

تعمیم تحلیل همبستگی کانونی چند متغیره به تابعی و مشکلات مربوط به آن

سید محمد ابراهیم حسینی نسب^۱

چکیده:

در حالت چند متغیره، هدف از تحلیل همبستگی کانونی برای دو مجموعه از متغیرهای x و y ، بدست آوردن ترکیب‌های خطی از آن‌ها به قسمی است که دارای بزرگترین همبستگی ممکن باشند. اما وقتی که x و y ذاتاً تابع پیوسته‌ای از یک متغیر دیگر (معمولاً زمان) باشند این دو تابع به فضاهای تابعی با بُعد نامتناهی متعلق هستند و تحلیل همبستگی برای آن‌ها باید با استفاده از ابزارهایی که از طریق تحلیل داده‌های تابعی فراهم می‌شود انجام پذیرد. در این مقاله، ضمن مرور تحلیل همبستگی کانونی برای داده‌های چند متغیره، به بیان تعاریف و مفاهیم مربوط به تحلیل همبستگی کانونی برای داده‌های تابعی با ذکر مشکلات بوجود آمده در تعمیم از حالت چند متغیره به تابعی می‌پردازیم. همچنین در ادامه یک مجموعه داده‌ی تابعی واقعی را تحلیل و نتایج بدست آمده را تفسیر می‌کنیم.

واژه‌های کلیدی: تحلیل داده‌های تابعی، تحلیل همبستگی کانونی تابعی، تحلیل همبستگی کانونی چند متغیره.

۱ مقدمه

نمی‌توان روش‌های موجود برای تحلیل داده‌های چند متغیره را برای آن‌ها به کار گرفت و تحلیل آن‌ها مستلزم فراهم ساختن بسترهای لازم است که در آن ماهیت تابعی داده‌ها در نظر گرفته شود (رمسی و سیلورمن، ۲۰۰۲ و ۲۰۰۵، تازیکی و همکاران، ۱۳۸۷). مطالعه‌ی همبستگی کانونی برای چنین داده‌هایی به یک روش شناسی جدید در آمار منتهی می‌شود که به آن تحلیل همبستگی کانونی تابعی (FCCA)^۲ گویند.

در این مقاله در بخش ۲ به بررسی اجمالی تحلیل همبستگی کانونی در حالت چند متغیره می‌پردازیم و در ادامه نیز در بخش ۳ به اختصار، تعاریف تحلیل همبستگی تابعی را با ذکر مشکلات بوجود آمده در تعمیم از حالت چند متغیره به تابعی بیان می‌کنیم. در انتها نیز در بخش ۴ یک مجموعه داده‌ی تابعی واقعی تحلیل و نتایج مربوط به آن تفسیر می‌شود.

بررسی ساختار همبستگی و ارتباط میان متغیرها یکی از مسایل مورد توجه در آمار است. روش‌های مختلفی برای بررسی این ارتباط وجود دارد که بسته به ویژگی‌های داده‌های مورد بررسی و هدف تحقیق می‌توان از هر کدام از آن‌ها استفاده نمود. چنانچه دو مجموعه از متغیرها شامل تعداد زیادی متغیر باشند و هدف مطالعه، بیان ارتباط میان این دو مجموعه در تعداد محدودی ترکیب خطی با برخی خواص ماکسیمال خلاصه شود، از تحلیل همبستگی کانونی استفاده می‌شود. تحلیل همبستگی کانونی متضمن افزایش کردن گردایه‌ای از متغیرها به دو مجموعه x (شامل p متغیر) و y (شامل q متغیر) است که هدف آن بدست آوردن ترکیب‌های خطی $u = a'y$ و $v = b'x$ به قسمی است که u و v دارای بیشترین همبستگی ممکن باشند.

اگر مشاهدات، توابع پیوسته‌ای از یک متغیر (معمولاً زمان) باشند،

^۱ عضو هیأت علمی گروه آمار دانشگاه شهید بهشتی

^۲Functional Canonical Correlation Analysis

۲ تحلیل همبستگی کانونی در داده‌های

چند متغیره

تحلیل همبستگی کانونی شناخت و کمی کردن رابطه‌ی میان دو مجموعه از متغیرها را بررسی می‌کند. در تحلیل همبستگی کانونی، همبستگی بین یک ترکیب خطی از متغیرهای یک مجموعه و یک ترکیب خطی از متغیرهای مجموعه دیگر مطالعه می‌شود. این کار با تعیین دو ترکیب خطی از دو مجموعه متغیر با بیشترین همبستگی آغاز می‌شود. سپس دو ترکیب خطی که در میان تمام زوج‌های ناهمبسته با زوج انتخاب شده‌ی اول، دارای بیشترین همبستگی باشد، بدست می‌آید و این فرآیند ادامه می‌یابد. هر زوج ترکیب خطی را متغیر کانونی و همبستگی بین آن‌ها را همبستگی کانونی نامند. همبستگی‌های کانونی شدت ارتباط بین دو مجموعه از متغیرها را اندازه می‌گیرد.

تحلیل همبستگی کانونی برخی خاصیت‌های ماکسیمال مشابه با آنچه که در بحث تحلیل مولفه اصلی مطرح است را دارا می‌باشد، با این تفاوت که تحلیل مولفه اصلی، همبستگی‌های درونی بین مجموعه‌ای از متغیرها را در نظر می‌گیرد، در حالی که، بحث همبستگی کانونی بر همبستگی بین دو گروه از متغیرها تمرکز می‌کند.

تحلیل همبستگی کانونی را می‌توان از جنبه‌ی دیگری نیز مورد توجه قرار داد، بدین صورت که این تحلیل را تعمیمی از تحلیل رگرسیونی در نظر گرفت. همان‌طور که می‌دانیم در رگرسیون در پی یافتن ترکیب خطی $b'x$ از متغیرهای مستقل x می‌باشیم که قویاً با y همبسته باشد، حال آنکه تحلیل همبستگی کانونی بر یافتن ترکیب‌های خطی $b'x$ و $a'y$ که قویاً با هم همبسته باشند متمرکز است. بنابراین این دو تحلیل تنها در تعداد متغیرهای مجموعه y تفاوت دارند؛ به این معنی که در تحلیل همبستگی کانونی تعداد متغیرهای مجموعه‌ی y می‌تواند بیش از یک باشد.

جنبه‌ی ماکسیم کردن در این روش کوششی برای متمرکز ساختن رابطه‌ی بین دو مجموعه از متغیرها، به چند زوج از متغیرهای کانونی است. اگر x به عنوان علت y تعبیر شود، آن‌گاه $b'x$ را می‌توان «بهترین پیشگو» و $a'y$ را «بهترین قابلیت پیشگویی» نامید. به عنوان مثال نمرات امتحانی ۱۰۰ دانشجو در ۵ درس را در نظر بگیرید که در آن‌ها ۳ درس جزوه باز و ۲ درس جزوه بسته امتحان گرفته شده است. در اینجا موضوع مورد علاقه آن است که چگونه استعداد یک دانشجو در

امتحانات جزوه بسته با توانایی‌های او در امتحانات جزوه باز قویاً همبسته است. با استفاده از این اطلاعات می‌توان از نتایج امتحان‌های جزوه باز برای پیشگویی نتایج امتحان‌های جزوه بسته استفاده نمود و بالعکس.

