دانشگاه تربیت مدرس

دانشکده **علوم ریاضی**

رساله دوره دکترای آمار

پیشگویی فضایی برای میدان تصادفی چوله گاوسی بسته

توسط

امید کریمی

استاد راهنما

دکتر محسن محمدزاده

اسفند ۱۳۸۸

قدردانی

منت خدای را عزوجل که طاعتش موجب قربت است و به شکر اندرش مزید نعمت، حمد و ستایش خدای را که همه از اوییم و هر چه هست از اوست. برخود لازم می‌دانم از استاد گرامیم جناب آقای دکتر محسن محمدزاده که در طی سالها، همواره راهنمای اینجانب بوده‌اند و از ابتدا تا انتهای این پایان‌نامه مرا صبورانه و صمیمانه یاری نمودند و همواره با راهنمایی‌ها و فکرهای ارزشمند خود راهگشای من در مشکلات بوده‌اند، قدردانی و سپاسگزاری نمایم. از اساتید محترم جناب آقای دکتر خالدی و جناب آقای دکتر گل علیزاده که با راهنمایی‌های ارزنده‌شان نقش موثری در بهبود این رساله داشته‌اند، کمال تشکر را دارم. همینطور از اساتید محترم جناب آقای دکتر پاشا که افتخار شاگردی ایشان را داشته‌ام و جناب آقای دکتر پزشک به سبب حضور در جمع داوران سپاسگزاری می‌نمایم و امیدوارم همواره در پناه الطاف الهی محفوظ و موفق باشند.

از تمام دوستانی که در این مرحله از زندگی مرا به هر شکل ممکن یاری نموده‌اند، مخصوصاً همسر گرامیم خانم فاطمه حسینی که در تدوین و نگارش این پایان‌نامه وقت خود را در اختیار من گذاشته و مرا مورد لطف خود قرار داده از صمیم قلب تشکر می‌کنم.

خدایا چنان کن سرانجام کار　　　تو خشنود باشی و ما رستگار

امید کریمی

تهران – اسفند ۱۳۸۸

تقدیم به

صفا بخش تمامی لحظه‌های خوشبختی‌ام

فاطمه

و

پسرم آرمان

چکیده

در اغلب تحلیل‌های آمار فضایی فرض بر این است که داده‌ها تحققی از یک میدان تصادفی گاوسی هستند، اما مشخصه‌های ناگاوسی مانند متغیرهای تصادفی نامنفی با توزیع چوله در اکثر زمینه‌های علمی دیده می‌شوند. مدل‌بندی این نوع داده‌ها با استفاده از میدان تصادفی چوله گاوسی، که براساس توزیع چوله نرمال چند متغیره تعریف شده و از انعطاف‌پذیری بیشتری برخوردار است، صورت می‌پذیرد. در این رساله خانواده توزیع‌های چوله نرمال بسته که نسبت به ترکیبات خطی و توزیع‌های شرطی بسته است، برای تحلیل داده‌های فضایی چوله مورد مطالعه قرار گرفته است. سپس پیشگویی فضایی بیزی برای میدان تصادفی چوله گاوسی بسته ارائه شده و مطالعات شبیه‌سازی برای بررسی مناسب بودن مدل با بکارگیری معیار میانگین توان دوم خطاها صورت پذیرفته است. بعلاوه مدل چوله گاوسی بسته روی داده‌های واتنش و حداقل دمای هوا با استفاده از معیار اعتبارسنجی متقابل مورد ارزیابی قرار گرفته و کاهش میانگین توان دوم خطاهای معیار اعتبار سنجی متقابل برای مدل چوله گاوسی بسته نشان داده شده است. سپس توزیع چوله نرمال بسته در دو مدل «رگرسیون فضایی با خطاهای خودهمبسته همراه با مشاهدات گمشدهٔ متغیر پاسخ» و «معکوس ارتعاشی» بکار گرفته شده و پارامترهای آنها با رهیافت بیزی و استفاده از الگوریتم‌های زنجیر مارکوف مونت کارلویی برآورد شده‌اند. آنگاه مدل رگرسیون پیشنهادی را برای تحلیل داده‌های آلودگی هوای تهران، که معمولاً چوله هستند، بکار برده و دقت پیشگویی با مدل‌های گاوسی و چوله گاوسی بسته توسط معیار اعتبار سنجی متقابل ارزیابی شده است. در ادامه مدل معکوس گاوسی بیزی را به «مدل معکوس چوله گاوسی بسته بیزی» تعمیم داده و نحوه مدل بندی چولگی متغیرهای مورد مطالعه برای تحلیل داده‌های ارتعاشی نشان داده شده است. سپس داده‌های

