

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده علوم ریاضی

گروه آمار

پایان نامه کارشناسی ارشد

مقیاس گذاری چندبعدی از دیدگاه بیزی در توزیع نرمال چند متغیره

سیده بصیرا حسینی

استاد راهنما

دکتر محمد آرشی

شهریور ۱۳۹۴

تقدیم بہ:
پدرم،

تکلیہ گاہ و حامی من در تمام مراحل زندگیم

مادرم،

کہ زندگیم را بی یون مہر و عطوفت اومی دانم
و، ہمسر،

اسطورهی زندگیم، پناہ خستگی ہام و امید بودم.

سپاس گزار می...

سپاس خدای را که سخنوران، در ستودن او بمانند و شمارندگان، شمردن نعمت‌های او ندانند و کوشندگان، حق او را گزاردن نتوانند. و سلام و مورد بر محمد و خاندان پاک او، طاهران معصوم، هم آنان که وجودمان وامدار وجودشان است. به نشانه سپاس از الطافش، دستان مهربان‌ترین بندگانش، پدر و مادر عزیزم را بوسه می‌زنم.

بدون شک جایگاه و منزلت معلم، اجل از آن است که در مقام قدردانی از زحمات بی شائبه‌ی او، با زبان قاصر و دست ناتوان، چیزی بنگاریم. اما از آنجایی که تجلیل از معلم، سپاس از انسانی است که هدف و غایت آفرینش را تامین می‌کند و سلامت امانت‌هایی را که به دستش سپرده‌اند، تضمین؛ بر حسب وظیفه و از باب ”من لم یشکر المنعم من المخلوقین لم یشکر الله عزّ و جلّ“ از استاد بزرگوارم، جناب دکتر محمد آرشی که همواره بر کوتاهی و درستی من، قلم عفو کشیده و کریمانه از کنار غفلت‌هایم گذشته‌اند، کمال تشکر را دارم. باشد که این خردترین، بخشی از زحمات آنان را سپاس گوید.

هچنین لازم می‌دانم که از دوستان عزیز و مهربانم که در انجام این رساله با کمک‌های بی دریغشان مرا مورد لطف و توجه خود قرار دادند، تشکر کنم و برای آن‌ها بهترین‌ها را آرزومندم.

سیده بصیرا حسینی
شهریور ۱۳۹۴

تعمدنامه

اینجانب سیده بصیرا حسینی دانشجوی کارشناسی ارشد رشته آمار دانشکده علوم ریاضی دانشگاه شاهرود، نویسنده پایان‌نامه با عنوان مقیاس‌گذاری چندبعدی از دیدگاه بیزی در توزیع نرمال چند متغیره، تحت راهنمایی دکتر محمد آرشی متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تاکنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه شاهرود متعلق دارد، و مقالات مستخرج با نام “ دانشگاه شاهرود “ یا “ Shahrood University “ به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده) شده است، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

سیده بصیرا حسینی
شهریور ۱۳۹۴

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان‌نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

اگر فاصله‌ی شهرداری، فرمانداری و ادارات مختلف یک شهر را داشته باشیم، با استفاده از روش مقیاس‌گذاری چندبعدی، می‌توانیم این مکان‌ها را در یک صفحه‌ی دو بعدی رسم کنیم. یعنی از ماتریس فاصله به ماتریس مشاهدات، نه به‌طور دقیق بلکه به‌طور تقریبی که خطای آن بسیار ناچیز است، برسیم. روش مقیاس‌گذاری چندبعدی در بسیاری از زمینه‌ها و موضوعات کاربرد دارد. و فقط به منظور یافتن مکان یا نقشه‌ی شهرها نمی‌باشد بلکه برای کاهش بعد متغیرها نیز به کار می‌رود. داده‌ها می‌توانند هر چیزی باشند. در این روش براساس داده‌های مشاهده شده یک ماتریس فاصله تشکیل می‌شود و سپس پیکره مورد نظر در بعد کمتر به دست می‌آید.

مقیاس‌گذاری چندبعدی، شامل چندین روش از تحلیل داده‌های چندمتغیره است که به مسئله‌ی بصری‌سازی در فضای اقلیدسی می‌پردازد. این تحلیل، اطلاعاتی را درباره‌ی فواصل بین متغیرها که ممکن است متری یا غیرمتری باشند، به دست می‌دهد که در آن فواصل الزاما اقلیدسی نیستند. مقیاس‌گذاری چندبعدی بیزی، یکی از روش‌های جدید مقیاس‌گذاری چندبعدی است که تشکیل پیکره در بعد کمتر را از دیدگاه بیزی بررسی می‌کند. در این پایان‌نامه روش مقیاس‌گذاری چندبعدی بیزی را با توزیع‌های پیشین نرمال و آمیخته برای فاصله مشاهدات مورد بررسی قرار داده و در این راستا برتری این روش را بر روش کلاسیک در قالب شبیه‌سازی و مثال واقعی مورد مطالعه قرار می‌دهیم.

کلمات کلیدی: مقیاس‌گذاری چندبعدی، مقیاس‌گذاری چندبعدی بیزی، نرمال آمیخته، توزیع پسین، عدم شباهت، فاصله اقلیدسی.

پیشگفتار

مقیاس‌گذاری چندبعدي (MDS)، روشی است که با استفاده از داده‌های شباهت یا عدم شباهت به بازیابی پیکره اشیا در فضای نه الزاما اقلیدسی، با بعدی کمتر از بعد واقعی داده‌ها، می‌پردازد. یکی از جنبه‌های مهم در مقیاس‌گذاری چندبعدي یافتن بعد مناسب است.

در این پایان‌نامه مقیاس‌گذاری چندبعدي کلاسیک را که یکی از انواع مقیاس‌گذاری بوده از دیدگاه بیزی مورد بررسی قرار می‌دهیم. از آنجایی که تعیین بعد مناسب از اهمیت بسزایی برخوردار است، با استفاده از نتایج حاصل از مقیاس‌گذاری چندبعدي بیزی، دو معیار انتخاب بعد بیزی *MDSIC* و *BDIM* را مورد بررسی قرار می‌دهیم.

این مجموعه شامل چهار فصل و پنج پیوست می‌باشد که مطالب هر فصل به طور مختصر عبارتند از:

- در فصل اول، مقدمه و تاریخچه‌ای از مقیاس‌گذاری آورده شده و بنا به ضرورت و آشنایی بیشتر، تعاریفی از مقیاس‌گذاری و انواع آن و همچنین معیار استرس برای تعیین بعد مناسب ذکر شده است.

- در فصل دوم، مقیاس‌گذاری چندبعدي را از دیدگاه بیزی با پیشین نرمال مورد بررسی قرار داده‌ایم و معیار انتخاب بعد *MDSIC* را براساس پیکره برآوردشده از *MDS* ارائه می‌دهیم. در انتها با انجام یک شبیه‌سازی به مقایسه مقیاس‌گذاری کلاسیک و بیزی می‌پردازیم.

- در فصل سوم، مقیاس‌گذاری چندبعدي بیزی با پیشین آمیخته را مورد بررسی قرار داده و معیار انتخاب بعد بیزی *BDIM* را برای انتخاب بعد مناسب معرفی و در یک شبیه‌سازی به مقایسه آن با معیار *MDSIC* پرداخته‌ایم.

- در فصل چهارم، با به‌کارگیری داده‌های واقعی به مقایسه و بررسی مقیاس‌گذاری چندبعدي کلاسیک و بیزی، و معیارهای ارائه شده‌ی انتخاب بعد مناسب می‌پردازیم. در انتها پیشنهادهایی برای ادامه‌ی کار در آینده مطرح گردیده‌اند.

- در پیوست ۱، مقدمه‌ایی بر جبر ماتریس‌ها به منظور درک بهتر روابط بکار رفته در این پایان‌نامه مطرح گردیده است.

- در پیوست ۲، برخی توابع آماری به کار رفته را یادآوری می‌کنیم.

- در پیوست ۳، به اثبات رابطه (۱۳.۲) پرداخته‌ایم.

- در پیوست ۴، توضیحاتی مختصر درباره‌ی استنباط بیزی و روش *MCMC* داده شده است.

- در پیوست ۵، برنامه‌های کامپیوتری با نرم‌افزار *R* ارائه گردیده است.

فهرست نشانه‌ها، نمادها و علائم اختصاری

| | |
|---|-----------------------|
| فضای اقلیدسی m بعدی | \mathbb{R}^m |
| ماتریس A با m سطر و n ستون | $A(m \times n)$ |
| ماتریس تصادفی | X |
| بردار تصادفی | \mathbf{x} |
| ماتریس همانی با بعد m | I_m |
| ماتریس واریانس - کوواریانس | Σ |
| ترانپوزیته ماتریس A | A^T |
| معکوس ماتریس A | A^{-1} |
| ریشه‌ی دوم ماتریس A | $A^{\frac{1}{2}}$ |
| ماتریس قطری A | $diag(A)$ |
| دترمینان A | $ A $ |
| شی i ام پیکره | x_i |
| فاصله‌ی بین دو شی i ام و j ام پیکره اشیا | d_{ij} |
| فاصله اقلیدسی | δ_{ij} |
| توزیع گامای معکوس با پارامترهای α و β | $IG(\alpha, \beta)$ |
| توزیع گاما با پارامترهای α و β | $G(\alpha, \beta)$ |
| توزیع بتا با پارامترهای α و β | $Beta(\alpha, \beta)$ |
| توزیع نرمال با پارامترهای μ و σ | $N(\mu, \sigma^2)$ |
| تابع چگالی توزیع نرمال استاندارد | $\phi(\cdot)$ |
| تابع توزیع تجمعی نرمال استاندارد | $\Phi(\cdot)$ |
| توزیع نرمال p متغیره با بردار میانگین μ و ماتریس واریانس - کوواریانس Σ | $N_p(\mu, \Sigma)$ |
| مقیاس‌گذاری چندبعدی | MDS |
| مقیاس‌گذاری چندبعدی کلاسیک | $CMDS$ |
| مقیاس‌گذاری چندبعدی بیزی | $BMDS$ |
| معیار انتخاب بعد بیزی | $BDIM$ |
| معیار اطلاع MDS | $MDSIC$ |
| مونت کارلوی زنجیره مارکوفی | $MCMC$ |
| تقریباً برابر | \approx |

فهرست مطالب

| | |
|----|---|
| س | لیست تصاویر |
| ص | لیست جداول |
| ۱ | ۱ مقیاس‌گذاری چندبعدي |
| ۱ | ۱.۱ مقدمه |
| ۳ | ۲.۱ تاریخچه |
| ۴ | ۳.۱ معرفی روش <i>MDS</i> |
| ۵ | ۱.۳.۱ مدل‌های داده‌های مجاورت |
| ۵ | ۲.۳.۱ مدل‌های فضایی برای مجاورت‌ها: مقیاس‌گذاری چندبعدي |
| ۶ | ۴.۱ مقیاس‌گذاری چندبعدي کلاسیک |
| ۷ | ۱.۴.۱ محاسبات تحلیلی |
| ۱۳ | ۵.۱ مقیاس‌گذاری چندبعدي نامتری |
| ۱۶ | ۶.۱ تنش |
| ۲۱ | ۲ مقیاس‌گذاری چند بعدی بیزی با پیشین نرمال |
| ۲۱ | ۱.۲ مقدمه |
| ۲۱ | ۲.۲ معرفی مدل |
| ۲۳ | ۳.۲ توزیع پیشین |
| ۲۴ | ۴.۲ توزیع پسین |
| ۲۷ | ۵.۲ محاسبات عددی |
| ۲۷ | ۱.۵.۲ الگوریتم <i>MCMC</i> |
| ۲۹ | ۲.۵.۲ انتخاب بعد مناسب |
| ۳۱ | ۳.۵.۲ شبیه‌سازی |
| ۳۲ | ۶.۲ نتیجه‌گیری |
| ۳۵ | ۳ مقیاس‌گذاری چند بعدی بیزی با پیشین آمیخته |
| ۳۵ | ۱.۳ مقدمه |

| | | |
|----|---------------------------------------|-------|
| ۳۵ | مدل آماری | ۲.۳ |
| ۳۶ | توزیع پیشین | ۳.۳ |
| ۳۷ | توزیع پسین | ۴.۳ |
| ۳۸ | محاسبات عددی | ۵.۳ |
| ۳۸ | الگوریتم <i>MCMC</i> | ۱.۵.۳ |
| ۴۱ | شبیه‌سازی | ۲.۵.۳ |
| ۴۲ | نتیجه‌گیری | ۶.۳ |
| ۴۵ | مثال‌های کاربردی | ۴ |
| ۴۵ | مقدمه | ۱.۴ |
| ۵۱ | پیشنهادات و آینده تحقیق | ۲.۴ |
| ۵۳ | مفاهیم اولیه | آ |
| ۵۳ | مقدمه | ۱.آ |
| ۵۳ | جبر ماتریس‌ها | ۲.آ |
| ۵۷ | توزیع‌های آماری | ب |
| ۵۷ | برخی توزیع‌های آماری | ب.۱ |
| ۵۷ | توزیع‌های یک متغیره | ب.۱.۱ |
| ۵۹ | اثبات | پ |
| ۵۹ | اثبات رابطه‌ی (۱۳.۲) | پ.۱ |
| ۶۱ | مقدماتی از استنباط بیزی و <i>MCMC</i> | ت |
| ۶۱ | مقدمه‌ای از استنباط بیزی | ت.۱ |
| ۶۱ | اصل بیز | ت.۲ |
| ۶۲ | مونت کارلو | ت.۳ |
| ۶۲ | زنجیره مارکف | ت.۴ |
| ۶۲ | توزیع مانا | ت.۵ |
| ۶۳ | روش شبیه‌سازی <i>MCMC</i> | ت.۶ |
| ۶۳ | متروپلیس - هستینگز | ت.۷ |
| ۶۴ | الگوریتم نمونه‌گیر گیبز | ت.۱.۷ |
| ۶۶ | الگوریتم متروپلیس-هستینگز تحت گیبز | ت.۲.۷ |
| ۶۷ | برنامه‌های کامپیوتری با نرم‌افزار R | ث |
| ۶۹ | کدهای مربوط به فصل ۴ | ث.۱ |

۷۷

مراجع

۸۱

واژه‌نامه فارسی به انگلیسی

لیست تصاویر

| | | | |
|----|--|------|--|
| ۱۴ | آبی | ۱۰.۱ | نمایش دوبعدی از مقیاس‌گذاری چندبعدی برای ماتریس فاصله‌ی موش‌های صحرایی- |
| ۱۹ | نقشه پیکره فکری رهبران سیاسی جنگ جهانی دوم | ۲.۱ | |
| ۴۲ | <i>MDSIC</i> و <i>BDIM</i> روش | ۱.۳ | فراوانی $p^* = 1$ در مقابل ρ از راست به ترتیب برای روش |
| ۴۳ | <i>BDIM</i> و <i>MDSIC</i> به ترتیب | ۲.۳ | فراوانی $p^* = 1$ در مقابل σ/ρ از راست به ترتیب |
| ۴۴ | روش | ۳.۳ | فراوانی $p^* = 1$ در مقابل σ/ρ برای دو روش |
| ۴۶ | <i>CMDS</i> و | ۱.۴ | فواصل هوایی مشاهده شده در مقابل فواصل برآورد شده برای دو روش <i>BMDS</i> |
| ۴۶ | برآورد مکان شهرها با استفاده از روش <i>BMDS</i> | ۲.۴ | |
| ۵۱ | احتمال معنی‌دار برای هر بعد در روش <i>BDIM</i> برای داده‌های دلستر | ۳.۴ | |
| ۵۱ | مقادیر <i>MDSIC</i> به ازای هر بعد برای داده‌های دلستر | ۴.۴ | |
| ۵۲ | پیکره اشیا دلسترها در فضای دو بعدی | ۵.۴ | |

لیست جداول

| | | | |
|----|-------|---|-----|
| ۱۲ | | ماتریس داده‌های عدم شباهت موش‌های صحرائی- آبی | ۲.۱ |
| ۱۵ | | میزان برآزش بر اساس معیار تنش | ۳.۱ |
| ۱۸ | | ماتریس فاصله داده‌های رهبران سیاسی جنگ جهان دوم | ۵.۱ |
| ۳۲ | | تحلیل شبیه‌سازی داده‌های تولید شده از $N_i(0, I)$ | ۱.۲ |
| ۴۶ | | جدول تحلیل فاصله‌ی شهرها | ۱.۴ |
| ۴۷ | | فاصله‌ی هوایی بین شهرها (برحسب مایل) | ۳.۴ |
| ۴۸ | | احتمال‌های حاصل از BDIM به ازای بعدها مختلف | ۴.۴ |
| ۴۹ | | ویژگی‌های بویایی ۱۹ دلستر | ۵.۴ |
| ۵۰ | | ویژگی‌های طعم و مزه ۱۹ دلستر | ۶.۴ |

فصل ۱

مقیاس‌گذاری چندبعدي

۱.۱ مقدمه

مقیاس‌گذاری چندبعدي^۱ (*MDS*)، روشی برای تولید یک طرح (نقشه^۲) با بعدی کمتر از بعد داده‌ها بوده که با استفاده از ماتریس فاصله که از روی ماتریس داده‌های چندمتغیره X (ماتریس‌های معروف به شباهت^۳ یا عدم‌شباهت^۴ که به روش‌های گوناگون از جمله قضاوت‌های محقق در مورد شباهت جفت‌های یکسان از اشیا^۵، افراد، محرک‌ها و موضوعات تولید می‌شود) به دست می‌آید، حاصل می‌شود. این اشیا معمولاً به اجسام بی‌جان (بطری دلستر)، افراد^۶ (مردم یا حیوانات) و محرک^۷ (موارد غیر ملموس از جمله طعم و مزه دلستر) اشاره دارد. اغلب اصطلاح مجاورت^۸ هم برای داده‌های شباهت و هم برای داده‌های عدم‌شباهت به کار برده می‌شود.

داده‌های شامل شباهت یا عدم‌شباهت بین جفت اشیا مشاهده شده در بسیاری از زمینه‌ها از جمله تحقیقات علمی در بازار داد و ستد کالا، روانشناسی، جامعه‌شناسی، تحقیقات محیطی، بازیابی اطلاعات وب و دیگر پایگاه داده‌ها کاربرد دارد.

با توجه به تعریف ساده فوق، حال می‌توان تعبیری دیگر را از *MDS* را ارائه کرد؛ به عبارتی دیگر *MDS* روشی است که با استفاده از داده‌های شباهت یا عدم‌شباهت به بازیابی مکان اشیا (پیکره اشیا^۹) در فضای اقلیدسی با بعدی کوچکتر از بعد واقعی می‌پردازد به طوری که فاصله بین اشیا در فضا به خوبی حفظ شود.

^۱Multidimensional scaling

^۲Map

^۳Similarity

^۴Dissimilarity

^۵Objects

^۶Subjects

^۷Stimuli

^۸Proximity

^۹Object configuration

روش MDS، در برگیرنده چندین روش تحلیل داده‌های چند متغیره می‌باشد که در نهایت هر روش نمایش گرافیکی اشیا از داده‌ها را ارائه می‌دهد؛ مثلاً یکی از کاربردهای MDS مجسم‌سازی است به طوری که مجموعه‌ای پیچیده از شباهت‌ها را به صورتی که برای مشاهده آسان‌تر باشد و اطلاعات مفیدی از روابط بین داده‌ها حاصل شود را ارائه می‌دهد و امکان تحلیل بهتر را فراهم می‌کند.

با توجه به انواع روش‌های تحلیل داده‌های مجاورت و نوع داده‌های چندمتغیره، می‌توان MDS را به انواع مقیاس‌گذاری تقسیم‌بندی کرد که در زیربخش ۱.۳.۱ عنوان خواهیم کرد. به طور کلی بر اساس کمی یا کیفی بودن داده‌های مجاورت، MDS به دو دسته متری^{۱۰} و نامتری^{۱۱} دسته‌بندی می‌شود.

روش MDS کلاسیک یکی از انواع مقیاس‌گذاری متری است که پیکره اشیا را به صورت مستقیم از ماتریس فاصله (عدم شباهت) که معمولاً آن را با D نشان می‌دهند، بازیابی می‌کند در حالی که در روش‌های نامتری، رتبه عدم شباهت داده‌ها در بازیابی پیکره مورد استفاده قرار می‌گیرد. در بخش‌های ۱.۴ و ۱.۵ به بررسی این دو مقیاس‌گذاری می‌پردازیم.

روش MDS ریشه در روانشناسی و به‌طور خاص در علم روابط بین روانشناسی و فیزیک دارد که بر پایه مقایسه بین مفهوم شباهت در روانشناسی و مفهوم هندسی فاصله است. همچنین به‌طور گسترده در علوم رفتاری و مسایل دیگر اجتماعی مورد استفاده بوده و مقالات زیادی با موضوع هوش (روان‌سنجی) در این راستا چاپ شده است. علاوه بر مواردی که در ابتدای مقدمه در خصوص کاربرد روش MDS به آن‌ها اشاره شد، می‌توان به کاربرد آن در موضوعاتی مانند زمین‌شناسی (تیبشیرانی و همکاران^{۱۲}، ۱۹۹۹)، بازیابی اطلاعات وب و دیگر پایگاه داده‌ها (اسچوتز و سیلورستین^{۱۳}، ۱۹۹۷)، بازیابی^{۱۴}، داده‌کاوی^{۱۵}، بوم‌شناسی^{۱۶}، علم اطلاعات^{۱۷}، سایکو فیزیک^{۱۸} و بصری‌سازی علمی^{۱۹} اشاره کرد (برای آگاهی بیشتر مراجع، ۴، ۶، ۸، ۱۱، ۲۸، ۳۱، ۳۹، ۴۲ را ببینید).

یکی از کاربردهای اصلی MDS، یافتن بعد مناسب (مفید) در فضا است، یعنی یافتن صفت‌هایی که در بیان شباهت‌ها یا عدم شباهت‌ها نقش معنی‌داری دارند. به عنوان مثال، کاربرد MDS در انتخاب کاندیدهای سیاسی را در نظر بگیرید. داده‌ها شامل قضاوت‌های فردی هستند که بر پایه صفاتی (متغیرهایی) از قبیل محل تولد، سطح تحصیلات، سن، عضویت در حزب سیاسی و ... کاندید شکل گرفته‌اند. لذا اگر فردی بخواهد ببیند که آیا دو کاندید از منظر صفات (متغیرهای) فوق شباهت دارند، می‌تواند از روش MDS استفاده کند. در این راستا، اگر تعداد متغیرهای مورد بررسی برای تعیین شباهت‌ها یا عدم شباهت‌ها بین کاندیدها زیاد باشد تفسیر ساختار پیکره این شباهت‌ها یا عدم شباهت‌ها

^{۱۰} Metric

^{۱۱} Nonmetric

^{۱۲} Tibshirani

^{۱۳} Schutze and Silverstein

^{۱۴} Marketing

^{۱۵} Data mining

^{۱۶} Ecology

^{۱۷} Science information or Information technology

^{۱۸} Psycho-physics

^{۱۹} Scientific visualization

مشکل خواهد بود.

از طرفی چنانچه این تعداد خیلی کم باشد ممکن است صفات مهم و تاثیرگذاری در ساخت پیکره از دست برود. لذا تعیین بعد مناسب که با کمترین تعداد صفات (متغیرها) بتواند تا حد بالایی پیکره را به درستی تعیین (ترسیم) کند از اهمیت به سزایی در MDS برخوردار است. در آخر، موضوع مهم دیگر، پس از پیاده‌سازی روش MDS بر روی داده‌ها و ساخت پیکره و تعداد بعد مناسب، تعیین مقادیر متغیرهای جدید (در بعد کمتر از بعد واقعی داده‌ها) برای رسم پیکره می‌باشد.

۲.۱ تاریخچه

در این بخش تاریخچه مقیاس‌گذاری چندبعدي را از ابتدا تا تحقیقات و مطالعات اخیر بررسی می‌کنیم. روش یافتن مختصات واقعی بر اساس فاصله اقلیدسی اولین بار توسط اسچوینبرگ^{۲۰} (۱۹۳۵) مطرح شد و یانگ و هولدر^{۲۱} (۱۹۳۸) نشان دادند که چطور با ماتریس فاصله بین نقاط در یک فضای اقلیدسی می‌توان مختصات نقاط را به طوری که فاصله‌ها حفظ شوند، یافت. پس از آن مقیاس‌گذاری چندبعدي کلاسیک توسط تورگوسون^{۲۲} (۱۹۵۲) توسعه داده شد.

افرادی همچون شپارد^{۲۳} (۱۹۶۲) و کروسکال^{۲۴} (۱۹۶۴) روش مقیاس‌گذاری چندبعدي را به الگوریتمی که مختصات نقاط را بر مبنای رتبه‌بندی عدم شباهت می‌سازد، توسعه دادند.

کامبس^{۲۵} (۱۹۶۴) یک طبقه‌بندی کلی از انواع داده‌های مقیاس‌گذاری را ارائه می‌دهد که پس از آن در سال ۱۹۸۰ توسط کارول و عربی^{۲۶} طبقه‌بندی داده‌ها و همچنین طبقه‌بندی انواع تحلیل‌های مقیاس‌گذاری چندبعدي به‌روز شد.

گریناکر و آندرهیل^{۲۷} (۱۹۸۲) از مقیاس‌گذاری چندبعدي برای بروز کردن نقشه شهرهای کشور آفریقای جنوبی، بر اساس زمان پرواز بین فرودگاه‌های شهرهای آفریقای جنوبی استفاده کردند.

اولین کتاب مقدماتی درباره مقیاس‌گذاری چندبعدي توسط کروسکال و ویش^{۲۸} در سال (۱۹۷۸) نوشته شد. مطالعات بیشتری را می‌توان در تحقیقات اسچیفمن^{۲۹} و همکاران (۱۹۸۱)، دیویسون^{۳۰} (۱۹۸۳)

^{۲۰}Schoenberg

^{۲۱}Young and Householder

^{۲۲}Torgerson

^{۲۳}Shepard

^{۲۴}Kruskal

^{۲۵}Coombs

^{۲۶}Carroll, Arabie

^{۲۷}Greenacre and Underhill

^{۲۸}Kruskal and Wish

^{۲۹}Schiffman

^{۳۰}Davidson

، گاور^{۳۱} (۱۹۸۴) ، يانگ^{۳۲} (۱۹۸۷) ، ميد^{۳۳} (۱۹۹۲) ، برگ و گرونن^{۳۴} (۱۹۹۷) ديد. جديدترين کتاب که در زمينه مقیاس‌گذاری چندبعدي نوشته شده است، حاصل تلاش کاکس و کاکس^{۳۵} در سال (۲۰۰۱) می‌باشد. روش‌های مقیاس‌گذاری چندبعدي بر پایه درست‌نمایی ماکسیمم توسط رامسی^{۳۶} (۱۹۸۲) ، تاکانه و کارول^{۳۷} (۱۹۸۱) ، گرونن (۱۹۹۳) و گرونن و همکاران (۱۹۹۵) مورد بررسی قرار گرفته است. اُه و رافتری^{۳۸} (۲۰۰۱) برای اولین بار مقیاس‌گذاری چندبعدي را از دیدگاه بیزی بررسی کردند. پس از آن اُه (۲۰۱۲) مطالعات را تعمیم داد. تمامی تحقیقات فوق‌الذکر در خارج از ایران بوده است. در ایران تنها می‌توان به مطالعات علی‌رضا علی‌احمدی و سپهر قاضی‌نوری با عنوان کاربرد روش آماری مقیاس‌گذاری چندبعدي در دسته‌بندی اقتصادی و فرهنگی کشورهای جهان اشاره کرد.

۳.۱ معرفی روش MDS

روش مقیاس‌گذاری چندبعدي بر روی مجموعه‌ای مانند O که شامل اشیا بوده بررسی می‌شود و عدم شباهت روی مجموعه حاصلضرب دکارتی $O \times O$ تعريف می‌شود. عدم شباهت واقعی بین i امین و j امین شی (که مجهول است) را با نماد δ_{ij} به ازای $i, j \in O$ نمایش می‌دهیم.

فرض کنید ϕ یک نگاشت دلخواه از O به E (یک فضای اقلیدسی) است. مثلاً نقشه $\phi(i) = x_i$ که در آن $i \in O$ و $x_i \in E$ ، نشان دهنده مکان نقطه است. فرض کنید $X = \{x_i : i \in O\}$ مجموعه تصاویر باشد. فاصله بین نقاط x_i و x_j در X (که بر پایه مشاهدات می‌باشد) را با d_{ij} نمایش می‌دهند. متر MDS در پی یافتن یک مجموعه نقاط در فضا است به طوری که

$$d_{ij} \approx f(\delta_{ij})$$

که در آن $f(\cdot)$ یک تابع پارامتری یکنواخت پیوسته و نماد (\approx) نشان‌دهنده تقریباً برابر است. تابع f می‌تواند یک تابع همانی یا تابعی که عدم شباهت را به یک فاصله برای اندازه‌گیری مجاورت تبدیل کند، باشد.

هدف یافتن یک نگاشت مانند ϕ برای d_{ij} است که برای همه $i, j \in O$ تقریباً برابر با $f(\delta_{ij})$ باشد.

^{۳۱}Gower

^{۳۲}Young

^{۳۳}Mead

^{۳۴}Borg and Groenen

^{۳۵}Cox and Cox

^{۳۶}Ramsay

^{۳۷}Takane and Carroll

^{۳۸}Oh and Raftery

۱.۳.۱ مدل‌های داده‌های مجاورت

مدل‌های مقیاس‌گذاری به داده‌های مجاورت^{۳۹} برازش داده می‌شوند تا به روشن شدن، نمایش دادن و فهم ساختارها یا الگوهای احتمالی بین مشاهدات یا اندازه‌های مجاورت محاسبه‌شده که در مجموعه مقادیر عددی به آسانی ظاهر نمی‌شوند، کمک کنند.

