دانشگاه شهید بهشتی

دانشکده علوم ریاضی

پایان‌نامه دوره کارشناسی ارشد آمار محض

آزمونهای نیکویی برازش در فضاهای نامتناهی‌بعد

نگارنده

سارا صفوی

استاد راهنما

دکتر محمدرضا فریدروحانی

استاد مشاور

دکتر مجتبی خزائی

تابستان ۱۳۸۸

تقدیم به
پیشگاه لایزال یگانه‌اش

مادر و پدر بزرگوارم

و همه کسانی که دوستشان دارم.

قدردانی و تشکر

آغازین سخن کلام زیبای وحی است که: علّمَ الانسان ما لم یَعلم و بدین‌سان خداوند عالم، خود را «معلم» انسان دانسته و گوید که: «ای انسان من به تو آموختم آنچه را که نمی‌دانستی» و چه خوب و با سعادت در مسیر خدایند آنانکه معلم‌اند و از خرمن دانش خود تشنه‌گان را سیراب و خسته‌گان را توان و سرگردانان وادی علم را «نجات» می‌بخشند.

دانش آموختگان را فرض است که با تأسی به کلام گهربار امیرالمؤمنین علی(ع): من عَلّمَنی حرفاً فقد سَیّرَنی عبداً «هر کس کلامی به من بیاموزد، مرا بنده و عبد خود گردانیده است»، آنرا جامهٔ عمل پوشانند؛ لذا به مصداق این بیان پرنغز امیرالمؤمنین علی(ع) همیشه بر خود واجب می‌دانم که از پدر و مادر عزیز و معلمین بزرگوار خود در طول تحصیل که انصافاً چشمهٔ دانش‌دوستی و دانش‌خواهی را در وجودم روشن نگه داشته‌اند و نیز از کلیهٔ اساتید محترم دانشگاه شهید بهشتی که از فرد فرد آنان دانش و خرد فرا گرفته و همچنین از داوران بزرگوار و محترم آقایان دکتر مجتبی گنجعلی و دکتر محمد ابراهیم حسینی‌نسب که زحمت داوری این پایان نامه را متقبل گردیده‌اند و نیز از استاد محترم و بزرگوار جناب آقای دکتر مجتبی خزایی که رهنمودهای ارزنده‌ای ارائه فرموده و نیز علی‌الخصوص استاد راهنمای زحمتکش و بزرگوار جناب آقای دکتر محمدرضا فریدروحانی که با صرف وقت پر و متانت و سعهٔ‌صدر، روش دانش‌پژوهی و دانش‌آموزی را به ما آموختند و نیز از استاد والامقام آقای دکتر پارسیان که اجازهٔ بهره‌مندی از خرمن دانش خود را در دانشگاه تهران بر من گشوده و سرکار خانم دکتر مونا نبیعی که با ما همگامی و همراهیهای فراوان نمودند، صمیمانه تشکر و قدردانی نموده و سعادت و بهروزی را برای همهٔ آنان آرزومندم.

چکیده

پیشرفتهای اخیر در حوزهٔ فن‌آوری به ما اجازه می‌دهند تا بتوانیم داده‌های بیشتری را در یک دوره زمانی ثبت کنیم. داده‌های بدست آمده در علوم پزشکی، مهندسی، علوم اجتماعی و امور مالی اغلب داده‌های تابعی هستند. از آنجا که برخی از تحلیل‌های آماری مطالعه شده در داده‌های تابعی، بر اساس مفروضات خاص توزیعی استوارند، آزمون نیکویی برازش چنین داده‌هایی از اهمیت ویژه‌ای برخوردار است. رهیافت اصلی مورد استفاده در این پایان‌نامه مبتنی بر روش تصویر تصادفی است. این روش تکنیکی است برای تصویر مجموعه‌ای از نقاط از فضایی با بعد بالا بر زیر فضایی با بعد کوچکتر، که بطور تصادفی انتخاب می‌شود. بدین معنی که اگر با مجموعه‌ای از مشاهدات dبعدی سروکار داشته باشیم، به کمک این شیوه به حل مسأله در فضای kبعدی (k در مقایسه با d بسیار کوچک است) پرداخته و نهایتاً با استناد به قضایایی که در این پایان‌نامه معرفی خواهند شد نتایج بدست آمده را به فضای dبعدی اصلی تبدیل می‌کنیم. می‌توان از این رهیافت برای آزمون نیکویی برازش داده‌های با بعد بالا و به ویژه نامتناهی‌بعد نیز بهره گرفت. بدین منظور قضایایی به تفصیل در این پایان‌نامه مورد بررسی قرار گرفته است.

