دانشگاه پیام نور

واحد تهران مرکز

دانشکده : علوم پایه وکشاورزی

پایان نامه کارشناسی ارشد

عنوان:

برآورد بیزی پارامترهای مدل ARFIMA

استاد راهنما:دکتر مسعود یار محمدی

استاد مشاور:دکتر پرویز نصیری

محقق:

راضیه قاسمی

زمستان 1390

تقدیم به

حامیان بی ادعای زندگی پدر و مادر عزیزم و تمام کسانی که به من علم آموختند.

سپاس گذاری:

صمیمانه از جناب آقای دکتر مسعود یار محمدی استاد ارجمندم تقدیر و تشکر می کنم که در راستای این پایان نامه از هیچ کمکی دریغ نفرمودند و در تمام مراحل با راهنمایی های ارزشمند و دلسوزانه شان مرا حمایت نمودند.

همچنین از جناب آقای دکتر پرویز نصیری که با کمک ها و راهنمایی هایشان مرا در انجام این پروژه یاری کردند کمال تشکر و قدردانی را به جای می آورم.

همچنین از تمام دوستانی که مرا کمک کردند تشکر می کنم .

چکیده:

سریهای زمانی با حافظه بلند در علوم مختلف کاربردهای فراوانی دارند.در اینگونه سریهای زمانی ،تابع خود همبستگی دوامی نشان میدهد که نه با فرآیندهای ARIMA(p,1,q) و نه با ARIMA(p,0,q) سازگار است.به عبارت دیگر ضرایب خود همبستگی ،مانایی سری را تایید نکرده و پس از یکبار تفاضل گیری هم به نظر می رسد که بیش تفاضل گیری شده باشند.سریهای زمانی ARFIMA (در حالی که فرآیندهای با حافظه بلند باشند)با تفاضل گیری کسری میتوان مانا کرد.به همین دلیل این سریها را اتورگرسیو میانگین متحرک تلفیق شده کسری[1] ARFIMA یا ARIMA با تفاضل گیری کسری می نامند. در این پایان نامه ضمن معرفی سریهای زمانی ARFIMA،مطالعات شبیه سازی را روی این مدلها انجام می دهیم .بر اساس مطالعاتی که قبلا صورت گرفته است مدل ARFIMA از قدرت پیش بینی کنندگی در مقایسه با مدلهای ARIMA برخوردار است.مثالهای متعددی از فرآیندهای با حافظه بلند در طبیعت وجود دارند..هاسکینگ(1984) ،علاوه بر تحقیق روی داده های نیل ،وجود حافظه بلند را در داده های میانگین درجه حرارت سالانه انگلستان از سال 1659 تا 1976 نیز تشخیص داده است.یکی از دلایل استفاده روز افزون از آمار بیزی در مقابل آمار کلاسیک ،بالا بودن دقت محاسبات در آمار بیزی است،و این به دلیل سهیم بودن اطلاعات و عقاید پیشین کاربرد در استنباط آماری می باشد.در این پایان نامه کاربردهای اسلوب شناسی بیزی را با تحلیل داده های سریهای زمانی بلند حافظه را بررسی میکنیم.روش زنجیر مارکوف مونت کارلو(MCMC) یک وسیله محاسباتی مهم برای به دست آوردن نمونه هایی از یک توزیع پسین میباشد.

[1]- Auto regressive Fractionally Integrated Moving Average

مقدمه:

سریهای زمانی در علوم مختلف کاربردهای بسیاری دارد.یکی از مدلهای سری زمانی سری زمانی با حافظه بلند میباشد که در بسیاری از علوم همچون اقلیم شناسی ،آب شناسی ،هوا شناسی و مخصوصا" در اقتصاد کاربرد فراوانی دارند.در دو دهه گذشته پیشرفتهای چشم گیری در زمینه اقتصاد سنجی مربوط به سریهای زمانی انجام شده که در تحلیل سریهای زمانی بلند مدت مستقلا" استفاده شده است.مدلهای ARFIMA ابتدا به وسیله گرنجر و جیوکس[2](1980) ارائه شده است. .توسط هسلت و رفتری[3] (1989) یکی از این مدلها برای داده های مربوط به نیروی باد پیشنهاد شده است.مدل های با حافظه بلند برای تحقیق درباره علت تغییرات آب و هوایی نیز مفید بوده اند.در اقتصاد مشاهده تجربی اثر حافظه بلند به گرنجر (1966) بر میگردد این مدلها بطور گسترده در زمینه های مختلفی همچون تحلیل پدیده ژئو فیزیک نواکز و دیگران[4](1988) و بلوم فیلد[5] (1992)،مدل بندی اقتصاد سنجی (رودبوش[6] (1989)و سوول[7] (1992))، تحلیل سریهای زمانی مالی (شی[8] (1991)و چونگ[9] (1993)) و پیش بینی بلند مدت (گوئک و پورتو–هداک [10] (1993)،ری[11] (1993) و ساتکلیف[12] (1993)) استفاده شده اند.برآورد پارامترهای مدلهای ARFIMA یک مسئله مهم میباشدیک سری زمانی با حافظه بلند را میتوان به وسیله تابع خود همبستگی آن که با نرخ هیپربولیک(هذلولی) کاهش می یابد ،مشخص کرد.نرخ کاهش هیپربولیک بسیار آهسته تر از نرخ کاهش تابع خود همبستگی سری زمانی ای که حافظه کوتاه مدت دارد است.برآورد پارامترهای مدل ARFIMA از ابتدا مورد توجه بوده است.روشهای ML دقیق بوسیلهٔ هاسلت و رافتری (1989)،هسکینگ(1984) ،لی (1986) و بیرن (1994) روش ML دقیق را ارائه کرده اند.چانگ و بایلی (1993) روش حداقل مربعات شرطی (CSS) را ارائه کردند.تحلیل بیزی مدلهای ARFIMA بوسیلهٔ کوپ ،لی،اسیوالسکی و استیل (1997) ،جفریز و راویشنکار (1996 و 1998) ارائه شد.توزیع پیشنهادی پارامتر تفاضل گیری کسری بوسیلهٔ نان-جانگ و بریدت(2003) ارائه شده است.

در فصل اول این پایان نامه مفاهیم اساسی در بخش اول آورده شده است ،در بخش دوم نامانایی در میانگین ، در بخش سوم به معرفی الگوی اتورگرسیو میانگین متحرک تلفیق شده می پردازد ، در بخش چهارم به معرفی

2- *Granger and Joyeux*
3- *Haslett & Raftery*
4- *Noakea et al*
5- *Bloomfield*
6- *Rudebush*
7- *Sowell*
8- *Shea*
9- *Cheung*
10- *Guek & Porter-Hudak*
11- *Ray*
12- *Sutcliffe*

الگوی اتورگرسیو میانگین متحرک تلفیق شده(ARIMA) ، در بخش پنجم قانون اعداد بزرگ ، در بخش ششم دریچه پارزن ، در بخش هفتم پیش بینی فرآیند های مانا و در بخش هشتم بعضی روشهای پیش بینی را مطرح می کنیم و همچنین الگوریتم داربین-لوینسن را معرفی میکنیم.

در فصل دوم به معرفی و تاریخچهٔ مدلهای اتورگرسیو میانگین متحرک تلفیق شده کسری[13] می پردازیم و تعاریف جایگزینی برای آن ارائه می دهیم ، فرآیندهای ARFIMA نامانارا معرفی می کنیم ، بخش2-9 به آزمونهای تشخیص وجود حافظه بلند می پردازد.در بخش 2-10 مانایی و علت پذیری و وارون پذیری را در فرآیندهای ARFIMA شرح می دهیم. همچنین بسط AR وMA نامتناهی و ویژگی های فرایند های ARFIMA (p, d, q) را بیان میکنیم.

در فصل سوم در بخش 3-3 روشهای زنجیرمارکوف مونت کارلو رابیان می کنیم والگوریتم متروپلیس-هستینگز و نمونه گیری گیبس و راه های تشخیص همگرایی آنها را بیان می کنیم ، در بخش 3-4 تعیین زمان توقف و دوره داغیدن و در بخش 3-7 همگرایی کنترل شده را بیان می کنیم.