در برخی حوزه‌ها به ویژه روان‌شناسی، هدف نهایی در تحلیل یک مجموعه از داده‌های مجاورت‌ها خاص‌تر است. بدین معنی که به دنبال نظریه‌هایی هستیم تا بتوانیم شباهت‌ها و عدم شباهت‌ها بین داوری‌های افراد را توجیح کنیم. به عبارت دقیق‌تر هدف از مدل‌بندی داده‌های مجاورت، تلاش برای یافتن پاسخی به این پرسش است که چه عاملی باعث می‌شود که برخی چیزها شبیه و برخی متفاوت به نظر برسند؟ به طور خلاصه مدل‌ها برای تحلیل داده‌های مجاورت به سه رده اصلی طبقه‌بندی می‌شوند:

- مدل‌های فضایی^{۴۰}
- مدل‌های درختی^{۴۱}
- مدل‌های دو رگه (تلفیقی)^{۴۲}

ما در این پایان‌نامه فقط به بررسی رده مدل‌های فضایی می‌پردازیم. در ادامه منظور از مجاورت‌ها همان داده‌های مجاورت است.

۲.۳.۱ مدل‌های فضایی برای مجاورت‌ها: مقیاس‌گذاری چندبعدي

یک نمایش فضایی از یک ماتریس مجاورت، شامل یک مجموعه از n نقطه (مختصات^{۴۳}) p بعدی می‌باشد که هر کدام، یکی از n واحد در داده‌ها را نمایش می‌دهند. به‌طور کلی مختصات مورد نظر برای ساخت پیکره‌ها با مینیم کردن چند اندازه از برازش بین فاصله‌هایی که از مختصات و مجاورت‌های مشاهده شده نتیجه شده‌اند، به دست می‌آید. به زبان ساده، در یک مدل هندسی علاقه‌مندیم هر چه فاصله‌ی مشاهده شده بین نقاط بزرگتر یا عدم شباهت بین دو واحد بیشتر (شباهت کمتر) باشد، نقاط نمایش دهنده‌ی آن‌ها در مدل، از هم جداتر باشند.

به‌طور کلی (اما نه منحصر) فاصله‌ی بین نقاط در مدل فضایی، اقلیدسی فرض می‌شود. مقیاس‌گذاری چندبعدي، شامل چندین روش از تحلیل داده‌های چند متغیره می‌باشد که هدف بسیاری از این روش‌ها، یافتن مجموعه بهترین برازش از مختصات (یعنی یافتن مختصات نقاط) و کاهش بعد p ، که به اندازه کافی بیانگر شباهت‌ها یا عدم شباهت‌ها باشد، است و در نهایت هر روش نمایش گرافیکی اشیا از داده‌های چندمتغیره را ارائه می‌دهد. علاقه‌مندیم که بعد داده‌ها یعنی p کوچک شود. در حالت ایده‌آل دو

^{۳۹}Proximity data

^{۴۰}Spatial models

^{۴۱}Tree models

^{۴۲}Hybrid models

^{۴۳}Coordinate

یا سه خواهد شد، به طوری که پیکره فضایی در نظر گرفته شده به آسانی رسم شود. تنوع روش‌های مقیاس‌گذاری چندبعدي به دلیل تفاوت زیاد در چگونگی ارزیابی تطابق بین فاصله برازش شده (d_{rs}) و مجاورت‌های مشاهده شده (δ_{rs}) می‌باشد. از انواع مقیاس‌گذاری می‌توان به موارد زیر اشاره کرد:

- مقیاس‌گذاری کلاسیک ^{۴۴}
- مقیاس‌گذاری متوسط کریتچلی ^{۴۵}
- مقیاس‌گذاری نامتری ^{۴۶}
- مقیاس‌گذاری تک‌بعدي ^{۴۷}
- مقیاس‌گذاری متری کمترین توان‌های دوم ^{۴۸}
- مقیاس‌گذاری بی‌زی ^{۴۹}

نوع آخر مقیاس‌گذاری که از دیدگاه نظریه بی‌زی مورد بررسی قرار می‌گیرد، موضوع اصلی این پایان‌نامه می‌باشد. لازم به ذکر است که پایه مقیاس‌گذاری بی‌زی، MDS است. در این پایان‌نامه تنها دو مقاله نوشته‌شده در این زمینه را مورد بررسی و مطالعه قرار می‌دهیم. در ادامه به بررسی جزئی‌تر مقیاس‌گذاری کلاسیک و نامتری می‌پردازیم.

۴.۱ مقیاس‌گذاری چندبعدي کلاسیک

مقیاس‌گذاری کلاسیک درصدد است که یک ماتریس مجاورت را به وسیله‌ی یک نقشه یا مدل هندسی ساده ارائه دهد. چنین مدلی با یک مجموعه از نقاط x_1, x_2, \dots, x_n در p بعد نشان داده می‌شود. هدف تعیین بعد مناسب p از مدل و n مختصات p بعدی x_1, x_2, \dots, x_n است به طوری که مدل، یک برازش خوب برای مجاورت‌های مشاهده شده را نتیجه دهد.

برازش، اغلب با چند شاخص داوری می‌شود که این شاخص‌ها میزان تطابق بین مجاورت‌ها و فاصله‌ها در مدل هندسی را اندازه می‌گیرند. به عبارت ساده‌تر برای یک عدم شباهت مشاهده شده d_{ij} (بزرگ‌تر باید نقاط نمایشگر آن‌ها در مدل) δ_{ij} (بیشتر از هم جدا باشند).

حال سوالی که پیش می‌آید این است که چگونه p و مقادیر مختصات x_1, x_2, \dots, x_n از ماتریس مجاورت مشاهده شده را برآورد کنیم؟

^{۴۴} Classical scaling

^{۴۵} Critchley intermediate

^{۴۶} Nonmetric scaling

^{۴۷} Unidimensional scaling

^{۴۸} Metric least squares scaling

^{۴۹} Bayesian MDS

به عنوان مثالی شهودی، اگر فواصل بین شهرهای ایران را داشته باشیم، آیا می‌توانیم موقعیت شهرها را روی یک نقشه دو بعدی تعیین کنیم؟ مقیاس‌گذاری چندبعدي کلاسیک (CMDS) پاسخ این سوال را به درستی می‌دهد.

۱.۴.۱ محاسبات تحلیلی

ماتریس داده‌های اولیه $X = (x_1^T, \dots, x_n^T)^T$ در اندازه $n \times p$ که در آن $x_i = (x_{i1}, \dots, x_{ip})$ به ازای $i = 1, \dots, n$ را در نظر بگیرید. اگر فاصله اقلیدسی بین مشاهده‌های i ام و j ام را با d_{ij} نشان دهیم، آنگاه داریم

$$\begin{aligned} d_{ij} &= d(x_i, x_j) \\ &= \|x_i - x_j\| \\ &= [(x_i - x_j)'(x_i - x_j)]^{\frac{1}{2}} \\ &= \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{ip} - x_{jp})^2} \end{aligned}$$

در این صورت ماتریس فاصله‌های اقلیدسی $D_{n \times n}$ به صورت زیر حاصل می‌شود:

$$\begin{aligned} D &= \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ & d_{22} & \dots & d_{2n} \\ & & \ddots & \vdots \\ & & & d_{nn} \end{bmatrix} \\ &= \begin{bmatrix} d(x_1, x_1) & d(x_1, x_2) & \dots & d(x_1, x_n) \\ & d(x_2, x_2) & \dots & d(x_2, x_n) \\ & & \ddots & \vdots \\ & & & d(x_n, x_n) \end{bmatrix}. \end{aligned}$$

بدیهی است از آن جایی که $d_{ij} = d_{ji}$ ، ماتریس فاصله‌های D متقارن است. حال، فرض کنید ماتریس فاصله‌های اقلیدسی D ماتریس مجاورت باشد، که از روی ماتریس داده‌ی $n \times p$ اولیه‌ی X ساخته شده است. مقیاس‌گذاری چندبعدي کلاسیک اساساً در رابطه با عکس حالت فوق می‌باشد یعنی با فاصله داده‌شده D چگونه می‌توانیم X را بازیابی کنیم.

لازم به ذکر است که هدف، یافتن مختصات X ای است که بعد آن کمتر از p باشد و بتوانیم داده‌های مجاورت واقعی را، که از جنس فاصله می‌باشند، با دقت بالایی تخمین بزنیم و پیکره اولیه را بازسازی کنیم. برای این منظور ابتدا فرض کنید X معلوم باشد، ماتریس $B_{n \times n}$ را به صورت زیر تعریف می‌کنیم:

$$B = XX' \quad (1.1)$$

که در آن

$$B = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{bmatrix}$$

$$= \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix}.$$

بنابراین عناصر B به صورت زیر می‌باشند:

$$b_{ij} = \sum_{k=1}^p x_{ik}x_{jk}. \quad (2.1)$$

توان دوم فاصله اقلیدسی بین سطرهای X می‌تواند برحسب عناصر B به صورت زیر محاسبه شوند:

$$\begin{aligned} d_{ij}^2 &= d^2(x_i, x_j) \\ &= (x_{i1} - x_{j1})^2 + \dots + (x_{ip} - x_{jp})^2 \\ &= \sum_{k=1}^p x_{ik}^2 + \sum_{k=1}^p x_{jk}^2 - 2 \sum_{k=1}^p x_{ik}x_{jk} \\ &= \sum_{k=1}^p x_{ik}x_{ik} + \sum_{k=1}^p x_{jk}x_{jk} - 2 \sum_{k=1}^p x_{ik}x_{jk} \\ &= b_{ii} + b_{jj} - 2b_{ij} \end{aligned} \quad (3.1)$$

اگر b ها برحسب d ها در معادله (۳.۱) به دست آیند، آنگاه مقادیر مختصات لازم می‌تواند با تجزیه‌ی B به صورت (۱.۱) حاصل شود. اما بدون این‌که یک محدودیت مکانی روی داده‌ها اعمال شود، هیچ جواب یکتایی حاصل نمی‌شود. برای این منظور، معمولاً مرکز نقاط \bar{x}_i در مبدا قرار داده می‌شود به طوری که برای هر k ، $\sum_{i=1}^n x_{ik} = 0$. با اعمال این محدودیت در رابطه (۳.۱)، مجموع عبارات در هر سطر B باید صفر باشد. در نتیجه با جمع کردن روابط داده شده در (۳.۱) روی i و j و در نهایت روی هر دو به معادلات زیر می‌رسیم.

$$\sum_{i=1}^n d_{ij}^2 = T + nb_{jj}, \quad \sum_{j=1}^n d_{ij}^2 = T + nb_{ii}, \quad \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2 = 2nT$$

که در آن $T = \sum_{i=1}^n b_{ii}$ اثر ماتریس B می‌باشد. حال عناصر B را برحسب مجذور فاصله اقلیدسی می‌توان به صورت زیر نوشت

$$b_{ij} = -\frac{1}{2}(d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2)$$

که در آن

$$d_{i.}^{\lambda} = \frac{1}{n} \sum_{j=1}^n d_{ij}^{\lambda}, \quad d_{.j}^{\lambda} = \frac{1}{n} \sum_{i=1}^n d_{ij}^{\lambda}, \quad d_{..}^{\lambda} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^{\lambda}$$

با داشتن عناصر B بر حسب فاصله اقلیدسی، کفایت تجزیه آن را برای یافتن مقادیر به دست آوریم. تجزیه طیفی B ^{۵۰} را به صورت زیر در نظر بگیرید:

$$B = V \Lambda V'$$

که در آن

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \lambda_3 & & \\ & & & \ddots & \\ & & & & \lambda_n \end{pmatrix}$$

ماتریس قطری مقادیر ویژه B و $V = (v_1, \dots, v_n)$ ماتریس بردارهای ویژه متناظر نرمال شده است، به طوری که $v_i v_i' = 1$. به علاوه فرض کنیم مقادیر ویژه به ترتیب $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0$ باشند. وقتی D از یک ماتریس رتبه‌ی کامل $n \times p$ گرفته می‌شود آنگاه رتبه‌ی B باید p باشد. پس $n - p$ مقدار ویژه آخر آن صفر خواهد شد. بنابراین B را می‌توان به صورت زیر بازنویسی کرد:

$$B = V_1 \Lambda_1 V_1'$$

که در آن V_1 شامل p بردار ویژه اول و Λ_1 شامل p مقدار ویژه غیر صفر هستند. بنابراین مقادیر مختصات لازم عبارتند از

$$X = V_1 \Lambda_1^{\frac{1}{2}}$$

که در آن $\Lambda_1^{\frac{1}{2}} = \text{diag}(\lambda_1^{\frac{1}{2}}, \dots, \lambda_p^{\frac{1}{2}})$. لذا X را از XX' بازیابی کردیم. اما باید دقت داشت که استفاده از همه‌ی p بعد منجر به بازیابی کامل ماتریس فاصله اقلیدسی اولیه خواهد شد در حالی که هدف ما استفاده از p بعد نمی‌باشد و به دنبال کاهش بعد هستیم. بنابراین فرض می‌کنیم m بعد $(m < p)$ بهترین برازش به وسیله‌ی m بردار ویژه B متناظر با m مقدار ویژه بزرگتر باشد. مناسب بودن m به وسیله معیار زیر داوری می‌شود به طوری که مقادیر P_m بزرگتر یا مساوی 0.8 یک برازش معقول را پیشنهاد می‌کنند، که در آن

$$P_m = \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i}$$

لازم به ذکر است که ماتریس داده‌های مجاورت، شامل فاصله اقلیدسی که از ماتریس داده‌های X محاسبه شده است، می‌باشد.

بنابراین با استفاده از روش CMDS، با داشتن فواصل بین شهرها (همان ماتریس مجاورت) می‌توانیم موقعیت شهرها (ماتریس داده‌های واقعی) را به دست آوریم و یک نقشه دو بعدی ترسیم کنیم (چنانچه

^{۵۰} Spectral decomposition

بعد دو مناسب باشد). وقتی ماتریس مجاورت مشاهده شده اقلیدسی نباشد، ماتریس B معین مثبت نیست. در چنین مواردی برخی از مقادیر ویژه B منفی خواهند شد و به طور متناظر برخی از مقادیر مختصات اعداد مختلط می‌شوند. اگر با وجود این B ، تنها تعداد کمی از مقادیر منفی کوچک وجود داشته باشند، یک نمایش مفید از ماتریس مجاورت با استفاده از بردارهای ویژه مرتب با m مقدار ویژه مثبت بزرگتر امکان‌پذیر است. مناسب بودن این بعد با یکی از دو معیار زیر ارزیابی می‌شود:

$$P_m^{(1)} = \frac{\sum_{i=1}^m |\lambda_i|}{\sum_{i=1}^n |\lambda_i|}$$

$$P_m^{(2)} = \frac{\sum_{i=1}^m \lambda_i^2}{\sum_{i=1}^n \lambda_i^2}$$

در این جا هم مقادیر بالای 0.8 یک برازش خوب را ارائه می‌کنند؛ و در صورتی که اندازه برازش زیر مینیمم شود، جواب مختصات اصلی m بعدی ($m < p$) بهترین است

$$S = \sum_{i=1}^n \sum_{j=1}^n (d_{ij}^2 - (d_{ij}^{(m)})^2)$$

که در آن d_{ij} فاصله اقلیدسی بین i و j بر اساس مقادیر p متغیر اولیه و $d_{ij}^{(m)}$ فاصله‌ی متناظر محاسبه شده از مقادیر m بعد اصلی می‌باشد. در نهایت تنها دو مقدار دقیق‌تر برای تصمیم‌گیری در خصوص تعداد ابعاد مناسب در مدل‌های فضایی به منظور دستیابی به نمایشی مناسب از مجاورت‌های مشاهده شده وجود دارد. این دو معیار عبارتند از:

معیار اثر 51 : تعداد مختصات (بعد مناسب) را طوری انتخاب کنید که مجموع مقادیر ویژه‌ی مثبت تقریباً مساوی با مجموع تمام مقادیر ویژه‌ها باشد.

معیار بزرگی 52 : تنها مقادیر ویژه‌ای که اندازه آن‌ها از بزرگترین مقدار ویژه منفی بزرگتر است را بپذیرید. اگر ماتریس B تعداد قابل ملاحظه‌ای مقادیر ویژه منفی داشته باشد، مقیاس‌گذاری کلاسیک حاصل از ماتریس مجاورت ممکن است مناسب نباشد و برخی روش‌های دیگر مقیاس‌گذاری به طور مثال مقیاس‌گذاری نامتری (در بخش بعد ببینید) ممکن است بهتر باشد.

مثال ۱.۴.۱. در سال ۱۹۷۰ طی تحقیقاتی که پیرامون جمعیت موش‌های صحرایی - آبی 53 (یک جوندۀ نیمه آبی از خانواده‌ی همیسترهای دم کوتاه، زیر خانواده آرویکوال 54 است و دارای دو گونه‌ی سپیداس 55 و ترستریس 56 می‌باشد) توسط کربرت 57 و همکارانش انجام شد. هدف از انجام این مطالعات، مقایسه جمعیت این حیوانات در انگلستان با سایر کشورهای اروپایی و تحقیق در خصوص امکان وجود بیش از یک گونه موش مذکور در انگلستان بود. داده‌های اصلی شامل مشاهداتی از وجود یا عدم وجود

51 Trace criterion

52 Magnitude criterion

53 Water voles

54 Arvicola

55 Sapidus

56 Terrestriis

57 Corbet

سیزده ویژگی درباره ۳۰۰ مجمه موش‌های صحرایی- آبی، ناشی از شش منطقه از کشور انگلستان و هشت منطقه از بقیه کشورهای اروپایی، هستند. ماتریس فاصله این داده‌ها در جدول ۱.۱ داده شده است. در ادامه مقیاس‌گذاری کلاسیک را برای این داده‌ها به‌کار می‌بریم.

| | Srry | Shrp | Yrks | Prth | Abrd | ElnG | Alps | Ygsl | Grmn | Nrwy | PyrI | PyrII | NrtS | SthS |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Surrey | ۰/۰۰۰ | | | | | | | | | | | | | |
| Shropshire | ۰/۰۹۹ | ۰/۰۰۰ | | | | | | | | | | | | |
| Yorkshire | ۰/۰۳۳ | ۰/۰۲۲ | ۰/۰۰۰ | | | | | | | | | | | |
| Perthshire | ۰/۱۸۳ | ۰/۱۱۴ | ۰/۰۴۲ | ۰/۰۰۰ | | | | | | | | | | |
| Aberdeen | ۰/۱۴۸ | ۰/۲۲۴ | ۰/۰۵۹ | ۰/۰۶۸ | ۰/۰۰۰ | | | | | | | | | |
| EleanGambna | ۰/۱۹۸ | ۰/۰۳۹ | ۰/۰۵۳ | ۰/۰۸۵ | ۰/۰۵۱ | ۰/۰۰۰ | | | | | | | | |
| Alps | ۰/۴۶۲ | ۰/۲۶۶ | ۰/۳۲۲ | ۰/۴۳۵ | ۰/۲۶۸ | ۰/۰۲۵ | ۰/۰۰۰ | | | | | | | |
| Yugoslavia | ۰/۶۲۸ | ۰/۴۴۲ | ۰/۴۴۴ | ۰/۴۰۶ | ۰/۲۴۰ | ۰/۱۲۹ | ۰/۰۱۴ | ۰/۰۰۰ | | | | | | |
| Germany | ۰/۱۱۳ | ۰/۰۷۰ | ۰/۰۴۶ | ۰/۰۴۷ | ۰/۰۳۴ | ۰/۰۰۲ | ۰/۱۰۶ | ۰/۱۲۹ | ۰/۰۰۰ | | | | | |
| Norway | ۰/۱۷۳ | ۰/۱۱۹ | ۰/۱۶۲ | ۰/۳۳۱ | ۰/۱۷۷ | ۰/۰۳۹ | ۰/۰۸۹ | ۰/۲۳۷ | ۰/۰۷۱ | ۰/۰۰۰ | | | | |
| PyreneesI | ۰/۴۳۴ | ۰/۴۱۹ | ۰/۳۳۹ | ۰/۵۰۵ | ۰/۴۶۹ | ۰/۳۹۰ | ۰/۳۱۵ | ۰/۳۴۹ | ۰/۱۵۱ | ۰/۴۳۰ | ۰/۰۰۰ | | | |
| PyreneesII | ۰/۷۶۲ | ۰/۶۳۳ | ۰/۷۸۱ | ۰/۷۰۰ | ۰/۷۵۸ | ۰/۶۲۵ | ۰/۴۶۹ | ۰/۶۱۸ | ۰/۴۴۰ | ۰/۵۳۸ | ۰/۶۰۷ | ۰/۰۰۰ | | |
| NSpain | ۰/۵۳۰ | ۰/۳۸۹ | ۰/۴۸۲ | ۰/۵۷۹ | ۰/۵۹۷ | ۰/۴۹۸ | ۰/۳۷۴ | ۰/۵۶۲ | ۰/۲۴۷ | ۰/۳۸۳ | ۰/۳۸۷ | ۰/۰۸۴ | ۰/۰۰۰ | |
| SSpain | ۰/۵۸۶ | ۰/۴۳۵ | ۰/۵۵۰ | ۰/۵۳۰ | ۰/۵۵۲ | ۰/۵۰۹ | ۰/۳۶۹ | ۰/۴۷۱ | ۰/۲۳۴ | ۰/۳۴۶ | ۰/۴۵۶ | ۰/۰۹۰ | ۰/۰۳۸ | ۰/۰۰۰ |

جدول ۲.۱: ماتریس داده‌های عدم شباهت موش‌های صحرایی-آبی

برای به کار بردن مقیاس‌گذاری کلاسیک کد ۱ (پیوست ۵) را به کار بردیم. مقادیر ویژه حاصل عبارتند از:

[1] 7.359910e-01 2.626003e-01 1.492622e-01 6.990457e-02 2.956972e-02
 [6] 1.931184e-02 1.387779e-16 -1.139451e-02 -1.279569e-02 -2.849924e-02
 [11] -4.251502e-02 -5.255450e-02 -7.406143e-02 -1.097833e-01

همان‌طور که مشاهده می‌شود، برخی مقادیر ویژه منفی هستند. لذا برای تعیین بعد مناسب که یک نمایش خوب از ماتریس فاصله‌ی مشاهده شده را داشته باشیم، مقادیر معیارهای $P_m^{(1)}$ و $P_m^{(2)}$ را محاسبه می‌کنیم (کد ۲ در پیوست ۵) مقادیر $P_m^{(1)}$:

[1] 0.4605000 0.6248056 0.7181970 0.7619353 0.7804367 0.7925199 0.7925199
 [8] 0.7996493 0.8076554 0.8254870 0.8520881 0.8849708 0.9313100 1.0000000

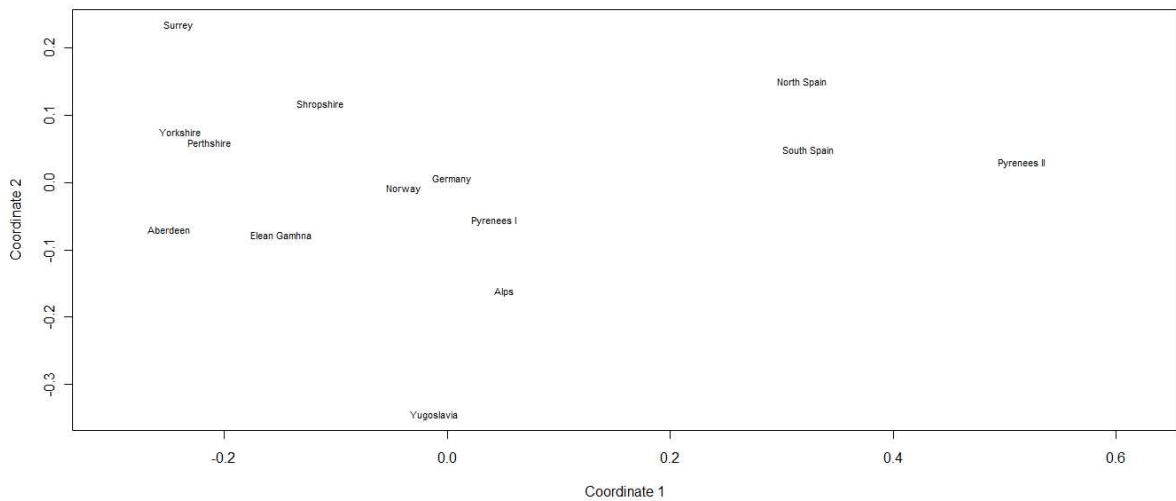
مقادیر $P_m^{(2)}$:

[1] 0.8179213 0.9220468 0.9556875 0.9630662 0.9643865 0.9649496 0.9649496
 [8] 0.9651457 0.9653929 0.9666193 0.9693486 0.9735191 0.9818014 1.0000000

با توجه به مقادیر معیار $P_m^{(2)}$ تعداد بعد دو برای یک نمایش خوب از ماتریس فاصله‌های مشاهده شده مناسب می‌باشد. اما مقادیر معیار $P_m^{(1)}$ تعداد سه و حتی چهار بعد را برای نمایش خوب پیشنهاد می‌کند. اما از آنجایی که هدف MDS دستیابی به بعد کوچکتر و نمایش بهتر داده‌ها است، لذا بعد دو را به عنوان بعد مناسب انتخاب می‌کنیم. لذا می‌توانیم پیکره مورد نظر را در یک نقشه دو بعدی به صورت زیر ترسیم کنیم (کد ۳ در پیوست ۵) با توجه به نمودار به نظر می‌رسد که جمعیت موش‌های صحرائی-آبی در شش منطقه انگلستان نزدیک به جمعیت این موش‌ها در آلپ، یوگسلاوی، آلمان، نروژ، پیرنه I و از نوع ترستریس بوده و جمعیت این موش‌ها در اسپانیای شمالی و جنوبی و پیرنه II نزدیک به هم و از نوع سپیداس می‌باشند.

۵.۱ مقیاس‌گذاری چندبعدی نامتری

در برخی مطالعات روان‌شناسی و در تحقیقات بازاری، ماتریس‌های مجاورت مربوط به اهداف بشری بر پایه‌ی داوری‌ها درباره‌ی شباهت‌ها یا عدم شباهت‌های اشیا می‌باشند. به عبارتی قضاوت‌های انسانی درباره‌ی ماتریس مجاورت را می‌سازند. هنگام جمع‌آوری و تحلیل بر روی داده‌هایی از این نوع، ممکن است محقق احساس کند که تنها می‌تواند ترتیبی برای داوری‌های خود در خصوص اشیا را در نظر بگیرد.



شکل ۱۰.۱: نمایش دوبعدی از مقیاس‌گذاری چندبعدی برای ماتریس فاصله‌ی موش‌های صحرائی-آبی

به طور مثال هنگام مقایسه‌ی مجموعه‌ای از رنگ‌ها، ممکن است تنها قادر باشیم، با درصدی از اطمینان، مشخص کنیم که یک رنگ روشن‌تر از دیگری است، اما به مراتب اطمینان این که بخواهیم مقداری برای این که چقدر روشن‌تر است در نظر بگیریم، کم خواهد بود.

چنین بررسی‌هایی منجر شد که در سال ۱۹۶۰ تحقیقی برای یک روش مقیاس‌گذاری چندبعدی که تنها از ترتیب مرتبه‌ی تقریب‌ها برای تولید یک نمایش فضایی از آن‌ها استفاده می‌کند، انجام شود. به عبارت دیگر، روشی که تحت تبدیلات یکنوا از ماتریس مجاورت مشاهده‌شده، ثابت باشد. به طور مثال، اگر مقادیر عددی مجاورت‌های مشاهده شده تغییر کنند، ترتیب رتبه‌های آن‌ها تغییر نکند و مختصات حاصل همان طور بدون تغییر باقی بماند.

برای این‌که روش کار در این نوع مقیاس‌گذاری را بهتر متوجه شویم، ابتدا فرض کنیم d_{ij} نشان دهنده فاصله اقلیدسی مشاهده شده بین مشاهدات x_i و x_j باشد که به عنوان مجاورت مشاهده شده در نظر می‌گیریم. فاصله مجاورت واقعی متناظر که در عمل مجهول است را با δ_{ij} نشان می‌دهیم. ایده اصلی این است که اگر یک ترتیب بین δ ها برقرار باشد این ترتیب تقریباً در بین d ها رعایت شود. حال سوال این است که آیا می‌توانیم فاصله‌های اقلیدسی را در فضایی با بعد کمتر طوری بیابیم که همچنان این ترتیب رعایت شود. پاسخ مثبت است. در ادامه فاصله‌های اقلیدسی در فضای تقلیل یافته (فضای با بعد کمتر) را با \hat{d} نشان داده و با اعمال قید یکنوایی و به‌کارگیری رگرسیون یکنوا سعی می‌کنیم آن‌ها را بیابیم. بنابراین هدف مقیاس‌گذاری نامتری به طور خلاصه عبارتست از:

ساخت پیکره‌ای از مشاهدات در بعد کوچکتر از بعد واقعی مشاهدات، به قسمی که رتبه‌بندی (ترتیب) فواصل آن‌ها در این پیکره تقریباً با ترتیب درایه‌های ماتریس واقعی یکی (منطبق) باشد.