در همبستگی کانونی هدف، اندازه‌گیری رابطه‌ی میان دو گروه از متغیرها است. برای این کار، اولین گروه شامل p متغیر را با بردار تصادفی y و دومین گروه شامل q متغیر را با بردار تصادفی x نشان می‌دهیم. فرض کنید برای بردارهای تصادفی y و x :

$$Cov(y) = \Sigma_{yy}, Cov(x) = \Sigma_{xx}, Cov(y, x) = \Sigma_{yx}.$$

عناصر ماتریس Σ_{yx} (یا Σ_{xy}) در واقع کوواریانس بین تمام زوج‌هایی از متغیرها که می‌توان از y و x انتخاب نمود را شامل می‌شود. در این صورت این ماتریس $p \times q$ عضو دارد که ارتباط بین این دو مجموعه را نشان می‌دهد. وقتی p و q نسبتاً بزرگ باشند، تفسیر اعضای Σ_{yx} ، با هم، معمولاً غیر ممکن و بی‌فایده است. علاوه بر این اغلب، ترکیب‌های خطی از این متغیرها هستند که برای اهداف تخمینی یا مقایسه‌ای جالب و مفیدند. بنابراین، کار اصلی تحلیل همبستگی کانونی آن است که روابط مجموعه‌های y و x را به جای $p \times q$ کوواریانس در ماتریس Σ_{yx} بر حسب چند ترکیب خطی که به دقت انتخاب شده‌اند، خلاصه کند. در حقیقت، این ترکیب‌های خطی، اندازه‌های خلاصه شده‌ی ساده‌ای از این دو مجموعه از متغیرها را فراهم می‌کنند.

برای یک زوج بردار ضرایب a و b قرار می‌دهیم:

$$u = a'y \quad v = b'x$$

داریم:

$$Var(u) = a'Cov(y)a = a'\Sigma_{yy}a,$$

$$Var(v) = b'Cov(x)b = b'\Sigma_{xx}b,$$

$$Cov(u, v) = a'Cov(y, x)b = a'\Sigma_{yx}b.$$

بردارهای ضرایب a و b را به گونه‌ای می‌یابیم که

$$Corr(u, v) = \frac{a'\Sigma_{yx}b}{\sqrt{a'\Sigma_{yy}a}\sqrt{b'\Sigma_{xx}b}}$$

تا حد امکان بزرگ باشد.

۱.۲ تعاریف

اولین زوج متغیر کانونی: زوج ترکیب خطی (u_1, v_1) که دارای واریانس‌های واحد بوده و همبستگی میان ترکیب‌های خطی را ماکسیم کند اولین زوج متغیر کانونی نامند.

درفرمول R^2 ، q کوواریانس (ضرب همبستگی) بین y و x ها در s_{yx} و (r_{yx}) ، معیار ساده‌ای از رابطه خطی بین y و x ها است (رنجر، ۲۰۰۲). حال فرض کنید بردار y بیش از یک مولفه داشته باشد یعنی $1 < p$. در این صورت اگر

$$(y_{n1}, \dots, y_{np}, x_{n1}, \dots, x_{nq}), \dots, (y_{11}, \dots, y_{1p}, x_{11}, \dots, x_{1q})$$

یک نمونه به حجم n از مشاهدات باشد آنگاه ماتریس کوواریانس آن‌ها به صورت زیر است:

$$S = \begin{pmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{pmatrix},$$

که در آن S_{yy} ماتریس واریانس-کوواریانس نمونه $p \times p$ از مولفه‌های y و S_{xx} ماتریس واریانس-کوواریانس نمونه‌ای $q \times q$ از مولفه‌های x است. همچنین S_{yx} ماتریس $p \times q$ ، واریانس-کوواریانس نمونه‌ای بین مولفه‌های y و مولفه‌های x است.

با تعمیم مفهوم ضرب همبستگی چندگانه R^2 ، می‌توان توان دوم همبستگی کانونی به صورت ماکسیم توان دوم همبستگی بین ترکیب‌های خطی از y و ترکیب‌های خطی از x در نظر گرفت که در این صورت بردارهای ضرایب a و b به گونه‌ای انتخاب می‌شوند که توان دوم همبستگی نمونه‌ای r_{uv}^2 بین $u = a'y$ و $v = b'x$ ماکسیم گردد. ضرب همبستگی نمونه‌ای بین ترکیب‌های خطی u و v به صورت زیر تعریف می‌شود:

$$r_{uv} = \text{Corr}(u, v) = \frac{a'S_{yx}b}{\sqrt{(a'S_{xx}a)(b'S_{yy}b)}}.$$

تعریف: مربع همبستگی کانونی به صورت $r^2 = \max_{a,b}(r_{uv}^2)$ تعریف می‌شود و ترکیب‌های خطی $u = a'y$ و $v = b'x$ که ماکسیم توان دوم همبستگی در آن‌ها رخ می‌دهد متغیرهای کانونی نامیده می‌شوند. اگر $p \leq q$ آنگاه

$$\max_{a,b} \text{Corr}(u, v) = r_1^*,$$

و این ماکسیم به وسیله ترکیب‌های خطی (اولین زوج متغیر کانونی) زیر بدست می‌آید:

$$u_1 = e_1' S_{yy}^{-1/2} y, \quad v_1 = f_1' S_{xx}^{-1/2} x.$$

همچنین زوج متغیر کانونی k ام برای $k = 2, 3, \dots, p$ عبارتند از:

$$u_k = e_k' S_{yy}^{-1/2} y, \quad v_k = f_k' S_{xx}^{-1/2} x,$$

دومین زوج متغیر کانونی: ترکیب‌های خطی (v_2, u_2) که دارای واریانس واحد هستند و همچنین همبستگی میان ترکیبات خطی را در میان تمام انتخاب‌هایی که با اولین زوج متغیر کانونی ناهمبسته است ماکسیم می‌کند، دومین زوج متغیر کانونی نامیده می‌شود.

k امین زوج متغیر کانونی: ترکیب‌های خطی u_k و v_k که دارای واریانس‌های واحد بوده و همبستگی میان ترکیب‌های خطی را در میان تمام انتخاب‌های ناهمبسته به $k-1$ زوج متغیر کانونی قبلی ماکسیم کند به عنوان k امین زوج متغیر کانونی شناخته می‌شود. همبستگی بین زوج k ام از متغیرهای کانونی را همبستگی کانونی k ام می‌نامند.

۲.۲ روش بدست آوردن همبستگی و متغیرهای کانونی در داده‌های چند متغیره

برای فهم بهتر محاسبات و نظریه‌ی همبستگی کانونی، همچنین وجود رابطه‌ی میان این همبستگی و ضرب همبستگی چندگانه، مناسب‌تر است ابتدا به توضیح مختصری از ضرب همبستگی چندگانه بپردازیم.

ضرب همبستگی چندگانه R که در بحث رگرسیون چند گانه مورد بحث قرار می‌گیرد، اندازه‌ای از رابطه‌ی (خطی) بین یک متغیر وابسته y و چندین متغیر مستقل x را نشان می‌دهد. فرض کنید n نمونه تصادفی از بردار مشاهدات به صورت $(y_i, x_{i1}, x_{i2}, \dots, x_{iq})$ در اختیار داریم. کوواریانس و همبستگی نمونه‌ای میان $y, x_1, \dots, x_q$ به صورت زیر بیان می‌شود:

$$S = \begin{pmatrix} S_y^2 & s'_{yx} \\ s_{yx} & S_{xx} \end{pmatrix}, \quad R = \begin{pmatrix} 1 & r'_{yx} \\ r_{yx} & R_{xx} \end{pmatrix},$$

که در آن بردار s'_{yx} شامل کوواریانس نمونه‌ای متغیر y با متغیرهای $x_1, \dots, x_q$ و S_{xx} ماتریس کوواریانس نمونه‌ای بین مولفه‌های x است. همچنین زیر ماتریس‌های موجود در ماتریس همبستگی افزا شده R به طور مشابه تعریف می‌شوند.

