حساس از توزیع‌های مجانبی منجر به آزمایش‌هایی می‌شود که به‌طور طبیعی ضدمحافظ کار هستند به خصوص با موارد کوچک نمونه‌های متوسط.

نرمال بودن داده‌ها، یکی از معمول‌ترین فرض‌های در نظر گرفته شده در استفاده از روش‌های آماری (آزمون‌های پارامتری) است. ساده‌ترین روش برای آزمون و بررسی فرض نرمال بودن داده‌ها، روش‌های نموداری است. آزمون‌های کلموگوروف، اسمیرنت، کرامر-ون-میزس، اندرسن-دارلینگ و ... از دیگر روش‌های بررسی فرض نرمال بودن داده‌هاست.

در انتخاب یک آزمون آماری برای تحقیق باید تصمیم بگیریم که آیا از آزمون‌های پارامتری یا ناپارامتری استفاده کنیم. یکی از اصلی‌ترین ملاک‌ها برای انتخاب، آزمون کلموگوروف اسمیرنت است. این آزمون نرمال بودن توزیع داده‌ها را نشان می‌دهد؛ یعنی توزیع یک صفت در یک نمونه را با توزیعی که برای جامعه مفروض است مقایسه می‌کند. اگر آزمون کلموگوروف-اسمیرنت رد شود، داده‌ها توزیع نرمال ندارند و امکان استفاده از آزمون‌های ناپارامتری برای تحقیق وجود ندارد. اما اگر آزمون کلموگوروف-اسمیرنت تایید شود یعنی داده‌ها دارای توزیع نرمال هستند پس باید از آزمون‌های ناپارامتری در تحقیق استفاده کنیم. در نرم‌افزار R اگر P کوچک‌تر از 0.05 بود یعنی آزمون معنی‌دار است و توزیع داده‌ها نرمال نیست و ما می‌توانیم از آزمون‌های ناپارامتری استفاده کنیم و بالعکس. چون تایید شدن این آزمون، نشانه پارامتری بودن داده‌هاست. این آزمون ناپارامتری و بدون توزیع است.

آزمون اندرسن دارلینگ و کرامر ون میزس نیز مانند این آزمون عمل می‌کنند.

فصل ۴

مطالعه شبیه‌سازی و مثال‌های کاربردی

در این فصل عملکرد آزمون‌های معرفی‌شده، یعنی آزمون درست‌نمایی تجربی و آزمون‌های مفصل مبتنی بر مسافت، را برای بررسی PQD بودن مشاهدات با مطالعه شبیه‌سازی و در مثال‌های کاربردی مقایسه می‌کنیم. افزون بر این، در مقایسه خود، آزمون‌های معرفی‌شده توسط کچار و گوپتا (۱۹۸۷)، که با KG نشان می‌دهیم، و یانیک و همکاران (۲۰۰۴)، که با JKL نمایش می‌دهیم، را نیز وارد خواهیم کرد. شایان یادآوری است که این آزمون‌ها ذاتاً ناپارامتری هستند. توجه داشته باشید که R موجود در سایت *Github*، <https://github.com/cftang9/pqd> کلیه شبیه‌سازی موجود در این پایان‌نامه را تولید می‌کند.

کچار و گوپتا (۱۹۸۷) یک رده از U -آماره‌ها، مثل U_{k+1} ، پیشنهاد دادند که برای برآورد

$$\Pi_{1k} = \int_{\mathbb{R}^2} H^k(x, y) dH(x, y) \quad \forall k > 0$$

استفاده می‌شوند. آن‌ها نشان دادند که آزمون فرضیه‌های H_0 در مقابل $H_1 - H_0$ ، به آزمون $\Pi_{1k} = (1+k)^{-2}$ تحت استقلال در مقابل $\Pi_{1k} > (1+k)^{-2}$ تحت PQD کامل تبدیل می‌شود. بنابراین مقادیر بزرگ U_{k+1} ، شهودی بر علیه H_0 است. آماره رتبه یک‌طرفه کندهال حالت خاصی از رده آزمون‌های KG به ازای $k = 1$ است. برای k های

بزرگ، محاسبه U-آماره آزمون KG برای n های بزرگ، مثل $n = 100$ یا بیشتر، نیاز به محاسبه $\binom{n}{k+1}$ بار توابع هسته دارد. در نتیجه واضح است که در این موارد، حجم محاسبات می‌تواند بیش از توان پردازنده‌های موجود باشد. به همین دلیل، در مطالعه شبیه‌سازی این فصل $k = 1$ تنظیم شده است.