کلمات کلیدی: تصویر تصادفی، فضای هیلبرت، آزمونهای نیکویی برازش، آزمون تصویرشدهٔ کولموگوروف—اسمیرنوف، آزمون کرامر فون—میزس.

پیشگفتار

پیشرفتهای اخیر در تکنولوژی به طرز چشمگیری به ما اجازه می‌دهند که بتوانیم داده‌های بیشتری را در یک دوره زمانی ثبت کنیم. در آمار که علم مدل‌بندی و پیش‌بینی پدیده‌های تصادفی است عموماً داده‌های بدست آمده از عناصر تصادفی مقادیری عددی، برداری و یا ماتریسی هستند. اما داده‌های بدست آمده درعلوم پزشکی، مهندسی، علوم اجتماعی و امورمالی اغلب داده‌های تابعی هستند. اغلب داده‌های تابعی را می‌توان مقادیری در یک فضای هیلبرت در نظر گرفت. بررسی اینگونه داده‌ها، در حقیقت بکاربردن روشهای آماری در فضای مشاهدات تابعی و یا به عبارتی همان فضای مشاهدات بی‌نهایت‌بعدی است. در سراسر این پایان‌نامه فرض بر این است که داده‌ها، نمونه‌ای مستقل و هم‌توزیع از یک فرایند تصادفی هستند که مقادیر خود را در یک فضای هیلبرت اختیار می‌کنند.

برای مثال می‌توان نمونه تصادفی $\{X_1(t), \ldots, X_n(t) : t \in T\}$ را در نظر گرفت که T بازه‌ای در $\mathcal{R}$ و داده‌ها مقادیر خود را در فضای هیلبرت $L^2(T)$ اختیار می‌کنند. با توسل به فرایندهای ریاضی به منظور بررسی و کاوش در چنین داده‌هایی دو مسأله را پیش رو خواهیم داشت.

۱) مسایل بسیاری وجود دارند که حل آنها از دیدگاه نظری شناخته‌شده بوده ولی حتی برای داده‌هایی از فضای با بعد متناهی، تحقق مسأله در عمل با مشکل روبرو است.

۲) مسائلی وجود دارند که بررسی نظری و عملی آنها برای داده‌های با بعد متناهی شناخته‌شده است ولی برای داده‌های تابعی ناشناخته است.

رهیافت پیشنهادشده دراین پایان‌نامه مبتنی بر روش تصویرسازی تصادفی است. بطور خلاصه این فرایند می‌تواند بصورت زیرتوصیف گردد:

فرض کنید با مسأله‌ای مرتبط با شئ‌ای d—بعدی سروکار داریم. روش تصویرسازی تصادفی شامل انتخاب —بطورتصادفی— زیرفضایی k—بعدی (k درمقایسه با d بسیارکوچک است) است که مسأله مورد نظر در این زیرفضا حل شده و پس ازآن پاسخ بدست آمده به فضای اصلی (d—بعدی) برگردانده می‌شود. بسیاری از کاربردهای این روش بر این حقیقت استوارند که تقریب تصاویر تصادفی، فواصل دوبدو را با احتمال بالایی حفظ می‌کند. نکتهٔ شگفت‌انگیز آنکه گرچه آمار در قلب روش تصویرسازی تصادفی قرار دارد، اما ایدهٔ استفاده از این روش بندرت در مسائل آماری بکاررفته است. اولین نگاه به این روش با انگیزه آماری توسط گوستا آلبرتوز و همکاران ($۲۰۰۶a$) صورت گرفته است. هدف آنها تحلیل این مسأله بوده است که تصاویر تصادفی تا چه حد می‌توانند یک توزیع احتمال را تعیین کنند؟! قضایای مطرح‌شده حاکی از آن است که یک تصویر که بطور تصادفی انتخاب شده، تعیین‌کننده یک توزیع خواهدبود.

از آنجا که بسیاری از تحلیل‌های آماری مطالعه‌شده در داده‌های تابعی، بر اساس مفروضات خاص توزیعی، استوارند، بحث نیکویی برازش چنین داده‌هایی از اهمیت خاصی برخوردار شده است. از این نتیجه که یک تصویر تصادفی انتخاب شده، تعیین کننده یک توزیع است،

می‌توان به عنوان پایه‌ای برای شروع تحلیل آماری در بدست آوردن آزمونهای نیکویی برازش یک یا دونمونه‌ای استفاده نمود. تأکید بر این نکته حائز اهمیت است که آزمونهای مطرح شده تنها بر پایه تصاویر یک‌بعدی استوار خواهند بود که این نکته با ذکر استدلالهای لازم اثبات خواهد شد. سازماندهی این پایان‌نامه به شرح زیر خواهد بود:

در فصل اول به مرور مفاهیم کلی مرتبط با فضای هیلبرت که فضای حاکم بر این پایان‌نامه است، پرداخته شده است. در فصل دوم نیز به معرفی داده‌های تابعی و روش تصویر تصادفی پرداخته و نیز به منظور ورودی به مبحث آزمونهای نیکویی برازش در داده‌های تابعی و نامتناهی بعد، گروه مهمی از آزمونهای نیکویی برازش داده‌های متناهی‌بعد یک و چندمتغیره معرفی شده‌اند. فصل سوم به عنوان بخش اصلی پایان‌نامه به بررسی جامع آزمونهای نیکویی برازش در داده‌های نامتناهی‌بعد اختصاص یافته است. در این فصل، آزمون نیکویی برازش مبتنی بر آمارهٔ کولموگوروف—اسمیرنوف تصویرشده مطالعه شده است. در فصل چهارم آمارهٔ آزمون کرامر—فون میزس، به عنوان آماره‌ای دیگر برای بررسی آزمونهای نیکویی برازش در داده‌های تابعی مورد بررسی قرار گرفته است. فصل پنجم به شبیه‌سازی و ذکر مثالهای کاربردی اختصاص یافته است.

فهرست مندرجات

فصل ۱

پیشنیازهای ریاضی

۱.۱ مقدمه

تحلیل تابعی یکی از مهم‌ترین زمینه‌های جدید در ریاضیات قرن بیستم است. تحلیل تابعی بیش از هر کس با نام ریاضیدانی فرانسوی به نام لوی (۱۹۷۱—۱۸۸٦) گره خورده است. از جمله پیشگامان این رشته را می‌توان، ریاضی‌دانانی همچون ولترا (۱۹۴۰—۱۸٦۰)، فردهولم (۱۹۲۷—۱۸٦٦)، هیلبرت (۱۹۴۳—۱۸٦۲) و ریس (۱۹۵٦—۱۸۸۰) نام برد. ریاضیدانی لهستانی به نام باناخ (۱۹۴۵—۱۸۹۲) نیز مفاهیم توپولوژیکی را در تحلیل تابعی وارد نمود. همچنین بررسی عملگرها در فضای هیلبرت نیز مدیون مطالعات فون نیومان (۱۹۵۷—۱۹۰۳) است.

از سوی دیگر الهام‌بخش نظریهٔ تحلیل تابعی، متأثر از پیشرفت مکانیک کوانتوم در فیزیک دهه ۱۹۲۰ توسط فیزیکدانانی همچون بوهر (۱۹٦۲—۱۸۸۵)، دیراک (۱۹٦۱—۱۸۸۷)

هایزنبرگ (۷۶—۱۹۰۱) و شرودینگر (۱۹۶۱—۱۸۸۷) بوده است.

هدف اصلی در این پایان نامه، آزمون نیکویی برازش برای داده‌های تابعی است که متعلق به یک فضای هیلبرت مناسب هستند. برای نیل به این منظور در ابتدا لازم می‌دانیم که به بررسی و تعریف خصوصیات، قضایای مهم و عملگرهای موجود در اینگونه فضاها بپردازیم.

مطالب این فصل عمدتاً برگرفته از کتاب هَنسن (۲۰۰۶) است.

۲.۱ فضای برداری

مجموعه ناتهی $\mathcal{V}$ از عناصر x،y،z،... را فضای خطی (یا فضای برداری) می‌گوییم اگر در اصول موضوع زیر صدق کند:

- اصول جمع

$\mathcal{V}$ نسبت به جمع بسته باشد، یعنی برای هر $x, y \in \mathcal{V}$ عضو $x + y \in \mathcal{V}$ بطوریکتایی تعریف شده و داشته باشیم:

$i)\quad x + y = y + x$

$ii)\quad (x + y) + z = x + (y + z)$

$iii)\quad \exists \circ \in \mathcal{V}: \quad \forall x \in \mathcal{V}; \quad x + \circ = x$

$iv)\quad \forall x \in \mathcal{V} \quad \exists -x \in \mathcal{V}: \quad x + (-x) = \circ$

• اصول ضرب

مجموعه $\mathcal{V}$ نسبت به عمل ضرب اسکالر بسته باشد، یعنی برای هر $x \in \mathcal{V}$ و هر $\alpha \in \mathcal{R}$، عضو $\alpha x \in \mathcal{V}$ بطور یکتایی تعریف شده و برای هر $\alpha, \beta \in \mathcal{R}$ داشته باشیم:

$i) \quad \alpha(\beta x) = (\alpha\beta)x$

$ii) \quad 1x = x$

به علاوه عملگرهای جمع و ضرب تعریف‌شده در بالا، از قوانین زیر پیروی می‌کنند:

$i) \quad (\alpha+\beta)x = \alpha x + \beta x$

$ii) \quad \alpha(x+y) = \alpha x + \alpha y$

همچنانکه می‌دانیم، $\mathcal{R}^n$ با جمع و ضرب معمولی یک فضای برداری است. فضاهای تابعی همان فضاهای برداری هستند که عناصرشان تابع‌اند. به عنوان مثال، گیریم $V = C([a,b])$ بیانگر فضای توابع پیوسته $f : [a,b] \longrightarrow \mathcal{R}$ تعریف‌شده بر بازه بسته $[a,b]$ باشد. از آنجا که $\mathcal{V}$ یک فضای برداری بوده که عناصر آن را توابع تشکیل می‌دهند به آن یک فضای تابعی گویند.

تعریف ۱. فرض کنید $\mathcal{V}$ یک فضای برداری باشد. در این صورت ضرب داخلی تابعی از $\mathcal{V} \times \mathcal{V}$ به $\mathcal{R}$ و با نماد $<x, y>$ است که در شرایط زیر صدق می‌کند:

$i) <x, x> \geq 0 \quad , \forall x \in \mathcal{V}$

$$ii) < x,x >= 0 \quad \iff \quad x = 0 \quad , \forall x \in \mathcal{V}$$

$$iii) < x,y >=< y,x > \quad \forall x,y \in \mathcal{V}$$

$$iv) < rx,y >= r < x,y >=< x,ry > \quad \forall r \in \mathcal{R} \quad , \quad \forall x,y \in \mathcal{V}$$

و نیز $\forall x,y,z \in \mathcal{V}$ داریم:

$$v) < x,y+z >=< x,y > + < x,z > \quad \forall x,y,z \in \mathcal{V}$$

تعریف ۲. یک فضای برداری که روی آن یک ضرب داخلی تعریف شود را یک فضای ضرب داخلی می‌گویند. مجدداً واضح است که $\mathcal{R}^n$ با جمع و ضرب معمولی یک فضای ضرب داخلی است.

تعریف ۳. گیریم $\mathcal{V}$ یک فضای برداری باشد. آنگاه یک نرم روی $\mathcal{V}$ تابعی حقیقی روی $\mathcal{V}$ با نماد $||x||$ است که در ویژگی‌های زیر صدق می‌کند:

$$i) ||x|| \geq 0 \quad \forall x \in \mathcal{V}$$

$$ii) ||x|| = 0 \quad \iff \quad x = 0$$

$$iii) ||\alpha x|| = |\alpha| \, ||x|| \quad \forall \alpha \in \mathcal{R} \quad , \quad \forall x \in \mathcal{V}$$

$$iv) ||x+y|| \leq ||x|| + ||y|| \quad \forall x,y \in \mathcal{V}$$

تعریف ۴. یک فضای برداری که روی آن یک نرم تعریف شود را یک فضای برداری نرم‌دار یا به اختصار فضای نرم‌دار گویند.

همچنانکه می‌دانیم، $\mathcal{R}^n$ با نرم اقلیدسی یک فضای نرم‌دار است. توجه کنید یک ضرب داخلی می‌تواند نرمی به صورت $||x|| = \sqrt{<x,x>}$ را القا کند.

تعریف ۵(دنبالهٔ کوشی). دنبالهٔ $\{x_n\}_n$ از نقاط فضای متریک (M,d) را یک دنبالهٔ کوشی می‌نامیم، اگر به ازای هر $\epsilon > 0$، عددی طبیعی مانند n_0 چنان موجود باشد بطوریکه به ازای هر دو عدد طبیعی $n, m \geq 0$، آنگاه $d(x_n, x_m) < \epsilon$.

تعریف ٦. فضای متریک (M,d) را کامل گوییم هرگاه هر دنبالهٔ کوشی در آن همگرا باشد.

تعریف ۷. زیرمجموعهٔ W از فضای متریک (M,d)، یا بطور کلی‌تر فضای توپولوژی $(M,\mathcal{T})$، را چگال در M گوییم اگر داشته باشیم $\overline{W} = M$.

تعریف ۸. فضای برداری نرم‌داری که به‌عنوان یک فضای متریک، کامل باشد فضای باناخ نامیده می‌شود.

در فضاهای با بعد متناهی که تعریف آن در صفحات بعدی خواهد آمد هیچ مشکلی با کامل‌بودن فضا نخواهیم داشت. درحقیقت هرفضای برداری نرم‌دار متناهی‌بعد یک فضای باناخ است.