ضرب همبستگی چندگانه R به صورت ماکسیم همبستگی بین y و یک ترکیب خطی از مولف‌های x یعنی $x = b'x$ تعریف می‌شود. همچنین می‌توان نشان داد که توان دوم ضرب همبستگی چندگانه به صورت زیر است:

$$R^2 = \frac{s'_{yx} S_{xx}^{-1} s_{yx}}{S_y^2} = r'_{yx} R_{xx}^{-1} r_{yx},$$

R تعریف شده‌اند و انتگرال توان دوم آن‌ها متناهی است. حاصلضرب داخلی مرتبط با هر فضا به صورت زیر تعریف می‌شود:

$$\langle u, v \rangle = \int u(s)v(s)ds.$$

همبستگی کانونی در حالت چند متغیره برای بردارهای بُعد متناهی $X \in R^{k_1}$ و $Y \in R^{k_2}$ و در حالت تابعی برای فرآیندهای تصادفی $X(\cdot) \in L_2(T_1)$ و $Y(\cdot) \in L_2(T_2)$ تعریف می‌شود. در این صورت k امین ($k > 1$) همبستگی کانونی و توابع وزن مربوط به آن که به ترتیب آن‌ها را با ρ_k ، $u_k(\cdot)$ و $v_k(\cdot)$ نشان می‌دهیم به صورت زیر تعریف می‌شود:

$$\begin{aligned} \rho_k &= \sup_{u \in H_1, \text{Var}(\langle u, X \rangle) = 1, \\ v \in H_2, \text{Var}(\langle v, Y \rangle) = 1} \text{Cov}(\langle u, X \rangle, \langle v, Y \rangle), \\ &= \text{Cov}(\langle u_k, X \rangle, \langle v_k, Y \rangle), \end{aligned}$$

که در آن k امین جفت از متغیرهای کانونی (U_k, V_k) با $k - 1$ جفت $\{(U_i, V_i), i = 1, 2, \dots, k - 1\}$ ناهمبسته است و $U_k = \langle u_k, X \rangle$ و $V_k = \langle v_k, Y \rangle$ در فرمول فوق H_i در حالت چند متغیره به صورت $H_i = R^{k_i}$ و در حالت تابعی به صورت $H_i = L_2(T_i)$ تعریف می‌شود.

لازم به ذکر است که با وجود تفاوت‌های اساسی در مفهوم و بحث وجودی آن، تعریف همبستگی کانونی در حالت تابعی مشابه با حالت چند متغیره است. این تشابه با تبدیل حاصلضرب داخلی در حالت چند متغیره که برای دو بردار تصادفی x و y به صورت $\langle x, y \rangle = x'y$ است به حاصلضرب داخلی در فضاهای هیلبرت مرتبط با دو فرآیند تصادفی $X(\cdot)$ و $Y(\cdot)$ یعنی $\langle X, Y \rangle = \int X(t)Y(t)dt$ ، نمود پیدا می‌کند. اما برای تعریف CCA برای داده‌های تابعی، تنها تبدیل حاصلضرب داخلی مربوط به فضاهای برداری به فضاهای هیلبرت مورد نظر راه گشا نخواهد بود زیرا عدم توجه به نکات مهم در هنگام تعریف، ممکن است به اراییه یک تعریف ناکارآمد منجر شود. در حالی که تعریف CCA برای داده‌های چند متغیره بدون هیچ شرط دست و پاگیری انجام می‌شود، ارائه‌ی تعریف CCA برای داده‌های تابعی ممکن است با مشکلاتی مواجه شود. داکسیس و همکاران (۲۰۰۴) یک تحلیل همبستگی کانونی بین دو متغیر تصادفی تابعی نسبت به یک متغیر دیگر معرفی کردند که تنها زمانی که $C_{XX}^{1/2}$ و $C_{YY}^{1/2}$ به ترتیب زیر فضاهای بسته‌ای

که این زوج متغیر مقدار $\text{Corr}(u_k, v_k)$ را در میان ترکیب‌های خطی ناهمبسته با $k - 1$ متغیر کانونی قبلی ماکسیمم می‌کند. اگر این مقدار ماکسیمم را با r_k^* نشان دهیم آنگاه p همبستگی کانونی اول به صورت $r_1^* \geq r_2^* \geq \dots \geq r_p^*$ خواهد بود که در واقع ویژه مقادیر ماتریس $S_{yy}^{-1/2} S_{yx} S_{xx}^{-1} S_{xy} S_{yy}^{-1/2}$ هستند. همچنین $e_1, e_2, \dots, e_p$ ویژه بردارهای $p \times 1$ بعدی متناظر با این ویژه مقادیر می‌باشند. (کمیت‌های $r_1^*, \dots, r_p^*$ ، اولین p ویژه مقدار ماتریس $S_{xx}^{-1/2} S_{xy} S_{yy}^{-1} S_{yx} S_{xx}^{-1/2}$ متناظر با ویژه بردارهای $q \times 1$ بعدی $f_1, f_2, \dots, f_p$ هستند که در آن هر f_i با e_i $S_{xx}^{-1/2} S_{xy} S_{yy}^{-1/2}$ متناسب است). در این صورت برای $k, l = 1, 2, \dots, p$ متغیر کانونی ساخته شده به صورت بالا خواص زیر را داریم:

- ۱) $\text{Var}(u_k) = \text{Var}(v_k) = 1$,
- ۲) $\text{Cov}(u_k, v_l) = \text{Corr}(u_k, v_l) = 0, k \neq l$,
- ۳) $\text{Cov}(v_k, v_l) = \text{Corr}(v_k, v_l) = 0, k \neq l$,
- ۴) $\text{Cov}(u_k, u_l) = \text{Corr}(u_k, u_l) = 0, k \neq l$.

برای جزئیات بیشتر در مورد تحلیل همبستگی کانونی چند متغیره به رنچر (۲۰۰۲) رجوع کنید.

۳ تحلیل همبستگی کانونی داده‌های تابعی

داده‌های چند متغیره براساس ماتریس مشاهدات تحلیل می‌شوند که در آن مولفه‌ی x_{ij} نشان‌دهنده‌ی مقادیر متغیر j ام برای آزمودنی i ام است. اما در تحلیل داده‌های تابعی^۳ (FDA)، مقدار مشاهده شده برای آزمودنی i ام به صورت $x_i(t)$ است که در آن $t \in I = [0, T]$. با وجود این تفاوت، با اقتباس از نگاشت‌های ماتریسی در حالت چند متغیره، نگاشت‌های تابعی برای مشاهداتی با بُعد بی‌نهایت نیز قابل تعریف هستند (رمسی ۱۹۸۲).
تعریف: فرآیند دو متغیره (X, Y) که در آن $X(\cdot) \in L_2(T_1)$ و $Y(\cdot) \in L_2(T_2)$ دارای یک توزیع توام می‌باشند را در نظر بگیرید:

$$\int E \|Z\|^2 = E[\langle ZZ \rangle] = E \int_T (Z(s))^2 ds < \infty;$$

$Z = X$ یا Y ,

که در آن T_1 و T_2 دو مجموعه‌ی اندیس‌گذار و $L_2(T_1)$ و $L_2(T_2)$ فضای هیلبرت هستند که توابع موجود در آن‌ها به ترتیب از T_1 و T_2 به

به سادگی می‌توان نشان داد که:

$$\begin{aligned} Cov(<u, X>, <v, Y>) &= \\ E\{[<u, X> - E(<u, X>)][<v, Y> - E(<v, Y>)]\} \\ &= E\{[<u, X - EX>][<v, Y - EY>]\} \\ &= <u, C_{XY}v>. \end{aligned}$$

همچنین به طور مشابه داریم:

$$\begin{aligned} Var(<u, X>) &= <u, C_{XX}u>, \\ Var(<v, Y>) &= <v, C_{YY}v>. \end{aligned}$$

با توجه به ویژگی‌ها و خواص عملگرهای کوواریانس همچنین روابط بالا می‌توان k امین همبستگی کانونی را به صورت زیر محاسبه کرد:

$$\begin{aligned} \rho_k &= \sup_{\substack{u \in L_T(T_1), <u, C_{XX}u>=1, \\ v \in L_T(T_2), <v, C_{YY}v>=1}} <u, C_{YX}v> \\ &= <u_k, C_{YX}v_k>, \end{aligned}$$

که در آن برای $k > 1$ ، (U_k, V_k) با (U_i, V_i) برای $i = 1, 2, \dots, k-1$ ناهمبسته است. برای جزئیات بیشتر به هی و همکاران (۲۰۰۳) مراجعه نمایید.