مختصات‌ها در نمایش فضایی مجاورت‌های مشاهده شده، منجر به تولید فاصله برآزش شده d_{ij} می‌شوند که این فاصله‌ها طبق فرمول زیر وابسته به یک مجموعه از اعداد که آن را اختلافات^{۵۸} نامیده و با \hat{d}_{ij}

^{۵۸}Disparities

| | |
|-----------------|-------|
| مقدار تنش | برازش |
| $S_m \geq 20\%$ | ضعیف |
| $S_m = 10\%$ | معقول |
| $S_m \leq 5\%$ | خوب |
| $S_m = 0$ | عالی |

جدول ۳.۱: میزان برازش بر اساس معیار تنش

نشان می‌دهیم، هستند:

$$d_{ij} = \hat{d}_{ij} + \epsilon_{ij},$$

که در آن عبارات ϵ_{ij} خطا هستند. این خطاها اشاره به کلمه تقریباً در هدف فوق دارند. با این فرمول‌بندی، هدف مقیاس‌گذاری نامتری دستیابی به \hat{d}_{ij} در فضایی با بعد کمتر از بعد واقعی مشاهدات است. خطاهای انحرافی به دلیل متناظر نبودن فاصله‌ها با یک پیکره در تعداد خاص بعدهای انتخاب شده، ایجاد شده است. انحرافات با مجاورت‌های مشاهده شده یکنوا است و به خاطر این محدودیت، مانند فاصله‌های برازش شده تا حد ممکن دقیق می‌باشند. به طور کلی تنها یک محدودیت یکنوایی ضعیف به کار رفته است. بنابراین اگر عدم شباهت‌های مشاهده شده δ_{ij} از کمترین به بیشترین به صورت زیر مرتب شوند

$$\delta_{i_1j_1} < \delta_{i_2j_2} < \dots < \delta_{i_Nj_N}$$

که در آن $N = \frac{1}{2}n(n-1)$ ، آنگاه داریم

$$\hat{d}_{i_1j_1} < \hat{d}_{i_2j_2} < \dots < \hat{d}_{i_Nj_N}$$

برای یافتن اختلافات بکار رفته و سپس مختصات لازم در نمایش فضایی عدم شباهت‌های مشاهده شده که با $\hat{X}_{n \times m}$ یا \hat{X} نشان می‌دهیم، از رگرسیون یکنوا استفاده می‌کنیم (برای آگاهی بیشتر فرهادی، ۱۳۹۳ را ببینید). در این روش معیاری به نام تنش^{۵۹} را که با S نشان می‌دهیم، مینیمم می‌کنیم که در آن

$$S(\hat{X}) = \min \frac{\sum_{i < j} (\hat{d}_{ij} - d_{ij})^2}{\sum_{i < j} d_{ij}^2}$$

در واقع تنش‌ها حدی را که ترتیب رتبه‌ی فاصله‌های برازش شده با ترتیب رتبه‌ی عدم شباهت‌های مشاهده شده مخالف است، نمایش می‌دهند. در فرمول فوق مخرج کسر به دلیل حذف تغییرات مقیاس در نمایش فضایی نهایی اضافه شده است. به ازای هر مقدار از تعداد بعدهای m ، در پیکره فضایی، پیکره‌ای که کوچکترین تنش را دارد پیکره بهترین برازش در بعد m ، S_m نامیده می‌شود و معیار داوری برازش به صورت جدول ۲.۱ است. برازش عالی تنها اگر ترتیب رتبه‌ی فاصله‌های برازش شده با ترتیب رتبه‌ی عدم شباهت‌های مشاهده شده سازگار باشد، اتفاق می‌افتد که البته در عمل خیلی نادر است.

به عبارت ساده‌تر به ازای هر بعد (m) ، از یک تا چند مقدار خاص (که به نظر متخصص بستگی دارد) کمیت S_m را محاسبه می‌کنیم و بر اساس جدول ۲۰۱ به ازای هر m میزان برازش مشخص می‌شود.

۶.۱ تنش

در این بخش تعریفی دقیق‌تر از تنش و ساختارش را مطرح می‌کنیم. تنش از مجموع توان دوم باقی‌مانده‌ها به دست می‌آید که آن را با نماد S نمایش می‌دهند، و در کاهش بعد تاثیرگذار است؛ هر چه مقدار S کوچکتر باشد بهتر است. مقدار S مثبت بوده و می‌توان تنش را به صورت درصد بیان کرد. همان‌طور که گفته شد عدم شباهت بین دو شی i ام و j ام را با δ_{ij} نشان می‌دهیم و فرض می‌کنیم $\delta_{ij} = \delta_{ji}$ و $\delta_{ii} = 0$. بنابراین برای n شی تعداد $n(n-1)/2$ عدم‌شباهت داریم. حال عدم‌شباهت‌های مشاهده شده δ_{ij} را به صورت صعودی رتبه‌بندی می‌کنیم:

$$\delta_{i_1j_1} < \delta_{i_2j_2} < \dots < \delta_{i_Nj_N}$$

می‌خواهیم n شی را با n نقطه در فضای m بعدی نمایش دهیم؛ برای این منظور نقاط را x_1, \dots, x_n در نظر گرفته و فرض می‌کنیم مقدار m را می‌دانیم. پس n نقطه در فضای m بعدی داریم که پیکره را می‌سازد. علاقه‌مندیم پیکره‌ای که برای نمایش داده‌ها بهتر است را بیابیم. حال \hat{d}_{ij} را دنباله‌ای یکنواخت که تقریباً برابر با d_{ij} ها است به صورت زیر تعریف می‌کنیم:

$$\hat{d}_{i_1j_1} \leq \hat{d}_{i_2j_2} \leq \dots \leq \hat{d}_{i_mj_m}$$

بنابراین با مقادیر برازش شده d_{ij} ، تنش را می‌توان به صورت زیر محاسبه کرد:

$$S^* = \sum_{i < j} (d_{ij} - \hat{d}_{ij})^2$$

فرمول تنش فوق فاقد خواص مطلوب بوده و تحت رفتارهای پیکره (دوران، انتقال، انعکاس)، کشیدگی و چولگی آن ثابت نیست. برای مثال اگر پیکره x_1, \dots, x_n را به وسیله عامل k کشیده کنیم پیکره به kx_1, \dots, kx_n تبدیل می‌شود یعنی هر نقطه $(x_{i_1}, \dots, x_{i_p})$ به $(kx_{i_1}, \dots, kx_{i_p})$ تغییر می‌کند (به تبع آن مقدار \hat{d}_{ij} نیز تغییر می‌کند). در نتیجه تنش از S^* به $k^2 S^*$ تغییر می‌کند. روشن است که برای رفع این عیب باید فرمول تنش فوق را بر یک فاکتور مقیاس تقسیم کرد، برای این منظور فاکتور مقیاس را $T^* = \sum d_{ij}^2$ در نظر می‌گیریم. بنابراین S^*/T^* بهترین اندازه برازش است زیرا همه‌ی خواص مطلوب S^* را داشته به علاوه نسبت به تغییر رفتارهای پیکره، کشیدگی و چولگی آن ثابت است. در نهایت مانند محاسبه انحراف استاندارد از واریانس برای کاهش بزرگی اثر توان دوم در فرمول، ریشه‌ی دوم را به صورت زیر در نظر می‌گیریم:

$$\text{Stress} = S = \sqrt{\frac{S^*}{T^*}} = \sqrt{\frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}}$$

از آن جا که هر چه مقدار تنش کمتر باشد برازش بهتری داریم، در نهایت تنش را در محاسبات به صورت زیر لحاظ می‌کنیم:

$$S(x_1, \dots, x_n) = \min \sqrt{\frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}}$$

حال، روش کمی ارزیابی هر پیکره را داریم. واضح است پیکره‌ای مطلوب است که کمترین تنش را داشته و بهترین برازش پیکره زمانی رخ می‌دهد که تنش حداقل باشد (کروسکال، ۱۹۶۴).

مثال ۱.۶.۱. در این مثال روش مقیاس‌گذاری نامتری را برای دستیابی به یک نمایش فضایی از عدم تشابه داوری‌ها در مکتب تعدادی از رهبران جهان و سیاست‌مداران برجسته در زمان جنگ جهانی دوم برای تشکیل یک پیکره فکری به کار می‌بریم. ماتریس فاصله در جدول ۳.۱ داده شده است.

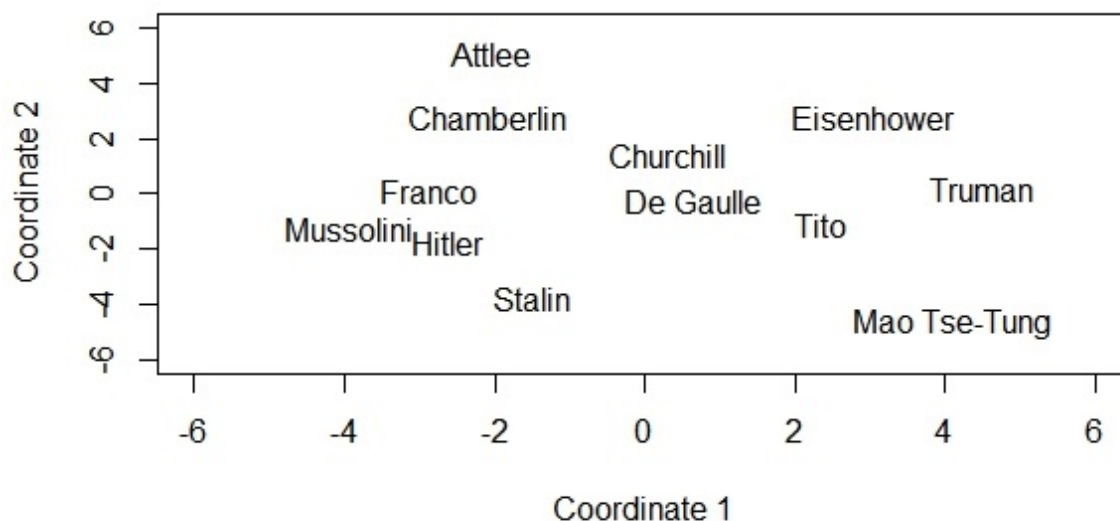
| | <i>Tit</i> | <i>Chm</i> | <i>Trm</i> | <i>MT</i> | <i>DGI</i> | <i>Frn</i> | <i>Att</i> | <i>Stl</i> | <i>Esn</i> | <i>Chr</i> | <i>Mss</i> | <i>Hit</i> |
|----------------------|------------|------------|------------|-----------|------------|------------|------------|------------|------------|------------|------------|------------|
| <i>Hitler</i> | ۰ | ۳ | ۴ | ۷ | ۳ | ۸ | ۳ | ۴ | ۸ | ۹ | ۴ | ۷ |
| <i>Mussolini</i> | ۳ | ۰ | ۶ | ۸ | ۵ | ۹ | ۲ | ۴ | ۹ | ۹ | ۵ | ۸ |
| <i>Churchill</i> | ۴ | ۶ | ۰ | ۴ | ۶ | ۳ | ۵ | ۳ | ۸ | ۵ | ۵ | ۲ |
| <i>Eisenhower</i> | ۷ | ۸ | ۴ | ۰ | ۸ | ۸ | ۷ | ۵ | ۹ | ۴ | ۴ | ۴ |
| <i>Stalin</i> | ۳ | ۵ | ۶ | ۸ | ۰ | ۹ | ۶ | ۶ | ۶ | ۷ | ۷ | ۷ |
| <i>Attlee</i> | ۸ | ۹ | ۳ | ۹ | ۸ | ۰ | ۷ | ۵ | ۹ | ۸ | ۲ | ۸ |
| <i>Franco</i> | ۳ | ۲ | ۵ | ۷ | ۶ | ۷ | ۰ | ۴ | ۸ | ۸ | ۲ | ۳ |
| <i>DeGaulle</i> | ۴ | ۴ | ۳ | ۵ | ۶ | ۵ | ۴ | ۰ | ۷ | ۴ | ۵ | ۲ |
| <i>MaoTse – Tung</i> | ۸ | ۹ | ۸ | ۹ | ۶ | ۹ | ۸ | ۷ | ۰ | ۴ | ۹ | ۴ |
| <i>Truman</i> | ۹ | ۹ | ۵ | ۴ | ۷ | ۸ | ۸ | ۴ | ۴ | ۰ | ۵ | ۵ |
| <i>Chamberlin</i> | ۴ | ۵ | ۵ | ۴ | ۷ | ۲ | ۲ | ۵ | ۹ | ۵ | ۰ | ۷ |
| <i>Tito</i> | ۷ | ۸ | ۲ | ۴ | ۷ | ۸ | ۳ | ۲ | ۴ | ۵ | ۷ | ۰ |

جدول ۵.۱: ماتریس فاصله داده‌های رهبران سیاسی جنگ جهان دوم

داوری‌ها برپایه یک مقیاس ۹ نقطه‌ای انجام شده‌اند. بدین معنی که نقاط مقیاس ۱ و ۹ به ترتیب نشانگر (افکار) خیلی مشابه و خیلی نامشابه هستند. با انجام مقیاس‌گذاری چندبعدی نامتری نتایج زیر حاصل می‌شوند (کد مثال ۱۰۶۰۱ در پیوست ۵)

```
initial value 20.400713
iter 5 value 15.219311
iter 5 value 15.207733
iter 5 value 15.207733
final value 15.207733
converged
$points
$stress
[1] 15.20773
```

از آنجایی که دو بعد برای ترسیم پیکره مورد نظر کافی است، می‌توان نقشه پیکره فکری رهبران سیاسی جنگ جهانی دوم را به صورت شکل ۲۰۱ رسم کرد.



شکل ۲۰۱: نقشه پیکره فکری رهبران سیاسی جنگ جهانی دوم

با توجه به نمودار واضح است که سه گروه رهبران فاشیست همانند نخست وزیران انگلیسی عمل کرده‌اند. دیدگاه فکری-سیاسی استالین^{۶۰} و ماوتسیتونگ^{۶۱} جدای از دیگر رهبران است. همچنین

^{۶۰}Stalin

^{۶۱}Mao Tse Tung

به نظر می‌رسد که ایزن‌هاور^{۶۲} به دولت انگلیسی بیشتر از ریس‌جمهور خود ترومن^{۶۳} وابسته باشد. به طور جالب دی‌گول^{۶۴} در مرکز جواب MDS قرار گرفته است.

^{۶۲}Eisenhower

^{۶۳}Truman

^{۶۴}de Gualle

فصل ۲

مقیاس‌گذاری چند بعدی بیزی با پیشین نرمال

۱.۲ مقدمه

در این فصل دو مسئله‌ی اصلی در MDS متری، یعنی تعیین پیکره اشیا و بعد مناسب را در یک چارچوب بیزی مورد مطالعه و بررسی قرار می‌دهیم. برای این منظور یک مدل فاصله اقلیدسی معرفی و یک خطای اندازه‌گیری گوسی (نرمال) برای عدم شباهت‌ها در نظر می‌گیریم. تحت این مدل یک الگوریتم مونت کارلو زنجیر مارکوفی^۱ (MCMC) را برای به‌دست آوردن یک جواب بیزی برای پیکربندی اشیا پیشنهاد کرده و بر اساس پیکره برآورد شده یک معیار اطلاع MDS^۲، (MDSIC) را برای انتخاب بعد مناسب ارائه می‌دهیم.

در پایان فصل نتایج شبیه‌سازی ارایه شده حاکی از آن است که روش MDS بیزی (BMDS) نتایج بهتری نسبت به روش MDS کلاسیک (CMDS) برای پیکره اشیا فراهم می‌کند. آنچه در این فصل انجام می‌دهیم یک تحلیل بیزی از MDS متری می‌باشد. مطالب این فصل برگرفته از مقاله آه و رافتی (۲۰۰۱) می‌باشند.

۲.۲ معرفی مدل

می‌دانیم داده‌های عدم شباهت از روش‌های مختلفی به‌دست می‌آیند، و از آن‌جا که فاصله‌ی اقلیدسی، نسبت به انتخاب بعد در مقایسه با دیگر اندازه‌های فاصله، غیرحساس است، در مواردی که بعد مجهول است بیشتر مورد استفاده قرار گرفته و به کارگیری آن آسان است (دیوسن، ۱۹۸۳). بدین ترتیب فرض کنید اندازه عدم شباهت δ_{ij} که نشان‌دهنده فاصله بین اشیا i و j در یک فضای اقلیدسی است، به صورت

^۱Markov Chain Monte Carlo

^۲MDS information criterion

زیر محاسبه شود

$$\delta_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (1.2)$$

که عدم شباهت δ_{ij} به p صفت مشاهده نشده اشیا وابسته بوده و x_{ik} ، نشان دهنده مقدار مشاهده نشده k امین صفت i امین شی پیکره اشیا X است که

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ik} & \cdots & x_{ip} \\ \vdots & & & & & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & & & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix} = (X_1, \dots, X_n)^T$$

و $X_i \in \mathbb{R}^p$. در واقعیت اغلب خطای اندازه‌گیری در مشاهدات وجود دارد بنابراین فرض می‌کنیم که اندازه عدم شباهت مشاهده شده d_{ij} برابر با اندازه درست δ_{ij} به اضافه یک خطای اندازه‌گیری است (لازم به ذکر است که باید اندازه عدم شباهت مشاهده شده همیشه مثبت باشد).

به طور دقیق‌تر، مدل d_{ij} را نرمال بریده شده با پارامترهای δ_{ij}, σ^2 به صورت زیر در نظر می‌گیریم:

$$d_{ij} \sim N(\delta_{ij}, \sigma^2) I(d_{ij} > 0), \quad i \neq j, i, j = 1, 2, \dots, n. \quad (2.2)$$

در نتیجه تابع درستنمایی پارامترهای مجهول $X = \{x_i\}$ و σ^2 به صورت زیر به دست می‌آید:

$$\begin{aligned} L(X, \sigma^2 | D) &= \prod_{\substack{i \neq j \\ i, j=1}}^n \frac{\frac{1}{\sigma} \phi\left(\frac{d_{ij} - \delta_{ij}}{\sigma}\right)}{1 - \Phi\left(\frac{-\delta_{ij}}{\sigma}\right)} \\ &= \prod_{\substack{i \neq j \\ i, j=1}}^n \left(\frac{1}{\sqrt{2\pi}\sigma}\right) \exp\left\{-\frac{1}{2\sigma^2}(d_{ij} - \delta_{ij})^2\right\} \frac{1}{1 - \Phi\left(\frac{-\delta_{ij}}{\sigma}\right)} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\frac{n(n-1)}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum (d_{ij} - \delta_{ij})^2\right\} \frac{1}{\prod_{\substack{i \neq j \\ i, j=1}}^n (1 - \Phi\left(\frac{-\delta_{ij}}{\sigma}\right))} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^{\frac{n(n-1)}{2}} \left(\frac{1}{\sigma}\right)^{\frac{n(n-1)}{2}} \exp\left\{-\frac{1}{2\sigma^2} \left(\sum (d_{ij} - \delta_{ij})^2\right)\right\} \frac{1}{\prod_{\substack{i \neq j \\ i, j=1}}^n \Phi\left(\frac{\delta_{ij}}{\sigma}\right)} \\ &\propto (\sigma^2)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2\sigma^2} SSR\right\} \exp\left\{\log\left(\prod_{i>j} \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right)^{-1}\right\} \\ &\propto (\sigma^2)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2\sigma^2} SSR - \sum_{i \neq j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\} \end{aligned} \quad (3.2)$$

که در آن D نشان‌دهنده مشاهدات، $m = n(n-1)/2$ تعداد عدم شباهت‌ها، $SSR = \sum_{i>j} (d_{ij} - \delta_{ij})^2$ مجموع توان‌های دوم باقی مانده‌ها، $\phi(\cdot)$ و $\Phi(\cdot)$ به ترتیب توابع چگالی احتمال و توزیع تجمعی نرمال استاندارد می‌باشند.

لازم به ذکر است که پارامترهای مورد نظر مشاهدات x_i و σ^2 می‌باشند و از آنجایی که δ_{ij} ها به مشاهدات وابسته هستند، تابع درستنمایی را برای δ_{ij} و σ^2 نوشتیم.

۳.۲. توزیع پیشین

برای انجام یک تحلیل بیزی لازم است که توزیع پیشین^۴ پارامترهای X و σ^2 را مشخص کنیم. برای این منظور فرض می‌کنیم

$$x_1, \dots, x_n \sim N_p(\circ, \Lambda) \quad (۴.۲)$$

که در آن

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p) = \begin{pmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \lambda_3 & & \\ & & & \ddots & \\ \circ & & & & \lambda_p \end{pmatrix}.$$

و برای ابر پارامتر $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_p)$ با بعد مفروض p ، فرض می‌کنیم

$$\lambda_j \sim IG(\alpha, \beta_j) \quad j = 1, \dots, p$$

در نهایت فرض می‌کنیم مولفه σ^2 دارای توزیع گامای معکوس با مد $1 + b/a$ باشد به عبارتی

$$\sigma^2 \sim IG(a, b).$$

به علاوه پیشین‌های توام X و Λ ، σ^2

را مستقل از یکدیگر در نظر می‌گیریم. در این صورت اگر $\pi(X, \sigma^2, \Lambda)$ ، $\pi(X)$ ، $\pi(\sigma^2)$ و $\pi(\Lambda)$ به ترتیب نشان‌دهنده توزیع‌های پیشین توام X, σ^2, Λ ، توزیع پیشین حاشیه‌ای X ، σ^2 و Λ باشند، چگالی (توزیع) پیشین توام همه‌ی پارامترهای مدل به صورت زیر حاصل می‌شود:

$$\pi(X, \sigma^2, \Lambda) = \pi(X|\Lambda)\pi(\Lambda)\pi(\sigma^2).$$

^۳Data

^۴Prior distribution

۴.۲ توزیع پسین

اگر $\pi(X, \sigma^2, \Lambda|D)$ نشان‌دهنده‌ی چگالی پسین پارامترهای مدل به شرط داده‌ها باشد. آنگاه با استفاده از تابع درستنمایی و توزیع پیشین توام پارامترها، تابع چگالی پسین^۵ توام همه‌ی پارامترهای مدل به صورت زیر نتیجه می‌شود:

$$\begin{aligned}
 \pi(X, \sigma^2, \Lambda|D) &\propto L(X, \sigma^2|D)\pi(X, \sigma^2, \Lambda) \\
 &\propto L(X, \sigma^2|D)\pi(\sigma^2)\pi(\Lambda)\pi(X) \\
 &\propto (\sigma^2)^{-\frac{m}{\nu}} \exp\left\{-\frac{1}{\nu\sigma^2}SSR - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\} \times \frac{b^a}{\Gamma a}(\sigma^2)^{-a-1} \exp\left\{-\frac{b}{\sigma^2}\right\} \\
 &\times \prod_{j=1}^p \frac{\beta_j^\alpha}{\Gamma \alpha} \lambda_j^{-\alpha-1} \exp\left\{-\frac{\beta_j}{\lambda_j}\right\} \times \prod_{i=1}^n \frac{1}{|\Lambda|^{\frac{1}{\nu}}} \exp\left\{-\frac{1}{\nu}x_i'\Lambda^{-1}x_i\right\} \\
 &\propto \exp\left\{-\frac{1}{\nu\sigma^2}SSR - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right) - \frac{b}{\sigma^2} - \frac{1}{\nu} \sum_{i=1}^n x_i'\Lambda^{-1}x_i - \sum_{j=1}^p \frac{\beta_j}{\lambda_j}\right\} \\
 &\times (\sigma^2)^{-\left(\frac{m}{\nu}+a+1\right)} \left(\prod_{j=1}^p \lambda_j\right)^{-\frac{n}{\nu}-\alpha-1}. \tag{۵.۲}
 \end{aligned}$$

بنابراین با استفاده از توزیع پسین توام X, σ^2, Λ ، می‌توان توزیع‌های شرطی کامل پارامترها را به دست آورد. چگالی پسین حاشیه‌ایی λ_j با استفاده از روابط فوق به صورت زیر نتیجه می‌شود:

$$\begin{aligned}
 \pi(\lambda_j|X, \sigma^2, D) &\propto \pi(x_i|\lambda_j)\pi(\lambda_j)L(X, \sigma^2|D) \\
 &\propto \prod_{i=1}^n \frac{1}{|\Lambda|^{\frac{1}{\nu}}} \exp\left\{-\frac{1}{\nu}x_i'\Lambda^{-1}x_i\right\} \times \prod_{j=1}^p \frac{\beta_j^\alpha}{\Gamma \alpha} \lambda_j^{-\alpha-1} \exp\left\{-\frac{\beta_j}{\lambda_j}\right\} \\
 &\times (\sigma^2)^{-\frac{m}{\nu}} \exp\left\{-\frac{1}{\nu\sigma^2}SSR - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\} \\
 &\propto \left(\prod_{j=1}^p \lambda_j\right)^{-\frac{n}{\nu}-\alpha-1} \exp\left\{-\frac{1}{\nu} \sum_{i=1}^n x_i'\Lambda^{-1}x_i - \sum_{j=1}^p \frac{\beta_j}{\lambda_j}\right\} \\
 &\propto \left(\prod_{j=1}^p \lambda_j\right)^{-\frac{n}{\nu}-\alpha-1} \exp\left\{-\frac{1}{\nu} \sum_{i=1}^n \sum_{j=1}^p \frac{x_{ij}^2}{\lambda_j} - \sum_{j=1}^p \frac{\beta_j}{\lambda_j}\right\} \\
 &\propto \left(\prod_{j=1}^p \lambda_j\right)^{-\frac{n}{\nu}-\alpha-1} \exp\left\{\sum_{j=1}^p -\frac{1}{\lambda_j} \left(\frac{S_j}{\nu} + \beta_j\right)\right\}
 \end{aligned}$$

^۵Posterior distribution

که در آن

$$S_j = \sum_{i=1}^n x'_i x = \sum_{i=1}^n x_{ij}^2 \quad (۶.۲)$$

بنابراین می‌توان نتیجه گرفت توزیع پسین حاشیه‌ای λ_j ها گامای معکوس با پارامترهای $\alpha + n/2$ و $\beta_j + S_j/2$ است. به عبارتی

$$\pi(\lambda_j | X, \sigma^2, D) \sim IG\left(\alpha + \frac{n}{2}, \beta_j + \frac{S_j}{2}\right)$$

چگالی شرطی کامل x_i ، به صورت زیر به دست می‌آید:

$$\begin{aligned} \pi(x_i | \sigma^2, \Lambda, D) &\propto L(x_i, \sigma^2 | D) \pi(x_i | \Lambda) \\ &\propto (\sigma^2)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2\sigma^2} SSR - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\} \\ &\times \prod_{i=1}^n \frac{1}{|\Lambda|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} x' \Lambda^{-1} x\right) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2} SSR - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right) - \frac{1}{2} \sum_{i>j} \Lambda^{-1} x' x\right\} \quad (۷.۲) \end{aligned}$$

همچنین چگالی پسین σ^2 به صورت زیر نتیجه می‌شود:

$$\begin{aligned} \pi(\sigma^2 | X, \Lambda, D) &\propto L(X, \sigma^2 | D) \pi(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2\sigma^2} SSR - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\} \times \frac{b^a}{\Gamma a} (\sigma^2)^{-a-1} \exp\left\{-\frac{b}{\sigma^2}\right\} \\ &\propto (\sigma^2)^{-\frac{m}{2}-a-1} \exp\left\{-\frac{1}{2\sigma^2} SSR - \frac{b}{\sigma^2}\right\} \\ &\propto (\sigma^2)^{-\left(\frac{m}{2}+a\right)-1} \exp\left\{-\frac{1}{\sigma^2} \left(\frac{SSR}{2} + b\right)\right\} \end{aligned}$$

واضح است که توزیع پسین حاشیه‌ای σ^2 ، توزیع گامای معکوس با پارامترهای $m/2 + a$ و $SSR/2 + b$ است، به عبارتی

$$\pi(\sigma^2 | X, \Lambda, D) \sim IG\left(\frac{m}{2} + a, \frac{SSR}{2} + b\right)$$

و چنانچه بعد p را به عنوان یک متغیر مجهول در نظر گرفته و احتمال پیشین برای تمام مقادیر p را برابر با p_{max} فرض کنیم، آنگاه پسین توام همه‌ی پارامترها به علاوه p به صورت زیر می‌باشد:

$$\begin{aligned} \pi(X, \sigma^2, \Lambda, p|D) &\propto L(X, \sigma^2, p|D)\pi(X|\Lambda)\pi(\sigma^2)\pi(\Lambda|p)\pi(p) \\ &= (\sigma^2)^{-\frac{m}{\nu}} \left(\frac{1}{\sqrt{\nu\pi}}\right)^{\frac{n(n-1)}{2}} \exp\left\{-\frac{1}{2\sigma^2}SSR - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\} \\ &\times \prod_{i=1}^n \prod_{j=1}^p \frac{1}{(\nu\pi)^{\frac{n}{\nu}}|\Lambda|^{\frac{1}{\nu}}} \exp\left\{-\frac{1}{\nu}x'\Lambda^{-1}x\right\} \frac{\beta_j^\alpha}{\Gamma(\alpha)}(\lambda_j)^{-\alpha-1} \exp\left\{-\frac{\beta_j}{\lambda_j}\right\} \\ &\times \frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1} \exp\left\{-\frac{b}{\sigma^2}\right\} p_{max} \\ &= (\nu\pi)^{-\frac{m}{\nu}} \sigma^{2\left(-\frac{m}{\nu}\right)} \exp\left\{-\frac{1}{2\sigma^2}SSR - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\} \\ &\times (\nu\pi)^{-\frac{np}{\nu}} \prod_{j=1}^p \lambda_j^{-\frac{n}{\nu}} \exp\left\{-\sum_{j=1}^p \sum_{i=1}^n \frac{1}{\nu\lambda_j}x'x\right\} \\ &\times \Gamma(a)^{-1} b^a (\sigma^2)^{-(a+1)} \exp\left\{-\frac{b}{\sigma^2}\right\} \times \Gamma(\alpha)^{-p} \prod_{j=1}^p \beta_j^\alpha \lambda_j^{-(\alpha+1)} \exp\left\{-\frac{\beta_j}{\lambda_j}\right\} p_{max} \\ &= A(p)h(\sigma^2, X)g(\Lambda, X, p) \end{aligned}$$

که در آن

$$A(p) = (\nu\pi)^{-\frac{(m+np)}{\nu}} \Gamma(a)^{-1} b^a \Gamma(\alpha)^{-p} \prod_{j=1}^p \beta_j^\alpha p_{max} \quad (۸.۲)$$

$$h(\sigma^2, X) = (\sigma^2)^{-\left(\frac{m}{\nu}+a+1\right)} \times \exp\left\{-\frac{\left(\frac{SSR}{\nu}+b\right)}{\sigma^2} - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\} \quad (۹.۲)$$

$$g(\Lambda, X, p) = \prod_{j=1}^p \lambda_j^{-\left(\frac{n}{\nu}+\alpha+1\right)} \exp\left\{-\frac{\frac{S_j}{\nu} + \beta_j}{\lambda_j}\right\} \quad (۱۰.۲)$$

بنابراین چگالی توزیع پسین حاشیه‌ایی (X, p) ، که در انتخاب بعد مناسب به‌کار می‌رود، به صورت زیر حاصل می‌شود:

$$\begin{aligned} \pi(X, p|D) &= cA(p) \int h(\sigma^2, X) d\sigma^2 \int g(\Lambda, X, p) d\Lambda \\ &\approx cA^*(p) \left(\frac{SSR}{m}\right)^{-\frac{m}{\nu}+1} \times \prod_{j=1}^p \left(\frac{S_j}{\nu} + \beta_j\right)^{-\left(\frac{n}{\nu}+\alpha\right)} \end{aligned} \quad (۱۱.۲)$$

که در محاسبه آن از روابط زیر استفاده کرده‌ایم:

$$\begin{aligned} \int g(\Lambda, X, p) d\Lambda &= \int \dots \int \prod_{j=1}^p \lambda_j^{-(\frac{n}{\nu} + \alpha + 1)} \exp\left\{-\frac{S_j + \beta_j}{\lambda_j}\right\} d\lambda_1, \dots, d\lambda_p \\ &= \prod_{j=1}^p \left(\frac{\Gamma(\frac{n}{\nu} + \alpha)}{(\frac{S_j}{\nu} + \beta_j)^{\frac{n}{\nu} + \alpha}} \right) \\ &= \Gamma^p\left(\frac{n}{\nu} + \alpha\right) \prod_{j=1}^p \left(\frac{S_j}{\nu} + \beta_j\right)^{-\left(\frac{n}{\nu} + \alpha\right)} \end{aligned} \quad (12.2)$$

و

$$A^*(p) = A(p) \cdot (2\pi)^{\frac{1}{\nu}} \Gamma^p\left(\frac{n}{\nu} + \alpha\right) \left(\frac{m}{\nu}\right)^{-\frac{1}{\nu}} \exp\left\{-\frac{m}{\nu}\right\}$$

و

$$\int h(\sigma^2, X) d\sigma^2 \approx (2\pi)^{\frac{1}{\nu}} \left(\frac{m}{\nu}\right)^{-\frac{1}{\nu}} \times \left(\frac{SSR}{m}\right)^{-\frac{m}{\nu} + 1} \exp\left\{-\frac{m}{\nu}\right\} \quad (13.2)$$

اثبات رابطه‌ی (۱۳.۲) در پیوست آمده است.