خاصیت‌های مواد الاستیکی با این دو مدل تحلیل شده و دقت پیشگویی مولفه‌های خاصیت‌های مواد الاستیکی توسط هر دو مدل از طریق معیار میانگین توان دوم خطاهای پیشگویی مورد ارزیابی قرار گرفته است.

واژه‌های کلیدی : توزیع چوله نرمال بسته، پیشگویی فضایی بیزی، داده‌های ارتعاشی، مدل معکوس چوله گاوسی بسته بیزی.

فهرست مندرجات

لیست اشکال

فصل ۱

مفاهیم مقدماتی

۱.۱ مقدمه

پیشگویی فضایی یک مساله مهم در علوم محیطی مانند هواشناسی، زمین شناسی، جغرافیا، کشاورزی و غیره است. در تئوری‌های پیشگویی فضایی اغلب فرض بر این است که داده‌ها تحققی از یک میدان تصادفی گاوسی هستند. اما در عمل اینگونه نیست، مشخصه‌های ناگاوسی مانند متغیرهای تصادفی نامنفی با توزیع چوله در بعضی زمینه‌های علمی دیده می‌شوند.

وقتی داده‌های فضایی ناگاوسی هستند اما تبدیلی از آنها گاوسی باشد، اولیویرا و همکاران (۱۹۹۷)، اولیویرا و ایکر (۲۰۰۲)، اولیویرا (۲۰۰۳) و محمدزاده و خالدی (۱۳۸۳) کریگیدن گاوسی تبدیل یافته را برای پیشگویی فضایی مورد مطالعه قرار دادند. اما در عمل تبدیل نرمال ساز داده‌ها نامعلوم است و نه تنها تعیین آن تحلیل داده‌ها را با مشکلاتی مواجه می‌سازد بلکه گاهی تفسیر داده‌های تبدیل یافته نسبت به داده‌های اصلی از دشواری بیشتری برخوردار است (آزالینی و کاپیتانیو، ۱۹۹۹ و کیم و مالیک، ۲۰۰۴).

وقتی توزیع مجموعه‌ای از داده‌ها واجد اکثر خواص توزیع نرمال باشد اما متقارن نباشند، بعبارت دیگر دارای چولگی باشند، توزیع چوله نرمال[1](SN) می‌تواند برای مدل بندی اینگونه داده‌ها مورد استفاده قرار گیرد . توزیع چوله نرمال چند متغیره توسط آزالینی و دالاواله (۱۹۹۶) معرفی گردید. آزالینی و کاپیتانیو (۱۹۹۹) خصوصیات این توزیع را در مسائل کاربردی بیان نمودند. سپس گوپتا و همکاران (۲۰۰۴) این توزیع را به حالت کلی‌تری بسط دادند. کلاس جدیدی از توزیع‌ها تحت عنوان چوله نرمال بسته[2](CSN) توسط دامینگوس وهمکاران (۲۰۰۳) معرفی شد که اکثر توزیع‌های چوله نرمال معرفی شده را در بر می‌گیرد و خصوصیات این توزیع همانند بسته بودن تحت تبدیلات خطی، حاشیه‌ای و شرطی کردن بصورت جامع توسط گنزالس و همکاران (۲۰۰۴) ارائه شده است. چون در آمار فضایی گاهی با مواردی مواجه می‌شویم که داده‌ها نامتقارن و چوله هستند، لذا در این رساله با استفاده از یک میدان تصادفی چوله گاوسی بسته[3](CSG) به تحلیل داده‌های فضایی چوله پرداخته می‌شود. برای اولین بار کیم و مالیک (۲۰۰۲) تحلیل داده‌های فضایی را برای یک میدان تصادفی چوله گاوسی[4](SG) مورد بررسی قرار دادند، همچنین کیم و مالیک (۲۰۰۴) پیشگویی فضایی بیزی را در یک مثال کاربردی برای این میدان تصادفی بکار گرفتند و پس از آنها آلارد و ناویو (۲۰۰۵) نحوه شبیه‌سازی یک میدان تصادفی CSG را برای داده‌های فضایی ارائه کردند و آلارد و ناویو (۲۰۰۷) اولین و دومین گشتاور توزیع CSN را برای برآورد پارامترهای مدل به روش گشتاوری محاسبه نمودند. آنها همچنین نشان دادند که برآورد تغییرنگار بر اساس داده‌های چوله گاوسی شبیه‌سازی شده نااریب است. با توجه به اینکه توزیع CSN از توزیع چوله نرمال کلی‌تر و

