

برآوردگر ناپارامتری موجکی - آستانه‌ای تابع چگالی و ساختار کوواریانس ضرائب موجکی

محمود افشاری^{1*}، نرگس عباسی²، علیرضا مهردوست³

1. دانشیار، گروه آمار، دانشگاه خلیج فارس بوشهر

2. دانشیار، گروه آمار، دانشگاه پیام نور

3. کارشناس ارشد، گروه آمار، دانشگاه پیام نور

تاریخ دریافت: 1396/11/18 تاریخ پذیرش: 1397/03/14

Wavelet - Threshold Nonparametric Density Estimator and Covariance Structure of Wavelet Coefficients

M. Afshari¹, N. Abbasi², A. Mehrdoost³

1. Associate Professor, Department of Statistics, Persian Gulf University, Bushehr

2. Associate Professor, Department of Statistics, Payam Noor University

3. M.Sc., Department of Statistics, Payam Noor University

Received: 2018/02/07 Accepted: 2018/06/04

Abstract

Wavelets are one of the newest achievements of mathematical science, which have many applications in other sciences especially statistics. In this paper, after introducing wavelet transforms, the nonparametric estimator of the density function is expressed by the nuclear wavelet and threshold wavelet method. Then variance-covariance of wavelet coefficient is investigated. At the end we survey the theoretical outcomes with numerical computation by using R software to compare purpose estimators.

Keywords

Wavelet, Wavelet Coefficients, Scale Function, Threshold.

چکیده

موجک‌ها یکی از جدیدترین دستاوردهای علم ریاضی هستند که کاربردهای زیادی در مخابرات و سایر علوم به‌ویژه آمار دارند. در این مقاله پس از معرفی تبدیلات موجک، برآوردگر ناپارامتری تابع چگالی به روش موجک و آستانه‌ای به دست آمده و سپس ساختار واریانس - کوواریانس ضرائب موجکی بررسی می‌شود. در پایان، مطالب تئوری به دست آمده، در عمل، با استفاده از نرم‌افزار R، به کمک شبیه‌سازی بررسی و دو نوع برآوردگر مقایسه می‌شوند.

واژگان کلیدی

موجک، ضرائب موجکی، تابع مقیاس، آستانه.

مقدمه

موجک‌ها یکی از دستاوردهای علم ریاضیات است که میتنی بر چندین دهه پژوهش در آنالیز هم‌ساز هست. در آنالیز موجک همانند آنالیز فوریه با بسط تابع‌ها سروکار داریم ولی این بسط برحسب موجک‌ها انجام می‌گیرد. موجک تابع مشخص مفروضی با میانگین صفر است و بسط برحسب انتقال و اتساع‌های این تابع انجام می‌گیرد. برخلاف چندجمله‌ای‌های مثلثاتی، موجک‌ها در فضا به صورت موضعی بررسی می‌شوند و به این ترتیب ارتباط نزدیک‌تری بین بعضی توابع و ضرایب آنها امکان‌پذیر می‌شود و پایداری عددی بیشتری در بازسازی و محاسبات فراهم می‌گردد. موجک‌ها توابع ریاضی هستند که سیگنال‌ها را به اجزا فرکانسی مختلفی می‌شکنند و سپس هر جزء را با تحلیل مقیاسی مربوطه مورد بررسی قرار می‌دهند. مزیت آنها نسبت به روش‌های قدیمی فوریه در تجزیه سیگنال‌ها به وسیله ناپویسته‌ها و توابع نوک‌تیز است.

اخیراً موضوع تحلیل موجکی مورد توجه زیاد ریاضی‌دانان و مهندسان قرار گرفته است. برخی به آن به عنوان پایه‌ای جدید برای نمایش توابع می‌نگرند وعده‌ای آنها به عنوان فن برای تحلیل زمان بسامدی به حساب می‌آورند و گروهی نیز آنها به عنوان شاخه جدید ریاضیات می‌دانند. در حال حاضر تبدیل‌های موجکی در حال رقابت با تبدیل‌های فوریه هست.

ضعف تبدیل‌های فوریه از آنجا ناشی می‌شود که توابع سینوسی و کسینوسی، با وجود اینکه پایه‌ای برای فضای $L^2[0, 2\pi]$ هستند به $L^2(\mathbb{R})$ تعلق ندارند، بنابراین برای برطرف کردن این نقیصه باید به دنبال پایه‌ای بود که $L^2(\mathbb{R})$ را تولید کند و ترجیح می‌دهیم که این کار تنها توسط یک تابع انجام شود.

این مسئله‌ای بود که مدت‌ها ذهن پژوهشگران را به خود مشغول کرده بود. آنها به دنبال خانواده‌ای از توابع بودند که دارای محمل فشرده بوده و بر یکدیگر عمود باشند. توأم بودن این دو خاصیت بدیهی به نظر نمی‌رسید. این کار بالاخره توسط اینگرید دایبچز¹ (1985) انجام شد.

نام‌گذاری موجک‌ها به پیروی از گراسمان و مورلت² (1982) انجام شده است. محققان برجسته دیگری همانند

مالات³ (1989)، به تحقیقات گسترده در رابطه با موجک پرداختند.

تعریف 1: برای یک بازه $a \leq t \leq b$ فضای $L^2([a, b])$ مجموعه همه توابع مربعی انتگرال‌پذیر روی $a \leq t \leq b$ است. به عبارت دیگر:

$$L^2([a, b]) = \{ f : [a, b] \rightarrow \mathbb{C}; \int_a^b |f(t)|^2 dt < \infty \}$$

ضرب داخلی L^2 روی $L^2([a, b])$ به صورت

$$\langle f, g \rangle_{L^2} = \int_a^b f(t) \overline{g(t)} dt \quad f, g \in L^2([a, b])$$

و نرم f به صورت $\|f\| = \sqrt{\int_a^b |f(t)|^2 dt}$ تعریف می‌شود.