۵.۲ محاسبات عددی

در این بخش جزئیات محاسبات عددی مربوط به الگوریتم MCMC و تعیین بعد بهینه را شرح می‌دهیم.

۱.۵.۲ الگوریتم MCMC

همانطور که در بخش قبل دیدیم برخی توزیع‌های پسین صورت بستهبندی ندارند، لذا تولید داده از آن‌ها یا استفاده از آن‌ها در یک مثال واقعی امکان‌پذیر نیست. لذا در این بخش با استفاده از روش (تکنیک) MCMC سعی می‌کنیم از توزیع‌های پسین، داده تولید کرده و بعد مناسب را برای پیکربندی تعیین کنیم. همان‌طور که گفتیم مقیاس‌گذاری چند بعدی بیزی، پیکره اشیا در بعد کمتر را در یک چارچوب بیزی بازیابی می‌کند.

در بخش‌های قبل برای پارامترهای مجهول توزیع‌های پیشین مناسب را تعریف کردیم و چگالی‌های پسین متناظر با آن‌ها را به دست آوردیم. حال برای دستیابی به برآوردهای بیزی پارامترها، نیازمند به انتگرال‌گیری عددی هستیم اما به دلیل صورت پیچیده تابع چگالی پسین (۵.۲)، از یک الگوریتم MCMC (گیلکز و همکاران^۶، ۱۹۹۶) استفاده می‌کنیم. الگوریتم مورد نظر مقادیر جدید X, σ^2, Λ را به شرط مقادیر فعلی تولید کرده و این فرآیند را تا رسیدن به درجه‌ای از همگرایی تکرار می‌کند. (برای آگاهی بیشتر پیوست ۴ را ببینید).

برای پیاده‌سازی یک الگوریتم MCMC باید مقادیر اولیه پارامترها را تعیین کنیم. بنابراین ابتدا روند مقداردهی اولیه پارامترهای مجهول که برای الگوریتم MCMC لازم است، را مطرح می‌کنیم.

^۶Gilks

مقدار $x_i^{(o)}$ ، که از یک مقیاس‌گذاری کلاسیک به دست آمده را به عنوان مقدار اولیه x_i ها در نظر می‌گیریم. حاصل می‌تواند در مبدا مرکزی شده و سپس با تجزیه طیفی طوری تبدیل شود که یک ماتریس کوواریانس نمونه‌ای قطری، مطابق پیشین ذکر شده (۴.۲) را داشته باشیم. از $x_i^{(o)}$ های اصلاح شده می‌توان مجموع توان‌های دوم باقی‌مانده‌ها $SSR^{(o)}$ و $\sigma^2 = SSR^{(o)}/m$ را محاسبه کرد، و از σ^2 به عنوان یک مقدار اولیه برای σ^2 در الگوریتم استفاده می‌کنیم. همچنین عناصر روی قطر ماتریس واریانس - کواریانس X اصلاح شده به عنوان مقادیر اولیه λ_j در نظر گرفته می‌شوند.

حال، جزییات تولید نمونه از الگوریتم MCMC را شرح می‌دهیم. در هر تکرار، یک مقدار جدید λ_j از توزیع پسین شرطی به شرط مجهولات دیگر تولید می‌کنیم. برای تولید X_i و σ^2 از الگوریتم قدم زدن تصادفی متروپولیس - هستینگز^۷ (هستینگز^۸، ۱۹۷۰) استفاده می‌کنیم. چگالی پیشنهادی برای تولید نمونه از x_i را نرمال فرض می‌کنیم. برای انتخاب واریانس این چگالی پیشنهادی نرمال، با توجه به چگالی پسین x_i ، (۷.۲) را به صورت زیر فرمول‌بندی می‌کنیم،

$$\pi(x_i|D) \propto \exp \left\{ -\frac{1}{\sigma^2}(Q_1 + Q_2) - \sum_{\substack{j \neq i, \\ j=1}}^n \log \Phi \left(\frac{\delta_{ij}}{\sigma} \right) \right\}$$

که در آن $Q_1 = \frac{1}{\sigma^2} \sum_{j=1}^n (\delta_{ij} - d_{ij})^2$ و $Q_2 = x_i' \Lambda^{-1} x_i$. از آنجایی که Q_1 ، به تعداد $n - 1$ تابع درجه دوم از x_i ها با ضریب $1/\sigma^2$ ، $(\delta_{ij} - d_{ij})^2/\sigma^2$ را دارد در حالی که Q_2 تنها یک عبارت درجه دوم با ضریب Λ دارد، تنها Q_1 را در نظر گرفته و واریانس پسین x_i را با $1 - \sigma^2/n$ تقریب زده و واریانس چگالی پیشنهادی نرمال را به صورت مضربی ثابت از $1 - \sigma^2/n$ انتخاب می‌کنیم.

اُه و رفتاری (۲۰۰۱) با یک مطالعه عددی مقدماتی نشان دادند که تابع چگالی شرطی σ^2 به خوبی می‌تواند به وسیله تابع چگالی $IG(m/2 + a, SSR/2 + b)$ تقریب زده شود. بنابراین الگوریتم قدم زدن تصادفی متروبولیس - هستینگز را با چگالی پیشنهادی نرمال با واریانس متناسب با واریانس توزیع $IG(m/2 + a, SSR/2 + b)$ در نظر می‌گیریم.

تکرار تولید $\{X_i\}$ ، σ^2 و $\{\lambda_j\}$ ، با صرف یک زمان طولانی، یک نمونه از توزیع پسین پارامترهای مجهول فراهم می‌کند که با استفاده از این نمونه می‌توان برآورد بیزی پارامترها را به دست آورد. از آنجایی که یک فاصله اقلیدسی را برای مدل اندازه عدم شباهت δ_{ij} در نظر گرفتیم، نمونه‌های پسین $\{X_i\}$ باید تحت تبدیل، دوران و انعکاس حول مبدا، همانند MDS کلاسیک، پایا باشد. بنابراین برای تایید همگرایی MCMC لازم است همگرایی δ_{ij} پیش از X بررسی شود. زیرا در صورت عدم همگرایی X ، استفاده از میانگین‌های نمونه به عنوان برآورد بیزی درست نیست، چون نمونه‌های MCMC از X حتی وقتی فواصل δ_{ij} پایدار هستند، ممکن است ناپایدار باشند. برای برطرف کردن این مشکل و استفاده نکردن از میانگین‌های نمونه به عنوان برآورد بیزی، یک تقریب مد پسین از X را به عنوان یک برآورد بیزی از X ، جواب BMDS پیکربندی اشیا، در نظر می‌گیریم. مد پسین موقعیت‌های متناسب از X_i ها را متناظر با ماکسیمم چگالی پسین فراهم می‌کند.

^۷Random walk Metropolis-Hastings

^۸Hastings

۲.۵.۲ انتخاب بعد مناسب

همان طور که می‌دانیم هدف دیگر MDS، یافتن بعد مناسب در فضا است. به عبارتی یافتن صفت‌هایی که در بیان شباهت یا عدم شباهت نقش معنی‌داری دارند. اگر تعداد صفات زیاد باشند می‌توانند تفسیر ساختار داده‌ها را مشکل کنند. از طرفی اگر تعداد صفات خیلی کم باشند می‌توانند منجر به از دست رفتن صفات مهم و اساسی عدم شباهت شوند؛ بنابراین تعیین بعدی مهم است که در صورت امکان، با کمترین تعداد، صفات مهم را حفظ کند. اما با وجود اهمیت زیادی که بعد در بسیاری از کاربردها دارد، تحقیقات زیادی روی آن انجام نشده است. با این وجود استفاده از روش یافتن یک شکستگی (شبهه خم آرنج) یعنی نقطه‌ایی که در آن اندازه برازش یا سطح عدم شباهت تغییر می‌کند، رایج‌تر است (تورگوسون، ۱۹۵۸ را ببینید). البته، اغلب یافتن شکستگی (خم آرنج) مشکل است. این را در نظر داشته باشید که جنبه بصری یک نمودار ممکن است گمراه‌کننده باشد چون ظاهر آن اغلب به مقیاس محورها بستگی دارد (اسپنس و جراف^۹، ۱۹۷۴، دیوسن، ۱۹۸۳، برگ و همکاران، ۱۹۹۷ را ببینید) در این بخش یک معیار بیزی ساده برای انتخاب بعد مناسب بر اساس پیکره به‌دست آمده از BMDS با نام اختصاری MDSIC را ارائه می‌دهیم.

برای انتخاب بعد مناسب بین دو بعد p و p' ، پیکره برآورد شده بر اساس ابعاد p و p' ، (برای p یعنی پسین حاشیه‌ایی $\pi(X, p|D)$ برای X ‌ایی که از روش BMDS به‌دست آمده) را بررسی می‌کنیم و بعدی را که بزرگترین مقدار پسین را نتیجه دهد انتخاب می‌کنیم.

$$\begin{aligned}\pi(x, p|D) &= \int \pi(x, p, \sigma^2|D) d\sigma^2 \\ &= \int L(x, p, \sigma^2|D) \pi(\sigma^2) d\sigma^2\end{aligned}$$

و پسین حاشیه‌ایی

$$\int \pi(X, \Lambda, p) d\Lambda \approx cL(X, p|D) \pi(X, p)$$

(در بخش توزیع‌های پسین محاسبه شده‌اند) که در آن c یک ثابت مستقل از X و p ، و $L(X, p|D)$ درستنمایی حاشیه‌ایی (X, p) و $\pi(X, p)$ پیشین حاشیه‌ایی (X, p) است را در نظر بگیرید. با افزایش p پیکره با دقت بیشتری بازیابی می‌شود و در نتیجه مقدار SSR کوچک و در نتیجه درستنمایی حاشیه‌ایی افزایش و پیشین حاشیه‌ایی کاهش می‌یابد.

در مقایسه دو بعد $(x^{(p)}, p)$ و $(x^{(p+1)}, p+1)$ یک مشکل وجود دارد، و آن این است که برای یک فاصله اقلیدسی معین، بعد p بازیابی خوبی از پیکره دارد در حالی که بعد $p+1$ نسبت به p تاثیر چندانی در بهبود بازیابی پیکره ندارد اما از طرفی با افزایش بعد از p به $p+1$ مختصات X به مبداء نزدیکتر و S_j (واریانس نمونه‌ای مختصات j ام، X) کوچکتر می‌شود. که با توجه به عبارت $\prod_{j=1}^p (S_j/2 + \beta_j)^{-(n/2+\alpha)}$ در

^۹Spence and Graef

پسین حاشیه‌ایی، منجر به بزرگ شدن مقدار پسین حاشیه‌ایی و انتخاب بعد $p + 1$ می‌شود در حالی که هیچ تغییر بهینه‌ایی در بازایی پیکره صورت نگرفته است. به عبارتی پسین حاشیه‌ایی وابسته به مقیاس X است. برای مثال فاصله اقلیدسی برای دو نقطه $\{1, -1\}$ در یک فضای یک بعدی با فاصله اقلیدسی بین $(1/\sqrt{2}, 1/\sqrt{2})$ و $(-1/\sqrt{2}, -1/\sqrt{2})$ در یک فضای دو بعدی برابر و مساوی دو می‌باشد. همان‌طور که می‌بینید فاصله اقلیدسی برای هر دو یکسان است اما با افزایش بعد از یک به دو مختصات‌ها به مبدا نزدیکتر شده‌اند در نتیجه واریانس هم در فضای دو بعدی کوچکتر شده که واریانس کوچکتر منجر به یک پسین بزرگتر و در نتیجه انتخاب بعد بالاتر می‌شود؛ اگر چه که فاصله ثابت است. برای رفع این مشکل، یک معیار انتخاب بعد باید X ‌ها را در بعد یکسان مقایسه کند. بدین منظور، $(X^{(p)} : \circ) = X^{*(p+1)}$ در فضای $p + 1$ بعدی که p مختصات اول $X^{(p)}$ و مختصات آخر مساوی صفر باشد را در نظر می‌گیریم که در واقع $X^{*(p+1)}$ فضای اقلیدسی یکسان و برازش یکسان مانند $X^{(p)}$ می‌دهد. در حالت ایده‌آل اگر p درست باشد، آنگاه جواب بهینه در فضای $p + 1$ بعد $X^{*(p+1)}$ خواهد بود. بنابراین $X^{(p+1)}$ و $X^{*(p+1)}$ را مقایسه و p را در صورتی به عنوان بعد انتخاب می‌کنیم که چگالی پسین حاشیه‌ای $X^{*(p+1)}$ از چگالی پسین حاشیه‌ای $X^{(p+1)}$ بزرگتر باشد. حال برای مقایسه بهتر نسبت پسین حاشیه‌ای R_p را به صورت زیر در نظر می‌گیریم:

$$\begin{aligned} R_p &\equiv \frac{\pi(X^{(p+1)}, p+1|D)}{\pi(X^{*(p+1)}, p+1|D)} \\ &= \left\{ \frac{SSR_{p+1}}{SSR_{p+1}^*} \right\}^{-\frac{m}{\nu}+1} \left\{ \prod_{j=1}^{p+1} \frac{\frac{S_j}{\nu} + \beta_j}{\frac{S_j^*}{\nu} + \beta_j} \right\}^{(-\frac{n}{\nu}+\alpha)} \\ &= \left\{ \frac{SSR_{p+1}}{SSR_p} \right\}^{-\frac{m}{\nu}+1} \left\{ \prod_{j=1}^p \frac{\frac{S_j^{(p+1)}}{\nu} + \beta_j}{\frac{S_j^{(p)}}{\nu} + \beta_j} \right\}^{(-\frac{n}{\nu}+\alpha)} \\ &\quad \times \left\{ \frac{\frac{S_{p+1}^{(p+1)}}{\nu} + \beta_{p+1}}{\beta_{p+1}} \right\}^{(-\frac{n}{\nu}+\alpha)} \end{aligned}$$

که در آن $\pi(X^{(p+1)}, p+1|D)$ چگالی پسین حاشیه‌ایی X و p با $p + 1$ بعد به شرط داده‌ها و $\pi(X^{*(p+1)}, p+1|D)$ چگالی پسین حاشیه‌ایی p و X با p بعد و ستون آخر برابر صفر به شرط داده‌ها می‌باشند. واضح است نسبت R_p وابسته به انتخاب ابرپارامترهای α و β_j از Λ است. هنگامی که هیچ اطلاع پیشین قوی از پارامترها نداریم، یک انتخاب معقول برای (α, β_j) در فضای $p + 1$ بعد می‌تواند انتخاب‌های $\alpha = \frac{1}{\nu}$ و $\beta_j = \frac{1}{\nu} S_j^{(p+1)} / n$ باشد. رافتری (۱۹۹۹) استدلال کرد که این یک پیشین مناسب معقول برای تقریب موقعیت‌هایی که اطلاعات پیشین کمی در دسترس است، می‌باشد. در نتیجه داریم

$$R_p = \left(\frac{SSR_{p+1}}{SSR_p} \right)^{-\frac{m}{\nu}+1} \left\{ \prod_{j=1}^p \frac{r_j^{(p+1)}(n+1)}{(n+r_j^{(p+1)})} \right\}^{-\frac{(n+1)}{\nu}} \times (n+1)^{-\frac{(n+1)}{\nu}}$$

که در آن $r_j^{(p+1)} = S_j^{(p+1)} / S_j^{(p)}$. در این صورت می‌توان نوشت:

$$LR_p \equiv -2 \log R_p = (m - 2) \log \left(\frac{SSR_{p+1}}{SSR_p} \right) \quad (14.2)$$

$$+ \left\{ (n + 1) \sum_{j=1}^p \log \left[\frac{r_j^{(p+1)}(n + 1)}{(n + r_j^{(p+1)})} \right] + (n + 1) \log(n + 1) \right\} \quad (15.2)$$

عبارت (۱۴.۲) در LR_p تقریباً نسبت لگاریتم درست‌نمایی است که معمولاً منفی می‌باشد (چون هر چه تعداد بعد بالاتر باشد پیکره را با دقت بیشتر می‌توان بازیابی کرد لذا مقدار SSR کوچکتر و در نتیجه $\log SSR_{p+1}$ کوچکتر بوده و بنابراین $\log SSR_{p+1} - \log SSR_p < 0$ و با افزایش بعد همواره زیاد می‌شود (با افزایش بعد SSR کوچک، درست‌نمایی بزرگ و منفی لگاریتم درست‌نمایی نیز بزرگ می‌شود). بنابراین هر چه تعداد بعد را بیشتر کنیم بهتر است اما از طرفی به دنبال بازیابی مناسب پیکره در بعد کوچک هستیم بنابراین عبارت (۱۵.۲) نقش جریمه را برای افزایش نابجا عبارت فوق دارد. عبارت جریمه بدین صورت عمل می‌کند که هرچه بعد بیشتر شود مقدار LR_p را، چنانچه افزایش بعد تاثیری در بازیابی X نداشته باشد، کم می‌کند. اگر افزایش p به $p + 1$ تاثیر چندانی در بازیابی X نداشته باشد، چنانچه این جریمه را در نظر نگیریم افزایش کاذب در مقدار عبارت فوق داریم که نادرست است. هرگاه افزایش بعد از p به $p + 1$ هیچ تغییر معناداری در بازیابی X نداشته باشد، آنگاه $r_j^{(p+1)} \approx 1$ و عبارت جریمه به طور تقریبی $(n + 1) \log(n + 1)$ می‌شود. به ازای LR_p مثبت، بعد p را به $p + 1$ ترجیح می‌دهیم و چنانچه LR_p منفی باشد بعد $p + 1$ مناسب‌تر از p است. بنابراین، به طور خلاصه می‌توان بعدی را انتخاب کرد که مقدار LR_p مثبت را نتیجه می‌دهد. لذا معیار اطلاع MDS (MDSIC) را برای انتخاب بعد مناسب می‌توان به صورت زیر تعریف کرد:

$$MDSIC_1 = (m - 2) \log SSR_1 \quad (16.2)$$

$$MDSIC_p = MDSIC_1 + \sum_{j=1}^{p-1} LR_j$$

در نهایت بعد بهینه بعدی است که در آن $MDSIC_p$ مینیمم شود.

۳.۵.۲ شبیه‌سازی

در این بخش با استفاده از شبیه‌سازی عملکرد روش BMDS را با روش CMDS مقایسه می‌کنیم. برای این منظور تعداد $n = 50$ نمونه تصادفی از توزیع نرمال چند متغیره ($p = 10$) با میانگین صفر و واریانس I (ماتریس همانی) تولید می‌کنیم ($X \sim N_{10}(\mathbf{0}, I)$). و برای اندازه‌ی عدم شباهت جفت اشیا x_i و x_j مدل فاصله اقلیدسی را به کار می‌بریم، و با استفاده از δ_{ij} ، فاصله مشاهده‌شده d_{ij} را از یک توزیع نرمال بریده‌شده در صفر با میانگین δ_{ij} و انحراف استاندارد 0.3 تولید می‌کنیم. برای مقداردهی اولیه، همان‌طور که پیش‌تر توضیح داده شد، از نتایج CMDS استفاده کردیم (کد مربوطه را در پیوست ۵ ببینید). نتایج حاصل در جدول ۱۰.۲ داده شده است.

| بعد | تنش $CMDS$ | تنش $BMDS$ | LRT | جریمه | MDSIC |
|-----|------------|------------|------------|-------|-------|
| ۱ | ۰/۹۷۷۹۳ | ۰/۶۵۲۳۹۴۶۸ | -۷۹۱/۸۰۲۷۴ | ۲۶۷/۳ | ۱۳۰۱۱ |
| ۲ | ۰/۹۰۷۹۳ | ۰/۵۷۵۵۸۷۴ | -۷۱۹/۴۸۰۴۲ | ۲۳۸/۲ | ۱۲۲۰۱ |
| ۳ | ۰/۶۰۳۴۱ | ۰/۳۲۲۷۹۰۳۱ | -۷۵۹/۳۶۴۴۰ | ۲۳۴/۶ | ۱۱۳۹۴ |
| ۴ | ۰/۳۹۸۵۰ | ۰/۱۹۰۹۰۶۸۹ | -۵۵۷/۴۶۲۷۱ | ۲۳۹/۷ | ۱۰۶۱۹ |
| ۵ | ۰/۲۷۵۷۲ | ۰/۱۴۹۰۸۶۸۳ | -۴۰۳/۱۷۶۰۵ | ۲۴۰/۵ | ۱۰۱۶۰ |
| ۶ | ۰/۱۸۰۲۳ | ۰/۰۹۰۳۷۱۰۷ | -۱۶۹/۴۰۳۷۰ | ۲۴۱/۳ | ۹۴۱۲ |
| ۷ | ۰/۱۱۷۲۷ | ۰/۰۷۱۰۷۶۴۴ | -۱۲۷/۶۷۱۰۹ | ۲۵۲/۳ | ۹۰۲۹ |
| ۸ | ۰/۷۲۵۵۸ | ۰/۰۷۸۷۳۳۴۰ | -۴۶/۳۱۲۱۱ | ۲۵۱/۴ | ۹۱۳۶ |
| ۹ | ۰/۷۳۸۹۲ | ۰/۰۷۹۱۱۰۳۱ | -۳۱/۹۱۳۴۹ | ۲۵۰/۲ | ۹۳۸۰ |
| ۱۰ | ۰/۹۰۳۷۷ | ۰/۰۸۶۴۹۰۰۶ | -۱۴/۸۲۷۳۲ | ۲۵۰/۴ | ۹۱۷۷ |
| ۱۱ | ۰/۹۰۳۰۸ | ۰/۰۸۷۷۱۱۵۵ | -۱۲/۴۲۴۴۳ | ۲۵۱/۳ | ۹۱۸۵ |
| ۱۲ | ۰/۹۰۴۶۱ | ۰/۰۸۶۱۶۴۹۸ | -۵/۷۵۹۲۴ | ۲۵۱/۵ | ۹۳۲۱ |
| ۱۳ | ۰/۹۰۷۱۷ | ۰/۰۸۵۶۱۹۷۹ | -۴/۴۶۶۷۴ | ۲۵۲/۳ | ۹۹۶۹ |
| ۱۴ | ۰/۹۰۹۲۳ | ۰/۰۹۲۹۶۳۹۳ | -۳/۵۱۳۰۱ | ۲۵۱/۱ | ۱۰۱۷۹ |

جدول ۱۰۲: تحلیل شبیه‌سازی داده‌های تولید شده از $N_i(\circ, I)$

همان‌طور که از جدول ۱۰۲ مشخص می‌شود روش $BMDS$ با یک تنش کوچک‌تر بهتر عمل می‌کند، همچنین نسبت لگاریتم درست‌نمایی LRT با افزایش بعد تا هفت به طور یکنواخت کاهش می‌یابد ولی بعد از آن تغییر قابل ملاحظه‌ای نمی‌کند. در نهایت معیار $MDSIC$ بعد هفت را به عنوان بعد مناسب انتخاب می‌کند.

۶.۲ نتیجه‌گیری

در این فصل یک روش بیزی برای تعیین پیکره اشیا در MDS ($BMDS$)، و یک معیار انتخاب بعد بیزی، بر اساس نتایج حاصل از $BMDS$ ، را مورد بررسی قرار دادیم. اندازه‌ی عدم شباهت‌های مشاهده شده d_{ij} را برابر با اندازه‌ی درست δ_{ij} (فواصل اقلیدسی) به اضافه‌ی یک جمله خطا مدل‌بندی کردیم. در مقیاس‌گذاری نامتری، عدم شباهت‌ها برابر با یک تابع غیر خطی از فواصل اقلیدسی به اضافه خطا مدل‌بندی می‌شوند، و به طور دقیق‌تر مدل d_{ij} به صورت زیر در نظر گرفته می‌شود:

$$d_{ij} \sim N(g(\delta_{ij}), \sigma^2) I(d_{ij} > \circ) \quad i \neq j, i, j = 1, \dots, n$$

که در آن $g(\cdot)$ تابعی غیرخطی اما یکنواخت است. استنباط بیزی با استفاده از MCMC مجدداً ممکن است منجر به مقیاس‌گذاری چند بعدی نامتری بیزی شود.

توزیع نرمال بریده‌شده را برای فاصله مشاهده‌شده به کار بردیم، که به طور تقریبی یک صورت مزدوج برای توزیع پیشین X را نتیجه داد. از توزیع‌های دیگر مانند لگ‌نرمال و کای دو غیرمرکزی می‌توان به جای توزیع نرمال بریده‌شده استفاده کرد، که در این صورت یک تغییر کوچک در روند الگوریتم متروپلیس

هستینگز برای تولید X لازم است (برای آگاهی بیشتر ساپس و زینس^{۱۰}، ۱۹۶۳، و رامسای^{۱۱}، ۱۹۶۹، ۱۹۷۷ را ببینید). از مزیت‌های اصلی BMDS این است که برازش خیلی بهتری نسبت به CMDS ارائه می‌دهد. در این فصل معیار انتخاب بعد MDSIC، که به آسانی محاسبه می‌شود، را روی پیکره حاصل از BMDS تعریف کردیم.

در مقایسه BMDS با CMDS یک ویژگی اصلی BMDS این است که وقتی بعد افزایش می‌ابد، مختصات برای بعدها‌ی پایین تغییر کرده در حالی که در CMDS مختصات برای یک بعد پایین‌تر همیشه یک زیر مجموعه از بالاتری‌ها است. مختصات به‌دست آمده از بعدها‌ی پایین‌تر الزاماً یک انتخاب بعد بهینه، وقتی بعد افزایش می‌ابد نیست و نگاه داشتن آن‌ها در ابعاد بالاتر ممکن است تاثیر بدی بر عملکرد CMDS داشته باشد.