۱.۳ برآوردهای تجربی

فرض کنید (X_i, Y_i) ، $i = 1, 2, \dots, n$ نمونه‌ای به حجم n باشد. آنگاه متغیرهای کانونی نمونه به وسیله جایگذاری عملگرهای کوواریانس با نسخه‌های نمونه‌ای آن مانند قبل بدست می‌آید. برآورد تجربی عملگرهای کوواریانس به صورت زیر است:

$$\begin{aligned} \hat{C}_{XX}u(s) &= \int_{T_1} \hat{c}_{XX}(s, t)u(t)dt, \quad u \in L_T(T_1), \\ \hat{C}_{XY}u(s) &= \int_{T_2} \hat{c}_{XY}(s, t)u(t)dt, \quad u \in L_T(T_2), s \in T_1, \\ \hat{C}_{YY}u(s) &= \int_{T_2} \hat{c}_{YY}(s, t)u(t)dt, \quad u \in L_T(T_2). \end{aligned}$$

در $L_T[0, T_1]$ و $L_T[0, T_2]$ هستند تعریف می‌شود که این خاصیت تنها زمانی که دو عملگر متناهی بعد هستند برقرار است. این پژوهش مشابه با دیگر کارهای مکتب فرانسه از قبیل داکسیس و پوز (۱۹۷۶) است که بر اساس نتایج اولیه‌ی هتلینگ که برای اولین بار CCA را برای داده‌های چند متغیره پیشنهاد داد بنا شده‌اند. اما کوپیدن و همکاران (۲۰۰۸) نشان دادند که جواب همبستگی کانونی ممکن است بر مرز دامنه‌ی $C_{YY}^{1/2}$ و $C_{XX}^{1/2}$ قرار گیرد به طوری که زوج متغیرهای کانونی در فضاها در نظر گرفته شده وجود نداشته باشد زیرا ممکن است سوپریم در تعریف ρ_k یک ماکسیم نباشد. هی و همکاران (۲۰۰۳) کوشش کردند تا این مشکل را با اعمال محدودیت‌های غیر ضروری بر مقادیر منفرد عملگر $R = C_{XX}^{-1/2} C_{XY} C_{YY}^{-1/2}$ بر طرف سازند اما چون این شرایط همیشه برقرار نیست بنابراین همبستگی‌های کانونی معمولاً وجود ندارند (ایوبنک و هسینگ، ۲۰۰۸). بالاخره کوپرسانین و همکاران (۲۰۱۰) تعریفی از FCCA بر اساس نمایش فضای هیلبرت با هسته‌ی بازآفرین از یک فرآیند تصادفی ارائه می‌کنند که مفهوم کلاسیک CCA را به طور موفقیت‌آمیزی به حالت تابعی تعمیم می‌دهد و بدین ترتیب همه‌ی مشکلات موجود را برطرف می‌سازد.

تعریف: برای فرآیندهای تصادفی $X \in L_T(T_1)$ و $Y \in L_T(T_2)$ توابع کوواریانس به صورت زیر تعریف می‌شود:

$$\begin{aligned} c_{XX}(s, t) &= Cov\{X(s), X(t)\}, \quad s, t \in T_1, \\ c_{YY}(s, t) &= Cov\{Y(s), Y(t)\}, \quad s, t \in T_2, \\ c_{XY}(s, t) &= Cov\{X(s), Y(t)\}, \quad s \in T_1, t \in T_2. \end{aligned}$$

همچنین عملگرهای کوواریانس $C_{XX} : L_T(T_1) \rightarrow L_T(T_1)$ ، $C_{YY} : L_T(T_2) \rightarrow L_T(T_2)$ و $C_{XY} : L_T(T_2) \rightarrow L_T(T_1)$ به صورت زیر تعریف می‌شوند:

$$\begin{aligned} C_{XX}u(s) &= \int_{T_1} c_{XX}(s, t)u(t)dt, \quad u \in L_T(T_1), \\ C_{XY}u(s) &= \int_{T_2} c_{XY}(s, t)u(t)dt, \quad u \in L_T(T_2), s \in T_1, \\ C_{YY}u(s) &= \int_{T_2} c_{YY}(s, t)u(t)dt, \quad u \in L_T(T_2). \end{aligned}$$

لازم به ذکر است که عملگرهای C_{XX} و C_{YY} فشرده، هیلبرت اشمیت، متقارن و معین نامنفی و عملگر C_{XY} فشرده و هیلبرت اشمیت است.

که در آن داریم:

$$\begin{aligned}\widehat{Cov}_{XX}(s, t) &= \widehat{Cov}\{X(s), X(t)\} = \\ &= \frac{1}{n-1} \sum_{i=1}^n \{X_i(s) - \bar{X}(s)\} \{X_i(t) - \bar{X}(t)\}, \\ & \quad s, t \in T_1,\end{aligned}$$

$$\begin{aligned}\widehat{Cov}_{YY}(s, t) &= \widehat{Cov}\{Y(s), Y(t)\} = \\ &= \frac{1}{n-1} \sum_{i=1}^n \{Y_i(s) - \bar{Y}(s)\} \{Y_i(t) - \bar{Y}(t)\}, \\ & \quad s, t \in T_2,\end{aligned}$$

$$\begin{aligned}\widehat{Cov}_{XY}(s, t) &= \widehat{Cov}\{X(s), Y(t)\} = \\ &= \frac{1}{n-1} \sum_{i=1}^n \{X_i(s) - \bar{X}(s)\} \{Y_i(t) - \bar{Y}(t)\}, \\ & \quad s \in T_1, t \in T_2,\end{aligned}$$

که در آن $\bar{X}(\cdot)$ و $\bar{Y}(\cdot)$ توابع میانگین نمونه هستند. همچنین تابع همبستگی^۴ برای X برابر است با:

$$\widehat{Corr}_X(t_1, t_2) = \frac{\widehat{Cov}_X(t_1, t_2)}{\sqrt{\widehat{Var}_X(t_1)\widehat{Var}_X(t_2)}}, \quad (1)$$