در فصل چهارم روشهای برآورد برآورد پارامتر تفاضل گیری در ARFIMA مانا را بیان می کنیم.در بخش 4-3 برآورد درستنمایی ماکسیمم ، در بخش 4-4 تقریبهای اتورگرسیو ، در بخش 4-5 روش هاسلت-رافتری و در بخش 4-6 روش بیرن ، در بخش 4-7 روش لگاریتم دوره نگار GPH ، در بخش 4-8 روش دوره نگار هموار شده ، در بخش 4-9 روش ویتل ، در بخش 4-10 روش شبه حداکثر درستنمایی و در 4-11 به برآورد پارامتر تفاضل گیری در ARFIMA نامانا می پردازیم.

در فصل پنجم روش بیزی را مطرح می کنیم ، در بخش5-3 مدل سازی بیزی در قرن 21 ، در بخش 5-4 مفاهیم و واژگان اختصاصی ، بخش 5-6 برآوردگرهای بیز ، در بخش 5-7 انتخاب توزیع پیشنهادی و در بخش 5-8 شبیه سازی پسین، در بخش 5-10 روشهای محاسبهٔ بیزی ، در بخش 5-11 روشهای بیزی در مدلهای ARFIMA ، در بخش 5-12 استنباط بیزی ، در بخش 5-13 مشکلات مدلسازی بیزی ، در بخش 5-14 چارچوب بیزی ، در بخش 5-15 شکل چگالی پسین ، در بخش 5 -16 افراز تابع درستنمایی ، در بخش 5-17 برآورد مستقیم درستنمایی گوسی ، در بخش 5-18 مدلسازی بیزی ، در بخش 5-18-1روشهای زنجیر مارکوف مونت کارلو و در بخش 5-20 به کاربرد روی داده های واقعی و مقایسهٔ روشها می پردازیم.

فهرست مطالب

فصل اول

آشنایی با سری زمانی ومفاهیم اولیه

برای مطالعه فرآیندهای با حافظه بلند لازم است مروری بر سری زمانی داشته باشیم که قسمتهایی از نمایش طیفی را نیز شامل میشود.

1–1–1 تعریف سری زمانی

یک سری زمانی دنباله ای مرتب شده از مشاهدات است به عبارت دیگر یک سری زمانی عبارت است از داده هایی که از مشاهده یک پدیده در طول زمان به دست آمده اند . گرچه معمولاً بر حسب زمان ،بویژه در فواصل زمانی مساوی مرتب میشوند،ولی مرتب شدن ممکن است با توجه به ابعاد دیگری چون فاصله نیز باشد.سریهای زمانی در زمینه های گوناگونی مثل کشاورزی ،بازرگانی،مهندسی ،صدا،ژئوفیزیک،هواشناسی،سرعت باد در ساعت،میزان باران سالانه،در مطالعات پزشکی(برای اندازه گیری ردهای الکتروانسفاوگرام و الکترو کاردیوگرام) ،نرخهای مرگ ومیر کاربرد دارد، البته زمینه هایی که در آن سریهای زمانی مشاهده و مطالعه میشود بی پایان است.در یک تقسیم بندی کلی سریهای زمانی را به دو دسته پیوسته و گسسته تقسیم میکنیم .یک سری زمانی نظیر علائم الکتریکی و ولتاژ که مشاهدات در آنها بطور پیوسته در زمان ایجاد میشوند، سریهای زمانی پیوسته میباشند.یک سری زمانی ،مانند نرخ های بهره وحجم فروش را که مشاهدات فقط در زمانهای معینی که معمولاً به فواصل مساوی از یکدیگر قرار دارند ایجاد شده باشند، سریهای زمانی گسسته میباشند. یک سری زمانی با توجه به الگوهای پارامتری ،از توابع خودهمبستگی و خود همبستگی جزئی استفاده میکند،تحلیل در قلمرو زمان می نامند.یک روش جایگزین را که برای مطالعه تجزیه ناپارامتری یک سری زمانی به مؤلفه های فرکانس مختلف آن از توابع طیفی استفاده میکند،تحلیل در قلمرو فرکانس می نامند.گرچه این دو روش از نظر

ریاضی معادلند،بدین مفهوم که تابع خودهمبستگی و تابع طیفی یک جفت تبدیل فوریه را تشکیل میدهند ولی مواردی وجود دارد که یک روش بر دیگری رجحان دارد.