رهیافت ارایه‌شده توسط یانیک و همکارانش (۲۰۰۴) بر این واقعیت تکیه دارد که X و Y ، PQD هستند، اگر و فقط اگر به ازای تمام توابع نازولی $s(\cdot)$ و $t(\cdot)$

$$\text{cov}\{s(F(x)), t(G(y))\} \geq 0.$$

یانیک و همکارانش (۲۰۰۴) برای نمایش پارامتری این مجموعه از توابع از یک دنباله از چندجمله‌ای‌های لژاندر^۱، $\{b_j^*\}_{j=1}^k$ ، استفاده کردند که برای مدل‌بندی تابع چگالی احتمال توام $(F(x), G(y))$ تحت PQD به کار می‌روند. با بهره گرفتن از این نوع بازپارامتری و برآورد F و G با توابع توزیع تجربی متناظرشان، رهیافت JKL به یک خانواده از آماره‌های رتبه منتهی می‌شوند. به‌عنوان مثال، برای $k = 1$ ، این آماره آزمون برای H_0 در مقابل $H_1 - H_0$ برابر است با

$$V(1, 1) = n^{-1/2} \sum_{i=1}^n b_1^* \{(R_i - \frac{1}{2})/n\} b_1^* \{(S_i - \frac{1}{2})/n\}$$

به‌طوری که

$$b_1^*(u) = \sqrt{3}(2u - 1)$$

و R_i ها و S_i ها به ترتیب رتبه‌های X_1, \dots, X_n و Y_1, \dots, Y_n هستند. جالب است بدانید رد کردن H_0 به ازای مقدار بزرگ $V(1, 1)$ معادل است با انجام آزمون رتبه یک‌طرفه اسپیرمن.

می‌توان آماره‌های آزمون را با استفاده از تعداد بزرگ‌تری از چندجمله‌ای‌ها (یعنی برای k بزرگ‌تر) فرمول‌بندی کرد، اما حجم محاسبات و سختی اجرای آزمون‌ها نیز به‌طور قابل ملاحظه‌ای افزایش خواهد یافت. به همین دلیل برای انجام آزمون JKL نیز توجه خود را به $k = 1$ معطوف می‌کنیم.

تمام آماره‌های آزمون که تاکنون تشریح شدند، مانند آماره‌های کندال و اسپیرمن، به توزیع‌های حاشیه‌ای F و G بستگی ندارند. بنابراین برای ارزیابی ویژگی‌های کوچک نمونه آزمون‌ها، از توابع مفصل برای شبیه‌سازی داده‌ها استفاده می‌کنیم. برای این کار از خانواده‌های مفصل معمول شامل مفصل‌های کلایتون^۲، فرنک^۳، گاوسی^۴، و گامبل^۵،

¹Legendre polynomials

²Clayton

³Frank

⁴Gaussian

⁵Gumbel

استفاده کردیم. این توابع توسط گیلز و همکارانش (۲۰۱۰) نیز مورد استفاده قرار گرفته‌اند. توابع مفصل کلایتون، فرنک و گامبل به رده معروف توابع مفصل ارشمیدسی تعلق دارند که به‌طور گسترده‌ای برای مدل‌بندی داده‌ها با وابستگی‌های دمی متفاوت در دم راست به کار گرفته می‌شوند. هر تابع مفصل در رده مفصل‌های ارشمیدسی با پارامتر τ کندال اندیس‌گذاری می‌شود که تحت فرض استقلال برابر صفر خواهد بود و برای حالتی که X و Y ، PQD هستند، بزرگ‌تر از صفر است. وضعیت مشابهی برای مفصل گاوسی برقرار است که با پارامتر همبستگی ρ اندیس‌گذاری می‌شود.