تعریف 2: تابع $f(x)$ را متعلق به فضای سوبولوف⁴ $W_2^s(\mathbb{R})$ می‌نامیم هرگاه:

$$\| (1 + |w|^2)^{\frac{s}{2}} \hat{f}(w) \| < \infty$$

که در آن $\hat{f}(w)$ تبدیل فوریه تابع $f(x)$ در بسامد w است.

تعریف 3: تابع $f(x)$ را متعلق به فضای هولدر⁵ $C^s(A)$ گوئیم هرگاه شرایط زیر برقرار باشد:

الف) اگر $s = n$ آنگاه $f(x)$ ، n بار مشتق‌پذیر باشد.
ب) اگر $0 < s < 1$ آنگاه

$$C^s(\mathbb{R}) = \{ f : f \in L^\infty(\mathbb{R}), \sup_x \frac{|f(x+h) - f(x)|}{|h|^s} < \infty \}$$

ج) اگر $0 < s < 1$ ، $s = n + s_0$ آنگاه

$$C^s(\mathbb{R}) = \{ f : f \in L^\infty(\mathbb{R}) \cap C^n(\mathbb{R}) \mid \frac{d^n f}{dx^n} \in C^{s_0}(\mathbb{R}) \}$$

3. Mallat
4. Soboluf space
5. Holder space

1. Ingrid Daubchies
2. Grasman and Morlet

$$y(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{سایر جاها} \end{cases} \quad (1)$$

برای $j \geq 0$ و $k = 0, 1, \dots, 2^j - 1$ انتقال‌های مقیاسی ϕ و ψ ، به صورت زیر تعریف می‌شوند

$$\begin{aligned} \phi_{j,k}(x) &= 2^{-j} \phi(2^j x - k), \\ \psi_{j,k}(x) &= 2^{-j} \psi(2^j x - k) \end{aligned} \quad (2)$$

توابع ϕ و $\psi(x)$ و $\phi_{j,k}(x)$ و $\psi_{j,k}(x)$ که قطعه‌وار ثابت هستند می‌توانند در برآورد توابع استفاده شوند و برای هر j ثابت $\psi_{j,k}$ را موجک‌های نسل j ام گوئیم. اگر برای هر $j \in \mathbb{Z}$ ، $V_j = \text{Span}\{\phi_{j,k}, k \in \mathbb{Z}\}$ ، در این صورت V_j ها دنباله‌ای از فضاهای تودرتو به صورت $V_j \subset V_{j+1}$ خواهند بود. در نهایت برای تقریب تابع f در مقیاس مورد نظر، لم زیر ارائه می‌شود.

تعریف 4: یک تابع $\psi \in L^2(\mathbb{R})$ ، یک آنالیز تجزیه چندگانه² (MRA) تولید می‌کند؛ هرگاه یک دنباله از زیر فضاهای تودرتو تولید کند به طوری که در شرایط زیر صدق کند:

$$\dots \subset V_{-1} \subset V_0 \subset V_1 \subset \dots \quad (1)$$

$$(U_{j \in \mathbb{Z}} V_j) = L^2(\mathbb{R}) \quad (2)$$

$$U_{j=-\infty}^{\infty} V_j = L^2(\mathbb{R}) \quad (3)$$

$$(\cap_{j \in \mathbb{Z}} V_j) = \{0\} \quad (4)$$

$$V_k = \text{span}\{\phi(x - k), k \in \mathbb{Z}\} \quad (5)$$

$$f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1} \quad (6)$$

این اصول یک دنباله از فضاهای تقریب V_j در $L^2(\mathbb{R})$ را توصیف می‌کند که بستر اجتماع آنها با فضای $L^2(\mathbb{R})$ برابر است (اصل سوم)، همچنین تصویر تابعی مانند

در این مقاله ابتدا آنالیز موجکی بیان می‌شود و سپس با معرفی آستانه، برآوردگر ناپارامتری تابع چگالی به روش موجک و آستانه‌ای را به دست آورده و سپس ساختار واریانس - کوواریانس ضرایب موجکی بررسی می‌شود. در پایان با استفاده از شبیه‌سازی در دو مثال جداگانه برآوردگرهای به دست آمده مقایسه می‌شود.

آنالیز موجکی

نظریه موجک‌ها ابتدا توسط آلفرد هار¹ در سال (1910) مطرح شد. اساس کار هار به این صورت بود که او برای تقریب یک تابع پیوسته f در فاصله $[0, 1]$ از یک تابع پله‌ای واحد مانند $j_0(x)$ شروع کرد و با انتقال هر زوج پله مجاور ه از توابع را به صورت زیر نوشت:

$$f_n(x) = \langle f, j_0 \rangle j_0(x) + \langle f, j_1 \rangle j_1(x) + \dots + \langle f, j_n \rangle j_n(x)$$

به طوری که $\langle f, j_i \rangle = \int_0^1 f(x) j_i(x) dx$ ، در این صورت اگر $\mathbb{R} \ni \mathbb{N}$ در این صورت f_n تا به طور یکنواخت به f همگرا هستند.

موجک‌ها ابزاری قابل انعطاف با محتوای غنی ریاضی هستند که در تجزیه و تحلیل یک تابع نوسانی از زمان و مکان به کار می‌روند. در تجزیه موجکی، موجک اصلی (تابع یا سیگنال) روی دسته‌ای از توابع پایه که موجک نامیده می‌شوند، تصویر شده و با استفاده از حاصل ضرب موج اصلی و تابع موجک، تجزیه موجکی انجام می‌شود. موجک، تابعی با میانگین صفر است که بسط بر اساس انتقال و اتساع آن انجام می‌شود. اساس موجک از موجک پدر $\phi(\cdot)$ و موجک مادر $\psi(\cdot)$ تشکیل شده که به ترتیب روند کلی داده‌ها و انحرافات را نشان می‌دهند و به صورت زیر تعریف می‌شوند:

$$\phi(x) = \begin{cases} 1, & 0 \leq x < 1, \\ 0, & \text{سایر نقاط} \end{cases}$$