1-1-2 فرآیند تصادفی

یک فرآیند تصادفی ،خانواده ای از متغیرهای تصادفی y(w,t) است که روی یک فضای احتمال تعریف شده است و w به فضای نمونه و t به یک مجموعهٔ شاخص ،متعلق میباشد.برای یک ثابت y(w,t)،t یک متغیر تصادفی است.برای یک w معلوم y(w,t) به عنوان تابعی از t ،یک نمونه یا یک مصداق نامیده میشود.جامعه ای متشکل از تمام مصادیق ممکن در فرایندهای تصادفی و تحلیل سریهای زمانی ،یک مجموعه کلی نامیده میشود.بنابراین ،یک سری زمانی یک مصداق یا یک تابع نمونه از یک فرآیند تصادفی معین است.در این بحث ،مجموعهٔ شاخص را مجموعهٔ تمام اعداد صحیح فرض میکنیم.مجموعه ای متناهی از متغیر های تصادفی $\{Y_{t_1},...., Y_{t_n}\}$ را ،از یک فرآیند تصادفی $\{Y(w,t) := t = 0, \pm 1, \pm 2, ...\}$ در نظر میگیریم.و برای آن تابع توزیع n- بعدی زیر را تعریف میکنیم:

$$F(Y_{t_1},....., Y_{t_n}) = p\{w : y(w,t_1) \le y_{t_1},..., y(w,t_n) \le y_{t_n}\}.$$

1-1- 3ضرب داخلی

فرض کنید M یک فضای برداری باشد.یک ضرب داخلی یک تابع بشکل $<.,.>: M \to R$ میباشد چنانکه برای همه $x, y, z \in M$ و $\alpha, \beta \in R$ در شرطهای زیر صدق کند:

1) $< \alpha x + \beta y, z >= \alpha < x, z > + \beta < y, z >$

2) $< x, y >=< y, x >$

3) $< x, x > \ge 0$ و $< x, x >= 0$ اگر و تنها اگر $x = 0$.

فضای برداری M با شرایط بالا یک فضای ضرب داخلی($<.,.>$)نامیده میشود.

نرم فضای برداری M تابعی است که بصورت $\|.\|: M \to [0, \infty)$ که برای همه $x, y \in M$ و $\alpha \in R$ در شرایط زیر صدق میکند:

1) $\| x \| = 0$ اگر و تنها اگر $x = 0$.

2) $\| \alpha x \| = | \alpha | \| x \|$

3) $\|x+y\| \le \|x\| + \|y\|$.

با توجه به تعریف ضرب داخلی نتیجه میگیریم که $\|x\| = \sqrt{<x,x>}$. $\{x_t\}$ در M نشان دهنده یک دنباله کوشی است اگر و تنها اگر برای همه $\varepsilon > 0$ عدد صحیح n وجود داشته باشد طوری که $\|x_t - x_s\| < \varepsilon$ (برای $t,s > n$). فضای برداری M کامل است اگر وتنها اگر هر دنباله کوشی یک حد در M داشته باشد.با این تعاریف ما فضای هیلبرت را معرفی میکنیم.یک فضای هیلبرت یک ضرب داخلی کامل فضای برداری است.این فضاها مخصوصا برای تحلیل سریهای زمانی با توجه به قضیه تصویر که در بخش های بعدی آمده است مهم هستند.

تعریف:(گستره[1]). گستره $\overline{sp}\{x_t : t \in T\}$ از هر زیر مجموعه $\{x_t : t \in T\}$ از یک فضای هیلبرت H کوچکترین زیر فضای بستهٔ H که شامل هر مولفهٔ $x_t : t \in T$ میباشد، تعریف میشود.

نکته:فضای بسته از یک مجموعهٔ متناهی $\{x_1,..,x_n\}$ مجموعه ای از همهٔ ترکیبهای خطی $y = \alpha_1 x_1 + ... + \alpha_n x_n, \alpha_1,...,\alpha_n \in C(or)R$ میباشد.برای مثال اگر $x_1, x_2 \in R^3$ و x_1 یک اسکالر چندگانه[2] از x_2 باشد آنگاه $\overline{sp}\{x_1, x_2\}$ شامل صفحه ای[3] از x_1 و x_2 میباشد.