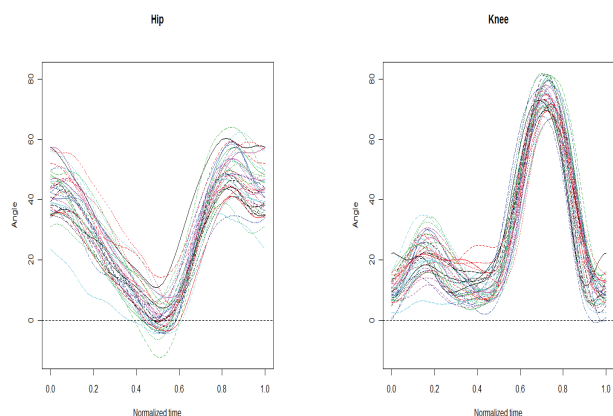
و تابع همبستگی متقاطع $X(t)$ و $Y(t)$ به صورت زیر تعریف می‌شود:

$$\begin{aligned}\widehat{r}_{XY}(t_1, t_2) &= \widehat{Corr}\{X(t_1), Y(t_2)\} \\ &= \frac{\widehat{Cov}\{X(t_1), Y(t_2)\}}{\sqrt{\widehat{Var}\{X(t_1)\}\widehat{Var}\{Y(t_2)\}}}.\end{aligned}$$

۴ تحلیل یک مجموعه داده‌ی واقعی

داده‌های مورد استفاده شامل زوایای به‌وجود آمده از ران و زانو در یک چرخه قدم زدن، برای ۳۹ کودک است. اطلاعات آن از لابراتوار تحلیل حرکت در بیمارستان سانتیاگو جمع‌آوری شده است. زمان برای هر چرخه به گونه‌ای مقیاس‌بندی شده که هر منحنی، مقادیر t (زمان) خود را در بازه‌ی $[0, 1]$ اختیار می‌کند. چرخه‌ی قدم زدن که شامل برداشتن دوگام متوالی برای هر کودک است به صورت شبکه‌ی منظمی از ۲۰ نقطه گسسته سازی شده است. مراحل چرخه قدم زدن به صورت زیر در نظر گرفته می‌شود: هنگامی که پاشنه پا به زمین اصابت می‌کند چرخه آغاز می‌شود و متناظر با آن ران خم می‌شود و وزن روی پا می‌افتد و به دنبال آن ران قائم می‌شود، زمانی که ران صاف شود نیمه‌ای از چرخه طی شده است. در نیمه دوم چرخه ران متمایل می‌شود، پا برداشته می‌شود و

جلوتر به زمین برخورد می‌کند، که در آن پا به سمت جلو تاب می‌خورد. البته داده‌ها در واقعیت به طور کامل دوره‌ای نیستند اما در نظر گرفتن آن به صورت دوره‌ای منطقی به نظر می‌رسد. برای جزئیات بیشتر می‌توان به اولشن و همکاران (۱۹۸۹) مراجعه کرد.



شکل ۱. زوایای ران و زانو ۳۹ کودک در یک چرخه قدم زدن.

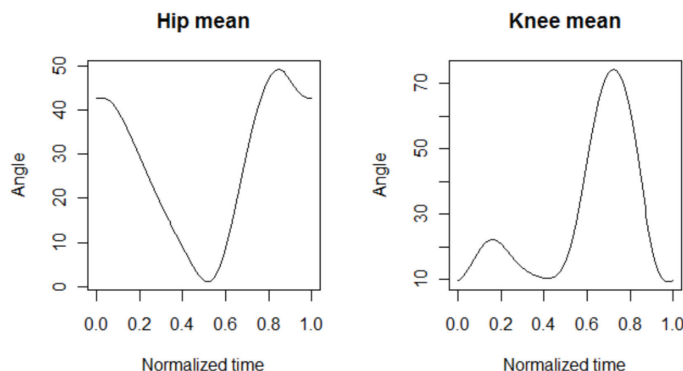
همان‌طور که قبلاً بدان اشاره شد، چون داده‌ها به متغیر زمان وابسته هستند برای تحلیل آن از تحلیل داده‌های تابعی استفاده شده است. شکل ۱ زوایای ران و زانو ۳۹ کودک را نشان می‌دهد. این منحنی‌های پیوسته با استفاده از یک سری فوریه با ۱۴ جمله به مشاهدات گسسته برازش شده‌اند. با مشاهده این نمودار می‌توان گفت که منحنی زانو دارای روندی دو فازی است درحالی که ران حرکت تک فازی دارد. یافتن ارتباط این دو به یکدیگر از روی منحنی‌های ران و زانو به تنهایی امکان‌پذیر نیست، زیرا جفت کردن منحنی‌های مربوط به زاویه زانو با منحنی‌های متناظر آن که مربوط به زاویه ران است از روی شکل میسر نمی‌شود.

شکل ۲ چرخه قدم زدن یک کودک و میانگین ۳۹ کودک را نشان می‌دهد که در آن زاویه زانو در مقابل زاویه ران که هر دو نسبت به خط قائم اندازه‌گیری شده‌اند، رسم شده است. با مشاهده منحنی می‌توان گفت بسته بودن منحنی، دوره‌ای بودن قدم زدن را نشان می‌دهد. همچنین همسان بودن فرم زاویه بین زانو و ران یک کودک با میانگین کودکان، ارتباط میان آن‌ها را مشخص می‌کند. این شکل نشان می‌دهد کودکی که قدم زدن آن با منحنی ممتد نشان داده شده در دو قسمت با میانگین متفاوت است، اولین قسمت بخش $C-D$ و دومین قسمت بخشی از چرخه است که ران به بیشترین مقدار خود خم می‌شود که در این نقطه زاویه ران به طور اساسی کمتر از میانگین مشاهده می‌شود. اما این مساله در مورد زانو

^۴Correlation function

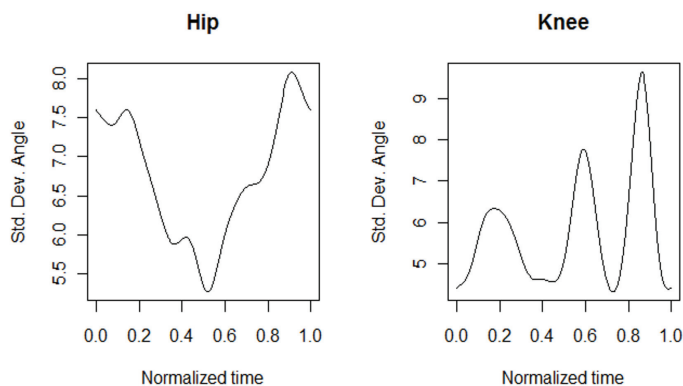
صادق نیست (رمسی و سیلورمن ۲۰۰۵).

استفاده نمود (برای جزییات بیشتر در مورد GCV و اندازه‌ی درجات آزادی به گو، ۲۰۰۲ و برای بحث هموارسازی به رمسی و سیلورمن، ۲۰۰۵ مراجعه نمایید). نمودار سمت راست شکل ۳ نشان می‌دهد که در این سطح از همواری، اندازه‌ی درجات آزادی مقدار ۱۳ دارد که از عدد ۱۴ که به عنوان تعداد توابع پایه برای برازش منحنی‌های پیوسته به داده‌های گسسته استفاده شد اختلاف معنی‌داری ندارد.



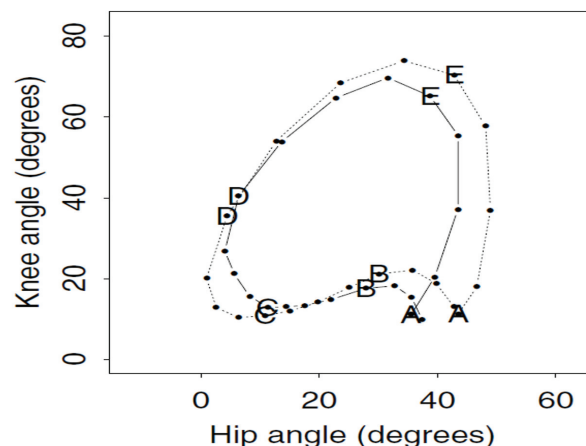
شکل ۲. تابع میانگین برای زوایای ران و زانوی کودک در یک چرخه قدم زدن.