---

[1] Skew Normal

[2] Closed Skew Normal

[3] Closed Skew Gaussian

[4] Skew Gaussian

دارای خواص بسته بودن تحت تبدیلات خطی و شرطی کردن است، مطالعه یک میدان تصادفی CSG می‌تواند شرایط ساده‌تری برای پیشگویی فضایی دقیق‌تر فراهم نماید.
در این رساله، مفاهیم اولیه توزیع چوله نرمال بسته و خواص آن در فصل ۱ ارائه می‌گردد. در فصل ۲ پیشگویی بیزی برای میدان تصادفی CSG معرفی و نحوه کاربست آن در دو مثال کاربردی نشان داده می‌شود. در فصل ۳ تحلیل بیزی مدل رگرسیون فضایی با خطاهای خود همبسته ارائه می‌گردد. همچنین نحوه مدل‌بندی و پیشگویی فضایی بیزی مشاهدات گمشده برای این مدل رگرسیونی نیز بیان می‌شود. در فصل ۴ مدل معکوس گاوسی بیزی به مدل معکوس چوله گاوسی بسته بیزی تعمیم داده شده و ضمن مقایسه دقت این دو مدل، نحوه بکارگیری مدل پیشنهادی در تحلیل داده‌های ارتعاشی نشان داده شده است.

۲.۱ توزیع چوله نرمال بسته

در این بخش فرم‌های مختلف توزیع چوله نرمال بسته که برای اولین بار توسط دامینگوس وهمکاران (۲۰۰۳) ارئه شد، بیان می‌گردد و خواص مهم آن از جمله بسته بودن تحت تبدیلات خطی و شرطی کردن نیز مورد بررسی قرار می‌گیرد. دلیل اصلی نامگذاری این توزیع به چوله نرمال «بسته» اینست که این توزیع نسبت به تبدیلات خطی، حاشیه‌ای و شرطی کردن بسته است. یعنی هر تبدیل خطی روی مولفه‌های این خانواده نیز عضوی از این خانواده است. بیشتر موضوعات این بخش برگرفته از مقالات دامینگوس وهمکاران (۲۰۰۳)، گنزالس و همکاران (۲۰۰۴)، دامینگوس و همکاران (۲۰۰۷) و کریمی و محمدزاده (۲۰۰۹) است.

توزیع چوله نرمال برای اولین بار توسط روبرتس (۱۹۶۶) بدست آمد، اما اولین فرم رسمی

آن توسط آزالینی (۱۹۸۵، ۱۹۸۶) معرفی شد. حالت چند متغیره این توزیع توسط آزالینی و دالاواله (۱۹۹۶) با یک تابع چگالی به فرم

$$f(\boldsymbol{y};\boldsymbol{\mu},\Sigma,\lambda)=2\phi_p(\boldsymbol{y};\boldsymbol{\mu},\Sigma)\Phi(\boldsymbol{\lambda}'\Sigma^{-\frac{1}{2}}(\boldsymbol{y}-\boldsymbol{\mu}))$$