^{۱۰}Suppes and Zinnes

^{۱۱}Ramsay

فصل ۳

مقیاس‌گذاری چند بعدی بیزی با پیشین آمیخته

۱.۳ مقدمه

در این فصل روش ساده و کارآمد بیزی BDIM^۱ برای انتخاب بعد مناسب در MDS را ارائه می‌دهیم و آن را با روش MDSIC ارائه شده در فصل قبل مقایسه می‌کنیم. در این فصل مشابه فصل دو، یک MDS بیزی با در نظر گرفتن یک فاصله اقلیدسی برای مدل آماری داده‌های مجاورت را مورد بررسی قرار می‌دهیم. همانند فصل قبل پیکربندی بیزی اشیا را با استفاده از روش MCMC ارزیابی می‌کنیم. تفاوت این فصل با فصل قبل در انتخاب توزیع پیشین می‌باشد. در این راستا، یک پیشین آمیخته از صفر و یک توزیع پیوسته برای پارامترهای مدل در نظر می‌گیریم. بنابراین توزیع پسین حاشیه‌ای هر پارامتر نیز یک توزیع آمیخته بدست خواهد آمد. برای برآورد پارامترهای توزیع پسین آمیخته، همانند فصل قبل، از روش MCMC استفاده می‌کنیم. مطالب این فصل برگرفته از مقاله اُه (۲۰۱۲) می‌باشد.

۲.۳ مدل آماری

مدل آماری در این فصل کاملاً مشابه با مدل آماری ارائه شده برای MDSIC در فصل دوم می‌باشد، لذا همانند بخش (۲.۲) فرض می‌کنیم به ازای $i, j = 1, \dots, n, i \neq j$ ، $d_{ij} \sim N(\delta_{ij}, \sigma^2)I(d_{ij} > 0)$ ، بنابراین تابع درست‌نمایی x, σ^2 همانند رابطه‌ی (۳.۲) حاصل می‌شود.

^۱Bayesian dimension

۳.۳ توزیع پیشین

برای انجام یک تحلیل بیزی، باید توزیع پیشین پارامترهای مجهول را مشخص کنیم. برای این منظور فرض می‌کنیم k امین بردار ستونی X دارای توزیع پیشین به صورت زیر باشد:

$$X_k | \lambda_k, w \sim wI(X_k = \circ) + (1 - w)I(X_k \neq \circ)N(X_k; \circ, \lambda_k^{-1}I_n) \quad (1.3)$$

که در آن w ، $(0 \leq w \leq 1)$ وزن، $I(\cdot)$ تابع مشخصه و $N_n(X_k; \circ, \lambda_k^{-1}I_n)$ توزیع نرمال n متغیره با بردار میانگین صفر و ماتریس واریانس - کوواریانس $\lambda_k^{-1}I_n$ می‌باشد. برای ابر پارامتر λ_k ، $k = 1, \dots, p$ فرض می‌کنیم:

$$\lambda_k \sim G(a, b_k).$$

باید دقت داشت که ابر پارامتر λ_k فقط زمانی که $X_k \neq \circ$ است، تعریف می‌شود. بنابراین اگر مقدار $X_k = \circ$ باشد، آنگاه ابر پارامتر λ_k در توزیع پیشین نقشی نداشته و می‌توان یک توزیع تباهیده در صفر با وزن مثلاً $w > 0$ را به عنوان توزیع پیشین در نظر گرفت. اما اگر $X_k \neq \circ$ ، آنگاه ابر پارامتر λ_k در ساختار توزیع پیشین ظاهر شده و همچنین خود آن نیز دارای توزیع پیشین خواهد بود. لذا می‌توان در حالت $X_k \neq \circ$ توزیع $G(a, b_k)N_p(\circ, \lambda_k^{-1}I_n)$ را به عنوان توزیع پیشین در نظر گرفت. در نهایت چنانچه بخواهیم حالات $X_k = \circ$ و $X_k \neq \circ$ را توأم در نظر بگیریم توزیع پیشین به صورت توزیع آمیخته زیر خواهد بود:

$$f(X_k, \lambda_k | w) = wI(x_k = \circ) + (1 - w)I(x_k \neq \circ)N(x_k; \circ, \lambda_k^{-1}I_n)G(\lambda_k; a, b_k)$$

و از آنجایی که $0 \leq w \leq 1$ نیز می‌تواند یک مقدار تصادفی باشد، فرض می‌کنیم

$$w \sim B(a_w, b_w).$$

در نهایت فرض می‌کنیم مولفه σ^2 دارای توزیع گامای معکوس با مد $1 + \beta/\alpha$ است؛ به عبارتی دیگر

$$\sigma^2 \sim IG(\alpha, \beta).$$

علاوه بر فرضیات فوق پیشین‌های توأم (X, λ) ، σ^2 و w را مستقل از یکدیگر در نظر می‌گیریم. در این صورت اگر $\pi(X, \sigma^2, \lambda, w)$ و $\pi(\sigma^2)$ و $\pi(X, \lambda | w)$ و $\pi(w)$ به ترتیب نشان‌دهنده توزیع توأم w, σ^2, λ, X و توزیع‌های پیشین حاشیه‌ایی (σ^2, X, λ) و w باشند توزیع پیشین توأم همه پارامترهای مدل به صورت زیر حاصل می‌شود:

$$\pi(X, \sigma^2, \lambda, w) = \pi(\sigma^2)\pi(X, \lambda | w)\pi(w)$$

۴.۳ توزیع پسین

در این بخش با استفاده از توزیع‌های پیشین و تابع درستنمایی، توابع چگالی پسین را بدست می‌آوریم. تابع چگالی پسین توام همه پارامترهای مدل به صورت زیر می‌باشد:

$$\begin{aligned} \pi(X, \sigma^2, \lambda, w|D) &\propto L(X, \sigma^2|D)\pi(\sigma^2)\pi(X, \sigma^2, \lambda, w) \\ \pi(X, \sigma^2, \lambda, w|D) &\propto L(X, \sigma^2|D)\pi(\sigma^2)\pi(X, \lambda|w)\pi(w) \\ &\propto L(X, \sigma^2|D)\pi(\sigma^2)\pi(w) \prod_{k=1}^p \pi(x_k, \lambda_k|w) \\ &\propto (\sigma^2)^{(-\frac{m}{\nu})} \exp\left\{-\frac{1}{\nu\sigma^2} SSR - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\} \frac{\beta^\alpha}{\Gamma\alpha} (\sigma^2)^{-\alpha-1} \exp\left\{-\frac{\beta}{\sigma^2}\right\} \\ &\times \frac{\Gamma(a_w + b_w)}{\Gamma a_w \Gamma b_w} w^{a_w-1} (1-w)^{b_w-1} \\ &\times \prod_{k=1}^p \{wI(X_k = \circ) + (1-w)I(X_k \neq \circ)N(\circ, \lambda_k^{-1}I_n)G(a, b_k)\} \\ &\propto (\sigma^2)^{-\frac{m}{\nu}-\alpha-1} \frac{1}{\prod_{i>j} \Phi\left(\frac{\delta_{ij}}{\sigma}\right)} w^{a_w-1} (1-w)^{b_w-1} \exp\left\{-\frac{1}{\nu\sigma^2} SSR - \frac{\beta}{\sigma^2}\right\} \\ &\times w^{\sum_{k=1}^p I(X_k=\circ)} \{(1-w)N(\circ, \lambda_k^{-1}I_n)G(a, b_k)\}^{\sum_{k=1}^p I(X_k \neq \circ)}. \end{aligned}$$

بنابراین با استفاده از توزیع پسین توام X, σ^2, λ, w می‌توان توزیع‌های پسین حاشیه‌ای پارامترها را بدست آورد. توزیع پسین حاشیه‌ای w به صورت زیر بدست می‌آید:

$$\begin{aligned} \pi(w|X, \sigma^2, \lambda, D) &\propto L(X, \sigma^2|D)\pi(\sigma^2)\pi(w)\pi(X, \lambda|w) \\ &\propto (\sigma^2)^{(-\frac{m}{\nu})} \exp\left\{-\frac{1}{\nu\sigma^2} SSR - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\} \pi(w) \prod_{k=1}^p \pi(x_k, \lambda_k|w) \quad \propto (\sigma^2)^{(-\frac{m}{\nu})} \\ &\times \prod_{k=1}^p \{wI(X_k = \circ) + (1-w)I(X_k \neq \circ)N(\circ, \lambda_k^{-1}I_n)G(a, b_k)\} \\ &\propto w^{a_w-1} (1-w)^{b_w-1} w^{\sum_{k=1}^p I(X_k=\circ)} \{(1-w)N(\circ, \lambda_k^{-1}I_n)G(a, b_k)\}^{\sum_{k=1}^p I(X_k \neq \circ)} \\ &\propto w^{a_w + \sum_{k=1}^p I(x_k=\circ)-1} (1-w)^{b_w + \sum_{k=1}^p I(x_k \neq \circ)-1} \end{aligned}$$

که می‌توان نتیجه گرفت w دارای توزیع پسین حاشیه‌ای بتا است، به عبارتی،

$$w|X, \sigma^2, \lambda \sim \text{Beta}\left(a_w + \sum_{k=1}^p I(x_k = \circ), b_w + \sum_{k=1}^p I(x_k \neq \circ)\right).$$

توزیع پسین شرطی σ^2 عبارتست از

$$\begin{aligned}\pi(\sigma^2|X, w, \lambda, D) &\approx \pi(\sigma^2)L(X, \sigma^2|Data) \\ &= IG(\sigma^2, \alpha, \beta) \frac{(\sigma^2)^{-\frac{m}{\nu}} \exp[-\frac{1}{\nu\sigma^2}SSR]}{\prod_{i>j}^n \Phi(\frac{\delta_{ij}}{\sigma})}\end{aligned}$$

اغلب $m = \frac{n(n-1)}{2}$ بزرگتر از α است (m تعداد عدم شباهت‌ها و α مجموع تعداد گامای معکوس‌ها است). بنابراین در چگالی پسین شرطی σ^2 جمله $(\sigma^2)^{-\frac{m}{\nu}} e^{-\frac{1}{\nu\sigma^2}SSR}$ غالب است. لذا چگالی پسین σ^2 را می‌توان به خوبی با تابع چگالی $IG(\frac{m}{\nu}, \frac{SSR}{\nu})$ تقریب زد. همچنین اگر m بزرگ باشد، توزیع گامای معکوس با توزیع نرمال به خوبی تقریب زده می‌شود. بنابراین می‌توان از الگوریتم قدم زدن تصادفی متروپلیس - هستینگز برای تولید σ^2 استفاده کرد. در نهایت برای تولید (x_k, λ_k) چگالی پسین شرطی (x_k, λ_k) به صورت زیر حاصل می‌شود:

$$\begin{aligned}\pi(x_k, \lambda_k|w, \sigma^2, D) &\approx \pi(x_k, \lambda_k|w).L(X, \sigma^2|Data) \\ &= [wI(x_k = \circ) + (1-w)I(x_k \neq \circ)]\phi(x_k; \circ, \lambda_k^{-1}I_n)g(\lambda; a, b_k) \\ &\frac{(\sigma^2)^{-\frac{m}{\nu}} \exp[-\frac{1}{\nu\sigma^2}SSR]}{\prod_{i>j}^n \Phi(\frac{\delta_{ij}}{\sigma})}\end{aligned}$$

۵.۳ محاسبات عددی

در این قسمت فرآیند تولید نمونه از توزیع‌های حاشیه‌ای پسین را با استفاده از الگوریتم MCMC (گاتردو و رافتری^۲، ۲۰۰۸) شرح می‌دهیم.

۱.۵.۳ الگوریتم MCMC

از آنجایی که توزیع پیشین آمیخته از یک جرم نقطه‌ای و یک توزیع پیوسته برای (x_k, λ_k) تشکیل شده است، توزیع پسین توام نیز آمیخته‌ای از توزیع‌های تکی است. لذا با به‌کارگیری یک الگوریتم MCMC برای آمیخته‌های توزیع‌های مورد نظر مکرراً نمونه‌های پسین از $X_k, \sigma^2, \lambda, w$ را تولید می‌کنیم. ابتدا روند الگوریتم و تولید نمونه برای σ^2 را شرح می‌دهیم. همان‌طور که در بخش ۲.۵.۱ مطرح شد می‌توان در این قسمت نیز برای تولید σ^2 از الگوریتم قدم زدن تصادفی متروپلیس - هستینگز استفاده کرد و یک توزیع نرمال با واریانس متناسب با واریانس توزیع $IG(\frac{m}{\nu}, \frac{SSR}{\nu})$ را به عنوان توزیع پیشنهادی بکار برد. بدین ترتیب با توجه به نمونه فعلی σ_{old} ، یک نمونه جدید σ_{new} از $IG(\sigma_{old}, covar)$ ، که در

^۲Gottardo and Raftery

آن گلمن^۳ و همکارانش (۱۹۹۶) نشان دادند که استفاده از ثابت الگوریتم MCMC را سرعت بخشیده و زمان اجرای آن را کم می‌کند. $c = ۲/۳۸۲$. احتمال پذیرش نمونه جدید σ_{new}^2 در الگوریتم متروپلیس-هستینگز با استفاده از معیار زیر بررسی می‌شود:

$$\alpha = \min \left\{ 1, \frac{\pi(\sigma_{new}^2 | else)}{\pi(\sigma_{old}^2 | else)} \right\}$$

با احتمال α ، σ_{new}^2 پذیرفته می‌شود در غیر این صورت دوباره از $IG(\sigma_{old}^2, cvar)$ نمونه تولید می‌کنیم. پس از تولید نمونه از σ^2 از (x_k, λ_k) نمونه تولید می‌کنیم. با توجه به ساختار چگالی پسین، تولید نمونه از آن به‌طور مستقیم ساده نبوده و لذا از الگوریتم قدم زدن تصادفی متروپلیس-هستینگز استفاده می‌کنیم. فرض می‌کنیم $\theta_k = (x_k, \lambda_k)$ در این حالت برای تولید مقدار اولیه $\theta_k^{(1)} = (x_k^{(1)}, \lambda_k^{(1)})$ از یک پیکره CMDS استفاده می‌کنیم. سپس مقدار جدید $\theta_k^{(2)} = (x_k^{(2)}, \lambda_k^{(2)})$ را از توزیع $\theta_k^{(2)} | \theta_k^{(1)}$ ، با فرمول زیر، تولید می‌کنیم.

$$q(\theta_k^{(2)} | \theta_k^{(1)}) \sim \circ / \delta I(x_k^{(2)} = \circ) + \circ / \delta I(x_k^{(2)} \neq \circ) \phi(x_k^{(2)}; x_k^{(1)}, \tau^{-1} I_n) g(\lambda_k^{(2)}; a + \frac{n}{\gamma}, b_k + \frac{S_k^{(2)}}{\gamma})$$

که در آن تابع چگالی نرمال با پارامترهای میانگین $x_k^{(1)}$ و ماتریس واریانس-کوواریانس $\tau^{-1} I_n$ در نقطه $x_k^{(2)}$ و $g(\lambda_k^{(2)}; a + \frac{n}{\gamma}, b_k + \frac{S_k^{(2)}}{\gamma})$ تابع چگالی گاما با پارامترهای $a + n/2$ و $b_k + S_k^{(2)}$ در نقطه $\lambda_k^{(2)}$ است که در آن $S_k^{(2)} = \sum_{i=1}^n (x_{ik}^{(2)})^2$ اما همان‌طور که دیده می‌شود برای تولید $\theta_k^{(2)}$ نیاز به تولید $\lambda_k^{(2)}$ و $x_k^{(2)}$ داریم. برای این منظور به روش زیر عمل می‌کنیم:

ابتدا $x_k^{(2)}$ را از $\circ / \delta I(x_k^{(2)} = \circ) + \circ / \delta I(x_k^{(2)} \neq \circ) \phi(x_k^{(2)}; x_k^{(1)}, \tau^{-1} I_n)$ تولید می‌کنیم. اگر $x_k^{(2)} = \circ$ ، دیگر نیازی به تولید $\lambda_k^{(2)}$ نیست و اگر $x_k^{(2)} \neq \circ$ باشد $\lambda_k^{(2)}$ را از $g(\lambda_k^{(2)}; a + \frac{n}{\gamma}, b_k + \frac{S_k^{(2)}}{\gamma})$ تولید می‌کنیم. توجه داشته باشید که $g(\lambda_k; a + \frac{n}{\gamma}, b_k + \frac{S_k^{(2)}}{\gamma})$ تقریب خوبی برای $\phi(x_k; \circ, \lambda_k^{-1} I_n) g(\lambda_k; a, b_k)$ است.

حال، مقدار جدید $\theta_k^{(2)}$ را با احتمال

$$\alpha(\theta_k^{(1)}, \theta_k^{(2)}) = \min \{ 1, \alpha^*(\theta_k^{(1)}, \theta_k^{(2)}) \}$$

می‌پذیریم که در آن

$$\alpha^*(\theta_k^{(1)}, \theta_k^{(2)}) = \frac{\pi(\theta_k^{(2)} | else) q(\theta_k^{(1)} | \theta_k^{(2)})}{\pi(\theta_k^{(1)} | else) q(\theta_k^{(2)} | \theta_k^{(1)})}$$

در این قسمت لازم است مقدار $\alpha^*(\theta_k^{(1)}, \theta_k^{(2)})$ را به ازای مقادیر مختلف $x_k^{(1)}$ و $x_k^{(2)}$ محاسبه کنیم. برای این منظور، ابتدا فرض می‌کنیم:

$$\Delta_k SSR = SSR(x_k = x_k^{(2)}) - SSR(x_k = x_k^{(1)})$$

در ادامه مقدار $\alpha^*(\theta_k^{(1)}, \theta_k^{(2)})$ را در حالت‌های مختلف زیر داریم:

$$\alpha^*(\theta_k^{(1)}, \theta_k^{(2)}) = 1 \text{ اگر } (i) \ x_k^{(1)} = x_k^{(2)} = \circ$$

^۳Gelman

(ii) اگر $x_k^{(1)} = \circ, x_k^{(2)} \neq \circ$ آنگاه

$$\alpha^*(\theta_k^{(1)}, \theta_k^{(2)}) = \exp \left\{ -\frac{1}{\nu \sigma^2} \Delta_k SSR \right\} \prod_{i>j} \left[\frac{\phi(\frac{\delta_{ij}^{(1)}}{\sigma})}{\phi(\frac{\delta_{ij}^{(2)}}{\sigma})} \right] \times \frac{(1-w) \phi(x_k^{(2)}; \circ, \lambda_k^{-(2)} I_n)}{w \phi(x_k^{(2)}; \circ, \tau^{-1} I_n)} \\ \times \frac{g(\lambda_k^{(2)}; a, b_k)}{g(\lambda_k^{(2)}; a + \frac{n}{\nu}, b_k + \frac{S_k^{(2)}}{\nu})}$$

(iii) اگر $x_k^{(1)} \neq \circ, x_k^{(2)} = \circ$ آنگاه

$$\alpha^*(\theta_k^{(1)}, \theta_k^{(2)}) = \exp \left\{ -\frac{1}{\nu \sigma^2} \Delta_k SSR \right\} \prod_{i>j} \left[\frac{\phi(\frac{\delta_{ij}^{(1)}}{\sigma})}{\phi(\frac{\delta_{ij}^{(2)}}{\sigma})} \right] \times \frac{w \phi(x_k^{(1)}; \circ, \tau^{-1} I_n)}{(1-w) \phi(x_k^{(1)}; \circ, \lambda_k^{-(1)} I_n)} \\ \times \frac{g(\lambda_k^{(1)}; a + \frac{n}{\nu}, b_k + \frac{S_k^{(1)}}{\nu})}{g(\lambda_k^{(1)}; a, b_k)}$$

(iv) اگر $x_k^{(1)} \neq \circ, x_k^{(2)} \neq \circ$ آنگاه

$$\alpha^*(\theta_k^{(1)}, \theta_k^{(2)}) = \exp \left\{ -\frac{1}{\nu \sigma^2} \Delta_k SSR \right\} \prod_{i>j} \left[\frac{\phi(\frac{\delta_{ij}^{(1)}}{\sigma})}{\phi(\frac{\delta_{ij}^{(2)}}{\sigma})} \right] \times \frac{\phi(x_k^{(2)}; \circ, \lambda_k^{-(2)} I_n)}{\phi(x_k^{(1)}; \circ, \lambda_k^{-(1)} I_n)} \\ \times \frac{\phi(x_k^{(1)}; x_k^{(2)}, \tau^{-1} I_n)}{\phi(x_k^{(2)}; x_k^{(1)}, \tau^{-1} I_n)} \times \frac{g(\lambda_k^{(2)}; a, b_k)}{g(\lambda_k^{(1)}; a, b_k)} \times \frac{g(\lambda_k^{(1)}; a + \frac{n}{\nu}, b_k + \frac{S_k^{(1)}}{\nu})}{g(\lambda_k^{(2)}; a + \frac{n}{\nu}, b_k + \frac{S_k^{(2)}}{\nu})}$$

با تکرار تولید $w, \sigma^2, \{x_k\}, \{\lambda_k\}$ ، نمونه‌هایی از توزیع‌های پسین حاشیه‌ای برای پارامترهای مجهول تولید شده و استنباط بیزی بر اساس این نمونه‌ها امکان‌پذیر خواهد بود. اگر مقدار احتمال $\gamma_k = p(x_k \neq \circ | Data)$ معنی‌داری k امین بردار ستونی (معنی‌داری بعد p) باشد، با برآورد $\hat{\gamma}$ ، به ازای هر k ، چنانچه این مقدار از مقداری که محقق در نظر می‌گیرد بزرگتر باشد می‌توان گفت k امین بردار ستونی معنی‌دار است در غیر این صورت از آن چشم‌پوشی می‌کنیم. برای مقدار اولیه X ، $\sigma^2 = SSR/m$ و $\lambda_k = 1/var(x_k)$ در MCMC، از یک پیکره MDS کلاسیک استفاده می‌کنیم.

اگر مرکز و جهت ماتریس کواریانس نمونه X دلخواه باشد آنگاه توزیع پسین، $N(\circ, \lambda_k^{-1} I_n)$ برای $x_k \neq \circ$ مناسب نخواهد بود. لذا نمونه‌های MCMC تولید شده بعد از هر تکرار را دوباره پردازش کرده و X را در هر تکرار با استفاده از تبدیل $X_i = D'_x(X_i - \bar{X})$ ، که در آن میانگین نمونه X_i ها و D_x ماتریسی است که ستون‌های آن بردارهای ویژه ماتریس کواریانس نمونه‌ای

$$S_x = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$$

می‌باشد تصحیح می‌کنیم. با استفاده از این تبدیل، x_i ها دارای میانگین نمونه‌ای \circ و ماتریس کواریانس قطری هستند که متناظر با مشخصات پیشین می‌باشد.

۲.۵.۳ شبیه‌سازی

در این بخش با یک شبیه‌سازی عملکرد BDIM و MDSIC را با هم مقایسه می‌کنیم. به تعداد $n = 50$ شی مستقل تولید می‌کنیم، بدین ترتیب که اولین مختصات i امین شی x_{i1}^* از $N(0, 1)$ و دومین مختصات i امین شی (به طور مستقل) x_{i1}^* ، از $N(0, \rho^2)$ (که در آن $0 < \rho < 1$) اندازه پراکندگی x_{i1}^* از صفر است) را تولید می‌کنیم. حال به ازای $i = 1, 2, \dots, 50$ ، فاصله اقلیدسی را برای هر جفت از اشیا $X^* = (x_{i1}^*, x_{i2}^*)$ به صورت زیر محاسبه می‌کنیم:

$$\delta_{ij}^* = \sqrt{(x_{i1}^* - x_{j1}^*)^2 + (x_{i2}^* - x_{j2}^*)^2}$$

سپس داده‌های عدم شباهت را از $N(\delta_{ij}^*, \sigma^2)I(d_{ij} \geq 0)$ تولید می‌کنیم. لازم به ذکر است که یک مقدار σ بزرگ موجب می‌شود بازایی X^* واقعی از اندازه عدم شباهت به تاخیر بیفتد که باعث خطایی بزرگ در برآورد پیکره X می‌شود، برای برطرف کردن این مشکل کافی است در مدل‌های پیچیده برای یک σ بزرگ بعد از کوچک انتخاب کنیم که در این قسمت یک در نظر می‌گیریم (برای کاهش اثر بد ناشی از بزرگی واریانس، که منجر به بازایی نادرست پیکره می‌شود بعد از کاهش می‌دهیم)

برای بررسی اثرات ρ و σ ، مقادیر $0.1, 0.2, \dots, 0.5$ را برای ρ و مقادیر $0.3, 0.5, 0.7$ را برای σ در این فرآیند شبیه‌سازی در نظر می‌گیریم. برای هر ρ مقادیر اندازه خطا بررسی و برای هر ترکیب (ρ, σ) ، روش BDIM را به کار می‌بریم. در شبیه‌سازی 5000 نمونه اول را در نظر نگرفته (به عبارتی آن را برای افزایش سرعت همگرایی می‌سوزانیم) و از 5000 نمونه بعدی برای برآورد

$$\gamma_k = P(x_k \neq 0 | \text{Data}) \quad k = 1, \dots, 5$$

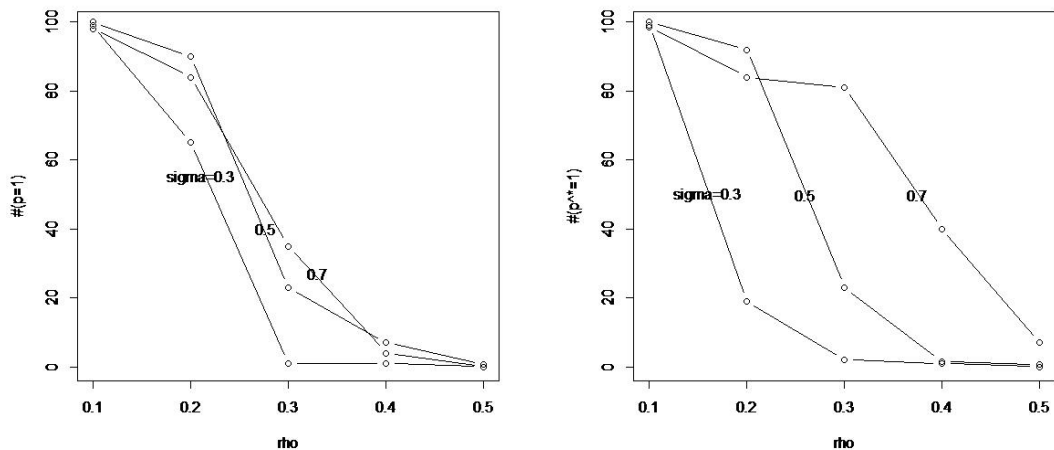
استفاده می‌کنیم. برای هر k اگر

$$\hat{\gamma}_k = \hat{P}(x_k \neq 0 | \text{Data}) > 0.5$$

آنگاه k امین بعد، بعد مناسب است. این شبیه‌سازی را برای هر ترکیب (ρ, σ) به تعداد صد بار تکرار می‌کنیم، با محاسبه فروانی‌ها به ازای $p^*, p^* = 1, \dots, 5$ را به عنوان بعد بهینه انتخاب می‌کنیم. به منظور مقایسه، همچنین این شبیه‌سازی را برای روش MDSIC با 3500 تکرار بکار می‌بریم. سپس نتش را برای MDSIC با استفاده از فرمول زیر محاسبه می‌کنیم:

$$STRESS = \sqrt{\frac{\sum_{i>j} (d_{ij} - \delta_{ij})^2}{\sum_{i>j} d_{ij}^2}}$$

در ادامه نمودار فراوانی $p^* = 1$ در مقابل ρ به ازای مقادیر مختلف σ را در هر دو روش مقایسه می‌کنیم. با توجه به شکل ۱۰.۳، در هر دو روش، به ازای مقدار معین σ ، چنانچه ρ کوچک باشد مقدار $p^* = 1$ و چنانچه ρ بزرگ باشد مقدار $p^* \neq 1$ ، که در اینجا $p^* = 2$ است، معنی‌دار می‌باشد. همچنین به ازای مقدار ρ ثابت برای σ های بزرگ، هر دو روش میل به انتخاب $p^* = 1$ دارند. لازم به ذکر است که تغییر p نسبت به σ در روش BDIM نسبت به MDSIC خیلی محسوس‌تر است. همانطور که در مقایسه دو نمودار شکل ۱۰.۳ مشخص است، نمودار مربوط به BDIM بر خلاف MDSIC یک تغییر جهت معنی‌داری با افزایش σ را نشان می‌دهد. به عبارتی در روش BDIM احتمال وجود خطا در اندازه عدم شباهت و بنابراین خطا در پیکره اشیا نسبت به MDSIC بیشتر است. به عبارت دیگر در BDIM

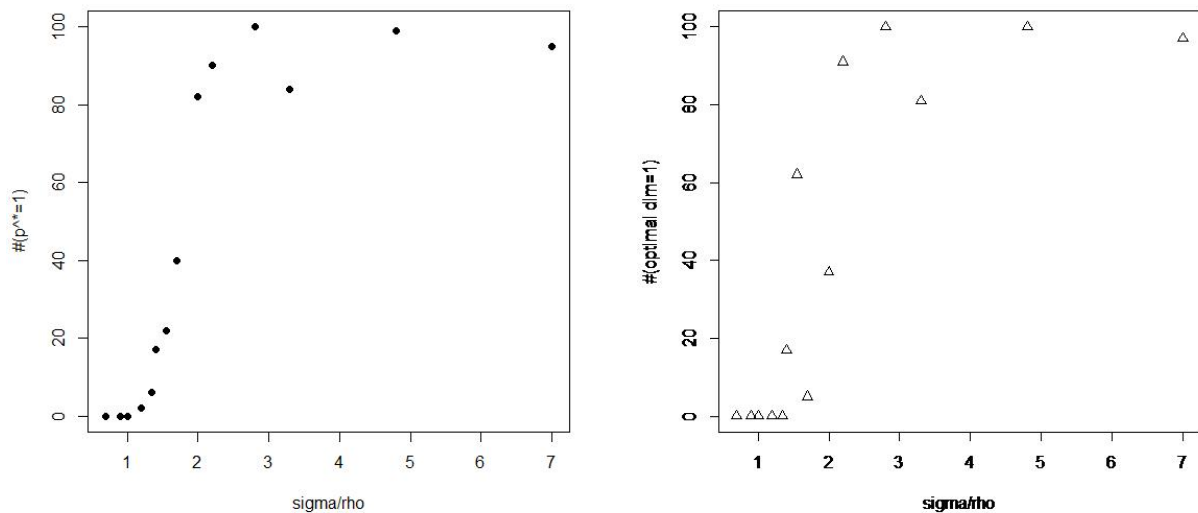


شکل ۱.۳: فراوانی $p^* = 1$ در مقابل ρ از راست به ترتیب برای روش $BDIM$ و $MDSIC$

انتخاب $p^* = 1$ نسبت به $MDSIC$ برای مقدار معین ρ وقتی σ بزرگ است محتمل‌تر می‌باشد. برای مثال هنگامی که $\rho = 0.3, \sigma = 0.7$ روش $BDIM$ حدود ۹۰ درصد و روش $MDSIC$ فقط با ۳۰ درصد $p^* = 1$ را به عنوان بعد بهینه انتخاب می‌کند. بنابراین ρ و σ در انتخاب بعد موثرند و دارای اثر متقابل بر یکدیگر نیز می‌باشند. حال در شکل ۲.۳ فراوانی $p^* = 1$ در مقابل σ/ρ را برای هر دو روش مقایسه می‌کنیم. همانطور که در شکل ۲.۳ می‌بینیم در هر دو روش فراوانی $p^* = 1$ با افزایش σ/ρ افزایش می‌یابد. در این شکل روش $BDIM$ یک منحنی ناهموار S شکل و یک افزایش خطی در فاصله $1/5 < \sigma/\rho < 3$ و در مقابل $MDSIC$ یک الگوی قاعده‌تر را نشان می‌دهد. شکل ۳.۳ ترکیب دو نمودار در یک نمودار است که نقطه‌ها نمایش فراوانی $BDIM$ و مثلث نمایش فراوانی $MDSIC$ می‌باشد. با مقایسه دو روش متوجه می‌شویم که جز در ۳ مورد دو روش مشابه هستند لازم به ذکر است $BDIM$ نسبت به $MDSIC$ در انتخاب $p^* = 1$ کمی محافظه کارتر است. سه موردی که تفاوت بزرگی را در دو روش نشان می‌دهد از چپ به راست $(\sigma, \rho) = (0.3, 0.2), (0.7, 0.4), (0.7, 0.3)$ هستند. برای مثال هنگامی که $\sigma = 0.3, \rho = 0.2$ روش $BDIM$ در ۲۰ مورد در حالی که روش $MDSIC$ حدود ۶۰ مورد $p^* = 1$ را به عنوان بعد مناسب انتخاب می‌کند. نتیجه برای $\sigma = 0.7, \rho = 0.5$ با همان $\rho = 0.2$ روش $BDIM$ یک تفاوت بزرگ را با افزایش σ نشان می‌دهد اما $MDSIC$ تفاوت کوچکی را نشان می‌دهد.