۱.۴ ارزیابی آزمون‌ها با شبیه‌سازی

در این بخش، عملکرد شش آزمون پیشنهادی، شامل آزمون‌های درست‌نمایی تجربی، (EL)، کولموگروف-اسمیرنوف، (KS)، کرامر ون میسرز، (CvM)، اندرسون-دارلینگ، (AD)، اسپیرمن و کندال، را برای آزمون استقلال در مقابل PQD بودن ارزیابی می‌کنیم. با توجه به این که توزیع‌های متناهی-نمونه^۶ آماره EL_n و آماره‌های مبتنی بر مسافت به توزیع‌های حاشیه‌ای F و G وابسته نیستند، بنابراین تکیه زدن بر نتایج جانبی بیان‌شده در فصل ۳ برای اجرای آزمون‌ها ضرورتی ندارد. در واقع تانگ و همکارانش (۲۰۱۹) دریافته‌اند که استفاده از مقادیر بحرانی محاسبه‌شده از روی توزیع‌های جانبی آماره‌های آزمون ذکرشده به آزمون‌هایی منتهی می‌شود که، به‌ویژه برای حجم‌های نمونه کوچک و متوسط، به شکلی ناروا ضد محافظه‌کار هستند. به همین دلیل، آن‌ها استفاده از مقادیر بحرانی حاصل از توزیع‌های متناهی-نمونه متناظر آماره‌ها را توصیه کردند. برای محاسبه این مقادیر بحرانی، با دنبال کردن شیوه پیشنهادی تانگ و همکاران (۲۰۱۹)، به ازای یک حجم نمونه مشخص n باید ابتدا ۱۰ هزار مشاهده از توزیع یکنواخت دومتغیره روی مربع واحد، $H = U_{[0,1]}^2$ ، تولید کرد. سپس برای هر مجموعه داده از ۱۰۰۰۰ مجموعه، مقدار هر آماره آزمون محاسبه شده و سپس بر اساس توزیع تجربی آماره‌های آزمون، چندک مرتبه α بالایی ثبت می‌شود. برای هر آزمون هرگاه مقدار آماره آزمون بزرگتر از چندک محاسبه‌شده باشد، فرضیه H_0 رد می‌شود.

در مطالعه شبیه‌سازی گزارش‌شده در این فصل، از مقادیر بحرانی متناهی-نمونه برای همه آزمون‌ها استفاده کرده‌ایم. همچنین مجموعه داده‌ها در همه شبیه‌سازی‌ها $B = 10000$ بار تکرار شده‌اند.

جدول ۱.۴ اندازه و توان آزمون برآوردشده برای شش آزمون پیشنهادی را برای $n = 100$ و $\alpha = 0.05$ نشان می‌دهد. با توجه به مقادیر گزارش‌شده در جدول می‌توان نتایج زیر را بر شمرد:

^۶Finite-sample

۵۰ مطالعه شبیه‌سازی و مثال‌های کاربردی

۱. اندازه آزمون تجربی برای هر شش آزمون، زمانی که H_0 درست است، تقریباً برابر با اندازه اسمی ۰/۰۵ است.
 ۲. اگرچه، برای همه سناریوها، نمی‌توان آزمونی را به‌عنوان پرتوان‌ترین آزمون یکنواخت نسبت به سایر رقبا معرفی کرد، اما می‌توان تقریباً به صراحت گفت که آزمون KS به‌طور کلی کم‌توان‌ترین است.
 ۳. برای توابع مفصل با دم سنگین‌تر، آزمون EL تحت مفصل کلایتون به وضوح بهترین عملکرد را داراست.
 ۴. آزمون AD به‌طور کلی، تحت مفصل گامبل، نسبت به بقیه ارجحیت دارد؛ به‌ویژه زمانی که وابستگی PQD ضعیف‌تر است.
 ۵. برای توابع مفصل فرنک و گاوسی، آزمون EL تقریباً با آزمون‌های کندال و اسپیرمن رقابت می‌کند و به‌طور کلی هم‌توان یا پرتوان‌تر از آزمون‌های مبتنی بر فاصله است.
- جدول ۱.۴: اندازه و توان آزمون برآوردشده شش آزمون پیشنهادی برای $n = ۱۰۰$