شکل ۴. تابع میانگین برای زوایای ران و زانوی کودک را نشان می‌دهد. بر اساس این شکل، میانگین زاویه‌ی ران نسبت به زانو مقادیر کوچکتری اتخاذ می‌کند. همچنین هر دو تابع در میانه‌ی چرخه تغییر جهت می‌دهند.



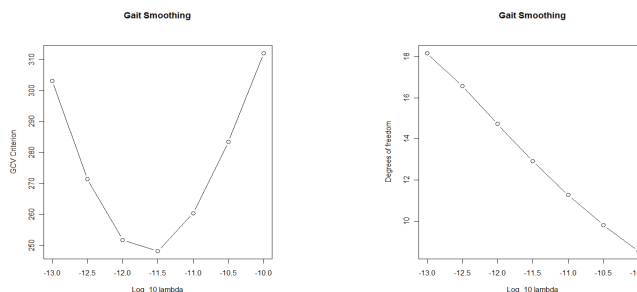
شکل ۳. نمودار سمت چپ: مقادیر GCV بر حسب $\log_{10}(\lambda)$ برای هموارسازی منحنی‌های برازش داده شده به داده‌های زوایای ران و زانوی ۳۹ کودک در یک چرخه قدم زدن. نمودار سمت راست: مقادیر اندازه‌ی درجات آزادی بر حسب $\log_{10}(\lambda)$ برای این داده‌ها که توسط گو (۲۰۰۲) پیشنهاد شده است را نشان می‌دهد.

شکل ۵. زوایای ران و زانوی ۳۹ کودک در یک چرخه قدم زدن. شکل ۵. تابع انحراف استاندارد برای زوایای ران و زانوی ۳۹ کودک را نشان می‌دهد. بر اساس این شکل، مقادیر انحراف استاندارد زاویه‌ی ران بین ۳/۵ و ۱/۸ در نوسان است و در میانه‌ی چرخه کمترین مقدار را اتخاذ می‌کند. همچنین مقادیر انحراف استاندارد زاویه‌ی زانو بین ۴/۴ و ۵/۹ در نوسان است. علاوه بر آن تغییرات زاویه‌ی زانو نسبت به ران در سراسر چرخه بیشتر است.



شکل ۴. خط توپر: زوایای ران که در مقابل زانو برای یک کودک رسم شده است، خط منقطع: نمودار مرتبط با میانگین زوایای ران و زانوی ۳۹ کودک در یک چرخه قدم زدن.

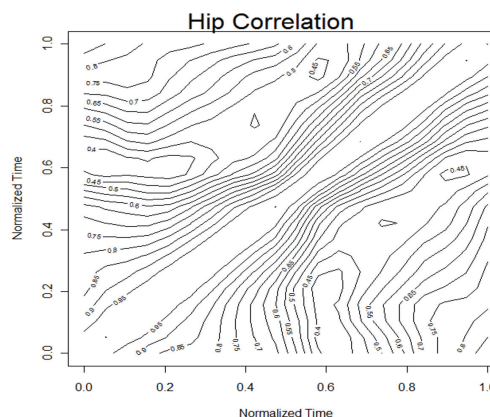
گو (۲۰۰۲) جزییاتی از بحث‌های نظری و محاسباتی در مورد انتخاب پارامتر همواری (λ) برای برازش منحنی‌های پیوسته‌ی هموار شده به نقاط گسسته‌ی مشاهده شده با استفاده از معیار همبستگی متقاطع تعمیم یافته (GCV) ارائه داد.



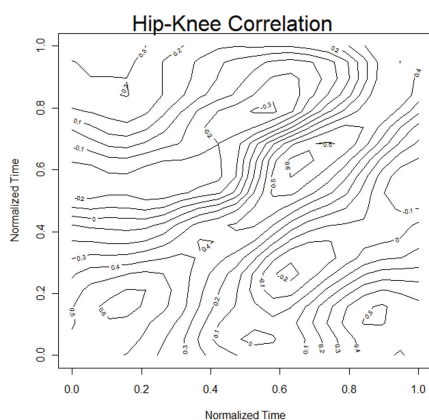
شکل ۳. نمودار سمت چپ: مقادیر GCV بر حسب $\log_{10}(\lambda)$ برای هموارسازی منحنی‌های برازش داده شده به داده‌های زوایای ران و زانوی ۳۹ کودک در یک چرخه قدم زدن. نمودار سمت راست: مقادیر اندازه‌ی درجات آزادی بر حسب $\log_{10}(\lambda)$ برای این داده‌ها که توسط گو (۲۰۰۲) پیشنهاد شده است را نشان می‌دهد.

شکل ۳ تغییرات آماره اعتبار سنجی متقابل را براساس مقیاس $\log_{10}(\lambda)$ نشان می‌دهد. همانطور که می‌بینیم GCV در $\log_{10}(\lambda) = -11.5$ - مینیمم شده است. چون هم تعداد توابع پایه که در برازش منحنی‌های پیوسته به داده‌های گسسته استفاده شده و هم مقدار λ می‌تواند باعث ایجاد همواری شود بنابراین برای جلوگیری از همواری بیش از حد می‌توان از اندازه‌ی درجات آزادی که توسط گو (۲۰۰۲) پیشنهاد شد

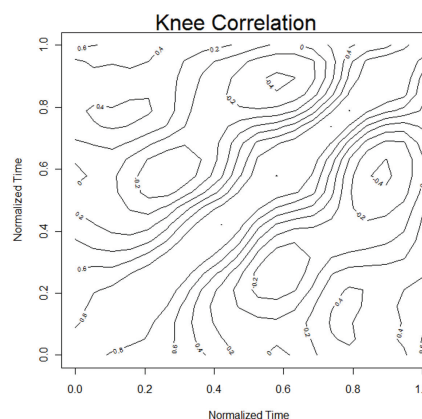
نشان می‌دهد نمی‌تواند مانا باشد، زیرا همبستگی بین زاویه در $t_1 = 0$ و $t_2 = 0/3$ تقریباً $0/4$ است در حالی که این همبستگی برای $t_1 = 0/3$ و $t_2 = 0/6$ مقداری منفی ($-0/2$) است. در میانه‌ی این چرخه وقتی که از قطر اصلی دور می‌شویم همبستگی به سرعت کاهش می‌یابد، در حالی که در انتهای این چرخه دامنه‌ی همبستگی‌ها طولانی‌تر است. این ران یک انحراف جزئی و نامحسوس از مانایی از این نوع دارند. این ویژگی‌ها ممکن است به اثر بزرگتر بر عوامل خارجی زانو از قبیل برخورد، پاشنه‌ی پا و وزن مربوطه که بر هر دوی آن‌ها قرار می‌گیرد مربوط باشد، همچنین ران تحت کنترل ماهیچه‌ای بیشتری در سراسر چرخه قرار دارد.



شکل ۶. تابع تراز برای همبستگی زوایه‌های ران ۳۹ کودک در یک چرخه قدم زدن.



شکل ۸. تابع تراز برای همبستگی مقاطع بین زوایه‌ی ران و زوایه‌ی زانوی ۳۹ کودک در یک چرخه قدم زدن.



شکل ۷. تابع تراز برای همبستگی زوایه‌های زانوی ۳۹ کودک در یک چرخه قدم زدن.