با توجه به روابط (1) و (2) نمایشی از بهترین مقیاس تابع f به صورت زیر است:

$$f(x) = \sum_{k \in \mathbb{Z}} c_{j_0 k} \phi_{j_0 k}(x) \sum_{j=j_0}^{\infty} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x). \quad (3)$$

اکنون با توجه به این که برای بردار داده‌ها ψ ، اولین مقیاس درشت $J-1$ است و با قرار دادن $k, j_0 = 0$ ، مربوط به $\phi_{j_0 k}$ صفر و k مربوط به $\psi_{j,k}$ از 0 تا $J-1$ خواهد بود، رابطه (3)، رابطه زیر که تجزیه موجکی تابع f نامیده می‌شود را نتیجه می‌دهد.

$$f(x) = c_{0,0} \phi(x) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x). \quad (4)$$

تبدیل موجک گسسته در بسیاری از تجزیه و تحلیل‌های سیگنال و پردازش و متراکم‌سازی تصاویر استفاده می‌شود. نمایش ریاضی این تبدیل را با فرض اینکه بردار داده‌ها باشد که در آن J عدد صحیح مثبت و نشان‌گر مقیاس داده‌هاست.

برآورد موجکی و آستانه‌ای تابع چگالی

یکی از مباحث مهم در آمار نظریه برآوردیابی است. اهمیت برآورد چگالی به آن دلیل است که کاربردهای فراوانی در تحلیل و تبدیل داده‌های آماری و ثبت تصاویر و غیره دارد. روش‌های مختلفی برای برآورد چگالی وجود دارد که به دو گروه اصلی پارامتری و ناپارامتری تقسیم می‌شود.

در روش برآورد پارامتری توزیع آماره $\mathbf{T}(x)$ بر اساس نمونه تصادفی X_1, \dots, X_n معلوم است و یا در صورت مجهول بودن، توزیع $\mathbf{T}(x)$ به صورت تقریبی بنا به قضیه حد مرکزی معلوم است و در هر دو این حالات استنباط به کمک آماره $\mathbf{T}(x)$ انجام می‌شود؛ ولی در برآورد ناپارامتری چون توزیع نمونه تصادفی نامعلوم است و همچنین به دلیل غیرخطی بودن آماره $\mathbf{T}(x)$ برحسب $\sum X_i$ و استفاده نکردن از قضیه حد مرکزی، امکان تعیین توزیع $\mathbf{T}(x)$

$f \in L^2(\mathbb{R})$ به درون فضای V_j ، f_j تقریبی از تابع f است که هنگامی که $j \rightarrow \infty$ آنگاه f_j به f همگراست. با توجه به اصل پنجم فضای V_j پایه‌ای دارد که از انتقال‌های صحیح تابع ϕ به دست می‌آید. در نهایت به دلیل نیازی که تابع f دارد تا از فضای V_j به V_{j+1} و یا برعکس منتقل شود، فضاها باید با یکدیگر در ارتباط باشند.

حال فرض کنید فضاها V_j داده شده‌اند. آنگاه به ازای هر تابع V_j یک زیرفضایی W_j که در تعامد با V_j است، وجود دارد به طوری که:

$$V_j \cap W_j = \{0\}, \quad V_{j+1} = V_j \oplus W_j$$

با توجه به خاصیت تودرتو بودن، می‌توان نوشت:

$$L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j = \bigoplus W_{-1} \oplus W_0 \oplus W_1 \oplus \dots$$

از آنجا که $V_j \cap W_j = \{0\}$ ، $W_j \subset V_{j+1}$ پس می‌توان یک مجموعه از توابعی مانند ψ پیدا کرد که $\psi \in V_{j+1}$ ولی $\psi \in V_j$ و مجموعه توابع ψ پایه‌ای برای W_j تشکیل دهد.

این رده از توابع را موجک می‌نامیم و فضای W_j را به صورت زیر برحسب توابع پایه‌ای تعریف می‌کنیم:

$$W_j = \text{span}\{\psi(2^j x - k), k \in \mathbb{Z}\} \quad j \in \mathbb{Z}$$

همان‌طور که $W_j \subset V_{j+1}$ پس می‌توان هر $\psi(2^j x - k)$ را به صورت ترکیب خطی از توابع پایه‌ای $\phi(2^{j+1} x - k)$ نوشت یا به عبارت دیگر:

$$\psi(x) = \sum_{k \in \mathbb{Z}} h_k \phi(2x - k)$$

یا در حالت کلی:

$$\psi(2^j x) = \sum_{k \in \mathbb{Z}} h_k \phi(2^{j+1} x - k)$$

ϕ و ψ را به ترتیب توابع موجکی پدر و مادر مرتبه r معرفی می‌کنیم که کران‌دار دارای محمل فشرده هستند و دارای خواص زیر هستند:

$$\int \phi^2 = \int \psi^2 = 1, \quad (7)$$

$$\int x^m \psi(x) = \begin{cases} 0 & 0 \leq m \leq r-1 \\ r! k & m = r \end{cases} \quad (8)$$

ϕ_k و ψ_{jk} پایه‌های یکمعامد از کلاس توابع $f \in L(R)$ به صورت زیر تعریف می‌شوند:

$$\phi_k(x) = p^{-\frac{1}{2}} \phi(px-k), \quad \psi_{jk}(x) = p_j^{-\frac{1}{2}} \psi(p_j x - k), \quad p > 0, p_j = p^j, j \geq 0$$

که در روابط زیر صادق هستند:

$$\int \phi_{k_1} \phi_{k_2} = \delta_{k_1 k_2}, \quad \int \psi_{j_1 k_1} \psi_{j_2 k_2} = \delta_{j_1 j_2} \delta_{k_1 k_2}, \quad \int \phi_{k_1} \psi_{j_2 k_2} = 0$$