	آزمون	$\tau = 0$	۰/۱۰	۰/۲۰	۰/۳۰	۰/۴۰
کلایتون	EL	۰/۰۵۲	۰/۴۷۵	۰/۹۲۳	۰/۹۹۹	۱/۰۰۰
	KS	۰/۰۵۳	۰/۳۱۱	۰/۷۸۰	۰/۹۸۳	۰/۹۹۹
	CvM	۰/۰۵۳	۰/۴۰۷	۰/۸۸۶	۰/۹۹۶	۱/۰۰۰
	AD	۰/۰۵۰	۰/۳۶۶	۰/۸۷۷	۰/۹۹۶	۱/۰۰۰
	اسپیرمن	۰/۰۵۱	۰/۴۴۰	۰/۹۰۲	۰/۹۹۷	۱/۰۰۰
فرنک	کندال	۰/۰۴۹	۰/۴۳۲	۰/۸۹۸	۰/۹۹۷	۱/۰۰۰
	EL	۰/۰۴۹	۰/۴۲۶	۰/۹۰۲	۰/۹۹۸	۱/۰۰۰
	KS	۰/۰۵۰	۰/۳۴۲	۰/۸۱۵	۰/۹۹۰	۱/۰۰۰
	CvM	۰/۰۵۰	۰/۴۲۹	۰/۹۰۵	۰/۹۹۸	۱/۰۰۰
	AD	۰/۰۴۹	۰/۳۱۴	۰/۸۱۷	۰/۹۹۲	۱/۰۰۰
گامبل	اسپیرمن	۰/۰۴۹	۰/۴۴۱	۰/۹۱۴	۰/۹۹۸	۱/۰۰۰
	کندال	۰/۰۴۸	۰/۴۳۵	۰/۹۱۳	۰/۹۹۸	۱/۰۰۰
	EL	۰/۰۴۹	۰/۴۴۶	۰/۹۰۲	۰/۹۹۷	۱/۰۰۰
	KS	۰/۰۵۱	۰/۳۲۶	۰/۷۸۰	۰/۹۷۸	۰/۹۹۹
	CvM	۰/۰۵۰	۰/۴۲۸	۰/۸۹۲	۰/۹۹۶	۱/۰۰۰
گاوسی	AD	۰/۰۵۰	۰/۵۰۶	۰/۹۱۹	۰/۹۹۷	۱/۰۰۰
	اسپیرمن	۰/۰۴۹	۰/۴۳۴	۰/۸۹۸	۰/۹۹۷	۱/۰۰۰
	کندال	۰/۰۴۹	۰/۴۳۲	۰/۸۹۶	۰/۹۹۷	۱/۰۰۰
	آزمون	$\rho = 0$	۰/۱۰	۰/۲۰	۰/۳۰	۰/۴۰
	EL	۰/۰۴۸	۰/۲۳۷	۰/۵۹۹	۰/۸۹۷	۰/۹۹۱
گاوسی	KS	۰/۰۵۴	۰/۱۸۹	۰/۴۴۵	۰/۷۴۰	۰/۹۳۷
	CvM	۰/۰۵۱	۰/۲۳۴	۰/۵۷۱	۰/۸۷۵	۰/۹۸۶
	AD	۰/۰۴۸	۰/۱۹۸	۰/۵۲۰	۰/۸۵۰	۰/۹۸۳
	اسپیرمن	۰/۰۴۹	۰/۲۴۵	۰/۶۰۵	۰/۸۹۷	۰/۹۹۱
	کندال	۰/۰۴۸	۰/۲۳۹	۰/۵۹۹	۰/۸۹۳	۰/۹۹۱

۲.۴ مثال‌های واقعی

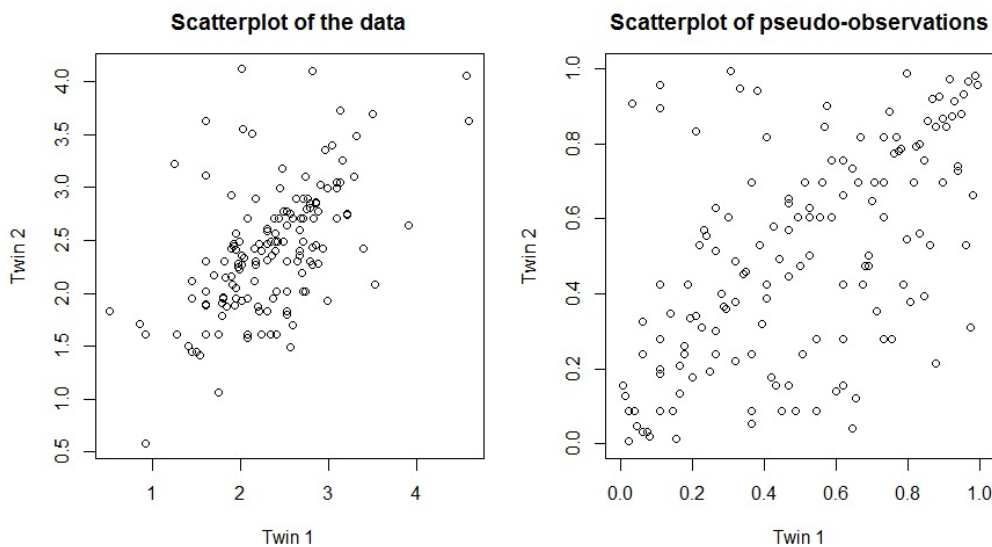
در این بخش، سه مجموعه داده واقعی را معرفی و همه آزمون‌های پیشنهادی را بر روی آن‌ها پیاده کرده‌ایم.