بیان گردید، که در آن $\phi_p(\cdot;\boldsymbol{\mu},\Sigma)$ چگالی نرمال p-متغیره با میانگین $\boldsymbol{\mu}$ و ماتریس واریانس کوواریانس Σ، $\Phi(\cdot)$ تابع توزیع تجمعی نرمال استاندارد و $\boldsymbol{\lambda}$ یک بردار p-بعدی از پارامترهای چولگی است. توزیع چوله نرمال بعنوان تعمیمی از توزیع نرمال برای مدل بندی چولگی ارائه شده است. در واقع این خانواده از توزیع‌ها علاوه بر اینکه شامل خانواده توزیع‌های نرمال هستند و در بعضی از خواص این توزیع‌ها مشترک هستند، دارای یک پارامتر چولگی می‌باشند که میزان چولگی توزیع را کنترل می‌کند. در عمل گاهی با داده‌هایی مواجه هستیم که توزیع آنها متقارن نیستند، مانند داده‌های مالی (کوزوبوسکی، ۱۹۹۹)، داده‌های بارندگی (کیم و مالیک، ۲۰۰۴) یا داده‌های ارتعاشی[۵] (کریمی و همکاران، ۲۰۰۹). بعضی از خواص چگالی‌های نامتقارن[۶] و چوله در آزالینی و کاپیتانیو (۱۹۹۹) ارائه شده است. لیزو و لوپرفیدو (۲۰۰۳) چگالی چوله نرمال چند متغیره را بعنوان تعمیمی از حالت چند متغیره اوهاگان و لئونارد (۱۹۷۶) بدست آوردند.

توزیع چوله نرمال بسته حالت کلی‌تری از توزیع چوله نرمال چند متغیره و تعمیمی از چگالی تعریف شده توسط گوپتا و همکاران (۲۰۰۴) است. این توزیع تحت تبدیلات خطی و عملگر شرطی کردن بسته است، و معمولا این خصوصیات‌ها در خیلی از موارد استنباطهای آماری از جمله پیشگویی مورد استناد قرار می‌گیرد. تعریف رسمی توزیع چوله نرمال بسته بصورت زیر است.

تعریف ۱.۱ (دامینگوس و همکاران، ۲۰۰۳) فرض کنید $p\geq 1$، $q\geq 1$، $\boldsymbol{\mu}\in I\!R^p$، $\boldsymbol{\nu}\in I\!R^q$، Γ

---

[۵] Seismic

[۶] Asymmetric

یک ماتریس دلخواه $q \times p$، Σ و Δ ماتریس‌های معین مثبت به ترتیب با بعدهای $p \times p$ و $q \times q$ باشند. آنگاه چگالی توزیع CSN بصورت

$$f_{p,q}(\boldsymbol{y};\boldsymbol{\mu},\Sigma,\Gamma,\boldsymbol{\nu},\Delta) = K\phi_p(\boldsymbol{y};\boldsymbol{\mu},\Sigma)\Phi_q[\Gamma(\boldsymbol{y}-\boldsymbol{\mu});\boldsymbol{\nu},\Delta], \quad (۱.۲.۱)$$

تعریف می‌شود، که در آن

$$K^{-۱} = \Phi_q(\mathbf{o};\boldsymbol{\nu},\Delta+\Gamma\Sigma\Gamma'), \quad (۲.۲.۱)$$

$\Phi_q(\cdot;\boldsymbol{\eta},\Psi)$ تابع توزیع تجمعی q-بعدی نرمال با میانگین $\boldsymbol{\eta}$ و ماتریس واریانس کوواریانس Ψ و $\mathbf{o}$ یک بردار q-بعدی با مولفه‌های صفر است.

متغیر تصادفی $\boldsymbol{y}$ دارای توزیع CSN با پارامترهای $\boldsymbol{\mu}$، Σ، Γ، $\boldsymbol{\nu}$ و Δ بصورت $\boldsymbol{y} \sim CSN_{p,q}(\boldsymbol{\mu},\Sigma,\Gamma,\boldsymbol{\nu},\Delta)$ نمایش داده می‌شود. در ادامه لِم ارائه شده توسط الیسون (۱۹۶۴) به حالت چند متغیره تعمیم داده می‌شود که با استفاده از آن می‌توان نشان داد (۱.۲.۱) یک تابع چگالی است.