فرض کنید $X_1, \dots, X_n \stackrel{iid}{\sim} f$ باشد، ضرایب موجکی تجربی به صورت

$$\hat{\alpha}_k = n^{-1} \sum_{i=1}^n \phi_k(x_i), \quad \hat{\beta}_{jk} = n^{-1} \sum_{i=1}^n \psi_{jk}(x_i).$$

تعریف می‌شوند که برآوردگرهای ناریب $\hat{\alpha}_k$ و $\hat{\beta}_{jk}$ هستند. اکنون با جایگزین کردن ضرایب موجکی تجربی در رابطه (6)، برآوردگر موجکی تابع f به دست می‌آید:

$$\hat{f} = \sum_k \hat{\alpha}_k \phi_k + \sum_k \sum_{j=0}^{\infty} \hat{\beta}_{jk} \psi_{jk}. \quad (9)$$

برآوردگر حاصل را برآوردگر موجکی غیرخطی می‌نامند. در حقیقت دلیل غیرخطی بودن برآوردگر بالا را می‌توان مجموع‌های دوگانه بیان کرد که ناپیوستگی‌های \hat{f} را میسر می‌سازد. برآوردگر موجکی خطی، مؤلفه اول رابطه (9) است که به صورت زیر تعریف می‌شود:

$$\hat{f} = \sum_k \hat{\alpha}_k \phi_k, \quad (10)$$

وجود ندارد. در این بخش، ابتدا، برآورد تابع چگالی را به روش موجکی بیان می‌کنیم و سپس با تشریح قوانین آستانه‌ای، شرایطی روی پارامتر آستانه در نظر می‌گیریم که برآوردگر موجکی نرخ همگرایی بهینه و قابل قبولی داشته باشد. موجک‌ها بنا به خواص متعددشان در تحقیقات آماری مورد توجه ویژه قرار گرفته‌اند. اولین بار، موجک‌ها را داناها¹ و جانستون² (1992) در تحقیقات آماری بکار بردند. در سال‌های اخیر، تحقیقات زیادی درباره کاربرد موجک‌ها در برآورد توابع آماری از جمله توابع چگالی، مشتق تابع چگالی، توابع رگرسیون انجام شده است که می‌توان به آنتونیادیس³ (1994) و داناها و همکاران⁴ (1996) و ویداکویک⁵ (1999) و هریک⁶ (2000) دوستی و همکاران (2006) و (2008)، نیسون (2011) و افشاری (2013, 2014, 2016, 2017) اشاره کرد.

فرض کنید $f \in L(R)$ کلاسی از توابع باشد که در شرایط زیر صدق کند:

$$f \geq 0 \text{ و } \int f = 1 \quad (\text{الف})$$

$$\int |f^{(m)}| \leq B \quad (\text{ب})$$

(د) $f^{(r)}$ در بازه $(-\infty, \infty)$ پیوسته یکنواخت است.

(ه) $f^{(r)}$ در فاصله $(-\infty, -B)$ و (B, ∞) یکنوا

است.

که در آن $1 \leq m \leq r$ و $B \geq 1$ است.

فرض بر این است که تابع f متعلق به فضای سوبولوف باشد و در این کلاس دارای r مشتق کران‌دار باشد و اکنون هدف برآورد موجکی تابع f است.

در تبدیل موجکی تابع f مجموع‌های دوگانه نامتناهی داریم که یکی روی وضوح (ا) و دیگری روی مکان (ک) بسته می‌شود. تبدیل موجکی $f \in L(R)$ به صورت زیر است:

$$f = \sum_k \alpha_k \phi_k + \sum_k \sum_{j=0}^{\infty} \beta_{jk} \psi_{jk}. \quad (6)$$

1 Donoho

2 Jonston

3 Antoniadis

4 Donoho et all

5 Vidakovic

6 Herrick

پارامتر آستانه

در این بخش می‌خواهیم وضعیت همگرایی برآورد موجکی غیرخطی به دست آمده در رابطه (9) را بررسی کنیم. این برآوردگر به دلیل مجموع‌های نامتناهی، فاقد همگرایی است. این مسئله با برش مجموع روی وضوح قابل حل است. بدین صورت که مجموع‌های دوگانه متناهی را خواهیم داشت که ضرایب موجکی $(\hat{\beta}_{jk})$ از یک مقدار خاص تجاوز می‌کنند. این مقدار خاص آستانه و این روش را هموارسازی گوییم.

بحث هموارسازی در بخش برآورد توابع بر اساس موجک‌ها، بدین معنا است که برخی از ضرایب را در بسط موجکی برابر صفر قرار دهیم. این کار به روش‌های مختلفی انجام می‌گیرد؛ یکی از این راه‌ها، روش انقباضی کردن است. برای مثال چندین جمله اول از بسط تابع را نگه داشته و بقیه را برابر صفر قرار می‌دهیم. این روش یک هموارسازی خطی است؛ راه دیگر نگه داشتن آن ضرایبی است که قدر مطلق اندازه آنها بیشتر از مقدار خاصی باشد، که نتیجه یک تابع غیرخطی از ضرایب خواهد بود. به این روش آستانه‌ای کردن می‌گوییم. در ادامه به بررسی این روش می‌پردازیم.

عمل برش باید به گونه‌ای باشد که ضرایب تجربی که در برآوردگر تابع وجود دارند بر اساس اطلاعات کافی باشند؛ بنابراین انتخاب آستانه برای این عملکرد بسیار مهم است. یکی از بحث‌های مهم در روابط موجک‌ها، این است که تصمیم بگیریم چه تعداد از سطوح (ضرایب موجکی تجربی) را در برآورد چگالی وارد کنیم. اگر سطوح زیادی برای برآورد انتخاب شده باشند، آنگاه برآمدگی‌های خیلی ریز تابع نشان داده می‌شود و از نظر هندسی قله‌های زیادی در منحنی توابع پیدا می‌شود که مناسب نیست و زمانی که تعداد کمی از سطوح در دسترس‌اند ممکن است اطلاعات مهمی از چگالی به دست نیاید؛ بنابراین ما کسیم سطح را تعیین می‌کنند. بعد از پیدا کردن ما کسیم سطح، باید تصمیم بگیریم که کدامیک از ضرایب موجکی را وارد مدل کنیم. یک روش برای تعیین این که کدامیک از ضرایب موجکی را وارد مدل کنیم، روش آستانه است. در واقع برای اغتشاش‌زدایی بهین، از آستانه استفاده می‌کنیم. منظور از اغتشاش‌زدایی بهین، حذف اغتشاش‌ها به گونه‌ای است که ویژگی‌های اصلی سیگنال حذف نشود.