تعریف ۹. فضای هیلبرت $\mathcal{H}$، یک فضای برداری ضرب داخلی است که نسبت به نرم القاشده توسط آن ضرب داخلی، فضای باناخ باشد.

کاربردهای تحلیل تابعی اغلب نیاز به عملگرها یا نگاشتهایی خطی بر عناصر داخل فضاهای برداری نرم‌دار دارند که مقادیر خود را در فضاهای برداری نرم‌دار دیگری اختیار می‌کنند. درچنین وضعیتی یک نگاشت خطی را اغلب یک عملگر خطی و یا گاهی به تنهایی عملگر می‌نامیم.

درباره پیوستگی عملگرهای خطی بین فضاهای برداری نرم‌دار در قضیه زیر عباراتی معادل

آورده شده است:

قضیه ۱. گیریم $T: V \to W$ یک عملگر خطی بین فضاهای برداری نرم‌دار $(V, ||.||_V)$ و $(W, ||.||_W)$ باشد. آنگاه عبارات زیر معادلند:

(۱) T پیوسته است.

(۲) T در $\circ \in V$ پیوسته است.

(۳) ثابت k چنان وجود دارد که برای تمامی بردارهای واحد $x \in V$ که برای آنها $||x||_V = ۱$، داریم $||T(x)||_W \leq k$.

(۴) ثابت k چنان وجود دارد که

$$||T(x)||_W \leq k\, ||x||_V \quad \forall x \in V$$

برهان. هنسن (۲۰۰٦)، صفحه ۱۸.

تعریف ۱۰. یک عملگرخطی $T: V \to W$ بین فضاهای برداری نرم‌دار $(V, ||.||_V)$ و $(W, ||.||_W)$ را عملگر خطی کراندار می‌نامیم اگر این عملگر، گوی بستهٔ واحد $C(V) = \{x \in V : \quad ||x||_V \leq ۱\}$ در V را به روی یک مجموعه کراندار (یک گوی) در W بنگارد.

نتیجه ۱. عملگر خطی T از فضای نرم‌دار $(V, ||.||_V)$ به فضای نرم‌دار $(W, ||.||_W)$ پیوسته است اگر و تنها اگر T کراندار باشد.

تمام قضایایی که در ادامه آورده می‌شوند، برای فضای برداری $C([a, b])$ از توابع پیوسته $f: [a, b] \to \mathcal{R}$ در بازه $[a, b]$ برقرار هستند.

ابتدا یادآوری می‌کنیم که دو فضای خطی X و Y روی میدان $\mathcal{F}$ را همسنج (طول‌پای) گویند

اگر عملگری یک‌به‌یک و پوشا از X به Y وجود داشته باشد.

قضیه ۲(قضیه تکمیل).

(۱) گیریم (M,d) یک فضای متریک باشد. آنگاه فضای متریک کامل (M',d') و زیرمجموعه چگال $\widetilde{M} \subseteq M'$ $(\overline{\widetilde{M}} = M')$ چنان وجود دارند که (M,d) و $(\widetilde{M},d')$ همسنج باشند، یعنی نگاشت دوسویی مانند $T : M \longrightarrow \widetilde{M}$ چنان وجود دارد که

$$d'(Tx,Ty) = d(x,y) \qquad \forall x,y \in M.$$

(۲) گیریم $(V,||.||)$ فضای برداری نرم‌دار باشد. آنگاه فضای برداری نرم‌دار کامل (فضای باناخ) $(V',||.||')$ و زیرمجموعه چگال $\widetilde{V} \subseteq V'$ ($\overline{\widetilde{V}} = V'$) چنان وجود دارند که $(V,||.||)$ و $(\widetilde{V},||.||')$ همسنج باشند، یعنی نگاشت خطی دوسویی مانند $T : V \longrightarrow \widetilde{V}$ چنان وجود دارد که

$$||T(x)||' = ||x|| \qquad \forall x \in V.$$

تمام تکمیلهای فضای برداری نرم‌دار $(V,||.||)$ همسنجند.

برهان. هنسن(۲۰۰۶)، صفحه ۲۶.

۳.۱ فضاهای L^p

نظریهٔ فضاهای L^p نقشی پایه‌ای در کاربرد تحلیل تابعی بازی می‌کند. فضای L^p به فضای برداری نرم‌دار « توابع توان p انتگرالپذیر » بر یک فضای اندازه اشاره دارد. این فضاهای خطی خاص را فضای توابع هم می‌نامند. فضای برداری تمام توابع توان p انتگرالپذیر لبگ