لم ۱.۱ اگر $\boldsymbol{T}$ یک بردار تصادفی p-بعدی با توزیع $N_p(\boldsymbol{\eta},\Lambda)$ باشد. آنگاه برای هر $q \geq ۱$

$$E_T[\Phi_q(\boldsymbol{a}+B\boldsymbol{T};\boldsymbol{\gamma},\Psi)] = \Phi_q(\boldsymbol{a}+B\boldsymbol{\eta};\boldsymbol{\gamma},\Psi+B\Lambda B'),$$

است، که در آن $\boldsymbol{a}$ و $\boldsymbol{\gamma}$ بردارهای q-بعدی، B ماتریس دلخواه $q \times p$ بعدی و Ψ ماتریس معین مثبت $q \times q$ بعدی است.

برهان: چون $\boldsymbol{T} \sim N_p(\boldsymbol{\eta},\Lambda)$ بنابراین

$$E_T[\Phi_q(\boldsymbol{a}+B\boldsymbol{T};\boldsymbol{\gamma},\Psi)] \;=\; \int_{-\infty}^{+\infty} \phi_p(\boldsymbol{t};\boldsymbol{\eta},\Lambda)\Phi_q(\boldsymbol{a}+B\boldsymbol{t};\boldsymbol{\gamma},\Psi)d\boldsymbol{t},$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{a+Bt}\phi_p(\boldsymbol{t};\boldsymbol{\eta},\Lambda)\phi_q(\boldsymbol{s};\boldsymbol{\gamma},\Psi)d\boldsymbol{t}d\boldsymbol{s},$$

که در آن $\boldsymbol{S}\sim N_q(\boldsymbol{\gamma},\Psi)$ و مستقل از $\boldsymbol{T}$ است، آن‌گاه

$$\begin{aligned} E_T[\Phi_q(\boldsymbol{a}+B\boldsymbol{T};\boldsymbol{\gamma},\Psi)] &= P_{S,T}(\boldsymbol{S}\leq \boldsymbol{a}+B\boldsymbol{T}),\\ &= P_{S,T}(\boldsymbol{S}-B\boldsymbol{T}\leq \boldsymbol{a}),\end{aligned}$$

چون $\boldsymbol{S}-B\boldsymbol{T}$ دارای توزیع $N_q(\boldsymbol{\gamma}-B\boldsymbol{\eta},\Psi+B\Lambda B')$ است، می‌توان نوشت:

$$\begin{aligned} E_T[\Phi_q(\boldsymbol{a}+B\boldsymbol{T};\boldsymbol{\gamma},\Psi)] &= \Phi_q(\boldsymbol{a};\boldsymbol{\gamma}-B\boldsymbol{\eta},\Psi+B\Lambda B'),\\ &= \Phi_q(\boldsymbol{a}+B\boldsymbol{\eta};\boldsymbol{\gamma},\Psi+B\Lambda B'). \quad \blacksquare\end{aligned}$$

اکنون بنابر لِم ۱.۱ داریم

$$\begin{aligned} \int f_{p,q}(\boldsymbol{y};\boldsymbol{\mu},\Gamma,\boldsymbol{\nu},\Delta)dy &= K\int_{\mathbb{R}^p}\phi_p(\boldsymbol{y};\boldsymbol{\mu},\Sigma)\Phi_q(\Gamma(\boldsymbol{y}-\boldsymbol{\mu});\boldsymbol{\nu},\Delta)dy,\\ &= KE_T[\Phi_q(\Gamma(\boldsymbol{T}-\boldsymbol{\mu});\boldsymbol{\nu},\Delta)], \qquad (\boldsymbol{T}\sim N_p(\boldsymbol{\mu},\Sigma))\\ &= K\Phi_q(\Gamma(\boldsymbol{\mu}-\boldsymbol{\mu});\boldsymbol{\nu},\Delta+\Gamma\Sigma\Gamma'),\\ &= K\Phi_q(\mathbf{o};\boldsymbol{\nu},\Delta+\Gamma\Sigma\Gamma'),\\ &= ۱\end{aligned}$$

تابع توزیع CSN

چنانچه $\boldsymbol{Y}\sim CSN_{p,q}(\boldsymbol{\mu},\Sigma,\Gamma,\boldsymbol{\nu},\Delta)$، آن‌گاه

$$F_{p,q}(\boldsymbol{y};\boldsymbol{\mu},\Sigma,\Gamma,\boldsymbol{\nu},\Delta) = P(\boldsymbol{Y}\leq \boldsymbol{y}),$$