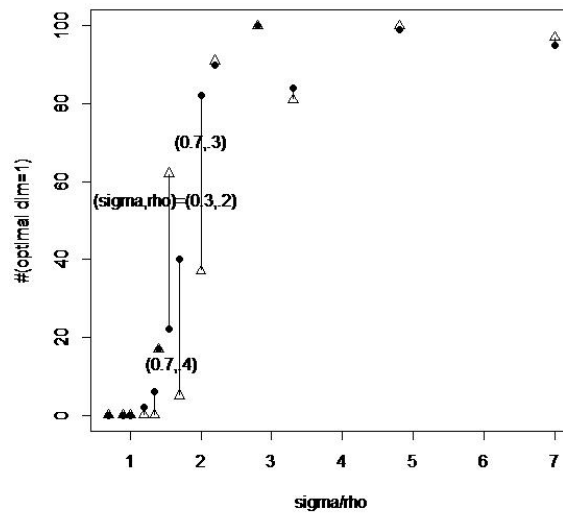
۶.۳ نتیجه‌گیری

در این فصل هم مانند فصل دوم، از یک $BMDS$ جهت تعیین پیکره اشیا استفاده کردیم و یک معیار انتخاب بعد بیزی $BDIM$ را بر اساس نتیجه‌ی حاصل از $BMDS$ مورد بررسی قرار دادیم. برای هر بردار ستونی از پیکره اشیا X ، یک آمیخته پیشین از یک چگالی نقطه‌ای و توزیع پیوسته با



شکل ۲.۳: فراوانی $p^* = 1$ در مقابل σ/ρ از راست به ترتیب $BDIM$ و $MDSIC$

وزن w فرض و چگالی پسین متناظر را بدست آوردیم. و با استفاده از روش‌های MCMC پیکره اشیا را بازیابی کردیم. برای بدست آوردن پیکره اشیا به دوباره پردازش کردن MCMC، پرداختیم که در نتیجه‌ی آن، ستون‌های برداری X حاصل متعامدند، لذا بررسی توزیع پسین حاشیه‌ای هر بردار کافی بوده و نیازی به بررسی توزیع توام همه‌ی ستون‌ها نمی‌باشد که این، انتخاب بعد را بسیار ساده‌تر می‌کند. داده‌های گمشده را نیز با یک تغییر کوچک در الگوریتم MCMC به آسانی می‌توان به‌کاربرد. برای مثال اگر d_{ij} داده گمشده ما باشد با اضافه کردن یک گام در الگوریتم MCMC، d_{ij} از $N(\delta_{ij}, \sigma^2)I(d_{ij} \geq 0)$ برای X و σ^2 معین تولید می‌کنیم.



شکل ۳.۳: فراوانی $p^* = 1$ در مقابل σ/ρ برای دو روش

فصل ۴

مثال‌های کاربردی

۱.۴ مقدمه

در این فصل عملکرد مقیاس‌گذاری چند بعدی بیزی را در قالب چند مثال واقعی مورد بررسی قرار می‌دهیم.

مثال ۱.۱.۴. (فواصل هوایی بین شهرها). هارتیگان^۱ (۱۹۷۵)، فواصل هوایی بین ۳۰ شهر اصلی جهان را که در جدول ۳.۴ نشان داده شده است، ارائه نمود. شهرها روی سطح زمین به صورت سه بعدی قرار گرفته و هواپیماها در بالای سطح زمین (روی سطح یک کره) حرکت می‌کنند. بنابراین فواصل هوایی دقیقاً فواصل اقلیدسی نیستند و ممکن است انتظار داشته باشیم که بعد x_i ها بین ۲ و ۳ باشد. برای داده‌های فواصل هوایی، مقیاس‌گذاری کلاسیک و بیزی را بکار بردیم. نتایج حاصل در جدول ۱.۴ ارائه شده اند.

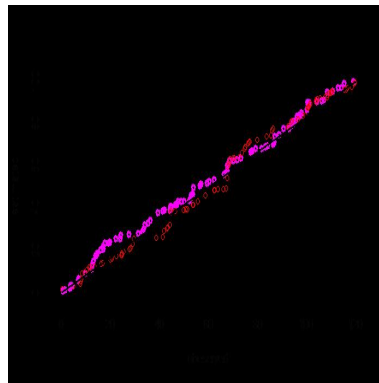
مشاهده می‌کنید همان‌طور که انتظار داشتیم استرس حاصل از روش BMDS کوچکتر از روش CMDS می‌باشد و خیلی سریع تا بعد سه کاهش یافته و از بعد چهار دوباره افزایش یافته که این حاکی از مناسب بودن بعد سه می‌باشد.

همان‌طور که درباره‌ی MDSIC در فصل ۲ شرح دادیم، کمترین مقدار MDSIC تعیین‌کننده بعد مناسب است، که در این جا بعد سه کمترین مقدار MDSIC را دارد. لذا بعد سه، بعد مناسب می‌باشد. شکل ۱.۴ نمودار فواصل هوایی مشاهده شده در مقابل فواصل اقلیدسی برآورد شده می‌باشد که نقاط حاصل از برازش به روش BMDS به رنگ قرمز و نقاط حاصل از برازش به روش CMDS به رنگ آبی نشان داده شده است. همان‌طور که در شکل می‌بینید برازش کامل در راستای یک خط ۴۵ درجه حاصل می‌شود که نقاط حاصل از BMDS به جز چند نقطه که مربوط به فواصل بزرگ هستند، به خط ۴۵ درجه نزدیک‌ترند (وقتی فاصله بزرگ‌تر می‌شود برازش خوبی حاصل نمی‌شود چون وقتی شهرها دور از هم هستند یک اختلاف بزرگ بین فاصله هوایی و فاصله اقلیدسی در سه بعد وجود دارد)، که این بیانگر برتری روش BMDS است.

^۱Hartigan

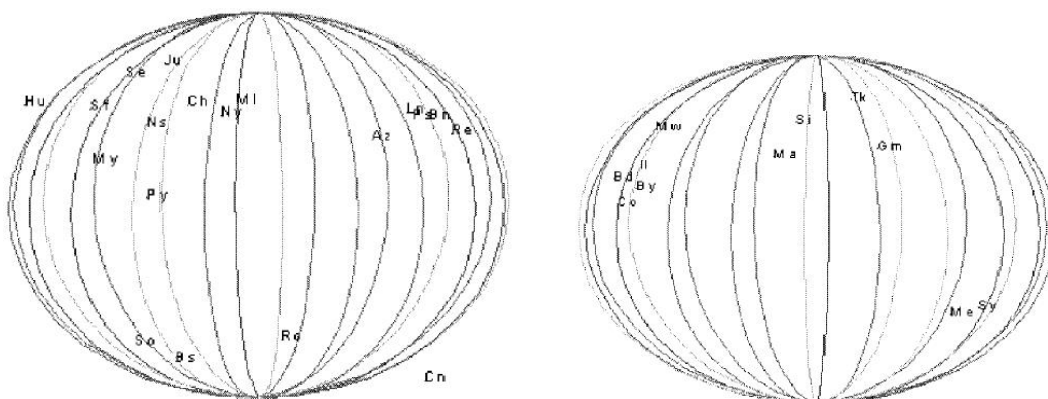
| بعد | تنش <i>BMDS</i> | تنش <i>CMDS</i> | LRT | جریمه | MDSIC |
|-----|-----------------|-----------------|------------|-----------|----------|
| ۱ | ۰/۶۳۳۹۵۸۷ | ۰/۸۷۲۷۶۹ | -۴۹/۸۲۹۳۷۴ | ۱/۱۳۷۰۶۰۷ | ۹۱۳۰/۱۰۷ |
| ۲ | ۰/۲۲۹۹۵۲۱ | ۰/۵۲۴۳۷۹ | ۱۸۱/۷۸۹۴۸۳ | ۰/۱۳۷۴۶۸۴ | ۹۲۶۶/۰۳۰ |
| ۳ | ۰/۱۳۱۳۵۹۵ | ۰/۴۰۹۵۴۲ | -۲۷/۲۱۹۳۳۴ | ۰/۸۸۰۸۴۵۵ | ۹۰۰۲/۴۷۴ |
| ۴ | ۰/۱۴۳۵۲۰۳ | ۰/۵۵۷۵۸۱ | -۱/۶۵۱۸۲۳ | ۲/۷۳۷۲۷۱۴ | ۹۲۳۱/۹۹۴ |
| ۵ | ۰/۱۵۱۹۵۰۰ | ۰/۵۶۲۶۲۷۵ | ۲/۵۱۶۵۰۳ | ۰/۶۱۱۳۴۵۰ | ۹۲۳۹/۱۸۶ |

جدول ۱.۴: جدول تحلیل فاصله‌ی شهرها



شکل ۱.۴: فواصل هوایی مشاهده شده در مقابل فواصل برآورد شده برای دو روش *BMDS* و *CMDS*

شکل ۲.۴ مکان شهرها، حاصل از روش *BMDS* با بعد سه را نمایش می‌دهد. می‌توان مشاهده کرد که شهرها به طور تقریبی روی سطح یک کره با شعاع زمین قرار گرفته‌اند و مکان شهرها به خوبی بازیابی شده‌اند.



شکل ۲.۴: برآورد مکان شهرها با استفاده از روش *BMDS*

| | AZ | BD | BN | BY | BS | CO | CN | CH | GM | HU | IL | JU | LN | MA | ME | MY | ML | MW | NS | NY | PY | PS | RO | RE | SF | SO | SE | SI | SY | TO | |
|-------------|-----|----|-----|-----|-----|----|-----|----|-----|----|----|----|-----|-----|-----|----|-----|----|----|-----|----|----|-----|----|----|----|----|----|----|----|--|
| Azores | ° | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Bagdad | ۳۹ | ° | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Berlin | ۵۱ | ۲۰ | ° | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Bombay | ۵۱ | ۲۰ | ۳۹ | ° | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Bue - Aires | ۵۴ | ۸۱ | ۷۴ | ۹۳ | ° | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Cairo | ۳۳ | ۸ | ۱۸ | ۲۷ | ۷۳ | ° | | | | | | | | | | | | | | | | | | | | | | | | | |
| Capetown | ۵۷ | ۴۹ | ۶۰ | ۵۱ | ۵۶ | ۶۱ | ۸۵ | ° | | | | | | | | | | | | | | | | | | | | | | | |
| Chicago | ۳۲ | ۶۴ | ۴۴ | ۸۱ | ۵۶ | ۸۱ | ۸۸ | ۷۴ | ° | | | | | | | | | | | | | | | | | | | | | | |
| Guam | ۸۹ | ۶۳ | ۷۱ | ۴۸ | ۱۰۴ | ۷۱ | ۸۸ | ۲۳ | ۳۸ | ° | | | | | | | | | | | | | | | | | | | | | |
| Honolulu | ۷۳ | ۸۴ | ۷۳ | ۸۰ | ۷۶ | ۸۸ | ۱۱۵ | ۲۳ | ۳۸ | ° | | | | | | | | | | | | | | | | | | | | | |
| Stanbul | ۲۹ | ۱۰ | ۱۱ | ۳۰ | ۷۶ | ۸ | ۵۲ | ۵۵ | ۶۹ | ۸۱ | ° | | | | | | | | | | | | | | | | | | | | |
| Juneau | ۴۶ | ۶۱ | ۴۶ | ۶۹ | ۷۷ | ۶۳ | ۱۰۳ | ۲۳ | ۵۱ | ۲۸ | ° | | | | | | | | | | | | | | | | | | | | |
| London | ۱۶ | ۲۵ | ۶ | ۴۵ | ۶۹ | ۲۲ | ۶۰ | ۴۰ | ۷۵ | ۷۲ | ۱۶ | ° | | | | | | | | | | | | | | | | | | | |
| Manila | ۸۳ | ۴۹ | ۶۱ | ۳۲ | ۱۱۱ | ۵۷ | ۷۵ | ۸۱ | ۱۶ | ۵۳ | ۵۷ | ۵۹ | ° | | | | | | | | | | | | | | | | | | |
| Melbourne | ۱۲۰ | ۸۱ | ۹۹ | ۶۱ | ۷۲ | ۸۷ | ۶۴ | ۹۷ | ۲۵ | ۵۵ | ۹۱ | ۸۱ | ۱۰۵ | ۳۹ | ° | | | | | | | | | | | | | | | | |
| Mex - city | ۲۵ | ۸۱ | ۶۱ | ۹۷ | ۴۶ | ۷۷ | ۸۵ | ۱۷ | ۷۵ | ۳۸ | ۷۱ | ۳۲ | ۵۶ | ۸۸ | ۸۴ | ° | | | | | | | | | | | | | | | |
| Montreal | ۲۴ | ۵۸ | ۳۷ | ۷۵ | ۵۶ | ۵۴ | ۷۹ | ۸ | ۷۷ | ۴۹ | ۴۸ | ۲۶ | ۸۲ | ۱۰۴ | ۱۰۴ | ۲۳ | ° | | | | | | | | | | | | | | |
| Moscow | ۳۲ | ۱۶ | ۱۰ | ۳۱ | ۸۴ | ۱۸ | ۶۳ | ۵۰ | ۶۱ | ۷۰ | ۱۱ | ۴۶ | ۱۶ | ۵۱ | ۹۰ | ۶۷ | ۲۴ | ° | | | | | | | | | | | | | |
| New - Orle | ۲۶ | ۷۲ | ۵۱ | ۸۹ | ۴۹ | ۶۸ | ۸۳ | ۸ | ۷۷ | ۴۲ | ۶۲ | ۲۹ | ۴۶ | ۸۷ | ۹۳ | ۹ | ۱۴ | ۵۸ | ° | | | | | | | | | | | | |
| New - York | ۲۵ | ۶۰ | ۴۰ | ۷۸ | ۵۳ | ۷۸ | ۷۸ | ۷ | ۸۰ | ۵۰ | ۵۰ | ۲۹ | ۳۵ | ۸۵ | ۱۰۴ | ۲۱ | ۳ | ۴۷ | ۱۲ | ° | | | | | | | | | | | |
| Pan - City | ۳۸ | ۷۸ | ۵۹ | ۹۷ | ۳۳ | ۷۱ | ۷۰ | ۲۳ | ۹۰ | ۵۳ | ۶۸ | ۴۵ | ۵۳ | ۱۰۳ | ۹۰ | ۱۵ | ۲۵ | ۶۷ | ۱۶ | ۲۲ | ° | | | | | | | | | | |
| Paris | ۱۶ | ۲۴ | ۵ | ۴۴ | ۶۹ | ۲۰ | ۵۸ | ۴۱ | ۷۶ | ۷۵ | ۱۴ | ۴۷ | ۲ | ۶۷ | ۱۰۴ | ۵۷ | ۳۴ | ۱۶ | ۴۸ | ۲۶ | ° | | | | | | | | | | |
| R.D.Janeiro | ۴۳ | ۶۹ | ۶۲ | ۸۳ | ۱۲ | ۶۱ | ۳۸ | ۵۳ | ۱۱۶ | ۸۳ | ۶۴ | ۷۶ | ۵۷ | ۱۱۳ | ۸۲ | ۴۸ | ۵۱ | ۷۲ | ۴۸ | ۴۸ | ۳۳ | ° | | | | | | | | | |
| Rome | ۲۱ | ۱۸ | ۷ | ۳۸ | ۶۹ | ۱۳ | ۵۲ | ۴۸ | ۷۶ | ۸۰ | ۹ | ۵۳ | ۹ | ۶۵ | ۹۹ | ۶۴ | ۴۱ | ۱۵ | ۵۵ | ۴۳ | ۵۹ | ۵۷ | ° | | | | | | | | |
| San - franc | ۵۰ | ۷۵ | ۵۷ | ۸۴ | ۶۴ | ۷۵ | ۱۰۳ | ۱۹ | ۵۸ | ۲۴ | ۶۷ | ۱۵ | ۵۴ | ۷۰ | ۷۹ | ۱۹ | ۵۴ | ۵۹ | ۱۹ | ۲۶ | ۳۳ | ۵۶ | ۶۳ | ° | | | | | | | |
| Santiago | ۵۷ | ۸۸ | ۷۸ | ۱۰۰ | ۶۴ | ۷۵ | ۱۰۳ | ۱۹ | ۵۸ | ۲۴ | ۶۷ | ۱۵ | ۵۴ | ۷۰ | ۷۹ | ۱۹ | ۵۴ | ۵۹ | ۱۹ | ۲۶ | ۳۳ | ۵۶ | ۶۳ | ° | | | | | | | |
| Seattle | ۴۶ | ۶۸ | ۵۱ | ۷۷ | ۶۹ | ۶۸ | ۱۰۲ | ۱۷ | ۵۷ | ۲۷ | ۶۱ | ۹ | ۴۸ | ۶۷ | ۸۲ | ۲۳ | ۲۳ | ۵۲ | ۲۱ | ۲۴ | ۳۷ | ۵۰ | ۶۹ | ° | | | | | | | |
| Shanghai | ۷۲ | ۴۴ | ۵۱ | ۳۱ | ۱۲۲ | ۵۲ | ۸۱ | ۷۰ | ۴۹ | ۴۹ | ۴۹ | ۵۷ | ۱۲ | ۵۰ | ۸۰ | ۷۰ | ۷۰ | ۴۲ | ۷۷ | ۷۳ | ۹۳ | ۵۷ | ۱۱۳ | ° | | | | | | | |
| Sydney | ۱۲۱ | ۸۳ | ۱۰۰ | ۶۳ | ۷۳ | ۹۰ | ۶۹ | ۹۲ | ۳۳ | ۵۱ | ۹۳ | ۷۷ | ۱۰۶ | ۳۹ | ۴ | ۸۱ | ۱۰۰ | ۹۰ | ۸۹ | ۱۰۰ | ۸۸ | ۸۴ | ۱۰۱ | ° | | | | | | | |
| Tokyo | ۷۳ | ۵۲ | ۵۶ | ۴۲ | ۱۱۴ | ۶۰ | ۹۲ | ۶۳ | ۱۶ | ۳۹ | ۵۶ | ۴۰ | ۶۰ | ۱۹ | ۵۱ | ۷۰ | ۶۵ | ۴۷ | ۶۹ | ۶۸ | ۸۴ | ۶۱ | ۱۱۵ | ° | | | | | | | |

جدول ۳.۴: فاصله‌ی هوایی بین شهرها (برحسب مایل)

| <i>dim</i> | <i>(prob)</i> |
|------------|---------------|
| ۱ | ۱ |
| ۲ | ۰/۴ |
| ۳ | ۰/۱۳ |
| ۴ | ۰/۱۲ |
| ۵ | ۰/۱۰ |
| ۶ | ۰/۰۸ |
| ۷ | ۰/۰۰۵ |
| ۸ | ۰/۰۰۴ |
| ۹ | ۰/۰۰۲ |
| ۱۰ | ۰/۰۰۲ |
| ۱۱ | ۰/۰۰۲ |
| ۱۲ | ۰/۰۰۱ |

جدول ۴.۴: احتمال‌های حاصل از BDIM به ازای بعدها مختلف

مثال ۲.۱.۴. در سال ۱۹۹۴ میلادی، در شرکت هولستن و باواریا، نوزده نوع دلستر از نظر بو، طعم و مزه با استفاده از مقیاس‌گذاری چند بعدی نامتری مورد بررسی قرار گرفت.

جدول ۵.۴ وجود یا عدم وجود یازده ویژگی بویایی را در رابطه با ۱۹ نوع دلستر می‌باشد که عدد یک نشان دهنده وجود ویژگی در آن دلستر است.

جدول ۶.۴ مشابه جدول ۵.۴ بوده و ویژگی‌های طعم و مزه را نشان می‌دهد.

در این مثال روش BDIM و MDSIC را برای انتخاب بعد مناسب برای ۱۱ ویژگی بویایی و ۱۲ ویژگی طعم و مزه برای ۱۹ نوع دلستر بکار بردیم.

مقدار آستانه برای تعیین بعد مناسب ۰.۵ در نظر گرفته شده است. همان‌طور که در فصل سه شرح داده شد روش BDIM احتمال‌های پسین را به ازای بعدها مختلف محاسبه و مقدار این احتمال‌ها را با آستانه ۰.۵ مقایسه می‌کند. در این‌جا همان‌طور که در جدول ۴.۴ مشاهده می‌کنید، با مقایسه احتمال‌های حاصل شده با آستانه، بعد یک به عنوان بعد مناسب به نظر می‌رسد (البته برای بعد دو با احتمال ۰.۴، باید در انتخاب بعد جانب احتیاط را رعایت کرد).

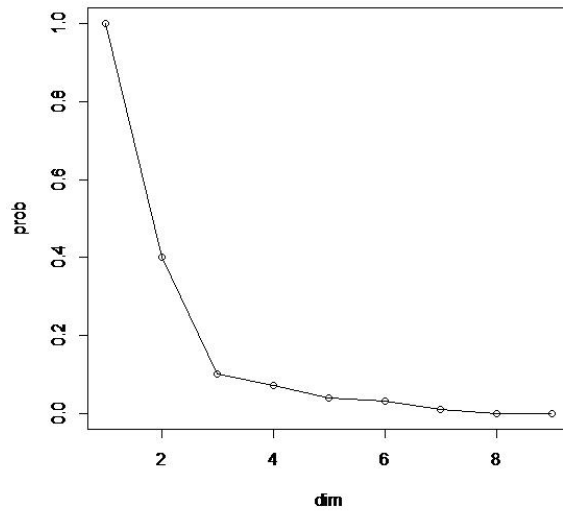
در روش MDSIC کمترین مقدار برای بعد دو حاصل می‌شود. لذا در این روش بعد دو به عنوان بعد مناسب انتخاب می‌شود.

| <i>Potable</i> | <i>Characteristics</i> | | | | | | | | | | | |
|---------------------|------------------------|---|---|---|---|---|---|---|---|----|----|----|
| | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۸ | ۹ | ۱۰ | ۱۱ | ۱۲ |
| <i>Glenburgie</i> | ۱ | ۱ | | | ۱ | | | | | | | |
| <i>Strathisla</i> | ۱ | | | | | | | | | | | |
| <i>Balblair</i> | | | | | ۱ | ۱ | | | | | | |
| <i>Clyrielsli</i> | ۱ | | | | | | | | ۱ | | | |
| <i>RoyalBrackla</i> | | ۱ | | | | ۱ | | ۱ | ۱ | | | |
| <i>Teaninich</i> | ۱ | | | | | ۱ | | ۱ | | | | |
| <i>GlenGarioch</i> | | | | | | ۱ | | ۱ | | | | |
| <i>Glenturret</i> | | | | ۱ | | ۱ | | | | ۱ | | |
| <i>Obari</i> | | | | | | ۱ | | ۱ | | | | |
| <i>Bladnoch</i> | ۱ | | ۱ | | | | | | | ۱ | | |
| <i>Littlernill</i> | ۱ | | ۱ | ۱ | | | | | | | | |
| <i>Ardbeg</i> | | | | | | ۱ | | ۱ | | | | |
| <i>Bournore</i> | | ۱ | | | | | | ۱ | ۱ | | | |
| <i>Lagavulin</i> | | | | | | | | ۱ | ۱ | | | ۱ |
| <i>Laphroaig</i> | | | | | | | | ۱ | ۱ | | ۱ | |
| <i>HighlandPark</i> | | | | ۱ | | | | ۱ | | | | |
| <i>IsleofJura</i> | | | | | | | | | ۱ | | | ۱ |
| <i>Toberritory</i> | ۱ | | ۱ | | | ۱ | | | | | | |
| <i>Bushmills</i> | | | | | | ۱ | | ۱ | | | | ۱ |

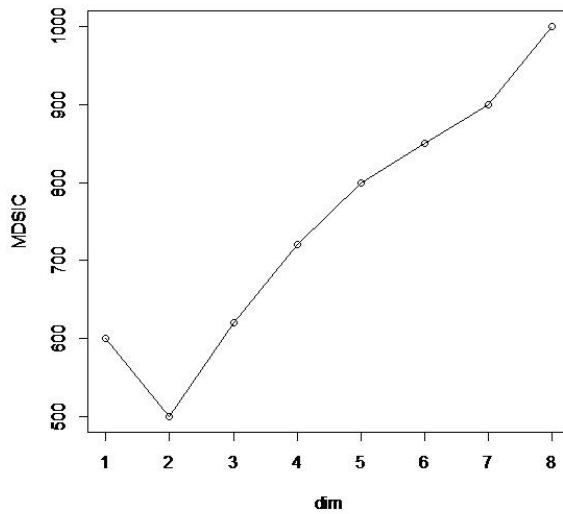
جدول ۵.۴: ویژگی‌های بویایی ۱۹ دلستر

| <i>Potable</i> | <i>Characteristics</i> | | | | | | | | | | | | | |
|---------------------|------------------------|---|---|---|---|---|---|---|---|----|----|----|----|----|
| | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۸ | ۹ | ۱۰ | ۱۱ | ۱۲ | ۱۳ | ۱۴ |
| <i>Glenburgie</i> | ۱ | | | ۱ | | | | | | | | | | |
| <i>Strathisla</i> | | ۱ | | | | | | | | | | | | |
| <i>Balblair</i> | | | | | | | | | | | | ۱ | ۱ | |
| <i>Clyrielsli</i> | | ۱ | | | | | | | | | | | | ۱ |
| <i>RoyalBrackla</i> | | | | | | | ۱ | ۱ | | | | ۱ | ۱ | |
| <i>Teaninich</i> | | | | | ۱ | | | | | | | | ۱ | |
| <i>GlenGarioch</i> | | | | | | | | | | ۱ | | | | |
| <i>Glenturret</i> | | | | | | | | | | | | | | ۱ |
| <i>Obari</i> | | | | | | | | | | | | ۱ | | |
| <i>Bladnoch</i> | | | | | | | | | | ۱ | | ۱ | | |
| <i>Littlernill</i> | | | | ۱ | | | | | | | | | | |
| <i>Ardbeg</i> | | | | | | | ۱ | | | ۱ | | | | |
| <i>Bournore</i> | | ۱ | | | | | ۱ | ۱ | | | | | | |
| <i>Lagavulin</i> | | | | | ۱ | | | | | ۱ | | | | |
| <i>Laphroaig</i> | | | | | | | | | | ۱ | | ۱ | | |
| <i>HighlandPark</i> | | | | | ۱ | | ۱ | | | ۱ | | | | |
| <i>IsleofJura</i> | | | ۱ | | | | | | | | | | | ۱ |
| <i>Toberritory</i> | | | ۱ | | | | | | | | | ۱ | ۱ | |
| <i>Bushmills</i> | | | | | | | | | | ۱ | | | | |

جدول ۶.۴: ویژگی‌های طعم و مزه ۱۹ دلستر



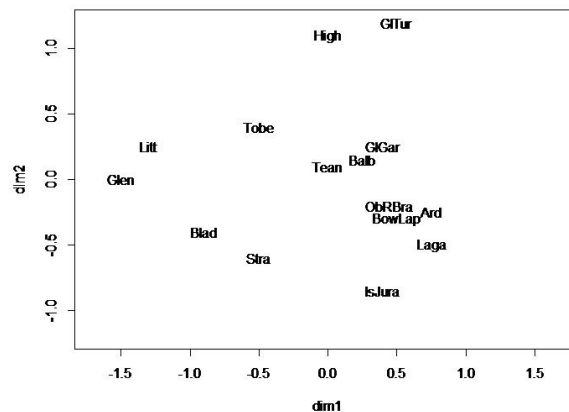
شکل ۳.۴: احتمال معنی دار برای هر بعد در روش BDIM برای داده‌های دلستر



شکل ۴.۴: مقادیر MDSIC به ازای هر بعد برای داده‌های دلستر

۲.۴ پیشنهادات و آینده تحقیق

همان‌طور که در فصول گذشته بحث شد، متأسفانه در زمینه مقیاس‌گذاری چند بعدی بیزی تحقیقات زیادی صورت نگرفته است. از طرفی با توجه به مطالب ارائه شده در فصل‌های دو و سه پی‌بردیم تولید نمونه پسین با استفاده از روش MCMC سخت بوده و در هر قسمت، فرضیات و تقریب‌هایی را برای پیاده‌سازی الگوریتم شبیه‌سازی در نظر گرفتیم. با توجه به این توضیحات می‌توان در آینده بر روی موضوعات پیشنهادی زیر کار کرده و نتایج آن‌ها را با روش‌های این پایان‌نامه مقایسه کرد:



شکل ۵.۴: پیکره اشیا دلسترها در فضای دو بعدی

۱. به جای توزیع نرمال برای حالت $x_k \neq 0$ در فصل سه از یک توزیع دم پهن‌تر مانند t -استیودنت استفاده کرده و به ازای درجات آزادی مختلف در توزیع t -استیودنت، سرعت همگرایی الگوریتم MCMC را با نرمال مقایسه کنیم. در این راستا می‌توان رابطه میزان تشخیص بعد مناسب را با درجه آزادی توزیع t -استیودنت مورد بررسی قرار داد.
۲. چون ساختار تابع درست‌نمایی در فصل‌های دو و سه صورت بسته‌ایی ندارد شاید بتوان از روش‌های محاسباتی بیزی تقریبی ^۲ (ABC) به جای الگوریتم MCMC در تولید نمونه از توزیع‌های پسین استفاده کرد.
۳. در فصول دو و سه، توزیع پیشین ابر پارامتر λ_j را گامای معکوس با پارامترهای ثابت و معلوم در نظر گرفتیم. در تحقیقی دیگر می‌توانیم این پارامترها را به روش‌های بیزی تجربی ^۳ یا بیزی سلسله مراتبی ^۴ برآورد کرده و مقادیر معنی‌دارتری برای آن‌ها در نظر بگیریم.
۴. در آخر، ماتریس واریانس-کواریانس پیکره اشیا X در هر دو فصل دو و سه، یک ماتریس قطری در نظر گرفته شد. به عنوان پیشنهادی به منظور مطالعه‌ی بیشتر، می‌توان ماتریس واریانس-کواریانس X را بدون ساختار ^۵ در نظر گرفته و توزیع ویشارت معکوس را به عنوان پیشین مناسب در مطالعات، برای آن به کار برد.