انواع آستانه گذار

بعد از این که مقدار آستانه مشخص شد، می‌توان ضرایب موجکی را به دو دسته، به صورت زیر تقسیم کرد:
الف) ضرایب موجکی که بیشتر از مقدار آستانه هستند، ضرایب مهم نام‌گذاری می‌شوند.
ب) ضرایب موجکی که از مقدار آستانه کمتر هستند، ضرایب کم‌اهمیت نامیده می‌شوند.
آستانه گذار، یک روش غیرخطی است که روی ضرایب موجکی عمل می‌کند. انواع متفاوتی از آستانه گذار وجود دارد که در اینجا به معرفی دو نوع از آنها می‌پردازیم:

الف) آستانه گذار سخت

این آستانه گذار توسط داناها و جانستون تعریف شده است و عملکرد آن بدین صورت است که ضرایب موجکی کم اهمیت را صفر و ضرایب مهم را بدون تغییر می‌گذارد. فرمول آن به صورت زیر است:

$$\delta_t^{\text{hard}} = \begin{cases} 0 & |\hat{\beta}_{jk}| \leq t \\ \hat{\beta}_{jk} & |\hat{\beta}_{jk}| > t \end{cases}$$

که در آن t مقدار آستانه است.

ب) آستانه گذار نرم

این آستانه گذار نیز توسط داناها و جانستون تعریف شده است و عملکرد آن بدین صورت است که ضرایب کم اهمیت را صفر و ضرایب مهم را به اندازه مقدار آستانه کاهش می‌دهد و به صورت زیر تعریف می‌شود:

$$\delta_t^{\text{soft}} = \begin{cases} 0 & |\hat{\beta}_{jk}| \leq t \\ \hat{\beta}_{jk} - t & \hat{\beta}_{jk} > t \\ \hat{\beta}_{jk} + t & \hat{\beta}_{jk} \leq -t \end{cases}$$

که در آن t مقدار آستانه است.

آستانه گذار سخت، ساده‌ترین نوع آستانه گذار و آستانه گذار نرم، تعمیم آستانه گذار سخت است.

$$\hat{f} = \sum_k \hat{\alpha}_k \phi_k + \sum_{j=0}^{q-1} \sum_k \hat{\beta}_{jk} I(|\hat{\beta}_{jk}| > t_{jk}) \psi_{jk} \quad (13)$$

که در آن I تابع مشخصه است. نمایش f و \hat{f} منحصر به فرد نیست؛ به طوری که به ازای هر عدد صحیح مثبت j^* تابع f را می‌توان به صورت زیر بیان کرد:

$$f = \sum_k \alpha_k^* \phi_k + \sum_{j=j^*}^{\infty} \sum_k \beta_{jk}^* \psi_{jk}^*$$

که در آن

$$\alpha_k^* = (\rho^{2k})^{\frac{1}{2}} \phi(\rho^{2k} x - k); \quad \psi_{jk}^* = \phi_{j+1/2k}; \quad \alpha_k^* = \int f \phi_k; \quad \beta_{jk}^* = \beta_{j+1/2k}$$

به علاوه اگر $0 \leq j \leq j^* - 1$ و $t_{jk} = 0$ باشد، برآوردگر \hat{f} تعریف شده در رابطه (13) با f تعریف شده در رابطه زیر یکسان است.

$$\hat{f} = \sum_k \hat{\alpha}_k^* \phi_k^* + \sum_{j=j^*}^{q-1} \sum_k \hat{\beta}_{jk}^* I(|\hat{\beta}_{jk}^*| > t_{j^*+j,k}^*) \psi_{jk}^* \quad (14)$$

که در آن

$$\hat{\alpha}_k^* = n^{-1} \sum_{i=1}^n \phi_k^*(X_i); \quad \hat{\beta}_{jk}^* = n^{-1} \sum_{i=1}^n \psi_{jk}^*(X_i) = \hat{\beta}_{j^*+j,k} \quad (15)$$

است.

ملاحظه 1: نوع آستانه انتخاب شده تأثیر زیادی در دقت برآوردگر حاصل دارد؛ بدین ترتیب که یک آستانه بزرگ ممکن است قسمت مهمی از تابع تحت بررسی را حذف کند و برعکس، مقدار کوچک آن باعث ورود اغتشاش در برآورد می‌شود. برای جلوگیری از این مشکلات در محاسبات مقدار آستانه را به صورت $\frac{s \sqrt{2 \log n}}{\sqrt{n}}$ در نظر می‌گیریم که به آستانه جهانی

برآوردگر موجکی آستانه‌ای تابع چگالی

در این بخش برآوردگر موجکی تابع چگالی را به روش آستانه به دست آورده و نرخ همگرایی آن را بررسی می‌کنیم.

مالات (1989)، برای اولین بار، روش آستانه‌ای کردن را در مسائل بررسی سیگنال‌ها و برآورد توابع به کار برد. داناو و جانستون (1996)، برای اولین بار موجک آستانه‌ای شده را در برآورد تابع چگالی احتمال معرفی کردند.