شکل ۸ تابع تراز همبستگی مقاطع بین $Hip(t_1)$ و $Knee(t_2)$ را نشان می‌دهد. در این شکل محور عمودی مقادیر t_1 و محور افقی مقادیر t_2 را مشخص می‌کنند. لبه‌ای که در طول قطر اصلی این منحنی تراز قرار دارد از وجود همبستگی‌های بزرگتر بین ران و زانو وقتی که t_1 و t_2 تقریباً برابرند حکایت دارد. هنگامی که از قطر اصلی دور می‌شویم این نمودار تأخیری را در یکی از این دو (ران یا زانو) نسبت به دیگری نشان می‌دهد. صرف‌نظر از این مطلب، تفاوت‌هایی در رفتار همبستگی‌های مقاطع در نقاط مختلف چرخه وجود دارد. به عنوان مثال وقتی که $0/2 < t_1 < 0/5$ ، همبستگی‌های مقاطع برای مقادیر مختلف t_2 ابتدا از $0/5$ شروع و به $0/6$ می‌رسد و سپس به $0/5$ و از آنجا به $0/4$ می‌رسد و در ادامه وقتی $0/6 < t_2 < 0/35$ ، ابتدا به $0/3$ و سپس به تدریج کاهش می‌یابد. وقتی که $0/6 < t_2 < 1$ ، مقدار همبستگی مقاطع از $0/1$ شروع و به تدریج در انتهای این بازه‌ی زمانی به $0/5$ می‌رسد.

شکل‌های ۶ و ۷ به ترتیب توابع تراز همبستگی برای داده‌های زوایای ران و زانو را نشان می‌دهند. در هر نمودار روی محور افقی مقادیر t_1 و روی محور عمودی مقادیر t_2 مشخص شده‌اند. توابع تراز برای $Corr_{Hip}$ و $Corr_{Knee}$ هر کدام به تنهایی الگوهای متفاوتی از تغییرپذیری برای زوایای ران و زانو را نشان می‌دهند. به طور کلی، $Corr_{Hip}$ همبستگی‌های مثبت را نشان می‌دهد به طوری که اگر زاویه‌ی ران در یک نقطه از چرخه بزرگتر از متوسط باشد این زاویه تمایل خواهد داشت تا در هر نقطه‌ی دیگر نیز بزرگتر از متوسط باشد. خطوطی که روی این منحنی تراز قرار گرفته‌اند به طور تقریباً موازی با قطر اصلی قرار دارند که دلالت بر این نکته دارد که همبستگی تقریباً تابعی از $t_1 - t_2$ است و تغییرات در زاویه‌های ران را می‌توان به عنوان یک فرآیند مانای تقریبی در نظر گرفت. از طرف دیگر، شکل ۷ که تغییرات زاویه‌ی زانو را

آن استفاده نمود:

$$\langle u, \hat{C}_{XY} v \rangle \approx \frac{1}{400} \sum_{k=1}^{400} \sum_{l=1}^{400} u\left(\frac{k}{400}\right) \hat{C}_{XY}\left(\frac{k}{400}, \frac{l}{400}\right) v\left(\frac{l}{400}\right).$$

لازم به ذکر است که در فرمول بالا چرخه به طول ۱ در نظر گرفته شده است. همچنین تمام همبستگی‌های نمونه‌ای هموار شده و غیرهموار به وسیله‌ی همین تقریب محاسبه شده است. تاوان همواری گسسته‌سازی شده به وسیله عبارت زیر تقریب زده می‌شود:

$$[u, u] \approx 8000 \sum_{k=1}^{400} \left\{ u\left(\frac{k-1}{400}\right) - 2u\left(\frac{k}{400}\right) + u\left(\frac{k+1}{400}\right) \right\}^2$$

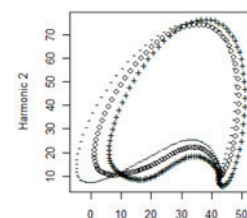
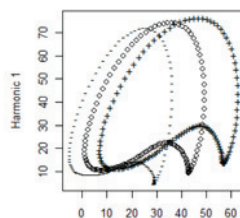
که در آن $u(0) = u(1)$ و $u(0.5) = u(0.5)$ که با فرض دوره‌ای بودن برقرار هستند. ثابت ۸۰۰۰ تضمین می‌کند که تقریب $[u, u] \approx \int u''^2$ برای توابع پیوسته متناوب هموار u برقرار باشد.

شکل ۱۰ توابع u و v را که همبستگی نمونه‌ای غیرهموار را ماکسیم می‌کند نشان می‌دهد. همبستگی نمونه‌ای بدست آمده به وسیله این توابع یک است. بردارهای نشان داده شده در این شکل اطلاعات معناداری در خصوص داده‌ها به دست نمی‌دهند و به طور واضح لزوم استفاده از هموارسازی را نشان می‌دهد. لازم به ذکر است که این رفتار تنها مختص این داده‌ها نبوده بلکه یک خصوصیت ذاتی از تحلیل همبستگی کانونی تابعی است. توابع وزنی اولین و دومین همبستگی کانونی هموار شده در شکل ۱۱ رسم شده است. نمودار سمت چپ از این شکل اولین متغیرهای کانونی هموار شده $\hat{u}_\alpha$ و $\hat{v}_\alpha$ را نشان می‌دهد که در آن پارامتر همواری به وسیله روش اعتبارسنجی متقابل که قبلاً بدان پرداخته شده، بدست آمده است (برای بحث هموارسازی در تحلیل همبستگی کانونی تابعی و چگونگی بدست آوردن برآوردهای هموار به گلشن و همکاران، ۱۳۸۸ مراجعه نمایید). از آنجایی که هدف اصلی، مقایسه‌ی منحنی‌ها است از نرمال شده‌ی آن‌ها استفاده کرده‌ایم بنابراین انتگرال توان دوم آن‌ها یک خواهد شد.

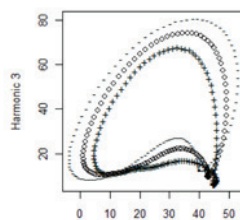
جدول ۱. جدول مقادیر توان دوم همبستگی کانونی هموار شده و نشده برای داده‌های زوایای ران و زانوی ۳۹ کودک در یک چرخه قدم زدن.

متغیرهای کانونی	هموار نشده	هموار شده
اولین متغیر	۰/۷۶	۰/۸۲
دومین متغیر	۰/۶۲	۰/۷۲
سومین متغیر	۰/۱۵	۰/۲۰

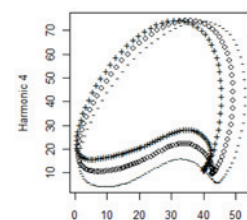
PCA function 1 (Percentage of variability 46.4) PCA function 2 (Percentage of variability 19.6)



PCA function 3 (Percentage of variability 11.8)



PCA function 4 (Percentage of variability 8.5)

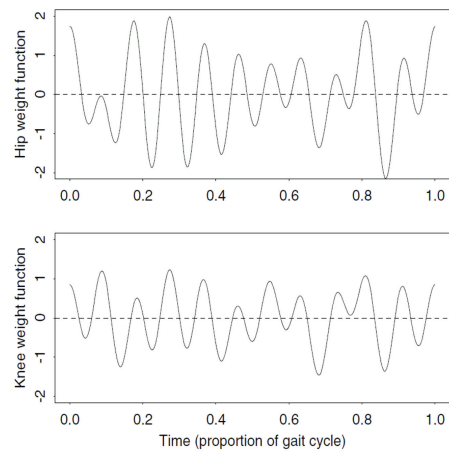


شکل ۹. نمودار اولین چهار تابع مؤلفه‌ی اصلی برای داده‌های زوایای ران و زانوی ۳۹ کودک در یک چرخه قدم زدن. در این نمودارها منحنی نقطه‌ای میانگین، منحنی‌های (+) و (-) به ترتیب اثر اضافه و کم کردن ضریبی از مؤلفه‌ی اصلی مربوطه به میانگین را نشان می‌دهند.