داده‌های دوقلوها

یک کاربرد طبیعی مساله آزمون PQD، در مطالعات دوقلوهاست. آشنفلتر و کروگر (۱۹۹۴) یک مطالعه مشاهده‌ای را در مورد ارزیابی دست‌مزدهای ساعتی $n = ۱۴۹$ دوقلوی همسان در کشور آمریکا منتشر کردند، به طوری که هر قلی (از جفت دوقلوها) سطح تحصیلات متفاوتی داشتند. افراد حاضر در این مطالعه سنی بالاتر از ۱۸ سال داشتند. هدف از این مطالعه بررسی تأثیر سطح تحصیلات بر دست‌مزدها بود. دلیل محدود کردن مطالعه فقط بر روی دوقلوهای همسان، کاهش تأثیر متغیرهایی مثل تفاوت‌های ژنتیکی و پیشینه خانوادگی است که قابل اندازه‌گیری نیستند یا به سختی در دسترس می‌باشند. در این مثال رابطه دومتغیره دست‌مزدها در هر دوقلو ارزیابی می‌شود.

فرض کنید X و Y دست‌مزد ساعتی برای دوقلوها بر حسب لگاریتم دلار آمریکا باشد. نمودار پراکنش داده‌های (x_i, y_i) ، به همراه نمودار پراکنش شبه-مشاهدات (u_i, v_i) ، در شکل ۱.۴ نمایش داده شده‌اند. کاملاً آشکار است که شبه-مشاهدات با استقلال مشاهدات هم‌خوانی ندارند، زیرا مقادیر (u_i, v_i) تمایل به تجمع حول قطر مربع واحد را دارند. بنابراین به نظر می‌رسد داده‌ها دارای وابستگی PQD باشند.

جدول ۲.۴ خلاصه نتایج مربوط به آماره‌های آزمون، مقادیر بحرانی و P - مقدار آزمون‌های پیشنهادی این پایان‌نامه را برای مجموعه داده دوقلوها نشان می‌دهد. هماهنگ با شکل ۱.۴، نتایج هر شش آزمون، شهود قوی به نفع PQD بودن داده‌ها را نشان می‌دهند. همان‌طور که مشهود است، P - مقدار همه آزمون‌ها کمتر از ۰/۰۰۱ است که به معنی رد استقلال مشاهدات جفتی دو متغیر X و Y یعنی جفت دوقلوهاست.



شکل ۱.۴: داده‌های دوقلوها: دستمزد ساعتی (به دلار آمریکا) برای $n = 149$ جفت از دوقلوهای همسان. نمودار سمت راست شبه-مشاهدات هستند.

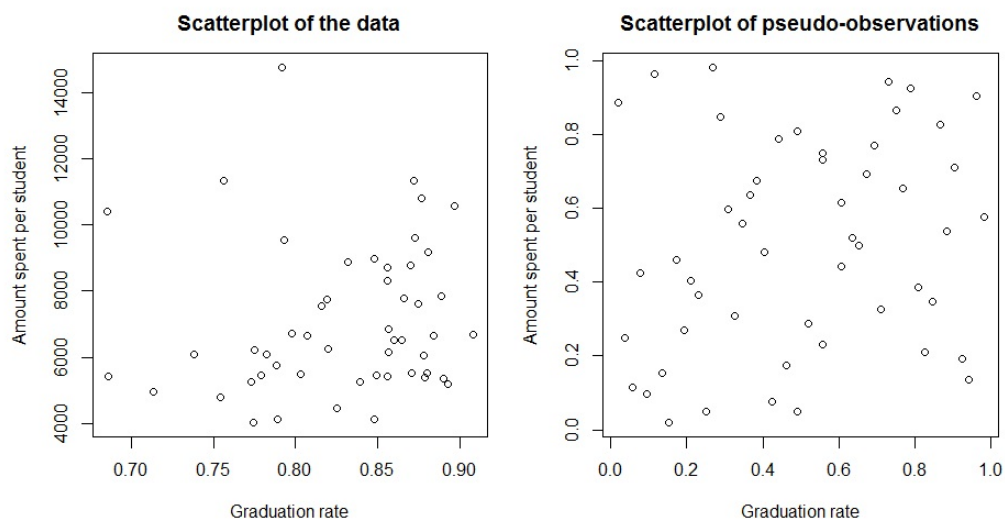
جدول ۲.۴: داده‌های دوقلوها: آماره آزمون، مقادیر بحرانی و P -مقدار

آزمون	آماره آزمون	مقدار بحرانی	P -مقدار
EL	۱۳/۱۷۶	۱/۳۷۴	$< 0/001$
KS	۱/۴۶۵	0/۶۸۳	$< 0/001$
CvM	0/۷۵۶	0/0۶۷	$< 0/001$
AD	۲۳/۴۷۶	۳/۱۶۳	$< 0/001$
اسپیرمن	0/۵۵۹	0/۱۳۵	$< 0/001$
کندال	0/۴۲۱	0/0۹۱	$< 0/001$