فرض کنید $f \sim \text{iid } X_1, \dots, X_n$ باشد، در این صورت حالت کلی برآورد موجکی f با استفاده از روش آستانه به صورت زیر است:

$$\hat{f} = \sum_k \hat{\alpha}_k \phi_k + \sum_{j=0}^{q-1} \sum_k \hat{\beta}_{jk} \omega(|\hat{\beta}_{jk}| / t_{jk}) \psi_{jk} \quad (11)$$

که در آن q پارامتر برش و t_{jk} آستانه است و تابع ω به صورت زیر تعریف می‌شود:

$$\omega(u) = \begin{cases} 0 & 0 < u < c_1 \\ \epsilon[0,1] & c_1 \leq u \leq c_2 \\ 1 & u > c_2 \end{cases} \quad (12)$$

c_1 و c_2 مقادیر ثابت هستند، به طوری که $0 < c_1 \leq c_2 < \infty$ اگر $c_1 = c_2 = 1$ آن گاه

$$\omega(u) = \begin{cases} 0 & 0 < u < 1 \\ 1 & u > 1 \end{cases}$$

هست که با جایگزین کردن آن در رابطه (11)، برآوردگر موجکی با آستانه گذار سخت به دست می‌آید.

اگر $c_1 \neq c_2$ آنگاه برآوردگر موجکی تحت آستانه گذار نرم به دست می‌آید؛ بنابراین برآورد تجربی موجکی f تحت آستانه گذار سخت به صورت زیر بیان می‌شود:

علاوه بر این می توان نوشت:

$$\int \psi_{jk}^2(x) f(x) dx = 2^{j/2} \alpha_{jk}$$

حال با استفاده از رابطه (16) و رابطه بالا واریانس ضرایب تجربی موجک هار می تواند به صورت زیر نوشته شود:

$$\text{var}[\hat{\beta}_{jk}] = \frac{1}{n} (2^{j/2} \alpha_{jk} - \beta_{jk}^2)$$

ملاحظه 2: اگر N_{jk} تعدادی نقاط داده ای باشد که در فاصله $[2^{-j}k, 2^{-j}(k+1)]$ قرار بگیرد می توان نوشت:

$$N_{jk} = n 2^{-j/2} \hat{\alpha}_{jk}$$

و احتمال یک مشاهده تصادفی که در فاصله $[2^{-j}k, 2^{-j}(k+1)]$ است به صورت زیر است:

$$p(X \in [2^{-j}k, 2^{-j}(k+1)]) = \int_{2^{-j}k}^{2^{-j}(k+1)} f(x) dx = 2^{-j/2} \alpha_{jk}$$

و بنابراین داریم:

$$N_{jk} \sim b(n, 2^{-j/2} \alpha_{jk})$$

حال توزیع شرطی از $\hat{\beta}_{jk}$ که توسط $\hat{\alpha}_{jk}$ ارائه شده است را در نظر می گیریم اگر M_{jk} تعدادی نقاط داده ای که در فاصله $[2^{-j}k, 2^{-j}(k+1)]$ است، داریم:

$$(M_{jk} | N_{jk} = n_{jk}) \sim b\left(n_{jk}, \frac{1}{2} \left(1 + \frac{\beta_{jk}}{\alpha_{jk}}\right)\right)$$

از رابطه (16) داریم:

$$n 2^{-j/2} \hat{\beta}_{jk} = M_{jk} - (n_{jk} - M_{jk})$$

در اینجا توزیع شرطی $\hat{\beta}_{jk}$ بروی $\hat{\alpha}_{jk}$ می تواند به اصطلاح یک توزیع دوجمله ای باشد.

معروف است. توجه داشته باشید که اگر S مجهول باشد، برآورد آن یعنی $\hat{\theta}$ جایگزین می شود.

ساختار کوواریانس ضرایب موجکی تجربی

وقتی که مشاهدات X_i هم توزیع با f فرض شوند به آسانی نشان داده می شود که:

$$\text{cov}[\hat{\beta}_{j_1 k_1}, \hat{\beta}_{j_2 k_2}] = \frac{1}{n} \left(\int \psi_{j_1 k_1}(x) \psi_{j_2 k_2}(x) f(x) dx - \beta_{j_1 k_1} \beta_{j_2 k_2} \right)$$

همچنین

$$\text{cov}[\hat{\alpha}_{j_1 k_1}, \hat{\alpha}_{j_2 k_2}] = \frac{1}{n} \left(\int \phi_{j_1 k_1}(x) \phi_{j_2 k_2}(x) f(x) dx - \alpha_{j_1 k_1} \alpha_{j_2 k_2} \right)$$

مشاهده می شود که حتی ضرایب به طور جداگانه در بسط انتگرال بالا صفر خواهند بود اگر مجانباً ناهمبسته باشند و از رابطه بالا نتیجه می گیریم که واریانس ضرایب تجربی حتی درون سطوح ثابت نیستند و به صورت زیر هستند:

$$\text{var}[\hat{\beta}_{jk}] = \frac{1}{n} \left(\int \psi_{jk}^2(x) f(x) dx - \beta_{jk}^2 \right) \quad (16)$$

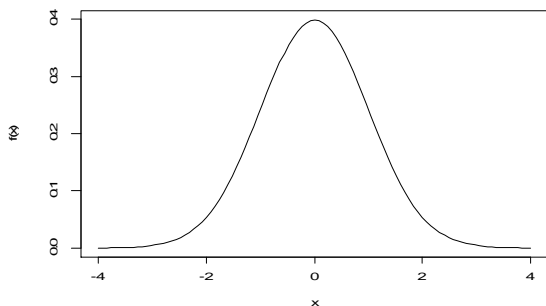
ضرایب پایین تجزیه پذیر تعدادی ترکیبات خطی نسبتاً زیادی از متغیرهای تصادفی مستقل هستند که این مورد با استفاده از قضیه حد مرکزی و فرض نرمال بودن ممکن است قابل قبول باشد؛ ولی به هر حال این ضرایب هنوز هم مستقل نیستند و فرض غیرنرمال بودن منطقیاً درباره ضرایب تجزیه پذیر بالا قابل قبول نیست؛ زیرا این ترکیبات خطی فقط چند متغیر تصادفی است.