^۲Approximate Bayesian computation

^۳Empirical Bayse

^۴Hierarchical Bayse

^۵Unstructured

پیوست آ

مفاهیم اولیه

آ.۱ مقدمه

در این پیوست یک سری تعاریف و مفاهیم اولیه از جبر ماتریس‌ها، توزیع‌ها و همچنین محاسبات مورد استفاده در این پایان‌نامه را بیان می‌کنیم.

آ.۲ جبر ماتریس‌ها

در این بخش تعاریف و قضایای مورد نیاز در این پایان‌نامه، از جبر ماتریس‌ها را بیان می‌کنیم. این مطالب را می‌توان در اکثر کتاب‌های جبر خطی و یا کتاب‌های تحلیل چند متغیره یافت. از جمله آندرسن^۱ (۲۰۰۳) و گوپتا و ناگار^۲ (۲۰۰۰) را ببینید.

تعریف آ.۲.۱. فرض کنید $A = (a_{ij})$ یک ماتریس مربع در اندازه $p \times p$ باشد آن‌گاه

i. A یک ماتریس قطری^۳ است و با نماد $\text{diag}(a_{11}, \dots, a_{pp})$ نشان داده می‌شود اگر به ازای $a_{ij} = 0, i \neq j$.

ii. A یک ماتریس متقارن^۴ است اگر به ازای $a_{ij} = a_{ji}, i \neq j$ ؛ به عبارت دیگر $A = A'$.

iii. A یک ماتریس متعامد^۵ است اگر $AA' = A'A = I_p$.

iv. A یک ماتریس معین مثبت^۶ (نیمه معین مثبت^۷) است و با نماد $A > 0$ ($A \geq 0$) نشان

^۱Anderson

^۲Gupta and Nagar

^۳Diagonal matrix

^۴Symmetric matrix

^۵Orthogonal

^۶Positive definite

داده می‌شود اگر A متقارن باشد و برای هر بردار غیر صفر $x_{p \times 1}$ داشته باشیم، $x'Ax > 0$ (۱).
 $(x'Ax \geq 0)$.

تعریف ۲.۲. اگر $A_{n \times p}$ دارای رتبه p باشد که در آن $p < n$ ، در این صورت A دارای رتبه‌ی ماکسیمم است و آن را با رتبه‌ی کامل گوئیم. به طور کلی، رتبه‌ی ماکسیمم ممکن یک ماتریس $n \times p$ برابر $\min(n, p)$ است.

تعریف ۳.۲. فرض کنید A ماتریس $p \times p$ باشد، آن‌گاه ریشه‌های معادله‌ی

$$\det(A - \lambda I_p) = 0$$

مقادیر ویژه λ ماتریس A نامیده می‌شوند.

قضیه ۴.۲. فرض کنید $\lambda_1, \dots, \lambda_p$ مقادیر ویژه‌ی ماتریس $A_{(p \times p)}$ باشند، آن‌گاه

$$\det(A) = \prod_{i=1}^p \lambda_i \quad \text{i.}$$

$$\text{tr}(A) = \sum_{i=1}^p \lambda_i \quad \text{ii.}$$

iii. رتبه‌ی A برابر تعداد مقادیر ویژه‌ی مخالف صفر ماتریس A است.

iv. ماتریس A معین مثبت (نیمه معین مثبت) است اگر و تنها اگر همه‌ی مقادیر ویژه‌ی آن مثبت (نامنفی) باشند.

تعریف ۵.۲. دو بردار $a_{n \times 1}$ و $b_{n \times 1}$ را متعامد^۹ می‌نامند، هرگاه

$$a'b = a_1b_1 + a_2b_2 + \dots + a_nb_n = 0.$$

تعریف ۶.۲. ماتریس $A = (c_1, c_2, \dots, c_p)$ متعامد است، هرگاه ستون‌های آن بردارهای متعامد باشند، که در آن c_i ها به ازای $i = 1, \dots, p$ بردارهای ستونی می‌باشند.

قضیه ۷.۲. (تجزیه طیفی^{۱۰}) فرض کنید $A_{(p \times p)}$ یک ماتریس متقارن باشد. آن‌گاه ماتریس متعامد H وجود دارد به طوری که

$$A = HDH'$$

و D ماتریس قطری است که درایه‌های روی قطر آن مقادیر ویژه‌ی ماتریس A است.

قضیه ۸.۲. اگر $A_{n \times n}$ معین مثبت (نیمه معین مثبت) با مقادیر ویژه $\lambda_1, \lambda_2, \dots, \lambda_n$ باشد، آن‌گاه:

$$(1) \text{ عناصر قطر اصلی } a_{ii} > 0 \text{ (} a_{ii} \geq 0 \text{) می‌باشند.}$$

$$(2) A \text{ ناویژه است.}$$

$$(3) A^{-1} \text{ معین مثبت است.}$$

$$(4) (\lambda_i \geq 0, \lambda_i > 0 \text{ برای } i = 1, 2, \dots, n).$$

^۷Positive semidefinite

^۸Eigenvalues

^۹Orthogonal

^{۱۰}Spectral decomposition

قضیه ۲.۹.۲. فرض کنید A یک ماتریس $n \times p$ است.

(۱) اگر $\text{rank}(A) = p$ ، آنگاه $A'A$ معین مثبت است.

(۲) اگر $\text{rank}(A) < p$ ، آنگاه $A'A$ نیمه معین مثبت است.

پیوست ب

توزیع‌های آماری

ب.۱ برخی توزیع‌های آماری

در این بخش برخی از توزیع‌های یک متغیره و چند متغیره مورد استفاده در این پایان‌نامه را به اختصار معرفی می‌کنیم.

ب.۱.۱ توزیع‌های یک متغیره

تعریف ب.۱.۱.۱. (توزیع بتا). گوئیم متغیر تصادفی X دارای توزیع بتا با پارامترهای α و β است و می‌نویسیم

$$X \sim B(\alpha, \beta)$$

اگر تابع چگالی احتمال آن به صورت زیر باشد:

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 \leq x \leq 1, \beta \geq 0$$

در این صورت

$$E(X) = \frac{\alpha}{\alpha + \beta}, \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)(\alpha + \beta + 1)^2}$$

تعریف ب.۲.۱.۱. (توزیع گاما). گوئیم متغیر تصادفی X دارای توزیع گامای یک متغیره با پارامترهای a و b است، اگر تابع چگالی احتمال آن به صورت زیر باشد:

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp\{-bx\}, \quad x > 0, a > 0, b > 0$$

و با نماد

$$X \sim G(a, b)$$

نشان داده می‌شود. در این صورت

$$E(X) = \frac{a}{b}, \quad \text{Var}(X) = \frac{a}{b^2}.$$

اگر $a = n/2$ و $b = 1/2$ باشد آن‌گاه X دارای توزیع کی‌دو با n درجه‌ی آزادی است.

تعریف ب.۳.۱. (توزیع گامای معکوس^۱) متغیر تصادفی W دارای توزیع گامای معکوس یک متغیره با پارامترهای α و σ است، اگر تابع چگالی احتمال آن به صورت زیر باشد:

$$f(w) = \frac{\sigma^\alpha}{\Gamma(\alpha)} w^{-\alpha-1} \exp\left\{-\frac{\sigma}{w}\right\}, \quad w > 0, \alpha > 0, \sigma > 0.$$

و با نماد

$$W \sim IG(\alpha, \sigma)$$

نشان داده می‌شود. در این صورت

$$E(X) = \frac{\sigma}{\alpha - 1}, \quad \alpha > 1, \quad \text{Mod}(X) = \frac{\sigma}{\alpha + 1}.$$

توزیع گامای معکوس، توزیعی پیوسته و دو پارامتری است که تنها روی متغیرها با مقدار مثبت تعریف می‌شود و در مدل‌سازی بیزی زمانی که واریانس توزیع نرمال نامعلوم است به عنوان توزیع پیشین حاشیه‌ایی مورد استفاده قرار می‌گیرد.

تعریف ب.۴.۱. (توزیع نرمال بریده شده). گوییم متغیر تصادفی X دارای توزیع نرمال بریده شده با پارامترهای μ و σ^2 است اگر تابع چگالی احتمال آن به صورت زیر باشد:

$$f(x; \mu, \sigma, a, b) = \frac{\frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \quad a \leq x \leq b$$

که در آن ϕ تابع چگالی احتمال توزیع نرمال استاندارد، و $\Phi(\cdot)$ توزیع تجمعی نرمال است. اگر $b = \infty$ باشد آنگاه

$$\Phi\left(\frac{b-\mu}{\sigma}\right) = 1$$

و به طور مشابه اگر $a = -\infty$ باشد آنگاه

$$\Phi\left(\frac{a-\mu}{\sigma}\right) = 0.$$

تعریف ب.۵.۱. (توزیع نرمال چند متغیره^۲) بردار تصادفی X دارای توزیع نرمال p -متغیره با بردار میانگین μ و ماتریس واریانس-کوواریانس Σ است (آن را با نماد $X \sim N_p(\mu, \Sigma)$ نشان می‌دهیم) اگر تابع چگالی احتمال آن به صورت زیر است

$$f_X(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{(x-\mu)' \Sigma^{-1} (x-\mu)}{2}\right\}$$

که در آن $x, \mu \in R^{p \times p}, \Sigma > 0$.

تعریف ب.۶.۱. (تابع نشانگر). یک تابع مشخصه است، که روی یک مجموعه X تعریف شده و عضویت را در یک زیر مجموعه (A) از X را به صورت زیر نشان می‌دهد:

$$I_A : X \rightarrow \{0, 1\}$$

$$I_A(X) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

^۱Inverse Gamma distribution

^۲Multivariate normal

پیوست

اثبات

تعریف پ.۷.۰. (تقریب لاپلاس). این نوع تقریب برای محاسبه‌ی انتگرال‌هایی به صورت

$$I = \int f(\theta) d\theta$$

یا

$$I = \int g(\theta) \exp\{-nh(\theta)\} d\theta$$

مناسب است که در آن $g: \mathbb{R}^p \rightarrow \mathbb{R}$ و $h: \mathbb{R}^p \rightarrow \mathbb{R}$ توابع همواری هستند که حداقل سه بار مشتق‌پذیرند.

فرض کنید $\hat{\theta}$ مقداری از θ است که h را کمینه می‌کند. روش لاپلاس I را به صورت زیر تقریب می‌زند

$$\hat{I} = g(\hat{\theta}) \left(\frac{2\pi}{n}\right)^{\frac{p}{2}} |\hat{\Sigma}|^{-\frac{1}{2}} \exp\{-nh(\hat{\theta})\}$$

که در آن

$$\hat{\Sigma} = \left[\frac{\partial^2 h(\hat{\theta})}{\partial \theta \partial \theta^T} \right]^{-1}$$

پ.۱ اثبات رابطه‌ی (۱۳.۲)

رابطه زیر را در نظر بگیرید:

$$h(\sigma^2, X) = (\sigma^2)^{-\left(\frac{m}{2} + a + 1\right)} \times \exp\left\{-\frac{\left(\frac{SSR}{2} + b\right)}{\sigma^2} - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\}$$

محاسبه انتگرال $h(\sigma^2, X)$ آسان نیست. در بیشتر موارد بزرگ بودن مقدار $m = n(n-1)/2$ منجر به غالب شدن عبارت درست‌نمایی می‌شود بنابراین می‌توانیم $h(\sigma^2, X)$ را به طور تقریبی متناسب با عبارت درست‌نمایی در نظر بگیریم.

$$L(X, \sigma^2) \equiv (\sigma^2)^{-\frac{m}{2}} \exp\left\{-\frac{SSR}{2\sigma^2} - \sum_{i>j} \log \Phi\left(\frac{\delta_{ij}}{\sigma}\right)\right\}$$

به علاوه به دلیل m بزرگ، عبارت درست‌نمایی $L(\sigma^2, X)$ را می‌توان با یک تابع چگالی نرمال تقریب زد.

حال، با استفاده از تقریب لاپلاس برای انتگرال $L(\sigma^2, X)$ داریم:

$$\int h(\sigma^2, X) d\sigma^2 \approx (2\pi)^{1/2} H^{-1/2} L(X, \hat{\sigma}^2)$$

ادامه‌ی اثبات فراتر از حد این پایان‌نامه می‌باشد. (برای آگاهی بیشتر اُه و رافتری، ۲۰۰۱ را ببینید).

پیوست

مقدماتی از استنباط بیزی و $MCMC$

ت. ۱. مقدمه‌ای از استنباط بیزی

در استنباط‌های آماری دو الگو عمده کلاسیک و بیزی مورد استفاده قرار می‌گیرند. در روش‌های بیزی، θ را کمیتی در نظر می‌گیریم که خود یک متغیر تصادفی است و تغییرات آن توسط یک تابع که به آن توزیع پیشین می‌گویند توجیه می‌شود. یکی از دلایل شهرت خوب رویکردهای بیزی، جواز استفاده از اطلاعات پیشین در بدست آوردن راه‌حلی در اعمال استنباط و مسئله تصمیم‌گیری است، این مسئله زمانی ارزشمندتر است که اطلاعات نمونه محدود باشد. روش‌های بیزی ترکیب فرضیه‌های علمی را در تحلیل (با استفاده از توزیع پیشین) ممکن می‌سازد.

ت. ۲. اصل بیز

در روش بیزی، θ به عنوان یک متغیر تصادفی با توزیع احتمال $\pi(\theta)$ در نظر گرفته می‌شود، که π را توزیع پیشین می‌نامند.

$f(x|\theta)$ تابع چگالی احتمال شرطی از متغیر تصادفی X است که در آن $\theta \in \Theta$ ثابت فرض شده، و با توجه به این‌که π توزیع θ است، در نتیجه تابع چگالی توام θ و X به صورت زیر محاسبه می‌شود:

$$f(x, \theta) \approx \pi(\theta)f(x|\theta).$$

لازم به ذکر است که توزیع پیشین، بر اساس اطلاعات قبلی، اعتقادات، باورها و تجربیات آزمایشگر تعیین می‌گردد. بر اساس نمونه جمع‌آوری شده از جامعه، توزیع پیشین تصحیح می‌گردد، توزیع پیشین تصحیح شده را توزیع پسین می‌نامند. به عبارتی ترکیبی از اطلاعات حاصل از مشاهدات بدست آمده از جامعه (تابع درست‌نمایی) و توزیع پیشین را بعنوان معیار توزیع پسین برای تعیین برآوردگر پارامتر نامعلوم بکار می‌برند.

ت.۳ مونت کارلو

مونت کارلو روشی برای تقریب انتگرال‌هایی است که به صورت مستقیم قابل محاسبه نمی‌باشند. در این خصوص برای محاسبه انتگرال پیچیده‌ایی مانند $\int_a^b h(x)dx$ اگر بتوان $h(x)$ را به صورت حاصلضرب یک تابع $f(x)$ در یک تابع چگالی احتمال $\pi(x)$ که روی فاصله (a, b) تعریف می‌شوند تجزیه کرد در این صورت می‌توان نوشت

$$\int_a^b h(x)d(x) = \int_a^b f(x)\pi(x)d(x) = E_{\pi(x)}[f(x)]$$

حال با تولید نمونه X_t از چگالی $\pi(x)$ داریم:

$$\int_a^b h(x)d(x) = E_{\pi(x)}[f(x)] \approx \frac{1}{n} \sum_{t=1}^n f(x_t)$$

این روش به انتگرال گیری مونت کارلو معروف است. وقتی نمونه‌های $\{X_t\}$ مستقل هستند، با افزایش n با توجه به قانون اعداد بزرگ^۱، تقریب از دقت بالاتری برخوردار خواهد بود و برآوردگر مونت کارلو به امید ریاضی $f(x)$ همگراست.

برای تولید نمونه‌ها از $\pi(x)$ ، تکنیک‌های مختلفی نظیر تبدیل متغیرها، نمونه‌گیری از نقاط مهم^۲، نمونه‌گیری رد و پذیرش^۳ و... بکار گرفته می‌شوند که تقریباً همه روش‌ها، نمونه‌های مستقلی تولید می‌کنند.

ت.۴ زنجیره مارکف

یک زنجیره مارکف $\{X_t\}$ یک فرآیند تصادفی زمان گسسته^۴ $\{X_0, X_1, \dots\}$ است به طوری که توزیع X_t به شرط مقادیر گذشته فرآیند، X_0, X_1, \dots, X_{t-1} فقط به X_{t-1} بستگی دارد یعنی برای هر A ،

$$P[X_t \in A | X_0, X_1, \dots, X_{t-1}] = P[X_t \in A | X_{t-1}]$$

که $P[.]$ هسته انتقال^۵ زنجیر نامیده می‌شود و به t (زمان) بستگی ندارد.

ت.۵ توزیع مانا

توزیع π را توزیع مانای زنجیر مارکف $\{X_t : t \geq 0\}$ با ماتریس احتمال‌های تغییر وضعیت $P = (P_{ij})$ گوئیم هر گاه به ازای هر j داشته باشیم:

$$\pi(j) = \sum_i \pi(i)P_{ij} \quad \forall j$$

^۱Law of large numbers

^۲Importance sampling

^۳Acceptance - rejection sampling

^۴Discrete time

^۵Transition Kernel

یعنی اگر توزیع فرآیند به این صورت باشد، از آن پس تمام X_t ها نیز دارای همین توزیع خواهند بود. به عبارت دیگر اگر زنجیر مارکف با توزیع آغازین شروع کن، آنگاه توزیع زنجیر، مشابه توزیع آغازین زنجیر خواهد بود. یک شرط کافی برای وجود یک توزیع مانای یکتا آن است که :

$$\pi(i)P_{ij} = \pi(j)P_{ji}$$

ت.۶ روش شبیه‌سازی MCMC

روش‌های مونت کارلوی زنجیره مارکفی (MCMC) برای نمونه‌گیری غیر مستقیم از توزیع‌های پیچیده استفاده می‌شود. در این روش توزیعی مانند $\pi(X)$ ، که توزیع هدف^۶ می‌شود را در نظر می‌گیریم. اگر π آنقدر پیچیده باشد که نتوان از آن نمونه‌گیری کرد، یک زنجیره مارکف با فضای حالت E ، که توزیع مانای آن $\pi(X)$ باشد می‌سازیم. اگر زنجیری به طول n تولید نماییم که حالت آغازین خود را فراموش کند بعد از m نمونه تولید شبیه‌سازی شده، $\{X_t : t = m + 1, \dots, n\}$ می‌تواند به عنوان نمونه‌ای همبسته از توزیع هدف رفتار کند و برای توصیف خصوصیات مهم π مورد استفاده قرار گیرد. حال می‌توان از نمونه‌های باقی‌مانده در زنجیر، برای برآورد $E[f(X)]$ که در آن X دارای توزیع $\pi(X)$ است استفاده کرد. (m نمونه اول چون تحت تاثیر نقطه آغازین زنجیر هستند کنار گذاشته می‌شوند که به سوزاندن معروف است).

ت.۷ متروپلیس - هستینگز

متداولترین روش MCMC روش نمونه‌گیری متروپلیس - هستینگز^۷ می‌باشد. فرض می‌کنیم هدف ما انتخاب نمونه از توزیع $\pi(X)$ باشد که در آن $\pi(X) = \frac{f(X)}{k}$ و k ثابت نرمال‌ساز نامیده می‌شود و ممکن است نا معلوم یا محاسبه آن دشوار باشد. الگوریتم متروپلیس - هستینگز زنجیری از مشاهدات این توزیع را به صورت زیر تولید می‌کند:

۱. با مقدار دلخواه اولیه^۸ x_0 که $f(x_0) > 0$ آغاز می‌کنیم.

۲. با استفاده از مقدار موجود x_0 ، یک نقطه پیشنهادی x^* از یک توزیع جهش^۹ $q(x_1, x_2)$ نمونه‌گیری می‌کنیم. که این توزیع عبارت است از احتمال رفتن به حالت x_2 به شرط آنکه مقدار قبلی x_1 باشد.

^۶Target

^۷Metropolis - Hastings

^۸Initial

^۹Jump

۳. با داشتن نقطه پیشنهادی x^* ، نسبت زیر را در نقاط x_{t-1} محاسبه می‌کنیم:

$$\alpha(x_{t-1}, x^*) = \frac{\pi(x^*)q(x^*, x_{t-1})}{\pi(x_{t-1})q(x_{t-1}, x^*)} = \frac{f(x^*)q(x^*, x_{t-1})}{f(x_{t-1})q(x_{t-1}, x^*)}$$

۴. اگر $\alpha \geq 1$ ، نقطه پیشنهادی x^* را می‌پذیریم و قرار می‌دهیم $x_t = x^*$ و به گام ۲ برمی‌گردیم. در غیر این صورت با احتمال α نقطه پیشنهادی را می‌پذیریم و با احتمال $1 - \alpha$ نقطه پیشنهادی را رد کرده و به گام دو برمی‌گردیم.

نمونه‌گیری متروپلیس - هستینگز را می‌توانیم به صورت زیر خلاصه کنیم:
ابتدا

$$\alpha(x_{t-1}, x^*) = \min \left(\frac{f(x^*)q(x^*, x_{t-1})}{f(x_{t-1})q(x_{t-1}, x^*)}, 1 \right)$$

را محاسبه کرده و آنگاه نقطه پیشنهادی را با احتمال α می‌پذیریم. برای این کار یک نقطه تصادفی u از توزیع یکنواخت $(0, 1)$ انتخاب می‌کنیم. آنگاه اگر $\alpha > u$ باشد نقطه پیشنهادی را می‌پذیریم و در غیر این صورت رد می‌کنیم. این روند یک زنجیره مارکف $x_0, x_1, \dots, x_k, \dots$ را تولید می‌کند، که احتمال انتقال از x_t به x_{t+1} تنها به x_t بستگی دارد و به x_0, x_1, \dots, x_{t-1} بستگی ندارد. پس از یک دوره سوختگی^{۱۰} m تایی زنجیر به حالت مانا می‌رسد و نمونه $x_{m+1}, x_{m+2}, \dots, x_n$ از $\pi(X)$ خواهد بود. تحت الگوریتم متروپلیس - هستینگز، از $q(x, y)$ نمونه گرفته و آنگاه انتقال را با احتمال $\alpha(x, y)$ می‌پذیریم. بنابراین

$$P(x \rightarrow y) = q(x, y)\alpha(x, y) = q(x, y)\min \left(\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1 \right)$$

در این روش انتخاب توزیع جهش در طول دوره سوختگی و سرعت همگرایی زنجیر تاثیر بسزایی دارد.

ت.۱.۷ الگوریتم نمونه‌گیر گیبز

در سال‌های اخیر، آمارشناسان به‌طور عمده‌ای روش‌های الگوریتم MCMC را به کار برده‌اند. الگوریتم نمونه‌گیر گیبز، یکی از بهترین روش‌های شناخته‌شده است. روش نمونه‌گیر گیبز یک ابزار شبیه‌سازی برای یافتن نمونه‌هایی از چگالی‌های توأم پیچیده می‌باشد. نمونه‌گیری گیبز در به‌دست آوردن توزیع‌های پسین کناری^{۱۱}، در مدل‌های بیزی می‌باشد (کسلا و همکاران، ۱۹۹۲). یکی از محدودیت‌ها در کاربرد گسترده روش‌های بیزی، یافتن توزیع پسین کناری است، که اغلب نیازمند انتگرال‌گیری از تابع‌هایی با ابعاد بالا بوده و این نیز مستلزم محاسبات بسیار پیچیده می‌باشد. معمولاً وقتی بردار پارامتر چندبعدی است، حتی روش‌های عددی انتگرال‌گیری مستقیم هم ممکن است برای حل چنین انتگرال‌هایی مفید نباشند. در این صورت، استفاده از روش نمونه‌گیر گیبز می‌تواند به‌عنوان یک روش جایگزین مؤثر واقع شود.

^{۱۰} Burn-in period

^{۱۱} Marginal Posterior Distribution

نمونه گیر گیبز دومرحله‌ای

نمونه‌گیر گیبز دومرحله‌ای یک زنجیر مارکوف را از توزیع توأم، به صورت زیر ایجاد می‌کند. اگر فرض کنیم متغیرهای تصادفی X و Y دارای تابع چگالی توأم $p(x, y)$ با چگالی‌های شرطی $p(x|y)$ و $p(y|x)$ هستند، نمونه‌گیر دومرحله‌ای گیبز مقادیر $(X^{(t)}, Y^{(t)})$ ، $t = 1, 2, \dots$ را به صورت زیر تولید می‌کند: با فرض $X_0 = x_0$ برای $t = 0, 1, 2, \dots$ تولید می‌کنیم:

$$Y^{(t+1)} \sim p(y|x^{(t)}) \quad ۱.$$

$$X^{(t+1)} \sim p(x|y^{(t+1)}) \quad ۲.$$

در صورتی که شبیه‌سازی در هر دو مرحله ساده و تولید نمونه از آن‌ها ساده باشد، اجرای الگوریتم آسان است. یا به عبارت دیگر، در صورتی که چگالی‌های شرطی فرم بسته‌ای داشته باشند، شبیه‌سازی به سادگی انجام می‌شود. بنابراین، در صورتی که نمونه‌گیری مستقیم قابل انجام نباشد، تقریب‌های مونت کارلو می‌تواند مورد استفاده قرار بگیرد. به راحتی در می‌یابیم که اگر $(X^{(t)}, Y^{(t)})$ از توزیع $p(x, y)$ باشد، در این صورت $(X^{(t+1)}, Y^{(t+1)})$ نیز از توزیع $p(x, y)$ خواهد بود. به عبارت دیگر توزیع مانای این زنجیر، $p(x, y)$ است.