داده‌های تحصیلات

در سال ۲۰۱۸ معلمان دبیرستان‌های ۱۸ ایالت در کشور آمریکا در یک راهپیمایی رسمی در اعتراض به کاهش‌های متوالی پشتیبانی‌های مالی از آموزش، شرکت کردند. نکته جالب در این مساله این بود که تمام ایالت‌های درگیر در این اعتراض، بر اساس انتخابات سال‌های ۲۰۱۶ و ۲۰۱۸، یا به شدت محافظه‌کار بودند یا به شدت تندرو. سوالی که می‌توان مطرح کرد این است که آیا خواسته‌های معلمین برحق است؟ می‌تواند آزمون‌های پیشنهادی این پایان‌نامه را بر روی یک مجموعه داده پیاده کرد که می‌تواند تا حدودی برای این سوال پاسخی ارائه دهد.

تانگ و همکارانش (۲۰۱۹) یک مجموعه داده برای ۱۸ ایالت درگیر به اضافه ۵۱ ناحیه منطقه کلمبیا فراهم کردند که شامل نرخ فارغ‌التحصیلی دبیرستان در سال ۲۰۱۶ (X) و متوسط هزینه‌ای که دولت ایالتی برای هر دانش‌آموز طی سال تحصیلی ۲۰۱۵-۲۰۱۶ (بر حسب لگاریتم دلار آمریکا) پرداخته است (Y)، می‌باشد. هر دو متغیر برای هر ایالت



شکل ۲.۴: مجموعه داده تحصیلات: نمودار پراکنش مبلغ هزینه‌شده برای هر دانشجو (به دلار آمریکا) در سال تحصیلی ۲۰۱۵-۲۰۱۶ در مقابل نرخ فارغ‌التحصیلی دبیرستان در سال ۲۰۱۶ (سمت چپ) و نمودار پراکنش شبه-مشاهدات (سمت راست)

جدول ۲.۴ نیز مشابه مثال قبل، مقادیر آماره‌های آزمون‌های مختلف به همراه مقادیر بحرانی متناهی- نمونه و P - مقدار آن‌ها را نشان می‌دهد. هیچکدام از آزمون‌ها در سطح $\alpha = 0.05$ استقلال مشاهدات را رد نمی‌کنند. البته آزمون‌های EL ، CvM ، کندال و اسپیرمن به‌طور مرزی، با کمی فاصله از 0.05 ، معنی‌دار هستند.

جدول ۳.۴: داده‌های تحصیلات: آماره آزمون، مقادیر بحرانی و P - مقدار

آزمون	آماره آزمون	مقدار بحرانی	P - مقدار
KS	۰/۴۷۰	۰/۶۷۲	۰/۳۳۰
CvM	۰/۰۸۳	۰/۰۹۱	۰/۰۶۸
AD	۲/۷۷۳	۴/۷۱۶	۰/۲۲۴
اسپیرمن	۰/۲۱۵	۰/۲۳۵	۰/۰۶۸
کندال	۰/۱۵۰	۰/۱۵۹	۰/۰۶۲

با توجه به نتایج به‌دست آمده، نمی‌توان وابستگی بین هزینه صرف‌شده و نرخ فارغ‌التحصیلی دانش‌آموزان دبیرستانی پیدا کرد. مثلاً نمی‌توان گفت که همراه با کاهش پرداختی‌های دولتی، نرخ فارغ‌التحصیلی نیز کم شده است. به عبارتی با این تحلیل نمی‌توان پرسش مطرح‌شده را پاسخ داد، اما می‌توان گفت نتیجه حاصل، به‌طور معنی‌داری، استدلالی قوی برای افزایش حمایت‌های پولی دولت‌های ایالتی نمی‌تواند باشد.