با استفاده از تعریف موجک مادر هار و توابع مقیاس به آسانی مشاهده می شود که ضرایب واقعی چگالی به صورت زیر است:

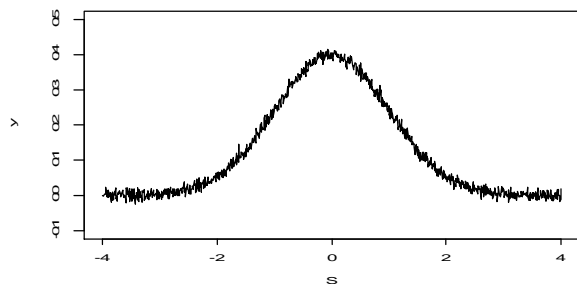
$$\beta_{jk} = 2^{j/2} \left\{ \int_{2^{-j}k}^{2^{-j}(k+\frac{1}{2})} f(x) dx - \int_{2^{-j}(k-\frac{1}{2})}^{2^{-j}(k+\frac{1}{2})} f(x) dx \right\}$$

$$\alpha_{jk} = 2^{j/2} \int_{2^{-j}k}^{2^{-j}(k+1)} f(x) dx$$

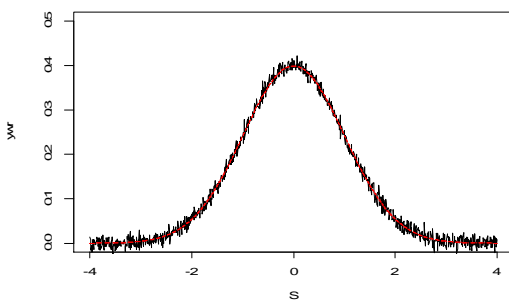
برآوردگر موجکی آستانه‌ای تابع چگالی رسم شده است. در این مثال فاصله x های بین $(-4 و 4)$ در نظر گرفته شده است.



شکل 1. تابع چگالی توزیع نرمال استاندارد



شکل 2. برآورد چگالی توزیع نرمال استاندارد به روش موجکی



شکل 3. برآورد تابع چگالی موجکی آستانه‌ای توزیع نرمال

مثال 2: فرض کنید $f(x) = \frac{1}{e^{-x^2} \sin(x^2)}$ برآورد موجکی این تابع به علت شکل خاص آن و تغییر

ملاحظه 3: آستانه‌ای کردن ضرایب موجکی تجربی یک تلاش برای پیدا کردن ضرایب واقعی غیر صفر است و این را می‌توان به‌عنوان انجام آزمون فرض به‌صورت زیر در نظر داشت:

$$H_0: \beta_{jk} = 0 \quad v \quad H_1: \beta_{jk} \neq 0$$

که این آزمون معادل این است که آیا احتمال دوجمله‌ای در رابطه فوق مساوی $\frac{1}{2}$ است. باید توجه کرد که مفهوم اعداد بزرگ‌تر آزمون در این روش، با توجه به اینکه $\hat{\beta}_{jk}$ مستقل نیست، بیانگر عدم استقلال است. مقادیر p -value به ازای هر (j,k) در این مورد شرطی متناسب با $\hat{\beta}_{jk}$ هستند. حال با ثابت نگه داشتن یک p -value استاندارد برای تمام آزمون‌ها از این مقدار احتمال استفاده می‌شود.

نتایج عددی

در این بخش دو مثال برای برآورد تابع چگالی و ضرایب موجکی ارائه می‌شود. تمامی کدهای مورد استفاده برای رسم نمودارهای این فصل با استفاده از نرم‌افزار R نوشته شده است. موجک استفاده شده در شبیه‌سازی از نوع دوپیچز است.

معادله $y_i = f(x_i) + \epsilon_i$ را در نظر بگیرید. فرض کنید $f(x)$ یک تابع چگالی دلخواه باشد؛ هدف برآورد تابع چگالی $f(x_i)$ است. با استفاده از روش موجک و موجک آستانه‌ای تحت آستانه جهانی تابع چگالی مورد نظر برآورد شده است. در انتها برآوردها به همراه تابع چگالی $f(x)$ روی نمودار نشان داده شده است؛ بنابراین می‌توان اختلاف بین هر برآورد و تابع اصلی را مشاهده کرد.

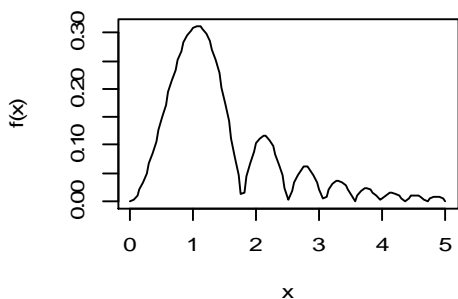
مثال 1: در این مثال هدف برآورد تابع چگالی توزیع نرمال استاندارد با استفاده از روش موجک و موجک آستانه‌ای است.

در شکل (1) نمودار تابع چگالی $N(0,1)$ ، در شکل (2) برآوردگر موجکی تابع $f(x)$ و در شکل (3)

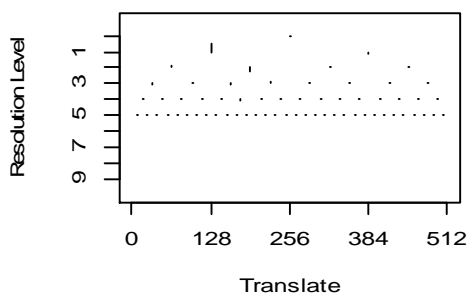
بحث و نتیجه گیری

موجکها در تحلیل دادهها، سیگنالها و توابع، وقتی که تبدیل فوریه از تحلیل زمان - فرکانسی دادهها ناتوان است،

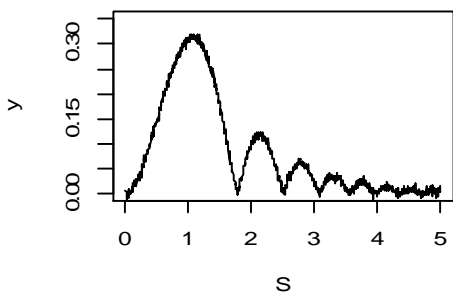
رفتارهای آنی و زیاد، اهمیت دارد. مشابه با مثالهای قبل نمودارهای زیر را داریم. ترتیب شکلها در نمودار زیر مشابه با مثالهای قبل است.