1–1–4 مجموعه های متعامد[4] :

A یک مجموعهٔ $\{e_t : t \in T\}$ ازمولفه های یک فضای ضرب داخلی متعامد گفته میشود اگر برای هر $s,t \in T$ داشته باشیم

$$<e_t, e_s> = \begin{cases} 1 & if\ s = t \\ 0 & if\ s \neq t \end{cases} \qquad (1\text{–}1)$$

قضیه1–1:

فرض کنید M یک زیر فضای بسته از فضای هیلبرت H باشد و فرض کنید $x \in H$.سپس

1)یک نقطه یکتای $y \in M$ وجود دارد چنانکه $\|x-y\| = \inf_{z \in M} \|x-z\|$

---

[1]- *span*
[2]- *multiple*
[3]- *plane*
[4]- *Orthonormal Sets*

2) $y \in M$ و $\| x-y \| = \inf_{z\in M} \| x-z \|$ اگر و تنها اگر $y \in M$ و $<x-y, z>=0$ برای همه $z \in M$
.

اثبات:[33]

مثال:فرض کنید n یک عدد صحیح مثبت باشد و فضای C^n با ضرب داخلی اقلیدسی زیر:

$$\langle x, y\rangle = x\bar{y} = \sum_{j=1}^{n} x_j \bar{y}_j$$

که در آن $x=(x_1,..,x_n)', y=(y_1,..,y_n)'$ و $\bar{y}$ مزدج مختلط از y میباشد.آنگاه، C^n یک فضای هیلبرت با اندازه $\| x \| = \left(\sum_{j=1}^{n} | x_j |^2 \right)^{1/2}$.

مثال :فرض کنید $\{y_t : t \in Z\}$ یک فرآیند تصادفی با میانگین 0 واقعی ،تعریف شده روی یک فضای احتمالی (Ω, F, p) باشد.آنگاه ، $L_2(\Omega, F, p)$ یک فضای هیلبرت با ضرب داخلی $\langle x, y\rangle = E(xy)$ و نرم $\| x \| = \sqrt{E(y^2)}$ (که در آن E(.) عملگر امید ریاضی است)،این فضا L_2 نامیده میشود.

مثال:فرض کنید F تابع توزیع باشد وفرض کنید $f(\lambda), g(\lambda)$ با توابع مختلط-مقدار با دامنه $[-\pi, \pi]$ باشد بطوری که:

$$\int_{-\pi}^{\pi} | f(\lambda) |^2 \, dF(\lambda) < \infty, \quad \int_{-\pi}^{\pi} | g(\lambda) |^2 \, dF(\lambda) < \infty$$

آنگاه

$$\langle f, g\rangle_F = \int_{-\pi}^{\pi} f(\lambda) \overline{g(\lambda)} dF(\lambda) \qquad (2\text{–}1)$$

یک ضرب داخلی تولید شده در فضای هیلبرت میباشد که با $L_2(F)$ نشان داده میشود.مشاهده میشود که نابرابری هولدر[5] روی این ضرب داخلی بشکل زیر تعریف میشود:

$$\left|\langle f, g\rangle\right|^2 \le \int_{-\pi}^{\pi} | f(\lambda) |^2 \, dF(\lambda) \int_{-\pi}^{\pi} \left| g(\lambda) \right|^2 dF(\lambda) < \infty \qquad (3\text{–}1)$$

[5] - *Holders*

1–1–5 تکینی و نظم

اگر $F_t = \overline{sp}\{y_s : s < t\}$ گذشته فرآیند در زمان t باشد و $F_{-\infty} = \bigcap_{t=-\infty}^{\infty} F_t$.فرآیند قطعی[۶] به معنی پیش بینی کننده موثر[۷] است ، اگر و تنها اگر $y_t \in F_{-\infty}$ (برای $t \in Z$). بعبارت دیگر ،یک فرآیند قطعی است اگر وتنها اگر $F_{-\infty} = F_{\infty}$.این فرآیندها همچنین تکین نامیده میشود.از طرف دیگر یک فرآیند غیر قطعی خالص[۸] یا نظم[۹] نامیده میشود اگر و تنها اگر $F_{-\infty} = \{0\}$.