حال به تحلیل همبستگی کانونی داده‌های قدم زدن می‌پردازیم. تحلیل همبستگی کانونی را می‌توان به عنوان مکملی از تحلیل مؤلفه‌ی اصلی در نظر گرفت. برای تحلیل داده‌های قدم زدن می‌توان از تحلیل مؤلفه‌های اصلی استفاده کرد. رایس و سیلورمن (۱۹۹۱) با استفاده از تحلیل مؤلفه‌ی اصلی دو متغیره دریافتند که اولین عامل تغییرات تقریباً به یک پرشیدگی ثابت از زاویه‌ی ران در سراسر چرخه به گونه‌ای که زاویه زانو به ندرت تحت تأثیر قرار می‌گیرد، مرتبط است. دومین عامل تغییرات، عاملی است که در آن چرخه‌ی بوجود آمده در فضای ران-زانو بدون تغییر باقی می‌ماند (شکل ۹ را ببینید). برای جزئیات بیشتر در مورد تحلیل و تفسیر این مؤلفه‌های اصلی به رایس و سیلورمن (۱۹۹۱) یا سیلورمن و سیلورمن (۲۰۰۵) مراجعه کنید.

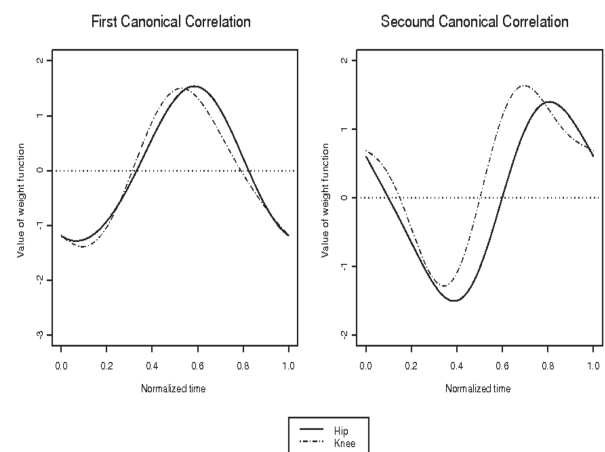
خواهیم دید که تحلیل همبستگی کانونی جوانب مختلفی از تغییرپذیری را عرضه خواهد کرد. باید توجه داشته باشیم که اولین عامل تغییرات بدست آمده توسط رایس و سیلورمن (۱۹۹۱)، به ندرت با اثرات متقابل متغیرها مرتبط است. در عامل دوم تغییرات، همبستگی قوی میان دو زاویه در هر نقطه‌ی زمانی وجود دارد، زیرا اگر یکی در چرخه بیشتر شود دیگری نیز به همین صورت عمل خواهد کرد. برای محاسبه‌ی مقدار تقریبی کوواریانس و واریانس نمونه‌ای، یعنی $\langle u, \hat{C}_{XY} v \rangle$ ، می‌توان مقدار آن را برای نقاط روی شبکه محاسبه و از فرمول زیر برای تقریب

شکل ۱۱. نمودار دو تابع وزنی همبستگی کانونی اول هموار شده برای داده‌های زوایای ران و زانوی ۳۹ کودک در یک چرخه قدم زدن.



شکل ۱۰. نمودار دو تابع وزنی همبستگی کانونی اول غیر هموار برای داده‌های زوایای ران و زانوی ۳۹ کودک در یک چرخه قدم زدن.

مقادیر توان دوم همبستگی غیر هموار و هموار شده در جدول ۱ آورده شده است. همچنین این جدول این دو مقدار را برای جفت‌های دوم و سوم از متغیرهای کانونی نشان می‌دهد. تفسیر کلی این متغیرها آن است که بین دو زاویه در هر زمان همبستگی وجود دارد. اما نکته‌ی قابل توجه آن است که بیشترین مقدار در منحنی مربوط به ران در میانه‌ی چرخه کمی بعد از منحنی مربوط به زانو اتفاق می‌افتد، در شرایطی که ترتیب بیشترین مقدار نزدیک آغاز چرخه در جهت عکس است. پس می‌توان گفت تغییرات زیاد ران باعث تغییرات زیاد در زانو می‌شود. نزدیک انتهای چرخه، این اثرها در جهت مختلف رخ داده است.



مراجع

[۱] تازیکه‌میاندره، ن.، خیراله زاده، ن.، موقری، ه. و حسینی نسب، م.ا. (۱۳۸۷)، تحلیل داده‌های تابعی: تعمیمی از تحلیل داده‌های چند متغیره،

مجله اندیشه‌ی آماری، ۲۵، ۶۸-۵۲.

- [۲] گلشن، آ.، حسینی نسب، م.ا. و فرید روحانی، م. (۱۳۸۸)، بررسی الگوهای دما، رطوبت و بارندگی در ایران با استفاده از تحلیل همبستگی کانونی، *مجله پژوهش‌های آماری ایران*، ۶، ۹۳-۱۱۹.
- [3] Cupidan, J., Eubank, R. Gilliam, D., and Ruymgaart, F. (2008). Some properties of canonical corelations and variates in infinite dimensions, *J. Multivariate Anal.* **99**, 1083-1104.
- [4] Dauxois, J., Nkiet, G. and Romain Y. (2004). Canonical analysis relative to a closed subspaces, *Linear Algebra Appl.* **388**, 119-145.
- [5] Dauxois, J. and Pousse, A. (1976). une extension de l'analyse canonique quelques applications, *Ann. Inst. H. Poincare Probab. Statist.* **11**, 355-379.
- [6] Eubank, R. and Hesing T. (2008). Canonical correlation for stochastic processes, *Stochastic Processe and their Applications* **118**, 1634-1681.
- [7] Gu, C. (2002). Smoothing Spline ANOVA MOdels. Springer, New York.
- [8] He, G. and Müller, H.-G. and Wang, J.-L. (2003). Functional canonical analysis for square integrable stochastic processes. *Multivariate Analysis.*, **85**, 54-77.
- [9] Jolliffe, I.T. (2002). *Principal Component Analysis*, 2nd Ed., Springer, New York.
- [10] Kupresanin, A., Shin, H., King D. and Eubank, R. (2010). An RKHS frmework for functional data analysis, *Journal of Statistical Planing and Infrence*, **85**, 3627-3637.
- [11] Leurgans, S.E. and Moyeed, R.A. and Silverman, B.W. (1993). Canonical corelation analysis when data are curves. *J.R. statist. Soc. B*, **55**, 725-740.
- [12] Mochon, S. and McMahon, T.A. (1980). Ballistic walking: an improved model. *Math. Biosci*, **52**, 241-260.
- [13] Philips, S.J., Robert, E.M. anf Huang, T.C. (1983). Quantification of intersegmental reactions during rapid swing motion. *J. Biomech.*, **16**, 411-417.
- [14] Ramsay, J.O. (1982). When the Data are Functions, *Psychometrika*, **47**, 4, 379-96.
- [15] Ramsay, J.O., Silverman, B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York.
- [16] Ramsay, J. O., Silverman, B. W. (2005). *Functional Data Analysis, 2nd edition*, Springer, New York.
- [17] Ramsay, J.O. (1982). When data are functions. *Psychometrika*, **47**, 4, 379-396.
- [18] Rencher, A.C. (2002). Method of Multivariate Analysis. *John Wiley and Sons, Inc. 2nd edition*.
- [19] Rice, J.A. and Silverman, B.W. (1991). Estimating the mean and covariance strucure nonparametrically when the data are curves. *J. Roy. statist. Soc. B*, **53**, 233-243.