داده‌های سهام

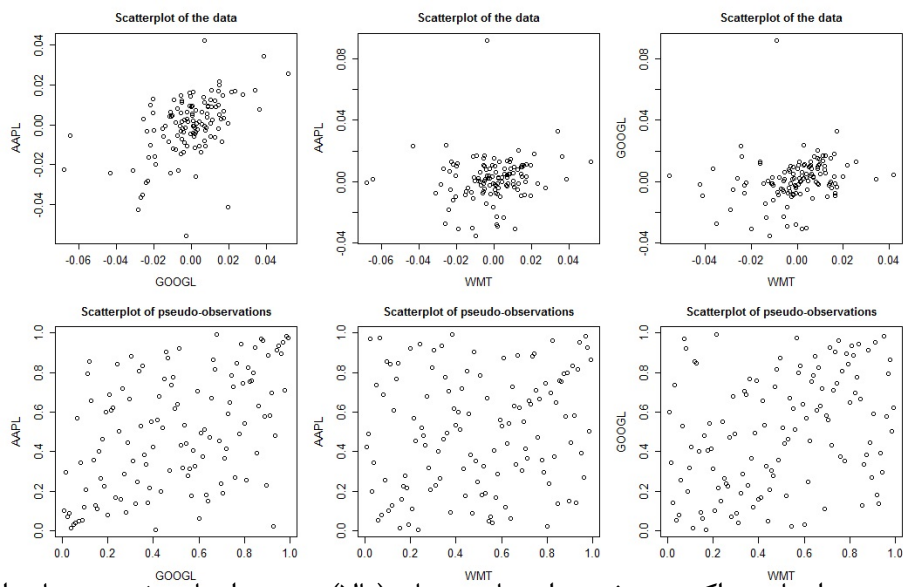
این مجموعه داده شامل قیمت‌های تمام‌شده سهام سه شرکت در آمریکا طی ۱۲۵ روز کاری متوالی بین ۴ ژانویه ۲۰۱۶ و ۳۰ ژوئن ۲۰۱۶ است که توسط تانگ و همکارانش (۲۰۱۹) تحلیل شده‌اند. این سه شرکت عبارتند از اپل (APPL)، گوگل (GOOGL) و المارت (WMT) برای هر شرکت، قیمت‌های تمام‌شده سهام ۱۲۵ روز مستقل نیستند، اما مقادیر تفاضلی مرتبه اول لگاریتم قیمت‌ها، یعنی ۱۲۴ مقدار لگاریتم بازگشت سرمایه، باید تقریباً مستقل باشند. با استفاده از آزمون لیونگ-باکس (لیونگ و باکس، ۱۹۷۸) استقلال مقادیر بازگشت سرمایه بررسی شدند که P -مقدار آماره‌های آزمون مربوط برای سه شرکت به ترتیب $۰/۵۷۷$ ، $۰/۶۳$ و $۰/۸۲$ به دست آمدند. یعنی استقلال بین این مقادیر در سطح ۵ درصد پذیرفته می‌شود. با توجه به طرح‌های تجاری مشترک بین این سه شرکت، بررسی ساختارهای وابستگی جفتی بین بازگشت سرمایه این سه شرکت می‌تواند مورد علاقه باشد.

جدول ۲.۴ مشابه دو مثال قبلی نتایج آزمون‌های سنجش وابستگی PQD را روی سه جفت متغیر این مجموعه داده نشان می‌دهد. همچنین شکل ۳.۴ نمودارهای پراکنش زوج مشاهدات و نمودارهای شبه-مشاهدات متناظر آن‌ها را نمایش می‌دهد. در نگاه اول به شکل ۳.۴ ممکن است تصور کنیم استقلال بین مشاهدات (بازگشت سرمایه) شرکت‌های اپل و المارت منطقی است؛ اما ۵ تا ۶ آزمون، از جمله آزمون EL، در اندازه $\alpha = ۰/۰۵$ فرضیه استقلال را رد می‌کنند. از طرفی، شهودی قوی و قابل دفاع برای PQD بودن دو جفت متغیر (اپل و گوگل) و (گوگل و المارت) وجود دارد. با این نتیجه، برای سرمایه‌گذارانی که به دنبال تنوع سهام و کاهش نوسان هستند، قرار دادن هر سه این شرکت‌ها در یک سبد سهام یک تصمیم غیرعقلانه است.

جدول ۴.۴: مجموعه داده سهام: آماره آزمون، مقادیر بحرانی و P -مقدار

	آزمون	آماره آزمون	مقدار بحرانی	P -مقدار
GOOGL v APPL	EL	۸/۲۵۳	۱/۴۰۳	$< ۰/۰۰۱$
	KS	۱/۰۱۸	۰/۶۷۹	$< ۰/۰۰۱$
	CvM	۰/۳۸۱	۰/۰۶۹	$< ۰/۰۰۱$
	AD	۱۷/۴۲۰	۳/۳۶۷	$< ۰/۰۰۱$
	اسپیرمن	۰/۴۷۶	۰/۱۴۸	$< ۰/۰۰۱$
	کندال	۰/۳۴۱	۰/۱۰۰	$< ۰/۰۰۱$
WMT v APPL	EL	۱/۴۱۰	۱/۴۰۳	$۰/۰۴۹$