قضیه1–1–5

فرض کنید $\{x_t\}$ یک فرآیند با میانگین 0 و فرض کنید $\{\varphi_j : j \in Z\}$ یک دنباله مطلقاً جمع پذیر باشد.آنگاه:

1)فرآیند $\{y_t\}$ بشکل $y_t = \sum_{j=-\infty}^{\infty} \varphi_j x_{t-j} = \varphi(B) x_t$ یک فرآیند مانا با میانگین 0 است.

2)اگر فرآیند مانا $\{x_t\}$ بشکل $x_t = \sum_{j=-\infty}^{\infty} \eta_j \varepsilon_{t-j} = \eta(B) \varepsilon_t$ باشد آنگاه:

$$y_t = \varphi(B)\eta(B)\varepsilon_t = \eta(B)\varphi(B)\varepsilon_t = \psi(B)\varepsilon_t$$

که در آن معادلات میانگین مربع معنی دار[۱۰] میباشد و $\psi(B) = \eta(B)\varphi(B)$.

اثبات: [33]

1–1–6 علت پذیری

فرآیند ARMA (p, q) تعریف شده با معادله $\phi(B)Y_t = \theta(B)\varepsilon_t$ را علت پذیر یاکازال گوییم اگر دنباله ای از ثابت های $\{\psi_j\}$وجود داشته باشد به طوری که $\sum_{j=0}^{\infty} |\psi_j| < \infty$, $Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$

مثال :

مانایی و علت پذیری دو مفهوم مرتبط هستند اما ضرورتاً معادل نیستند .با نشان دادن این نکته فرض میکنیم که $\{\varepsilon_t\}$ یک دنباله نوفهٔ سفید با $Var(\varepsilon_t) < \infty$ میباشد، برای مثال فرآیند $y_t = \varepsilon_t + \varepsilon_{t+1}$ را در نظر

6- deterministic process
7- *perfectly predictable*
8- *purely nondeterministic*
9- *regular*
10- *sense*

بگیرید.این فرآیند مانا است اما علت پذیر نیست زیرا y_t وابسته به مقادیر آینده دنباله $\{\varepsilon_t\}$میباشد.از طرف دیگر ،فرآیند $y_t = \varepsilon_t + \varepsilon_{t-1}$ مانا و علت پذیر است.

قضیه1– 3 (تجزیهٔ ولد[11])

هر فرآیند مانا مجموع یک فرآیند نظم و یک فرآیند تکین میباشد.این دو فرآیند متعامدند و تجزیه یکتا دارند.

طبق قضیه تجزیه ولد ،یک فرآیند غیر قطعی محض ممکن است بشکل زیر بیان شود:

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \psi(B)\varepsilon_t \qquad (4\text{–}1)$$

که در آن $\psi_0 = 1$ و $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ ، و $\{\varepsilon_t\}$ یک فرآیند نوفهٔ سفید با واریانس σ^2 میباشد.بسط ولد یکتا است و $\varepsilon_t \in F_{t+1}$ برای همهٔ $t \in Z$.

اثبات:[33]

1–1–7وارون پذیری :

یک فرآیند منظم خطی 1–4 یک فرآیند وارون پذیر نامیده میشود اگر یک دنباله از ضرایب $\{\pi_j\}$ وجود داشته باشد چنانکه

$$\varepsilon_t = -\sum_{j=0}^{\infty} \pi_j y_{t-j} \qquad (5\text{–}1)$$

که همگرا به L_2 میباشد.از1–4 و با فرض $\pi_0 = -1$ فرآیند $\{y_t\}$ ممکن است بشکل زیر بیان شود:

$$y_t = \varepsilon_t + \sum_{j=1}^{\infty} \pi_j y_{t-j}$$

1–1–8 ارگودیکی

فرض کنید $y_1,..,y_n$ دنباله ای از متغیر های تصادفی مستقل و هم توزیع باشند.اگر بخواهیم تابع توزیع F(y) مربوط به آنها را برای تمام y ها برآورد کنیم،میتوانیم این کار را با مشاهده تعدادی متناهی از y ها

[11]- Wold Decompsition