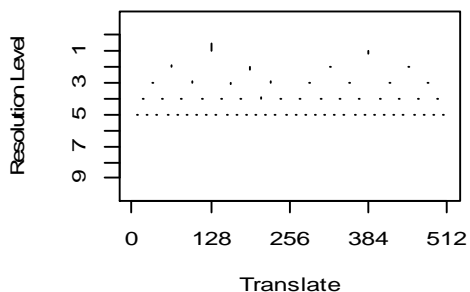
Wavelet Decomposition Coefficient



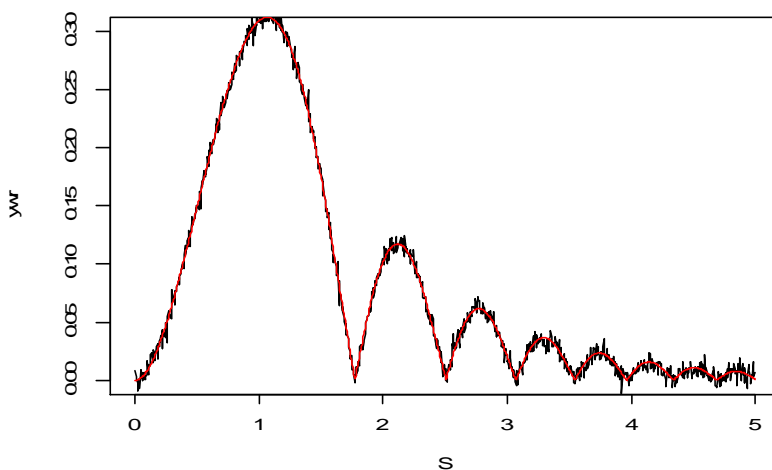
Standard transform Daub cmpct on ext. phase



Wavelet Decomposition Coefficient



Standard transform Daub cmpct on ext. phase



شکل 4. برآور گر موجکی تابع $f(x)$

روش‌ها برآورد می‌کند. به‌منظور ارزیابی توانایی این برآوردگرها، در بخش شبیه‌سازی از دو تابع به‌عنوان توابع هدف استفاده شد. نمودارهای رسم شده نشان می‌دهد که روش موجکی آستانه‌ای همواره نسبت به روش موجکی تقریب بهتری ارائه می‌دهد.

بسیار مناسب عمل می‌کند و ناپیوستگی توابع و تغییر فرکانس آنها را به بهترین شکل ممکن آشکار می‌سازد. در این مقاله برآوردگر موجکی و موجکی آستانه‌ای تابع چگالی بررسی شد. روش موجکی، بنا به خاصیت پایه‌های موجکی و مقیاسی، نقاط تیز و جهشی را بهتر از سایر

References

- [1] Afshari, M. (2013). A fast wavelet algorithm for analyzing one-dimensional signal processing and asymptotic distribution of wavelet coefficients with numerical example and simulation. *Communications in Statistics-Theory and Methods*, 42(22), 4156-4169.
- [2] Afshari, M. (2014). Wavelet density estimation of censoring data and evaluate of mean integral square error with convergence ratio and empirical distribution of given estimator. *Applied Mathematics*, 5(13), 20-62.
- [3] Afshari, M., Lak, F., & Gholizadeh, B. (2016). A new Bayesian wavelet thresholding estimator of nonparametric regression. *Journal of Applied Statistics*, 1-18.
- [4] Afshari, M. (2017). Nonlinear wavelet Shrinkage Estimator of Nonparametric Regression Function Via Cross Validation. *International Journal of Wavelets, Multiresolution and Information*, 15(6), 1-16.
- [5] Antoniadis, A., Gregoire, G. and McKeague, I. (1994). Wavelet methods for curve estimation. *Journal of the Americal Statistical Association* 89: 1340-1353.
- [6] Donoho, D.L., Johnstone, I.M., Krekyacharian, G., & Picard, D. (1996). Density Estimation by Wavelet Thresholding. *The Annals of Statistics*, 24(2), 508-539.
- [7] Daubechies, I. (1986). Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7), 909-996.
- [8] Doosti, H., H. A. Niroumand, and M. Afshari. (2006). Wavelet Based Estimation of the Derivatives of a Density for a Discrete-Time Stochastic Process: Lp-losses. *Journal of Sciences, Islamic Republic of Iran* 17(1), 75-81.
- [9] Doosti, H., Afshari, M., & Niroumand, H. A. (2008). Wavelets for nonparametric stochastic regression with mixing stochastic process. *Communications in Statistics — Theory and Methods*, 37(3), 373-385.
- [10] Grossmann, A., Morlet, J. (1984). Decomposition of Hardly functions into square integrable wavelets of constant shape. *SIAM J. Math.* 15:723-736.
- [11] Haar, A. (1910). Zur thorie der orthogonal Functioned - system, *Mathematics annalysis*, 69: 331-371.
- [12] Herrick, D.R.M. (2000). Wavelet Methods for Curve and Surface Estimation. Ph.D. thesis, University of Bristol.
- Grossmann, A., Morlet, J. (1984). Decomposition of Hardly functions into square integrable wavelets of constant shape. *SIAM J. Math.* 15:723-736.
- [13] Kovac, A. and Silverman, B. W. (2000). Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *J. Am. Statist.*
- [14] Mallat, S. (1989). Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.* 315, 69-87.
- [15] Nason, G. (2011). Wavelet Methods in Statistic with R.
- [16] Vidakovic, B. (1998). Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *Journal of the American Statistical Association*, 93(441), 173-179.