	KS	۰/۶۶۲	۰/۶۷۹	۰/۰۵۹
	CvM	۰/۰۸۹	۰/۰۶۹	۰/۰۲۱
	AD	۳/۸۰۱	۳/۳۶۷	۰/۰۳۲
	اسپیرمن	۰/۱۵۵	۰/۱۴۸	۰/۰۴۲
	کندال	۰/۱۱۲	۰/۱۰۰	۰/۰۳۱
WMT v GOOGL	EL	۴/۲۶۶	۱/۴۰۳	< ۰/۰۰۱
	KS	۱/۰۲۸	۰/۶۷۹	< ۰/۰۰۱
	CvM	۰/۲۴۷	۰/۰۶۹	< ۰/۰۰۱
	AD	۶/۸۶۳	۳/۳۶۷	۰/۰۰۲
	اسپیرمن	۰/۳۳۶	۰/۱۴۸	< ۰/۰۰۱
	کندال	۰/۲۳۶	۰/۱۰۰	< ۰/۰۰۱



شکل ۳.۴: نمودارهای پراکنش جفتی داده‌های سهام (بالا) و نمودارهای شبه-مشاهدات متناظر (پایین)

مراجع

- [1] Ahmad, L.A. and Li, Q. (1997) Testing symmetry of an unknown density function by kerne Nonparametr. Statist., 7.279-293.
- [2] Ashenfelter, O. and Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. American Economic Review 84,1157-1173.
- [3] Barlow. RE, Proschan F(1981) Stntistical Theory of Reliability and Life Test- ing. To Begin With , Silver Spring, MD.
- [4] Denuit, M. and Scaillet, O. (2004). Nonparametric tests for positive quadrant dependence. Journal of Financial Econometrics, 2(3), 422-450.
- [5] Einmahl, J. H. and Mckeague, l.w. (2003). Empirical likelihood based hypoth- esis testing . Bernoulli, 9(2), 267-290.
- [6] Gijbels, I., Omelka, M., and Sznajder, D. (2010). Positive quadrant depen- dence tests for copulas. Canadian Journal of Statistics, 38(4), 555-581.
- [7] Hill, D.L. and Rao, P.V. (1977) Tests of symmetry based on Cramer-von Mises statistics. Blumetrika, 64, 489-494
- [8] Hutchinson TP. Lai CD (1990) Continuous Bivariate Distributions. Empha- sising Applications, Rumsby Scientific Publishing, Adelaide.
- [9] Janic-Wroblewska, A. KALLENBERG, w, and Ledwina, T. (2004). Detect- ing positive quadrant depenence and positive function dependence. insur- ance:Mathematics and Economics 34, 467-487.
- [10] Kochar, S. C., and Gupta, R. P. (1987). Competitors of the Kendall-tau test for testing independence against positive quadrant dependence. Biometrika, 74(3), .664-666

-
- [11] Lehmann EL (1966) Some concepts of dependence. *Ann Math Statist* 37:1137-1153.
- [12] Lai, C. and Xie, M. (2006). *Stochastic Ageing and Dependence for Reliability*. Springer, New York.
- [13] Ljung, G. M., and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303.
- [14] Malevengne, Y. and Sornette, D. (2006). *Extreme Financial Risks: From Dependence to Risk Management*. Springer, Berlin.
- [15] Nelsen, R.B. (2007). *An introduction to copulas*. Springer Science and Business Media.
- [16] Nikitin, Y.Y. (1996b) Buldur efficiency of a test of exponentiality based on a lass-of-memory type functional equation *Nonparametr. Statist.*, 6, 13-26.
- [17] Nikitin, Y.Y. (1996a) On Baringhaus-Henze test for symmetry: Bahadur efficiency and local optimality for sluit alternatives, *Math. Medods Statis*, 5, 214-226.
- [18] Scott, D. (1992): "Multivarinte Density Estimation: Theory ; Prattice and Visualisation", John Wiley and Sons, New-York.
- [19] Tang, C. F. Wang, D. El Barmi, H. and Tebbs, . M. (2019). Testing for positive quadrant dependence. *the American Statistician*, 1-15.

Aabstract

Positive quadratic dependence (PQD) is related to the joint behavior of random variables; when small (or large) values are likely to be observed for them simultaneously. Recognition of the presence of this dependency structure in the data can lead to improved computations and decisions. Therefore, in this thesis, the independence test of two random variables versus having PQD dependence has been studied. Given that the actual distribution of random variables is practically unknown, we will use their empirical versions in a nonparametric approach based on the copula function. The simulation results show that the proposed empirical likelihood test performs well, having the ability to reveal (X,Y) are PQD.

Keywords: Positive quadratic dependence, independence test, copula function, empirical likelihood test

This thesis studies on writing projects, theses and dissertations using Shahrood_thesis Class. It ...



Faculty Of Mathematical Sciences

MSc Thesis in Mathematical Statistics

Nonparametric tests for positive quadratic dependence

By: Zahra Seidabadi

Supervisor:

Negar Eghbal

Advisor:

Hossein Baghishani

February 2021