نمونه‌گیر گیبز چندمرحله‌ای

نمونه‌گیر گیبز چندمرحله‌ای تعمیمی از نمونه‌گیر گیبز دومرحله‌ای است. فرض کنید که برای $m > 1$ ، متغیر تصادفی Y به صورت $Y = (Y_1, \dots, Y_m)$ نوشته شود. مؤلفه‌های Y_i ممکن است تک بعدی یا چند بعدی باشند. می‌توان برای $m, 1, 2, \dots$ به صورت زیر شبیه‌سازی کرد:

$$Y_i | y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_m \sim p_i(y_i | y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_m)$$

الگوریتم نمونه‌گیر گیبز تحت انتقال از $Y^{(t)}$ به $Y^{(t+1)}$ با تکرار $t = 1, 2, \dots$ و با فرض کردن $y^{(t)} = (y_1^{(t)}, y_2^{(t)}, \dots, y_m^{(t)})$ طی گام‌های زیر تولید می‌شود:

$$Y_1^{(t+1)} \sim p_1(y_1 | y_2^{(t)}, \dots, y_m^{(t)}) \quad ۱.$$

$$Y_2^{(t+1)} \sim p_2(y_2 | y_1^{(t+1)}, y_3^{(t)}, \dots, y_m^{(t)}) \quad ۲.$$

$$\vdots$$

$$Y_m^{(t+1)} \sim p_m(y_m | y_1^{(t+1)}, \dots, y_{m-1}^{(t+1)}) \quad m.$$

چگالی‌های p_1, \dots, p_m چگالی شرطی کامل^{۱۲} نامیده می‌شوند. یک خصوصیت نمونه‌گیر گیبز این است که تنها چگالی‌ها برای شبیه‌سازی مورد استفاده قرار می‌گیرند. بنابراین، حتی مسئله‌ای با ابعاد بالا را می‌توان به چند مسئله با بعدهای کوچکتر تقسیم کرد که این مزیت روش نمونه‌گیری گیبز به حساب می‌آید.

^{۱۲}Full Conditional Density

ت.۲.۷ الگوریتم متروپلیس-هستینگز تحت گیبز

نمونه‌گیر گیبز را در صورتی که برخی از چگالی‌های شرطی کامل را نتوان به‌سادگی شبیه‌سازی کرد، به آسانی با تنظیماتی می‌توان تعمیم داد. اگر تحت یک مجموعه از چگالی‌های شرطی کامل p_1, \dots, p_m برخی از p_i ها غیرمتداول باشند، نتایج نمونه‌گیر گیبز در صورتی که راهکار متروپلیس تحت گیبز را به کار بگیریم، به مخاطره نمی‌افتد. مثلاً فرض کنید شبیه‌سازی از

$$Y_i^{(t+1)} \sim p_i(y_i | y_1^{(t+1)}, \dots, y_{i-1}^{(t+1)}, y_{i+1}^{(t)}, \dots, y_m^{(t)})$$

مشکل باشد. در این حالت می‌توان برای تولید نمونه از آن (در یک مرحله از مراحل نمونه‌گیری گیبز) از الگوریتم MH استفاده کرد. در این حالت این گام با یک الگوریتم MH که توزیع مانای آن $p(y_i | y_1^{(t+1)}, \dots, y_{i-1}^{(t+1)}, y_{i+1}^{(t)}, \dots, y_m^{(t)})$ است، جایگزین می‌شود. این نوع الگوریتم ترکیبی از دو الگوریتم گیبز و MH است که به آن الگوریتم متروپلیس-هستینگز تحت گیبز گویند. در حالتی که شبیه‌سازی بیشتر از یک چگالی شرطی کامل مشکل باشد، می‌توان به همین ترتیب مراحل مختلفی از نمونه‌گیر گیبز را با الگوریتم‌های MH ترکیب کرد.

نرم‌افزارهای $JAGS$ ^{۱۳}، $WinBUGS$ ^{۱۴} و $OpenBUGS$ (اندرو و فینلی، ۲۰۱۳) به‌طور خودکار تولید نمونه از توزیع‌های پسین را با روش نمونه‌گیری گیبز یا MH تحت گیبز انجام می‌دهند و چگالی‌های شرطی کامل را شبیه‌سازی می‌کنند. (برای آگاهی بیشتر کیهانی ۱۳۹۴ را ببینید).

^{۱۳}Just Another Gibbs Sampler

^{۱۴}Win Bayesian inference Using Gibbs Sampler

پیوست

برنامه‌های کامپیوتری با نرم‌افزار R

در این پیوست تمام دستورهای مورد نیاز برای رسم شکل‌ها و جداول با نرم‌افزار R آمده است.

برنامه مربوط به مثال ۱.۴.۱: کد ۱:

```
> data("watervoles", package = "HSAUR2")
> voles_mds <- cmdscale(watervoles, k = 13, eig = TRUE)
Warning message:
In cmdscale(watervoles, k = 13, eig = TRUE) :
  only 7 of the first 13 eigenvalues are > 0
```

کد ۲:

```
> cumsum(abs(voles_mds$eig))/sum(abs(voles_mds$eig))

> cumsum((voles_mds$eig)^2)/sum((voles_mds$eig)^2)
```

کد ۳:

```
> x <- voles_mds$points[,1]
> y <- voles_mds$points[,2]
> plot(x, y, xlab = "Coordinate 1", ylab = "Coordinate 2",
+ xlim = range(x)*1.2, type = "n")
> text(x, y, labels = colnames(watervoles), cex = 0.7)
```

برنامه مربوط به ۱.۶.۱: کد:

```
> Htl = c(0, 3, 4, 7, 3, 8, 3, 4, 8, 9, 4, 7)
> Mss = c(3, 0, 6, 8, 5, 9, 2, 4, 9, 9, 5, 8)
> Chr = c(4, 6, 0, 4, 6, 3, 5, 3, 8, 5, 5, 2)
> Esn = c(7, 8, 4, 0, 8, 9, 7, 5, 9, 4, 4, 4)
> Stl = c(3, 5, 6, 8, 0, 8, 6, 6, 6, 7, 7, 7)
> Att = c(8, 9, 3, 8, 9, 0, 7, 5, 9, 8, 2, 8)
> Frn = c(3, 2, 5, 7, 6, 7, 0, 4, 8, 8, 2, 3)
> DGl = c(4, 4, 3, 5, 6, 5, 4, 0, 7, 4, 5, 2)
> MT = c(8, 9, 8, 9, 6, 9, 8, 7, 0, 4, 9, 4)
> Trm = c(9, 9, 5, 4, 7, 8, 8, 4, 4, 0, 5, 5)
> Chm = c(4, 5, 5, 4, 7, 2, 2, 5, 9, 5, 0, 7)
> Tit = c(7, 8, 2, 4, 7, 8, 3, 2, 4, 5, 7, 0)
> WWIleaders = cbind(Htl, Mss, Chr, Esn, Stl, Att, Frn, DGl, MT, Trm, Chm, Tit)
> rownames(WWIleaders) = c('Hitler', 'Mussolini', 'Churchill', 'Eisenhower', 'Stal
  'Tito')
> (WWII_mds <- isoMDS(WWIleaders))
initial value 20.400713
iter 5 value 15.219311
iter 5 value 15.207733
iter 5 value 15.207733
final value 15.207733
converged
$points
      [,1]      [,2]
Hitler -2.6142309 -1.7379740
Mussolini -3.9069970 -1.2052278
Churchill 0.3240108 1.4415523
Eisenhower 3.0235774 2.8142761
Stalin -1.4970418 -3.7542676
Attlee -2.0361086 5.0863312
Franco -2.8641212 0.1038147
De Gaulle 0.6474253 -0.2385760
Mao Tse-Tung 4.0870170 -4.6631954
```

```
Truman      4.4895837  0.2057644
Chamberlin -2.1052548  2.7707306
Tito        2.3601946 -1.0613770
```

```
$stress
```

```
[1] 15.20773
```

```
> plot(WWII_mds$points[,1], WWII_mds$points[,2],
+       type="n",xlab="Coordinate 1",ylab="Coordinate 2",
+       xlim=c(-6,6), ylim=c(-6,6))
> text(WWII_mds$points[,1], WWII_mds$points[,2],
+       labels=rownames(WWIIleaders))
```

ث.۱. کدهای مربوط به فصل ۴

```
> F<-function(D){
+
+ SSR<-function(X1,sigma){
+ X1<-as.matrix(X1)
+ p<-dim(X1)[2];n<-dim(X1)[1]
+ s<-d<-matrix(0,n,n)
+ for(ii in 1:n) for(jj in 1:n){
+ ss<-0; for( k in 1:p) ss<-ss+(X1[ii,k]-X1[jj,k])^2
+ s[ii,jj]=sqrt(ss)
+ d[ii,jj]=rnorm(1,s[ii,jj],sigma)
+ if(d[ii,jj]<0)d[ii,jj]=0 }
+ s1<-(d-s)^2;s2<-d^2
+ s11<-0->s22;for(ii in 1:n) for(jj in 1:i) {
+ s11=s11+s1[ii,jj]}
+ s11}
+
+ dd<-D
+ n<-dim(as.matrix(dd))[1]
+ sigma<-CMDS.STRESS<-BMDS.STRESS<-rep(0,n)
+ XX<-array(0,c(n,n,n))
+ P=which(cmdscale(D,k =(n-1),eig=TRUE)$eig<0)[1]-1
```

```

+ for( i in 1:P){
+ X1<-cmdscale(D, k = i, eig = TRUE)$points
+ CMDS.STRESS[i]<-sqrt((sum((D-dist(X1))^2)/sum(dist(X1)^2)))
+ p1<-dim(X1)[2];n1<-dim(X1)[1]
+ for( r in 1:1000){
+ Cx<-scale(X1)
+ Vax<-eigen(Cx%*%t(Cx))$values[1:p1]
+ Vex<-eigen(Cx%*%t(Cx))$vectors[,1:p1]
+ if(i==1) Xa<-as.matrix(Vex)*(Vax^.5) else Xa<-Vex%*%(diag(Vax)^.5)
+ SSR<-sum((D-dist(Xa))^2)
+ m<-n1*(n1-1)/2;sigma[i]<-SSR/m
+ if(i==1) landa0<-Vax else landa0<-diag(Vax)
+ x2<-mvrnorm(n1,rep(0,i),landa0)
+ X1<-x2}
+ for(ii in 1:10) for(jj in 1:i) XX[ii,jj,i]<-X1[ii,jj]
+ BMDS.STRESS[i]=sqrt((sum((D-dist(X1))^2)/sum(dist(X1)^2)))}
+
+ sigma<-sigma[1:P];XX<-XX[, ,1:P]
+ RR<-SS<-array(0,c(1,n,P))
+ for(ii in 1:P) { SS[, ,ii]<-apply(XX[, ,ii]^2,2,sum)
+ if(ii>1) RR[, ,ii]<-SS[, ,ii]/SS[, ,ii-1]}
+ LRT<-Penalty<-MDSIC<-rep(0,(P-1))
+ for( i in 1:(P-1)){
+ LRT[i]<-(m-2)*log(SSR(XX[,1:(i+1),(i+1)],sigma[i+1])/SSR(XX[,1:i,i],sigma[i]))
+ DF<-0;for(j in 1:i) DF<-DF+log((RR[1,j,i+1]*(n1+1))/(n1+RR[1,j,i+1]))
+ Penalty[i]<-DF
+ if(i==1) MDSIC[1]=(m-2)*log(SSR(XX[,1,1],sigma[1]))
+ else MDSIC[i]=sum(LRT[1:(i-1)])+ sum(Penalty[1:(i-1)])+(m-2)*log(SSR(XX[,1,1],sigma[1]))
+ Dim<-1:(P-1);CMDS.STRESS<-CMDS.STRESS[1:(P-1)];BMDS.STRESS<-BMDS.STRESS[1:(P-1)]
+ cbind(Dim,CMDS.STRESS,BMDS.STRESS,LRT,Penalty,MDSIC)}

```

وارد کردن داده ها

X۱ داده های مربوط به دلستر شماره ۱

X۲ داده های مربوط به دلستر شماره ۲

X۳ داده های مربوط فاصله شهرها

```

> x1<-matrix(scan(),19,12)
1: 1 1 0 1 0 1 0 0 0 1 1 0 0 0 0 0 0 1 0
20: 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0

```

```

39: 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0
58: 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 1 0 0 0
77: 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
96: 0 0 1 0 1 1 1 1 1 0 0 1 0 0 0 0 0 1 1
115: 0 0 1 0 1 1 1 0 0 0 0 0 1 1 1 1 0 0 1
134: 0 0 0 1 1 0 0 0 1 0 0 1 1 1 1 0 1 0 0
153: 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0
172: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
191: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1
210: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
229:
Read 228 items
> D1<-dist(X1)
>
> x1<-matrix(scan(),19,14)
1: 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
20: 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
39: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
58: 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0
77: 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 0
96: 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0
115: 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0
134: 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0
153: 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 1
172: 0 0 0 0 0 0 1 0 0 0 0 1 1 0 1 0 0 0 0
191: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0
210: 0 1 1 0 1 0 0 0 1 1 0 0 0 0 1 0 0 1 0
229: 0 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0
248: 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
267:
Read 266 items
>
> D2<-dist(x2)
> D2<-dist(x2)
> x3<-matrix(scan(),30,30,T)
1: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
31: 39 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
61: 22 20 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
91: 59 20 39 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

```



```

for(ii in 1:n) for(jj in 1:n){
ss<-0; for( k in 1:p) ss<-ss+(X1[ii,k]-X1[jj,k])^2
s[ii,jj]=sqrt(ss)
d[ii,jj]=rnorm(1,s[ii,jj],sigma)
if(d[ii,jj]<0)d[ii,jj]=0 }
s1<-(d-s)^2;s2<-d^2
s11<-0->s22;for(ii in 1:n) for(jj in 1:i) {
s11=s11+s1[ii,jj]}
s11}

```

```

STRESS<-function(x,sigma){
p<-dim(x)[2];n<-dim(x)[1]
s<-d<-matrix(0,n,n)
for(i in 1:n) for(j in 1:n){
ss<-0; for( k in 1:p) ss<-ss+(x[i,k]-x[j,k])^2
s[i,j]=sqrt(ss)
d[i,j]=rnorm(1,s[i,j],sigma)
if(d[i,j]<0)d[i,j]=0 }
s1<-(d-s)^2;s2<-d^2
s11<-0->s22;for(i in 1:n) for(j in 1:i) {
s11=s11+s1[i,j];s22<-s22+s2[i,j]}
sqrt(s11/s22)}

```

```

F<-function(D){
dd<-D
n<-dim(as.matrix(dd))[1]

P=which(cmdscale(D,k =(n-1),eig=TRUE)$eig<0)[1]-1

```

```

XX<-array(0,c(n,P,P))
sigma<-CMDS.STRESS<-BMDS.STRESS<-rep(0,P)
for( i in 1:P){
X1<-cmdscale(D, k = i, eig = TRUE)$points

```

```

CMDS.STRESS[i]<-sqrt((sum((D-dist(X1))^2)/sum(dist(X1)^2)))
p1<-dim(X1)[2];n1<-dim(X1)[1]
for( r in 1:10){
Cx<-scale(X1)
Vax<-eigen(Cx%*%t(Cx))$values[1:p1]
Vex<-eigen(Cx%*%t(Cx))$vectors[,1:p1]
if(i==1) Xa<-as.matrix(Vex)*(Vax^.5) else Xa<-Vex%*%(diag(Vax)^.5)
SSR1<-sum((D-dist(Xa))^2)
m<-n1*(n1-1)/2;sigma[i]<-SSR1/m
if(i==1) landa0<-Vax else landa0<-diag(Vax)
x2<-mvrnorm(n1,rep(0,i),landa0)
X1<-x2}
for(ii in 1:n1) for(jj in 1:i) XX[ii,jj,i]<-X1[ii,jj]
BMDS.STRESS[i]=STRESS(X1,sigma[i])}

RR<-SS<-array(0,c(1,n1,n1))
for(ii in 1:P) { SS[,1:p1,ii]<-apply(XX[, ,ii]^2,2,sum)[1:p1]
if(ii>1) RR[, ,ii]<-SS[, ,ii]/SS[, ,ii-1]}
LRT<-Penalty<-MDSIC<-rep(0,(P-1))
for( i in 1:(P-1)){
LRT[i]<-(m-2)*log(SSR(XX[,1:(i+1),(i+1)],sigma[i+1])/SSR(XX[,1:i,i],sigma[i]))
DF<-0;for(j in 1:i) DF<-DF+log((RR[1,j,i+1]*(n1+1))/(n1+RR[1,j,i+1]))
Penalty[i]<-DF
if(i==1) MDSIC[1]=(m-2)*log(SSR(XX[,1,1],sigma[1]))
else MDSIC[i]=sum(LRT[1:(i-1)])+ sum(Penalty[1:(i-1)])+(m-2)*log(SSR(XX[,1,1],sigma[i]))
Dim<-1:(P-1);CMDS.STRESS<-CMDS.STRESS[1:(P-1)];BMDS.STRESS<-BMDS.STRESS[1:(P-1)]
T=cbind(Dim,CMDS.STRESS,BMDS.STRESS,LRT,Penalty,MDSIC)
colnames(T) <- c("Dim","CMDS STRESS","BMDS STRESS","LRT","Penalty","MDSIC")
list(X=XX,Table=T)}

X4<-rnorm(50,0,1)
for( i in 1:10) X4=cbind(X4,rnorm(50,0,1))
D4=dist(X4)
F(D4)$Table[1:14,]

```


مراجع

- [۱] فرهادی ز، (۱۳۹۳). ”خوشه‌بندی توسط روش رده‌بندی جنگل‌های تصادفی“، پایان‌نامه کارشناسی ارشد، دانشگاه شاهرود.
- [۲] علی احمدی ع. ر، قاضی‌نوری س. (۱۳۸۴). ”کاربرد روش آماری مقیاس‌گذاری چند بعدی در دسته‌بندی اقتصادی و فرهنگی کشورهای جهان“، مقاله، دانشگاه الزهراء.
- [۳] کیهانی ح، (۱۳۹۴). ”مدل‌بندی پاسخ‌های نرخ و نسبت با رگرسیون مستطیلی بتا“، پایان‌نامه کارشناسی ارشد، دانشگاه شاهرود.
- [۴] نظام دوست س، (۱۳۹۰)، ”شبیه‌سازی زنجیره‌های بیولوژیکی نادر با استفاده از MCMCMC“، پایان‌نامه کارشناسی ارشد، دانشگاه مازندران.
- [۵] Anderson, T. W. (۲۰۰۳) An Introduction to Multivariate Statistical Analysis. Third edition, Wiley Series in Probability and Statistics. Wiley – Interscience [John Wiley and Sons], Hoboken, N. J. .
- [۶] Borg I. and Groenen P. (۱۹۹۷) Modern Multidimensional Scaling. New York: Springer-Verlag.
- [۷] B. Battenfield R.F. Reitsma (۲۰۰۲) Loglinear and multidimensional scaling models of digital library navigation, Internation Journal of Human-Computer Studies, ۵۷, ۱۰۱ – ۱۱۹.
- [۸] Carroll J.D. and Arabie P. (۱۹۸۰) Multidimensional scaling, Annu. Rev. Psychol, ۳۱, ۶۰۷ – ۶۴۹.
- [۹] Condon E. Golden B. Lele S. Raghavan S. Wasil E. (۲۰۰۲) A visualization model based on adjacency data, Decision Support Systems, ۳۳ ۳۴۹ – ۳۶۲.
- [۱۰] Coombs C.H. (۱۹۶۴) A Theory of Data, New York: Wiley.
- [۱۱] Corbet G. B. Cummins J. Hedges S. R. and Krzanowski W. J. (۱۹۷۰) The taxonomic structure of British water voles, genus Arvicola, Journal of Zoology, ۶۱, ۳۰۱ – ۳۱۶.
- [۱۲] Courrieu P. (۲۰۰۲) Straight monotonic embedding of data sets in Euclidean spaces, Neural Networks, ۱۵, ۱۱۸۵ – ۱۱۹۶.

- [١٣] Cox T. F. and Cox M. A. A. (٢٠٠١) Multidimensional Scaling, London: Chapman & Hall.
- [١٤] Davison M. L. (١٩٨٣) Multidimensional Scaling, New York: Wiley.
- [١٥] Elvevag B. Storms G. (٢٠٠٢) Scaling and clustering in the study of semantic disruptions in patients with schizophrenia: a re-evaluation, Schizophrenia Research.
- [١٦] Gelman A. Roberts G. O. and Gilks W. R. (١٩٩٦) Efficient Metropolis Jumping Rules, Bayesian Statistics, ٥, ٥٩٩ – ٦٠٨.
- [١٧] Gilks W. R. Richardson S. and Spiegelhalter D. J. (١٩٩٦) Markov Chain Monte Carlo in Practice, London: Chapman & Hall.
- [١٨] Gottardo A.E. Raftery (٢٠٠٨) Markov chain Monte Carlo with mixtures of mutually singular distributions, Journal of Computational and Graphical Statistics ١٧, ٩٤٩ – ٩٧٥.
- [١٩] Gower J.C. (١٩٨٤) Multivariate analysis: ordination, multidimensional scaling and allied topics. In Lloyd E.H. (ed.) Handbook of Applicable Mathematics, New York: Wiley.
- [٢٠] Greenacre M.J. and Underhill L.G. (١٩٨٢) Scaling a data matrix in a low dimensional Euclidean space. In Hawkins, D.M. (ed.), Topics in Applied Multivariate Analysis, Cambridge: Cambridge University Press, ١٨٣ – ٢٤٨.
- [٢١] Groenen P. J. F. (١٩٩٣) The Majorization Approach to Multidimensional Scaling: Some Problems and Extensions, Lieden, The Netherlands: DSWO.
- [٢٢] Groenen P. J. F. Mathar R. and Heisser W. J. (١٩٩٥) The Majorization Approach to Multidimensional Scaling for Minkowski Distances, Journal of Classification, ١٢, ٣ – ١٩.
- [٢٣] Gupta A. K. and Nagar D. K. (٢٠٠٠) Matrix Variate Distributions Chapman & Hull/ CRC, Boca Raton.
- [٢٤] Hartigan J. A. (١٩٧٥) Clustering Algorithms, New York: Wiley
- [٢٥] Hastings W. K. (١٩٧٠) Monte Carlo Sampling Methods Using Markov Chains and Their Applications, Biometrika, ٥٧, ٩٧ – ١٠٩.
- [٢٦] Kruskal J. B. (١٩٦٤) Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis, Psychometrika, ٢٩, ١ – ٢٨.
- [٢٧] Kruskal J.B. and Wish M. (١٩٧٨) Multidimensional Scaling, Beverly Hills, CA: Sage Publications.

- [٢٨] Mead A. (١٩٩٢) Review of the development of multidimensional scaling methods, *The Statistician*, ٤١, ٢٧ – ٣٩.
- [٢٩] Oh M. S. Raftery A. E. ,(٢٠٠١) Bayesian multidimensional scaling and choice of dimension, *Journal of the American Statistical Association*, ٩٦, ١٠٣١ – ١٠٤٤.
- [٣٠] Oh M. S. ,(٢٠١٢) A simple and efficient Bayesian procedure for selecting dimensionality in multidimensional scaling, *Journal of Multivariate Analysis*, ١٠٧, ٢٠٠ – ٢٠٩.
- [٣١] Okada K. Shigemasu K. ,(٢٠٠٩) BMDS: a collection of R functions for Bayesian multidimensional scaling, *Applied Psychological Measurement* ٣٣, ٥٧٠ – ٥٧١.
- [٣٢] Priem R. L. Love L. Shaffer M. A. ,(٢٠٠٢) Executives' perceptions of uncertainty sources: a numerical taxonomy and underlying dimensions, *Journal of Management* ٢٨, ٧٢٥ – ٧٤٤.
- [٣٣] Raftery A. E. ,(١٩٩٩) Bayes Factors and BIC—Comment on A Critique of the Bayesian Information Criterion for Model Selection, *Sociological Methods and Research*, ٢٧, ٤١١ – ٤٢٧.
- [٣٤] Ramsay J.O. ,(١٩٨٢) Some statistical approaches to multidimensional scaling data, *J. R. Stat. Soc. A*, ١٤٥, ٢٨٥ – ٣١٢.
- [٣٥] Ren S. Frymier P. D. ,(٢٠٠٣) Use of multidimensional scaling in the selection of wastewater toxicity test battery components, *Water Research*, ٣٧, ١٤٥٥ – ١٤٦١.
- [٣٦] Schiffman S.S Reynolds M.L. and Young F.W. ,(١٩٨١) Introduction to Multidimensional Scaling: Theory Methods and Applications, New York: Academic Press.
- [٣٧] Schoenberg I.J. ,(١٩٣٥) Remarks to Maurice Frechet's article Sur la definition axiomatique d'une classe d'espaces vectoriels distances applicables vectoriellement sur l'espace de Hilbert, *Ann. Math.*, ٣٦, ٧٢٤ – ٧٣٢.
- [٣٨] Schutze H. and Silverstein C. ,(١٩٩٧) Projections for Efficient Document Clustering, *ACM SIGIR*, ٩٧, ٧٤ – ٨١.
- [٣٩] Shepard R.N. ,(١٩٦٢) The analysis of proximities: multidimensional scaling with an unknown distance function I, *Psychometrika*, ٢٧, ١٢٥ – ١٤٠.

- [٤٠] Spence I. and Graef J. C. ,(١٩٧٤) The Determination of the Underlying Dimensionality of an Empirically Obtained Matrix of Proximities, *Multivariate Behavioral Research*, ٩, ٣٣١ – ٣٤١.
- [٤١] Suppes P. and Zinnes J. L. ,(١٩٦٣) Basic Measurement Theory, in *Handbook of Mathematical Psychology Vol. I* eds. R. D. Luce R. R. Bush and E. Galanter, New York: Wiley, pp. ٢ – ٧٦.
- [٤٢] Takane Y. and Carroll J. D. ,(١٩٨١) Nonmetric Maximum Likelihood Multidimensional Scaling from Directional Rankings of Similarities, *Behaviormetrika*, ١١, ٣٧ – ٤٨.
- [٤٣] Tibshirani R. Lazzeroni L. Hastie T. Olshen A. and Cox D. ,(١٩٩٩) The Global Pairwise Approach to Radiation Hybrid Mapping, Technical Report, Department of Statistics, Stanford University.
- [٤٤] Torgerson W. S. ,(١٩٥٢) Multidimensional Scaling: I. Theory and Method, *Psychometrika*, ١٧, ٤٠١ – ٤١٩.
- [٤٥] Torgerson W. S. ,(١٩٥٨) *Theory and Methods of Scaling*, New York: Wiley.
- [٤٦] Welchew D. E. Honey G. D. Sharma T. Robins T. W. Bullmore E. T. ,(٢٠٠٢) Multidimensional scaling of integrated neurocognitive function and schizophrenia as a disconnection disorder, *NeuroImage* , ١٧, ١٢٢٧ – ١٢٣٩.
- [٤٧] Young G. and Householder A. S. ,(١٩٣٨) Discussion of a set of points in terms of their mutual distances, *Psychometrika*, ٣, ١٩ – ٢٢.
- [٤٨] Young F. W. ,(١٩٨٧) *Multidimensional Scaling—History, Theory, and Applications*, Hauer R. M. (editor), Hillsdale Erlbaum Association, Hillsdale NJ: Lawrence Erlbaum.

واژه‌نامه فارسی به انگلیسی

| | |
|----------------------------------|-------------------------|
| Objects | اشیا |
| Dimension | بعد |
| Configuration | پیکره |
| Likelihood function | تابع درست‌نمایی |
| Spectral decomposition | تجزیه طیفی |
| Beta distribution | توزیع بتا |
| Posterior distribution | توزیع پسین |
| Prior distribution | توزیع پیشین |
| Gamma distribution | توزیع گاما |
| Inverse Gamma distribution | توزیع گامای معکوس |
| normal distribution | توزیع نرمال |
| multivariate normal distribution | توزیع نرمال چند متغییره |
| Approximation | تقریب |
| Rank | رتبه |
| Similarity | شباهت |
| Dissimilarity | عدم شباهت |
| Distance matrix | ماتریس فاصله |
| Diagonal matrix | ماتریس قطری |
| Symmetric matrix | ماتریس متقارن |
| Orthogonal matrix | ماتریس متعامد |
| Positive definite | ماتریس معین مثبت |
| Variance-covariance matrix | ماتریس واریانس-کواریانس |
| Metric | متری |
| Proximity | مجاورت |
| Stimuli | محرک |

| | |
|---|---|
| Coordinate | مختصات |
| Eigenvalues | مقادیر ویژه |
| multidimensional scaling | مقیاس‌گذاری چند بعدی |
| Besyan multidimensional scaling | مقیاس‌گذاری چند بعدی بیزی |
| Unidimensional scaling | مقیاس‌گذاری تک بعدی |
| Classical multidimensional scaling | مقیاس‌گذاری چند بعدی کلاسیک |
| Metric least squares multidimensional scaling | مقیاس‌گذاری چند بعدی متری کمترین توان‌های دوم |
| Critchley intermediate multidimensional scaling | مقیاس‌گذاری چند بعدی متوسط کریتچلی |
| Nonmetric multidimensional scaling | مقیاس‌گذاری چند بعدی نامتری |
| Criterion | معیار |

Abstract

If we know the distances between municipality and offices within a city, we can draw the locations on a 2D map using multi-dimensional scaling. Approximately distance matrix can approach to observation matrix. This approach has been traditionally used in a host of different applications. Not only can it be used for finding the location of cities but also for reducing variants dimensions. This approach works based on the observed data, a distance matrix is formed from which the designated configuration can be achieved in a lower dimension. multidimensional scaling include several methods from analyzing multi variable data that deals with observable Euclidian space. This method provides data on distances between variables that could be either metric or nonmetric. Such distances should not necessarily be Euclidian. Bayesian multidimensional scaling is one of the most recent of multidimensional scaling that At this thesis the method which retrieves the configuration in a lower dimension from a Bayesian perspective. In this thesis examinedis Bayesian multidimensional scaling method with prior distribution normal and mixed ,and in this regard, a simulation studey compares the superiority proposed method over classical method .

keywords: multidimensional scaling, Bayesian multidimensional scaling, mixed normal, prior distribution, Euclidean distance, dissimilarity.



University of Shahrood

Faculty Of Mathematical Sciences

**Bayesian multidimensional scaling in
multivariate normal distribution**

Seyede Basira Hosseini

Supervisor

Dr.M.Arashi

September 2015